The Genetic Basis of Lipopolysaccharide Biosynthesis in

_Campylobacter jejuni._

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

by

Neil J. Oldfield B.Sc.(Hons).
Department of Genetics, University of Leicester

2000
The Genetic Basis of Lipopolysaccharide Biosynthesis in *Campylobacter jejuni*

By

Neil J. Oldfield

Lipopolysaccharide (LPS) and lipooligosaccharide (LOS) molecules are important in the pathogenesis of many Gram negative bacteria. In the enteric pathogen *Campylobacter jejuni*, LPS and LOS molecules are endotoxic, have been suggested as adhesins and are implicated in the development of the auto-immune disorder Guillain Barré syndrome. However, the genetic basis of biosynthesis and structural variation of these molecules in *C. jejuni* is poorly understood. This was in contrast to other Gram negative bacteria in which the genes and pathways involved had been extensively studied. The aim of this study was to therefore investigate the genetic basis of LPS/LOS biosynthesis, structural variation and function in *C. jejuni*. A bioinformatic approach was used to identify and characterize LPS/LOS biosynthesis genes in the genome sequence of *C. jejuni* NCTC 11168 to enable a model of core oligosaccharide biosynthesis to be proposed. Structural variation occurs between LOS/LPS molecules produced by different strains of *C. jejuni*. The genetic basis of this inter-strain variation was investigated and both conserved and polymorphic genes were identified. Cloning and DNA sequence analysis of some polymorphic genes enabled tentative functions to be proposed. The functions of two genes, one conserved (*waaF*) and one polymorphic (*wlcuJ*) were further investigated. Studies on *wlcuJ* involving the construction of chimeric strains into which the normally absent *wlaJ* gene was inserted, did not reveal the precise role of this gene in LPS/LOS biosynthesis. However, the function of *waaF* as encoding a heptosyltransferase involved with inner core biosynthesis was confirmed using complementation and mutational analysis. Significantly, analysis of *waaF* in *C. jejuni* NCTC 11828 confirmed that the O-chain molecule produced by this strain is not linked to the core oligosaccharide, and is therefore an independent lipid bound capsular polysaccharide.
Acknowledgements

First of all, thanks to all members of Lab 145, both past and present for all their help and advice over the last few years. In particular, thanks to Anne, Clióna, Arnie, Mike, Rocky and Lorna, and of course to my supervisor Julian who was always ready with good advice and another student to supervise.

Thanks also to R. Thwaites, N. Gregson, D. Wareing, A. Swann and A. Lastovica for strains, and to Ben Fry and Anthony Moran for useful discussions. This work was funded by the Biotechnology and Biological Sciences Research Council.

Finally, many thanks to my parents, all four of them, for their financial and emotional support over the course of this Ph.D. I could not have done it without them.
Contents

List of figures viii
List of tables xiii
Abbreviations xv

Chapter 1: General Introduction

1.1. Background 1

1.2. The genus Campylobacter 1

1.3. Transmission of Campylobacter to humans 2

1.4. Campylobacter enteritis 3
   1.4.1. The epidemiology of Campylobacter enteritis 3
   1.4.2. Clinical features of Campylobacter enteritis 4
   1.4.3. The pathogenesis of Campylobacter enteritis 5
      1.4.3.1. Animal models of Campylobacter enteritis 5
      1.4.3.2. General characteristics of Campylobacter pathogenesis 6
      1.4.3.3. Colonization of the mucosal surface of the intestine 6
      1.4.3.4. Adherence 7
      1.4.3.5. Mechanisms for the perturbation of the absorptive capacity of the intestine 8

1.5. Guillain Barre Syndrome (GBS) 12
   1.5.1. Clinical features of Guillain Barre Syndrome 12
   1.5.2. The epidemiology of Guillain Barre Syndrome 13
   1.5.3. The link between C. jejuni and GBS 13
   1.5.4. Campylobacter serotypes associated with GBS 14
   1.5.5. Anti-ganglioside antibodies and GBS 15
   1.5.6. Molecular mimicry and GBS 17

1.6. Lipopolysaccharide (LPS) - Overview 18

1.7. Escherichia coli and Salmonella LPS 20
   1.7.1. The Kdo2-lipid A moiety: structure and biosynthesis 20
   1.7.2. The Kdo2-lipid A moiety: genetics 23
1.7.3. The Kdo₂-lipid A moiety: function 23
1.7.4. Core oligosaccharide: structure and biosynthesis 24
1.7.5. Core oligosaccharide: genetics 28
1.7.6. Core oligosaccharide: function 29
1.7.7. O-chain: structure and function 32
1.7.8. O-chain: biosynthesis 33
  1.7.8.1. The Wzy-dependent biosynthesis of O-chain 33
  1.7.8.2. The ABC transporter dependent synthesis of O-chain 34
  1.7.8.3. The synthetase dependent biosynthesis of O-chain 35
  1.7.8.4. The ligation of O-chain to the lipid A-core moiety 35
1.7.9. O-chain: genetics 36
1.8. Haemophilus influenzae LOS 37
  1.8.1. The Kdo-lipid A moiety: structure, function and biosynthesis 39
  1.8.2. The Kdo-lipid A moiety: genetics 40
  1.8.3. The oligosaccharide moiety: structure 41
  1.8.4. The oligosaccharide moiety: phase variation 41
  1.8.5. The oligosaccharide moiety: genetics 44
    1.8.5.1. The opsX/kdkA cluster 45
    1.8.5.2. The licI operon 45
    1.8.5.3. Other H. influenzae loci containing 5’-CAAT-3’ repeats 45
    1.8.5.4. Other oligosaccharide biosynthesis genes 46
1.9. Campylobacter jejuni LOS/LPS 47
  1.9.1. The lipid A moiety 47
  1.9.2. Core oligosaccharide 48
  1.9.3. O-chain 50
  1.9.4. LPS/LOS structures from isolates from patients with neurological diseases 51
1.10. Aims of this study 53

Chapter 2: Materials and Methods

2.1. Bacterial strains and growth conditions 54
2.2. Storage of bacterial strains 54
2.3. Media

2.4. Antibiotic supplements

2.5. Miscellaneous buffers and solutions

2.6. Extraction of DNA from bacterial cells
   2.6.1. Preparation of plasmid DNA from *E. coli* and *S. typhimurium* cells
   2.6.2. Preparation of *Campylobacter* chromosomal DNA

2.7. Quantification of DNA using spectrophotometry

2.8. Electrophoresis of DNA

2.9. Purification of DNA
   2.9.1. Extraction of DNA from agarose gels using the NUCLEOTRAP extraction kit
   2.9.2. Phenol/chloroform extraction and ethanol precipitation

2.10. Enzymatic modification of plasmid or chromosomal DNA
   2.10.1. Restriction endonuclease digestion of DNA
   2.10.2. Modification of a 5' overhang to a blunt end terminus
   2.10.3. Dephosphorylation of restriction digested DNA
   2.10.4. DNA ligation

2.11. Electrotrectransformation of *E. coli* and *S. typhimurium* cells
   2.11.1. Preparation of electrocompetent cells
   2.11.2. Electroporation of competent cells

2.12. Electrotrectransformation of *Campylobacter jejuni* strains
   2.12.1. Preparation of electrocompetent *C. jejuni* cells
   2.12.2. Electroporation of competent *C. jejuni* cells

2.13. Natural transformation of *C. jejuni* cells

2.14. Amplification of DNA using the Polymerase Chain Reaction
   2.14.1. Amplification of DNA using plasmid or chromosomal template
   2.14.2. Amplification of DNA using the Expand™ High Fidelity PCR system
   2.14.3. Amplification of DNA using the Expand™ Long Template PCR system

2.15. DNA sequencing

2.16. DNA and protein sequence analysis

2.17. Analysis of plasmid or chromosomal DNA by Southern hybridization
Chapter 3: Identification of lipopolysaccharide biosynthesis genes in the genome sequence of *C. jejuni* NCTC 11168

3.1. Introduction

3.2. Sequencing of the *C. jejuni* NCTC 11168 genome

3.3. Computer analysis of the *C. jejuni* genome sequence

3.4. Kdo-lipid A biosynthesis genes
   3.4.1. Summary of Kdo-lipid biosynthesis in *C. jejuni*

3.5. Core oligosaccharide biosynthesis genes
   3.5.1. The *gmhA, D and E* genes (ADP-L-glycero-D-manno-heptose biosynthesis)
   3.5.2. Transfer of ADP-L-glycero-D-manno-heptose residues
   3.5.3. Biosynthesis and transfer of N-acetylneuraminic acid (sialic acid)
   3.5.4. Inter-conversion between UDP-glucose and UDP-galactose
   3.5.5. Potential glycosyltransferases
   3.5.6. Open reading frames in the cluster which encode proteins of unknown function
3.5.7. The role of LOS biosynthesis cluster genes in protein glycosylation 119
3.6. O-chain biosynthesis and other saccharide biosynthesis genes 120
3.7. Homo-polymeric G tracts and phase variation 121
3.8. Discussion 122

Chapter 4: Identification and characterization of gene content polymorphisms in C. jejuni LOS clusters

4.1. Introduction 128
4.2. Gene content polymorphisms between wlaM and waaM 129
4.3. Identification and characterization of gene content polymorphisms between waaM and wlaSA 137
   4.3.1. The waaC - waaF region of C. jejuni NCTC 11828 141
4.4. Identification and characterization of gene content polymorphisms between wlaSA and wlaT 155
   4.4.1. Isolation and characterization of the C. jejuni NCTC 11828 waaF - gmhA insert 159
   4.4.2. Isolation and characterization of the C. jejuni 2523/90 waaF - gmhA insert 166
   4.4.3. Isolation and characterization of the C. jejuni 8F 169 and NCTC 11351 waaF - gmhA inserts 166
4.5. Discussion 167
   4.5.1. Identification of areas of gene content polymorphism in the LOS cluster 167
   4.5.2. Further characterization of the gene content polymorphisms 172

Chapter 5: Functional analysis of the polymorphic gene wlaJ

5.1. Introduction 176
5.2. Construction of plasmids to enable the relocation of wlaJ to C. jejuni NCTC 11828 177
5.3. Transformation of C. jejuni NCTC 11828 with pCOD2 and pCOD3 182
5.4. Phenotypic analysis of 11828COD2 and 11828COD3 LPS 188
5.5. Mutation of wlaJ in the chimeric strains 11828COD2 and 11828COD3 188
Chapter 6: Characterization of the waaF genes of C. jejuni NCTC 11168 and NCTC 11828

6.1. Introduction 230

6.2. Isolation and cloning of Cj1148 from C. jejuni NCTC 11168 231

6.3. Complementation of Salmonella typhimurium SL3789 (waaF511) with pNOL17
   6.3.1. Phenotypic analysis of SL3789 (pNOL17) LPS 234

6.4. Mutation of waaF in C. jejuni NCTC 11168 236
   6.4.1. Inverse PCR mutagenesis of C. jejuni NCTC 11168 waaF 236
   6.4.2. Transformation of C. jejuni NCTC 11168 with pNOL22 and pNOL23 240
   6.4.3. Phenotypic analysis of 11168NOL22 and 11168NOL23 LOS 240

6.5. Mutation of waaF in C. jejuni NCTC 11828 248
   6.5.1. Isolation of waaF from C. jejuni NCTC 11828 248
   6.5.2. Inverse PCR mutagenesis of C. jejuni NCTC 11828 waaF 252
   6.5.3. Transformation of C. jejuni NCTC 11828 with pNOL11 and pNOL12 252
   6.5.4. Phenotypic analysis of 11828NOL11 and 11828NOL12 LPS 256

6.6. Discussion 263
   6.6.1. Confirmation of Cj1148 as being the waaF gene of C. jejuni NCTC 11168 263
   6.6.2. Evidence that C. jejuni NCTC 11828 O-chain is not covalently attached to the core oligosaccharide 264
Chapter 7: General discussion

7.1. Introduction 268
7.2. Building a model of core oligosaccharide biosynthesis in C. jejuni NCTC 269
7.3. O-chain versus capsule 272
7.4. Distribution studies on LOS cluster genes 273

Appendix 1 278
Appendix 2 281
Appendix 3 283
Appendix 4 293
Appendix 5 297
Appendix 6 300
Appendix 7 302
Bibliography 305
List of Figures

Chapter 1: General Introduction

1.1: The chemical structures of common human gangliosides 16
1.2: Schematic representation of the generalized structure of LPS and LOS 19
1.3: Diagrammatic representation showing the biosynthesis of Kdo₂-lipid A in *E. coli* and *Salmonella* serovars 21
1.4: Generalized structure of the sugar backbone of the *E. coli* and *Salmonella* core oligosaccharide 25
1.5: Chemical structures and biosynthetic enzymes of the outer core oligosaccharides of *E. coli* and *Salmonella* 26
1.6: Organization and gene content of *E. coli* and *Salmonella* core oligosaccharide biosynthetic clusters 30
1.7: Schematic diagram showing the organization and gene content of the *E. coli* O:113 O-chain biosynthetic cluster and the structure of the O-chain repeat unit 38
1.8: Chemical structures of the major oligosaccharides produced by *H. influenzae* strains Eagan and Rd 42
1.9: Schematic diagram of core oligosaccharide structures produced by different HS serostrains of *C. jejuni* 49
1.10: Schematic diagram of core oligosaccharide structures produced by the *C. jejuni* HS:19 serostrain and two HS:19 isolates from patients with GBS 52

Chapter 3: Identification of lipopolysaccharide biosynthesis genes in the genome sequence of *C. jejuni* NCTC 11168

3.1: Alignment of the amino acid sequences of LpxA proteins from various Gram negative bacteria 85
3.2: Alignment of the carboxy-terminal amino acid sequences of various MsbA proteins 90
3.3: Schematic diagram of the LOS biosynthesis cluster of *C. jejuni* NCTC 11168

3.4: Alignment of the N-terminal amino acid sequences of various GmhD proteins

3.5: Alignment of the amino acid sequences of the Cj1139-encoded protein and *C. jejuni* OH 4384 CgtB

3.6: A model for LOS biosynthesis in *C. jejuni* NCTC 11168

### Chapter 4: Identification and characterization of gene content polymorphisms in *C. jejuni* LOS clusters

4.1: PCR analysis of strain NCTC 11168 to confirm the presence and orientation of the *wlaL* - *wlaF* genes

4.2: Schematic diagram showing the gene content and primer positions of the *wlaL* - *wlaF* region of strain NCTC 11168

4.3: PCR analysis of strain NCTC 11351 to confirm the presence and orientation of the *wlaL* - *wlaF* genes

4.4: Schematic diagram showing the gene content and primer positions of the *wlaL* - *wlaF* region of strain NCTC 11351

4.5: Schematic diagram showing the conservation of genes in the *wlaM* - *waaM* region of the LOS cluster in 38 *C. jejuni* strains

4.6: Schematic diagram showing the conservation of genes in the *waaM* - *wlaSA* region of the LOS cluster in 38 *C. jejuni* strains

4.7: Comparison of the long range PCR products obtained from *C. jejuni* NCTC 11168 and *C. jejuni* NCTC 11828

4.8: Schematic diagram of the *waaC* - *waaF* regions of *C. jejuni* NCTC 11168 and NCTC 11828

4.9: Alignment of the amino acid sequences from the *C. jejuni* NCTC 11168 WlaNA and *C. jejuni* NCTC 11828 WlaUA proteins

4.10: Alignment of the amino acid sequences from the *C. jejuni* NCTC 11828 WlaUD and WlaUE proteins

4.11: PCR analysis to illustrate the approximate size of gene insertions between *waaF* and *gmhA*
4.12: Schematic diagram summarizing the gene content in the wlaSA - wlaT regions of five C. jejuni strains

4.13: Diagrammatic representation showing pNOL5 and pNOL6

4.14: Diagrammatic representation showing pNOL7 and pNOL21

4.15: Alignment of the amino acid sequences from the C. jejuni 8F 169 and NCTC 11351 WlaZ proteins

Chapter 5:

Functional analysis of the polymorphic gene wlaJ

5.1: Diagrammatic representation showing the construction of pCOD1, pCOD2 and pCOD3

5.2: Agarose gel electrophoresis to show the restriction enzyme digestion of pCOD1, pCOD2 and pCOD3

5.3: Schematic diagram showing the gene content (and primer positions) for the wlaG - wlaK region of strains 11828COD2 and 11828COD3

5.4: PCR analysis of strains 11828COD2 and 11828COD3

5.5: Southern hybridization analysis of strains 11828COD2 and 11828COD3

5.6: LPS profiles of the chimeric C. jejuni strains, 11828COD2 and 11828COD3

5.7: Schematic diagram of plasmids pAWL611 and pAWL612

5.8: Schematic diagram showing the gene content (and primer positions) for the wlaG - wlaK region of strains COD2KF, COD2KR, COD3KF and COD3KR

5.9: PCR analysis of strains COD2KF and COD2KR

5.10: PCR analysis of strains COD3KF and COD3KR

5.11: Southern hybridization data from strains COD2KF, COD2KR, COD3KF and COD3KR

5.12: LPS profiles of the four C. jejuni strains COD2KF, COD2KR, COD3KF and COD3KR

5.13: Diagrammatic representation showing the construction of pCOD4 and pCOD5
5.14: Diagrammatic representation showing the annealing sites of WLAJR3 and WLAJF3 - primers used to delete the *wlaJ* gene from pCOD2 and pCOD3

5.15: Agarose gel electrophoresis (1% gel) to show the restriction enzyme digestion of pCOD4 and pCOD5

5.16: Schematic diagram showing the gene content (and primer positions) for the *wlaG* - *wlaK* region of strains 11828COD4 and 11828COD5

5.17: PCR analysis of strains 11828COD4 and 11828COD5

5.18: Southern hybridization data from strains 11828COD4 and 11828COD5

5.19: LPS profiles of the *C. jejuni* strains. 11828COD4 and 11828COD5

5.20: Schematic diagram showing the gene content (and primer positions) for the *wlah* - *wlaL* region of strains 11828AWL611 and 11828AWL612

5.21: PCR analysis of strains 11828AWL611 and 11828AWL612

5.22: Southern hybridization data from strains 11828AWL611 and 11828AWL612

5.23: LPS profiles of the *C. jejuni* strains. 11828AWL611 and 11828AWL612

---

**Chapter 6: Characterization of the *waaF* genes of *C. jejuni* NCTC 11168 and NCTC 11828**

6.1: Diagrammatic representation showing pNOL17

6.2: Agarose gel electrophoresis (0.8% gel) to show the restriction enzyme digestion of pNOL17

6.3: Silver staining of the LPS produced by the complemented *S. typhimurium* strain SL3789 (pNOL17)

6.4: Diagrammatic representation showing the annealing sites of WAAFF2 and WAAFR4 - primers used for the inverse PCR mutagenesis of the *waaF* gene of *C. jejuni* NCTC 11168

6.5: Diagrammatic representation showing pNOL22 and pNOL23

6.6: Agarose gel electrophoresis (0.8% gel) to show the restriction enzyme digestion of pNOL22 and pNOL23

6.7: PCR analysis of strains 11168NOL22 and 11168NOL23
6.8: Southern hybridization data from strains 11168NOL22 and 11168NOL23

6.9: Silver staining of the LOS produced by the \textit{waaF} mutant strains 11168NOL22 and 11168NOL23

6.10: Analysis of 11168NOL22 and 11168NOL23 LOS by immunoblotting

6.11: Diagrammatic representation showing pNOL8

6.12: Agarose gel electrophoresis (0.8% gel) to show the restriction enzyme digestion of pNOL8

6.13: Diagrammatic representation showing pNOL11 and pNOL12

6.14: Agarose gel electrophoresis (0.8% gel) to show the restriction enzyme digestion of pNOL11 and pNOL12

6.15: PCR analysis of strains 11828NOL11 and 11828NOL12

6.16: Southern hybridization data from strains 11828NOL11 and 11828NOL12

6.17: Silver staining of the LPS produced by the \textit{waaF} mutant strains 11828NOL11 and 11828NOL12

6.18: Analysis of 11828NOL11 and 11828NOL12 LPS by immunoblotting
Chapter 2: Materials and Methods

2.1: Sources and heat-stable (HS) serotypes of the C. jejuni strains used in this study
2.2: Antibiotic supplements
2.3: Tricine SDS-PAGE gel constituents

Chapter 3: Identification of lipopolysaccharide biosynthesis genes in the genome sequence of C. jejuni NCTC 11168

3.1: Details of the Kdo-lipid A biosynthesis genes identified in the genome of C. jejuni NCTC 11168
3.2: Details of the C. jejuni NCTC 11168 genes involved with the biosynthesis and transfer of ADP-L-glycero-D-manno-heptose
3.3: Details of the C. jejuni NCTC 11168 genes involved with the biosynthesis and transfer of N-acetylneuraminic acid to the core oligosaccharide
3.4: Details of the C. jejuni NCTC 11168 gene involved with the interconversion of UDP-glucose and UDP-galactose
3.5: Details of the C. jejuni NCTC 11168 genes potentially encoding enzymes involved with the transfer of glycosyl residues
3.6: Details of the C. jejuni NCTC 11168 genes found in the LOS biosynthesis cluster which encode proteins of unknown function

Chapter 4: Identification and characterization of gene content polymorphisms in C. jejuni LOS clusters

4.1: The distribution of the wlaJ gene in 38 strains of C. jejuni
4.2: The distribution of the wlaNA - wlaSB genes in 38 strains of C. jejuni
4.3: Details of the novel genes identified in the waaC - waaF region of C. jejuni NCTC 11828

xiii
4.4: The distribution and approximate size of inserted sequences between \textit{waaF} and \textit{gmhA} in 38 strains of \textit{C. jejuni}

4.5: Details of the genes identified in the \textit{waaF} - \textit{gmhA} insertions from four strains of \textit{C. jejuni}
# Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Ampère</td>
</tr>
<tr>
<td>AP</td>
<td>Alkaline phosphatase</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>bp</td>
<td>Base pairs</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>c</td>
<td>Centi ($10^2$)</td>
</tr>
<tr>
<td>°C</td>
<td>Degrees Celsius</td>
</tr>
<tr>
<td>CIP</td>
<td>Calf intestinal phosphatase</td>
</tr>
<tr>
<td>CSA</td>
<td>Campylobacter blood free selective agar base</td>
</tr>
<tr>
<td>CTAB</td>
<td>Hexadecyltrimethyl ammonium bromide</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>Deoxyribonucleotide triphosphate</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>Disodium ethylenediaminetetraacetate</td>
</tr>
<tr>
<td>FD</td>
<td>Faraday</td>
</tr>
<tr>
<td>g</td>
<td>Angular velocity as multiples of the acceleration due to gravity ($g = 9.8 \text{ m s}^{-2}$)</td>
</tr>
<tr>
<td>g</td>
<td>Grams</td>
</tr>
<tr>
<td>hr</td>
<td>Hour (s)</td>
</tr>
<tr>
<td>H₂O</td>
<td>Water</td>
</tr>
<tr>
<td>HS</td>
<td>Heat stable</td>
</tr>
<tr>
<td>J</td>
<td>Joules</td>
</tr>
<tr>
<td>k</td>
<td>Kilo ($10^3$)</td>
</tr>
<tr>
<td>kb</td>
<td>Kilobase pairs</td>
</tr>
<tr>
<td>l</td>
<td>Litre (s)</td>
</tr>
<tr>
<td>LA</td>
<td>Luria-Bertani agar</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani broth</td>
</tr>
<tr>
<td>LOS</td>
<td>Lipooligosaccharide</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>LPS</td>
<td>Lipopolysaccharide</td>
</tr>
<tr>
<td>μ</td>
<td>Micro ($10^{-6}$)</td>
</tr>
<tr>
<td>m</td>
<td>Metre</td>
</tr>
<tr>
<td>m</td>
<td>Milli ($10^{-3}$)</td>
</tr>
<tr>
<td>M</td>
<td>Mole (s)</td>
</tr>
<tr>
<td>MHB</td>
<td>Mueller-Hinton broth</td>
</tr>
<tr>
<td>min</td>
<td>Minute (s)</td>
</tr>
<tr>
<td>n</td>
<td>Nano ($10^{-9}$)</td>
</tr>
<tr>
<td>Ω</td>
<td>Ohm (s)</td>
</tr>
<tr>
<td>OD&lt;sub&gt;600&lt;/sub&gt;</td>
<td>Optical Density at a wave length of 600 nm</td>
</tr>
<tr>
<td>p</td>
<td>Pico ($10^{-12}$)</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>s</td>
<td>Second (s)</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium lauryl sulphate</td>
</tr>
<tr>
<td>tRNA</td>
<td>Transfer RNA</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>v/v</td>
<td>Volume per volume</td>
</tr>
<tr>
<td>V</td>
<td>Volts</td>
</tr>
<tr>
<td>VAIN</td>
<td>Variable atmosphere incubator</td>
</tr>
<tr>
<td>w/v</td>
<td>Weight per volume</td>
</tr>
<tr>
<td>X-Gal</td>
<td>Chloro-3-indoyl-β-D-galactoside</td>
</tr>
</tbody>
</table>
Chapter 1: General Introduction

1.1. Background

Campylobacter species, in particular *C. jejuni*, are the leading cause of foodborne bacterial diarrhoeal disease worldwide (Blaser, 1997). In addition, infection with *Campylobacter* is the most common antecedent to the neuromuscular polyneuropathy Guillain Barré Syndrome, which although rare is a common cause of generalized paralysis (Moran & O'Malley, 1995). Despite the large economic and social burden (ACMSF, 1993; Tauxe, 1992), relatively little is known about the pathogenesis and virulence mechanisms of these pathogens - a matter of government and public concern. The need for effective control of *Campylobacter* in the food chain and other disease prevention strategies has led to an increase in scientific research on these organisms which will undoubtedly be further stimulated by the recent release of the complete genome sequence of *C. jejuni* strain NCTC 11168 (Parkhill et al., 2000a).

1.2. The genus *Campylobacter*

Described as ‘vibrios’ or ‘spirilla’ (McFadyean & Stockman, 1913), campylobacters were initially assigned to the genus *Vibrio* (Smith & Taylor, 1919). Although showing some physiological and morphological similarity to *Vibrio* species, they were placed in a distinct genus due to fundamental differences such as differing nucleotide base composition (G+C ratio) and their different growth requirements (Sebald & Veron, 1963). Initially campylobacters were almost exclusively considered to be veterinary pathogens. However, the development of selective culture techniques in the early 1970’s that allowed the isolation of campylobacters from faecal samples showed their true importance as being responsible for significant cases of human gastroenteritis (Butzler et al., 1973; Dekeyser et al., 1972; Skirrow, 1977).
Campylobacters are Gram negative curved or spiral rods, 1.5 - 6.0 \( \mu \)m long and 0.2 - 0.5 \( \mu \)m wide (Ketley, 1997), although in older cultures the cells may appear coccoid (Karmali & Skirrow, 1984). They produce a characteristic unsheathed flagellum at one or occasionally both ends of the cell (Simbert, 1978), which aided by the 'corkscrew' morphology of the cell gives a high degree of motility to the bacterium. This allows the cell to move within the viscose environment of the intestinal mucus (Szymanski et al., 1995). Most *Campylobacter* species are microaerophilic, requiring an \( \text{O}_2 \) concentration of between 3 and 15\%, and a \( \text{CO}_2 \) concentration of between 3 and 5\% (Bolton & Coates, 1983), but some are aerotolerant (Griffiths & Park, 1990). Campylobacters grow optimally at a variety of temperatures from 15°C for *C. cryoaerophilia* to 42°C for the thermophilic species *C. jejuni* and *C. coli*.

The main human pathogen of the genus is *C. jejuni* which is thought to be responsible for the majority of enteric infections (80 - 90\%), but *C. coli*, and to a lesser extent, *C. upsaliensis*, *C. hypointestinalis* and *C. lari* can also cause human intestinal disease - although the true incidence of disease caused by these species is still not clear (ACMSF, 1993). *C. fetus* rarely infects man, and usually only occurs in immunocompromised patients, but causes reproductive tract infection and abortion in animals (Mishu et al., 1992).

1.3. Transmission of *Campylobacter* to humans

Campylobacters are present in many wild and domestic animals such as sheep, cows, avians and swine, where they form part of the natural intestinal flora (Penner, 1988). They may also be carried by pets such as cats and dogs, and because of this large animal reservoir, contamination of soil and water often occurs (Blaser, 1997). Infected animals are generally asymptomatic, but transmission of the bacteria to animal meat during the slaughtering process can lead to infection of man if the meat is undercooked or cross-contamination occurs. Transmission can also occur due to drinking untreated water or the handling of infected animals (Ketley, 1997). Human-to-human transmission rarely occurs (Blaser, 1997).
1.4. *Campylobacter* enteritis

1.4.1. The epidemiology of *Campylobacter* enteritis

The epidemiology of *Campylobacter* enteritis varies greatly across the globe. In the developed world most cases appear to be sporadic, with a highest incidence in children under 4 years old and young adults aged 15 to 44 (Butzler & Skirrow, 1979; Friedman et al., 2000; Tauxe, 1992). Below the age of 45 there is a consistent, but as yet unexplained higher incidence of infection in males than in females. There is also a marked seasonality of cases, with the number reported in the summer being more than double the number reported in the winter months (ACMSF, 1993; Friedman et al., 2000; Tauxe, 1992). The numbers of reported cases of *Campylobacter* enteritis have been growing steadily in the developed world. This increase perhaps reflects the changing eating habits of the population, such as the increased consumption of poultry products and 'fast food' - often bought from establishments where hygiene standards are not of a consistently high standard (Griffiths & Park, 1990). Despite this increase in numbers it is still likely that the number of cases reported is a gross underestimation. Many patients do not feel the need to visit a doctor and many doctors do not request stool samples (Blaser, 1997; Ketley, 1995).

In developing countries the epidemiology of *Campylobacter* enteritis is markedly different to that evident in many industrialized nations (Oberhelman & Taylor, 2000; Taylor, 1992). Asymptomatic *Campylobacter* carriage is high, but symptomatic infection seems almost exclusively restricted to children. Whilst inflammatory diarrhea is the main clinical manifestation of *Campylobacter* infection in the developed world, in developing countries it is invariably milder, with a watery, non-inflammatory diarrhoea predominating. The case-to-infection rate and duration of intestinal excretion both decline with the age of the patient in developing countries (Taylor et al., 1993) suggesting the development of immunity - probably caused by the higher exposure and infection rate early in life. The observation that travellers to developing countries who acquire *Campylobacter* infections still show the same symptoms as those seen in
developed countries (Taylor, 1992) supports this argument. Despite some evidence that strain differences can correlate to clinical symptoms (Everest et al., 1992; Ruiz-Palacios et al., 1992; Ruiz-Palacios et al., 1985; Ruiz-Palacios et al., 1983), it appears that the strains present in the developed and developing countries possess common virulence mechanisms suggesting that there is no difference in the relative virulence of strains from the two areas. The development of immunity is therefore likely to be responsible for the differences in disease manifestation evident between *Campylobacter* infections in developed and developing countries.

1.4.2. Clinical features of *Campylobacter* enteritis

The incubation period after ingestion of *Campylobacter* is thought to be 24 - 72 hours (Skirrow & Blaser, 1995). However, difficulties in identifying the source of infection can often make the exact timing difficult to establish. Human volunteer studies have shown that as few as 500 organisms can cause disease in some patients, with the rates of infection increasing with dose (Black et al., 1988; Robinson, 1981). Infection in the developed world usually presents as an inflammatory diarrhoea, with a prodrome of non-specific symptoms including headache, chills and fever (reviewed by Butzler & Skirrow, 1979; Walker et al., 1986). The diarrhoea is frequent, profuse and may at first be watery, but frequently becomes bloody, containing mucus and polymorphonuclear leukocytes. This acute diarrhoea commonly lasts 2 - 3 days, during which time the patient may become dehydrated. Abdominal pain persists during and sometimes after the diarrhoea has stopped and may be so severe as to mimic appendicitis. Overall, symptoms usually resolve over the period of about a week, but can be more prolonged. Complications following *Campylobacter* enteritis are generally rare (Ketley, 1997), however extra-intestinal infections affecting adjacent organs causing cholecystitis (Darling et al., 1979), pancreatitis (Ezpeleta et al., 1992), cystitis and septic abortion have been reported (Blaser, 1997). Bacteraemia, especially in immuno-compromised patients (Blaser et al., 1986), reactive arthritis (Bremell et al., 1991) and Guillain Barré Syndrome (Nachamkin et al., 1998; Rhodes & Tattersfield, 1982) may also occur.
Upon inspection by sigmoidoscopy or rectal biopsy, the intestines of infected patients can show a broad range of effects caused by Campylobacter infection varying from slight changes in the mucosa to colitis. Sites of tissue injury include the jejunum, ileum and colon where degeneration of the mucosa, crypt absences in the epithelial glands and ulceration of the mucosal epithelium occur. In addition, neutrophils, mononuclear cells and eosinophils infiltrate the lamina propria as part of the inflammatory response (Blaser & Reller, 1981).

Whilst in developed countries Campylobacter enteritis usually presents as an inflammatory diarrhoea, in developing countries infection is more likely to result in a non-inflammatory watery diarrhoea. No significant mucosal changes are evident in this milder clinical presentation, and no blood, mucus or leukocytes are present in the stool (Ketley, 1997).

1.4.3. The pathogenesis of Campylobacter enteritis

1.4.3.1. Animal models of Campylobacter enteritis

The virulence factors associated with Campylobacter infection are likely to be complex and multi-factorial in nature, and their study has been hampered by the lack of a suitable animal model to closely mimic the clinical manifestations of Campylobacter enteritis (reviewed in Young et al., 2000). Many animal models have been proposed including congenitally athymic mice in a germ-free condition (Yrios & Balish, 1985), the removable intestinal tie adult rabbit diarrhoea (RITARD) model (Caldwell et al., 1983), the rabbit ileal loop model (Everest et al., 1993a), new-born, colostrum-deprived piglets (Babakhani et al., 1993) and orally challenged (Blaser et al., 1983), intraperitoneally challenged (Abimiku & Dolby, 1987), or intranasally challenged mice (Baqar et al., 1993). In addition, chicken models are available to investigate colonization (Meinersmann et al., 1991; Young et al., 1999), but none of these animals are naturally susceptible, without surgery or pre-conditioning, to the inflammatory-type enteritis seen in man. Only young Rhesus monkeys (Russell et al., 1989) and young
ferrets (Bell & Manning, 1990) are model animals known to be naturally susceptible to experimental infection with *Campylobacter* which produce the clinical manifestations that most closely mimic the self-limiting acute colitis seen in man.

1.4.3.2. General characteristics of *Campylobacter* pathogenesis

*C. jejuni* forms part of the natural intestinal flora of many farm and domestic animals, especially poultry (Penner, 1988). Studies using animal models show that campylobacters can readily colonize the intestine, but that infection usually progresses no further (unless particular interventions such as surgery or pre-conditioning are involved) and asymptomatic carriage or clearance of the organism occurs. However, when campylobacters enter the human intestine the clinical outcome is often different with patients developing an inflammatory or non-inflammatory diarrhoea or, where protective immunity is established asymptomatic carriage (Ketley, 1997). Experimental evidence shows that campylobacters can adhere to and invade epithelial cells, initiate the inflammatory response and/or produce toxins. The resulting damage to cell function and the disruption of the normal fluid absorption of the intestine leads to the clinical presentation of inflammatory diarrhoea. The relative contribution of the three factors; invasion, toxins and the inflammatory response make to the pathological changes that occur during *Campylobacter* infection remains contentious but understanding them is crucial to the development of disease prevention strategies.

1.4.3.3. Colonization of the mucosal surface of the intestine

Following transmission to man, the organism gains access to the intestine via the stomach, having overcome the hydrochloric acid barrier and proteolytic enzymes. Upon entering the intestine, the bacterium must be able to locate, move to and associate with the thick mucus layer that overlays the epithelial cells of the intestine. The subsequent attachment to the epithelial cells of the distal ileum and colon is presumably necessary for invasion, but adherence and invasion are not necessary for colonization *per se* (Lee
An obviously vital requirement for colonization is motility. This has been shown to be absolutely necessary for campylobacters to colonize the intestine of humans and experimental animals (Black et al., 1988; Morooka et al., 1985; Pavlovskis et al., 1991). Motility is imparted by the flagellum - perhaps the best characterized virulence determinant of *Campylobacter* (reviewed in Guerry, 1997). The structural genes encoding the two flagella proteins, FlaA and FlaB, have now been extensively characterized (Guerry et al., 1992; Nuijten et al., 1992), and defined mutants in *flaA*, encoding the major flagellin subunit, have shown the essential role of the flagellum in colonization (Nachamkin et al., 1993; Wassenaar et al., 1993b).

Besides motility, chemotaxis is also required for colonization. *Campylobacter* has been shown to respond to chemoattractants such as mucin, L-serine and L-fucose, and chemorepellants such as bile salts (Flugdahl et al., 1988). Chemically mutagenized, non-chemotactic, but fully motile mutants were unable to colonize the neonatal mouse intestine (Takata et al., 1992). Recent work characterizing the *cheY* gene of *C. jejuni* has shown that when this important regulator of chemotaxis is mutated, campylobacters become non-chemotactic (Marchant, 1999), but show an increase in adherence and invasiveness *in vitro* (Yao et al., 1997). However the *cheY* mutant was unable to colonize or cause disease in the ferret disease model (Yao et al., 1997).

1.4.3.4. Adherence

Studies on adhesion have primarily used epithelial-like cell lines to identify potential adhesins in what is thought to be a complex and multi-factorial process (reviewed in Wooldridge & Ketley, 1997). The surface of the bacterium is where the initial interaction with a host cell occurs, and so surface structures have been the focus of many of these studies. The flagellum was initially investigated as a potential adhesin because *Campylobacter* flagellin preparations were shown to bind intestinal cells (McSweegan & Walker, 1986). However, this pre-treatment did not effect the subsequent binding of bacterial cells (McSweegan & Walker, 1986), and later studies using mutant strains with inactivated flagellin genes have confirmed that the flagellum
is not an adhesin of biological significance (Grant et al., 1993; Moser & Schroder, 1995; Wassenaar et al., 1991; Yao et al., 1994). Other potential adhesins are outer membrane proteins (de Melo & Pechere, 1990; Kelle et al., 1998; Moser & Schroder, 1997) including PEB1. Isolated PEB1 blocks the adherence of campylobacters to intestinal cells, as does antibody raised against PEB1 (Kervella et al., 1993). In addition, inactivation of the pebLA gene of *C. jejuni* leads to a 50- to 100-fold reduction of adherence to cultured epithelial cells (Pei et al., 1998). Lipopolysaccharide is another surface molecule that has been suggested as a potential adhesin (McSweegan & Walker, 1986). However, this was questioned by studies that showed that the pre-incubation of campylobacters with LPS-specific monoclonal antibodies reduced adhesion to intestinal cells only slightly, whilst pre-incubation with *C. jejuni*-specific antiserum reduced adhesion significantly. This suggests that adhesins other than LPS are more crucial for adherence (Moser & Schroder, 1995).

Although the adherence of campylobacters to intestinal cells has been well documented, the importance of adhesion to the extra-cellular matrix may also be significant (Konkel et al., 1997). Various proteins have been shown to adhere to components of the extra-cellular matrix. These include the major outer membrane protein (MOMP), which may also be involved in binding to epithelial cells (Moser et al., 1997), and CadF, a 37 kDa outer membrane protein which binds fibronectin - a large multi-functional protein found in the extra-cellular matrix. Inactivation of the cadF gene renders *C. jejuni* incapable of colonizing the cecum of newly hatched chicks (Ziprin et al., 1999).

1.4.3.5. Mechanisms for the perturbation of the absorptive capacity of the intestine

Campylobacters disturb the normal absorptive capacity of the intestine by damaging epithelial cells by cell invasion, the production of toxins and by initiating the inflammatory response. Cellular invasion has been reported in colonic biopsy samples from patients with *Campylobacter* induced colitis (van Spreeuwel et al., 1985). Several studies have shown that the degree to which campylobacters invade is strain dependent
Chapter 1: General Introduction

(Everest et al., 1992; Harvey et al., 1999; Konkel et al., 1992a; Konkel & Joens, 1989; Oelschlaeger et al., 1993). The study by Everest et al. (1992) found that there was a significant correlation between an isolates ability to invade host cells in vitro and the clinical origin of the strain; invasive strains were associated with inflammatory disease. Some strains did invade in vitro even though they came from patients classified as having non-inflammatory disease. This situation may mimic that found in developing countries, where the hosts immune system limits the progression of infection, even though the Campylobacter strains are capable of causing inflammatory disease. Therefore, host factors such as immune status are also important.

Invasion requires new bacterial proteins to be produced and secreted by campylobacters (Konkel & Cieplak, 1992; Konkel et al., 1999; Konkel et al., 1993). This secretion seems to be dependent on the CiaB protein. C. jejuni ciaB null mutants show a reduced ability to secrete proteins thought to be involved in invasion and showed a significant reduction in internalization in vitro (Konkel et al., 1999).

Motility and expression of the flagellum are additional factors necessary for invasion. Non-motile mutants have a severely reduced ability to invade epithelial cells in vitro (Grant et al., 1993; Wassenaar et al., 1991; Yao et al., 1994), and successful entry of the bacterium is greatly enhanced when the two components of flagellin, FlaA and FlaB are expressed (Wassenaar et al., 1994).

In contrast to the bacterial cell, induction of new protein synthesis in the host cell is not required for invasion (Konkel & Cieplak, 1992; Oelschlaeger et al., 1993). The method of entry into host cells is still a matter of conjecture and yet to be firmly established (reviewed in Wooldridge & Ketley, 1997). Microfilament dependent internalization has been suggested (Konkel et al., 1992a; Konkel & Joens, 1989), as has a microtubule dependent process (Hu & Kopecko, 1999; Oelschlaeger et al., 1993). In addition, a novel mechanism has been reported in which bacterial cells interact with receptors in the caveolae. This results in signal transduction events across the membrane which culminate in membrane ruffling and subsequent phagocytosis (Wooldridge et al., 1996). It is possible that all of these different mechanisms occur in vivo, but different
Chapter 1: General Introduction

strain characteristics and perhaps the presence and/or abundance of certain host receptors may influence which invasion pathway predominates.

Once inside the cell, *C. jejuni* cells appear to be mostly confined to membrane bound vesicles (de Melo et al., 1989; Konkel et al., 1992a; Russell & Blake, 1994; Russell et al., 1993). However, campylobacters have been observed free in the cytoplasm (Konkel et al., 1992a; Russell et al., 1993). These free forms have been associated with cytopathic effects including cell swelling, loss of microvilli and exfoliation into the lumen (Russell et al., 1993).

Although tissue invasion could be responsible for the clinical picture, there is growing evidence for the role of toxins in the disease process of *Campylobacter* enteritis (reviewed in Pickett, 2000; Wassenaar, 1997). Many toxins have been reported to be produced by campylobacters including a cholera-like toxin and several cytotoxins. However, the variety of different bacterial strains, host cell lines, and assays used by different researchers has led to considerable confusion about the nature of the toxins, and their role in pathogenicity (Wassenaar, 1997). The only toxin for which there is genetic data to support its existence is the cytolethal distending toxin (CLDT). This toxin has been identified from strains causing inflammatory diarrhoea (Johnson & Lior, 1985) and its production requires the transcription of three adjacent genes *cdtA*, *cdtB* and *cdtC* (Pickett et al., 1996). The three polypeptides of the toxin reveal a unique structure compared to other known proteins (Pickett et al., 1996), and combine to confer upon the toxin a novel mode of action (Aragon et al., 1997; Whitehouse et al., 1998). It is believed that the protein causes an irreversible block in the G2 phase of the host cell cycle, and thus the production of the toxin adjacent to intestinal cells prevents proliferation into functional villous cells, eroding the villous epithelium and damaging the normal absorptive function of the intestine (Aragon et al., 1997; Whitehouse et al., 1998).

Host cell invasion and the production of toxins are likely to be key elements in the pathogenesis of *Campylobacter* enteritis and could well be responsible for the net fluid loss and eventual disruption of epithelial integrity that leads to diarrhoea (Ketley, 1997). However, inflammation of the intestine is one of the pathological changes that is
apparent during *Campylobacter* enteritis, and this too could be responsible for net fluid loss (Black *et al.* 1988; Everest *et al.* 1993a). The inflammatory response is triggered by host cell damage and the presence of bacteria in the lamina propria. One way campylobacters can reach the lamina propria is by a process of translocation, although the exact mechanism by which campylobacters accomplish this is not yet understood. Transcytosis of M cells (Walker *et al.* 1988), translocation through the cytoplasm of invaded cells (transcellular; Everest *et al.* 1992; Brás *et al.* 1999) and translocation via intercellular spaces (paracellular; Russell *et al.* 1993; Welkos, 1984; Everest *et al.* 1992; Konkel *et al.* 1992b) have all been described.

As the inflammatory response develops, local vasodilation occurs and polymorphonuclear leukocytes (PMNLs) and monocytes infiltrate the lamina propria in order to counter the invading organism (Black *et al.* 1988; Duffy *et al.* 1980; Ruiz-Palacios *et al.* 1981; Russell *et al.* 1989). Vasodilation increases the net fluid loss from the intestine that has already been caused by cell damage, and this may further be exacerbated by a secretary response induced by inflammatory mediators (Everest *et al.* 1993b; Wooldridge & Ketley, 1997).

In addition to the non-specific inflammatory response, the host also develops a specific antibody response to help counter the bacterium. Upon recurrent challenge this can lead to a fully protective immunity which leads to infected people becoming asymptomatic - a situation often seen in the developing world. The acquisition of immunity to campylobacters has been demonstrated in human volunteers (Black *et al.* 1988), where initial infection led to inflammatory disease and antibody response, but subsequent re-challenge led only to colonization and not illness (Black *et al.* 1988). Whilst the specific antibody response is undoubtedly an important component in deciding the outcome of a *Campylobacter* infection, there is growing evidence that it is also associated with the development of the neurological condition Guillain Barré Syndrome (GBS) - an auto-immune disorder of the peripheral nervous system (Adams & Victor, 1993; reviewed in Nachamkin *et al.*, 1998; Nachamkin *et al.*, 2000). Previous infections, particularly of the respiratory and intestinal tracts are a common feature, but infection with *C. jejuni* is the single most commonly associated antecedent (Kaldor &
Chapter 1: General Introduction

Speed, 1984; Ropper, 1988; Winer et al., 1988). Significantly, the time lapse between Campylobacter enteritis and the development of GBS suggests that humoral immunopathogenic mechanisms are involved (Nachamkin et al., 1998).

1.5. Guillain Barré Syndrome (GBS)

1.5.1. Clinical features of Guillain Barré Syndrome

Patients with Guillain Barré Syndrome develop weakness of the limbs and respiratory muscles, as well as areflexia (loss of reflexes). The condition is self-limiting, with symptoms usually reaching their peak within 2 - 3 weeks, followed by complete or partial recovery over the next few months. Most patients recover completely, but a proportion (15 - 20%) are left with severe neurological problems and death can occur (Cole & Matthew, 1987; Ropper, 1986). Two beneficial treatments are plasmapheresis and intravenous human immunoglobulin administration which have lowered the mortality rates in the developed world to approximately 2 - 3%, but in the developing world the absence of such treatment means that the mortality rate is much higher (Van der Meche, 1996; Winer & Evans, 1993).

GBS was until recently recognized as one clinical entity, but now has been divided into several forms based on electro-physical and pathological differences (Griffin et al., 1996a; reviewed in Nachamkin et al., 1998). The most common type encountered in Europe and North America is acute inflammatory demyelinating polyneuropathy (AIDP) which is characterized by an immune attack and lymphocyte infiltration of the myelin. Severe cases may also show axonal degeneration in addition to demyelination. Less frequently seen in Europe and North America, but more common in the Far East are the axonal forms, where the axons are the primary targets of the immune system. These axonal forms can be further sub-divided into two types: acute motor-sensory axonal neuropathy (AMSAN) and acute motor axonal neuropathy (AMAN) which are extremes of a continuous spectrum of disease with AMAN being the more innocuous involving just motor fibres, whilst AMSAN is more severe
involving both motor and sensory fibres. A final, but related form of the disease is Miller Fisher Syndrome (Fisher, 1956) which is characterized by unsteadiness of gait (ataxia), areflexia, and an inability to move the eyes.

1.5.2. The epidemiology of Guillain Barré Syndrome

The annual incidence of clinically defined GBS in the developed world ranges from 0.4 - 4 cases per 100 000 people and is more common in Japan and the Far East than North America or Europe (Hughes & Rees, 1997; Nachamkin et al., 1998). Unlike Campylobacter enteritis, GBS does not show a marked seasonality of cases. However, recent data does suggest some summertime peaks in China, Mexico, Spain and Korea (reviewed in Nachamkin et al., 1998). GBS is slightly more common in males than females (Hughes & Rees, 1997) and incidence increases with age. In addition, in the West there appears to a correlation between age and the pathology of GBS; the AIDP form of GBS appears to affect the older population, whereas the AMAN form predominates in children and young adults (McKhann et al., 1993).

1.5.3. The link between C. jejuni and GBS

It has been known for many years that preceding infections are important antecedents to the development of GBS. Campylobacter infection was first reported as a potential cause in 1982 in a man who developed severe GBS two weeks after suffering Campylobacter gastroenteritis (Rhodes & Tattersfield, 1982). Following this first report, more have followed, and it has become clear that Campylobacter associated GBS cases are more severe and more likely to involve axonal injury than GBS associated with any other preceding infection (Constant et al., 1983; Molnar et al., 1982; Rhodes & Tattersfield, 1982; Ropper, 1986; Yuki et al., 1990).

Attempting to associate Campylobacter infection with the development of GBS can be problematical. The secretion of campylobacters in the stools can often last for several weeks following the resolution of symptoms (Ketley, 1995), but by the time
GBS has been diagnosed and faecal samples taken. Many GBS patients with preceding *Campylobacter* infections give falsely negative stool cultures (Nachamkin, 1997). Because of these limitations serological techniques have also been used to study the association of *Campylobacter* with GBS. IgM and IgG antibody responses peak at about 2 weeks post infection, but remain elevated for several weeks, thus it is theoretically possible to determine whether a GBS patient had a preceding *Campylobacter* infection (Blaser & Duncan, 1984). Using serological data combined with culture studies it has been estimated that at least 30 - 40% of patients are infected with *Campylobacter* 10 days to 2 weeks prior to the development of GBS symptoms (Kuroki et al., 1993; Mishu & Blaser, 1993). However, this figure is likely to be underestimation because of the limitations of both techniques (Herbrink et al., 1988; Nachamkin et al., 1998).

1.5.4. *Campylobacter* serotypes associated with GBS

In addition to confirming that there is an association between *Campylobacter* infection and the development of GBS, culture studies have also revealed that specific groups of strains may be involved. Two major systems are used world-wide for serotyping campylobacters. One is the Lior scheme based on heat-labile antigens (HL), particularly flagellin antigens, which detects over 100 *C. jejuni* serotypes (Lior, 1984). The second is the heat-stable (HS) serotyping scheme which detects over 60 types of *C. jejuni* and *C. coli* and is thought to be based on the detection of lipopolysaccharide (LPS) antigens (Frost et al., 1998; Penner et al., 1983). By HS serotyping the campylobacters isolated from patients with GBS, and comparing them with the serotypes of strains from cases of uncomplicated gastroenteritis, various studies have found that certain serotypes seem to be over represented in the GBS isolates. In Japan it appears that HS:19 strains are highly associated with GBS cases (Kuroki et al., 1993; Yuki et al., 1997), whilst in South Africa HS:41 strains predominate (Goddard et al., 1997; Lastovica et al., 1997). However, these associations do not mean that other serotypes more frequently associated with uncomplicated gastroenteritis are not also isolated from GBS patients. HS serotypes that have been reported from GBS patients.
Chapter 1: General Introduction

include HS:1, HS:2, HS:4, HS:5, HS:10, HS:16, HS:23, HS:37, HS:44 and HS:64 (reviewed in Nachamkin et al., 1998). In contrast, no association between Lior (HL) serotypes and GBS has been found (Nachamkin et al., 1998; Yuki et al., 1997).

1.5.5. Anti-ganglioside antibodies and GBS

Current opinion suggests that GBS is an auto-immune disease involving antibodies, a view supported by the fact that plasma exchange elicits a beneficial response. Studies on patients who were given bovine brain ganglioside to treat various neurological disorders have shown that a significant proportion of patients go on to develop GBS (reviewed in Yuki, 1997). This would suggest that anti-ganglioside antibodies are a cause of GBS. Gangliosides are sialic acid containing glycopospholipids which are present in the plasma membrane of vertebrate tissues and are the major surface molecules in the peripheral nervous system (Ledeen, 1985; Figure 1.1). In support of this argument, it has become clear that these ganglioside antigens are present at the sites of known antibody binding in nerve cells. For example, the GM1 ganglioside is concentrated at nodal regions as well as the paranodal myelin loops (Corbo et al., 1993; Sheikh et al., 1998), the same regions where IgG and complement have been shown to bind in autopsied AMAN patients (Griffin et al., 1996b; Hafer-Macko et al., 1996).

Serum antibodies against predominantly ganglioside GM1 have been observed in around 30% of GBS patients (Van der Meche et al., 1992), whereas more than 90% of Miller Fisher patients have IgG antibodies against gangliosides GT1a and GQ1b (Chiba et al., 1993; Willison et al., 1994). Patients with antecedent Campylobacter infections have been reported to contain anti-ganglioside antibodies too, especially anti-GM1 antibodies (Nobile-Orazio et al., 1992; Obayashi et al., 1993; Yuki et al., 1991; Yuki et al., 1993; Yuki et al., 1990), whilst no anti-GM1 antibodies were detected in patients who had C. jejuni enteritis but did not go on to develop GBS (Yuki et al., 1990). Therefore, it is hypothesized that there is molecular mimicry between infectious agents and human gangliosides i.e. some campylobacters must synthesize molecules with
### Chapter 1: General Introduction

<table>
<thead>
<tr>
<th>Chemical Structure</th>
<th>Ganglioside</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gal - (β1-3) - GalNAc - (β1-4) - Gal - (β1-4) - Glc - (β1-1) - Ceramide (α 2-3) NeuNAc</td>
<td>GM1</td>
</tr>
<tr>
<td>Gal - (β1-3) - GalNAc - (β1-4) - Gal - (β1-4) - Glc - (β1-1) - Ceramide (α 2-3) NeuNAc</td>
<td>GD1a</td>
</tr>
<tr>
<td>Gal - (β1-3) - GalNAc - (β1-4) - Gal - (β1-4) - Glc - (β1-1) - Ceramide (α 2-3) NeuNAc</td>
<td>GT1a</td>
</tr>
<tr>
<td>Gal - (β1-3) - GalNAc - (β1-4) - Gal - (β1-4) - Glc - (β1-1) - Ceramide (α 2-3) NeuNAc</td>
<td>GQ1b</td>
</tr>
<tr>
<td>GalNAc - (β1-4) - Gal - (β1-4) - Glc - (β1-1) - Ceramide (α 2-3) NeuNAc</td>
<td>GM2</td>
</tr>
<tr>
<td>Gal - (β1-4) - Glc - (β1-1) - Ceramide (α 2-3) NeuNAc</td>
<td>GD3</td>
</tr>
</tbody>
</table>

**Figure 1.1: The chemical structures of common human gangliosides.** Those structures presented here have identity or close similarity to the core oligosaccharide structures produced by some isolates of *C. jejuni*. Diagram adapted from Nachamkin *et al.* (1998). Linkages between sugars are also shown, Glc, glucose; Gal, galactose; GalNAc, N-acetylgalactosamine; NeuNAc, N-acetylneuraminic acid.
structures identical or closely similar to human gangliosides, these elicit the production of antibodies by the host which can cross-react with nerve gangliosides leading to the development of GBS.

1.5.6. Molecular mimicry and GBS

The association between the development of GBS and _C. jejuni_ isolates of certain HS serotypes (e.g. HS:19) suggested that it may be the LPS molecules produced by certain campylobacters that have structural identity to human gangliosides. Cross reactions between _Campylobacter_ LPS and anti-ganglioside antisera support this hypothesis (Prendergast _et al._, 1999; Schwerer _et al._, 1995; Wirguin _et al._, 1994; Yuki _et al._, 1993; Yuki _et al._, 1990). Furthermore, electro-physiological studies have shown that the application of monoclonal antibodies raised against _C. jejuni_ LPS to nerve cells leads to a block in neuro-transmission associated with extensive deposition of antibodies at nerve terminals (Goodyear _et al._, 1999).

The structural characterization of _Campylobacter_ LPS molecules has confirmed the striking homology, if not identity that exists between the core oligosaccharides of _Campylobacter_ LPS and human gangliosides. Characterization of the core moieties produced by _C. jejuni_ HS:19 strains show that they have structural similarity to GM1, GD1a, GT1a and GD3 gangliosides (Aspinall _et al._, 1994a; Moran & O'Malley, 1995; Yuki _et al._, 1993). Additional studies on serotypes HS:1, HS:23 and HS:36 show that the core molecules produced by these serotypes have similarity to GM2, whilst the core molecules of _C. jejuni_ HS:4 and HS:41 mimic gangliosides GD1a and GM1 respectively (Aspinall _et al._, 1994b; Prendergast _et al._, 1998). However, it is clear that mimicry of gangliosides is not just limited to those strains associated with GBS, and that enteritis isolates may also mimic gangliosides (Moran & O'Malley, 1995; Nachamkin _et al._, 1999; Prendergast _et al._, 1999; Sheikh _et al._, 1998). This observation, coupled with the fact that only approximately 1 in 1000 cases of _C. jejuni_ enteritis is followed by GBS (Nachamkin _et al._, 1998) suggests that other bacterial or host factors are probably involved with the pathogenesis of GBS in addition to molecular mimicry.
Chapter 1: General Introduction

1.6. Lipopolysaccharide (LPS) - Overview

Whilst potentially important in the development of GBS, this is by no means the only significant role LPS plays in both Campylobacter pathogenesis and in the wider picture of Gram negative bacterial pathogenesis. LPS is an essential and characteristic component of most Gram negative outer membranes. As a major surface component it plays a non-specific role as a physical barrier to the entry of deleterious compounds and phages (Hancock et al., 1994), but also plays a more distinct role as a virulence determinant in many pathogens (Raetz, 1996). LPS consists of three covalently linked domains; lipid A, core oligosaccharide and O-chain (Figure 1.2). The hydrophobic lipid A moiety is composed of fatty acids and sugars and anchors the molecule in the outer leaflet of the outer membrane. This is linked to the oligosaccharide core region, which contains sugars and their derivatives such as 3-deoxy-D-manno-octulosonic acid (Kdo). The core oligosaccharide can be conceptually divided into two regions; the inner (lipid A proximal), and the outer core which links to the outermost O-chain. This extends from the cell and consists of repeating oligosaccharide units generally composed of 3 - 6 sugars (Raetz, 1996). Some species of Gram negative bacteria, for example Neisseria menigitidis and Haemophilus influenzae, produce an LPS lacking this polymeric O-antigen; this is termed lipooligosaccharide (LOS; Figure 1.2).

Elements of the LPS/LOS molecule such as the lipid A and inner core moieties are highly structurally conserved between species, whilst the outer core and O-chain can vary considerably, even between members of the same species. These structural differences can convey different immunological and biological properties on the cell and hence can effect pathogenic potential. Understanding the structure, genetics and biosynthesis of LPS/LOS is therefore an important step in understanding the pathogenesis of Gram negative bacteria. Several species have been studied to try and understand these elements of LPS/LOS, the following sections will discuss this work but will concentrate on a few more intensely studied species - these include Escherichia coli and Salmonella serovars which produce LPS, and H. influenzae which produces LOS.
Figure 1.2: Schematic representation of the generalized structure of (A) LPS and (B) LOS. Adapted from Moran & Penner (1999). Both structures contain lipid A and core moieties, but LPS contains the additional polymorphic O-chain moiety.
1.7. *Escherichia coli* and *Salmonella* LPS

1.7.1. The Kdo$_2$-lipid A moiety: structure and biosynthesis

Structurally, the lipid A moiety of *E. coli* and *Salmonella* is a hexa-acylated disaccharide of glucosamine substituted at the 1' and 4' positions with phosphate (Raetz, 1996; Rietschel *et al.*, 1994). Lipid A does not exist as such in cells (Galloway & Raetz, 1990). The minimum LPS structure required for growth of *E. coli* consists of lipid A containing five acyl chains (Somerville *et al.*, 1996), glycosylated with two 3-deoxy-D-manno-octulosonic acid (Kdo) residues. Kdo is an eight carbon sugar, the biosynthesis of which is necessary for growth (Goldman *et al.*, 1987; Hammond *et al.*, 1987).

The biosynthesis of the Kdo$_2$-lipid A in *E. coli* is now well documented and is shown in Figure 1.3 (reviewed in Raetz, 1998; Schnaitman & Klena, 1993; Wyckoff *et al.*, 1998). Synthesis begins with the acylation of UDP-N-acetylglucosamine, which is catalyzed by LpxA, to form UDP-3-hydroxymyristoyl-N-acetylglucosamine. The next reactions involve N-deacylation and the transfer of a second acyl group to form UDP-2,3-diacylglucosamine. These steps are catalyzed by two enzymes, LpxC and LpxD respectively. Condensation of one molecule of UDP-2,3-diacylglucosamine with one molecule of lipid X (UDP-2,3-diacylglucosamine-1-phosphate) is the next reaction, catalyzed by LpxB. A membrane bound kinase, LpxK then incorporates a 4'-monophosphate to generate an intermediate known as lipid IV$_A$. The presence of the 4'-monophosphate moiety is an absolute requirement for the next step, which is the addition of two Kdo residues to the lipid IV$_A$ molecule. Kdo is synthesized via the action of two essential proteins, KdsA and KdsB (Schnaitman & Klena, 1993). The KdsA enzyme (Kdo-8-phosphate synthetase) combines D-arabinose-5-phosphate and phosphoenolpyruvate to generate Kdo-8-phosphate (Ray, 1980; Strohmaier *et al.*, 1995). Following the removal of the phosphate, Kdo is activated to CMP-Kdo by the CMP-Kdo synthetase KdsB (Goldman *et al.*, 1986; Strohmaier *et al.*, 1995). Once activated, two Kdo residues are transferred to the lipid A moiety - both of these
Figure 1.3: Diagrammatic representation showing the biosynthesis of Kdo₂-lipid A in *E. coli* and *Salmonella* serovars. Adapted from Wyckoff *et al.* (1998). Gray ovals indicate glucosamine residues. 'zig-zag' lines represent fatty acyl chains and white boxes are Kdo moieties. The enzymes that catalyze each reaction are indicated. Ac. acetyl group; ACP. acyl carrier protein; ATP. adenosine triphosphate; Kdo. 3-deoxy-D-manno-octulosonic acid; lipid IV₅. tetraacyl-disaccharide 1,4'-bis-phosphate; lipid X. UDP-2,3-diacylglicosamine-1-phosphate; 3-OH-C₁₄. R-3-hydroxymyristate; UDP. undecaprenol; UDP-DAG. UDP-2,3-diacylglicosamine; UDP-GlcNAc. UDP-N-acetylglicosamine.
additions are catalyzed by the same enzyme WaaA (Belunis et al., 1995; Belunis & Raetz, 1992). In the final steps of Kdo₂-lipid A biosynthesis, the WaaM and WaaN proteins add the final acyl chains (Khan et al., 1998; Somerville et al., 1999).

