Regulation of *Escherichia coli* K5 Capsular Polysaccharide Expression

Thesis submitted for the degree of
Doctor of Philosophy

by

Mark Paul Stevens
Department of Microbiology and Immunology
University of Leicester

December 1995
Regulation of *Escherichia coli* K5 Capsular Polysaccharide Expression

Mark P. Stevens

Abstract

The role of known regulatory genes in expression of the *Escherichia coli* K5 (group II) capsular polysaccharide was investigated. Capsule expression was found to require the *rfaH* gene, which regulates several virulence and fertility genes in *E. coli*. None of the mutations studied induced capsule production at 18°C. By using chromosomal reporter gene fusions and by direct detection of RNA, RfaH was shown to activate transcription of the serotype-specific region 2 genes. Expression of region 1 and 3 genes, which are conserved in group II capsule gene clusters, was not regulated by RfaH. A model is presented in which RfaH prevents the termination of region 3 transcripts such that they continue into region 2. Transcription antitermination as this is called, may serve to boost the amount of region 2 products to a level at which efficient polysaccharide biosynthesis can occur. By reverse transcriptase-PCR a transcript spanning the gap between regions 2 and 3 was detected. RNase protection and primer extension analysis were used to locate the promoter upstream of region 3. Transcription was found to initiate from the same site in an *rfaH* mutant. Approximately 1.2 kb of DNA sequence 5' of region 3 was obtained and the end of the capsule gene cluster defined.

A short nucleotide sequence termed JUMPstart, located upstream of region 3 in group II capsule gene clusters, was found to be homologous to sequences 5' of RfaH-regulated operons. A deletion of this sequence abolished expression of the K5 antigen at 37°C and produced a pattern of transcription of the capsule genes similar to that seen in an *rfaH* mutant. This suggests that the JUMPstart sequence may be needed for the action of RfaH. The JUMPstart sequence is also found upstream of many other polysaccharide biosynthesis genes in enteric bacteria and this may indicate a common mode of regulation of these genes by RfaH.
Acknowledgements

I am very grateful to Professor Ian Roberts for his careful supervision and for critically reading the manuscript. I wish to thank Professor Klaus Jann and Dr Barbara Jann and members of their laboratory for analysis of bacterial strains and for the interest they have shown in this study. Special thanks to everyone in the laboratory for the support, advice and practical help I received. This thesis is dedicated to my family for the constant encouragement they have given me.
Contents

1 Introduction ............................................................................................... 1

1.1 General features of bacterial polysaccharides .......................................... 1
1.2 Polysaccharide antigens of *Escherichia coli* .............................................. 2
1.3 Functions of bacterial polysaccharides ....................................................... 4
1.3.1 Prevention of desiccation ............................................................................ 4
1.3.2 Attachment to surfaces .................................................................................. 4
1.3.3 Virulence .......................................................................................................... 5
1.3.3.1 Resistance to complement-mediated killing ............................................ 6
1.3.3.2 Protection against phagocytosis ................................................................. 7
1.3.3.3 Immunogenicity .............................................................................................. 7
1.4 K antigens of *Escherichia coli* ....................................................................... 8
1.4.1 Structure of E. coli K antigens ...................................................................... 9
1.4.2 Mode of expression of E. coli K antigens .................................................... 10
1.4.3 Cell surface attachment of E. coli K antigens ........................................... 12
1.4.4 Genetic determination of E. coli K antigens ............................................... 13
1.4.5 Group III K antigens ........................................................................................ 13
1.5 Genetics and biochemistry of group II capsule production in E. coli . . . . . . 13
1.5.1 Genetic organisation of group II capsule gene clusters . . . . . . . . . . . . . . . 13
1.5.2 Region 1 of the E. coli K5 capsule gene cluster ......................................... 15
1.5.2.1 Transcriptional organisation of region 1 ..................................................... 15
1.5.2.2 KpsF .................................................................................................................. 17
1.5.2.3 KpsE .................................................................................................................. 17
1.5.2.4 KpsD .................................................................................................................. 18
1.5.2.5 KpsU .................................................................................................................. 19
1.5.2.6 KpsC .................................................................................................................. 21
1.5.2.7 KpsS .................................................................................................................. 21
1.5.3 Region 3 of the E. coli K5 capsule gene cluster ......................................... 22
1.5.4 Region 2 of the E. coli K5 capsule gene cluster ......................................... 24
1.5.4.1 Transcriptional organisation of region 2 ..................................................... 24
1.5.4.2 KfiA and KfiB ................................................................................................. 24
1.5.4.3 KfiC .................................................................................................................. 25
1.5.4.4 KfiD .................................................................................................................. 25
1.5.5 Genetic organisation of *H. influenzae* and *N. meningitidis* capsule gene clusters .................................................................................................................. 26
1.5.6 Genetic organisation of group III capsule gene clusters .......................... 29
1.5.7 Genetic basis of diversity at the serA-linked capsule locus .................... 30
1.6 Biosynthesis of E. coli group II K antigens ............................................ 31
1.6.1 Model for expression of the E. coli K5 antigen .................................. 32
1.7 Regulation of gene expression in bacteria .............................................. 34
1.7.1 Transcriptional control of gene expression ......................................... 35
1.7.1.1 Influence of DNA topology ............................................................. 35
1.7.1.2 Alternative RNA polymerase σ factors ........................................... 37
1.7.1.3 DNA rearrangements ........................................................................ 38
1.7.1.4 Transcript processing and stability .................................................... 39
1.7.1.5 Attenuation and antitermination ....................................................... 39
1.7.2 Translational control of gene expression ............................................. 42
1.7.3 Two-component regulatory systems ..................................................... 43
1.7.4 Temperature regulation of gene expression .......................................... 43
1.8 Regulation of extracellular polysaccharide expression in bacteria .......... 45
1.8.1 Regulation of colanic acid biosynthesis in E. coli ............................... 45
1.8.2 Regulation of alginate biosynthesis in Pseudomonas aeruginosa .......... 47
1.8.3 Regulation of lipopolysaccharide biosynthesis in E. coli and S. typhimurium ................................................................. 48
1.9 The aims of this thesis ............................................................................. 50

2 Materials and methods ..................................................................51

2.1 Bacterial strains and plasmids .............................................................. 51
2.1.1 Growth media and conditions ............................................................... 51
2.2 Transformation of E. coli ................................................................. 56
2.2.1 Preparation of competent cells ............................................................. 56
2.2.2 Transformation with plasmid DNA ...................................................... 56
2.2.3 Transformation with bacteriophage DNA ........................................... 57
2.2.4 Transformation of E. coli by electroporation ....................................... 57
2.3 Transduction of E. coli using bacteriophage P1vir .................................. 58
2.3.1 Preparation of bacteriophage P1vir lysates ......................................... 58
2.4. Conjugation of E. coli ........................................................................ 58
2.5 Transposon mutagenesis of the F factor .............................................. 59
2.6 Routine DNA manipulation ................................................................. 59
2.6.1 Agarose gel electrophoresis ............................................................... 59
2.6.2 Phenol extraction and ethanol precipitation of DNA ......................... 60
3 Involvement of known regulatory genes in expression of the
E. coli K5 antigen ................................................................. 78
3.1 Introduction ........................................................................ 78
3.2 Results ................................................................................ 79
3.2.1 Construction of an E. coli K5/K-12 hybrid ......................... 79
3.2.2 The effect of mutations in known regulatory genes on expression of the K5 antigen ........................................ 79
3.2.3 Expression of the K5 antigen is directly regulated by RfaH ...... 82
3.3 Discussion ........................................................................... 85

4 RfaH regulates transcription of the K5 capsule genes ................ 91
4.1 Introduction ........................................................................ 91
4.2 Results ................................................................................ 92
4.2.1 Construction of a luxAB fusion in kfiA in MS101 ............. 92
4.2.2 Construction of a luxAB fusion in kpsT in MS101 ............. 97
4.2.3 Construction of a luxAB fusion in kpsE in MS101 ............. 97
4.2.4 The effect of growth temperature on transcription of kpsE, kfiA and kpsT ......................................................... 100
4.2.5 The effect of RfaH on transcription of kpsE, kfiA and kpsT ...... 103
4.2.6 RfaH regulates the expression of all region 2 genes .......... 109
4.2.7 Plasmid pAS30 restores capsule production in MS106 ....... 109
4.3 Discussion ......................................................................... 111

5 Model for regulation of the K5 capsule gene cluster by RfaH 114
5.1 Introduction ......................................................................... 114
5.2 Results ............................................................................... 115
5.2.1 Transcription from the kfiA promoter is not regulated by RfaH .... 115
5.2.2 The region between kpsT and kfiA is transcribed .............. 116
5.2.3 Determination of the nucleotide sequence upstream of region 3 ... 118
5.2.4 Mapping of the region 3 promoter by RNase protection analysis ... 118
5.2.5 Determination of the 5' end of region 3 transcripts by primer extension ...................................................... 122
5.3 Discussion ........................................................................... 125
6 Involvement of cis-acting sequences in regulation of the K5 capsule gene cluster .......... 131

6.1 Introduction .................................................. 131
6.2 Results .......................................................... 131
6.2.1 The JUMPstart sequence is conserved upstream of RfaH-regulated operons .......... 131
6.2.2 Insertion of an oligonucleotide linker in the JUMPstart sequence in MS101 .......... 133
6.2.3 Deletion of the JUMPstart sequence in MS101 .......... 138
6.2.4 JUMPstart mutations affect biosynthesis of the K5 antigen .......... 142
6.2.5 Transcription of kfiA requires the JUMPstart sequence .......... 144
6.2.6 Plasmid pAS30 restores capsule production in MS137 .......... 147
6.3 Discussion .................................................. 148

7 Discussion .................................................. 153

References .................................................. 162
List of abbreviations

Ap  Ampicillin
bp  Base pairs
Cm  Chloramphenicol
cfu Colony forming unit
EDTA Disodium ethylene diaminetetra-acetic acid
IPTG Isopropyl-β-D-thiogalactopyranoside
kb  Kilobase pairs
kDa Kilodalton
Km Kanamycin
PBS Phosphate buffered saline
PEG Polyethylene glycol
pfu Plaque forming unit
rlu Relative light unit
rpm Revolutions per minute
PCR Polymerase chain reaction
SDS Sodium dodecyl sulphate
Sm Streptomycin
SSC Standard sodium citrate
Tc Tetracycline
TEMED N, N, N', N' - tetramethylethylenediamine
X-gal 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
Chapter 1

Introduction

Polysaccharide capsules are made by many bacteria, some of which cause serious invasive infections of man. Little is known of how the genes for capsule expression are regulated or how polysaccharides are synthesised and translocated to the cell surface. An understanding of these processes will be important both for the rational design of chemotherapeutic agents against encapsulated pathogens and to facilitate the production of polysaccharides of medical or industrial importance in bacteria. In many respects, the genetics and biochemistry of group II capsular polysaccharide production in *Escherichia coli* provides a paradigm for capsule expression in other bacteria. The aim of this study is to determine how the genes for expression of the *E. coli* K5 (group II) capsule are regulated.

1.1 General features of bacterial polysaccharides

Polysaccharides are highly hydrated polymers composed of repeating single units (monosaccharides) joined by glycosidic bonds. A large number of monosaccharides exist, including neutral sugars, uronic acids, polyols and amino sugars, and these may be joined in a variety of combinations (Sutherland, 1985; Jann and Jann, 1990, 1992). In addition, any two monosaccharides may be linked in several ways as glycosidic bonds can be formed at different positions on the sugar ring. Further diversity is achieved through branching of the polysaccharide chain and by substitution of the polysaccharide with both organic and inorganic molecules. Common substituents include phosphate, formate, pyruvate and acetyl groups, although sulphonation has not been observed (Sutherland, 1985). Some bacteria, including *E. coli*, *Klebsiella* spp. and *Streptococcus pneumoniae*, synthesise a large number of polysaccharides with different repeat structures. Little is known of the genetic basis for this diversity or how the synthesis and export of such diverse molecules is achieved by different bacteria. Despite this, different bacteria may synthesise chemically identical polysaccharides, for example the *E. coli* K1 and *Neisseria meningitidis* group B capsular polysaccharides are both homopolymers of α2,8-linked *N*-acetylneuraminic acid (sialic acid) (Bhattacharjee et al., 1975).

In bacteria, polysaccharide often constitutes the outermost layer of the cell exposed to the environment, therefore its properties affect the interactions of the bacterium. The physical nature of polysaccharides affects dehydration, ion transport,
and the colonisation of surfaces (Dudman, 1977; Costerton et al., 1981, 1987). In addition, interactions between bacterial polysaccharides and the immune system are vital in deciding the outcome of certain infections (Cross, 1990; Moxon and Kroll, 1990).

Bacterial polysaccharides differ in their mode of attachment to the cell surface. In Gram negative bacteria, polysaccharide chains may be covalently linked to either lipid A or phospholipid in the outer membrane (Whitfield and Valvano, 1993), whereas in Gram positive bacteria they may be linked to cell wall teichoic acids or the peptide moiety of peptidoglycan (Yeung and Mattingly, 1986; Sorensen et al., 1990). Alternatively, polysaccharides may be loosely associated with the cell surface, forming an amorphous hydrated layer, sometimes referred to as slime, that encompasses not only the producing cell but also cells nearby. Slime polysaccharides are therefore important in the formation of biofilms, creating a microenvironment in which many cells coexist (Costerton et al., 1981, 1987). Examples of slime polysaccharides include the alginate polysaccharide of *Pseudomonas aeruginosa* and colanic acid (M antigen) produced by certain members of the Enterobacteriaceae.

1.2 Polysaccharide antigens of *Escherichia coli*

There are four types of polysaccharide antigen produced by *E. coli*: lipopolysaccharide (O antigens), capsular polysaccharides (K antigens), enterobacterial common antigen (EGA) and colanic acid (reviewed in Whitfield et al., 1994). The structure and composition of the O and K antigens is variable and this has facilitated the serological classification of *E. coli* (Orskov et al., 1977). There are at least 173 O antigens and 80 K antigens of *E. coli*, and these occur naturally in selected combinations (Orskov and Orskov, 1992). Although the O and K antigens may be considered to be serotype-specific antigens within the species, structurally identical polysaccharides may be made in other bacteria. In contrast, ECA and colanic acid are not serotype-specific. ECA is produced by all members of the Enterobacteriaceae except *Erwinia chrysanthemi* and is a heteropolymer of amino-sugars that may be linked to L-glycerophosphatidyl residues in the outer membrane or in some cases core-lipid A (Kuln et al., 1988). In *Shigella sonnei*, cyclic forms of ECA containing 4 to 6 repeating units that are not linked to lipid have also been described (Dell et al., 1984). It is not clear if cyclic forms of ECA are synthesised by other enteric bacteria or how these forms are attached to the cell surface.

Colanic acid is also not unique to *E. coli* and is produced by some *Klebsiella*, *Salmonella* and *Aerobacter* species (Whitfield and Valvano, 1993). Colanic acid is a high molecular weight polymer of glucose, galactose, fucose and glucuronic acid, substituted with acetyl and pyruvate residues (Markovitz, 1977). Evidence for linkage
of colanic acid to lipid A or phospholipid has not been reported and it is believed that the polysaccharide may be held at the cell surface by ionic interactions. The production of colanic acid is increased under conditions of environmental stress, particularly desiccation (Ophir and Gutnick, 1994), low growth temperatures (e.g. 20°C) and limitation of certain nutrients (Markovitz, 1977; Gottesman and Stout, 1991).

Lipopolysaccharide is a complex glycolipid that is an essential component of the outer membrane of Gram negative bacteria (Whitfield and Valvano, 1993). In enteric bacteria, LPS is composed of polysaccharide chains (O antigens) of variable length and composition, linked by a core oligosaccharide to lipid A in the outer leaflet of the membrane (Whitfield and Valvano, 1993; Whitfield et al., 1994). In other Gram negative bacteria, for example Neisseria species, the polysaccharide moiety may be much shorter (Griffiss et al., 1988). This type of LPS is referred to as lipooligosaccharide (LOS).

In E. coli, lipid A consists of a β-glucosaminyl-1,6-glucosamine backbone which is substituted with 4 β-hydroxyristoyl residues (Qureshi and Takayama, 1990). Two of the fatty acid chains are attached directly to each glucosamine residue. A further two myristoyl residues are esterified to the non-reducing end of the hydroxyristoyl residues attached to one of the sugars. The glucosamine residues of lipid A are also substituted with phosphate and phosphomonooester groups (Qureshi and Takayama, 1990).

The LPS core oligosaccharide comprises a conserved inner region nearest lipid A which is common among most E. coli strains and other members of the Enterobacteriaceae, and a more variable outer region to which the O antigen is attached (Whitfield et al., 1994). The inner core contains 2-keto-3-deoxy-D-manno-octulosonic acid (KDO) and L-glycero-D-manno-heptose. KDO was originally thought to be unique to LPS but is now known to occur in certain E. coli capsular polysaccharides (Jann and Jann, 1990). The outer core contains a number of hexose residues and variation in this region gives rise to five core types in E. coli (Jansson et al., 1981). Mutations affecting the composition of the core oligosaccharide often prevent the attachment of O antigen and therefore cause LPS truncation (Schnaitman and Klena, 1993). Most clinical isolates of E. coli produce 'smooth' LPS which contains O antigen. In contrast, the LPS of laboratory strains of E. coli K-12 terminates with the core oligosaccharide and is termed 'rough'. This terminology historically describes the morphology of bacterial colonies expressing the respective LPS types.

Serological studies on E. coli indicated that many strains exhibit polysaccharide antigens distinct from O antigens (Ørskov et al., 1977). These are called capsular polysaccharides, or K antigens, and form an envelope, or capsule, around the cell. E. coli K antigens differ in their mode of attachment to the cell surface, being linked either to phospholipid or core-lipid A (section 1.4.3) (Jann and Jann, 1990). The
thickness of the K antigen layer varies widely (0.1-10μm) (Bayer and Thurow, 1977; Beveridge and Graham, 1991). Some capsules are visible by light microscopy using a negative stain such as India ink, whereas others can only be seen by electron microscopy (Bayer, 1990). Capsules are highly hydrophilic and tend to collapse under the dehydrating conditions used in the preparation of samples for electron microscopy. However, the structure can be stabilised by pretreatment with capsule-specific antibody or by using the freeze-substitution method for sample preparation (Bayer, 1990; Beveridge and Graham, 1991). Due to their anionic nature, capsules can be stained for electron microscopy using electron dense cationic proteins, such as ferritin (Bayer, 1990). Capsular polysaccharide may also be detected using specific antibodies conjugated to colloidal gold particles (immunoelectron microscopy). This has been useful in elucidating the role of proteins involved in capsule expression in *E. coli* (Kröncke *et al.*, 1990a, 1990b; Bronner *et al.*, 1993a, 1993b).

1.3 Functions of bacterial polysaccharides

1.3.1 Prevention of desiccation

It is estimated that at least 95% of the bulk of capsular polysaccharides is water (Costerton *et al.*, 1981). Thus, capsules form a hydrated gel at the cell surface that may offer protection against desiccation (Roberson and Firestone, 1992; Ophir and Gutnick, 1994). Mucoid strains of *E. coli*, *Acinetobacter calcoaceticus* and *Erwinia stewartii* have been shown to be much more resistant to drying than isogenic non-mucoid strains (Ophir and Gutnick, 1994). Production of colanic acid in *E. coli* is induced by desiccation and this has been correlated with elevated transcription of the capsule genes (Ophir and Gutnick, 1994). One might expect that desiccation is sensed by bacteria as an increase in external osmolarity. High external osmolarity is known to induce alginate biosynthesis in *Pseudomonas aeruginosa* (Berry *et al.*, 1989).

1.3.2 Attachment to surfaces

Capsular polysaccharides and slime facilitate the colonisation of surfaces by bacteria and the formation of biofilms (Costerton *et al.*, 1981, 1987). This is largely due to electrostatic attraction between the polysaccharide and surfaces and may benefit bacteria living in oligotrophic environments because nutrients are also attracted to surfaces (Costerton *et al.*, 1981). Biofilms provide protection from predation by phagocytic protozoa and slime moulds and can prevent infection by some bacteriophages (Dudman, 1977). The formation of microbial biofilms is important in the development of dental caries and has many other consequences, ranging from the
fouling of pipes to the colonisation of indwelling catheters (Costerton et al., 1987). Whilst some polysaccharides may promote the interaction of bacteria with surfaces, the opposite may be true in other cases. Recently, it was postulated that a cell-surface polysaccharide of *Proteus mirabilis* facilitates the migration of swarm cells by reducing surface friction (Gygi et al., 1995).

Bacterial polysaccharides may also be involved in establishing symbiotic interactions with a plant. The formation of nitrogen-fixing root nodules on leguminous plants by *Rhizobium* species involves several bacterial polysaccharides (Gray and Rolfe, 1990; Noel, 1992). The major acidic polysaccharide of *R. meliloti* is succinoglycan, a polymer of octasaccharide subunits composed of one galactose residue, seven glucose residues and acetyl, succinyl and pyruvyl substituents (Glucksmann et al., 1993). Mutant strains of *R. meliloti* that do not synthesise succinoglycan induce nodule formation on alfalfa, but do not penetrate or colonise the nodule, indicating that succinoglycan is required for nodule invasion and development (Noel, 1992). *Rhizobium* LPS mutants which lack O antigen or have altered O chain lengths also form empty nodules that do not fix nitrogen (Noel, 1992). Secreted oligosaccharides made by *Rhizobium* species may also be involved in nodulation, by inducing changes in the root hairs that facilitate entry of the bacteria (Gray and Rolfe, 1990).

### 1.3.3 Virulence

Extracellular polysaccharides often confer resistance to non-specific host immunity by several mechanisms (Moxon and Kroll, 1990; Cross, 1990). This is discussed below in relation to the *E. coli* O and K antigens but may occur similarly in other bacteria. There are no reports of the enterobacterial antigen or colanic acid being involved in the virulence of *E. coli* or other Gram negative bacteria. Recently, Russo _et al._ (1995) reported that the expression of colanic acid by an extraintestinal isolate of *E. coli* (O4:K54:H5) did not affect its serum resistance or virulence in mice. It is possible that these polysaccharides are more important for survival of the bacterium outside the host. This does not mean that slime polysaccharides have no role in the pathogenesis of infections by Gram negative bacteria. The production of alginate by strains of *Pseudomonas aeruginosa* causing lung infections is the major cause of morbidity and mortality among patients with cystic fibrosis (Govan, 1988). Alginate is believed to confer resistance to phagocytosis, opsonisation and reactive oxygen intermediates.

In some bacteria, the production of certain cell surface polysaccharides may be turned on or off from one generation to the next. Phase variation, as this is called, occurs with the lipopolysaccharide of *Haemophilus influenzae* type b (Kimura and
Hansen, 1986), and may be important in allowing the bacteria to evade immune responses. This kind of variation has not been observed with the O and K antigens of *E. coli* (Whitfield et al., 1994), although variants of the K1 antigen which differ in O-acetylation of the polysaccharide have been observed (Ørskov et al., 1979). The basis and function of this variation is not known.

1.3.3.1 Resistance to complement-mediated killing

Bacteria can activate complement by either the classical or alternative pathway (Frank et al., 1987). Although the initial steps differ, the final product of both pathways is the membrane attack complex (MAC), a multi-protein complex which causes lysis of Gram negative bacteria by piercing the outer membrane (Joiner, 1985). The classical pathway is activated by the binding of antibodies to bacterial surface components, whereas the alternative pathway (AP) is important in the pre-immune stage of infection when specific antibodies are absent. The AP is initiated by the non-specific binding of the serum protein C3b to the bacterial cell surface (Horstmann, 1992). Once bound, C3b may be activated by interacting with factor B to form the C3 convertase, C3bBb. This results in the deposition of more C3b and initiates MAC assembly. Alternatively, bound C3b can be inactivated by factor I to iC3b, with factor H acting as a cofactor. Bacterial polysaccharides may confer serum resistance by interfering with activation of the AP, however they provide little protection once specific antibody is generated (Moxon and Kroll, 1990).

