Molecular Analysis of Capsular Polysaccharide Biosynthesis in *Escherichia coli* K5

Annabel N. Smith

Thesis presented for the degree of Doctor of Philosophy

University of Leicester

1995
Abstract

Molecular Analysis of Capsular Polysaccharide Biosynthesis in Escherichia coli K5

Annabel N. Smith

Escherichia coli can produce over 70 structurally distinct acidic capsular polysaccharides (K antigens) and these have been divided into two groups. The genetic determinants required for the expression of group II K antigens map at kps and have previously been shown to have a conserved genetic organization consisting of three functional regions. The nucleotide sequence of region 3 of the K5 kps locus was determined and this revealed the presence of two genes, kpsM and kpsT. KpsM and KpsT have characteristics in common with members of the ABC family of membrane transporters and since region 3 had previously been implicated in polymer translocation across the inner membrane, it is proposed that KpsM and KpsT constitute an inner-membrane capsular polysaccharide export complex. Proteins homologous to KpsM and KpsT are encoded by the capsule loci of Haemophilus influenzae and Neisseria meningitidis, possibly suggesting that capsule production in these unrelated species has a common evolutionary origin. Region 2 of the kps locus has previously been shown to comprise a cassette of polymer synthesis genes. Determination of the nucleotide sequence of region 2 of the K5 kps locus revealed the presence of at least four genes, kfiA, kfiB, kfiC and kfiD. The KfiC protein may act as a glycosyltransferase and KfiD may be a UDP-glucose dehydrogenase. The functions of KfiA and KfiB are unknown. It was demonstrated that region 2 could direct the synthesis and polymerization of the K5 polymer in the absence of region 1 and region 3 encoded proteins. Region 2 has an atypically low G+C content and may have been acquired from another microorganism, possibly through recombination events occurring in the conserved flanking genes, kpsS and kpsT. These genes have a variable 3' end, which may have arisen as a result of in-frame fusions with incoming region 2 elements.
Statement

The accompanying thesis submitted for the degree of Doctor of Philosophy entitled "Molecular Analysis of Capsular Polysaccharide Biosynthesis in *Escherichia coli K5*" is based on work conducted by the author in the Department of Microbiology at the University of Leicester mainly during the period between July 1988 and December 1991.

All the work recorded in this thesis is original unless otherwise acknowledged in the text or by references.

None of the work has been submitted for another degree in this or any other University.

signed. [signature] date. 18th October 1995
Acknowledgements

I sincerely thank Dr. Graham Boulnois and Prof. Ian Roberts for their supervision of this work. I also thank Prof. Klaus and Dr. Barbara Jann and the members of their group for the useful collaboration with regards to this project. I would like to thank the other members of the capsule group, Sumita Ganguli, Rowan Pearce and especially Rachel Drake and Carlo Pazzani, for their contribution to my work and for all of the useful discussions we had during the three years. I must also thank other people in the laboratory for their encouragement, particularly Tim Wallis and Mark Coleman. I am also very grateful to John Sutherland for all of his support during this project. Since I have been in Oxford, a number of people have given me much help and encouragement. I would particularly like to thank Scott Tebbutt, Catherine Wardle, Ann Harris and especially Nicky High, Karen Carpenter and Andrew Walley. I also thank Dr. Bill Cookson and his group for allowing me to use their computing facilities. My thanks also go to Andrew Blakeley and Alan Fowler for their support during these last difficult months. Finally, I would like to dedicate this thesis to my parents and thank them for all their patience and support over the last few years.
CONTENTS

1 Introduction ................................. 1
  1.1 A general introduction to bacterial exopolysaccharides ................................. 1
    1.1.1 The physical and chemical nature of exopolysaccharides ............................ 1
    1.1.2 The function of exopolysaccharides .................................................. 4
    1.1.3 Exopolysaccharides as virulence factors ............................................ 5
    1.1.4 Why study EPS synthesis? ................................................................. 8
  1.2 The polysaccharide antigens of Escherichia coli ............................................. 9
    1.2.1 The O antigen and lipopolysaccharide ................................................. 10
    1.2.2 The K antigen (colanic acid) ............................................................ 12
    1.2.3 The enterobacterial common antigen ................................................... 13
  1.3 The K antigens of Escherichia coli .................................................................. 13
    1.3.1 Group 1 and group II K antigens ......................................................... 14
    1.3.2 Group I/II K antigens ............................................................................. 21
  1.4 The biosynthesis of group II K antigens ......................................................... 22
    1.4.1 Cloning and analysis of the group II kps locus ....................................... 22
    1.4.2 The mechanisms of bacterial polysaccharide biosynthesis ............................ 24
    1.4.3 Region 2 and the synthesis of group II K antigens ..................................... 27
    1.4.4 The role of regions 1 and 3 in the export of group II K antigens ................... 34
    1.4.5 The role of membrane adhesion sites and porins in the export of group II K antigens. 42
  1.5 Other bacterial species which express group II capsular polysaccharides ...................... 43
    1.5.1 The Haemophilus influenzae cap locus ................................................... 43
    1.5.2 The Neisseria meningitidis cps locus ................................................... 44
  1.6 The generation of group II K antigen diversity ................................................. 46
    1.6.1 The clonal population structure of E. coli and genetic diversity ..................... 46
    1.6.2 Evolution of the kps locus ....................................................................... 49
  1.7 Temperature regulation of E. coli group II capsule biosynthesis .............................. 51
  1.8 The aims of this project ................................................................................. 54
## 2 Materials and Methods

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Chemicals, media and enzymes</td>
<td>55</td>
</tr>
<tr>
<td>2.2 Bacterial strains and plasmids</td>
<td>55</td>
</tr>
<tr>
<td>2.3 Transformation of bacterial cells</td>
<td>56</td>
</tr>
<tr>
<td>2.4 Procedures used for DNA extraction</td>
<td>57</td>
</tr>
<tr>
<td>2.5 Procedures used in routine DNA manipulation</td>
<td>61</td>
</tr>
<tr>
<td>2.6 Screening for recombinant plasmids</td>
<td>62</td>
</tr>
<tr>
<td>2.7 DNA hybridization techniques</td>
<td>63</td>
</tr>
<tr>
<td>2.8 DNA sequencing</td>
<td>67</td>
</tr>
<tr>
<td>2.9 Overexpression of proteins</td>
<td>70</td>
</tr>
<tr>
<td>2.10 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)</td>
<td>71</td>
</tr>
<tr>
<td>2.11 Minicells</td>
<td>74</td>
</tr>
<tr>
<td>2.12 Extraction and detection of capsular polysaccharide</td>
<td>76</td>
</tr>
<tr>
<td>2.13 Preliminary promoter analysis</td>
<td>78</td>
</tr>
</tbody>
</table>

## 3 Region 3 of the K5 Capsule Gene Cluster: Identification of an Inner-Membrane Capsular Polysaccharide Transport System

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Introduction</td>
<td>81</td>
</tr>
<tr>
<td>3.2 Results</td>
<td>82</td>
</tr>
<tr>
<td>3.2.1 The precise localization of region 3 of the K5 capsule gene cluster</td>
<td>82</td>
</tr>
<tr>
<td>3.2.2 Cloning and nucleotide sequence analysis of region 3</td>
<td>83</td>
</tr>
<tr>
<td>3.2.3 Comparison of region 3 from the K1 and K5 antigen gene clusters</td>
<td>85</td>
</tr>
<tr>
<td>3.2.4 KpsM and KpsT are members of the ABC family of transporters</td>
<td>86</td>
</tr>
<tr>
<td>3.2.5 Identification of a common mechanism for polysaccharide export in bacteria expressing group II capsules</td>
<td>91</td>
</tr>
<tr>
<td>3.2.6 Visualization of the KpsM and KpsT proteins.</td>
<td>93</td>
</tr>
<tr>
<td>3.2.7 Preliminary studies of the region 3 promoter</td>
<td>98</td>
</tr>
<tr>
<td>3.3 Discussion</td>
<td>101</td>
</tr>
</tbody>
</table>
4 Region 2 of the K5 Capsule Gene Cluster: Preliminary
Analysis of the Polymer Synthesis Genes 106
4.1 Introduction .....................................106
4.2 Results ...........................................107
  4.2.1 Cloning and nucleotide sequence analysis
        of region 2 ...................................107
  4.2.2 Southern hybridization to determine whether
        the K4 and K5 kps loci have any homologous
        polymer synthesis genes .......................111
  4.2.3 Minicell analysis ............................114
  4.2.4 Region 2 alone may be able to direct the
        synthesis and polymerization of the K5
        polysaccharide ...............................116
  4.2.5 Preliminary studies of the promoters of the
        kfiA, kfiB and kfiE genes .......................118
4.3 Discussion .......................................120

5 Genetic Variability in Group II K Antigen Gene Clusters 126
5.1 Introduction .....................................126
5.2 Results ..........................................127
  5.2.1 The kpsE genes from E. coli K5 and K7 are
        highly conserved ................................127
  5.2.2 KpsS and KpsT have a variable C-terminus ...130
  5.2.3 G+C content analysis of the K5 kps locus ...132
5.3 Discussion .......................................133

6 Discussion ..........................................136

References ..........................................143
LIST OF ABBREVIATIONS

ABC  ATP-binding cassette
AP   alternative pathway
APS  ammonium persulphate
bp   basepair
BSA  bovine serum albumin
CCCP carbonyl cyanide m-chlorophenylhydrazone
CP   classical pathway
CPS  capsular polysaccharide
CKS  CMP-KDO synthetase
ECA  enterobacterial common antigen
EDTA ethylenediaminetetra-acetic acid (disodium salt)
EPS  exopolysaccharide
ET   electrophoretic type
GCG  University of Wisconsin Genetics Computer Group sequence analysis software
GlcA glucuronic acid
GlcNAc N-acetylglicosamine
HEPES N-[2-hydroxyethyl]piperazine-N'-(2-ethane sulphonic acid)
IPTG isopropyl-β-D-thiogalactoside
Kb   kilobase pair
kDa  kiloDalton
KDO  3-deoxy-D-manno-octulosonic acid or 2-keto-3-deoxy-octulosonic acid
LOS  lipooligosaccharide
LPS  lipopolysaccharide
MAC  membrane attack complex
ManNac N-acetylmannosamine
MCS  multiple cloning site
mins minutes
MLEE multilocus enzyme electrophoresis
NeuAc N-acetylenuraminic acid (sialic acid)
ORF  open reading frame
PBS phosphate buffered saline (OXOID - 1 tablet dissolved in 100ml distilled water)
PEG  polyethylene glycol
Pfu  plaque forming units
PMSF  phenylmethylsulphonyl fluoride
SDS  sodium dodecyl sulphate
SDS-PAGE  SDS-polyacrylamide gel electrophoresis
SSC  standard saline citrate (20X: 3M NaCl, 0.3M trisodium citrate)
TE  Tris-EDTA (10mM Tris-Cl pH 8.0, 1mM EDTA pH 8.0)
TEMED  N,N,N',N'-tetramethylethylene-diamine
TRIS  Tris-[hydroxymethyl]-methylamine
und-P  undecaprenyl phosphate
X-GAL  5-bromo-4-chloro-3-indolyl-β-D-galactoside
CHAPTER 1

INTRODUCTION

The bacterial cell is capable of synthesizing a number of polysaccharides. Some of these provide mechanisms for storing carbon or energy and others are components of cell wall structures such as peptidoglycan and teichoic acids. However, bacteria may also synthesize polysaccharides which are located at the cell surface and these were referred to as exopolysaccharides (EPS) by Sutherland (1972). Such polysaccharides are found associated with diverse bacterial genera in a wide range of natural and pathogenic environments.

1.1 A GENERAL INTRODUCTION TO BACTERIAL EXOPOLYSACCHARIDES

1.1.1 The Physical And Chemical Nature Of Exopolysaccharides

Bacterial exopolysaccharides can occur in two basic forms (Whitfield, 1988). The polysaccharide may be organized into a distinct structure intimately associated with the cell surface, referred to as a capsule or capsular polysaccharide (CPS). Alternatively, polysaccharide may be present as an extracellular slime, which is only loosely associated with the surface of the bacterium. Distinguishing between these two forms of EPS can be difficult, especially if CPS material is released from the cell surface, giving the appearance of slime production. Distinction between CPS and slime is often determined by the degree of cell association of the polysaccharide following centrifugation.

After a brief introduction to the general features of exopolysaccharides, the biosynthesis of capsular polysaccharide (K antigen) in the Gram-negative bacterium *Escherichia coli* will represent the main subject of this thesis.
Originally light microscopy and negative staining techniques, using for example India ink (Duguid, 1951), were successfully used to visualize EPS and could generally distinguish between encapsulated cells and those bacteria which produced extracellular slime. However, since exopolysaccharides usually consist of more than 95% water (Bayer, 1990), problems have arisen during the development of further techniques for the visualization of EPS by both light and electron microscopy. Water must be retained in the preparation or be substituted with other suitable molecules during dehydration and embedding procedures, to prevent the collapse of the polysaccharide. Stabilization of EPS for microscopy has now been successfully achieved by the use of specific antibodies, plant lectins, cationic dyes and cationic proteins (Bayer, 1990).

Microscopy has shown that in the case of CPS, the capsule generally totally surrounds the bacterium, although the presence of an asymmetric capsule, covering one portion of the bacterium only, has been reported for Rhizobium japonicum (Vasse et al., 1984). Thin polysaccharide fibres which span from the outer membrane to the periphery of the capsule have been identified in the capsular polysaccharides of a variety of bacteria. In E. coli K29, measurements revealed the fibres to have a diameter of 2-4 nm and approximately 3500 capsule fibres were estimated to be present on each cell (Bayer, 1990). Unfortunately, the nature of the linkage between capsular polysaccharides and the cell wall has been difficult to identify in most cases. However, phospholipids have been implicated in the attachment of CPS to the outer membrane in a number of Gram-negative bacteria, since in Neisseria meningitidis group A, B and C and in some E. coli capsular polysaccharides, phosphatidic acid was found to be covalently attached to the reducing end of the polysaccharide via a phosphodiester bridge (Gotschlich et al., 1981; Schmidt and Jann, 1982). Kuo et al. (1985) also found a phospholipid moiety covalently associated with the Haemophilus influenzae type b capsular polysaccharide. In contrast, in some E. coli strains the capsular polysaccharide is reported to be linked to core-lipid A (Jann and Jann, 1990). In the case of Gram-positive bacteria, there are some reports of covalent attachment of CPS
to the peptidoglycan. In serotype III group B streptococci the covalent linkage is to the peptide moiety of the peptidoglycan (Yeung and Mattingly, 1986), whereas in Micrococcus lysodeikticus a phosphodiester linkage appears to be involved (Nasir-ud-Din et al., 1985).

Exopolysaccharides are linear or branched polymers of varying complexity formed from repeating oligosaccharide subunits, an exception being alginate, the EPS produced by Azotobacter vinelandii and Pseudomonas aeruginosa, where the polymer consists of irregular blocks of polymannuronic acid and polyguluronic acid interspersed with regions of alternating mannuronic and guluronic acid (Sutherland, 1985). Common components of EPS include neutral hexoses, uronic acids and amino sugars. Non-carbohydrate substituents such as phosphate, formate, pyruvate, succinate and acetate may also be present (Sutherland, 1972). The wide range of available monosaccharides, their sequence and linkage in the repeating subunit and the presence of non-carbohydrate components results in the production of chemically diverse exopolysaccharides. As one would expect, EPS generally varies between species. However, strains within a single species can also produce a diverse array of polysaccharides; for example, Streptococcus pneumoniae elaborates 84 serologically distinct capsular polysaccharides. Interestingly, despite the diversity found in EPS, serological cross-reactivity and chemical identity does occur between polysaccharides produced by different bacterial genera (Table 1.1).

Many exopolysaccharides are negatively charged homo- or heteropolymers. However, the production of neutral homopolysaccharides such as mutans, dextrans and levans is common in bacteria; Erwinia amylovora, for example, is able to synthesize both a levan and the anionic heteropolysaccharide amylovorin (Bennett and Billing, 1980). Exopolysaccharides are synthesized in different growth phases and under a variety of growth conditions, depending on the organism studied (Sutherland, 1982, 1985). The biosynthetic processes involved can essentially be divided into two principal categories based on the site of synthesis and the nature of the precursors. The
Table 1.1 Some examples of serological cross-reactivity between bacterial capsular polysaccharides

<table>
<thead>
<tr>
<th>Cross-reaction</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Escherichia coli</em> K1</td>
<td>Grados and Ewing (1970)</td>
</tr>
<tr>
<td>Neisseria meningitidis group B</td>
<td></td>
</tr>
<tr>
<td><em>Escherichia coli</em> K92</td>
<td>Glode et al. (1977)</td>
</tr>
<tr>
<td>Neisseria meningitidis group C</td>
<td></td>
</tr>
<tr>
<td><em>Escherichia coli</em> K100</td>
<td>Schneerson et al. (1972)</td>
</tr>
<tr>
<td>Haemophilus influenzae type b</td>
<td></td>
</tr>
<tr>
<td><em>Escherichia coli</em> K7</td>
<td>Robbines et al. (1975)</td>
</tr>
<tr>
<td>Streptococcus pneumoniae type 3</td>
<td></td>
</tr>
<tr>
<td><em>Escherichia coli</em> K30</td>
<td>Chakraborty et al. (1980)</td>
</tr>
<tr>
<td>Klebsiella K20</td>
<td>Homonylo et al. (1988)</td>
</tr>
</tbody>
</table>
neutral homopolysaccharides mentioned above are often synthesized outside of the cell by extracellular enzymes and sucrose is generally used as the precursor. In contrast, most of the anionic exopolysaccharides studied to date appear to be synthesized at the cytoplasmic membrane utilizing sugar nucleotide precursors. Troy (1979) reviewed the general model which has been proposed for synthesis of these polymers. A brief outline of this model will be given here, although biosynthesis will be discussed in more detail later, with particular reference to the capsular polysaccharides of *E. coli* (Section 1.4). The model proposes that synthesis of EPS generally involves the stepwise assembly of repeating oligosaccharide subunits, facilitated by the sequential transfer of monosaccharides from sugar nucleotide precursors to a carrier lipid, undecaprenyl phosphate, located in the cytoplasmic membrane. These lipid-linked intermediates are subsequently polymerized into the mature polysaccharide. However, in some cases there is no evidence for the involvement of lipid-linked intermediates, even though synthesis occurs at the cytoplasmic membrane and utilizes sugar nucleotide precursors, for example cellulose biosynthesis in *Acetobacter xylinum* (Aloni et al., 1983).

**1.1.2 The Function Of Exopolysaccharides**

Exopolysaccharides are not essential for bacterial viability in the laboratory. Capsules or slime can be removed physically or enzymatically without adverse affect on bacterial growth and exopolysaccharide negative mutants are readily isolated, either as a result of spontaneous mutations or after mutagenesis (Sutherland, 1972). However, EPS may be important for survival under the competitive conditions encountered in a natural habitat. Bacteria which produce exopolysaccharides are present in many ecological niches such as the soil, aquatic environments, the plant rhizosphere and animal tissues, but the role of EPS in such environments has not always been clearly established. Most of the functions of EPS are thought to be of a protective nature. The highly hydrated polysaccharide layer may delay dessication of the bacterium in adverse conditions and protect against antibacterial agents such as bacteriocins,
bacteriophage, antibodies and antibiotics (Costerton et al., 1981). However, EPS production may be disadvantageous in some circumstances, since certain bacteriophage only recognize and infect bacteria encapsulated by a specific exopolysaccharide (Lindberg, 1977). The anionic nature of EPS may enable it to act as an ion exchange resin, concentrating minerals and nutrients in the vicinity of the cell (Costerton et al., 1981). EPS is also thought to facilitate adhesion of bacteria to inert and biological surfaces, resulting in adherent biofilms which are the predominant form of bacterial growth in nature and in many diseases (Costerton et al., 1981). The extent of the involvement of EPS in adhesion is under debate and other surface structures, such as pili, may be involved in the initial adhesive interaction, with EPS being responsible for the subsequent stabilization and persistence of the biofilm. The occurrence of biofilms is important in microbiological problems as diverse as the corrosion of pipelines and the onset of dental caries (Costerton et al., 1987). In addition, EPS has been shown to play a role in cellular recognition events; for example, the specificity in a *Rhizobium*-legume symbiotic relationship results from the interaction of plant lectins and bacterial exopolysaccharides (Bohlool and Schmidt, 1974). However, the most studied aspect of EPS function has been its association with the virulence of a number of pathogenic bacteria and this will be discussed below.

### 1.1.3 Exopolysaccharides As Virulence Factors

Since many encapsulated bacteria are harmless, the elaboration of a capsule per se does not ensure virulence in a bacterial species. However, the capsules of some Gram-negative and Gram-positive bacteria possess structural and functional characteristics which are important in the pathogenesis of infections in animals, plants and insects (Sutherland, 1977). In plants, for example, the EPS synthesized by *E. amylovora* is thought to cause vascular occlusion, resulting in the characteristic wilting observed in apple and pear fireblight (Bennett and Billing, 1978). Also many of the most serious invasive infections which affect man, including meningitis, septicaemia, pneumonia and pyelonephritis, are caused by
encapsulated bacteria (Moxon and Kroll, 1990). However, since decimation of the host is unlikely to be of any benefit to the bacterium, it must be remembered that such infectious diseases are probably an incidental, or perhaps accidental, occurrence in the co-evolution of man and microbe, reflecting the struggle of each to survive (Moxon and Kroll, 1990). Thus, although capsular polysaccharides often constitute important virulence factors, their primary role may be to protect the bacterium against antibacterial agents, facilitating the exploitation of the host as a nutrient source. In addition, by preventing dessication, capsules may enhance survival during spread of the bacterium from one host to another.

Bacteria that cause invasive infections must be able to evade host defence mechanisms such as complement-mediated bacteriolysis and opsonophagocytosis. For efficient phagocytic ingestion, the bacterium must be opsonized with immunoglobulin and/or complement, notably C3b. C3b can be generated by either the classical pathway (CP) or the alternative pathway (AP), with the latter being of particular importance in the non-immune host, because it is generally antibody-independent. The AP is activated by many surface-exposed polysaccharides such as lipopolysaccharides (LPS), teichoic acids and peptidoglycan (Cross, 1990). However, CPS may sterically hinder C3b deposition on these underlying structures or mask bound C3b from phagocyte receptors. In addition, some capsular polysaccharides are poor activators of the AP. Generally when C3b is deposited on the bacterial surface, amplification of the response occurs through a C3 convertase generated by the action of serum factors B, D and P on C3b. Factor H can compete with factor B for surface bound C3b, forming H-C3b complexes which are degraded by factor I, terminating the amplification loop and preventing opsonization. A number of different mechanisms whereby capsular polysaccharides limit the activation of the AP have been identified. Sialic acid-containing capsules such as those of E. coli K1, N. meningitidis group B and type III group B streptococci enhance the affinity of factor H for cell-bound C3b, the type 7 and 12 pneumococcal capsules result in a decreased binding affinity for factor B and the ribosyl-ribitol phosphate capsule of H. influenzae type b is a poor acceptor.
for covalent C3b deposition (Joiner, 1988; Moxon and Kroll, 1990). Stevens et al. (1983) examined 15 different capsular polysaccharides of E. coli for their ability to activate the alternative pathway. K types 1, 3, 5, 12 and 92 were found to be poorly opsonized and K types 2 and 13 moderately opsonized by C3b. This correlates well with those capsular antigens found associated with invasive extraintestinal isolates of E. coli. Unfortunately, the mechanisms by which the non-sialic acid-containing E. coli K antigens limit activation of the alternative pathway have not been identified.

Activation of the classical pathway generally requires the presence of specific antibodies. However, some capsular polysaccharides are poorly immunogenic. This results from their similarity to host molecules; for example, in E. coli the structure of the K1 polysaccharide is identical to the terminal carbohydrate region of the embryonic form of the neural cell adhesion molecule and the K5 polysaccharide is identical to the first polymeric intermediate in the biosynthesis of heparin (Jann and Jann, 1990). Such molecular mimicry enables the bacterium to evade the host's specific immune response, thus preventing the activation of the classical complement pathway. Finally, capsular polysaccharides are generally hydrophilic and negatively charged, characteristics which are believed to be intrinsically antiphagocytic (Stendahl, 1983). Therefore it appears that certain capsular polysaccharides are able to protect the bacterium against opsonophagocytosis. In addition, in Gram-negative bacteria, encapsulation may also confer an increased resistance to the bactericidal activity of complement. Cytotoxicity is mediated by insertion of the complement C5b-9 membrane attack complex (MAC) into the lipid bilayer, which disrupts the integrity of the cell. Unfortunately, the evidence advocating a major role for CPS in resistance to complement-mediated serum killing is contradictory (Johnson, 1991; Orskov and Orskov, 1985). LPS, however, has been shown to be important in serum resistance, whereby long-chain LPS molecules result in the formation of the MAC at a location distant from the outer membrane. Such MAC complexes can be sloughed off the surface of the bacterium before damage occurs (Joiner, 1988). A similar mechanism could conceivably
occur with certain capsular polysaccharides (Cross, 1990). Another potential mechanism for avoidance of host defences is the shedding of capsular material from the surface of the bacterium. This may facilitate the removal of attached host opsonins and also possibly divert serum components away from the cell (Moxon and Kroll, 1990).

In conclusion, certain capsular polysaccharides appear to play an important role in evasion of the host’s specific and non-specific immune defences. However, the multifactorial nature of bacterial virulence must not be forgotten and thus, although capsular polysaccharides appear to contribute significantly to the pathogenicity of invasive bacteria, they only represent one of several potential virulence factors.

1.1.4 Why Study EPS Synthesis?

A number of bacterial exopolysaccharides have important applications in the fields of medicine and industry. Polysaccharides are commonly used in the food, pharmaceutical, cosmetic, oil, paper, agricultural and textile industries, mainly as stabilizers and as agents for suspension, gelling and viscosity control. The capsular polysaccharides of some pathogenic bacteria are used in vaccines, for example the 23-valent pneumococcal polysaccharide vaccine (Jennings, 1990). The disruption of capsule synthesis by the use of appropriate inhibitors may represent a future approach to the treatment of invasive infections caused by encapsulated organisms (Jann and Jann 1983). The *E. coli* K5 capsular polysaccharide is identical to an intermediate in the biosynthetic pathway of heparin and may be useful in the commercial production of this important anti-coagulant. In order to maximize the commercial exploitation of exopolysaccharides, a thorough understanding of the processes involved in the synthesis and export of EPS is required. Such knowledge would allow the genetic manipulation of the physical characteristics and quantity of EPS synthesized, and possibly facilitate the production of novel polysaccharides. The K antigens of *E. coli* are amongst the best studied exopolysaccharides and the aim of this project was to further investigate their biosynthesis, by carrying out a
molecular analysis of the K5 antigen gene cluster, in conjunction with Carlo Pazzani (Pazzani, 1992).

1.2 THE POLYSACCHARIDE ANTIGENS OF Escherichia coli

*Escherichia coli*, first described by Theodor Escherich in 1885, is a member of the Enterobacteriaceae, which are Gram-negative, rod shaped, facultative anaerobes. *E. coli* is generally found as a harmless member of the commensal flora associated with the intestinal tract of humans and warm blooded animals. However, *E. coli* can also be regarded as an opportunistic pathogen and is often the prime agent in surgical and other nosocomial infections in compromised patients. Furthermore, certain strains of *E. coli* are inherent pathogens, being commonly associated with both intestinal and extraintestinal disease in humans and animals.

Those strains which cause intestinal infections are generally divided into four main groups: enteropathogenic *E. coli* (EPEC), associated with infantile gastroenteritis; enterotoxigenic *E. coli* (ETEC), responsible for "Traveller's diarrhoea" and other food or water borne diarrhoeal disease; enteroinvasive *E. coli* (EIEC), which induce a dysentery-like disease; and enterohaemorrhagic *E. coli* (EHEC), which cause initially watery and then bloody diarrhoea. The virulence factors associated with these strains are often plasmid encoded and include, for example, colonization factors, enterotoxins, cytotoxins, haemolysin and aerobactin. With regards to extraintestinal infection in man, *E. coli* is the major causative organism of urinary tract infections, which include asymptomatic bacteriuria, cystitis and pyelonephritis (Johnson, 1991). In addition, *E. coli* is a common cause of neonatal septicemia and meningitis, and may also cause septicemia in adults, particularly if they are immunocompromised. The gastrointestinal tract appears to act as a reservoir for the strains which cause these extraintestinal infections (Orskov and Orskov, 1985) and the virulence factors associated with such isolates include fimbrial and non-fimbrial adhesins, haemolysin, aerobactin and certain cell surface polysaccharide antigens (Johnson, 1991; Orskov and Orskov, 1985).
E. coli elaborates a number of different cell surface polysaccharide antigens. Serological techniques developed for typing Salmonella enterica O antigens were first successfully used to analyse the E. coli cell surface by Kauffmann in the 1940s. Employing O antiserum, produced using a boiled culture, he was able to establish well defined E. coli O antigen groups. The O antigen is the thermostable polysaccharide surface antigen found in all smooth Enterobacteriaceae and it represents the outermost part of the cell wall LPS. During his investigations, Kauffmann also demonstrated that many strains isolated from extraintestinal infections elaborated surface antigens that could inhibit the agglutination of live, but not boiled culture in O antiserum. Such surface antigens were believed to surround the cell in an envelope or capsule-like structure, thus masking the O antigen. On the basis of differing thermostability, three classes of these antigens were described, the so called L, A and B antigens. However, because of their many similarities, Kauffmann and Vahlne (1945) proposed to name them collectively as K antigens (K referred to the German word for capsule, Kapsel). Orskov et al. (1977) decided to abandon the L, A and B labels, since the agglutination technique used for their original assignment would sometimes give ambiguous results which were not in agreement with analysis carried out using other more precise techniques, such as immunoprecipitation in agarose gels. Before discussing the K antigen in more detail (Section 1.3), the other cell surface polysaccharide antigens of E. coli will be briefly considered.

1.2.1 The O Antigen And Lipopolysaccharide

A major component of the E. coli cell surface is the lipopolysaccharide, which consists of three regions: lipid A, core oligosaccharide and the O antigen. Lipid A constitutes the toxic and immunostimulatory component of the LPS. It is the hydrophobic moiety which anchors the LPS in the outer membrane, it is probably essential for cell viability and is generally well conserved in the Enterobacteriaceae. In E. coli, lipid A consists of a β-1,6-linked glucosamine disaccharide substituted
with long chain fatty acids (Raetz, 1987). In contrast, *E. coli* may express one of at least five different core oligosaccharides, however they do have a common basal structure. The sugar acid 3-deoxy-D-manno-octulosonic acid (KDO) links the 6' position of the lipid A to the remainder of the core oligosaccharide, which generally consists of heptose, glucose and galactose residues (Orskov et al., 1977; Rick, 1987). The core may be substituted with other sugars and also with phosphate, phosphorylethanolamine and pyrophosphoryl-ethanolamine (Orskov et al., 1977). The O antigen and most of the core sugars can be deleted without adverse effects on cell growth, but the KDO residues seem to be essential for viability of the cell (Raetz, 1987). However, the O antigen does appear to be important in interactions with the host, since strains which do not synthesize O polysaccharide are generally avirulent, especially if they are non-encapsulated (Jann and Jann, 1985). Over 150 O antigens have been identified in *E. coli* and these may be neutral or acidic (Orskov et al., 1977). They consist of oligosaccharide repeating units, composed of up to six different sugar constituents, which in the case of the acidic O antigens include components such as hexuronic acids, N-acetylneuraminic acid, glycerol phosphate and lactyl-substituted sugars (Orskov et al., 1977). Most O antigens are heteropolysaccharides and only two homopolysaccharides have been described in *E. coli*, namely the mannans 08 and 09 (Orskov et al., 1977).

The biosynthesis of LPS involves many genes, often located in clusters at different positions on the bacterial chromosome. In *E. coli*, the genes required for synthesis of the O antigen repeating unit are found in the *rfb* locus (44-45 mins), closely linked to *his*, except for any necessary housekeeping genes (Bachmann, 1990; Orskov et al., 1977). To date, two different mechanisms for O antigen biosynthesis have been described (reviewed in Whitfield, 1995). The two biosynthetic pathways can be referred to as being *rfc*-dependent and *rfc*-independent and will be discussed further in Section 1.4.2. In the case of *rfc*-dependent O antigens, it has been found that the *rfc* gene can be located either within or outside of the *rfb* gene cluster (Reeves, 1993). Most of the genes determining the biosynthesis
of the core region are present in the rfa cluster linked to mtl (80-81 mins) (Bachmann, 1990; Orskov et al., 1977). However, the synthesis of the KDO region of the core requires genes located outside of the rfa cluster, including kdsA, kdsB and kdtA (this gene is actually closely linked to the rfa locus) (Clementz and Raetz, 1991; Goldman et al., 1986; Wöisetschlager and Hogenauer, 1987). In addition, certain housekeeping genes are involved, for example galE (Rick, 1987). Finally, the biosynthesis of lipid A involves at least one operon, ipx, located at 4-5 mins on the E. coli chromosome (Bachmann, 1990; Raetz, 1987).

1.2.2 The M Antigen (Colanic Acid)

In addition to serotype-specific capsular polysaccharide (K antigen), some strains of E. coli can also produce an extracellular slime layer called colanic acid or M antigen. This can be regarded as a common antigen of the Enterobacteriaceae, since many strains elaborate serologically identical or very similar M antigens. The M antigen consists of a hexasaccharide repeating unit composed of glucose, galactose, fucose and glucuronic acid, in which the terminal galactose of the side chain may be substituted with pyruvate, formaldehyde or acetaldehyde (Orskov et al., 1977). Unlike many bacterial capsular polysaccharides, colanic acid does not significantly increase the resistance of E. coli to serum killing or phagocytosis by the host (Allen et al., 1987). In fact, colanic acid is generally only synthesized under stressful physiological conditions, such as relatively low temperature or high osmolarity of the medium (Orskov et al., 1977). Thus, the role of the M antigen may be to aid survival of the bacterium outside of the animal/human intestinal tract, possibly by protecting against dessication. The genes necessary for colanic acid biosynthesis in E. coli K-12 (cps genes) appear to be mainly located near rfb and his at 44-45 mins on the E. coli linkage map (Bachmann, 1990; Trisler and Gottesman, 1984). In addition, a number of housekeeping genes are involved in the synthesis of the required nucleotide sugar precursors, for example galE.
1.2.3 The Enterobacterial Common Antigen

Another cell surface antigen common to the Enterobacteriaceae is the enterobacterial common antigen (ECA), which is located in the outer leaflet of the outer membrane (Kuhn et al., 1988). ECA is a glycolipid, the polysaccharide component of which consists of the amino sugars: N-acetyl-D-glucosamine (GlcNAc), N-acetyl-D-mannosaminuronic acid (ManNAcA) and 4-acetamido-4,6-dideoxy-D-galactose (Fuc4NAc), linked in a trisaccharide repeat unit 3)-α-D-Fuc4NAc-(1,4)-β-D-ManNAcA-(1,4)-α-D-GlcNAc-(1, (Lugowski et al., 1983). The polysaccharide is attached through a phosphodiester linkage to a phosphatidic acid molecule, which probably serves to anchor the molecule in the outer membrane (Kuhn et al., 1988). However, in addition to phospholipid-linked ECA, rough E. coli strains (lacking O antigen) with a complete R1, R4 or K-12 core may have a small proportion of their core-lipid A (about 5%) substituted with ECA polysaccharide (ECA^pg) (Kuhn et al., 1988). In general, apart from housekeeping genes, the ECA biosynthesis genes are located in the rfe-rff region at 85 mins on the E. coli chromosome (Bachmann, 1990; Kuhn et al., 1988). However, in the case of bacteria expressing ECA and a polysaccharide antigen that has one or more precursors/components in common with ECA, some of the ECA genes may be found at alternative loci, for example the rfb locus in S. enterica serovar Typhimurium (Lew et al., 1986). In addition, an intact rfaL gene, which encodes the enzyme responsible for the translocation of O antigen onto core-lipid A, is required for the synthesis of ECA^pg in appropriate rough strains (Kuhn et al., 1988). Although there have been a few reports that ECA may act as a virulence factor, the biological significance of ECA for the enterobacterial cell is not yet fully understood.

1.3 THE K ANTIGENS OF Escherichia coli

Acidic capsular polysaccharides (K antigens) are frequently encountered on the surface of extraintestinal isolates of E. coli and over 70 chemically and serologically distinct K antigens have been described (Orskov and Orskov, 1990; Orskov et al., 1977). The structures of some of these are shown in
Table 1.2. A single isolate only produces one of these antigens and switching of capsular type has not been observed (Boulnois and Jann, 1989; Boulnois and Roberts, 1990). On the basis of physical, biochemical and genetic characteristics, these capsular polysaccharides have been divided into two broad groups, I and II (Jann and Jann, 1987, 1990). The distinctions between the two groups will be discussed below and are summarized in Table 1.3. In addition, recent findings have indicated that a third capsule group should be introduced (Drake et al., 1993; Finke et al., 1990; Pearce, Roberts and Boulnois, personal communication) and the evidence which supports this proposal will be outlined.

1.3.1 Group I And Group II K Antigens

1.3.1.1 Physical And Biochemical Characteristics

Group I includes all the former thermostable A antigens and some B and L antigens, whereas group II represents most of the former thermolabile L antigens (Section 1.2). Group I polymers are reported to have a higher molecular weight (>100 000) than group II polymers (<50 000) and in electron micrographs group I polysaccharides can be seen to be present as thick copious capsules, whereas those of group II form thin patchy capsules (Jann and Jann, 1983). Group I capsular polysaccharides are predominantly co-expressed with the 08 and 09 antigens and occasionally with 020 and 0101. Polysaccharides in group II, however, are found associated with a variety of different O antigens, except 08, 09, 020 and 0101 (Orskov et al., 1977). There are characteristic differences in the nature of the acidic component present in these two groups of capsular polysaccharides (Jann and Jann, 1983). Group I polysaccharides generally contain hexuronic acids (predominantly glucuronic acid) and in some cases pyruvate, whereas group II polysaccharides contain more unusual acidic components, such as KDO, N-acetyleneuraminic acid, N-acetylmannosaminuronic acid and phosphate, in addition to hexuronic acids. Group I polymers generally consist of complex oligosaccharide repeating units with up to six different sugar constituents and often include short side chains. In contrast, the repeating units of group II
Table 1.2 The repeating units of some E. coli K antigens

<table>
<thead>
<tr>
<th>Group</th>
<th>K antigen</th>
<th>Repeating unit</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>II</td>
<td>K1</td>
<td>8)-α-NeuAc-(2,1)</td>
<td>McGuire and Binkley (1964)</td>
</tr>
<tr>
<td></td>
<td>K4</td>
<td>4)-β-GlcA-(1,3)-β-GalNAc-(1,3,2)β-Fru</td>
<td>Rodriguez et al. (1988)</td>
</tr>
<tr>
<td></td>
<td>K5</td>
<td>4)-β-GlcA-(1,4)-α-GlcNAc-(1,3)</td>
<td>Vann et al. (1981)</td>
</tr>
<tr>
<td></td>
<td>K7</td>
<td>3)-β-ManNAcA-(1-4)-β-Glc-(1,6-OAc)</td>
<td>Tsui et al. (1982)</td>
</tr>
<tr>
<td></td>
<td>K92</td>
<td>8)-α-NeuAc-(2,9)-α-NeuAc-(2,1)</td>
<td>Glode et al. (1977)</td>
</tr>
<tr>
<td>I/II</td>
<td>K10</td>
<td>3)-α-Rha-(1,3)-β-Qu4NMal-(1</td>
<td>Sieberth et al. (1993)</td>
</tr>
<tr>
<td>I</td>
<td>K30</td>
<td>2)-α-Man-(1,3)-β-Gal-(1,3,1)β-GlcA-(1,3)-α-Gal</td>
<td>Chakraborty et al. (1980)</td>
</tr>
<tr>
<td></td>
<td>K40</td>
<td>4)-β-GlcA-(1,4)-α-GlcNAc-(1,6)-α-GlcNAc-(1,3,1)CO.NH (serine)</td>
<td>Dengler et al. (1986)</td>
</tr>
</tbody>
</table>

Abbreviations: Fru, fructose; Gal, galactose; GalNAc, N-acetylgalactosamine; Glc, glucose; GlcA, glucuronic acid; GlcNAc, N-acetylglucosamine; Man, mannose; ManNAcA, N-acetylmannosaminuronic acid; NeuAc, N-acetyleneuraminic acid; OAc, O-acetyl; Qu4NMal, 4,6-dideoxy-4-malonylaminoglucone; Rha, rhamnose.
Table 1.3 A summary of the distinctions between *E. coli* group I and group II K antigens

<table>
<thead>
<tr>
<th>Property</th>
<th>Group I</th>
<th>Group II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular weight</td>
<td>&gt;100,000</td>
<td>&lt;50,000</td>
</tr>
<tr>
<td>Electrophoretic mobility</td>
<td>low</td>
<td>high</td>
</tr>
<tr>
<td>Acidic component</td>
<td>GlcA, GalA, pyruvate</td>
<td>GlcA, NeuAc, KDO, phosphate, ManNAcA</td>
</tr>
<tr>
<td>Lipid at reducing end</td>
<td>core-lipid A&lt;sup&gt;®&lt;/sup&gt;</td>
<td>phosphatidic acid</td>
</tr>
<tr>
<td>Stability at pH 5-6/100°C</td>
<td>stable</td>
<td>labile</td>
</tr>
<tr>
<td>Expressed below 20°C</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>CMP-KDO synthetase activity at 37°C</td>
<td>low</td>
<td>high</td>
</tr>
<tr>
<td>Coexpression with</td>
<td>08,09,020,0101</td>
<td>many O antigens</td>
</tr>
<tr>
<td>Chromosomal determination (close to)</td>
<td>cps (his)(trp?)</td>
<td>kps (serA)</td>
</tr>
<tr>
<td>Intergeneric relationship with</td>
<td><em>Klebsiella</em></td>
<td><em>H. influenzae</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>N. meningitidis</em></td>
</tr>
</tbody>
</table>

<sup>®</sup> This substitution has been verified for only a few polysaccharides.