1.7.2. The Kdo₂-lipid A moiety: genetics

The genes encoding the enzymes responsible for the biosynthesis of the Kdo₂-lipid A moiety of *E. coli* and *Salmonella* have been known for some time (Garrett et al., 1997; Raetz, 1996; Schnaitman & Klena, 1993). Several of the genes, including *lpxA, B* and *D* are found together in an eleven gene cluster at 4 min on the *E. coli* K-12 gene map (Schnaitman & Klena, 1993). Most of the other Kdo₂-lipid A biosynthesis genes including *lpxC, waaN, waaM* and *kdsA* are unlinked (Schnaitman & Klena, 1993). The remaining lipid A biosynthesis gene, *lpxK*, is found at 21 min on the *E. coli* linkage map, where it forms part of an operon with the *kdsB* and *msbA* genes. The function of the MsbA protein will be discussed in subsequent sections.

1.7.3. The Kdo₂-lipid A moiety: function

The lipid A moiety is noteworthy in pathogenesis because it is often responsible for the toxic manifestations and generalized inflammation caused by Gram negative infections (Darveau, 1998). Many cell types, including epithelial cells and macrophages, respond to the presence of lipid A by synthesizing cytokines and inflammatory mediators such as TNFα and IL-1β (Raetz, 1990; Rietschel et al., 1994). The regulated expression of such responses is meant to eliminate bacterial infections, but in debilitated or immunosuppressed patients, the rapid release of such factors can lead to a shock syndrome characterized by increased vascular permeability, severe hypotension, multiple organ failure and death (Moran, 1995b).

Studies on the role of particular epitopes from the Kdo₂-lipid A moiety in pathogenesis have been hampered by the fact that mutation of the *lpx, kds, waaA* or *waaM* genes leads to a conditionally lethal phenotype (Darveau, 1998). However,
mutation of the \textit{waaN} gene in \textit{E. coli} and \textit{S. typhimurium} has been possible and leads to loss of one of the secondary acyl chains present on the Kdo$_2$-lipid A moiety (Khan \textit{et al.}, 1998; Somerville \textit{et al.}, 1996). Such mutants are impaired in their ability to induce TNF\(\alpha\) and IL-1\(\beta\) \textit{in vitro} and \textit{in vivo}, suggesting that the secondary fatty acid chains that have a pivotal role in host - bacterium interactions (Khan \textit{et al.}, 1998; Somerville \textit{et al.}, 1996). It has also been reported that \textit{S. typhimurium} is able to alter the fatty acid substitutions, thus modifying the structure of the lipid A moiety (Guo \textit{et al.}, 1997). These structural modifications caused a reduction in the LPS-mediated expression of TNF\(\alpha\) by host monocytes and may represent a mechanism for the bacterium to gain a selective advantage in the hosts tissues. It was also shown that the modifications were regulated by the PhoP-PhoQ two component regulatory system, which regulates numerous genes in response to host micro-environments. This may imply that lipid A structure can be controlled along with other virulence determinants as part of an integrated regulatory system (Guo \textit{et al.}, 1997).

1.7.4. Core oligosaccharide: structure and biosynthesis

The core oligosaccharides of \textit{E. coli} and \textit{Salmonella} share a common organization with the inner core containing three Kdo and three L-glycero-D-manno-heptose residues, whilst the outer core contains a tri-hexose backbone (Figure 1.4). Substitutions to this tri-hexose backbone distinguish the five core types of \textit{E. coli}; R1, R2, R3, R4 and K12 (reviewed in Heinrichs \textit{et al.}, 1998c). Until recently, only one core structure had been described for \textit{Salmonella} species (Holst & Grade, 1992), however a second structure has now been described in serovar Arizonae IIIa O:62 (Olsthoorn \textit{et al.}, 1998). The five \textit{E. coli} and two \textit{Salmonella} structures are shown in Figure 1.5.

The biosynthesis of the core is thought to involve a coordinated complex of membrane associated glycosyltransferases which are found on the cytoplasmic face of the inner membrane (Raetz, 1996). These add the core residues to a pre-formed lipid A moiety, but unlike the process of Kdo$_2$-lipid A biosynthesis in which each enzyme can only act on the product of the preceding step in the pathway, the order of core domain
Figure 1.4: Generalized structure of the sugar backbone of the \textit{E. coli} and \textit{Salmonella} core oligosaccharide. Adapted from Heinrichs \textit{et al.} (1998c). The inner core is highly conserved containing three Kdo residues (3-deoxy-D-manno-octulosonic acid) and three Hep (L-glycero-D-manno-heptose) residues and is phosphorylated. The outer core consists of a three hexose (hex) backbone substituted with varying hexose sugars (detailed in Figure 1.5). The conserved inner core biosynthetic enzymes are also shown. P, phosphate; PEtn, 2-aminoethyl phosphate; PPEtn, pyrophosphoethanolamine.
Figure 1.5: Chemical structures and biosynthetic enzymes of the outer core oligosaccharides of *E. coli* and *Salmonella*. Adapted from Heinrichs et al. (1998c) and Amor et al. (2000). The asterisks indicate the points of attachment to O-chain, but this has only been experimentally determined for the R1, R2, *S. enterica* sv. Typhimurium and *S. enterica* sv. Arizonae core types. The linkages between sugars are also shown, and the enzymes that catalyze each linkage are indicated where known. Glc, glucose; Gal, galactose; GlcNAc, N-acetylglucosamine; Hep, L-glycero-D-manno-heptose.
biosynthesis is less rigid. The major sugars of the core backbone are added sequentially, but substitutions are added independently of one another (Brozek & Raetz, 1990; Raetz, 1996).

As would be expected from the common inner core structure seen in *E. coli* and *Salmonella* (Figure 1.4), the enzymes responsible for its biosynthesis are also highly conserved. The WaaA enzyme adds the two Kdo residues to the lipid IV<sub>A</sub> molecule as part of the biosynthesis of lipid A (discussed in section 1.7.1; Belunis & Raetz, 1992; Belunis et al., 1995), but the identity of the enzyme responsible for the addition of the non-stoichiometric third Kdo residue is still unclear (Heinrichs et al., 1998c). The addition of the first L-glycero-D-manno-heptose (Hep I) is catalyzed by the WaaC enzyme (Kadrmas & Raetz, 1998), the second by WaaF (Sirisena et al., 1994), whilst the third is added by WaaQ (Yethon et al., 1998). The decoration of the inner core with some of the phosphoryl substitutions is attributed to the WaaP and WaaY proteins (Yethon et al., 1998). The enzyme involved with the addition of the 2-aminoethyl phosphate substitution to Kdo II is as yet unknown (Heinrichs et al., 1998c).

The enzymes involved with the assembly of the outer core vary in different strains, which leads to the synthesis of the different core types reported for *E. coli* and *Salmonella* (Heinrichs et al., 1998c; shown in Figure 1.5). The presence of different transferases is the result of differences in the gene content of strains from each core type (Heinrichs et al., 1998c). This will be discussed in the following section. Finally, following the completion of the outer core, the complete lipid A-core structure is transferred to the periplasmic face of the inner membrane prior to O-chain attachment (Whitfield et al., 1997). This transfer process requires energy (Marino et al., 1985) and is thought to involve the ATP-binding cassette (ABC) transporter MsbA (Polissi & Georgopoulos, 1996).

1.7.5. Core oligosaccharide: genetics

In *E. coli* and *Salmonella*, most of the genes encoding for the biosynthesis of core oligosaccharide are clustered together in the *waa* region of the chromosome. This
Chapter 1: General Introduction

region is found at 81 - 82 min on the E. coli K-12 and Salmonella linkage maps (reviewed in Schnaitman & Klena, 1993; Heinrichs et al., 1998c). In E. coli K-12, the cluster contains fifteen genes which are organized into three operons (Figure 1.6). The first contains gmhD, waaC and waaF - enzymes required for the biosynthesis and transfer of the inner core L-glycero-D-manno-heptose residues. In some strains the operon also contains waaL, which is responsible for the ligation of the lipid A-core molecule to O-chain (Whitfield et al., 1997). The central operon (beginning with waaQ) contains the genes responsible for the biosynthesis of the outer core, and for the modification of the oligosaccharide with phosphoryl substitutions. Differing gene content in this operon is responsible for the presence or absence of particular sugar transferases, which in turn results in the differing core types shown in Figure 1.5 (reviewed in Heinrichs et al., 1998c). The third operon contains the waaA gene which encodes the bi-functional Kdo transferase (Belunis & Raetz, 1992).

1.7.6. Core oligosaccharide: function

The core oligosaccharide is not considered to be a virulence factor per se, but plays a role in establishing the essential barrier function of the outer membrane (Nikaido & Vaara, 1985). The inner heptose region is crucial for outer membrane stability, as phosphorylation of this region is proposed to enable cross-linking of adjacent LPS molecules by divalent cations or polyamines, and interactions with proteins via positively charged amino acids (Raetz, 1990; Schnaitman & Klena, 1993). This crucial function may explain why the inner core moieties of E. coli and Salmonella are so highly structurally conserved. The inability to produce or incorporate the heptose residues, or loss of phosphoryl groups leads to substantial structural and compositional changes in the outer membrane. Proteins, especially porins, become profoundly deficient in the outer membrane of such strains (Ames et al., 1974; Raetz, 1990), and the amount of phospholipid inserted into the outer leaflet of the membrane increases (Raetz, 1996). These changes result in an increased sensitivity to detergents and antibiotics (Nikaido & Vaara, 1987; Parker et al., 1992; Raetz, 1996).
Figure 1.6: Organization and gene content of *E. coli* and *Salmonella* core oligosaccharide biosynthetic clusters. Based on Heinrichs et al. (1998c). Arrows represent open reading frames, and indicate the direction of transcription. The colour of each arrow indicates at which stage of the core biosynthesis pathway the gene product is involved: black, inner core assembly; red, outer core assembly; blue, ligation and green, unknown. Sequence data from *S. enterica* sv. Arizonae and upstream of *waaC* in the *E. coli* R-stains is unavailable. This diagram is not shown to scale.
disruption of the outer membrane also causes a ‘leaky’ phenotype in which periplasmic proteins are released into the extra-cellular environment, and results in the secretion of an inactive form of *E. coli* haemolysin (Bauer & Welch, 1997; Parker *et al.*, 1992).

### 1.7.7. O-chain: structure and function

The O-chain moiety of LPS is composed of repeating oligosaccharide units, with each unit commonly composed of three to six sugars (reviewed in Whitfield, 1995). More than 50 and 170 different O-chain structures have now been recognized in *Salmonella* and *E. coli* respectively (Orskov *et al.*, 1977; Popoff & Minor, 1997). This variation is caused not only by differences in monosaccharide constituents, but also by a diversity of linkages, and the addition of non-sugar moieties such as amino acids (Reeves *et al.*, 1996). These differences can be detected serologically and allow the classification of strains according to O serotype. As in core biosynthesis, the enzymes for O-chain biosynthesis are encoded by a set of genes which are organized in a cluster (Schnaitman & Klena, 1993). Differing gene content in this cluster is responsible for the presence or absence of differing sugar transferases, which in turn results in the differing O-chain types (Whitfield, 1995).

The O-chain component of LPS provides a hydrophilic surface layer to the cell, which may mask underlying conserved core epitopes (Whitfield *et al.*, 1997). Many strains that produce O-chain are serum resistant. This is due to the fact that O-side chains preferentially bind C3, leading to the formation of weakly bound C5b-9 complexes which are sterically hindered from integrating into and disturbing the hydrophobic outer membrane (Hackett *et al.*, 1987; Joiner *et al.*, 1986). It appears that the length of the chains is critical for this resistance (Burns & Hull, 1998; Goldman *et al.*, 1984; Porat *et al.*, 1992; Whitfield *et al.*, 1997). O-chain has also been implicated in the ability of some strains to evade ingestion by human polymorphonuclear leukocytes (Stinavage *et al.*, 1989; Svanborg-Edén *et al.*, 1987).
1.7.8. O-chain: biosynthesis

The biosynthesis of O-chain occurs independently from that of the lipid A-core moiety, and involves the sequential transfer of sugars onto a lipid carrier, undecaprenol phosphate (und-P) by sugar transferases. Three pathways for O-chain biosynthesis in E. coli and Salmonella have now been described which vary in the components required for polymerization, in the cellular location of polymerization, and in the way the O-chain is transferred across the inner membrane (reviewed in Whitfield et al., 1997). Following biosynthesis, the O-chain is ligated to the core-lipid A molecule in the periplasm, and translocated out of the cell. Unfortunately, the mechanisms for this later process are not yet known (Whitfield et al., 1997; Whitfield & Valvano, 1993).

1.7.8.1. The Wzy-dependent biosynthesis of O-chain

The Wzy-dependent pathway is responsible for the synthesis of the majority of E. coli and Salmonella hetero-polymeric O-chains, and involves the assembly of und-P linked O-chain units at the cytoplasmic face of the inner membrane. These units are constructed by the sequential transfer of sugar residues to the non-reducing end of the molecule (Whitfield, 1995). In E. coli, the first residue of Wzy-dependent O-chains is often N-acetylglucosamine (GlcNAc), the transfer of which is catalyzed by the WecA enzyme (Rick et al., 1994), whilst in Salmonella the first sugar is often galactose - transferred by the WbaP enzyme (Wang et al., 1996; Wang & Reeves, 1994). Following the attachment of the remaining sugar residues, these units are then translocated across the inner membrane by a process of trans-membrane ‘flipping’ catalyzed by the cytoplasmic membrane protein Wzx (Liu et al., 1996).

Following translocation, the O-chain sub-units are polymerized by the Wzy protein by a mechanism involving the successive addition of the reducing end of the polymer to the non-reducing end of the und-P linked sub-units (Lukomski et al., 1996). The chain length of the complete O-chain polymer (i.e. the modal number of O-chain units per complete O-chain polymer) is regulated by the Wzz protein (Batchelor et al., 1996).
Chapter 1: General Introduction

The precise mode of action of Wzz is unknown, but it is proposed to interact with the O-chain polymerase (Wzy) in the periplasm to control Wzy-mediated polymerization (Bastin et al., 1993; Morona et al., 1995; Whitfield et al., 1997).

1.7.8.2. The ABC transporter dependent synthesis of O-chain

The ABC transporter dependent pathway is responsible for the synthesis of mono-polymeric O-chains, such as the *E. coli* O:8 and O:9 O-chains. Like the Wzy-dependent pathway, this process also depends on the transfer of sugars by transferases onto und-P at the cytoplasmic face of the inner membrane (Whitfield et al., 1997). The process begins with the WecA enzyme transferring an N-acetylglucosamine (GlcNAc) residue to und-P (Rick et al., 1994). All O-chain molecules produced by the ABC transporter dependent pathway, but only some of the Wzy-dependent pathway, are initiated by the WecA enzyme (Whitfield, 1995). The transferred GlcNAc residue subsequently serves as the acceptor molecule for the sequential transfer of sugars, but does not become part of the repeating O-chain units (Whitfield, 1995). In this pathway only one und-P molecule is used per O-chain polymer and there is no requirement for a polymerase (Wzy). Instead the specificity of the transferases dictates the repeating unit structure. The complete polysaccharide is then transported across the inner membrane by an ATP-binding cassette transport system (Bronner et al., 1994; Kido et al., 1995; Whitfield, 1995). There is therefore no requirement for the Wzx protein because there is no requirement for the export of individual und-P linked O-chain sub-units (Whitfield et al., 1997). The ATP-binding cassette transport system is composed of two proteins; Wzm, an integral membrane protein and Wzt, an ATP-binding protein. Two sub-units of each associate to form a functional transporter (Fath & Kolter, 1993). The Wzt proteins use ATP hydrolysis to drive transport, whilst the integral membrane proteins are thought to form a pore-like structure through which the completed O-chain is transported (Higgins, 1992; Kerppola et al., 1991).
1.7.8.3. The synthetase dependent biosynthesis of O-chain

Recently, a third mechanism for O-chain biosynthesis has been reported in *S. enterica* serovar Borreze for the homo-polymeric O:54 antigen (Keenleyside & Whitfield, 1996). This pathway is similar to the ABC transporter dependent pathway in that it requires WecA to provide a GlcNAc residue to initiate the pathway. However, export of the polymer does not require a Wzx `flippase' or an Wzt/Wzm ABC transporter. Indeed, no dedicated ABC transporter system or Wzx homologs exist in the O:54 system. Instead, the glycosyltransferase (synthetase) WbbF protein, which functions to polymerize the N-acetylmannosamine (ManNAc) homopolymer may also function as an exporter (Keenleyside & Whitfield, 1996). The C-terminal transmembrane region of the protein is thought to form a pore through which the polysaccharide is transported (Keenleyside & Whitfield, 1996).

1.7.8.4. The ligation of O-chain to the lipid A-core moiety

Once the completed O-chain has been transferred to the periplasmic face of the inner membrane, it is attached the lipid A-core molecule (Whitfield *et al.*, 1997). The mechanism for this process remains unresolved, and only one enzyme has been shown to be needed for ligation. This is the WaaL protein, mutation of which leads to a ligation deficient phenotype (Heinrichs *et al.*, 1998a; Heinrichs *et al.*, 1998b). It remains to be seen exactly how the WaaL protein operates, and if there are any as yet unidentified proteins involved in the ligation reaction (Heinrichs *et al.*, 1998a), but the ligase enzyme shows no obvious specificity for the structure of the ligated polysaccharide (Whitfield *et al.*, 1997). However, the nature of the accepting core molecule is certainly important for ligation (Heinrichs *et al.*, 1998c). Mutation of WaaK, a sugar transferase which attaches a GlcNAc residue to the outer core of *E. coli* R2 and *S. enterica* serovar Typhimurium eliminates ligation (MacLachlan *et al.*, 1991; Heinrichs *et al.*, 1998a; Klena *et al.*, 1992).
In addition to being the acceptor molecule for O-chain, it is becoming increasingly obvious that the lipid A-core molecule can also serve as an acceptor for other independently synthesized polymers (Whitfield et al., 1997). In E. coli, the heteropolysaccharide enterobacterial common antigen (ECA) can be attached to lipid A-core molecules to form ECA\textsubscript{LPS}, and the capsular K antigens can also be attached forming K\textsubscript{LPS} (Whitfield et al., 1997; Rick et al., 1994; Whitfield & Roberts, 1999). Salmonella strains can attach O-chain to the core-lipid structure or alternatively attach the minor surface polymers T1 or T2 antigens (Whitfield et al., 1997). The WaaL protein catalyses the attachment of all these polymers to the lipid A-core moiety. Furthermore, there are other similarities between the biosynthesis of O-chain and some of these alternative polymers such as the involvement of the Wzy and Wzx proteins. These issues have been extensively reviewed elsewhere and are beyond the immediate scope of this study (Whitfield et al., 1997; Whitfield & Roberts, 1999).

1.7.9. O-chain: genetics

Similar to the genes coding for core synthesis, those coding for O-chain biosynthesis are mainly clustered, and map to 44 min on the E. coli linkage map (reviewed in Raetz, 1996; Schnaitman & Klena, 1993). Less than ten E. coli O-chain synthesis clusters have been partially, or completely sequenced, but considerable variation in the sequence and gene content between these clusters occurs (Marolda & Valvano, 1993; Paton & Paton, 1999; Reeves et al., 1996; Wang & Reeves, 1998). This variation is responsible for the generation of the sometimes radically different O-chain structures produced by strains of differing serotype (Reeves et al., 1996). However, various generalizations about the O-chain biosynthesis clusters of E. coli and Salmonella can be made. They are usually between 10-15 kilobases in length and encode enzymes necessary for the synthesis of sugar residues, those involved with sugar transfer, and additional enzymes needed for completion of the LPS molecule (Reeves et al., 1996). Furthermore, O-chain clusters appear to have a low G+C content, in some cases below 30% G+C, which is considerably lower than that of the flanking DNA and
the rest of the *E. coli* genome. This may suggest that O-chain gene clusters were acquired by horizontal transfer from another species (Jiang *et al.*, 1991; Reeves, 1993).

An example of a O-chain biosynthesis cluster is shown in Figure 1.7A. This recently characterized example is from *E. coli* O:113, and contains nine closely linked genes (Paton & Paton, 1999). Briefly, the cluster contains four genes encoding glycosyltransferases (*wbnA, B, D* and *E*) which are required to transfer four of the five sugars of the O:113 O-chain unit (shown in Figure 1.7B; Paton & Paton, 1999). The missing transferase is presumed to be *wecA*, which transfers the first N-acetylglucosamine residue to und-P to initiate synthesis of the O-chain unit. The cluster also contains an O-subunit flippase and polymerase, which are needed for transport of the completed O-chain units across the inner membrane, and polymerization of multiple units to form a completed O-chain molecule. The *wbnC* gene has homology to known O-acetylation proteins, and is presumably needed for the O-acetylation of *E. coli* O:113 O-chain units. These are known to be acetylated, although the precise nature and position of the modification is yet to be identified (Parolis & Parolis, 1995). The *galE* gene encodes for a UDP-glucose-4-epimerase which is required for the inter-conversion of UDP-glucose to UDP-galactose, and ensures adequate supplies of UDP-galactose for O-chain biosynthesis (Paton & Paton, 1999). The final gene of the cluster is *gnd*, which is found at the 3' end of many O-chain biosynthesis clusters (Wang & Reeves, 1998). Gnd is a 6-phospho-gluconate dehydrogenase, but there is no evidence for this enzyme having a role in O-chain biosynthesis.

1.8. *Haemophilus influenzae* LOS

*H. influenzae* is a commonly found commensal of the human respiratory tract, but can be a major cause of diseases such as meningitis, bacteraemia, pneumonia and otitis media (Hood *et al.*, 1996; Turk, 1984). Electrophoretic analyses of *H. influenzae* LPS molecules show that they lack the repeating polysaccharide O-chain (Campagnari *et al.*, 1990; Gibson *et al.*, 1993) - these molecules are therefore known as lipooligosaccharide (LOS). *Haemophilus* LOS plays an important part in virulence and
Figure 1.7: Schematic diagram showing the organization and gene content of the
*E. coli* O:113 O-chain biosynthetic cluster (A) and the structure of the O-chain
repeat unit (B). Adapted from Paton & Paton (1999) and Parolis & Parolis (1995). The
open reading frames and direction of their transcription are indicated by arrows. The O-
chain repeat unit contains O-acetyl groups, but the position of these is currently unclear.
Gal, galactose; GalA, galacturonic acid, GlcNAc, N-acetylglucosamine and GalNAc, N-
acetylgalactosamine. Linkage types between sugar residues are also shown.
has been implicated in every stage of \textit{H. influenzae} pathogenesis from colonization, systemic dissemination, to the cytotoxic injury of target tissues (Cope \textit{et al.}, 1990; DeMaria \textit{et al.}, 1997; Hood \textit{et al.}, 1996; Moxon & Maskell, 1992). This role in pathogenesis is a direct result of the structural, and therefore biosynthetic characteristics of \textit{H. influenzae} LOS molecules (Hood \textit{et al.}, 1996). The study of these characteristics has been greatly enhanced by the completion of the genome sequence of \textit{H. influenzae} strain Rd (Fleischmann \textit{et al.}, 1995). This allowed the identification and subsequent mutation of many genes thought to be involved in LOS biosynthesis (Hood \textit{et al.}, 1996). However, thorough structural and biochemical analyses of these LOS mutants is still required to confirm the role of particular gene products in the biosynthesis of LOS. As such, our knowledge of the biosynthesis and function of LOS in \textit{H. influenzae} lags behind our understanding of \textit{E. coli} LPS. The work described in the following sections will discuss important aspects of \textit{H. influenzae} LOS such as structure, function and genetics. In addition, the similarities and differences between \textit{H. influenzae} LOS and \textit{E. coli} LPS will be highlighted.

1.8.1. The Kdo-lipid A moiety: structure, function and biosynthesis

The lipid A moiety of \textit{H. influenzae} is structurally very similar to that of \textit{E. coli} and thus is assumed to have similar functions. The only slight differences between the two structures are the acylation and phosphorylation patterns (Helander \textit{et al.}, 1988; Masoud \textit{et al.}, 1997; Moran, 1995b). This conservation in structure is thought to reflect similarities in biosynthesis. The most striking difference between the Kdo-lipid A regions of \textit{E. coli} and that of \textit{H. influenzae} is the presence of differing numbers of Kdo residues. \textit{E. coli} LPS contains two Kdo residues and can also contain an additional non-stoichiometric third Kdo residue (Heinrichs \textit{et al.}, 1998c). In contrast, structural analysis of the \textit{H. influenzae} LOS shows the presence of only one Kdo residue (Helander \textit{et al.}, 1988; Phillips \textit{et al.}, 1992). Furthermore, this sole Kdo residue is phosphorylated at the 4-OH position that is occupied by the second Kdo residue in \textit{E. coli} (Phillips \textit{et al.}, 1992). Recent work characterizing the WaaA enzyme of \textit{H. influenzae}
influenzae has shown that in this species, WaaA is a mono-functional Kdo transferase (White *et al.*, 1997), rather than the bi-functional Kdo transferase identified in *E. coli* (Belunis *et al.*, 1995). In addition, the enzyme responsible for the phosphorylation of the Kdo residue in *H. influenzae* has been identified (White *et al.*, 1997; White *et al.*, 1999). This activity of this enzyme (designated KdkA - Kdo kinase A) is absent from *E. coli* (White *et al.*, 1999). Mutation of the *kdkA* gene in *H. influenzae* yields mutants which show an attenuated virulence in the infant rat model of infection, suggesting that phosphorylation of the Kdo residue is important in pathogenesis (Hood *et al.*, 1996).

1.8.2. The Kdo-lipid A moiety: genetics

All of the Kdo-lipid A biosynthesis genes are readily identifiable in the complete genome sequence of *H. influenzae* strain Rd (Fleischmann *et al.*, 1995; available at [http://www.tigr.org/tdb/CMR/ghi/htmls/SplashPage.html](http://www.tigr.org/tdb/CMR/ghi/htmls/SplashPage.html)). The genes involved with the biosynthesis of lipid A, including *lpxC, D, waaM, waaN* and *lpxK* are all unlinked with the exception of the *lpxA* and *lpxB* genes which are contiguous (Fleischmann *et al.*, 1995). The genes involved with the biosynthesis and transfer of the single Kdo residue, *waaA, kdsA* and *kdsB* are also unlinked (Fleischmann *et al.*, 1995; Hood *et al.*, 1996). The recently identified *kdkA* gene is part of a cluster of LOS related biosynthesis genes (White *et al.*, 1999). This cluster will be discussed in subsequent sections.

Because of the conditionally lethal phenotype that mutants with inactivated Kdo-lipid A biosynthesis genes exhibit, the analysis of these genes in *H. influenzae* has been limited. However, *waaM* mutants, which lack some of the secondary acyl chains present on the lipid A moiety, have been obtained (Lee *et al.*, 1995b). These mutants were unaffected in *in vivo* colonization, but showed a reduced ability to induce otitis media and persist in the inner ear of chinchillas (DeMaria *et al.*, 1997). This suggests that the acyl chains of lipid A are important for the virulence of *H. influenzae* (DeMaria *et al.*, 1997).
1.8.3. The oligosaccharide moiety: structure

The first complete structural characterizations of *H. influenzae* oligosaccharide molecules have only recently been reported (Masoud *et al.*, 1997; Risberg *et al.*, 1999; Risberg *et al.*, 1997). These studies have shown that the inner core moiety of *H. influenzae* LOS is highly conserved between strains and is composed of three Hep (L-glycero-D-manno-heptose) residues attached in a chain to one Kdo residue, whilst the outer core contains glucose and galactose residues (Figure 1.8). Although only a few strains have been analyzed, these studies confirmed earlier work indicating marked inter-strain heterogeneity in the outer portion of the LOS molecule (Gulig *et al.*, 1987; Kimura *et al.*, 1987; Zamze & Moxon, 1987). At least ten serotypes of *H. influenzae* LOS have now been identified (Campagnari *et al.*, 1989; van Alphen *et al.*, 1990).

1.8.4. The oligosaccharide moiety: phase variation

The structures of *H. influenzae* LOS molecules are difficult to determine due to molecular heterogeneity caused by phase variation and variable substitutions like phosphorylation, acetylation and sialylation (Masoud *et al.*, 1997; Risberg *et al.*, 1999; Risberg *et al.*, 1997). Therefore, any one strain may be producing a variety of similar, but distinct LOS structures at any one time, and thus the reported structures often only represent the most commonly found LOS produced by a given strain (Rahman *et al.*, 1999). The phase variation of *H. influenzae* LOS is likely to be an adaptive mechanism which is advantageous for the survival of the bacterium in differing host micro-environments, and when confronted with differing aspects of the host immune response (Weiser & Pan, 1998; Weiser *et al.*, 1990b). This spontaneous loss or gain of residues occurs at a high frequency \((10^{-2} - 10^{-3} \text{ per generation})\), and has complicated our understanding of the role of LOS in commensal and pathogenic interactions with the host (Weiser *et al.*, 1990a).
Figure 1.8: Chemical structures of the major oligosaccharides produced by *H. influenzae* strains Eagan and Rd. Adapted from Lysenko et al. (2000). Both strains elaborate mixtures of glycoforms in which further chain extension can occur from Hep II or Hep III respectively. Glc, glucose; Gal, galactose; Hep, L-glycero-D-manno-heptose; Kdo, 3-deoxy-D-manno-octulosonic acid; PEt, phosphoethanolamine. Linkage types between sugar residues are also shown.
Chapter 1: General Introduction

One of the most commonly seen variations is the presence of phosphorylcholine (choP) attached to the terminal hexose residues extending from either Hep I or Hep III (Schweda et al., 1997). This choP substitution is the target of C-reactive protein (CRP) - a mediator of the host innate immune response (Weiser et al., 1998). Expression of choP therefore confers sensitivity to killing mediated by complement because CRP binds choP resulting in the deposition of C1q and the activation of the classical complement pathway (Weiser et al., 1998). Despite this increased sensitivity to CRP-mediated killing, the expression of choP seems to be important for persistence of the bacterium in the respiratory tract (Weiser et al., 1997). Genetic analysis of strains present on the mucosal surface of human respiratory tract indicated that >90% of organisms had a phase-on choP-expressing phenotype (Weiser et al., 1998).

Another phase variable substitution in H. influenzae LOS is the epitope Galα (1-4)βGal (High et al., 1993; Weiser & Pan, 1998). This structure is also a component of P blood group antigens and related glycoconjugates in humans (High et al., 1993; Mandrell et al., 1992; Virji et al., 1990). Expression of this epitope by H. influenzae seems to protect the bacterium from killing mediated by naturally acquired antibody and complement (Weiser & Pan, 1998). This effect suggests that H. influenzae may incorporate LOS structures against which humans do not normally produce antibody to evade immune mediated clearance (Weiser & Pan, 1998).

In addition to the structural variation already discussed, H. influenzae can also modify LOS structures by the addition of sialic acid (5-acetylneuraminic acid) to terminal sugar residues (Mandrell et al., 1992). Approximately one half of H. influenzae strains include sialic acid in their LOS molecules (Hood et al., 1999). Sialic acid is commonly found in higher animals, but is rarely seen in microorganisms and is often associated with virulence (Schauer et al., 1995; Takahashi et al., 1999). The incorporation of sialic acid can be exogenous (i.e. an external source of sialic acid is required) in some bacteria (for example N. gonorrhoeae), but is endogenous in H. influenzae (i.e. sialic acid is synthesized within the cell), although when sialic acid is added to the growth media it can enhance sialylation (Hood et al., 1999). Mutant sialylation deficient strains of H. influenzae showed a increase in susceptibility to
Chapter 1: General Introduction

killing by human serum, but attach to, and invade epithelial cells normally (Hood et al., 1999). The addition of sialic acid to H. influenzae LOS is likely to be a major factor influencing the survival of the bacterium within the host (Hood et al., 1999).

1.8.5. The oligosaccharide moiety: genetics

Prior to the release of the complete genome sequence, very few LOS biosynthesis genes from H. influenzae had been identified and characterized. Those that had included isn, galE, gmhD and gmhE, the protein products of which are involved in the formation of heptose residues for incorporation into LOS (Lee et al., 1995a; Maskell et al., 1992; Nichols et al., 1997; Preston et al., 1996b). The availability of the complete genome sequence allowed the systematic searching of the genome with sequences from known LPS/LOS biosynthesis enzymes. Using this approach Hood et al. (1996) identified another 25 candidate LOS biosynthesis genes. These were cloned, insertionally inactivated and the LOS from the mutant strains analyzed using SDS-PAGE and electrospray mass spectroscopy (Hood et al., 1996). Using this technique, the role of the majority of these genes in LOS biosynthesis was confirmed (Hood et al., 1996). However, thorough structural and biochemical analyses of these LOS mutants is still required to define the role of particular gene products in biosynthesis.

Following the identification and mutation of these candidate LOS biosynthesis genes, the virulence of mutant strains was tested using the infant rat model of H. influenzae infection (Hood et al., 1996). These results suggested that deletion of LOS epitopes effects the intravascular dissemination of H. influenzae (Hood et al., 1996). Overall, this report by Hood et al. (1996), reinforced how understanding the genetics behind oligosaccharide biosynthesis can be used to enhance our understanding of the role of these molecules in virulence. The following sections will combine some of the data from the Hood et al. (1996) report with others, and will illustrate the major elements of the genetics behind core oligosaccharide biosynthesis in H. influenzae.
1.8.5.1. The \textit{opsX}/\textit{kdkA} cluster

The \textit{kdkA} gene which, as mentioned earlier, encodes the enzyme which adds a phosphoryl substitution to the 4-OH position of the Kdo residue, is found clustered with other LOS related genes (Hood et al., 1996; White et al., 1999). The \textit{kdkA} gene is part of an operon which also contains the \textit{orfM} and \textit{IgtC} genes (Hood et al., 1996). These genes are thought to be glycosyltransferases which add outer core sugars to the core oligosaccharide (Hood et al., 1996). Divergently transcribed from the \textit{kdtA} operon is the \textit{opsX} gene. Mutational analysis of this gene suggests that it encodes the heptosyltransferase I, which adds the first L-\textit{glycero-D-manno}-heptose residue to \textit{H. influenzae} LOS (Hood et al., 1996).

1.8.5.2. The \textit{l icl} operon

The \textit{licl} operon is responsible for the incorporation of choline, which is acquired from the extra-cellular environment, into LOS as phosphorylcholine (choP; Weiser et al., 1997). The operon is composed of four genes; \textit{licl}:\textit{licA} which has homology to choline kinases; \textit{licB}, thought to be choline permease; \textit{licC}, a pyrophosphorylase; and \textit{licD}, a diphospho-nucleoside choline transferase (Lysenko et al., 2000; Weiser et al., 1997; Weiser et al., 1990a). The phase variable loss or gain of choP is associated with the \textit{licl}:\textit{licA} gene (Weiser et al., 1989). This gene contains multiple tandem repeats of 5'-'CAAT-3' which creates a translational on-off switch by a mechanism involving slip-stranded mispairing (Weiser et al., 1989). Mutant strains containing an in-frame deletion of the 5'-'CAAT-3' region constitutively express choP and exhibit no phase variation of choP (Lysenko et al., 2000).

1.8.5.3. Other \textit{H. influenzae} loci containing 5'-'CAAT-3' repeats

In addition to \textit{licl}, two other chromosomal loci contain the repeat sequence 5'-'CAAT-3' and are involved with the biosynthesis of LOS. \textit{Lic2} is a putative galactosyl
transferase which transfers the terminal galactose to complete the generation of the phase variable Galα(1-4)βGal epitope (High et al., 1993; High et al., 1996). In frame expression of this gene is necessary but not sufficient for the synthesis of this structure (High et al., 1996). Expression also requires the lgtC gene. This appears to be responsible for the addition of the proximal galactose residue onto glucose, which generates Galβ(1-4)Glu (Hood et al., 1996; Weiser & Pan, 1998). The lgtC gene contains multiple tandem repeats of 5'-GACA-3', suggesting that expression of the proximal galactose is also phase variable (Hood et al., 1996). The third site where the repeat sequence 5'-CAAT-3' is found, is in the first gene (lic3A) of the lic3 operon. The function of this gene is still unclear, and no evidence exists to suggest a role in LOS biosynthesis (Masked et al., 1991). However, the second gene in the operon is an epimerase gene which provides galactose for incorporation into the LOS molecule (Masked et al., 1991).

1.8.5.4. Other oligosaccharide biosynthesis genes

Approximately one half of H. influenzae strains include sialic acid in their LOS molecules (Hood et al., 1999). Unfortunately, the genetic basis for this synthesis is still unclear because the genome strain, H. influenzae strain Rd, is one of those strains which does not incorporate sialic acid (Risberg et al., 1999). No sialic acid biosynthesis genes are found in this strain, with the exception of the CMP-sialic acid synthetase gene (siaA) - a homologue of a gene involved in the biosynthesis of the sialic acid containing capsule of Neisseria meningitidis (Swartley et al., 1996). Mutation of this gene in H. influenzae strains that do incorporate sialic acid results in mutants showing an increase in susceptibility to killing by human serum (Hood et al., 1999).

The analysis of the genome sequence of H. influenzae strain Rd has also identified other putative LOS biosynthesis genes. These include; waaF and orfH, which transfer the second and third L-glycero-D-manno-heptose residues to LOS, and several genes, such as lpsA and lgtA, thought to encode outer core glycosyl transferases (Hood et al., 1996). The study also identified genes involved in the supply of sugar precursors,
for example, *pgmB* and *galU*, and a cluster of eight genes with homology to those involved in O-chain biosynthesis (Hood *et al.*, 1996). This is surprising considering that *H. influenzae* does not produce O-chain repeats. The role of these genes is still not understood (Hood *et al.*, 1996).

1.9. *Campylobacter jejuni* LOS/LPS

*C. jejuni* strains were thought capable of producing both LOS and LPS. This conclusion was based on biochemical reports that suggested that some *C. jejuni* strains produced just the lipid A and core moieties, whilst others produced, in addition, a core linked O-chain - these reports will be discussed below. Therefore, the experiments described in this study were largely conceived, undertaken and written up on this basis. However very recently, reports have cast doubt on this hypothesis, and indicate that the O-chain-like repeat units produced by *C. jejuni* strains are core-independent capsular structures (Karlyshev *et al.*, 2000). The results described in Chapter 6 have a bearing on the question of O-chain and capsule, and this topic will be discussed in some depth in that chapter. However, for the purposes of this chapter and the results Chapters 3, 4 and 5, *C. jejuni* strains will be assumed to produce LOS or LPS.

1.9.1. The lipid A moiety

The lipid A moieties of most Gram negative species are generally highly conserved and non-variable (Rietschel *et al.*, 1994). However, it appears that the lipid A of *C. jejuni* is more changeable, with three different backbone disaccharide structures occurring (Moran, 1997; Moran *et al.*, 1991). In the *C. jejuni* HS:2 serostrain, approximately three quarters of the lipid A molecules have a disaccharide backbone of diaminoglucose (2,3-diamino-2,3-dideoxy-D-glucose) and glucosamine; 15% have a backbone where both sugars are diaminoglucose, and the remaining disaccharide molecules contain two glucosamine residues - the typical composition of enterobacterial lipid A molecules. All three types are phosphorylated and acylated in the
same manner (Moran, 1997; Moran et al., 1991). Analysis of other C. jejuni strains has indicated that they also contain these three types of disaccharide backbone, but the relative proportions of each type differ (Moran, 1993). Thus, minor inter-strain variation in C. jejuni lipid A may occur (Moran, 1997).

C. jejuni is responsible for an inflammatory type diarrhoea in immunologically naïve patients (Ketley, 1997). The lipid A moiety of LPS is a potent stimulator of the inflammatory response in humans (Raetz, 1990; Rietschel et al., 1994), and so C. jejuni lipid A could potentially be responsible for the pathology of Campylobacter enteritis. However, studies on C. jejuni LOS/LPS molecules have shown that they are less potent than their Salmonella counterparts, exhibiting 50% lower lethal toxicity in mice, 30- to 50-fold lower pyrogenicity in rabbits, and a 100-fold lower ability to induce TNFα from mouse peritoneal macrophages (Moran, 1995a). This may suggest that factors other than C. jejuni lipid A are the major stimulators of the host inflammatory response.

1.9.2. Core oligosaccharide

The core oligosaccharide structures from seven different HS serostrains have now been elucidated (Figure 1.9; Aspinall et al., 1993a; Aspinall et al., 1993b; Aspinall et al., 1994a; Aspinall et al., 1995). Like H. influenzae inner core structures they contain only one Kdo residue, but this is unphosphorylated in C. jejuni. The Kdo residue links to two L-glycero-D-manno-heptose sugars which are substituted by glucose, phosphate or phosphoethanolamine depending on the strain. The outer core is more varied containing galactose and N-acetylglucosamine in varying arrangements. These variations are thought to be responsible for the serotypic diversity that enables differentiation of C. jejuni strains that only produce LOS (Moran & Penner, 1999). The outer core residues can be substituted with sialic acid, or in the case of the HS:3 strain substituted with 3-amino-3,6-dideoxy-D-glucose. The presence of sialic acid is another similarity to the LOS structures of H. influenzae (Preston et al., 1996a). Sialic acid linked to galactose residues makes the terminal structures of some C. jejuni oligosaccharides similar to the terminal regions of human gangliosides.
Figure 1.9: Schematic diagram of core oligosaccharide structures produced by different HS serostrains of *C. jejuni*. ‘P’ is the abbreviation for phosphate, other chemical constituents are given above (based on data from Aspinall *et al.*, 1995; Aspinall *et al.*, 1994a; Aspinall *et al.*, 1993b; Aspinall *et al.*, 1993a).
(Moran et al., 1996b; Moran et al., 1996a). The core oligosaccharide structures produced by some strains of differing HS serotypes (e.g. HS:23 and HS:36) are identical (Figure 1.9; Aspinall et al., 1993a). These strains are presumed to be differentiated serologically on the basis of their O-chain polymers (Moran & Penner, 1999).

1.9.3. O-chain

Immunoblotting with homologous antisera has shown that 16 of 38 C. jejuni serostrains produce an O-chain-like repeating unit (Preston & Penner, 1987a; Preston & Penner, 1989). However, in contrast to the molecules produced by E. coli and Salmonella, the C. jejuni O-chain is not visible using silver stain detection (Preston & Penner, 1987a; Preston & Penner, 1989). Of the seven serostrains so far chemically characterized, only the HS:4, HS:19, HS:23 and HS:36 strains produce O-chain-like polymers (Aspinall et al., 1994c; Aspinall et al., 1992a; Aspinall et al., 1994b). The nature of the HS:4 polymer is as yet unknown, but the other three have been chemically determined (Aspinall et al., 1994b; Aspinall et al., 1992a; Aspinall et al., 1994c). The serotype HS:19 polymer consists of a disaccharide repeat of β-D-glucuronic acid amidated with 2-amino-2-deoxyglycerol linked to N-acetylglucosamine (Aspinall et al., 1994b; Aspinall et al., 1994c), and has been suggested as having a role in pathogenesis. It is curious that the HS:4 and HS:19 core oligosaccharide structures are identical and can mimic the same gangliosides (Aspinall et al., 1994a), yet HS:4 strains are rarely isolated from GBS patients. The major difference between the two serotypes, and thus likely to be responsible for the differences in pathogenesis is the O-chain polymer. The HS:19 O-chain resembles hyaluronic acid - a component of human connective tissue and a component of group A haemolytic streptococcal capsules, where it is thought to have a role as an anti-phagocytic factor (Moran & Penner, 1999). Thus, the O-chain polymer of HS:19 strains could be an additional example of molecular mimicry where the bacterium induces the production of antibodies which cross react with host structures (Moran & O'Malley, 1995).
The O-chain polymers of HS:23 and HS:36 are more complex than the HS:19 example. HS:36 produces an O-chain of trisaccharide repeating units consisting of N-acetylglucosamine, galactose and an additional heptose (Aspinall et al., 1992a). This additional heptose can be either 6-deoxy-D-altro-heptose, D-glycero-D-altro-heptose, 6-deoxy-D-methyl-D-altro-heptose or 3-methyl-D-glycero-D-altro-heptose (Aspinall et al., 1992a). This could suggest that HS:36 strains produce O-chain with heterogeneous units (Aspinall et al., 1992a), or produce four distinct O-polymers simultaneously (Moran & Penner, 1999).

The HS:23 O-chain polymer is identical to the HS:36 polymer except that only three of the four additional heptoses are present (Aspinall et al., 1992a). D-glycero-D-altro-heptose was not detected in the O-chain preparations from the HS:23 serostrain. The structural data from the core oligosaccharides and O-chain polymers does not explain how HS:23 and HS:36 strains can be serologically differentiated. The core oligosaccharides of the two are identical, and the O-chain polymers have common repeating units. Further investigations in to this area are required, and also to ascertain whether C. jejuni strains can control the synthesis of different O-chain polymers in differing host environments, perhaps in order to evade host immune responses.

1.9.4. LPS/LOS structures from isolates from patients with neurological diseases

In addition to chemically characterizing the LOS/LPS molecules produced by C. jejuni serostrains, recent attempts have focused on C. jejuni isolates from patients who developed GBS or Miller Fisher Syndrome. For example, two HS:19 isolates (OH 4382 and OH 4384) from patients who developed GBS have been chemically characterized (Aspinall et al., 1994b; Aspinall et al., 1994c; Aspinall et al., 1994a). These studies showed that both strains produced identical O-chain structures to the serostrain, but very different core structures (Aspinall et al., 1994b; Aspinall et al., 1994c; Aspinall et al., 1994a). The core oligosaccharide of C. jejuni OH 4384 has a disaccharide neuraminobiose present instead of the terminal sialic acid (Figure 1.10), whereas in C. jejuni OH 4382, the core lacks the terminal galactose and N-acetylgalactosamine, but
Chapter 1: General Introduction

Figure 1.10: Schematic diagram of core oligosaccharide structures produced by the C. jejuni HS:19 serostrain, and two HS:19 isolates from patients with GBS. ‘P’ is the abbreviation for phosphate, other chemical constituents are given above. Based on data from Aspinall et al. (1994a).
contains a neuraminobiose disaccharide linked to an inner galactose. Both GBS isolates therefore have a trisaccharide terminus of sialic acid - sialic acid - galactose, a structure which also occurs in the human ganglioside GD3 (Aspinall et al., 1994a). Furthermore, the outer core region of C. jejuni OH 4384 mimics the ganglioside GT1a (Aspinall et al., 1994a). The trisaccharide terminal structure of sialic acid - sialic acid - galactose is also seen in a HS:10 isolate from a patient with Miller Fisher syndrome (Nam Shin et al., 1998), and preliminary analysis of a HS:23 isolate, again from a Miller Fisher syndrome patient, also suggests the presence of this same trisaccharide (Aspinall et al., 1998). The presence of the trisaccharide repeat in neurological strains, but not in serostrains has been suggested as being important for the pathogenesis of C. jejuni-mediated neurological diseases (Salloway et al., 1996). However, structures have been reported from GBS associated C. jejuni strains where the trisaccharide is not present. These include a further two HS:19 strains (Moran & O'Malley, 1995), and a HS:41 strain (Prendergast et al., 1998).

1.10. Aims of this study

C. jejuni LOS/LPS molecules are important for the pathogenesis of the organism - they are endotoxic and have been suggested as potential adhesins. Structural work on C. jejuni LOS/LPS molecules have shown that structural variation occurs between strains of the same, and different HS serotypes. It also has shown the high level of similarity, if not identity that exists between the core oligosaccharides produced by some C. jejuni strains, and human ganglioside molecules. Current knowledge about the genetics and the biosynthesis of these biologically significant molecules is severely limited. The aim of this study was to enhance our understanding of these areas, by identifying and characterizing genes thought to encode key enzymes in LOS/LPS biosynthesis. Furthermore, by analyzing the genetic differences in LOS/LPS gene content, the basis of structural variation between strains of the same, and different, HS serotypes could be better understood.
2.1. Bacterial strains and growth conditions

The E. coli strain DH5α (supE44 lacU169 (80lacZ M15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1; Hanahan, 1983) was used as the host strain for cloning experiments in this study. The Salmonella typhimurium strains SL3789 (waaF511; Roantree et al., 1977) and SL3770 (Roantree et al., 1977) were used in the waaF complementation experiments detailed in Chapter 6. Both E. coli and S. typhimurium cells were grown at 37°C in Luria-Bertani (LB;) broth with shaking, or on LB media supplemented with 1% (w/v) agar (LA). Antibiotics were added when appropriate.

Campylobacter strains were grown on Campylobacter Blood Free Selective Agar Base (CSA; Oxoid, Unipath, UK), or in Mueller-Hinton broth (MHB; Oxoid) with shaking in a Variable Atmosphere Incubator (VAIN; Don Whitley Scientific Ltd., UK) containing 5% O2, 10% CO2, and 85% N2 at 37°C. A full list of all the Campylobacter strains utilized in this study can be found in Table 2.1. All strains were identified as belonging to the genus Campylobacter using PCR system described by Linton et al. (1996) and were further confirmed as being C. jejuni isolates using the PCR strategy of Gonzalez et al. (1997).

2.2. Storage of bacterial strains

Strains were grown overnight in LB or MHB broth containing antibiotics when appropriate. Cells were pelleted by centrifugation (3200 g for 10 min) and resuspended in 0.75 ml of LB or MHB also containing the relevant antibiotics and transferred to a 1.5 ml cryotube. 0.75 ml of sterile glycerol was added and the culture stored at -80°C.
Table 2.1: Sources and heat-stable (HS) serotypes of the C. jejuni strains used in this study. All strains were identified as belonging to the genus Campylobacter using PCR system described by Linton et al. (1996), and were further confirmed as being C. jejuni isolates using the PCR strategy of Gonzalez et al. (1997). Serotyping was kindly done by R. Thwaites, Central Public Health Laboratory, Colindale, London, UK. NK: not known. UT: untypable. Asterisks indicate strains isolated from patients with Guillian Barré Syndrome.
# Chapter 2: Materials and Methods

## C. jejuni strain

<table>
<thead>
<tr>
<th>C. jejuni strain</th>
<th>HS serotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCTC 11168</td>
<td>2</td>
<td>R. Owen, National Collections of Type Cultures and Pathogenic Fungi, Colindale, London, UK.</td>
</tr>
<tr>
<td>NCTC 11351</td>
<td>23</td>
<td>R. Owen, National Collections of Type Cultures and Pathogenic Fungi, Colindale, London, UK.</td>
</tr>
<tr>
<td>O:3</td>
<td>3</td>
<td>R. Owen, National Collections of Type Cultures and Pathogenic Fungi, Colindale, London, UK.</td>
</tr>
<tr>
<td>O:1</td>
<td>1</td>
<td>R. Owen, National Collections of Type Cultures and Pathogenic Fungi, Colindale, London, UK.</td>
</tr>
<tr>
<td>2561/90</td>
<td>2</td>
<td>D. Wareing, Public Health Laboratory, Royal Preston Hospital, Preston, UK.</td>
</tr>
<tr>
<td>2608/90</td>
<td>4</td>
<td>D. Wareing, Public Health Laboratory, Royal Preston Hospital, Preston, UK.</td>
</tr>
<tr>
<td>2523/90</td>
<td>2</td>
<td>D. Wareing, Public Health Laboratory, Royal Preston Hospital, Preston, UK.</td>
</tr>
<tr>
<td>306/90</td>
<td>1</td>
<td>D. Wareing, Public Health Laboratory, Royal Preston Hospital, Preston, UK.</td>
</tr>
<tr>
<td>2258/90</td>
<td>23</td>
<td>D. Wareing, Public Health Laboratory, Royal Preston Hospital, Preston, UK.</td>
</tr>
<tr>
<td>1915/91</td>
<td>2</td>
<td>D. Wareing, Public Health Laboratory, Royal Preston Hospital, Preston, UK.</td>
</tr>
<tr>
<td>27F 155</td>
<td>NK</td>
<td>A. Swann, Public Health Laboratory, Leicester Royal Infirmary, Leicester, UK.</td>
</tr>
<tr>
<td>18F 5</td>
<td>11</td>
<td>A. Swann, Public Health Laboratory, Leicester Royal Infirmary, Leicester, UK.</td>
</tr>
<tr>
<td>21F 242</td>
<td>NK</td>
<td>A. Swann, Public Health Laboratory, Leicester Royal Infirmary, Leicester, UK.</td>
</tr>
<tr>
<td>4F 182</td>
<td>NK</td>
<td>A. Swann, Public Health Laboratory, Leicester Royal Infirmary, Leicester, UK.</td>
</tr>
<tr>
<td>2F 68</td>
<td>50</td>
<td>A. Swann, Public Health Laboratory, Leicester Royal Infirmary, Leicester, UK.</td>
</tr>
<tr>
<td>2F 90</td>
<td>UT</td>
<td>A. Swann, Public Health Laboratory, Leicester Royal Infirmary, Leicester, UK.</td>
</tr>
<tr>
<td>2F 8</td>
<td>5</td>
<td>A. Swann, Public Health Laboratory, Leicester Royal Infirmary, Leicester, UK.</td>
</tr>
<tr>
<td>4F 225</td>
<td>NK</td>
<td>A. Swann, Public Health Laboratory, Leicester Royal Infirmary, Leicester, UK.</td>
</tr>
<tr>
<td>8F 169</td>
<td>5</td>
<td>A. Swann, Public Health Laboratory, Leicester Royal Infirmary, Leicester, UK.</td>
</tr>
<tr>
<td>21F 185</td>
<td>15</td>
<td>A. Swann, Public Health Laboratory, Leicester Royal Infirmary, Leicester, UK.</td>
</tr>
<tr>
<td>K85</td>
<td>2</td>
<td>Everest et al. (1992)</td>
</tr>
<tr>
<td>B404</td>
<td>NK</td>
<td>Everest et al. (1992)</td>
</tr>
<tr>
<td>N82</td>
<td>UT</td>
<td>Everest et al. (1992)</td>
</tr>
<tr>
<td>O73</td>
<td>11</td>
<td>Everest et al. (1992)</td>
</tr>
<tr>
<td>J75</td>
<td>28</td>
<td>Everest et al. (1992)</td>
</tr>
<tr>
<td>E206</td>
<td>UT</td>
<td>Everest et al. (1992)</td>
</tr>
<tr>
<td>NCTC 11828 (81116)</td>
<td>6</td>
<td>T. Wassenaar, University of Mainz, Germany.</td>
</tr>
<tr>
<td>81-176</td>
<td>23/36</td>
<td>Russell et al. (1989)</td>
</tr>
<tr>
<td>16971.94GSH*</td>
<td>41</td>
<td>A. Lastovica, Red Cross Children’s Hospital, Cape Town, S. Africa (Prendergast et al., 1998)</td>
</tr>
<tr>
<td>28134.94GSH*</td>
<td>41</td>
<td>A. Lastovica, Red Cross Children’s Hospital, Cape Town, S. Africa (Prendergast et al., 1998)</td>
</tr>
<tr>
<td>260.94RXH*</td>
<td>41</td>
<td>A. Lastovica, Red Cross Children’s Hospital, Cape Town, S. Africa (Prendergast et al., 1998)</td>
</tr>
<tr>
<td>176.83</td>
<td>41</td>
<td>A. Lastovica, Red Cross Children’s Hospital, Cape Town, S. Africa (Prendergast et al., 1998)</td>
</tr>
<tr>
<td>G1*</td>
<td>1</td>
<td>N. Gregson, Guy’s Medical School, London, UK.</td>
</tr>
<tr>
<td>G2*</td>
<td>UT</td>
<td>N. Gregson, Guy’s Medical School, London, UK.</td>
</tr>
<tr>
<td>G3*</td>
<td>1</td>
<td>N. Gregson, Guy’s Medical School, London, UK.</td>
</tr>
<tr>
<td>G4*</td>
<td>UT</td>
<td>N. Gregson, Guy’s Medical School, London, UK.</td>
</tr>
<tr>
<td>F3*</td>
<td>UT</td>
<td>N. Gregson, Guy’s Medical School, London, UK.</td>
</tr>
<tr>
<td>P19</td>
<td>19</td>
<td>N. Gregson, Guy’s Medical School, London, UK.</td>
</tr>
</tbody>
</table>
2.3. Media

LB (Roth, 1970) was prepared by dissolving 10 g of bacto-tryptone (Oxoid), 5 g of bacto-yeast (Oxoid) and 10 g NaCl in 950 ml distilled water. The pH of the media was then adjusted to pH 7.0 with 10 M NaOH and the volume increased to 1 litre with water. The broth was sterilized by autoclaving at 121°C, 15 lb. inch\(^{-2}\) for 15 min. LA was made in an identical fashion except for the addition of 1% (w/v) agar (Oxoid) prior to sterilization. Following cooling to approximately 40°C, any necessary antibiotics were added and the agar transferred to petri dishes (Sterilin, UK).

MHB was prepared by dissolving 21 g media (Oxoid) in 1 litre distilled water. The broth was sterilized by autoclaving at 121°C, 15 lb. inch\(^{-2}\) for 15 min.

CSA was prepared by dissolving 45.4 g media (Oxoid) in 1 litre of distilled water. The media was sterilized by autoclaving at 121°C, 15 lb. inch\(^{-2}\) for 15 min. Following cooling to approximately 40°C, any necessary antibiotics were added and the CSA transferred to petri dishes.

SOC medium was prepared by dissolving 20 g of tryptone, 5 g of yeast extract (Oxoid), and by adding 0.01 M NaCl, 0.005 M KCl, 0.01 M MgCl\(_2\) and 0.01 M MgSO\(_4\) to a final volume of 1 litre water. The broth was sterilized by autoclaving at 121°C, 15 lb. inch\(^{-2}\) for 15 min, and stored at 4°C until required. Immediately prior to use, 0.02 M filter sterilized glucose (0.22 μM pore, Gelman Sciences, MI, USA) was added.

2.4. Antibiotic supplements

All antibiotics were purchased from Sigma. Antibiotic stock solutions prepared with distilled water were filter sterilized (0.22 μM pore, Gelman Sciences) before use. All stock solutions were stored at 4°C in the dark until used. In experiments involving DH5α strains harboring the plasmid pUC19 (Yanisch-Perron et al., 1985) chloro-3-indoyl-β-D-galactoside (X-Gal; 20 μg ml\(^{-1}\) media concentration) was also added to the medium. This addition facilitated the effective identification of recombinant colonies due to the presence of the lac\(_Z\) gene on the pUC19 plasmid.
Table 2.2: Antibiotic supplements.

<table>
<thead>
<tr>
<th>Antibiotic</th>
<th>Solvent</th>
<th>Stock concentration (mg ml(^{-1}))</th>
<th>Media concentration (μg ml(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ampicillin</td>
<td>water</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Chloramphenicol</td>
<td>50% ethanol, 50% water</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Kanamycin</td>
<td>water</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Novobiocin</td>
<td>water</td>
<td>25</td>
<td>5</td>
</tr>
</tbody>
</table>

2.5. Miscellaneous buffers and solutions

All solutions were made using analytical grade reagents supplied by Fisons Scientific Equipment, Loughborough unless otherwise stated.

**CTAB/NaCl solution:** 0.7 M NaCl was dissolved in distilled water, followed by the addition of 10% (w/v) CTAB (hexadecyltrimethyl ammonium bromide; Sigma). After heating the mixture to 60°C to dissolve the CTAB, the volume was adjusted with water. The solution was not autoclaved.

**0.5 M EDTA:** 0.5 M disodium diaminoethanetetra-acetic acid was added to 950 ml of distilled water. The pH was then adjusted to pH 8.0 using 10 M NaOH to allow the EDTA to dissolve. The solution was then made up to 1 litre using distilled water and sterilized by autoclaving at 121°C, 15 lb. inch\(^{-2}\) for 15 min.

**GES solution:** 6 g guanidium thiocyanate, 2 ml 0.5 M EDTA (pH 8.0) and 2 ml distilled water were heated at 65°C with mixing until dissolved. After cooling 500 μl 10% (w/v) sarkosyl was added, the solution made up to 10 ml with distilled water and then filter sterilized (0.45 μM pore, Gelman Sciences). The solution was stored at room temperature.
Chapter 2: Materials and Methods

**Phenol/chloroform:** 200 ml of liquefied phenol and an identical volume of chloroform were mixed. 0.4 g 8-hydroxyquinoline (Sigma) was added as an antioxidant. The pH of the solution was adjusted to pH 7.5 by extracting twice with 150 ml of 1 M Tris pH 7.5 followed by one extraction with 0.1 M Tris pH 7.5 and finally one extraction with 0.01 M Tris pH 7.5. The phenol/chloroform was stored at 4°C under 0.01 M Tris pH 7.5.

**1 x Phosphate buffered saline (PBS):** 0.137 M NaCl, 0.027 M KCl, 0.0015 M KH₂PO₄ and 0.008 M Na₂HPO₄ was added to a final volume of 1 litre. After adjusting the pH of the solution to pH 7.3 the PBS was sterilized by autoclaving at 121°C, 15 lb. inch⁻² for 15 min.