The length, composition and distribution of the O and K polysaccharides are important in the resistance of *E. coli* to complement-mediated killing. The complement proteins C3b and C5b-9 tend to be bound preferentially by long O polysaccharide chains at a distance from the outer membrane (Forat et al., 1992). This may prevent the MAC from inserting into the outer membrane and/or alter the configuration of the MAC complex. *E. coli* K antigens are also thought to provide a permeability barrier to complement proteins, masking underlying surface antigens that may otherwise activate the AP (Horwitz and Silverstein, 1980). *E. coli* causing invasive infections exhibit selected combinations of O and K antigens and it likely that these polysaccharides act in concert to confer serum resistance (Kim et al., 1986).

Polysaccharides containing sialic acid, including K1, K9 and K92, all contribute to serum resistance by inhibiting AP activation (Van Dijk et al., 1979; Michalek et al., 1988). This is believed to result from their ability to bind factor H, thereby favouring the degradation of C3b deposited on the cell surface to iC3b and breaking the amplification loop of the AP (Michalek et al., 1988). In addition, the K1 polysaccharide has a low affinity for factor B and this inhibits the formation of C3 convertase (Van Dijk et al., 1979). The lipooligosaccharide of certain *Neisseria* species is also substituted
with sialic acid and confers resistance to serum killing (Hammerschmidt et al., 1994), and phagocytosis (Kim et al., 1992).

1.3.3.2 Protection against phagocytosis

K antigens are the most important antiphagocytic factor produced by E. coli, although the basis for this is not well understood. E. coli expressing the K29 antigen are not ingested by phagocytes, even when attached to the phagocyte by lectins, whereas isogenic strains lacking the capsule are efficiently ingested under these conditions (Horwitz and Silverstein, 1980). The negative charge and hydrophilicity conferred by the K antigen are thought to inhibit interaction of phagocytes with the bacteria (Allen et al., 1987; Roberts et al., 1989). K antigens that are highly charged or particularly hydrophilic tend to confer greater resistance to phagocytosis (Moxon and Kroll, 1990).

Phagocytes have receptors for antibody and complement on their surface, therefore the binding of antibody and/or complement to the bacterial surface (opsonisation) facilitates phagocytosis (Joiner, 1985). It follows that O and K antigens which fail to activate complement confer resistance to phagocytosis. K antigens may also mask C3b deposited on O antigens or the cell surface from C3b receptors on the phagocyte. This is not always the case however as antibodies to the O18 antigen can opsonise encapsulated E. coli O18:K1 (Kim et al., 1986). The K1 antigen is a potent inhibitor of phagocytosis and this may be correlated to its failure to activate complement and to bind C3b efficiently (Van Dijk et al., 1979; Allen et al., 1987; Michalek et al., 1988). The K5 capsule does not appear to confer protection against phagocytosis (Cross et al., 1986), even though E. coli K5 strains are highly pathogenic and are frequently isolated from cases of neonatal meningitis, septicemia, pyelonephritis and urinary tract infections (Ørskov and Ørskov, 1992; Nataro and Levine, 1994). The K5 serotypes associated with these infections are limited (O75:K5:H5, O6:K5:H1 and O18ac:K5:H1), and their invasive nature of may reflect the function of other surface antigens, including the O antigen.

1.3.3.3 Immunogenicity

The protection offered by K antigens against non-specific host immunity is generally lost with the development of specific antibody. However, the structures of certain E. coli K antigens prevent an effective antibody response from being generated. This is because they mimic structures that are found within the host. The K1 antigen for example, is structurally identical to carbohydrate moieties of the embryonic form of the neural cell adhesion molecule (N-CAM) (Finne, 1982). Similarly the K5
polysaccharide, a polymer of glucuronic acid and N-acetylg glucosamine, is identical to the first polymeric intermediate in heparin biosynthesis (Vann et al., 1981). Such structural mimicry is not confined to K antigens, for example the E. coli O2 antigen is serologically cross-reactive with kidney tissue (Ørskov et al., 1977).

With the exception of capsular polysaccharides that are structurally similar to host components, most polysaccharides are immunogenic and may be used for the generation of vaccines (Lee, 1987; Jennings, 1990). Vaccines based on purified capsular polysaccharides have been used to prevent a number of bacterial infections of man. The vaccine against meningitis caused by Neisseria meningitidis consists of the purified polysaccharides from serogroups A, C, W135 and Y, which cause 90% of all infections (Cadoz et al., 1985). The situation is more complex with Streptococcus pneumoniae where 23 capsular polysaccharides are needed to provide effective coverage (Jennings, 1990).

A problem with the use of polysaccharide vaccines is their poor immunogenicity in infants (Robbins, 1978). Purified polysaccharides appear to act as T-cell independent antigens in infants, being capable of eliciting only short-lived IgM responses that cannot be boosted by further exposure (Gotschlich et al., 1977). To obviate this problem, polysaccharides can be conjugated to proteins (Robbins and Schneerson, 1990). Infants respond to polysaccharide-protein conjugates by producing a wide range of antibodies of the IgG isotype that may be augmented by repeated exposure (Robbins and Schneerson, 1990). Vaccines against meningitis caused by Haemophilus influenzae type b have been generated using this approach and are now used successfully in many countries (Moxon and Rappuoli, 1990). The immunogenicity of polysaccharide vaccines may also be improved by chemical modification of the polysaccharide (Jennings et al., 1993). It was found that the antibody response of mice to the polysialic acid capsules of E. coli K1 and N. meningitidis group B was improved if the N-acetyl groups were replaced with N-proprionyl residues (Jennings, 1990). This response was most effective when the modified polysaccharide was conjugated to protein and did not affect the specificity of the antibodies for E. coli K1 and group B meningococci (Jennings, 1990).

1.4 K antigens of Escherichia coli

Escherichia coli can express at least 80 different K antigens (Ørskov and Ørskov, 1992). Individual isolates produce only one K antigen and switching of K antigen types is not observed (Ørskov et al., 1977). Serotype specificity of the K antigen may be conferred by the type of linkage between monosaccharides or by substituents, even if the polysaccharide backbone remains invariant. The K1 and K92 antigens for example, are both homopolymers of N-acetylmuraminic acid but are distinguished by the
alternating α2,9 linkage found in the K92 antigen (Egan et al., 1977). Similarly, the K13, K20 and K23 polysaccharides have the same sugar backbone, but their antigenicity is determined by the presence and location of O-acetyl residues (Varna et al., 1983). Chemical and genetic criteria have been used to divide E. coli K antigens into two broad groups, I and II (Table 1.1) (Jann and Jann, 1990). The existence of a third group of K antigens (III or I/II) has been suggested (Finke et al., 1990; Pearce and Roberts, 1995), and this is discussed later. It is important to note that the K-12 designation for E. coli laboratory strains is unrelated to capsule type. These strains lack the genes for production of the group II K12 antigen.

1.4.1 Structure of E. coli K antigens

Group I K antigens typically have higher molecular weights than those of group II and produce thicker capsules (Bayer and Thurow, 1977). Group I K antigens have large repeating units comprising mostly of hexoses and methyl pentoses, although pyruvate residues are common. They may be further divided by the presence of amino sugars. Group IA K antigens do not contain amino sugars and resemble polysaccharides produced by Klebsiella spp. For example, the E. coli K28 and K55 antigens are identical to the K54 and K5 antigens of Klebsiella respectively (Jann and Jann, 1992). Group IB K antigens contain amino sugars and have no obvious counterparts in other bacteria. Colanic acid is sometimes referred to as a group IA capsular polysaccharide owing to structural similarities and the fact that genes for the biosynthesis of colanic acid (cps) and group I K antigens are allelic (Keenleyside et al., 1992). However, colanic acid is loosely associated with the cell surface and should therefore be referred to as a slime polysaccharide.

Group II K antigens contain more acidic components than those of group I, including N-acetylneuraminic acid, phosphate and KDO, and therefore exhibit higher charge densities. This is reflected in the higher electrophoretic mobility of group II K antigens compared with those of group I. In general, group II K antigens resemble the capsules made by H. influenzae and N. meningitidis in their physical and chemical properties, mode of attachment to the cell surface and the organisation of the capsule genes (Jann and Jann, 1990). Group II capsular polysaccharides containing ribitol-phosphate (e. g. K18) or glycerol-phosphate (e. g. K2a) are similar to the cell wall teichoic acids of Gram positive bacteria. The repeat unit structure of E. coli K antigens representative of each group is given in Table 1.2.
### Table 1.1. Grouping of *Escherichia coli* 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 kDa</td>
<td>&lt; 50 kDa</td>
</tr>
<tr>
<td>Acidic component</td>
<td>GlcA, GalA, Pyr</td>
<td>GlcA, NeuNAc, KDO, ManNAcA, Phosphate</td>
</tr>
<tr>
<td>Intergeneric relationship with</td>
<td><em>Klebsiella</em> spp.</td>
<td><em>H. influenzae</em></td>
</tr>
<tr>
<td>Coexpression with</td>
<td>O8, O9, and O20</td>
<td>Many O antigens</td>
</tr>
<tr>
<td>Expressed below 20°C</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Elevated levels of CMP-KDO synthetase</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Lipid at reducing end</td>
<td>Core-lipid A</td>
<td>Phosphatidic acid</td>
</tr>
<tr>
<td>Removal of lipid at pH 6/100°C</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Chromosomal location of genes</td>
<td>Near <em>his</em> and <em>rfb</em></td>
<td>Near <em>serA</em></td>
</tr>
</tbody>
</table>

*Abbreviations:* GalA, galacturonic acid; GlcA, glucuronic acid; KDO, 2-keto-3-deoxy-D-manno-octulosonic acid; ManNAcA, N-acetylmannosaminuronic acid; NeuNAc, N-acetylmuramic acid; Pyr, pyruvate. † Adapted from Jann and Jann (1990).

### 1.4.2 Mode of expression of *E. coli* K antigens

Epidemiological studies have indicated that group I K antigens are expressed by a limited number of O serotypes of *E. coli*, usually O8, O9 and O20 (Ørskov et al., 1977). Group II K antigens can be expressed with many O antigen types, although selected combinations of O and K antigens are associated with pathogenic *E. coli*. An important distinction between the capsular polysaccharides of *E. coli* is that group II K antigens are expressed at 37°C but not below 20°C, whereas group I K antigens are made at all growth temperatures (Ørskov et al., 1984).

The expression of several other virulence factors of *E. coli*, including certain pili and fimbrial adhesins, is also regulated by temperature (Maurelli, 1989). Many human pathogens are able to coordinate expression of their virulence factors in response to temperature (Mekalanos, 1992). This may ensure that factors needed for survival within the host are made only when it is appropriate.

*E. coli* expressing group II capsules have elevated levels of the enzyme CMP-KDO synthetase, which supplies activated KDO for incorporation into polysaccharide (Finke et al., 1989, 1990). The CMP-KDO synthetase activity of *E. coli* with group II K
Table 1.2. Repeating units of some *Escherichia coli* K antigens.

<table>
<thead>
<tr>
<th>K Antigen</th>
<th>Group</th>
<th>Structure</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>K30</td>
<td>IA</td>
<td>$-2\text{-Man}(1,3)\text{-Gal}(1\text{-3}$</td>
<td>Chakraborty et al., (1980)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta\text{-GlcA}(1,3)\text{-Gal}$</td>
<td></td>
</tr>
<tr>
<td>K40</td>
<td>IB</td>
<td>$4\text{-}\beta\text{-GlcA}(1,4)\text{-}\alpha\text{-GlcNAc}(1,6)\text{-}\alpha\text{-GlcNAc}(1\text{-}$</td>
<td>Dengler et al., (1986)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\text{CO.NH (serine)}$</td>
<td></td>
</tr>
<tr>
<td>K1</td>
<td>II</td>
<td>$-8\text{-}\alpha\text{-NeuNAc}(2\text{-}$</td>
<td>McGuire and Binckley, (1964)</td>
</tr>
<tr>
<td>K5</td>
<td>II</td>
<td>$-4\text{-}\beta\text{-GlcA}(1,4)\text{-}\alpha\text{-GlcNAc}(1\text{-}$</td>
<td>Vann et al., (1981)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\text{CO.NH (serine)}$</td>
<td></td>
</tr>
<tr>
<td>K10</td>
<td>III</td>
<td>$-3\text{-}\alpha\text{-Rha}(1,3)\text{-}\beta\text{-QuiNMal}(1\text{-}$</td>
<td>Sieberth et al., (1993)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\text{OAc}$</td>
<td></td>
</tr>
<tr>
<td>K54</td>
<td>III</td>
<td>$-3\text{-}\beta\text{-GlcA}(1,3)\text{-}\alpha\text{-Rha}(1\text{-}$</td>
<td>Hofmann et al., (1985)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\text{CO.NH threonine (serine)}$</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: Gal, galactose; GlcA, glucuronic acid; GlcNAc, N-acetylglicosamine; Man, mannose; NeuNAc, N-acetylneuraminic acid; OAc, O-acetyl group; QuiNMal, 4, 6-dideoxy-4-malonylaminogluucose; Rha, rhamnose.

antigens is at least 5 fold higher than that of unencapsulated strains or strains expressing group I K antigens (Finke et al., 1990). In addition, the CMP-KDO synthetase activity of group II strains is much higher at 37°C than at 18°C (Finke et al., 1989, 1990). All *E. coli* strains possess a CMP-KDO synthetase encoded by the *kdsB* gene, which is involved in core lipopolysaccharide biosynthesis (Goldman and Kohlbremer, 1985). *E. coli* expressing group II K antigens have a second CMP-KDO synthetase, which is encoded by the capsule gene cluster (section 1.5.2.5) (Pazzani et al., 1993a).
1.4.3 Cell surface attachment of E. coli K antigens

E. coli K antigens also differ in their mode of attachment to the cell surface. Group I K antigens are thought to be attached to the cell surface by core-lipid A and therefore resemble LPS (Jann et al., 1992). However, not all of the polysaccharide on the cell surface terminates in lipid A (Jann et al., 1992). SDS polyacrylamide gel electrophoresis of the E. coli K40 (group IB) antigen has indicated that low and high molecular weight species of the K antigen exist. The low molecular weight species comprises short oligosaccharides of the K antigen linked to core-lipid A, whereas the high molecular weight species are longer polysaccharide chains that are not linked to lipid (Jann et al., 1992). This is also true for the E. coli K30 (group IA) antigen (Homonylo et al., 1988). E. coli expressing group I K antigens where little or no polysaccharide is associated with lipid A still make well developed capsules (MacLachlan et al., 1993). In addition, E. coli K30 mutants with truncated lipid A cores make capsules that are indistinguishable from the wild-type when examined by electron microscopy (MacLachlan et al., 1993). This implies that the surface expression of group I K antigens can occur independently of linkage to lipid A. It is possible that capsular polysaccharide not attached to core-lipid A may remain associated with the cell surface via ionic interactions with attached molecules.

In contrast to group I K antigens, group II capsular polysaccharides are thought to be anchored in the outer membrane by phospholipid. Several group II K antigens have an L-glycerophosphatidyl residue attached to their reducing terminus via a phosphodiester linkage (Jann and Jann, 1990). The lipid moiety was first detected on the capsular polysaccharides of E. coli K92 and N. meningitidis groups A, B and C (Gotschlich et al., 1981; Frosch and Müller, 1993), and is also present on the capsular polysaccharides of H. influenzae (Kuo et al., 1985). The enterobacterial common antigen is also linked to phospholipid (Kuhn et al., 1988). In E. coli K12 and K82, the K antigen is linked to phospholipid by a KDO residue, even though KDO is not part of the repeat structure of these polysaccharides (Schmidt and Jann, 1982). The K5 antigen is also linked to phospholipid by a KDO residue at the reducing terminus (Finke et al., 1991). Phospholipid substitution of N. meningitidis and H. influenzae capsular polysaccharides has been shown to be necessary for association of the polysaccharide with isolated outer membrane vesicles (Arakere et al., 1994). In addition, the linkage of N. meningitidis capsular polysaccharides to lipid is required for their translocation to the cell surface (Frosch and Müller, 1993). It is estimated however, that only 20-50% of group II polysaccharide chains are lipid substituted (Jann and Jann, 1990). Again, polysaccharide lacking the membrane anchor may be retained at the cell surface via ionic interactions. The linkage between group II K antigens and phospholipid is acid labile and easily destroyed by heating.
1.4.4 Genetic determination of E. coli K antigens

Genetic analysis, involving Hfr crosses, revealed that the genes for production of group I and II K antigens are not allelic (Ørskov and Ørskov, 1962). Group I capsule genes map at 44 minutes on the E. coli chromosome near his and rfb (Ørskov et al., 1977), and have not been studied in detail. Studies with the K27 antigen indicate that a second trp-linked locus is necessary for the expression of a complete capsule (Schmidt et al., 1977), although this is not the case for the K30 antigen (Laasko et al., 1988). Group II capsule genes are located at 64 minutes on the E. coli chromosome near serA (Ørskov et al., 1977; Vimr, 1991), and are designated kps (formerly kpsA (Ørskov and Nyman, 1974)). The genes for production of several group II capsules have been cloned (Silver et al., 1981; Roberts et al., 1986; Drake et al., 1990), and subjected to detailed analysis.

1.4.5 Group III K antigens

A number of E. coli capsular polysaccharides possess characteristics of both group I and group II K antigens and cannot easily be assigned to either group. These include for example the K2, K10, K11, K19, K54 and K98 antigens. The K10 and K54 antigens were originally classified as group II capsules on the basis of their composition and charge density (Ørskov et al., 1977). This classification was supported by the mapping of the genes for production of the K10 and K54 antigens to serA (Ørskov and Nyman, 1974). However, both the K10 and K54 antigens are made at all growth temperatures, which is typical of group I K antigens (Ørskov et al., 1984). Also, strains expressing these capsules do not have elevated levels of CMP-KDO synthetase (Finke et al., 1990). DNA probes from group II capsule gene clusters do not hybridise to DNA from group III strains (Drake et al., 1993), suggesting that the group III capsule genes are distinct from those of group II even though they map at the same location. Analysis of the cloned K10 and K54 capsule genes supports the existence of at least three groups of capsule gene clusters (section 1.5.6) (Pearce and Roberts, 1995).

1.5 Genetics and biochemistry of group II capsule production in E. coli

1.5.1 Genetic organisation of group II capsule gene clusters

The genes for several group II K antigens have been cloned. These include the genes for the K1 (Silver et al., 1981; Echarti et al., 1983), K4 (Drake et al., 1990), K5, K7, K12 and K92 antigens (Roberts et al., 1986) (reviewed in Boulnois and Roberts, 1990). The K1 and K5 capsule gene clusters occupy approximately 18 kb and have been
extensively studied. The complete nucleotide sequence of the K5 capsule gene cluster is now available (Pazzani et al., 1993a; Petit et al., 1995; Smith et al., 1990; D. Simpson, unpublished results; M. Stevens, this study).

Detailed molecular analysis of group II capsule gene clusters revealed that they have a conserved organisation consisting of three functional regions (Figure 1.1) (Boulnois et al., 1987; Roberts et al., 1988). The capsulation loci of H. influenzae and N. meningitidis have a similar organisation (section 1.5.5). The central region, region 2, is serotype-specific (Roberts, M. et al., 1988). Mutations in region 2 abolish polysaccharide biosynthesis, suggesting that it encodes enzymes for synthesis of the K antigen (Silver et al., 1984; Boulnois et al., 1987; Roberts et al., 1988). In the case of group II K antigens that contain sugars not ubiquitous to E. coli, region 2 also encodes the enzymes for sugar synthesis and activation. Region 2 of the E. coli K1 capsule gene cluster for instance, encodes three genes for the synthesis and activation of N-acetyleneuraminic acid (Vann et al., 1993). There is a broad correlation between the size of region 2 and the chemical complexity of the K antigen. Thus, region 2 encoding for synthesis of the K4 antigen, which has a backbone of -(3)-β-glucuronic acid-(1,4)-β-N-acetylglucosamine-1(- that is substituted at each glucosamine with a β-(3,2) linked fructose residue (Rodriquez et al., 1988), is much larger than the region 2 for the K1 antigen, which is a polymer of just one sugar in which all the glycosidic linkages are the same (Drake et al., 1990). In the case of K antigens of similar structure, such as K1 and K92, the region 2 determinants are very similar (Roberts et al., 1986), whereas for structurally diverse polymers, such as K1 and K5, there is little (if any) homology in this region (Roberts et al., 1988).

Region 2 is flanked by regions 1 and 3 which are common to all group II capsule gene clusters so far studied. Mutations in regions 1 and 3 cause the accumulation of polysaccharide inside the cell, suggesting that they are needed for translocation of the K antigen to the cell surface (Boulnois et al., 1987; Kröncke et al., 1990a, 1990b; Bronner et al., 1993a, 1993b). Regions 1 and 3 are highly conserved among group II capsule gene clusters and can be aligned on the basis of shared restriction endonuclease cleavage sites (Roberts et al., 1986). By minicell analysis, region 1 from E. coli K1, K5 and K7 was found to encode several proteins of similar molecular weight (Roberts et al., 1986). Indeed, mutations in regions 1 or 3 can even be complemented in trans by the corresponding region from other group II capsule gene clusters (Roberts et al., 1988). These observations indicate that regions 1 and 3 encode a conserved set of proteins that are capable of exporting chemically different polysaccharides to the cell surface, independent of the repeat structure of the polysaccharide.

In keeping with their conservation among group II capsule gene clusters, genes in regions 1 and 3 are designated kps. Region 2 genes in E. coli K5 are designated kfi.
Figure 1.1. Cassette-like organisation of group II capsule genes. Regions 1 and 3 are conserved in all group II capsule gene clusters. The figure shows the region 2 for production of the K1 antigen inserted between regions 1 and 3. In *E. coli* expressing the K4 through K92 antigens, the region 2 for *E. coli* K1 would be replaced by the region 2 determinant shown. The hatched regions indicate the extent of homology with the sequences flanking group III capsule gene clusters (section 1.5.6). The shaded region adjacent to region 1 is found only in group II capsule gene clusters.

(K five antigen, Petit *et al.*, 1995) and those in *E. coli* K1 are designated *neu* because the capsule comprises only N-acetylmuraminic acid (Silver *et al.*, 1993).

Studies in this laboratory have focused on expression of the *E. coli* K5 antigen as a paradigm for other group II capsules. The genetic organisation of the K5 *kps* locus is shown in Figure 1.2. Homologues of the proteins encoded in regions 1 and 3 are made by *E. coli* expressing other group II capsule gene clusters and may have conserved functions in polysaccharide processing and export.

1.5.2 Region 1 of the *E. coli* K5 capsule gene cluster

1.5.2.1 Transcriptional organisation of region 1

Region 1 of the K5 *kps* locus was originally thought to comprise five genes, *kpsE, D, U, C, S* (Pazzani *et al.*, 1993a, 1993b). The polarity of insertions in region 1 indicates that these genes may be organised in a single transcriptional unit (Pazzani, 1992; Cieslewicz *et al.*, 1993). An additional open reading frame upstream of *kpsE* has been identified in *E. coli* K1 (Cieslewicz *et al.*, 1993), and K5 (Pazzani, 1992), and is
Figure 1.2. Genetic organisation of the *E. coli* K5 kps locus. Boxes labelled 1-3 represent the three functional regions described in the text. The genes of the conserved regions 1 and 3 are designated kps whilst K5-specific genes in region 2 are designated kfi. The approximate size of the encoded proteins is shown in kDa. Arrows indicate the direction of transcription. A partial map of restriction endonuclease cleavage sites pertinent to this study is given (B, BstI107I; Bg, BglIII; Bs, BsaBI; Bx, BstXI; C, Clal; E, EcoRI; Ev, EcoRV; F, FspI; H, HindIII; Hc, HincII; S, Styl; Sm, SmaI; St, StuI; V, VspI; X, XmnI).
designated \textit{kpsF}. In \textit{E. coli} K5, \textit{kpsF} was not thought to be part of region 1 as subclones lacking this open reading frame still direct capsule expression (Pazzani, 1992). However, by RNase protection analysis, region 1 transcripts in \textit{E. coli} K5 were found to originate from upstream of \textit{kpsF} (D. Simpson, unpublished results), indicating that it is the first gene of region 1. The promoter for region 1 has been mapped 5' of \textit{kpsF} and a single transcript of c. 8 kb spanning region 1 detected by Northern blotting (D. Simpson, unpublished results). Promoter activity has not been detected in the region between \textit{kpsF} and \textit{kpsE} (D. Simpson, unpublished results), suggesting that in subclones lacking \textit{kpsF}, transcription of region 1 is directed by a vector promoter. By direct detection of RNA in slot blot hybridisation experiments, transcription of region 1 was shown to be reduced at 18°C (D. Simpson, unpublished results).