Abbreviations: GalA, galacturonic acid; GlcA, glucuronic acid; KDO, 3-deoxy-D-manno-octulosonic acid or 2-keto-3-deoxy-octulosonic acid; ManNAcA, N-acetylmannosaminuronic acid; NeuAc, N-acetylneuraminic acid.
polymers are typically composed of linear mono- di- or tri-
saccharides and are frequently substituted with O acetyl groups
(Jann and Jann, 1983). The high charge density of the smaller
repeating units of the group II polysaccharides, together with
their lower molecular weight, means that they show a greater
electrophoretic mobility than group I polysaccharides, which
have a higher molecular weight and lower charge density.
Consequently, immunoelectrophoresis has been successfully used
to differentiate between the two groups (Jann and Jann, 1983;
Orskov et al., 1977).

In all group II polysaccharides so far analysed, phosphatidic
acid appears to be attached to the reducing end via an acid-
labile phosphodiester bridge (Gotschlich et al., 1981; Jann and
Jann, 1990; Schmidt and Jann, 1982) (Section 1.1.1). The lipid
moiety may serve to anchor the polysaccharide in the outer
membrane. However, only approximately 40-70% of the poly-
saccharide is substituted with lipid, with the remaining
unsubstituted polymer possibly being retained at the cell
surface by hydrogen bonding and/or charge interactions with
those molecules attached via the lipid (Jann et al., 1992).
Group I polysaccharides are heat stable at pH 5-6 and 100°C.
However, group II polysaccharides are generally released from
the cell under these conditions and this may result from the
extreme acid-lability of the phosphodiester linkage between the
polysaccharide and its lipid anchor (Jann and Jann, 1990).

In group I, the capsular polysaccharides can be subdivided on
the basis of the absence (group IA) or presence (group IB) of
amino sugars in the repeating unit (Jann and Jann, 1983; Orskov
et al., 1977). The amino sugar-containing group IB K antigen,
K40, is substituted at the reducing end with core-lipid A and
it has been proposed that this moiety acts as a membrane anchor
for the polysaccharide (Jann and Jann, 1990; Jann et al.,
1992). Thus, strains with such K antigens elaborate two cell
wall lipopolysaccharides, a neutral one being the O antigen
(08, 09, 020 or 0101) and an acidic one being the capsular K
antigen (Jann and Jann, 1990). Although high molecular weight
K40 polysaccharide has been shown to be attached to core-lipid
A, only a small proportion of the total polysaccharide was
found to be substituted with the lipid moiety (Jann et al., 1992). However, this substituted polymer may interact with and retain unsubstituted K40 polysaccharide at the cell surface (Jann et al., 1992). Interestingly, it was found that some amino sugar-containing group IB polysaccharides could be the K antigen of one E. coli strain and the O-specific polysaccharide moiety of the LPS in another; for example, the structures of the K87, K85 and K9 antigens are identical to those of the O32, O141 and O104 antigens, respectively (Jann and Jann, 1990).

The K30 antigen is an example of an amino sugar-free group IA K antigen and core-lipid A substituted with one or very few repeat units of this capsular polysaccharide has been isolated (K30\textsubscript{LPS}) (Homyolo et al., 1988; Orskov et al., 1977). However, in contrast to the K40 antigen, it has not yet been possible to identify a core-lipid A moiety associated with high molecular weight K30 polysaccharide (Homyolo et al., 1988). In fact, it has been shown recently that K30\textsubscript{LPS} is not required for the cell surface expression of high molecular weight K30 capsular polysaccharide (MacLachlan et al., 1993). In addition, it was found that whilst all strains with group IB K antigens appear to express some K\textsubscript{LPS}, a number of strains expressing group IA K antigens appear to lack K\textsubscript{LPS} (MacLachlan et al., 1993). Thus, the means by which high molecular weight K30 polymer is anchored at the cell surface remain unknown.

Orskov et al. (1984) observed that group I K antigens are generally expressed at both 18°C and 37°C, whereas most of those in group II are produced at 37°C, but not at 18°C (for exceptions see Section 1.3.2). The latter K antigens are commonly associated with invasive, extraintestinal isolates of E. coli, which is in contrast to those K antigens that are expressed at all growth temperatures (Orskov et al., 1984). The primary role of the temperature-dependent group II K antigens may be to provide protection against host defences during infection, whilst colanic acid (Section 1.2.2), which is known to be produced by these bacteria under certain circumstances (Keenleyside et al., 1992, 1993), may aid survival outside of the host. Recently, it has been shown that E. coli strains expressing temperature-independent group IA K
antigens are unable to synthesize colanic acid and there is preliminary evidence suggesting that the K30 and colanic acid biosynthesis genes are allelic (Jayaratne et al., 1993; Keenleyside et al., 1992). Thus, in addition to conferring some protection against defences inside the host, a major role of the capsular polysaccharide in such strains may be to facilitate the survival of the bacterium outside the host in the same manner as colanic acid. In contrast, strains expressing temperature-independent group IB K antigens appear to be able to synthesize both serotype-specific K antigen and colanic acid (Jayaratne et al., 1993).

KDO was recently found to be the reducing sugar in the group II K5 capsular polysaccharide, a polymer which does not contain KDO in its repeating unit (Finke et al., 1989, 1991). The enzyme which catalyses the synthesis of the activated form of KDO, CMP-KDO, is CMP-KDO synthetase (CKS) and a low level of CKS activity is always present in Gram-negative bacteria, since KDO is essential for the biosynthesis of lipid A and cell viability (Section 1.2.1). However, it was found that the activity of this enzyme was significantly elevated in E. coli K5 strains at 37°C as compared to unencapsulated E. coli and rough E. coli K-12 strains (Finke et al., 1989). The elevated CKS activity was observed in bacteria incubated at 37°C, but not after growth at 18°C. Thus, both polysaccharide biosynthesis and the elevated CKS activity appear to be regulated by temperature in E. coli K5 (Finke et al., 1989). Similar observations were made with other group II K antigen expressing strains, although there were some exceptions (see Section 1.3.2) (Finke et al., 1990). Therefore, elevated levels of CMP-KDO synthetase activity may be important for the synthesis of these group II polysaccharides (Section 1.4.3.2). However, this does not appear to be the case for strains expressing group I capsular polysaccharides, since CKS activity was always low in these bacteria (Finke et al., 1990).

1.3.1.2 Genetic Characteristics

All E. coli capsular polysaccharides so far analysed are determined by chromosomal genes (Jann and Jann, 1990). The
genetic determinants required for expression of the K1 and K4 group II K antigens (Orskov et al., 1976) map at a locus termed kps (Silver et al., 1984; Vimr et al., 1989), formerly kpsA (Orskov and Nyman, 1974), which is located near serA (63 mins) on the E. coli K-12 linkage map (Bachmann, 1990). Recently the K1 kps gene cluster has been mapped more precisely to an accretion domain located near 64 mins (Vimr, 1991). The kps genes encoding the biosynthetic functions for a number of group II capsular polysaccharides have been cloned (Drake et al., 1990; Echarti et al., 1983; Roberts et al., 1986; Silver et al., 1981) and subsequent molecular analysis has revealed a common genetic organization (Boulnois et al., 1987; Roberts et al., 1988a). This will be considered in more detail in Section 1.4.

The group I capsular polysaccharide biosynthesis genes have not yet been studied in detail. However, the genetic determinants for the following K antigens have been mapped near to the rfb and his loci (44-45 mins): K26 (Orskov and Orskov, 1962), K27 (Schmidt et al., 1977), K30 (Laakso et al., 1988), and K8, K9, K17 and K57 (Orskov et al., 1977). In addition, a second trp-linked locus is thought to be essential for the full expression of the K27 antigen (Schmidt et al., 1977). The absence of this locus resulted in bacteria which expressed K-antigenicity, but did not synthesize a full capsular layer and this phenotype was termed K1 or K intermediate (Schmidt et al., 1977). It was suggested that the K27i specificity may be due to the presence of single oligosaccharide repeating units linked to an acceptor (possibly core-lipid A). Thus, it was proposed that the trp-linked locus may encode a K antigen-specific polymerase, analogous to the rfc gene product found in certain E. coli and S. enterica strains which is responsible for the polymerization of O antigen repeating units (Section 1.4.2) (Schmidt et al., 1977). However, there is evidence that such a trp-linked locus may not be required for the expression of the K30 antigen. Whitfield et al. (1989) isolated both K30 acapsular and K30i (one or very few K30 repeating units linked to core-lipid A) mutants and demonstrated that all of the genetic determinants required to complement these mutants were present in a his-linked locus, which could be transferred from
the wild type *E. coli* K30 strain. No such complementation occurred in trp' transconjugants and thus a trp-linked locus does not appear to be involved in the expression of the K30 capsule (Whitfield et al., 1989).

These apparently contradictory results, regarding the distribution of loci required for the expression of the K27 and K30 antigens, have not been fully explained, but they may indicate that the genes involved in the biosynthesis of different group I capsular polysaccharides do not share a common genetic organization. It has been proposed that the his-linked group I K antigen biosynthesis functions are encoded by the *rfb* gene cluster responsible for O antigen biosynthesis (Jann and Jann, 1987). However, although genetic crosses have shown that there is a close genetic linkage between the O and K antigen genes (Laakso et al., 1988; Orskov et al., 1977; Schmidt et al., 1977), there is no direct evidence to suggest that they form a single cluster. Some enzymes may be common to the two biosynthetic pathways, but only 2-3% of the K30 mutants isolated by Whitfield et al. (1989) had simultaneously lost both O9 and K30 antigens and this suggests that the number of shared loci is small. Nonetheless, it has recently been shown that *rfbM* and also possibly *rfbK* are required for the synthesis of the K30 antigen in *E. coli* O9:K30 (Jayaratne et al., 1994). These two genes are encoded by the *rfbO9* gene cluster and are involved in the formation of GDP-mannose, which is required for the biosynthesis of both the O9 and K30 antigens.

Whitfield et al. (1989) initially proposed the designation kst (K antigen, thermoSTable) for the genes specifically associated with the expression of group I capsules, to distinguish them from the *rfb* O antigen genes and the group II K antigen *kps* genes. However, since the group IA K30 antigen biosynthesis genes are thought to be allelic with the *cps* genes required for colanic acid synthesis (Keenleyside et al., 1992), the K30 antigen gene cluster is now generally designated *cpsK30*. In the case of strains expressing group IB K antigens, the situation is more complicated. These strains are able to synthesize colanic acid (Jayaratne et al., 1993) and consequently, they must have functional *cps* gene clusters responsible for the
production of both serotype-specific K antigen and colanic acid.

1.3.1.3 Intergeneric Relationships

The capsular polysaccharides of *E. coli* show similarities with those of other genera. Thus, the amino sugar-free group IA capsules share many properties with those of *Klebsiella*, which also never contain amino sugars (Jann and Jann, 1990). The polysaccharides are often structurally very similar and some, for example the *E. coli* K30 and *Klebsiella* K20 antigens, are chemically and serologically indistinguishable (Chakraborty et al., 1980; Homonylo et al., 1988). The genetic determinants required for expression of both of the aforementioned capsules are located near the *his* and *rfb* loci (Laakso et al., 1988) and there is preliminary evidence suggesting that the polysaccharides are synthesized by comparable mechanisms (Jann and Jann, 1983, 1990). Furthermore, the bacteria expressing these related capsular polysaccharides may cause similar extra-intestinal infections, particularly in immunocompromised hosts.

In contrast, the *E. coli* group II capsular polysaccharides have characteristics in common with those of *N. meningitidis* and *H. influenzae*. A number of these polymers are structurally related, for example the polyribosyl-ribitol phosphate containing capsules of *E. coli* K100 and *H. influenzae* type b (Jann and Jann, 1990). Moreover, the polysialic acid capsules of *E. coli* K1 and *N. meningitidis* group B are chemically and serologically identical. In addition, the identification of a phospholipid at the reducing end of several of these polymers suggests that they are linked to the cell surface in a similar manner (Gotschlich et al., 1981; Kuo et al., 1985; Schmidt and Jann, 1982) (Section 1.1.1). Bacteria expressing these closely related capsules are often implicated in similar invasive infections. Recently, the capsular polysaccharide biosynthesis genes from *N. meningitidis* and *H. influenzae* have been cloned and subjected to detailed molecular analysis (Frosch et al., 1989, 1991; Hosieth et al., 1986; Kroll et al., 1989). The capsule gene clusters were found to have a similar genetic organization to that of the *kps* loci of *E. coli* and this will
be discussed further in Section 1.5. It has been proposed that the term "group II capsular polysaccharides" should be used to collectively refer to the capsules of these three unrelated species (Frosch et al., 1991).

### 1.3.2 Group I/II K Antigens

The capsular polysaccharides of *E. coli* can be divided into two groups on the basis of the aforementioned parameters. However, there are exceptions, such as the K2, K3, K10, K11, K19 and K54 antigens, which do not fit easily into either group and have been tentatively classified as group I/II K antigens (Finke et al., 1990). These capsules are considered as group II K antigens on the basis of their molecular weight, composition, substitution with phosphatidic acid, thermostability at pH 5-6 and co-expression with many O antigens (Jann and Jann, 1990; Jann and Jann, personal communication). Furthermore, the genetic determinants for the K10 and K54 antigens appear to map at the group II *kps* locus near *serA* on the *E. coli* chromosome (Orskov and Nyman, 1974). However, the group I/II K antigens also appear to share some biochemical properties with group I capsules. Firstly, the K3, K10, K11 and K54 antigens are expressed at all growth temperatures (Orskov et al., 1984). With regards to the K2 and K19 antigens, temperature-regulated expression has been reported by Orskov et al. (1984), but this has been contradicted by Jann and Jann (personal communication). This discrepancy has not yet been resolved. Secondly, strains expressing the K2, K3, K10, K11, K19 and K54 antigens do not have elevated CKS activity at 37°C (Finke et al., 1990). Thus, on the basis of these criteria the group I/II K antigens have been proposed to represent a new group of *E. coli* capsules (Finke et al., 1990) and this has been further supported by genetic studies. Southern hybridization analysis has shown that the K3, K10, K11, K19 and K54 capsule genes must only have limited, if any, DNA homology with those of group II strains (Drake et al., 1993). In contrast, it was shown that the K2 antigen may in fact be a group II K antigen, albeit one with unusual properties (Drake et al., 1993).

Recently the K10 biosynthesis genes were cloned and preliminary
analysis has shown that DNA highly homologous to the K10 genes is also present on the chromosomes of strains expressing the K3 and K54 antigens (Pearce, Roberts and Boulnois, personal communication). Thus, the biosynthesis genes of these three group I/II K antigens may be closely related. A similar level of DNA homology was not observed in the case of the K11 antigen (K19 was not examined), indicating that the group I/II K antigens may in fact represent more than one group of E. coli capsules. Interestingly, the DNA flanking the genetic determinants of the K10 (group I/II) and K4 (group II) antigens is homologous, which suggests that they are located at the same position on the chromosome (Pearce, Roberts and Boulnois, personal communication). This in agreement with the earlier findings of Orskov and Nyman (1974) and Orskov et al. (1976).

1.4 THE BIOSYNTHESIS OF GROUP II K ANTIGENS

1.4.1 Cloning And Analysis Of The Group II kps Locus

The kps genes encoding functions necessary for the biosynthesis of the E. coli group II K antigens K1, K4, K5, K7, K12 and K92, have been cloned using a cosmid cloning strategy and laboratory strains of E. coli K-12, for example LE392, which are not encapsulated and lack all of the group II capsule determinants (Drake et al., 1990; Echarti et al., 1983; Roberts et al., 1986; Silver et al., 1981). This has shown that all of the requisite genes are clustered and in the case of the K1 antigen approximately 17kb of DNA were found to be essential for capsule production (Boulnois et al., 1987). Detailed molecular analysis, including restriction endonuclease cleavage, DNA hybridization, deletion and transposon mutagenesis and complementation experiments, has revealed that the capsule genes have a conserved genetic organization consisting of three functional regions (Boulnois et al., 1987; Roberts et al., 1988a) (Figure 1.1).

Region 2 is serotype specific and has been proposed to encode the enzymes necessary for the synthesis and polymerization of the K antigen in question, since transposon insertions in this region abolish polysaccharide production (Boulnois et al.,
Figure 1.1 The genetic organization of *E. coli* group II capsule gene clusters

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>43</td>
<td>60</td>
<td>27</td>
</tr>
<tr>
<td>76</td>
<td>46</td>
<td></td>
</tr>
</tbody>
</table>

*kps* E D U C S

1 kb

---

**K5**

| 47 | 28 | 44 | 39 |

**K1**

*neu* S E C A B D

---

Figure 1.1 The genetic organization of the K1 and K5 capsule gene clusters is shown, drawn to scale. The lines labelled with 1, 2 and 3 above correspond to regions 1, 2 and 3, respectively. Regions 1 and 3 are common to each locus, whereas region 2 is serotype specific. The broken lines highlight the difference in the size of region 2 from each capsule gene cluster. The boxes labelled *kps* E D U C S represent the genes identified in region 1 of the K5 locus. Similarly, the boxes labelled *neu* S E C A B D represent the genes identified in region 2 of the K1 locus. The predicted size of the encoded gene products is indicated above in kDa. The arrows denote the direction of transcription.
1987; Roberts et al., 1988a). The size of region 2 is variable and this broadly reflects the complexity of the specific K antigen, for example biosynthesis of the fructose-substituted K4 polysaccharide requires a region 2 of approximately 14kb, the largest so far described (Drake et al., 1990). Thus, region 2 appears to comprise a cassette of biosynthetic determinants which are unique to a particular K antigen. Furthermore, there is no evidence for the existence of other silent region 2 cassettes located elsewhere on the chromosome (Roberts et al., 1988b). This unique region is flanked by regions 1 and 3, which are common to different K antigen gene clusters. Regions 1 and 3 are proposed to encode products mainly involved in translocation of the polysaccharide to the cell surface, since mutations in these regions result in the production of intracellular, immunoreactive polymer (Boulnois et al., 1987; Roberts et al., 1986, 1988a). Hybridization studies have shown that there is extensive DNA homology (generally over 95%) between the conserved regions (1 and 3) of different group II K antigen gene clusters (Roberts et al., 1988a). In addition, mutations in a conserved region of one cluster can be complemented in trans by the equivalent region from a different capsule gene cluster (Roberts et al., 1986, 1988a). Vimr et al. (1989) have proposed that conserved genes (regions 1 and 3) should retain their kps designation, whereas genes unique to a particular K antigen (region 2) should have a designation reflecting the nature of the polysaccharide. Thus the term neu was proposed for the K1-specific biosynthetic genes necessary for the synthesis and polymerization of N-acetylneuraminic acid.

Although the biosynthesis genes for a number of group II capsules have been cloned, those of the medically important K1 and K5 antigens have been the ones most extensively studied. The steps involved in the biosynthesis of these group II K antigens will now be considered in the order in which they occur, that is polymer synthesis (Section 1.4.3) followed by transport of the polysaccharide to the cell surface (Section 1.4.4). However, since relatively little is known about the mechanism of K antigen biosynthesis in E. coli, the synthesis of other bacterial polysaccharides which have been studied in
more detail will be considered briefly first.

1.4.2 The Mechanisms Of Bacterial Polysaccharide Biosynthesis

The first step in the biosynthesis of anionic bacterial polysaccharides involves the synthesis of activated glycosyl donors in the form of nucleotide derivatives of the component sugars. Some of the enzymes involved in this process are unique to the biosynthesis of a particular polysaccharide and others are encoded by housekeeping genes. Such enzymes are generally cytoplasmic, although in some cases they may be loosely associated with the inner membrane, allowing spatial organization with the membrane-bound glycosyl transferase enzymes involved in the next stage of biosynthesis, transfer and polymerization of the sugar components (Sutherland and Norval, 1970). Much of the current knowledge regarding the polymerization of bacterial polysaccharides results from extensive study of O antigen synthesis in *S. enterica* and *E. coli*. To date two different mechanisms of O antigen biosynthesis have been identified: one being *rfc*-dependent and the other *rfc*-independent (reviewed in Whitfield, 1995).

1.4.2.1 *rfc*-Dependent O Antigens

The *rfc*-dependent mechanism was first studied in *S. enterica* and is now known to occur in other bacteria, including some *E. coli* O groups and species of *Shigella*. The biosynthetic mechanism can essentially be divided into three steps (Rick, 1987; Whitfield, 1995). The first step involves the synthesis of a single oligosaccharide repeating unit covalently linked to a membrane-bound carrier lipid, undecaprenyl phosphate (und-P). This process, which is catalysed by specific glycosyl transferases, involves the formation of a high energy pyrophosphoryl linkage between the first sugar and the carrier lipid, followed by the sequential incorporation of the remaining sugars to complete the repeat unit. The initial transferase reaction often involves the transfer of galactose-1-phosphate, catalysed by the galactosyltransferase RfbP. At the present time, the only other known initiating enzyme is Rfe, an N-acetylglucosamine-1-phosphotransferase,
which also functions in the initial step of ECA biosynthesis. Apart from housekeeping genes and the rfe gene if required, all the genes encoding the enzymes necessary for the first step appear to be clustered in the rfb locus. The next step involves the blockwise polymerization of the individual repeat units to form the O-specific polysaccharide still linked to carrier lipid. In this process, O-antigen polymerase, encoded by the rfc gene, catalyses the formation of a glycosidic linkage between the reducing terminal sugar of the nascent O antigen and the non-reducing terminal moiety of a newly synthesized repeat unit. Thus, elongation of the O antigen occurs by growth at the reducing terminus. The final step involves the transfer of the completed polymer from the carrier lipid to the non-reducing terminus of core-lipid A, in a process requiring the product of the rfaL gene. The distribution of O-antigen chain lengths appears to be under the control of a protein known as either Rol (Batchelor et al., 1992) or Cld (Bastin et al., 1993).

The synthesis of oligosaccharide repeat units is thought to occur at the cytoplasmic face of the inner membrane, whereas polymerization of the O antigen and its subsequent transfer to core-lipid A have been shown to take place at the periplasmic face of the inner membrane (Marino et al., 1991; McGrath and Osborn, 1991). Thus, the lipid-linked oligosaccharide repeat units have to be translocated across the inner membrane and it is thought that this process may involve a "flippase" protein, possibly encoded by rfbX (Whitfield, 1995).

1.4.2.2 rfc-Independent O Antigens

The rfc-independent mechanism is currently only known to be involved in the synthesis of the homopolymer O antigens of E. coli O8 and O9 and Klebsiella pneumoniae O1 (Jann and Jann, 1984; Whitfield, 1995). Although synthesis of the E. coli O8 and O9 antigens has been studied in some depth (Jann and Jann, 1984), the finer details of this biosynthetic pathway are only now being elucidated, (Bronner et al., 1994; Kido et al., 1995; Rick et al., 1994). Interestingly, the initial reaction of this pathway is analogous to the first step in the rfc-dependent
mechanism and involves the formation of a pyrophosphoryl linkage between und-P and a sugar-1-phosphate residue, which in this pathway does not form part of the repeating unit of the O antigen. The lipid-linked moiety subsequently acts as the acceptor for monomers of the repeating unit during the polymerization process. The initiation reaction is dependent on the product of the \textit{rfe} gene, but there is some confusion regarding the precise identity of the residue transferred to und-P. In \textit{E. coli} O9, the acceptor had previously been characterized as und-P-P-glucose (Jann et al., 1982). In contrast, the acceptor was found to be und-P-P-N-acetyl-glucosamine in \textit{E. coli} O8/K-12 hybrids, which would be consistent with the function of Rfe in the synthesis of other O antigens and ECA (Rick et al., 1994). Recently, it has been suggested that Rfe may be able to direct the transfer of both sugars (Kido et al., 1995).

In \textit{rfc}-independent O antigen biosynthesis, polymerization occurs by a totally different mechanism to that observed in the \textit{rfc}-dependent pathway and involves the sequential transfer of individual sugar residues from their nucleotide derivative to the non-reducing end of the growing polymer, without the participation of an \textit{rfc}-encoded O-antigen polymerase. At the present time it is not known how the repetitive oligosaccharide sequence is maintained or how the distribution of O-antigen chain lengths is determined. Polymerization is proposed to occur at the cytoplasmic face of the inner membrane. Thus, the nascent polymer has to be translocated across the inner membrane to the periplasmic face, where it subsequently undergoes ligation to core-lipid A. Recently, it has been shown that genes present in the \textit{rfb} loci of \textit{E. coli} O9 and \textit{K. pneumoniae} O1 encode a putative inner-membrane O-antigen export system, which has characteristics in common with members of the ATP-binding cassette (ABC) superfamily of transporters (Bronner et al., 1994; Kido et al., 1995). Functionally analogous transporters appear to be involved in the export of other polysaccharides, including the group II capsular polysaccharides of \textit{E. coli}, \textit{N. meningitidis} and \textit{H. influenzae} (see Chapter 3).
1.4.2.3 Other E. coli Polysaccharide Antigens

The biosynthesis of other E. coli cell surface polysaccharide antigens has not been studied in as much detail. Thus, even though the regulation of colanic acid synthesis has been fairly well characterized (Gottesman and Stout, 1991), relatively little is known about the biosynthetic pathway, although it does appear that lipid-linked intermediates are involved (Sutherland, 1982). The early steps involved in the biosynthesis of ECA have been determined in some detail, wherein the trisaccharide repeat unit (Section 1.2.3) is assembled as an und-P-P-linked intermediate, lipid III. Unfortunately, although it is clear that the subsequent steps utilise lipid III as the substrate for polysaccharide chain elongation, the actual mechanism involved remains to be established (Meier-Dieter et al., 1992). Again, little is known about the synthesis of the Klebsiella-like group I K antigens. However, the biosynthesis of capsular polysaccharide in Klebsiella aerogenes has been shown to occur via the polymerization of lipid-linked intermediates in the same manner as with rfc-dependent O antigens (Sutherland and Norval, 1970; Troy et al., 1971) and there is preliminary evidence indicating that the E. coli K30 antigen is synthesized by a similar mechanism (Jann and Jann, 1990; Orskov et al., 1977). The identification of an intermediate Ki form of the E. coli K27 and K30 antigens, with only one oligosaccharide repeating unit linked to core-lipid A, is in agreement with this, since a defect in an rfc-like polymerase gene would account for this phenotype (Schmidt et al., 1977; Whitfield et al., 1989). Relatively more information is available regarding the synthesis and polymerization of E. coli group II capsular polysaccharides and this will now be considered with reference to the most studied examples, the K1 and K5 antigens.

1.4.3 Region 2 And The Synthesis Of Group II K Antigens

1.4.3.1 The K1 Antigen

The biochemical pathway involved in the synthesis of the K1 polysaccharide has been fairly well characterized. The K1
antigen is composed of up to 200 sialic acid (N-acetyl neuraminic acid, NeuAc) residues joined internally through α-2,8 linkages (Rohr and Troy, 1980). Molecular analysis of the K1 capsule gene cluster has shown that the central segment, region 2, encodes the enzymes required for the synthesis, activation and polymerization of NeuAc (Boulnois et al., 1987; Silver et al., 1984; Vimr et al., 1989). Many of the genes encoding these enzymes have now been characterized and their location in region 2 determined (Silver et al., 1993) (Figure 1.1). The synthesis of NeuAc involves the following reactions:

1. \[ \text{UDP-N-acetylglucosamine} \rightarrow \text{N-acetylmannosamine} + \text{UDP} \]

2. \[ \text{N-acetylmannosamine} + \text{phosphoenolpyruvate} \rightarrow \text{NeuAc} + \text{Pi} \]

It was found that strains harbouring mutations in the region 2 neuB and neuC genes are able to express a wild type capsule when supplied with exogenous NeuAc (Boulnois et al., 1987; Silver et al., 1984; Vimr et al., 1989; Zapata et al., 1992). Thus, the products encoded by these genes are likely to function in NeuAc synthesis. It has recently been proposed that the 39kDa NeuB protein is NeuAc synthase, the enzyme responsible for the second step in NeuAc synthesis, since it has significant homology with the NeuAc synthase encoded by the siaC gene located in the capsule gene cluster of \textit{N. meningitidis} group B (Annunziato et al., 1995; Edwards et al., 1994; Ganguli et al., 1994). This may indicate that the 44kDa neuC gene product (Zapata et al., 1992) is responsible for the conversion of N-acetylglucosamine to N-acetylmannosamine (ManNAC) in the first step of NeuAc synthesis. In support of this, NeuC has been found to have significant homology with the SiaA protein thought to be involved in ManNAC synthesis in group B meningococci (Edwards et al., 1994; Ganguli et al., 1994). Interestingly, Zapata et al. (1992) argue against a role for NeuC in ManNAC synthesis, since they found that mutations in \textit{neuC} could not be complemented by the addition of exogenous ManNAC. However, it has been suggested that exogenous ManNAC is modified during uptake into the cell and consequently, it may be unable to act as a substrate for NeuAc synthase in such complementation experiments (Vimr, 1992). CMP-NeuAc is the
activated form of NeuAc required for participation in the polymerization reaction, and the 49kDa product of the region 2 neuA gene, CMP-NeuAc synthetase, catalyses the formation of this nucleotide monophosphate derivative (Vann et al., 1987; Zapata et al., 1989).

\[ 3 \text{ NeuAc} + \text{CTP} \rightarrow \text{CMP-NeuAc} + \text{PPi} \]

Polymerization of the K1 capsular polysaccharide has been shown to be catalysed by a membranous polysialyltransferase complex (polyST) (Troy, 1979). Determination of the transmembrane organization of this complex has shown that it is asymmetrically oriented at the cytoplasmic face of the inner membrane (Troy et al., 1993). The polyST complex catalyses the transfer of sialosyl residues from CMP-NeuAc to endogenous acceptors in the membrane or to exogenously added oligomers of polysialic acid, with elongation of the polymer occurring at the non-reducing end (Rohr and Troy, 1980; Troy, 1979). The nature of the endogenous acceptor has not yet been determined, however there have been suggestions that it may be either undecaprenyl phosphate (Steenbergen et al., 1992) or a protein (Rodriguez-Aparicio et al., 1988; Troy, 1979; Weisgerber and Troy, 1990). It has also been suggested that undecaprenyl phosphate (und-P) may function as a carrier of NeuAc during polymerization, and the following reactions were put forward to summarize the biosynthetic pathway (Sutherland, 1982; Troy, 1979; Troy et al., 1993):

\[ 4 \text{CMP-NeuAc} + \text{P-und} \rightarrow \text{NeuAc-P-und} + \text{CMP} \]

\[ 5 n(\text{NeuAc-P-und}) \rightarrow (\text{NeuAc})_n\text{-P-und} + n-1(\text{P-und}) \]

\[ 6 (\text{NeuAc})_n\text{-P-und} + \text{endogenous} \rightarrow (\text{NeuAc})_n\text{-acceptor} + \text{P-und} \text{ acceptor} \]

This biosynthetic mechanism would suggest the involvement of possibly up to three enzymes (Sutherland, 1982; Troy, 1979). However, preliminary evidence has recently been put forward indicating that chain elongation occurs by the sequential transfer of monomeric sialosyl units from CMP-NeuAc to the
growing polymer attached to an acceptor, although unfortunately the role of undecaprenyl phosphate in this process was not determined (Steenbergen and Vimr, 1990). These results are in conflict with the polymerization mechanism shown above which is proposed to involve oligosaccharide lipid-linked intermediates. To date, the role of undecaprenyl phosphate in K1 polymer synthesis has not been resolved.

Studies carried out by Vimr et al. (Steenbergen and Vimr, 1990; Vimr et al., 1989) have shown that a transposon insertion in the region 2 neuS gene results in undetectable endogenous or exogenous sialyltransferase activity in the host strain. In addition, expression of neuS in a strain lacking the kps gene cluster resulted in the presence of exogenous sialyltransferase activity. Thus, it was initially proposed that the 47.5kDa neuS gene product was the only sialyltransferase enzyme necessary for polymer synthesis (Steenbergen and Vimr, 1990). Further evidence for the role of NeuS came from the demonstration that the neuS gene from E. coli K92, a bacterium which synthesizes an α-2,8 α-2,9-linked polysialic acid polymer, could complement the K1 neuS insertion mutant, resulting in the production of a K92 capsule (Steenbergen et al., 1992).

It is important to note that if only one sialyltransferase was encoded by the kps gene cluster, it would have to be a multifunctional enzyme capable of catalysing both polysialic acid initiation and elongation reactions. However, it has recently been shown that the neuS gene product alone cannot initiate de novo polysialic acid synthesis (Steenbergen et al., 1992). Moreover, a quantitatively minor sialyltransferase activity has been demonstrated that could function to initiate polymer synthesis (Steenbergen et al., 1992). This activity may be due to the product of the region 2 neuE gene and it has been suggested that the NeuE protein (approximately 28kDa) initiates de novo synthesis by transferring one or more sialosyl residues to an endogenous acceptor, possibly undecaprenyl phosphate, which is then followed by chain elongation catalysed by NeuS (Steenbergen et al., 1992). Alternatively, the minor sialyltransferase activity could be encoded by another region 2 gene, neuD. Although the biological role of the 23kDa NeuD protein is
not known, it has been suggested that it may be a component of the sialyltransferase complex, since it does not appear to be required for NeuAc synthesis and bacteria harbouring mutations in neuD do not synthesize polymer (Silver et al., 1993). However, it has also recently been suggested that NeuD may be an acetyltransferase which exerts its influence through the modification of other region 2 proteins (Annunziato et al., 1995).

The polyST complex has been proposed to involve the participation of other gene products in addition to sialyltransferase(s), since mutations in many of the region 1 and region 3 kps genes resulted in reduced endogenous polymerase activity (Steenbergen et al., 1992; Vimr et al., 1989). The roles of many of the kps encoded proteins have not been fully elucidated and it is not always clear whether they are involved in translocation of the polymer to the cell surface or whether they act as structural and/or catalytic components of the polyST complex (Pavelka et al., 1991; Pazzani et al., 1993; Smith et al., 1990; Steenbergen et al., 1992; Vimr et al., 1989) (Section 1.4.4).

Thus, in summary, an rfc-dependent 0 antigen-like mechanism of polymerization, involving oligosaccharide lipid-linked intermediates, was originally proposed for K1 biosynthesis, albeit with polymerization occurring at the non-reducing terminus. However, it now appears that K1 biosynthesis may use an rfc-independent 0 antigen-like mechanism, involving initiation on an endogenous acceptor, possibly undecaprenyl phosphate, followed by the sequential addition of monomeric sialosyl residues to the non-reducing terminus.

1.4.3.2 The K5 Antigen

Unfortunately, less information is available with regards to the biosynthesis of the E. coli K5 antigen, a heteropolysaccharide with the primary structure 4)-β-glucuronosyl-(1,4)-α-N-acetylglucosaminyl-(1, (Vann et al., 1981). In agreement with the studies of K1 biosynthesis, molecular analysis of the K5 capsule gene cluster has shown that region 2
encodes the enzymes required for polymer synthesis (Roberts et al., 1988a). However, in contrast to the presence of the rare bacterial sugar, NeuAc, in the K1 antigen, the constituent sugars of the K5 polysaccharide are key components of other polysaccharides found in the cell: N-acetylglucosamine (GlcNAc) is present in ECA, peptidoglycan and some LPS molecules, whereas glucuronic acid (GlcA) is found in colanic acid. It must be noted that colanic acid is synthesized by many, but not all strains of E. coli. However, wild type E. coli K1 and K5 strains have recently been shown to be capable of synthesizing colanic acid under certain conditions (Keenleyside et al., 1993). Thus, one may expect the determinants for UDP-GlcNAc and UDP-GlcA synthesis to be encoded by loci distinct from the K5 kps cluster, although alternatively the genes may be duplicated, with one copy residing in region 2 of the capsule gene cluster and the other located elsewhere on the chromosome. In contrast, one would expect region 2 to encode the necessary sugar transferase enzymes, glucuronosyltransferase and N-acetylglucosaminytransferase, and possibly an additional enzyme to catalyse the initiation reaction. Unfortunately, in contrast to E. coli K1, a detailed genetic analysis of the K5 polymer synthesis genes has not yet been undertaken, although the biochemical pathway involved in the synthesis of the K5 polysaccharide has recently been investigated (Finke et al., 1991).

Membrane preparations from an E. coli K5 wild type strain and from a recombinant K-12 strain expressing the K5 capsule were used to study in vitro polymerization of the polysaccharide. Polymer synthesis was found to occur at the cytoplasmic face of the inner membrane and involved the sequential transfer of the two sugars from their UDP derivatives to the non-reducing end of the growing polysaccharide, without the participation of lipid-linked intermediates (Finke et al., 1991). Thus, the K5 polymer appears to be synthesized by a mechanism comparable to that used in rfc-independent O antigen synthesis. As mentioned earlier, the K5 polysaccharide is identical to the first polymeric intermediate in heparin biosynthesis (Section 1.1.3). Interestingly, it appears that the heparin precursor is also polymerized by a stepwise addition of GlcA and GlcNAc to the
non-reducing terminus of the nascent polymer, without lipid participation (Lidholt and Lindahl, 1992). A similar mechanism has also been reported to occur with the streptococcal polysaccharide, hyaluronic acid, a polymer which has the structure \( \beta \)-glucuronosyl-(1,3)-\( \beta \)-N-acetylglucosaminyl-(1, and thus differs from the K5 polysaccharide only in one of the sugar linkages (Stoolmiller and Dorfman, 1969). It would be interesting to determine whether the prokaryotic and eukaryotic genes involved in the synthesis of these structurally similar polysaccharides are closely related.

The reducing sugar of the K5 polysaccharide was found to be KDO (Finke et al., 1989, 1991). This sugar has also been found at the reducing terminus of a number of group II K antigens which have KDO in their repeating units (Schmidt and Jann, 1982). The activated form of the sugar is CMP-KDO and an elevated CMP-KDO synthetase activity was found in E. coli strains expressing group II capsules (Finke et al., 1990) (Section 1.3.1.1). The elevated activity of this enzyme in E. coli K5 was found to correlate with the temperature regulated expression of the K5 polysaccharide and this may indicate that CMP-KDO has an important role in polymer biosynthesis (Finke et al., 1989). Initially, it was thought that CMP-KDO may be involved in the reaction initiating biosynthesis of the K5 polymer and possibly other group II K antigens (Finke et al., 1989, 1990). Thus, the following mechanism was proposed: polymerization is initiated by substitution of a carrier, possibly undecaprenyl phosphate, with KDO from CMP-KDO and this then functions as the acceptor for subsequent chain elongation (Finke et al., 1991). However, recent studies, using recombinant E. coli K-12 strains carrying the K5 capsule genes, have shown that bacteria with mutations in certain region I genes produce intracellular polysaccharide which does not appear to have KDO at the reducing terminus (Bronner et al., 1993a) (Section 1.4.4.1). This suggests that KDO is not required for the initiation of polymerization and may in fact be added to the K5 polymer post-polymerization, possibly together with the phosphatidic acid moiety.

Thus, it can be seen that studies on the synthesis of the K5 polymer are not as advanced as with the K1 antigen. In order to
facilitate a better understanding of the processes involved, the genetic organization of the K5 region 2 needs to be elucidated and one of the aims of this project was the preliminary characterization of this region, in conjunction with C. Pazzani (Pazzani, 1992).

1.4.4 The Role Of Regions 1 And 3 In The Export Of Group II K Antigens

Regions 1 and 3 are of particular interest since these regions are common to different group II K antigen gene clusters and their products appear to be capable of interacting with a wide range of chemically and structurally diverse polysaccharides.