**10% (w/v) SDS:** 10 g of sodium lauryl sulphate was added to 90 ml distilled water. The solution was stored at room temperature.

**3 M NaAc (pH 5.2):** 0.3 M sodium acetate was dissolved in 90 ml distilled water. The pH of the solution was adjusted to pH 5.2 using glacial acetic acid and the volume corrected to 100 ml. The solution was then sterilized by autoclaving at 121°C, 15 lb. inch⁻² for 15 min.

**TAE buffer:** 0.04 M Tris-[hydroxymethyl]-aminomethane (Tris) and 0.001 M EDTA were dissolved in 950 ml distilled water. The pH of the solution was corrected to pH 7.8 with glacial acetic acid and the volume corrected to 1 litre.

**TE buffer:** 0.01 M Tris-HCl and 0.001 M EDTA (pH 8.0) were dissolved in 900 ml of distilled water. The volume was then adjusted to 1 litre and the buffer sterilized by autoclaving at 121°C, 15 lb. inch⁻² for 15 min.

**X-Gal:** 5-bromo-4-chloro-3-indoyl-β-D-galactoside was prepared by dissolving 20 mg ml⁻¹ in dimethylformamide. This stock solution was stored at -20°C in the dark.
2.6. Extraction of DNA from bacterial cells

2.6.1. Preparation of plasmid DNA from *E. coli* and *S. typhimurium* cells

The methodology used to isolate plasmid DNA varied depending on the quantity and purity of DNA required. For the initial analysis of potential recombinant clones a modification of the alkaline extraction method of Birnboim & Doly (1979) was used. This generally yields 3 - 5 μg of DNA of sufficient quality for restriction digestion. Briefly, a 1.5 ml broth culture was centrifuged (3200 g for 10 min) and the pellet resuspended in 100 μl of solution P1 (100 μg ml⁻¹ RNase A, 50 mM Tris/HCl, 10 mM EDTA). Following the addition of 200 μl solution P2 (200 mM NaOH, 1% SDS) the tube was mixed by inversion and left at room temperature for 5 min. 150 μl of pre-chilled solution P3 (3 M KOAc, pH 4.8) was then added, the tube mixed by inversion and returned to ice for a further 10 min. The preparation was centrifuged at 11600 g for 20 min and the supernatant removed to a fresh micro-centrifuge tube. One volume of phenol/chloroform was then added and after mixing thoroughly, the mixture was centrifuged (11600 g, 10 min). The upper aqueous phase was removed and 1/10th volume 3 M NaAc (pH 5.2) and 2 - 3 volumes of 100% ethanol added. The tube was then returned to ice for 30 min before being centrifuged (11600 g, 20 min). Following removal of the supernatant, the pellet was washed in 1 ml 70% ethanol and dried at room temperature before being resuspended in sterile distilled water.

Qiagen-tip 20 columns (Qiagen, UK) were used for larger yields (approximately 20 μg) of DNA which was of sufficient quality for sequencing and Southern hybridization. The protocol was carried out according to the manufacturers instructions. Briefly, 3 ml of overnight broth culture was centrifuged (3200 g, 10 min) and resuspended in 300 μl solution P1. Following the addition of 300 μl solution P2, the tube was incubated at room temperature for 5 min. 300 μl of pre-chilled solution P3 was added and after mixing the tube placed on ice for 5 min. The cell debris was then pelleted by centrifugation (11600 g, 10 min) and the clear supernatant transferred to a Qiagen tip-20 column, pre-equilibrated with 1 ml buffer QBT (supplied by Qiagen; 0.75
Chapter 2: Materials and Methods

M NaCl, 50 mM MOPS pH 7.0, 15% isopropanol, 0.15% Triton X-100). The column was washed with 4 x 1 ml of buffer QC (supplied by Qiagen; 1 M NaCl, 50 mM MOPS pH 7.0, 15% isopropanol), before elution of the DNA using 800 µl of buffer QF (supplied by Qiagen; 1.25 M NaCl, 50 mM Tris-HCl pH 8.5, 15% isopropanol). The eluted DNA was precipitated using 0.7 volumes of isopropanol and centrifuged (11600 g, 30 min) before being washed in 1 ml 70% ethanol. After drying the pellet at room temperature for 5 min, the purified DNA was resuspended in sterile distilled water.

For large yields of high quality plasmid (up to 100 µg) Qiagen-tip 100 columns were used. Briefly, 50 ml of overnight culture was centrifuged (3020 g, 20 min) and the cells resuspended in 4 ml of solution P1. After the addition of 4 ml solution P2, the tube was mixed by inversion and incubated at room temperature for 5 min. 4 ml of solution P3 was then added and the tube mixed, before being placed on ice for 15 min. Following centrifugation (3020 g, 30 min) the clear supernatant was transferred to a Qiagen-tip 100 column (pre-equilibrated with 4 ml buffer QBT). The column was then washed with 2 x 10 ml of buffer QC and the DNA eluted with 5 ml buffer QF. The plasmid DNA was precipitated using 0.7 volumes of isopropanol, pelleted by centrifugation (11600 g, 30 min) and washed with 2 ml 70% ethanol. The purified plasmid DNA was then air-dried at room temperature for 10 min and finally resuspended in sterile, distilled water.

2.6.2. Preparation of *Campylobacter* chromosomal DNA

Chromosomal DNA was obtained from *C. jejuni* strains using one of two methods. A modification of the guanidium thiocyanate method (Pitcher et al., 1989) was used for those *Campylobacter* strains of Lior biotype 2 (DNase positive; Lior, 1984). Cells were harvested from an overnight plate and resuspended in 125 µl TE. After the addition of 500 µl GES solution, the tube was mixed by inversion and incubated at room temperature for 10 min. 250 µl pre-chilled 2.5 M ammonium acetate was then added. Following incubation on ice for 15 min, 1 volume of phenol/chloroform was added. This was mixed by inversion and then centrifuged (11600 g, 30 min). The aqueous phase was gently removed to a fresh micro-centrifuge tube and the chromosomal DNA
precipitated using 0.54 volumes of isopropanol. The precipitated DNA was pelleted by centrifugation (11600 g, 5 min), washed with 1 ml 70% ethanol before being air-dried for 10 min. Purified chromosomal DNA was resuspended in sterile, distilled water and stored at -20°C.

Extraction of chromosomal DNA from DNase negative strains of Campylobacter was achieved using a method based on that of Murray & Thompson (1980). Cells were harvested from three overnight plates and resuspended in 9 ml MHB. Following centrifugation (3020 g, 20 min) the cells were resuspended in 9.5 ml 1 x PBS. 2 ml 10% SDS and 100 µl 20 mg ml⁻¹ proteinase K were then added. This was incubated at 37°C for 1 hr before 1.8 ml 5 M NaCl and 1.5 ml CTAB/NaCl were added. The tubes were mixed gently by inversion, and then incubated at 65°C for 20 min. An equal volume of phenol/chloroform was added and the tubes centrifuged (3020 g, 20 min). The upper aqueous phase was then removed to a fresh tube and the DNA precipitated using 0.7 volumes of isopropanol. The precipitate was hooked out using a Pasteur pipette into 1 ml 70% ethanol. After pelleting (3020 g, 10 min) the purified chromosomal DNA was air-dried for 5 min, resuspended in distilled water and stored at -20°C.

2.7. Quantification of DNA using spectrophotometry

DNA concentration was determined by adding 10 µl of sample to 990 µl sterile distilled water. This was mixed before being transferred to a 1 ml quartz cuvette. The absorbence of the sample was then measured at 260 and 280 nm using a Pharmacia Ultraspec III spectrophotometer. The concentration of DNA could be determined using the assumption that one A₂₆₀ unit equals 50 µg ml⁻¹ of double stranded DNA or 40 µg ml⁻¹ single stranded DNA. The ratio of A₂₆₀/A₂₈₀ was also determined to assess the purity of the DNA. A ratio of approximately 1.8 was considered to indicate high purity.
2.8. Electrophoresis of DNA

DNA fragments were separated using horizontal gel electrophoresis. The gel was prepared by dissolving 0.5 - 1.5% (w/v) agarose (SeaKem, FMC Bioproducts, UK) in TAE buffer. Following heating of the mixture to ensure the complete solubilisation of the agarose, ethidium bromide was added to a final concentration of 0.5 μg ml⁻¹ and the gel poured into a casting tray. A well comb was then inserted into the still molten agarose and the gel left to set. Samples were prepared by adding 2 μl loading buffer (5 x TAE buffer, 0.05% orange G (Sigma) and 15% v/v glycerol) to each 10 μl of sample. After solidification, the well comb was removed, the gel placed in a running tank and submerged in 1 x TAE. The DNA samples were then loaded alongside molecular markers (λ DNA restricted with HindIII (Gibco-BRL; 23130 bp, 9416 bp, 6557 bp, 4361 bp, 2322 bp, 2027 bp, 564 bp, and 125 bp) and ΦX174 DNA restricted with HaeIII (Gibco-BRL; 1353 bp, 1078 bp, 872 bp, 603 bp, 310 bp, 281 bp, 271 bp, 234 bp, 174 bp, 118 bp and 72 bp). The gel was run at a constant voltage of 11 V/cm until the orange dye front had travelled through approximately 75% of the gel. The DNA fragments were then visualized using ultra-violet light (290 nm) and photographed.

2.9. Purification of DNA

Several techniques for the purification of DNA were utilized during this study. The conventional phenol/chloroform extraction followed by ethanol precipitation was used for the purification of PCR products and removal of proteins and excess salt from DNA solutions, but DNA was extracted from agarose gels using the NUCLEOTRAP extraction kit (Macherey-Nagel, Germany).

2.9.1. Extraction of DNA from agarose gels using the NUCLEOTRAP extraction kit

All buffers used in this procedure were provided with the extraction kit but details of their composition were not given. The agarose containing the DNA to be
purified was excised using a clean scalpel blade and weighed. 300 µl of solubilisation buffer NT1 and 10 µl NUCLEOTRAP beads were added to each 100 ng of agarose. After mixing thoroughly, the sample was incubated at 50°C for 15 min, during which time the sample was regularly mixed to ensure the NUCLEOTRAP beads remained in suspension. Following incubation, the sample was centrifuged (11600 g, 30 s) and the DNA containing pellet resuspended in 500 µl buffer NT2. After mixing, the sample was re-centrifuged (11600 g, 30 s) and the pellet resuspended in 500 µl buffer NT3. This wash step was repeated before the pellet was air-dried for 15 min. The beads were then resuspended in 50 µl TE (pH 8.3) and left at room temperature for 10 min to allow elution of the DNA. The sample was centrifuged (11600 g, 30 s) and the DNA containing solution removed to a fresh tube.

2.9.2. Phenol/chloroform extraction and ethanol precipitation

The volume of sample was corrected to at least 100 µl and an equal volume of phenol/chloroform added. This was mixed thoroughly before being centrifuged (11600 g, 30 min). The aqueous upper phase was then removed to a fresh tube and 1/10th volume 3 M sodium acetate (pH 5.2) and 2 volumes of 100% ethanol added. 1 µl tRNA (10 mg ml⁻¹; type A from *E. coli*, Sigma) was also added to improve precipitation if the DNA concentration was expected to be low. The sample was then incubated at -20°C for 20 min, the precipitated DNA pelleted by centrifugation (11600 g, 15 min) and washed with 1 ml 70% ethanol. The purified DNA was then air-dried for 10 min before being resuspended in a suitable volume of distilled water.

2.10. Enzymatic modification of plasmid or chromosomal DNA

2.10.1. Restriction endonuclease digestion of DNA

All restriction enzymes and their corresponding reaction buffers were obtained from Gibco-BRL. All restriction digestion was carried out in accordance with the
Chapter 2: Materials and Methods

manufacturers instructions. Typically 0.1 - 1 μg of DNA was restricted in 1 x reaction buffer at 37°C for 1 hr, but DNA needed for subsequent cloning or Southern hybridization experiments was incubated at 37°C overnight to ensure complete digestion. Restriction was usually terminated by the addition of 1/5th volume of loading buffer, allowing the subsequent visualization of restriction fragments by gel electrophoresis. When further manipulation of the restricted products was needed, DNA was extracted with phenol/chloroform before being ethanol precipitated (section 2.9.2) and resuspended in a suitable volume of sterile water.

2.10.2. Modification of a 5' overhang to a blunt end terminus

The 3' - 5' exodeoxyribonuclease activity of the large fragment of *E. coli* DNA polymerase I (Klenow fragment) was used to remove the 5' overhangs created by the digestion of DNA with certain restriction enzymes (e.g. *BamH*I). Following restriction digestion (section 2.10.1) the DNA was extracted with phenol/chloroform before being ethanol precipitated (section 2.9.2). The purified DNA was then resuspended in a suitable volume of distilled water containing a final concentration of 40 μM of dNTP’s and 0.1 mg ml⁻¹ of BSA (bovine serum albumin). One unit of Klenow polymerase (Pharmacia; 6435 units ml⁻¹) was then added per microgram of DNA. The reaction was incubated at room temperature for 10 min. Termination of the reaction was achieved by heating at 75°C for 10 min. When further manipulation of the products was needed, DNA was extracted with phenol/chloroform before being ethanol precipitated (section 2.9.2) and resuspended in a suitable volume of sterile water.

2.10.3. Dephosphorylation of restriction digested DNA

In order to prevent the re-ligation of linearized vector DNA with compatible cohesive ends, the terminal 5' phosphate group was removed using calf intestinal phosphatase (CIP; Pharmacia). Typically, 0.1 - 1 μg DNA was dephosphorylated in a 20 μl volume containing 2 μl 10 x CIP buffer (0.01 M ZnCl₂, 0.01 M MgCl₂, 0.1 M Tris-
Chapter 2: Materials and Methods

HC1; pH 8.3) and 0.5 µl of CIP (1000 units ml⁻¹) made up to 20 µl with sterile water. The reaction was incubated at 37°C for 1 hr. Termination of the reaction was achieved by adding 80 µl sterile water and extracting the DNA with phenol/chloroform. Dephosphorylated DNA was then ethanol precipitated and resuspended in a suitable volume of sterile water ready for subsequent manipulation.

2.10.4. DNA ligation

Ligation reactions were typically carried out in a 10 µl volume containing purified insert and vector DNA in a 2:1 molar ratio, 2 µl 5 x ligation buffer (0.25 M Tris-HCl pH 7.6, 50 mM MgCl₂, 5 mM ATP, 5 mM DTT and 25% PEG-8000; Gibco-BRL), 1 µl 10 mM ATP and 0.3 µl T4 ligase (5000 units ml⁻¹; Gibco-BRL) made up to volume with distilled water. Ligation reactions were incubated at 16°C overnight. Following ligation, DNA was ethanol precipitated in the presence of tRNA (section 2.9.2) and resuspended in 10 µl sterile distilled water in preparation for transformation into the E. coli host strain DH5α (section 2.11.2).

2.11. Electrotransformation of E. coli and S. typhimurium cells

The method described is based on that of Dower et al. (1988).

2.11.1. Preparation of electrocompetent cells

In order to prepare electrocompetent cells, 100 µl of an overnight culture was used to inoculate a 50 ml culture of pre-warmed LB broth. This was incubated at 37 °C with vigorous shaking until the culture had grown to a OD₆₀₀ of 0.6. The culture was then placed on ice for 10 min before the cells were pelleted by centrifugation at 4°C (3020 g, 20 min). The supernatant was discarded and the cells resuspended in 50 ml ice-cold sterile water before the centrifugation step was repeated. The pellet was then resuspended in 1 ml ice-cold sterile water and transferred to a 1.5 ml micro-centrifuge
Chapter 2: Materials and Methods

tube. The cells were then re-centrifuged (11600 g, 3 min) and after discarding the supernatant resuspended in 1 ml ice-cold sterile water. This washing procedure was repeated another 2 - 3 times before the cells were finally resuspended in 400 μl cold sterile water and stored on ice until needed.

2.11.2. Electroporation of competent cells

The precipitated ligation mixture (section 2.9.2) was added to a ice-cold electroporation cuvette (0.2 cm inter-electrode distance; Bio-Rad, UK) along with 40 μl of competent cells. This was then subjected to electroporation in a Bio-Rad Gene Pulser at a voltage of 1.5 kV, a resistance of 1000 Ω and a capacitance of 25 μFD. The contents of the tube were then resuspended in 1 ml of SOC and the cells removed to a 1.5 ml micro-centrifuge tube. This was incubated at 37°C for at least one hour to allow recovery of the cells and expression of the resistance gene on the transformed plasmid. Following recovery, transformants were plated out onto LA plates containing any appropriate antibiotics and incubated overnight at 37°C.

2.12. Electrotransformation of Campylobacter jejuni strains

This was performed using the method of Wassenaar et al. (1993a).

2.12.1. Preparation of electrocompetent C. jejuni cells

The overnight growth from four CSA plates was harvested into 10 ml of MHB and centrifuged (3020 g, 20 min). The campylobacters were then resuspended in 20 ml of ice-cold W&E buffer (272 mM sucrose, 15% glycerol) before being re-centrifuged (3020 g, 20 min). This washing procedure was repeated three times before the bacterial cells were finally resuspended in 1 ml W&E buffer and left on ice until required.
2.12.2. Electroporation of competent *C. jejuni* cells

Purified DNA (1 - 5 µg) was added to an electroporation cuvette (0.2 cm inter-electrode distance; Bio-Rad, UK) along with 40 µl of competent *Campylobacter* cells and placed on ice for 10 min. This was then subjected to electroporation in a Bio-Rad Gene Pulser at a voltage of 2.5 kV, a resistance of 200 Ω and a capacitance of 25 µFD. The contents of the tube were then resuspended in 0.1 ml of SOC and spread out on to CSA plates (containing no antibiotics). These plates were incubated overnight at 37°C in microaerophilic conditions to allow the recovery of transformants. The following day, overnight growth was removed and resuspended in 0.1 ml MHB. Aliquots were plated out onto CSA plates containing any appropriate selective antibiotics and incubated for up to 5 days in microaerophilic conditions at 37°C.

2.13. Natural transformation of *C. jejuni* cells

Purified chromosomal DNA (1 - 5 µg) was spotted onto a lawn of *C. jejuni* grown overnight on MH agar in microaerophilic conditions at 37°C. The cells were then incubated for a further 5 hr under these same conditions, before the recipients were harvested into 200 µl of MHB. Aliquots were then plated out onto MH plates containing any appropriate selective antibiotics and incubated for up to 5 days in microaerophilic conditions at 37°C.

2.14. Amplification of DNA using the Polymerase Chain Reaction

The polymerase chain reaction has been used for a variety of purposes during this work. All primers used in this study can be found in Appendix 1. The normal conditions for amplification are given below in section 2.14.1. However in certain instances, for example cloning of a PCR fragment or its subsequent mutagenesis by inverse PCR mutagenesis, extremely high replication fidelity was required. In these circumstances, the Expand™ High Fidelity PCR system (Boehringer Mannheim) was
used. This kit contains a mixture of DNA polymerases, some of which have a 3' - 5' exonuclease proofreading activity resulting in a high fidelity of DNA synthesis. This High Fidelity PCR system was used on DNA fragments of < 7 kb, but when amplifying fragments of > 7 kb in length for cloning or mutagenesis the Expand™ Long Template PCR system (Boehringer Mannheim) was used. This kit also contains a mixture of DNA polymerases, some of which have a 3' - 5' exonuclease proofreading activity resulting in a high fidelity of DNA synthesis, but the mixture is optimized to give a high yield of PCR products up to 40 kb in length.

2.14.1. Amplification of DNA using plasmid or chromosomal template

PCR reactions were carried out in a final volume of 10 μl, and consisted of 1 μl 10 x PCR buffer (100 mM Tris-HCl, 500 mM KCl, 15 mM MgCl₂), template DNA (1 - 10 ng plasmid; 10 - 50 ng chromosomal), 2 pmol of each primer, 1 μl 2 mM dNTP's, 0.1 μl Taq polymerase (Advanced Biotechnologies, UK) made up to volume with sterile water. Reactions were then overlaid with 50 μl mineral oil (Sigma), and subjected to amplification in an automatic thermal cycle machine (Omnigene) along with appropriate positive and negative controls.

PCR amplification typically involved initial denaturation conditions of 95°C for 3 min, followed by a denaturation step of 95°C for 30 s, primer annealing at 55°C for 30 s and extension at 72°C for 1 - 7 min for 30 cycles. The extension time was adjusted depending on the size of product expected. A final cycle consisting of annealing at 55°C for 1 min and extension at 72°C for 10 min was then carried out to ensure the completion extension of DNA products. Amplification products were visualized by agarose gel electrophoresis.

2.14.2. Amplification of DNA using the Expand™ High Fidelity PCR system

This kit was used according to the manufacturers instructions. Briefly, two master mixes of 75 μl were set up. The first contained template DNA (15 - 150 ng
plasmid; 150 - 750 ng chromosomal), 15 µl of each primer (2 pm µl⁻¹), 15 µl 2 mM dNTPs made up to volume with sterile water. The second contained 15 µl 10 x buffer (supplied by Boehringer Mannheim), 2 µl Expand™ DNA polymerase mix and 58 µl sterile water. The preparation of two master mixes avoids the partial degradation of primer and template through the 3' - 5' exonuclease activity of the Expand™ Taq mix which may occur in the absence of dNTPs. The two master mixes were then gently combined and before 10 µl aliquots were dispensed into 0.5 ml micro-centrifuge tubes. Reactions were then overlaid with 50 µl mineral oil (Sigma) and subjected to amplification in an automatic thermal DNA cycle machine (Omnigene).

Amplification using the Expand™ High Fidelity PCR system typically involved initial denaturation conditions of 94°C for 2 min, followed by a denaturation step of 94°C for 15 s, primer annealing at 55°C for 30 s and extension at 68°C for 1 - 7 min for 30 cycles. The extension time was adjusted depending on the size of product expected. A final cycle consisting of annealing at 55°C for 1 min and extension at 68°C for 7 min was then carried out to ensure the completion of DNA products. Following amplification, products were pooled and the DNA phenol /chloroform extracted and ethanol precipitated (section 2.9.2) in preparation for further manipulation.

2.14.3. Amplification of DNA using the Expand™ Long Template PCR system

This kit was used according to the manufacturers instructions. Briefly, two master mixes of 75 µl were set up. The first contained template DNA (15 - 150 ng plasmid; 150 - 750 ng chromosomal), 15 µl of each primer (2 pm µl⁻¹), 15 µl 2 mM dNTPs made up to volume with sterile water. The second contained 15 µl 10 x buffer No.1 (containing 17.5 mM MgCl₂; supplied by Boehringer Mannheim), 2 µl Expand™ Long Template DNA polymerase mix and 58 µl sterile water. The two master mixes were then gently mixed and combined before 10 µl aliquots were dispensed into 0.5 ml micro-centrifuge tubes. Reactions were then overlaid with 50 µl mineral oil (Sigma) and subjected to amplification in an automatic thermal DNA cycle machine (Omnigene). Amplification using the Expand™ Long Template PCR system typically involved...
initial denaturation conditions of 94°C for 2 min, followed by a denaturation step of 94°C for 10 s, primer annealing at 60°C for 30 s and extension at 68°C for 10-18 min for 30 cycles. The extension time was adjusted depending on the size of product expected. A final cycle consisting of annealing at 60°C for 1 min and extension at 68°C for 18 min was then carried out to ensure the completion of DNA products. Following amplification, products were pooled and the DNA phenol/chloroform extracted and ethanol precipitated (section 2.9.2) in preparation for further manipulation.

2.15. DNA sequencing

The DNA to be sequenced was prepared using the BigDye Terminator Cycle Sequencing ready reaction kit (ABI, Applied Biosystems). Briefly, 1 µg of DNA was mixed with 3.2 pmol primer, 8 µl BigDye Terminator Ready Reaction Mix (ABI; composition not available) made up to 20 µl with distilled water. The reaction was overlaid with 50 µl mineral oil and subjected to amplification under the following conditions: 26 cycles of denaturing at 96°C for 30 s, annealing at 50°C for 15 s and extension at 60°C for 4 min. Following amplification, the products were removed to a fresh 0.5 ml micro-centrifuge tube and precipitated by the addition of 2 µl 3 M NaAc (pH 5.2) and 40 µl ethanol. This was incubated on ice for 10 min before being centrifuged (11600 g, 20 min) and the pellet washed with 1 ml 70% ethanol. The extension products were then air-dried for 10 min. Alternatively, the amplified products were purified using CENTRI-SEP columns (Princeton Separations, NJ, USA). These were used as per manufacturers instructions. Briefly, 0.8 ml sterile distilled water was added to each column, these were then left at room temperature for 30 min to allow hydration of the column contents. The column was then placed into a 1.5 ml micro-centrifuge tube and centrifuged (750 g for 2 min) to allow excess fluid to drain. The amplification products were then dispensed onto the top of the gel bed in the column before the column was placed into a fresh 1.5 ml micro-centrifuge tube and re-centrifuged at 750 g for 2 minutes. The purified extension products that had collected at the bottom of the micro-centrifuge tube were then dried in a vacuum centrifuge for 30
min. Following purification the extension products were sent to the Protein and Nucleic Acid Chemistry Laboratory at the University of Leicester for analysis using an ABI 377 DNA sequencer (Perkin-Elmer).

2.16. DNA and protein sequence analysis

Sequence analysis was carried out using the computer program Gene Runner (version 3.02; available from Hastings Software, Inc.), whilst the schematic diagrams of plasmids presented in this study were constructed using the Clone Manager 5 program (version 5.02; available from Scientific and Educational Software, Inc.). The following were some of most useful Web sites visited during the course of this study:


http://alpha2.bmc.uu.se/sbnet/seq_anal.html (Protein Sequence Analysis Launcher).


http://microbios1.mds.qmw.ac.uk/campylobacter/ (C. jejuni genome web-site).


2.17. Analysis of plasmid or chromosomal DNA by Southern hybridization

Methods based on those described by Southern (1975) were used for the detection of specific DNA sequences separated by agarose gel electrophoresis. These experiments were carried out using the non-radioactive Gene Images random prime
labelling and detection kits supplied by Amersham. This protocol is based on the
detection of fluorescein-dUTP (Fl-dUTP) labelled DNA using an anti-fluorescein-
alkaline phosphatase conjugate (anti-Fl-AP).

2.17.1. Preparation of a Fl-dUTP labelled probe

Prior to labelling, 50 ng of template DNA was denatured by heating for 5 min in
a boiling water bath. This was then briefly cooled before being mixed with 5 µl of
primer, 10 µl of dNTP labelling mix and 1 µl of Klenow (5 units µl⁻¹) made up to a
volume of 50 µl with distilled water. This reaction was incubated at 37°C for 1 hr before
the reaction was terminated by the addition of EDTA to a final concentration of 20 mM.
The labelled probe was denatured by boiling for 5 min before being used.

2.17.2. Transfer of DNA to a nylon membrane for Southern hybridization analysis

Approximately 5 - 10 µg of chromosomal or 200 ng of high quality plasmid
DNA (the preparation of which is described in section 2.6) was digested overnight with
the appropriate enzyme(s) before being subjected to agarose gel electrophoresis. The gel
was then photographed before being washed twice for 10 min in depurinating solution
(0.2 M HCl), twice for 10 min in denaturing solution (1.5 M NaCl, 1.5 M NaOH) and
twice for 10 min in neutralizing solution (1 M Tris pH 7.4, 1.5 M NaCl). Following a
final wash for 2 min in 20 x SSC (3 M NaCl, 0.3 M Na-citrate pH 7.0) the DNA was
transferred to Hybond-N membranes (Amersham, UK) by capillary transfer for a
minimum of 4 hr. The DNA was cross-linked to the membrane by exposure to UV
radiation for 30 s at an intensity of 70 mJ cm⁻².

2.17.3. Hybridization of labelled probe to specific DNA sequences

The membrane was washed in pre-heated pre-hybridization buffer (5 x SSC,
0.1% SDS, 5% w/v dextran sulphate, 5% blocking agent) for 1 hr at 68°C in a Hybaid
Chapter 2: Materials and Methods

dual hybridization oven. The denatured probe was then added to the pre-hybridization buffer and incubated with the membrane at 68°C overnight. Following hybridization, the membrane was washed at 68°C in 100 ml of pre-heated stringency solution 1 (1% w/v SDS, 0.1 x SSC) and then 100 ml of stringency solution 2 (0.5% (w/v) SDS, 0.1 x SSC).

2.17.4. Detection of probe - target hybrids

The membrane was rinsed in diluent buffer (100 mM Tris-HCl, 300 mM NaCl pH 7.0) before being incubated for 1 hr with shaking in a 1% (w/v) solution of blocking reagent in diluent buffer. Anti-FI AP conjugate (1 in 10 000 diluted) and 0.5% (w/v) BSA in diluent buffer were then added and the membrane incubated for a further 1 hr with shaking. Unbound conjugate was removed by rinsing three times in a 0.3% (v/v in diluent) Tween-20 solution before detection reagent was applied to the DNA covered surface. After incubation for 5 min, the excess reagent was removed and the membrane subjected to autoradiography to detect bound conjugate. Typically, the autoradiography film was exposed for 30 min before developing.

2.18. Analysis of lipopolysaccharides (LPS)

2.18.1. Preparation of LPS

*C. jejuni* (or *S. typhimurium*) strains were grown overnight in 5 ml of MHB (or LB) at 37°C with shaking. Late exponential phase cells (1 ml OD$_{600}$ 0.6) were harvested by centrifugation (3020 g, 20 min) and resuspended in 200 µl of SDS-PAGE loading buffer (SDS polyacrylamide gel electrophoresis; 100 mM Tris-Cl pH 8.0, 2% β-mercaptoethanol, 4% SDS, 0.2% bromophenol blue, 0.2% xylene cyanol, 20% glycerol) to allow cell lysis. Following boiling for 10 min, proteinase K was added to a final concentration of 0.5 mg ml$^{-1}$ and the sample incubated at 65°C for 2 hr. Samples were stored at -20°C until needed.
Chapter 2: Materials and Methods

2.18.2. Analysis of LPS using Tricine SDS-PAGE

This procedure was carried out using a method modified from Sprott et al. (1994). SDS-PAGE was performed using either a Protean II apparatus (20 x 16 cm gel; Bio-Rad) or the smaller Mini-Protean II (7.2 x 10.2 cm; Bio-Rad) apparatus. Both stacking (4% (v/v) final acrylamide concentration) and separating gels (either 15 or 18% (v/v) final acrylamide concentration) were made using 30%:0.8% (w/v) acrylamide mix (Protogel, supplied by National Diagnostics in the ratio 30:0.8 acrylamide:bis-acrylamide). The constituents of both stacking and separating gels are shown in Table 2.3. LPS preparations were boiled for 10 min before loading on to the SDS-PAGE gel.

Table 2.3: Tricine SDS-PAGE gel constituents

<table>
<thead>
<tr>
<th>Constituent</th>
<th>Protean II apparatus (20 x 16 cm gel)</th>
<th>Mini-Protean II (7.2 x 10.2 cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30% Protogel acrylamide</td>
<td>1.5 ml</td>
<td>375 µl</td>
</tr>
<tr>
<td>4 x Gel Buffer (4 M Tris-HCl; 0.4% (w/v) SDS, pH 8.45)</td>
<td>1.86 ml</td>
<td>465 µl</td>
</tr>
<tr>
<td>Glycerol</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Distilled water</td>
<td>6.64 ml</td>
<td>1.66 ml</td>
</tr>
<tr>
<td>Ammonium persulphate (10 mg ml⁻¹; Sigma)</td>
<td>35 µl</td>
<td>12 µl</td>
</tr>
<tr>
<td>TEMED (N, N, N', N'-tetramethyl-ethylenediamine; Sigma)</td>
<td>18 µl</td>
<td>10 µl</td>
</tr>
</tbody>
</table>

75
Chapter 2: Materials and Methods

Samples were loaded (10 μl per well Protean II apparatus, 5 μl per well Mini Protean II apparatus) and electrophoresed in the presence of SDS-PAGE running buffer (0.1 M Tris, 0.1 M Tricine, 0.1 M SDS). Protean II gels were run at 15 mA per gel for 16 hr. Mini Protean II gels were also run at 15 mA per gel but for only 3 hr. Samples were run alongside biotinylated SDS-PAGE molecular weight standards (Bio-Rad); 200 kDa (myosin), 116.25 kDa (β-galactosidase), 97.4 kDa (phosphorylase b), 66.2 kDa (bovine serum albumin), 45 kDa (ovalbumin), 31 kDa (carbonic anhydrase), 21.5 kDa (soybean trypsin inhibitor), 14.4 kDa (lysozyme) and 6.5 kDa (aprotinin).

2.18.3. Silver staining of LPS

Following SDS-PAGE the separating gel was removed and the LPS visualized by silver staining using a method based on that Tsai & Frasch (1982). All incubations were at room temperature with gentle shaking. Briefly, the gel was fixed by incubating overnight in a solution of 40% (v/v) ethanol and 5% (v/v) glacial acetic acid. The gel was then transferred to an oxidative solution (0.7% w/v periodic acid, 40% (v/v) ethanol and 5% (v/v) glacial acetic acid) for 5 min before being washed for 10 min in distilled water. This wash stage was repeated three more times before the gel was transferred to a pre-staining solution (28 ml 0.1 M NaOH, 2 ml NH₄OH and 115 ml distilled water). Whilst the gel was slowly being shaken in the pre-staining solution, 5 ml of freshly made 20% (w/v) AgNO₃ was gradually added. Following this addition the gel was incubated for 10 min to allow staining. The separating gel was then washed twice for 5 min in distilled water to remove excess stain before being transferred to developing solution (10 mg citric acid and 100 μl formaldehyde in 200 ml distilled water). The gel was incubated in this solution in the dark until bands could be visualized. The developing reaction was halted by rinsing the gel several times in distilled water. The silver stained gel was then photographed or dried for long term storage using a gel-air dryer (Bio-Rad).
2.18.4. Immunoblot analysis of LPS using Penner (HS) antisera

2.18.4.1. Transfer of LPS to a nitrocellulose membrane

The transfer of LPS was carried out electrophoretically according to the protein transfer method described by Sambrook et al. (1989). Translocation from the separating Tricine SDS-PAGE gel to a nitrocellulose membrane (Schleicher & Schuell, Dassel, Germany) was carried out at 100 V for 2 hr in the presence of western transfer buffer (0.033 M Tris, 0.19 M glycine, 200 ml methanol, 800 ml distilled water). The nitrocellulose membrane was then Ponceau S (Sigma) stained to confirm transfer and to mark the molecular weight standards.

2.18.4.2. Hybridization of primary antibody to LPS

The nitrocellulose membrane was washed twice for 10 min in 1 x TBS (20 mM Tris (pH 7.5), 0.5 M NaCl) before being incubated at 4°C for 16 hr in blocking solution (5% non-fat dry milk in 1 x TBS). The membrane was then washed in 1 x TTBS (20 mM Tris (pH 7.5), 0.5 M NaCl, 0.1% Tween-20) for 10 min before primary antibody was added (diluted 1/5000 in a solution of 5% non-fat dry milk in 1 x TTBS). Following a 2 hr incubation at room temperature with shaking, the membrane was washed twice for 5 min in TTBS to remove unbound antibody.

2.18.4.3. Detection of bound primary antibody

Detection of primary antibody was achieved using the Immuno-Star chemiluminescent detection system supplied by Bio-Rad. This protocol is based on the detection of the bound rabbit IgG primary antibody using a goat anti-rabbit IgG secondary antibody conjugate. The secondary antibody is conjugated to alkaline phosphatase (anti-rabbit IgG-AP). Briefly, the anti-rabbit IgG-AP (1 in 3000 diluted in a solution of 5% non-fat dry milk in 1 x TTBS) was added and the membrane incubated
for 1 hr with shaking. Unbound conjugate was removed by rinsing three times in 1 x TTBS before a detection reagent and enhancer solution (20:1 ratio reagent:enhancer) was applied to the LPS covered surface. After incubation for 5 min, the excess reagent/enhancer was removed and the membrane subjected to autoradiography. Typically, the autoradiography film was exposed for 30 min before developing.

2.18.5. Immunoblot analysis of LPS using cholera toxin - biotin conjugate

This method is identical to the immunoblot protocol using Penner (HS) antiserum (section 2.18.4) except for two differences. The first was that cholera-biotin conjugate is used in place of the primary antibody. This was added to the membrane at a concentration of 12.5 μg ml⁻¹ in a solution of 5% non-fat dry milk in 1 x TTBS. The second difference was due to the fact that the cholera toxin is not a rabbit IgG-antibody, but instead conjugated to biotin. Detection is therefore not carried out using the secondary antibody anti-rabbit IgG-AP, but with a streptavidin-alkaline phosphatase conjugate (diluted 1 in 100 000 diluted in a solution of 5% non-fat dry milk in 1 x TTBS).
Chapter 3: Identification of lipopolysaccharide biosynthesis genes in the genome sequence of *C. jejuni* NCTC 11168

3.1. Introduction

The General Introduction to this study (Chapter 1) discussed current knowledge concerning the structure, biosynthesis, genetics and function of lipopolysaccharide (LPS) and lipooligosaccharide (LOS) molecules. This description was based on reports relating to a few intensely studied species - *E. coli* and *Salmonella* serovars which produce LPS, and *H. influenzae* which produces LOS. These examples illustrate the complex biochemical structures, varying mechanisms of biosynthesis and variety of functions which LPS/LOS molecules exhibit. In contrast with this wealth of information are the scant reports concerning the LPS/LOS molecules produced by *C. jejuni*. It is clear from the examples of *E. coli* and *H. influenzae* that a study of the genetic basis of LPS/LOS biosynthesis has greatly enhanced our knowledge of these biologically important molecules. In particular, the study by Hood *et al.* (1996) has shown how powerful a genome-wide approach to identifying LPS/LOS biosynthesis genes can be. The aim of this study was to apply a similar bioinformatic analysis to the genome of *C. jejuni* NCTC 11168 in order to identify putative LPS/LOS genes. The information from such a study could then provide the basis for an experimental approach to further characterize LPS/LOS biosynthesis in *C. jejuni*.

3.2. Sequencing of the *C. jejuni* NCTC 11168 genome

The *C. jejuni* project was carried out by the Sanger Centre (funded via Beowulf Genomics; [http://www.beowulf.org.uk/home/html](http://www.beowulf.org.uk/home/html)). The strain used was NCTC 11168, which was originally isolated from a human patient with gastroenteritis. This strain is from the HS:2 serotype group and therefore is not thought to produce O-chain (Aspinall *et al.*, 1993b). The genome was sequenced using a small insert random shotgun library, is 1,641,481 bp in length and contains 1,654 predicted open-reading frames (Parkhill *et al.* 1997).
Chapter 3: Identification of LPS biosynthesis genes in *C. jejuni* NCTC 11168

*al.*., 2000a). The complete sequence and annotation can be obtained from EMBL with the ID CJ11168 (accession number AL111168) or from the Sanger Centre (http://www.sanger.ac.uk/Projects/C_jejuni/).

3.3. Computer analysis of the *C. jejuni* genome sequence

During the sequencing of the *C. jejuni* genome, the data obtained by the Sanger Centre was regularly put into the public domain via the Internet. This information was used in BLAST homology searches with known sequences of interest from species such as *E. coli*, *H. influenzae*, *Neisseria meningitidis* and *Helicobacter pylori*, in order to identify similar sequences in *C. jejuni*. BLAST searches were normally carried out with protein, rather than nucleotide sequences to search against the translated *C. jejuni* sequence. Once a *C. jejuni* protein sequence of interest was identified, the corresponding nucleotide sequence and flanking sequence (5 kb either side) were analyzed further using the Gene Runner computer program (details of this program and some of the other bioinformatic tools used in this Chapter are provided in section 2.16). This program was used to identify the complete open reading frame encoding the protein sequence responsible for the original BLAST match. Once the whole protein sequence from this open reading frame was obtained it was used in a second BLAST search in order to obtain a putative identification. Analysis of nucleotide sequences using the Gene Runner program also allowed a picture of operon organization and the clustering of genes with related function to be obtained.

However, whilst informative, homology data should be treated with some caution as it cannot confirm the identity of a given gene, and can only act as a guide. In several cases, further characterization of the encoded protein sequence was also undertaken to support the putative identification. This further characterization involved several factors, such as investigating the secondary structure of a protein of interest. The conservation of secondary structure between a protein and its homologs implies it has a similar function to that of its homologs. The presence of active sites or conserved residues and motifs was also investigated. Once again, the conservation of these
features between a protein and its homologs implies a similar or identical function. The potential sub-cellular location of a given protein was also determined by considering the presence of potential trans-membrane domains. This again, can provide evidence of a functional relatedness between similar proteins.

Following the completion of the above analysis, the information it provided was incorporated into the annotation of the completed sequence and into the public announcement of the finished genome (Parkhill et al., 2000a). It will be presented here in three major sections based on the structural moieties of LPS (i.e. Kdo-lipid A, core and O-chain). Detailed information about promoter sequences, ribosome binding sites, and translation start sites will not be given - the analysis will concentrate on which genes are present (or apparently absent), their relative position and organization in the genome, and their similarity to other LPS/LOS biosynthesis genes.

3.4. Kdo-lipid A biosynthesis genes

\( lpxA \)

The protein encoded by the open reading frame Cj0274 is highly homologous to various UDP-N-acetylglucosamine \( O \)-acyltransferase (LpxA) proteins (Table 3.1). The \( E. \ coli \) LpxA enzyme is cytoplasmic and homotrimeric, with an active site probably located between the sub-units (Raetz & Roderick, 1995). It contains a left handed parallel \( \beta \) helix secondary structure which is specified by a series of hexapeptide motifs (Raetz & Roderick, 1995). These motifs are also apparent in the LpxA proteins from other Gram negatives, and are characterized by an isoleucine, leucine or valine in the first position, and often a glycine residue in the second (Vuorino et al., 1994). The predicted protein sequence from Cj0274 contains similar motifs (Figure 3.1), reinforcing the putative identification of this open reading frame as \( lpxA \). Cj0274 is not linked to other LOS biosynthesis genes, this is in contrast to the \( E. \ coli \) and \( H. \ influenzae \) \( lpxA \) genes which are located in proximity to other lipid A biosynthesis genes (Fleischmann et al., 1995; Schnaitman & Klena, 1993).
Table 3.1: Details of the Kdo-lipid A biosynthesis genes identified in the genome of C. jejuni NCTC 11168. Open reading frame numbers are those provided in the C. jejuni NCTC 11168 genome annotation (http://www.sanger.ac.uk/Projects/C_jejuni/). Homology searches were carried out using the SWISS-PROT protein sequence database - the three protein sequences with highest similarity are shown here along with a probability score. This gives an indication of the confidence with which the identification was made; lower numbers indicate higher probability. The proposed function and gene names are suggested on the basis of homology and in some cases by the further characterization of the encoded protein sequence (see text for details).
<table>
<thead>
<tr>
<th>Open reading frame no.</th>
<th>Proteins with highest similarity (probability score)</th>
<th>Putative function of gene product in C. jejuni</th>
<th>Proposed gene name in C. jejuni</th>
</tr>
</thead>
</table>
| Cj0274                | *Aquifex aeolicus* LpxA (8e⁻⁻⁵⁵)  
*Haemophilus influenzae* LpxA (9e⁻⁻⁵²)  
*Proteus mirabilis* LpxA (9e⁻⁻⁵⁰) | UDP-N-acetylglucosamine O-acyltransferase | lpxA |
| Cj0132                | *Haemophilus influenzae* LpxC (1e⁻⁻⁷¹)  
*Escherichia coli* LpxC (1e⁻⁻⁶³)  
*Aquifex aeolicus* LpxC (1e⁻⁻³⁷) | UDP-3-O-[hydroxymyristoyl]-N-acetylglucosamine deacetylase | lpxC |
| Cj0576                | *Helicobacter pylori* 210 LpxD (7e⁻⁻⁶⁵)  
*Helicobacter pylori* J99 LpxD (7e⁻⁻⁶⁵)  
*Rickettsia rickettsii* LpxD (5e⁻⁻⁵²) | UDP-3-O-[hydroxymyristoyl]-N-acetylglucosamine acyltransferase | lpxD |
| Cj0288                | *Helicobacter pylori* 210 LpxB (1e⁻⁻⁶⁸)  
*Haemophilus influenzae* LpxB (6e⁻⁻²⁷)  
*Escherichia coli* LpxB (1e⁻⁻²⁴) | Lipid A disaccharide synthetase | lpxB |
| Cj0811                | *Haemophilus influenzae* LpxK (3e⁻⁻¹⁵)  
*Francisella tularensis* var. novicida LpxK (5e⁻⁻¹²)  
*Escherichia coli* LpxK (2e⁻⁻⁰⁶) | Tetra-acyl-disaccharide 4' kinase | lpxK |
<table>
<thead>
<tr>
<th>Open reading frame no.</th>
<th>Proteins with highest similarity (probability score)</th>
<th>Putative function of gene product in <em>C. jejuni</em></th>
<th>Proposed gene name in <em>C. jejuni</em></th>
</tr>
</thead>
</table>
| Cj0384                | *Helicobacter pylori* 26695 KdsA (2e⁻⁹⁵)  
*Helicobacter pylori* J99 KdsA (2e⁻⁹⁵)  
*Aquifex aeolicus* KdsA (2e⁻⁵⁶) | Kdo-8-phosphate synthetase | *kdsA* |
| Cj0813                | *Helicobacter pylori* J99 KdsB (5e⁻⁴⁴)  
*Helicobacter pylori* 26695 KdsB (5e⁻⁴⁴)  
*Haemophilus influenzae* KdsB (1e⁻⁴¹) | CMP-2-keto-3-deoxyoctulosonic acid synthetase | *kdsB* |
| Cj0707                | *Haemophilus influenzae* WaaA (6e⁻³¹)  
*Escherichia coli* WaaA (5e⁻²⁶)  
*Chlamydia trachomatis* WaaA (4e⁻²⁵) | 3-deoxy-D-manno-octulosonic acid (Kdo) transferase | *waaA* |
| Cj1134                | *Escherichia coli* WaaM (6e⁻¹²)  
*Haemophilus influenzae* WaaN (8e⁻¹⁰)  
*Escherichia coli* WaaN (1e⁻³⁰) | lauroyl acyltransferase | *waaM* |
| Cj0803                | *Haemophilus influenzae* MsbA (9e⁻⁸⁷)  
*Rhizobium* sp. NGR234 Y4gm (MsbA homolog; 2e⁻⁸³)  
*Escherichia coli* MsbA (4e⁻⁸¹) | ATP-binding cassette (ABC) transporter | *msbA* |
Chapter 3: Identification of LPS biosynthesis genes in *C. jejuni* NCTC 11168

H. *inf lpxA*, *P. mir lpxA*, *A. aeo lpxA* and *Cj 0274* were proposed as encoding the *C. jejuni* LpxA protein on the basis of homology (Table 3.1), and the presence of hexapeptide repeat motifs which are characteristic of acyltransferase proteins. The first residue of each repeat is highlighted by an asterisk (*) and is either an isoleucine, leucine or valine (Dicker & Seetharam, 1992; Vuorino et al., 1994). Identical residues are highlighted in red. H. *inf lpxA*, *H. influenzae* LpxA; *P. mir lpxA*, *P. mirabilis* LpxA; *A. aeo lpxA*, *Aquifex aeolicus* LpxA.

![Alignment of the amino acid sequences of LpxA proteins from various Gram negative bacteria.](image)

**Figure 3.1:** Alignment of the amino acid sequences of LpxA proteins from various Gram negative bacteria. *Cj0274* is proposed as encoding the *C. jejuni* LpxA protein on the basis of homology (Table 3.1), and the presence of hexapeptide repeat motifs which are characteristic of acyltransferase proteins. The first residue of each repeat is highlighted by an asterisk (*) and is either an isoleucine, leucine or valine (Dicker & Seetharam, 1992; Vuorino et al., 1994). Identical residues are highlighted in red. H. *inf lpxA*, *H. influenzae* LpxA; *P. mir lpxA*, *P. mirabilis* LpxA; *A. aeo lpxA*, *Aquifex aeolicus* LpxA.

85
Chapter 3: Identification of LPS biosynthesis genes in _C. jejuni_ NCTC 11168

**lpxC**

The LpxC protein acts as a UDP-3-0-[hydroxymyristoyl]-N-acetylglucosamine deacetylase and is an essential cytoplasmic enzyme (Wyckoff et al., 1998). Cj0132 has a high level of identity to other Gram negative _lpxC_ genes, suggesting that it encodes the _C. jejuni_ LpxC protein (Table 3.1). Cj0132 is not found clustered with genes of related function - a characteristic shared with the _lpxC_ genes of _E. coli_ and _H. influenzae_ (Fleischmann et al., 1995; Schnaitman & Klena, 1993).

**lpxD**

The LpxD protein of _E. coli_ shares significant sequence homology with LpxA, consistent with their related functions as acyltransferases (Raetz, 1996). Just like the LpxA protein, LpxD is cytoplasmic, essential, and contains multiple hexapeptide motifs which are presumed to give the two proteins a similar secondary structure (Dicker & Seetharam, 1992; Raetz, 1996). The protein encoded by open reading frame Cj0576 is closely related to other Gram negative LpxD proteins, and contains similar hexapeptide motifs (data not shown). Unlike the _E. coli_ _lpxD_ gene, but similar to the _H. influenzae_ _lpxD_ gene, Cj0576 is not clustered with other lipid A biosynthesis genes (Fleischmann et al., 1995; Schnaitman & Klena, 1993).

**lpxB**

The _lpxB_ gene in _E. coli_ and other Gram negatives encodes the essential cytoplasmic enzyme, lipid A disaccharide synthetase. In _C. jejuni_, this enzyme is most probably encoded by Cj0288 (Table 3.1). Unlike in _E. coli_, this open reading frame is not found in the same operon as the _lpxA_ gene but is unlinked like the _H. influenzae_ _lpxB_ gene (Fleischmann et al., 1995; Schnaitman & Klena, 1993).

**lpxK**

The protein encoded by the _C. jejuni_ open reading frame Cj0811 is homologous to Gram negative tetra-acyl-disaccharide 4' kinases (LpxK). The LpxK protein of _E. coli_ is an essential membrane bound enzyme (Garrett et al., 1998) - the Cj0811
predicted protein sequence contains one trans-membrane domain at the N-terminus (data not shown). In both *E. coli* and *H. influenzae*, the *lpxK* gene is located in close proximity to the *msbA* and *kdsB* genes (Fleischmann *et al.*, 1995; Garrett *et al.*, 1997). This arrangement is partially conserved in *C. jejuni*, where the *lpxK* gene is part of an operon which also contains the *kdsB* gene, but does not include *msbA*.

**kdsA/kdsB**

In *C. jejuni*, the cytoplasmic enzymes KdsA and KdsB are probably encoded by the open reading frames Cj0384 and Cj0813 respectively (Table 3.1). Cj0384 (*kdsA*) is not located in close proximity to other LOS genes, a situation mirrored in *E. coli* and *H. influenzae* (Fleischmann *et al.*, 1995; Schnaitman & Klena, 1993). As previously mentioned, Cj0813 (*kdsB*) is located in an operon with *lpxK*, and not with *msbA* as it is in *E. coli* and *H. influenzae*.

**waaA**

The protein encoded by the *C. jejuni* open reading frame Cj0707 is homologous to Gram negative 3-deoxy-D-manno-octulosonic acid (Kdo) transferases (*WaaA*; Table 3.1). These enzymes are found anchored to the inner membrane and are responsible for the addition of Kdo to the lipid A moiety (Schnaitman & Klena, 1993). Two transmembrane regions (data not shown) are evident in the predicted protein sequence of Cj0707, reinforcing the identification of this open-reading frame as *waaA*. In *H. influenzae*, WaaA transfers one Kdo residue to the lipid A moiety, whilst the *E. coli* WaaA transfers two, and *Chlamydia trachomatis* WaaA transfers three (Belunis *et al.*, 1995; Belunis *et al.*, 1992; Fleischmann *et al.*, 1995). The protein product of Cj0707 shares the highest similarity to *H. influenzae* WaaA, and like this enzyme, only transfers one Kdo residue - all characterized *C. jejuni* core oligosaccharides contain one Kdo residue (reviewed in Moran & Penner, 1999). The structural basis for the ability of different Kdo transferases to incorporate different numbers of Kdo residues is still unclear, but it is not due to the presence of additional catalytic domains since all WaaA proteins are of similar length and exhibit high levels of sequence homology along their
entire sequence (Isobe et al., 1999). In *E. coli* and *Salmonella* strains, the *waaA* gene is located adjacent to two operons which contain the majority of core biosynthesis genes (reviewed in Heinrichs et al., 1998c). In contrast, the *C. jejuni waaA* is not located near such operons.

**waaM/waaN**

In *E. coli*, the WaaM enzyme is a membrane bound lauroyl acyltransferase which adds a 12 carbon containing fatty acid side chain to the lipid A moiety (Khan et al., 1998). Following this addition, a 14 carbon chain is transferred by the WaaN protein to complete the Kdo₂-lipid A moiety (Somerville et al., 1999). The WaaN and WaaM proteins of Gram negatives show a high degree of homology, presumably because the proteins have similar functions (Raetz, 1996). In the *C. jejuni NCTC 11168* genome, only one open reading frame (Cj1134) with similarity to either the *waaM* or *waaN* genes is present. The protein product of this open reading frame is predicted to contain one trans-membrane region, and because it is most similar to the *E. coli* WaaM protein will be designated WaaM (Table 3.1). The apparent absence of a *waaN* gene in *C. jejuni* will be discussed in section 3.4.1.

In *E. coli* and *H. influenzae*, the *waaM* and *waaN* genes are found unlinked to other genes involved in the biosynthesis of LPS/LOS (Fleischmann et al., 1995; Schnaitman & Klena, 1993). This arrangement is not conserved in *C. jejuni NCTC 11168*, where the *waaM* gene is located in a large cluster of genes. The protein products of these genes are thought to be involved with the synthesis of the core oligosaccharide moiety. This cluster will be described in section 3.5 of this chapter.

**msbA**

The protein encoded by the *C. jejuni* open reading frame Cj0803 is homologous to Gram negative lipid A-core ABC transporter proteins (MsbA; Table 3.1). These proteins are inner membrane bound and thought to be responsible for the export of the completed lipid A-core moiety to the periplasm, prior to O-chain attachment (Polissi & Georgopoulos, 1996). This process requires energy (Marino et al., 1985) - the predicted
Chapter 3: Identification of LPS biosynthesis genes in \textit{C. jejuni} NCTC 11168

protein encoded by Cj0803 contains an ATP/GTP-binding site motif, and also an ABC transporter family signature (Figure 3.2). Five trans-membrane regions are also apparent in the protein (data not shown). In \textit{E. coli} and \textit{H. influenzae}, the \textit{msbA} gene is found as part of an operon which also contains the \textit{kdsB} and \textit{lpxK} genes (Fleischmann \textit{et al.}, 1995; Schnaitman & Klena, 1993). This arrangement is not conserved in \textit{C. jejuni}, where the \textit{msbA} gene is found unlinked to other LPS/LOS biosynthesis genes.

3.4.1. Summary of Kdo-lipid biosynthesis in \textit{C. jejuni}

Homology searches show that homologs of the \textit{E. coli} genes involved with the biosynthesis of the Kdo-lipid A moiety can all been identified in the \textit{C. jejuni} genome, with the exception of the \textit{waaN} gene. Thus, the pathway of Kdo-lipid A biosynthesis in \textit{C. jejuni} is presumed to be similar to that of other Gram negative bacteria. The only possible deviation from this biosynthetic pathway could occur at the point where the secondary fatty acyl chains are added. In \textit{E. coli}, the WaaN protein completes the biosynthesis of the Kdo\textsubscript{2}-lipid A moiety by adding a 14 carbon chain to the disaccharide backbone of lipid A (Somerville \textit{et al.}, 1999). The apparent absence of this gene in \textit{C. jejuni} is intriguing, but could simply be because it is sufficiently diverged from other \textit{waaN} genes that homology searches fail to identify it. In contrast to \textit{E. coli}, where a 12- and then a 14-carbon acyl chain are added to the lipid A moiety, both the corresponding acyl molecules in the \textit{C. jejuni} HS:2 type strain appear to be 16-carbon chains (Moran, 1995a; Moran \textit{et al.}, 1991). Therefore, it is possible that the \textit{C. jejuni} WaaM enzyme is bi-functional and responsible for the addition of both the acyl chains. Functional studies will have to be undertaken to investigate this possibility.

The Kdo-lipid A biosynthesis genes identified are generally unlinked to others involved in the biosynthesis of LPS/LOS - a situation mirrored in \textit{E. coli} and \textit{H. influenzae} (Fleischmann \textit{et al.}, 1995; Schnaitman & Klena, 1993). The major exception is the \textit{waaM} gene which is located in a cluster of genes thought to be involved in the biosynthesis of the core moiety (section 3.5). The significance of this is not clear, but this arrangement is not conserved in \textit{E. coli} or \textit{H. influenzae} which possess both \textit{waaN}
Chapter 3: Identification of LPS biosynthesis genes in *C. jejuni* NCTC 11168

---

**msbA E.col:**

VPAIRINLKIPAKTVALKVSKSTIASCSTRFYDIDEKELMDGDLREYTLAS 415

**msbA H.inf:**

ELAVNNISFVAPAKTVALKVSKSTIASCSTRFYDIEQEGLLDGNIQYRLSNL 420

**msbA C.jej:**

KSVKRGVNFDFVKMLVITSQGGRSSTILNMFFYEKQRKLLLQEDISTTIEESJ 412

**msbA A.aeo:**

---1KNVNLVYRNQSVGIKHDTSKSTFVDRPRLLAE-Y[DF]FVDHELREFKLKS: 399

---

**msbA E.col:**

RNQVALVSNAVHLDNDTVIDTAYARTEYSRQIEEHARMYAMDMINKMDNLTVIG: 475

**msbA H.inf:**

RNCVAVSVQVHLDNDTIANIAYAAQDKYRSREIIKAKAYALEQIEKLDQVFTIG: 480

**msbA C.jej:**

HAKGVLTVNIYLDSDFAATAYS--EEELERVQILKLNAAYEYVKEEM-GIWAEVK: 469

**msbA A.aeo:**

REKIGFLSFVMMVRGSHVRFILLIRARSAELEELMLKIA-SCKTVLNSGFEKLDFLE: 461

---

**msbA E.col:**

NVLMLGGGQRQRIIHLNALLRDSPIIIADATSALUTESRAIQAALDELQKNRTSLLVI: 535

**msbA H.inf:**

RNITSLSGQRQRIIHLNALLRSPVLIILDEATSALUTESRAIQAALDELEKRDQRTV: 540

**msbA C.jej:**

THKMLSGGQRQRIIHLNALLKYNPDVLIDDEATSALUTESRAIQAALDELEKRDQRTV: 529

**msbA A.aeo:**

GCKNLSGGEKQRIHNLIDLKFLKNEIVLVDEATSALUTESRAIQAALDELEKRDQRTV: 519

---

**msbA E.col:**

AHLSTIKAELVVEDTVIREGTHNDLLEHGGVYAOQLHKKMFQ---: 582

**msbA H.inf:**

AHLSTIENADELVIDREIERGENHKTLEDQENGAKQLHSMQFT----: 587

**msbA C.jej:**

AHLSTIENADNEVVLKXEGKRVLLIGDDEELLQTCSLQKFKSKEKPSFS: 580

**msbA A.aeo:**

AHTPIFNFDONILVIEPEGKLKEFSTSKENIKL-------------------: 553

---

**Figure 3.2:** Alignment of the carboxy-terminal amino acid sequences of various MsbA proteins. Cj0803 is proposed as encoding the *C. jejuni* MsbA protein on the basis of homology (Table 3.1), the presence of an ATP/GTP-binding site motif, and the presence of an ABC transporter family signature (Higgins *et al.*, 1990; Hung *et al.*, 1998). The ATP/GTP-binding site is indicated by *---*** and has the consensus ‘G/A-x-x-x-G-K-S/T’ (Walker *et al.*, 1982). The ABC transporter family signature is represented by **abcdefgijkl** where: a, [LIVMFYC]; b, [SA]; c, [SAPGLVFYKQH]; d, [G]; e, [DENQMW]; f, [KRQASPLCMIMFW]; g, [KRQSTAVM]; h, [KRACLVM]; i, [LIVMFYPAN]; j, [LIVAMFW]; k, [SAGCLVP]; l, [LIVMFWYAST]. Identical residues are highlighted in red. E.col, *E. coli*; H.inf, *H. influenzae*; C.jej, *C. jejuni*; A.aeo, *Aquifex aeolicus*
and waaM genes (Fleischmann et al., 1995; Schnaitman & Klena, 1993).

3.5. Core oligosaccharide biosynthesis genes

The genes described so far, whose protein products are involved with the synthesis of the Kdo-lipid A moiety, are mostly unclustered and appear in operons with genes of unrelated function. This is not the case for the genes involved with the biosynthesis of the core oligosaccharide which are clustered together in a approximately 30 kb long section of the genome known as the LOS cluster. This group includes the putative waaM gene described above, and genes involved in the biosynthesis and transfer of sugar residues to the growing core molecule. The sequence of the cluster also contains homo-polymeric tracts which may play a role in the phase variation of LOS epitopes (Parkhill et al., 2000a). This will be discussed in later sections of this chapter (section 3.7).

Some of the open reading frames present in the cluster had been described prior to the sequencing of the genome (Fry et al., 1998; Klena et al., 1998; Wood et al., 1999), and several have been characterized since (Fry et al., 2000a; Gilbert et al., 2000; Linton et al., 2000; Szymanski et al., 1999). The information from these studies will be incorporated into this chapter to provide a comprehensive view of the core biosynthesis genes of C. jejuni. The following sections will therefore describe the cluster and the genes it contains, and will attempt to group these into several functional categories. A map of the cluster is shown in Figure 3.3.