1.5.2.2 \textbf{KpsF}

By minicell analysis, \textit{kpsF} has been shown to encode a protein of c. 35 kDa (Cieslewicz et al., 1993; D. Simpson, unpublished results). A deletion of \textit{kpsF} from subclones directing synthesis of the K5 antigen does not affect capsule expression (Pazzani, 1992). In contrast, an insertion in the \textit{E. coli} K1 \textit{kpsF} gene altered the morphology of plaques obtained with a K1-specific bacteriophage (Cieslewicz et al., 1993). This was claimed by the authors to reflect a reduction in cell surface expression of the polysaccharide, however it is not clear if the insertion had a polar effect on expression of the downstream genes. The nucleotide sequence of \textit{kpsF} from \textit{E. coli} K1 and K5 is 98% identical (D. Simpson, unpublished results), implying that it has a conserved function. \textit{KpsF} is 72% similar to \textit{GutQ}, a 23.6 kDa putative ATP-binding protein of unknown function encoded by the glucitol utilisation operon of \textit{E. coli} (Yamada et al., 1990). \textit{E. coli gutQ} mutants transformed with subclones of the K5 capsule genes lacking \textit{kpsF} are sensitive to K5-specific bacteriophage (D. Simpson, unpublished results). This indicates that for subclones lacking \textit{kpsF}, \textit{gutQ} does not supply \textit{in trans} the functions needed for capsule expression.

1.5.2.3 \textbf{KpsE}

Mutations in \textit{kpsE} cause accumulation of polysaccharide in the periplasm (Bronner et al., 1993a, 1993b), suggesting that it may be involved in export of the polysaccharide across the outer membrane. The \textit{kpsE} genes from \textit{E. coli} K1 and K5 have been cloned and sequenced (Cieslewicz et al., 1993; Pazzani et al., 1993a). By minicell analysis, \textit{kpsE} from \textit{E. coli} K1 was shown to encode a protein of 39 kDa (Cieslewicz et al., 1993), whereas for \textit{E. coli} K5 a 43 kDa protein was detected (Pazzani et al., 1993a). The proposed translational start site for \textit{kpsE} in \textit{E. coli} K5 has been disputed
Cieslewicz et al., 1993), but was confirmed by purification and N-terminal sequencing of the protein (Rosenow et al., 1995a). A prediction of the hydropathy profile of KpsE according to the method of Kyte and Doolittle (1982), indicated that KpsE has hydrophobic regions of at least 20 residues located near the N- and C-termini (Pazzani et al., 1993a). It is believed that these regions form α-helices that span the cytoplasmic membrane and that KpsE has a periplasmic domain of about 300 amino acids (Pazzani et al., 1993a; Rosenow et al., 1995a). Western blotting of KpsE from membranes and spheroplasts, before and after treatment with proteinase K, has verified the association of KpsE with the cytoplasmic membrane and shown that a large periplasmic domain exists (Rosenow et al., 1995a; F. Esuneh, unpublished results). This topology has been confirmed by TnphoA mutagenesis and the construction of β-lactamase fusions to KpsE (Rosenow et al., 1995a).

The predicted amino acid sequence of KpsE is 73.2% similar over 359 amino acids to the BexC protein of H. influenzae and 73.2% similar over 355 amino acids to CtrB of N. meningitidis (Pazzani et al., 1993a). Both of these proteins are believed to participate in the translocation of capsular polysaccharide across the inner membrane (Kroll et al., 1990; Frosch et al., 1991). The hydropathy profiles of BexC and CtrB are similar to that of KpsE and TnphoA mutagenesis has revealed that both of these proteins have a periplasmic domain (Kroll et al., 1990; Frosch et al., 1991). It is postulated that these proteins may act in concert with other transport proteins in the cytoplasmic membrane (section 1.5.3) (Reizer et al., 1992), perhaps by accepting polysaccharide on the outer face of the cytoplasmic membrane.

1.5.2.4 KpsD

KpsD is a 60 kDa periplasmic protein postulated to be involved in the translocation of polysaccharide across the outer membrane (Silver et al., 1987). Analysis of the predicted amino acid sequence of KpsD from E. coli K1 and K5 revealed a typical N-terminal signal sequence (Pazzani et al., 1993a; Wunder et al., 1994). A kpsD insertion mutation in the K1 capsule gene cluster resulted in accumulation of polysaccharide in the periplasm (Wunder et al., 1994). KpsD is 67% similar over 68 amino acids to the E. coli FgpB protein (phosphatidyl glycerophosphate (PGP) B phosphatase), which is capable of hydrolysing PGP, lysophosphatidic acid and phosphatidic acid (Icho, 1988). This homology may be significant because it is localised to the hydrophilic domain of PgpB which is thought to interact with phosphatidic acid (Icho, 1988). As K5 polysaccharide in the periplasm is linked to phosphatidyl-KDO (Bronner et al., 1993a), it is possible that KpsD interacts with the polysaccharide via the lipid moiety. KpsD is also 67% homologous over 100 amino acids to the ExoF protein of Rhizobium meliloti, a periplasmic protein
involved in the export of succinoglycan (Muller et al., 1993). A homologue of the E. coli kpsD gene has not been identified in H. influenzae or N. meningitidis (Pazzani et al., 1993b; Wunder et al., 1994). However, the translocation of capsular polysaccharides across the outer membrane is poorly understood and may occur by a different mechanism in these bacteria (section 1.6.1).

1.5.2.5 KpsU

The kpsU gene encodes a protein of c. 27 kDa that is 44.3% identical and 70% similar to the product of the E. coli kdsB gene (Pazzani et al., 1993a). KdsB is a CMP-KDO synthetase, which supplies activated KDO for LPS biosynthesis (Goldman and Kohlbrenner, 1985), and is present at basal levels in all E. coli, independent of encapsulation (Finke et al., 1990). Comparison of the nucleotide sequences of the kpsU and kdsB genes revealed 63.5% identity over 540 bp (Pazzani et al., 1993a), suggesting that they may be the products of a gene duplication. The identification of a second CMP-KDO synthetase in E. coli expressing group II K antigens explains why these strains have elevated levels of this enzyme (Finke et al., 1989, 1990). The notion that kpsU encodes a CMP-KDO synthetase is supported by the finding that kpsU mutants exhibit only the basal level of CMP-KDO synthetase activity that is attributed to KdsB (Pazzani, 1992; Bronner et al., 1993a), and that the purified KpsU protein has CMP-KDO synthetase activity in vitro (Rosenow et al., 1995b). In addition, the predicted N-terminal amino acid sequence of KpsU is in agreement with that obtained using the purified capsule-specific CMP-KDO synthetase (Rosenow et al., 1995b). The predicted amino acid sequence for KpsU is also 66% similar to NeuA, a CMP-N-acetylneuraminic acid synthetase encoded in region 2 of the E. coli K1 capsule gene cluster (Zapata et al., 1989).

A kpsU mutant of E. coli K5 was found to be sensitive to K5-specific phage (Pazzani, 1992). By immunoelectron microscopy this mutant was shown to produce patches of polysaccharide at the cell surface and to accumulate polysaccharide in electron dense aggregates in the cytoplasm (Bronner et al., 1993a). Cytoplasmic inclusions of polysaccharide are also made by kpsC and kpsS mutants (section 1.5.2.7) (Bronner et al., 1993a), and it is possible that the mutation has a polar effect on the expression of these genes. The mutation was caused by the insertion of an oligonucleotide linker containing a stop codon and results in the production of a truncated form of KpsU (Pazzani, 1992). The translation initiation codon for kpsC overlaps the end of kpsU, indicating that these genes may be translationally coupled (Pazzani, 1992). Thus, if translation of kpsU is terminated prematurely, KpsC (and perhaps KpsS) may not be made.
The observation that KDO is present at the reducing terminus of the K5 polysaccharide and that polymerisation occurs from the non-reducing end (Finke et al., 1989, 1991), prompted speculation that KpsU may provide KDO for the synthesis of an acceptor for the sugar components of the polysaccharide. The finding that the CMP-KDO synthetase activity is elevated at capsule-permissive temperatures is consistent with this notion (Finke et al., 1989, 1990). However, a role for KpsU in the initiation of polysaccharide biosynthesis is unlikely as kpsU mutants can synthesise the K5 polysaccharide, albeit in reduced amounts (Bronner et al., 1993a). Further, it has been reported that the polysaccharide made by kpsU mutants is not substituted with KDO (Bronner et al., 1993a), suggesting that KdsB does not complement the kpsU defect. It is possible however, that the linkage between KDO and the polysaccharide is labile and was destroyed during extraction of the polysaccharide.

It has also been suggested that KpsU may supply KDO for the synthesis of 2-phosphatidyl-KDO, for ligation to polysaccharide chains (Bronner et al., 1993a). In E. coli K5, polysaccharide at the cell surface and in the periplasm (from kpsE and kpsD mutants), is linked to phosphatidyl-KDO, whereas polysaccharide in the cytoplasm is not (Bronner et al., 1993a). This has led to the suggestion that substitution of the K5 polysaccharide with phosphatidyl-KDO is required for its translocation to the cell surface, as is the case with N. meningitidis capsular polysaccharides (section 1.4.3) (Frosch and Müller, 1993). This would imply that KpsU is required for translocation of the polysaccharide. However, kpsU mutants express some polysaccharide at the cell surface, suggesting that this may not be the case. Thus, the role of KpsU in expression of the K5 antigen remains poorly defined.

The presence of two CMP-KDO synthetase enzymes in E. coli expressing group II capsules may allow the expression of LPS and the K antigen to be independently regulated. LPS is an essential component of the outer membrane of Gram negative bacteria, therefore KdsB must be expressed at all growth temperatures. In contrast, group II capsules are not expressed below 20°C, so KpsU is not always needed. The CMP-KDO synthetase activity in E. coli expressing group II capsules is much lower at 18°C than at 37°C (Finke et al., 1989, 1990). This is because region 1 transcription is reduced at 18°C and does not reflect cold sensitivity of the enzyme, as purified KpsU is active at this temperature in vitro (Rosenow et al., 1995b). It may also be that two CMP-KDO synthetase enzymes are needed in group II strains because of biochemical differences between the enzymes that are suited to their respective roles in LPS and capsule production. This assumes however, that CMP-KDO made by KdsB and KpsU is used exclusively for LPS or capsule biosynthesis, respectively. The Michaelis constant ($K_m$) for KpsU is ten fold higher than for KdsB, indicating that KpsU has a weaker affinity for its substrate (Rosenow et al., 1995b). In addition, the maximum rate of catalysis ($V_{max}$) is much lower for KpsU than for KdsB.
Thus, KdsB is a more efficient CMP-KDO synthetase than KpsU and this may ensure that LPS is expressed preferentially when CTP or KDO are limited. This is consistent with the essential role of LPS in outer membrane integrity.

1.5.2.6 KpsC

Analysis of the nucleotide sequence of kpsC revealed several possible open reading frames, with predicted proteins ranging from c. 55 to 76 kDa (Pazzani, 1992). By minicell analysis kpsC was shown to encode a protein of c. 76 kDa in E. coli K5 (Pazzani et al., 1993a), and proteins of a similar size are also encoded in region 1 of E. coli K1 and K7 (Roberts et al., 1986). The hydrophathy profile of KpsC and lack of an obvious signal sequence suggest that it is located in the cytoplasm. KpsC is 76% similar to LpsZ, a cytoplasmic protein of Rhizobium meliloti which has a poorly defined role in the modification of rhizobial LPS and K polysaccharides (Brzoska and Signer, 1991; Reuhs et al., 1995). The K polysaccharides of R. meliloti contain KDO and are structurally similar to the group II K antigens of E. coli (Reuhs et al., 1993). Mutations in lpsZ alter the size range of polysaccharide chains expressed at the cell surface and confer resistance to capsule-specific bacteriophages (Reuhs et al., 1995). Unlike kpsC mutations, lpsZ defects do not prevent surface expression of the polysaccharide. A kpsC mutation in the E. coli K5 genes could not be complemented by the cloned lpsZ gene from R. meliloti (M. Collins, unpublished results). The phenotype of kpsC and kpsS mutations is identical and is discussed below.

1.5.2.7 KpsS

In E. coli K1 and K5, kpsS encodes a protein of c. 44 kDa that is predicted to be located in the cytoplasm (Silver et al., 1984; Pazzani et al., 1993a). Mutations in both kpsC and kpsS cause the accumulation of polysaccharide in electron dense aggregates in the cytoplasm (Bronner et al., 1993a, 1993b). This polysaccharide has a high molecular weight, comparable to that of surface polymer, but is not linked to KDO or phospholipid (Bronner et al., 1993a).

It has also been reported that membranes from kpsC and kpsS mutants have reduced endogenous transferase activities in vitro (Vimr et al., 1989; Bronner et al., 1993a, 1993b). In E. coli K1 for example, the endogenous sialyltransferase activity of a kpsS mutant is reduced by at least 10 fold (Vimr et al., 1989), whereas membranes from E. coli K5 kpsC and kpsS mutants have endogenous transferase activities of about 2% that of wild-type membranes (Bronner et al., 1993a, 1993b). It was shown by Bronner et al. (1993b), that addition of exogenous K5 polysaccharide to the membranes of kpsC and kpsS mutants, increased the transferase activity to about 25% of the wild-type. It is
possible therefore, that \textit{kpsC} and \textit{kpsS} mutations reduce either the initiation of polysaccharide synthesis, or the amount of existing endogenous acceptor on the membrane. The notion that KpsC and KpsS may be required to hold growing polysaccharide chains on the membrane is appealing, since it may explain why \textit{kpsC} and \textit{kpsS} mutants release polysaccharide into the cytoplasm \textit{in vivo}. It has been postulated that KpsC and KpsS may form a complex with the enzymes for polysaccharide biosynthesis at the cytoplasmic face of the inner membrane (Vimr \textit{et al.}, 1989; Bronner \textit{et al.}, 1993a, 1993b). It is unlikely that the loss of KpsC or KpsS affects the ability of this complex to extend polysaccharide chains because \textit{kpsC} and \textit{kpsS} mutants still make polysaccharide.

KpsC and KpsS are homologous to the LipA and LipB proteins of \textit{N. meningitidis} respectively (Frosch and Müller, 1993). These proteins are encoded in region B of the \textit{N. meningitidis} capsule gene cluster, which has been implicated in the translocation of capsular polysaccharide to the cell surface (Frosch \textit{et al.}, 1989). As with \textit{kpsC} and \textit{kpsS} mutations, deletion of \textit{lipA} or \textit{lipB} results in cytoplasmic accumulation of polysaccharide which is not linked to phospholipid (Frosch and Müller, 1993). It is believed that LipA and LipB may determine the phospholipid substitution of \textit{N. meningitidis} capsular polysaccharides and that this substitution is a prerequisite for export of the polysaccharide (Frosch and Müller, 1993). It is possible that the KpsC and KpsS proteins have an analogous function in the expression of group II K antigens.

\textbf{1.5.3 Region 3 of the \textit{E. coli} K5 capsule gene cluster}

Mutations in region 3 result in accumulation of cytoplasmic polysaccharide which is shorter than surface polymer and lacks phospholipid substitution (Kröncke \textit{et al.}, 1990a; Bronner \textit{et al.}, 1993b). This phenotype is also induced by the addition of the membrane-decoupling agent, carbonyl-cyanide-M-chlorophenyl hydrzone, to wild-type \textit{E. coli} K5 prior to shifting from a capsule-restrictive to a capsule-permissive temperature (Kröncke \textit{et al.}, 1990b). These results suggest that the region 3 products are involved in the energy-dependent translocation of polysaccharide across the inner membrane.

In \textit{E. coli} K1 and K5, region 3 comprises two genes, \textit{kpsM} and \textit{kpsT}, organised in a single transcriptional unit (Figure 1.2) (Smith \textit{et al.}, 1990; Pavelka \textit{et al.}, 1991). Analysis of the predicted amino acid sequences of KpsM and KpsT indicates that they belong to the \textit{Δ}TP-binding cassette (ABC) family of transport proteins, also referred to as traffic ATPases (Higgins, 1992; Fath and Kolter, 1993). This family comprises both prokaryotic and eukaryotic proteins that transport specific substances across the cell membrane using the energy derived from ATP hydrolysis. Most ABC-transporters comprise a hydrophobic membrane component, which mediates the transport process,
coupled to a separate hydrophilic ATP-binding component (Higgins, 1992; Fath and Koiter, 1993). The kpsM and kpsT genes appear to be translationally coupled (Smith et al., 1990; Pavelka et al., 1991). This may ensure that the region 3 proteins are made in the correct stoichiometry.

Analysis of the hydropathy profile of KpsM from E. coli K1 indicates that it contains six hydrophobic domains of at least 20 residues separated by short stretches of hydrophilic amino acids (Pavelka et al., 1991). The hydrophobic domains are postulated to form α-helices which span the cytoplasmic membrane, with both the N- and C-termini in the cytoplasm (Pavelka et al., 1991). This topology has been confirmed by the generation of β-lactamase and alkaline phosphatase sandwich fusions in KpsM (Pigeon and Silver, 1994; Silver, 1994). Protease accessibility studies have also confirmed the parts of KpsM that are exposed to the periplasm. Site-directed mutagenesis and linker insertions have defined certain regions of the protein that are important in polysaccharide export (Pigeon and Silver, 1994).

KpsT is predicted to be a cytoplasmic protein and contains an adenine nucleotide binding fold (Smith et al., 1990; Pavelka et al., 1991). Support for the interaction of KpsT with ATP was derived by photochemical labelling of KpsT in vitro with analogues of ATP (Pavelka et al., 1994). Further, mutations affecting conserved amino acids within the ATP-binding sequence of KpsT prevent this labelling and prevent polysaccharide export (Pavelka et al., 1991; Pavelka et al., 1994; Silver, 1994).

The predicted amino acid sequences and hydropathy profiles of KpsM and KpsT are homologous to the H. influenzae BexB and BexA proteins (Kroll et al., 1990), the N. meningitidis CtrC and CtrD proteins (Frosch et al., 1991), and the NodJ and NodI proteins of Rhizobium species (Vázquez et al., 1993), respectively. These proteins are all involved in the translocation of polysaccharide across the cytoplasmic membrane and appear to be ABC-transporters (Reizer et al., 1992). KpsM and KpsT are also homologous to two open reading frames, ORF1 and ORF2 respectively, involved in capsular polysaccharide transport in Actinobacillus pleuropneumoniae (Ward and Inzana, 1995). Mutations in region 3 of the E. coli K5 capsule gene cluster can be complemented by the homologous genes from A. pleuropneumoniae (Ward and Inzana, 1995), indicating that capsular polysaccharides may be exported by a common mechanism in Gram negative bacteria.

A distinct subfamily of ABC-transporters involved in polysaccharide export has been proposed (ABC-2 subfamily) (Reizer et al., 1992). Some members of the ABC-2 subfamily are postulated to have two integral membrane components involved in polysaccharide export. It is suggested that in H. influenzae the BexC protein interacts with the BexAB transporter and that in N. meningitidis CtrB interacts with CtrCD (Reizer et al., 1992). By analogy, the KpsE protein, which is homologous to BexC and CtrB, might interact with KpsMT. Whilst it is tempting to speculate about the function...
of these inner membrane proteins, there is no evidence that they interact with other components of the polysaccharide export apparatus. The isolation of extragenic suppressor mutations may provide evidence for such interactions.

1.5.4 Region 2 of the *E. coli* K5 capsule gene cluster

1.5.4.1 Transcriptional organisation of region 2

Biosynthesis of the K5 antigen, 4)-β-glucuronic acid-(1,4)-α-N-acetylglucosamine-(1-, is determined by genes in region 2 (Roberts *et al*., 1988). Analysis of the nucleotide sequence of region 2 revealed that there are four open reading frames, termed *kfiA* through *kfiD* (Figure 1.2) (Petit *et al*., 1995). An additional open reading frame is located between *kfiB* and *kfiC*, and is predicted to be transcribed counter to the other region 2 genes and to encode a protein of c. 16 kDa. However, the protein has not been visualised in minicells (B. Clarke, unpublished results), and an insertion in this open reading frame did not affect expression of K5 antigen (Petit *et al*., 1995).

By Northern blotting three overlapping region 2 transcripts of c. 8, 6.5 and 3 kb were detected and the promoters located 5' of *kfiA*, *kfiB* and *kfiC* by primer extension and promoter probe analysis (Petit *et al*., 1995). The capsule biosynthesis genes in region A of the *N. meningitidis* cps cluster are also organised in a single transcriptional unit (Edwards *et al*., 1994), and it appears that the region 2 genes in *E. coli* K1 are also transcribed in the same direction (Silver, 1994). It is not clear why three overlapping transcripts are necessary or why large intergenic regions are transcribed. The gap between the *kfiB* and *kfiC* genes for instance is 1293 bp and contains no open reading frames on the positive sense strand. In support of the Northern blot data, transcripts spanning the gaps between the region 2 genes were detected by reverse-transcriptase PCR (C. Petit, unpublished results). Promoter probe analysis indicates that the region 2 genes are expressed at a very low level (Petit *et al*., 1995).

1.5.4.2 KfiA and KfiB

The *kfiA* and *kfiB* genes encode proteins of c. 27 kDa and 66 kDa respectively, which are believed to be located in the cytoplasm (Petit *et al*., 1995). An additional 35 kDa protein is associated with the expression of *kfiB* in minicells (Petit *et al*., 1995). This could be a breakdown product of KfiB or may arise by initiation of translation at a site within the gene. The predicted amino acid sequences of KfiA and KfiB are not strongly homologous to any proteins in the database. Secondary structure analysis of
KfiB indicates that it contains an α-helical coiled-coil motif (Cohen and Parry, 1994), which is shared by eukaryotic cytoskeletal proteins such as α-tropomyosin.

A kfiA mutant has been found to exhibit a low endogenous transferase activity in vitro, but can extend exogenous polysaccharide (Sieberth, 1994). This suggests that KfiA may participate in the initiation of polysaccharide biosynthesis, or may be required to hold the endogenous acceptor on the membrane. The possible function of KfiB is also poorly understood. A TnphoA insertion c. 300 bp from the 3' end of kfiB abolished expression of the K5 antigen at 37°C, as determined by sensitivity to a capsule-specific bacteriophage (M. Stevens, unpublished results). Polar effects of this insertion are unlikely as a separate transcript for the kfiC and kfiD genes has been identified (Petit et al., 1995).

1.5.4.3 KfiC

KfiC is a protein of c. 60 kDa which exhibits homology to a number of glycosyltransferase enzymes (Petit et al., 1995). These include the ExoO, ExoU, and ExoW proteins of R. meliloti, involved in the biosynthesis of succinoglycan (Glucksman et al., 1993), and the HasA protein of group A Streptococci, which is involved in the biosynthesis of a hyaluronic acid capsule (DeAngelis et al., 1993). This is significant because hyaluronic acid, like the K5 antigen, is composed of glucuronic acid and N-acetylglucosamine.