1.4.4.1 Region 1

A region of the kps locus which appeared to be responsible for the extracellular appearance of polymer was first identified by Echarti et al. (1983). This region was later referred to as region 1 (Timmis et al., 1985) (Figure 1.1). DNA hybridization studies and complementation analysis have demonstrated that the region is common to all group II capsule gene clusters so far analysed (Drake et al., 1990; Roberts et al., 1986, 1988a). Transposon and deletion mutagenesis of region 1 of the K1 capsule locus resulted in the presence of intracellular polysaccharide, which could be released from the bacterium by osmotic shock suggesting that it was primarily located in the periplasmic space (Boulnois et al., 1987). This polymer reacted with anti-K1 polyclonal and monoclonal antibodies and had an immunoelectrophoretic mobility indistinguishable from that of the mature cell surface antigen (Boulnois and Roberts, 1990; Boulnois et al., 1987). Also, since the material formed micelles, it was proposed that the polymer was attached to a phospholipid anchor molecule (Boulnois and Roberts, 1990). Thus, region 1 appears to encode the functions necessary for translocation of mature, lipid-linked, polysaccharide from the periplasmic space to the cell surface. Immunoelectron microscopy and biochemical analysis of a K5 region 1 deletion mutant supported this interpretation, since full length
polymer, substituted with phosphatidic acid, was found to be located in the periplasm following plasmolysis of the bacterium (Kroncke et al., 1990a). However, Vimr et al. (1989) have proposed that region 1 may also encode proteins which do not function directly in polymer translocation per se, but act as structural and/or catalytic components of the membrane bound glycosyltransferase complex. Furthermore, the gene responsible for the elevated CMP-KDO synthetase activity associated with E. coli strains expressing group II capsules was shown to be located in region 1 (Finke et al., 1989). Thus, proteins encoded by region 1 may function not only in polymer translocation, but also in the process of K antigen assembly.

Region 1, initially reported to be approximately 8kb (Boulnois and Jann, 1989) to 11.6kb (Silver et al., 1988) in size, was subjected to preliminary protein analysis using E. coli minicells (Boulnois and Roberts, 1990; Roberts et al., 1986; Silver et al., 1984, 1988). This demonstrated that a similar set of polypeptides was encoded by region 1 in the K1, K5 and K7 capsule gene clusters, although there were some discrepancies. However, the recent sequencing and analysis of region 1 of the K5 capsule locus, carried out by C. Pazzani in this laboratory, has helped to resolve these (Pazzani et al., 1993). The cosmid pGB110 encodes the functions required for biosynthesis and export of the K5 capsular polysaccharide (Roberts et al., 1986). Sequence analysis of a 7.7kb HindIII-EcoRV fragment of pGB110, shown by complementation analysis to encode region 1, revealed the presence of a minimum of five open reading frames (ORFs): 5' kpsE kpsD kpsU kpsC kpsS 3', possibly organized in a single transcriptional unit (Pazzani et al., 1993) (Figure 1.1). The ORFs which will now be described are defined by utilization of the first AUG start codon. However, to allow the accurate assignment of the correct translational start codon, the N-terminal amino acid sequence of the encoded protein must be determined. Such information has recently become available for KpsE (Rosenow et al., 1995a) and KpsU (Rosenow et al., 1995b). In addition, the N-terminal amino acid sequence of the mature form of the KpsD protein has been determined (Wunder et al., 1994). These findings are in agreement with the predictions made previously from the DNA sequence.
The first ORF, \textit{kpsE}, is 1146bp long and encodes a protein with a predicted molecular mass of 43.0kDa. Computer analysis, using the GCG (Devereux \textit{et al.}, 1984) sequence alignment program GAP, revealed 27.5\% identity and 49.3\% similarity between KpsE and the 42.3kDa BexC protein encoded by region 1 of the capsule locus (\textit{cap}) of \textit{H. influenzae} type b (Kroll \textit{et al.}, 1990) (Section 1.5.1; Figure 1.2). KpsE is also 26.3\% identical and 48.7\% similar to the 41.9kDa CtrB protein encoded by region C of the capsule locus (\textit{cps}) of \textit{N. meningitidis} group B (Frosch \textit{et al.}, 1991) (Section 1.5.2; Figure 1.2). CtrB also has 60.7\% identity and 76.7\% similarity with BexC and both proteins have been proposed to be involved in capsular polysaccharide export (Frosch \textit{et al.}, 1991; Kroll \textit{et al.}, 1990). The isolation of PhoA\textsuperscript{+} fusion proteins has indicated that KpsE, BexC and CtrB each have one or more periplasmically orientated domains and, in the case of KpsE and CtrB, the proteins have been shown to be associated with the inner membrane (Frosch \textit{et al.}, 1991; Kroll \textit{et al.}, 1990; Rosenow \textit{et al.}, 1995a). The hydropathy profiles (Kyte and Doolittle, 1982) of the KpsE, BexC and CtrB proteins are very similar and resemble that of an oligotopic membrane protein (Dalbey, 1990), with potential membrane-spanning domains located near both the N- and C-termini. Thus, it may be possible that these proteins are anchored in the inner membrane with their large hydrophilic central region orientated towards the periplasm. The topology of the KpsE protein within the inner membrane has recently been determined and appears to be in agreement with these predictions (Rosenow \textit{et al.}, 1995a). Since the KpsE, BexC and CtrB proteins appear to have some primary sequence homology and a similar secondary structure, it has been suggested that they may have a common function in capsule biogenesis, possibly being involved in polysaccharide export (Pazzani \textit{et al.}, 1993).

The recent analysis of a \textit{kpsE\textsuperscript{-}} mutant has begun to shed some light on the possible role of the KpsE protein. The plasmid pPC8 contains the whole K5 capsule gene cluster in which an oligonucleotide encoding an in-frame translational stop codon has been inserted into the \textit{kpsE} gene (Pazzani, 1992). Immunoelectron microscopy of \textit{E. coli} LE392 (pPC8) showed that
the K5 polysaccharide synthesized by this mutant was located in the periplasm (Bronner, unpublished results). This suggests that KpsE functions in the translocation of polysaccharide from the periplasmic face of the inner membrane to the outer surface of the cell, which would be consistent with its predicted location in the cell and would be in agreement with the proposed roles of BexC and CtrB. However, the possibility that the mutation in pPC8 has a polar effect on genes downstream of kpsE cannot be ruled out and consequently, the phenotype observed with this mutant must be interpreted with caution.

The second ORF, kpsD, is 1674bp long and encodes a protein with a predicted molecular mass of 60.4kDa. The hydropathy profile of KpsD is characteristic of a hydrophilic protein and analysis of the predicted amino acid sequence has identified the presence of a potential N-terminal signal peptide. Together, these observations suggest that KpsD may be a periplasmic protein. The plasmid pPC4 encodes the K5 capsule gene cluster in which there is a deletion of part of the kpsD gene (Pazzani, 1992). Immunoelectron microscopy of E. coli LE392 (pPC4) showed that the K5 polysaccharide synthesized by this mutant was located in the periplasm (Bronner, unpublished results). This would suggest that, like KpsE, KpsD also functions in the translocation of polysaccharide from the periplasm to the bacterial cell surface. However, since it is possible that the mutation may have a polar effect on downstream genes, the phenotype must be interpreted with caution. These results are in good agreement with the observation that a 60kDa protein is encoded by region 1 of the K1 capsule gene cluster (Silver et al., 1987; Wunder et al., 1994). This protein was shown to be synthesized as a larger precursor, with the processed form being located in the periplasm. Furthermore, in the absence of the 60kDa protein, polysaccharide was found to accumulate in the periplasmic space (Silver et al., 1987; Wunder et al., 1994).

The third ORF, kpsU, is 738bp long and encodes a protein with a predicted molecular mass of 27.1kDa. Computer analysis revealed 44.3% identity and 76.8% similarity between KpsU and the 27.5kDa CMP-KDO synthetase enzyme encoded by the kdsB gene of
*E. coli*, which is known to catalyse the activation of KDO to CMP-KDO during LPS biosynthesis (Goldman et al., 1986). In addition, the KpsU protein has recently been purified and shown to have CMP-KDO synthetase activity (Rosenow et al., 1995b). Thus, the elevated level of CMP-KDO synthetase activity associated with *E. coli* strains expressing group II capsules would appear to be due to the presence of an additional CMP-KDO synthetase enzyme encoded by *kpsU*. Why a separate CMP-KDO synthetase enzyme is required for group II capsule biosynthesis is not yet understood. However, it may reflect important biochemical differences between the two enzymes which are specific to their individual roles. Another possibility is that it may allow independent regulation of LPS and capsule biosynthesis, since KdsB may catalyse the rate-limiting step in KDO incorporation into LPS (Goldman and Kohlbrenner, 1985) and KpsU appears to be regulated by temperature (Finke et al., 1989).

The fourth ORF, *kpsC*, is 2025bp long and encodes a protein with a predicted molecular mass of 75.7kDa. Computer analysis has revealed 46.6% identity between KpsC and the ORF5 protein encoded by region 3 of the *cap* locus of *H. influenzae* type b (Brophy et al., personal communication) (Section 1.5.1; Figure 1.2). KpsC is also 51.9% identical and 67.1% similar to the LipA protein encoded by region B of the *cps* locus of *N. meningitidis* group B (Frosch and Müller, 1993) (Section 1.5.2; Figure 1.2). The fifth ORF, *kpsS*, is 1167bp long and encodes a protein with a predicted molecular mass of 46.3kDa. Comparison of the primary structure of KpsS and the ORF6 protein, also encoded by region 3 of the capsule locus of *H. influenzae* type b, revealed 40.6% identity (Brophy et al., personal communication). KpsS is also 39.3% identical and 62.5% similar to the LipB protein encoded by region B of the meningococcal group B *cps* locus (Frosch and Müller, 1993). As expected, LipA and LipB also have homology with the ORF5 and ORF6 proteins of *H. influenzae*, respectively (Brophy et al., personal communication) (Table 1.4).

The plasmid pPC7 and the cosmids pGB118::1 and pGB110::28 encode the K5 capsule locus in which there is a deletion of
part of the kpsC gene, a Tn1000 insertion in kpsS and a Tn5phoA
insertion which maps either at the end of kpsC or at the
beginning of kpsS, respectively (Bronner et al., 1993a; Pazzani,
1992). Immunoelectron microscopy of E. coli LE392
harbouring pFC7, pGB118::1 or pGB110::28 revealed the presence
of intracellular polysaccharide, associated with areas of the
cytoplasm which appear to have low electron density (Bronner
et al., 1993a; Bronner, unpublished results). Interestingly,
purification and biochemical analysis of the polysaccharide
synthesized by LE392 (pGB118::1) has shown that there is no KDO
or phosphatidic acid present at the reducing end of the polymer
in this mutant (Bronner et al., 1993a). These findings are in
good agreement with the recent proposal that the meningococcal
proteins, LipA and LipB, direct the substitution of the group B
capsular polysaccharide with phospholipid, a process which may
be essential for the subsequent or concomitant translocation of
the polymer across the inner membrane (Frosch and Müller,
1993). Since the KpsC and KpsS proteins of E. coli and the ORF5
and ORF6 proteins of H. influenzae appear to have homology with
LipA and LipB, it is possible that they have a similar role in
capsule biogenesis. The absence of KDO at the reducing terminus
of the polymer synthesized by LE392 (pGB118::1) suggests that
KDO is not required for the initiation of K5 polysaccharide
biosynthesis, an observation which is in disagreement with the
hypothesis put forward by Finke et al. (1989) (Section
1.4.3.2). However, since KDO appears to be added post-
polymerization, possibly together with the phosphatidic acid
moiety, it may function in a later stage of capsule biogenesis,
for example polysaccharide export.

In addition to the aforementioned five ORFs encoded by region
1, a further ORF has been identified upstream of the kpsE gene.
This ORF appears to start before the available sequence and
encodes a protein with a minimum molecular mass of 25.6kDa.
However, since E. coli LE392 harbouring the K5 capsule gene
cluster with a deletion located in this ORF synthesize an
apparently wild type capsule, it would appear that this protein
is not essential for capsule biogenesis (Bronner, unpublished
results; Pazzani, 1992). A number of smaller ORFs were
identified on the opposite strand which would encode proteins
with predicted molecular masses of less than 12kDa, but analysis of the proteins encoded by region 1 did not provide any evidence to suggest that these ORFs are expressed. However, analysis using either minicells or an *in vitro* transcription-translation system did identify proteins with molecular masses of 43kDa, 60kDa, 27kDa, 75kDa and 44kDa, which correspond to the predicted products of the *kpsE, kpsD, kpsU, kpsC* and *kpsS* genes, respectively (Pazzani *et al.*, 1993).

In summary, region 1 appears to encode five proteins: KpsE and KpsD, which may function in the translocation of capsular polysaccharide from the periplasm to the cell surface, KpsU which has CMP-KDO synthetase activity and finally KpsC and KpsS, which may have roles in directing the substitution of the polysaccharide with phosphatidic acid. Interestingly KpsE, KpsD, KpsC and KpsS have homology, although fairly limited, with a number of proteins associated with reactions involving phosphatidic acid or sn-glycerol-3-phosphate, which is a component of phosphatidic acid (Pazzani, 1992). This is in agreement with the proposed role of KpsC and KpsS and may also imply that proteins involved in polysaccharide export recognize the phosphatidic acid present at the reducing end of group II polymers as a common component, thus facilitating the translocation of capsular polysaccharides regardless of their chemical structure.

### 1.4.4.2 Region 3

Region 3 (Figure 1.1), initially reported to be approximately 2.5kb in size, was shown by DNA hybridization and complementation studies to be common to different *E. coli* group II K antigen gene clusters (Roberts *et al.*, 1988a). Mutations in this region resulted in the presence of intracellular immunoreactive polysaccharide, which was released from the bacterium by cell lysis, but not by osmotic shock (Boulnois *et al.*, 1987). This suggests that the polymer was located in the cytoplasm. The polysaccharide isolated from region 3 mutants of the K1 capsule locus was shown by immunoelectrophoretic analysis to have a low electrophoretic mobility compared with the mature K1 polysaccharide (Boulnois *et al.*, 1987). It was
suggested that the material consisted of polysaccharide linked to either undecaprenyl phosphate or endogenous acceptor and that region 3 functioned in a post-polymerization modification of the polymer, such as the attachment of phosphatidic acid (Boulnois et al., 1987). Unfortunately, the polysaccharide synthesized by these mutants has never been chemically characterized. The intracellular polysaccharide synthesized by bacteria harbouring the K5 capsule gene cluster with a transposon insertion in region 3, was shown by immunoelectron microscopy to be located in the cytoplasm (Kroncke et al., 1990a). This material was found to be devoid of phospholipid and had the characteristic greater immunoelectrophoretic mobility which is associated with unsubstituted polysaccharides (Kroncke et al., 1990a). This was in contrast to the situation found previously with equivalent mutations in the K1 capsule genes. However, differences in the procedures used to extract the polysaccharide may account for the apparent variation in the phenotypes of the K1 and K5 region 3 mutants. The polymer isolated from the cytoplasm of a K5 region 3 mutant was originally reported to be shorter than the wild type K5 polysaccharide (Kroncke et al., 1990a), but recent analysis has shown that the material is in fact full length and lacks KDO in addition to phosphatidic acid (Bronner, Sieberth and Jann, personal communication).

Interestingly, the addition of the membrane energy uncoupling agent, carbonyl cyanide m-chlorophenylhydrazone (CCCP), to wild type E. coli K5, prior to shifting from a capsule restrictive (18°C) to permissive (37°C) temperature, resulted in the same phenotype as seen with a K5 region 3 mutant (Jann, personal communication; Kroncke et al., 1990b). These results may suggest that region 3 products are involved in the energy-dependent translocation of capsular polysaccharide across the inner membrane. The attachment of phosphatidic acid in a (trans)lipidation reaction may be associated with this process in a vectorial manner (Boulnois and Jann, 1989).

One of the aims of this project was to continue the molecular analysis of region 3 of the E. coli K5 antigen gene cluster, thus facilitating the preliminary characterization of the
1.4.5 The Role Of Membrane Adhesion Sites And Porins In The Export Of Group II K Antigens

The term membrane adhesion site (Bayer junction) has been used to refer to sites where the inner and outer membrane of Gram-negative bacteria come into close apposition (Bayer, 1979). These sites have been implicated in the export of newly synthesized cell surface components, such as LPS and the group I capsular polysaccharide of *E. coli* K29, (Bayer, 1979, 1990). Furthermore, recent evidence suggests that they may also have a role in the surface expression of *E. coli* group II K antigens (Kroncke *et al.*, 1990b). When wild type *E. coli* K1, K5 and K12 were shifted from a capsule restrictive to a capsule permissive temperature, the newly exported surface polymer was shown by electron microscopy to be initially present in the form of tufts, which appeared to be located above membrane adhesion sites. These patches of polysaccharide then rapidly dispersed to form a continuous capsule. Thus, the *kps* encoded proteins, proposed to be involved in the translocation of capsular polysaccharide to the cell surface, may function in association with these membrane adhesion sites. However, at the present time it is not clear whether such sites actually do have a role in group II capsule expression.

A 40kDa protein, termed protein K, commonly found in encapsulated *E. coli* isolates, but rarely in unencapsulated strains, was also suggested to be involved in capsule biogenesis (Paakkanen *et al.*, 1979). This protein, which has been shown to function as a porin (Sutcliffe *et al.*, 1983; Whitfield *et al.*, 1983), is absent in recombinant *E. coli* K-12 strains expressing the K1 capsular polysaccharide (Timmis *et al.*, 1985) and thus protein K cannot be essential for capsule expression. However, a structurally and functionally related *E. coli* K-12 porin may be able to substitute for protein K in this situation. Whether porin proteins play any role in capsule development has yet to be determined, but the calculated pore size of *E. coli* porins would tend to argue against a direct role in the translocation of capsular polysaccharide through
the outer membrane (Whitfield et al., 1983). In fact, the apparent correlation of protein K expression and encapsulation in E. coli may simply be due to a common response to selective pressure (Whitfield et al., 1983).

1.5 OTHER BACTERIAL SPECIES WHICH EXPRESS GROUP II CAPSULAR POLYSACCHARIDES

The capsular polysaccharides of H. influenzae and N. meningitidis have a number of properties in common with the group II K antigens of E. coli (Section 1.3.1.3). Furthermore, molecular analysis of the cap locus of H. influenzae and the cps locus of N. meningitidis, which encode the functions required for capsule synthesis, has shown that they have a similar genetic organization to the kps locus of E. coli (Figure 1.2) and this will be discussed below.

1.5.1 The Haemophilus influenzae cap Locus

H. influenzae is a common cause of meningitis and other life-threatening invasive infections in young children. Six capsular serotypes have been described (a-f), with type b strains being responsible for more than 95% of the cases of serious infection (Turk, 1982). The type b capsular polysaccharide, a polymer of ribose and ribitol-5-phosphate, has been shown to be an important determinant of virulence (Zwahlen et al., 1989). Analysis of the capsular polysaccharide biosynthesis genes (cap) of serotypes a, b, c and d has revealed a conserved genetic organization consisting of three functional regions (Figure 1.2): a central serotype-specific cassette (region 2) flanked by DNA common to all serotypes (regions 1 and 3). As in the E. coli kps locus, region 2 encodes the enzymes for the synthesis and polymerization of the specific polysaccharide, whereas regions 1 and 3 are proposed to be involved in non-serotype-specific functions such as polymer export (Kroll et al., 1989).

The population of encapsulated H. influenzae has been divided into two phylogenetic divisions (Musser et al., 1988a). The capsule gene cluster of division I strains has been shown to
Figure 1.2 The capsule gene clusters of *E. coli*, *H. influenzae* and *N. meningitidis*

**N. coli K5**

<table>
<thead>
<tr>
<th>export</th>
<th>biosynthesis</th>
<th>export</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

**H. influenzae type b**

<table>
<thead>
<tr>
<th>export</th>
<th>biosynthesis</th>
<th>export</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

box A B C D

orf 5 6

---

**N. meningitidis group B**

<table>
<thead>
<tr>
<th>regulation</th>
<th>export</th>
<th>biosynthesis</th>
<th>LOS synthesis</th>
<th>PI-substitution</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>C</td>
<td>A</td>
<td>D</td>
<td>B</td>
</tr>
</tbody>
</table>

ctr D C B A

lip A B

2kb

---

Figure 1.2 The genetic organization of the *E. coli K5* kps locus, the *H. influenzae* type b cap locus and the *N. meningitidis* group B cps locus is shown, drawn approximately to scale. The different functional regions of each locus are represented by the boxes labelled 1 to 3 and A to E, with the proposed function of each region written above. LOS refers to lipooligosaccharide and PI refers to phospholipid. The smaller boxes labelled hex A B C D and orf 5 6 correspond to the genes identified in regions 1 and 3 of the *H. influenzae* cap locus, respectively. Similarly, the boxes labelled ctr D C B A and lip A B represent the genes identified in regions C and B of the meningococcal cps locus, respectively. The size of the predicted gene products is indicated in kDa, except in the case of lip A B. There has been some controversy regarding the size of these proteins. The arrows indicate the direction of transcription.
lie between direct repeats of an insertion element, IS1016, and this is proposed to facilitate the reversible amplification of the cap locus through unequal homologous recombination (Kroll et al., 1991). However, a direct repeat of cap, with a 1.2kb deletion at one end of the duplicated locus, has become fixed in the population of division I type b strains (Hosieth et al., 1986; Kroll et al., 1991). The deletion is located at the 3' end of region 1 and removes part of a gene, bexA, which is critical for polysaccharide export (Kroll et al., 1988). The capsulate phenotype of these type b strains is dependent on preservation of the direct repeat structure in order to avoid recombination-mediated loss of the other copy of bexA (Kroll et al., 1991). Thus, it appears that the cap locus in division I strains of H. influenzae has the structure of a compound transposon. At the present time there is no evidence to suggest that similar IS-like elements and duplications exist at the capsulation loci of E. coli and N. meningitidis, although an uncharacterized repetitive element may be present in the meningococcal cps locus (Frosch et al., 1989). The nucleotide sequence of regions 1 and 3 of the capb locus has been determined (Brophy et al., unpublished results; Kroll et al., 1990). Region 1 contains a cluster of four genes, bexDCBA, possibly organized in a single transcriptional unit and region 3 encodes two proteins ORF5 and ORF6 (Figure 1.2). These will be considered below, together with the corresponding meningococcal proteins.

1.5.2 The Neisseria meningitidis cps Locus

N. meningitidis is a major causative agent of bacterial septicaemia and meningitis. The capsular polysaccharide is an important virulence determinant and up to 12 capsular serogroups have been described, with groups A, B, C, Y and W135 being the most clinically significant (Jennings, 1990). Analysis of the capsule gene clusters (cps) of N. meningitidis serogroups B and C has revealed a genetic organization similar to that found in E. coli and H. influenzae, although here the complex is divided into five functional regions, A-E (Frosch et al., 1989, 1991) (Figure 1.2). As in the kps locus of E. coli and the cap locus of H. influenzae, the central region, region
A, encodes the enzymes necessary for synthesis of the serotype-specific polysaccharide and two regions, B and C, are proposed to encode non-serotype-specific functions involved in the transport of the polymer to the cell surface. However, the two remaining regions, D and E, do not appear to have functional equivalents in the capsule gene clusters of *E. coli* and *H. influenzae*. Region D has recently been found to contain genes which are involved in lipooligosaccharide (LOS) synthesis, although some of these genes do not appear to be functional (Hammerschmidt et al., 1994; Jennings et al., 1993). However, mutations in the *galE* gene located in this region have been shown to result in an altered LOS phenotype. Finally, region E is believed to play a regulatory role in capsule expression (Frosch et al., 1989).

Region C of the meningococcal group B *cps* locus has been sequenced and shown to consist of four genes, *ctrABCD*, possibly organized in a single transcriptional unit (Frosch et al., 1991) (Figure 1.2). Comparison of the proteins encoded by the meningococcal region C and region 1 of the *H. influenzae capB* locus has shown that CtrA, CtrB, CtrC and CtrD have extensive homology with BexD, BexC, BexB and BexA, respectively (Frosch et al., 1991) (Table 1.4). In addition, some of these proteins also show homology with gene products of the *E. coli kps* locus. BexA and BexB, and likewise CtrD and CtrC, are proposed to direct polysaccharide export across the inner membrane and may be functionally equivalent to proteins encoded by region 3 of the *kps* locus (Frosch et al., 1991; Kroll et al., 1990; Pavelka et al., 1991; Smith et al., 1990). This will be discussed at greater length in Chapter 3. As mentioned previously (Section 1.4.4.1), BexC and CtrB have homology with the *kps* region 1 encoded KpsE protein, which is thought to play a role in the translocation of capsular polysaccharide from the periplasm to the cell surface (Pazzani et al., 1993). CtrA was shown to be located in the outer membrane and is thus proposed to be involved in polysaccharide export across the outer membrane (Frosch et al., 1991). Unfortunately, the function of BexD has not yet been defined (Kroll et al., 1990). Interestingly, the *E. coli kps* locus does not appear to encode a protein with homology to BexD and CtrA. However, it has been suggested that
Table 1.4 Sequence homology between proteins encoded by the *cps* locus of *N. meningitidis* and the *cap* locus of *H. influenzae*

<table>
<thead>
<tr>
<th>N. meningitidis</th>
<th>H. influenzae</th>
<th>% identity</th>
<th>% similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>CtrA</td>
<td>BexD</td>
<td>52.6</td>
<td>71.0</td>
</tr>
<tr>
<td>CtrB</td>
<td>BexC</td>
<td>60.7</td>
<td>76.7</td>
</tr>
<tr>
<td>CtrC</td>
<td>BexB</td>
<td>69.1</td>
<td>84.9</td>
</tr>
<tr>
<td>CtrD</td>
<td>BexA</td>
<td>81.0</td>
<td>89.4</td>
</tr>
<tr>
<td>LipA</td>
<td>ORF5</td>
<td>60.5</td>
<td>$$</td>
</tr>
<tr>
<td>LipB</td>
<td>ORF6</td>
<td>56.3</td>
<td>$$</td>
</tr>
</tbody>
</table>

The amino acid sequences of the proteins encoded by the non-serotype specific regions of the *cps* and *cap* loci were subjected to pairwise comparisons using the GCG program GAP.

$\$ This information was unavailable.
CtrA has a secondary structure similar to that associated with porin proteins and thus perhaps protein K (Section 1.4.5) fulfils an equivalent function in capsule biogenesis in E. coli (Frosch et al., 1992; Paakkanen et al., 1979).

Recently, the nucleotide sequence of region B of the meningococcal group B cps locus has been determined and this region has been shown to encode two proteins LipA and LipB (Frosch and Müller, 1993) (Figure 1.2). As discussed in Section 1.4.4.1, LipA and LipB have homology with the KpsC and KpsS proteins encoded by region 1 of the E. coli kps locus and the ORF5 and ORF6 proteins encoded by region 3 of the H. influenzae capB locus (Table 1.4), respectively. LipA and LipB have been proposed to direct the substitution of the group B capsular polysaccharide with phosphatidic acid (Frosch and Müller, 1993) and the related E. coli and H. influenzae proteins may have a similar role in capsule biosynthesis.

Thus, it can be seen that the capsule gene clusters of E. coli, H. influenzae and N. meningitidis have a remarkably similar genetic organization. In addition, the regions involved in non-serotype-specific functions, such as polymer export, appear to encode a number of homologous proteins, although it is also apparent that some of the proteins do not have functional equivalents in each of the loci and the relative positions of the genes encoding these proteins is not always the same. Nevertheless, as proposed by Frosch et al. (1991), the homology which exists between the capsule genes of these three unrelated Gram-negative bacterial species does perhaps suggest a common evolutionary origin of their molecular mechanisms of encapsulation and this will be discussed in more detail in Chapter 6.

1.6 THE GENERATION OF GROUP II K ANTIGEN DIVERSITY

1.6.1 The Clonal Population Structure Of E. coli And Genetic Diversity

In recent years, the genetic characteristics of bacterial populations have been studied by measuring the amount of allelic diversity at a number of different loci by multilocus
enzyme electrophoresis (MLEE) (Selander et al., 1987). In this technique, electromorphs (allozymes) of metabolic enzymes are equated with alleles at the corresponding structural gene loci. Since nearly all amino acid sequence variation can be detected using electrophoresis, MLEE is a useful method for studying allele frequencies and genotypes in bacterial populations. Isolates can be characterized by their combinations of alleles at the different chromosomal loci and assigned a distinctive electrophoretic type (ET).

Individual loci in *E. coli* appear to be highly polymorphic (Selander et al., 1987). Thus, if all alleles assorted randomly and independently of each other, one would expect a vast array of ETs to be present in the *E. coli* population. However, MLEE has demonstrated that natural populations of *E. coli* are basically clonal in structure and comprise a relatively limited number (in the order of $10^2$-$10^3$) of independently evolving lineages which are temporally and spatially stable (Selander et al., 1987). A principal component analysis of the allele profiles of the 302 ETs identified in one study of natural populations of *E. coli* (including *Shigella* spp.) showed that they fell into three overlapping clusters (Selander et al., 1987). This observation of a strong and complex non-random association of alleles (linkage disequilibrium) implies a low rate of recombination within the *E. coli* population, thereby allowing linkage to be maintained throughout the chromosome. A similar clonal population structure appears to exist in other bacterial species, including the naturally transformable species *H. influenzae* (Musser et al., 1988b), where recombination might be expected to occur more frequently. *N. meningitidis*, another naturally transformable species, was also thought to have a clonal population structure (Caugant et al., 1987), but recently it has been shown that this may not be strictly true (Maynard Smith et al., 1993). Nevertheless, intraspecific recombination events do occur in bacteria, superimposed on the stable chromosomal background. Recent studies have shown that recombination sometimes takes place within genes; for example, comparison of the nucleotide sequence of nine alleles of the *E. coli* phoA gene indicated that intragenic recombination had occurred amongst these
alleles, giving rise to a mosaic gene structure (Dubose et al., 1988). Larger recombinational events are also thought to occur, whereby a favourable allele together with neutral adjacent hitch-hiking loci replace preexisting alleles (Krawiec and Riley, 1990). Consequently, Milkman and Bridges (1990) have referred to the E. coli chromosome as a clonal frame within which individual segments can have independent phylogenetic histories. It is important to note that MLEE only examines genotype at a limited number of loci and since recombinational events occur at a low rate and generally involve short segments of DNA, they probably only rarely alter the ET of the individual, thus maintaining a clonal population structure as detected by MLEE (Maynard Smith et al., 1991).

Genetic diversity in E. coli is also generated by spontaneous mutations occurring within the chromosome. These also occur at a low rate and range from point mutations to larger rearrangements such as duplications, inversions and transpositions (Krawiec and Riley, 1990). In addition, interspecific recombination may provide a further source of genetic variation. However, a balance must be maintained between the benefits of acquiring new genetic information and the need to preserve the integrity of the species and thus such events are probably extremely rare (Krawiec and Riley, 1990). There is a growing understanding of the mechanisms which help to limit genetic variation in the bacterial chromosome and some of these are described below. The restriction-modification system of the host may result in cleavage of incoming foreign DNA. Mismatch repair systems which normally function to maintain fidelity during DNA replication, may also play a role in preventing intrachromosomal, intraspecific and interspecific recombination between imperfectly matched DNA sequences (Rayssiguier et al., 1989). Some genetic rearrangements may disrupt a critical gene order in the chromosome and thus have a deleterious effect on the cell (Krawiec and Riley, 1990). In addition, genetic variation within the bacterial population as a whole may be reduced by the random extinction of cell lines, occurring either stochastically or through the "periodic selection" of lineages with greater fitness (Selander et al., 1987). Thus, although natural populations of E. coli are clonal in
structure, there is still scope for the generation of genetic diversity, whilst at the same time maintaining the basic stability of the clonal frame.

MLEE detects electromorphic variants of metabolic enzymes and the allelic variation observed with these and other housekeeping functions is generally deemed to be selectively neutral or nearly so (Selander et al., 1987). However, with certain genes, such as those associated with the synthesis of cell surface antigens, genetic variation may be subject to stronger selection pressures, since the generation of a new or modified phenotype may confer a survival advantage on the bacterium. In such cases, the variant genotypes may become fixed in the population more rapidly than allelic variants of housekeeping genes. This may help to account for the great genetic diversity associated with cell surface structures such as the K antigens of \textit{E. coli}.

1.6.2 Evolution Of The \textit{kpS} Locus

Group II capsular polysaccharides are associated with many \textit{E. coli} O serotypes (Section 1.3.1.1). Thus, it has been suggested that an intact or partial \textit{kpS} gene cluster was acquired relatively early in the evolutionary history of \textit{E. coli}, since the clonal nature of the \textit{E. coli} population would tend to argue against a more recent acquisition of capsule genes and their subsequent dissemination amongst different O serotypes (Boulnois and Jann, 1989). The origin of the \textit{kpS} cluster is unknown, but it may have been acquired by horizontal gene transfer from an unrelated species. As discussed in Section 1.5, the \textit{kpS} locus of \textit{E. coli} and the capsule gene clusters of \textit{H. influenzae} and \textit{N. meningitidis} have been shown to have a similar genetic organization and this possibly indicates that the genes have a common evolutionary origin. Interestingly, the capsule locus present in division I strains of encapsulated \textit{H. influenzae} appears to have the structure of a compound transposon (Kroll et al., 1991). This could suggest that if the capsule loci of these three distantly related bacterial species have originated from a common ancestor, it may have been via a transposition-like event. Unfortunately, it is not known
whether the capsule gene clusters of *E. coli* and *N. meningitidis* are also flanked by insertion sequences. Transposable elements are in fact ideal candidates for facilitating horizontal gene transfer, since once inside the cell they enable the foreign DNA to integrate into the host chromosome. Moreover, the streptococcal conjugative transposon Tn916 has recently been shown to transfer naturally between a variety of phylogenetically distant bacterial species (Bertram *et al.*, 1991). However, it remains to be seen whether similar elements are common in Gram-negative bacteria.

There is a vast structural diversity of group II capsular polysaccharides and how this has arisen poses an interesting problem. However, despite their diversity, group II K antigens can be subdivided into families, since many contain identical sugars, but have different linkages and modifications (Jann and Jann, 1990). As one might expect, the kps loci which direct the synthesis of the two closely related sialic acid-containing polymers, K1 (α-2,8 linkages) and K92 (alternating α-2,8 α-2,9 linkages), appear to have very similar region 2 cassettes (Roberts *et al.*, 1986). In fact, recent analysis indicates that the different linkages found in these two polymers result solely from a relatively small number of amino acid differences in the respective polysialyltransferase enzymes (Vimr *et al.*, 1992). In contrast, in the case of bacteria synthesizing capsular polysaccharides which have very different structures, the region 2 cassettes appear to have little, if any, DNA sequence homology (Roberts *et al.*, 1986, 1988a). Thus, it may be possible to loosely group the region 2 cassettes into families in the same manner as related polysaccharides (Boulnois and Jann, 1989). These observations have led Boulnois and Jann (1989) to put forward the following model to account for the generation of group II K antigen diversity.

An *E. coli* isolate expressing a group II capsule may acquire new region 2-like determinants as a result of, for example, duplication of LPS biosynthesis genes or genetic exchange with other bacteria. If the incoming region 2-like segment is very different from the resident region 2 cassette then the whole region may be replaced en bloc. This may involve recombination
between the conserved flanking regions, 1 and 3, or alternatively specific sequences may be present at the junctions between regions 1 and 2 and regions 2 and 3 which facilitate a site-specific recombinational event. Such an exchange may serve to fix a specific capsule type into a different clonal lineage and in fact, horizontal transfer of region 2 genes between strains with highly disparate genetic backgrounds has been reported in *H. influenzae* (Kroll and Moxon, 1990). In contrast, if the resident and newly acquired region 2-like sequences are similar yet distinct, recombination may occur within region 2, resulting in the generation of a hybrid cassette which directs the synthesis of a related, but novel polymer. This may be the mechanism by which families of related region 2 cassettes have arisen. Furthermore, region 2 cassettes may also be modified by point mutations, duplications, deletions and transposition-like events. Thus, there are a number of possible mechanisms which may account for the generation of capsule diversity. This subject will be discussed further in Chapter 5, in the light of information generated by the recent genetic analysis of several different group II K antigen gene clusters.

1.7 TEMPERATURE REGULATION OF *E. coli* GROUP II CAPSULE BIOSYNTHESIS

*E. coli* strains synthesizing group II K antigens were found to express capsule in a temperature dependent manner (Bortolussi et al., 1983; Orskov et al., 1984) (Section 1.3.1.1). Thus, capsular material can be detected after growth at 37°C, but not after incubation at 18°C. The regulation of an important virulence determinant by an environmental factor such as temperature is not unusual. Bacteria continually sense and respond to factors such as temperature, osmolarity, pH, oxygen, CO₂, iron concentration and nutrient supply, and modify their gene expression accordingly (Mekalanos, 1992). Such environmental parameters can signal the entry of a bacterial pathogen into host tissues, thus enabling the bacterium to mount an appropriate response to the numerous stresses encountered upon transition from a free-living to a host-associated state. At subsequent stages in the infection process further adaptations may be made by the pathogen in response to different
microenvironments within the host. In human pathogens, as one might expect, the increase in growth temperature from <37°C to 37°C upon infection has often been found to induce the expression of virulence determinants; for example, virulence gene expression in *Shigella flexneri* (Maurelli, 1989) and expression of P fimbriae in *E. coli* (Göransson et al., 1990) are subject to thermoregulation. Thus, it is perhaps not surprising that the expression of *E. coli* group II capsular polysaccharides is regulated by temperature, since these K antigens are commonly associated with extraintestinal disease isolates (Orskov et al., 1984).

The molecular basis of signal transduction, that is how the bacterium senses environmental change and subsequently regulates gene expression, is beginning to be understood. In recent years it has become evident that bacteria possess global regulatory networks, wherein a given environmental signal results in the co-ordinated induction or repression of a number of unlinked, yet functionally related, genes and operons (Gottesman, 1984). This may involve the action of a two-component sensor-kinase/response-regulator signal transduction system (Gross et al., 1989). After receiving an environmental stimulus, the sensor-kinase protein phosphorylates the response-regulator protein, which then becomes an active regulator altering the level of transcription of specific target genes. It must be noted that there are variations on this theme, where the system consists of a single polypeptide or alternatively, three or more components are involved (Gross et al., 1989). Interestingly, environmentally induced changes in the level of DNA supercoiling in the cell, which are thought to be mediated through the action of topoisomerases and histone-like proteins such as the H1 (H-NS) protein, also appear to play an important role in the regulation of gene expression (Dorman, 1995; Higgins et al., 1990a). It has been suggested that alterations in DNA topology may affect the accessibility of promoters to the transcriptional apparatus, including the response-regulator components of two-component signal transduction systems, thereby modulating the level of transcription (Dorman, 1991; Kroll, 1991). Thus, there appears to be a regulatory hierarchy wherein changes in DNA super-
coiling provide a crude underlying global regulatory network upon which more specific regulatory processes are superimposed (Dorman, 1995; Ni Bhriain et al., 1989). The moment to moment co-ordination of gene expression represents the interplay of these overlapping factors (Kroll, 1991).

At the present time, the means by which *E. coli* group II capsule biosynthesis is regulated by temperature is not fully understood. It might be expected that the aforementioned regulatory processes, which act at the level of transcription, are involved. Indeed there is preliminary evidence suggesting that transcription of the genes present in regions 1 and 3 of the capsule gene cluster is subject to thermoregulation (Roberts, personal communication; Silver, personal communication). However, it is not known whether this involves changes in DNA topology and/or the action of the response-regulator component of a two-component signal transduction system or another type of transcriptional regulator. Unfortunately, the situation is far from simple, since it has also been proposed that the temperature dependency of K1 antigen biosynthesis is mediated by the reversible cold-inactivation of the enzymes involved in the synthesis and polymerization of sialic acid, particularly NeuAc synthase (Merker and Troy, 1990; Troy et al., 1993).

One of the aims of this project was to establish the detailed genetic organization of the K5 antigen gene cluster, in conjunction with the work of C. Pazzani (Pazzani, 1992). Such information may be useful in helping to determine precisely how capsule biosynthesis is regulated by temperature and possibly by other environmental factors.
1.8 THE AIMS OF THIS PROJECT

The mechanisms by which bacterial capsular polysaccharides are synthesized, transported across the cell wall and organized at the cell surface are only now beginning to be characterized in detail. Such knowledge may be useful in the commercial exploitation of bacterial exopolysaccharides and may allow the development of novel therapeutic interventions for use in the treatment of diseases caused by encapsulated bacteria. The K antigens of *E. coli*, in particular those which are classified as group II K antigens, have been studied in some detail and the aim of this project was to further investigate the processes involved in their biosynthesis, by carrying out a detailed molecular analysis of the *E. coli* K5 antigen gene cluster, in conjunction with C. Pazzani (Pazzani, 1992).
CHAPTER 2

MATERIALS AND METHODS

2.1 CHEMICALS, MEDIA AND ENZYMES

Chemicals were routinely purchased from BDH/Merck Ltd. and Fisons Scientific Equipment. Media were purchased from Difco Laboratories and BBL/Becton Dickinson UK Ltd. Antibiotics were purchased from Sigma Chemical Company Ltd. Restriction endonucleases and DNA modifying enzymes were purchased from GIBCO-BRL Ltd. and Pharmacia LKB Biotechnology Ltd.