3.5.1. The gmhA, D and E genes (ADP-L-glycero-D-manno-heptose biosynthesis)

The inner core oligosaccharides of all Gram negative LPS/LOS molecules contain at least two L-glycero-D-manno-heptose residues (Heinrichs et al., 1999). The biochemical pathway for the biosynthesis of this sugar has not been completely elucidated (Valvano et al., 2000), although several key steps have now been biochemically and genetically characterized. The first committed step of biosynthesis
Figure 3.3: Schematic diagram of the LOS biosynthesis cluster of *C. jejuni* NCTC 11168. Open reading frames and the direction of their transcription are indicated by arrows. Proposed gene names and open reading frame numbers are also given. The cluster has been split into two sections for illustration purposes, because of the number of open reading frames it contains. The proposed function of each gene product is indicated by colour coding. The gene names are suggested on the basis of homology or functional studies (details in text). Where the specific function of a gene is unclear, a name is proposed based on the existing *wla* nomenclature (Fry *et al.*, 1998; Fry *et al.*, 2000b). This nomenclature provides sufficient gene designations for the polymorphic open reading frames described in Chapter 4. This diagram is not shown to scale.
Chapter 3: Identification of LPS biosynthesis genes in C. jejuni NCTC 11168

Functional categories:

- **Green**: Biosynthesis of lipid A
- **Red**: Biosynthesis and transfer of ADP-L-glycero-D-manno-heptose residues
- **Yellow**: Biosynthesis and transfer of N-acetylneuraminic acid (sialic acid)
- **Blue**: Inter-conversion between UDP-glucose and UDP-galactose
- **Purple**: Putative glycosyltransferases
- **Black**: Other functions

Proposed gene name:

- Green: *wlaM wlaL wlaK wlaJ wlaH wlaG wlaF wlaE wlaD wlaC wlaB galE wlaA*

Cj number: 1119 1120 1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132

- Red: *waaC waaM wlaNA wlaNB wlaO wlaP wlaQ wlaR neuBl neuCl neuA1 wlaSC wlaSB wlaSA waaF gmhA gmhE gmhD wlaT*

Cj number: 1133 1134 1135 1136 1137 1138 1139 1140 1141 1142 1143 1144 1145 1146/7 1148 1149 1150 1151 1152
involves the conversion of sedoheptulose 7-phosphate into D-glycero-D-manno-heptose 7-phosphate by a cytoplasmic protein, phosphoheptose isomerase, which is encoded by the \( \text{gmhA} \) gene (Brooke & Valvano, 1996a; Brooke & Valvano, 1996b; Valvano, 1999). Mutation of this gene leads to the absence of L-glycero-D-manno-heptose, and all subsequently attached residues from the wild type LPS/LOS molecule (Bauer et al., 1998; Brooke & Valvano, 1996b; Preston et al., 1996b). The protein product of open reading frame Cj1149 shows high similarity to other GmhA molecules and is presumed to encode the \( C. \) jejuni GmhA protein (Table 3.2).

The intermediate steps in the biosynthesis of ADP-L-glycero-D-manno-heptose are still unclear (Valvano et al., 2000), but the GmhE protein is thought to be involved (Valvano et al., 2000). In \( E. \) coli, the GmhE protein has two domains, the first is a putative kinase from the ribokinase family (Valvano et al., 2000), whilst the second has strong homology to the cytidylyltransferase superfamily (Valvano et al., 2000). This domain structure is conserved in the \( C. \) jejuni GmhE protein - presumed to be encoded by Cj1150 (Table 3.2).

The final step in the biosynthetic pathway is the conversion of ADP-D-glycero-D-manno-heptose to ADP-L-glycero-D-manno-heptose by an epimerase encoded by the \( \text{gmhD} \) gene (Coleman, 1983). This gene has been isolated and characterized in several species including \( N. \) gonorrhoeae (Drazek et al., 1995), \( H. \) influenzae (Nichols et al., 1997), \( S. \) typhimurium (Sirisena et al., 1994) and \( V. \) cholerae (Stroher et al., 1995b; Vimont et al., 1997). In \( C. \) jejuni, the open reading frame Cj1151 is thought to encode the GmhD protein (Table 3.2). The protein product of this gene contains a \( \beta\alpha\beta \) ADP-binding fold (Gly-x-Gly-x-x-Gly), which is common to GmhD proteins, at the N-terminus of the protein (Figure 3.4; Rossmann et al., 1974; Wierenga et al., 1986; Stroher et al., 1995b).

In other Gram negatives, the \( \text{gmhA}, \text{D} \) and \( \text{E} \) genes are unlinked to each other. For example, in \( E. \) coli, the \( \text{gmhA} \) gene is found at about 5.9 minutes on the K-12 chromosome, in between \( \text{rnhQ} \) and \( \text{pepD} \) (Brooke & Valvano, 1996a). The \( \text{gmhE} \) gene is several minutes from the \( \text{gmhD} \) gene (Valvano et al., 2000), which is found in the core biosynthesis cluster (81-82 min), where it is the first gene in an operon containing
Table 3.2: Details of the *C. jejuni* NCTC 11168 genes involved with the biosynthesis and transfer of ADP-L-glycero-D-manno-heptose. Open reading frame numbers are those provided in the *C. jejuni* NCTC 11168 genome annotation (http://www.sanger.ac.uk/Projects/C_jejuni/). Homology searches were carried out using the SWISS-PROT protein sequence data base - the three protein sequences with highest similarity are shown here along with a probability score. This gives an indication of the confidence with which the identification was made; lower numbers indicate higher probability. The proposed function and gene names are suggested on the basis of homology and in some cases by the further characterization of the encoded protein sequence (see text for details).
<table>
<thead>
<tr>
<th>Open reading frame no.</th>
<th>Proteins with highest similarity (probability score)</th>
<th>Putative function of gene product in <em>C. jejuni</em></th>
<th>Proposed gene name in <em>C. jejuni</em></th>
</tr>
</thead>
</table>
| Cj1149                | *Helicobacter pylori* J99 GmhA (9e⁻⁵⁷)  
*Helicobacter pylori* 26695 GmhA (3e⁻⁵⁰)  
*Methanococcus jannaschii* LpcA (probable GmhA; 3e⁻⁴⁰) | Phosphoheptose isomerase                          | gmhA                             |
| Cj1150                | *Helicobacter pylori* 26695 GmhE (2e⁻¹ⁱ³)  
*Haemophilus influenzae* GmhE (4e⁻⁹¹)  
*Escherichia coli* GmhE (4e⁻⁹¹) | ADP-heptose synthetase                           | gmhE                             |
| Cj1151                | *Haemophilus influenzae* GmhD (8e⁻⁴⁰)  
*Salmonella typhimurium* GmhD (3e⁻³⁹)  
*Escherichia coli* GmhD (8e⁻³⁸) | ADP-L-glycero-D-manno-heptose-6-epimerase         | gmhD                             |
| Cj1133                | *Escherichia coli* WaaC (1e⁻¹⁹)  
*Salmonella typhimurium* WaaC (3e⁻¹⁶)  
*Escherichia coli* WaaQ (6e⁻⁰⁷) | Heptosyltransferase I                           | waaC                             |
| Cj1148                | *Haemophilus influenzae* WaaF (5e⁻²⁰)  
*Salmonella typhimurium* WaaF (2e⁻¹⁹)  
*Escherichia coli* WaaF (2e⁻¹⁸) | Heptosyltransferase II                          | waaF                             |
Chapter 3: Identification of LPS biosynthesis genes in *C. jejuni* NCTC 11168

GmhD E.coli: -KIVGAGVTGAINKAIWNLDDGTIKMYLVLVRIYADYMDKEDLIQ:59
GmhD H.infl: -KIVGAGVTGAINKAIWNLDDGTIKMYLVLVRIYADYMDKEDLIQ:59
GmhD V.chol: -KIVGAGVTGAINKAIWNLDDGTIKMYLVLVRIYADYMDKEDLIQ:59
*Cj1151*: MQLALNDKGITLKDGRKELGRKIEK---HEITIIEPRSSATPEGNQLQSFHFKNLLELIG:58

Figure 3.4: Alignment of the N-terminal amino acid sequences of various GmhD proteins. *Cj1151* is proposed as encoding the *C. jejuni* GmhD protein on the basis of homology (Table 3.2), and the presence of an βαβ ADP-binding fold (Gly-x-Gly-x-x-Gly) indicated by *-*-*--* (Rossmann *et al.*, 1974; Stroeher *et al.*, 1995b; Wierenga *et al.*, 1986). Identical residues are highlighted in red. E.coli, *E. coli*; H.infl, *H. influenzae*; V.chol, *Vibrio cholerae* O:1.
the \textit{waaC}, \textit{waaF} and \textit{waaL} genes (Heinrichs \textit{et al.}, 1998c). This is not the case in \textit{C. jejuni}, where the three \textit{gmh} genes are found in an operon which also contains an additional open reading frame of unknown function (Figure 3.3). Further characterization of this operon is needed to reveal whether the genes, although positionally linked, are transcriptionally linked.

3.5.2. Transfer of ADP-L-glycero-D-manno-heptose residues

Once synthesized, at least two L-glycero-D-manno-heptose residues are added to the core oligosaccharide of Gram negative LOS/LPS molecules (Heinrichs \textit{et al.}, 1999). The first L-glycero-D-manno-heptose is added to Kdo by the WaaC enzyme (Kadrmas & Raetz, 1998; Sirisena \textit{et al.}, 1992). This is a peripheral membrane protein which is thought to function as part of a coordinated complex of glycosyltransferases on the cytoplasmic face of the inner membrane (Raetz, 1996). Mutation of the \textit{waaC} gene in \textit{S. typhimurium} leads to truncation of the LPS caused by the loss of L-glycero-D-manno-heptose and all subsequently attached sugars from the molecule (Sirisena \textit{et al.}, 1992).

The WaaC enzyme in \textit{C. jejuni} NCTC 11168 is encoded by Cj1133 (Table 3.2). This open reading frame has already been isolated and partially characterized in the \textit{C. coli} strain M275, and \textit{C. jejuni} strain F38011 (Klena \textit{et al.}, 1998). The suggestion that this gene encodes for the WaaC protein is supported by similarity (Table 3.2) and the finding that the \textit{C. coli} gene was able to complement a \textit{S. typhimurium waaC} mutant strain and thus restore biosynthesis of a complete LPS structure (Klena \textit{et al.}, 1998). However, this identification should be confirmed by gene mutation and analysis of the LPS/LOS molecules produced by the resulting mutant strain. Unfortunately attempts to mutate the \textit{waaC} gene in \textit{Campylobacter} have not succeeded (Ben Fry, personal communication; John Klena, personal communication). In contrast to the \textit{E. coli waaC} gene which is found in an operon with \textit{gmhD} and \textit{waaFL}, the \textit{C. jejuni waaC} is not associated with other genes involved with the biosynthesis or transfer of ADP-L-
glycero-D-manno-heptose (Figure 3.3). Instead, C. jejuni waaC is the first gene of an operon which also contains the waaM gene and two putative glycosyltransferases.

In Gram negatives, the second L-glycero-D-manno-heptose residue is transferred by the WaaF protein (Allen et al., 1998; Bauer et al., 1999; Jennings et al., 1995; Kievit & Lam, 1997; Nichols et al., 1997; Sirisena et al., 1994). In C. jejuni, this enzyme is probably encoded by the open reading frame Cj1148 (Table 3.2). The predicted protein from Cj1148 contains one putative trans-membrane domain, which supports its identification as being the peripheral membrane protein WaaF (data not shown). The Cj1148 predicted protein also shows significant similarity to the putative C. jejuni WaaC protein (data not shown) - this similarity presumably reflects the close functional similarity between the two proteins. The putative identification of Cj1148 as the C. jejuni waaF gene is investigated further in Chapter 6 of this study.

In E. coli, and in other species for example, H. influenzae strain 2019, the waaF gene is located in close proximity to other genes involved with the biosynthesis and transfer of L-glycero-D-manno-heptose (Brabetz et al., 1997; Nichols et al., 1997). This arrangement is conserved in C. jejuni NCTC 11168, where Cj1148 is located adjacent to the gmhDEA operon (Figure 3.3).

3.5.3. Biosynthesis and transfer of N-acetylneuraminic acid (sialic acid)

Sialic acid is a common constituent of C. jejuni core oligosaccharide molecules (Moran & Penner, 1999). It is also found present in Haemophilus and Neisseria core moieties (Hood et al., 1999; Mandrell et al., 1991; Melaugh et al., 1994), and in the α-2,8-linked polysialic acid capsules produced by E. coli and N. meningitidis (Troy, 1992). In E. coli, sialic acid biosynthesis requires the action of two cytoplasmic proteins. The first is NeuC, which is involved in the biosynthesis of N-acetylmannosamine - a substrate required for N-acetylneuraminic acid biosynthesis (Zapata et al., 1992). The second is NeuB, (N-acetylneuraminic acid synthetase) which is responsible for the condensation of N-acetylmannosamine with phosphoenolpyruvate to form sialic acid (Annunziato et al., 1995; Vann et al., 1997). Mutation of either of
these genes yields strains which are defective in the synthesis of the α-2,8-linked capsule (Vimr et al., 1989; Zapata et al., 1992). Following the synthesis of sialic acid, an additional enzyme, NeuA, is required to activate the sialic acid (Vann et al., 1987; Zapata et al., 1989). This is then polymerized by the action of the polysialyltransferase NeuS (Steenbergen & Vimr, 1990; Vimr et al., 1995).

In *C. jejuni*, homologs of the genes that encode the *E. coli* NeuA, B and C proteins are all evident in the cluster of genes involved with core biosynthesis. The NeuC enzyme is probably encoded by Cj1142 (Table 3.3). The predicted protein sequence from this open reading frame has a high level of similarity to the SiaA/SynA proteins of *N. meningitidis*, which like the *E. coli* NeuC protein, are involved with the biosynthesis of *N*-acetylmannosamine from *N*-acetylglicosamine-6-phosphate (Edwards et al., 1994). Cj1142 also shows homology to the MnaA protein of *Legionella pneumophila* (Table 3.3). The MnaA protein is involved in the biosynthesis of sugar residues for incorporation into O-chain. It has similar function (UDP-*N*-acetylglicosamine 2-epimerase), and hence protein sequence similarity, to NeuC proteins. The NeuB protein in *C. jejuni* is encoded by Cj1141. The predicted protein from this open reading frame has high levels of similarity to SiaC from *N. meningitidis*, which like NeuB, is a *N*-acyetylneuraminic acid synthetase (Table 3.3; Edwards et al., 1994). The function of this gene has recently been confirmed in a study by Linton et al. (2000). This report showed that expression of Cj1141 in an *E. coli neuB*-deficient strain restored polysialic acid capsule production and thus sensitivity to the K1 bacteriophage (Linton et al., 2000). This phage only infects encapsulated strains (Vimr & Troy, 1985). Furthermore, *N*-acyetylneuraminic acid synthetase activity could be detected in *E. coli neuB*-deficient strains expressing the Cj1141-encoded protein (Linton et al., 2000). The report also confirmed that the sialic acid produced by the action of the NeuB protein was incorporated into the core oligosaccharide by insertionally inactivating Cj1141 in *C. jejuni* NCTC 11168 (Linton et al., 2000). This resulted in the production of a truncated LOS molecule which no longer contained *N*-acyetylneuraminic acid (Linton et al., 2000).
Table 3.3: Details of the *C. jejuni* NCTC 11168 genes involved with the biosynthesis and transfer of N-acetylneuraminic acid to the core oligosaccharide. Open reading frame numbers are those provided in the *C. jejuni* NCTC 11168 genome annotation (http://www.sanger.ac.uk/Projects/C_jejuni/). Homology searches were carried out using the Genbank protein sequence data base - the three protein sequences with highest similarity are shown here along with a probability score. This gives an indication of the confidence with which the identification was made; lower numbers indicate higher probability. The proposed function and gene names are suggested on the basis of homology and in some cases by the further characterization of the encoded protein sequence (see text for details).
<table>
<thead>
<tr>
<th>Open reading frame no.</th>
<th>Proteins with highest similarity (probability score)</th>
<th>Putative function of gene product in ( C. ) jejuni</th>
<th>Proposed gene name in ( C. ) jejuni</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cj1142</td>
<td>\textit{Neisseria meningitidis} SiaA ( (3e^{-84}) ) \newline \textit{Neisseria meningitidis} SynA ( (4e^{-84}) ) \newline \textit{Legionella pneumophila} MnaA ( (4e^{-50}) )</td>
<td>Involved in the synthesis of ( N )-acetylmannosamine (ManNAc)</td>
<td>\textit{neuC1}</td>
</tr>
<tr>
<td>Cj1141</td>
<td>\textit{Neisseria meningitidis} SiaC ( (1e^{-11}) ) \newline \textit{Methanococcus jannaschii} hypothetical protein MJ1065 ( (1e^{-55}) ) \newline \textit{Streptococcus agalactiae} SiaC ( (2e^{-42}) )</td>
<td>( N )-acetylneuraminic acid synthetase</td>
<td>\textit{neuB1}</td>
</tr>
<tr>
<td>Cj1143</td>
<td>\textit{Helicobacter pylori} J99 hypothetical protein JHP0203 ( (4e^{-64}) ) \newline \textit{Helicobacter pylori} 26695 hypothetical protein HP0217 ( (1e^{-63}) ) \newline \textit{Haemophilus ducreyi} NeuA ( (2e^{-35}) ) \newline \textit{Escherichia coli} NeuA ( (7e^{-33}) )</td>
<td>Acylneuraminate cytidylyltransferase</td>
<td>\textit{neuA1}</td>
</tr>
<tr>
<td>Cj1140</td>
<td>\textit{Haemophilus influenzae} hypothetical protein HI0352 ( (4e^{-53}) ) \newline \textit{Plasmodium falciparum} hypothetical protein MAL4P2.50 ( (0.27) )</td>
<td>( \alpha)-2,3 sialyltransferase</td>
<td>\textit{wlaR}</td>
</tr>
</tbody>
</table>
Chapter 3: Identification of LPS biosynthesis genes in *C. jejuni* NCTC 11168

The gene necessary for the activation of sialic acid prior to incorporation into the core oligosaccharide, NeuA, is probably encoded by Cj1143 (Table 3.3). The Cj1143-encoded protein has homology to two hypothetical proteins from *H. pylori*, which are thought to be acylneuraminate cytidylyltransferases (NeuA), but in the absence of functional data have not been confirmed as such yet. The Cj1143-encoded protein also has similarity to the NeuA proteins of *H. ducreyi* and *E. coli*. In *H. ducreyi*, the NeuA protein probably functions as a dimer, and in contrast to other acylneuraminate cytidylyltransferases can utilize N-glycolylneuraminic acid as well as N-acetylneuraminic acid (Tullius *et al.*, 1996). The substrate specificity of the *C. jejuni* NeuA protein remains to be determined.

In *E. coli*, once sialic acid has been activated it is polymerized to form the α-2,8-linked polysialic acid capsule by the action of the NeuS protein (Steenbergen & Vimr, 1990; Vimr *et al.*, 1995). In *C. jejuni*, the activated sialic acid is not polymerized, but transferred to the core oligosaccharide by the action of the Cj1140-encoded transferase (Gilbert *et al.*, 2000). This open reading frame is a good example of how homology searches alone can fail to indicate the potential function of a gene. The only protein in the Genbank sequence database with similarity to the predicted protein sequence of Cj1140 is a protein of unknown function from *H. influenzae* (Table 3.3). However, by employing an activity screening assay and through NMR studies, it has been revealed that Cj1140 encodes for a α-2,3 sialyltransferase which adds sialic acid to the core to form a α-NeuNAc-(2→3)-β-D-Gal epitope (Gilbert *et al.*, 2000).

3.5.4. Inter-conversion between UDP-glucose and UDP-galactose

In Gram negative bacteria, galactose is a common constituent of core oligosaccharide and O-chain moieties. Galactose is provided for incorporation into these structures by the activity of the GalE enzyme (Huong *et al.*, 1990; Maskell *et al.*, 1991; Pierson & Carlson, 1996; Robertson *et al.*, 1993). GalE is a UDP-galactose-4-epimerase, which mediates the conversion of UDP-glucose to UDP-galactose, and *visa versa*, in a process involving the co-factor NAD+ (Bauer *et al.*, 1992). In *C. jejuni*,...
GalE is encoded by Cj1131 - this identification is based on homology and the presence of conserved residues, common to UDP-galactose-4-epimerases, which interact with NADH and either UDP-glucose or UDP-galactose (Table 3.4; Bauer et al., 1992; Liu et al., 1997; Swanson & Frey, 1993; Thoden et al., 1996).

In Salmonella strains, mutation of galE results in the loss of UDP-galactose leading to truncation of the core oligosaccharide and loss of the O-chain moiety (Hone et al., 1987). The identification of Cj1131 as galE was confirmed by the study of Fry et al. (2000a) who used the homologous gene from C. jejuni NCTC 11828 to complement a S. typhimurium galE mutant strain, thus restoring wild-type core and O-chain biosynthesis (Fry et al., 2000a). The report also detailed the construction of a galE mutant strain of C. jejuni NCTC 11828 which expressed a truncated core oligosaccharide moiety - a phenotype consistent with the loss of galactose residues from the outer core (Fry et al., 2000a).

3.5.5. Potential glycosyltransferases

Eleven genes encoding potential glycosyltransferases are identifiable by homology in the LOS cluster. These are highlighted in Figure 3.3 and shown in Table 3.5. Several of the predicted protein sequences from these genes have extensive similarity to each other - presumably because the proteins have similar functions. Some of these genes have been characterized, both in C. jejuni NCTC 11168 and other strains, whilst others still await functional studies to confirm their role in core biosynthesis.

The Cj1120-encoded protein shows similarity to the CapD and CpsE proteins of Staphylococcus aureus and Streptococcus suis, respectively (Table 3.5). The function of the CapD protein is uncertain (Lin et al., 1994, Sau et al., 1997), but CpsE is thought to be a glycosyltransferase (Kolkman et al., 1996; Kolkman et al., 1997). However, the WlaL protein also contains an NAD binding motif (GxxGxxG; Macpherson et al., 1994) and has similarity in the central part of the molecule to UDP-glucose-4-epimerases (data not shown). Functional studies have shown that mutation of wlaL in C. jejuni NCTC 11168 or 11828 causes no detectable change to the core
Table 3.4: Details of the *C. jejuni* NCTC 11168 gene involved with the interconversion of UDP-glucose and UDP-galactose. The open reading frame number is that provided in the *C. jejuni* NCTC 11168 genome annotation (http://www.sanger.ac.uk/Projects/C_jejuni/). Homology searches were carried out using the Genbank protein sequence data base - the three protein sequences with highest similarity are shown here along with a probability score. This gives an indication of the confidence with which the identification was made; lower numbers indicate higher probability. The proposed function and gene name are suggested on the basis of homology and the further characterization of the encoded protein (Fry et al., 2000a).
<table>
<thead>
<tr>
<th>Open reading frame no.</th>
<th>Proteins with highest similarity (probability score)</th>
<th>Putative function of gene product in <em>C. jejuni</em></th>
<th>Proposed gene name in <em>C. jejuni</em></th>
</tr>
</thead>
</table>
| Cj1131                 | *Lactobacillus casei* GalE (1e⁻⁸⁰)  
*Azospirillum brasilense* ExoB (GalE homolog; 1e⁻⁷⁵)  
*Streptococcus mutans* GalE (4e⁻⁷²) | UDP-galactose-4-epimerase | galE                             |

Chapter 3: Identification of LPS biosynthesis genes in *C. jejuni* NCTC 11168
Table 3.5: Details of the *C. jejuni* NCTC 11168 genes potentially encoding enzymes involved with the transfer of glycosyl residues. Open reading frame numbers are those provided in the *C. jejuni* NCTC 11168 genome annotation (http://www.sanger.ac.uk/Projects/C_jejuni/). Homology searches were carried out using the Genbank protein sequence data base - the protein sequences with highest similarity are shown here along with a probability score. This gives an indication of the confidence with which the identification was made; lower numbers indicate higher probability. The proposed functions are suggested by the author on the basis of homology and in some cases by the further characterization of the encoded protein sequence (see text for details). The *wla* gene nomenclature (Fry *et al.*, 1998; Fry *et al.*, 2000b) is used for these putative glycosyltransferase genes until an unambiguous function for the encoded proteins can be established.
<table>
<thead>
<tr>
<th>Open reading frame no.</th>
<th>Proteins with highest similarity (probability score)</th>
<th>Putative function of gene product in C. jejuni</th>
<th>Proposed gene name in C. jejuni</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cj1120</td>
<td><em>Staphylococcus aureus</em> Cap1D (involved in capsule biosynthesis; (4e^{-06}))</td>
<td>Glycosyltransferase/NAD-dependent epimerase?</td>
<td><em>wlaL</em></td>
</tr>
<tr>
<td></td>
<td><em>Streptococcus suis</em> Cps9E (glycosyltransferase; (2e^{-91}))</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Staphylococcus aureus</em> Cap5D (involved in capsule biosynthesis; (1e^{-88}))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cj1124</td>
<td><em>Campylobacter hyoilei</em> RfbP (WbaP; (e^{-106}))</td>
<td>Galactosyltransferase</td>
<td><em>wlaH</em></td>
</tr>
<tr>
<td></td>
<td><em>Bacteroides fragilis</em> WcgN (putative WbaP homolog; (2e^{-64}))</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Bacillus subtilis</em> YvfC (involved in capsule biosynthesis (4e^{-62}))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cj1125</td>
<td><em>Campylobacter hyoilei</em> RfbF (galactosyltransferase; 0.0)</td>
<td>Galactosyltransferase</td>
<td><em>wlaG</em></td>
</tr>
<tr>
<td></td>
<td><em>Neisseria meningitidis</em> strain Z2491 PglA (glycosyltransferase; (1e^{-35}))</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Neisseria meningitidis</em> strain MC58 hypothetical protein NMB0218 ((2e^{-35}))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cj1127</td>
<td><em>Aquifex aeolicus</em> putative capsular polysaccharide protein ((9e^{-34}))</td>
<td>Glycosyltransferase</td>
<td><em>wlaE</em></td>
</tr>
<tr>
<td></td>
<td><em>Aquifex aeolicus</em> capsular polysaccharide protein M homolog ((2e^{-27}))</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Campylobacter jejuni</em> NCTC 11168 Cj1129 (putative glycosyltransferase; (4e^{-26}))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cj1128</td>
<td><em>Pyrococcus abyssi</em> hypothetical protein PAB0772 (putative glycosyltransferase; (5e^{-17}))</td>
<td>Glycosyltransferase</td>
<td><em>wlaD</em></td>
</tr>
<tr>
<td></td>
<td><em>Pyrococcus horikoshii</em> hypothetical protein PH0430 (putative glycosyltransferase; (8e^{-16}))</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Streptococcus suis</em> strain 6555 Cps11 (putative glycosyltransferase; (3e^{-15}))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open reading frame no.</td>
<td>Proteins with highest similarity (probability score)</td>
<td>Putative function of gene product</td>
<td>Proposed gene name in <em>C. jejuni</em></td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----------------------------------------------------</td>
<td>----------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Cj1129</td>
<td><em>Campylobacter jejuni</em> NCTC 11168 Cj1127 (4e⁻²⁶) <em>Yersinia enterocolitica</em> TrsD (putative galactosyltransferase; 5e⁻²⁴)</td>
<td>Glycosyltransferase</td>
<td>wlaC</td>
</tr>
<tr>
<td></td>
<td><em>Haemophilus influenzae</em> Rd HI0653 (putative glycosyltransferase; 2e⁻³⁶) <em>Klebsiella pneumoniae</em> WaaE (glycosyltransferase; 8e⁻³¹) <em>Serratia marcescens</em> WaaE (glycosyltransferase; 1e⁻³⁰)</td>
<td>Glycosyltransferase</td>
<td>wlaNA</td>
</tr>
<tr>
<td>Cj1136</td>
<td><em>Campylobacter jejuni</em> NCTC 11168 Cj1138 (putative glycosyltransferase; e⁻¹¹³) <em>Campylobacter jejuni</em> NCTC 11168 Cj1440 (putative glycosyltransferase; 1e⁻⁸⁹) <em>Streptococcus agalactiae</em> CpsK (putative glycosyltransferase; 2e⁻²⁴)</td>
<td>Glycosyltransferase</td>
<td>wlaNB</td>
</tr>
<tr>
<td>Cj1138</td>
<td><em>Campylobacter jejuni</em> NCTC 11168 Cj1139 (putative glycosyltransferase; e⁻¹¹³) <em>Campylobacter jejuni</em> NCTC 11168 Cj1140 (putative glycosyltransferase; e⁻¹⁰⁴) <em>Streptococcus agalactiae</em> CpsI (galactosyltransferase; 2e⁻²⁴)</td>
<td>Glycosyltransferase</td>
<td>wlaP</td>
</tr>
<tr>
<td>Cj1139</td>
<td><em>Campylobacter jejuni</em> NCTC 11168 Cj1146/7 (2e⁻⁴⁰) <em>Streptococcus suis</em> strain 6555 Cps1J (putative glycosyltransferase; 8e⁻²⁰) <em>Streptomyces coelicolor</em> A3 (2) SCE50.11 (putative glycosyltransferase; 8e⁻²⁰)</td>
<td>Glycosyltransferase</td>
<td>wlaQ</td>
</tr>
<tr>
<td>Cj1146/7</td>
<td><em>Campylobacter jejuni</em> NCTC 11168 Cj1139 (putative glycosyltransferase; 2e⁻⁴⁰) <em>Streptococcus suis</em> strain 10 Cps2K (putative glycosyltransferase; 1e⁻²¹) <em>Streptococcus suis</em> strain 6555 Cps1K (putative glycosyltransferase; 2e⁻²⁰)</td>
<td>Glycosyltransferase</td>
<td>wlaSA</td>
</tr>
</tbody>
</table>
Chapter 3: Identification of LPS biosynthesis genes in *C. jejuni* NCTC 11168

oligosaccharide moiety. This evidence would appear to argue against the Cj1120-encoded protein being a glycosyltransferase (Wood *et al.*, 1999). However, the lack of detectable change could be due to the use of polyclonal antiserum in the immunoblot that does not detect minor changes in the structure of the LPS. Further studies are needed to clarify the function of WlaL as either a glycosyltransferase or a NAD-dependent epimerase.

The protein product encoded by Cj1124 (*wlaH*) shows similarity to various WbaP proteins (Table 3.5) including that of *S. typhimurium* (probability score; 2e⁻¹⁵). In *Salmonella* and other Gram negatives, WbaP transfers the first galactose residue to undecaprenol phosphate during the Wzy-dependent biosynthesis of O-chain (Wang *et al.*, 1996; Wang & Reeves, 1994). The protein product of Cj1124 is therefore suggested as a galactosyltransferase. Functional studies in *C. jejuni* NCTC 11168 and NCTC 11828 also support this identification (Fry *et al.*, 2000b; Wood *et al.*, 1999). The potential role of WlaH in the biosynthesis of *C. jejuni* LOS will be discussed in more detail in subsequent sections.

Cj1125 (*wlaG*) is also proposed as a galactosyltransferase on the basis of homology (Table 3.5). Mutation of the homologous open reading frame in *C. jejuni* NCTC 11828 causes a reduction in the reactivity, but not mobility of the core oligosaccharide to homologous antiserum, suggesting a slight alteration to the core - a phenotype consistent with this identification (Fry, 1997). The predicted protein sequence from Cj1125 indicates the absence of any trans-membrane domains (data not shown) suggesting that WlaG is found in the cytosolic fraction of the cell.

The protein products of Cj1127, 1128, 1129, 1135, 1136, 1138 and 1146/7 are all similar to various glycosyltransferases, including the TrsD protein of *Yersinia enterocolitica* (Skurnik *et al.*, 1995), and to each other (Table 3.5). The protein products of Cj1128 and Cj1146/7 both contain one putative trans-membrane domain, whilst the others do not (data not shown). Mutation of the homologous gene to Cj1127 (*wlaE*) in *C. jejuni* NCTC 11828 causes a reduction in the reactivity of the core oligosaccharide to homologous antisera - a phenotype consistent with the identification of the gene as encoding a core glycosyltransferase (Fry, 1997).
The remaining potential core glycosyltransferase is encoded by Cj1139 (Table 3.5). The homologous gene to Cj1139 has been isolated from *C. jejuni* strain OH 4384 (named *cgtB*) and shown to have β-1,3-galactosyltransferase activity (Gilbert *et al.*, 2000). This function was assigned by over-expressing the CgtB protein in *E. coli* and then assaying transferase activity and the type of glycosidic bond formed using synthetic fluorescent oligosaccharides as acceptors (Gilbert *et al.*, 2000). Although the *C. jejuni* OH 4384 CgtB protein is similar to the *C. jejuni* NCTC 11168 Cj1139-encoded protein, particularly in the N-terminal region, the proteins only exhibit approximately 60% identity overall (Figure 3.5). This may suggest that the Cj1139-encoded protein has a similar but different catalytic activity to CgtB. Functional studies are needed to resolve this issue and to confirm the β-1,3-galactosyltransferase activity of the Cj1139-encoded protein.

3.5.6. Open reading frames in the cluster which encode proteins of unknown function

The remaining proteins encoded by open reading frames of the LOS biosynthesis cluster are difficult to categorize functionally. In some cases this is because their putative function, as judged by homology, is uncertain or seems unrelated to LOS biosynthesis, whilst in others functional studies have not revealed a definite role in LOS biosynthesis.

**Cj1119/wlaM**

WlaM is an integral membrane protein which shows similarity to *H. pylori* proteins of unknown function, and has a low level of similarity to the AcfB protein of *V. cholerae* (Table 3.6; Everiss *et al.*, 1994). AcfB contains a methyl-accepting chemotaxis motif (Alley *et al.*, 1992), and is thought to encode a environmental sensor/signal transducing protein which interacts with elements of the *V. cholerae* chemotaxis machinery (Everiss *et al.*, 1994). However, WlaM does not contain a methyl-accepting chemotaxis motif, and thus the function of this protein is unclear.
Figure 3.5: Alignment of the amino acid sequences of the Cj1139-encoded protein and *C. jejuni* OH 4384 CgtB. The CgtB protein has β-1,3-galactosyltransferase activity (Gilbert et al., 2000). The Cj1139-encoded protein is proposed by the author as a glycosyltransferase, with an as yet unknown specificity and activity. Identical residues are highlighted in red.
Table 3.6: Details of the *C. jejuni* NCTC 11168 genes found in the LOS biosynthesis cluster which encode proteins of unknown function. Open reading frame numbers are those provided in the *C. jejuni* NCTC 11168 genome annotation (http://www.sanger.ac.uk/Projects/C_jejuni/). Homology searches were carried out using the Genbank protein sequence data base - the protein sequences with highest similarity are shown here along with a probability score. This gives an indication of the confidence with which the identification was made: lower numbers indicate higher probability. The *wla* gene nomenclature (Fry *et al.*, 1998; Fry *et al.*, 2000b) is used for these genes until an unambiguous function for the encoded proteins can be established.
<table>
<thead>
<tr>
<th>Open reading frame no.</th>
<th>Proteins with highest similarity (probability score)</th>
<th>Putative function of gene product in LOS biosynthesis</th>
<th>Proposed gene name in C. jejuni</th>
</tr>
</thead>
</table>
| Cj1119                | *Helicobacter pylori* strain 26695 hypothetical protein HP0158 (5e⁻⁵²)  
*Helicobacter pylori* strain J99 hypothetical protein JHP0146 (1e⁻⁵¹)  
*Vibrio cholerae* AcfB (2e⁻⁰⁶) | Unknown | *wlaM* |
| Cj1121                | *Bacteroides fragilis* WegP (putative aminotransferase; 1e⁻⁸⁵)  
*Bacillus subtilis* YvfE (spore coat polysaccharide biosynthesis homolog; 1e⁻⁷⁵)  
*Legionella pneumophila* hypothetical protein ORF21 (7e⁻⁵⁷) | Unknown | *wlaK* |
| Cj1122                | *African Swine Fever virus* polyprotein pp220 precursor (0.19)  
*Toxoplasma gondii* ribosomal protein L6 (0.33) | Unknown | *wlaJ* |
| Cj1123                | *Neisseria meningitidis* PglB (3e⁻²⁸)  
*Caulobacter crescentus* LpsB (putative acyltransferase; 2e⁻²⁵)  
*Escherichia coli* NeuD (4e⁻¹⁹) | Unknown | *wlaL* |
| Cj1126                | *Methanobacterium thermoautotrophicum* MTH1906 (STT3 homolog; 1e⁻¹⁶)  
*Drosophila melanogaster* CG1518 gene product (9e⁻¹⁰)  
*Caenorhabditis elegans* STT3 sub-unit homolog (1e⁻⁰⁶) | Oligosaccharyl-transferase | *wlaF* |
<table>
<thead>
<tr>
<th>Open reading frame no.</th>
<th>Proteins with highest similarity (probability score)</th>
<th>Putative function of gene product in LOS biosynthesis</th>
<th>Proposed gene name in <em>C. jejuni</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cj1130</td>
<td><em>Helicobacter pylori</em> strain J99 ABC transporter/ATP-binding protein ($e^{-115}$)</td>
<td>Membrane transporter</td>
<td>wlaB</td>
</tr>
<tr>
<td></td>
<td><em>Helicobacter pylori</em> strain 26695 multidrug resistance protein ($e^{-113}$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Helicobacter pylori</em> strain 60190 multidrug resistance protein ($e^{-93}$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cj1132</td>
<td><em>Helicobacter pylori</em> strain J99 hypothetical protein JHP0628 ($e^{-61}$)</td>
<td>Unknown</td>
<td>wlaA</td>
</tr>
<tr>
<td></td>
<td><em>Neisseria meningitidis</em> strain MC58 hypothetical protein NMB1417 ($2e^{-37}$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Neisseria meningitidis</em> strain Z2491 hypothetical protein NMA1628 ($3e^{-36}$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cj1137</td>
<td><em>Streptococcus pneumoniae</em> capsular polysaccharide protein Cps14K ($5e^{-19}$)</td>
<td>Unknown</td>
<td>wlaO</td>
</tr>
<tr>
<td></td>
<td><em>Streptococcus pneumoniae</em> capsular polysaccharide protein Capp33F1 ($2e^{-15}$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Streptococcus pneumoniae</em> capsular polysaccharide protein Cap37I ($2e^{-15}$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cj1144</td>
<td><em>Arabidopsis thaliana</em> hypothetical protein F16M14.8 ($0.002$)</td>
<td>Unknown</td>
<td>wlaSC</td>
</tr>
<tr>
<td></td>
<td><em>Plasmodium falciparum</em> Rp16 ($1.1$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cj1145</td>
<td><em>Arabidopsis thaliana</em> hypothetical protein AT2G38680 ($0.97$)</td>
<td>Unknown</td>
<td>wlaSB</td>
</tr>
<tr>
<td></td>
<td><em>Helicobacter pylori</em> strain 26695 urease protein UreC ($3.8$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cj1152</td>
<td><em>Haemophilus influenzae</em> hypothetical protein HI0621 ($1e^{-34}$)</td>
<td>Unknown</td>
<td>wlaT</td>
</tr>
<tr>
<td></td>
<td><em>Helicobacter pylori</em> strain J99 hypothetical protein JHP0794 ($5e^{-24}$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Helicobacter pylori</em> strain 26695 hypothetical protein HP0860 ($3e^{-33}$)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 3: Identification of LPS biosynthesis genes in *C. jejuni* NCTC 11168

**Cj1121/wlaK**

WlaK shows similarity to various members of the DegT family of aminotransferases such as WcgP from *Bacteroides fragilis* (Table 3.6; Comstock et al., 1999). Similarity also exists to FlmB from *Caulobacter crescentus* which is involved in flagellum glycosylation (Leclerc et al., 1998), and RfbE from *Vibrio cholerae*, a putative perosamine synthetase involved in the biosynthesis of dideoxy-hexose sugars (Manning et al., 1995; Stroeher et al., 1995a). Mutation of WlaK in *C. jejuni* NCTC 11168 and 11828 alters the reactivity, but not mobility of the core oligosaccharide to homologous antiserum, suggesting a slight alteration to the core (Wood et al., 1999). The function of the WlaK protein is still unclear.

**Cj1122/wlaJ**

The WlaJ protein is a putative integral membrane protein with no significant homology to any other sequence in the Genbank database. Insertional inactivation of the *wlaJ* gene in *C. jejuni* NCTC 11168 does not affect core oligosaccharide reactivity to homologous antiserum (Wood et al., 1999). Preliminary biochemical studies have confirmed that *wlaJ* mutant strains produce LOS molecules with identical structure to the wild-type NCTC 11168 strain (Anthony Moran, personal communication). However, Northern hybridization results (Anne Wood, personal communication) show that the gene is transcribed implying that the *wlaJ* gene is expressed and so is likely to have a function. Further functional studies on the *wlaJ* gene are described in Chapter 5 of this study.

**Cj1123/wlal**

This protein is most similar to the PglB protein of *N. meningitidis* which is involved with the glycosylation of pilin (Table 3.6; Parkhill et al., 2000b). The N-terminal part of the Wlal protein is also similar, although to a lesser extent, to Gram negative LpxD and LpxA proteins, which are acyltransferases involved in lipid A biosynthesis. Like these proteins, Wlal contains hexapeptide motifs (data not shown; Vuorino et al., 1994; Dicker & Seetharam, 1992). Wlal also shows homology to the
NeuD protein of *E. coli* which is involved in synthesis of the α-2,8-linked polysialic acid capsule (Annunziato *et al.*, 1995). The precise function of the NeuD protein is not clear, but it is thought to be an acetyltransferase which may modify other proteins involved with the biosynthesis of sialic acid (Annunziato *et al.*, 1995). *WlaI* may therefore have a similar, but as yet undefined role, in the biosynthesis of sialic acid in *C. jejuni*. Mutation of the *wlaI* gene has been reported in *C. jejuni* NCTC 11828 (Fry, 1997). The resultant mutant strain produced core oligosaccharide which showed no alteration in the mobility or reactivity from the wild-type. However, the core oligosaccharide of *C. jejuni* NCTC 11828 is not thought to contain sialic acid (Anthony Moran, personal communication), making an evaluation of the role of *WlaI* in sialic acid biosynthesis impossible.

**Cj1126/wlaF**

The WlaF protein is an integral membrane protein containing eleven putative trans-membrane domains (data not shown), and has similarity to oligosaccharyl-transferases (Table 3.6). These proteins are involved in transferring oligosaccharides to proteins or other saccharide containing molecules (Yoshida *et al.*, 1995).

One of the galactosyltransferase genes identified in the cluster (*wlaH*) is a homolog of *E. coli* WbaP. This transfers the first galactose residue of an O-chain subunit onto the lipid carrier, undecaprenol phosphate (Wang *et al.*, 1996; Wang & Reeves, 1994). Most *C. jejuni* core oligosaccharides contain galactose as the next residue after the second L-glycero-D-manno-heptose (reviewed in Moran & Penner, 1999). It is therefore possible that the *C. jejuni* outer core is synthesized on a lipid carrier, independently of the inner-core lipid A structure. The function of the WlaF integral membrane protein, could therefore be the transfer of the completed outer core moiety to the inner-core lipid A molecule. This biosynthetic mechanism will be discussed in later sections.
Chapter 3: Identification of LPS biosynthesis genes in *C. jejuni* NCTC 11168

**Cj1130/wlab**

The WlaB protein contains six putative trans-membrane domains, and is homologous to various export proteins including multi-drug resistance proteins (Table 3.6). These bind to, and transport a variety of drugs out of the cell, thereby conferring resistance (Endicott & Ling, 1989). The WlaB protein also has lower levels of similarity to proteins involved with the export of haemolysin, for example, HlyB from *E. coli* (probability score $1e^{-32}$).

Analysis of the WlaB protein sequence reveals an ATP-binding site and a ABC transporter family signature (data not shown; Hung *et al.*, 1998; Higgins *et al.*, 1990; Walker *et al.*, 1982). These same motifs are present in MsbA which suggests that like MsbA, WlaB is involved with the transport of LOS across a membrane. The export of LPS/LOS involves two steps; the first is the transfer of the completed lipid A-core complex across the inner membrane which is catalyzed by the MsbA protein (Polissi & Georgopoulos, 1996). The second involves the transfer of the completed LPS/LOS molecule across the outer membrane. The mechanism by which this occurs in Gram negatives is not yet understood (Whitfield *et al.*, 1997; Whitfield & Valvano, 1993). Since a MsbA homolog has already been identified in *C. jejuni* NCTC 11168 (Cj0803), WlaB could perhaps be involved in the later transport step. Insertional inactivation of the *wlaB* gene has been attempted in *C. jejuni* NCTC 11828, but no mutants were obtained (Fry, 1997). This suggests that the WlaB protein is essential for cell viability. Mutation of the *msbA* gene in *E. coli* is also not possible because of the lethal phenotype of the resultant strains (Polissi & Georgopoulos, 1996).

**Cj1132/wlaA**

WlaA shows no similarity to any known LPS/LOS biosynthesis proteins (Table 3.6). Low levels of similarity are seen to various DNA polymerases (Argos, 1988), but the function of this protein in *C. jejuni* is unknown. Nevertheless, the *wlaA* gene is the first gene in an operon of up to fourteen genes which does contain known LPS/LOS biosynthesis genes such as *galE*. 

118
The remaining genes in the LOS cluster

The WlaO protein has low homology to several hypothetical proteins in polysaccharide biosynthesis clusters (Table 3.6). However, the functions of these proteins are not yet understood, so interpretations of the role of WlaO in \textit{C. jejuni} are equally problematical. The WlaO sequence contains no putative trans-membrane domains suggesting the protein is found in the cytosolic fraction of the cell.

The WlaSC, WlaSB and WlaT proteins have no significant homology to any other sequence in the Genbank database (Table 3.6). Without any functional data this makes evaluating the role of these proteins in \textit{C. jejuni} impossible. None of these proteins contain any putative trans-membrane domains suggesting that they are cytoplasmic proteins.

3.5.7. The role of LOS biosynthesis cluster genes in protein glycosylation

Recently, it has been reported that several proteins encoded by genes of the LOS biosynthesis cluster may also have a role in protein glycosylation (Szymanski \textit{et al.}, 1999). The genes \textit{wlaFGHIK} and \textit{L} were isolated from \textit{C. jejuni} 81-176 and individually insertionally inactivated (Szymanski \textit{et al.}, 1999). Analysis of the LPS molecules produced by the resultant mutants showed no alteration in core mobility or intensity using detection with homologous antiserum (Szymanski \textit{et al.}, 1999). Instead, there were major changes to the immuno-reactivity of various proteins, including flagellin, when immuno-detected using homologous antiserum. Furthermore, \textit{in vitro} deglycosylation of wild-type proteins resulted in changes in immuno-reactivity with the same set of proteins whose immuno-reactivity was affected by the \textit{wla} mutations (Szymanski \textit{et al.}, 1999).

Following this work, the homologous genes were isolated from another \textit{C. jejuni} strain - a HS:10 isolate from a patient with Miller Fisher Syndrome (Christine Szymanski, personal communication). Mutation of the \textit{wlaF, G} and \textit{I} genes in this strain resulted in changes to the immuno-reactivity of various proteins - the same phenotype as observed for the corresponding 81-176 mutants (Christine Szymanski,
personal communication). However, in addition, analysis of the core oligosaccharide molecules produced by these mutants revealed changes in mobility, indicating that in this strain, the WlaG, F and I proteins have a role in both core biosynthesis and protein glycosylation.

From this work, it therefore appears that some of the Wla proteins have a dual role in some strains. Mutation of these genes affects both protein glycosylation and core biosynthesis in some, whilst in other strains, mutation only affects protein glycosylation. Whether a similar dual role for these genes exists in C. jejuni NCTC 11168 has not yet been investigated, only the role in core biosynthesis has been confirmed (Wood et al., 1999).

3.6. O-chain biosynthesis and other saccharide biosynthesis genes

Attempts were also made to find homologs of E. coli O-chain specific biosynthesis genes in the C. jejuni NCTC 11168 genome. No wzx, wzy or wzz genes were found - this is not surprising since this strain is not thought to produce a O-chain. However, in addition to the genes that have already been described involved with the biosynthesis of the Kdo-lipid A and core moieties, analysis of the genome of C. jejuni NCTC 11168 also revealed genes thought to be involved in the biosynthesis of other saccharide containing structures (Parkhill et al., 2000a). As mentioned in section 3.5, in C. jejuni, flagellin, the major constituent of the flagellum, is glycosylated (Doig et al., 1996). The nature of these glycosylations is not yet clear, but sialic acid is certainly present (Doig et al., 1996). Two genes, ptmA (Cj1332) and ptmB (Cj1331), have already been described which are involved in the glycosylation process (Guerry et al., 1996). PtmB is highly similar to the E. coli NeuA protein, and inspection of the genome sequence reveals the presence of neuB and neuC homologs in close proximity to ptmAB (Linton et al., 2000; Parkhill et al., 2000a). Therefore, the LOS biosynthesis cluster neuABC genes are not the only sialic acid biosynthesis genes present in the C. jejuni genome. The ptmAB region of the genome also contains other open reading frames with
similarity to known sugar biosynthesis genes which are presumed to have a role in flagellin glycosylation (Parkhill et al., 2000a).

Another large cluster of saccharide biosynthesis genes are found flanked by open reading frames with similarity to E. coli kps genes (Karlyshev et al., 2000; Parkhill et al., 2000a). Kps proteins are involved in the transport of some E. coli capsules (reviewed in Whitfield & Roberts, 1999). The presence of biosynthetic genes flanked by transport genes suggests that C. jejuni NCTC 11168 is capable of synthesizing and exporting a capsular polysaccharide. The possibility that C. jejuni produces a capsule has already been eluded to in Chapter 1, and will be discussed in depth in Chapters 6 & 7.

3.7 Homo-polymeric G tracts and phase variation

One significant feature of the C. jejuni NCTC 11168 genome is the presence of homo-polymeric tracts, which are mainly, but not exclusively polyG:C tracts (Parkhill et al., 2000a). During the sequencing of the genome, it was noted that the lengths of some, but all, of these tracts varied between otherwise identical sequencing clones. This phenomenon has been shown not to be an artifact of the sequencing process suggesting the high frequency slipped-strand mispairing of DNA during replication (Moxon et al., 1994). Such a mechanism did not occur when the clones are present in E. coli, indicating that the process occurs in C. jejuni NCTC 11168 (Parkhill et al., 2000a).

The variations in the length of the polymeric tracts are thought to lead to alterations in protein translation, resulting in the ‘switching on or off’ of genes (Moxon et al., 1994). Most of these hyper-variable regions map to the three regions of saccharide biosynthesis genes; the LOS cluster, the capsule polysaccharide cluster and the flagellin modification cluster (Parkhill et al., 2000a). It is therefore thought that variations in the length of the polymeric tracts may lead to phase and/or antigenic variation of these surface structures (Parkhill et al., 2000a).

In the LOS cluster two homo-polymeric G tracts are evident. The first is present within the coding sequence of Cj1139 - which encodes a putative glycosyltransferase of
Chapter 3: Identification of LPS biosynthesis genes in *C. jejuni* NCTC 11168

303 amino acids in length. Variations in the number of residues in the tract could lead to truncation of the protein after approximately 120 amino acids (data not shown). The second polyG tract is present at the end of the coding sequence of Cj1145. Depending on the number of residues the tract changes by, variations could cause a slight truncation of the encoded protein or translationally fuse Cj1145 and Cj1144 - both genes of unknown function. The significance of these polyG tracts in LOS biosynthesis is not yet clear, but certainly merits further investigation.

3.8. Discussion

This chapter has detailed the bioinformatic analysis of the complete genome sequence of *C. jejuni* NCTC 11168 to identify genes involved with the biosynthesis of LOS. Some of the genes described here had been characterized prior to the sequencing of the genome (Fry *et al.*, 1998; Klena *et al.*, 1998; Wood *et al.*, 1999), and several have been characterized since (Fry *et al.*, 2000a; Gilbert *et al.*, 2000; Linton *et al.*, 2000; Szymanski *et al.*, 1999). The information from these studies has been incorporated into this chapter in an attempt to provide a comprehensive view of LOS biosynthesis in *C. jejuni* NCTC 11168.

As summarized in section 3.4.1, homologs of the *E. coli* genes involved with the biosynthesis of the Kdo-lipid A moiety can all been identified in the *C. jejuni* genome, with the exception of the *waaN* gene. Thus, the pathway of Kdo-lipid A biosynthesis in *C. jejuni* is presumed to be similar to that of other Gram negative bacteria, with the possible exception of the final acylation step.

The genome also contains homologs of genes involved with core oligosaccharide biosynthesis. In contrast to the genes involved with the biosynthesis of Kdo-lipid A which are mainly unclustered, these genes are found in a cluster of approximately 30 open reading frames (Figure 3.3). This cluster contains genes which encode enzymes necessary for the biosynthesis and transfer of L-glycero-D-manno-heptose residues (Klena *et al.*, 1998) and *galE* which is required for the conversion of glucose to galactose for incorporation into core oligosaccharide (Fry *et al.*, 2000a).
Sialic acid is a common constituent of C. jejuni core oligosaccharide molecules (Moran & Penner, 1999). In C. jejuni NCTC 11168, three genes that encode proteins necessary for the biosynthesis of sialic acid are also found in the LOS cluster (Linton et al., 2000). These three genes are found in an operon with the transferase responsible for the addition of the α-2,3-linked sialic acid residue to the core (Gilbert et al., 2000). The other genes in the cluster are all more difficult to identify on the basis of homology. Some have been classified as encoding glycosyltransferases, whilst others encode proteins with no obvious function in the biosynthesis of core oligosaccharide.

Now that the genes involved with the biosynthesis of the core have been identified, and some functionally characterized, a tentative model to provide testable predictions about core oligosaccharide biosynthesis in C. jejuni NCTC 11168 can be developed. Although the chemical structure of the NCTC 11168 core oligosaccharide has not been elucidated, preliminary analysis (Anthony Moran, personal communication) indicates that it contains identical residues to the HS:2 type strain core structure (Aspinall et al., 1993b). However, micro-heterogeneity is also apparent, and it is possible that additional sugar residues are present in sub-populations of NCTC 11168 core oligosaccharide, which are not present in the HS:2 type strain (Anthony Moran, personal communication). However, in the absence of firm chemical data to show the sequence, linkage, overall composition and possible micro-heterogeneity of the NCTC 11168 core, for the purposes of the proposed model, the NCTC 11168 core structure will be assumed to be identical to the HS:2 type strain structure.

The model assumes that the outer core is synthesized on a lipid carrier, independently of the inner core lipid A structure (Figure 3.6). Such a pattern is supported by the presence of the wlaH gene which probably encodes a transferase which adds the initial outer core galactose to a lipid carrier. Most C. jejuni strains analyzed to date produce a core oligosaccharide which contains galactose as the first residue of the outer core (reviewed in Moran & Penner, 1999). Such a model has also been suggested for Yersinia enterocolitica, where a similar Gal to lipid carrier transferase gene has been identified (Skurnik et al., 1995). Following the action of several other galactosyltransferases - possibly WlaG and WlaQ, and the sialic acid
Figure 3.6: A model for LOS biosynthesis in C. jejuni NCTC 11168. The inner core-lipid A and outer core moieties are synthesized separately before being transferred to the periplasm. They are then ligated, and exported to the outer face of the outer membrane. Protein names are indicated at their suggested stage of involvement. For the purposes of this diagram, the structure of the NCTC 11168 core oligosaccharide is presumed to be identical to that of the HS:2 type strain. Asterisks indicate proteins whose role has been confirmed by functional studies. GTS; glycosyltransferase.
transferase WlaR (Gilbert et al., 2000), the completed outer core is ‘flipped’ across the inner membrane to the periplasm. No functional data is available to identify the proteins involved in this transport step, but as an integral membrane protein the WlaM protein may take part.

Mutation of the wlaH gene has been attempted in C. jejuni NCTC 11168 but no mutants were obtained, suggesting that the gene is essential for the viability and growth of the cell (Wood et al., 1999). According to the model proposed here, mutation of wlaH would prevent the synthesis of the outer core perhaps leading to a build up of lipid carrier and/or galactose intermediates to toxic levels. However, mutants lacking WlaH were obtained by causing a large deletion which removed the 3’ end of wlaF, wlaG, wlaH, wlaI, wlaJ and the 5’ end of wlaK suggesting that one of these genes acts as a suppressor mutation for the inactivation of wlaH (Wood et al., 1999). Further studies are required to identify which gene mutation is required to obtain wlaH mutants. According to the proposed model, the core oligosaccharide molecules produced by such a large deletion mutant would lack the entire outer core due to the absence of the initiating galactosyltransferase (WlaH), a second putative galactosyltransferase (WlaG) and the oligosaccharyltransferase (WlaF). Preliminary structural analysis of the core oligosaccharide molecules produced by this mutant indicates the absence of several sugar residues consistent with loss of the outer core (Anthony Moran, personal communication).

The inner core is probably synthesized on the face of the inner membrane by the activity of several sugar transferases including WaaC (Klena et al., 1998) and WaaF which add the two L-glycero-D-manno-heptose residues, and two other transferases which add two glucose residues - the identity of the enzymes which catalyze these steps are not clear in the absence of functional data. The identity of the enzyme responsible for the addition of the phosphoethanolamine residue to the core is also unclear. One intriguing point about the LOS cluster is that there appears to be eleven putative glycosyltransferase genes present, but only five corresponding sugars in the core oligosaccharide of C. jejuni NCTC 11168 (assuming the structure is identical to the HS:2 serostrain). This would indicate that either the identification of some of these
genes is incorrect, or that some of the genes encode transferases which are involved with the biosynthesis of other structures.

Following the completion of the inner core, it is transferred to the periplasmic face of the membrane by the action of the MsbA protein. The outer and inner core are then ligated, possibly by the WlaF protein, which has similarity to other oligosaccharyltransferases. The completed LOS molecule is then exported out of the cell, perhaps by the action of the WlaB protein (Figure 3.6).

The proposed model is based on the putative identification of genes in the cluster by homology, and the functional characterization of some genes. However, mutational analysis and thorough analysis of the resulting LOS phenotypes has not been, or has only been partially completed for many of the critical steps of the pathway. Functional studies are required to identity which of the eleven putative glycosyltransferases contribute to the LOS structure and the sugars transferred by each enzyme. The study by Gilbert et al. (2000) showed how the approach of over-expressing putative glycosyltransferases in *E. coli* and using synthetic fluorescent oligosaccharides as acceptors can be used to determine the sugar residue transferred, and the type of linkage created (Gilbert et al., 2000). Also, the roles of the WlaM, MsbA, WlaF and WlaB proteins in the transport and transfer stages needs to be rigorously defined. The systematic mutation of the LOS cluster genes, and the careful evaluation of resultant mutants is a vital pre-requisite in clearly defining the role of genes in LOS biosynthesis. In addition, complementation studies can be used to confirm the role of genes. The *Campylobacter waaC, galE* and *neuB1* genes have already been used in such studies (Fry et al., 2000a; Klena et al., 1998; Linton et al., 2000). This approach could also be used for the *neuA1* and *neuC1* genes involved with sialic acid biosynthesis. Chapter 6 of this study details the use of complementation analysis to confirm the function of the *waaF* gene.

The role of some of the LOS biosynthesis genes in protein glycosylation also needs to be investigated further. The results of the work by Szymanski and co-workers (personal communication) on the HS:10 isolate, indicates a role for some of the *wla* genes in both protein glycosylation and core biosynthesis. Such a dual role would be
possible if glycosylation oligosaccharides are synthesized in a similar manner, utilizing some of the same proteins, as the *C. jejuni* outer core. However, if some of the *wla* genes do have a dual function, the phenotype of the 81-176 mutants in which protein glycosylation but not core biosynthesis were affected is curious (Szymanski *et al.*, 1999). One of the 81-176 strains reported by Szymanski *et al.* (1999) contained an insertional inactivated *wlaH* gene. However, no individual *wlaH* mutants have been obtained in *C. jejuni* NCTC 11828 or 11168 (Wood *et al*., 1999; Fry, 1997). This may imply a differing role of the WlaH protein in 81-176. In *C. jejuni* NCTC 11168, the WlaH protein is suggested as a galactosyltransferase which initiates outer core synthesis. *C. jejuni* 81-176 is from the HS:23/36 serogroup, and examination of the core oligosaccharides synthesized by the HS:23/36 serostrains reveals that the first outer core sugar of these structures is glucose, not galactose (Aspinall *et al.*, 1993a). Therefore, it is possible that in 81-176 another enzyme may initiate outer core synthesis, but WlaH still has a role in protein glycosylation. This may explain why mutation of *wlaH* in 81-176 is possible, and it does not appear to affect core biosynthesis. The apparent conservation of genes, but the differing role of these genes in different strains needs to be examined further. The next chapter of this study investigates the conservation of LOS cluster genes in different strains of *C. jejuni*. 
Chapter 4: Identification and characterization of gene content polymorphisms in \textit{C. jejuni} LOS clusters

4.1. Introduction

The previous chapter detailed the analysis of the complete genome sequence of \textit{C. jejuni} NCTC 11168 to identify genes involved in the biosynthesis of LOS/LPS. One significant finding of this work was the identification of a cluster of approximately thirty genes thought to be involved in the biosynthesis of the core moiety. Proteins encoded by genes in this cluster include those involved with the synthesis and transfer of sugar residues and transport of LOS/LPS structures across membranes. In \textit{E. coli} and \textit{Salmonella}, most of the genes encoding for the biosynthesis of core oligosaccharide are also clustered together (reviewed in Schnaitman & Klena, 1993; Heinrichs \textit{et al.}, 1998c). Differing gene content in this cluster is responsible for the five different \textit{E. coli}, and two different \textit{Salmonella} core types which have been characterized. The aim of the work presented in this chapter was to determine the gene content of the LOS cluster in different isolates of \textit{C. jejuni}. Any polymorphic regions identified could then be characterized, in an attempt to identify genes responsible for the core structures produced by a range of \textit{C. jejuni} strains. The initial analysis of the LOS cluster in different strains of \textit{C. jejuni} was achieved using a simple PCR approach (described in section 2.14.1) involving amplification from open reading frame to open reading frame. The presence or absence and size of the resultant amplification product can be interpreted to indicate any deletions, insertions or rearrangements of genes. Primers were designed from the \textit{C. jejuni} NCTC 11168 genome sequence and are detailed in Appendix 1. The analysis was initially applied to NCTC 11168 to validate the integrity of the PCR-based approach and to confirm the structure of the LOS cluster as detailed in the \textit{C. jejuni} NCTC 11168 genome. A total of 37 \textit{C. jejuni} strains (excluding NCTC 11168) were analyzed using this technique - these are listed in Table 2.1.