In vitro transferase assays using membranes from a strain which overexpresses kfiC have indicated that KfiC is a bifunctional glycosyltransferase, capable of transferring both glucuronic acid and N-acetylglucosamine to exogenous K5 polysaccharide (Sieberth, 1994). The S. pyogenes HasA protein is also a bifunctional transferase, but the linkages it forms are different from those of KfiC (DeAngelis et al., 1993). Regions of strong conservation between the predicted amino acid sequence of KfiC and other glycosyltransferases have been identified (Petit et al., 1995; Saxena et al., 1995). Site-directed mutagenesis of these regions may enable their role in the activity of KfiC to be elucidated.

1.5.4.4 KfiD

KfiD is believed to be an NAD-dependent UDP-glucose dehydrogenase, catalysing the formation of UDP-glucuronic acid for incorporation into the K5 polysaccharide (Petit et al., 1995; Sieberth et al., 1995). Evidence that KfiD is a UDP-glucose dehydrogenase comes from the observation that overexpression of kfiD results in elevated UDP-glucose dehydrogenase activity (Petit et al., 1995). Further, KfiD has been purified and shown to convert UDP-glucose to UDP-glucuronic acid in
vitro (Sieberth et al., 1995). The nucleotide sequence of the kflD gene shares significant homology to genes encoding NAD-dependent sugar dehydrogenases in other bacteria (Petit et al., 1995). It is 59% identical over 478 bp to the hasB gene, involved in synthesis of the hyaluronic acid capsule of group A Streptococci (Dougherty and van der Rijn, 1993), and 60% identical over 574 bp to the cap3A gene which encodes a UDP-glucose dehydrogenase required for expression of the type 3 capsule of S. pneumoniae (Arrecubieta et al., 1994). This conservation implies that the genes for UDP-glucose dehydrogenases have been transmitted between diverse bacteria relatively recently in evolution.

1.5.5 Genetic organisation of H. influenzae and N. meningitidis capsule gene clusters

*Haemophilus influenzae* elaborates six serologically distinct capsular polysaccharides designated a through f. Of the six serotypes, mostly type b strains are associated with invasive disease in humans (Moxon and Kroll, 1990). The capsule genes (cap) of *H. influenzae* have a conserved organisation that is similar to that of *E. coli* group II capsule gene clusters (Figure 1.3) (Kroll et al., 1989). There is a central serotype-specific region which encodes the biosynthetic functions (Van Eldere et al., 1995), which is flanked by regions that are common to all serotypes that determine polysaccharide export (Kroll et al., 1989; Kroll et al., 1990). Region 1 of the *H. influenzae* type b cap locus has been sequenced and contains four genes, bexA through bexD, probably organised in a single transcriptional unit (Kroll et al., 1990). Region 2 also comprises four genes, one of which appears to encode CDP-ribitolpyrophosphorylase, an enzyme needed for the synthesis of a precursor of the type b polysaccharide (Van Eldere et al., 1995).

In type b strains of phylogenetic division I there is an extensive duplication of the cap genes, with two directly repeated copies of c. 18 kb flanking a c. 1.2 kb bridge region (Hoiseth et al., 1986). The duplication is not perfect, since cap contains only one functional copy of bexA (Kroll et al., 1988). The bexA gene is needed for polysaccharide export and is located in the bridge region (Kroll et al., 1990). *H. influenzae* type b strains revert to an acapsular phenotype at high frequency. This instability is due to recombination events between homologous sites in each repeat which reduce cap to a single copy and result in the loss of bexA (Kroll et al., 1988).

*H. influenzae* type b strains may contain multiple copies of the cap genes (Kroll and Moxon, 1988). The cap locus is flanked by direct repeats of an insertion sequence-like element, IS1016 (Kroll et al., 1991). Recombination events mediated by IS1016 are responsible for amplification of the cap locus (Kroll et al., 1991). Amplification of cap is accompanied by an increase in capsule synthesis (Kroll and Moxon, 1988). The advantage of the cap instability and gene-dosage phenomena in vivo is not clear.
Figure 1.3. Genetic organisation of the capsulation loci of H. influenzae and N. meningitidis. The organisation of the E. coli K1 capsule gene cluster is shown for comparison. Large boxes represent the conserved functional regions described in the text. Open reading frames are represented by small boxes.
It has been suggested that capsule loss may promote adhesion to epithelia and invasion (St Geme III and Falkow, 1991), and may therefore be required at certain stages of infection.

The genes for capsule production in *N. meningitidis* group B have been cloned and are designated *cps* (Frosch et al., 1989). The *cps* locus is composed of five regions, designated A through E (Figure 1.3). The central region A encodes the enzymes necessary for polysaccharide biosynthesis (Edwards et al., 1994; Ganguli et al., 1994). As with group II capsule gene clusters, transport functions are assigned to the flanking regions. Region B appears to direct the translocation of polysaccharide across the cytoplasmic membrane and region C is required for transport of the polysaccharide from the periplasm to the cell surface (Frosch et al., 1989). Region B encodes two proteins, LipA and LipB, that are thought to determine the phospholipid substitution of the polysaccharide (Frosch and Müller, 1993). Region C of the *cps* locus comprises four genes, *ctrA* through *ctrD*, in a single operon (Frosch et al., 1991). Region D encodes a UDP-glucose-4-epimerase (GalE) that is required for the biosynthesis of lipooligosaccharide (Hammerschmidt et al., 1994). In addition, region D encodes proteins that are homologous to the RfbB, RfbC and RfbD enzymes of *E. coli* and *S. typhimurium*, which are involved in the biosynthesis of rhamnose for incorporation into LPS (Hammerschmidt et al., 1994). Region D is not thought to be involved in biosynthesis of the group B capsular polysaccharide, as mutations in this region do not affect capsule production (Hammerschmidt et al., 1994). Region E is believed to be involved in the regulation of polysaccharide biosynthesis or export, since mutations in this region increase capsule expression (Frosch et al., 1989). Regions analogous to regions D and E have not been identified in *E. coli* group II capsule gene clusters.

The capsular polysaccharide of *N. meningitidis* group B is identical to the *E. coli* K1 antigen (Bhattacharjee et al., 1975). Biosynthesis of these polymers involves several identical steps in both bacteria (section 1.6) (Vann et al., 1993). Although initial studies indicated that little homology exists between the *E. coli* K1 capsule genes and *N. meningitidis* group B DNA (Echarti et al., 1983), subsequent analysis has revealed that the proteins involved in synthesis of the polysaccharide are similar. For example, the NeuB protein of *E. coli* K1, which is postulated to encode an N-acetylneuraminic acid synthetase, is 57% similar to the CpsB protein of *N. meningitidis* (Annunziato et al., 1995). Significant homology is also shared by the CMP-N-acetylneuraminic acid synthetase and sialyltransferase enzymes in these bacteria (Frosch et al., 1991; Ganguli et al., 1994).

The homologous organisation of the capsule genes in *H. influenzae*, *N. meningitidis* and *E. coli* expressing group II K antigens suggests that they have a common evolutionary origin. Evidence for this comes from demonstrations of DNA and protein homology between the functional regions of the capsule gene clusters of
these organisms. The \textit{kpsM} and \textit{kpsT} genes of region 3 in group II capsule gene clusters for example, share extensive sequence homology with the \textit{H. influenzae} \textit{bexB} and \textit{bexA} genes (Smith \textit{et al}., 1990; Pavelka \textit{et al}., 1991; Kroll \textit{et al}., 1990), and with the \textit{N. meningitidis} \textit{ctrC} and \textit{ctrD} genes (Frosch \textit{et al}., 1991), respectively. In addition, \textit{kpsE} is similar to \textit{bexC} and \textit{ctrB} (Cieslewicz \textit{et al}., 1993; Pazzani \textit{et al}., 1993a). It is believed that these genes determine the translocation of polysaccharide across the inner membrane (section 1.5.3) (Reizer \textit{et al}., 1992). Not only does this homology imply that the genes have a common origin, but it indicates that there may be a common mechanism for polysaccharide export in these bacteria.

1.5.6 Genetic organisation of group III capsule gene clusters

Hybridisation studies using DNA probes from group II capsule gene clusters have been performed to investigate the organisation of group III capsule genes (Boulnois \textit{et al}., 1992; Drake \textit{et al}., 1993; Pearce and Roberts, 1995). Probes from regions 1 and 3 of group II capsule gene clusters do not hybridise to DNA from group III strains (Drake \textit{et al}., 1993). DNA immediately adjacent to region 1 is also specific to group II capsule gene clusters. However, DNA further upstream of region 1 and adjacent to region 3 hybridises to DNA from all group III strains (Figure 1.1). This implies that group III capsule production involves genes at or near \textit{serA} which are different from group II capsule genes.

The genes for production of the group III K10 and K54 antigens have been cloned (Pearce and Roberts, 1995). DNA probes taken from within the K10 capsule gene cluster hybridise to DNA from most group III capsule-producing \textit{E. coli}, whilst others hybridise only to \textit{E. coli} K10 DNA (Pearce and Roberts, 1995). This indicates that group III capsule gene clusters, like those of group II, share conserved regions but are internally variable. Certain subclones of the K10 and K54 capsule genes are able to complement mutations in region 1 of the K5 (group II) capsule gene cluster which affect polysaccharide export. This indicates that despite the lack of detectable DNA homology between region 1 and the group III capsule genes (Drake \textit{et al}., 1993), certain stages in polysaccharide export may occur by a conserved mechanism. However, the genetic organisation of the group II and group III capsule genes is different. For instance, the \textit{kpsE}, \textit{kpsD} and \textit{kpsC} genes in \textit{E. coli} K5 are part of the same operon (Pazzani \textit{et al}., 1993a), whereas in the K10 capsule gene cluster the homologue of \textit{kpsC} is separated from the \textit{kpsE} and \textit{kpsD} homologues by central K10-specific sequences (Pearce and Roberts, 1995). Some DNA probes taken from within the K10 capsule gene cluster failed to hybridise to DNA from \textit{E. coli} expressing certain other group III capsules (Pearce and Roberts, 1995), indicating that further diversity exists among strains expressing group III K antigens.
1.5.7 Genetic basis of diversity at the serA-linked capsule locus

It has been suggested that group II capsule diversity may have been achieved through the acquisition of new region 2 determinants, *en bloc*, from other bacteria (Boulnois and Jann, 1989). Some support for this comes from the observation that the G+C content of the DNA in region 2 is lower than that of the flanking regions. In *E. coli* K5 the G+C content of region 2 is 33.4% (Petit *et al.*, 1995), whereas for region 1 it is 50.6% (Pazzani *et al.*, 1993a) and for region 3 it is 42.3% (Smith *et al.*, 1990). It is also lower than the 50% G+C ratio typical of *E. coli* (Ørskov, 1984). Low G+C ratios have been reported for the capsule genes of other Gram negative bacteria (Frosch *et al.*, 1991; Arakawa *et al.*, 1995; Van Eldere *et al.*, 1995), and it possible that this reflects a common ancestry.

It is postulated that a resident region 2 may be displaced by a newly acquired region 2 in a site-specific recombination event mediated by specific sequences at the ends of region 2 (Boulnois and Jama, 1989). However the lack of conserved sequences at the boundaries between region 2 and regions 1 and 3 suggests this may not be the case (Drake, 1991). Alternatively, new region 2 determinants may arise by homologous recombination between the flanking regions 1 and 3 of a resident and incoming capsule gene cluster. Analysis of the predicted amino acid sequences of KpsS and KpsT from the K1, K4 and K5 capsule gene clusters shows that there is a marked divergence in the amino acid sequence at the C-terminus of the proteins (Drake, 1991). Since the 3' ends of the *kpsS* and *kpsT* genes border region 2, these differences may have arisen through recombinational exchanges with the capsule gene clusters of other bacteria. Further diversity may be achieved by recombination between region 2 genes of the incoming and resident clusters and between other polysaccharide biosynthesis genes. Region 2 cassettes may be also modified by point mutations, insertions, deletions or duplications.

Hybridisation studies have indicated that group III capsule gene clusters are also internally variable (Pearce and Roberts, 1995). The finding that group II-specific sequences are not present in group III capsule gene clusters and that sequences flanking the group II capsule genes are also present in group III strains (section 1.5.6) (Drake *et al.*, 1993), implies that group III capsule gene clusters may have evolved by replacement of group II capsule genes with an entirely different set of genes. The *serA*-linked capsule locus is therefore highly polymorphic.
1.6 Biosynthesis of E. coli group II K antigens

Polymerisation of the E. coli K1 and K5 antigens involves the sequential addition of sugar residues from their nucleotide-activated forms to the non-reducing terminus of the polysaccharide (Steenbergen and Vimr, 1990; Finke et al., 1991). By assaying the biosynthesis of the K5 polysaccharide in membrane vesicles with defined orientations, polymerisation was found to occur on the cytoplasmic face of the inner membrane (Finke et al., 1991). The biosynthesis of polysialic acid in E. coli K1 is best understood and begins with the synthesis of N-acetylneuraminic acid in a condensation reaction between phosphoenolpyruvate and N-acetylmannosamine (Vann et al., 1993). N-acetylneuraminic acid (NeuNAc) is then activated by conversion to CMP-NeuNAc (Vann et al., 1993). The genes for synthesis and activation of NeuNAc are located in region 2 (Figure 1.3) (Silver, 1994). The neuB gene encodes NeuNAc synthetase (Annunziato et al., 1995), and neuA encodes CMP-NeuNAc synthetase (Zapata et al., 1989). The neuS gene encodes the sialyltransferase enzyme which polymerises NeuNAc from CMP-NeuNAc (Weisgerber et al., 1991).

The polysialyltransferase enzyme (polyST) elongates polysialyl polymers within the membrane, as well as exogenous acceptors consisting of polysialic acid (Kundig et al., 1971; Troy and McCloskey, 1979). A single enzyme catalyses both of these reactions (Steenbergen et al., 1992). The polyST enzymes from E. coli K1 and K92 are very similar (Vimr et al., 1992), however the K92 polyST is a bifunctional enzyme, capable of introducing both α2,8 and α2,9 linkages between NeuNAc residues (Steenbergen et al., 1992). The polyST enzyme is unable to initiate K1 polysaccharide synthesis by itself (Steenbergen et al., 1992), and the reaction which initiates polysialic acid synthesis is unknown. Sialic acid residues linked to undecaprenol phosphate have been detected (Troy et al., 1975), but it is not known if these molecules act as intermediate donors of sialic acid, or as acceptors for the growing polysaccharide. It has been postulated that linkage of polysialic acid chains to a lipid carrier such as undecaprenol phosphate may be important for their translocation (Troy et al., 1993). A 20 kDa membrane protein has also been reported at the reducing terminus of polysialic acid chains in E. coli K1 (Weisgerber and Troy, 1990). One model suggests that this protein may accept single sialic acid residues or short oligomers from an undecaprenol phosphate carrier (Whittfield and Valvano, 1993).

The predicted amino acid sequence of NeuE, which is also encoded by region 2 of the K1 capsule gene cluster, revealed the presence of a potential polyisoprenol-binding site in a membrane-spanning domain of the protein (Steenbergen et al., 1992; Troy et al., 1993). This motif has also been identified in yeast glycosyltransferases implicated in the transfer of sugars, from their nucleotide-activated form, to derivatives of dolichol phosphate (Albright et al., 1989). This raised the hypothesis
that NeuE is involved in the initiation of polysialic acid synthesis by transferring NeuNAc residues to undecaprenol (Steenbergen et al., 1992). However, polysialic acid can be made in strains with a defined neuE defect, suggesting that it is not involved in initiation (Vimr and Steenbergen, 1993). Interestingly, KpsM also contains a putative polyisoprenol-binding site (Troy et al., 1993). This has led to speculation that lipid may provide the scaffolding to maintain the multi-enzyme complex needed for polysaccharide biosynthesis and export on the cytoplasmic membrane (Troy et al., 1993). Mutations affecting proteins from region 1 (KpsC and KpsS) and region 3 (KpsT) impair biosynthesis of both the K1 and K5 antigens in vitro (Vimr et al., 1989; Bronner et al., 1993a, 1993b), indicating that region 2 proteins alone do not suffice for the biosynthesis of group II K antigens.

Biosynthesis of the K5 antigen is poorly understood. No lipid-linked intermediates have been identified and the initiation reaction and endogenous acceptor have not been defined (Finke et al., 1991). Because KDO is present at the reducing terminus of the K5 antigen, it has been postulated that initiation may involve the transfer of KDO to an acceptor molecule (Finke et al., 1991). However, E. coli K5 kpsU, kpsC and kpsS mutants produce polysaccharide that lacks KDO at the reducing terminus (Bronner et al., 1993a), suggesting that KDO is added to the polysaccharide after polymerisation. However, it is possible that the linkage between a KDO-substituted carrier and the polysaccharide could be labile and this may prevent the isolation of intact intermediates. It follows that a role for KDO in the initiation of K5 polysaccharide biosynthesis cannot be precluded.

In E. coli K1, the temperature-dependent expression of the capsule has been proposed to reflect cold sensitivity of the enzyme sialic acid synthetase (Merker and Troy, 1990). The availability of sialic acid for biosynthesis of the K1 antigen may also be regulated by N-acetylneuraminic acid pyruvate-lyase (Vimr and Troy, 1985). However, all group II K antigens share a common mode of expression, even if they do not contain sialic acid, indicating that a conserved mechanism for regulating group II capsule expression may exist.

### 1.6.1 Model for expression of the E. coli K5 antigen

A model for expression of the K5 antigen has been proposed (Roberts, 1995). Polymerisation is believed to involve a multi-enzyme complex assembled at the cytoplasmic face of the inner membrane. KfIC catalyses the sequential transfer of alternate glucuronic acid and N-acetylglucosamine residues, from their UDP-activated forms, to the non-reducing terminus of the polysaccharide. A UDP-glucose dehydrogenase, KfID, supplies the UDP-glucuronic acid for K5 antigen biosynthesis. The KfIA and KfIB proteins may be involved in the synthesis of an acceptor on which
the polysaccharide is assembled. The translocation of completed polysaccharide chains may then require their substitution with phosphatidyl-KDO. It may be that phospholipid substitution is the motif recognised by the polysaccharide export apparatus that the polysaccharide is mature and ready for export. This is an appealing notion since it might explain how the conserved region 1 and 3 proteins can export chemically different polysaccharides independent of their repeat structure. Based on homology to proteins which determine phospholipid substitution in *N. meningitidis*, KpsC and KpsS are proposed to be involved in the attachment of KDO to phosphatidic acid and the subsequent ligation of phosphatidyl-KDO to the reducing terminus of the polysaccharide. KpsU may furnish the CMP-KDO for this process. Polysaccharide would then be translocated across the cytoplasmic membrane by KpsM and KpsT, which constitute an ABC-transporter. It is proposed that KpsT couples the energy derived from ATP hydrolysis to the transport process mediated by KpsM. The KpsE and KpsD proteins may accept polysaccharide as it enters the periplasm and deliver it to the cell surface.

The translocation of polysaccharide across the outer membrane and its assembly at the cell surface are poorly understood. There is no known outer membrane component involved in translocation of group II K antigens. In *N. meningitidis*, an outer membrane protein, CtrA, is believed to be part of the polysaccharide export apparatus (Frosch et al., 1992). Although this has yet to be demonstrated, CtrA is highly conserved among *N. meningitidis* serogroups. No homologues of CtrA have been identified in *E. coli* expressing group II K antigens (Frosch et al., 1991). The predicted secondary structure of CtrA suggests that it has eight amphipathic beta strands which span the outer membrane (Frosch et al., 1992). A similar secondary structure has been predicted for outer membrane porins in *E. coli* (Vogel and Jähnig, 1986). Porins form aqueous channels across the outer membrane that facilitate the transport of small hydrophilic solutes. In *E. coli*, a porin (protein K) which is rarely found in unencapsulated strains, has been correlated with capsule production (Paakanen et al., 1979; Whitfield et al., 1983). Indeed, the appearance of protein K in the outer membrane even coincides with the cell surface expression of capsular polysaccharide in temperature up-shift experiments (Whitfield et al., 1985). It is unlikely that protein K has a specific role in capsule expression, as *E. coli* K-12 strains, which lack protein K, are able to express group II K antigens when transformed with the capsule genes on a plasmid.

In *E. coli* expressing group II K antigens newly exported polysaccharide is associated with areas in which the cytoplasmic and outer membranes are closely apposed (Bayer junctions) (Bayer and Thurow, 1977; Whitfield et al., 1984; Kröckcke et al., 1990b). This was revealed in temperature up-shift experiments by immunoelectron
microscopy. The K5 polysaccharide was found to appear above Bayer junctions in tufts, before rapidly dispersing over the cell surface (Kröncke et al., 1990b). It is possible that the polysaccharide translocation apparatus could interact directly with the outer membrane at Bayer junctions, obviating the need for polysaccharide to cross the periplasm. However the existence of Bayer junctions is contested and they may be artefacts of the method of preparation of samples for electron microscopy.

There are several aspects of how E. coli capsular polysaccharides are synthesised and translocated to the cell surface which remain to be addressed. What is the acceptor on which the K5 polysaccharide is synthesised? How is the complex of proteins involved in polysaccharide biosynthesis maintained on the membrane? How is the polysaccharide translocated across two membranes to the cell surface? What is the role of phosphatidyl-KDO substitution of the polymer in this process? The model for expression of the K5 antigen presented here represents a starting point for the future investigation of these problems.

1.7 Regulation of gene expression in bacteria

Bacteria must adapt to sudden changes in temperature, osmotic pressure, pH and the availability of oxygen, nutrients and ions. This requires the concerted expression of a large number of often unlinked genes and may be achieved by grouping genes that are needed for a particular response into regulons (Miller et al., 1989). A regulon is a collection of genes under the control of a common regulator. The regulator is often a DNA binding protein which alters the expression of genes by binding to specific DNA sequences in their control regions. The ability of the regulator to carry out its task may be under environmental control, through mechanisms such as covalent modification and/or cofactor binding. Most regulons are concerned with functions essential in commensal life. These include systems to cope with carbon, nitrogen or phosphate starvation, starvation of amino acids (the stringent response), survival of DNA damage (the SOS response), survival of oxidative and osmotic stress and the switch to anaerobic growth. The influence of these regulators is confined to genes which possess the specific sequences to which they can bind. However, other regulatory factors have emerged which affect the expression of a large number of genes without interacting directly with their control sequences. This group includes histone-like proteins, which affect gene expression by altering DNA topology. Some of the mechanisms used to control gene expression are discussed below, mostly with reference to Gram negative bacteria.
1.7.1 Transcriptional control of gene expression

1.7.1.1 Influence of DNA topology

A number of environmental cues, such as temperature, osmolarity and anaerobiosis alter the superhelicity of DNA (Drlica, 1992; Dorman and Ní Bhriain, 1992). This has been shown to affect the expression of a large number of genes, at least some of which may contribute to virulence (Dorman, 1995). DNA supercoiling is principally controlled by the activity of two enzymes, DNA gyrase, which introduces supercoils, and topoisomerase I, which relaxes them (Drlica, 1992). DNA is further organised by histone-like proteins, the most abundant of which is protein HU, a basic heterodimeric protein of 19 kDa which binds to and wraps DNA non-specifically (Drlica and Rouvière-Yaniv, 1987). The subunits of HU are encoded by the unlinked *hupA* and *hupB* genes (Bachmann, 1990). Protein HU is involved in chromosome replication, transposition and recombination events, and is thought to act by organising DNA in a manner which lowers the activation energy for these processes (Rouvière-Yaniv et al., 1990).