2.2 BACTERIAL STRAINS AND PLASMIDS

The bacterial strains and plasmids used in this study are listed in Tables 2.1 and 2.2.

2.2.1 Growth Conditions And Media

Bacteria were grown in Luria broth (per litre: 10g peptone/tryptone, 5g yeast extract, 5g NaCL) at 37°C unless otherwise stated. When necessary, media was solidified by the addition of agar to a final concentration of 1.5%. B-agar (per litre: 10g peptone/tryptone, 8g NaCl, 15g agar), B-top agar (per 100ml: 1g peptone/tryptone, 0.8g NaCl, 0.6g agar) and brain heart infusion broth (OXOID, 37g per litre) were used as stated.

Antibiotics were added to the media as required at the following final concentrations: Ampicillin 100µg/ml; Chloramphenicol 25µg/ml; Kanamycin 25µg/ml; Tetracycline 20µg/ml.
### Table 2.1

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>LE392</td>
<td>F' (rK', mK') supE44 supF88 hsdR514 galK2 galT22 metH1 trpR55 lacY1</td>
<td>Sambrook et al. (1989)</td>
</tr>
<tr>
<td>JM101</td>
<td>supE thi Δ(lac-proAB) F' (traD36 proA8 lacZAM15)</td>
<td>Sambrook et al. (1989)</td>
</tr>
<tr>
<td>DS410</td>
<td>minA minB</td>
<td>Wilkins et al. (1981)</td>
</tr>
</tbody>
</table>

### Table 2.2

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Characteristics</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pUC18/19</td>
<td>Cloning vector, Ap&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Yanisch-Perron et al. (1985)</td>
</tr>
<tr>
<td>M13mp18/19</td>
<td>Bacteriophage cloning/sequencing vector</td>
<td>Yanisch-Perron et al. (1985)</td>
</tr>
<tr>
<td>pACYC184</td>
<td>Cloning vector, Cm&lt;sup&gt;R&lt;/sup&gt; Tc&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Chang and Cohen (1978)</td>
</tr>
<tr>
<td>pMW1</td>
<td>lacI&lt;sup&gt;+&lt;/sup&gt; gene in pACYC184, Tc&lt;sup&gt;R&lt;/sup&gt;</td>
<td>I.S. Roberts, Leicester</td>
</tr>
<tr>
<td>pBR328</td>
<td>Cloning vector, Ap&lt;sup&gt;+&lt;/sup&gt; Cm&lt;sup&gt;R&lt;/sup&gt; Tc&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Soberon et al. (1980)</td>
</tr>
<tr>
<td>pGB110</td>
<td>K5 antigen gene cluster, Ap&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Roberts et al. (1986)</td>
</tr>
<tr>
<td>pKT274</td>
<td>K1 antigen gene cluster, Ap&lt;sup&gt;+&lt;/sup&gt; Km&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Bcharti et al. (1983)</td>
</tr>
<tr>
<td>pRD1</td>
<td>K4 antigen gene cluster, Ap&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Drake et al. (1990)</td>
</tr>
<tr>
<td>pGB148</td>
<td>Subclone from K7 antigen gene cluster in pUC19, Ap&lt;sup&gt;+&lt;/sup&gt;</td>
<td>I.S. Roberts, Leicester</td>
</tr>
<tr>
<td>pGB118::1</td>
<td>S8 (Tn1000) insertion in K5 antigen gene cluster, K5&lt;sup&gt;+&lt;/sup&gt;, Ap&lt;sup&gt;+&lt;/sup&gt;</td>
<td>I.S. Roberts, Leicester</td>
</tr>
<tr>
<td>pGB118::34</td>
<td>Tn5phoA insertion outside of K5 antigen gene cluster, K5&lt;sup&gt;+&lt;/sup&gt;, Ap&lt;sup&gt;+&lt;/sup&gt; Km&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Pazzani, 1992</td>
</tr>
</tbody>
</table>

Ap: Ampicillin  
Cm: Chloramphenicol  
Km: Kanamycin  
Tc: Tetracycline
2.3 TRANSFORMATION OF BACTERIAL CELLS
(Cohen et al., 1972)

2.3.1 Preparation Of Competent Cells

10ml Luria broth were inoculated with 100μl of an overnight culture and incubated at 37°C until the OD_{600nm} reached 0.5 (mid log phase). The bacteria were harvested by centrifugation at 3000g for 10 mins at 4°C, resuspended in 0.4 volumes of ice cold 10mM NaCl and collected by centrifugation (1850g, 5 mins, 4°C). The cells were resuspended in 0.4 volumes of ice cold 100mM CaCl₂ (Sigma) and stored on ice for 30 mins, before being collected by gentle centrifugation (1285g, 5 mins, 4°C). The cell pellet was finally resuspended in 0.1 volumes of ice cold 100mM CaCl₂ and stored on ice until required, (cells may be stored in CaCl₂ solution at 4°C for up to 48 hours).

2.3.2 Transformation Of Competent Cells With Plasmid DNA

100μl of competent cells and up to 20μl of the plasmid DNA were mixed, stored on ice for 1 hour and temperature shocked at 42°C for 3 mins. Luria broth (500μl) was added to the cells, which were then incubated at 37°C for 1 hour to allow the bacteria to recover and express the antibiotic resistance marker encoded by the plasmid. The cells were plated out onto Luria agar (100μl per plate) containing the appropriate antibiotics and the plates were incubated overnight at 37°C. The resulting colonies were screened for the presence of recombinant plasmids by one or more of the following techniques: testing for α-complementation (Section 2.6); restriction analysis (Section 2.5) of small scale preparations of plasmid DNA (Section 2.4); colony hybridization (Section 2.7).
2.3.3 Transfection Of Competent Cells With M13 Bacteriophage DNA

100μl of competent *E. coli* JM101 cells were mixed with an appropriate volume of M13 DNA (up to 20μl), stored on ice for 1 hour and temperature shocked at 42°C for 3 mins. The transformed cells and 200μl of a mid log phase JM101 culture (plating cells) were added to a tube containing 3ml of molten B-top agar (at 42°C), 20μl of 100mM IPTG and 50μl of 2% X-Gal in dimethyl formamide. The suspension was mixed and poured immediately onto a Bagar plate. The plate was swirled gently to distribute the top agar and once set, was incubated overnight at 37°C. Due to α-complementation (Section 2.6) non-recombinant bacteriophage gave rise to blue plaques, whereas recombinant phage yielded white plaques. Double stranded replicative form M13 DNA and single stranded (template) M13 DNA were prepared from white plaques using the procedures described in Section 2.4.

2.4 PROCEDURES USED FOR DNA EXTRACTION

The following solutions were used in the extraction of both plasmid DNA and the replicative form of M13 bacteriophage DNA.

**Solution I:**
- 2% glucose
- 25mM Tris-Cl pH 8.0
- 10mM EDTA pH 8.0
- 4mg/ml lysozyme

**Solution II:**
- 1% SDS
- 0.2M NaOH

**Solution III (per 100ml):**
- 60ml 5M potassium acetate
- 11.5ml glacial acetic acid
- 28.5ml distilled water

Solution III was 3M with respect to potassium and 5M with respect to acetate.
2.4.1 Small Scale Extraction Of Plasmid DNA
(Birnboim and Doly, 1979)

This method was used for both plasmid DNA and the replicative form of M13 bacteriophage DNA (Section 2.4.3).

1.5ml of stationary phase cells taken from an overnight culture were harvested by centrifugation at 12000g for 1 min in a microfuge. The pellet was resuspended in 100μl of ice cold solution I and stored on ice for 30 mins. 200μl of solution II (room temperature) were added and the tube stored on ice for 5 mins, before the addition of 150μl of ice cold solution III. The tube was kept on ice for 5 mins and then after centrifugation for 5 mins at 12000g, the supernatant was transferred to a fresh tube. The solution was extracted once with phenol:chloroform:isoamylalcohol and once with chloroform: isoamylalcohol (Section 2.4.4), before the plasmid DNA was precipitated using 2 volumes of absolute ethanol. The tube was left at room temperature for 2 mins and the DNA collected by centrifugation for 5 mins at 12000g. The pellet was resuspended in 50μl of TE pH 8.0 containing DNAase-free pancreatic RNAase (20μg/ml). The DNA was stored at -20°C.

2.4.2 Large Scale Extraction Of Plasmid DNA
(Birnboim and Doly, 1979)

The cells from a 400ml stationary phase overnight culture were harvested by centrifugation (3330g, 10 mins, 4°C), resuspended in 10ml of ice cold solution I and stored on ice for 30 mins. 20ml of solution II (room temperature) were added and the mixture stored on ice for 10 mins, before the addition of 15ml of ice cold solution III. After 10 mins on ice, cell debris was removed by centrifugation at 36900g for 30 mins at 4°C. To precipitate the DNA, 0.6 volumes of isopropanol were added to the supernatant, mixed well and left at room temperature for 15 mins. The nucleic acids were recovered by centrifugation at 3320g for 30 mins at 20°C. The plasmid DNA was then purified by one of the following two methods.
Method 1- Equilibrium Centrifugation In CsCl Ethidium Bromide Gradients

The nucleic acid pellet was resuspended in sterile nanopure water to a final volume of 17ml, to which 17g of CsCl were added. The solution was transferred to an ultracentrifuge tube containing 1ml of 10mg/ml ethidium bromide. The remainder of the tube was filled with paraffin oil. The closed circular plasmid DNA was separated from other nucleic acids by centrifugation at 40000 rpm, using a Sorvall TV850 rotor in a Sorvall OTD-65 centrifuge, for 20 hours at 20°C. The DNA was visualized using UV light and the lower band containing the plasmid DNA collected using a hypodermic needle and syringe. The ethidium bromide was extracted using an equal volume of CsCl-saturated isopropanol. This was repeated until the red colouration had disappeared. The CsCl was removed from the DNA solution by exhaustive dialysis against sterile distilled water or TE pH 8.0, at room temperature. The DNA was stored at -20°C.

Method 2- Precipitation With Polyethylene Glycol (PEG) (R.Treisman, pers.commun.- Sambrook et al., 1989)

The nucleic acid pellet was resuspended in 3ml of TE pH 8.0. 3ml of an ice cold solution of 5M LiCl were added to precipitate the high molecular weight RNA, which was collected by centrifugation at 9220g for 10 mins at 4°C. An equal volume of isopropanol was added to the supernatant and the precipitated nucleic acids were recovered by centrifugation at 9220g for 10 mins at room temperature. The pellet was dissolved in 500μl of TE pH 8.0 containing DNAase-free pancreatic RNAase (20μg/ml) and stored at room temperature for 30 mins. 500μl of 1.6M NaCl/13% PEG 6000 were added and the plasmid DNA recovered by centrifugation at 12000g for 5 mins in a microfuge. The pellet was resuspended in 400μl of TE pH 8.0 and extracted once with phenol:chloroform:isoamylalcohol and once with chloroform:isoamylalcohol (Section 2.4.4). To precipitate the plasmid DNA, 100μl of 10M ammonium acetate were added to the aqueous phase, together with 2 volumes of absolute ethanol. After 10 mins at room temperature, the DNA was recovered by centrifugation at 12000g for 5 mins. The pellet was resuspended in 500μl of TE pH 8.0 and stored at -20°C.
2.4.3 Extraction Of Single Stranded M13 Bacteriophage DNA

5mL of Luria broth were inoculated with 100μl of a JM101 overnight culture. A white M13 bacteriophage plaque (Section 2.3.3) was picked off a plate using a sterile toothpick, which was dropped into the broth. The culture was incubated at 37°C with constant agitation for 5 hours. 1.5mL of infected cells were harvested by centrifugation at 12000g for 3 mins in a microfuge. The pellet was used to extract the double stranded replicative form of M13 DNA (Section 2.4.1) for restriction analysis (Section 2.5) and the supernatant was used to prepare single stranded template DNA as follows. To precipitate the bacteriophage particles, 200μl of 2.5M NaCl/20% PEG 6000 were added to 800μl of the supernatant and after 30 mins at room temperature, the particles were collected by centrifugation at 12000g for 5 mins. The supernatant was discarded and any residual liquid was removed after a further 2 mins centrifugation at 12000g. The pellet was resuspended in 100μl of 1.1M sodium acetate pH 7.0 and extracted once with phenol:chloroform:isoamylalcohol and once with chloroform:isoamylalcohol (Section 2.4.4). To precipitate the DNA, 2 volumes of absolute ethanol were added to the aqueous phase and after 30 mins at -20°C, the DNA was collected by centrifugation at 12000g for 10 mins. The pellet was resuspended in 20μl of TE pH 8.0 and stored at -20°C. To check the template DNA, 2μl were visualized using agarose gel electrophoresis (Section 2.5) and up to 7μl were used in a sequencing reaction (Section 2.8).

2.4.4 Phenol:Chloroform Extraction And Ethanol Precipitation Of DNA

Phenol:chloroform:isoamylalcohol (25:24:1) containing 8-hydroxyquinoline (1:0.001) was equilibrated by extraction with 0.1M Tris-Cl pH 7.6 and was stored in the dark at 4°C under 0.01M Tris-Cl pH 7.6.

The following procedure was used to remove contaminating proteins from DNA preparations. An equal volume of phenol:chloroform:isoamylalcohol was mixed with the sample and after centrifugation at 12000g for 3 mins in a microfuge, the
aqueous phase containing the DNA was transferred to a fresh tube. These steps were repeated until no protein precipitate was visible at the interface. Generally, one extraction with an equal volume of chloroform:isoamylalcohol (24:1) was performed using the above procedure, to remove traces of phenol from the DNA preparation.

DNA was recovered by ethanol precipitation using the following procedure, unless otherwise indicated. Sodium acetate was added to the sample at a final concentration of 300mM, together with 2-3 volumes of -20°C absolute ethanol. After mixing, the sample was placed at -20°C for a minimum of 30 mins and the DNA collected by centrifugation at 12000g for 5 mins in a microfuge. The supernatant was removed by aspiration and once dry, the pellet was resuspended in the desired volume of sterile nanopure water or TE pH 8.0.

2.5 PROCEDURES USED IN ROUTINE DNA MANIPULATION

In order to subclone specific DNA fragments into a vector, the DNA was cleaved by the appropriate restriction endonucleases and the relevant fragments ligated together. Cleavage by restriction endonucleases was also used to analyse DNA preparations to identify recombinant plasmids containing the correct insert. Restriction endonucleases were used in accordance with the manufacturers recommendations. In a typical reaction the DNA sample was mixed with sterile nanopure water to give a final volume of 18μl, to which 2μl of 10X reaction buffer and 1-2 units of enzyme were added. The reaction was incubated at the appropriate temperature (generally 37°C) for approximately 1 hour.

Restriction fragments were separated by agarose gel electrophoresis using 0.7% SeaKem agarose (FMC BioProducts) in 1X TAE buffer (50X per litre: 242g Tris base, 100ml 0.5M EDTA pH 8.0, 57.1ml glacial acetic acid). Small fragments (less than 1kb) were separated using a 1% gel. 4μl of 6X gel loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol FF, 15% Ficoll type 400, 10μg/ml DNAase-free pancreatic RNAase) were mixed with the DNA sample before loading. Electrophoresis was
carried out in 1X TAE buffer containing ethidium bromide at a final concentration of 0.5µg/ml, at 120V for approximately 2 hours or 10-20V overnight. The restriction fragments were visualized using a long wave UV transilluminator. Appropriate DNA size markers were used: 1kB ladder (GIBCO-BRL).

For the purpose of subcloning, the appropriate restriction fragments were purified from the agarose using the following procedure. Electrophoresis was carried out in the dark to protect the DNA from UV induced damage caused by the intercalated ethidium bromide. Agarose slices containing the fragments of interest were excised from the gel and placed in dialysis tubing containing 300µl of 1X TAE buffer without ethidium bromide. The DNA was electroeluted out of the gel at 120V for 1 hour and then the polarity was reversed for 30 seconds to release the DNA bound to the dialysis tubing. The DNA was recovered from the buffer by precipitation with ethanol (Section 2.4.4) and collected by centrifugation at 12000g for 5 mins in a microfuge. The pellet was resuspended in 10µl of sterile nanopure water.

Ligation of the purified DNA fragments was carried out as follows. 1µl of 10X ligation buffer (0.5M Tris-Cl pH 7.5, 0.1M MgCl₂, 0.01M ATP), up to 8µl of the DNA fragments, 1µl of T4 DNA ligase (1u/µl, GIBCO-BRL) and sterile nanopure water to give a final volume of 10µl were incubated at 14°C overnight. 2µl of the reaction were visualized using agarose gel electrophoresis to check that ligation had occurred, before the DNA was used to transform a suitable strain (Section 2.3).

2.6 SCREENING FOR RECOMBINANT PLASMIDS

The method used for screening for recombinant plasmids depended on which vector was used. pUC and M13 bacteriophage vectors contain a segment of DNA coding for the amino terminal fragment of β-galactosidase and can complement strains which only express the carboxyl terminal portion of this protein, a phenomenon known as α-complementation. This results in the production of blue colonies or plaques on media containing the chromogenic substrate X-Gal (at a final concentration of
40μg/ml) after induction with IPTG (present in the medium at a final concentration of 23.8μg/ml). When a fragment of DNA is cloned into the multiple cloning site of these vectors, α-complementation generally cannot occur and the resulting colonies or plaques are white. Bacteria transformed with these vectors were plated on Luria agar containing X-Gal and IPTG to test for α-complementation. Small scale DNA preparations were made from the white colonies or plaques (Section 2.4) and restriction analysis (Section 2.5) was used to determine whether the correct insert was present. With vectors where there was no selection procedure, either DNA isolated from individual colonies was analysed by restriction endonuclease cleavage or the colonies were screened by hybridization using a radiolabelled nucleic acid probe (Section 2.7).

2.7 DNA HYBRIDIZATION TECHNIQUES

2.7.1 Preparation Of Nylon Filters For Colony Hybridization

A method based on that described by Grunstein and Hogness (1975) was used for in situ lysis of bacterial colonies on nylon filters, attachment of the released DNA to the filter and its subsequent hybridization to a radiolabelled probe.

A nylon filter (Hybond-N, Amersham) was placed on a Luria agar plate. The colonies to be screened were transferred using sterile toothpicks onto a master agar plate and onto the filter, in the form of small streaks in an identical position on both plates. A suitable negative control was included. The plates were incubated overnight at 37°C, in the presence of the appropriate antibiotics.

The nylon filter was placed on Whatman 3mm papers (approximately 5) saturated with denaturing solution (0.5M NaOH, 1.5M NaCl) for 5 mins, before being transferred to Whatman soaked in neutralizing solution (0.5M Tris-Cl pH 7.5, 3M NaCl) for a further 5 mins. The filter was air dried at room temperature for 15-30 mins, wrapped in Saran wrap and the DNA fixed to the filter by 5 mins irradiation with longwave UV light from a transilluminator. To remove cell debris, the
filter was scrubbed using polymer wool soaked in 5X SSC. The filter was air dried and stored in the dark at room temperature until required for hybridization (Section 2.7.4).

2.7.2 Transfer Of DNA To Nylon Filters By Southern Blotting

DNA was transferred from agarose gel to a solid support using the method described by Southern (1975).

The DNA was cleaved by appropriate restriction endonucleases and the fragments separated according to size by agarose gel electrophoresis at 10-20V overnight (Section 2.5). The gel was photographed alongside a ruler. To prevent cross-linking of the DNA to the agarose, exposure to UV was kept to a minimum. To partially depurinate the DNA (to aid transfer of large fragments), the gel was washed in 250ml 0.25M HCl for 7 mins. The gel was rinsed in distilled water and placed in 250ml of denaturing solution (0.5M NaOH, 1.5M NaCl) for 30 mins. After rinsing in distilled water, the gel was placed in 250ml of neutralizing solution (0.5M Tris-Cl pH 7.5, 3M NaCl) for 30 mins. Gentle agitation was used in these steps. The gel was again rinsed in distilled water, before being placed on 6 pieces of Whatman 3mm filter paper cut to gel size and soaked in 20X SSC, avoiding the generation of air bubbles. A precut piece of nylon filter (Hybond-N, Amersham) soaked in 3X SSC was placed on the gel, followed by a precut sheet of Whatman soaked in 3X SSC, three sheets of dry Whatman and a stack of paper towels. Again care was taken to avoid air bubbles. Finally a plastic gel plate and a 500g weight were placed on top. During transfer the lower sheets of Whatman were regularly soaked with 20X SSC and the wet paper towels removed. After 4-20 hours the apparatus was dismantled, the nylon filter rinsed briefly in 3X SSC and air dried thoroughly. To fix the DNA, the filter was wrapped in Saran wrap and exposed for 5 mins to longwave UV light from a transilluminator. The filter was then stored in the dark at room temperature until required for hybridization (Section 2.7.4).
2.7.3 Synthesis Of $^{32}$P Labelled DNA Probes Using Random Hexadeoxynucleotide Primers

The method described by Feinberg and Vogelstein (1983, 1984) was used to radiolabel the DNA in the presence of melted agarose.

The DNA fragments generated by cleavage with the appropriate restriction endonucleases, were separated by agarose gel electrophoresis (Section 2.5) using 1% low melting point agarose (GIBCO-BRL). The fragment required for labelling was excised and placed in sterile nanopure water (1.5ml water/g agarose). The sample was boiled for 7 mins and then stored at -20°C. Before a labelling reaction, the sample was reboiled and kept at 37°C until required. The labelling reaction was set up as follows.

5μl sterile nanopure water (to a total volume of 25μl)
5μl 5X OLB
xμl DNA fragment (25ng, up to 18μl)
1μl Klenow fragment E.coli DNA polymerase I (6u/μl, GIBCO-BRL)
1μl [α-$^{32}$P] dCTP (Amersham) specific activity - 10μCi/μl

The reaction was incubated at room temperature for 5-20 hours.

T.E. Buffer pH 7.0:
3mM Tris-Cl
0.2mM EDTA

Solution O pH 8.0:
1.25M Tris-Cl
0.125M MgCl₂
stored at 4°C

dNTP's (Pharmacia):
dATP ]
dTTP ] 0.1M in T.E. buffer, stored at -20°C
dGTP ]
Solution A:
1ml solution 0
18μl β-mercaptoethanol
5μl each of dATP, dTTP, dGTP
stored at -20°C

Solution B:
2M HEPES (titrated to pH 6.6 with 4M NaOH)
stored at 4°C

Solution C:
Hexadecoxynucleotides (Pharmacia), 50 A units resuspended in
550μl T.E. buffer to give a final concentration of 90 A
units/ml, stored at -20°C

5X OLB:
Solutions A, B and C were mixed in the ratio 10:25:15
respectively and stored at -20°C.

2.7.4 Hybridization Of Radiolabelled Probes To Immobilized
Nucleic Acids

50X Denhardts:
1% each Ficoll type 400, BSA and polyvinylpyrrolidone,
stored at -20°C

Prehybridization Solution:
3X SSC
5X Denhardts
200μg/ml salmon sperm DNA (Sigma)
0.1% SDS
6% PEG 6000

Hybridization Solution:
same as above except 2X Denhardts

Prehybridization and hybridization solutions were made up in
advance, minus the salmon sperm DNA and stored in 25ml aliquots
at -20°C. The salmon sperm DNA (5mg/ml in sterile nanopure
water), which was stored at -20°C, was sheared by repeatedly
forcing it through a narrow gauge needle and was denatured by boiling (5 mins) immediately before use.

The nylon filter (Sections 2.7.1 and 2.7.2) was placed in a bottle containing 25ml of prehybridization solution (prewarmed to 65°C) and rotated in a hybridization oven (HYBAID) at 65°C for 1-2 hours. The prehybridization solution was replaced by 25ml of hybridization solution (prewarmed to 65°C) containing the radiolabelled probe (denatured by boiling for 5 mins immediately before use). The hybridization reaction was rotated at 65°C for 4-20 hours. Unless stated otherwise, after hybridization, the filter was washed twice for 20 mins in 250ml of 2X SSC, 0.1% SDS (prewarmed to 65°C) in a shaking waterbath at 65°C. The filter was then washed using a more stringent wash solution. Alteration of the concentration of SSC in the wash solution affected the Tm of the hybrid DNA and consequently its stability at a particular temperature. 2X SSC allowed hybrids with approximately 70% homology to remain intact at 65°C, whereas with 0.5X SSC approximately 80% homology was required and with 0.1X SSC approximately 95% homology was required. The SDS concentration and the washing temperature were kept at 0.1% and 65°C, respectively. The filter was washed twice in 250ml of the appropriate solution (prewarmed to 65°C) for 20 mins in a shaking waterbath at 65°C. The filter was wrapped in Saran wrap and placed in a cassette carrying intensifying screens. Kodak X-Omat S film was exposed to the filter at -70°C and then developed using an Agfa-Gevaert automatic film processor.

2.8 DNA SEQUENCING

DNA was sequenced using the dideoxy chain termination method described by Sanger et al. (1977).

2.8.1 DNA Sequencing Reactions

The bacteriophage cloning vectors M13mpl8 and M13mpl9 were used to generate single stranded DNA templates (Section 2.4.3). -40 universal primer or specific oligonucleotide primers were used to initiate synthesis. ^35S dATP was used to radiolabel the DNA fragments. Sequencing reactions were carried out using the
Sequenase and Sequenase Version 2.0 kits produced by the United States Biochemical Corporation. The manufacturers protocol was followed. To anneal the primer to the template, 1 μl of primer (0.5 pmol/μl), 2 μl of 5X reaction buffer (200 mM Tris-Cl pH 7.5, 100 mM MgCl₂, 250 mM NaCl) and 7 μl DNA template (approximately 1 μg) were incubated at 65°C for 2 mins, allowed to cool slowly to 30°C on the bench and placed on ice. To set up the labelling reaction 1 μl of 0.1M dithiothreitol, 2 μl of a 1 in 10 dilution of labelling mix, 0.5 μl [³²S] dATP (Amersham, 10 μCi/μl) and 2 μl of Sequenase diluted 1:8 in ice cold enzyme dilution buffer (10 mM Tris-Cl pH 7.5, 5 mM dithiothreitol, 0.5 mg/ml BSA) were added to the annealing reaction. After mixing, the labelling reaction was incubated at room temperature for 2-5 mins. 2.5 μl of each dideoxynucleotide termination mix were placed in separate tubes and incubated at 37°C for at least 1 min. 3.5 μl of the labelling reaction were added to each of the four termination tubes, mixed and incubated at 37°C for 3-5 mins. 4 μl of stop solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF) were added to each termination reaction. The sequencing reactions were stored at -20°C until required (up to 1 week).

Labelling Mix:
7.5 μM dTTP, 7.5 μM dCTP, 7.5 μM dGTP

Dideoxynucleotide Termination Mixes:
800 μM dATP, 800 μM dTTP, 800 μM dCTP, 800 μM dGTP, 50 mM NaCl and either 80 μM ddATP, ddTTP, ddCTP or ddGTP

2.8.2 DNA Sequencing Gel

The radiolabelled DNA fragments were separated by denaturing buffer gradient polyacrylamide gel electrophoresis (Biggin et al., 1983).

40% Acrylamide (per litre):
380 g acrylamide (BDH)
20 g N,N'-methylenebisacrylamide (BDH)
deionised using 50 g of Amberlite (BDH) per litre stored at 4°C in the dark
10X TBE (per litre):
109g Tris
55g boric acid
9.3g EDTA
pH 8.3 without adjustment

0.5X TBE 6% Acrylamide Urea Mixture (per litre):
430g urea
50ml 10X TBE
150ml 40% acrylamide
stored at 4°C in the dark

2.5X TBE 6% Acrylamide Urea Mixture (per litre):
430g urea
250ml 10X TBE
150ml 40% acrylamide
50g sucrose
50mg bromophenol blue
stored at 4°C in the dark

Gel Solution 1:
40ml 0.5X mixture
180µl 10% APS (Bio-Rad)
7.5µl TEMED (Sigma)

Gel Solution 2:
7ml 2.5X mixture
45µl 10% APS
2.5µl TEMED

Two glass plates (20cm x 50cm), one of them notched, were cleaned and the surface of one plate was siliconized using dimethyldichlorosilane solution (BDH). The plates were taped together, separated by 0.4mm spacers. 10ml of gel solution 1, followed by 7ml of gel solution 2, were drawn into a 25ml pipette and four air bubbles introduced to generate a rough gradient. The solution was expelled between the gel plates, which were held at an angle of approximately 45° to the horizontal. The remainder of solution 1 was used to complete the gel and the flat edge of the sharkstooth comb (0.4mm) was
inserted approximately 0.5cm into the gel matrix. Bull dog clips were positioned at both sides of the gel, which was left to polymerize for a minimum of 2 hours.

After polymerization the comb was removed and aluminium plates were placed on either side of the gel (to facilitate even heat distribution). The gel was clamped in position in a vertical electrophoresis system with the upper reservoir filled with 0.5X TBE and the lower one with 1X TBE. The gel was pre-run at 40W for 20 mins, the surface of the gel rinsed with 0.5X TBE buffer and the comb placed in position with the teeth penetrating the surface. The sequencing reactions were heated at 85°C for 5 mins before being loaded. Electrophoresis was performed at 40W until the first dye front ran off the gel (approximately 2 hours 15 mins). This enabled DNA closest to the primer to be sequenced. For more distal sequence information the gel was run for increasingly longer periods, generally 5 hours and 8 hours.

After electrophoresis the apparatus was dismantled and the plates prised apart. The gel was soaked in sequence fix (10% methanol, 10% glacial acetic acid) for 10 mins, before being rinsed with distilled water. The gel was then transferred from the glass plate to a wet piece of Whatman 3mm paper, covered with Saran wrap and dried under vacuum at 80°C for 1 hour. Dupont Cronex film was exposed to the gel in a cassette placed at room temperature. The autoradiograph was developed using an Agfa-Gevaert automatic film processor.

The nucleotide sequence read from the gels was analysed using the University of Wisconsin Genetics Computer Group sequence analysis software (Devereux et al., 1984) on the VAX cluster at the University of Leicester and the University of Oxford.

2.9 OVEREXPRESSION OF PROTEINS

The expression vector pKK223-3 (Brosius and Holy, 1984) contains the strong tac promoter, a hybrid trp-lac promoter regulated by the lac repressor. When derepressed by the addition of IPTG, this strong promoter directs the expression
of genes cloned in the vector. The plasmid pMW1 (lacr^ repressor gene cloned into pACYC184) can be used to help maintain repression until the addition of IPTG.

Recombinant pKK223-3 DNA was transformed into E. coli LE392 harbouring the pMW1 plasmid (Section 2.3). 100ml cultures of this strain and LE392 harbouring non-recombinant pKK223-3 and pMW1, supplemented with the appropriate antibiotics, were incubated at 37°C until an OD_{600nm} of 0.5 was reached, at which point IPTG was added to a final concentration of 1mM. Incubation at 37°C was continued and 10ml aliquots were removed at 0 mins, 30 mins, 60 mins and 90 mins after the addition of IPTG. Samples were also taken from cultures where no IPTG had been added. The bacteria were harvested by centrifugation at 3000g for 10 mins at 4°C and resuspended in 1ml of 50mM Tris-Cl pH 6.8. The samples were sonicated (4X 15 seconds with 15 seconds in between, on ice, at 50W) and then boiled for 3 mins in an equal volume of 2X SDS-PAGE loading buffer (Section 2.10.1). Cell debris was collected by centrifugation at 12000g for 2 mins and the supernatant was stored at -20°C, until required for analysis by SDS-polyacrylamide gel electrophoresis (Section 2.10).

2.10 SDS-POLYACRYLAMIDE GEL ELECTROPHORESIS (SDS-PAGE)

Denaturing SDS-PAGE (Laemmli, 1970) was used to analyse the proteins in samples taken from the overexpression (Section 2.9) and minicell (Section 2.11) experiments.

2.10.1 Pouring And Running SDS-PAGE Gels

44% Acrylamide (per litre):
440g acrylamide (Bio-Rad)
8g N,N'-methylenebisacrylamide (BDH)
stored at 4°C in the dark

Resolving Gel Buffer pH 8.8:
0.75M Tris base
0.2% SDS
pH 8.8 with HCl
Stacking Gel Buffer pH 6.8:
0.25M Tris base
0.2% SDS
pH 6.8 with HCl

Resolving Gel (10ml):
- 12.5% 15.0%
- Resolving gel buffer 4.94ml 4.94ml
- 44% acrylamide 2.84ml 3.40ml
- Sterile nanopure water 1.84ml 1.28ml
- 1% APS (Bio-Rad) 348µl 348µl
- TEMED (Sigma) 40µl 40µl

Stacking Gel (9.5ml):
- 3.5%
- Stacking gel buffer 4.50ml
- 44% acrylamide 0.75ml
- Sterile nanopure water 3.99ml
- 1% APS 225µl
- TEMED 30µl

2X SDS-PAGE Loading Buffer:
- 125mM Tris-Cl pH 6.8
- 1.44M β-mercaptoethanol
- 4% SDS
- 0.1% bromophenol blue
- 40% glycerol

Tris-Glycine Electrophoresis Buffer (per litre):
- 3.02g Tris base
- 1.0g SDS
- 14.4g glycine
- pH 8.5 without adjustment

SDS-PAGE was generally carried out using the Bio-Rad mini protean II system. The glass plates were assembled according to the manufacturer’s instructions. A mark was made 0.5cm below the teeth of the comb and the resolving gel was poured to this level. A thin layer of isobutanol was loaded onto the surface and the gel was allowed to polymerize at room temperature for 15 mins. The gel surface was rinsed with water before the
stacking gel was poured. Then the comb was inserted and the gel left to polymerize at room temperature for 30 mins. Samples were boiled in an equal volume of 2X SDS-PAGE loading buffer for 3 mins before loading. The gel was placed in the vertical electrophoresis system containing Tris-glycine buffer, the samples loaded (up to 15µl) and electrophoresis carried out at 200V until the dye front reached the bottom of the gel. The apparatus was dismantled and the gels either stained with Coomassie blue (Section 2.10.2) or used for autoradiography (Section 2.10.3).

2.10.2 Staining With Coomassie Blue

SDS-PAGE Coomassie Blue Stain:
0.5% Coomassie Brilliant Blue R-250 (Bio-Rad)
30% methanol
10% glacial acetic acid

SDS-PAGE Destain:
30% methanol
10% glacial acetic acid

The gel was simultaneously fixed with methanol:glacial acetic acid and stained with Coomassie blue for 15 mins, before being destained in methanol:glacial acetic acid for 4-8 hours. The gel was then either photographed or placed on Whatman 3mm paper, covered with Saran wrap and dried under vacuum at 60°C for 30 mins.

2.10.3 Autoradiography

The SDS-PAGE gel was fixed for 15 mins in 25% isopropanol:10% glacial acetic acid, before being dried (Section 2.10.2). Hyperfilm (Amersham) or Kodak X-Omat S film was exposed to the gel in a cassette placed at room temperature. The film was developed using an Agfa-Gevaert automatic film processor. An optional step after fixation was to place the gel in Amplify (Amersham) for 30 mins, to enhance the autoradiographic signal.
2.10.4 Large SDS-PAGE Gels

For SDS-PAGE using a large gel, an 11% polyacrylamide resolving gel and a 5.2% polyacrylamide stacking gel were used. 25μl of sample were loaded and electrophoresis carried out at 25mA until the sample reached the boundary between the stacking and the resolving gel, at which point a current of 50mA was applied until the dye front reached the bottom of the gel.

2.11 MINICELLS

The E. coli strain DS410 (Wilkins et al., 1981) carries a mutation which results in the production of spherical anucleate minicells. However, these minicells may harbour plasmid DNA. Genes encoded by the plasmid can be expressed in minicells in the presence of 35S methionine and the radiolabelled proteins detected by SDS-PAGE (Section 2.10) and autoradiography (Section 2.10.3).

2.11.1 Isolation Of Minicells

400ml of brain heart infusion broth, supplemented with the appropriate antibiotics, were inoculated with either DS410 alone, DS410 harbouring the vector or DS410 harbouring the recombinant vector and incubated at 37°C overnight. The bacteria were subjected to gentle centrifugation at 1000g for 5 mins at 4°C and the resulting pellet contained the vegetative cells. The minicells were collected from the supernatant by centrifugation at 8500g for 15 mins at 4°C. The pellet was resuspended in 6ml of ice cold 1X M9 salts (10X, per litre: 60g Na₂HPO₄, 30g KH₂PO₄, 10g NH₄Cl, 5g NaCl) and loaded onto two 20ml sucrose gradients (20% in 1X M9 salts, with the gradient made by freezing and subsequent thawing at 4°C). Centrifugation at 4000g for 20 mins at 4°C resulted in the separation of the minicells from the pellet of remaining vegetative cells. The minicells were collected from each gradient, harvested by centrifugation at 9220g for 10 mins at 4°C and resuspended in 3ml of ice cold 1X M9 salts, before being loaded onto another 20ml sucrose gradient. After centrifugation (4000g, 20 mins, 4°C), the minicells were collected from the gradient, harvested
by centrifugation (9200g, 10 mins, 4°C) and resuspended in ice
cold 70% 1X M9 salts/30% glycerol to give a final OD$_{600nm}$ of
2.0. Aliquots (50μl) were stored at -20°C (for up to 6 months)
until required for labelling (Section 2.11.2).

2.11.2 Radiolabelling Proteins Using L-$[^35]$S$\text{] Methionine

**Methionine Assay Medium** (Difco):
0.105g in 1ml sterile nanopure water
boiled for 2 mins

**Solution I** (per ml):
15μl methionine assay medium (fresh)
10μl 40% glucose
100μl 10X M9 salts

**Protease Inhibitor Mix A:**
1μg/ml Pepstatin A
1μg/ml Antipain
1μg/ml Chymostatin
1μg/ml Leupeptin
stored at -20°C

**Solution II** (per ml):
835μl protease inhibitor mix A
20μl 5mg/ml PMSF (stored at -20°C)
20μl 0.5M EDTA
100μl 10X M9 salts
10μl 40% glucose
15μl methionine assay medium

**Protease Inhibitor Mix B:**
1ml protease inhibitor mix A
20μl 5mg/ml PMSF
20μl 0.5M EDTA

50μl of minicells were harvested by centrifugation at 12000g
for 2 mins in a microfuge, resuspended in 100μl of 1X M9 salts
and collected by centrifugation (12000g, 2 mins). The cells
were resuspended in 100μl of solution I. 10μl of 2.2mg/ml D-
cycloserine were added and the cells incubated at 37°C for 90 mins. The cells were harvested by centrifugation at 12000g for 2 mins and resuspended in 100μl of solution II. 15μCi of L-[15S] methionine (Amersham, 10μCi/μl) were added and the cells incubated for 45 mins at 37°C. The cells were collected by centrifugation (12000g, 2 mins), resuspended in 90μl of solution II and 10μl of 2mg/ml cold methionine (Sigma) and incubated at 37°C for 15 mins. The cells were harvested by centrifugation (12000g, 2 mins) and resuspended in 12.5μl of 2X SDS-PAGE loading buffer (Section 2.10.1) and 12.5μl protease inhibitor mix B. The samples were stored at -20°C (up to 1 week) until required for analysis by SDS-PAGE (Section 2.10).

2.12 EXTRACTION AND DETECTION OF CAPSULAR POLYSACCHARIDE

2.12.1 Extraction Of Polysaccharide

Total cell polysaccharide was extracted from a 10ml stationary phase bacterial culture. The cells were harvested by centrifugation at 3000g for 10 mins at 4°C and were vigorously resuspended in 5ml of absolute ethanol. This was repeated twice to dehydrate the preparation and the final pellet was resuspended in 1ml of ethanol. The sample was transferred to a 1.5ml microfuge tube and collected by centrifugation at 12000g for 1 min. The pellet was resuspended in 1ml of acetone and then collected by centrifugation (12000g, 1 min). All traces of acetone were removed and the pellet air dried at 37°C for 20 mins, before being resuspended in 50μl of 1M MgCl₂. The sample was incubated at 37°C for 2-3 hours. After centrifugation at 12000g for 5 mins, the supernatant containing a crude extract of polysaccharide was retained and stored at -20°C.

2.12.2 Double Immunodiffusion

Barbitone Buffer pH 8.2: 6g 5′5′ diethylbarbituric acid sodium salt was dissolved in 400ml distilled water. 2.2g 5′5′ diethylbarbituric acid was dissolved in 75ml distilled water at 95°C. The two solutions were mixed immediately and adjusted to pH 8.2 using 5M NaOH. The volume was made up to 500ml.
The technique of double immunodiffusion (Ouchterlony) was used to analyse polysaccharide extracts (Section 2.12.1) using anti-capsular polysaccharide antibodies. 4ml of 1% SeaKem agarose in barbitone buffer pH 8.2, were pipetted into a 35mm diameter petri dish and once set, 2.5mm diameter holes were cut into the agarose using a punch. Generally 10-15µl of anti-capsular polysaccharide antibody were loaded into the central well and a similar volume of polysaccharide antigen loaded into the 6 peripheral wells. The plates were left in a moist chamber at room temperature overnight. If the antibody recognized the antigen, a precipitin arc would form at their point of equivalence.