To facilitate interpretation, the majority of the results from this study will be
presented in tabulated form, however as an example, Figure 4.1 shows the amplification products from *C. jejuni* NCTC 11168 using a reverse primer which anneals to *wlaL* and forward primers to each of the following six genes (*wlaK, wlaJ, wlaL, wlaH, wlaG* and *wlaF*; primer positions and orientation are also indicated in Figure 4.2). This analysis yields six products of increasing size corresponding to the presence and organization of the seven genes in NCTC 11168. In contrast, Figure 4.3 shows the amplification products from strain NCTC 11351. The absence of a band in lane 5 indicates the absence of *wlaJ* in this strain (Wood *et al.*, 1999; shown in Figure 4.4). In addition, the sizes of the amplification products in lanes 1 - 4 are approximately 600bp smaller than the corresponding bands in NCTC 11168 (Figure 4.1) again correlating to the absence of *wlaJ*. The work relating to the presence or absence of the *wlaJ* gene forms part of the publication by Wood *et al.* (1999).

This PCR approach was subsequently applied to all of the strains, across the whole of the cluster, and led to the identification of three areas of gene content polymorphism. These three regions of the cluster will be discussed in turn, along with details of the further characterization of each polymorphism.

### 4.2. Gene content polymorphisms between *wlaM* and *waaM*

In the 38 *C. jejuni* strains tested, all the genes between *wlaM* and *waaM* were conserved in the same arrangement as the genome strain, with the exception of the *wlaJ* gene (Table 4.1; Figure 4.5). The *wlaJ* gene was present in NCTC 11168, and 17 other strains of varying HS serotype. The remaining 20 strains, including NCTC 11828, NCTC 11351 and 81-176 (Fry *et al.*, 1998; Szymanski *et al.*, 1999; Wood *et al.*, 1999), lacked the *wlaJ* gene, with the *wlaL* and *wlaK* genes being contiguous in the cluster (Figure 4.4). The absence of *wlaJ* in these strains was confirmed by Southern hybridization (data not shown). Serotyping of the strains (Table 4.1) revealed that the presence or absence of the *wlaJ* gene did not correlate with serotype - strains of HS:1, HS:2, HS:5 and HS:11 serotype are represented in both *wlaJ*-present and *wlaJ*-absent groups. Furthermore, LOS and LPS producing strains are found in both groups, so the
Figure 4.1: PCR analysis of strain NCTC 11168 to confirm the presence and orientation of the \textit{wlaL} - \textit{wlaF} genes. PCR products were separated by horizontal gel electrophoresis (0.8\% agarose). Marker sizes, in kilobase pairs, are shown on the right hand side of the figure. Lane 1. WLALR1/WLAF1 primer combination produces a 6.2 kb product; Lane 2. WLALR1/WLAGF1 yields a product of 4.2 kb; Lane 3, WLALR1/WLAHF1 produces a fragment of 3.1 kb; Lane 4, WLALR1/WLAF1 yields a fragment of 2.5 kb; Lane 5, WLALR1/WLAF1 yields a 1.9 kb product; Lane 6, WLALR1/WLAF1 primer combination produces a product of 1.2 kb; Lane 7. \textit{\lambda. Hind} III. $\phi/x Hae$III marker.

Figure 4.2: Schematic diagram showing the gene content and primer positions for the \textit{wlaL} - \textit{wlaF} region of strain NCTC 11168. Open reading frames and the direction of their transcription are indicated by black arrows. Primer positions and their orientation are indicated by blue arrows. This diagram is not shown to scale.
Chapter 4: Gene content polymorphisms in *C. jejuni* LOS clusters

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Gene name</th>
</tr>
</thead>
<tbody>
<tr>
<td>WlALR1</td>
<td>wlaL</td>
</tr>
<tr>
<td>WlAKF1</td>
<td>wlaK</td>
</tr>
<tr>
<td>WlAJF1</td>
<td>wlaJ</td>
</tr>
<tr>
<td>WlAF1</td>
<td>wlaI</td>
</tr>
<tr>
<td>WlAHF1</td>
<td>wlaH</td>
</tr>
<tr>
<td>WLAGF1</td>
<td>wlaG</td>
</tr>
<tr>
<td>WALFF1</td>
<td>wlaF</td>
</tr>
</tbody>
</table>
Figure 4.3: PCR analysis of strain NCTC 11351 to confirm the presence and orientation of the \textit{wlaL - wlaF} genes. The \textit{wlaJ} gene is absent in the NCTC 11351 LOS cluster, as indicated by the lack of an amplification product in lane 5, and because the fragments in lanes 1 - 4 are approximately 600 bp smaller than the corresponding bands in NCTC 11168. PCR products were separated by horizontal gel electrophoresis (0.8% agarose). Marker sizes, in kilobase pairs, are shown on the right hand side of the figure. Lane 1. WLALR1/WLAFF1 primer combination produces a 5.6 kb; Lane 2, WLALR1/WLAGF1 combination yields a product of 3.6 kb; Lane 3, WLALR1/WLAHF1 produces a fragment of 2.5 kb; Lane 4. WLALR1/WLAIF1 produces a fragment of 1.9 kb; Lane 5, WLALR1/WLAJF1 yields no product confirming the absence of the \textit{wlaJ} gene in this strain; Lane 6. WLALR1/WLAKF1 primer combination produces a product of 1.2 kb; Lane 7. \textit{Hind} III, \textit{phiX HaeIII} marker.

Figure 4.4: Schematic diagram showing the gene content and primer positions for the \textit{wlaL - wlaF} region of strain NCTC 11351. Open reading frames and the direction of their transcription are indicated by black arrows. Primer positions and their orientation are indicated by blue arrows. This diagram is not shown to scale.
Chapter 4: Gene content polymorphisms in *C. jejuni* LOS clusters

![Genes and Primers Diagram]

**Primer name**

WLALR1  WLAKF1  WLAIF1  WLAHF1  WLAGF1  WLAFF1

**Gene name**
wlaL  wlaK  wlaI  wlaH  wlaG  wlaF
Table 4.1: The distribution of the \textit{wlaJ} gene in 38 strains of \textit{C. jejuni}. The other genes between \textit{waaM} and \textit{wlaM} are present in all strains in the same arrangement as in NCTC 11168 (data not shown). Strain sources were shown in Table 2.1. Asterisks indicate a strain isolated from a patient with Guillain Barré Syndrome. The predicted phenotype (LOS/LPS) is based on structural data from several \textit{C. jejuni} strains which suggested that serotype HS:1, HS:2, HS:3 and HS:41 strains produce LOS, whilst strains of serotype HS:4, HS:19, HS:23 and HS:36 produce an O-chain (LPS) - this is described in section 1.9. NK, not known. UT, untypable.
### Chapter 4: Gene content polymorphisms in *C. jejuni* LOS clusters

<table>
<thead>
<tr>
<th><em>C. jejuni</em> strain</th>
<th>HS serotype</th>
<th>Predicted phenotype (LOS/LPS)</th>
<th>Presence of <em>wlaJ</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>O:1</td>
<td>1</td>
<td>LOS</td>
<td>+</td>
</tr>
<tr>
<td>G1*</td>
<td>1</td>
<td>LOS</td>
<td>+</td>
</tr>
<tr>
<td>G3*</td>
<td>1</td>
<td>LOS</td>
<td>+</td>
</tr>
<tr>
<td>NCTC 11168</td>
<td>2</td>
<td>LOS</td>
<td>+</td>
</tr>
<tr>
<td>1915/91</td>
<td>2</td>
<td>LOS</td>
<td>+</td>
</tr>
<tr>
<td>2561/90</td>
<td>2</td>
<td>LOS</td>
<td>+</td>
</tr>
<tr>
<td>K85</td>
<td>2</td>
<td>LOS</td>
<td>+</td>
</tr>
<tr>
<td>2608/90</td>
<td>4</td>
<td>LPS</td>
<td>+</td>
</tr>
<tr>
<td>2F 8</td>
<td>5</td>
<td>NK</td>
<td>+</td>
</tr>
<tr>
<td>18F 5</td>
<td>11</td>
<td>NK</td>
<td>+</td>
</tr>
<tr>
<td>P19</td>
<td>19</td>
<td>LPS</td>
<td>+</td>
</tr>
<tr>
<td>2F 68</td>
<td>50</td>
<td>NK</td>
<td>+</td>
</tr>
<tr>
<td>N82</td>
<td>UT</td>
<td>NK</td>
<td>+</td>
</tr>
<tr>
<td>G2*</td>
<td>UT</td>
<td>NK</td>
<td>+</td>
</tr>
<tr>
<td>G4*</td>
<td>UT</td>
<td>NK</td>
<td>+</td>
</tr>
<tr>
<td>F3*</td>
<td>UT</td>
<td>NK</td>
<td>+</td>
</tr>
<tr>
<td>27F 155</td>
<td>NK</td>
<td>NK</td>
<td>+</td>
</tr>
<tr>
<td>4F 182</td>
<td>NK</td>
<td>NK</td>
<td>+</td>
</tr>
<tr>
<td>306/90</td>
<td>1</td>
<td>LOS</td>
<td>-</td>
</tr>
<tr>
<td>2523/90</td>
<td>2</td>
<td>LOS</td>
<td>-</td>
</tr>
<tr>
<td>O:3</td>
<td>3</td>
<td>LOS</td>
<td>-</td>
</tr>
<tr>
<td>8F 169</td>
<td>5</td>
<td>NK</td>
<td>-</td>
</tr>
<tr>
<td>NCTC 11828</td>
<td>6</td>
<td>LPS</td>
<td>-</td>
</tr>
<tr>
<td>O73</td>
<td>11</td>
<td>NK</td>
<td>-</td>
</tr>
<tr>
<td>21F 185</td>
<td>15</td>
<td>NK</td>
<td>-</td>
</tr>
<tr>
<td>2258/90</td>
<td>23</td>
<td>LPS</td>
<td>-</td>
</tr>
<tr>
<td>NCTC 11351</td>
<td>23</td>
<td>LPS</td>
<td>-</td>
</tr>
<tr>
<td>81-176</td>
<td>23/36</td>
<td>LPS</td>
<td>-</td>
</tr>
<tr>
<td>J75</td>
<td>28</td>
<td>NK</td>
<td>-</td>
</tr>
<tr>
<td>16971.94GSH*</td>
<td>41</td>
<td>LOS</td>
<td>-</td>
</tr>
<tr>
<td>28134.94GSH*</td>
<td>41</td>
<td>LOS</td>
<td>-</td>
</tr>
<tr>
<td>260.94.RXH*</td>
<td>41</td>
<td>LOS</td>
<td>-</td>
</tr>
<tr>
<td>176.83</td>
<td>41</td>
<td>LOS</td>
<td>-</td>
</tr>
<tr>
<td>2F 90</td>
<td>UT</td>
<td>NK</td>
<td>-</td>
</tr>
<tr>
<td>E206</td>
<td>UT</td>
<td>NK</td>
<td>-</td>
</tr>
<tr>
<td>21F 242</td>
<td>NK</td>
<td>NK</td>
<td>-</td>
</tr>
<tr>
<td>4F 225</td>
<td>NK</td>
<td>NK</td>
<td>-</td>
</tr>
<tr>
<td>B404</td>
<td>NK</td>
<td>NK</td>
<td>-</td>
</tr>
</tbody>
</table>
Figure 4.5: Schematic diagram showing the conservation of genes in the \textit{wlaM} - \textit{waaM} region of the LOS cluster in 38 \textit{C. jejuni} strains. Open reading frames and the direction of their transcription are indicated by arrows. Proposed gene names are also given. Black text and arrows indicate genes conserved in all 38 (including NCTC 11168) strains of \textit{C. jejuni} tested. Red text and arrows indicate genes absent in some strains - in this case \textit{wlaJ}. This gene is absent in 20 out of 38 strains of \textit{C. jejuni}, with the \textit{wlaI} and \textit{wlaK} genes being contiguous in the cluster. This diagram is not shown to scale.
production of an O-chain-like polymer does not correlate with the presence of the *wlaJ* gene. Finally, strains isolated from patients with Guillain Barré Syndrome were also present in both groups, suggesting there is no correlation between the presence of *wlaJ* and the ability of a *C. jejuni* strain to cause GBS. The function of the protein encoded by the *wlaJ* gene in NCTC 11168 as judged by homology is uncertain (section 3.5.6), but is investigated further in Chapter 5.

4.3. Identification and characterization of gene content polymorphisms between *waaM* and *wlaSA*

PCR analysis of the *C. jejuni* strains across the *waaM* to *wlaSA* region of the LOS cluster allowed the classification of the strains into two groups (summarized in Table 4.2; Figure 4.6). The first group, consisting of ten strains, contained an identical gene content and arrangement to the genome strain between *waaM* and *wlaSA*. A further 23 strains, which comprised the second group, seemed to have a radically different gene content in this central area of the LOS cluster. In these strains no amplification products were obtained using primers which anneal to the *wlaNB, O, P, Q, R, neuB1, C1, A1, wlaSC* and *wlaSB* genes. However, all these strains did yield amplification products using primers which anneal to the *wlaNA* and *wlaSA* genes, with the exception of NCTC 11828. This strain had, in addition to the lack of the ten genes listed above, no *wlaNA* gene, as judged by PCR analysis (data not shown). Table 4.2 summarizes the above information along with the HS serotype and LOS/LPS phenotype of each strain. No obvious correlation between these characteristics and the strains in each group is evident. Strains from serotype groups HS:1, HS:2 and HS:23 are evident in both groups, as are LOS and LPS producing strains. Finally, strains isolated from patients with Guillain Barré Syndrome were also present in both groups, suggesting there is no direct correlation between the gene content of the *waaM* - *wlaSA* region and the ability of a *C. jejuni* strain to cause GBS.

A second set of amplification primers were subsequently designed to the genes in *C. jejuni* NCTC 11168 which were missing in the second group of strains
Table 4.2: The distribution of the \textit{wlaNA} - \textit{wlaSB} genes in 38 strains of \textit{C. jejuni}.

Strain sources were shown in Table 2.1. Asterisks indicate a strain isolated from a patient with Guillan Barré Syndrome. The predicted phenotype (LOS/LPS) is based on structural data from several \textit{C. jejuni} strains which suggested that serotype HS:1, HS:2, HS:3 and HS:41 strains produce LOS, whilst strains of serotype HS:4, HS:19, HS:23 and HS:36 produce an O-chain (LPS) - this is described in section 1.9. NK, not known, UT, untypable, ND, not done.
### Chapter 4: Gene content polymorphisms in *C. jejuni* LOS clusters

<table>
<thead>
<tr>
<th><em>C. jejuni</em> strain</th>
<th>HS serotype</th>
<th>Predicted phenotype (LOS/LPS)</th>
<th>Presence of wlaNA</th>
<th>Presence of wlaNB-wlaSB</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1*</td>
<td>1</td>
<td>LOS</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>G3*</td>
<td>1</td>
<td>LOS</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2561/90</td>
<td>2</td>
<td>LOS</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>NCTC 11168</td>
<td>2</td>
<td>LOS</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>K85</td>
<td>2</td>
<td>LOS</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>81-176</td>
<td>23/36</td>
<td>LPS</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>N82</td>
<td>UT</td>
<td>NK</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>G4*</td>
<td>UT</td>
<td>NK</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>4F 182</td>
<td>NK</td>
<td>NK</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>27F 155</td>
<td>NK</td>
<td>NK</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>O:1</td>
<td>1</td>
<td>LOS</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>2523/90</td>
<td>2</td>
<td>LOS</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>O:3</td>
<td>3</td>
<td>LOS</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>2608/90</td>
<td>4</td>
<td>LPS</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>8F 169</td>
<td>5</td>
<td>NK</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>O73</td>
<td>11</td>
<td>NK</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>21F 185</td>
<td>15</td>
<td>NK</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>P19</td>
<td>19</td>
<td>LPS</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>2258/90</td>
<td>23</td>
<td>LPS</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>NCTC 11351</td>
<td>23</td>
<td>LPS</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>J75</td>
<td>28</td>
<td>NK</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>16971.94GSH*</td>
<td>41</td>
<td>LOS</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>28134.94GSH*</td>
<td>41</td>
<td>LOS</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>260.94.RXH*</td>
<td>41</td>
<td>LOS</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>176.83</td>
<td>41</td>
<td>LOS</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>2F 68</td>
<td>50</td>
<td>NK</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>2F 90</td>
<td>UT</td>
<td>NK</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>E206</td>
<td>UT</td>
<td>NK</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>G2*</td>
<td>UT</td>
<td>NK</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>F3*</td>
<td>UT</td>
<td>NK</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>4F 225</td>
<td>NK</td>
<td>NK</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>B404</td>
<td>NK</td>
<td>NK</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>NCTC 11828</td>
<td>6</td>
<td>LPS</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>306/90</td>
<td>1</td>
<td>LOS</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>1915/91</td>
<td>2</td>
<td>LOS</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>2F 8</td>
<td>5</td>
<td>NK</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>18F 5</td>
<td>11</td>
<td>NK</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>21F 242</td>
<td>NK</td>
<td>NK</td>
<td>ND</td>
<td>ND</td>
</tr>
</tbody>
</table>
Figure 4.6: Schematic diagram showing the conservation of genes in the \textit{waaM} - \textit{wlaSA} region of the LOS cluster in 38 \textit{C. jejuni} strains. Open reading frames and the direction of their transcription are indicated by arrows. Proposed gene names are also given. Black text and arrows indicate genes conserved in all 38 (including NCTC 11168) strains of \textit{C. jejuni} tested. Red text and arrows indicate genes apparently absent in some strains, as judged by PCR (details in text). This diagram is not shown to scale.
Chapter 4: Gene content polymorphisms in C. jejuni LOS clusters

\(i.e. \ wlaNA - wlaSB\). This was undertaken to confirm that nucleotide sequence polymorphisms at the primer annealing sites in these 23 strains were not responsible for the absence of amplification products. No amplification products were obtained using this second set of primers either (data not shown). This suggested that if the genes were present they had a radically different nucleotide sequence or overall gene organization to NCTC 11168. Alternatively, it was possible that the genes were not present at all in these strains. This would mean that the \(wlaNA\) and \(wlaSA\) genes would be contiguous in the cluster (except in NCTC 11828 which also lacks the \(wlaNA\) gene). To test this hypothesis, attempts were made to amplify from \(wlaNA\) to \(wlaSA\), using a standard PCR methodology (described in section 2.14.1), in the 22 strains missing the \(wlaNB - wlaSB\) genes. No amplification products were obtained from any of the strains indicating that the two genes are not contiguous (data not shown). This data supported an additional possibility - that the ‘missing’ genes are replaced with other, as yet unidentified genes. To test this hypothesis it was decided to try and characterize the gene content of the \(waaC - waaF\) region in a representative strain - \(C. jejuni\) NCTC 11828. This strain was chosen because the cluster from \(wlaM - waaC\) had already been described and partly characterized (Fry et al., 1998). If the gene content of the \(waaC\) to \(waaF\) region could also be investigated it would mean that the majority of the cluster would therefore have been characterized in two strains of \(C. jejuni\) - the genome strain, and also NCTC 11828. In addition, \(C. jejuni\) NCTC 11828 is highly amenable to genetic manipulation such as natural and electro-transformation. This is in contrast to some strains of \(C. jejuni\), and would allow the subsequent construction of specific mutants should the need arise. \(C. jejuni\) NCTC 11828 is from HS group 6, and produces an O-chain like repeating unit.

4.3.1. The \(waaC - waaF\) region of \(C. jejuni\) NCTC 11828

The \(waaC - waaF\) region was amplified from NCTC 11828 by the Expand™ Long Template PCR methodology described in section 2.14.3 using primers LRF1 and LRR1 (Appendix 1). The resulting 15.1 kb fragment was then purified by horizontal gel
electrophoresis (section 2.8) followed by extraction using the NUCLEOTRAP kit (section 2.9.1). Gel electrophoresis revealed that the amplified fragment was larger than the corresponding fragment from C. jejuni NCTC 11168 (Figure 4.7), confirming the hypothesis that the 'missing genes' are replaced in NCTC 11828 and that waaM and wlaSA were not contiguous. The sequence of the long range PCR product was then obtained using a combination of primer walking and sub-cloning of Bgl II fragments, and is shown in Appendix 3. Subsequent analysis of this sequence using the Gene Runner computer program revealed the presence of 15 novel open-reading frames. This was in addition to homologs of the NCTC 11168 waaM and wlaSA genes and the two incomplete genes, waaF and waaC. A schematic diagram showing the gene organization of the long range product in comparison to the same region of the NCTC 11168 cluster is shown in Figure 4.8. In the absence of experimental data to confirm the function of these genes, wla designations have been used to identify the NCTC 11828 open reading frames (Fry et al., 1998; Fry et al., 2000b).

In order to assign a putative function to the proteins encoded by these genes, each protein sequence was used in a BLAST search against the Genbank protein sequence database. This data is shown in Table 4.3, and will be briefly discussed in the following sections. Detailed information about promoter sequences, ribosome binding sites, and translation start sites will not be given here - the analysis will concentrate on assigning a putative function to the proteins based on their similarity to other LOS/LPS biosynthesis enzymes.

**wlaUA**

The protein sequence encoded by this open-reading frame shows similarity to β 1,4 glucosyltransferases such as the LgtF proteins of *Aquifex aeolicus* (Deckert et al., 1998) and *Neisseria meningitidis* (Kahler et al., 1996). In *N. meningitidis*, LgtF is responsible for the addition of a glucose substitution to a L-glycero-D-manno-heptose residue in the inner core moiety (Kahler et al., 1996). Therefore, WlaUD is suggested as having a similar function in *C. jejuni* NCTC 11828. Unfortunately, the chemical structure of the core oligosaccharide of *C. jejuni* NCTC 11828 has not been elucidated,
Chapter 4: Gene content polymorphisms in *C. jejuni* LOS clusters

Figure 4.7: Comparison of the long range PCR products obtained from *C. jejuni* NCTC 11168 and *C. jejuni* NCTC 11828. Products were obtained using primers LRF1 and LRR1. PCR products were separated by horizontal gel electrophoresis (0.5% agarose). Marker sizes, in kilobase pairs, are shown on the left hand side of the figure. Lanes 1 & 4, *λ* Hind III, φ/x Hae III marker; Lane 2, NCTC 11168 yields a fragment of 14 kb; Lane 3, NCTC 11828 yields a fragment of 15.1 kb so confirming the differing gene content in this strain between *waaC* and *waaF*. 
Figure 4.8: Schematic diagram of the waaC - waaF regions of C. jejuni NCTC 11168 and NCTC 11828. Open reading frames and the direction of their transcription are indicated by arrows. The proposed function of each gene product is indicated by colour coding. The gene names waaC, waaM, wlaSA and waaF are suggested for the NCTC 11828 genes which show high levels of similarity to the corresponding NCTC 11168 genes, and are therefore assumed to have an identical function. Where no homology exists between the two strains, a name is proposed based on the existing wla nomenclature (Fry et al., 1998; Fry et al., 2000b). This diagram is not shown to scale. Stars indicate genes containing homo-polymeric G/C tracts.
Chapter 4: Gene content polymorphisms in C. jejuni LOS clusters

Functional categories:
- Green: Biosynthesis of lipid A
- Red: Biosynthesis and transfer of ADP-L-glycero-D-manno-heptose residues
- Yellow: Biosynthesis and transfer of N-acetyl neuraminic acid (sialic acid)
- Purple: Putative glycosyltransferases
- Black: Other functions

C. jejuni NCTC 11168

Proposed gene name

C. jejuni NCTC 11828

Proposed gene name
Table 4.3: Details of the novel genes identified in the \textit{waaC} - \textit{waaF} region of \textit{C. jejuni} NCTC 11828. Homology searches were carried out using the Genbank protein sequence data base - the three protein sequences with highest similarity are shown here along with a probability score. This gives an indication of the confidence with which the identification was made: lower numbers indicate higher probability. The proposed functions are suggested on the basis of homology (see text for details).
<table>
<thead>
<tr>
<th>Open reading frame</th>
<th>Proteins with highest similarity (probability score)</th>
<th>Putative function of gene product in <em>C. jejuni</em> NCTC 11828</th>
</tr>
</thead>
</table>
| wlaUA              | *Aquifex aeolicus* LgtF (β-1,4 glucosyltransferase; 2e⁻⁴³)  
*Campylobacter jejuni* NCTC 11168 WlaNA (putative glucosyltransferase; 1e⁻⁴⁰)  
*Neisseria meningitidis* LgtF (β-1,4 glucosyltransferase; 3e⁻³⁴) | Glucosyltransferase                                      |
| wlaUB              | *Escherichia coli* RmlA (e⁻¹⁰¹)  
*Enterococcus faecalis* strain OG1RF putative RmlA (OrfDe6; e⁻¹⁰⁰)  
*Salmonella enterica* RmlA (7e⁻⁹⁹) | Glucose-1-phosphate thymidyltransferase                  |
| wlaUC              | *Legionella pneumophila* RmlB (1e⁻⁹⁰)  
*Xanthomonas campestris* pv. campestris RmlB homolog (1e⁻⁹⁰)  
*Leptospira interrogans* serovar copenhageni RmlB (1e⁻⁹⁰) | dTDP-glucose-4,6-dehydratase                            |
| wlaUD              | *Vibrio anguillarum* Orf15x4 (unknown; 5e⁻²⁴)  
*Leptospira borgpetersenii* OrfH19 (unknown; 3e⁻²⁰)  
*Leptospira interrogans* OrfJ19 (unknown; 8e⁻²⁰) | Unknown                                                 |
| wlaUE              | *Leptospira interrogans* OrfJ19 (unknown; 3e⁻³¹)  
*Vibrio anguillarum* Orf15x4 (unknown; 1e⁻³⁰)  
*Leptospira borgpetersenii* OrfH19 (unknown; 1e⁻³⁰) | Unknown                                                 |
<table>
<thead>
<tr>
<th>Open reading frame</th>
<th>Proteins with highest similarity (probability score)</th>
<th>Putative function of gene product in <em>C. jejuni</em> NCTC 11828</th>
</tr>
</thead>
</table>
| **wlaUF**         | *Vibrio anguillarum* Orf16x2 (putative acetyltransferase; $3 \times 10^{-55}$)  
*Pyrocococcus abyssi* strain Orsay PAB0773 (putative acetyltransferase; $1 \times 10^{-31}$)  
*Pseudomonas aeruginosa* WbpD ($1 \times 10^{-21}$) | Putative acetyltransferase |
| **wlaUG**         | *Synechocystis sp.* strain PCC 6803 Fmt ($1 \times 10^{-11}$)  
*Ureaplasma urealyticum* Fmt ($6 \times 10^{-11}$)  
*Borrelia burgdorferi* Fmt ($4 \times 10^{-10}$) | Methionyl-tRNA formyltransferase |
| **wlaUH**         | *Rickettsia prowazekii* LgtD (glycosyltransferase; $3 \times 10^{-21}$)  
*Streptococcus suis* Cps7H (putative glycosyltransferase; $6 \times 10^{-20}$)  
*Campylobacter jejuni* OH 4384 wlaQ (β-1,3 galactosyltransferase; $1 \times 10^{-17}$) | Glycosyltransferase |
| **wlaUI**         | *Streptococcus suis* strain 6555 Cps11 (putative glycosyltransferase; $4 \times 10^{-24}$)  
*Streptococcus suis* strain 10 Cps2K (putative glycosyltransferase; $1 \times 10^{-21}$)  
*Streptococcus suis* strain 6555 Cps1K (putative glycosyltransferase; $3 \times 10^{-21}$) | Glycosyltransferase |
| **wlaUJ**         | *Vibrio anguillarum* Orf41x4 (putative amino transferase; $1 \times 10^{-14}$)  
*Bacteroides fragilis* WcgH (putative amino transferase; $1 \times 10^{-11}$)  
*Leptospira interrogans* OrfJ20 (putative amino transferase; $1 \times 10^{-3}$) | Putative amino transferase |
<table>
<thead>
<tr>
<th>Open reading frame</th>
<th>Proteins with highest similarity (probability score)</th>
<th>Putative function of gene product in <em>C. jejuni</em> NCTC 11828</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>wlaVB</strong></td>
<td><em>Campylobacter jejuni</em> NCTC 11168 NeuC1 (3e⁻³⁵)</td>
<td>Unknown</td>
</tr>
<tr>
<td></td>
<td><em>Campylobacter jejuni</em> strain TGH9011 invasion phenotype protein CipA (5e⁻²²)</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Campylobacter jejuni</em> NCTC 11168 NeuB1 (2e⁻²⁰)</td>
<td></td>
</tr>
<tr>
<td><strong>wlaVA</strong></td>
<td><em>Campylobacter jejuni</em> OH4384 CgtA (β 1,4 N-acetylgalactosaminyltransferase; 1e⁻⁹¹)</td>
<td>β 1,4-N-acetylgalactosaminyltransferase</td>
</tr>
<tr>
<td></td>
<td><em>Campylobacter jejuni</em> NCTC 11168 NeuA1 (1e⁻⁷³)</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Helicobacter pylori</em> strain J99 hypothetical protein JHP0203 (1e⁻⁶⁴)</td>
<td></td>
</tr>
<tr>
<td><strong>wlaWA</strong></td>
<td><em>Actinobacillus pleuropneumoniae</em> putative LPS biosynthesis protein (2e⁻¹⁴)</td>
<td>Unknown</td>
</tr>
<tr>
<td></td>
<td><em>Haemophilus ducreyi</em> LosA/LgbA (4e⁻¹³)</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Homo sapiens</em> hypothetical protein KIAA0584 (2e⁻¹²)</td>
<td></td>
</tr>
<tr>
<td><strong>wlaWB</strong></td>
<td><em>Campylobacter jejuni</em> NCTC 11168 WlaO (6e⁻⁷⁸)</td>
<td>Unknown</td>
</tr>
<tr>
<td></td>
<td><em>Streptococcus pneumoniae</em> cps14K (6e⁻²³)</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Streptococcus pneumoniae</em> capp33f1 (3e⁻¹⁸)</td>
<td></td>
</tr>
<tr>
<td><strong>wlaWC</strong></td>
<td><em>Escherichia coli</em> HtrL (involved in LPS biosynthesis; 2e⁻²⁰)</td>
<td>Unknown</td>
</tr>
<tr>
<td></td>
<td><em>Caenorhabditis elegans</em> hypothetical protein T11F9.12 (7e⁻¹⁴)</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Plasmodium falciparum</em> hypothetical protein C0810c (0.035)</td>
<td></td>
</tr>
</tbody>
</table>
but glucose is a common constituent of other *C. jejuni* core oligosaccharides (reviewed in Moran & Penner, 1999). The WlaUA protein also has similarity to WlaNA of *C. jejuni* NCTC 11168 (Table 4.3; Figure 4.9). Both the *wlaUA* (11828) and *wlaNA* (11168) genes are located in an identical position and orientation in the LOS clusters of their respective strains (Figure 4.8), and are similar in the N-terminal region, but the WlaUA protein appears to be truncated compared to the WlaNA protein. The significance of this is unclear.

**wlaUB and wlaUC**

The protein products from two genes in the NCTC 11828 cluster, *wlaUB* and *wlaUC* are highly similar to RmlA and RmlB proteins, respectively (Table 4.3). The RmlA protein of *E. coli* is a glucose-1-phosphate thymidyltransferase which converts D-glucose-1-phosphate to dTDP-D-glucose. This is then converted to dTDP-6-deoxy-L-threo-D-glycero-hexulose by the action of the RmlB protein (reviewed in Raetz, 1996). This molecule can then act as a precursor for several enzymatic pathways which lead to the biosynthesis of activated sugars such as dTDP-N-acetylfucosamine, dTDP-L-rhamnose, dTDP-4-acetamido-4,6-dideoxy-D-glucose and dTDP-6-deoxy-altrose (reviewed in Reeves *et al.*, 1996). These sugars are common components of Gram negative O-chain molecules, suggesting that the WlaUB and WlaUC proteins in *C. jejuni* NCTC 11828 are involved in the biosynthesis of sugars for incorporation into O-chain. Unfortunately, no structural analysis of the O-chain produced by this strain has so far been reported, so it is not clear which sugar they may be involved in synthesizing. However, the O-chain structures reported from other *C. jejuni* strains contain similar sugars to some of those mentioned above (Aspinall *et al.*, 1992a; Aspinall *et al.*, 1994c).

**wlaUD and wlaUE**

The protein products of the two genes *wlaUD* and *wlaUE* are highly similar (Figure 4.10) suggesting a similar or identical function. Both proteins have homology to LPS associated proteins of unknown function such as OrfH19 from *Leptospira*
Chapter 4: Gene content polymorphisms in *C. jejuni* LOS clusters

**Figure 4.9:** Alignment of the amino acid sequences from the *C. jejuni* NCTC 11168 WlaNA and *C. jejuni* NCTC 11828 WlaUA proteins. Identical residues are highlighted in red. Similar residues are indicated by a colon (*i.e.* :). Both proteins are suggested as glycosyltransferases on the basis of homology. The WlaUA protein appears to be a similar but truncated homolog of the WlaNA protein. The genes encoding the two genes are found in identical positions and orientation in the LOS cluster of their respective strain (downstream of the *waaC* and *waaM* genes).
Chapter 4: Gene content polymorphisms in \textit{C. jejuni} LOS clusters

WlaUD : ---LYT/YKF/TKNSKSNHAVHQKDVPNCFEELKHYVYFPLDDSINLPNLSSSEIF:58
wlaUE : MIRKCIINLRAEDRNRGSIALENNKEVPEEIKVVFYITFDPNQPRLAHKLQVL:60

\begin{verbatim}
WlaUD : LALNCGCNIILGDQGTKKILNKTAGLYIDTMIRKQVNEFCDIC1LVLNTDYDEKK :118
wlaUE : YMSKCIKILNLTREYKCKRPDLGYIDSTNQREALPHYAGAKLVLASDETYDEKK :120
\end{verbatim}

\begin{verbatim}
WlaUD : YDYKCYCELKNIVWGGYAIKMP :145
wlaUE : YRNDEYLRT-------- :135
\end{verbatim}

Figure 4.10: Alignment of the amino acid sequences from the \textit{C. jejuni} NCTC 11828 WlaUD and WlaUE proteins. Identical residues are highlighted in red. Similar residues are indicated by a colon (\textit{i.e.} :). The above alignment shows that both proteins have a similar primary sequence suggesting a similar or identical function. However, this function is as yet unclear due to a lack of informative homologs (Table 4.3).
Chapter 4: Gene content polymorphisms in C. jejuni LOS clusters

Therefore, the role of WlaUD and WlaUE in the LPS biosynthesis of C. jejuni NCTC 11828 is not clear and studies are needed to establish the functions of these proteins. One significant feature of the LOS cluster in C. jejuni NCTC 11168 is the presence of two homo-polymeric tracts which are located in the wlaQ and wlaSC genes (indicated in Figure 4.8; Parkhill et al., 2000a). Examination of the DNA sequence of WlaUD reveals the presence of a similar homo-polymeric tract at the C-terminal end of the open-reading frame. Variations in the numbers of G/C residues would lead to a short truncation or extension (six amino acids) of the protein (data not shown).

wlaUF

Like wlaUD, this open-reading frame also contains a homo-polymeric tract of G/C residues at the C-terminal end of the gene (indicated in Figure 4.8). Variations in the numbers of G/C residues would lead to a short truncation or extension (18 amino acid residues) of the protein (data not shown). The WlaUF protein shows similarity to putative acetyltransferase proteins (Table 4.3). Functional studies are needed to confirm whether WlaUF performs a similar function in the biosynthesis of LPS in C. jejuni NCTC 11828.

wlaUG

The WlaUG protein shows low levels of similarity to various Gram negative Fmt proteins (methionyl-tRNA formyltransferases; Table 4.3). These proteins have no apparent role in LPS biosynthesis, therefore it appears that the cluster in NCTC 11828 may not exclusively contain LOS/LPS associated genes. However, the level of similarity is low and thus the WlaUG protein may not act as a methionyl-tRNA formyltransferase.

wlaUH and wlaUI

Both of these proteins show similarity to several glycosyltransferases (Table 4.3) including the Cps7H and Cps2K proteins of Streptococcus suis (Smith et al.,
Chapter 4: Gene content polymorphisms in C. jejuni LOS clusters

1999a; Smith et al., 1999b). WlaUH and WlaUI are therefore suggested as having a glycosyltransferase function in C. jejuni NCTC 11828. Studies are needed to determine the specific sugar residues transferred and the glycosidic bonds formed by the action of these proteins.

wlaUJ

The protein encoded by the wlaUJ gene shows similarity to various amino-transferases such as WcgH from Bacteroides fragilis (Comstock et al., 1999). This protein is suggested to be responsible for the addition of an amino group (i.e. positive charge) to the B. fragilis polysaccharide repeating unit, thus mediating the overall charge characteristics of the molecule (Comstock et al., 1999). WlaUJ is suggested to have a similar role in C. jejuni NCTC 11828, but the receptor molecule (i.e. core or O-chain) is not clear.

wlaVB

The WlaVB protein shows low levels of similarity along its entire length to both the NeuC1 and NeuB1 proteins of C. jejuni NCTC 11168 (Table 4.3). These proteins are involved in the biosynthesis of sialic acid for incorporation into core oligosaccharide (Linton et al., 2000). However, preliminary structural analysis (Anthony Moran, personal communication) indicates that no sialic acid is present in the core oligosaccharide of NCTC 11828. This would suggest that the WlaVB protein is involved in the biosynthesis of a similar but alternative sugar.

The NeuC protein of E. coli is involved with the biosynthesis of N-acetylmannosamine - a precursor of sialic acid (Zapata et al., 1992). No N-acetylmannosamine residues have been detected in the C. jejuni core oligosaccharides which have been structurally determined (Moran & Penner, 1999), but perhaps the core of NCTC 11828 does contain this sugar. Alternatively, the WlaVB protein may be involved in the biosynthesis of a related sugar for example N-acetylgalactosamine. These sugars are commonly found in the outer core moieties of C. jejuni (Moran & Penner, 1999).
Chapter 4: Gene content polymorphisms in C. jejuni LOS clusters

**wlaVA**

This protein has highest similarity to the CgtA protein of C. jejuni OH 4384 which has been shown to have β-1,4-N-acetylgalactosaminyltransferase activity (Gilbert et al., 2000). This function was assigned by over-expressing the CgtA protein in E. coli and then assaying transferase activity and the type of glycosidic bond formed using synthetic fluorescent oligosaccharides as acceptors (Gilbert et al., 2000). Although the presence of N-acetylgalactosamine in the core oligosaccharide of NCTC 11828 has not been established, this sugar is common constituent of the core moiety of other C. jejuni strains (Moran & Penner, 1999). Work on the C. jejuni OH 4384 LOS cluster will be discussed in later sections.

**wlaWA, wlaWB and wlaWC**

The proteins encoded by these three genes all have similarity to LPS or capsular associated proteins of unknown function such as the LosA/LbgA proteins from H. ducreyi (Gibson et al., 1997; Stevens et al., 1997). Unfortunately without further functional studies, or more informative homologs, the role of these proteins in LPS/LOS biosynthesis is uncertain.

**4.4. Identification and characterization of gene content polymorphisms between wlaSA and wlaT**

PCR analysis across the wlaSA to wlaT region of the LOS cluster allowed the classification of strains into two groups (Table 4.4). The first group, comprising of 19 strains, had an identical gene content and arrangement to the genome strain. The second group of 13 strains contained the same six genes (wlaSA, waaF, gmhA, E, D and wlaT) in the same organization as NCTC 11168, but had in addition, insertions of various lengths located between the waaF and gmhA genes. Eleven of these were estimated to be approximately 0.7 kb in length, whilst two others, from strains 2523/90 and NCTC 11828 were estimated at 1.4 kb and 2.7 kb, respectively (Figure 4.11). As with the other gene content polymorphisms described in previous sections, no obvious correlations...
Table 4.4: The distribution and approximate size of inserted sequences between \textit{waaF} and \textit{gmhA} in 38 strains of \textit{C. jejuni}. Strain sources were shown in Table 2.1. Asterisks indicate a strain isolated from a patient with Guillain Barré Syndrome. The inserts from those strains indicated by \# were characterized further by cloning and sequencing (details in text). The predicted phenotype (LOS/LPS) is based on structural data from several \textit{C. jejuni} strains which suggested that serotype HS:1, HS:2, HS:3 and HS:41 strains produce LOS, whilst strains of serotype HS:4, HS:19, HS:23 and HS:36 produce an O-chain (LPS) - this is described in section 1.9. NK, not known, UT, untypable. NA, not applicable. ND, not done.
### Chapter 4: Gene content polymorphisms in *C. jejuni* LOS clusters

<table>
<thead>
<tr>
<th><em>C. jejuni</em> strain</th>
<th>HS serotype</th>
<th>Predicted phenotype (LOS/LPS)</th>
<th>Presence of insert between <em>waaF</em> and <em>gmhA</em></th>
<th>Estimated size of insert (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1*</td>
<td>1</td>
<td>LOS</td>
<td>-</td>
<td>NA</td>
</tr>
<tr>
<td>G3*</td>
<td>1</td>
<td>LOS</td>
<td>-</td>
<td>NA</td>
</tr>
<tr>
<td>O:1</td>
<td>1</td>
<td>LOS</td>
<td>-</td>
<td>NA</td>
</tr>
<tr>
<td>2561/90</td>
<td>2</td>
<td>LOS</td>
<td>-</td>
<td>NA</td>
</tr>
<tr>
<td>NCTC 11168</td>
<td>2</td>
<td>LOS</td>
<td>-</td>
<td>NA</td>
</tr>
<tr>
<td>K85</td>
<td>2</td>
<td>LOS</td>
<td>-</td>
<td>NA</td>
</tr>
<tr>
<td>O:3</td>
<td>3</td>
<td>LOS</td>
<td>-</td>
<td>NA</td>
</tr>
<tr>
<td>21F 185</td>
<td>15</td>
<td>NK</td>
<td>-</td>
<td>NA</td>
</tr>
<tr>
<td>P19</td>
<td>19</td>
<td>LPS</td>
<td>-</td>
<td>NA</td>
</tr>
<tr>
<td>J75</td>
<td>28</td>
<td>NK</td>
<td>-</td>
<td>NA</td>
</tr>
<tr>
<td>2F 68</td>
<td>50</td>
<td>NK</td>
<td>-</td>
<td>NA</td>
</tr>
<tr>
<td>N82</td>
<td>UT</td>
<td>NK</td>
<td>-</td>
<td>NA</td>
</tr>
<tr>
<td>G4*</td>
<td>UT</td>
<td>NK</td>
<td>-</td>
<td>NA</td>
</tr>
<tr>
<td>2F 90</td>
<td>UT</td>
<td>NK</td>
<td>-</td>
<td>NA</td>
</tr>
<tr>
<td>E206</td>
<td>UT</td>
<td>NK</td>
<td>-</td>
<td>NA</td>
</tr>
<tr>
<td>G2*</td>
<td>UT</td>
<td>NK</td>
<td>-</td>
<td>NA</td>
</tr>
<tr>
<td>F3*</td>
<td>UT</td>
<td>NK</td>
<td>-</td>
<td>NA</td>
</tr>
<tr>
<td>4F 182</td>
<td>NK</td>
<td>NK</td>
<td>-</td>
<td>NA</td>
</tr>
<tr>
<td>27F 155</td>
<td>NK</td>
<td>NK</td>
<td>-</td>
<td>NA</td>
</tr>
<tr>
<td>B404</td>
<td>NK</td>
<td>NK</td>
<td>-</td>
<td>NA</td>
</tr>
<tr>
<td>2523/90 #</td>
<td>2</td>
<td>LOS</td>
<td>+</td>
<td>1.4</td>
</tr>
<tr>
<td>2608/90</td>
<td>4</td>
<td>LPS</td>
<td>+</td>
<td>0.7</td>
</tr>
<tr>
<td>8F 169 #</td>
<td>5</td>
<td>NK</td>
<td>+</td>
<td>0.7</td>
</tr>
<tr>
<td>NCTC 11828 #</td>
<td>6</td>
<td>LPS</td>
<td>+</td>
<td>2.7</td>
</tr>
<tr>
<td>O73</td>
<td>11</td>
<td>NK</td>
<td>+</td>
<td>0.7</td>
</tr>
<tr>
<td>2258/90</td>
<td>23</td>
<td>LPS</td>
<td>+</td>
<td>0.7</td>
</tr>
<tr>
<td>NCTC 11351 #</td>
<td>23</td>
<td>LPS</td>
<td>+</td>
<td>0.7</td>
</tr>
<tr>
<td>81-176</td>
<td>23/36</td>
<td>LPS</td>
<td>+</td>
<td>0.7</td>
</tr>
<tr>
<td>16971.94GSH*</td>
<td>41</td>
<td>LOS</td>
<td>+</td>
<td>0.7</td>
</tr>
<tr>
<td>28134.94GSH*</td>
<td>41</td>
<td>LOS</td>
<td>+</td>
<td>0.7</td>
</tr>
<tr>
<td>260.94.RXH*</td>
<td>41</td>
<td>LOS</td>
<td>+</td>
<td>0.7</td>
</tr>
<tr>
<td>176.83</td>
<td>41</td>
<td>LOS</td>
<td>+</td>
<td>0.7</td>
</tr>
<tr>
<td>4F 225</td>
<td>NK</td>
<td>NK</td>
<td>+</td>
<td>0.7</td>
</tr>
<tr>
<td>306/90</td>
<td>1</td>
<td>LOS</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>1915/91</td>
<td>2</td>
<td>LOS</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>2F 8</td>
<td>5</td>
<td>NK</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>18F 5</td>
<td>11</td>
<td>NK</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>21F 242</td>
<td>NK</td>
<td>NK</td>
<td>ND</td>
<td>ND</td>
</tr>
</tbody>
</table>
Figure 4.11: PCR analysis to illustrate the approximate size of gene insertions between \textit{waaF} and \textit{gmhA}. All products were obtained using primers GMHAF1 and WAAFR5 (detailed in Appendix 1). PCR products were separated by horizontal gel electrophoresis (0.8% agarose). Marker sizes, in kilobase pairs, are shown on the left hand side of the panel. Lanes containing irrelevant information have been deleted. Lane 1, \textit{Hind} III, \textit{phi} x \textit{Hae}III marker; Lane 2, NCTC 11168 produces a product of 0.59 kb; Lanes 3 - 13, strains 2258/90, 4F 225, 8F 169, O73, 81-176, 16971.94GSH, 28134.94GSH, 260.94RXH, 176.83, 2608/90 and NCTC 11351, respectively, yield products of approximately 1.3 kb - confirming the presence of a 0.7 kb insertion between \textit{waaF} and \textit{gmhA}; Lane 14, 2523/90 yields a 2.0 kb fragment, indicating the presence of a 1.4 kb insertion between \textit{waaF} and \textit{gmhA} in this strain; Lane 15, NCTC 11828 DNA produces a fragment of approximately 3.3 kb, indicating the presence of 2.7 kb insert between \textit{waaF} and \textit{gmhA} in this strain; Lane 16, Negative control.
between serotype or LOS/LPS phenotype were evident (Table 4.4), and strains isolated from patients with Guillain Barré Syndrome are found in both groups.

In order to investigate the gene content of these polymorphisms, it was decided to clone and sequence the waaF - gmhA insertions from several representative strains (summarized in Figure 4.12). The methodology used to achieve this was similar for all the fragments described in the following sections, but will only be described in depth for the C. jejuni NCTC 11828 fragment.

4.4.1. Isolation and characterization of the C. jejuni NCTC 11828 waaF - gmhA insert

The waaF - gmhA insert was amplified by the Expand™ High Fidelity PCR methodology described in section 2.14.2 using primers GMHAF1 and WLASAR1 (Appendix 1). The amplified fragment was then purified by horizontal gel electrophoresis followed by extraction using the NUCLEOTRAP kit. The purified DNA was digested with BamH1 overnight before being ligated into BamH1 digested pUC19 (section 2.10.4; a schematic diagram of pUC19 is shown in Appendix 2) and electroporated into E. coli DH5α cells (section 2.11). The plasmid DNA from potential positive clones was extracted using the alkaline extraction method (section 2.6.1) and used for PCR (section 2.14.1; data not shown) to confirm that one of the clones contained the desired construct, which was named pNOL5. The sequence of the pNOL5 insert was obtained by primer walking (shown in Appendix 4). Analysis of this sequence using the Gene Runner computer program revealed the presence of two incomplete (gmhA and wlaSA), and three complete open-reading frames - waaF and two others, named wlaXA and wlaXB. Both these novel genes are in the same orientation, and appear to be in the same operon as the waaF gene. A schematic diagram of pNOL5 is shown in Figure 4.13A. In order to assign a putative function to the proteins encoded by these two genes, each protein sequence was used in a BLAST search against the Genbank protein sequence database. This data is summarized in Table 4.5. Briefly, the WlaXA protein shows similarity to various Gram negative alginate O-acetylation proteins such as AlgI from Pseudomonas aeruginosa
Figure 4.12: Schematic diagram summarizing the gene content in the \textit{wlaSA - wlaT} regions of five \textit{C. jejuni} strains. Open reading frames and the direction of their transcription are indicated by arrows. Proposed gene names are also given. Black text and arrows indicate genes conserved in all 38 strains of \textit{C. jejuni}. Red text and arrows indicate polymorphic genes inserted between the \textit{gmhA} and \textit{waaF} genes. Details of these polymorphic genes and the putative functions of the proteins they encode can be found in the text. This diagram is not shown to scale.
Chapter 4: Gene content polymorphisms in *C. jejuni* LOS clusters

NCTC 11168

\[ \text{Gene content} \]

\[ wlaSA \ waaF \ gmhA \ gmhE \ gmhD \ wlaT \]

NCTC 11828

\[ \text{Gene content} \]

\[ wlaSA \ waaF \ wlaXA \ wlaXB \ gmhA \ gmhE \ gmhD \ wlaT \]

2523/90

\[ \text{Gene content} \]

\[ wlaSA \ waaF \ wlaY \ gmhA \ gmhE \ gmhD \ wlaT \]

8F 169

\[ \text{Gene content} \]

\[ wlaSA \ waaF \ wlaZ \ gmhA \ gmhE \ gmhD \ wlaT \]

NCTC 11351

\[ \text{Gene content} \]

\[ wlaSA \ waaF \ wlaZ \ gmhA \ gmhE \ gmhD \ wlaT \]
Chapter 4: Gene content polymorphisms in *C. jejuni* LOS clusters

Figure 4.13: Diagrammatic representation showing pNOL5 (A) and pNOL6 (B).
The *waaF*/*gmhA* insert and flanking sequence was amplified from *C. jejuni* NCTC 11828 using the primers GMHAF1 and WLASAR1 (Appendix 1). The PCR product was digested with *Bam*HI before being ligated into pUC19 to form pNOL5. The insert from *C. jejuni* 2523/90 was amplified and cloned using the same methodology, but using the primers GMHAF1 and WAAFR1 (Appendix 1) to form pNOL6.
Chapter 4: Gene content polymorphisms in *C. jejuni* LOS clusters

**A**

pNOL5

7347 bps

**B**

pNOL6

4840 bps
Table 4.5: Details of the genes identified in the *waaF* - *gmhA* insertions from four strains of *C. jejuni*. Gene names are suggested based on the existing *wla* nomenclature (Fry *et al*., 1998; Fry *et al*., 2000b). Homology searches were carried out using the Genbank protein sequence database - the three protein sequences with highest similarity are shown here along with a probability score. This gives an indication of the confidence with which the identification was made: lower numbers indicate higher probability. The proposed functions are suggested on the basis of homology (see text for details).
<table>
<thead>
<tr>
<th>Open reading frame</th>
<th>\textit{C. jejuni} strain</th>
<th>Proteins with highest similarity (probability score)</th>
<th>Putative function of gene product</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{wlaXA}</td>
<td>NCTC 11828</td>
<td>\textit{Helicobacter pylori} strain 26695 Algl ((e^{-130}))  \  \textit{Pseudomonas aeruginosa} Algl ((5e^{-76}))  \  \textit{Azotobacter vinelandii} Algl ((5e^{-73}))</td>
<td>\textit{O-acetylation protein}</td>
</tr>
<tr>
<td>\textit{wlaXB}</td>
<td>NCTC 11828</td>
<td>\textit{Helicobacter pylori} strain 26695 hypothetical protein HP0856 ((1e^{-12}))  \  \textit{Borrelia burgdorferi} exodeoxyribonuclease V, gamma chain (RecC; 0.006)  \  \textit{Ureaplasma urealyticum} hypothetical protein ((0.069))</td>
<td>Unknown</td>
</tr>
<tr>
<td>\textit{wlaY}</td>
<td>2523/90</td>
<td>\textit{Borrelia burgdorferi} P115 protein ((0.042))  \  \textit{Plasmodium falciparum} hypothetical protein PFC1015c ((0.055))  \  \textit{Caenorhabditis elegans} hypothetical protein TR:O01963 ((0.095))</td>
<td>Unknown</td>
</tr>
<tr>
<td>\textit{wlaZ}</td>
<td>NCTC 11351</td>
<td>\textit{Haemophilus influenzae} Lex2B ((5e^{-26}))  \  \textit{Haemophilus somnus} Lob1 ((1e^{-19}))  \  \textit{Helicobacter pylori} strain 26695 Lex2B homolog ((2e^{-17}))</td>
<td>Unknown</td>
</tr>
<tr>
<td>\textit{wlaZ}</td>
<td>8F 169</td>
<td>\textit{Haemophilus influenzae} Lex2B ((3e^{-28}))  \  \textit{Haemophilus somnus} Lob1 ((1e^{-23}))  \  \textit{Helicobacter pylori} strain 26695 Lex2B homolog ((3e^{-19}))</td>
<td>Unknown</td>
</tr>
</tbody>
</table>
(Franklin & Ohman, 1996). The precise function of the Algl protein is not clear, but it is required for the O-acetylation of P. aeruginosa alginate. This is a linear exopolysaccharide composed of D-mannuronate and L-guluronate (Franklin & Ohman, 1996). The WlaXA protein of C. jejuni NCTC 11828 is therefore proposed to have a similar O-acetylation function, perhaps involved with O-chain biosynthesis. BLAST searches using the protein sequence from the second gene in the NCTC 11828 gmhA - waaF insertion, wlaXB, failed to identify any similar sequences in the Genbank database. The role of this protein in C. jejuni NCTC 11828 is therefore unknown.

4.4.2. Isolation and characterization of the C. jejuni 2523/90 waaF - gmhA insert

The 1.4 kb insert from C. jejuni 2523/90 was amplified and cloned in the same manner as the NCTC 11828 insert, but using primers GMHAF1 and WAAFR1 (Appendix 1). The sequence of the resulting clone - pNOL6, was obtained by a strategy of primer walking (Appendix 5). Analysis of this sequence revealed the presence of two incomplete (gmhA and waaF), and one complete open reading frame, named wlaY. This open reading frame appears to be located in an operon with the gmhA gene (shown in Figure 4.13B). In order to assign a putative function to the protein encoded by wlaY, the protein sequence was used in a BLAST search against the Genbank protein sequence database (Table 4.5). Unfortunately, this search failed to identify any similar sequences, making a prediction of the function of the WlaY protein problematical.

4.4.3. Isolation and characterization of the C. jejuni 8F 169 and NCTC 11351 waaF - gmhA inserts

Eleven C. jejuni strains were determined to have an insert of approximately 0.7 kb located between waaF and gmhA. To determine the gene content of this insert it was decided to clone and sequence it from two representative strains; 8F 169 and NCTC 11351. These two inserts were amplified and cloned in exactly the same way as those previously mentioned, using primers GMHAF1 and WAAFR1, to obtain two constructs; pNOL7 (containing the NCTC 11351 insert) and pNOL21 (containing the
Chapter 4: Gene content polymorphisms in C. jejuni LOS clusters

8F 169 insert). The sequence of these clones was obtained by primer walking, and is shown in Appendices 6 and 7 respectively (data not shown). Analysis of the sequence revealed the presence of two incomplete (gmhA and waaF), and one complete open-reading frame, named wlaZ in both clones. Schematic diagrams of pNOL7 and pNOL21 are shown in Figure 4.14. Homology searches (Table 4.5) showed that both WlaZ proteins had similarity to the Lex2b protein of H. influenzae (Jarosik & Hansen, 1994), and the Losb1 protein of H. sommus (Inzana et al., 1997). The precise role of both of these proteins is unknown, but both the lex2 locus and losb1 gene are thought to be involved with the phase variation of lipooligosaccharide epitopes due to the presence of tetra-nucleotide repeats such as 5'-CAAT-3' (Inzana et al., 1997; Jarosik & Hansen, 1994). Variations in the numbers of these repeats creates a translational on-off switch by a mechanism involving slip-stranded mispairing (Inzana et al., 1997; Jarosik & Hansen, 1994). Examination of the C. jejuni 8F 169, and NCTC 11351 sequence for the presence of tetra-nucleotide motifs failed to identify any such repeats (data not shown).

Although both the NCTC 11351 and 8F 169 proteins show similarity to the same proteins, further examination of their respective protein sequences shows that the similarity between them is mainly restricted to the N-terminal region (Figure 4.15). It therefore appears that the proteins may have slightly differing functions in the two different strains. Functional studies are needed to confirm the role of these proteins in LOS/LPS biosynthesis.

4.5. Discussion

4.5.1. Identification of areas of gene content polymorphism in the LOS cluster

The initial aim of this chapter was to identify areas of gene content polymorphism in the LOS biosynthesis cluster in different isolates of C. jejuni. This was undertaken in an attempt to identify genes responsible for the differing core structures produced by different strains of C. jejuni. To achieve this, the clusters from various strains were analyzed using a simple PCR approach to pin-point areas of gene content which were different from the genome strain. This analysis indicated three areas
Figure 4.14: Diagrammatic representation showing pNOL7 (A) and pNOL21 (B).
The waaF - gmhA inserts and flanking sequence were amplified from C. jejuni NCTC 11351 (A) and 8F 169 (B) using the primers GMHAF1 and WAAFR1. These PCR products were digested with BamHI before being ligated into pUC19 and transformed into E. coli DH5α. Both inserts contain a gene (wlaZ) which encodes a protein with similarity to Lex2b from H. influenzae (Jarosik & Hansen, 1994), and the Losb1 protein of H. sommus (Inzana et al., 1997).
Chapter 4: Gene content polymorphisms in *C. jejuni* LOS clusters

**A**

```
3737/3463
3510/HindIII
3441/HindIII
3365/HindIII
3392/HindIII
2990/HindIII
2843/HindIII
2585
```  

**B**

```
3763/3463
3757/NcoI
3751/AvrII
3747/NcoI
3742/HindIII
```
Chapter 4: Gene content polymorphisms in *C. jejuni* LOS clusters

A

3772\(^{EcoR\mathbf{I}}\),
3760\(^{KmI}\),
3751\(^{BamHI}\)

3510\(^{Hin\mathbf{d}III}\)

3367\(^{Hin\mathbf{d}III}\);
3321\(^{EcoR\mathbf{I}}\);
3300\(^{BclI}\)

3188\(^{Hin\mathbf{d}III}\)

3034\(^{Hin\mathbf{d}III}\)

2990\(^{Hin\mathbf{d}III}\)

2843\(^{Hin\mathbf{d}III}\)

2265

B

3763\(^{EcoR\mathbf{I}}\),
3757\(^{SacI}\),
3751\(^{KmI}\),
3747\(^{KpnI}\),
3742\(^{BamHI}\)

2235

3432\(^{Hin\mathbf{d}III}\)

3358\(^{Hin\mathbf{d}III}\)

3350\(^{BclI}\),
3312\(^{EcoR\mathbf{I}}\),
3265\(^{SacI}\)

3179\(^{Hin\mathbf{d}III}\)

3023\(^{Hin\mathbf{d}III}\)

2981\(^{Hin\mathbf{d}III}\)

2798\(^{Hin\mathbf{d}III}\)

2265\(^{BamHI}\)

2259\(^{SacI}\)

2253\(^{EcoI}\)

2247\(^{PstI}\)

2235\(^{BglII}\)

pNOL7

4172 bps

pNOL21

4163 bps
Chapter 4: Gene content polymorphisms in *C. jejuni* LOS clusters

**Figure 4.15: Alignment of the amino acid sequences from the *C. jejuni* 8F 169 and NCTC 11351 WlaZ proteins.** Identical residues are highlighted in red. Similar residues are indicated by a colon (i.e. :). The above alignment shows that both proteins have an almost identical primary sequence in the N-terminal region, but the level of similarity is considerably less in the C-terminal portion of the proteins. This may imply a differing function for the protein in the two strains. WlaZ_8F 169, WlaZ from *C. jejuni* 8F 169; WlaZ_11351, WlaZ from *C. jejuni* NCTC 11351.
of polymorphism. The first was the presence or absence of the \textit{wlaJ} gene. Approximately half of the strains tested lacked the \textit{wlaJ} gene, with the \textit{wlaI} and \textit{wlaK} genes being contiguous in the cluster. The second polymorphism identified was in the \textit{wlaNA} - \textit{wlaSB} region of the locus. PCR analysis of the \textit{C. jejuni} strains across this region allowed the classification of the strains into two groups. The first group contained an identical gene content and arrangement to the genome strain. The second group, were thought to have a radically different gene content in this central area of the LOS cluster. This was suggested by the inability to obtain amplification products using primers which annealed to the \textit{C. jejuni} NCTC 11168 genes found in this region. A third and final polymorphic region was identified between \textit{gmhA} and \textit{waaF}. Of the strains tested, 13 had insertions of various lengths located between these genes.