Several other histone-like proteins exist, including H-NS, integration host factor and FIS (Factor for Inversion Stimulation). These proteins have been shown to regulate the initiation of transcription and site-specific recombination events by altering DNA topology (Hulton et al., 1990; Freundlich et al., 1992; Finkel and Johnson, 1992). It is thought that these proteins affect transcription by altering the architecture of DNA in the promoter region of genes, either facilitating or inhibiting the formation of an open complex and/or the association of other factors with the DNA (Pérez-Martín et al., 1994; Dorman, 1995; Goosen and Van de Putte, 1995). There are several other DNA binding proteins which affect gene expression in this manner. These include the leucine-responsive regulatory protein (Newman et al., 1992; Calvo and Matthews, 1994) and the cyclic AMP-receptor protein (CRP) (Botsford and Harman, 1992), which bind to specific sequences in the promoter region of the genes they regulate. Another sequence-specific DNA binding protein, FNR, which is structurally similar to CRP, regulates gene expression in Gram negative bacteria under anaerobic conditions (Iuchi and Lin, 1991). It is believed that a large proportion of promoters are sensitive to DNA topology (Pérez-Martín et al., 1994), therefore mutations affecting histone-like proteins are highly pleiotropic. These regulators are discussed in more detail below.

H-NS is a 15.5 kDa basic protein that exists in three isoforms (a, b and c) (Drlica and Rouvière-Yaniv, 1987). The H-NS (a) isoform predominates in stationary phase cells and can strongly compact DNA into structures resembling nucleosomes (Drlica and Rouvière-Yaniv, 1987). The binding of H-NS to DNA occurs non-specifically,
although a preference for curved DNA has been observed (Dorman, 1995). H-NS plays important roles in the response of bacteria to environmental cues during both pathogenic and commensal life (Hulton et al., 1990). For example, H-NS is involved in the adaptation of *E. coli* to osmotic stress (Higgins et al., 1988). Mutations in *hns* alter the expression of certain outer membrane porins and increase transcription of the *proU* locus, which encodes a system for the uptake of the osmoprotectant glycine betaine (Higgins et al., 1988). H-NS also controls the expression of several genes in response to temperature (section 1.7.4) (Dorman et al., 1990; Göransson et al., 1990).

Integration host factor (IHF) was first identified in *E. coli* as a protein required for lysogeny by bacteriophage λ, and is a heterodimeric protein of 21.8 kDa (Friedman, 1988). The α and β subunits of IHF are encoded by the *himA* and *himD* genes respectively (Bachmann, 1990). By two-dimensional polyacrylamide gel electrophoresis of proteins made in *himA* mutant and wild-type *E. coli*, IHF was found to alter the expression of a large number of proteins (Freundlich et al., 1992). This is also reflected in the pattern of transcription in an IHF mutant (Chuang et al., 1993). IHF differs from HU and H-NS in that it interacts with DNA at specific sites (Drlica and Rouvière-Yaniv, 1987). It is believed that IHF binds a site of about 40 bp which contains a 13 bp consensus sequence (WATCAANNNNTTR, where W=A or T, R= A or G, and N= any nucleotide) (Goodrich et al., 1990). The region flanking IHF binding sites is often rich in A+T and this may facilitate DNA bending and strand separation (Goosen and Van de Putte, 1995). IHF is able to bend DNA through 140° and is thought to stimulate some σ70 promoters by assisting enhancer-like elements located upstream of the promoter to contact the promoter region (Collado-Vides et al., 1991; Goosen and Van de Putte, 1995).

The leucine-responsive regulatory protein (LRP), like IHF, is a sequence-specific DNA binding protein (Calvo and Matthews, 1994). LRP binds to the consensus sequence YAGHAWATTWDCTR, where Y=C or T, H= not G, W= A or T, D= not C, and R= A or G (Cui et al., 1995). In *E. coli*, LRP regulates the metabolism of several amino acids, peptide transport, and the expression of certain outer membrane porins (Calvo and Matthews, 1994). Regulation by LRP may be positive or negative and can be potentiated or inhibited by exogenous leucine (Calvo and Matthews, 1994). It is not clear what aspect of the environment leucine represents or how the interaction of LRP with leucine affects the specificity of DNA binding.

The binding of LRP may be modulated by methylation of the DNA. This has been shown to regulate phase variation of Pap pili in *E. coli* K-12 (Blyn et al., 1990; Smyth and Smith, 1992). LRP binds to a deoxyadenosine methylase (Dam) site in the promoter region for the pilin subunit gene, occluding the RNA polymerase recognition site (Braaten et al., 1994). Methylation of this site prevents the binding of LRP and therefore enables pilus expression (Braaten et al., 1994). LRP regulates the
expression of several other types of pili in *E. coli* by a similar mechanism (Low, 1994). Hale *et al.*, (1994) have identified several chromosomal Dam sites that are differentially methylated in response to environmental stimuli. It is possible these sites are involved in the regulation of genes by LRP or other DNA binding proteins.

The cyclic AMP (cAMP)-cAMP receptor protein system regulates the induction of catabolic enzymes in response to the availability of carbohydrates (Botsford and Harman, 1992). In the presence of cAMP, CRP binds to specific sequences in the promoter region of genes and may activate or repress their transcription by locally bending the DNA (Pérez-Martín *et al.*, 1994). The consensus sequence for the binding of CRP is an 11 bp inverted repeat (AAATGTGATCT-6 bp-AGATGAGATTT) (Berg and von Hippel, 1988). The cAMP-cAMP receptor protein system is believed to control over 200 genes and has been implicated in many processes, including cell division, flagellum synthesis and the response of *E. coli* to starvation and anaerobiosis (Botsford and Harman, 1992). In addition, the cAMP-CRP system is required for the virulence of *Salmonella typhimurium* (Curtiss and Kelly, 1987), and negatively regulates the transcription of the plasmid borne virulence genes (O’Byrne and Dorman, 1994).

It is clear that bacteria possess several proteins capable of regulating the transcription of genes by altering DNA topology. For proteins which bind to specific sequences, such as IHF, LRP and CRP, this effect is largely confined to genes which contain the requisite sequences. In contrast, H-NS is able to influence the expression of a much larger number of genes through its role in the organisation of the bacterial nucleoid. Each of the histone-like proteins described here has been shown to be required for the expression of certain virulence genes. This emphasises the importance of the control of DNA topology in the adaptation of bacteria to adverse environments.

1.7.1.2 Alternative RNA polymerase σ factors

Another strategy for the regulation of gene expression is the synthesis of alternative σ factors. The σ subunit of bacterial RNA polymerase confers promoter specificity on the holoenzyme. Most genes in *E. coli*, including those concerned with housekeeping functions, are transcribed by RNA polymerase containing the σ70 subunit. Several alternative σ factors exist which direct RNA polymerase to promoters with different -10 and -35 regions (Gross *et al.*, 1992). The synthesis of alternative σ factors therefore invokes the transcription of a specific collection of genes. Thus, σH and σP direct the transcription of genes that respond to heat-shock (Yura *et al.*, 1993), σN regulates genes involved in nitrogen metabolism (Merrick, 1993), and σE controls gene expression in the starvation response (Loewen and Hengge-Aronis, 1994). In *Bacillus*
subtilis, the sequential expression of genes involved in sporulation is controlled by
the production of a series of alternative \( \sigma \) factors (Errington, 1993).

Alternative \( \sigma \) factors may also control the expression of virulence genes. The
transcription of plasmid borne virulence genes in \( S. \) typhimurium for instance, is
controlled by \( \sigma^d \) (Fang et al., 1992). In Pseudomonas aeruginosa, alternative \( \sigma \) factors
control pilus formation (\( \sigma^m \)) (Ishimoto and Lory, 1989), flagellin synthesis (\( \sigma^e \))
(Starnbach and Lory, 1992), and the biosynthesis of alginate (\( \sigma^t \)) (Deretic et al., 1994), all
of which contribute to virulence. Alternative \( \sigma \) factors that regulate stress responses,
such as \( \sigma^h \), \( \sigma^s \) and \( \sigma^a \) are also important for survival of the bacteria within the host
(Yura et al., 1993; Loewen and Hengge-Aronis, 1994).

The activity of \( \sigma \) factors may be negatively regulated by the binding of specific
proteins (anti-\( \sigma \) factors) (Brown and Hughes, 1995). Anti-\( \sigma \) factors sequester the \( \sigma \)
factor, preventing it from interacting with RNA polymerase or affecting its ability to
recognise its cognate promoter sequence. Bacteriophage T4 uses an anti-\( \sigma \) factor (AsiA)
to sequester the major \( E. \) coli \( \sigma \) subunit (\( \sigma^m \)), while a bacteriophage-encoded \( \sigma \) factor,
gp55, redirects the RNA polymerase to transcribe from T4 late promoters (Orsini et al.,
1993). Anti-\( \sigma \) factors also regulate sporulation in \( B. \) subtilis (Errington, 1993) and
flagellum synthesis in \( S. \) typhimurium (Hughes et al., 1993). Further complexity is
added by the existence of anti anti-\( \sigma \) factors, which positively regulate transcription of
specific genes by preventing the anti-\( \sigma \) factor from interacting with its cognate \( \sigma \) factor
(Brown and Hughes, 1995). An anti anti-\( \sigma \) factor positively regulates flagellum
synthesis in Salmonella by exporting the anti-\( \sigma \) factor out of the cell (Hughes et al.,
1993).

1.7.1.3 DNA rearrangements

The expression of some genes is controlled by DNA rearrangements affecting
their control sequences. Phase variation of type 1 fimbriae in \( E. \) coli K-12 for example,
involves the inversion of a 314 bp segment (switch) carrying the promoter for \( fim A \),
the structural gene for the major fimbrial subunit (Abraham et al., 1985). Thus, \( fim A \)
is transcribed when the switch is in one orientation, but not the other. This inversion
occurs by site-specific recombination between 9 bp inverted repeats located at the ends
of the segment, and is controlled by H-NS (Higgins et al., 1988), IHF, and the FimB and
FimD recombinases (Dorman and Higgins, 1987). Fimbrial phase variation is also
positively regulated by LRP and is potentiated by exogenous leucine (Blomfield et al.,
1993; Gally et al., 1993). Inversion of the \( fim \) switch is sensitive to DNA supercoiling
(Dove and Dorman, 1994), and it is believed that IHF, H-NS and LRP act by altering the
topology of DNA in the \( fim \) region to facilitate or inhibit recombination (Dorman,
Temperature also affects the frequency of fimbrial phase variation, perhaps by affecting DNA supercoiling (Gally et al., 1993).

DNA rearrangements may also affect the coding sequence of a gene. The antigenicity of *Neisseria gonorrhoeae* pili for instance, is controlled by recombination events involving the gene for the pilin subunit (Koomey, 1994). A number of different versions of the pilin gene are located at different sites on the chromosome, but only one has a promoter and is expressed at any one time. Homologous recombination between a silent copy of the gene and the version currently being expressed may place a different version of the gene downstream of the promoter and result in the expression of pilin with a different antigenicity (Koomey, 1994).

The antigenicity of gonococcal pili may also be affected by frameshift mutations in the *pilC* gene, which is involved in processing of the pilin subunit (Jonsson et al., 1991). The amino-terminal end of *pilC* contains multiple repeats of a pentameric DNA sequence. Slipped-strand mispairing of the DNA strands during DNA replication may alter the number of repeats in this region, generating a frameshift mutation which prevents the functional PilC protein from being made (Jonsson et al., 1991). A similar process is believed to control the expression of *N. gonorrhoeae* opacity proteins (Makino et al., 1991), and LPS phase variation in *H. influenzae* type b (Szabo et al., 1992).

1.7.1.4 Transcript processing and stability

In *E. coli* two 3'-5' exoribonucleases (polynucleotide phosphorylase and ribonuclease II), degrade mRNA by processively removing nucleotides from the 3' end (Belasco and Higgins, 1988). The processing of transcripts by site-specific endoribonucleases produces new 3' ends that are susceptible to degradation. Processed transcripts may be stabilised by secondary structures within the mRNA which inhibit the activity of 3'-5' exoribonucleases, thereby allowing the differential expression of cotranscribed genes (Belasco and Higgins, 1988). This occurs in the *E. coli* *pap* operon, where processing of *papBA* transcripts favours the accumulation of *papA* mRNA, which encodes the major pilin subunit (Bága et al., 1988; Smyth and Smith, 1992).

1.7.1.5 Attenuation and antitermination

Several genes in *E. coli* rely on readthrough transcription for their expression (reviewed in Greenblatt et al., 1993). The control of transcription termination may involve translating ribosomes which mask termination signals (attenuation) (Yanofsky, 1981), or specific proteins that either bind to a discrete site on the RNA or
modify RNA polymerase so that it fails to stop at termination signals (antitermination) (Houman et al., 1990; Roberts, 1993).

Attenuation is common in operons which direct the biosynthesis of amino acids (YanoFSky, 1981). In these operons, transcripts pass through a short region (the 'leader'), before entering the structural genes. The leader region encodes a peptide that is rich in the amino acid specified by the operon, as well as secondary structures capable of terminating transcription (an 'attenuator'). As the leader region is transcribed, ribosomes begin to translate the message. If the amino acid is in short supply, the ribosome will stall at the relevant codons. Ribosomes paused in the leader region alter the secondary structure of the RNA and prevent the attenuator from forming. This enables transcription to continue into the structural genes and allows the amino acid to be made.

Transcription antitermination in bacteria is poorly understood and occurs at both Rho-dependent and Rho-independent terminators (Greenblatt et al., 1993). Antitermination in the E. coli bgl operon, which determines the uptake and utilisation of β-glucosides, is mediated by BglG, the product of the first gene of the operon (Schnetz and Rak, 1988). It is believed that BglG binds to a specific sequence in the nascent bgl mRNA such that the formation of a terminator is inhibited (Houman et al., 1990). The activity of BglG requires the formation of a dimer and is modulated by phosphorylation by BglF, an integral membrane sensor kinase that detects β-glucosides (Amster-Choder and Wright, 1992). The Bacillus subtilis SacT protein is thought to act in a similar way to BglG to control transcription of the sacPA operon, which is involved in the utilisation of sucrose (Debarbouille et al., 1990).

The best characterised antitermination factors are the N and Q proteins of bacteriophage λ. These proteins modify RNA polymerase such that it does not recognise termination signals (reviewed in Friedman and Court, 1995). The action of N and Q determines the temporal sequence in which genes are expressed during λ development (Friedman and Gottesman, 1983; Campbell, 1994). The N and Q proteins act in different ways and require host factors and cis-acting sequences for antitermination (Roberts, 1993; Friedman and Court, 1995). Antitermination by N requires a number of host proteins (NusA, NusB, NusG (Nus, N utilisation substance) and ribosomal protein S10), as well as a short sequence nut, which functions in the nascent RNA (Mason et al., 1992; DeVito and Das, 1994). It is believed that N interacts with RNA polymerase as it transcribes the nut site and catalyses the formation of a termination resistant transcription complex (Das, 1992). The Nus factors are thought to stabilise this complex (DeVito and Das, 1994), which can readthrough terminators up to 10 kb from the nut site. The nut site is highly conserved among lambdoid phages and comprises a stem-loop structure (boxB) to which N binds, and a conserved 12 bp motif (boxA), which facilitates the interaction of N (Campbell, 1994).
Modification of RNA polymerase by the Q protein does not require host factors, although it is enhanced by NusA (Grayhack et al., 1985), and occurs while the polymerase is paused at a short DNA sequence (quit) that overlaps the late gene promoter pR' (Yarnell and Roberts, 1992). It is thought that Q acts by accelerating RNA polymerase through a pause site so quickly that it does not interact with termination factors (Jin et al., 1992).

Transcription antitermination also occurs in the ribosomal RNA operons of *E. coli* (reviewed in Morgan, 1986). In *E. coli* there are seven ribosomal RNA operons, *rrnA-G*, which encode the 16S, 23S and 5S rRNAs and some transfer RNAs (Lindahl and Zengel, 1986). The overall organisation of these operons is the same (Figure 1.4).

![Figure 1.4. Generalised organisation of *E. coli* rrn operons.](image)

Figure 1.4. Generalised organisation of *E. coli* *rrn* operons. Open boxes represent the ribosomal RNA and transfer RNA genes. The *rrn* promoters are designated P₁ and P₂. The arrow represents the 30S RNA that is processed to give the different RNA species. Filled boxes represent the boxA-like sequences.

In all *rrn* operons, tandem promoters and a conserved 'leader' region precede the 16S rRNA gene. The region between the 16S rRNA and 23S rRNA genes (the 'spacer') is also conserved and contains one or two tRNA genes. The 23S rRNA genes are followed by 5S rRNA genes and, in some cases, one or two tRNA genes (Lindahl and Zengel, 1986). Transcription of *rrn* operons generates a single 30S RNA that is processed by ribonuclease III to release the individual rRNA and tRNA species (Lindahl and Zengel, 1986).

Ribosomal RNA operons are thought to contain multiple transcription terminators, and yet premature termination is not observed (Morgan, 1986). Early evidence for antitermination in *rrn* operons came from the observation that Tn9, Tn10 and IS1 insertions, which introduce strong Rho-dependent terminators, are not polar (Morgan, 1980; Brewster and Morgan, 1981). In common with the λ N system, antitermination in *rrn* operons requires NusA, NusB, NusG and the ribosomal protein S10 (Squires *et al.*, 1993; Nodwell and Greenblatt, 1993). At least one other protein is also required, but this has yet to be characterised (Squires *et al.*, 1993). This protein may catalyse the formation of a termination resistant elongation complex in...
an analgous way to the λ N protein. Antitermination in rrr operons also requires short sequences in the leader region that are homologous to boxA of the λ nut site (Figure 1.4) (Li et al., 1984; Berg et al., 1989, Albrechtsen et al., 1990; Heinrich et al., 1995). BoxA-like sequences are also located in the spacer region, and can cause antitermination in an in vitro system (Berg et al., 1989).

Antitermination also occurs in the region between the rplKAJL and rpoBC genes in E. coli (Barry et al., 1980). Transcription of the rpoBC genes, which encode subunits of RNA polymerase, relies on the readthrough of transcripts originating from promoters located in rplK and rplJ and is regulated, in part, by an attenuator in the rplL-rpoB intergenic region (Ralling and Linn, 1984). Antitermination in the rplL-rpoB intercistronic region, is increased by NusA and the RfaH protein (Ralling and Linn, 1987). RfaH regulates several virulence and fertility genes in E. coli and has been suggested to act by modifying RNA polymerase in a manner analagous to the N or Q proteins (Beutin et al., 1981; Gaffney et al., 1983; Bailey et al., 1992; Schnaitman and Klena, 1993).

### 1.7.2 Translational control of gene expression

The initiation of translation involves the binding of ribosomes to a sequence in the mRNA, located just upstream of the start codon, that is complementary to the sequence 3'-UCCUCCA-5' at the 3' end of the 16S ribosomal RNA (Shine and Dalgarno, 1974). Statistical analysis has indicated that the extent of homology between the Shine-Dalgarno sequence and the 16S rRNA dictates, in part, the efficiency of translation initiation (Gold et al., 1981). The secondary structure of the mRNA may also affect translation initiation, by altering the accessibility of the Shine-Dalgarno sequence and start codon to ribosomes (De Smit and Van Duin, 1990). The stability of these secondary structures under varying environmental conditions may control expression of the gene. A stem-loop structure in the Yersinia pestis lcrF mRNA is thought to regulate translation of the gene in response to temperature (section 1.7.4) (Hoe and Goguen, 1993). The translational coupling of genes on a polycistronic mRNA may also be regulated by secondary structures which inhibit re-initiation of ribosomes once the upstream gene has been translated (Gold, 1988). This would permit the differential expression of genes within an operon.

Translation initiation may also be regulated by the binding of an antisense RNA or protein to the region where initiation normally occurs. This may occlude the Shine-Dalgarno sequence and/or affect the stability of secondary structures within the mRNA. The response of E. coli to high osmolarity for instance, involves the binding of an antisense RNA (micF) to the region which initiates translation of the OmpF porin (Inouye, 1988). In addition, the expression of bacteriophage T4 DNA polymerase
is autogenously regulated by binding of the protein to the ribosome binding site of its own mRNA (Gold, 1988).

1.7.3 Two-component regulatory systems

In bacteria, the expression of genes in response to environmental cues is often regulated by two-component systems (Miller et al., 1989; Gross, 1993; Parkinson, 1993). This family is typically composed of two different types of protein, a sensor protein which spans the membrane and a response regulator located in the cytoplasm. Environmental signals are perceived by the periplasmic domain of the sensor and invoke a histidine kinase activity that is furnished by the cytoplasmic domain. This requires a change in the conformation of the sensor and causes autophosphorylation. The phosphate group is then transferred from the sensor to a conserved N-terminal domain of the response regulator (RR), which is typically a transcriptional activator. Phosphorylation of the RR modulates its activity, and usually activates the transcription of genes needed for a particular response. Response regulators can be divided into families on the basis of homology between their C-terminal domains (Gross et al., 1989).

Two-component systems regulate the expression of genes needed for both pathogenic and commensal life. In E. coli, two-component systems regulate genes for adaptation to osmotic stress, colanic acid production, chemotaxis and the metabolism of nitrogen, phosphate and certain sugars (Gross, 1993). The system which regulates the chemotactic response of E. coli is unusual in that the sensor and response regulator proteins are located in the cytoplasm (Parkinson, 1993). Activation of the histidine kinase depends on the methylation state of membrane proteins (chemoreceptors) that sense amino acids, peptides or sugars, and does not involve the transfer of phosphate (reviewed in Parkinson, 1993).

In order to ensure rapid adaptation of two-component systems in response to changes in the environment, some sensor proteins have a phosphatase activity and can dephosphorylate the RR (Gross, 1993). In other cases the modified RR has a short half-life. Sensor proteins from different two-component systems can sometimes phosphorylate the same RR (cross-talk) (Parkinson, 1993). This means that several environmental cues can affect the expression of specific genes.

1.7.4 Temperature regulation of virulence genes

Temperature appears to be an important stimulus for the expression of virulence genes in bacteria (Maurerelli, 1989; Mekalanos, 1992). The induction of virulence genes at temperatures normally encountered within an animal host
ensures that the genes are only expressed when it is appropriate. In *E. coli*, several virulence factors are expressed at the cell surface at 37°C but not at lower temperatures (typically below 25°C). These include group II K antigens (section 1.4.2), pyelonephritis associated (Pap) pili (Göransson and Uhlin, 1984), K88 and K99 pili (Guinee and Jansen, 1979; Roosendaal *et al.*, 1986), the colonisation factor antigen (Evans *et al.*, 1977), and the TraT protein encoded by the F factor, which is involved in resistance to phagocytosis (Sukupolvi and O'Connor, 1990).

Thermoregulation of virulence genes may depend on DNA supercoiling (Dorman and Ní Bhriain, 1992). An example is provided by the regulation of invasion genes in *Shigella flexneri* (Dorman, 1995). In *S. flexneri*, the genes needed for invasion of the intestinal mucosa are regulated by temperature and osmolarity (Porter and Dorman, 1994), being expressed at 37°C but not 30°C. Transcription of the invasion genes is activated by the VirB protein and is reduced at 30°C because *virB* transcription is repressed (Tobe *et al.*, 1991). It has been shown that the repression of *virB* transcription is due to the binding of H-NS in the *virB* promoter region (Dorman *et al.*, 1990), and involves a change in the local DNA topology (Tobe *et al.*, 1993). The repression of *virB* transcription by H-NS is relieved by the binding of a positive regulator (VirF) in the promoter region for *virB*, although it is not clear what triggers this (Tobe *et al.*, 1993). The thermal induction of *virB* transcription does not occur in low osmolarity media (Porter and Dorman, 1994), indicating that a combination of environmental signals is needed for expression of the invasion genes.

Transcription of the *E. coli pap* operon is also repressed by H-NS at low growth temperatures (Göransson *et al.*, 1990). By transposon mutagenesis, a second locus involved in thermoregulation of the *pap* operon was identified (White-Ziegler *et al.*, 1990). This locus (tcp) was later shown to be an allele of *rimJ*, which encodes the N-terminal acetylase of ribosomal protein S5 (White-Ziegler *et al.*, 1992). The tcp locus is not thermoregulated, since transcripts were detected at both 23°C and 37°C. It is not clear what role, if any, acetylation of ribosomal protein S5 has in the assembly or function of ribosomes, or how this may affect transcription of the *pap* operon. It is possible that RimJ may modulate the activity of another regulator of the *pap* operon by acetylation.