2.12.3 Infection With Capsule Specific Bacteriophage

Bacteriophage have been isolated which only recognize E. coli encapsulated by a specific polysaccharide. The K5 antigen specific bacteriophage was used to detect the presence of a K5 polysaccharide capsule.

A 10ml culture was incubated at 37°C until the OD_{600nm} was 0.4. The cells were harvested by centrifugation at 3000g for 10 mins at 4°C, resuspended in 10ml of ice cold 10mM MgSO₄ and stored on ice until required. The K5 specific bacteriophage lysate (approximately 10⁸-10⁹ Pfu/ml) was diluted in phage dilution buffer (10mM Tris-Cl pH 7.5, 10mM MgSO₄, 0.01% gelatin) to give a serial dilution up to 10⁻⁸. 100µl of bacteria and 100µl of diluted bacteriophage were mixed together and incubated at room temperature for 20 mins to allow adsorption. 3ml of molten B-top agar (at 42°C) were added and the mixture poured onto a Luria agar plate, supplemented with antibiotics when necessary. The plate was incubated overnight at 37°C. Infection with bacteriophage resulted in the appearance of plaques on the bacterial lawn.

To prepare a bacteriophage lysate, the B-top agar from an infected plate was scraped off into 3ml of Luria broth. 50µl of chloroform were added and the cell debris and agar collected by centrifugation (3000g, 15 mins, 4°C). The clear supernatant containing the bacteriophage was retained and stored at 4°C.
2.13 PRELIMINARY PROMOTER ANALYSIS

Chloramphenicol acetyl transferase (CAT) acetylates the antibiotic chloramphenicol, resulting in its inactivation.

\[
\text{Chloramphenicol + acetylCoA} \rightarrow \text{chloramphenicol 3-acetate + CoASH}
\]

The plasmid pKK232-8 (Brosius, 1984) contains a promoterless chloramphenicol acetyl transferase (cat) gene. When DNA containing a promoter is cloned into the multiple cloning site, the cat gene is expressed and the host strain becomes resistant to chloramphenicol. The minimum concentration of chloramphenicol required to inhibit growth can be determined (Section 2.13.1) and CAT specific activity can be measured using a spectrophotometric assay (Shaw, 1975) (Section 2.13.2), to give an indication of the strength of the promoter.

2.13.1 Minimum Inhibitory Concentration Of Chloramphenicol

Chloramphenicol was added to 5ml aliquots of Luria broth to give a range of final concentrations from 0μg/ml up to 90μg/ml. Any other appropriate antibiotics were added at their normal concentration. The broths were inoculated with 50μl of an overnight culture of E. coli LE392 harbouring either recombinant or non-recombinant pKK232-8. A vector encoding a constitutive cat gene was used as a positive control, either pBR328 (Soberon et al., 1980) or pACYC184 (Chang and Cohen, 1978). The cultures were incubated at 37°C and growth at the different concentrations of chloramphenicol was observed and recorded at regular intervals.

2.13.2 Spectrophotometric Assay For CAT Activity

The acetylation of chloramphenicol results in the generation of a free CoA sulphydryl group. This can react with 5,5’ dithiobis-2-nitrobenzoic acid (DTNB) to yield amongst other products, a molar equivalent of 5-thio-2-nitrobenzoate, which is yellow and has a molar extinction coefficient of 13600 at 412nm. The rate of production of this coloured product at 37°C...
can be followed by measuring the change in absorbance at 412nm before and after the addition of chloramphenicol. The net change in absorbance per min is divided by 13.6 to give the μmoles of DTNB reacted per min, which is equal to the μmoles of chloramphenicol acetylated per min. 1μmole of chloramphenicol acetylated per min at 37°C equals 1 unit of CAT activity. Thus, if the concentration of protein in the sample is known, the specific activity in units of CAT activity per mg of total cell protein can be calculated.

10ml of Luria broth, supplemented with appropriate antibiotics, were inoculated with 100μl of an overnight culture of LE392 harbouring either pBR328 (positive control), recombinant or non-recombinant pKK232-8. The cultures were incubated at 37°C until they had reached mid-log phase, whereupon the OD<sub>600nm</sub> was recorded. 3ml of the culture were used to make a small scale plasmid DNA preparation (Section 2.4.1), to enable the DNA to be checked by restriction analysis (Section 2.5). The bacteria from 1ml of culture were harvested by centrifugation at 12000g for 2 mins in a microfuge and resuspended in 1ml of sterile PBS (OXOID). The cells were washed once more in PBS and then the sample was sonicated (4X 15 seconds with 15 seconds in between, on ice, at 50W). Cell debris was collected by centrifugation at 12000g for 5 mins and the supernatant was retained for use in the protein assay and the CAT assay.

Total cell protein was determined using the Bradford Assay (Bradford, 1976), which is based on the observation that the absorbance maximum for an acidic solution of the dye Coomassie Brilliant Blue G-250 changes from 465nm to 595nm when it binds to proteins. The Bio-Rad microassay procedure was followed. Lysozyme, at concentrations ranging from 2μg/ml to 80μg/ml in sterile PBS, was used as the protein standard and 200μl of the sample were diluted 1 in 5 in sterile PBS. 0.2ml of dye reagent concentrate (Bio-Rad) were mixed with either 0.8ml of standards, diluted sample or sterile PBS. After 5 mins the OD<sub>595nm</sub> versus the reagent blank was measured. The concentration of the diluted sample in μg/ml was read from a standard curve of OD<sub>595nm</sub> versus the concentration of the lysozyme standards.
DTNB:
4mg/ml 1M Tris-Cl pH 7.8

AcetylCoA:
5mM in sterile nanopure water

Chloramphenicol:
5mM in sterile nanopure water

100μl of neat sample were mixed with 1ml of freshly prepared substrate solution (0.4mg/ml DTNB, 100mM Tris-Cl pH 7.8, 0.1mM acetylCoA) and incubated at 37°C for 10 mins. Then the rate of increase in absorbance at 412nm was recorded over a period of 10 mins at 37°C before and after chloramphenicol was added at a final concentration of 0.1mM. Specific activity in units of CAT activity per mg of total cell protein was calculated as described above.

The spectrophotometric assay was also carried out for cultures grown at 18°C.
CHAPTER 3

REGION 3 OF THE K5 CAPSULE GENE CLUSTER: IDENTIFICATION OF AN INNER-MEMBRANE CAPSULAR POLYSACCHARIDE TRANSPORT SYSTEM

3.1 INTRODUCTION

Region 3 of the \( \text{kps} \) locus was previously shown by DNA hybridization and complementation studies to have a common function in the biosynthesis of chemically diverse group II K antigens (Roberts et al., 1988a). Mutations in this region result in the presence of cytoplasmic, full length, immunoreactive polymer, which is devoid of both KDO and phospholipid (Bronner, Sieberth and Jann, personal communication; Kroncke et al., 1990a). This phenotype can be mimicked by the addition of the membrane energy uncoupling agent CCCP to wild type cells prior to shifting from a capsule restrictive (18°C) to permissive (37°C) temperature (Jann, personal communication; Kroncke et al., 1990b). As a result of these findings, it has been tentatively proposed that the products encoded by region 3 are involved in the energy-dependent translocation of polymer across the inner membrane (Boulnois and Jann, 1989; Kroncke et al., 1990a). It was also suggested that the attachment of phosphatidic acid (and KDO) to the polysaccharide may be associated with this process in a vectorial manner (Boulnois and Jann, 1989). However, it is possible that the proteins encoded by region 3 are responsible for the addition of phosphatidic acid (and KDO) to the reducing terminus of the polysaccharide and that this post-polymerization modification is then necessary for the subsequent export of the polymer (Kroncke et al., 1990a).

In order to try to establish the function of region 3 in capsule biogenesis, the nucleotide sequence of this region in the \( E. \ coli \) K5 antigen gene cluster was determined and a preliminary characterization of the two encoded proteins undertaken. These experiments are described in this chapter.
3.2 RESULTS

Region 3 has been partly defined on the basis of the extensive DNA homology (approximately 95%) observed in this region between different group II capsule gene clusters (Roberts et al., 1988a). Previous preliminary Southern hybridization experiments, using a fragment corresponding to region 3 of the K1 antigen gene cluster as a probe, localized the region 3 functions of the K5 antigen gene cluster to a 4.7kb EcoRI-HindIII fragment of the cosmid pGB118, which encodes the K5 kps locus (Roberts et al., 1988a) (Figure 3.1). Furthermore, when a 1.8kb BglII fragment located at one end of this region was deleted from pGB118, it was found that bacteria harbouring the resulting plasmid had a typical region 3 mutant phenotype (Roberts et al., 1988a). However, to facilitate the cloning and nucleotide sequence analysis of the K5 region 3, further hybridization experiments were required to define more precisely the location and extent of this region and these are described below.

3.2.1 The Precise Localization Of Region 3 Of The K5 Capsule Gene Cluster

The cosmids pKT274 and pGB110 encode the K1 and K5 antigen gene clusters, respectively (Echarti et al., 1983; Roberts et al., 1986) (Figure 3.1). pGB110 DNA was cleaved with the restriction endonucleases BglII, ClaI, EcoRI and HindIII in single and double digests. The resulting DNA fragments were separated by agarose gel electrophoresis and transferred to a nylon filter by Southern blotting. A 2.7kb HindIII-XhoI fragment, which encodes region 3 of the K1 antigen gene cluster and the end of transposon Tn5, was isolated from cosmid pKT274 and used as a radiolabelled probe to determine the precise location of the K5 region 3 in pGB110, based on the extent of DNA homology between the two regions. After hybridization the filter was washed at 65°C using 0.5X SSC and 0.1% SDS, conditions which should allow sequences with approximately 80% or greater homology to hybridize. The pattern of hybridization is shown in Figure 3.2a. The probe hybridized to the 1.8kb BglII fragment previously shown to be associated with the K5 region 3, but not
to flanking BglII fragments and this indicated that region 3 does not extend beyond this 1.8kb fragment (Figure 3.1). Furthermore, since the probe hybridized to the 1.5kb EcoRI-BglII fragment, but not to the 0.3kb BglII-EcoRI fragment, it would appear that region 3 does not extend past the EcoRI site. Thus, on the basis of the observed DNA homology with the K1 region 3 probe, region 3 of the K5 capsule gene cluster was localized to the 1.5kb EcoRI-BglII fragment of pGB110 (Figure 3.1).

3.2.2 Cloning And Nucleotide Sequence Analysis Of Region 3

To facilitate the further characterization of region 3, the 1.8kb BglII fragment was subcloned from pGB110 into the BamHI site of the vector pUC19. Cleavage of pGB110 with the restriction endonuclease BglII generated several fragments and positive recombinant clones were therefore identified by colony hybridization using the 2.7kb HindIII-Xhol fragment K1 region 3 probe (Figure 3.2b). One such recombinant, termed pAS1, was chosen for further study and the nucleotide sequence of the 1.8kb insert present in pAS1 was determined on both strands as follows. First, a detailed restriction endonuclease cleavage map of the insert was generated (Figure 3.3). Then a series of overlapping DNA fragments were subcloned from pAS1 into the vectors Ml3mp18 and Ml3mp19 and the single stranded DNA templates prepared from these were sequenced using the M13 -40 primer and the dideoxy chain-termination method (Sanger et al., 1977). Where necessary oligonucleotide primers were used to complete the sequence on both strands. The DNA sequencing strategy used can be seen in Figure 3.3. The nucleotide sequence of the 1.8kb BglII fragment was analysed using the University of Wisconsin Genetics Computer Group (GCG) sequence analysis software (Devereux et al., 1984) on the VAX cluster at the University of Leicester and the University of Oxford.

The BglII fragment, which encodes region 3 of the K5 antigen gene cluster, was found to be 1864bp in length and analysis of the DNA sequence revealed the presence of two open reading frames, ORF1 and ORF2 (Smith et al., 1990) (Figure 3.4). If the rare translation initiation codon GUG is taken into account,
ORF1 starts at coordinate 75, is 876bp long and encodes a putative polypeptide of 292 amino acids with a predicted molecular weight of 33,278. However, if the typical initiation codon AUG is used, ORF1 starts at coordinate 177, is 774bp long and encodes a polypeptide of 258 amino acids with a predicted molecular weight of 29,526. Putative Shine-Dalgarno ribosome binding sites (Shine and Dalgarno, 1974) have been identified 10bp 5' to both potential initiation codons (Figure 3.4). However, since the AUG codon is used to initiate prokaryotic translation in approximately 90% of cases (McCarthy and Gualerzi, 1990), the open reading frame that starts with AUG at coordinate 177 was used in the computer analysis of ORF1. The UGA stop codon of ORF1 (coordinate 951) overlaps with the AUG initiation codon of ORF2 (coordinate 950) (Figure 3.4), which suggests that ORF1 and ORF2 are organized in a single transcriptional unit. ORF2 (coordinates 950 to 1622) is 672bp long and encodes a putative polypeptide of 224 amino acids with a predicted molecular weight of 25,452. A potential Shine-Dalgarno ribosome binding site has been identified 13bp 5' to the initiation codon of ORF2 (Figure 3.4). The 176bp of DNA sequence determined upstream of the start of ORF1 were examined for the presence of potential E. coli -35 and -10 promoter sequences. Unfortunately no such sequences were identified in this region and this may indicate that the region 3 promoter is located upstream of the available sequence.

Interestingly, on the basis of the observed DNA homology with the K1 region 3 probe, the DNA hybridization experiment indicated that region 3 of the K5 capsule gene cluster does not extend beyond the EcoRI site at coordinate 1498 (Section 3.2.1). However, on examination of the nucleotide sequence of the K5 region 3, it can be seen that ORF2 extends 118bp beyond this EcoRI site (Figure 3.4). The reason for this discrepancy was only identified once the sequence of region 3 of the K1 antigen gene cluster became available (Pavelka et al., 1991) and this will be discussed in Section 3.2.3.
3.2.3 Comparison Of Region 3 From The K1 And K5 Antigen Gene Clusters

The nucleotide sequence of region 3 of the E. coli K1 antigen gene cluster became available during the undertaking of this project (Pavelka et al., 1991). As expected two open reading frames were identified and these have been named kpsM and kpsT. In accordance with their terminology, ORF1 and ORF2, encoded by region 3 of the K5 capsule locus, have been named kpsM and kpsT, respectively. To distinguish between the genes in the different kps loci they will be referred to in the following manner: kpsM\textsubscript{K1}, kpsM\textsubscript{K5} and so on.

DNA hybridization and complementation studies have shown that region 3 is conserved between different group II K antigen gene clusters. In agreement with this, kpsM\textsubscript{K1} and kpsT\textsubscript{K1} appear to be organized in a single transcriptional unit, in the same manner as the K5 genes, with the UGA stop codon of kpsM\textsubscript{K1} overlapping with the AUG initiation codon of kpsT\textsubscript{K1} (Pavelka et al., 1991). However, when the nucleotide sequence of the K1 genes was compared to that of the K5 genes, using the GCG program GAP, it was found that although kpsM\textsubscript{K1} and kpsM\textsubscript{K5} are 94.9% identical, kpsT\textsubscript{K1} and kpsT\textsubscript{K5} are only 69.4% identical. At the amino acid level, KpsM\textsubscript{K1} and KpsM\textsubscript{K5} are 97.7% identical and 99.2% similar, whereas KpsT\textsubscript{K1} and KpsT\textsubscript{K5} have only 72.6% identity and 84.0% similarity (Figure 3.5). Furthermore, the putative KpsT\textsubscript{K5} protein is 5 amino acids longer at the carboxyl terminus than the predicted product of the K1 gene. Thus, it can be seen that the high level of DNA homology thought to exist between the K1 and K5 region 3 does not extend across the whole region and this will be discussed further in Chapter 5. This finding explains the earlier discrepancy regarding the Southern hybridization experiment and the nucleotide sequence analysis of the K5 region 3, since the level of homology which exists between kpsT\textsubscript{K1} and kpsT\textsubscript{K5} beyond the EcoRI site would not allow hybridization to occur under the washing conditions used in the experiment (Sections 3.2.1 and 3.2.2).
3.2.4 KpsM And KpsT Are Members Of The ABC Family Of Transporters

3.2.4.1 An Introduction To The ABC Family Of Transporters

The designation ABC refers to the highly conserved ATP-binding cassette, which is the most characteristic feature of this superfamily (Higgins et al., 1990b; Hyde et al., 1990). These transporters have also been referred to as traffic ATPases (Ames et al., 1990). Over 50 ABC transporters have been identified and although the majority have been found in prokaryotic species, an increasing number of eukaryotic examples are being reported. The majority of ABC transporters are proposed to utilize the energy of ATP hydrolysis to transport substrate across the (inner) membrane against a concentration gradient. Each transport system appears to be relatively specific for a single substrate (or group of related substrates). However, the variety of substrates handled by ABC transporters appears to be vast and includes amino acids, peptides, proteins, inorganic ions, vitamins, sugars and polysaccharides. Some ABC transporters are involved in the uptake of substrate into the cell, whereas others facilitate the export of substrate from the cell.

Typically, ABC transporters consist of four (inner) membrane-associated domains (Higgins, 1992) (Figure 3.6). Two of these domains are highly hydrophobic, integral membrane components, which are proposed to bind the substrate and mediate its translocation across the membrane, while the other two domains, which are located at the cytoplasmic face of the membrane, bind ATP and couple ATP hydrolysis to the transport process. In many prokaryotic ABC transporters the four domains are present as separate polypeptides, for example the oligopeptide permease in *S. enterica* serovar Typhimurium (Hiles et al., 1987) (Figure 3.6). However, there are also numerous examples where the domains are fused into larger, multifunctional proteins, especially in eukaryotes (Figure 3.6). In some cases, an ABC transporter may lack a full complement of domains; for example, the operon encoding the histidine transporter of *S. enterica* serovar Typhimurium includes only a single gene encoding an
ATP-binding component, HisP (Higgins et al., 1982). However, HisP has been shown to exist in a 2:1 ratio with the other domains in the transport complex and therefore probably functions as a homodimer (Kerppola et al., 1991) (Figure 3.6).

The ATP-binding domains of ABC transporters share considerable sequence identity (around 30% or more) over a cassette of about 200 amino acids (Higgins et al., 1986; Hyde et al., 1990). This region of homology includes the two short motifs, G--G-GKS/T and R/K---G---XXXXD (X represents a hydrophobic amino acid residue), which are often referred to as the Walker motifs. These motifs are found in many adenine nucleotide binding proteins and appear to contribute to the formation of a nucleotide binding fold (Higgins et al., 1985, 1986; Walker et al., 1982). However, since the region of sequence homology extends beyond the two Walker motifs, the ATP-binding domains of the ABC transporters can be clearly distinguished from other nucleotide binding proteins which generally share no additional homology outside of these motifs. Proteins containing these conserved ATP-binding cassettes are referred to as ABC proteins (Hyde et al., 1990) and are presumed to share a common evolutionary origin (Higgins et al., 1990b).

The ATP-binding domains are hydrophilic and the available data suggest that they are peripherally associated with the cytoplasmic face of the membrane, in a complex with the transmembrane components (Higgins, 1992). Structural models of the ATP-binding cassettes have been generated (Hyde et al., 1990; Mimura et al., 1991). These predict that the ABC proteins have a tightly folded core structure that binds and hydrolyzes ATP. Loops which extend from this core are proposed to interact with the transmembrane components and couple, presumably via conformational changes, the energy of ATP hydrolysis to the transport process (Higgins, 1992). Although it has been demonstrated that ATP hydrolysis occurs concomitantly with transport (Bishop et al., 1989; Mimmack et al., 1989), the stoichiometry of ATP hydrolysis has not been firmly established. However, a stoichiometry of close to two molecules of ATP hydrolyzed per molecule of substrate transported has been reported (Mimmack et al., 1989) and this is consistent
with the presence of two ATP-binding domains per transporter.

The two integral membrane components of ABC transporters are highly hydrophobic and each is generally predicted to consist of six potential membrane spanning α-helices, separated by short stretches of hydrophilic sequence. The amino and carboxyl termini are predicted to lie at the cytoplasmic face of the membrane. The membrane topology of OppB and OppC, the integral membrane components of the oligopeptide permease of *S. enterica* serovar Typhimurium, has been determined experimentally and confirms these predictions (Pearce *et al.*, 1992). However, some ABC transporters do not conform to this two-times-six transmembrane α-helix pattern, for example the histidine transporter of *S. enterica* serovar Typhimurium. In this case, the two transmembrane components, HisQ and HisM, have each been shown to have only five transmembrane segments and this places the amino terminus at the periplasmic face of the membrane (Kerppola and Ames, 1992). Thus, two-times-five transmembrane segments may provide the minimal unit required to form the translocation pathway (Higgins, 1992).

Comparison of the amino acid sequences of the two transmembrane components of a single ABC transporter generally reveals significant sequence similarity. However, unlike the ATP-binding domains, the transmembrane domains of different transporters typically have little or no significant sequence similarity, even though they seem to be structurally related. This is perhaps not surprising, since different ABC transporters handle different substrates and the transmembrane domains are proposed to play an important role in determining substrate specificity (Higgins, 1992). Thus, the transmembrane domains may have arisen from a common ancestor which has subsequently undergone evolutionary divergence to accommodate different substrates, whilst at the same time maintaining the structural constraints required for function of the domains. Notwithstanding the lack of general sequence conservation, a short conserved sequence motif has been identified in the transmembrane components of a number of bacterial ABC transporters (Dassa and Hofnung, 1985). This is located on the cytoplasmic loop between transmembrane segments four and five.
and is appropriately positioned to interact with the ATP-binding domains (Kerppola and Ames, 1992; Pearce et al., 1992).

It is not known whether this motif is present in the transmembrane components of all ABC transporters, since sequences homologous to the motif are not always immediately obvious, especially in eukaryotic systems (Kerppola and Ames, 1992).

The molecular mechanism by which ABC transporters facilitate transport of the substrate across the membrane is not fully understood. However, a model has been put forward (Higgins, 1992) whereby the substrate interacts with its binding site located within a pore-like translocation pathway formed by the transmembrane domains. Subsequently, the hydrolysis of ATP by the ATP-binding domains induces a conformational change which results in the reorientation of the substrate binding site at the opposite face of the membrane and release of the substrate. Alternatively, reorientation of the binding site may be facilitated by substrate binding, with the hydrolysis of ATP being required to reset the system.

3.2.4.2 KpsM And KpsT Have Properties In Common With ABC Transporters

Analysis of the predicted amino acid sequence of the KpsT<sub>15</sub> protein revealed the presence of potential adenine nucleotide binding fold sequences (Walker et al., 1982) (Figure 3.7). These motifs are also found in KpsT<sub>15</sub> and to determine whether ATP-binding is important for the function of KpsT<sub>15</sub>, Pavelka et al. (1991) used site-directed mutagenesis to change the conserved lysine residue at position 44 (part of the first motif, see Figure 3.7) to a glutamic acid residue. In contrast to a cloned wild type kpsT<sub>15</sub> gene, the mutated gene was unable to complement the mutation present in a KpsT<sub>15</sub><sup>−</sup> strain. This indicated that the proposed adenine nucleotide binding fold plays an essential role in KpsT<sub>15</sub> activity (Pavelka et al., 1991). Furthermore, when a search was made against the available protein sequence databases using the GCG program FASTA, KpsT<sub>15</sub> was found to have significant amino acid homology with a number of ABC proteins. In many cases, sequence identities of around 23-30% were found to extend over a region
of approximately 200 amino acids corresponding to the conserved ATP-binding cassette (two proteins showing greater homology with KpsT were also identified and these will be described in Section 3.2.5). Thus, KpsT appears to have the characteristics of a typical ABC protein/domain. A hydropathy plot, calculated by the method of Kyte and Doolittle (1982) using a 15-residue window, showed KpsT to be a relatively hydrophilic protein (Figure 3.8a). Similarly, the hydropathy profiles of other ABC proteins/domains have been found to be hydrophilic.

The hydropathy profile of KpsM is characteristic of a highly hydrophobic integral membrane protein, with six potential membrane spanning domains (Figure 3.8b). In agreement with this, preliminary Tn5phoA fusion analysis (Manoil and Beckwith, 1985), carried out by Pavelka et al. (1991), indicated that at least two regions of the virtually identical KpsM protein are exposed to the periplasm. Thus, KpsM has properties in common with the transmembrane components of ABC transporters. As expected, a search of the protein sequence databases showed that KpsM has no obvious homology with the transmembrane domains of other ABC transporters (exceptions to this will be discussed in Section 3.2.5). Unfortunately, a sequence homologous to the conserved sequence motif described by Dassa and Hofnung (1985) has not been identified in KpsM. However, this motif may not be present in all ABC transporters.

Region 3 of the kps locus has been tentatively proposed to play a role in the energy dependent export of capsular polysaccharide across the inner membrane (Boulnois and Jann, 1989). Thus, since KpsM and KpsT appear to have many characteristics in common with ABC transporters, it has been proposed that these region 3 proteins constitute an inner-membrane capsular polysaccharide export complex (Smith et al., 1990). The KpsT protein may couple ATP-hydrolysis to the KpsM mediated transport of polysaccharide across the inner membrane. The archetypal ABC transporter appears to require an inner-membrane complex consisting of two ATP-binding domains and two transmembrane domains. To fulﬁl these criteria, the KpsM and KpsT proteins would each need to be present as a homodimer in the putative export complex.
3.2.5 Identification Of A Common Mechanism For Polysaccharide Export In Bacteria Expressing Group II Capsules

Region 1 of the capsule locus (cap) of *H. influenzae* and region C of the capsule locus (cps) of *N. meningitidis* have both been proposed to be essential for polysaccharide export (Frosch et al., 1989; Kroll et al., 1988, 1989). Thus, the role these regions play in capsule biogenesis may be similar to that of region 3 of the *E. coli* kps locus. During the undertaking of this project the nucleotide sequence of these regions became available. Four genes, *bexDCBA*, were identified in region 1 of the *H. influenzae* cap locus (Kroll et al., 1990) and similarly four genes, *ctrABCD*, were found in region C of the cps locus of *N. meningitidis* (Frosch et al., 1991) (Figure 1.2). Analysis of the putative proteins encoded by *bexD*, *bexC*, *ctrA* and *ctrB* has shown that they have no homology with KpsM and KpsT and these proteins have been briefly described in Section 1.5.2. The remaining proteins, *BexB*, *BexA*, *CtrC* and *CtrD*, will be discussed below.

The *bexA* and *ctrD* genes encode putative polypeptides with molecular masses of 24.7kDa and 24.6kDa, respectively (Frosch et al., 1991; Kroll et al., 1990). Analysis of the predicted amino acid sequence of *BexA* and *CtrD* revealed the presence of the two short motifs associated with the formation of an adenine nucleotide binding fold (Walker et al., 1982) (Figure 3.7). This indicated that *BexA* and *CtrD* may have characteristics in common with KpsT. A comparison of the amino acid sequences of *BexA*, *CtrD* and KpsT was carried out using the GCG program GAP and the results of this are shown in Table 3.1a. A multiple alignment of these proteins was generated using the GCG programs FILEUP and PRETTYPLOT and this is shown in Figure 3.9a. KpsT, BexA and CtrD appear to have extensive sequence identity which extends along the full length of the proteins, with BexA and CtrD being more closely related to each other than to KpsT. This suggests that BexA and CtrD are further examples of typical ABC proteins/domains. Interestingly, the adenine nucleotide binding fold motifs are found in approximately the same position in each of the three proteins.
and the two cysteine residues of KpsT align exactly with cysteine residues present in BexA and CtrD. Also the hydropathy profiles of these three proteins are very similar (Figure 3.8a). Thus, KpsT, BexA and CtrD not only have extensive primary sequence homology, but may also have similar secondary and tertiary structures.

The putative polypeptides encoded by the \textit{bexB} and \textit{ctrC} genes have molecular masses of 30.2kDa and 30.1kDa, respectively (Frosch et al., 1991; Kroll et al., 1990). Analysis of the predicted amino acid sequence of BexB and CtrC has revealed that the hydropathy profiles of the two proteins are almost identical to the profile of the KpsM\textsubscript{Ecoli} protein (Figure 3.8b). Furthermore, \textit{Tn5}\textit{phoA} mutagenesis carried out by Frosch et al. (1991) has shown that the CtrC protein, like KpsM\textsubscript{Ecoli} (Pavelka et al., 1991), has domains exposed to the periplasm. Cell fractionation and western blot analysis of the \textit{phoA} fusion proteins revealed CtrC to be associated with the inner membrane. Thus, like KpsM, BexB and CtrC have characteristics in common with the transmembrane components of ABC transporters. The predicted amino acid sequences of the KpsM\textsubscript{Ecoli}, BexB and CtrC proteins were compared and the results are shown in Table 3.1b. A multiple alignment of these proteins is shown in Figure 3.9b. BexB appears to have extensive sequence identity with CtrC. However, with regard to identical amino acids, the homology these two proteins have with KpsM is rather more limited, although if conservative substitutions are taken into account the homology becomes significantly more extensive. In each case, the homology extends along the full length of the proteins.

Thus, the BexA and BexB proteins of \textit{H. influenzae} and the CtrD and CtrC proteins of \textit{N. meningitidis} appear to have homology with the KpsT and KpsM proteins of \textit{E. coli}. This suggests that their role in capsule biogenesis may be similar to that of the \textit{E. coli} proteins and it has been proposed that they too constitute an inner-membrane capsular polysaccharide export complex, wherein the BexA/CtrD protein couples ATP-hydrolysis to the BexB/CtrC mediated translocation of polysaccharide across the inner membrane (Frosch et al., 1991; Kroll et al., 1992).
Therefore it appears that a common mechanism may exist for the energy-dependent translocation of capsular polysaccharide across the inner membrane of distantly related bacterial species expressing group II capsules.

As discussed above, these group II capsular polysaccharide export systems have characteristics in common with members of the ABC transporter family. To fulfill the requirements of an archetypal ABC transporter, the component proteins would need to be present as homodimers in a tetrapeptide inner-membrane complex. Interestingly, Frosch et al. (1991) have suggested that the CtrB protein (Section 1.5.2) is also a transmembrane component of the complex, indicating that it forms a heterodimer with CtrC. In agreement with this, TnSphoA mutagenesis has shown that CtrB has periplasmic domains and is associated with the inner membrane (Frosch et al., 1991). However, the hydropathy profiles of CtrB and the homologous proteins BexC and KpsE (Section 1.4.4.1) are not characteristic of an integral membrane protein. In fact, these proteins only have two potential membrane spanning domains, which are located near their amino and carboxyl termini. Furthermore, the polysaccharide synthesized by a kpsE+ E. coli strain was found to be located in the periplasm (Bronner, unpublished results; Pazzani, 1992; Section 1.4.4.1), which suggests that the KpsE protein functions in the translocation of polymer from the periplasm to the cell surface, rather than across the inner membrane. Unfortunately, it is not yet known whether CtrB has a similar role in capsule biogenesis, but it would appear that the protein is unlikely to form a transmembrane heterodimer with CtrC in a tetrapeptide inner-membrane capsular polysaccharide export complex.

3.2.6 Visualization Of The KpsM And KpsT Proteins

The visualization of a protein, of an appropriate molecular mass, provides further evidence for the existence of an open reading frame which has been predicted by nucleotide sequence data. In order to visualize both the KpsM protein and the KpsT protein, a number of strategies had to be employed and these will be described below.
3.2.6.1 Overexpression Experiments Using The Vector pKK223-3

The vector pKK223-3 contains the strong tac promoter, a hybrid trp-lac promoter which is regulated by the lac repressor. Genes cloned into the multiple cloning site (MCS) of this vector can be expressed at a high level upon induction of the promoter using isopropyl β-D-thiogalactoside (IPTG). In order to ensure tight regulation of the promoter, the vector can be used in conjunction with the plasmid pMW1 (provided by I.S. Roberts). This plasmid encodes the LacI? repressor and was constructed by cloning a 1.7kb EcoRI fragment containing the lacI? gene from pMC9 (Calos et al., 1983) into the vector pACYC184.

Initially, the 1.8kb BgIII fragment from pGB110, encoding both KpsM and KpsT, was cloned into the vector pKK223-3. This was carried out as follows. The 1.8kb BgIII fragment was cut out of the plasmid pAS1, using the restriction endonucleases PstI and SmaI, and subsequently cloned into pKK223-3, cleaved with PstI and SmaI (Figure 3.10). Positive recombinant clones were identified by colony hybridization using the 1.8kb BgIII fragment as a probe, since α-complementation cannot be used with the pKK223-3 vector. One such recombinant, termed pAS2, was chosen for further study.

pAS2 was transformed into E. coli LE392 harbouring the plasmid pMW1. LE392 harbouring non-recombinant pKK223-3 and pMW1 provided a negative control. Denaturing polyacrylamide gel electrophoresis (SDS-PAGE) was used to analyse E. coli lysates for the IPTG-induced expression of proteins. Samples taken at 0 mins, 30 mins, 60 mins and 90 mins after the addition of IPTG were compared to samples taken from cultures which had not been treated with IPTG. Figure 3.11a shows total E. coli proteins, from the samples taken at 90 mins, after SDS-PAGE. An induced protein with a molecular mass of approximately 26.5kDa can be seen only in the sample prepared from LE392(pAS2 and pMW1) treated with IPTG. This corresponds well to the molecular mass of 25.5kDa predicted from the amino acid sequence of the KpsT protein. A protein with a molecular mass of 29.5kDa corresponding to that predicted for the KpsM protein was not
detected. It is possible that such a protein was masked by other *E. coli* proteins of the same molecular mass. Alternatively, the KpsM protein may be expressed at extremely low levels. Interestingly, Hiles and Higgins (1986) have reported difficulty in detecting the expression of other integral membrane proteins.

In order to try to visualize both KpsM and KpsT, it was decided that a more sensitive technique should be used. One such technique involves the isolation of *E. coli* minicells and this will be described below.

### 3.2.6.2 Minicell Analysis

The *E. coli* strain DS410 produces spherical anucleate minicells, which can be isolated from the normal vegetative cells. These minicells often harbour plasmid DNA and any genes carried on the plasmid can be expressed in the minicells in the presence of ^35S methionine and the radiolabelled proteins detected by SDS-PAGE and autoradiography.

Minicells were isolated from *E. coli* DS410 harbouring pAS2 and pMW1. Minicells from DS410(pKK223-3), DS410(pMW1) and DS410 (pKK223-3 and pMW1) were used to distinguish between vector and non-vector encoded proteins. Labelling of the expressed proteins with ^35S methionine was carried out both with and without induction of the tac promoter using IPTG. However, it must be noted that the LacI⁹ repressor encoded by pMW1 was unable to ensure total repression of the tac promoter in the absence of IPTG. Furthermore, subsequent induction of the promoter with IPTG only led to a small increase in the level of protein expression (Figure 3.11b). In Figure 3.11b, pAS2 can be seen to encode a non-vector protein with an approximate molecular mass of 27kDa. This is in good agreement with the observations made in the previous overexpression experiment (Section 3.2.6.1) and the size of the protein is consistent with the predicted molecular mass of the KpsT protein. As before, a protein with a molecular mass similar to that predicted for KpsM was not detected. However, the possibility that the observed protein was in fact KpsM, running on the SDS-
PAGE gel with a faster mobility than expected, could not be discounted. Hydrophobic proteins frequently migrate anomalously fast during SDS-PAGE (Pearce et al., 1992). Therefore, in order to determine whether the 27kDa protein corresponded to KpsM or KpsT, it was decided that the kpsM gene and the kpsT gene should be cloned separately into pKK223-3 and then subjected to minicell analysis.

In order to clone kpsM into pKK223-3, pAS1 was cut with the restriction endonuclease Rsal, generating a 0.8kb fragment which spans the kpsM gene (Figure 3.10). This fragment was subsequently cloned into pKK223-3 cleaved with SmaI. Positive recombinant clones, with the insert in the correct orientation, were identified by restriction endonuclease cleavage of small scale plasmid DNA preparations and one such recombinant, termed pAS3, was chosen for further study. In the case of the kpsT gene, pAS1 was cleaved with the restriction endonuclease SspI, which cuts 294bp upstream of kpsT and HindIII, which cuts in the pUC19 multiple cloning site immediately outside of the 1.8kb region 3 insert. This generated a 1.2kb fragment spanning the kpsT gene (Figure 3.10), which was cloned into pKK223-3 cleaved with SmaI and HindIII. Positive recombinant clones were identified as above and one, pAS4, was chosen for further study.

Minicells were isolated from DS410(pKK223-3 and pMW1), DS410 (pAS2 and pMW1), DS410(pAS3 and pMW1) and DS410(pAS4 and pMW1). Labelling of the expressed proteins with 35S methionine was carried out both with and without IPTG induction of the tac promoter. However as before, repression of the promoter by the LacI repressor encoded by pMW1 was not absolute and its subsequent induction with IPTG resulted in only a modest increase in expression of the encoded protein(s) (Figure 3.11c). In Figure 3.11c, it can be seen that pAS4 encodes a non-vector protein with an approximate molecular mass of 27kDa. This corresponds well to the protein identified in the earlier overexpression and minicell studies and since the insert in pAS4 contains only the kpsT gene, the 27kDa protein can be finally identified as the KpsT protein.
A non-vector protein of approximately 24kDa can be seen to be very weakly expressed by pAS3 (Figure 3.11c) and since only the kpsM\textsubscript{k3} gene is present in the insert in this plasmid, the observed protein must be KpsM\textsubscript{k3}. Thus, KpsM\textsubscript{k3} appears to have a faster mobility on SDS-PAGE gels than would be expected from its predicted molecular mass of 29,526. Alternatively, translation of KpsM may initiate from the second methionine codon at position 160 in the nucleotide sequence. This would result in synthesis of a protein with a predicted molecular mass of approximately 23.5kDa, which is close to the value estimated from SDS-PAGE. However, the tendency for highly hydrophobic proteins to migrate anomalously fast in SDS-PAGE gels would suggest that KpsM is more likely to be synthesized as a 29.5kDa protein. Ideally, in order to ascertain which AUG codon is used to initiate KpsM synthesis, the protein should be purified and its N-terminal amino acid sequence determined.

As one would expect pAS2, which contains both kpsM\textsubscript{k3} and kpsT\textsubscript{k3}, can be seen to express two non-vector proteins, with molecular masses of approximately 24kDa and 27kDa (Figure 3.11c). The band at 24kDa corresponding to the KpsM protein is rather faint and in the first minicell experiment this band was initially overlooked (Figure 3.11b). However, with hindsight and further study of the autoradiograph in Figure 3.11b, an extremely faint band can in fact be seen at this position. In the first minicell experiment the ^{35}S labelled proteins were run out on a large 11\% SDS-polyacrylamide gel, whereas in the later experiment the labelled proteins were run on a small 12.5\% gel using the Bio-Rad mini protein II system. The resolution of the protein bands was greatly improved in the latter case.

Thus in conclusion, the KpsM and KpsT proteins, encoded by region 3 of the E. coli K5 kps locus, were visualized as proteins with apparent molecular masses of 24kDa and 27kDa, respectively. Similar findings have been reported for the KpsM\textsubscript{k1} and KpsT\textsubscript{k1} proteins (Pavelka et al., 1991). The E. coli K1 region 3 proteins were studied using an in vitro transcription-translation system and two proteins with apparent molecular masses of 24.2kDa and 25.8kDa were observed.
3.2.7 Preliminary Studies Of The Region 3 Promoter

The expression of *E. coli* group II capsular polysaccharides is subject to temperature regulation with the capsule being expressed at 37°C, but not at 18°C (Section 1.3.1.1). Whether such regulation is mediated at the level of transcription or translation poses an interesting question. In order to try to address this, a preliminary investigation of the *E. coli* K5 region 3 promoter was carried out.

The vector pKK232-8 contains a promoterless chloramphenicol acetyl transferase gene (*cat*). When DNA fragments containing a promoter are cloned into the upstream multiple cloning site, the *cat* gene is expressed and the host strain becomes resistant to chloramphenicol. It was decided that this vector should be used to analyse putative region 3 promoter sequences. Firstly, DNA fragments cloned into pKK232-8 could be examined for promoter activity by determining the minimum concentration of chloramphenicol required to inhibit growth of the host bacterium. Secondly, in order to determine whether transcription from a potential promoter is subject to regulation by growth temperature, the level of *Cat* activity could be assayed after growth of the bacteria at both 37°C and 18°C. Such studies are described below.

As discussed in Section 3.2.2, there are no obvious candidate *E. coli* -35 and -10 promoter sequences present in the 176bp of DNA sequence immediately upstream of the AUG translational start codon of *kpsM^K5_. Furthermore, preliminary findings reported by Pavelka *et al.* (1991) suggest that *kpsM^K5_ and *kpsT^K5_ are transcribed from a promoter located 743bp upstream of the putative initiation codon of *kpsM^K5_. Thus, since region 3 in *E. coli* K1 and K5 appears to be highly conserved, it may be expected that the K5 region 3 promoter is located in a similar position. Therefore, in order to ensure that a DNA fragment cloned into pKK232-8 contains the K5 region 3 promoter, it may be necessary to include DNA up to at least 800-900bp upstream of *kpsM^K5_.