Once identified, it would have been extremely informative to consider the presence of these genetic polymorphisms in the light of structural data from the core oligosaccharide molecules produced by each strain. This might have allowed correlations between the presence of genetic polymorphisms with the presence of certain core epitopes to be made. However, the defined chemical structures for the core oligosaccharides produced by strains used in this study are not available. In the absence of this information, HS serotyping data was used as an alternative. This was assumed to reflect core structural similarities and differences - \textit{i.e.} strains of the same HS serotype are assumed to produce highly similar (but not necessarily identical) core structures. However, correlations between the HS data and the presence of the three polymorphisms were largely uninformative - it was clear that in the majority of cases, strains of the same HS serotype were evident in groups of strains that contained gene content polymorphisms and those that did not. This could be because the polymorphic areas identified are not involved with the biosynthesis of core or any other HS antigen and therefore do not affect HS serotype. This hypothesis is undermined by the fact that the genes characterized in the gene content polymorphisms show similarity to LOS/LPS related genes, and so probably do have a role in affecting core structure. Alternatively, this could be due to the fact that strains which do not produce identical core structures, and therefore have gene content differences, can still be classed as belonging to the
same serogroup. Therefore, the HS serotyping system is not discriminatory between core structures and correlations with gene content cannot be made. Finally, an alternative hypothesis is that the specificity of the *C. jejuni* HS serotyping scheme relates predominately to a heat stable antigen other than LOS/LPS. Therefore, differences in gene content, and thus core structure are not reflected in HS serotype differences. This final suggestion is discussed in a wider context in Chapter 7 of this study.

In addition to correlations with HS serotype, attempts were also made to correlate the presence of the three polymorphisms with the ability of the strain to produce O-chain (as judged by HS serotyping). No correlation was observed suggesting that the production of an O-chain is not affected by gene content polymorphisms in the cluster. Finally, strains isolated from patients with Guillain Barré Syndrome were also evident in groups of strains that contained gene content polymorphisms and those that did not. This suggests that gene content polymorphisms in the cluster do not directly affect the ability of a *C. jejuni* strain to induce GBS. If the polymorphisms are responsible for core structural differences, then this supports evidence that the ability of a *C. jejuni* strain to induce GBS is not just linked to the ability to synthesis core structures which mimic human gangliosides, but in addition, also involves host and other bacterial factors (Moran & O'Malley, 1995; Nachamkin *et al.*, 1999; Prendergast *et al.*, 1999; Sheikh *et al.*, 1998).

4.5.2. Further characterization of the gene content polymorphisms

Once located, further characterization of the gene content differences was required to establish the identity of novel or polymorphic genes. The further characterization of the *wlaJ* polymorphism is dealt with in Chapter 5. To investigate the nature of the *wlaNA - wlaSB* polymorphic region, this area was characterized from a representative strain - *C. jejuni* NCTC 11828. Analysis of this sequence revealed a totally different gene content and organization to the corresponding area of strain NCTC 11168. The NCTC 11828 region contains 15 novel open-reading frames, which
Chapter 4: Gene content polymorphisms in C. jejuni LOS clusters

encode proteins with various putative functions, as judged by homology. However, these suggested functions should be treated with caution, and need to be confirmed by the construction of defined mutants and the careful evaluation of the resulting phenotypes. In particular it needs to be established which structure the five putative glycosyltransferases encoded by the region, are involved with synthesizing. The presence of the \textit{waaC} and \textit{waaF} genes at either end of the region suggest the genes are involved with core biosynthesis. However, some of the other genes found in the cluster seem more indicative of an O-chain cluster. For example the presence of \textit{rmlA} and \textit{rmlB} homologs, which in \textit{E. coli} encode proteins involved with the biosynthesis of complex sugars for incorporation into O-chain. It appears that the cluster may therefore contain genes involved with the biosynthesis of both structures.

The cluster also contains the \textit{fmt} gene that encodes a protein which has function unrelated to LOS/LPS biosynthesis. The significance of the presence of this gene in the cluster is not clear. Some genes of note that are present in 11168 are absent in the NCTC 11828 cluster. For example, the \textit{wlaR, neuA1, B1, and C1} genes are absent which may suggest that sialic acid is not a component of the core oligosaccharide in this strain. Preliminary structural characterization of the \textit{C. jejuni} NCTC 11828 core oligosaccharide supports this suggestion (Anthony Moran, personal communication).

Whilst this study was being undertaken, a report was published which detailed the analysis of the same region of the LOS cluster in \textit{C. jejuni} OH 4384 - a strain from HS:41 serogroup (Gilbert et al., 2000). The \textit{waaC - waaF} region of this strain was cloned and characterized in a similar approach to that used here. Sequence analysis of the region in \textit{C. jejuni} OH 4384 revealed that in many ways it was similar to the corresponding region of the genome strain NCTC 11168. The major differences between the two were the absence of the \textit{wlaO, wlaP} and \textit{wlaSB} genes in \textit{C. jejuni} OH 4384, and the appearance of a novel gene with similarity to a putative acetyltransferase from \textit{Rhodobacter capsulatus} (Gilbert et al., 2000). The other significant finding was that the \textit{neuA1} gene of NCTC 11168 is in fact a in-frame fusion of two genes which are separate in OH 4384 (Gilbert et al., 2000). One of these has the expected similarity to the \textit{neuA} gene of \textit{E. coli}, whilst the other has been functionally characterized as
encoding a β-1,4-N-acetylgalactosaminyltransferase (Gilbert *et al.*, 2000). The significance of the fusion in *C. jejuni* NCTC 11168 is not clear. In *C. jejuni* 11828, the *neuA* part of the 11168 fusion is absent, but the β-1,4-N-acetylgalactosaminyltransferase gene is present (named *wlaVA*). This would suggest that *N*-acetylglucosamine is present in the core oligosaccharide of NCTC 11828, but structural analysis is needed to confirm this.

The three strains that have been characterized; NCTC 11168, NCTC 11828 and OH 4384 illustrate considerable gene content variation in the *waaC*-*waaF* region, and there is probably significantly more variation to be seen in other strains. Future work should involve the further characterization of these regions, but from strains which have had the core oligosaccharides they synthesize thoroughly chemically analyzed. Therefore, gene content polymorphisms may be correlated with the appearance of particular epitopes or residues. If there is considerably more gene content variation observed in these strains, this could be utilized for the detection, differentiation and tracking of strains in epidemiological studies. From the results already presented here, it appears that strains of the same HS serotype can in some cases be differentiated by gene content polymorphisms. Therefore, a PCR approach based on these polymorphisms might provide a better differentiation of strains than the current antibody based HS serotyping scheme (*Frost* *et al.*, 1998; *Penner* *et al.*, 1983).

One significant feature of the *waaC*-*waaF* region in the genome strain NCTC 11168 is the G+C ratio of the DNA. From the start of the *wlaNA* gene to the end of *wlaSA*, the G+C ratio is approximately 25%, in contrast to the overall G+C ratio of 30.6% for the complete genome (*Parkhill et al.*, 2000a). This region is of course the site of considerable gene content variation between strains. In NCTC 11828, the G+C ratio is even lower at approximately 24% for this corresponding region. The mechanism by which cluster variation has evolved has not been considered here in depth. The chapter has focused simply on the characterization of polymorphisms, not determining the mechanism by which they occur, but the low G+C ratio of these polymorphic regions may imply acquisition by horizontal transfer and recombination. This has also been suggested as a possible mechanism for the acquisition and evolution of *Salmonella* O-
Chapter 4: Gene content polymorphisms in C. jejuni LOS clusters

chain gene clusters (Reeves, 1993). Future work could involve mapping more closely the G+C ratio across these LOS clusters - this may indicate that only certain genes or parts of the cluster have a significantly lower G+C ratio than the genome average. This information may give further indications of how the cluster has evolved in different strains.

The final polymorphic regions identified and then characterized were the inserts identified between waaF and gmhA in some strains. Four representative inserts were cloned and sequenced. The insert from NCTC 11828 contains two genes which are located in an operon with the waaF gene. The protein product of one of these genes, wlaXA shows homology to various O-acetylation proteins, and is suggested as having a role in LPS biosynthesis, whilst the second, wlaXB encodes a protein of unknown function. The insert from 2523/90 was also cloned and sequenced - this contains one open-reading frame which encodes a protein of unknown function. Finally, the inserts from two representative strains containing a 0.7 kb fragment inserted between the gmhA and waaF genes were characterized. Both of these inserts contained one putative open reading frame, named wlaZ. Both WlaZ proteins have similarity to the same LOS biosynthesis proteins such as Lex2b from H. influenzae (Jarosik and Hansen, 1994), but are not identical along the entire length. They share a high level of identity in the N-terminal region, but not so in the C-terminal region. This may indicate that recombination events at this site generate variation in the WlaZ proteins of different strains, thus perhaps altering the function of the WlaZ proteins in different strains of C. jejuni. Future work could involve the cloning and sequencing of the other 0.7 kb inserts found in some strains. This may indicate whether other strains contain genes encoding for WlaZ or WlaZ-like proteins. Furthermore, insertional inactivation of these waaF and gmhA insert genes is needed to confirm the role of these in C. jejuni LOS/LPS biosynthesis.
Chapter 5: Functional analysis of \textit{wlaJ}

5.1. Introduction

The function of the \textit{C. jejuni} NCTC 11168 \textit{WlaJ} protein is unclear. The protein sequence shows only weak similarity to sequences in the Genbank database (Table 3.6) and insertional inactivation of \textit{wlaJ} in NCTC 11168 has no detectable effect on the mobility or intensity of core oligosaccharide as judged by immunoblotting (Wood \textit{et al.}, 1999). Preliminary biochemical studies have confirmed that \textit{wlaJ} mutant strains produce core molecules with identical sugar content to the wild-type NCTC 11168 strain (Anthony Moran, personal communication). However, Northern hybridization results (Anne Wood, personal communication) show that the gene is transcribed in \textit{C. jejuni} NCTC 11168 implying that the \textit{wlaJ} gene is expressed and so is likely to have a function. The gene also represents the only difference in gene content present in the \textit{wlaM-waaM} region of 38 isolates of \textit{C. jejuni}, which also suggests a significant function for this gene. Ultimately the presence or absence of \textit{wlaJ} in the \textit{C. jejuni} strains did not correlate with LPS/LOS phenotype (Table 4.1), but initial results from the study suggested that the \textit{wlaJ} gene was only found in LOS producing strains, and not in LPS producing strains. This suggested a putative role for \textit{WlaJ} in influencing the production of O-chain.

In order to gain further information as to the function of the \textit{wlaJ} gene a strategy was designed to insert the gene into the LPS producing \textit{C. jejuni} strain NCTC 11828. This strain is one of those which lacks the \textit{wlaJ} gene, with \textit{wlaI} and \textit{wlaK} being contiguous in the cluster (Fry, 1997; Fry \textit{et al.}, 1998). If the \textit{wlaJ} gene did affect O-chain biosynthesis, such a change would be easily detectable using immunoblotting. This chapter therefore details the construction of \textit{C. jejuni} NCTC 11828 chimeric strains in which the \textit{wlaJ} gene (and a selectable marker) were inserted and also shows the resulting change in LPS phenotype. The initial experiments in this Chapter were conducted by an M.Sc student Clióna O’Dwyer under the supervision of the author and
are reported in the paper by Wood et al. (1999). The chapter then goes on to describe additional experiments conducted by the author, to establish that it was the \textit{wlaJ} insertion that was responsible for this phenotypic change.

Chimeric strains are readily made in \textit{C. jejuni} by transformation with an appropriate plasmid construct. Such a construct would contain the gene of interest and a suitable selective marker and approximately 500 bp of flanking sequence either side cloned into the vector pUC19. pUC19 acts as a suicide vector in \textit{C. jejuni} and can be introduced by electroporation. Subsequent homologous recombination and gene replacement results in the insertion of the gene of interest and selectable marker into the \textit{C. jejuni} chromosome with the loss of the vector (reviewed in van Vliet et al., 1998).

5.2. Construction of plasmids to enable the relocation of \textit{wlaJ} to \textit{C. jejuni} NCTC 11828

Plasmid pAWL63 was chosen as the starting point of these experiments as it contains the intact \textit{wlaJ} gene (Figure 5.1). This plasmid was digested with \textit{BamH}I and then \textit{BssH}II to remove unwanted flanking sequence. The 4 kb fragment from this digestion, containing pUC19, \textit{wlaJ} and flanking sequence from \textit{wlaI} and \textit{wlaK} was purified by horizontal gel electrophoresis (section 2.8) followed by extraction using the NUCLEOTRAP extraction kit (section 2.9.1). The 5' overhangs created by digestion were converted into blunt-end termini (section 2.10.2), the DNA re-ligated (section 2.10.4), and transformed into \textit{E. coli} DH5\(\alpha\) cells by electroporation (section 2.11). The plasmid DNA from potential positive clones was extracted using the alkaline extraction method (section 2.6.1) and used for restriction analysis (section 2.10.1). This data confirmed that one of the clones contained the desired construct, which was named pCOD1 (Figure 5.1). A restriction enzyme digestion of pCOD1 is shown in Figure 5.2A.

It was then necessary to introduce a suitable antibiotic resistance gene in between the \textit{wlaJ} gene and the \textit{wlaK} flanking sequence present on pCOD1. The resulting constructs could then be transformed into \textit{C. jejuni} NCTC 11828. Clones which do not contain the \textit{wlaJ} gene and selectable marker gene successfully recombined into the chromosome will be selected against by the use of the appropriate
Figure 5.1: Diagrammatic representation showing the construction of pCOD1, pCOD2 and pCOD3. Digestion of pAWL63 with *BamH*I (·) and *BssH*II (**) removed unwanted insert DNA and a *SspI* site present in *wlaH*. The 5' overhangs created by this digestion were converted to blunt end termini and the DNA re-ligated to form pCOD1. A *cat* cassette was inserted into the *SspI* (***) site of pCOD1 in both forward (pCOD3) and reverse (pCOD2) orientations.
Chapter 5: Functional analysis of \textit{wlaJ}

\textbf{pAWL63}  
6213 bps

\textbf{pCOD1}  
4045 bps

\textbf{pCOD2}  
4887 bps

\textbf{pCOD3}  
4887 bps

179
Figure 5.2: Agarose gel electrophoresis (1% gel) to show the restriction enzyme digestion of pCOD1, pCOD2 and pCOD3. Marker sizes, in kilobase pairs, are shown on the left hand side of each panel.

A: Restriction enzyme digestion of pCOD1 confirming the removal of unwanted flanking sequence from pAWL63 and the SpI site the flanking sequence contained. Lane 1. λ Hind III. φ/x HaelIII marker; Lane 2. Uncut pCOD1 (4045 bp); Lane 3. Hind III digest yielding fragments of 3581 bp, 440 bp and 49 bp (not visible); Lane 4. SpI digest yielding fragments of 2469 bp and 1576 bp; Lane 5. Pst I and EcoRI digest yielding fragments of 2647 bp and 1398 bp.

B: Restriction enzyme digestion of pCOD2 confirming the reverse orientation of the inserted cat cassette. The resistance cassette was inserted in the SpI site (nucleotide number 2650) of pCOD1 in a three-way ligation. Lane 1. λ Hind III. φ/x HaelIII marker; Lane 2. Hind III digest yielding fragments of 3581 bp, 997 bp, 285 bp and 24 bp (not visible); Lane 3. Pst I digest yielding fragments of 3651 bp, 824 bp and 412 bp; Lane 4. Uncut pCOD2 (4887 bp); Lane 5. λ Hind III. φ/x HaelIII marker.

C: Restriction enzyme digestion of pCOD3 confirming the forward orientation of the inserted cat cassette. The resistance cassette was inserted in the SpI site (nucleotide number 2650) of pCOD1 in a three-way ligation. Lane 1. λ Hind III. φ/x HaelIII marker; Lane 2. Hind III digest yielding fragments of 3581 bp, 675 bp, 607 bp and 24 bp (not visible); Lane 3. Pst I digest yielding fragments of 3651 bp, 824 bp and 412 bp; Lane 4. Uncut pCOD3 (4887 bp); Lane 5. λ Hind III. φ/x HaelIII marker.
Chapter 5: Functional analysis of \textit{wlaJ}

A

B

C
antibiotic. In this case the chloramphenicol acetyl transferase (cat) antibiotic resistance cassette was used (Wang & Taylor, 1990a). pAV35 which contains the cat cassette (see Appendix 2) was digested overnight at 25°C with Smal and the 842 bp fragment containing the cat cassette purified. pCOD1 was digested overnight with SspI resulting in two fragments. The 842 bp cat cassette was then ligated to the two pCOD1 fragments in a three-way ligation. The resulting products were electroporated into E. coli DH5α cells and positive clones selected using chloramphenicol and ampicillin. The plasmid DNA from potential recombinants was extracted and used for PCR analysis (data not shown). This analysis showed that two of the clones contained the correct constructs, the first contained the cat cassette in the reverse orientation (pCOD2; Figure 5.1), whilst the second contained the cat cassette in the forward orientation (pCOD3; Figure 5.1). Restriction digestion of both plasmids confirmed the orientation of the cat cassette (Figure 5.2B and Figure 5.2C).

5.3. Transformation of C. jejuni NCTC 11828 with pCOD2 and pCOD3

pCOD2 and pCOD3 were transformed into C. jejuni NCTC 11828 by electroporation (section 2.12) to create strains 11828COD2 and 11828COD3 (Figure 5.3). Positive clones were selected on CSA plates containing chloramphenicol. The chromosomal DNA from putative positive clones was extracted (section 2.6.2) and used for PCR and Southern analysis (section 2.17). PCR using WLAHF1, WLAKR3 and primers which anneal to the cat cassette identified several correct recombinants which contained wlaJ + cat recombined in the correct position. The PCR data for one of the clones containing wlaJ + cat in the forward orientation (strain 11828COD3) and one clone containing wlaJ + cat in the reverse orientation (strain 111828COD2) is shown in Figure 5.4. Southern hybridization using the wlaJ and cat genes as probes also confirmed that strains 11828COD3 and 11828COD2 were correct (Figure 5.5), and due to the presence of an internal ClaI restriction site in the cat cassette the orientation of the cassette could be confirmed. The chromosomal DNA from both strains was also probed with pUC19 to ensure that no plasmid sequence remained. No pUC19 sequence was detected (data not shown).
Chapter 5: Functional analysis of \textit{wlaJ}

![Diagram showing gene content and primer positions for \textit{wlaG} - \textit{wlaK} region of strains 11828COD2 and 11828COD3.]

**Figure 5.3:** Schematic diagram showing the gene content (and primer positions) for the \textit{wlaG} - \textit{wlaK} region of strains 11828COD2 and 11828COD3. The parental strains NCTC 11828 and NCTC 11168 are shown for reference. Blue arrows represent NCTC 11828 gene sequence, red arrows represent NCTC 11168 gene sequence and black arrows represent chloramphenicol acetyl transferase genes (\textit{cat}). Genes represented by both red and blue arrows (\textit{i.e.} \textit{wlaI} and \textit{wlaK}) represent sites of recombination. In the absence of DNA sequence data from both strains to pinpoint the exact sites, for the purposes of this diagram, the maximum possible extent of genetic exchange is shown. Primer positions are also indicated. This diagram is not shown to scale.
Figure 5.4: PCR analysis of strains 11828COD2 and 11828COD3. This confirmed the correct recombination of pCOD2 and pCOD3 into the chromosome of *C. jejuni* NCTC 11828 and also confirmed the orientation of the *cat* cassette. PCR products were separated by horizontal gel electrophoresis (0.8% agarose). Primer position and orientation is shown in Figure 5.3. Marker sizes, in kilobase pairs, are shown on the left hand side of each panel.

**A: PCR analysis of strain 11828COD2.** Lane 1. λ *Hind* III. φ/x *Hae*III marker: Lane 2. WLAHF1/WLAKR3 primer combination using 11828COD2 DNA produces a product of 3.13 kb, 1.45 kb greater than the product from the wild-type strain (Lane 4). This size difference equals the size of the *wlaJ* + *cat* insertion: Lane 3. WLAHF1/CAT13' primer combination using 11828COD2 DNA produces a product of 1.9 kb confirming the presence and position of the *wlaJ* + *cat* insertion. This also confirms the reverse orientation of the *cat* cassette: Lane 4. WLAHF1/WLAKR3 primer combination using NCTC 11828 DNA produces a fragment of 1.68 kb: Lane 5. WLAHF1/CAT13' primer combination using NCTC 11828 DNA produces no fragment confirming the absence of the *cat* cassette in this wild-type strain: Lane 6. Negative control: Lane 7. λ *Hind* III. φ/x *Hae*III marker.

**B: PCR analysis of strain 11828COD3.** Lane 1. λ *Hind* III. φ/x *Hae*III marker: Lane 2. WLAHF1/WLAKR3 primer combination using 11828COD3 DNA produces a product of 3.13 kb, 1.45 kb greater than the product from the wild-type strain. This size difference equals the size of the *wlaJ* + *cat* insertion: Lane 3. WLAHF1/CAT15' primer combination using 11828COD3 DNA produces a product of 1.95 kb confirming the presence and position of the *wlaJ* + *cat* insertion. This also confirms the forward orientation of the *cat* cassette: Lane 4. WLAHF1/WLAKR3 primer combination using NCTC 11828 DNA produces a fragment of 1.68 kb: Lane 5. WLAHF1/CAT15' primer combination using NCTC 11828 DNA produces no fragment confirming the absence of the *cat* cassette in this wild-type strain: Lane 6. Negative control: Lane 7. λ *Hind* III. φ/x *Hae*III marker.
Chapter 5: Functional analysis of \textit{wlaJ}

A

B
Figure 5.5: Southern hybridization analysis of strains 11828COD2 and 11828COD3. This confirmed the correct recombination of pCOD2 and pCOD3 into the chromosome of *C. jejuni* NCTC 11828 and also confirmed the orientation of the *cat* cassette. Chromosomal DNA was restricted with *Cla* I overnight before being separated by horizontal gel electrophoresis (0.8% agarose). Lanes containing irrelevant information have been deleted. Marker sizes, in kilobase pairs, are shown on the left hand side of each panel.

A: Southern hybridization analysis of strains 11828COD2 and 11828COD3 using the *cat* cassette (and *λ Hind III* marker) as probe. Two hybridizing bands are expected for each mutant as the *cat* cassette contains an internal *Cla* I site. Lane 1, *λ Hind III* marker; Lane 2, 11828COD2 yields hybridizing *Cla* I fragments of 3.5 kb and 1.1 kb so confirming the presence and reverse orientation of the *cat* cassette; Lane 3, 11828COD3 yields hybridizing *Cla* I fragments of 4 kb and 0.55 kb confirming the presence and forward orientation of the *cat* cassette; Lane 4, no hybridizing band is present in the wild-type strain NCTC 11828; Lane 5, no hybridizing band is present in NCTC 11168.

B: Southern hybridization analysis of strains 11828COD2 and 11828COD3 using the *wlaJ* gene (and *λ Hind III* marker) as probe. This probe was produced by PCR from the NCTC 11168 *wlaJ* gene using primers WLAJF1/WLAJR1 (see Appendix 1). Lane 1, *λ Hind III* marker; Lane 2, 11828COD2 yields a hybridizing *Cla* I fragment of 3.5 kb so confirming the presence and position of the *wlaJ* gene; Lane 3, 11828COD3 yields a hybridizing *Cla* I fragment of 4 kb confirming the presence and position of the *wlaJ* gene; Lane 4, no hybridizing band is present in the wild-type strain NCTC 11828; Lane 5, *C. jejuni* NCTC 11168 produces a hybridizing band of 3.5 kb.
Chapter 5: Functional analysis of \textit{wla}J

A

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>23.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>23.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 5: Functional analysis of \textit{wlaJ}

5.4. Phenotypic analysis of 11828COD2 and 11828COD3 LPS

The LPS from the two chimeric strains was extracted (section 2.18.1) and analyzed for changes in LPS phenotype and serotype specificity compared to wild-type \textit{C. jejuni} NCTC 11828 and 11168. Analysis was carried out by SDS-PAGE gel electrophoresis (section 2.18.2) followed by silver staining (section 2.18.3) and immunoblotting (section 2.18.4) using both HS:6 and HS:2 antiserum. Silver staining (Figure 5.6A) showed that there was no change in the mobility of the core molecules produced by chimeric strains compared with those produced by both wild-type strains. Immunoblotting with HS:6 antiserum revealed that both chimeric strains no longer produced a detectable O-chain, but also showed that core mobility and intensity was unchanged compared to wild-type \textit{C. jejuni} NCTC 11828 (Figure 5.6B). The altered LPS produced by 11828COD2 and 11828COD3 did not react with HS:2 antiserum (data not shown) indicating no change in serotype specificity.

5.5. Mutation of \textit{wlaJ} in the chimeric strains 11828COD2 and 11828COD3

The phenotypic data from 11828COD2 and 11828COD3 suggests that insertion of \textit{wlaJ} from the LOS producer \textit{C. jejuni} NCTC 11168 into the chromosome of NCTC 11828 affects the biosynthesis of O-chain, but does not affect core oligosaccharide. To confirm this, a strategy was devised to mutate \textit{wlaJ} in the two chimeric strains by the insertion of a kanamycin resistance cassette derived from \textit{C. coli} (Trieu-Cuot \textit{et al.}, 1985). This was achieved by electro-transforming the two chimeric strains with plasmids pAWL611 and pAWL612 (Figure 5.7). These plasmids had been constructed previously to mutate \textit{wlaJ} in \textit{C. jejuni} NCTC 11168 (Wood \textit{et al.}, 1999). Thus, four new strains would be created from 11828COD2 and 11828COD3; these were called COD2KF, COD2KR, COD3KF and COD3KR respectively (Figure 5.8). Transformants were selected on CSA plates containing both chloramphenicol and kanamycin. The chromosomal DNA from several putative positive clones was extracted and used for PCR and Southern analysis. PCR using WLAF1, WLAKR3 and primers which anneal to the \textit{cat} cassette identified several correct recombinants which contained the kan
Figure 5.6: LPS profiles of the chimeric *C. jejuni* strains, 11828COD2 and 11828COD3. Samples were electrophoresed on a 15% SDS-PAGE gel and subsequently detected using (A) Silver staining or (B) Immunoblotting using HS:6 antiserum. Marker sizes are shown on the left hand side of each panel.

A: Silver stain. Lane 1, 11828COD2; Lane 2, 11828COD3; Lane 3, *C. jejuni* NCTC 11828; Lane 4, *C. jejuni* NCTC 11168. No change in the mobility of the core molecules synthesized by the chimeric strains is apparent compared with the wild-type molecules. *C. jejuni* O-chain is not detectable using silver staining.

B: Immunoblot with HS:6 antiserum. Lane 1, 11828COD2; Lane 2, 11828COD3; Lane 3, *C. jejuni* NCTC 11828; Lane 4, *C. jejuni* NCTC 11168. Both chimeric strains no longer produced a detectable O-chain. Core mobility and intensity is unchanged compared to wild-type *C. jejuni* NCTC 11828. The LOS produced by *C. jejuni* NCTC 11168 is undetectable using HS:6 antiserum.
Chapter 5: Functional analysis of \textit{wlaJ}

Figure 5.7: Schematic diagram of plasmids pAWL611 and pAWL612 (Wood \textit{et al.}, 1999). Both contain the \textit{wlaH} - \textit{wlaL} region from NCTC 11168 with the \textit{wlaJ} gene insertionally inactivated by a kanamycin cassette. pAWL611 and pAWL612 were used to mutate \textit{wlaJ} in the chimeric strains 11828COD2 and 11828COD3.
Figure 5.8: Schematic diagram showing the gene content (and primer positions) for the \textit{wlaG} - \textit{wlaK} region of strains COD2KF, COD2KR, COD3KF and COD3KR. Blue arrows represent NCTC 11828 gene sequence, red arrows represent NCTC 11168 gene sequence and black arrows represent antibiotic resistance cassettes. Genes represented by both red and blue arrows (\textit{i.e. wlaH} and \textit{wlaK}) represent sites of recombination. In the absence of sequence data from both strains to pinpoint the exact sites, for the purposes of this diagram, the maximum possible extent of genetic exchange is shown. Primer positions are also indicated. This diagram is not shown to scale.
Chapter 5: Functional analysis of \textit{wlaJ}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure5.png}
\caption{Diagram showing the primer positions and PCR amplification of the \textit{wlaJ} gene.}
\end{figure}

\begin{table}
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
Primer Position & COD2KF & COD2KR & COD3KF & COD3KR \\
\hline
\textit{wlaG} & \textit{wlaH} & \textit{wlaJ} & \textit{cat} & \textit{wlaK} \\
\hline
\end{tabular}
\caption{Primer positions for PCR amplification.}
\end{table}
cassette recombined in the correct position, and which had the \textit{wlaJ} and \textit{cat} genes still present. PCR analysis of these four new strains is shown in Figures 5.9 and 5.10 respectively. Southern hybridization using the \textit{wlaJ} gene and the kan cassette as probes also established that the four strains were correct (Figure 5.11), and confirmed the orientation of the kan cassette. The chromosomal DNA from the four strains was also probed with pUC19 to ensure that no plasmid sequence remained. No pUC19 sequence was detected (data not shown).

5.6. Phenotypic analysis of COD2KF, COD2KR, COD3KF and COD3KR LPS

The LPS from the four chimeric strains was extracted and analyzed for changes compared to \textit{C. jejuni} NCTC 11828 and NCTC 11168. Analysis was carried out by SDS-PAGE followed by silver staining and immunoblotting using HS:6 antiserum. Silver staining showed that there was no change in the mobility or intensity of the core molecules produced by the four strains compared with those produced by \textit{C. jejuni} NCTC 11828 (data not shown). Immunoblotting with HS:6 antiserum also showed that core mobility and intensity was unchanged, but in addition, showed that none of the four strains produced a detectable O-chain (Figure 5.12). Therefore, mutation of \textit{wlaJ} in 11828COD2 and 11828COD3 by insertional inactivation did not restore the O-chain producing phenotype.

Assuming that insertion of the kan cassette successfully inactivated the \textit{wlaJ} gene in the four chimeric strains, the possibility remained that it was the insertion of the \textit{cat} cassette into the locus which was causing the loss of O-chain biosynthesis in the strains 11828COD2 and 11828COD3, and not the insertion of the \textit{wlaJ} gene. Other \textit{cat}-insertion mutants in \textit{C. jejuni} NCTC 11828 (for example \textit{wlaL}; Fry \textit{et al.}, 2000b) have no altered O-chain which suggests that it is not the product of the \textit{cat} cassette that leads to this effect.

The \textit{cat} cassette contains a promoter to ensure high levels of expression of resistance determinants, but does not contain a transcriptional terminator. The absence of a terminator is to help minimize possible effects on the expression of downstream
Chapter 5: Functional analysis of the polymorphic gene *wla*J

Figure 5.9: PCR analysis of strains COD2KF and COD2KR. This confirmed the insertional inactivation of the *wla*J gene of 11828COD2 using the kan cassette. PCR products were separated by horizontal gel electrophoresis (0.8\% agarose). Primer position and orientation is shown in Figure 5.8. Marker sizes, in kilobase pairs, are shown on the left hand side of each panel.

**A: PCR analysis of strain COD2KF.** Lane 1. *Hind* III. φ x *Hae*III marker; Lane 2. WLAHF1 WLAKR3 primer combination using COD2KF DNA produces a product of 4.63 kb, 1.5 kb greater than the product from the parent strain (11828COD2; Lane 4). This size difference corresponds to the size of the kan cassette; Lane 3. WLAHF1 CAT13' primer combination using COD2KF DNA produces a product of 3.4 kb. This product is 1.5 kb larger than that from the parent strain (11828COD2; Lane 5) confirming the insertion of the kan cassette; Lane 4. WLAHF1/WLAKR3 primer combination using 11828COD2 DNA produces a fragment of 3.13 kb; Lane 5. WLAHF1 CAT13' primer combination using 11828COD2 DNA produces a fragment of 1.9 kb; Lane 6. Negative control; Lane 7. *Hind* III. φ x *Hae*III.

**B: PCR analysis of strain COD2KR.** Lane 1. *Hind* III. φ/x *Hae*III marker; Lane 2. WLAHF1 WLAKR3 primer combination using COD2KR DNA produces a product of 4.63 kb, 1.5 kb greater than the product from the parent strain (11828COD2; Lane 4). This size difference corresponds to the size of the kan cassette; Lane 3. WLAHF1 CAT13' primer combination using COD2KR DNA produces a product of 3.4 kb. This product is 1.5 kb larger than that from the parent strain (11828COD2; Lane 5) confirming the insertion of the kan cassette; Lane 4. WLAHF1/WLAKR3 primer combination using 11828COD2 DNA produces a fragment of 3.13 kb; Lane 5. WLAHF1/CAT13' primer combination using 11828COD2 DNA produces a fragment of 1.9 kb; Lane 6. Negative control; Lane 7. *Hind* III. φ/x *Hae*III marker.
Chapter 5: Functional analysis of \textit{wlaJ}

A

\begin{tabular}{cccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 \\
23.13 & 9.42 & 6.56 & 4.36 & 2.32 & 2.03 & 1.35 & 1.08 & 0.87 & 0.60
\end{tabular}

B

\begin{tabular}{cccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 \\
23.13 & 9.42 & 6.56 & 4.36 & 2.32 & 2.03 & 1.35 & 1.08 & 0.87
\end{tabular}
Figure 5.10: PCR analysis of strains COD3KF and COD3KR. This confirmed the insertional inactivation of the wlaJ gene of 11828COD3 using the kan cassette. PCR products were separated by horizontal gel electrophoresis (0.8% agarose). Primer position and orientation is shown in Figure 5.8. Marker sizes, in kilobase pairs, are shown on the left hand side of each panel.

**A: PCR analysis of strain COD3KF.** Lane 1. $\lambda$ Hind III, $\phi$ x HaelIII marker; Lane 2. W1.AHF1 W1.AKR3 primer combination using COD3KF DNA produces a product of 4.63 kb, 1.5 kb greater than the product from the parent strain (11828COD3: Lane 4). This size difference corresponds to the size of the kan cassette: Lane 3. W1.AHF1 CAT15' primer combination using COD3KF DNA produces a product of 3.45 kb. This product is 1.5 kb larger than that from the parent strain (11828COD3: Lane 5) confirming the insertion of the kan cassette: Lane 4. W1.AHF1/W1.AKR3 primer combination using 11828COD3 DNA produces a fragment of 3.13 kb: Lane 5. W1.AHF1/CAT15' primer combination using 11828COD3 DNA produces a fragment of 1.95 kb: Lane 6. Negative control: Lane 7. $\lambda$ Hind III, $\phi$ x HaelIII.

**B: PCR analysis of strain COD3KR.** Lane 1. $\lambda$ Hind III, $\phi$ x HaelIII marker; Lane 2. W1.AHF1 W1.AKR3 primer combination using COD3KR DNA produces a product of 4.63 kb, 1.5 kb greater than the product from the parent strain (11828COD3: Lane 4). This size difference corresponds to the size of the kan cassette: Lane 3. W1.AHF1/CAT15' primer combination using COD3KR DNA produces a product of 3.45 kb. This product is 1.5 kb larger than that from the parent strain (11828COD3: Lane 5) confirming the insertion of the kan cassette: Lane 4. W1.AHF1/W1.AKR3 primer combination using 11828COD3 DNA produces a fragment of 3.13 kb: Lane 5. W1.AHF1/CAT15' primer combination using 11828COD3 DNA produces a fragment of 1.95 kb: Lane 6. Negative control: Lane 7. $\lambda$ Hind III, $\phi$ x HaelIII marker.
Chapter 5: Functional analysis of \textit{wlaJ}

A

\begin{tabular}{cccccc}
1 & 2 & 3 & 4 & 5 & 7 \\
23.13 & 9.42 & 6.56 & 4.36 & 2.32 & 2.03 & 1.35 & 1.08 & 0.87 & 0.60
\end{tabular}

B

\begin{tabular}{cccccc}
1 & 2 & 3 & 4 & 5 & 7 \\
23.13 & 9.42 & 6.56 & 4.36 & 2.32 & 2.03 & 1.35 & 1.08 & 0.87 & 0.60
\end{tabular}
Figure 5.11: Southern hybridization data from strains COD2KF, COD2KR, COD3KF and COD3KR. This confirmed the correct recombination of pAW1.611 and pAW1.612 into the chromosome of C. jejuni strains 11828COD2 and 11828COD3. DNA was restricted with Cla I overnight before being separated by horizontal gel electrophoresis (0.8% agarose). Lanes containing irrelevant information have been deleted. Marker sizes, in kilobase pairs, are shown on the left hand side of each panel.

A: Southern hybridization analysis of strains COD2KF, COD2KR, COD3KF and COD3KR using the kan cassette (and λ Hind III marker) as probe. Lane 1, λ Hind III marker; Lane 2, COD3KF yields a hybridizing Cla I fragment of 2.75 kb so confirming the presence and forward orientation of the kan cassette; Lane 3, COD3KR yields a hybridizing Cla I fragment of 4 kb confirming the presence and reverse orientation of the kan cassette; Lane 4, no hybridizing band is present in strain 11828COD3 confirming the absence of the kan cassette in this strain; Lane 5, COD2KF yields a hybridizing Cla I fragment of 2.2 kb so confirming the presence and forward orientation of the kan cassette; Lane 6, COD2KR yields a hybridizing Cla I fragment of 4 kb confirming the presence and reverse orientation of the kan cassette; Lane 7, no hybridizing band is present in strain 11828COD2 confirming the absence of the kan cassette in this strain.

B: Southern hybridization analysis of strains COD2KF, COD2KR, COD3KF and COD3KR using the wlaJ gene (and λ Hind III marker) as probe. This probe was produced by PCR from the NCTC 11168 wlaJ gene using primers WLAJF1/WLAJR1 (see Appendix 1). Lane 1, λ Hind III marker; Lane 2, COD3KF yields a hybridizing Cla I fragment of 2.75 kb; Lane 3, COD3KR yields a hybridizing Cla I fragment of 1.4 kb; Lane 4, strain 11828COD3 yields a 4 kb Cla I fragment; Lane 5, COD2KF yields a hybridizing Cla I fragment of 2.2 kb; Lane 6, COD2KR yields a hybridizing Cla I fragment of 0.8 kb; Lane 7, strain 11828COD2 yields a 3.5 kb Cla I fragment.
Chapter 5: Functional analysis of wlaJ

A

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>23.13</td>
<td>9.42</td>
<td>6.56</td>
<td>4.36</td>
<td>2.32</td>
<td>2.03</td>
<td>0.60</td>
</tr>
</tbody>
</table>

B

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>23.13</td>
<td>9.42</td>
<td>6.56</td>
<td>4.36</td>
<td>2.32</td>
<td>2.03</td>
<td>0.60</td>
</tr>
</tbody>
</table>
Figure 5.12: LPS profiles of the four *C. jejuni* strains COD2KF, COD2KR, COD3KF and COD3KR. Samples were electrophoresed on a 15% SDS-PAGE gel and subsequently detected by immunoblotting using HS:6 antiserum. Marker sizes are shown in to the left of the figure. Lane 1, COD2KF; Lane 2, COD2KR; Lane 3, *C. jejuni* NCTC 11828; Lane 4, COD3KF; Lane 5, COD3KR; Lane 6, *C. jejuni* NCTC 11168. Mutation of *wlaJ* in 11828COD2 and 11828COD3 by insertional inactivation to produce COD2KF, COD2KR, COD3KF and COD3KR does not restore the O-chain producing phenotype. Core mobility and intensity is unchanged compared to wild-type *C. jejuni* NCTC 11828. The LOS produced by *C. jejuni* NCTC 11168 is undetectable using HS:6 antiserum.
Chapter 5: Functional analysis of wlaJ

genes by the insertion of a resistance cassette in the same orientation as the target operon. When the cassette is in the opposite orientation to the direction of transcription of the operon the lack of O-chain could possibly be due to effects on the expression of the downstream genes (wlaK, L and M), but since the same phenotype is observed whether the cassette is in the same transcriptional direction as the operon (and hence hopefully not affecting downstream genes) this is unlikely. Having eliminated the product and orientation of the cat cassette as being responsible, there remains the possibility that is the position of the cat insertion that is leading to the loss of O-chain. To test this hypothesis plasmids were constructed to allow the relocation of only the cat cassette to the same position of the NCTC 11828 locus.

5.7. Construction of plasmids to allow the relocation of the cat cassette to the NCTC 11828 locus

Plasmids pCOD4 and pCOD5 were derived from pCOD2 and pCOD3 respectively and their construction is shown in Figure 5.13. Briefly, the wlaJ gene was deleted from both pCOD2 and pCOD3 plasmids using the Expand™ High Fidelity PCR methodology described in section 2.14.2 with primers WLAJF3 and WLAJR3 (Appendix 1). These primers were designed such that the wlaJ gene could be deleted without effecting the adjacent open reading frames (Figure 5.14). The 4.3 kb amplified fragments containing pUC19, cat and flanking sequence from wlaI and wlaK was purified by horizontal gel electrophoresis followed by extraction using the NUCLEOTRAP extraction kit. The purified DNA was then digested with Bgl II overnight, precipitated and re-ligated, before being transformed into E. coli DH5α cells by electroporation. Positive clones were selected on LA containing ampicillin and chloramphenicol. The plasmid DNA from potential positive clones was extracted using the alkaline extraction method and used for restriction analysis and sequencing (section 2.15; data not shown). This data confirmed that one of the clones contained the desired construct with the cat cassette in the forward orientation (named pCOD5), and one containing the desired construct with the cat cassette in the reverse orientation (named
Chapter 5: Functional analysis of the polymorphic gene *wlaJ*

Figure 5.13: Diagrammatic representation showing the construction of pCOD4 and pCOD5. The *wlaJ* gene was deleted from plasmids pCOD2 and pCOD3 by inverse PCR using the primers WLAJF3 and WLAJR3. The 4.3 kb amplified fragments containing pUC19, *cat* and flanking sequence from *wlaJ* and *wlaK* were digested with *Bgl* II and re-ligated, thus reforming the *Bgl* II (*) site, before being transformed into *E. coli* DH5α cells by electroporation.
Chapter 5: Functional analysis of \textit{wlaJ}

\textbf{pCOD2} 
4887 bps

\textbf{pCOD3} 
4887 bps

\textbf{pCOD4} 
4312 bps

\textbf{pCOD5} 
4312 bps
Figure 5.14: Diagrammatic representation showing the annealing sites of WLAJ3 (A) and WLAJF3 (B) - primers used to delete the \textit{wlaJ} gene from pCOD2 and pCOD3. Annealing sites are represented by bold nucleotides; \# represent 'stop' codons. The position of both primers (shown by arrows) enables the deletion of as much as possible of the \textit{wlaJ} gene without effecting adjacent open reading frames.
Chapter 5: Functional analysis of the polymorphic gene wlaJ
pCOD4). Restriction enzyme digestions confirming the correct construction of pCOD4 and pCOD5 are shown in Figure 5.15.

5.8. Transformation of C. jejuni NCTC 11828 with pCOD4 and pCOD5

pCOD4 and pCOD5 were transformed into C. jejuni NCTC 11828 by electroporation to create strains 11828COD4 and 11828COD5 (Figure 5.16). Positive clones were selected on CSA plates containing chloramphenicol. The chromosomal DNA from putative positive clones was extracted and used for PCR and Southern hybridization analysis. PCR using WLAHF1, WLAKR3 and primers which anneal to the cat cassette identified several correct recombinants which contained cat recombined in the correct position between wlaI and wlaK. The PCR data for one of the clones containing cat in the forward orientation (strain 11828COD5) and one clone containing cat in the reverse orientation (strain 11828COD4) is shown in Figure 5.17. Southern hybridization using the wlaJ and cat genes as probes also confirmed that strains 11828COD5 and 11828COD4 were correct and contained the cat gene but not the wlaJ gene (Figure 5.18). Due to the presence of an internal ClaI restriction site in the cat cassette the orientation of the cassette could be also confirmed. The chromosomal DNA from both strains was also probed with pUC19 to ensure that no plasmid sequence remained. No pUC19 sequence was detected (data not shown).

5.9. Phenotypic analysis of 11828COD4 and 11828COD5 LPS

The LPS from these two strains was extracted and analyzed for changes compared to the parent strain C. jejuni NCTC 11828. Analysis was carried out by SDS-PAGE gel followed by silver staining and immunoblotting using HS:6 antiserum. Silver staining showed that there was no change in the mobility or intensity of the core molecules produced by strains containing the cat cassette inserted between wlaI - wlaK compared with those produced by the wild-type strain C. jejuni NCTC 11828 (data not shown). Immunoblotting with HS:6 antiserum also confirmed that the core molecules produced by the strains 11828COD4 and 11828COD5 were unchanged.
Figure 5.15: Agarose gel electrophoresis (1% gel) to show the restriction enzyme digestion of pCOD4 and pCOD5. Marker sizes, in kilobase pairs, are shown on the left hand side of each panel.

**A: Restriction enzyme digestion of pCOD4 confirming the removal of the wlaJ gene from pCOD2 using inverse PCR.** Lane 1. \( \lambda \) Hind III, \( \phi \times \) HaeIII marker; Lane 2. Bgl II digested pCOD4 yielding a fragment of 4312 bp; Lane 3. EcoRI digest yielding fragments of 3064 bp, 814 bp and 434 bp; Lane 4. Hind III digest yielding fragments of 3030 bp, 997 bp and 285 bp.

**B: Restriction enzyme digestion of pCOD5 confirming the removal of the wlaJ gene from pCOD3 using inverse PCR.** Lane 1. \( \lambda \) Hind III, \( \phi \times \) HaeIII marker; Lane 2. Bgl II digested pCOD5 yielding a fragment of 4312 bp; Lane 3. EcoRI digest yielding fragments of 3064 bp, 814 bp and 434 bp; Lane 4. Hind III digest yielding fragments of 3030 bp, 675 bp and 607 bp.
Chapter 5: Functional analysis of wlaJ

A

B

23.13  9.42  6.56  4.36  2.32  2.03  1.35  1.08  0.87  0.60

23.13  9.42  6.56  4.36  2.32  2.03  1.35  1.08  0.87  0.60
**Figure 5.16:** Schematic diagram showing the gene content (and primer positions) for the *wlaG* - *wlaK* region of strains 11828COD4 and 11828COD5. Blue arrows represent NCTC 11828 gene sequence, red arrows represent NCTC 11168 gene sequence and black arrows represent the chloramphenicol acetyl transferase gene (*cat*). Genes represented by both red and blue arrows (i.e. *wlaI* and *wlaK*) represent sites of recombination. In the absence of sequence data from both strains to pinpoint the exact sites, for the purposes of this diagram, the maximum possible extent of genetic exchange is shown. Primer positions are also indicated. This diagram is not shown to scale.
Figure 5.17: PCR analysis of strains 11828COD4 and 11828COD5. This confirmed the correct recombination of pCOD4 and pCOD5 into the chromosome of *C. jejuni* NCTC 11828 and also confirmed the orientation of the inserted *cat* cassette. PCR products were separated by horizontal gel electrophoresis (0.8% agarose). Primer position and orientation is shown in Figure 5.16. Marker sizes, in kilobase pairs, are shown on the left hand side of each panel.

A: PCR analysis of strain 11828COD4. Lane 1. *Hind* III. φ x *HaeIII* marker; Lane 2. WLH1F1WLAKR3 primer combination using 11828COD4 DNA produces a product of 2.56 kb, 0.88 kb greater than the product from the wild-type strain (Lane 4). This size difference equals the size of the *cat* insertion; Lane 3. WLH1F1/CA113' primer combination using 11828COD4 DNA produces a product of 1.33 kb confirming the presence and position of the *cat* insertion. This also confirms the reverse orientation of the *cat* cassette; Lane 4. WLH1F1WLAKR3 primer combination using NCTC 11828 DNA produces a fragment of 1.68 kb; Lane 5. WLH1F1/CA113’ primer combination using NCTC 11828 DNA produces no fragment confirming the absence of the *cat* cassette in this wild-type strain; Lane 6. Negative control; Lane 7. *Hind* III. φ x *HaeIII* marker.

B: PCR analysis of strain 11828COD5. Lane 1. *Hind* III. φ/x *HaeIII* marker; Lane 2. WLH1F1WLAKR3 primer combination using 11828COD5 DNA produces a product of 2.56 kb, 0.88 kb greater than the product from the wild-type strain (Lane 4). This size difference equals the size of the *cat* insertion; Lane 3. WLH1F1/CA115’ primer combination using 11828COD5 DNA produces a product of 1.38 kb confirming the presence and position of the *cat* insertion. This also confirms the forward orientation of the *cat* cassette; Lane 4. WLH1F1/WLAKR3 primer combination using NCTC 11828 DNA produces a fragment of 1.68 kb; Lane 5. WLH1F1/CA115’ primer combination using NCTC 11828 DNA produces no fragment confirming the absence of the *cat* cassette in this wild-type strain; Lane 6. Negative control; Lane 7. *Hind* III. φ/x *HaeIII* marker.
Chapter 5: Functional analysis of \textit{wlaJ}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figureA}
\caption{A}
\end{figure}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figureB}
\caption{B}
\end{figure}
Chapter 5: Functional analysis of the polymorphic gene \textit{wlaJ}

**Figure 5.18: Southern hybridization data from strains 11828COD4 and 11828COD5.** This confirmed the correct recombination of pCOD4 and pCOD5 into the chromosome of \textit{C. jejuni} NCTC 11828 and also confirmed the orientation of the \textit{cat} cassette. DNA was restricted with \textit{Cla I} overnight before being separated by horizontal gel electrophoresis (0.8\% agarose). These panels have been split into two, reflecting the requirement for differing exposure times to produce the best results from all lanes. Marker sizes, in kilobase pairs, are shown on the left hand side of each panel.

**A: Southern hybridization analysis of strains 11828COD4 and 11828COD5 using the \textit{cat} cassette (and \textit{\lambda Hind III} marker) as probe.** Two hybridizing bands are expected for each mutant as the \textit{cat} cassette contains a \textit{Cla I} site. Lane 1. \textit{\lambda Hind III} marker: Lane 2. 11828COD4 yields hybridizing \textit{Cla I} fragments of 2.7 kb and 1.1 kb so confirming the presence and reverse orientation of the \textit{cat} cassette; Lane 3. 11828COD5 yields hybridizing \textit{Cla I} fragments of 3.25 kb and 0.56 kb confirming the presence and forward orientation of the \textit{cat} cassette; Lanes 4 & 5. no hybridizing bands are present in the strains NCTC 11828 (Lane 4) and NCTC 11168 (Lane 5); Lanes 6 & 7. digestion of pCOD4 and pCOD5 with \textit{Cla I} produces hybridizing bands of 4.3 kb; Lane 8. \textit{\lambda Hind III} marker.

**B: Southern hybridization analysis of strains 11828COD4 and 11828COD5 using the \textit{wlaJ} gene (and \textit{\lambda Hind III} marker) as probe.** This probe was produced by PCR from the NCTC 11168 \textit{wlaJ} gene using primers WLAJF1/WLAJR1 (see Appendix 1). Lane 1. \textit{\lambda Hind III} marker: Lanes 2 & 3. 11828COD4 and 11828COD5 yield no hybridizing bands so confirming the absence of the \textit{wlaJ} gene; Lane 4. no hybridizing band is present in the wild-type strain NCTC 11828; Lane 5. \textit{C. jejuni} NCTC 11168 produces a hybridizing band of 3.5 kb; Lanes 6 & 7. digestion of pCOD4 and pCOD5 with \textit{Cla I} produces no hybridizing bands confirming the absence of \textit{wlaJ} from these plasmids; Lane 8. \textit{\lambda Hind III} marker.
Chapter 5: Functional analysis of \textit{wlaJ}

A

\begin{tabular}{cccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
\end{tabular}

\begin{tabular}{cccccccc}
23.13 & 9.42 & 6.56 & 4.36 & 2.32 & 2.03 & 0.60 & \\
\end{tabular}

B

\begin{tabular}{cccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
\end{tabular}

\begin{tabular}{cccccccc}
23.13 & 9.42 & 6.56 & 4.36 & 2.32 & 2.03 & 0.60 & \\
\end{tabular}
Chapter 5: Functional analysis of \textit{wlaJ}

compared to the parent strain, and revealed that both recombinant strains produced a detectable O-chain which was identical in intensity and mobility to the wild-type O-chain (Figure 5.19). Therefore, insertion of a \textit{cat} cassette in either orientation between the \textit{wlaJ} and \textit{wlaK} genes of \textit{C. jejuni} NCTC 11828 does not affect O-chain biosynthesis. This evidence combined with the fact that other \textit{cat}-insertion mutants in \textit{C. jejuni} NCTC 11828 (for example \textit{wlaL}; Fry, 1997; Fry \textit{et al.}, 2000b) have no altered O-chain suggests that it is not the product, position or orientation of the \textit{cat} cassette that leads to the loss of O-chain in the chimeric strains so far constructed.

The phenotypes of the 11828COD4 and 11828COD5 recombinants also eliminated another possibility. One potential problem with transferring \textit{wlaJ} from NCTC 11168 to NCTC 11828 was the effect this would have on the upstream and downstream genes. In order to recombine \textit{wlaJ} into the NCTC 11828 locus, approximately 500 bp of flanking sequence was needed each side of the gene to facilitate homologous recombination. Therefore, by recombining in \textit{wlaJ}, the \textit{wlaL} and \textit{wlaK} genes could also be effected, because although the in-coming NCTC 11168 DNA has a high level of identity to the recipient NCTC 11828 DNA, the sequence is not identical (data not shown). Therefore, small alterations to the \textit{wlaL} and \textit{wlaK} genes in the recombinant strains could have effected O-chain biosynthesis. However, the fact that the 11828COD4 and 11828COD5 strains produce O-chain means that it is unlikely that the loss of O-chain in the strains 11828COD2 and 11828COD3 was due to changes in \textit{wlaL} or \textit{wlaK}.

The fact that strains 11828COD4 and 11828COD5 do produce O-chain would therefore suggest that it is the presence of the \textit{wlaJ} gene in strains 11828COD2 and 11828COD3 that is leading to the loss of O-chain. Such a conclusion is not however supported by the chimeric strains COD2KF, COD2KR, COD3KF and COD3KR, as these strains contain a insertionally inactivated \textit{wlaJ} gene. The phenotypes of these four strains could perhaps be explained if the insertional mutagenesis of \textit{wlaJ} did not completely abolish \textit{wlaJ} activity - hence O-chain would be still absent in these strains. The plasmids used to mutate the \textit{wlaJ} gene in strains 11828COD2 and 11828COD3 were pAWL611 and pAWL612. These plasmids contain a kanamycin cassette inserted after amino acid 28 of the predicted protein sequence of the \textit{wlaJ} gene, thus dividing
Figure 5.19: LPS profiles of the *C. jejuni* strains, 11828COD4 and 11828COD5. Samples were electrophoresed on a 15% SDS-PAGE gel and subsequently detected by immunoblotting using HS:6 antiserum. Marker sizes are shown to the left of the figure. Lane 1, 11828COD4; Lane 2, 11828COD5; Lane 3, *C. jejuni* NCTC 11828; Lane 4, *C. jejuni* NCTC 11168. The LPS produced by 11828COD4 and 11828COD5 is identical compared to the wild-type strain NCTC 11828. Therefore, insertion of a cat cassette in either orientation between the wlaI and wlaK genes of *C. jejuni* NCTC 11828 does not affect O-chain or core biosynthesis. The LOS produced by *C. jejuni* NCTC 11168 is undetectable using HS:6 antiserum.
Chapter 5: Functional analysis of \textit{wlaJ}

the open-reading frame into two parts. Either of these two partial open-reading frames could be responsible for the hypothesized \textit{wlaJ} activity in strains COD2KF, COD2KR, COD3KF and COD3KR. However it does seem unlikely that this hypothesized \textit{wlaJ} activity is due to the 5' portion of \textit{wlaJ}, given that this portion of the gene produces a product only 28 amino acids long.

It is also unlikely that the 3' portion could be responsible, because it would require a second transcriptional start codon and Shine-Dalgarno sequence being present within this portion of \textit{wlaJ}. It is also doubtful that the 3' portion is responsible given that the same phenotype is observed in all four strains, irrespective of the orientation of the resistance cassettes. It is possible that in strains with the kan cassette in the same orientation as the \textit{wlaJ} gene, that read-through of the 3' portion of the \textit{wlaJ} gene could occur. The fact that strains which have the kan cassette in the opposite orientation, where read-through of the 3' portion of \textit{wlaJ} should not occur, have the same phenotype suggests that expression of the 3' portion of \textit{wlaJ} is not responsible for the lack of O-chain.

Another possible explanation for the lack of O-chain in the chimeric strains COD2KF, COD2KR, COD3KF and COD3KR is that whilst the involvement of changes to the \textit{wlaJ} and \textit{wlaK} genes brought about by the recombining of NCTC 11168 sequence in to the NCTC 11828 locus has been eliminated as being a reason for the loss of O-chain, plasmids pAWL611 and pAWL612 contain NCTC 11168 sequence from \textit{wlaH - wlaL}. Using these plasmids to mutate \textit{wlaJ} in 11828COD2 and 11828COD3 could have led to the resulting strains - COD2KF, COD2KR, COD3KF and COD3KR also containing NCTC 11168 DNA sequence from \textit{wlaH - wlaK} (Figure 5.8). The NCTC 11168 DNA sequence in this region has a high level of identity to the NCTC 11828 DNA sequence, and any differences between the sequences rarely lead to differences in the amino acid sequence (data not shown). However, it remained a possibility that by altering the \textit{wlaH - wlaK} region by recombining in DNA sequence from a \textit{C. jejuni} strain that wasn’t thought to produce O-chain (NCTC 11168), and thus removing DNA from NCTC 11828 - a strain that does produce O-chain, that O-chain biosynthesis could be effected.
Yet another possibility for the absence of O-chain in strains COD2KF, COD2KR, COD3KF and COD3KR is that it is the product, orientation or position of the kan cassette that is interfering with O-chain biosynthesis and thus is masking the effect of having insertionally inactivated the \textit{wlaJ} gene.

5.10. Transformation of \textit{C. jejuni} NCTC 11828 with pAWL611 and pAWL612

To confirm what effects the insertionally inactivated \textit{wlaJ} gene, the kan cassette and the alteration of the NCTC 11828 \textit{wlaH - wlaK} region may have on O-chain, plasmids pAWL611 and pAWL612 were electroporated into \textit{C. jejuni} NCTC 11828 to create strains 11828AWL611 and 11828AWL612 (Figure 5.20). Positive clones were selected on CSA plates containing kanamycin. The chromosomal DNA from putative positive clones was extracted and used for PCR and Southern hybridization analysis. PCR using WLAHF1/WLAJR3 and WLAIF3/WLALR3 primer combinations identified several correct recombinants which contained the insertionally inactivated \textit{wlaJ} gene recombined in the correct position. The PCR data for one of the clones containing kan in the forward orientation (strain 11828AWL611) and one clone containing kan in the reverse orientation (strain 11828AWL612) is shown in Figure 5.21. Southern hybridization using the \textit{wlaJ} gene and kan as probes also confirmed that strains 11828AWL611 and 11828AWL612 were correct and contained the \textit{wlaJ} gene insertionally inactivated by a kan cassette (Figure 5.22). The chromosomal DNA from both strains was also probed with pUC19 to ensure that no plasmid sequence remained. No pUC19 sequence was detected (data not shown).

5.11. Phenotypic analysis of 11828AWL611 and 11828AWL612 LPS

The LPS from these two strains was extracted and analyzed for changes compared to the parent strain \textit{C. jejuni} NCTC 11828. Analysis was carried out by SDS-PAGE gel followed by silver staining and immunoblotting using HS:6 antiserum. Silver staining showed that there was no change in the mobility or intensity of the core molecules produced by strains containing the insertionally inactivated \textit{wlaJ} gene (data
Figure 5.20: Schematic diagram showing the gene content (and primer positions) for the \textit{wlaH} - \textit{wlaL} region of strains 11828AWL611 and 11828AWL612. Blue arrows represent NCTC 11828 gene sequence, red arrows represent NCTC 11168 gene sequence and black arrows represent kanamycin resistance cassettes. Genes represented by both red and blue arrows (\textit{i.e. wlaH} and \textit{wlaL}) represent sites of recombination. In the absence of sequence data from both strains to pinpoint the exact sites, for the purposes of this diagram, the maximum possible extent of genetic exchange is shown. Primer positions are also indicated. This diagram is not shown to scale.
Chapter 5: Functional analysis of the polymorphic gene \textit{wlaJ}

Figure 5.21: PCR analysis of strains 11828AW1.611 and 11828AW1.612. This confirmed the correct recombination of pAW1.611 and pAW1.612 into the chromosome of \textit{C. jejuni} NCTC 11828. PCR products were separated by horizontal gel electrophoresis (0.8\% agarose). Primer position and orientation is shown in Figure 5.20. Marker sizes, in kilobase pairs, are shown on the left hand side of each panel.

\textbf{A: PCR analysis of strain 11828AW1.611.} Lane 1. \textit{\textgamma} Hind III, \phi \times Haelll marker; Lane 2. W1.AlIF1 W1.AJR3 primer combination using 11828AW1.611 DNA produces a product of 1.12 kb confirming the presence of the \textit{wlaJ} gene; Lane 3. W1.AlIF3 W1.AJR3 primer combination using 11828AW1.611 DNA produces a product of 4.35 kb, 2.11 kb greater than the product from the wild-type strain (Lane 5). This size difference equals the size of the \textit{wlaJ} - kan insertion; Lane 4. W1.AlIF1 W1.AJR3 primer combination using NCTC 11828 DNA produces no fragment confirming the absence of \textit{wlaJ} in this strain; Lane 5. W1.AlIF3 W1.AJR3 primer combination using NCTC 11828 DNA produces a fragment of 2.24 kb; Lane 6. Negative control; Lane 7. \textit{\textgamma} Hind III, \phi \times Haelll marker.