In *Bordetella pertussis*, a two-component sensory system regulates the expression of several virulence genes in response to temperature (Gross, 1993). This system, encoded by the *bvgA* and *bvgS* genes, also responds to sulphate and nicotinate. BvgS is the sensor protein and is located in the membrane. The isolation of mutations in *bvgS* which prevent the temperature-dependent expression of virulence genes has shown that BvgS is capable of sensing temperature (Miller *et al.*, 1992). BvgS is peculiar in so far that it contains, in addition to the histidine-kinase domain, a highly conserved receiver domain (Gross, 1993). This domain may be used for
communication with other two-component sensors, or may directly regulate gene expression. An analogous system has not been identified in *E. coli*.

Temperature regulation of virulence genes may also be achieved at the level of translation. In *Yersinia pestis*, translation of the LcrF protein is thermally regulated (Hoe and Goguen, 1993). LcrF is a transcriptional activator which induces the expression of several virulence genes at 37°C (Hoe et al., 1992). Much lower levels of the LcrF protein are found at 26°C than at 37°C, even though the *lcrF* gene is transcribed at similar levels at both temperatures. Secondary structure predictions for the *lcrF* mRNA have indicated that the Shine-Dalgarno sequence is sequestered within a stem-loop (Hoe and Goguen, 1993). It is proposed that with increasing temperature the stem-loop becomes unstable, allowing translation to initiate more efficiently.

### 1.8 Regulation of extracellular polysaccharide expression in bacteria

#### 1.8.1 Regulation of colanic acid biosynthesis in *E. coli*

Production of colanic acid is increased under conditions of environmental stress, in particular desiccation (Ophir and Gutnick, 1994), low growth temperature and limitation of certain nutrients (Markovitz, 1977; Gottesman and Stout, 1991). Transcription of the *E. coli* K-12 *cps* gene cluster for the biosynthesis of colanic acid is regulated by a two-component sensory system encoded by the *rcsB* and *rcsC* genes (*res*, regulation of capsule synthesis) (Figure 1.5) (Stout and Gottesman, 1990). It is proposed that RcsB is the response regulator and RcsC is the sensor. The phosphorylation of RcsB and its ability to bind DNA have not been demonstrated experimentally and the prediction is based on the homology of RcsB to other response regulators. It has been suggested that another protein, RcsF, activates RcsB by phosphorylation and that RcsC serves not as a kinase, but as an environmentally responsive phosphatase (Gervais and Drapeau, 1992). It is possible that the phosphorylation state of RcsB may be modulated by other proteins, allowing capsule expression to be regulated by many factors. Gervais et al., (1992) have shown that RcsB also activates the expression of *ftsZ*, a gene involved in cell division, suggesting that the *rcs* system regulates other cellular functions.

Genetic data indicate that RcsB activates transcription of the *cps* genes by acting in concert with another positive regulator, RcsA (Stout et al., 1991). RcsA belongs to the LuxR family of transcriptional activators and has a helix-turn-helix DNA binding motif, although DNA binding has not been demonstrated (Stout et al., 1991). RcsA is an unstable protein and is rapidly degraded by Lon, an ATP-dependent protease which
controls induction of the SOS response (Gottesman and Maurizi, 1992). The formation of an active RcsA-RcsB complex may require the chaperone DnaK (Zuber et al., 1995).

The transcription of rcsA is repressed by H-NS, suggesting a role for DNA supercoiling in the control of colanic acid production (Sledjeski and Gottesman, 1995). This notion is supported by the finding that an imbalance in the synthesis of protein HU subunits increases colanic acid biosynthesis (Painbéri et al., 1993). The repression of rcsA transcription by H-NS is relieved by a small (85 nucleotide) RNA, dsrA, which is encoded by a gene downstream of rcsA (Sledjeski and Gottesman, 1995). The dsrA RNA is a positive regulator of other H-NS repressed operons (e.g., pap and proU), and is believed to act by preventing the association of H-NS with DNA in the promoter region (Sledjeski and Gottesman, 1995).

The transcription of rcsB also appears to be regulated in a complex manner. The promoter region for rcsB contains potential \( \sigma^70 \) and \( \sigma^N \) recognition sites and a putative binding site for LexA, a repressor of operons involved in the SOS response (Stout and Gottesman, 1990; Gervais et al., 1992). Transcription of rcsB is not \( \sigma^N \) dependent, but it is suggested that the promoter used may depend on the prevailing environmental conditions and interactions with other effectors (Gervais et al., 1992).
similar to the $\sigma^N$ recognition site have been identified upstream of many operons for polysaccharide biosynthesis in enteric bacteria (Arawaka et al., 1995).

The rcs system also regulates the expression of E. coli group IA K antigens (Keenleyside et al., 1992; Jayaratne et al., 1993). E. coli expressing the group IA K30 antigen do not produce colanic acid and the cps and K30 capsule genes appear to be allelic (Keenleyside et al., 1992). Cloning of the genes for rcsA, rcsB and rcsC from E. coli O9:K30 has revealed that they are almost identical to their counterparts in E. coli K-12 (Keenleyside et al., 1992; Jayaratne et al., 1993). The RcsA and RcsB proteins are not essential for expression of the K30 antigen, but are required for high levels of synthesis (Jayaratne et al., 1993). Increasing the copy number of rcsA and rcsB dramatically increases synthesis of the K30 polysaccharide, as do mutations in the lon and rcsC genes (Keenleyside et al., 1992; Jayaratne et al., 1993).

E. coli expressing group IB and group II K antigens are able to simultaneously express colanic acid with the K antigen and contain functional copies of rcsA and rcsB (Keenleyside et al., 1992, 1993). It is not thought that expression of group II K antigens is positively regulated by the rcs system, as the presence of multicy copy rcsA or rcsB does not affect synthesis of the K1 and K5 antigens in vitro (Keenleyside et al., 1992, 1993). Russo and Singh (1993) reported that RcsA negatively regulates expression of the K54 antigen. This effect required RcsB and varied with the growth temperature. The K54 antigen was described by the authors as a group II K antigen, however it is made at all growth temperatures and genetic data suggest that it belongs to group III (Pearce and Roberts, 1995). Thus, a clearly defined role for the rcs system in the expression of group II capsular polysaccharides has not been demonstrated.

The rcs system also exists in other bacteria and regulates the expression of structurally distinct extracellular polysaccharides (reviewed in Whitfield and Valvano, 1993). In K. pneumoniae for example, RcsA regulates expression of the serotype-specific K antigen (McCallum and Whitfield, 1991), whereas in Citrobacter freundii and Salmonella enterica serovars typhi, paratyphi C and dublin, RcsB is required for expression of the Vi antigen (Houng et al., 1992). Expression of the S. typhi Vi antigen, which resembles group II capsular polysaccharides, also requires the OmpR-EnvZ two-component system, which regulates the response to osmotic stress (Pickard et al., 1994).

1.8.2 Regulation of alginate biosynthesis in Pseudomonas aeruginosa

The first committed step in the biosynthesis of alginate in P. aeruginosa is catalysed by AlgD, a GDP-mannose dehydrogenase, and is subject to complex regulation at the level of algD transcription (reviewed in Deretic et al., 1994). This involves a two-component sensory system, where AlgR is the response regulator and
AlgQ is the sensor (Gross, 1993). AlgQ is unusual in that it is a cytoplasmic protein. An additional response regulator, AlgB, is needed for the high-level expression of alginate (Goldberg and Dahnhke, 1992). The transcription of algD is directed by an alternative σ factor which is homologous to σE of E. coli (Deretic et al., 1994). The activity of this σ factor is modulated by two proteins MucA and MucB, which appear to be environmentally responsive and may act as anti-σ factors (Deretic et al., 1994). Transcription of algD is also controlled by histone-like proteins including AlgP (Kato et al., 1990), integration host factor (Wozniak, 1994), and the cAMP-receptor protein (DeVault et al., 1991). It is believed these proteins act by altering DNA topology in the promoter region for algD and may facilitate the interaction of other effectors.

1.8.3 Regulation of lipopolysaccharide biosynthesis in E. coli and S. typhimurium

The structure of LPS, in particular the size and number of O chains, has been shown to be affected by growth conditions (Nelson et al., 1991). In E. coli K-12, an increase in growth temperature from 30°C to 42°C reduces both the O chain length and the total number of LPS cores capped with O antigen (Pradel and Schnaitman, 1991). The O polysaccharide chain length in E. coli and Salmonella is regulated by a gene termed rol (regulator of O-chain length), which is distinct from the O antigen polymerase and ligase genes (Batchelor et al., 1992). It is not known how the activity of Rol is regulated. LPS biosynthesis also requires the rfaH (sfrB) gene (Creeger et al., 1984). The rfaH gene encodes a protein of 18.3 kDa and is located at 87 minutes on the E. coli chromosome, outside the gene clusters for biosynthesis of the core oligosaccharide (rfa) and O antigen (rfb) (Rehemtulla et al., 1986; Bachmann, 1990). RfaH mutants of S. typhimurium produce rough LPS with a heterogeneous core structure (Lindberg and Hellerqvist, 1980).

In E. coli K-12, the genes for biosynthesis of the core oligosaccharide are organised in three transcriptional units (Figure 1.6) (reviewed in Schnaitman and Klena, 1993). This was deduced from complementation studies (Austin et al., 1990), and by the insertion of polar O cassettes, using TnlacZ transcriptional fusions to monitor the expression of downstream genes (Roncero and Casadaban, 1992). The rfa genes in Salmonella spp. are believed to be similarly organised (Schnaitman and Klena, 1993).

By primer extension analysis, divergent promoters for kdtA and the rfaQ-K (long) operon have been mapped (Clementz, 1992). The region between kdtA and rfaQ is highly conserved between E. coli and S. typhimurium and may be important for regulation of the rfa genes (Klena et al., 1993; Brazas et al., 1991). There is no evidence for internal promoters within the long rfa operon (Roncero and Casadaban, 1992).
Figure 1.6. Genetic organisation of the *E. coli* K-12 *rfa* genes. Boxes represent the open reading frames. Arrows indicate the direction of transcription. Adapted from Schnaitman and Klena (1993).

By using *lacZ* transcriptional fusions in the *S. typhimurium* *rfaGBIJ* operon, Brazas *et al.* (1991) have shown that RfaH activates the transcription of the entire *rfa* long operon. Similarly, the expression of TnLacZ translational fusions along the *E. coli* K-12 long *rfa* operon is activated by RfaH (Pradel and Schnaitman, 1991). In both cases, the extent to which the fusions were regulated by RfaH increased with their distance from the promoter. It has been proposed that RfaH acts to allow transcription through multiple Rho-dependent terminators in the long *rfa* operon (antitermination) (Farewell *et al.*, 1991; Pradel and Schnaitman, 1991; Brazas *et al.* 1991). Indirect support for this notion was provided by the isolation of extragenic suppressors of the *rfaH* phenotype in *S. typhimurium* in the genes encoding termination factor Rho and subunits of RNA polymerase (Farewell *et al.*, 1991). Thus, RfaH may interfere with the ability of RNA polymerase to recognise Rho-dependent termination signals. Brazas *et al.*, (1991) reported evidence that a terminator exists in or after *rfaG*, but before *rfaB*, in *S. typhimurium*.

RfaH is also required for the expression of F factor *tra* genes (Beutin and Achtman, 1979; Sanderson and Stocker, 1981), and the *E. coli* *hly* operon for the synthesis and secretion of α-haemolysin (Bailey *et al.*, 1992). Hybridisation and gene fusion studies have indicated that RfaH is a transcriptional activator of both these operons, and may act by causing antitermination (Beutin *et al.*, 1981; Gaffney *et al.*, 1983; Bailey *et al.*, 1992). RfaH also increases readthrough transcription in the *rplL*-rpoB intergenic region (Ralling and Linn, 1987).
1.9 The aims of this thesis

It may be predicted that expression of the capsule genes will be tightly controlled in order to ensure that the proteins involved in biosynthesis and export of the polysaccharide are made in the correct stoichiometry. The aim of this study is to identify proteins that regulate the expression of *E. coli* group II capsule genes. Using expression of the K5 capsule gene cluster as a paradigm, the role of known regulatory proteins will be investigated. Involvement of RfaH, which regulates core LPS biosynthesis, and Rcs proteins, which control the expression of *E. coli* group I K antigens and other polysaccharides in enteric bacteria, will also be examined. In addition, it is intended to determine the mechanism by which the regulatory protein(s) act. Finally, it is hoped that the experiments performed in this study will increase our understanding of how group II capsule expression is regulated by temperature.
Chapter 2

Materials and methods

2.1 Bacterial strains and plasmids

The bacterial strains and plasmids used in this study are listed in Tables 2.1 and 2.2 respectively. Strains given the abbreviation CGSC, were obtained from B. Bachmann of the E. coli Genetic Stock Centre, Yale University, U.S.A. Tables 2.3 and 2.4 list the strains and plasmids generated in this study.

2.1.1 Growth conditions and media

Luria broth (L-broth) was used for routine growth of bacterial strains. L-broth of lower osmolarity (50mM NaCl) was used to grow topoisomerase I mutants RED31 and MS103 (Tables 2.1 and 2.3) to reduce the selection for compensatory mutations. Minimal medium M9 supplemented with thiamine (10μg/ml), glucose (or other carbon source) (0.4% (w/v)), and amino acids (50μg/ml) was used where appropriate. Blood agar was prepared using blood agar base number 2 (Oxoid) and contained 7% (v/v) horse blood (Difco). Antibiotics were used at the following concentrations; ampicillin (Ap), 100μg/ml; chloramphenicol (Cm), 25μg/ml; kanamycin (Km), 25μg/ml; streptomycin (Sm), 25μg/ml and tetracycline (Tc), 25μg/ml. Media were sterilised by autoclaving at 121°C (15 pounds/square inch) for 15 minutes.

Luria broth (1L)
10g Tryptone
5g Yeast extract
5g NaCl
Distilled water to 1L

For Luria agar, Bacto-agar (Difco) was added at 1.5% (w/v).

Minimal medium M9 (1L)
800ml sterile distilled water or molten 1.5% (w/v) agar (<60°C)
100ml 10x M9 salts (0.96M Na₂HPO₄, 0.22M KH₂PO₄, 0.09M NaCl, 0.19M NH₄Cl)
1ml 100mM CaCl₂
2ml 1M MgSO₄
Additional supplements and sterile distilled water to 1L.
<table>
<thead>
<tr>
<th>Strain</th>
<th>Relevant genotype</th>
<th>Source/reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bi8337-41</td>
<td>O10:K5:H4 supE44 supF58 lacY1 galK2 galT22 metB1 trpR55</td>
<td>I. Ørskov</td>
</tr>
<tr>
<td>LE392</td>
<td>[F hsdR74 (r5-, mcr-)] supE44 supF58 lacY1 galK2 galT22 metB1 trpR55</td>
<td>Murray et al., (1977)</td>
</tr>
<tr>
<td>JM101</td>
<td>[F traD36 proAB lacI2 ΔM15 lac-proAB] supE thi</td>
<td>Yanisch-Perron et al., (1985)</td>
</tr>
<tr>
<td>BW96F</td>
<td>F+ tdk deoA deoC</td>
<td>Boulnois and Wilkins, (1979)</td>
</tr>
<tr>
<td>PA360</td>
<td>F* serA1 rpsL9 thr1 leuB6 hisG1 argH1 thi1 rfbD1 malT1 (ΔfhuA2 lacY1 gal6 galP63 xyf7 himA2 supF44)</td>
<td>CGSC#288</td>
</tr>
<tr>
<td>GM230</td>
<td>hns205::Tn10</td>
<td>Higgins et al., (1988)</td>
</tr>
<tr>
<td>RED31</td>
<td>topA20::Tn10</td>
<td>Dorman et al., (1989)</td>
</tr>
<tr>
<td>HN1491</td>
<td>ΔhimA::CmR</td>
<td>M. Freundlich</td>
</tr>
<tr>
<td>CH1569</td>
<td>ΔhimD::CmR</td>
<td>Dorman and Higgins, (1987)</td>
</tr>
<tr>
<td>1.8</td>
<td>rfaH::Tn5</td>
<td>Bailey et al., (1992)</td>
</tr>
<tr>
<td>CV1008</td>
<td>trp35::Tn10</td>
<td>Platko et al., (1990)</td>
</tr>
<tr>
<td>VS20184</td>
<td>ΔrcsA51::KmR</td>
<td>C. Whitfield</td>
</tr>
<tr>
<td>VS20185</td>
<td>ΔrcsB62::KmR</td>
<td>C. Whitfield</td>
</tr>
<tr>
<td>CA8445</td>
<td>Δcrp ΔcytA</td>
<td>CGSC#7042</td>
</tr>
<tr>
<td>KY1429</td>
<td>rpoH</td>
<td>CGSC#6934</td>
</tr>
<tr>
<td>ET8045</td>
<td>rpoN::Tn10</td>
<td>M. Merrick</td>
</tr>
<tr>
<td>UM56-64</td>
<td>rpoS</td>
<td>CGSC#7157</td>
</tr>
<tr>
<td>OHP109</td>
<td>hupA::CmR</td>
<td>Painbéní et al., (1993)</td>
</tr>
<tr>
<td>OHP96</td>
<td>hupB::CmR</td>
<td>Painbéní et al., (1993)</td>
</tr>
<tr>
<td>DL652</td>
<td>tcp1::mTn10</td>
<td>White-Ziegler and Low, (1992)</td>
</tr>
<tr>
<td>ED3869</td>
<td>sfrB11</td>
<td>CGSC#6990</td>
</tr>
<tr>
<td>CA10</td>
<td>galK9</td>
<td>CGSC#4973</td>
</tr>
<tr>
<td>A586</td>
<td>spoC</td>
<td>CGSC#4925</td>
</tr>
<tr>
<td>SM10λpir</td>
<td>λpir recA::RP4-2·Tc::Mu KmR</td>
<td>Donnenberg and Kaper, (1991)</td>
</tr>
</tbody>
</table>
Table 2.2. Plasmids used in this study.

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Relevant genotype</th>
<th>Source/reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pUC19</td>
<td>Ap^ high copy number cloning vector</td>
<td>Yanisch-Perron et al., (1985)</td>
</tr>
<tr>
<td>pACYC184</td>
<td>Cm^Te^ medium copy number cloning vector</td>
<td>Chang and Cohen, (1978)</td>
</tr>
<tr>
<td>M13mp19</td>
<td>Bacteriophage cloning/sequencing vector</td>
<td>Yanisch-Perron et al., (1985)</td>
</tr>
<tr>
<td>pPC6</td>
<td>Cm^ K5 kps in pACYC184</td>
<td>Pazzani, (1992)</td>
</tr>
<tr>
<td>pAS30</td>
<td>Cm^ 9.2 kb XhoI-BgII fragment from K5 region 2 in pACYC184</td>
<td>A. Smith</td>
</tr>
<tr>
<td>pAS1</td>
<td>Ap^ 1.8 kb BglII fragment from K5 region 3 in pUC19</td>
<td>Smith et al., (1990)</td>
</tr>
<tr>
<td>pCR6</td>
<td>Ap^ 2.2 kb SmaI-HincII fragment from K5 region 1 in expression vector pCE30</td>
<td>Rosenow et al., (1995a)</td>
</tr>
<tr>
<td>pCP1</td>
<td>Ap^ 7.8 kb HindIII fragment spanning K5 region 3 in pTZ19R</td>
<td>C. Petit</td>
</tr>
</tbody>
</table>
Table 2.3. *Escherichia coli* strains generated in this study.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Relevant genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS101</td>
<td>As PA360 (Table 2.1) but serA+ K5 kps+</td>
</tr>
<tr>
<td>MS102</td>
<td>MS101 hns::Tn10</td>
</tr>
<tr>
<td>MS103</td>
<td>MS101 topA::Tn10</td>
</tr>
<tr>
<td>MS104</td>
<td>MS101 ΔhimA::CmR</td>
</tr>
<tr>
<td>MS105</td>
<td>MS101 ΔhimD::CmR</td>
</tr>
<tr>
<td>MS106</td>
<td>MS101 rfaH::Tn5</td>
</tr>
<tr>
<td>MS107</td>
<td>MS101 lrp::Tn10</td>
</tr>
<tr>
<td>MS108</td>
<td>MS101 kfsA::luxAB</td>
</tr>
<tr>
<td>MS109</td>
<td>MS101 ΔrcsA::KmR</td>
</tr>
<tr>
<td>MS110</td>
<td>MS101 ΔrcsB::KmR</td>
</tr>
<tr>
<td>MS111</td>
<td>MS101 hupA::CmR</td>
</tr>
<tr>
<td>MS112</td>
<td>MS101 hupB::KmR</td>
</tr>
<tr>
<td>MS113</td>
<td>MS101 tcp::mTn10</td>
</tr>
<tr>
<td>MS115</td>
<td>MS108 rfaH::Tn5</td>
</tr>
<tr>
<td>MS117</td>
<td>MS108 ΔhimA::CmR</td>
</tr>
<tr>
<td>MS120</td>
<td>MS101 kpsT::luxAB</td>
</tr>
<tr>
<td>MS121</td>
<td>MS120 rfaH::Tn5</td>
</tr>
<tr>
<td>MS122</td>
<td>MS120 hns::Tn10</td>
</tr>
<tr>
<td>MS124</td>
<td>MS120 tcp::mTn10</td>
</tr>
<tr>
<td>MS130</td>
<td>MS101 kpsE::luxAB</td>
</tr>
<tr>
<td>MS131</td>
<td>MS130 rfaH::Tn5</td>
</tr>
<tr>
<td>MS135</td>
<td>JM101 rfaH::Tn5</td>
</tr>
<tr>
<td>MS136</td>
<td>MS101 with linker insertion in JUMPstart</td>
</tr>
<tr>
<td>MS137</td>
<td>MS101 with JUMPstart deletion</td>
</tr>
</tbody>
</table>
Table 2.4. Plasmids generated in this study.

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pMS1</td>
<td>kfiA::luxAB fusion in pAS21. 2.4 kb Smal-PvuII fragment of pHV100 cloned into XmnI (partial) cut pAS21.</td>
</tr>
<tr>
<td>pMS2</td>
<td>kfiA::luxAB fusion in pCVD442. 4.2 kb Smal-StuI fragment of pMS1 cloned into Smal cut pCVD442.</td>
</tr>
<tr>
<td>pMS3</td>
<td>kpsT::luxAB fusion in pAS1. 2.4 kb Smal-PvuII fragment of pHV100 cloned into EcoRV cut pAS1.</td>
</tr>
<tr>
<td>pMS4</td>
<td>kpsT::luxAB fusion in pCVD442. 4.3 kb SacI-SphI fragment pMS3 cloned into SacI-SphI cut pCVD442.</td>
</tr>
<tr>
<td>pMS5</td>
<td>kpsE::luxAB fusion in pCR6. 2.4 kb Smal-PvuII fragment of pHV100 cloned into EcoRV cut pCR6.</td>
</tr>
<tr>
<td>pMS6</td>
<td>kpsE::luxAB fusion in pCVD442. 4.5 kb Smal-Bst1107I fragment of pMS5 cloned into Smal cut pCVD442.</td>
</tr>
<tr>
<td>pMS8</td>
<td>kfiA::lacZ fusion in pCVD442. 5.7 kb XmnI-Sall fragment of pMS7 cloned into Smal-Sall cut pCVD442.</td>
</tr>
<tr>
<td>pMS9</td>
<td>kfiA::lacZ fusion in pCP1. 4.7 kb PstI fragment of pKOK6 cloned into NsiI cut pCP1.</td>
</tr>
<tr>
<td>pMS10</td>
<td>pMS9 with linker insertion in JUMPstart. 30-mer (oligonucleotides MS9 and MS10) cloned into StyI cut pMS9.</td>
</tr>
<tr>
<td>pMS11</td>
<td>pAS1 with linker insertion in JUMPstart. 30-mer (oligonucleotides MS9 and MS10) cloned into StyI cut pAS1.</td>
</tr>
<tr>
<td>pMS12</td>
<td>Region 3 with linker insertion in JUMPstart in pCVD442. 1.9 kb SacI-SphI fragment of pMS11 cloned into SacI-SphI cut pCVD442.</td>
</tr>
<tr>
<td>pMS13</td>
<td>2.2 kb XmnI-StuI region 3 fragment from pCP1 cloned into HincII cut pUC19.</td>
</tr>
<tr>
<td>pMS14</td>
<td>pMS13 with JUMPstart deletion. Cut pMS13 with BsaBI and BstXI, blunted BstXI end with T4 DNA polymerase and ligated.</td>
</tr>
</tbody>
</table>
Table 2.4. Plasmids generated in this study (continued).