98
A number of Tn5phoA insertions in the cosmid pGB110 (encodes the \textit{E. coli} K5 \textit{kps} locus) have been generated by C. Pazzani (Pazzani, 1992). Interestingly, the Tn5phoA insertion in pGB110::34, which maps approximately 900bp upstream of the \textit{kpsM} gene, resulted in the host strain having a wild type K5 phenotype (Bronner et al., 1993a). This was detected using K5 specific bacteriophage and immunoelectron microscopy. These results imply that the insertion is located outside of the \textit{kps} cluster and consequently, the region 3 promoter must lie downstream of the insertion site. Thus, the Tn5phoA insertion in pGB110::34 provided a source of restriction endonuclease cleavage sites, which could be used to facilitate the cloning of the whole of region 3 including potential promoter sequences.

The cosmid pGB110::34 was cleaved with the restriction endonucleases HindIII (which cuts within ISS50 of Tn5phoA) and StuI (which cuts downstream of region 3) to generate a 4.2kb fragment spanning region 3 (Figure 3.12). This was cloned into the vector pUC18 cleaved with HindIII and HincII, generating the plasmid pAS7. The region upstream of the \textit{kpsM} gene in pAS7 was mapped for the presence of restriction endonuclease cleavage sites suitable for use in cloning putative promoter sequences into pKK232-8. The enzyme \textit{SspI} was found to cut near the start of the partial ISS50 element and also 400bp downstream of the initiation codon of \textit{kpsM}, generating a 1.55kb fragment (Figure 3.12). This fragment was initially cloned into pUC19 cut with HincII. After checking for the desired orientation using restriction analysis, the fragment was cut out of pUC19 using BamHI and HindIII and cloned into pKK232-8 cleaved with BamHI and HindIII, generating plasmid pAS10. Thus, pAS10 contained putative region 3 promoter sequences cloned upstream of the \textit{cat} gene. In addition, the construct also contained the first 400bp of \textit{kpsM}. However, since the vector pKK232-8 contains stop codons, in all three reading frames, between the MCS and the initiation codon of the \textit{cat} gene, translational read-through from \textit{kpsM} into the \textit{cat} gene would be prohibited.

In order to determine whether the fragment cloned in pAS10
contained promoter sequences, LE392(pKK232-8) and LE392(pAS10) were incubated at 37°C in the presence of concentrations of chloramphenicol ranging from 0μg/ml to 90μg/ml. LE392 harbouring the plasmid pACYC184, which contains a constitutive cat gene, were used as a positive control. The results are shown in Table 3.2. As expected, LE392 harbouring pKK232-8 were unable to grow in the presence of chloramphenicol. However, the minimum concentration of chloramphenicol required to inhibit growth of LE392(pAS10) was between 45μg/ml and 60μg/ml. Thus, the fragment cloned in pAS10 appears to have promoter-like activity, which suggests that it may contain the K5 region 3 promoter.

To try to determine whether the putative region 3 promoter in pAS10 was subject to regulation by growth temperature, a spectrophotometric assay was used to measure CAT activity after growth of the bacteria at both 37°C and 18°C. In order to calculate specific CAT activity in units of CAT activity/mg total cell protein, the total protein concentration in mg/ml was determined for each sample. The assay was carried out using LE392 harbouring either pKK232-8, pAS10 or pBR328 (a plasmid containing a constitutive cat gene, used as a positive control). The CAT activity values obtained in the assay for LE392(pAS10) grown at 37°C were approximately 50-100 fold lower than those seen with the positive control LE392(pBR328) and were only just above background levels. After growth at 18°C, the LE392(pAS10) values were even lower and seemed comparable to those seen with the sample blank and negative control LE392(pKK232-8). These preliminary findings suggest that the putative region 3 promoter present in pAS10 is subject to regulation by growth temperature. However, these results must be interpreted with caution, since the CAT activity values obtained with LE392(pAS10) grown at 37°C were only marginally higher than those observed after growth at 18°C and in addition, the measurements of total cell protein were inconsistent.
3.3 DISCUSSION

Region 3 of the *E. coli* K5 kps locus is approximately 1.8kb in size and is composed of two genes, *kpsM* and *kpsT*, possibly organized in a single transcriptional unit (Smith et al., 1990). The encoded KpsM and KpsT proteins were visualized, using minicells and SDS-PAGE, as proteins with apparent molecular masses of 24kDa and 27kDa, respectively. The size observed for the KpsT protein is consistent with the molecular mass of 25.5kDa predicted from its nucleotide sequence. In contrast, the KpsM protein appears to have a much faster mobility in SDS-polyacrylamide gels than would be predicted from its calculated molecular mass of 29.5kDa. However, this was not totally unexpected, since KpsM has the characteristics of a highly hydrophobic integral membrane protein and such proteins frequently migrate anomalously fast during SDS-PAGE. As one would expect, similar observations were made for the virtually identical KpsMgi protein (Pavelka et al., 1991; Pigeon and Silver, 1994).

As a result of previous DNA hybridization studies, it has been proposed that extensive DNA homology (approximately 95%) exists between the region 3 genes of different group II K antigen gene clusters (Roberts et al., 1988a). However, when the nucleotide sequence of the *kpsM*k and *kpsTk* genes (Pavelka et al., 1991) was compared with that of the K5 region 3 genes, it was found that although *kpsMk* and *kpsMg* have an expected DNA homology of 94.9%, the *kpsTk* and *kpsTk* genes only have a DNA homology of 69.4%. As would be expected, the level of DNA homology is reflected in the level of homology observed between the deduced amino acid sequences of the encoded proteins. Thus, whilst the KpsMk and KpsMg proteins are almost identical, the KpsTk and KpsTk proteins have only 72.6% identity and 84.0% similarity. Nevertheless, despite this, the KpsTk and KpsTk proteins are known to be functionally interchangeable (Roberts et al., 1988a). Unfortunately, the reason for the observed sequence divergence in the *kpsTk* and *kpsTk* genes is not known, but one possible explanation is discussed in Chapter 5.

A preliminary characterization of the KpsMk and KpsTk
proteins has been undertaken and this has shown that they have many characteristics in common with members of the ABC family of membrane transporters. These findings support the earlier proposal that region 3 encodes the functions necessary for the energy-dependent translocation of capsular polysaccharide across the inner membrane (Boulnois et al., 1989; Kroncke et al., 1990a). The hydropathy profile of KpsM suggests that it is an integral membrane protein, whilst KpsT appears to contain a conserved ATP-binding cassette. Thus, the region 3 proteins have been proposed to constitute an inner-membrane capsular polysaccharide export complex, wherein the KpsT protein couples ATP-hydrolysis to the KpsM mediated transport process. In order to fulfil the requirements of a typical ABC transporter, these proteins may need to be present as homodimers in a tetrapeptide export complex. At the present time, the model put forward for the export of capsular polysaccharide across the inner membrane proposes that the polymer passes directly through a translocation pathway formed by the KpsM homodimer. However, there is currently no direct evidence to support this proposal and thus, one cannot exclude the possibility that the KpsMT complex may function indirectly in the export process, by transporting another component that may be critical for polymer export by an as yet unidentified mechanism.

Since the completion of the work described in this thesis, a number of further analyses have been made on the KpsM and KpsT proteins (Pavelka et al., 1994; Pigeon and Silver, 1994). Firstly, the topology of KpsM within the inner membrane has been investigated (Pigeon and Silver, 1994). These experiments have provided evidence for a model of KpsM having six membrane-spanning regions, as predicted from its hydropathy profile, with the N- and C-termini facing the cytoplasm and a short domain within the third periplasmic loop, the SV-SVI linker, also localizing within the membrane. Interestingly, in vivo cross-linking studies have indicated that KpsM forms dimers within the inner membrane, which is in agreement with the earlier proposal that KpsM and KpsT are present as homodimers in the putative KpsMT polysaccharide export complex. Secondly, the interaction of ATP with the KpsT protein has been studied (Pavelka et al., 1994). It has been shown that KpsT can
be photolabelled by an ATP analogue and saturation mutagenesis of the first Walker motif has confirmed its importance in KpsT function. Thus, these experiments provide further support for the view that ATP binding and presumably hydrolysis, although this has not yet been demonstrated, are important for the function of KpsT in capsular polysaccharide export.

The KpsM and KpsT proteins of *E. coli* appear to have significant homology with the BexB and BexA proteins of *H. influenzae* and the CtrC and CtrD proteins of *N. meningitidis*. These proteins have been proposed to comprise an inner-membrane capsular polysaccharide export system that is functionally analogous to the *E. coli* KpsMT complex (Frosch *et al.*, 1991; Kroll *et al.*, 1990). Thus, a conserved mechanism for polysaccharide export appears to exist in bacteria expressing group II capsules and this may indicate that capsule production in these unrelated species has a common evolutionary origin. Interestingly, the Bex and Ctr proteins are much more closely related to each other than to the corresponding *E. coli* proteins, which may suggest that their divergence from a common ancestral sequence has occurred relatively recently. The proteins involved in the export of group II capsular polysaccharides in *E. coli*, *H. influenzae* and *N. meningitidis* have been shown to be very highly conserved between the different serotypes of each species (Kroll *et al.*, 1989; Frosch *et al.*, 1989, 1991; Roberts *et al.*, 1988a). Unfortunately, the means by which these conserved proteins are able to recognize and transport the wide range of chemically diverse capsular polysaccharides expressed by the different serotypes have not yet been elucidated. Perhaps the transport apparatus recognizes a common component of the polysaccharides, such as the phospholipid moiety shown to be present at the reducing end of group II polymers. In agreement with this, Frosch and Müller (1993) have recently proposed that the attachment of the phospholipid moiety to the polysaccharide is an absolute requirement for the transport process. However, this matter has yet to be fully resolved.

The KpsMT, BexBA and CtrCD polysaccharide export systems, together with the drug resistance transporter of *Streptomyces*
peucetius (DrrAB) and two nodulation gene products (NodIJ) of Rhizobium leguminosarum, have been proposed, on the basis of primary structure homologies, to comprise a new subfamily within the ABC family of transporters, termed the ABC-2 subfamily (Reizer et al., 1992). Recently, a number of ABC transporters have been identified which appear to be more closely related to the capsular polysaccharide exporters of the ABC-2 subfamily than to other members of the family. Interestingly, these transporters are also involved in the export of cell surface polysaccharides and include: RfbHI of Vibrio cholerae O1 (Stroehrer et al., 1992), RfbDE of Yersinia enterocolitica O:3 (Zhang et al., 1993), RfbAB of K. pneumonialae O1 (Bronner et al., 1994) and ORF261 ORF431 of E. coli O9 (Kido et al., 1995), which have all been implicated in the transport of O antigen across the inner membrane; and TagGH of Bacillus subtilis 168 (Lazarevic and Karamata, 1995), which has been proposed to participate in the export of teichoic acids. In general, the polymers transported by these related polysaccharide export systems appear to be polymerized at the cytoplasmic face of the inner membrane by a rfc-independent O antigen-like processive mechanism. Thus, there appears to be a correlation between this type of polymerization mechanism and the requirement for an ABC-2 type of polysaccharide transporter. In contrast, polysaccharides synthesized by the rfc-dependent O antigen-like blockwise polymerization of und-P-linked repeating units appear to have no requirement for ABC-2 type transporters.

During this project a preliminary investigation of the E. coli K5 region 3 promoter was carried out. Since the region 3 promoter in E. coli K1 had been proposed to be located 743bp upstream of kpsM (Pavelka et al., 1991), the corresponding region in the K5 kps locus was examined for promoter activity. A fragment including approximately 900bp of DNA located upstream of the kpsM<sub>a</sub> gene, cloned into the promoter probe plasmid pKK232-8, was found to have promoter-like activity. However, in order to be able to identify the precise location of the region 3 promoter, transcript mapping techniques such as primer extension should now be used. In addition, northern blotting should be performed to determine the size of the
region 3 transcript(s) and to confirm that the \textit{kpsM} and \textit{kpsT} genes are in fact organized in a single transcriptional unit, as predicted from their genetic organization. Northern blotting could also be conducted using RNA extracted from bacteria grown at both 18°C and 37°C, in order to determine whether transcription of the \textit{kpsM} and \textit{kpsT} genes is subject to regulation by growth temperature. This may help to confirm the preliminary findings of the CAT assay experiments carried out during this project, which indicate that the promoter of region 3 is regulated by temperature. Such regulation may be expected, since \textit{E. coli} group II capsular polysaccharides are expressed at a growth temperature of 37°C, but not at 18°C.

Since the completion of this project, a number of known global regulators of gene expression in \textit{E. coli} have been studied in order to determine whether they play any role in the regulation of group II capsule expression (Stevens et al., 1994). None of the mutations in the regulatory genes examined were found to induce capsule expression at 18°C. However, mutations affecting RfaH, a regulator of several virulence and fertility genes in \textit{E. coli}, were found to abolish the expression of capsule at 37°C. Interestingly, a 39bp conserved DNA sequence termed \textit{JUMPstart} has been identified 28bp upstream of the initiation codon of the \textit{kpsM}_{1} and \textit{kpsM}_{3} genes (Hobbs and Reeves, 1994). The \textit{JUMPstart} sequence, which is found upstream of a number of polysaccharide biosynthesis gene clusters in enteric bacteria, has been shown to have extensive homology with sequences present in operons known to be regulated by RfaH (Stevens et al., 1994). This may suggest that the \textit{JUMPstart} sequence plays a role in the co-ordinate regulation of these genes by RfaH. Thus, the mechanisms involved in the regulation of \textit{E. coli} group II capsule expression are beginning to be elucidated, but the regulatory pathway is likely to be complex and involve many interacting factors, some of which may also be involved in the regulation of other polysaccharide biosynthetic processes.
Figure 3.1 Physical maps of the recombinant cosmids pKT274 and pGB110 carrying the biosynthesis genes for the Kl and K5 antigens, respectively. The restriction endonuclease cleavage maps were aligned on the basis of several common cleavage sites. The boxes above pKT274 labelled 1-3 refer to the three functional regions involved in the production of the Kl antigen. The open boxes above and below the lines refer to Tn5 and vector sequences, respectively. The Kl region 3 probe taken from pKT274 is represented by the black box. The hashed box corresponds to the 4.7kb EcoRI-HindIII fragment of cosmid pGB118 (a derivative of pGB110) to which the K5 region 3 functions had been previously been localized (Roberts et al., 1988a). The dotted box indicates the location and extent of the DNA homology which exists between region 3 of the Kl and K5 antigen gene clusters and corresponds to a 1.5kb EcoRI-BglII fragment of pGB110. Restriction endonuclease sites: B, BamHI; C, ClaI; E, EcoRI; G, BglII; H, HindIII; X, XhoI. The scale is shown in kb.
Figure 3.2  The localization and cloning of region 3 of the 
E. coli K5 capsule gene cluster

a) Southern hybridization analysis of pGB110 using the Kl region 3 probe shown in Figure 3.1. pGB110 DNA was cleaved with the following restriction endonucleases: Lane 1, BgIII; 2, ClaI; 3, EcoRI; 4, HindIII; 5, BgIII and ClaI; 6, BgIII and EcoRI; 7, BgIII and HindIII; 8, ClaI and EcoRI; 9, ClaI and HindIII; 10, EcoRI and HindIII. The numbers at the left hand side represent the migration of DNA size markers (1kb ladder, GIBCO BRL) in kb. In lane 1, the probe can be seen to hybridize solely to a 1.8kb BgIII fragment and in lane 6, it hybridizes to a single 1.5kb BgIII-EcoRI fragment. These fragments are discussed in the text. The DNA present in lanes 4, 5, 8 and 10 has not been completely digested.

b) The identification of JM101 clones harbouring recombinant pUC19 containing region 3 of the K5 capsule gene cluster, using the technique of colony hybridisation. The colonies were screened using the Kl region 3 probe shown in Figure 3.1. The dark lines indicate positive clones.
Figure 3.2

a)

b)
Figure 3.3 Region 3 of the *E. coli* K5 capsule gene cluster

The physical map of pGB110 carrying the biosynthesis genes for the K5 antigen is shown. The open box below the line indicates vector sequence. The boxes labelled 1-3 refer to the three functional regions involved in the production of the K5 antigen. The broken lines highlight the enlargement (on a scale of 10:1) of the 1.0kb *BglII* fragment previously shown to contain region 3 of the K5 capsule gene cluster (see text). This fragment corresponds to the insert cloned in the plasmid pAS1. The short horizontal lines in brackets labelled Forward and Reverse illustrate the strategy used to sequence region 3. The two open reading frames identified in region 3 are represented by the boxes labelled ORF1 (*kpsM*) and ORF2 (*kpsT*) and the arrow indicates the direction of their transcription. Restriction endonuclease sites: B, BamHI; C, ClaI; E, EcoRI; G, BglII; N, NdeI; F, PvuI; V, EcoRV. The scale is shown in kb.
Figure 3.4 The nucleotide sequence of region 3 of the E. coli K5 capsule gene cluster