\textbf{B: PCR analysis of strain 11828AW1.612.} Lane 1. \textit{\textgamma} Hind III, \phi \times Haelll marker; Lane 2. W1.AlIF1 W1.AJR3 primer combination using 11828AW1.612 DNA produces a product of 1.12 kb confirming the presence of the \textit{wlaJ} gene; Lane 3. W1.AlIF3 W1.AJR3 primer combination using 11828AW1.612 DNA produces a product of 4.35 kb, 2.11 kb greater than the product from the wild-type strain (Lane 5). This size difference equals the size of the \textit{wlaJ} - kan insertion; Lane 4. W1.AlIF1 W1.AJR3 primer combination using NCTC 11828 DNA produces no fragment confirming the absence of \textit{wlaJ} in this strain; Lane 5. W1.AlIF3 W1.AJR3 primer combination using NCTC 11828 DNA produces a fragment of 2.24 kb; Lane 6. Negative control; Lane 7. \textit{\textgamma} Hind III, \phi \times Haelll marker.
Chapter 5: Functional analysis of \textit{wlaJ}

A

B
Figure 5.22: Southern hybridization data from strains 11828AWL.611 and 11828AWL.612. This analysis confirmed the correct recombination of pAWL.611 and pAWL.612 into the chromosome of C. jejuni NCTC 11828. DNA was restricted with Clal overnight before being separated by horizontal gel electrophoresis (0.8% agarose). Marker sizes, in kilobase pairs, are shown on the left hand side of each panel.

**A: Southern hybridization analysis of strains 11828AWL.611 and 11828AWL.612 using the kan cassette (and λ Hind III marker) as probe.** Lane 1, λ Hind III marker; Lane 2, 11828AWL.611 yields a hybridizing Clal fragment of 2.4 kb so confirming the presence and forward orientation of the kan cassette; Lane 3, 11828AWL.612 yields a hybridizing Clal fragment of 4 kb confirming the presence and reverse orientation of the kan cassette; Lanes 4 & 5, no hybridizing bands are present in strains NCTC 11828 (lane 4) or NCTC 11168 (lane 5) confirming the absence of the kan cassette in this strains; Lanes 6, Clal digested pAWL.611 yields a hybridizing fragment of 2.4 kb; Lane 7, Clal digested pAWL.612 yields a hybridizing fragment of 6.5 kb; Lane 8, λ Hind III marker.

**B: Southern hybridization analysis of strains 11828AWL.611 and 11828AWL.612 using the wlaJ gene (and λ Hind III marker) as probe.** This probe was produced by PCR from the NCTC 11168 wlaJ gene using primers WLAJI 1WLAJR1 (see Appendix 1). Lane 1, λ Hind III marker; Lane 2, 11828AWL.611 yields a hybridizing Clal fragment of 2.4 kb confirming the presence of the disrupted wlaJ gene in this strain; Lane 3, 11828AWL.612 yields a hybridizing Clal fragment of 1 kb which confirms the presence of the disrupted wlaJ gene; Lane 4, strain NCTC 11828 produces no hybridizing fragments due to the absence of the wlaJ gene in this strain; Lane 5, NCTC 11168 yields a hybridizing Clal fragment of 3.5 kb; Lane 6, Clal digested pAWL.611 yields a hybridizing fragment of 2.4 kb; Lane 7, Clal digested pAWL.612 yields a hybridizing fragment of 1 kb; Lane 8, λ Hind III marker.
Chapter 5: Functional analysis of *wlaJ*

A

```
1  2  3  4  5  6  7  8
23.13
9.42
6.56
4.36
2.32
2.03
0.60
```

B

```
1  2  3  4  5  6  7  8
23.13
9.42
6.56
4.36
2.32
2.03
0.60
```
Chapter 5: Functional analysis of wlaJ

not shown). Immunoblotting with HS:6 antiserum also confirmed that the core molecules produced by the strains 11828AWL611 and 11828AWL612 were unchanged compared to the parent strain, and showed that both recombinant strains produced a detectable O-chain which was identical in intensity and mobility to the wild-type O-chain (Figure 5.23).

These results combined with the fact that other kan-insertion mutants in C. jejuni NCTC 11828 (for example wlaK; Fry, 1997) have no altered O-chain suggests that it is not the product, position or orientation of the kan cassette that leads to the lack of O-chain biosynthesis seen in strains COD2KF, COD2KR, COD3KF and COD3KR. The 11828AWL611 and 11828AWL612 strains also confirm that the inactivated wlaJ gene is not responsible for the lack of O-chain in these four strains since it is also present in 11828AWL611 and 11828AWL612 and in these strains O-chain is detected. The fact that strains 11828AWL611 and 11828AWL612 synthesize O-chain also makes it unlikely that it is the possible alteration of the NCTC 11828 wlaH - wlaK region due to the recombination of NCTC 11168 sequence that is leading to lack of O-chain biosynthesis seen in strains COD2KF, COD2KR, COD3KF and COD3KR. However, in the absence of sequencing data from across the wlaH - wlaK region in strains COD2KF, COD2KR, COD3KF, COD3KR, 11828AWL611 and 11828AWL612 to locate the sites of recombination, this explanation cannot be totally excluded.

5.12. The reconstruction of strains 11828COD2 and 11828COD3

The strains described in sections 5.7 - 5.11 were constructed to try and understand why after the insertion of wlaJ into the NCTC 11828 cluster and loss of O-chain biosynthesis, the subsequent insertional inactivation of the gene did not restore O-chain producing phenotype. The elimination of the potential effects of the cat and kan cassettes and the possible consequences of recombining NCTC 11168 wlaH - wlaK sequence into the NCTC 11828 cluster left only the possibility that strains containing the insertionally inactivated wlaJ gene still retained some wlaJ activity which was interfering with O-chain biosynthesis. This possibility was tested by transforming NCTC 11828 with the same constructs (pAWL611 and pAWL612) that were used to
Figure 5.23: LPS profiles of the *C. jejuni* strains, 11828AWL611 and 11828AWL612. Samples were electrophoresed on a 15% SDS-PAGE gel and subsequently detected by immunoblotting using HS:6 antiserum. Marker sizes are shown to the left of the figure. Lane 1, 11828AWL611; Lane 2, 11828AWL612; Lane 3, *C. jejuni* NCTC 11828; Lane 4, *C. jejuni* NCTC 11168. The LPS produced by 11828AWL611 and 11828AWL612 is identical compared to the wild-type strain NCTC 11828. Therefore, insertion of a kan-inactivated *wlaJ* gene into the LPS locus of *C. jejuni* NCTC 11828 does not affect O-chain or core biosynthesis. The LOS produced by *C. jejuni* NCTC 11168 is undetectable using HS:6 antiserum.
introduce an insertionally inactivated copy of the \textit{wlaJ} gene into 11828\textit{COD}2 and 11828\textit{COD}3. The resulting recombinant strains produced O-chain, thus eliminating the possibility of the insertionally inactivated \textit{wlaJ} gene still retaining some function. This leaves the possibility that during the construction of 11828\textit{COD}2 and 11828\textit{COD}3, the strains spontaneously lost the ability to produce O-chain due to a factor completely unrelated to the \textit{wlaJ} insertion. This would explain why the subsequent insertional inactivation of the \textit{wlaJ} gene made no difference to the O-chain phenotype - because it wasn’t the presence of the \textit{wlaJ} gene that caused the change in phenotype in the first place. To test for this possibility, several unsuccessful attempts were made to re-make strains 11828\textit{COD}2 and 11828\textit{COD}3 by electro-transforming \textit{C. jejuni} NCTC 11828 with plasmids p\textit{COD}2 and p\textit{COD}3. Attempts were therefore made to remake the strains by naturally transformation (section 2.13). Most \textit{C. jejuni} strains are naturally highly transformable using homologous DNA (Wang & Taylor, 1990b), so \textit{C. jejuni} NCTC 11828 was transformed with chromosomal DNA from strains 11828\textit{COD}2 and 11828\textit{COD}3. These attempts proved successful and the recombinant strains 11828\textit{COD}2\textit{N} and 11828\textit{COD}3\textit{N} obtained by selection on CSA containing chloramphenicol. These strains were confirmed as containing the \textit{wlaJ} and \textit{cat} genes recombined in the correct position by PCR (data not shown) and in addition, the \textit{wlaH} - \textit{wlaK} regions of both strains was sequenced - this data showed that both strains contained an intact copy of the \textit{wlaJ} gene (data not shown). The same region was also sequenced from the original 11828\textit{COD}2 and 11828\textit{COD}3 strains to confirm that these also contained the correctly inserted \textit{wlaJ} gene - the sequence data confirmed that this was the case (data not shown).

The LPS produced by the two recombinant strains 11828\textit{COD}2\textit{N} and 11828\textit{COD}3\textit{N} was then analyzed using SDS-PAGE followed by silver staining or immunoblotting using HS:6 antiserum and compared to the wild-type NCTC 11828 LPS. This analysis revealed that both strains produced core and O-chain indistinguishable from that of the parent strain (data not shown). This data confirmed that insertion of the intact \textit{wlaJ} gene into the \textit{C. jejuni} NCTC 11828 cluster does not effect O-chain biosynthesis, and suggested that the original 11828\textit{COD}2 and
11828COD3 strains must have lost the ability to produce O-chain in a wlaJ-independent fashion.

5.13. Discussion

The wlaJ gene represents the only difference in the gene content of 38 isolates of C. jejuni in the wlaM-waaM region. Initial results from these distribution studies seemed to suggest that there was a correlation between the presence of the gene and the LOS phenotype, and the absence of the gene and the LPS phenotype. This suggested a putative role for WlaJ in influencing the production of O-chain. In order to gain further information as to the function of the wlaJ gene, a strategy was designed to insert the gene into the LPS producing C. jejuni strain NCTC 11828. This strain lacks the wlaJ gene, with wlaI and wlaK being contiguous in the cluster (Fry, 1997; Fry et al., 1998). Plasmids pCOD2 and pCOD3 were therefore constructed to facilitate the transfer of wlaJ, a chloramphenicol resistance cassette and the necessary flanking sequence to C. jejuni NCTC 11828. The presence of the flanking sequence was required to allow the homologous recombination of the wlaJ gene + cat into the correct position of the NCTC 11828 cluster. When the LPS produced by the resulting strains 11828COD2 and 11828COD3 was analyzed by SDS-PAGE electrophoresis followed by immunoblotting with HS:6 antiserum, no O-chain was detected suggesting that the presence of wlaJ did in fact prevent O-chain biosynthesis.

To confirm this role, the wlaJ gene was mutated in strains 11828COD2 and 11828COD3 by the insertion of a kanamycin cassette after position 28 of the predicted protein sequence of wlaJ. Unexpectedly, detection by immunoblotting with HS:6 antiserum revealed that the LPS molecules produced by these strains also contained no O-chain. This result cast doubt on the role of wlaJ in preventing O-chain biosynthesis, therefore a series of strains were constructed to try and eliminate other factors which could be responsible for the loss of O-chain.

The first of these strains to be constructed were 11828COD4 and 11828COD5. These strains both produced O-chain therefore eliminated the possible role of the cat gene and the NCTC 11168 wlaI and wlaK flanking sequence. It therefore appeared that
it was the presence of the \textit{wlaJ} gene in strains 11828COD2 and 11828COD3 that was leading to the loss of O-chain. If this was the case, then the ability to produce O-chain should have been restored in strains COD2KF, COD2KR, COD3KF and COD3KR, in which the intact \textit{wlaJ} gene had been insertionally inactivated by a kanamycin resistance cassette.

One possible explanation was that the mutagenesis of \textit{wlaJ} did not completely abolish \textit{wlcuJ} activity - hence O-chain would be still absent in these strains. To test this hypothesis, plasmids pAWL611 and pAWL612 were used to confirm whether the kanamycin resistance cassette or the larger amount of NCTC 11168 flanking sequence (from \textit{wlaH} - \textit{wlaK}) could influence O-chain biosynthesis. The LPS produced by the two recombinant strains 11828AWL611 and 11828AWL612 was indistinguishable from the wild-type NCTC 11828 LPS therefore eliminating the influence of these factors on O-chain biosynthesis.

The data from these strains, and previous ones outlined above, suggested that none of the potential explanations for the four strains COD2KF, COD2KR, COD3KF and COD3KR not producing O-chain were valid - the four strains should produce O-chain but do not. This leaves the possibility that strains 11828COD2 and 11828COD3 - the parent strains of COD2KF, COD2KR, COD3KF and COD3KR, had spontaneously lost the ability to produce O-chain due to a factor completely unrelated to the \textit{wlaJ} insertion. This would explain why the subsequent insertional inactivation of the \textit{wlaJ} gene made no difference to the O-chain phenotype - because it wasn’t the presence of the \textit{wlaJ} gene that caused the change in phenotype in the first place. This hypothesis could be tested by re-making the original 11828COD2 and 11828COD3 strains. Unfortunately, electroporation of NCTC 11828 with the plasmids pCOD2 and pCOD3 did not yield any transformants - the reasons for the failure to obtain recombinants using electroporation are unclear, but are not thought to be technical. Only by naturally transforming NCTC 11828 with chromosomal DNA from 11828COD2 and 11828COD3 could the strains be re-made. After verifying that the new strains 11828COD2N and 11828COD3N contained an intact copy of the \textit{wlaJ} gene, phenotypic analysis of the LPS produced by the strains was undertaken. Comparison of the LPS profiles showed no difference between that of the chimeric strains and wild-
type NCTC 11828 - the two recombinant strains produced O-chain, confirming that the \textit{wlaJ} gene was not responsible for the loss of O-chain in the original strains 11828COD2 and 11828COD3. This conclusion supports that drawn from the distribution studies in Chapter 4 - that the presence of \textit{wlaJ} does not correlate with the LPS/LOS phenotype, and thus the gene is not involved with the production of O-chain.

It remains uncertain why the original strains containing the \textit{wlaJ} gene, 11828COD2 and 11828COD3, lost the ability to produce O-chain, apparently spontaneously. This phenomenon is clearly unrelated to the genetic locus targeted since the other strains constructed had similar genetic manipulations and still retained the O-chain producing phenotype. The spontaneous loss of the O-chain phenotype has not been observed in our laboratory to date (Julian Ketley, personal communication). It is possible that the O-chain deficient phenotype is linked to variations in the numbers of G/C residues found present in the homo-polymeric tracts of the \textit{wlaUD} and \textit{wlaUF} genes (discussed in Chapter 4) or those present in other polysaccharide biosynthesis genes. Variations in the length of the polymeric tracts may lead to phase and/or antigenic variation of LOS/LPS structures. Further experiments are needed to investigate this theory, but since the O-chain deficient phenotype appears to be stable rather than phase variable it is perhaps unlikely that the G/C polymeric tracts are involved.

The question still remains as to the function of the \textit{wlaJ} gene. The gene may encode for a protein that is involved with LOS biosynthesis. However, insertional inactivation of \textit{wlaJ} in NCTC 11168 has no detectable effect on the mobility or intensity of the core oligosaccharide as judged by immunoblotting (Wood \textit{et al.}, 1999). This has been reinforced by preliminary biochemical studies that have confirmed that \textit{wlaJ} mutant strains produce core molecules with identical sugar content to the wild-type NCTC 11168 strain (Anthony Moran, personal communication). This would suggest that the \textit{wlaJ} gene is not involved with LOS biosynthesis. However, Northern hybridization results (Anne Wood, personal communication) show that the gene is transcribed in \textit{C. jejuni} NCTC 11168 implying that the \textit{wlaJ} gene is expressed and so is likely to have a function. The gene also represents the only difference in gene content
Chapter 5: Functional analysis of \textit{wlaJ}

present in the \textit{wlaM- waaM} region of 38 isolates of \textit{C. jejuni}, which also suggests a significant function for this gene.

Alternatively the \textit{wlaJ} gene could have a role in protein glycosylation. The genes \textit{wlaF, G, H, I, K, L} and \textit{M} of \textit{C. jejuni} 81-176 have been shown to be involved with the glycosylation of multiple \textit{C. jejuni} 81-176 proteins including flagellin (Szymanski et al., 1999). From this publication, and others (Wood et al., 1999; Fry et al., 1998; Fry et al., 2000b), it appears that the \textit{wlaF- wlaM} genes could therefore have dual roles in LOS/LPS biosynthesis and protein glycosylation. The \textit{C. jejuni} strain 81-176 does not contain the \textit{wlaJ} gene as part of the LOS/glycosylation cluster (see Chapter 4; Szymanski et al., 1999), but since the surrounding genes have a role in glycosylation it is not unreasonable to suggest that \textit{wlaJ} could be involved in this function too. Whether the corresponding genes in \textit{C. jejuni} NCTC 11168 also have function in protein glycosylation has not yet been investigated, only the role in core biosynthesis has been confirmed (Wood et al., 1999). However, the NCTC 11168 \textit{wlaJ} mutant strains could be used to see if the glycosylation of proteins is in any way altered.
Chapter 6: Characterization of the \textit{waaF} genes of \textit{C. jejuni} NCTC 11168 and NCTC 11828

6.1. Introduction

Analysis of the complete genome sequence of \textit{C. jejuni} NCTC 11168 has led to the identification of a number of LOS/LPS related open-reading frames, the majority of which are found in the 30 kb cluster described in Chapter 3. In particular, this cluster contains several open reading frames with similarity to genes involved with inner core biosynthesis including \textit{gmhA}, \textit{gmhD}, \textit{gmhE}, \textit{waaC}, \textit{waaF} and \textit{galE}. The putative roles of both the \textit{galE} and \textit{waaC} genes of \textit{C. jejuni} have now been confirmed in functional studies (Klena \textit{et al}., 1998; Fry \textit{et al}., 2000a). This chapter will discuss the functional characterization of the open reading frame Cj1148, which is thought to encode the WaaF protein of \textit{C. jejuni}.

In \textit{E. coli} and \textit{S. typhimurium} the \textit{waaF} gene encodes a heptosyltransferase which catalyses the transfer of the second L-glycero-D-manno-heptose residue to the core oligosaccharide (Brabetz \textit{et al}., 1997; Sirisena \textit{et al}., 1994). Mutation of this gene stops the incorporation of this second L-glycero-D-manno-heptose which subsequently blocks the addition of other sugar residues. Thus a greatly altered LPS molecule is produced, consisting of a truncated core oligosaccharide and no O-chain (Schnaitman \& Klena, 1993). Such ‘deep rough’ \textit{Salmonella} mutants have increased sensitivity to antibiotics, detergents and bile salts (Raetz, 1996). In particular, they are highly susceptible to hydrophobic antibiotics such as novobiocin. However, mutants can become more resistant to the antibiotic if they are transformed with a plasmid containing a suitable complementing gene. Functional complementation restores the ability of the mutant to synthesize a complete core oligosaccharide and thus attach O-chain. This leads to an increased level of novobiocin resistance which allows positive selection for those cells containing the complementing plasmid. This complementation approach has led to the isolation of several \textit{waaF} genes from various bacterial species including \textit{Pseudomonas aeruginosa} (Kievit \& Lam, 1997), \textit{Bordetella pertussis} (Allen
Chapter 6: Characterization of the \textit{waaF} gene in \textit{C. jejuni}

et al., 1998) \textit{Haemophilus influenzae} (Nichols et al., 1997). \textit{Haemophilus ducreyi} (Bauer et al., 1999), and has been used to isolate the \textit{waaC} gene from \textit{C. coli} (Klena et al., 1998). The initial aim of this study was therefore to isolate Cj1148 and to support its putative identification as \textit{waaF} by using the cloned gene to complement a \textit{waaF} mutant strain of \textit{S. typhimurium}.

6.2. Isolation and cloning of Cj1148 from \textit{C. jejuni} NCTC 11168

Cj1148 was amplified from \textit{C. jejuni} NCTC 11168 by the Expand\textsuperscript{TM} High Fidelity PCR methodology described in section 2.14.2 using primers GMHAF3 and WLASAR3 (Appendix 1). The amplified fragment was then purified by horizontal gel electrophoresis (section 2.8) followed by extraction using the NUCLEOTRAP kit (section 2.9.1). The purified DNA was digested with \textit{Kpn} I and \textit{Pst} I overnight before being ligated into \textit{Kpn} I and \textit{Pst} I digested pUC19 (section 2.10.4) and electroporated into \textit{E. coli} DH5\textalpha{} cells (section 2.11). The plasmid DNA from potential positive clones was extracted using the alkaline extraction method (section 2.6.1) and used for PCR (section 2.14.1; data not shown) and restriction analysis (section 2.10.1). This data confirmed that one of the clones contained the desired construct, which was named pNOL17 (Figure 6.1). A restriction enzyme digestion of pNOL17 is shown in Figure 6.2.

6.3. Complementation of \textit{Salmonella typhimurium} SL3789 (\textit{waaF511}) with pNOL17

The \textit{S. typhimurium} strain SL3789 (\textit{waaF511}; a gift from Tomoko Isobe, Centre for Veterinary Science, University of Cambridge; Roantree et al., 1977) was transformed by electroporation with the plasmid pNOL17 containing the putative \textit{waaF} gene of \textit{C. jejuni} NCTC 11168. Transformants were selected on LA containing ampicillin (100 \textmu{}g ml\textsuperscript{-1}) and novobiocin (5 \textmu{}g ml\textsuperscript{-1}). This concentration of novobiocin was used, as it was previously found to permit the growth of the wild-type \textit{S. typhimurium} SL3770 (produces smooth LPS and so is more resistant to novobiocin),
Chapter 6: Characterization of the waaF gene in C. jejuni

Figure 6.1: Diagrammatic representation showing pNOL17. The waaF gene and approximately 500 bp of flanking sequence either side was amplified using the primers GMM1AF3 and WLASAR3. This PCR product was digested with Pst I and Kpn I before being ligated into pUC19 and transformed into E. coli DH15α.

Figure 6.2: Agarose gel electrophoresis (0.8% gel) to show the restriction enzyme digestion of pNOL17. Marker sizes, in kilobase pairs, are shown on the left hand side of the figure. Lane 1. †. Hind III. φ x Hae III marker. Lane 2. Bgl II digest yielding fragments of 3979 bp, 590 bp and 63 bp (not visible). Lane 3. Hind III digest yielding fragments of 2888 bp, 955 bp, 566 bp and three fragments of 80 bp, 74 bp and 69 bp which are not visible. Lane 4. uncut pNOL17 (4632 bp). Lane 5. †. Hind III. φ x Hae III marker.
Chapter 6: Characterization of the \textit{waaF} gene in \textit{C. jejuni}

\textbf{pNOL17}  
4632 bps
Chapter 6: Characterization of the \textit{waaF} gene in \textit{C. jejuni}

but to efficiently prevent growth of the more sensitive strain SL3789 (produces rough LPS, and so is more sensitive to novobiocin). Therefore it was expected that this concentration would permit growth of complemented cells. The plasmid DNA from potential positive clones was extracted using the alkaline extraction method and used for restriction analysis. This data confirmed that one of the clones contained pNOL17 (data not shown) and thus the clone was named SL3789 (pNOL17).

6.3.1. Phenotypic analysis of SL3789 (pNOL17) LPS

Growth of SL3789 (pNOL17) on 5 \textmu g ml\textsuperscript{-1} novobiocin implied that functional complementation of the \textit{waaF} mutation had occurred so allowing the strain to produce smooth LPS and hence become more resistant to the antibiotic. This was confirmed by analysis of the LPS from the complemented \textit{Salmonella} strain SL3789 (pNOL17). The LPS was extracted (section 2.18.1) and compared to the LPS produced by the wild-type SL3770 and mutant SL3789 strains. Analysis was carried out by SDS-PAGE gel electrophoresis (section 2.18.2) followed by silver staining (section 2.18.3). Unlike the case of \textit{C. jejuni}, both the core and O-chain of \textit{S. typhimurium} can be visualized by silver staining. This analysis (Figure 6.3) revealed the restoration of the O-chain phenotype in the complemented strain, and thus supported the identification of Cj1148 as \textit{waaF}. In addition, the rough core molecule of the mutant strain was not present in significant amounts in the LPS of the complemented strain showing that the complementation is complete. This is not always the case, when the \textit{waaF} gene of \textit{B. pertussis} was used in similar experiments, the SL3789 truncated core molecule was still seen in the LPS of the complemented strain suggesting that the functional complementation was partial (Allen \textit{et al.}, 1998).

Attempts were also made to complement \textit{S. typhimurium} SL3789 (\textit{waaF511}) with pNOL22 and pNOL23 - plasmids identical to pNOL17 except for the deletion and replacement of \textit{waaF} with a \textit{cat} cassette (section 6.4). No clones were obtained using these plasmids, confirming that it was indeed the \textit{waaF} gene that was responsible for the complementation and not any other sequence present on pNOL17.
Figure 6.3: Silver staining of the LPS produced by the complemented *S. typhimurium* strain SL3789 (pNOL17). Samples were electrophoresed on a 15% SDS-PAGE gel. Marker sizes are shown on the left hand side of the figure. Lane 1, wild-type *S. typhimurium* strain SL3770; Lane 2, mutant *S. typhimurium* strain SL3789 (*waaF*511); Lane 3, *S. typhimurium* strain SL3789 (pNOL17). Complementing a *S. typhimurium waaF* mutant with *C. jejuni* Cj1148 restores the production of O-chain supporting the identification of the gene as *waaF*. 
Chapter 6: Characterization of the \textit{waaF} gene in \textit{C. jejuni}

6.4. Mutation of \textit{waaF} in \textit{C. jejuni} NCTC 11168

Complementation of \textit{S. typhimurium} SL3789 with Cj1148 from \textit{C. jejuni} NCTC 11168 supports the identification of the open reading frame as \textit{waaF}. The next aim of this study was to mutate \textit{waaF} in \textit{C. jejuni}. This was carried out in \textit{C. jejuni} NCTC 11168 which produces LOS and is from serogroup HS:2 and was also subsequently achieved in \textit{C. jejuni} NCTC 11828 which produces LPS and is from serogroup HS:6 (section 6.5).

6.4.1. Inverse PCR mutagenesis of \textit{C. jejuni} NCTC 11168 \textit{waaF}

The \textit{waaF} gene was deleted from pNOL17 by inverse PCR using the Expand™ High Fidelity PCR methodology and primers WAAFF2 and WAAFR4 (Appendix 1). These primers were designed such that the \textit{waaF} gene could be deleted without effecting the adjacent open reading frames Cj1147 and Cj1149 (Figure 6.4). The amplified fragment was purified by horizontal gel electrophoresis followed by extraction using the NUCLEOTRAP extraction kit. The purified DNA was then digested with \textit{BamHI} overnight and treated with calf intestinal alkaline phosphatase (section 2.10.3) before being precipitated ready for ligation. The chloramphenicol acetyl transferase cassette was obtained by digesting pAV35 (Appendix 2) with \textit{BamHI} overnight. The 850 bp fragment containing the \textit{cat} cassette was then purified and ligated to the pNOL17 inverse PCR product before being transformed into \textit{E. coli} DH5α cells by electroporation. Positive clones were selected on LA containing ampicillin and chloramphenicol. The plasmid DNA from potential positive clones was extracted using the alkaline extraction method and used for PCR (data not shown) and restriction analysis. This data confirmed that one of the clones contained the desired construct with the \textit{cat} cassette in the forward orientation (named pNOL22), and one containing the desired construct with the \textit{cat} cassette in the reverse orientation (named pNOL23). Schematic diagrams of both pNOL22 and pNOL23 are shown in Figure 6.5.
Figure 6.4: Diagrammatic representation showing the annealing sites of WAAFF2 (A) and WAAFR4 (B) - primers used for the inverse PCR mutagenesis of the \textit{waaF} gene of \textit{C. jejuni} NCTC 11168. Annealing sites are represented by bold nucleotides: $\neq$ represent 'stop' codons. The position of both primers (shown by arrows) enables the deletion of as much as possible of the \textit{waaF} gene without affecting open reading frames Cj1147 and Cj1149.
Chapter 6: Characterization of the waaF gene in C. jejuni

A

Translational start of Cj1147

Translational start of Cj1148 (waaF)

B

WAAFR4

Translational end of Cj1148 (waaF)

Translational end of Cj1149
Chapter 6: Characterization of the \textit{waaF} gene in \textit{C. jejuni}

Figure 6.5: Diagrammatic representation showing pNOL22 and pNOL23. The \textit{waaF} gene was removed from pNOL17 by inverse PCR using the primers WAAFF2 and WAAFR4. The purified DNA was then digested with \textit{BamH}I before being ligated to the 850 bp \textit{BamH}I fragment from pAV35 containing the \textit{cat} cassette.
enzyme digestions confirming the correct construction of pNOL22 and pNOL23 are shown in Figure 6.6.

6.4.2. Transformation of C. jejuni NCTC 11168 with pNOL22 and pNOL23

pNOL22 and pNOL23 were transformed into C. jejuni NCTC 11168 by electroporation (section 2.12). Recombinants were selected on CSA plates containing chloramphenicol. The chromosomal DNA from several putative positive clones was extracted (section 2.6.2) and used for PCR and Southern hybridization analysis (section 2.17). PCR using GMHEF1, WLASAR4 and primers which anneal to the cat cassette identified several correct recombinants which contained the cat cassette in place of the deleted waaF gene. The PCR data for one correct clone containing cat in the forward orientation (strain 11168NOL22) and one clone containing cat in the reverse orientation (strain 11168NOL23) is shown in Figure 6.7. Southern hybridization using the waaF and cat genes as probes also confirmed that strains 11168NOL22 and 11168NOL23 were correct (Figure 6.8), and due to the presence of an internal Cla I restriction site in the cat cassette the orientation of the cassette could be confirmed. The chromosomal DNA from both strains was also probed with pUC19 to ensure that no plasmid sequence remained. No pUC19 sequence was detected (data not shown).

6.4.3. Phenotypic analysis of 11168NOL22 and 11168NOL23 LOS

The lipooligosaccharide from the two mutant strains, 11168NOL22 and 11168NOL23 was extracted and analyzed for changes compared to wild-type C. jejuni NCTC 11168. Analysis was carried out by SDS-PAGE gel followed by silver staining and immunoblotting (section 2.18.4) using HS:2 antiserum or cholera toxin. Silver staining (Figure 6.9) showed that the core molecules produced by the two mutant strains were significantly truncated compared with those produced by the wild-type, and thus migrated faster under SDS-PAGE. Immunoblotting with HS:2 anti-serum
A: Restriction enzyme digestion of pNOL22 confirming the removal of *waaF* and insertion of the *cat* cassette in the forward orientation. Lane 1. *Hind* III, φ X *HaeIII* marker; Lane 2. *BamHI* digested pNOL22 yielding fragments of 3769 bp and 850 bp. This digest confirmed the insertion of the *cat* cassette; Lane 3. *Hind* III digest yielding fragments of 2888 bp, 1179 bp, 409 bp, 74 bp and 69 bp (these latter two fragments are not visible). The fragments produced by this digestion confirmed the insertion and forward orientation of the *cat* cassette; Lane 4. *Pst I* digest yielding fragments of 3201 bp, 824 bp, and 594 bp; Lane 5. uncut pNOL22 (4619 bp); Lane 6. *Hind* III, φ X *HaeIII* marker.

B: Restriction enzyme digestion of pNOL23 confirming the removal of *waaF* and insertion of the *cat* cassette in the reverse orientation. Lane 1. *Hind* III, φ X *HaeIII* marker; Lane 2. *BamHI* digested pNOL23 yielding fragments of 3769 bp and 850 bp. This digest confirmed the insertion of the *cat* cassette; Lane 3. *Hind* III digest yielding fragments of 2888 bp, 857 bp, 731 bp, 74 bp and 69 bp (these latter two fragments are not visible). The fragments produced by this digestion confirmed the insertion and reverse orientation of the *cat* cassette; Lane 4. *Pst I* digest yielding fragments of 3201 bp, 824 bp, and 594 bp; Lane 5. uncut pNOL23 (4619 bp); Lane 6. *Hind* III, φ X *HaeIII* marker.
Chapter 6: Characterization of the \textit{waaF} gene in \textit{C. jejuni}

A

\begin{tabular}{ccccccc}
\hline
& 1 & 2 & 3 & 4 & 5 & 6 \\
\hline
\text{Marginal} & 23.13 & 9.42 & 6.56 & 4.36 & 2.32 & 2.03 \\
& 1.35 & 1.08 & 0.87 & 0.60 & 0.50 & 0.40 \\
\text{Peak} & 23.13 & 9.42 & 6.56 & 4.36 & 2.32 & 2.03 \\
\end{tabular}

B

\begin{tabular}{ccccccc}
\hline
& 1 & 2 & 3 & 4 & 5 & 6 \\
\hline
\text{Marginal} & 23.13 & 9.42 & 6.56 & 4.36 & 2.32 & 2.03 \\
& 1.35 & 1.08 & 0.87 & 0.60 & 0.50 & 0.40 \\
\text{Peak} & 23.13 & 9.42 & 6.56 & 4.36 & 2.32 & 2.03 \\
\end{tabular}
Figure 6.7: PCR analysis of strains 11168NOL22 and 11168NOL23. This confirmed the correct deletion of the \textit{waaF} gene and replacement with the \textit{cat} cassette. PCR products were separated by horizontal gel electrophoresis (0.8\% agarose). Marker sizes, in kilobase pairs, are shown on the left hand side of the figure.

\textbf{A: PCR analysis of strain 11168NOL22.} Lane 1. \textit{Hind} III. $\phi \times \text{Hae}III$ marker; Lane 2. GMHEF1 W1ASAR4 primer combination using 11168NOL22 DNA produces a product of 2.24 kb. The size of this product is virtually identical to the wild-type product (Lane 4) due to the deletion of the 863 bp \textit{waaF} fragment and insertion of the 850 bp \textit{cat} cassette; Lane 3. GMHEF1 CAT15' primer combination using 11168NOL22 DNA produces a product of 0.9 kb confirming the presence and position of the \textit{cat} cassette. This also confirmed the forward orientation of the \textit{cat} cassette; Lane 4. GMHEF1 W1ASAR4 primer combination using NCTC 11168 DNA produces a fragment of 2.25 kb; Lane 5. GMHEF1 CAT15' primer combination using NCTC 11168 DNA produces no fragment confirming the absence of the \textit{cat} cassette in this wild-type strain; Lane 6. Negative control; Lane 7. \textit{Hind} III. $\phi \times \text{Hae}III$ marker.

\textbf{B: PCR analysis of strain 11168NOL23.} Lane 1. \textit{Hind} III. $\phi \times \text{Hae}III$ marker; Lane 2. GMHEF1 W1ASAR4 primer combination using 11168NOL23 DNA produces a product of 2.24 kb. The size of this product is virtually identical to the wild-type product (Lane 4) due to the deletion of the 863 bp \textit{waaF} fragment and insertion of the 850 bp \textit{cat} cassette; Lane 3. GMHEF1 CAT15' primer combination using 11168NOL23 DNA produces a product of 0.95 kb confirming the presence and position of the \textit{cat} cassette. This also confirmed the reverse orientation of the \textit{cat} cassette; Lane 4. GMHEF1 W1ASAR4 primer combination using NCTC 11168 DNA produces a fragment of 2.25 kb; Lane 5. GMHEF1 CAT15' primer combination using NCTC 11168 DNA produces no fragment confirming the absence of the \textit{cat} cassette in this wild-type strain; Lane 6. Negative control; Lane 7. \textit{Hind} III. $\phi \times \text{Hae}III$ marker.
Chapter 6: Characterization of the \textit{waaF} gene in \textit{C. jejuni}

A

B
Figure 6.8: Southern hybridization data from strains 11168NOL.22 and 11168NOL.23. This data confirmed the correct deletion of the \textit{waaF} gene and replacement with the \textit{cat} cassette. Chromosomal DNA was restricted with \textit{Cla I} and plasmid DNA with \textit{Hind III} before being separated by horizontal gel electrophoresis (0.8\% agarose). Marker sizes, in kilobase pairs, are shown on the left hand side of the figure.

\textbf{A: Southern hybridization analysis using the \textit{waaF} gene (and \textit{Hind III} marker) as probe.} The probe was produced from the 590 bp \textit{Bgl II} fragment of pNOI.17 which contains part of the \textit{waaF} gene. Lane 1, \textit{Hind III} marker; Lanes 2 & 3, 11168NOL.22 and 11168NOL.23 yield no hybridizing fragments confirming the deletion of the \textit{waaF} gene; Lane 4, wild-type strain NCTC 11168 yields a hybridizing \textit{Cla I} fragment of 3.5 kb; Lanes 5 & 6, pNOI.22 and pNOI.23 yield no hybridizing fragments confirming the deletion of \textit{waaF}; Lane 7, \textit{Hind III} marker.

\textbf{B: Southern hybridization analysis using the \textit{cat} cassette (and \textit{Hind III} marker) as probe.} This panel has been split into two, reflecting the requirement for differing exposure times to produce the best results from all lanes. Two hybridizing bands are expected for each mutant as the \textit{cat} cassette contains a \textit{Cla I} site. Lane 1, \textit{Hind III} marker; Lane 2, 11168NOL.22 yields hybridizing \textit{Cla I} fragments of 3.0 kb and 0.9 kb so confirming the presence and forward orientation of the \textit{cat} cassette; Lane 3, 11168NOL.23 yields hybridizing \textit{Cla I} fragments of 3.5 kb and 0.4 kb (not visible) confirming the presence and reverse orientation of the \textit{cat} cassette; Lane 4, no hybridizing band is present in the wild-type strain NCTC 11168; Lane 5, digestion of pNOI.22 with \textit{Hind III} yields hybridizing fragments of 1.2 kb and 0.4 kb; Lane 6, digestion of pNOI.23 with \textit{Hind III} yields hybridizing fragments of 0.8 kb and 0.7 kb; Lane 7, \textit{Hind III} marker.
Chapter 6: Characterization of the waaF gene in C. jejuni
Chapter 6: Characterization of the *waaF* gene in *C. jejuni*

Figure 6.9: Silver staining of the LOS produced by the *waaF* mutant strains 11168NOL22 and 11168NOL23. Samples were electrophoresed on a 15% SDS-PAGE gel. Marker sizes are shown on the left hand side of the figure. Lane 1, 11168NOL22; Lane 2, 11168NOL23; Lane 3, *C. jejuni* NCTC 11168. Both *waaF* mutants produce a significantly truncated core molecule compared to *C. jejuni* NCTC 11168.
Chapter 6: Characterization of the waaF gene in C. jejuni

(Figure 6.10A) also revealed that the mutant LOS molecules were truncated, in fact to such an extent that they were no longer detectable. Deletion of the outer core has presumably removed the majority of anti-HS:2 reactive epitopes resulting in no detectable hybridization. Immunoblotting using cholera toxin (Figure 6.10B) showed that the LOS molecules produced by the two mutants could no longer be detected using this technique, again providing additional confirmation of the truncated core phenotype in which the GM1-like epitope to which cholera toxin binds has been removed.

Mutation of waaF in C. jejuni NCTC 11168 therefore leads to truncation of the core oligosaccharide, a phenotype entirely consistent with those of waaF mutants in other bacterial species. However, the effect such a mutation might have on O-chain cannot be shown using NCTC 11168, a strain reported to produce no O-chain. Therefore it was decided to mutate waaF in C. jejuni NCTC 11828, a strain from serotype group HS:6 which produces O-chain.

6.5. Mutation of waaF in C. jejuni NCTC 11828

6.5.1. Isolation of waaF from C. jejuni NCTC 11828

The waaF gene was amplified from C. jejuni NCTC 11828 by the Expand™ High Fidelity PCR methodology using primers WLAXAF3 and WLASAR3 (Appendix 1). The amplified fragment was purified by horizontal gel electrophoresis followed by extraction using the NUCLEOTRAP kit. The purified DNA was digested with Kpn I and Pst I overnight before being ligated into Kpn I and Pst I digested pUC19 and electroporated into E. coli DH5α cells. The plasmid DNA from potential positive clones was extracted and used for PCR (data not shown) and restriction analysis. This data confirmed that one of the clones contained the desired construct which was named pNOL8 (a schematic diagram of this plasmid is shown in Figure 6.11). A restriction enzyme digestion of pNOL8 is shown in Figure 6.12.
Figure 6.10: Analysis of 11168NOL22 and 11168NOL23 LOS by immunoblotting. Samples were electrophoresed on a 15% SDS-PAGE gel before being transferred to nitrocellulose and subsequently detected using (A) HS:2 antiserum or (B) cholera toxin. Marker sizes are shown on the left hand side of each panel.

A: Detection using O:2 antiserum. Lane 1, 11168NOL22; Lane 2, 11168NOL23; Lane 3, C. jejuni NCTC 11168. Mutation of \textit{waaF} in 11168NOL22 and 11168NOL23 has led to truncation of the core oligosaccharides, presumably removing reactive epitopes resulting in the mutant molecules being undetectable using HS:2 antiserum.

B: Detection using cholera toxin. Lane 1, 11168NOL22; Lane 2, 11168NOL23; Lane 3, C. jejuni NCTC 11168. Again, mutation of \textit{waaF} in 11168NOL22 and 11168NOL23 has led to truncation of the core oligosaccharides, removing the reactive GM1-like epitope resulting in the mutant molecules being undetectable using cholera toxin.
Figure 6.11: Diagrammatic representation showing pNOL8. The \textit{waaF} gene and approximately 500 bp of flanking sequence either side was amplified using the primers WLAAXF3 and WLAASAR3. This PCR product was digested with \textit{Pst I} and \textit{Kpn I} before being ligated into pUC19 and transformed into \textit{E. coli} DH5\(\alpha\).

Figure 6.12: Agarose gel electrophoresis (0.8\% gel) to show the restriction enzyme digestion of pNOL8. Marker sizes, in kilobase pairs, are shown on the right hand side of the figure. Lane 1. \textit{Hind} III digest yielding fragments of 2679 bp, 954 bp, 593 bp, 519 bp and 80 bp (not visible); Lane 2. \textit{Pst I} & \textit{Kpn I} double digest yielding fragments of 2659 bp and 2166 bp; Lane 3. uncut pNOL8 (4825 bp); Lane 4. \(\phi\times Hae\) III marker.
Chapter 6: Characterization of the \textit{waaF} gene in \textit{C. jejuni}
6.5.2. Inverse PCR mutagenesis of C. jejuni NCTC 11828 waaF

The waaF gene was deleted from pNOL8 by inverse PCR using the primers WAAFF2 and NOL8F1 (Appendix 1). These primers were designed such that the waaF gene could be deleted without effecting the adjacent open reading frames (data not shown). The amplified fragment was purified by horizontal gel electrophoresis followed by extraction using the NUCLEOTRAP extraction kit. The purified DNA was then digested with BamH1 overnight and treated with calf intestinal alkaline phosphatase. The chloramphenicol acetyl transferase cassette was obtained by digesting pAV35 (Appendix 2) with BamH1 overnight. The 850 bp fragment containing the cat cassette was then purified and ligated to the pNOL8 inverse PCR product before being transformed into E. coli DH5α cells by electroporation. Positive clones were selected on LA containing ampicillin and chloramphenicol. The plasmid DNA from potential positive clones was extracted using the alkaline extraction method and used for PCR (data not shown) and restriction analysis. This data confirmed that one of the clones contained the desired construct with the cat cassette in the forward orientation (named pNOL11), and one containing the desired construct with the cat cassette in the reverse orientation (named pNOL12). Schematic diagrams of both pNOL11 and pNOL12 are shown in Figure 6.13. Restriction enzyme digestions confirming the correct construction of pNOL11 and pNOL12 are shown in Figure 6.14.

6.5.3. Transformation of C. jejuni NCTC 11828 with pNOL11 and pNOL12

C. jejuni NCTC 11828 was transformed by electroporation with the plasmids pNOL11 and pNOL12. Positive clones were selected on CSA plates containing chloramphenicol. The chromosomal DNA from several putative positive clones was extracted and used for PCR and Southern hybridization analysis. PCR using WLAXBR1, WLASAR4 and primers which anneal to the cat cassette identified several correct recombinants which contained the cat cassette in place of the deleted waaF gene. The PCR data for one correct clone containing cat in the forward orientation
The *waaF* gene was removed from pNOL8 by inverse PCR using the primers WAAFF2 and NOL8F1. The purified DNA was then digested with *BamH*I before being ligated to the 850 bp *BamH*I fragment from pAV35 containing the *cat* cassette.
Figure 6.14: Agarose gel electrophoresis (0.8% gel) to show the restriction enzyme digestion of pNOL.11 and pNOL.12. Marker sizes, in kilobase pairs, are shown on the right hand side of the figure.

A: Restriction enzyme digestion of pNOL.11 confirming the removal of waaF and insertion of the cat cassette in the forward orientation. Lane 1. BamHII digested pNOL.11 yielding fragments of 3978 bp and 850 bp. This digest confirmed the insertion of the cat cassette; Lane 2. Hind III digest yielding fragments of 2679 bp, 1178 bp, 519 bp and 452 bp. The fragments produced by this digestion confirmed the insertion and forward orientation of the cat cassette; Lane 3. uncut pNOL.11 (4828 bp); Lane 4. β. Hind III. φ x HaeIII marker.

B: Restriction enzyme digestion of pNOL.12 confirming the removal of waaF and insertion of the cat cassette in the reverse orientation. Lane 1. BamHII digested pNOL.12 yielding fragments of 3978 bp and 850 bp. This digest confirmed the insertion of the cat cassette; Lane 2. Hind III digest yielding fragments of 2679 bp, 856 bp, 774 bp and 519 bp. The fragments produced by this digestion confirmed the insertion and reverse orientation of the cat cassette; Lane 3. uncut pNOL.12 (4828 bp); Lane 4. β. Hind III. φ x HaeIII marker.
Chapter 6: Characterization of the \textit{waaF} gene in \textit{C. jejuni}
(strain 11828NOL11) and one clone containing *cat* in the reverse orientation (strain 11828NOL12) is shown in Figure 6.15. Southern hybridization using the *waaF* and *cat* genes as probes also confirmed that strains 11828NOL11 and 11828NOL12 were correct (Figure 6.16), and due to the presence of an internal *Cla* I restriction site in the *cat* cassette the orientation of the cassette could be confirmed. The chromosomal DNA from both strains was also probed with pUC19 to ensure that no plasmid sequence remained. No pUC19 sequence was detected (data not shown).

6.5.4. Phenotypic analysis of 11828NOL11 and 11828NOL12 LPS

The lipopolysaccharide from the two mutant strains, 11828NOL11 and 11828NOL12 was extracted and analyzed for changes compared to wild-type *C. jejuni* NCTC 11828. Analysis was carried out by SDS-PAGE gel electrophoresis followed by silver staining and immunoblotting using HS:6 antiserum. Unlike NCTC 11168, cholera toxin does not bind to the core oligosaccharide produced by *C. jejuni* NCTC 11828, so preventing analysis by this methodology. Silver staining (Figure 6.17) and immunoblotting (Figure 6.18) with HS:6 antiserum showed that the core molecules produced by the two mutant strains were significantly truncated compared with those produced by the wild-type. This phenotype is entirely consistent with that observed for the *C. jejuni* NCTC 111168 *waaF* mutants. However, the most important observation was that the NCTC 11828 O-chain ladder was unaffected both in mobility or intensity by mutation of *waaF*. This observation is in contrast to the phenotype observed in other bacterial species where O-chain is no longer observed following mutation of *waaF* (Schnaitman & Klena, 1993; Allen *et al.*, 1998). This result is of major significance because it indicates that the ladder-like polysaccharide produced by *C. jejuni* NCTC 11828 is not covalently attached to the outer core oligosaccharide.
Figure 6.15: PCR analysis of strains 11828NOL11 and 11828NOL12. This confirmed the correct deletion of the waaF gene and replacement with the cat cassette. PCR products were separated by horizontal gel electrophoresis (0.8% agarose). Marker sizes, in kilobase pairs, are shown on the left hand side of the figure.

A: PCR analysis of strain 11828NOL11. Lane 1, $\phi$ Hind III marker; Lane 2, WLA XBR1 WLASAR4 primer combination using 11828NOL11 DNA produces a product of 3.2 kb. The size of this product is virtually identical to the wild-type product (Lane 4) due to the deletion of the 847 bp waaF fragment and insertion of the 850 bp cat cassette; Lane 3, WLASAR4 CAT115' primer combination using 11828NOL11 DNA produces a product of 0.86 kb confirming the presence and position of the cat cassette. This also confirmed the forward orientation of the cat cassette; Lane 4, WLA XBR1 WLASAR4 primer combination using NCTC 11828 DNA produces a fragment of 3.2 kb; Lane 5, WLASAR4 CAT115' primer combination using NCTC 11828 DNA produces no fragment confirming the absence of the cat cassette in this wild-type strain; Lane 6, Negative control; Lane 7, $\phi$ Hind III marker.

B: PCR analysis of strain 11828NOL12. Lane 1, $\phi$ Hind III marker; Lane 2, WLA XBR1 WLASAR4 primer combination using 11828NOL12 DNA produces a product of 3.2 kb. The size of this product is virtually identical to the wild-type product (Lane 4) due to the deletion of the 847 bp waaF fragment and insertion of the 850 bp cat cassette; Lane 3, WLASAR4 CAT113' primer combination using 11828NOL12 DNA yields a product of 0.82 kb confirming the presence and position of the cat cassette. This also confirmed the reverse orientation of the cat cassette; Lane 4, WLA XBR1 WLASAR4 primer combination using NCTC 11828 DNA produces a fragment of 3.2 kb; Lane 5, WLASAR4 CAT113' primer combination using NCTC 11828 DNA produces no fragment confirming the absence of the cat cassette in this wild-type strain; Lane 6, Negative control; Lane 7, $\phi$ Hind III marker.
Chapter 6: Characterization of the wuaF gene in *C. jejuni*
Figure 6.16: Southern hybridization data from strains 11828NOL11 and 11828NOL12. This confirmed the deletion of the \textit{waaF} gene and replacement with the \textit{cat} cassette. DNA was restricted with \textit{Cla} I before being separated by horizontal gel electrophoresis (0.8\% agarose). Marker sizes, in kilobase pairs, are shown on the left hand side of the figure.

A: Southern hybridization analysis using the \textit{waaF} gene (and \textit{\lambda Hind III marker}) as probe. The probe was produced from the 590 bp \textit{Bgl II} fragment of pNOL.8 which contains part of the \textit{waaF} gene. Lane 1, \textit{\lambda Hind III} marker; Lanes 2 & 3, 11828NOL11 and 11828NOL12 yield no hybridizing fragments confirming the deletion of the \textit{waaF} gene; Lane 4, wild-type strain NCTC 11828 yields a hybridizing \textit{Cla} I fragment of 1.4 kb; Lanes 5 & 6 pNOL11 and pNOL12 yield no hybridizing fragments confirming the deletion of \textit{waaF}; Lane 7, \textit{\lambda Hind III} marker.

B: Southern hybridization analysis using the \textit{cat} cassette (and \textit{\lambda Hind III marker}) as probe. This panel has been split into two, reflecting the requirement for different exposure times to produce the best results from all lanes. Two hybridizing bands are expected for each mutant as the \textit{cat} cassette contains a \textit{Cla} I site. Lane 1, \textit{\lambda Hind III} marker; Lane 2, 11828NOL11 yields hybridizing \textit{Cla} I fragments of 0.88 kb and 0.9 kb so confirming the presence and forward orientation of the \textit{cat} cassette; Lane 3, 11828NOL12 yields hybridizing \textit{Cla} I fragments of 1.4 kb and 0.4 kb (not visible) confirming the presence and reverse orientation of the \textit{cat} cassette; Lane 4, no hybridizing band is present in the wild-type strain NCTC 11828; Lane 5, digestion of pNOL11 with \textit{Cla} I yields hybridizing fragments of 0.88 kb and 0.9 kb; Lane 6, digestion of pNOL12 yields hybridizing fragments of 1.4 kb and 0.4 kb; Lane 7, \textit{\lambda Hind III} marker.
Chapter 6: Characterization of the waaF gene in C. jejuni

A

B
Figure 6.17: Silver staining of the LPS produced by the waaF mutant strains 11828NOL11 and 11828NOL12. Samples were electrophoresed on a 15% SDS-PAGE gel. Marker sizes are shown on the left hand side of the figure. Lane 1, 11828NOL11; Lane 2, 11828NOL12; Lane 3, C. jejuni NCTC 11828. Both waaF mutants produce a significantly truncated core oligosaccharide compared to C. jejuni NCTC 11828.
Figure 6.18: Analysis of 11828NOL11 and 11828NOL12 LPS by immunoblotting. Samples were electrophoresed on a 15% SDS-PAGE gel before being transferred to nitrocellulose and subsequently detected using HS:6 antiserum. Marker sizes are shown on the left hand side of the figure. Lane 1, 11828NOL11; Lane 2, 11828NOL12; Lane 3, C. jejuni NCTC 11828. Mutation of waaF in 11828NOL11 and 11828NOL12 has led to truncation of the core oligosaccharide, but significantly the O-chain has remained unaffected both in intensity and mobility compared to the wild-type NCTC 11828 molecule. This suggests that C. jejuni O-chain is not covalently attached to the core oligosaccharide as previously supposed.
Chapter 6: Characterization of the waaF gene in C. jejuni

6.6. Discussion

6.6.1. Confirmation of Cj1148 as being the waaF gene of C. jejuni NCTC 11168

Analysis of the complete genome sequence of C. jejuni NCTC 11168 identified open reading frame Cj1148 as being a candidate for the waaF gene (detailed in Chapter 3). The waaF gene in E. coli encodes the heptosyltransferase which catalyses the transfer of the second L-glycero-D-manno-heptose residue during the biosynthesis of the core oligosaccharide moiety of LPS (Brabetz et al., 1997). This putative identification was supported by the ability of the cloned open reading frame to complement a S. typhimurium waaF mutant and restore the ability of this strain to produce a complete core oligosaccharide and thus attach O-chain. The functional complementation was possible due to the similarity of the inner core structures of S. typhimurium and C. jejuni (discussed in Chapter 1), and to the similarity of the two waaF proteins (Table 3.2).

Once confirmed that Cj1148 could complement a S. typhimurium waaF mutant the gene was mutated in C. jejuni NCTC 11168 by deletion using inverse PCR and insertion of a chloramphenicol acetyl transferase gene (cat). Following allelic exchange of the mutated allele into the chromosome of C. jejuni NCTC 11168, phenotypic analysis of the mutant LOS was undertaken by SDS-PAGE gel electrophoresis followed by silver staining and immunoblotting using HS:2 antiserum or cholera toxin. This analysis confirmed that the mutants produced a truncated LOS molecule which migrated faster than wild-type LOS. This finding is consistent with the phenotype reported for waaF mutants of other Gram negative bacteria including E. coli (Brabetz et al., 1997). B. pertussis (Allen et al., 1998), H. influenzae (Nichols et al., 1997; Hood et al., 1996). H. ducreyi (Bauer et al., 1999), H. gonorrhoeae (Schwan et al., 1995) and N. meningitidis (Jennings et al., 1995). The open reading frame Cj1148 therefore has the ability to complement a S. typhimurium waaF mutant, and mutation of the gene leads to a truncated core oligosaccharide phenotype thus strongly supporting the putative identification of the gene as waaF.
6.6.2. Evidence that *C. jejuni* NCTC 11828 O-chain is not covalently attached to the core oligosaccharide

Mutation of *waaF* in NCTC 11168, although instructive, could not provide information about the effect the mutation might have on O-chain. *C. jejuni* NCTC 11168 is reported to produce no O-chain, therefore an identical mutagenesis methodology was used to generate *waaF* mutants in *C. jejuni* NCTC 11828, a strain from HS serogroup 6 which does produce O-chain. Analysis of these mutants by SDS-PAGE electrophoresis and immunoblotting using HS:6 antiserum revealed the same effect of truncation on the core oligosaccharide, but significantly the O-chain ladder visible by immunoblotting was unaffected both in intensity and mobility by the *waaF* mutation.

This lack of effect on the O-chain of NCTC 11828 observed in *waaF* mutants corresponds with that described when the *galE* gene is mutated (Fry et al., 2000a). The *galE* gene encodes a UDP-galactose-4-epimerase which catalyses the inter-conversion between UDP-glucose and UDP-galactose (see Chapter 3). The mutation of this gene in *N. meningitidis* results in truncation of LOS by absence of the outer galactose containing units (Hammerschmit et al., 1994). Galactose is also a common component of the *C. jejuni* core oligosaccharide structures so far elucidated (reviewed in Moran & Penner, 1999), and mutation of *galE* in *C. jejuni* NCTC 11828 leads to the similar phenotype of core oligosaccharide truncation. This phenotype is consistent with the loss of galactose containing units from the core (Fry et al., 2000a). The effect of *galE* mutation on O-chain though is identical to that seen in the *waaF* mutants. Neither mutation has an effect suggesting that the O-chain ladder produced by NCTC 11828 is not covalently attached to the outer core since this is absent in both mutants.

Conclusions about the site or actuality of O-chain attachment to the core oligosaccharide based on the *galE* mutants are hampered by the fact that the precise chemical structure of the NCTC 11828 core oligosaccharide and O-chain has yet to be elucidated. Although galactose is likely to be a major constituent of the outer core, the existence of O-chain attachment to a sugar residue unaffected by the *galE* mutation can
Chapter 6: Characterization of the \textit{waaF} gene in \textit{C. jejuni}

not be ruled out. However, the \textit{waaF} mutants described in this chapter are more informative, as in these strains the whole core oligosaccharide structure after the first L-glycero-D-manno-heptose is removed irrespective of composition or linkage. Thus, O-chain attachment must be through this first heptose residue or else the polysaccharide ladder seen on immunoblot is a core-independent structure. This structure must be heat stable to be present in LOS/LPS samples which have been heated to 100°C (section 2.18.1), and contain a lipid moiety to enable migration in the SDS-PAGE gels that are designed for the analysis of LPS (Tsai & Frasch, 1982; Preston & Penner, 1987b).

Various evidence, in addition to that presented in this Chapter, exists to suggest that the 11828 polysaccharide is a core independent polysaccharide. Firstly, none of the ladder-like polysaccharides produced by isolates of \textit{C. jejuni} are visible by silver staining. This is in contrast to the situation in for example \textit{S. typhimurium} and \textit{E. coli} where silver staining (Tsai & Frasch, 1982) can be used to detect both the core and O-chain components of LPS. Silver nitrate does not react with the lipid A moiety of LPS (Hitchcock \textit{et al.}, 1986), but can be used to detect the Re-LPS form of \textit{S. typhimurium} which only contains lipid A and Kdo - therefore Kdo is presumed to be the reactive group in the silver staining reaction (Kropinski, 1986). The \textit{E. coli} and \textit{S. typhimurium} O-chains are therefore detectable by silver staining because they are linked to the Kdo residues of the core oligosaccharide. The fact that \textit{C. jejuni} ‘O-chain’ is not detectable by silver staining may suggest that it is not linked to the Kdo-containing (and hence silver stainable) core oligosaccharide.

Additional evidence to support the view that the polysaccharide of \textit{C. jejuni} NCTC 11828 is not core-linked O-chain comes from the completed genome sequence of \textit{C. jejuni} NCTC 11168 and a recent report from Karlyshev \textit{et al.} (2000). In addition to containing the large cluster of genes involved with the biosynthesis of LOS, the genome also contains a cluster of genes highly similar to those involved in the biosynthesis of capsular polysaccharide in other Gram negative bacteria (Parkhill \textit{et al.}, 2000a; Karlyshev \textit{et al.}, 2000). The presence of capsular genes was an unexpected finding, as no capsule had been detected in \textit{C. jejuni}. In particular, the genes are homologous to, and were organized in a similar fashion to, the group II capsule
Chapter 6: Characterization of the \textit{waaF} gene in \textit{C. jejuni}

biosynthesis genes of \textit{E. coli} (Roberts, 1996; Whitfield & Roberts, 1999). Group II capsule biosynthesis genes are found in clusters which are composed of three areas; the first contains genes involved with the biosynthesis of the capsule, whilst the second and third regions, which flank the first, contain the genes involved with transport of the capsule to the cell surface (reviewed in Roberts, 1996). This organization is conserved in \textit{C. jejuni} NCTC 11168, with the ABC transporter genes, \textit{kpsM} and \textit{kpsT}, and the \textit{kpsS} and \textit{kpsC} genes which are involved with the attachment of the capsule to a lipid moiety found flanking a central region. This central area contains a large number of genes with similarity to known polysaccharide biosynthesis genes (Karlyshev \textit{et al.}, 2000).

Using Southern hybridization, Karlyshev and co-workers showed that the \textit{kpsM}, \textit{kpsC} and \textit{kpsS} genes were conserved in \textit{C. jejuni} strains of varying HS serotype including NCTC 11828 (Karlyshev \textit{et al.}, 2000). In addition, mutation of the \textit{kpsM} gene in this strain resulted in the loss of O-chain as judged by immunoblotting with homologous antiserum. Therefore, mutation of the \textit{waaF} gene (involved in core biosynthesis) has no effect on NCTC 11828 O-chain, but mutation of \textit{kpsM} (apparently involved in capsular transport) abolished O-chain biosynthesis. This would suggest that the \textit{C. jejuni} NCTC 11828 'O-chain' is in fact a heat stable, core independent, lipid bound polysaccharide.