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pMS15</td>
<td>pMS9 with JUMPstart deletion. Replaced 1.8 kb region 3 BglII fragment with the region 3 BglII fragment from pMS14.</td>
</tr>
<tr>
<td>pMS17</td>
<td>2.5 kb CiaI 5’ region 3 fragment from pCP1 cloned into AccI cut M13mp19.</td>
</tr>
<tr>
<td>pMS18</td>
<td>As pMS17 but with insert in opposite orientation.</td>
</tr>
<tr>
<td>pMS23</td>
<td>2.3 kb FspI-Smal fragment from pMS18 cloned into HincII cut pTZ18R.</td>
</tr>
<tr>
<td>pMS24</td>
<td>1.8 kb BglII region 3 fragment from pMS14 cloned into BamHI cut pTZ18R.</td>
</tr>
<tr>
<td>pMS25</td>
<td>Region 3 with JUMPstart deletion in pCVD442. 1.8 kb SacI-SalI fragment of pMS24 cloned into SacI-SalI cut pCVD442.</td>
</tr>
<tr>
<td>pMS26</td>
<td>As pMS23 but with a deletion from the HincII site 5’ region 3 to the SalI site in polylinker.</td>
</tr>
</tbody>
</table>

2.2 Transformation of E. coli

2.2.1 Preparation of competent cells

Competent cells were prepared as described by Sambrook et al., (1989). Bacteria were grown to mid-logarithmic phase (A600 ~ 0.5) in L-broth (10ml) and collected by centrifugation at 2750 x g for 10 minutes at 4°C. The cells were washed once with 10mM NaCl then resuspended in 4ml ice-cold 100mM CaCl2 and incubated on ice for 30 minutes. Competent cells were then collected by gentle centrifugation (1800 x g) at 4°C for 5 minutes and resuspended in 1ml ice-cold 100mM CaCl2.

2.2.2 Transformation with plasmid DNA

Competent cells (100µl) and typically 0.1 to 1µg plasmid DNA were mixed and placed on ice for 1 hour. The cells were then incubated at 42°C for 3 minutes and 0.5ml L-broth added. Transformed cells were incubated at 37°C for 1 hour and dilutions then plated onto selective medium.
2.2.3 Transformation with bacteriophage DNA

*E. coli* JM101 (Table 2.1) competent cells were mixed with recombinant M13mp19 bacteriophage DNA and placed on ice for 1 hour. Following incubation at 42°C for 3 minutes, the transformed cells were mixed with 100μl JM101 cells grown to mid-logarithmic phase. Next, 3ml molten soft top B-agar (at 42°C) containing 20μl IPTG (100mM) and 50μl X-gal (2% (w/v) in dimethylformamide) was added and the suspension mixed and poured onto a B-agar plate. Once set, plates were incubated at 37°C for 12 to 18 hours.

B-agar (1L)
1g Peptone
8g NaCl
15g Agar
Distilled water to 1L.
B-soft top agar contains 8g agar per litre.

2.2.4 Transformation of *E. coli* by electroporation

Electro-competent *E. coli* were prepared by the method of *Dower et al.*, (1988). Bacteria were grown to mid-logarithmic phase in L-broth (10ml) and then cooled on ice for 15 minutes. The cells were collected by centrifugation (2750 x g for 10 minutes at 4°C) and washed three times in ice-cold sterile distilled water then once in 10% (v/v) glycerol. The cells were then resuspended in 100μl 10% (v/v) glycerol and 40μl mixed with 0.1 to 1μg plasmid DNA (usually not more than 2μl plasmid DNA solution) in a chilled 0.2cm electroporation cuvette (Bio-Rad). A Bio-Rad Gene Pulser and Pulse Controller apparatus was used to apply a potential difference of 2.4 kV at 25μF and 200Ω, as recommended by *Dower et al.*, (1988). SOC medium (1ml) was then added and the cells incubated at 37°C for 1 hour before plating dilutions onto selective medium.

SOC Medium (1L)
20g Tryptone
5g Yeast extract
0.5g NaCl
0.19g KCl
Distilled water was added to 975ml. After autoclaving, 20ml sterile 1M glucose and 5ml 2M MgCl₂ were added.
2.3 Transduction of E. coli using bacteriophage P1vir

Recipient bacteria were grown to late logarithmic phase (A_{600} \sim 1) in L-broth (5ml) supplemented with 5mM CaCl_2. Bacteria (1ml) were mixed with 100μl P1vir lysate and incubated statically at 30°C for 30 minutes. Next, 4ml L-broth containing 5mM sodium citrate was added and incubation continued at 37°C for 1 hour to allow phenotypic expression. Citrate prevents reinfection of the cells by chelating divalent calcium ions needed for adsorption of the bacteriophage. Cells were collected by centrifugation (2750 x g for 10 minutes at 4°C), resuspended in L-broth containing 5mM citrate and plated onto selective medium.

2.3.1 Preparation of bacteriophage P1vir lysates

Bacteriophage P1vir lysates were prepared by the method of Silhavy et al., (1984). Donor bacteria were grown to early logarithmic phase (A_{600} = 0.2) in L-broth (5ml) supplemented with 5mM CaCl_2 and 0.2% (w/v) glucose at 37°C. Next, 100μl of a P1vir lysate (c. 10^5 pfu/ml) was added and incubation continued at 37°C until the cells lysed (about 3 hours). Chloroform (100μl) was then added, mixed by vortexing and the lysate transferred to 1.5ml Eppendorf tubes. Cell debris was removed by centrifugation (11600 x g for 5 minutes in a microfuge) and the supernatants recovered to fresh tubes. P1vir lysates were stored at 4°C and contained 10^5 to 10^8 pfu/ml.

2.4 Conjugation of E. coli

For conjugal transfer of plasmids, donor and recipient strains were grown separately to early logarithmic phase in L-broth (5ml). Bacteria from 1ml of recipient culture were transferred to a 1.5ml Eppendorf tube and collected by centrifugation at 11600 x g for 5 minutes in a microfuge. To the pellet of recipient cells, 1ml of donor culture was added and centrifugation repeated. The combined pellet was resuspended in 50μl L-broth and the suspension spread over a 2cm diameter 0.2μm nitrocellulose filter (Millipore) placed on the surface of an L-agar plate. Following incubation at 37°C for 5 hours, cells were washed from the filters using 1ml PBS in sterile Petri dishes and dilutions plated onto selective medium. To prevent the carry-over of nutrients when using minimal media for selection of transconjugants, cells were washed twice with PBS.
2.5 Transposon mutagenesis of the F factor

To select for transfer of the F factor between bacteria, transposon Tn10, which confers resistance to tetracycline, was inserted into F. This was done by infecting *E. coli* BW96F (Table 2.1), which contains the F factor, with the bacteriophage λ840::Tn10. Strain BW96F, which requires thymine for growth, was grown to mid-logarithmic phase in 10ml L-broth containing 0.02% (w/v) thymine. The bacteria were collected by centrifugation (2750 x g for 10 minutes at 4°C), resuspended in 1ml 10mM MgSO\(_4\) and 100μl mixed with 100μl (c. 10^9 pfu) λ840::Tn10 lysate (kindly supplied by B. Wilkins, University of Leicester) in a sterile test tube. Following incubation at ambient temperature for 30 minutes to allow adsorption, the mix was plated onto L-agar containing 0.02% (w/v) thymine and tetracycline to select for cells with Tn10 insertions. After incubation at 37°C for 12 to 18 hours the bacteria were washed from the plate using 1ml PBS and 100μl used to inoculate 10ml L-broth. Transposon insertions in the F factor were then isolated by mating the pool of tetracycline resistant BW96F cells with *E. coli* LE392 (Table 2.1), with selection on L-agar containing tetracycline but lacking thymine to prevent growth of the donor.

2.6 Routine DNA manipulation

Restriction endonucleases and DNA modifying enzymes were purchased from GIBCO/Bethesda Research Laboratories or New England Biolabs and used according to the manufacturer’s instructions. Restriction endonuclease digests were performed in 20μl volumes using 0.1 to 1μg DNA and 1 to 10 units of enzyme. Ligation of DNA fragments was performed in 20μl reactions using 1 unit of bacteriophage T4 DNA ligase with incubation at 14°C for 12 to 18 hours. DNA fragments were separated by agarose gel electrophoresis using 1 kb ladder or HindIII fragments of bacteriophage λ DNA (GIBCO/BRL) as size markers and were isolated from agarose gels using a Sephaglas™ BandPrep kit (Pharmacia) as described by the manufacturer.

2.6.1 Agarose gel electrophoresis

Agarose was used at 0.7% (w/v) for resolution of DNA fragments larger than 3 kb, or 1% (w/v) for smaller fragments, in TAE buffer pH 7.7 (40mM Tris-acetate, 1mM EDTA) containing 0.5μg/ml ethidium bromide. Prior to loading, samples were mixed with 1/10 volumes of 10x loading buffer (0.25% (w/v) bromophenol blue; 0.25% (w/v) xylene cyanol; 15% (w/v) ficoll type 400, in water). Electrophoresis was performed at constant voltage (<20V/cm) in TAE buffer containing 0.5μg/ml ethidium bromide. DNA was visualised under ultraviolet light using a long wave transilluminator.
2.6.2 Phenol extraction and ethanol precipitation of DNA

An equal volume of phenol:chloroform:isoamyl alcohol (25:24:1), equilibrated to pH 7.5 with Tris-HCl, was added to the DNA and mixed to form an emulsion. Following centrifugation at 11600 x g for 5 minutes in a microfuge, the aqueous phase was recovered and mixed with an equal volume of chloroform:isoamyl alcohol (24:1) and centrifugation repeated. DNA was then precipitated by the addition of 1/10 volumes of 3M sodium acetate (pH 5.2) and 3 volumes -20°C absolute ethanol, with incubation at -20°C for at least 30 minutes. Precipitated DNA was collected by centrifugation at 11600 x g for 10 minutes in a microfuge, washed once in 70% (v/v) ethanol to remove residual salt, dried \textit{in vacuo} and resuspended in sterile distilled water.

2.6.3 Removal of protruding 3' termini

The 3' to 5' exonuclease activity of bacteriophage T4 DNA polymerase was used to remove protruding nucleotides from the 3' termini of DNA fragments to allow blunt-end ligation. Restriction endonuclease fragments (0.2 to 1μg) were incubated with T4 DNA polymerase (1 unit/μg DNA) for 30 minutes at 12°C. Reactions were in 20μl volumes containing deoxynucleotides (Pharmacia) at 0.1mM and used the buffer supplied with the enzyme. The T4 DNA polymerase was inactivated by heating to 75°C for 10 minutes and the DNA ethanol precipitated for subsequent ligation.

2.6.4 Preparation of an oligonucleotide linker

To prepare a linker, 10ng of two complementary oligonucleotides were mixed in a final volume of 10μl. The mix was heated at 72°C for 10 minutes and then allowed to cool slowly to ambient temperature to enable the oligonucleotides to anneal. The linker (10μl) was then ligated with 0.1 to 1μg of digested plasmid DNA in a final volume of 20μl. Oligonucleotides were prepared using an Applied Biosystems DNA synthesiser by Debra Langton (Protein and Nucleic Acid Chemistry Laboratory, University of Leicester) and were ethanol precipitated and washed twice with 70% ethanol (v/v) prior to use. The concentration was determined by measuring absorbance at 260nm (an absorbance of 1 at 260nm is equivalent to 33μg/ml).
2.7 Procedures for extraction of DNA and RNA

The following solutions were used for extraction of DNA,

<table>
<thead>
<tr>
<th>Solution I</th>
<th>Solution II</th>
<th>Solution III</th>
</tr>
</thead>
<tbody>
<tr>
<td>50mM glucose</td>
<td>0.2N NaOH</td>
<td>3M Potassium acetate</td>
</tr>
<tr>
<td>25mM Tris-HCl pH 8.0</td>
<td>1% (w/v) SDS</td>
<td>5M Acetic acid</td>
</tr>
<tr>
<td>10mM EDTA pH 8.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5mg/ml lysozyme</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.7.1 Small scale extraction of plasmid DNA

Rapid extraction of plasmid DNA was performed by the alkaline lysis method of Birnboim and Doly (1979). Bacteria were grown to stationary phase in L-broth and 1.5ml collected by centrifugation at 11600 x g for 2 minutes in a microfuge. Cells were resuspended in 100μl ice-cold solution I and placed on ice for 30 minutes. Next, 200μl solution II (at ambient temperature) was added, mixed by inversion and incubation continued on ice for 5 minutes. Solution III (150μl) was then added and following incubation on ice for a further 5 minutes, cell debris was removed by centrifugation at 11600 x g for 5 minutes in a microfuge. The aqueous phase was recovered, phenol-chloroform extracted and the DNA ethanol precipitated. Plasmid DNA was resuspended in 50μl sterile distilled water and stored at -20°C.

2.7.2 Large scale extraction of plasmid DNA

Bacteria from 400ml stationary phase cultures were collected by centrifugation at 4800 x g for 10 minutes at 4°C using a Sorvall GSA rotor and Sorvall RC-5B centrifuge. Cells were resuspended in 10ml ice-cold solution I, incubated on ice for 30 minutes, then 20ml solution II added. After 10 minutes on ice, 15ml solution III was added and following a further 10 minutes on ice, the suspension was transferred to 30ml polyallomer tubes and cell debris removed by centrifugation at 37000 x g for 20 minutes at 4°C using a Sorvall SS-34 rotor. The aqueous phase was recovered and the DNA precipitated by adding 0.6 volumes isopropanol at room temperature. The precipitate was collected by centrifugation at 37000 x g for 20 minutes at 20°C, dried in air and resuspended in 17ml sterile distilled water. Insoluble material was removed by centrifugation at 2750 x g for 10 minutes at 4°C then caesium chloride and ethidium bromide added to 1g/ml and 50μg/ml respectively. The solution was transferred to a 30ml ultracentrifuge tube (DuPont), overlaid with paraffin oil and the tube sealed by crimping. Plasmid DNA was separated from chromosomal DNA and nicked plasmids.
by centrifugation at 129000 x g for 20 hours at 20°C using a Sorvall TV850 rotor and Sorvall OTD-50B ultracentrifuge. The DNA was visualised under ultraviolet illumination and the lower band comprising intact plasmid DNA collected. Ethidium bromide was extracted from the plasmid solution by mixing with an equal volume of caesium chloride-saturated isopropanol. The aqueous phase was extracted twice more as above then caesium chloride removed by dialysis against distilled water for at least 12 hours. The DNA concentration was determined by measuring absorbance at 260nm (an absorbance of 1 at 260nm is equivalent to 50μg/ml DNA).

Some plasmids, including those used for automated sequencing, were prepared using a Plasmid Maxi-kit (Qiagen) according to the manufacturer’s instructions.

2.7.3 Extraction of chromosomal DNA

The method used was based on that of Saito and Miura (1963). Bacteria from 5ml stationary phase cultures were washed once with 10mM NaCl, resuspended in 5ml ice-cold solution I and placed on ice for 30 minutes. Next, SDS and EDTA were added to 1% (w/v) and 50mM respectively and the suspension mixed by inversion and incubated at ambient temperature until clear. An equal volume of phenol:chloroform was added and gently mixed by inversion for 5 minutes. Following centrifugation at 3000 x g for 20 minutes at 4°C the aqueous phase was recovered and extracted twice more with phenol:chloroform. Sodium acetate pH 5.2 was then added to 0.3M and the mix poured into 20ml -20°C absolute ethanol in a conical flask. The DNA was recovered using spools made from glass Pasteur pipettes, resuspended in 1ml distilled water and stored at -20°C.

2.7.4 Extraction of bacteriophage DNA

Single stranded DNA from recombinant M13mp19 bacteriophages was isolated as described by Sambrook et al., (1989). Following transformation (section 2.2.3), bacteriophage from white plaques was transferred to 5ml L-broth containing 100μl stationary phase E. coli JM101 cells using a toothpick. After incubation at 37°C for 5 hours with shaking, 1.5ml was transferred to an Eppendorf tube and the bacteria removed by centrifugation at 11600 x g for 5 minutes. Double stranded (replicative form) bacteriophage DNA was extracted from the cell pellet as described in section 2.7.1 and digested with restriction endonucleases to verify the insertion. The supernatant (800μl) was recovered to a fresh tube, mixed with 200μl of a solution containing 2.5M NaCl and 20% (w/v) PEG 6000 and incubated for 30 minutes at ambient temperature to precipitate bacteriophage particles. The precipitate was collected by centrifugation at 11600 x g for 5 minutes and all traces of the supernatant
removed before resuspending in 100μl 1.1M sodium acetate pH 7.0. Single stranded DNA was then purified by phenol:chloroform extraction, precipitated with ethanol and resuspended in 15μl sterile distilled water.

2.7.5 Extraction of RNA

Total RNA was isolated from bacteria by hot acid phenol extraction (Arba et al., 1981). Bacteria were grown to mid-logarithmic phase in L-broth (10ml), collected by centrifugation (2750 x g for 10 minutes at 4°C) and resuspended in 0.5ml phenol equilibrated with sodium acetate pH 5.2 (at 60°C). The suspension was transferred to a 1.5ml Eppendorf tube and 0.8ml 60°C lysis buffer (20mM sodium acetate pH 5.2, 0.5% (w/v) SDS, 1mM EDTA) added. Samples were incubated at 60°C for 5 minutes, mixed periodically by inversion, then placed on ice for 10 minutes. Following centrifugation at 11600 x g for 5 minutes in a microfuge the aqueous phase was recovered to a fresh tube and the extraction repeated with a further 0.5ml 60°C phenol. The aqueous phase was then extracted with chloroform and the RNA precipitated with ethanol at -20°C for at least 1 hour. The RNA was collected by centrifugation at 11600 x g for 10 minutes in a microfuge, dried in vacuo and resuspended in 20μl sterile distilled water. The RNA concentration was determined by measuring absorbance at 260nm (an absorbance of 1 at 260nm is equivalent to 40μg/ml RNA). All solutions used for extraction and dilution of RNA were treated with diethyl pyrocarbonate (DEPC) to inactivate ribonucleases. DEPC was added at 0.05% (v/v) and the solution incubated at 37°C for 12 to 18 hours prior to autoclaving.

2.8 Procedures for DNA and RNA hybridisation

2.8.1 Southern blotting

DNA fragments were transferred from agarose gels to nylon membranes by the method of Southern (1975). After separation of restriction endonuclease fragments by agarose gel electrophoresis, the gel was photographed alongside a ruler with minimal exposure to ultraviolet light. The gel was soaked in 0.25M HCl for 7 minutes to depurinate the DNA then rinsed with distilled water. Next, denaturing solution (0.5N NaOH, 1.5M NaCl) was added and the gel submerged for 30 minutes with gentle rocking. The gel was rinsed again with distilled water and then submerged in neutralising solution (0.5M Tris-HCl pH 7.5, 3M NaCl) for 30 minutes. After rinsing with distilled water, the gel was placed on a stack of 10 sheets of Whatman 3mm paper pre-soaked with 20x SSC (3M NaCl; 0.3M tri-sodium citrate, pH 7.0). A sheet of nylon membrane (GeneScreen, DuPont) cut to gel-size and soaked in 3x SSC, was placed on
the gel and any air bubbles removed. Three sheets of Whatman 3mm soaked in 3x SSC were placed on the nylon membrane, followed by six sheets of dry Whatman 3mm paper and a stack of paper towels. A 500g weight was placed on the assembly and transfer allowed to proceed for 12 to 18 hours. During this time the paper towels above the gel were replaced regularly and the lower sheets kept soaked with 20x SSC. After transfer, the membrane was rinsed in 2x SSC and exposed to ultraviolet light for 4 minutes on a long wave transilluminator to fix DNA to the filter. Fixed membranes were stored in the dark at ambient temperature until needed for hybridisation.

2.8.2 Transfer of RNA to nylon membranes

A Bio-Rad Biodot-SF apparatus was used to apply RNA samples onto nylon membranes over a uniform area. The apparatus was first soaked for one hour in 0.5M sulphuric acid to inactivate ribonucleases and then rinsed with sterile distilled water. A sheet of nylon membrane and three Whatman 3mm filter papers cut to the same size were soaked in 10x SSC and placed in the apparatus. RNA extracts were diluted to 30µl with sterile distilled water and 90µl loading buffer (50% (v/v) formamide, 7% (v/v) formaldehyde, 1x SSC) added. The samples were heated at 65°C for 10 minutes then 120µl ice-cold 20x SSC added. Samples were applied to the nylon membrane under vacuum and the slots used rinsed with 100µl 10x SSC. The membrane was then exposed to ultraviolet light for 4 minutes on a long wave transilluminator to fix RNA to the filter. The blotting apparatus and all solutions were treated with DEPC and autoclaved prior to use. To confirm that DNA probes hybridise to RNA and not contaminating DNA, RNA extracts were treated with 20 units RNase-ITM RNase cocktail (Stratagene), at 37°C for 1 hour before applying to the membrane.

2.8.3 Colony hybridisation

For rapid identification of bacteria containing cloned DNA fragments, colony hybridisation was used (Grunstein and Hogness, 1975). Bacterial colonies were transferred to nylon membranes (Colony/Plaque Screen™, DuPont) by overlaying filters on the agar surface. Once the orientation of the membrane on the plate had been recorded, each filter was placed on a sheet of Whatman 3mm paper soaked in denaturing solution (0.5N NaOH, 1.5M NaCl) for 5 minutes. Filters were then placed on Whatman 3mm paper soaked with neutralising solution (0.5M Tris-HCl pH 7.5, 1.5M NaCl) for 5 minutes and then dried in air. Cell debris was removed by gentle scrubbing using polyallomer wool moistened with 6x SSC and the membranes dried in air prior to hybridisation.
2.8.4 Preparation of a radiolabelled DNA probe

Restriction endonuclease fragments or PCR products for use as probes were radiolabelled using random hexanucleotide primers as described by Feinberg and Vogelstein (1983). DNA fragments (10 to 100ng in 10μl sterile distilled water) were denatured by heating in a boiling water bath for 5 minutes and then placed on ice. Next, 3μl of 5x labelling buffer, 4 units of the Klenow fragment of DNA polymerase I and 10μCi [α-32P] dCTP (10mCi/ml, Amersham) were added and the reaction incubated at ambient temperature for 5 hours. Prior to use, labelled probe was denatured by heating as above.

To remove nucleotides not incorporated into the probe and to check that labelling was successful, some probes were purified by column chromatography. After the labelling reaction, the probe solution was loaded onto a column of fine Sephadex G-50 (Pharmacia), equilibrated with TE buffer pH 8.0 (10mM Tris-HCl pH 8.0, 1mM EDTA pH 8.0) in a glass Pasteur pipette plugged with polyallomer wool. The probe was eluted from the column with TE buffer containing 0.75M NaCl and 20 fractions of approximately 200μl collected in 0.5ml Eppendorf tubes. The radioactivity of each fraction was determined using a Minaxi 4000 series scintillation counter (Packard). The radiolabelled probe is eluted before unincorporated nucleotides and fractions of the highest activity were pooled and denatured prior to use.