**BglII**

```
1 AGAICTTATTATCCCGGAGAATATTTCGCTGAAATTTTTGCGACATATTAAAAAG
60
SD  kpsM
61 GTCCACCGCTGACGTTAATTATACACGTTCCGCGACCTTGATGAGTTGACCCGACGATTGACCT
120
  VNFIIQQLQISITSSSVLV
SD
121 TAGCTTACGACCAAGGCGGCTAGCCGACTGAGAGATTAGTCTGAGAACATCTCAACAAATGG
180
  AVKPRAVAVYILKRLGSHQMA
SD
181 CAGAGATTGCTGATGACTCAGAAGAGCTTACGCTATCGCAATGCGACACAT
240
  RSGFLEVQKVTVEALFLREIR
SD
241 GAACACGCGTTTGCTGATACCGCCATCCGAGTATTTGCGGCTGAGCTGTGATGACCCTCGGC
300
  TRFGKFRLGLYWIALEPSAH
SD
301 ATTTGTCTGACCGCTGATACCGCCATCCGAGTATTTGCGGCTGAGCTGTGATGACCCTCGGC
360
  LLILLGIGIFYIMHRTPDIS
SD
361 CATTCCGGTGTCCCTACCTACTACTTACGCGCTGACTCCCTTTTTTACTTCTTACGCGCTGACT
420
  FPPFLLNLGGLIPFIFSSISN
SD
361 ATCGGCTCTGAGGCGGCTGATACCGCCATCCGAGTATTTGCGGCTGAGCTGTGATGACCCTCGGC
480
  RSVGAIEANQGLYRNYPVKEP
CIal
481 CCGACAGAACCGCTGACCGCCATCCGAGTATTTGCGGCTGAGCTGTGATGACCCTCGGC
540
  IDMIARALLETLYVAYVI
SD
541 TGGTCCTCGGCTGATACCGCCATCCGAGTATTTGCGGCTGAGCTGTGATGACCCTCGGC
600
  LLMLIVWMAEGYFSITSINFLQ
SD
601 AACTACTGCCTGAGCGGGATTTTATTTATAGCAGGGCAATCGCGCTAATACGACTCCCTCGG
660
  LLTVWSELILSCGGILFM
SD
661 TGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
720
  VYGKTKFPFMQKVLPILLKPL
SD
721 TGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
780
  YFISCIMFFPFLHSIPKQYWWSY
SD
781 ATTTGCTCCTGAGCGGGATTTTATTTATAGCAGGGCAATCGCGCTAATACGACTCCCTCGG
840
  LLTVWSELILSCGGILFM
SD
841 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
900
  YISEGVSLNYLAMFPTLVTPL
SD  kpsT
901 TGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
960
  IGLALYRTREEMHLTS*MIKIKI
Figure 3.4 The nucleotide sequence of the 1.8kb Bglll fragment, with the amino acid sequence of the two ORFs, *kpsM* (ORF1) and *kpsT* (ORF2), shown underneath. Initiation codons and the encoded valine or methionine are shown in bold. Stop codons, indicated by the asterisks, are also shown in bold. Putative Shine-Dalgarno ribosome binding sites are double underlined and are labelled SD. Restriction sites are shown in bold italics. The potential adenine nucleotide binding fold sequences identified in *kpsT* (Section 3.2.4.2) are underlined.
Figure 3.5 Alignment of the KpsM and KpsT proteins encoded by the region 3 of the E. coli K1 and K5 capsule gene clusters

(a) KpsM

1 MARSFEGKVTVXALPURRTRPGKFRGLYWAXLLPSAHLILGIL 50

1 MARSFGEVQTVVEALPFRTRPGKFRGLYWAXLPSAHLILGIL 50

51 GYTHMTFDPFPFLNQLGRPFSSISNRSVEAEQGLFNYRPV 100

51 GYTHMTFPFPLNNQLGRPFSSFISNRSVEAEQGLFNYRPV 100

101 KPIDTIARALLTLLNLYMVLWAGEEYFEITNLQGLVLTWLL 150

101 KPIDTIARALLTLLNLYMVLWAGEEYFEITNLQGLVLTWLL 150

151 IILSCGGLFVGVKSFEPKVLPSLIPLSFILICSIMFPLHSGKQYW 200

151 IILSCGGLFVGVKSFEPKVLPSLIPLSFILICSIMFPLHSGKQYW 200

201 SYLVWNLVYAVLREAVMPGYISEGSLVNLAMPTLTFICLGLYRT 250

201 SYLVWNLVYAVLREAVMPGYISEGSLVNLAMPTLTFICLGLYRT 250

251 REEAMTIS 258

KpsN 251 REEAMTIS 258

(b) KpsT

1 MIKIANLTKSRTPFGRHVFKDLNIEIPSQSKAVFIRNGAGKSTLLRM 50

1 MIKIANLTKSRTPFGRHVFKDLNIEIPSQSKAVFIRNGAGKSTLLRM 50

51 IGGIDRPDSGGKRTNKSISPVGLAGGFQGSITGRENVKFVARLYAKQEE 100

51 IGGIDRPDSGGKRTNKSISPVGLAGGFQGSITGRENVKFVARLYAKQEE 100

101 LKEKIEFVRELAEGYFDMPIKTYSQGNSDRLGFGLSMFADFYYID 150

101 LNERVDFVEFSELGKGFDMPIKTYSQGNSQDLGFGLSMFADFYYID 150

151 VTVAGDARFKECAQLFKEHKEESSFLMVSHLSLKEFCDVAIFKNSY 200

151 VTVAGDARFKECAQLFKEHKEESSFLMVSHLSLKEFCDVAIFKNSY 200

201 IIGYENQSGIDYEKMQQDLIE 224

201 IIGYENQSGIDYEKMQQDLIE 224

219 KpsN 219 QSPFKVKNTEIAIDYK.KDL...

Figure 3.5 Alignment of a) KpsM and b) KpsT encoded by the E. coli K1 and K5 capsule gene clusters using the GCG program GAP. Vertical lines indicate amino acid identity, a pair of dots indicates strong similarity and a single dot indicates weak homology. Gaps inserted to maximize the alignment are represented by dots. The asterisk and star indicate the C-terminal 30 and 60 amino acids of the KpsT protein, respectively (see Section 5.2.2).
A typical ABC transporter consists of four domains, two highly hydrophobic membrane-spanning domains (white), which form the translocation pathway and two peripheral membrane domains (dotted), which couple ATP hydrolysis to the transport process. A number of transporters have additional domains (vertical stripes). The four domains may be present as separate polypeptides as in a) above or alternatively they may be fused together in larger, multifunctional proteins as shown in b). In some cases, an ABC transporter may lack a full complement of domains, for example, in the histidine transporter shown in c), the HisP protein functions as a homodimer.
Figure 3.7 Adenine nucleotide binding fold sequences

Motif A

Consensus

KpsT (38-45) GRNGAGKS
BexA (38-45) GRNGAGKS
CtrD (38-45) GRNGAGKS

Motif B

Consensus

KpsT (137-149) LSNMAPKFDYIYVD
BexA (135-147) LLSLSVEFDCYLID
CtrD (135-147) LSLAVEFDCYLID

Figure 3.7 Alignment of the potential adenine nucleotide binding fold sequences identified in the KpsT, BexA and CtrD proteins, using the GCG program GAP. The numbers in brackets denote the position of the motifs in the amino acid sequence of each protein. Vertical lines indicate amino acid identity, a pair of dots indicates strong similarity and a single dot indicates weak homology. The consensus sequences (Walker et al., 1982) for the A and B motifs are shown above. x refers to a hydrophobic amino acid residue and the dashes represent any amino acid residue.
Figure 3.8a) Hydropathy profiles of KpsT, BexA and CtrD, generated according to the method of Kyte and Doolittle (1982) using a window size of 15. The upper half of a hydropathy profile indicates hydrophobicity and the lower half indicates hydrophilicity. With a window size of 15, a value of $\geq 1.6$ is indicative of a membrane-spanning region.
Figure 3.8b) Hydropathy profiles of KpsM, BexB and CtrC, generated as described in a).
Table 3.1 Summary of amino acid sequence comparisons

a)

<table>
<thead>
<tr>
<th>Protein</th>
<th>BexA</th>
<th>CtrD</th>
</tr>
</thead>
<tbody>
<tr>
<td>KpsT&lt;sub&gt;55&lt;/sub&gt;</td>
<td>44.2 (71.9)</td>
<td>43.1 (69.4)</td>
</tr>
<tr>
<td>BexA</td>
<td></td>
<td>81.0 (89.4)</td>
</tr>
</tbody>
</table>

b)

<table>
<thead>
<tr>
<th>Protein</th>
<th>BexB</th>
<th>CtrC</th>
</tr>
</thead>
<tbody>
<tr>
<td>KpsM&lt;sub&gt;35&lt;/sub&gt;</td>
<td>25.6 (54.3)</td>
<td>23.7 (56.0)</td>
</tr>
<tr>
<td>BexB</td>
<td></td>
<td>69.1 (84.9)</td>
</tr>
</tbody>
</table>

The amino acid sequences of a) KpsT<sub>55</sub>, BexA and CtrD and b) KpsM<sub>35</sub>, BexB and CtrC were subjected to pairwise comparisons using the GCG program GAP. The percent identity for each comparison is shown, with the percent similarity taking into account conservative amino acid substitutions shown in brackets.
Figure 3.9 Multiple sequence alignments

a) BexA, CtrD, and KpsT.

b) BexB, CtrC, and KpsM.

This figure was created using the GCG programs PILEUP and PRETTYPLOT. Boxes highlight where two or more amino acids in the aligned sequences are identical.
Figure 3.10 Region 3 subclones used for the visualization of the KpsM and KpsT proteins

Figure 3.10 The physical map of pGB110 is shown, with the open box below the line indicating vector sequence. The boxes labelled 1-3 refer to the three functional regions that direct the synthesis of the K5 antigen. Vertical and oblique lines indicate restriction endonuclease sites, with the oblique lines referring to sites present in the multiple cloning site of the vector pUC19. The broken lines highlight the enlargement (on a scale of 10:1) of the 1.8kb BglII fragment containing region 3 which is cloned in plasmid pAS1. Subclones derived from pAS1 are shown below and are labelled pAS2, pAS3 and pAS4, respectively. The boxes labelled M and T refer to the kpsM and kpsT genes identified in region 3 of the K5 capsule gene cluster and the arrow indicates the direction of their transcription. Restriction endonuclease sites: B, BamHI; C, ClaI; R, EcoRI; G, BglII; H, HindIII; M, Smal; P, PstI; R, NotI; S, SspI. The scale is shown in kb.
Figure 3.11 The visualization of the KpsM and KpsT proteins

a) Expression of the kpsM and kpsT genes using the vector pK223-3. SDS-PAGE was used to analyse E. coli lysates for the IPTG-induced expression of proteins. Samples taken from cultures 90 mins after the addition of IPTG (at a final concentration of 1mM) were compared to those taken from cultures which had not been treated with IPTG. Lysates were analysed on a 15% SDS-polyacrylamide gel stained with Coomassie Brilliant Blue. Lane 1, LE392(pMW1 and pK223-3); 2, LE392(pMW1 and pK223-3); 3, LE392(pMW1 and pAS2); 4, LE392(pMW1 and pAS2). The samples in lanes 2 and 3 had been treated with IPTG, whereas those in lanes 1 and 4 were untreated. The numbers at the left hand side correspond to the MWs of protein markers in kDa. The arrow indicates the position of an induced protein, with an approximate molecular mass of 26.5kDa, which can be seen in lane 3.

b) Minicell analysis. Labelling of the expressed proteins with ^35S methionine was carried out both with and without induction using IPTG (at a final concentration of 10mM). The radiolabelled proteins were analysed using SDS-PAGE and autoradiography. Lane M, molecular weight markers; 1, DS410; 2, DS410(pMW1); 3, DS410(pK223-3); 4, DS410(pMW1 and pK223-3); 5, DS410(pMW1 and pAS2); 6, DS410(pMW1 and pK223-3); 7, DS410(pMW1 and pAS2). The samples in lanes 6 and 7 had been treated with IPTG, whereas the samples in the remaining lanes were untreated. The arrow indicates the position of a non-vector encoded protein, with an approximate molecular mass of 27kDa, which can be seen in lanes 5 and 7.

c) Minicell analysis. As in b) above, protein labelling was carried out both with and without IPTG induction. Lane M, molecular weight markers; 1, DS410(pMW1 and pK223-3); 2, DS410(pMW1 and pK223-3); 3, DS410(pMW1 and pAS2); 4, DS410(pMW1 and pAS2); 5, DS410(pMW1 and pAS3); 6, DS410(pMW1 and pAS3); 7, DS410(pMW1 and pAS4); 8, DS410(pMW1 and pAS4). The samples in lanes 2, 4, 6 and 8 had been induced with IPTG. The upper arrow indicates the position of a non-vector encoded protein, with an approximate molecular mass of 27kDa, which can be seen in lanes 3, 4, 7 and 8. This protein corresponds to the KpsT protein. The lower arrow indicates the position of a non-vector encoded protein of approximately 24kDa, which can be seen in lanes 3, 4, 5 and 6. This protein can be identified as the KpsM protein.
Figure 3.11

a) 

b) 

M 1 2 3 4 5 6 7

46.0 -
30.0 -
14.3 -

M 1 2 3 4 5 6 7

46.0 -
30.0 -
14.3 -

M 1 2 3 4 5 6 7

46.0 -
30.0 -
14.3 -
Figure 3.12 The physical map of pGB110::34 is shown, with vertical and oblique lines indicating either restriction endonuclease sites (capital letters) or the site of the TnSphoA insertion (filled triangle). The open box below the line indicates vector sequence and the hashed boxes refer to parts of the ISS54 element of TnSphoA. The boxes labelled 1-3 refer to the three functional regions required for the synthesis of the K5 antigen. The broken lines highlight the enlargement (on a scale of 4:1) of the fragments cloned in the plasmids pAS7 and pAS10. The boxes labelled M and T refer to the kpsM and kpsT genes identified in region 3 of the K5 capsule gene cluster and the arrow indicates the direction of their transcription. Restriction endonuclease sites: B, BamHI; C, CiaI; E, EcoRI; G, BglII; H, HindIII; S, SspI; T, StuI. The scale is shown in kb.
Table 3.2 The minimum inhibitory concentration of chloramphenicol

<table>
<thead>
<tr>
<th>Strain</th>
<th>Chloramphenicol concentration in µg/ml</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>LE392(pACYC184)</td>
<td>+</td>
</tr>
<tr>
<td>LE392(pKK232-8)</td>
<td>+</td>
</tr>
<tr>
<td>LE392(pAS10)</td>
<td>+</td>
</tr>
</tbody>
</table>

Assay of the minimum concentration of chloramphenicol required to inhibit the growth of LE392(pAS10). The cultures were observed after approximately 18 hours incubation at 37°C.

Abbreviations: +, growth; -, no growth; +/-, limited growth.
CHAPTER 4

REGION 2 OF THE K5 CAPSULE GENE CLUSTER: PRELIMINARY ANALYSIS OF THE POLYMER SYNTHESIS GENES

4.1 INTRODUCTION

Region 2 of the kps locus is serotype specific and appears to encode the enzymes necessary for the synthesis and polymerization of the K antigen in question, since transposon insertions in this region abolish polysaccharide production (Boulnois et al., 1987; Roberts et al., 1988a). In the case of the K1 antigen gene cluster, the genetic organization of region 2 and the biochemical pathway involved in synthesis of the K1 polymer have been fairly well characterized (Section 1.4.3.1). Unfortunately, less information is available with regards to the biosynthesis of the K5 antigen. Although the biochemical pathway has recently been investigated (Section 1.4.3.2), a detailed genetic analysis of region 2 of the K5 antigen gene cluster has not yet been undertaken.

Determination of the nucleotide sequence of region 2 of the K5 kps locus would allow the identification of the number of genes involved in polymer synthesis and computer analysis of the encoded proteins may provide preliminary information about their possible function. Furthermore, once the genetic organization of the region was established, mutations could be introduced into individual genes, in order to try to elucidate the role of their encoded products in the biochemical pathway. Thus, in order to facilitate a better understanding of the processes involved in the synthesis of the K5 polymer, the nucleotide sequence of region 2 of the K5 kps locus was determined and a preliminary characterization of the encoded proteins undertaken, in conjunction with C. Pazzani (Pazzani, 1992). These experiments are described in this chapter.
4.2 RESULTS

4.2.1 Cloning And Nucleotide Sequence Analysis Of Region 2

In order to clone and sequence region 2 of the K5 kps locus, it was necessary to establish the precise location of the region in the capsule gene cluster. Previously, DNA hybridization and transposon mutagenesis had shown region 2 to be located between regions 1 and 3 and to be approximately 8kb in size (Roberts et al., 1988a). The recent molecular analysis of region 1 (Pazzani et al., 1993) and region 3 (Smith et al., 1990; Chapter 3) of the K5 capsule gene cluster has allowed the location and extent of region 2 to be defined more precisely (Figure 4.1) and these results are in good agreement with the earlier findings.

To facilitate the sequencing of part of region 2, a 2.1kb EcoRI-SstI fragment (Figure 4.1) was subcloned from pGB110 (encodes the K5 kps locus) into the vector pUC19, which had been cleaved with EcoRI and SstI. Positive recombinant clones were identified by α-complementation and restriction analysis of small scale plasmid DNA preparations. One such clone, termed pAS20, was chosen for further study. Similarly, the adjacent 2.4kb EcoRI fragment (Figure 4.1) was subcloned from pGB110 into pUC19, generating the plasmid pAS21. A detailed restriction endonuclease cleavage map was generated for the inserts in pAS20 and pAS21. Then a series of overlapping DNA fragments were subcloned from both plasmids into M13mp18 and M13mp19 and sequenced using the M13 -40 primer. Oligonucleotide primers were used to complete the sequence in both strands. In addition, it was necessary to sequence across the EcoRI site thought to be common to the inserts in both pAS20 and pAS21, to demonstrate that the sequence was in fact contiguous. The 4.2kb BglII-SstI fragment (Figure 4.1) spanning this region was subcloned from pGB110 into pUC19, generating plasmid pAS22 and then a suitable subfragment of pAS22 was cloned into M13mp18, to allow the sequence to be determined across the EcoRI site. The DNA sequencing strategy used in this study can be seen in Figure 4.1.
The nucleotide sequence generated in this study overlapped with the sequence determined for region 3 by 367bp (Figure 4.1). Thus, only the 4219bp of unique sequence extending from the BgIII site (3' to kpsT) to the second of the two closely situated SstI sites is shown in Figure 4.2. Analysis of the DNA sequence revealed the presence of two open reading frames, ORF1 and ORF2 (Figure 4.2). ORF1 starts with an AUG initiation codon at coordinate 751, is 714bp long and encodes a putative polypeptide of 238 amino acids with a predicted molecular weight of 27,300. A potential Shine-Dalgarno ribosome binding site could not be identified 5' to ORF1. However, the Shine-Dalgarno region is apparently not absolutely essential for the initiation of translation, although it is important for translational efficiency (Calogero et al., 1988; McCarthy and Gualerzi, 1990). Thus, translation may initiate at the first AUG codon in ORF1 in the absence of a Shine-Dalgarno sequence, or alternatively a downstream initiation codon may be used. In order to ascertain which codon is used, the N-terminal amino acid sequence of ORF1 would have to be determined. ORF2 starts 432bp downstream of the stop codon of ORF1, at the AUG initiation codon at coordinate 1899 or at the adjacent one at coordinate 1902. The first AUG codon was chosen for use in the computer analysis of ORF2. ORF2 is 1689bp long and encodes a putative polypeptide of 563 amino acids with a predicted molecular weight of 65,732. A potential Shine-Dalgarno sequence has been identified 4bp 5' to the initiation codon of ORF2 (Figure 4.2).

The DNA sequence upstream of ORF1 and ORF2 was examined for the presence of candidate E. coli -10 and -35 promoter sequences. Two sets of possible promoter sequences were identified 5' to ORF1, with the first being located at coordinates 340 and 366 and the second at coordinates 350 and 374 (Figure 4.2). With regard to ORF2, candidate -10 and -35 promoter sequences were identified 5' to the open reading frames at coordinates 1753 and 1778 (Figure 4.2). However, transcript mapping techniques such as primer extension need to be carried out in order to determine whether any of these potential promoter sequences function *in vivo.*
The remaining part of region 2, a 3.7kb fragment from SstI to an EcoRV restriction site 3' to kpsS, was sequenced by C. Pazzani (Pazzani, 1992). The sequence of 3676bp was found to contain two open reading frames, ORF3 and ORF4 (Figure 4.3). ORF3 encodes a putative polypeptide of 520 amino acids with a predicted molecular weight of 59,896. ORF4 starts 42bp downstream of the stop codon of ORF3 and encodes a putative polypeptide of 392 amino acids with a predicted molecular weight of 44,057. A rather poor potential Shine-Dalgarno sequence was identified 4bp 5' to the initiation codon of ORF3, but unfortunately such a sequence could not be identified 5' to the initiation codon of ORF4. Potential E. coli -35 and -10 promoter sequences have been identified 5' to ORF3, but no transcriptional termination signal was found 3' to the ORF. However, an inverted repeat, potentially capable of forming a stem-loop structure characteristic of a termination signal, was identified 3' to ORF4. These findings, together with the short gap between the two ORFs, may suggest that ORF3 and ORF4 are organized in a single transcriptional unit (Pazzani, 1992).

The gap between the stop codon of ORF2 and the first AUG initiation codon of ORF3 is 1291bp. A small open reading frame, ORF5, has been identified in this region on the opposite strand to ORFs1-4 (Figure 4.3). ORF5 starts with the rare initiation codon GUG, is 423bp long and encodes a putative polypeptide of 141 amino acids with a predicted molecular weight of 16,199. A potential Shine-Dalgarno sequence could not be identified 5' to ORF5, but a number of candidate E. coli -10 and -35 promoter sequences have been identified upstream of the ORF. Although ORF5 is small, there is some preliminary evidence that it may be expressed (Section 4.2.5). However, a protein of this size was not seen in the minicell analysis of region 2 (Section 4.2.3). Smaller open reading frames, encoding putative polypeptides with molecular masses of less than 10kDa, were also identified, but these will not be considered here.

It was decided that the genes identified in region 2 of the K5 kps locus should be named in keeping with the nomenclature scheme proposed by Vimr et al. (1989), where genes unique to a particular K antigen are given a designation reflecting the
nature of the polysaccharide. Thus the term $kfi$, referring to K five, has been proposed for the K5-specific biosynthesis genes, with ORF1 being designated as $kfiA$, ORF2 as $kfiB$, ORF3 as $kfiC$ and ORF4 as $kfiD$ (Petit et al., in press) (Figure 4.3). In addition, for the remainder of this thesis I shall refer to ORF5 as $kfiE$.

A GCG FASTA search, using the amino acid sequence of the putative KfiA, KfiB and KfiE proteins, was made against the available protein sequence databases. Unfortunately, the identified homologies were mostly restricted to fairly short amino acid stretches and did not provide any useful indication of the potential roles of these three proteins in K5 polysaccharide biosynthesis. However, computer analysis did reveal 21.4% identity and 64.1% similarity, over 192 amino acids, between KfiC and the RodC protein of *Bacillus subtilis* 168 (Honeyman and Stewart, 1989; Pazzani, 1992). RodC is a CDP-glycerol:poly(glycerolphosphate) glycerophosphotransferase (CGPTase) which is involved in synthesis of the poly(glycerolphosphate) [poly(groP)] teichoic acid of *B. subtilis* 168 (Pooley et al., 1992). CGPTase catalyses extension of the main poly(groP) backbone through sequential transfer of glycerolphosphate units from CDP-glycerol to an acceptor consisting of the growing chain attached to a linkage unit lipid. KfiC was also found to have 21% identity and 66% similarity, over 100 amino acids, with the rat liver bilirubin UDP-glucuronosyltransferase (Pazzani, 1992; Sato et al., 1990). These findings suggest that the $kfiC$ gene may encode a protein with a glycosyltransferase-like activity. Region 2 of the K5 *kps* locus would be expected to encode at least two sugar transferase activities, glucuronosyltransferase and N-acetylglucosaminyltransferase (Section 1.4.3.2). Thus, it is possible that the $kfiC$ gene may encode one of these enzymes.

Analysis of the KfiD protein showed that it has 27% identity and 69.1% similarity, over 359 amino acids, with the GDP-mannose dehydrogenase (AlgD) of *Pseudomonas aeruginosa* (Deretic et al., 1987; Pazzani, 1992). AlgD is a key enzyme in the biosynthesis of alginate (a linear copolymer of mannuronic acid and guluronic acid), catalysing the NAD$^+$-dependent oxidation of
GDP-mannose to form the alginate precursor, GDP-mannuronic acid (Roychoudhury et al., 1989). Interestingly, biosynthesis of the E. coli K5 antigen requires the sugar nucleotide precursor UDP-glucuronic acid (Section 1.4.3.2). Thus, region 2 of the K5 kps locus may be expected to encode the enzyme UDP-glucose dehydrogenase, which catalyses the conversion of UDP-glucose to UDP-glucuronic acid. In view of this, the homology which exists between KfiD and the GDP-mannose dehydrogenase AlgD may suggest that the kfiD gene encodes the related enzyme UDP-glucose dehydrogenase. Further evidence for this hypothesis comes from the fact that the amino acid sequence surrounding a conserved cysteine residue in KfiD and AlgD has homology with the putative catalytic site of other NAD⁺-linked dehydrogenases, including the UDP-glucose dehydrogenase from bovine liver (Feingold and Franzen, 1981).

Thus, in summary, the database searches have provided some preliminary evidence which suggests that KfiC may act as a glycosyltransferase and KfiD may be a UDP-glucose dehydrogenase. Unfortunately, they did not provide any useful information regarding the potential roles of the proteins encoded by the remaining putative region 2 genes kfiA, kfiB and kfiE. Further information about the possible functions of the K5 region 2 proteins, which has become available since the completion of the work described in this thesis, will be discussed in Section 4.3.

4.2.2 Southern Hybridization To Determine Whether The K4 and K5 kps Loci Have Any Homologous Polymer Synthesis Genes

E. coli group II K antigens are structurally very diverse. However, despite this, they can be subdivided into families, since many contain one or more identical sugars, but have different linkages and modifications (Jann and Jann, 1990). It might be expected that kps loci directing the synthesis of closely related polymers would have similar region 2 cassettes. This appears to be the case with the K1 and K92 antigens, which are both polymers of sialic acid, with the former having α-2,8 linkages (McGuire and Binkley, 1964) and the latter having...
alternating α-2,8 α-2,9 linkages (Glode et al., 1977). The partial restriction maps of region 2 of the K1 and K92 kps loci appear to be identical (Roberts et al., 1986) and extensive DNA homology apparently exists between the K1 and K92 genes in this region (Echarti et al., 1983). This suggests that the K1 and K92 kps loci have many, if not all, of their polymer synthesis genes in common. In support of this, it has been shown that an E. coli K1- mutant, with a null mutation in the region 2 polysialyltransferase gene neuS (Section 1.4.3.1), can be complemented in trans by the K92 neuS gene, resulting in the synthesis of K92 polymer (Steenbergen et al., 1992). Sequence analysis of the neuS genes has shown that they are 87.3% identical at the nucleotide level and a comparison of the predicted amino acid sequence of the encoded proteins revealed 82.9% identity and 91.7% similarity (Vlmr et al., 1992). Thus, the neuS genes of E. coli K1 and K92 appear to be highly conserved, with only a limited number of encoded amino acid differences accounting for the dual linkage specificity of the K92 polysialyltransferase.

The E. coli K4 capsular polysaccharide, which consists of 4)-β-glucuronic acid-(1,3)-β-N-acetylgalactosamine-(1, substituted at each glucuronic acid residue with a single fructose moiety (Rodriguez et al., 1988), has the glucuronic acid component in common with the K5 antigen. Thus, region 2 of the K4 and K5 kps loci may be expected to have some polymer synthesis genes in common. Region 2 of the K4 kps locus, which directs the synthesis of the more complex K4 polysaccharide, is larger than that of the K5 kps locus and their partial restriction maps are different (Drake et al., 1990; Roberts et al., 1988a) (Figure 4.4). However, this does not exclude the possibility that some of the genes present in this region may be similar. If homologous genes are identified in region 2 of these two capsule gene clusters, the information may help to give an indication of the possible function of one or more of the proteins encoded by the K5 region 2 kfi genes (Section 4.2.1). In order to analyse region 2 of the K4 and K5 kps loci for the presence of homologous genes, a Southern hybridization was carried out as described below.
The cosmid pRD1 encodes the E.coli K4 antigen gene cluster (Drake et al., 1990). pRD1 DNA was cleaved with the restriction endonucleases BamHI, ClaI, SmaI and XhoI in single and/or double digests. The cosmid pGB110, cleaved with EcoRI, was used as a positive control and the vectors pUC19 and pACYC184, cleaved with EcoRI, were used as negative controls. The DNA fragments were separated by agarose gel electrophoresis and transferred to nylon filters in duplicate by Southern blotting. The plasmid pAS26, which consists of the whole of region 2 from the K5 kps locus cloned into the vector pUC19 (see Section 4.2.4), was used to generate suitable K5 region 2 radiolabelled probes. Probe A consisted of a 4.0kb AccI fragment and probe B was the adjacent 3.2kb SaI fragment (Figure 4.4). The filters were hybridized and washed using conditions which should allow sequences with approximately 70% or greater homology to hybridize.

The pattern of hybridization after 14 days exposure is shown in Figure 4.5. The expected fragments of pGB110 gave a strong signal with both probes. However, only a very weak signal was seen with fragments of pRD1, even though approximately 10X more DNA had been used in the Southern blot (see Figure 4.5). A faint signal was also seen for the negative control pUC19 and this may indicate that the probe fragments isolated from pAS26 were slightly contaminated with vector DNA. In agreement with this, analysis of the pRD1 hybridization pattern showed that most of the fragments which gave a signal with both probes contained the related vector cos4. A faint signal was also seen with pACYC184, which suggests that this vector may also contain some related sequences. Interestingly, probe B appeared to hybridize very weakly to some additional fragments of pRD1 and these were mainly found to be located in region 2 of the K4 kps locus. Thus, it appears that there may be some limited DNA homology between region 2 of the K4 and K5 capsule gene clusters. However, since the level of DNA homology was barely detectable under the hybridization and washing conditions used in this experiment, it is likely to be much less than 70%.

A small fragment of region 2 of the K4 capsule gene cluster has been sequenced (Drake, 1991) and the location of this fragment
is shown in Figure 4.4. An open reading frame was found to extend throughout the 458bp of sequence and a comparison of this with the nucleotide sequence of the entire K5 capsule gene cluster identified a 310bp region with 56% DNA homology. This stretch of homology lies within the K5 region 2 kflC gene and when the proteins predicted to be encoded by kflC and the K4 region 2 ORF were compared, they were found to have 35% identity and 77% similarity over a 143 amino acid overlap (Drake, 1991). These results are in agreement with the preliminary findings of the Southern hybridization experiment, confirming that some limited DNA homology does exist between region 2 of the K4 and K5 kps loci. It is possible that other genes with homology to the K5 kfi genes may also be present in region 2 of the K4 kps locus, but their identification will probably require further nucleotide sequence analysis.

Thus, in contrast to the neuS genes of the K1 and K92 kps loci, none of the genes present in region 2 of the K4 and K5 capsule gene clusters appear to be highly conserved. However, the level of amino acid homology which does exist between KfiC and the K4 region 2 protein suggests that they may have a similar function in capsule biogenesis, possibly being involved in the synthesis or polymerization of glucuronic acid, the sugar component common to the K4 and K5 antigens. Alternatively, the relatively modest level of homology which exists between the two proteins may suggest that they catalyse the synthesis or polymerization of the similar, but K antigen-specific sugar components N-acetylglucosamine and N-acetylgalactosamine.

4.2.3 Minicell Analysis

E. coli DS410 minicells (Section 3.2.6.2) were used to visualize the proteins predicted to be encoded by the region 2 kfi genes of the E. coli K5 kps locus. Minicells were isolated from DS410 harbouring the following plasmids: pAS25 and pAS26, which consist of the whole of region 2 from the K5 kps locus cloned into the vectors pUC18 and pUC19, respectively (Section 4.2.4) and pAS22, which consists of the kfiA and kfiB genes cloned into pUC19 (Section 4.2.1). The fragments cloned in these plasmids are shown in Figure 4.6. In addition, minicells
were isolated from DS410(pUC18) in order to be able to distinguish between vector and non-vector encoded proteins. Genes carried by the plasmids were expressed in the minicells in the presence of $^{35}$S methionine and the radiolabelled proteins were detected using SDS-PAGE and autoradiography.

In Figure 4.7a, pAS22 can be seen to encode two non-vector proteins with approximate molecular masses of 64kDa and 38kDa. In addition, a non-vector protein of approximately 28kDa can also be seen to be weakly expressed by the plasmid. pAS25 and pAS26 encode these proteins and also two additional non-vector proteins with approximate molecular masses of 56kDa and 45kDa (Figure 4.7a). The observed molecular masses of 64kDa, 56kDa and 45kDa correspond well to those of 65.7kDa, 59.9kDa and 44.1kDa predicted for the KfiB, KfiC and KfiD proteins, respectively (Section 4.2.1) and the pattern of protein expression supports these findings. Unfortunately, it was not possible to determine which of the two remaining proteins corresponded to the KfiA protein. KfiA has a predicted molecular mass of 27.3kDa and this corresponds well to the 28kDa protein observed in the minicells. However, if a mistake had been made in the sequence analysis of the kfiA gene and the gene actually encodes a larger protein, then the 38kDa protein may represent KfiA. In order to resolve this issue, minicells were isolated from DS410 harbouring pAS21 (Section 4.2.1), a plasmid which consists of the kfiA gene cloned into pUC19 (Figure 4.6). A non-vector protein of 28kDa was found to be encoded by this plasmid (Figure 4.7b) and consequently this must represent the KfiA protein. Interestingly, the 38kDa protein only appears to be expressed when the kfiB gene is present and therefore it may represent a breakdown product of the 65.7kDa KfiB protein. Alternatively, translation of KfiB may be able to start at a downstream initiation codon resulting in the expression of a truncated product. The predicted 16.2kDa product of the putative kfiE gene would be expected to be expressed in minicells harbouring the plasmids pAS25 and pAS26. However, no such protein was identified in these minicell experiments. This may indicate that the proposed kfiE gene does not represent a real open reading frame. However, it is also possible that the protein is either running with the dye front
at the bottom of the SDS-PAGE gel or is expressed at levels which are undetectable using this technique.

Thus, in summary, KfiA, KfiB, KfiC and KfiD, encoded by region 2 of the *E. coli* K5 *kps* locus, were visualized as proteins with apparent molecular masses of 28kDa, 64kDa, 56kDa and 45kDa, respectively. These molecular masses correspond well to the values predicted by the nucleotide sequence data. A protein with a molecular mass corresponding to that predicted for the product of the putative *kfiE* gene was not seen in these experiments.

### 4.2.4 Region 2 Alone May Be Able To Direct The Synthesis And Polymerization Of The K5 Polysaccharide

Region 2 of the *kps* locus has previously been shown to encode the enzymes necessary for the synthesis and polymerization of a particular K antigen (Boulnois et al., 1987; Roberts et al., 1988a). However, it is not known whether the proteins encoded by region 2 can direct polymer synthesis in the absence of other *kps* gene products. Vimr et al. (1989) have suggested that some of the proteins encoded by regions 1 and 3 may be non-polymerase components of the glycosyltransferase complex and may function in the process of K antigen assembly, rather than in the translocation of the polymer to the cell surface. This suggests that region 2 alone may not be able to synthesize K antigen. In order to try to resolve this issue, region 2 of the K5 *kps* locus was cloned separately into a suitable vector and bacteria harbouring this construct were analysed for their ability to synthesize K5 polymer. These experiments are described below.

Region 2 of the K5 *kps* locus is approximately 7.9kb in size. Unfortunately, the region could not be cloned into the vectors pUC18 and pUC19 as a single DNA fragment, due to the lack of suitable flanking restriction endonuclease sites. Thus, a more complex cloning strategy had to be employed. The cosmid pGB118, a deletion derivative of pGB110 (Roberts et al., 1986), encodes the entire K5 *kps* locus. A number of Tn1000 (γ6) insertion mutants of pGB118 have been generated (Roberts et al., 1988a).
and the insertion in one such mutant pGB118::1 is now known to map approximately 200bp before the end of the kpsS gene (Figure 4.8a). Bacteria harbouring pGB118::1 have been shown to synthesize full length K5 polymer, which lacks KDO and phosphatidic acid and is located in areas of the cytoplasm which appear to have low electron density (Bronner et al., 1993a). Since the Tn1000 insertion in pGB118::1 is located at the 3' end of region 1, it provided a source of restriction endonuclease cleavage sites which could be used to facilitate the cloning of region 2.

The cosmid pGB118::1 was cleaved with the restriction endonucleases XhoI (which cuts 1.4kb inside Tn1000) and ClaI (which cuts within region 2), generating a 5.5kb fragment spanning half of region 2 (Figure 4.8a). Cleavage of pGB118::1 at the same ClaI site and at a BglII site, located between region 2 and region 3, generated a 4.0kb fragment spanning the remaining part of region 2 (Figure 4.8a). The 5.5kb XhoI-ClaI and 4.0kb ClaI-BglII fragments were cloned together into the vectors pUC18 and pUC19 cleaved with SalI and BamHII, generating the plasmids pAS25 and pAS26, respectively. Positive recombinant clones were identified by α-complementation and restriction analysis of small scale plasmid DNA preparations. pAS25 and pAS26 would be expected to contain the whole of region 2 from the K5 kps locus and the results of the minicell experiments support this (Section 4.2.3).

Total cell polysaccharide was extracted from E. coli JM101 (pAS25) and JM101(pAS25). E. coli LE392(pGB110) and LE392 (pGB118::1) were used as positive controls, whereas LE392 and JM101 provided negative controls. Double immunodiffusion of the extracted polysaccharide against anti-K5 monoclonal antibody (provided by K. Jann) was carried out and the results can be seen in Figure 4.8b. The positive and negative controls gave the expected results, but in the case of pAS25 and pAS26 the results were rather variable, with some of the clones giving a strong positive signal and others giving either a weak signal or none at all. Interestingly, the amount of DNA present in the small scale plasmid DNA preparations, made from the cultures used for polysaccharide extraction, appeared to correlate quite
well with the strength of the signal in the double immuno-
diffusion experiment. Thus, it appears that bacteria harbouring
region 2 alone can synthesize K5 polymer, but whether it can be
detected by double immunodiffusion may depend on how well the
culture has grown. Unfortunately, due to the unavailability of
further anti-K5 monoclonal antibody, the double immunodiffusion
experiments could not be repeated. In order to determine the
acellular location of the K5 polymer synthesized by JM101(pAS25)
and JM101(pAS26), the bacteria were tested for their sens-
itivity to infection with K5 capsule-specific bacteriophage. As
would be expected, the bacteria were found to be insensitive to
the phage and this indicated that the polymer was located
within the cell and not at the cell surface.

4.2.5 Preliminary Studies of The Promoters Of The kfiA, kfiB
And kfiE Genes

The vector pKK232-8, which contains a promoterless
chloramphenicol acetyl transferase (cat) gene, was used to
analyse putative promoter sequences for the kfiA, kfiB and kfiE
genes in the same manner as that described for the putative
region 3 promoter in Section 3.2.7. As discussed in Section
4.2.1, candidate E. coli -10 and -35 promoter sequences have
been identified upstream of the translational start codon of
each of the kfiA, kfiB and kfiE genes. DNA fragments, spanning
these potential promoter sequences, were cloned into the MCS
located upstream of the cat gene in pKK232-8 using the
strategies described below.

In the case of the kfiA gene, pAS21 was cleaved with the
restriction endonuclease DraI, which cuts both 659bp upstream
and 467bp downstream of the AUG initiation codon, generating a
1.13kb fragment (Figure 4.9). This fragment was initially
cloned into pUC19 cleaved with HincII. After checking for the
desired orientation using restriction analysis, the fragment
was cut out of pUC19 using BamHI and HindIII and cloned into
pKK232-8 cleaved with BamHI and HindIII, generating plasmid
pAS32. Using a similar approach for the kfiB gene, pAS22 was
cleaved with the enzyme RsaI, which cuts both 349bp upstream
and 730bp downstream of the AUG initiation codon, generating a
1.08kb fragment (Figure 4.9). After being initially cloned into the HindII site of pUC19 in the desired orientation, the fragment was cut out of the vector using BamHI and HindIII and cloned into pKK232-8 cleaved with BamHI and HindIII, generating plasmid pAS33. The plasmid CH2 (provided by C. Pazzani), which consists of a 1.17kb HindIII-ClaI fragment cloned into pUC18 (Figure 4.9), was used to clone potential promoter sequences for the kfiE gene. CH2 was cleaved with the restriction endonucleases SspI, which cuts 522bp upstream of the initiation codon of kfiE and BamHI, which cuts outside of the insert in the pUC18 multiple cloning site, generating a 0.79kb fragment (Figure 4.9). This fragment was cloned into pKK232-8 cleaved with SmaI and BamHI, creating plasmid pAS34. In each case, positive recombinant clones were identified by colony hybridization, using probe A (Figure 4.4) from Section 4.2.2 as the radiolabelled probe, together with restriction analysis of small scale plasmid DNA preparations. Thus, pAS32, pAS33 and pAS34 contain potential promoter sequences for the kfiA, kfiB and kfiE genes, respectively, cloned upstream of the cat gene. In addition, the constructs also contain some coding sequence, but since the vector pKK232-8 contains a series of stop codons, in all three reading frames, located between the MCS and the initiation codon of the cat gene, any translational read-through into the cat gene would be prohibited.

In order to determine whether the DNA fragments cloned in pAS32, pAS33 and pAS34 had any promoter activity, LE392(pAS32), LE392(pAS33) and LE392(pAS34) were incubated at 37°C in the presence of concentrations of chloramphenicol ranging from 0µg/ml to 90µg/ml. LE392(pKK232-8) and LE392(pACYC184) were used as negative and positive controls, respectively. The results are shown in Table 4.1. As expected, LE392 harbouring pKK232-8 were unable to grow in the presence of chloramphenicol. However, LE392(pAS32) and LE392(pAS34) were still able to grow even in the presence of 90µg/ml chloramphenicol and the minimum concentration required to inhibit growth of LE392(pAS33) was between 60µg/ml and 75µg/ml. Thus, the DNA fragments cloned in pAS32, pAS33 and pAS34 appear to have some promoter-like activity, which suggests that they may contain the promoters of the kfiA, kfiB and kfiE genes, respectively.

119
To try to determine whether the putative promoters in pAS32, pAS33 and pAS34 were subject to regulation by growth temperature, a spectrophotometric assay was used to measure CAT activity after growth of the host bacteria at either 37°C or 18°C. In order to calculate specific CAT activity in units of CAT activity/mg total cell protein, the total protein concentration in mg/ml was determined for each sample. The assay was carried out using LE392 harbouring either pKK232-8 (negative control), pAS32, pAS33, pAS34 or pBR328 (positive control). Unfortunately, the results of these experiments were inconclusive, since the CAT activity values observed for LE392(pAS32), LE392(pAS33) and LE392(pAS34), after growth of the bacteria at both 37°C and 18°C, were too close to the limits of detection to be reliable and in addition, the measurements of total cell protein were inconsistent.

4.3 DISCUSSION

Region 2 of the E. coli K5 kps locus is approximately 7.9kb in size and appears to contain five ORFs (Figure 4.3), which have been named kflA-D (Petit et al., in press) and kflE. The putative KfIA, KfIB, KfIC and KfID proteins were visualized, using minicells and SDS-PAGE, as proteins with apparent molecular masses of 28kDa, 64kDa, 56kDa and 45kDa, respectively. These values are consistent with the molecular masses predicted from the nucleotide sequence data. In addition, when the kflB gene was present, a 38kDa protein was also observed and this protein may represent a breakdown product or a truncated version of the KfIB protein.

Unfortunately, a protein with a molecular mass of approximately 16kDa, corresponding to that predicted for the product of the putative kflE gene, was not seen in these minicell experiments. Thus, kflE may not represent a functional open reading frame. In support of this, analysis carried out since the completion of this project has shown that the disruption of kflE, by the insertion of pUC19 into the ORF, has no detectable effect on K5 capsule expression (Petit et al., in press). However, there is some preliminary evidence which suggests that the putative kflE
gene may in fact be expressed, since a DNA fragment containing the potential promoter sequences identified upstream of the ORF was found to have promoter-like activity. Therefore, although it appears that KfiE is not required for capsule biogenesis per se, one cannot exclude the possibility that the putative protein has an indirect role in this process; for example, it may be involved in the regulation of capsule expression. Ideally, in order to determine whether kfiE does actually represent a functional open reading frame, techniques such as northern blotting should be used to look for possible RNA transcripts.

A search of the available protein sequence databases has shed some light on the possible functions of the proteins encoded by region 2 of the K5 kps locus. The predicted amino acid sequence of the KfiD protein was found to have fairly extensive homology with AlgD, the NAD+-dependent GDP-mannose dehydrogenase of *P. aeruginosa*. This enzyme has been shown to catalyse the formation of GDP-mannuronic acid, a precursor in alginate biosynthesis. Interestingly, the related nucleotide sugar UDP-glucuronic acid is known to be essential for the biosynthesis of the K5 polymer (Section 1.4.3.2). Thus, on the basis of the homology observed between KfiD and AlgD, it seems likely that the *kfiD* gene may encode the UDP-glucose dehydrogenase that is responsible for the synthesis of this important precursor. Since the completion of the work described in this thesis, the KfiD protein has been found to have extensive homology with a number of NAD+-dependent nucleotide sugar dehydrogenase enzymes from a diverse range of bacterial species. These include the UDP-glucose dehydrogenases encoded by the capsule gene clusters responsible for the synthesis of the hyaluronic acid capsule in group A streptococci (Dougherty and van de Rijn, 1993) and the type 3 capsule in *S. pneumoniae* (Arrecubieta *et al*., 1994), both of which contain glucuronic acid. In addition, it has recently been shown that overexpression of the *kfiD* gene results in detectable levels of UDP-glucose dehydrogenase activity (Petit *et al*., in press). Thus, these findings appear to confirm the earlier suggestion that the KfiD protein may function as a UDP-glucose dehydrogenase.
The predicted amino acid sequence of the KfiC protein was found to have significant homology with several glycosyltransferase enzymes and this may suggest that the kfiC gene encodes one of the sugar transferases involved in the biosynthesis of the K5 antigen. In agreement with this, database searches carried out since the completion of this project have shown that the KfiC protein has homology with many more glycosyltransferases, including hyaluronan synthase (DeAngelis et al., 1993), which directs the synthesis of hyaluronic acid in group A streptococci. Interestingly, as discussed in Section 1.4.3.2, hyaluronic acid is closely related to the K5 antigen, differing only in one of the sugar linkages. Therefore, the homology observed between KfiC and hyaluronan synthase may suggest that they catalyse similar biosynthetic reactions during capsule biogenesis. Recently, overexpression of the kfiC gene has enabled the biochemical function of KfiC to be determined. In these experiments, the overexpression of kfiC resulted in detectable levels of K5 sugar transferase activity, in which there was a low, but significant incorporation of both glucuronic acid and N-acetylglucosamine into exogenously added K5 polysaccharide (Petit et al., in press). Thus, it appears that the kfiC gene may encode a bifunctional glycosyltransferase, which has both glucuronosyl- and N-acetylglucosaminyltransferase activities. In support of this, hyaluronan synthase is also thought to have a similar function in hyaluronic acid biosynthesis (DeAngelis et al., 1993). Several other bifunctional glycosyltransferase enzymes have also recently been identified, including the polysialyltransferase of E. coli K92, which is able to synthesize polysialic acid with alternating α-2,8 α-2,9 linkages (Steenbergen et al., 1992) and the E. coli KDO transferase, which catalyses the incorporation of two stereochemically distinct KDO residues during LPS biosynthesis (Belunis and Raetz, 1992).

None of the database searches carried out using the predicted amino acid sequence of the KfiA and KfiB proteins have given any useful indication of the possible functions of these proteins in the biosynthesis of K5 polysaccharide. However, some preliminary evidence has recently become available which suggests that the KfiA protein catalyses the initial sugar
transferase reaction in K5 polymer synthesis (Roberts, personal communication). Interestingly, it has been found that a distinct enzyme is required to initiate de novo synthesis of polysialic acid in *E. coli* K1 and K92, after which, the polysialyltransferase enzyme catalyses the polymerization reaction (Steenbergen *et al.*, 1992). Perhaps biosynthesis of the K5 polysaccharide has a similar requirement for a distinct initiating enzyme. Unfortunately, the role of the KfiB protein in K5 polymer synthesis still remains unclear. One possibility is that KfiB functions in the synthesis of UDP-N-acetylglucosamine, the other nucleotide sugar precursor essential for biosynthesis of the K5 polymer (Section 1.4.3.2). However, this is unlikely to be the case, since N-acetylglucosamine is an essential component of the cell and consequently, it would be expected that the enzyme required for the synthesis of its nucleotide derivative is encoded elsewhere on the *E. coli* chromosome.

During this project, it was demonstrated that bacteria harbouring region 2 alone were able to synthesize K5 polymer, as detected by the technique of double immunodiffusion. As one would expect, this polymer was shown to be located within the cell, since the recombinant bacteria were found to be resistant to infection by K5 capsule-specific bacteriophage. These findings conflict with those reported by Bronner *et al.* (1993b). In their study, it was found that bacteria harbouring region 2 alone (pAS26) were unable to produce any K5 polysaccharide that could be detected by immunoelectron microscopy. In addition, membranes prepared from the recombinant bacteria were unable to synthesize K5 polymer and did not elongate K5 polysaccharide added as an exogenous acceptor. Thus, it was concluded that the presence of region 2 alone is not sufficient for the synthesis of K5 polysaccharide and that the process requires the participation of certain *kps* encoded gene products from regions 1 and 3. These proteins may be required to hold the growing polysaccharide and the biosynthetic apparatus together in an efficiently functioning multicomponent membrane-associated complex. If this is the case, then it may provide an explanation for the conflicting results discussed above. If synthesis of the K5 polymer can only occur very inefficiently
in the absence of the requisite kps encoded proteins, then it may only be possible to detect the presence of the K5 polysaccharide using techniques such as double immunodiffusion, where the extracted polysaccharide has been greatly concentrated (50µl of polysaccharide extract is derived from 10ml of stationary phase bacterial culture). In support of this, radial immunodiffusion, a technique similar to double immunodiffusion, has recently been used to demonstrate that region 2 of the K1 kps locus is also able to direct the synthesis of intracellular polymer in the absence of region 1 and region 3 encoded proteins (Annunziato et al., 1995).

The promoter probe vector pKK232-8 was used during this project to carry out a preliminary investigation of the potential promoters of the kflA, kfiB and kfiE genes. Candidate E. coli -10 and -35 promoter sequences had previously been identified upstream of the translational initiation codon of each of the genes and DNA fragments containing these sequences were found to have promoter-like activity when cloned into pKK232-8. However, since the completion of this project, the transcriptional organization of the genes in region 2 of the K5 kps locus has been subjected to a more detailed analysis by northern blotting and transcript mapping (Petit et al., in press). The transcript mapping experiments have identified three transcriptional start sites within region 2 (it must be noted that no data was presented with regard to transcription and the putative kfiE gene). The first start site is 68bp upstream of the predicted initiation codon of the kflA gene and the promoter sequences identified 5' to this site do not correspond to the candidate -10 and -35 sequences described previously (Figure 4.2). The second start site is 101bp upstream of the predicted initiation codon of the kfiB gene and in this case, the associated -10 and -35 promoter sequences do correspond to the candidate sequences identified previously (Figure 4.2). The third start site is 361bp upstream of the predicted initiation codon of the kfiC gene and the promoter sequences located 5' to this site do not correspond to those previously described (Pazzani, 1992).

Three major overlapping transcripts of 8.0kb, 6.5kb and 3.0kb
have been identified by northern blotting and the size of these transcripts is in agreement with that predicted from the location of the promoters, assuming that transcription terminates after the kfiD gene in each case. Ideally, in order to determine whether transcription of the region 2 genes is subject to regulation by growth temperature, in a similar manner to K5 capsule expression, northern blots should now be carried out using RNA extracted from bacteria grown at 18°C. Alternatively, transcriptional fusions could be made with, for example, the luciferase gene and the bacteria could then be assayed for bioluminescence after growth at both 37°C and 18°C. Unfortunately, CAT assays conducted during this project to try to determine whether the putative promoters of the kfiA, kfiB and kfiE genes were regulated by growth temperature proved to be inconclusive. This was mainly due to the extremely low levels of CAT activity detected in these experiments. Petit et al. (in press) also observed very low levels of CAT activity when they used promoter-probe analysis to demonstrate that the promoters of the kfiA, kfiB and kfiC genes function at 37°C. These findings suggest that the region 2 genes may not be highly expressed.
Figure 4.1 Analysis of the first half of region 2 of the *E. coli* K5 capsule gene cluster

Figure 4.1 The physical map of pGB110 is shown, with the open box below the line indicating vector sequence. The boxes labelled 1-3 refer to the three functional regions that direct the synthesis of the K5 antigen. The boxes labelled E D U C S and M T represent the *kps* genes identified in region 1 and region 3, respectively. The broken lines highlight the enlargement (on a scale of 4:1) of the fragments cloned in plasmids pAS20, pAS21 and pAS22. The short horizontal lines in brackets labelled Forward and Reverse illustrate the strategy used to sequence the first half of region 2. The two open reading frames identified in this region are represented by the boxes labelled ORF1 (*kflA*) and ORF2 (*kflB*) and the arrows indicate the direction of their transcription. The end of the region 3 *kpsT* gene is also shown. Restriction endonuclease sites: B, BamHI; C, ClaI; E, EcoRI; G, BglII; H, HindIII; P, PstI; S, SstI; T, StuI; U, PvuII. The *SstI* site shown above represents two very closely situated sites. The scale is shown in kb.
Figure 4.2  The nucleotide sequence of the 4.2kb BglII-SstI fragment of region 2 of the E. coli K5 capsule gene cluster

**BglII**

```
1 ACATCTCTAAATACATACAAATATATTGTTAGATACCGTCTCTGTACACTAGACTAGC 60
61 TCCCTGAACTCTTGGCAGCTATCATATCTTAATACAGTGAGTACGAACTACACAGCACAC 120
121 TTAGCTAGTGAGTGGAGAATAGATAATTGGTGTCTTAGAAGCAGAGAAACGCAGACAGC 180
181 GTGCCACACAGGAAAATCGCTATTTTTCAGCAGACCTTGAAACCGAAAATGGCGGTCTTC 240
241 CTGTTGCAAGCAACCATGCTGCTGCGAGCTCCAGTTATCTTTTACGCGTAAATAC 300
301 (-35) (-35) (-10) (-10) (-10)
361 CTCTATATAATTAATATATATATATCTATATTATTTTATTAGTCAAGGAAAATGTTAGGT 420
421 TTGAACTAGGTAATTATGAATTTGATCGTGATCTCGTAATACGTTGCTGTTATTCTTTA 480
481 TTAATTATCTGCCAATTTATTTATGATAGTTACAGGAAATGTTTATGCAAAGAGTGGTT 540
541 TGATATGGTAAGAGTAAATATACGGGTAGATGAAGATAAATATATCAAACGTACACCCTAG 600
601 (-35) (-35) (-10) (-10)
661 TTCCTAAATATGACCTTTCTATATAATATTATACGCTTGAATTCTTTTCCATT 720
721 AACATGCAATATTAATTAAAATATTACCCCATGATTGTTGCAAATATGTCATCATACCCA 780
781 MIVANMSSYP

841 CTCGAAAAAGAAGTAGGCGACATTGCTACAAATACAGTGAGTACGAACTACACAGCACAC 900
841 ATCTTTGCTGGTAAGAATTATAGTGAATAGTTACAGGAAATGTTTATGCAAAGAGTGGTT 960
901 TGAACCTTTATATGACCTTTCTATATAATATTATACGCTTGAATTCTTTTCCATT 1020
961 AAAATGATATGATCGTACTTACAGATGATGATATTATTTCCGCGTTATGAGAA 1080
1021 KNDMIVLTDIYPPDYVE

1140 CIYIDADFQKSGSKKVFSTT
```
Figure 4.2 The nucleotide sequence of the 4.2kb BglII-SstI fragment, with the amino acid sequence of the two ORFs, \textit{kflA} (ORF1) and \textit{kflB} (ORF2), shown underneath. Initiation codons and the encoded methionine are shown in bold. Stop codons, indicated by the asterisks, are also shown in bold. Putative Shine-Dalgarno ribosome binding sites are double underlined and are labelled SD. Restriction sites are shown in bold italics. Potential \textit{E. coli} -10 and -35 promoter sequences are labelled in bold and those which have been identified by transcript mapping experiments carried out since the completion of this project (Section 4.3) are underlined.
Figure 4.3 The genetic organization of region 2 of the *E. coli* K5 capsule gene cluster

---

**Figure 4.3** The physical map of pGB110 is shown, with the open box below the line indicating vector sequence. The boxes labelled 1-3 refer to the three functional regions that direct the synthesis of the K5 antigen. The boxes labelled **K D U C S** and **T M** represent the genes identified in region 1 and region 3, respectively. The broken lines highlight the enlargement (on a scale of 4:1) of region 2. The open reading frames identified in region 2 are represented by the boxes labelled ORF1-5 and the **kfi** designation is shown above. The hatched areas correspond to non-coding sequence. The arrows indicate the direction of transcription. Restriction endonuclease sites: **B**, BamHI; **C**, ClaI; **E**, EcoRI; **G**, BglII; **S**, SstI; **V**, EcoRV. Additional EcoRV sites present in pGB110 are not shown and the SstI site shown above represents two very closely situated sites. The scale is shown in kb.
Figure 4.4 Physical maps of the cosmids pGB110 and pRD1

Figure 4.4 Physical maps of the cosmids pGB110 and pRD1 carrying the biosynthesis genes for the K5 and K4 antigens, respectively. The open boxes below the line indicate vector sequence. The boxes labelled 1-3 refer to the three functional regions involved in synthesis of the respective K antigen. The restriction endonuclease cleavage maps were aligned on the basis of the conserved BamHI cleavage sites in region 1. The box labelled C represents the kfiC gene identified in region 2 of the K5 antigen gene cluster and the black box corresponds to the partial ORF identified in region 2 of the K4 antigen gene cluster (see text). The arrows indicate the direction of transcription. The hashed box labelled A and the dotted box labelled B refer to the two K5 region 2 probes used in the Southern hybridization (see text). Restriction endonuclease sites: B, BamHI; C, ClaI; E, EcoRI; G, BglII; M, SmaI; S, SstI; V, EcoRV; X, XhoI. Additional EcoRV sites present in pGB110 are not shown and the SstI site shown above represents two very closely situated sites. The scale is shown in kb.
Figure 4.5  Southern hybridization analysis of pRD1 using probes taken from region 2 of the E. coli K5 capsule gene cluster

a) Ethidium bromide-stained 0.7% agarose gel showing the following: Lane M, DNA size markers; 1, pSV110 cleaved with EcoRI; 2, pUC19 cleaved with EcoRI; 3, pACYC184 cleaved with EcoRI; 4, pRD1 cleaved with BamHI; 5, pRD1 cleaved with ClaI; 6, pRD1 cleaved with SmaI; 7, pRD1 cleaved with BamHI and ClaI; 8, pRD1 cleaved with ClaI and SmaI; 9, pRD1 cleaved with BamHI and XhoI. DNA size markers (kb): 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1.6, 1, 0.5. Duplicate gels, identical to the one illustrated, were used in the Southern hybridization analysis shown in b) and c).

b) and c) Southern hybridization analysis of the duplicate gels using probe A (b) and probe B (c), taken from region 2 of the K5 capsule gene cluster. The origin of the probes is shown in Figure 4.4. The numbers at the left hand side represent the migration of DNA size markers in kb. The asterisks indicate restriction fragments of pRD1 that hybridized to both probes and contain cos4 vector sequence. The stars indicate restriction fragments of pRD1 which only hybridized to probe B and are located in region 2 of the K4 capsule gene cluster.
Figure 4.5

a)

```
M 1 2 3 4 5 6 7 8 9 M
```

b)

c)
**Figure 4.6** The physical map of pGB110 is shown. The open box below the line indicates vector sequence and the hashed box refers to part of the transposon Tn1000 (see below). The boxes labelled 1-3 refer to the three functional regions that direct the synthesis of the K5 antigen. The boxes labelled E D U C S and T M represent the kps genes identified in region 1 and region 3, respectively. The broken lines highlight the enlargement (on a scale of 4:1) from a site in the kpsS gene corresponding to the position of the Tn1000 insertion present in pGB118:1 (pGB118 is a deletion derivative of pGB110, see figure 4.8a) to the EcoRI site located in region 3. The subclones used in the minicell experiments are shown below and are labelled pAS25/pAS26, pAS22 and pAS21. The construction of plasmids pAS25 and pAS26 is discussed in the text (Section 4.2.4) and is represented in figure 4.8a. The boxes labelled A B C D E represent the kfl genes identified in region 2 and the arrows indicate the direction of their transcription. The 3' ends of the kpsS and kpsT genes present in plasmids pAS25/pAS26 and pAS21, respectively, are also shown. Restriction endonuclease sites: B, BamHI; C, ClaI; E, EcoRI; G, BglII; S, SstI; X, XhoI. The SstI site shown represents two very closely situated sites. The scale is shown in kb.
Figure 4.7 Minicell analysis of the proteins encoded by region 2 of the *E. coli* K5 capsule gene cluster

a) The proteins expressed in the minicells were labelled with $^{35}$S methionine, analysed using SDS-PAGE on a 12.5% gel and detected by autoradiography. Lane M, molecular weight markers; 1, DS410; 2, DS410(pUC18); 3, DS410(pAS22); 4, DS410(pAS25); 5, DS410(pAS26). Non-vector encoded proteins, with approximate molecular masses of 64kDa, 56kDa, 45kDa, 38kDa and 28kDa, are indicated by the arrows (see text).

b) The proteins were analysed as in a) above. Lane M, molecular weight markers; 1, DS410; 2, DS410(pUC18) 3, DS410(pAS26); 4, DS410(pAS22); 5, DS410(pAS21). The arrow indicates the position of a non-vector encoded protein, with an approximate molecular mass of 28kDa, which can be seen in lanes 3, 4 and 5. This protein corresponds to the KfiA protein.
Figure 4.7

a)

b)
a) The physical map of pGB118::l is shown, with vertical and oblique lines indicating either restriction endonuclease sites (capital letters) or the site of the Tn1000 insertion (filled triangle). The open box below the line indicates vector sequence and the hashed box refers to part of the Tn1000 transposon. The boxes labelled 1-3 refer to the three functional regions that direct the synthesis of the K5 antigen. The boxes labelled E D U C S and T M represent the kps genes identified in region 1 and region 3, respectively. The broken lines highlight the enlargement (on a scale of 4:1) from a XhoI site, present in the Tn1000 transposon located in the 3' end of the kpsS gene, to the end of region 2. This fragment corresponds to insert cloned in the plasmids pAS25 and pAS26 and its construction is represented below. The boxes labelled A B C D E represent the kfi genes identified in region 2 and the arrows indicate the direction of their transcription. The 3' end of the kpsS gene which is present in plasmids pAS25 and pAS26 is also shown. Restriction endonuclease sites: B, BamHI; C, CiaI; E, EcoRI; G, BglII; S, SstI; X, XhoI. The SstI site shown above represents two very closely situated sites. The scale is shown in kb.

b) A sketch illustrating the double immunodiffusion analysis of total cell polysaccharide extracted from E. coli JM101 harbouring the plasmids pAS25 (clones a-e) and pAS26 (clones a-c). E. coli LH392(pGB110) and LH392 (pGB118::l) were used as positive controls, whereas LH392 and JM101 provided negative controls. 10µl of anti-K5 monoclonal antibody were placed in the central well and 10µl of polysaccharide extract were placed in the peripheral wells as follows. Plate i): 1, LH392; 2, LH392(pGB118::l); 3, JM101(pAS25a); 4, JM101(pAS25b); 5, JM101(pAS25c); 6, JM101(pAS25d) and plate ii): 1, JM101; 2, LH392(pGB110); 3, JM101(pAS25e); 4, JM101(pAS26a); 5, JM101(pAS26b); 6, JM101(pAS26c).
Figure 4.8

a)

![Diagram of DNA restriction sites and plasmids pGB118::1 K5* and pAS25/pAS26]

b)

plate i