In addition to the construction of mutants in \textit{C. jejuni} NCTC 11828, the authors also constructed \textit{kpsM}, \textit{C} and \textit{S} mutants in \textit{C. jejuni} NCTC 11168. This strain of course was not thought to produce a polysaccharide, however using a novel immunoblotting protocol involving the use of a polyvinylidene difluoride (PVDF), rather than the traditional nitrocellulose membrane, it was shown that \textit{C. jejuni} NCTC 11168 and other strains which were thought not to produce a heat stable polysaccharide, did in fact produce one. Furthermore, mutation of the \textit{kps} genes in NCTC 11168 and other strains abolished the biosynthesis of these polysaccharide molecules (Karlyshev \textit{et al.}, 2000). To show the nature of the polysaccharide, LOS/LPS samples from various \textit{C. jejuni} strains including NCTC 11168 and NCTC 11828 were treated with phospholipase. The presence of a phospholipid moiety to anchor the molecule in the outer membrane is a
characteristic of a group II capsular polysaccharide, whilst O-chain polysaccharides are anchored via a lipid A-core moiety. The bond between the phospholipid and capsular polysaccharide can be disrupted by phospholipase, whilst no phospholipase susceptible bonds are present in an LPS molecule which has the polysaccharide linked through a non-repeating core oligosaccharide (Karlshev et al., 2000). Therefore, treatment with phospholipase can distinguish between capsular polysaccharides and lipopolysaccharides. In the case of C. jejuni, treatment with phospholipase abolished the presence of polysaccharide as detected by immunoblotting using homologous antiserum (Karlshev et al., 2000) suggesting that the polysaccharide is capsular. Removal of the phospholipid moiety by the action of the phospholipase, prevents the migration of the polysaccharide through the SDS-PAGE gel. It therefore appears that the commonly held view that C. jejuni strains can produce LOS or LPS is incorrect, but in fact they all produce LOS, and in addition can produce a core-independent capsular polysaccharide. This view is supported by the waaF mutants described in the chapter and will be discussed further in the following chapter.
Chapter 7: General Discussion

7.1. Introduction

LPS and LOS molecules are important components of Gram negative outer membranes. As major surface components they play a non-specific role as a physical barrier to the entry of deleterious compounds, but also play a more distinct role as a virulence determinant in many pathogens (Raetz, 1996). Furthermore, elements of the LPS/LOS molecule are highly structurally conserved between species, whilst others can vary considerably, even between members of the same species. These structural differences can convey different immunological and biological properties on the cell and hence can effect pathogenic potential. Due to these important functions and characteristics, the structure, biosynthesis and genetics of these molecules has been studied in great depth in a number of organisms and has greatly aided our understanding of the pathogenesis of these bacteria. However in contrast to these species, knowledge about the genetics and the biosynthesis of LOS/LPS molecules in C. jejuni is severely limited. In C. jejuni, LOS/LPS molecules have been suggested as potential adhesins and are likely to be important mediators of the inflammatory response during the development of Campylobacter-induced gastroenteritis. Furthermore, they show structural similarity to human gangliosides, and are thus considered important in the development of the auto-immune disorder GBS. The initial aim of this study was to enhance our understanding of the genetics and biosynthesis of these biologically important molecules in C. jejuni by identifying genes thought to encode key enzymes in LOS/LPS biosynthesis. This was achieved and allowed a model to provide testable predictions about core oligosaccharide biosynthesis in C. jejuni NCTC 11168 to be developed (Chapter 3). Once the genes had been identified, an experimental approach to further characterize biosynthesis in C. jejuni could be undertaken (Chapters 5 & 6), and in addition, by analyzing the genetic differences in gene content, the basis of LOS/LPS structural variation between different strains could be better understood (Chapter 4).
Chapter 7: General discussion

7.2. Building a model of core oligosaccharide biosynthesis in C. jejuni NCTC 11168

The identification of LOS/LPS related genes was achieved using a bioinformatic analysis of the genome sequence of C. jejuni NCTC 11168. However, the various genes identified and discussed in this study should not be considered to represent the sum total of LOS/LPS related genes in C. jejuni NCTC 11168. It is possible that there are LOS/LPS related open reading frames present which can not be identified using a homology based approach. Furthermore, it should not be assumed that all the genes identified are definitely involved with LOS/LPS biosynthesis. Notwithstanding these limitations, many genes which are likely to be involved in LOS/LPS biosynthesis and which had not been previously reported in C. jejuni have now been identified.

The identification of these genes suggested that the biosynthesis of the Kdo-lipid A moiety in C. jejuni was likely to be similar to that of other Gram negative bacteria, with the possible exception of the addition of the fatty acid chains in the final steps of biosynthesis. However, the core oligosaccharide biosynthesis genes identified suggested that C. jejuni core oligosaccharide was synthesized via a different pathway to that of bacteria such as E. coli and Salmonella. A model summarizing this pathway was put forward in order to allow testable predictions about core oligosaccharide biosynthesis in C. jejuni NCTC 11168 to be developed. The further characterization of core oligosaccharide genes by mutation and complementation analysis will hopefully shed further light on this biosynthetic pathway and may lead to a refinements of the model.

The construction of outer core mutants in C. jejuni NCTC 11168 should be unproblematic, indeed several mutant strains containing inactivated LOS cluster genes have already been constructed (Wood et al., 1999), but precisely defining the mutant phenotype is more difficult. Complex chemical analysis using techniques such as electrospray injection mass spectroscopy (ESI-MS), gas chromatography mass spectroscopy (GC-MS) or fast atom bombardment mass spectroscopy (FAB-MS) is expensive in both cost and time, but can be a powerful method of revealing alterations to core molecules. The immunoblotting approach using absorbed polyclonal sera
Chapter 7: General discussion

utilized in this study is inadequate for detecting minor alterations to core structures, but can rapidly reveal gross changes. For example, this technique clearly showed the truncation of the core molecules produced by the _waaF_ mutant strains detailed in Chapter 6. However, it is unlikely that the loss or gain of one or two sugar residues would be detected unambiguously by this method. Chemical analysis of core produced by the large deletion mutant of _C. jejuni_ NCTC 11168, in which the 3' end of _wlaF_, _wlaG_, _wlaH_, _wlaI_, _wlaJ_ and the 5' end of _wlaK_ were removed, indicates the loss of several residues consistent with the loss of the outer core (Anthony Moran, personal communication). However, analysis of the same core oligosaccharide molecules by immunoblotting using homologous antiserum showed only a slight reduction in reactivity (Wood _et al._, 1999). Therefore, more sensitive techniques than using polyclonal antiserum are needed to help define mutant core phenotypes. Ideally the method needs to be cheaper and less labour intensive than rigorous chemical analysis. An ideal compromise would be ligands with more specificity than polyclonal antisera, which could be used for the detection of certain structures or epitopes using the cheap and labour un-intensive immunoblotting technique. Such ligands include cholera toxin, which binds to GM1-like epitopes which some _C. jejuni_ strains produce as part of the core oligosaccharide, and peanut agglutinin which binds to the disaccharide structure Gal (β1-3) GalNAc. Peanut agglutinin has not yet been used in our laboratory to evaluate mutant LOS/LPS molecules, but cholera toxin was used in this study to confirm that the majority of the outer core in strain NCTC 11168 was no longer synthesized following the mutation of the _waaF_ gene. The use of these two ligands should be extended to evaluate the core phenotypes of the LOS cluster mutants already produced and any future strains constructed. However, these two ligands only bind to particular epitopes and not all _C. jejuni_ strains produce GM1-like or Gal (β1-3) GalNAc epitopes. For example, the core synthesized by NCTC 11828 does not contain a GM1-like, and hence cholera toxin binding epitope, therefore the use of these two ligands will be limited to certain strains. Potentially the most useful ligands would be a panel of monoclonal antibodies with specificities against different epitopes of the core structure. The use of such monoclonal antibodies has already been used to study _H._
influenza LOS (Hood et al., 1996) and Bordetella pertussis LPS (Allen & Maskell, 1996) and has greatly enhanced the understanding of the roles of various genes in the biosynthesis of these structures. Screening the core oligosaccharide molecules produced by mutant \textit{C. jejuni} strains with a panel of monoclonal antibodies would quickly and easily distinguish strains with altered core structures and would allow the selection of particular molecules of interest for more complete chemical analysis.

In order to fully understand the biosynthesis of core oligosaccharide and to refine the model, it will be necessary to investigate the precise chemical structure of the \textit{C. jejuni} NCTC 11168 core oligosaccharide. This is already underway (Anthony Moran, personal communication), and significantly, preliminary results reveal micro-heterogeneity in the core oligosaccharides produced by this strain. These differences may explain why slight changes to core molecules are not easily detected using polyclonal sera. Trying to distinguish molecules which are lacking a sugar residue as a result of a specific mutant in a population of molecules which are normally already heterogeneous would be difficult. The use of more sensitive monoclonal antibodies or reagents such as cholera toxin and peanut agglutinin will help to circumvent this problem. It would be extremely interesting to examine whether the slight variations in core structure may correlate to differences in the number of residues present in the two homo-polymeric tracts identified in the LOS cluster of NCTC 11168. One of the polymorphic tracts is present in a putative glycosyltransferase gene (\textit{wlaQ}). Variations in the number of residues in this tract could result in truncation, and perhaps inactivation of the WlaQ transferase - perhaps leading to the loss of residues from the core. Therefore, a population of \textit{C. jejuni} NCTC 11168 in which some cells are expressing the WlaQ transferase, whilst others are not, may well result in the micro-heterogeneity observed by Moran and co-workers. Investigations into the significance of the homo-polymeric tracts in core biosynthesis are already in progress in our laboratory, and hopefully may correlate the presence of variable epitopes with the translational status of the transferase proteins.
7.3. O-chain versus capsule

Following the identification of genes involved with the biosynthesis of LOS, the next logical step was to initiate functional studies to confirm the role of each gene product in the biosynthesis pathway. One method of achieving this is by functional complementation. Chapter 6 of this study illustrates how this methodology was used to support the identification of open reading frame Cj1148 as the C. jejuni waaF gene. Complementation of a waaF mutant strain of S. typhimurium with Cj1148, therefore allowing the complemented strain to produce a complete core oligosaccharide and attach O-chain provided evidence that Cj1148 did encode the C. jejuni WaaF protein. Further evidence came from insertionally inactivating open reading frame Cj1148. This led to the production of a truncated LOS molecule - a phenotype consistent with that of other Gram negative waaF mutants.

Following the construction of waaF mutant strains of C. jejuni NCTC 11168, the same approach was used to make similar mutants in strain NCTC 11828. Analysis of these mutants revealed the same effect of truncation on the core oligosaccharide, but significantly the polysaccharide ladder visible by immunoblotting was unaffected both in intensity and mobility by the waaF mutation. If the polysaccharide produced by NCTC 11828 was linked to the outer core, the appearance of the polysaccharide following immunoblotting should have been abolished following waaF mutation. This significant finding supports the conclusion of Karlyshev et al. (2000) and may suggest that all C. jejuni strains produce LOS, and in addition can elaborate a lipid bound capsular polysaccharide (Karlyshev et al., 2000). The work emphasizes some of the limitations of structural chemistry in evaluating the nature of LOS/LPS molecules. The presence of repeating polysaccharide units visible by immunoblotting using homologous antisera in some strains, and the structural similarity of some of these polysaccharide units to other Gram negative O-chains led to assumptions that the polysaccharide was O-chain and linked to the core oligosaccharide (Moran & Penner, 1999). This led to the erroneous assumption that some C. jejuni strains produced LOS, whilst others synthesized LPS. However, the linkage of the polysaccharide to core was
assumed rather than confirmed, and this conjecture has only been shown to be incorrect by the work of Karlyshev and co-workers and the waaF mutant strains described in this study. However, it should be apparent that even if the Karlyshev report had not been published, the waaF strains described combined with the inability to detect the polysaccharide using silver staining clearly shows that the ‘O-chain’ is a core independent lipid bound polysaccharide.

The waaF mutant strains described in this study will be useful tools in evaluating the contribution of C. jejuni LOS structures to pathogenesis. After all, the underlying aim of this study was to further current knowledge about the genetics, and biosynthesis of C. jejuni LOS/LPS molecules in order to better understand the contribution these molecules make to pathogenesis. The waaF mutant strains could be used in both in vivo and in vitro studies to study the effects of core truncation on processes such as colonization, adherence, invasion and the ability to stimulate the host inflammatory response. However, such studies may be hampered by the lack of a suitable animal model to closely mimic the clinical manifestations of Campylobacter enteritis (reviewed in Young et al., 2000). The waaF strains could also be used to investigate the contribution of core structures to the development of GBS. However, yet again, no suitable animal models are currently available which would allow the ability of such strains to induce GBS to be compared to wild-type isolates.

7.4. Distribution studies on LOS cluster genes

The majority of the genes involved with the biosynthesis of the core oligosaccharide were found to be grouped together in a large cluster known as the LOS cluster. Once identified, attempts were made to characterize gene content variations at this locus in different strains of C. jejuni. This was undertaken to identify genes responsible for the different core oligosaccharide structures produced by different strains. This study highlighted three areas of gene content polymorphism, and allowed the isolation and sequence analysis of several of these polymorphic areas. Further work in this area should involve the mutagenesis of the novel genes identified in the study in
order to understand the role such genes have in core biosynthesis. However, this could be problematical because some *C. jejuni* strains (e.g. NCTC 11351; Wood *et al.*, 1999) appear not to be transformable using current protocols. In addition, mutational analysis is most useful in strains which have had the structure of the core oligosaccharide they synthesize chemically determined. Very few of the strains used in this study have been analyzed in this way. In addition, many more polymorphic areas from different strains need to isolated, sequenced and functionally characterized. This information will be useful in designing possible genotyping methodologies to allow the detection, differentiation and tracking of strains in epidemiological studies.

An additional point to consider is that although the distribution of the LOS cluster genes in many *C. jejuni* strains was investigated in this study, no experiments have been undertaken to establish the presence or absence of other LOS/LPS genes which are located in other regions of the genome. Since most of these are involved with the biosynthesis of the Kdo-lipid A moiety, a pathway which is presumably conserved in all strains of *C. jejuni*, one would assume that they are also highly conserved. However, work has revealed that *C. jejuni* lipid A molecules show more structural heterogeneity than other those from other species, with three different backbone disaccharide structures occurring (Moran, 1997; Moran *et al.*, 1991). In different *C. jejuni* strains, the proportions of these three types varies, resulting in inter-strain lipid A diversity (Moran, 1993). No reports have yet been published which investigate how this variation occurs, but now that the genes involved with the biosynthesis of this moiety have been identified such a study may now be possible.

The recent revelation that the ‘O-chain’ of *C. jejuni* is capsular rather than linked to core oligosaccharide means that some of the results presented in Chapter 4 need re-evaluating. In the first place, the genetic distribution studies attempted to correlate the presence of genetic polymorphisms in the LOS cluster with the HS serotype of the strains. In the absence of structural information, HS serotyping data was assumed to reflect core structural similarities and differences - *i.e.* strains of the same HS serotype are assumed to produce highly similar (but not necessarily identical) core structures. However, no correlations were apparent between the polymorphisms and HS
serotype. This is probably due to the fact that the HS serotyping system is not discriminatory between core structures and correlations with gene content cannot be made - i.e. strains which do not produce identical core structures, and therefore have gene content differences, can still be classed as belonging to the same serogroup. An alternative hypothesis is that the specificity of the *C. jejuni* HS serotyping scheme relates predominately to a heat stable antigen other than LOS and therefore differences in the LOS cluster gene content which presumably lead to differences in core structure do not affect HS serotype. Evidence to support this later hypothesis comes from the work by Karlyshev *et al.* (2000). Mutation of the *kpsM* gene in several *C. jejuni* strains led to the absence of capsular polysaccharide production, but also led to HS serotype changes. For example, NCTC 11168, a strain from HS serogroup 2, became untypable following *kpsM* mutation. This would suggest that the capsular polysaccharide accounts for HS serotype specificity. However, it remains to be determined whether the capsular polysaccharide is the only determinant of HS serotype specificity, and LOS is not involved (or not significantly involved). Analysis by immunoblotting shows that HS antibodies clearly do bind to LOS molecules, therefore HS-reacting epitopes are present. Whether these epitopes are discriminatory between some or all strains is unclear. To determine whether LOS does have a role in HS serotype specificity, the HS serotypes of the *waaF* mutants described in this study should be compared to that of the parent (wild-type strains). The *waaF* strains produce a severely truncated LOS molecule - if these mutant strains retain the same serotype then that suggests that the capsular polysaccharide is the major sero-specifying determinant. However, mutation of the *waaF* gene, and thus truncation of LOS may also lead to HS serotyping changes, thus showing that the LOS heat stable structure also contributes to HS serospecificity.

In addition to attempting to correlate gene content polymorphisms to HS serotype, attempts were also made to correlate the presence of polymorphisms with the LOS/LPS phenotype of the strains. The predicted phenotype was based on the structural work on seven *C. jejuni* serostrains which suggested that serotype HS:1, HS:2 and HS:3 strains produce LOS, whilst strains of serotype HS:4, HS:19, HS:23 and HS:36 produce an O-chain (LPS). The correlation was undertaken in an attempt to highlight genes
which may encode proteins with a role in O-chain biosynthesis. No correlation between the polymorphisms and the LOS/LPS phenotype was evident - this is not surprising if one assumes that all strains are in fact LOS producers. Also the genes necessary for the production of the polysaccharide appear to be located in a separate and distinct cluster therefore it is unlikely that LOS cluster polymorphisms could affect polysaccharide biosynthesis anyway.

After locating areas of gene content polymorphism, several of these regions from different strains of *C. jejuni* were amplified, cloned and sequenced. Bioinformatic analysis of these sequences allowed the identification of potential open reading frames, and putative functions to be assigned to some of the encoded proteins. The analysis suggested that some of the encoded proteins from *C. jejuni* NCTC 11828 were most similar to O-chain or polysaccharide, rather than core, biosynthesis enzymes. These genes include *wlaUB* and *wlaUD* which encode proteins involved in the biosynthesis of sugars for incorporation into O-chain, *wlaUT*, a potential aminotransferase, and *wlaXA* which probably encodes an O-acetylation protein. In light of the work by Karlyshev *et al.* (2000), it may be reasonable to assume that these enzymes in fact function in capsular rather than O-chain biosynthesis. However, in the genome strain, *C. jejuni* NCTC 11168, the genes encoding for enzymes involved in LOS and capsular biosynthesis are found in distinct and separate clusters. Assuming that this arrangement is conserved in NCTC 11828, then it is unlikely that these LOS cluster genes are involved in capsular biosynthesis, but are probably involved in LOS biosynthesis after all. The systematic mutagenesis of both capsular and LOS cluster genes is required to confirm that each cluster is functionally distinct and encodes for enzymes involved with the two separate structures.

Initial results from the distribution studies indicated that the presence or absence of the *wlaJ* gene correlated with LOS/LPS phenotype. Therefore a series of experiments were undertaken involving the construction of chimeric NCTC 11828 strains containing *wlaJ* to show if the gene could affect O-chain biosynthesis. Ultimately, when the distribution studies were completed it became apparent that the initial correlation was erroneous, and in addition, the functional studies described in Chapter 4 confirmed that
the *wlaJ* gene did not have a role in affecting O-chain biosynthesis. The discovery that the O-chain is in fact a core independent polysaccharide, and is probably synthesized via the action of gene products encoded by a separate genetic locus provides further indications that *wlaJ* is not involved with polysaccharide biosynthesis. However, this series of experiments still yielded some useful information, for example, they showed how chimeric *C. jejuni* strains containing heterologous LOS genes can be constructed. Such an approach could be useful in functional studies on polymorphic genes, especially if the genes are from strains which are not transformable or strains for which no chemical analysis of the LOS molecules has been undertaken. By inserting polymorphic genes into a strain, for example NCTC 11168, which is transformable, and for which chemical data will shortly be available, any alterations to the LOS molecule can be attributed to the enzymatic activity of the encoded protein. However, the contribution of the antibiotic resistance marker and the flanking sequence needed for recombination to any LOS alterations needs to be eliminated.

In summary, this study has greatly increased current knowledge about the genetics and biosynthesis of *C. jejuni* LOS molecules. The identification of many of the key genes involved has allowed a model of core biosynthesis to be proposed and allows the construction of defined mutants which lack particular epitopes. Consequently the role of each epitope to the pathogenesis of *C. jejuni* can be more fully investigated. Now that such mutants can be easily constructed, the lack of sensitive, but inexpensive and labour un-intensive methods of determining altered LOS structures is apparent. In addition, there is also a need for suitable animal models which closely mimic the inflammatory-type diarrhoea and development of GBS which can be the result of *C. jejuni* infection. This work has also contributed to the discovery that the hypothesis that *C. jejuni* strains are capable of synthesizing LOS and LPS is erroneous, and that in fact the ‘O-chain’ is a core independent polysaccharide. Furthermore, genetic polymorphisms which presumably lead to the LOS structural differences between strains have been identified and preliminarily characterized.
Appendix 1: Primers used in this study. Asterisks indicate primers used for the distribution studies detailed in Chapter 4. Nucleotide sequence (5' - 3') and target gene are also shown.
## Appendix I

<table>
<thead>
<tr>
<th>Primer</th>
<th>Nucleotide sequence (5’-3’)</th>
<th>Target gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>CATIS’</td>
<td>CACCTATCGATTCAAGTGATCATG</td>
<td>cat</td>
</tr>
<tr>
<td>CATIS’</td>
<td>TAGTGGTCGAAATACTCTTTTCG</td>
<td>cat</td>
</tr>
<tr>
<td>GALEF1*</td>
<td>CGCGGATCCGCGTAAAAAGAAGTGATGAAGCG</td>
<td>galE</td>
</tr>
<tr>
<td>GALER1*</td>
<td>CGCGGATCCGCGTAAAAATCGATAAGTGATG</td>
<td>galE</td>
</tr>
<tr>
<td>GMHAF1*</td>
<td>CGCGGATCCCAATGCTAAGTAGTGAG</td>
<td>gmhA</td>
</tr>
<tr>
<td>GMHAF3</td>
<td>CGGGTATCCAAATGCTAAGTAGTGAG</td>
<td>gmhA</td>
</tr>
<tr>
<td>GMHDF1*</td>
<td>CGCGGATCCGCTAATTTATCGCCTTTGTTG</td>
<td>gmhD</td>
</tr>
<tr>
<td>GMHEF1*</td>
<td>CGCGGATCCGCTAATTTATCGCCTTTGTTG</td>
<td>gmhE</td>
</tr>
<tr>
<td>LRF1</td>
<td>CGCGGATCCGACATTAGTGACATTGCCACTCG</td>
<td>waaF</td>
</tr>
<tr>
<td>LRR1</td>
<td>CGCGGATCCGACATTAGTGACATTGCCACTCG</td>
<td>waaC</td>
</tr>
<tr>
<td>NEUA1F1*</td>
<td>CGCGGATCACCCTGGAATTTAAGGAG</td>
<td>neuA1</td>
</tr>
<tr>
<td>NEUB1F1*</td>
<td>CGCGGATCCGCGTCACTCCGTCAAATGAGG</td>
<td>neuB1</td>
</tr>
<tr>
<td>NEUC1F1*</td>
<td>CGCGGATCCCGCTAAGGAGCCCTTTGTTG</td>
<td>neuC1</td>
</tr>
<tr>
<td>NOL8F1</td>
<td>CGCGGATCCGATGAAAGATCTAAAGCCTG</td>
<td>waaF</td>
</tr>
<tr>
<td>WAACF1*</td>
<td>CGCGGATCCGAGCGCAATGTTGTTTGAAGAG</td>
<td>waaC</td>
</tr>
<tr>
<td>WAAFF1*</td>
<td>CGCGGATCCCGTTGGAAGATCGTTGTTG</td>
<td>waaF</td>
</tr>
<tr>
<td>WAAFF2</td>
<td>CGCGGATCCCGCTAAGGCGCTCCTTTGTTG</td>
<td>waaF</td>
</tr>
<tr>
<td>WAAFR1</td>
<td>CGCGGATCCGCGTAAAGGAGGCCCTTTGTTG</td>
<td>waaF</td>
</tr>
<tr>
<td>WAAFR4</td>
<td>CGCGGATCCGCGTAAAGGAGGCCCTTTGTTG</td>
<td>waaF</td>
</tr>
<tr>
<td>WAAFR5</td>
<td>CGCGGATCCGCGTAAAGGAGGCCCTTTGTTG</td>
<td>waaF</td>
</tr>
<tr>
<td>WAAMF1*</td>
<td>CGCGGATCCGCTGCAAATTGAGGCGTGAAGAG</td>
<td>waaM</td>
</tr>
<tr>
<td>WLAAF1*</td>
<td>CGCGGATCCGATGCTGCGGCTGCGG</td>
<td>wlaA</td>
</tr>
<tr>
<td>WLAR1*</td>
<td>CGCGGATCCGCAATGCTGGAATTTAAGGAG</td>
<td>wlaA</td>
</tr>
<tr>
<td>WLBR1*</td>
<td>CGCGGATCCAGCTCCTGGAATTTAAGGAG</td>
<td>wlaB</td>
</tr>
<tr>
<td>WLACR1*</td>
<td>CGCGGATCCGCGTAAAGGAGGCCCTTTGTTG</td>
<td>wlaC</td>
</tr>
<tr>
<td>WLADR1*</td>
<td>CGCGGATCCGCGTAAAGGAGGCCCTTTGTTG</td>
<td>wlaD</td>
</tr>
<tr>
<td>WLERA1*</td>
<td>CGCGGATCCGCGTAAAGGAGGCCCTTTGTTG</td>
<td>wlaE</td>
</tr>
<tr>
<td>WLAF1*</td>
<td>TATCTACGCTTACTTTAAGAGG</td>
<td>wlaF</td>
</tr>
<tr>
<td>WLAFR1*</td>
<td>CGCGGATCCGCGTAAAGGAGGCCCTTTGTTG</td>
<td>wlaF</td>
</tr>
<tr>
<td>WLAGF1*</td>
<td>GCCGCTAGGGTGCCGAAGATGAATACAGG</td>
<td>wlaG</td>
</tr>
<tr>
<td>WLHAF1*</td>
<td>CGCGGATCCGCGTAAAGGAGGCCCTTTGTTG</td>
<td>wlaH</td>
</tr>
<tr>
<td>WLHAF1*</td>
<td>CGCGGATCCGCGTAAAGGAGGCCCTTTGTTG</td>
<td>wlaH</td>
</tr>
<tr>
<td>WLAF1*</td>
<td>CGCGGATCCGCGTAAAGGAGGCCCTTTGTTG</td>
<td>wlaI</td>
</tr>
<tr>
<td>WLAF3</td>
<td>GGCCTAGGGGTGCGGAGCTAATGAGGAG</td>
<td>wlaI</td>
</tr>
<tr>
<td>WLAFJ1*</td>
<td>CGCGGATCCGCGTAAAGGAGGCCCTTTGTTG</td>
<td>wlaJ</td>
</tr>
<tr>
<td>WLAFJ3</td>
<td>GGCCTAGGGGTGCGGAGCTAATGAGGAG</td>
<td>wlaJ</td>
</tr>
<tr>
<td>WLKAF1*</td>
<td>CCAGGATCCCTATAAAGTAAATACACC</td>
<td>wlaK</td>
</tr>
<tr>
<td>WLKRC3</td>
<td>CTTTAAAGGAGCTATGAGGAGGCCCTTTG</td>
<td>wlaK</td>
</tr>
<tr>
<td>WLALR1*</td>
<td>GGAAGATCTGGGAGATATCTGCGGAGTCAATGAGG</td>
<td>wlaL</td>
</tr>
<tr>
<td>WLALR3</td>
<td>GGGCGGATCCGGCGGAGCTAATGAGGAG</td>
<td>wlaL</td>
</tr>
<tr>
<td>WLMRF3*</td>
<td>CGCGGATCCGCGTAAAGGAGGCCCTTTGTTG</td>
<td>wlaL</td>
</tr>
<tr>
<td>WLMR1*</td>
<td>CGCGGATCCGCGTAAAGGAGGCCCTTTGTTG</td>
<td>wlaM</td>
</tr>
<tr>
<td>Primer</td>
<td>Sequence (5’-3’)</td>
<td>Target gene</td>
</tr>
<tr>
<td>-----------</td>
<td>----------------------------------------------------------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>WLANAF1*</td>
<td>CGCGGATCCCAAGCACTTCATCAGCATCAATGC</td>
<td>wlaNA</td>
</tr>
<tr>
<td>WLANBF1*</td>
<td>CGCGGATCCGCTTATCACTCGCTAAAGC</td>
<td>wlaNB</td>
</tr>
<tr>
<td>WLANBR1*</td>
<td>CGCGGATCCCAAGGCAAATGTGTCAGAG</td>
<td>wlaNB</td>
</tr>
<tr>
<td>WLAOF1*</td>
<td>CGCGGATCCGCTTTACCTATGATACAAACAATG</td>
<td>wlaO</td>
</tr>
<tr>
<td>WLAPF1*</td>
<td>CGCGGATCCCCCTACATTTCTAGC</td>
<td>wlaP</td>
</tr>
<tr>
<td>WLAQF1*</td>
<td>CGCGGATCCGATCTAAATAGCAATTCTATAAC</td>
<td>wlaQ</td>
</tr>
<tr>
<td>WLARF1*</td>
<td>CGCGGATCCCCATTACCGCATACAAAGC</td>
<td>wlaR</td>
</tr>
<tr>
<td>WLARR1*</td>
<td>CGCGGATCCGATACCAGATAAGTATGC</td>
<td>wlaR</td>
</tr>
<tr>
<td>WLSAF1*</td>
<td>CGCGGATCCGTTATGGCGAAGATGTG</td>
<td>wlaSA</td>
</tr>
<tr>
<td>WLSAR1*</td>
<td>CGCGGATCCACTTAGCCCAAAACCGACGC</td>
<td>wlaSA</td>
</tr>
<tr>
<td>WLSAR3</td>
<td>AAACTGCAGCATTAGCCCAAAACCGACGC</td>
<td>wlaSA</td>
</tr>
<tr>
<td>WLSAR4</td>
<td>CGCGGATCCGACATCTTGCCCAAAC</td>
<td>wlaSA</td>
</tr>
<tr>
<td>WLSBF1*</td>
<td>CGCGGATCCCTGAGTACGGTATGTG</td>
<td>wlaSB</td>
</tr>
<tr>
<td>WLSCF1*</td>
<td>CGCGGATCCCTGTACAGTCCATG</td>
<td>wlaSC</td>
</tr>
<tr>
<td>WLATF1*</td>
<td>CGCGGATCCCTAAGTACAGTGAAC</td>
<td>wlaT</td>
</tr>
<tr>
<td>WLAXAF3</td>
<td>CGGGGTACGGAGCTAAAGCATAAAC</td>
<td>wlaXA</td>
</tr>
<tr>
<td>WLAXBR1</td>
<td>CGCGGATCCGAAATTACGAGGCAAAGGC</td>
<td>wlaXB</td>
</tr>
</tbody>
</table>
Appendix 2: Diagrammatic representation showing pUC19 and pAV35. pUC19 is a high copy number cloning vector (Yanisch-Perron et al., 1985) which was used routinely in this study. pAV35 contains the chloramphenicol acetyl transferase (cat) cassette (Wang & Taylor, 1990a) which was used as a selectable marker in the construction of mutant and chimeric strains of C. jejuni (see text for details).
Appendix 3: Nucleotide sequence of the waaF - waaC region of the *C. jejuni* NCTC 11828 LOS biosynthesis cluster. Two homo-polymeric G tracts are found within this sequence and are highlighted in bold (nucleotide numbers 10259 -10269 & 11065-11076).
Appendix 3

1751 CTATATTGTT TGGAATTTTT TTATTTTATA GGTATCTATA GTTCTCTAAA GTATCTATTA TTTTTTTTAT
1821 ATCAACGGCT GAAGAGAGG GGAATTTTTT TTATTTTTTT ATATTTTTTT ATATTTTTTT
1891 CTTTTTCTAA AAAAAAAAG ATATTTTATA TCTTGGGCTGT ATATATACGT TTTTTTTTGT
1961 AAAAAATTTAA AATATTTTAC TACTCTTTT AAAAAATTTTT ATATATTTTT ATATATTTTT
2031 ATAAAAAGCT GTCTTTAGCG TTTTTTACGT TTTTTTACGT TTTTTTACGT TTTTTTACGT
2101 TGGAAATTTT TCTATTATTT TTTAAAAA TTTAAAAA TTTAAAAA TTTAAAAA
2171 TAAATTTATC ATATTTTTAT ATTTTTTTAT TTTTTTTTT TTTTTTTTT TTTTTTTTT
2241 AAATTTAAAT TAAAAATCTT TTTTTTTAAG GGGCTAAACA ATATTTGAAA CAAAATATTT TTATTTTTAT
2311 TATTTTCTCA TTTTTTTTCA TTTTTTTTCA TTTTTTTTCA TTTTTTTTCA TTTTTTTTCA
2381 TAATTTAAAA AAAAAATTTA AAAAAATTTA AAAAAATTTA AAAAAATTTA AAAAAATTTA
2451 ATTTTTGTTG TTTTTTCTTG GTATCTTTTA AAAATTTAAA AAAATTTAAA AAAATTTAAA
2521 CTCTTTTCTT TTTTTTTTCT TTTTTTTTCT TTTTTTTTCT TTTTTTTTCT TTTTTTTTCT
2591 AACCTTTTTA AAAAAATTTA AAAATTTA AAAATTTA AAAATTTA AAAATTTA
2661 ATAAAAAGCG AAAATAAAAA ATATATTTTT ATATATTTTT ATATATTTTT ATATATTTTT
2731 TAATTTTTTT AAACAGATTTA TTTTTTTTTTA TTTTTTTTTTA TTTTTTTTTTA TTTTTTTTTTA
2801 TTTTTTCTTT AAAAAATTTA AAAAAATTTA AAAATTTA AAAATTTA AAAATTTA
2871 TTTTTTTTTTT AAAAAATTTA AAAAAATTTA AAAATTTA AAAATTTA AAAATTTA
2941 TACAACTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT
3011 AAAATTTATC ATATTTTTTT TTTTTTTTTA AAAATTTAAA AAAATTTAAA AAAATTTAAA
3081 TAATTTTTTT AAAATTTTTT AAAAAATTTA AAAAAATTTA AAAAAATTTA AAAAAATTTA
3151 CCAATATTTT TTTTTTTTTT AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA
3221 TACAACTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT
3291 TACAACTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT
3361 TTTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT
3431 TTTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT
Appendix 3

8751 CAAGTATTTT TTTAATGGCA ATTTCTTCCCA ATATTGGAATGT TACCTTGCATG TTTTAACTTG
8821 ATCTTTCGTT TATATCTTGA AGGCTACCAG TACCTTTTGAC CTGACATGAA TATATTTG
8891 TATAGGATCT ATACCTGCAA TATATCTTGA AGGCTACCAG TACCTTTTGAC CTGACATGAA
8961 TTCATCAAAC TATATCTTGA AGGCTACCAG TACCTTTTGAC CTGACATGAA TATATTTG
9031 ATAGTATACT ATACCTGCAA TATATCTTGA AGGCTACCAG TACCTTTTGAC CTGACATGAA
9101 GAGTTTGGGA TTTTAAATTT TACCTTGTAT TACCTTTTGAC CTGACATGAA TATATTTG
9171 TTTGACGTCT AATTTGGGAA TTTTAAATTT TACCTTGTAT TACCTTTTGAC CTGACATGAA
9241 AAAAACTGAA AAAAACTGAA TTTTAAATTT TACCTTGTAT TACCTTTTGAC CTGACATGAA
9311 GGGTATAAAA TTTTAAATTT TACCTTGTAT TACCTTTTGAC CTGACATGAA TATATTTG
9381 TATATTATTTAGGAGTTAAA TTTTAAATTT TACCTTGTAT TACCTTTTGAC CTGACATGAA TATATTTG
9451 AACCTTGTAA AAAAACTGAA TTTTAAATTT TACCTTGTAT TACCTTTTGAC CTGACATGAA TATATTTG
9521 ATATGCTCATATA TTTTAAATTT TACCTTGTAT TACCTTTTGAC CTGACATGAA TATATTTG
9591 TATTTTTTCC CAAAAAGAACAT TTTTAAATTT TACCTTGTAT TACCTTTTGAC CTGACATGAA TATATTTG
9661 GTATTTTTTGA TTTTAAATTT TACCTTGTAT TACCTTTTGAC CTGACATGAA TATATTTG
9731 TCATTATTAT CACCTTTTTT TTTTAAATTT TACCTTGTAT TACCTTTTGAC CTGACATGAA TATATTTG
9801 GCCAAGGAC AAAAACTGAA TTTTAAATTT TACCTTGTAT TACCTTTTGAC CTGACATGAA TATATTTG
9871 AGTTGACAAAT CACCTTTTTT TTTTAAATTT TACCTTGTAT TACCTTTTGAC CTGACATGAA TATATTTG
9941 ATATTTTTAT CACCTTTTTT TTTTAAATTT TACCTTGTAT TACCTTTTGAC CTGACATGAA TATATTTG
1001 ATGCACTCAG TTTTAAATTT TACCTTGTAT TACCTTTTGAC CTGACATGAA TATATTTG
1008 AGGTTTGGAA TTTTAAATTT TACCTTGTAT TACCTTTTGAC CTGACATGAA TATATTTG
1015 CAGTATTTT TTTTAAATTT TACCTTGTAT TACCTTTTGAC CTGACATGAA TATATTTG
1022 TTTTAAATTT TACCTTGTAT TACCTTTTGAC CTGACATGAA TATATTTG
10291 TTTTAAATTT TACCTTGTAT TACCTTTTGAC CTGACATGAA TATATTTG
10361 TTTTAAATTT TACCTTGTAT TACCTTTTGAC CTGACATGAA TATATTTG
10431 TTTTAAATTT TACCTTGTAT TACCTTTTGAC CTGACATGAA TATATTTG
Appendix 3

10501 TCA TTTTTCTA TAAAGCAATG AGAACATATA TTGCAATTAT CTCCTATTTT AGCATTAGGT AAAACTACGC
AGTAAGAGAT ATTGGTGTAC TCTGTGATAT AAGCTTAAAT GAGTAAATAA CTGTATACCA TTTTGATGCG

10571 TCA TTTTTCTA TAAAGCAATG AGAACATATA TTGCAATTAT CTCCTATTTT AGCATTAGGT AAAACTACGC
AGT AAAAGAT ATTTCGTTAC TCTTGTATAT AACGTTAATA GAGGATAAAA TCGTAATCCA TTTTGATGCG

10641 TCA TTTTTCTA TAAAGCAATG AGAACATATA TTGCAATTAT CTCCTATTTT AGCATTAGGT AAAACTACGC
ATAAGGAATT TTTAAGCTATG ATACTGAGAA AAGCTTAAAT GAGTAAATAA CTGTATACCA TTTTGATGCG

10711 TCA TTTTTCTA TAAAGCAATG AGAACATATA TTGCAATTAT CTCCTATTTT AGCATTAGGT AAAACTACGC
AGT AAAAGAT ATTTCGTTAC TCTTGTATAT AACGTTAATA GAGGATAAAA TCGTAATCCA TTTTGATGCG

10781 TCA TTTTTCTA TAAAGCAATG AGAACATATA TTGCAATTAT CTCCTATTTT AGCATTAGGT AAAACTACGC
CCTGTCAAAT TCACTGCTAT TCACTGCTAT TCACTGCTAT TCACTGCTAT TCACTGCTAT TCACTGCTAT TCACTGCTAT

10851 TCA TTTTTCTA TAAAGCAATG AGAACATATA TTGCAATTAT CTCCTATTTT AGCATTAGGT AAAACTACGC
AATATTGCAA TTAAGGAATT TTTAAGCTATG ATACTGAGAA AAGCTTAAAT GAGTAAATAA CTGTATACCA TTTTGATGCG

10921 TCA TTTTTCTA TAAAGCAATG AGAACATATA TTGCAATTAT CTCCTATTTT AGCATTAGGT AAAACTACGC
GTTAAAACGA GGTATTCTTT TTAAAAAGTA AAGAGAGGTG TATAAAAATG GATATATTTC AGGATATAGA

10991 TCA TTTTTCTA TAAAGCAATG AGAACATATA TTGCAATTAT CTCCTATTTT AGCATTAGGT AAAACTACGC
AGT AAAAGAT ATTTCGTTAC TCTTGTATAT AACGTTAATA GAGGATAAAA TCGTAATCCA TTTTGATGCG

11061 TCA TTTTTCTA TAAAGCAATG AGAACATATA TTGCAATTAT CTCCTATTTT AGCATTAGGT AAAACTACGC
GATTACCCAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC

11131 TCA TTTTTCTA TAAAGCAATG AGAACATATA TTGCAATTAT CTCCTATTTT AGCATTAGGT AAAACTACGC
AAAGTTACTG GAGTAAATAA CTGTATACCA TTTTGATGCG TTTTGATGCG TTTTGATGCG TTTTGATGCG TTTTGATGCG

11211 TCA TTTTTCTA TAAAGCAATG AGAACATATA TTGCAATTAT CTCCTATTTT AGCATTAGGT AAAACTACGC
GTTAAAACGA GGTATTCTTT TTAAAAAGTA AAGAGAGGTG TATAAAAATG GATATATTTC AGGATATAGA

11271 TCA TTTTTCTA TAAAGCAATG AGAACATATA TTGCAATTAT CTCCTATTTT AGCATTAGGT AAAACTACGC
AGT AAAAGAT ATTTCGTTAC TCTTGTATAT AACGTTAATA GAGGATAAAA TCGTAATCCA TTTTGATGCG

11341 TCA TTTTTCTA TAAAGCAATG AGAACATATA TTGCAATTAT CTCCTATTTT AGCATTAGGT AAAACTACGC
GATTACCCAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC

11411 TCA TTTTTCTA TAAAGCAATG AGAACATATA TTGCAATTAT CTCCTATTTT AGCATTAGGT AAAACTACGC
GTTAAAACGA GGTATTCTTT TTAAAAAGTA AAGAGAGGTG TATAAAAATG GATATATTTC AGGATATAGA

11481 TCA TTTTTCTA TAAAGCAATG AGAACATATA TTGCAATTAT CTCCTATTTT AGCATTAGGT AAAACTACGC
GTTAAAACGA GGTATTCTTT TTAAAAAGTA AAGAGAGGTG TATAAAAATG GATATATTTC AGGATATAGA

11551 TCA TTTTTCTA TAAAGCAATG AGAACATATA TTGCAATTAT CTCCTATTTT AGCATTAGGT AAAACTACGC
GATTACCCAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC

11621 TCA TTTTTCTA TAAAGCAATG AGAACATATA TTGCAATTAT CTCCTATTTT AGCATTAGGT AAAACTACGC
GATTACCCAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC

11691 TCA TTTTTCTA TAAAGCAATG AGAACATATA TTGCAATTAT CTCCTATTTT AGCATTAGGT AAAACTACGC
GATTACCCAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC

11761 TCA TTTTTCTA TAAAGCAATG AGAACATATA TTGCAATTAT CTCCTATTTT AGCATTAGGT AAAACTACGC
GATTACCCAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC

11831 TCA TTTTTCTA TAAAGCAATG AGAACATATA TTGCAATTAT CTCCTATTTT AGCATTAGGT AAAACTACGC
GATTACCCAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC

11901 TCA TTTTTCTA TAAAGCAATG AGAACATATA TTGCAATTAT CTCCTATTTT AGCATTAGGT AAAACTACGC
GATTACCCAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC

11971 TCA TTTTTCTA TAAAGCAATG AGAACATATA TTGCAATTAT CTCCTATTTT AGCATTAGGT AAAACTACGC
GATTACCCAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC

12041 TCA TTTTTCTA TAAAGCAATG AGAACATATA TTGCAATTAT CTCCTATTTT AGCATTAGGT AAAACTACGC
GATTACCCAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC

12111 TCA TTTTTCTA TAAAGCAATG AGAACATATA TTGCAATTAT CTCCTATTTT AGCATTAGGT AAAACTACGC
GATTACCCAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC

12181 TCA TTTTTCTA TAAAGCAATG AGAACATATA TTGCAATTAT CTCCTATTTT AGCATTAGGT AAAACTACGC
GATTACCCAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC TCCCCACAC

290
Appendix 3

12251 TCTCATGTTG ATTCTGCTGC AAAATTTATA ACCGCATTTA TTTTATATTT GCTGAAAATT TCACCAACTA
AGATGTCACG TAAAGCAGCT TTTTAATAAT TGGGCTAATAT AAATAATAAA CGACTTTTAA AGTGTTGAT

12321 AGGAAAATCTG CTTTAATACC CTTGCTGAAA ATACATAATT AGGATTTATT TTTAATATTAT TAAATTGGA
TTTCTTTGAG CGAATATAGG GGAACCTATT TTTGATATAA TCTTAAATAA AAAAAATAAT AATATTCAA

12391 ATATGCTGTA CGATAATCATA AAAATCTCAA ATTTATAATT TATGAGATTT GATTTTTCCT AAAAAAATAT
ATATAGCGAT CGTATACAAT TTTTTAGATT TAAATATTAA AATCGTAAAC CTAAAAAACCAA GAGCTGATTC

12461 AAATAAAAAT TACCTACCAT AAAAAACGCT CGGCCATGAA CAAATAATTATA ATACGCAAAAT TTTTTTTTA
TTATTTTTAT AAATATTGTC CATAATTACT TTTTGAAAGT AACAATGCTC TTTTTTCTAG GATTTCATTA

12531 ATATTTTTAT AAATATTGTC CATAATTACT TTTTGAAAGT AACAATGCTC TTTTTTCTAG GATTTCATTA
ATAAAAATAA TTTATACATG GTAATATTTT GCTATTTAAA AAAATATAAA CGACTTTTAA AGTGGTTGAT

12601 TCTATCGAGT TGTATATTGA GCCATTTCTC TCTAAGCAGG CAATCTTATA ACCCTGCCTT AATTCTATGG
AGATAGTCAG TTTGTGTTCT ATACAGACCA ATCCAGCTCC AACGTGTCATG TGGTGGAAAG

12671 TTGACCAAAA CGAATCTTCAT TCTTGAACCA TATGAGATTT GATTTTTCCT AAAAAAATAT AATATTCAA
AAATAAAAAT TACCTACCAT AAAAAACGCT CGGCCATGAA CAAATAATTATA ATACGCAAAAT TTTTTTTTA

12741 TAAATGTTTG GATTTTTCAT TATTAACCTT AAAATATATT ATATTTACATT GTTTTTATTTT AAAAAAATAT
ATATTTTTAT AAATATTGTC CATAATTACT TTTTGAAAGT AACAATGCTC TTTTTTCTAG GATTTCATTA

12811 CTGAGATTTG TTTTGGAGTTA TTTTGGAGTTA TTTTGGAGTTA TTTTGGAGTTA TTTTGGAGTTA TTTTGGAGTTA
GAACGCTTACG CAAATGCTTACG CAAATGCTTACG CAAATGCTTACG CAAATGCTTACG CAAATGCTTACG CAAATGCTTACG

12881 CAAATGCTTACG CAAATGCTTACG CAAATGCTTACG CAAATGCTTACG CAAATGCTTACG CAAATGCTTACG CAAATGCTTACG
AGATAGTCAG TTTGTGTTCT ATACAGACCA ATCCAGCTCC AACGTGTCATG TGGTGGAAAG

12951 CACCCCAAAA CTCTTCGAGT CTCTTCGAGT CTCTTCGAGT CTCTTCGAGT CTCTTCGAGT CTCTTCGAGT
AGATAGTCAG TTTGTGTTCT ATACAGACCA ATCCAGCTCC AACGTGTCATG TGGTGGAAAG

13021 TTGACCAAAA CGAATCTTCAT TCTTGAACCA TATGAGATTT GATTTTTCCT AAAAAAATAT AATATTCAA
AAATAAAAAT TACCTACCAT AAAAAACGCT CGGCCATGAA CAAATAAATTATA ATACGCAAAAT TTTTATTTTA

13091 CAAATGCTTACG CAAATGCTTACG CAAATGCTTACG CAAATGCTTACG CAAATGCTTACG CAAATGCTTACG CAAATGCTTACG
AGATAGTCAG TTTGTGTTCT ATACAGACCA ATCCAGCTCC AACGTGTCATG TGGTGGAAAG

13161 ATACGTGGAAT TCGCTTACCA GAAAAAGAC ATCTTTGAAAT AGCTATATCT ATCTTTGAAAT AGCTATATCT ATCTTTGAAAT
TAGACTTTAA ACCTTCTTGT AATTGAAAAG TGACGAATTAG TAGACTTTAA ACCTTCTTGT AATTGAAAAG TGACGAATTAG

13231 GTAGAAATAAT TCGAAACTAC TTTAATTTTT TCTCTTCTTA AAAACATCTAT ATTATCTTTTT ATTATCTTTTT ATTATCTTTTT
CATCTTTTAT ATCTTTATTT TCAGATAGAT ATCTTTATTT TCAGATAGAT ATCTTTATTT TCAGATAGAT ATCTTTATTT

13301 ATACGGCATAT AGTAAAGAGT TGCTCAGATA ATACGGCATAT AGTAAAGAGT TGCTCAGATA ATACGGCATAT AGTAAAGAGT
CATCTTTTAT ATCTTTATTT TCAGATAGAT ATCTTTATTT TCAGATAGAT ATCTTTATTT TCAGATAGAT ATCTTTATTT

13371 AGCTTAAATAA ATACCACTTCA TTGTACTAC CTTTTTATTT TTTTTTATTT TTTTTTATTT TTTTTTATTT TTTTTTATTT
GTTTTTATTT TTTTTTATTT TTTTTTATTT TTTTTTATTT TTTTTTATTT TTTTTTATTT TTTTTTATTT TTTTTTATTT

13441 CTTTTCTTGT AATAAACCTCT TTTTACCTTC TTTTACCTTC TTTTACCTTC TTTTACCTTC TTTTACCTTC TTTTACCTTC
GAACGCTTACG CAAATGCTTACG CAAATGCTTACG CAAATGCTTACG CAAATGCTTACG CAAATGCTTACG CAAATGCTTACG CAAATGCTTACG

13511 CAATATATATG TTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT
GTTTTTCTTA AAAATGTTTA AAAAAATTTT ATTTTTTTTTTT ATTTTTTTTTTT ATTTTTTTTTTT ATTTTTTTTTTT ATTTTTTTTTTT

13581 TATCTAATAA TTTGTGTTCT TTTTTTATTT TTTTTTATTT TTTTTTATTT TTTTTTATTT TTTTTTATTT TTTTTTATTT
ATAGAATTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT

13651 TTGCGCAATAT ATCGCTTCTC TTTTTTATTT TTTTTTATTT TTTTTTATTT TTTTTTATTT TTTTTTATTT TTTTTTATTT
AGCTTAAATAA ATACCACTTCA TTGTACTAC CTTTTTATTT TTTTTTATTT TTTTTTATTT TTTTTTATTT TTTTTTATTT

13721 TATATATATTG TATAGACCCG CAAATACATGT TTTTTTATTT TTTTTTATTT TTTTTTATTT TTTTTTATTT TTTTTTATTT
GTTTTTCTTA AAAATGTTTA AAAAAATTTT ATTTTTTTTTTT ATTTTTTTTTTT ATTTTTTTTTTT ATTTTTTTTTTT ATTTTTTTTTTT

13791 ATGTTTTTTCA AATACCACTTC TTTTTTATTT TTTTTTATTT TTTTTTATTT TTTTTTATTT TTTTTTATTT TTTTTTATTT
TACCGGAGAGT TATAGACCCG CAAATACATGT TTTTTTATTT TTTTTTATTT TTTTTTATTT TTTTTTATTT TTTTTTATTT

13861 ATCGTATACAA CTACGATTTA ATGCATTTTGA ATGTTTTTCA AATACCACTTC TTTTTTATTT TTTTTTATTT TTTTTTATTT
TACCGGAGAGT TATAGACCCG CAAATACATGT TTTTTTATTT TTTTTTATTT TTTTTTATTT TTTTTTATTT TTTTTTATTT

13931 AGCCCAATTCTT GATGGAATTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT
TGGACCAAAA TTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT

Appendix 4: Nucleotide sequence of the \textit{wlaS} - \textit{gmhA} region of the \textit{C. jejuni NCTC 11828 LOS biosynthesis cluster}. This 4.6 kb fragment was amplified using the primers GMHAF1 and WLASAR1 (Appendix 1). The PCR product was digested with \textit{BamHI} before being ligated into pUC19 to form pNOL5 (details in Chapter 4).
Appendix 4

5' 1 ACTTAGCCCA AACCGGACG CAAAAATGCC TTTGTTTCCT TAAAATCTCT TTTAGTTAAA
5' 3 TGATCCTGCT GCTTCCTGTC GTTTTACCG TATTTCACCG TAAAAAATCT AAAAATTCCT TTTGATTAAA

1 ACTTAGCCCA AAAAAAACACG TTTGTTTCCT TAAAATCTCT TTTAGTTAAA
5' 1 TGAATCGGGT TTGGCTGGTC GTTTTTACGG AAACAAACGA ATTTTTTAGA TTTTTAAGGA AAACTAATTT

3 AACTTCATCT TGTTTAAAAC GATAAAATTG TTTGGTTTTT ACCCTATGTA CAAAGGCATC AAAACAAAGC
3 1 TTGAAGTAGA ACAAATTTTG CTATTTTAAC AAA AAAA TGGGATACAT GTTTCCGTAG TTTTGTTTCG

4AAAT C AAAAC CTT TTT TC A TCTCTTTAAAT GCTATTTCAC AAGCATCAGG TGTTAAAAAA TCATCACTAT
4 1 TTTGCTAGCT GACAGCTGCT CTACCGAAGT GGACGAAATA TGCGATATTT TCTTGTAAAA TTTTTATGGG

5 ACTTGGCTAG GCGATGCAGT GATGGCTTCA CCTGCTTTAT ACGCTATAAA AGAACATTTT AAAAATACCC
5' 1 TGAACCGATC CGCTACGTCA CTACCGAAGT GGACGAAATA TGCGATATTT TCTTGTAAAA TTTTTATGGG

6 AATTTATGTT TTAATGCTCT TTGGTTTTT ACCCTATGTA CAAAGGCATC AAAACAAAGC
6 1 TTGAAGTAGA ACAAATTTTG CTATTTTAAC AAA AAAA TGGGATACAT GTTTCCGTAG TTTTGTTTCG

7 AACTTCATCT TGTTTAAAAC GATAAAATTG TTTGGTTTTT ACCCTATGTA CAAAGGCATC AAAACAAAGC
7 1 TTGAAGTAGA ACAAATTTTG CTATTTTAAC AAA AAAA TGGGATACAT GTTTCCGTAG TTTTGTTTCG

8 CCCTGGAAT AAAAAACACG TTTTAAATGC GAAAAAAATTT TTTTTATGGG
8 1 TTTTGAAGTT AAAATTTTTTTCTT TTTTTAATGCT TTTTAAATGC GAAAAAAATTT TTTTTATGGG

9 CCCTTGAAAT AAAATCGCAA GAATTAAAAA GCGGGGAGG TTTTTATGGG
9 1 AAATTAAGCAC CTTAATTCTTT TTTAAAATTC TTTTTTAAAAGCT TTTTTATGGG

10 ATTTTGCTTA CATCGCTGGT TTTTCTCTTA TTTTTTTGGG
10 1 AATGGTAAAA AAATTCTAGG ACTCAACCCT GGCGCAAGCT TTGGAAGTGC AAAAAGATGG GATGCGAGTT

11 CCCTGTCCTT TAAAGCACCA TAAATGCATG AAAGATCTAA AGCCTGAAAA GGTGATAGAA GAAATTAAAAA
11 1 ACCCTGTCCTT TAAAGCACCA TAAATGCATG AAAGATCTAA AGCCTGAAAA GGTGATAGAA GAAATTAAAAA

12 TTCACCTTGG CAAAATCAAA ACGCAAAATT AGTGCATTTA GATCTAGCTT GTATGCCTTG TATGCAAAAA
12 1 AAATTG ATTTAACAAA GGAATATTTA ATTTTTTAAAACG CTACCGAATTTTT TTTTTTTATGG

13 TTCACCTTGG CAAAATCAAA ACGCAAAATT AGTGCATTTA GATCTAGCTT GTATGCCTTG TATGCAAAAA
13 1 AAATTG ATTTAACAAA GGAATATTTA ATTTTTTAAAACG CTACCGAATTTTT TTTTTTTATGG

14 ATTTTGCTTA CATCGCTGGT TTTTCTCTTA TTTTTTTGGG
14 1 AATGGTAAAA AAATTCTAGG ACTCAACCCT GGCGCAAGCT TTGGAAGTGC AAAAAGATGG GATGCGAGTT

15 CCCTGTCCTT TAAAGCACCA TAAATGCATG AAAGATCTAA AGCCTGAAAA GGTGATAGAA GAAATTAAAAA
15 1 ACCCTGTCCTT TAAAGCACCA TAAATGCATG AAAGATCTAA AGCCTGAAAA GGTGATAGAA GAAATTAAAAA

16 TTCACCTTGG CAAAATCAAA ACGCAAAATT AGTGCATTTA GATCTAGCTT GTATGCCTTG TATGCAAAAA
16 1 AAATTG ATTTAACAAA GGAATATTTA ATTTTTTAAAACG CTACCGAATTTTT TTTTTTTATGG

294
Appendix 4

3501 TATATCTTTC AGTGGCTTTT ATATTATTCT TATCAAAAAGA CAAATATGTA TAGGAAAGAG
ATATGAGAG TTGCACAAAA TAATATACCA GATATAAAAAGAGATTTTTT TTGTTAAAAAT GCTCTTTA

3571 TAATTAAGAT GCTTTAACAA ATTGGTTTTT AGAAGAAAATA AACCACCTTTG GAGGATAGGA ATATCGGGAT
ATTAAATCTA TGAAAATTTG TAACTTTATT TTGGAATGGT ATTGAATAAC TATGACCC

3641 AAGAGATATG GCGATGATAA AATTTCCTAA AATGAAATTG ATCCGATTTT GAAATTTCCAA CCAAAATTG
TTTCTTACCA CGCGCTTTTT ATCTTTTTCT ATTTGAAAAAT ATATGGTTTT AATATACG

3711 ATATCTAGAT TAACTACAAA AAAAAATTAA CAAAAAGATA TCTCTTAAAG TTTATATATAT TTATATTAAAT AATATAGC

3781 AACATGTTTT TTTATACAAA TACCCAACTA TACACATTTA AACTACGTTA GGGGAATATT CTGCAATGCA ATTATACAC

3851 TATATAAAAG ATATGTGTTCT ATTTTCAAAC TATTATGCTA TTTTAAGTTG GATTATACAA GAAACTCAAA
ATATTATATC TATCCTAAAAA AAAAAATATG TTATATACAAAA TATATACACT

3921 TATGTTTACG TATACAAAAA TATGAAACAA ATAGCGCTAC AGGAAATATT AATCTGAAAA ATATTATACAA

4061 ATACAAAAAA ATATGCGTTA ATATCGTTCA AGAGAAGATA AAAAAATAGA TTATATACAAAA TATATACACT

4131 TTATATGATA TATATAAAAAAT GAAATTTAAA GAAATATTAA AAACCTCTGAC TATATACGTTA CAAATATTAA

4201 GTATGTTAAA AATGCGGATT TCTGTTGATG ATGGCAATTCT ATGACGATTG TGTTTCTTGG AGAACAGATG TATGCGAC

4271 AAGGTTTACG ATTACATTAAAA TTTGCGTTTCT CCTCCTTCTT AAAAAGCGTT TTTGCTTCT ATTTGCGTTT

4341 TTAAGAGTTT CTAAACAGTT AGGGCTTTTT CCCGGCTGTT AATACCCGTT TAAAAAGATC TTATATATAT ATGGATATGA

4411 CTAAGACCTG CATCTGTGTG GAAAGAAAAA CACTACAAAA ATAGTGTATT TTTTATACAAA TATGCAATTG GAGATATGG

4481 AGATAGGCTG GTGACAAGCT GAGCTGTTCTA TCTTCTTAT AAACGCTACG AGTATTAAAC ATGACGGAAG

4551 GCATATCTGC TTGCTGAGT GCAGCTGCAAA AAAAAATATG TTATTTCATG TTATATACAC TATATGCTTT

4621 ACTACACAAA AAGCTCTACAA ATCTTTACTA TGTTAATGTT TTGCAAGTAG AGAAATATG

296
Appendix 5: Nucleotide sequence of the *waaF* - *gmhA* region of the *C. jejuni* 2523/90 LOS biosynthesis cluster. This 2.1 kb fragment was amplified using the primers GMHAF1 and WAAFR1 (Appendix 1). The PCR product was digested with *BamHI* before being ligated into pUC19 to form pNO1.6 (details in Chapter 4).
Appendix 5

1821 AAACAAAGCA TATTAAGTTC TTTTCTTTTT TTTAAAGCCT TAAAACATT AGGGCTTTTT CCACCTGTTG
TTGGTTGCTG ATAAATCAAG AAAAAAGAAA AAATTTTCGGA GATTTTTGAA TCCCCAAAAA GGGAGGCAAAC
1891 AAATACCGAT TAAAAACATCA TTTTCAATTCTA CTAAGCTTCC CACTTGCTCTTTT GAAAACAAA ACTCAAACC
TTATGGCTA ATTTTTGATG AAAAGTAAAGG GATTTCCGAAAGG TGAACAGAAA CTTTTTTGTTA TGGTTTTTGG
1961 ATAGTCATATT CCTATGCGCA TAAAGTCGTA AGTACCGTT GTAAGTCGTA TGGCTGCAAA AGCCCCGCGT
TATCAGTAAA GGCATCCTG AGTTACCGACT TACACGACT CTATAAGCCAA CATTACCGAT AGGGAGGATT TGGGAGGCAA
2031 TCTTTTTTAT AACGCCACT AGTTTCACGT GCAAAAATGCT GAGCATCAGC TGCCTCCTAA AGCCTTGCGT
TATCAGTAAA GGCATCCTG AGTTACCGACT TACACGACT CTATAAGCCAA CATTACCGAT AGGGAGGATT TGGGAGGCAA
2101 AAATCATAAT TTGGCCACGT TTTTTAAGAGC ACTTGCAACAA AAGCTCAC 3'
TTTAGTTTTA GAAGGCTGGA AAAAAATTGT TGAGGCGTGT TTGGAGTG 5'

299
Appendix 6: Nucleotide sequence of the waaF - gmhA region of the C. jejuni NCTC 11351 LOS biosynthesis cluster. This 1.5 kb fragment was amplified using the primers GMHAF1 and WAAFR1 (Appendix 1). The PCR product was digested with BamHI before being ligated into pUC19 to form pNOL.7 (details in Chapter 4).
Appendix 7: Nucleotide sequence of the gmhA - waaF region of the C. jejuni 8F 169 LOS biosynthesis cluster. This 1.5 kb fragment was amplified using the primers GMIHF1 and WAAFR1 (Appendix 1). The PCR product was digested with BamHI before being ligated into pUC19 to form pNOI21 (details in Chapter 4).
Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