```
1
6
5
4
3
2
```

plate ii

```
1
6
5
4
3
2
```
Figure 4.9 Subclones used in the preliminary studies of the $kfiA$, $kfiB$ and $kfiE$ promoters

Figure 4.9 The physical map of pGB110 is shown, with the open box below the line indicating vector sequence. The boxes labelled 1-3 refer to the three functional regions that direct the synthesis of the K5 antigen. The boxes labelled EDUCS and TM represent the $kps$ genes identified in regions 1 and 3, respectively. The broken lines highlight the enlargement (on a scale of 4:1) of region 2 and the end of region 3. The subclones used in the promoter analysis experiments are shown below, labelled pAS32, pAS33 and pAS34 and these were derived from plasmids pAS21, pAS22 and CH2, respectively. The boxes labelled A B C D E represent the $kfi$ genes identified in region 2 and the arrows indicate the direction of their transcription. The vertical arrows indicate the position of the initiation codon in the $kfiA$, $kfiB$ and $kfiE$ genes. Restriction endonuclease sites: A, AccI; B, BamHI; C, ClaI; D, DraI; E, EcoRI; G, HgiII; H, HindIII; P, SspI; R, Rsal; S, SstI; V, EcoRV. Only the relevant sites are shown for AccI, DraI, EcoRV, Rsal and SspI. The SstI site shown above represents two very closely situated sites. The scale is shown in kb.
Table 4.1 The minimum inhibitory concentration of chloramphenicol

<table>
<thead>
<tr>
<th>Strain</th>
<th>Chloramphenicol concentration in µg/ml</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>LE392(pACYC184)</td>
<td>+</td>
</tr>
<tr>
<td>LE392(pKK232-8)</td>
<td>+</td>
</tr>
<tr>
<td>LE392(pAS32)</td>
<td>+</td>
</tr>
<tr>
<td>LE392(pAS33)</td>
<td>+</td>
</tr>
<tr>
<td>LE392(pAS34)</td>
<td>+</td>
</tr>
</tbody>
</table>

Assay of the minimum concentration of chloramphenicol required to inhibit the growth of LE392(pAS32), LE392(pAS33) and LE392(pAS34). The cultures were observed after approximately 18 hours incubation at 37°C.

Abbreviations: +, growth; -, no growth; +/-, limited growth.
CHAPTER 5

GENETIC VARIABILITY IN GROUP II K ANTIGEN GENE CLUSTERS

5.1 INTRODUCTION

As discussed previously, molecular analysis of the E. coli kps locus has shown that the capsule genes are organized into three functional regions (Boulnois et al., 1987; Roberts et al., 1988a) (Section 1.4.1). Region 2 appears to comprise a cassette of polymer synthesis genes which are unique to a particular K antigen, whereas regions 1 and 3, which flank region 2, are proposed to mainly encode products involved in translocation of the polysaccharide to the cell surface, via a mechanism common to the production of different K antigens.

Regions 1 and 3 have been partly defined on the basis of the extensive DNA homology (approximately 95%) observed in these regions between different group II capsule gene clusters (Roberts et al., 1988a). In addition, mutations present in regions 1 or 3 in one cluster can be complemented in trans by the equivalent region from a different capsule gene cluster (Roberts et al., 1986, 1988a). Thus, regions 1 and 3 appear to be functionally interchangeable between different kps loci and this, together with the extensive DNA homology which exists, suggests that the proteins encoded by these regions must be identical or nearly so. However, confirmation of this requires sequence information to be determined from a number of different capsule gene clusters. Unfortunately, although the nucleotide sequence of the entire K5 kps locus is now known (Pazzani et al., 1993; Petit et al., in press; Smith et al., 1990), very little sequence information is available for other group II capsule loci, the exception being regions 2 (reviewed in Silver et al., 1993; Section 1.4.3.1) and 3 (Pavelka et al., 1991) of the K1 antigen gene cluster and neuS<sub>42</sub> (Vimr et al., 1992). The nucleotide sequence of most of the region 1 gene kps<sub>5</sub> has also been published (Steenbergen et al., 1992). In
addition, since the completion of this project, the nucleotide sequence of the \textit{kpsE} \textsuperscript{K1} (Cieslewicz et al., 1993) and \textit{kpsD} \textsuperscript{K1} (Wunder et al., 1994) genes has been reported and this will be discussed in Section 5.3.

Interestingly, comparison of the nucleotide sequence of region 3 from the K1 and K5 \textit{kps} loci has shown that \textit{kpsM} \textsuperscript{K1} and \textit{kpsM} \textsuperscript{K5} have 94.9\% DNA homology as predicted, whereas \textit{kpsT} \textsuperscript{K1} and \textit{kpsT} \textsuperscript{K5} have only 69.4\% DNA homology (Section 3.2.3). Thus, on the basis of DNA homology, the \textit{kpsT} \textsuperscript{K1} and \textit{kpsT} \textsuperscript{K5} genes do not conform to the earlier definition of a region 3 gene, although they are known to be functionally interchangeable (Roberts et al., 1988a). Until recently, it was not known whether any of the genes present in region 1 showed such variation between different capsule gene clusters or whether \textit{kpsT} provided the exception to the rule. However, a limited amount of sequence information from several different \textit{kps} loci has now become available for the \textit{kpsE} and \textit{kpsS} genes of region 1 and, as with region 3, this has provided some interesting findings. These findings, together with a G+C content analysis of the K5 capsule gene cluster, are presented in this chapter and both are discussed with regard to possible mechanisms involved in the generation of group II K antigen diversity.

5.2 RESULTS

5.2.1 The \textit{kpsE} Genes From \textit{E. coli} K5 And K7 Are Highly Conserved

The nucleotide sequence of the entire K5 \textit{kps} locus has now been determined. However, at the start of the project described in this thesis, it had initially been decided that the nucleotide sequence of region 1 of the K7 \textit{kps} locus should be determined, since the K7 region 1 genes had previously been more intensively studied than those of other \textit{kps} loci (Roberts et al., 1988a). In contrast, our collaborators (Prof. K. Jann and his group at the Max-Planck-Institute, Freiburg, Germany) were studying the biochemical pathway involved in the synthesis of the K5 antigen. Consequently, it was later decided that all future nucleotide sequence analysis should utilize the K5 \textit{kps}
locus, so that biochemical and genetic studies could be carried out simultaneously for the same K antigen. The nucleotide sequence of region 1 from the K5 kps locus was subsequently determined by C. Pazzani (Pazzani et al., 1993). Meanwhile, the nucleotide sequence of a 1.6kb fragment from region 1 of the K7 kps locus had been determined as described below.

The cosmid pGB108 encodes the K7 antigen gene cluster (Roberts et al., 1986) and a 6.2kb KpnI-SalI fragment, encoding part of region 1, was previously subcloned from this cosmid into the vector pUC19, generating the plasmid pGB148 (Boulnois and Roberts, unpublished data) (Figure 5.1). pGB148 was used to facilitate the sequencing of a 1.6kb HindIII-ClaI fragment, which had previously been shown to be important for the export of K antigen (Roberts et al., 1988a). A detailed restriction endonuclease cleavage map was generated for the insert in pGB148. Then a series of DNA fragments were subcloned from pGB148 into M13mp18 and M13mp19 and sequenced using the M13-40 primer and appropriate oligonucleotide primers. The sequencing strategy used can be seen in Figure 5.1. It must be noted that the nucleotide sequence was mostly only determined in one strand.

The HindIII-ClaI fragment was found to be 1617bp in length and analysis of the DNA sequence revealed the presence of two open reading frames, ORF1 and ORF2 (Figure 5.2). ORF1 may start upstream of the sequence determined in this study and analysis of the available sequence shows that the ORF is at least 705bp long, encoding a putative polypeptide of 235 amino acids, with a predicted molecular mass of 25.6kDa. The nucleotide sequence of ORF1 was found to have 99.7% DNA homology (two mismatches out of 705bp) with an open reading frame located upstream of kpsE in the equivalent region of the K5 kps locus (Figure 5.2; Section 1.4.4.1). One of the two mismatches identified was translationally silent, whereas the other one resulted in a non-conservative amino acid substitution. Thus, the products encoded by the two ORFs were 99.6% identical over 235 amino acids. Unfortunately, the function of the product encoded by the ORF is not known. However, it does not appear to be essential for expression of the K5 capsule, since strains
harbouring the K5 kps locus with a 328bp deletion in this region, from the HindIII site to the SmaI site (Figure 5.1), are encapsulated (Pazzani, 1992). Nevertheless, this does not rule out the possibility that the protein may still play a role in capsule biogenesis, for example, it may have a regulatory function. Alternatively, the gene may represent an adjacent hitch-hiking locus (Section 1.6.1), which happened to be acquired by the bacterium at the same time as the capsule gene cluster. Interestingly, the amino acid sequence of the K5 ORF was found to have 42.3% identity and 83.9% similarity, over 137 amino acids with the GutQ protein of *E. coli* (Pazzani, 1992; Yamada et al., 1990). The gutQ gene is thought to be part of the glucitol operon associated with glucitol transport and fermentation, but unfortunately the function of the GutQ protein is not known.

ORF2 starts 71bp downstream of the stop codon of ORF1, at the AUG initiation codon at coordinate 781 and extends beyond the available sequence (Figure 5.2). Thus, ORF2 is at least 837bp long and encodes a putative polypeptide consisting of a minimum of 279 amino acids, with a predicted molecular mass of 31.4kDa. A potential Shine-Dalgarno sequence could not be identified 5' to ORF2. The nucleotide sequence of ORF2 was found to have 98.7% DNA homology (11 mismatches out of 837bp) with the *kpsE* gene located in region 1 of the K5 kps locus (Figure 5.2; Section 1.4.4.1). Unfortunately, the K7 nucleotide sequence determined in this study did not extend far enough to include DNA homologous to the last 309bp of the *kpsE* gene. Only two of the observed mismatches altered the amino acid sequence, with one resulting in a functionally conservative amino acid substitution and the other in a non-conservative change. Thus, the partial *KpsE* and *KpsE* proteins were 99.3% identical and 99.6% similar over 279 amino acids. The precise role of the *KpsE* protein has not yet been established. However, the available evidence would suggest that the protein functions in the translocation of capsular polysaccharide from the periplasm to the outer surface of the cell (Section 1.4.4.1).

Thus, it can be seen that the DNA sequence of the two ORFs present in the 1.6kb HindIII-ClaI fragment from the K7 kps
locus is highly conserved with that of the two ORFs found in the corresponding region of the K5 kps locus. In addition, the intervening sequence between the ORFs is also highly conserved, with the only change being the deletion of a single base pair in the K7 sequence (Figure 5.2). The level of DNA homology observed between $kpsE_5$ and $kpsE_7$ is in keeping with that seen with the $kpsE_1$ and $kpsE_3$ genes and is in agreement with the definition of a region 1 gene.

5.2.2  $kpsS$ and $kpsT$ Have a Variable C-Terminus

How the genetic organization of the E. coli group II capsule gene cluster has arisen poses an interesting problem (Section 1.6.2). Analysis of the DNA sequence at the junctions between the variable region 2 cassette and the conserved flanking regions 1 and 3 may shed some light on the mechanisms involved. Perhaps the junction regions contain specific sequences which are able to mediate site-specific recombination events, thereby allowing different region 2-like sequences to be inserted between the common regions 1 and 3. Since the sequence of the entire K5 kps locus has now been determined, the junction regions can be defined more precisely, thereby allowing a detailed analysis to be undertaken. Such a study was carried out by R. Drake (Drake, 1991), wherein the junction regions between regions 1 and 2 and regions 2 and 3 within the K5 kps locus were analysed for the presence of conserved sequences or any other common feature. The junction regions between regions 1 and 2 of the K1 and K4 capsule loci were also analysed and compared with that of the K5 locus. Unfortunately, no significant DNA homology was found in the junction regions either within or between capsule gene clusters (Drake, 1991) and thus they do not appear to represent sites for homologous recombination. However, this does not rule out the possibility that the junction regions may be involved in genetic exchange via some as yet unidentified mechanism.

Although the above study did not reveal anything significant in the junction regions, it did result in some interesting findings regarding the $kpsS$ gene. This gene is located at the 3' end of region 1 (Figure 1.1), adjacent to the junction
region between region 1 and region 2. A BamHI site, located 854bp 5' to the stop codon of kpsS\textsubscript{K1}, is conserved between the K1, K4 and K5 kps loci and during the analysis of the junction regions, the nucleotide sequence of kpsS\textsubscript{K1} and kpsS\textsubscript{K4} was determined from this site to the end of the gene. As would be expected, a comparison of the DNA sequence of the last 854bp of the kpsS\textsubscript{K3} gene with that of the kpsS\textsubscript{K1} and kpsS\textsubscript{K4} genes revealed approximately 96.5% homology between any two of the kpsS sequences (Drake, 1991). However, the kpsS\textsubscript{K1} and kpsS\textsubscript{K4} genes extend beyond the end of the kpsS\textsubscript{K3} gene, resulting in KpsS\textsubscript{K1} being 12 amino acids longer and KpsS\textsubscript{K4} being 14 amino acids longer, at the C-terminus, than the KpsS\textsubscript{K3} protein (Drake, 1991) (Figure 5.3). Furthermore, there is no similarity between the extended C-termini of KpsS\textsubscript{K1} and KpsS\textsubscript{K4}. Thus, the C-terminus of KpsS encoded by the K1, K4 and K5 kps loci appears to be unique for each K antigen. Therefore, since the 3' end of the kpsS gene appears to be serotype specific, the gene does not strictly conform to the definition of a conserved region 1 or 3 kps gene and even has some of the characteristics of a region 2 gene.

As discussed earlier (Section 5.1), the kpsT gene also does not appear to conform to the definition of a conserved region 1 or 3 kps gene, since the kpsT\textsubscript{K1} and kpsT\textsubscript{K3} genes only have 69.4% DNA homology. On closer inspection, it was found that the kpsT gene can be divided into two regions on the basis of the DNA homology which exists between kpsT\textsubscript{K1} and kpsT\textsubscript{K3}. The first 492bp of the two genes are fairly conserved and have 74.6% DNA homology, whereas the last 183bp have only 54.2% DNA homology. This is reflected in the deduced amino acid sequence of the KpsT\textsubscript{K1} and KpsT\textsubscript{K3} proteins. The proteins are relatively highly conserved over the first 164 amino acids, having 82.9% identity and 92.1% similarity, but they are only 38.3% identical and 55.0% similar over the remaining 60 amino acids of KpsT\textsubscript{K3} (Figure 3.5). When the last 30 amino acids of KpsT\textsubscript{K3} are considered, the proteins are only 26.7% identical and 40.0% similar. Furthermore, the KpsT\textsubscript{K3} protein is five amino acids longer than the KpsT\textsubscript{K1} protein at the C-terminus. Thus, like KpsS, KpsT also appears to have a variable C-terminus which is unique for each K antigen. With KpsS there is a precise point
in the C-terminus at which homology between the proteins encoded by different kps loci breaks down and the sequence becomes K serotype-specific. Unfortunately, the situation is not so clear cut for the KpsT protein, although the level of homology is clearly greatly reduced towards the C-terminus of the protein. A comparison of the KpsT proteins from other kps loci may help to clarify the situation.

In summary, kpsT is less highly conserved than any of the other kps genes studied so far. Also kpsS and kpsT have been found to have variable 3’ ends which appear to be unique for each K antigen. Thus, in many respects the kpsS and kpsT genes may be atypical members of the supposedly conserved regions of the kps locus. However, the kps genes from a greater number of different capsule gene clusters must be compared before any conclusions can be drawn. The possible mechanisms by which the variability observed in the kpsS and kpsT genes may have arisen will be discussed at the end of this chapter.

5.2.3 G+C Content Analysis Of The K5 kps Locus

Genomic G+C content varies between different bacterial species and this is thought to be due to differences in the mutation rates of (A.T)→(G.C) and (G.C)→(A.T), where (A.T) refers to an AT or TA base pair and (G.C) refers to a GC or CG base pair. This gives rise to a bias in genetic drift to give the whole genome a G+C content which is characteristic for a given species or group (Sueoka, 1988).

Analysis of the E. coli K5 kps locus has shown that the G+C content of the whole capsule gene cluster is 41.5%. This is lower than the 48–52% G+C content characteristic of the E. coli genome (Orskov, 1984). Interestingly, the G+C content varies considerably throughout the locus and is 50.2% in region 1, 33.4% in region 2 and 40.0% in region 3 (Figure 5.4). Thus, region 1 appears to have a G+C content typical of E. coli and this is fairly evenly distributed across the region, with kpsS having the lowest value of 45.6%. The kflA, kflB, kflC and kflD genes of region 2 have a very low G+C content, atypical of E. coli, and this may suggest that these genes did not originate
in *E. coli*, but were acquired instead from an organism with a similarly low G+C content. In contrast, the putative kfiE gene appears to have a G+C content of 48.4%, which is more typical of *E. coli* and may suggest a different origin for this region 2 gene. In region 3, the G+C content is 44.8% in kpsM and 39.4% in kpsT. However, a more detailed analysis of the kpsT gene has shown that the first 492bp have a G+C content of 43.9%, which is comparable to that found in kpsM, whereas the remaining 183bp have a G+C content of only 27.3%. Thus, the 3' end of the kpsT gene has a G+C content which appears to be consistent with the low G+C content observed in region 2, suggesting that this part of kpsT may have an origin more in common with region 2 than with the remainder of region 3. This will be discussed below.

5.3 DISCUSSION

A limited amount of nucleotide sequence has now been determined for a number of kps genes from different *E. coli* group II capsule gene clusters. On the whole the available sequence information supports the earlier proposal that the genes present in regions 1 and 3 are highly conserved between different kps loci and generally have at least 95% DNA homology (Roberts et al., 1988a). This was certainly found to be the case with the kpsE_{11,17}, kpsE_{6,13+15} (excluding the variable 3' end) and kpsM_{6,13} genes. In addition, since the completion of this project, the nucleotide sequence of the kpsE_{6,11} (Cieslewicz et al., 1993) and kpsD_{11} (Wunder et al., 1994) genes has been reported and comparison of these sequences with those previously available has shown that the kpsD_{11} and kpsD_{13} genes are equally highly conserved, as are kpsE_{11}, kpsE_{6,13} and kpsE_{6,17}. In contrast, it can be seen that the kpsT gene is not nearly so highly conserved, since kpsT_{11} and kpsT_{6,13} only have 69.4% DNA homology. Interestingly, the two genes are fairly conserved at the 5' end, with the first 492bp having 74.6% DNA homology, whereas the 3' end of the genes appears to be rather more variable, with the remaining 183bp having only 54.2% DNA homology. Furthermore, the product of the kpsT_{13} gene is five amino acids longer at the C-terminus than the kpsT_{11} gene product. The kpsS gene has also been found to have a variable
3' end, since $kps_{S4}$ and $kps_{S6}$ encode proteins which are 12 and 14 amino acids longer at the C-terminus, respectively, than the $kps_{S5}$ protein.

Thus, the $kpsS$ and $kpsT$ genes do not strictly conform to the earlier definition of a region 1 or region 3 gene, since the 3' end of these genes appears to be unique for each K antigen. Whether the variable C-terminus of the KpsS and KpsT proteins has any functional significance is unclear. It is possible that the unique C-termini of these proteins are required for the expression of a particular K antigen. However, this is unlikely to be the case, since the genes present in regions 1 and 3 from different capsule gene clusters appear to be functionally interchangeable (Roberts et al., 1986, 1988a). Interestingly, the $kpsS$ and $kpsT$ genes are the two genes which immediately flank region 2 and their variable 3' ends are orientated towards this serotype-specific region. Since it appears that the junction regions between regions 1 and 2 and regions 2 and 3 do not function as sites for homologous recombination, it is tempting to speculate that the different region 2 cassettes may have been generated in $E. coli$ through either homologous or illegitimate recombination events occurring within the 3' ends of the conserved flanking $kpsS$ and $kpsT$ genes and also at sites within region 2 itself. The variable C-termini of KpsS and KpsT may thus have arisen as a result of in-frame fusions between the genes and incoming region 2 elements. Similar speculations have recently been made by Pavelka et al. (1994) and Petit et al. (in press).

Analysis of the K5 $kps$ locus has shown that region 2 has an average G+C content of 33.4%, which is much lower than the 48-52% expected for $E. coli$. Uncharacteristically low G+C contents have also been reported for region 2 or its equivalent in the capsule loci of $E. coli$ K1, $H. influenzae$ type b and $N. meningitidis$ group B (Edwards et al., 1994; Ganguli et al., 1994; Van Eldere et al., 1995). This may suggest that many or all of the polymer synthesis genes in these organisms have been acquired by horizontal gene transfer events from species which typically have a low G+C content. A similar situation may have arisen with the $rfb$ genes of $S. enterica$ serovar Typhimurium.
The rfb gene cluster has a low G+C content varying from 32-45% (the expected G+C content of the Salmonella chromosome is about 51%) and is proposed to consist of several groups of genes, each with a different origin from outside of the genus Salmonella (Jiang et al., 1991). Interestingly, the G+C content of the variable 3' ends of kpsS (analysis of the extended 3' end of kpsS\textsuperscript{vi} and kpsS\textsuperscript{vi} revealed a very low G+C content) and kpsT (as with kpsT\textsuperscript{vi}, the 3' end of kpsT\textsuperscript{vi} has a very low G+C content) is similar to that observed in region 2 and is considerably lower than that found in the rest of regions 1 and 3. This is in agreement with the above suggestion that the variable 3' ends of these genes may in fact originate from region 2 cassettes which have been inserted between the conserved regions 1 and 3 via recombination events within kpsS and kpsT. However, before any further speculations are made regarding the variability of kps genes and the organization and evolution of the different kps loci, more sequence information should be determined for the kps genes of different group II capsule gene clusters.
Figure 5.1 Physical maps of the cosmids pGB110 and pGB108 carrying the biosynthesis genes for the K5 and K7 antigens, respectively. The open boxes below the lines refer to vector sequence. The boxes above pGB110 labelled 1-3 refer to the three functional regions involved in the production of the K5 antigen and the boxes labelled E D U C S represent the kps genes identified in region 1. The restriction endonuclease cleavage maps of pGB110 and pGB108 were aligned on the basis of the common HindIII and BamHI sites present in region 1 of each capsule locus. pGB148, a subclone of pGB108 is shown, with the broken lines highlighting the enlargement (on a scale of 10:1) of the 1.6kb HindIII-ClaI fragment within pGB148 which was subjected to nucleotide sequence analysis. The sequencing strategy used is illustrated by the short horizontal lines in brackets labelled Forward and Reverse. The two partial open reading frames identified in this region are represented by the open boxes labelled ORF1 and ORF2 (kpsE) and the arrows indicate the direction of their transcription.

Restriction endonuclease sites: B, BamHI; C, ClaI; E, EcoRI; H, HindIII; K, KpnI; M, SmaI; S, SalI; V, EcoRV. Only the relevant SalI site is shown. The scale is shown in kb.
Figure 5.2 The nucleotide sequence of the 1.6kb HindIII-ClaI fragment of region 1 of the E. coli K7 capsule gene cluster
Figure 5.2 The nucleotide sequence of the 1.6kb HindIII-ClaI fragment, with the amino acid sequence of the two partial ORFs, ORF1 and kpsE (ORF2), shown underneath. The initiation codon of kpsE and the encoded methionine are shown in bold. The stop codon of ORF1, indicated by the asterisk, is also shown in bold. Restriction sites are shown in bold italics. Nucleotide differences which exist in the sequence from the homologous region of the K5 capsule gene cluster are indicated above the K7 nucleotide sequence. The amino acids which result from these differences are indicated using the three-letter amino acid abbreviation. The double underlined dash present in the K7 nucleotide sequence at coordinate 726 indicates the position of a 1bp deletion found in this non-coding region when compared to the K5 sequence.
Figure 5.3 The variable C-terminus of the KpsS protein

K4 FRCYLLVKTQVNAVYGMIFNKIKKVR
K1 FRCYLLMKTQVNNVYFGNTTNCQHNIY
K5 FRCYLLVKTQVNAVYY

:  +  :+++  ++++

Figure 5.3 The C-termini of KpsS_{47}, KpsS_{55}, and KpsS_{55} were aligned using the criteria of the GCG program GAP. A plus sign indicates a non-conservative amino acid substitution and a pair of dots denotes a conservative amino acid change.
Figure 5.4 The G+C content of the *E. coli* K5 capsule gene cluster

<table>
<thead>
<tr>
<th>kps</th>
<th>kfi</th>
<th>kps</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>50.2</td>
<td>33.4</td>
<td>40.0</td>
</tr>
</tbody>
</table>

G+C %

Figure 5.4 The genetic organization of the *E. coli* K5 capsule gene cluster is shown, drawn to scale. The boxes labelled 1-3 refer to the three functional regions that direct the synthesis of the K5 antigen. The boxes labelled with capital letters represent the genes identified in each region and the appropriate gene designation is shown above the respective region. The average G+C% content is shown for regions 1-3 and also for individual genes and any intervening non-coding DNA (hashed areas). Restriction endonuclease sites: G, *BglII*; H, *HindIII*; V, *EcoRV*. Only relevant sites are shown. The scale is shown in kb.
CHAPTER 6
DISCUSSION

Previous molecular analysis of the *E. coli* kps locus has shown that the capsule genes are organized into three functional regions (Boulnois et al., 1987; Roberts et al., 1988a). Region 2 is serotype specific and encodes the enzymes necessary for polymer synthesis, whereas regions 1 and 3 are conserved between different capsule gene clusters and are proposed to mainly encode products involved in translocation of the polysaccharide to the cell surface. In order to further investigate the biosynthesis of group II capsular polysaccharides, the nucleotide sequence of the whole of the *E. coli* K5 capsule gene cluster has been determined and a preliminary characterization of the encoded proteins undertaken, in conjunction with C. Pazzani (Pazzani et al., 1993; Petit et al., in press; Smith et al., 1990). This is the first *E. coli* kps locus to be sequenced in its entirety.

The *E. coli* K5 capsule gene cluster is approximately 17.5kb in size and encodes at least 11 proteins (Figure 6.1a). Region 1, which is around 7.7kb in size, was found to consist of a minimum of five open reading frames: 5' *kpsE kpsD kpsU kpsC kpsS* 3', possibly organized in a single transcriptional unit (Pazzani et al., 1993) (Section 1.4.4.1). The putative products encoded by these genes have been visualized as proteins with molecular masses consistent with those predicted from the nucleotide sequence data. The recent analysis of recombinant *E. coli* strains with defined defects in region 1 genes has helped to give some indication of the possible functions of the encoded proteins (Bronner et al., 1993a). Thus, since mutations in *kpsE* and *kpsD* result in the accumulation of periplasmic polysaccharide, it has been suggested that the KpsE and KpsD proteins may play a role in the translocation of the polysaccharide from the periplasm to the bacterial cell surface. With respect to the *kpsU* gene, biochemical analysis
has shown that it encodes a capsule specific CMP-KDO synthetase enzyme (Bronner et al., 1993a; Rosenow et al., 1995b). This enzyme presumably accounts for the elevated level of CMP-KDO synthetase activity found to be associated with E. coli strains expressing group II capsules (Section 1.3.1.1). Unfortunately, the roles of the KpsC and KpsS proteins in capsule biogenesis are not clear. Mutations in kpsC and kpsS result in the presence of intracellular polysaccharide, which lacks KDO and phosphatidic acid and is associated with areas of the cytoplasm that appear to have low electron density (Bronner et al., 1993a). These findings could suggest that KpsC and KpsS function in the substitution of the polymer with KDO and/or phosphatidic acid and this would be in agreement with the proposed role of the LipA and LipB proteins of N. meningitidis. These meningococcal proteins, which have homology with KpsC and KpsS, are thought to direct the phospholipid substitution of the group B capsular polysaccharide (Frosch and Müller, 1993).

In addition to the aforementioned five ORFs, a partial open reading frame has been identified upstream of kpsE. Interestingly, this ORF does not appear to be essential for K5 capsule expression, but it does appear to be highly conserved between different capsule gene clusters (Chapter 5). The significance of this is not yet understood.

Region 2 of the K5 kps locus is approximately 7.9kb in size and consists of at least four genes kfiA, kfiB, kfiC and kfiD (Petit et al., in press) (Chapter 4). The transcriptional organization of this region has recently been investigated and several overlapping transcripts have been detected (Petit et al., in press). Proteins with molecular masses consistent with those predicted from the nucleotide sequence data have been identified using E. coli minicells. Using the technique of double immunodiffusion, region 2 alone has been shown to be able to direct the synthesis and polymerization of K5 polymer. Similarly, region 2 of the K1 kps locus was recently found to direct the synthesis of polysialic acid in the absence of region 1 and region 3 encoded proteins (Annuñziato et al., 1995). Searches of the available protein sequence databases, together with biochemical studies carried out since the completion of this project (Petit et al., in press), have
helped to identify the function of two of the proteins encoded by region 2 of the K5 kps locus. Firstly, the KfiD protein has been demonstrated to be a UDP-glucose dehydrogenase responsible for the synthesis of UDP-glucuronic acid, one of the nucleotide sugar precursors required for biosynthesis of the K5 polymer. Secondly, the KfiC protein appears to be a bifunctional K5 glycosyltransferase, capable of directing the transfer of both N-acetylglucosamine and glucuronic acid, the two sugar components of the K5 polymer. Unfortunately, the function of the KfiA and KfiB proteins in capsule biogenesis is unknown, although some preliminary evidence has recently become available which suggests that the KfiA protein may catalyse the sugar transferase reaction required to initiate K5 polymer synthesis (Roberts, personal communication). In addition to the four genes described above, region 2 of the K5 kps locus may encode a further small open reading frame, which I have termed kfiE. Whether kfiE represents a functional ORF is unclear. A DNA fragment containing potential promoter sequences located upstream of the ORF was found to have promoter-like activity, which suggests that the ORF is expressed. However, a protein with a molecular mass corresponding to that predicted for the KfiE protein was not seen in the minicell experiments and recent analysis (Petit et al., in press) indicates that the protein is not required for K5 capsule expression. The situation with regards to this putative open reading frame remains to be resolved.

Region 3 of the K5 kps locus is approximately 1.8kb in size and consists of two genes kpsM and kpsT, possibly organized in a single transcriptional unit (Smith et al., 1990) (Chapter 3). The KpsM and KpsT proteins were visualized using E. coli minicells and these experiments demonstrated that KpsM has a much faster mobility in SDS-polyacrylamide gels than would be expected from its predicted molecular mass. This is in keeping with the hydrophobic nature of the KpsM protein (Section 3.2.4.2), since such proteins frequently migrate anomalously fast during SDS-PAGE. Preliminary characterization of the region 3 encoded proteins has shown that KpsM has the characteristics of a highly hydrophobic integral membrane protein, whilst KpsT has been found to have a potential
conserved ATP-binding cassette. Thus, KpsM and KpsT appear to have many properties in common with members of the ABC family of membrane transporters. Consequently, since region 3 had previously been implicated in the translocation of polymer across the inner membrane (Boulnois and Jann, 1989), it has been proposed that the region 3 proteins constitute an inner-membrane capsular polysaccharide export complex, wherein the KpsT protein couples ATP-hydrolysis to the KpsM mediated transport process (Smith et al., 1990). In order to fulfil the criteria of an archetypal ABC transporter, the proteins would need to be present as homodimers in a tetrapeptide inner-membrane complex.

Thus, the genetic organization of the *E. coli* K5 kps locus has been established and putative functions have been proposed for many of the encoded proteins. A schematic diagram showing the potential arrangement of the Kps and Kfi proteins within the cell, together with their possible function based on the information available to date is shown in Figure 6.1b. It seems likely that many, if not all of the proteins involved in K5 capsule biogenesis will function together in an inner-membrane associated multicomponent complex, in which the biosynthetic machinery is intricately linked with the transport apparatus. Unfortunately, biosynthesis of the group II capsular polysaccharides of *E. coli* is a complex process and many questions remain unanswered. However, knowledge of the genetic organization of the K5 kps locus is now being used to generate strains carrying defined mutations in individual genes, in order to try to fully elucidate the biochemical pathway involved in the biosynthesis of the K5 polymer (Bronner et al., 1993a; Roberts, personal communication).

There is a vast structural diversity of *E. coli* group II capsular polysaccharides and this appears to have arisen as a result of the insertion of different region 2 polymer synthesis gene cassettes between the conserved flanking regions 1 and 3. Unfortunately, the mechanisms by which such genetic exchange of region 2-like elements has occurred have not yet been determined. However, analysis of kpsS and kpsT, the two genes which immediately flank region 2, has shown that they both have
a variable 3' end (Chapter 5) and this may suggest that the
different region 2 cassettes have been acquired by the *E. coli*
*kps* locus through either homologous or illegitimate recombi-
ation events occurring within the 3' ends of the *kpsS* and *kpsT*
genes and also possibly at sites within region 2 itself. The
variable 3' ends of these genes may thus have arisen as a
result of in-frame fusions between the genes and incoming
region 2 elements.

The nucleotide sequence analysis of the K5 *kps* locus has shown
that region 2 has an uncharacteristically low G+C content and
similar findings have been reported for the K1 *kps* locus
(Edwards et al., 1994). This may suggest that many or all of
the polymer synthesis genes present in region 2 of these two
*kps* loci have been acquired by horizontal gene transfer events
from other species which typically have a low G+C content.
Uncharacteristically low G+C contents have also recently been
observed for the polymer synthesis genes of the *H. influenzae*
type b and *N. meningitidis* group B capsule gene clusters
(Edwards et al., 1994; Ganguli et al., 1994; Van eldere et al.,
1995), perhaps suggesting a similar origin for these capsule
genes. Interestingly, analysis of the *cap* locus of *H.
influenzae* and the meningococcal *cps* locus has shown that they
have a similar genetic organization to the *kps* locus of *E. coli*
(Frosch et al., 1989; Kroll et al., 1989) (Figure 1.2). In each
case, a serotype-specific cassette of polymer synthesis genes
is flanked by regions proposed to be involved in non-serotype-
specific functions such as polymer export. Furthermore, the
recent nucleotide sequence analysis of the capsule gene loci of
these three unrelated bacteria has shown that the non-serotype-
specific regions appear to encode a number of homologous
proteins (Brophy, unpublished results; Frosch and Müller, 1993;
Frosch et al., 1991; Kroll et al., 1990; Pazzani et al., 1993;
Smith et al., 1990) (Section 1.5.2; Section 3.2.5). However, it
is also apparent that some of the proteins do not have
functional equivalents in each of the loci and the relative
position of the genes encoding these proteins is not always the
same. These findings are summarized in Figure 6.2.
Thus, it can be seen that the genetic organization of the capsule loci of *H. influenzae* type b and *N. meningitidis* group B is virtually identical, except for the fact that a cassette of LOS biosynthesis genes appears to have been inserted between regions A and B of the meningococcal *cps* locus, and the proteins encoded by the non-serotype-specific regions of the two loci have been shown to have extensive homology (Table 1.4). In contrast, although the products of the *E. coli kpsE*, *kpsC*, *kpsS*, *kpsT* and *kpsM* genes appear to have homology with proteins encoded by the *cap* locus of *H. influenzae* and the *cps* locus of *N. meningitidis*, the level of homology observed is generally much lower (Table 3.1; Section 1.4.4.1) and the organization of the genes within the *kps* locus is very different to that seen with the homologous genes in the *cap* and *cps* loci (Figure 6.2). Furthermore, the *kps* locus does not appear to encode a Kps equivalent to the BexD protein of *H. influenzae* and the CtrA protein of *N. meningitidis* and similarly, genes encoding proteins with homology to KpsD and KpsU have not been identified in the *cap* and *cps* loci. Whether these differences are solely the result of genera-specific requirements or are partly due to the acquisition of different, but functionally related genes is not entirely clear. Certainly, the function of the capsule specific CMP-KDO synthetase encoded by *kpsU* appears to be unique to the biosynthesis of *E. coli* group II capsular polysaccharides. Nevertheless, despite the apparent differences in the genetic organization of the *kps*, *cap* and *cps* loci, the homology which exists between the capsule genes of *E. coli*, *H. influenzae* and *N. meningitidis* does perhaps suggest that capsule production in these unrelated species has a common evolutionary origin. Furthermore, the observation that the *cap* locus of *H. influenzae* and the *cps* locus of *N. meningitidis* are much more closely related to each other than to the *E. coli kps* locus may suggest that their divergence from a possible common ancestral locus has occurred relatively recently.

In conclusion, the work described in this thesis, together with the work of C. Pazzani (Pazzani, 1992), has hopefully contributed towards a better understanding of the processes involved in the biosynthesis of *E. coli* group II capsular
polysaccharides. The genetic organization of the K5 kps locus has been established and some preliminary information has been generated regarding the possible functions of many of the encoded proteins. However, many challenges still lie ahead and continued investigation into the specific functions of the kps encoded proteins and their interaction with each other is required in order to attain a thorough understanding of the mechanisms of capsule biogenesis. It is hoped that this work will be of use in such future experiments.
Figure 6.1 The genetic organization of the *E. coli* K5 capsule gene cluster and biosynthesis of the K5 antigen

a) The genetic organization of the *E. coli* K5 capsule gene cluster is shown, drawn to scale. The boxes labelled 1-3 refer to the three functional regions that direct the synthesis of the K5 antigen. The coloured boxes labelled with capital letters represent the genes identified in each region and the relevant gene designation is written above the respective region. The arrows indicate the direction of transcription and the molecular weight of the encoded gene products is shown in kDa. The unlabelled white box represents the putative *kfiE* gene which does not appear to be required for K5 capsule expression. The following genes were sequenced by C. Pazzani: *kpsE, kpsD, kpsU, kpsC, kpsS, kfiD* and *kfiC* (Pazzani, 1992).

b) A schematic diagram showing the possible location and function of the proteins encoded by the *E. coli* K5 capsule gene cluster, based on the knowledge available to date. UDP-N-acetylglucosamine, represented by the red triangles, is synthesized by a housekeeping gene, whereas the synthesis of UDP-glucuronic acid (yellow squares) is catalysed by the *KfiD* protein. *KfiA* and/or *KfiB* may catalyse the initial sugar transferase reaction, with the subsequent polymerization reaction being directed by the *KfiC* protein. The function of the *KpsC* and *KpsS* proteins is unclear, but they may catalyse the addition of KDO (green circles) and/or phosphatidic acid (black circles) to the K5 polymer. This process requires the nucleotide derivative of KDO, CMP-KDO, which is synthesized by the *KpsU* protein. The K5 polymer then appears to be transported across the inner membrane via an ABC transporter consisting of the *KpsM* and *KpsT* proteins. Subsequently, the *KpsE* and *KpsD* proteins may direct the translocation of the polymer through the periplasm and possibly also onto the cell surface. Unfortunately, the means by which the K5 polymer is transported across the outer membrane remain unclear.

Abbreviations: OM, outer membrane; P, periplasm; IM, inner membrane; C, cytoplasm.
Figure 6.1

a)

<table>
<thead>
<tr>
<th>kps</th>
<th>kfi</th>
<th>kps</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

**K5**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>D</td>
<td>U</td>
<td>C</td>
</tr>
<tr>
<td>43</td>
<td>60</td>
<td>27</td>
<td>76</td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>D</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>44</td>
<td>60</td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>66</td>
<td>27</td>
<td>25</td>
<td>30</td>
</tr>
</tbody>
</table>

2kb

b)

OM

P

IM

C
Figure 6.2 The genetic organization of the capsule gene clusters of *E. coli*, *H. influenzae* and *N. meningitidis*

**kps**

<table>
<thead>
<tr>
<th>export</th>
<th>biosynthesis</th>
<th>export</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

**cap**

<table>
<thead>
<tr>
<th>export</th>
<th>biosynthesis</th>
<th>export</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

**cps**

<table>
<thead>
<tr>
<th>regulation</th>
<th>export</th>
<th>biosynthesis</th>
<th>LOS synthesis</th>
<th>PI-substitution</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>C</td>
<td>A</td>
<td>D</td>
<td>B</td>
</tr>
</tbody>
</table>

2kb

**Figure 6.2** The genetic organization of the *E. coli* K5 kps locus, the *H. influenzae* type b cap locus and the *N. meningitidis* group B cps locus is shown, drawn approximately to scale. The different functional regions of each locus are represented by the boxes labelled 1 to 3 and A to E, with the proposed function of each region written above. LOS refers to lipooligosaccharide and PI refers to phospholipid. The coloured boxes labelled with capital letters represent the genes encoded by the non-serotype specific regions of the different loci and the arrows indicate the direction of their transcription. The appropriate gene designation is written alongside. Boxes with the same colour indicate that the encoded gene products have been shown to have amino acid sequence homology.
REFERENCES


for polymerization resembling that of ribosomes and fatty acid synthetase, with a novel mechanism for determining chain length. Mol. Microbiol. 7:725-734.


Birnboim, H.C. and Doly, J. (1979) A rapid alkaline extraction
procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 1:1513-1523.


Jann, K. and Jann, B. (1985) Cell surface components and


154


Lindberg, A.A. (1977) Bacterial surface carbohydrates and


McGuire, E.J. and Binkley, S.B. (1964) The structure and


**Escherichia coli** K5, which is involved in polysaccharide export. *J. Bacteriol.* 177:1137-1143.


Schneerson, R., Bradshaw, M., Whisnant, J.K., Myerowitz, R.L., Parke, J.C. and Robbins, J.B. (1972) An *Escherichia coli* antigen cross-reactive with capsular polysaccharide of


Vann, W.F., Schmidt, M.A., Jann, B. and Jann, K. (1981) The structure of the capsular polysaccharide (K5 antigen) of