5x Labelling buffer
Solutions A, B and C were prepared using sterile distilled water and mixed in a ratio of 10:25:15 respectively. 5x labelling buffer was stored at -20°C.

Solution A  1ml 1.25M Tris-HCl pH 8.0, 0.125M MgCl₂
            18μl β-mercaptoethanol
            5μl each of 100μM dATP, dTTP and dGTP

Solution B  2M HEPES (N-[2-hydroxyethyl] piperazine-N'-[ethanesulfonic acid]) pH 6.6

Solution C  90 absorbance units/ml random hexanucleotides (Pharmacia).
2.8.5 Hybridisation conditions

Fixed membranes were incubated in 25ml prehybridisation solution at 60°C for 3 hours with rotation. This solution was then replaced with 25ml hybridisation solution (as prehybridisation solution but with 6x SSC and lacking denatured heterologous DNA) and the denatured probe added. Incubation was continued at 60°C for 12 to 24 hours. For some hybridisations to detect RNA, the buffers also contained 10% (w/v) dextran sulphate and 50% (v/v) deionised formamide and in these cases incubation was at 42°C.

Prehybridisation solution
5x SSC
0.1% (w/v) SDS
5x Denhardt's solution
200μg/ml Denatured herring sperm DNA
A 100x Denhardt's solution comprises 2% (w/v) each of ficoll type 400, bovine serum albumin and polyvinylpyrrolidone in distilled water. Herring sperm DNA was prepared at 10mg/ml in distilled water and the DNA sheared to give small fragments by repeatedly heating and forcing it through a narrow gauge needle.

After hybridisation, the membranes were washed twice in a solution containing 2x SSC and 0.1% (w/v) SDS then twice in 0.1x SSC, 0.1% (w/v) SDS at 60°C for 15 minutes each wash. Membranes were wrapped in Saran wrap (DuPont) and exposed to Kodak X-Omat film at -70°C in cassettes containing intensifying screens. Autoradiographs were developed using an Agfa-Geveart automatic film processor. Quantitative measurements of the relative amounts of RNA detected by hybridisation with DNA probes were made using a Molecular Dynamics model 425E PhosphorImager with ImageQuant v. 5.6 software.

2.8.6 Removal of bound probe from nylon membranes

Following autoradiography, bound radiolabelled probe was removed from some membranes to allow hybridisation with other probes. This was possible only if membranes were kept moist at all times after hybridisation. Membranes were placed in boiling 0.1% (w/v) SDS and the solution allowed to cool to ambient temperature. Removal of the probe was monitored with a Geiger counter and by autoradiography.
2.9 DNA Sequencing

2.9.1 Manual DNA sequencing

Manual DNA sequencing was performed using chain termination inhibitors as described by Sanger et al., (1977). The DNA to be sequenced was first cloned into bacteriophage M13mp19 and a single stranded template generated (section 2.7.4). Sequencing reactions were performed using approximately 1pmol template DNA and 0.5pmol oligonucleotide primer with a Sequenase™ version 2.0 T7 DNA polymerase sequencing kit (United States Biochemical Corporation) according to the manufacturer's instructions. Reaction products were radiolabelled by incorporation of [α-35S] thio-dATP and separated by gradient gel electrophoresis.

2.9.2 Gradient gel electrophoresis

Buffer gradient gels for the separation of products from DNA sequencing reactions were prepared as described by Biggin et al., (1983) using the following solutions;

Solution I (1L)
430g Urea
50ml 10x TBE
150ml 40% (w/v) Acrylamide (19:1 acrylamide:bis-acrylamide) (Accugel, National Diagnostics)
Distilled water to 1L

Solution II (1L)
430g Urea
150ml 10x TBE
150ml 40% (w/v) Acrylamide
50g Sucrose
50mg Bromophenol blue

Solution A
65ml Solution I
300μl 10% (w/v) Ammonium persulfate (Electrophoresis grade, Bio-Rad)
70μl TEMED (N, N', N''-tetramethylethylenediamine, Sigma)
**Solution B**

12ml Solution II
60µl 10% (w/v) Ammonium persulfate
20µl TEMED

10x TBE (1L)
108g Tris-HCl
55g boric acid
40ml 0.5M EDTA pH 8.0
Distilled water to 1L.

Gels were cast using 20cm x 50cm glass plates, cleaned thoroughly before use, and 0.4mm spacers. To one plate, dimethyldichlorosilane solution (2% (w/v) in 1,1,1-trichloroethane) was applied to ensure the gel adhered to just one plate after electrophoresis. Immediately following the addition of TEMED to solutions A and B, all of solution B then 12ml solution A were drawn up into a 25ml pipette. A few air bubbles were drawn into the pipette to form a crude gradient and the solution run between the plates. The remaining space was filled with solution A and a comb inserted to give the top of the gel a straight edge. Gels were allowed to polymerise for 12 to 18 hours before use.

A vertical electrophoresis system was used, with 0.5x TBE in the upper tank and 1x TBE in the lower tank. To ensure even heat distribution, 5mm aluminium sheets cut to gel size, were clamped either side of the gel. The gel was then pre-run for 30 minutes at a constant power of 40W. Samples were denatured by heating at 90°C for 5 minutes prior to loading and electrophoresis performed at 40W for 3 to 9 hours.

After electrophoresis the siliconised plate was removed and the gel soaked in fixing solution (10% (v/v) methanol, 10% (v/v) acetic acid) for 15 minutes. The gel was rinsed briefly with distilled water then transferred to a sheet of Whatman 3mm paper and dried under vacuum at 80°C for 45 minutes. Gels were exposed to Kodak X-Omat film for 24-48 hours at ambient temperature.

2.9.3 Automated DNA sequencing

Automated DNA sequencing was performed using an Applied Biosystems PRISM™ Ready Reaction DyeDeoxy™ Cycle Sequencing kit. As with the method of Sanger *et al.*, (1977), the DNA sequence is derived using chain termination inhibitors except the dideoxynucleotides used are labelled with different dyes. The dye-labelled products of chain termination are then separated by electrophoresis and detected using an Applied Biosystems 373A DNA sequencing system.
Reactions were performed using 500ng double stranded plasmid DNA and 3pmoles of an oligonucleotide primer. A mix of dye-labelled deoxynucleotides as well as dATP, dTTP, dCTP and dITP was added to each reaction together with AmpliTaq™ DNA polymerase (Applied Biosystems) to a final volume of 20μl. Reactions were overlaid with mineral oil (Sigma) and placed in a Perkin Elmer Cetus 480 thermal cycler preheated to 95°C. After 30 seconds, 30 cycles of denaturation (95°C for 30 seconds), annealing (50°C for 15 seconds) and extension (60°C for 4 minutes) at a ramp rate of 1°C/s, were performed.

The reaction products were purified by extraction three times with phenol:water:chloroform (68:18:14, Applied Biosystems), then precipitated by the addition of 1/10 volumes 2M sodium acetate pH 4.5 and 3 volumes absolute ethanol for at least 1 hour at -20°C. The precipitate was then collected by centrifugation at 11600 x g for 10 minutes in a microfuge, washed once with 70% (v/v) ethanol and dried in vacuo. The reaction products were analysed by Dr Katherine Lilly (Protein and Nucleic Acid Chemistry Laboratory, University of Leicester).

Nucleotide sequences were analysed using the SeqEd program (Version 1.0.3, Applied Biosystems) and the University of Wisconsin Genetics Computer Group programs on the VAX VMS cluster (Devereux et al., 1984).

2.10 Polymerase chain reaction

Polymerase chain reactions (PCR) were performed in 100μl volumes, using 100ng template DNA, 0.25μM oligonucleotide primers and each deoxynucleotide at 200μM. Reactions also contained 10mM Tris-HCl pH 8.8, 50mM KCl, 1.5mM MgCl₂ and 0.1% (v/v) Triton X-100. Prior to addition of the template DNA, reactions were irradiated for 15 minutes with ultraviolet light to destroy contaminating DNA. Reactions were overlaid with 50μl mineral oil then heated at 92°C for 5 minutes in a Perkin Elmer Cetus 480 thermal cycler. Biotaq™ DNA polymerase (Bioline) was then added (one unit per 100μl reaction) and 30 cycles of amplification performed. Templates were denatured at 92°C for 2 minutes, primers annealed at c. 50°C and extension performed at 72°C (allowing 1 minute for each kilobase of DNA to be synthesised), with a final incubation at 72°C for 5 minutes.

2.10.1 Colony PCR

For rapid screening of bacteria for specific sequences, colony PCR was performed (Clackson et al., 1991). A reaction mix containing primers, dNTP's, buffer and Taq polymerase was prepared and 25μl dispensed into 0.5ml Eppendorf tubes. To each tube
a small amount of bacteria, from stationary phase cultures or a colony, was then added using a sterile toothpick. Reactions were overlaid with mineral oil and heated to 95°C for 5 minutes to lyse the cells then 30 cycles of PCR performed.

2.10.2 Reverse transcriptase-PCR

Reverse transcriptase-PCR (RT-PCR) was used to demonstrate the transcription of specific sequences. Total RNA (0.2 to 5µg) was mixed with an oligonucleotide which binds at the 3' end of the RNA at a final concentration of 1.25µM in 50µl 1x PCR buffer (50mM KCl, 10mM Tris-HCl pH 8.3, 1.5 mM MgCl₂ and 200µM deoxynucleotides). The mix was then incubated at 90°C for 2 minutes then rapidly cooled on ice to anneal the primer to the RNA. Next, 10 units of avian myeloblastosis virus (AMV) reverse transcriptase (Promega) were added together with 10 units of human placental ribonuclease inhibitor (Amersham) in a total of 10µl 1x PCR buffer. Reactions were then incubated at 42°C for 1 hour to allow first strand cDNA synthesis. The reverse transcriptase was then denatured by heating at 90°C for 2 minutes and an oligonucleotide which anneals at the 3' end of the cDNA added to a final concentration of 1.25µM in 20µl 1x PCR buffer. Taq polymerase (one unit, Bioline) was then added, reactions overlaid with mineral oil and the cDNA amplified by 30 to 40 cycles of PCR as described in section 2.10. All solutions for RT-PCR were treated with DEPC to inactivate ribonucleases.

2.11 RNase protection assay

To locate sites of transcription initiation an RNase protection assay was performed (Sambrook et al., 1989). In this method, a radiolabelled antisense RNA is hybridised to the transcript of interest and the hybrid then digested with a ribonuclease which only cleaves single stranded RNA. The size of the protected RNA hybrid is determined by polyacrylamide gel electrophoresis and reflects the extent of the transcript which is recognised by the probe.

2.11.1 Preparation of a radiolabelled antisense RNA

Radiolabelled antisense RNA probes were synthesised by in vitro transcription with T7 polymerase using Promega's Riboprobe™ system. A DNA fragment containing the promoter of interest was first cloned into plasmid pTZ18R (Table 2.2) such that transcription from the T7 promoter generated an RNA of the opposite sense to the transcript to be mapped. The plasmid was then digested with an enzyme which leaves either a 5' overhang or blunt end and linear DNA isolated. A 20µl reaction
containing 0.5μg template DNA, 125μM rNTP’s (rUTP at 12μM) and 12.5μCi [α-32P] rUTP (Amersham) was prepared using the buffer supplied with the kit. T7 polymerase (20 units) and 20 units of an RNase inhibitor (RNasin™, Promega) were added and the reaction incubated at 37°C for 1 hour. The DNA template was then removed by treatment with 10 units RNase-free DNase (Pharmacia) for 30 minutes at 37°C to prevent it hybridising with the probe. Next, the run-off transcripts were extracted with acid phenol and chloroform, ethanol precipitated and purified by polyacrylamide gel electrophoresis. Ideally, RNA probes should be 300-500 bp long and have a specific activity of 1-3 x 10^8 cpm/μg.

The precipitated probe was resuspended with 5μl nuclease free water and 5μl gel loading buffer (80% (v/v) deionised formamide, 1mM EDTA, 0.1% (w/v) SDS, 0.1% (w/v) bromophenol blue and 0.1% (w/v) xylene cyanol). The probe was then denatured by incubation at 85°C for 5 minutes and separated on a denaturing polyacrylamide gel (section 2.9.2). After electrophoresis for 3 hours the gel was covered with plastic wrap and exposed to X-ray film for 2 minutes to locate the full length probe. The probe was excised from the gel using a sterile razor blade and eluted by addition of 300μl gel elution buffer (0.5M ammonium acetate, 1mM EDTA, 0.2% (w/v) SDS) with incubation at 37°C for 12-18 hours. Following centrifugation at 11600 x g in a microfuge for 5 minutes the supernatant was recovered and stored at -70°C until needed.

2.11.2 Hybridisation of the RNA probe to target RNA and RNase digestion

The RNA probe (1-5 x 10^6 cpm at a specific activity of 1-3 x 10^8 cpm/μg) was mixed with 100μg total RNA from a strain containing the transcript of interest and precipitated with ethanol. The precipitate was washed with 70% (v/v) ethanol, dried in air then resuspended in 30μl hybridisation buffer (80% (v/v) deionised formamide, 1mM EDTA, 40mM PIPES pH 6.4, 0.2M sodium acetate). After incubation at 85°C for 5 minutes to denature the RNA, the mix was incubated at 45°C for 12 to 18 hours to allow the RNA probe to hybridise with the target RNA. Next, 270μl RNase digestion buffer (10mM Tris-HCl pH 7.5, 5mM EDTA, 20mM sodium acetate) was added and single stranded RNA removed by digestion with 50 units RNase ONE™ (Promega) at ambient temperature for 1 hour. The RNase was inactivated by adding 30μl stop solution (10% (w/v) SDS, 1mg/ml yeast tRNA) and the protected fragments precipitated with ethanol. The precipitate was washed with 70% (v/v) ethanol, dried in air and resuspended in 10μl gel loading buffer. Samples were denatured at 85°C for 5 minutes and then analysed on denaturing polyacrylamide gels alongside a DNA sequence ladder to enable the size of the protected fragments to be estimated.
2.12 Primer extension analysis

To precisely map sites of transcription initiation an AMV reverse transcriptase primer extension kit (Promega) was used. In this method, reverse transcriptase is used to extend an end-labelled oligonucleotide hybridised to the transcript of interest. The size of the cDNA reflects the distance between the primer and the 5' end of the transcript and is determined by polyacrylamide gel electrophoresis.

The end-labelled primer was generated by phosphorylation using \([\gamma-^{32}P]\) ATP and T4 polynucleotide kinase. The primer (10pmol) was mixed with 3μCi \([\gamma-^{32}P]\) ATP (Amersham) and 10 units T4 polynucleotide kinase in a 10μl reaction containing 50mM Tris-HCl pH 7.5, 10mM MgCl\(_2\), 5mM dithiothreitol and 0.1mM spermidine.

The reaction was incubated at 37°C for 10 minutes then heated at 90°C for 2 minutes to inactivate the enzyme. The primer concentration was then adjusted to 200fmol/μl by adding 40μl nuclease free water.

Primer extension reactions were performed in a final volume of 20μl using the buffer supplied with the kit and contained 200fmol end-labelled primer and 50μg total RNA. The primer and RNA were annealed by heating the mix at 58°C for 20 minutes then allowing it to cool at ambient temperature for 10 minutes. Next, sodium pyrophosphate (pre-warmed to 37°C) was added to 2.8mM along with one unit of AMV reverse transcriptase. The reaction was incubated at 42°C for one hour then 20μl gel loading buffer (98% (v/v) formamide, 10mM EDTA, 0.1% (w/v) bromophenol blue, 0.1% (w/v) xylene cyanol) added. The extension products were heated at 90°C for 10 minutes before analysing 5μl on a denaturing polyacrylamide gel alongside a DNA sequence ladder obtained with the same primer.

2.13 Assays of capsular polysaccharide production

2.13.1 Assay of sensitivity to capsule-specific bacteriophage

Bacterial strains under test were grown to mid-logarithmic phase in L-broth (10ml). The cells were collected by centrifugation (2750 x g at 4°C for 10 minutes), resuspended in 1ml 10mM MgSO\(_4\) and 100μl mixed with 100μl of dilutions of the bacteriophage in a sterile test tube. Following incubation at ambient temperature for 15 minutes to allow adsorption, soft top agar at 42°C (3ml) was added and the mix poured onto an L-agar plate supplemented with antibiotics as appropriate. Once set, the plates were incubated at 37°C or 18°C. Plaques were normally visible after 5 or 24 hours incubation respectively.
Soft top agar (1L)
10g Tryptone
8g NaCl
8g Agar
Distilled water to 1L

2.13.2 Preparation of capsule-specific bacteriophage lysates

Bacteriophage from several plaques was transferred to a tube containing 100µl of stationary phase host cells using sterile toothpicks. Following incubation at ambient temperature for 10 minutes to allow adsorption, 2ml L-broth containing 10mM MgSO₄ was added and incubation continued for 5 hours at 37°C with shaking. The lysate was divided into two 1.5ml Eppendorf tubes then 100µl chloroform added and mixed by vortexing. Cell debris was removed by centrifugation (5 minutes at 11600 x g in a microfuge) and supernatants recovered to fresh tubes. Bacteriophage lysates were stored at 4°C and contained 10⁸ to 10¹⁰ pfu/ml. Dilutions were prepared using phage dilution buffer (10mM Tris-HCl pH 7.4, 10mM MgSO₄, 0.01% (w/v) gelatin).

2.13.3 Detection of the K5 polysaccharide by immunoelectron microscopy

I am grateful to Peter Hänfling (Max-Planck-Institut für Immunbiologie, Freiburg, Germany) for analysis of K5 antigen expression by immunoelectron microscopy. Sample embedding, preparation and labelling were as described by Kröncke et al., (1990a). A mouse monoclonal K5 specific antibody (IgM) and gold conjugated anti-mouse IgM were used.

2.13.4 In vitro membrane transferase assay

Biosynthesis of the K5 antigen (4)-β-glucuronic acid-(1,4)-α-N-acetylglucosamine-(1-) in isolated membranes, was assayed by incorporation of UDP-[¹⁴C] glucuronic acid as described by Finke et al., (1991). Membranes were prepared from bacteria grown to late logarithmic phase in L-broth (75ml). Bacteria were collected by centrifugation at 4800 x g for 10 minutes at 4°C, resuspended in 75ml ice-cold T buffer pH 8.0 (50mM Tris-HCl, 30mM magnesium acetate, 2mM dithiothreitol) and centrifuged again. The cell pellet was resuspended in 3ml ice-cold T buffer and the bacteria disrupted by sonications eight times on ice for 15 seconds at 30 second intervals using a Braun Labsonic 2000 sonicator. The homogenate was centrifuged at 6000 x g for 10 minutes at 4°C to remove large fragments of cell debris and the supernatant transferred to a 2.5ml polycellomer ultracentrifuge tube (Beckman). Membranes were
collected at 166000 x g using a Beckman Optima™ TL-100 ultracentrifuge and TLS-55 rotor at 4°C for 15 minutes, resuspended in 0.5ml T buffer and stored at -70°C. Protein concentrations of membrane preparations were determined as described in section 2.14.3 and typically were 3 to 5 mg/ml.

Assays of K5 polysaccharide biosynthesis were performed in 100μl reactions containing membranes (100 to 200μg protein/ml), 50μM UDP-[14C] glucuronic acid (DuPont NEN) and 50μM UDP-N-acetylglucosamine (Sigma) in T buffer. Reactions were incubated at 37°C for 30 minutes then stopped by the addition of 0.5ml ice-cold 12% (v/v) glacial acetic acid and placed on ice. The precipitate was collected by filtration through a 2.5cm diameter 0.45μm cellulose acetate filter (Sartorius) under vacuum. Filters were washed three times with 1ml ice-cold 12% (v/v) glacial acetic acid, placed in 10ml scintillation fluid (Optiphase Safe™, LKB Scintillation Products) in 25ml plastic vials and radioactivity on the filters measured with a scintillation counter. The radioactivity of a reaction containing no membranes was also recorded (maximum counts) and incorporation of UDP-[14C] glucuronic acid into K5 polysaccharide derived as below.

\[
\text{μMol UDP-[14C] glucuronic acid} = \frac{5 \times \text{Sample cpm}}{\text{Max. cpm} \times 0.05 \times \text{Protein conc. (mg/ml)}}
\]

Assays were performed in duplicate using membranes from three separate cultures.

2.14 Assays for reporter gene products and protein

2.14.1 Luciferase assay

For detection of bioluminescent colonies, a few drops of n-decanal were placed in the lid of the Petri dish and after 2 to 3 minutes the plates were exposed to Kodak X-Omat film. Quantitative measurements were made using a Luminoskan luminometer (Labsystems) at 30°C. Strains under test were grown to mid-logarithmic phase at 18°C or 37°C in L-broth and eight 100μl aliquots dispensed into black 96 well plates (Labsystems) and placed in the luminometer. To each sample, 10μl of a 1% (v/v) solution of n-decanal (in absolute ethanol) was added, mixed for 1 second and the relative light units (rlu) produced measured over 5 seconds. The protein concentration of the cultures used was determined (section 2.14.3) and luciferase activity expressed as mean rlu (per second) per mg total cell protein. Measurements were performed in duplicate and repeated at least three times.
2.14.2 \(\beta\)-galactosidase assay

\(\beta\)-galactosidase was assayed as described by Miller (1972). Bacterial strains under test were grown to mid-logarithmic phase in L-broth and absorbance of the culture at 600nm recorded. To 100\(\mu\)l cells, 900\(\mu\)l Z buffer, 40\(\mu\)l chloroform and 20\(\mu\)l 0.1% (w/v) SDS were added and mixed by vortexing for 10 seconds. Reactions were equilibrated to 37°C for 10 minutes then 200\(\mu\)l 4mg/ml O-nitrophenyl-\(\beta\)-D-galactopyranoside (in Z buffer at 37°C) was added and incubation continued at 37°C. The time taken for a faint yellow colour to develop was recorded and reactions then stopped by the addition of 500\(\mu\)l 1M sodium carbonate. Absorbance of the reactions was measured at 420nm and 550nm and units of \(\beta\)-galactosidase activity derived from the equation below. Assays were performed in duplicate using at least three separate cultures.

Units of \(\beta\)-galactosidase activity = \(1000 \times \frac{A_{420} - (1.75 \times A_{550})}{t \times v \times A_{600}}\)

\((t = \text{time of the reaction in minutes}; v = \text{volume of culture used in the assay in ml})\)

Z Buffer (1L)
10.7g Na\(_2\)HPO\(_4\)
6.2g NaH\(_2\)PO\(_4\)
0.7g KCl
0.25g MgSO\(_4\)
3.47ml 2-\(\beta\)-mercaptoethanol
Sterile distilled water to 1L

2.14.3 Protein assay

Protein concentration was derived using a Bio-Rad protein assay kit according to the manufacturer's instructions. A range of dilutions of the sample were assayed in duplicate using bovine serum albumin as standard. For determination of total cell protein, bacteria were collected by centrifugation, washed once with PBS then disrupted by sonication four times for 15 seconds at 30 second intervals on ice.

2.15 SDS-polyacrylamide gel electrophoresis

Protein separations by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) were as described by Laemmli (1970). The following solutions were used,
Resolving gel (12% (v/v) acrylamide)
3.35ml Distilled water
2.5ml 1.5M Tris-HCl (pH 8.8)
0.1ml 10% (w/v) SDS
4.0ml Acrylamide:Bisacrylamide (19:1) 30% (v/v) (Protogel, National Diagnostics)
50µl 10% (w/v) Ammonium persulfate
5µl TEMED

Stacking gel (4% (v/v) acrylamide)
6.1ml Distilled water
2.5ml 0.5M Tris-HCl (pH 6.8)
0.1ml 10% (w/v) SDS
1.3ml Acrylamide:Bisacrylamide (19:1) 30% (v/v)
50µl 10% (w/v) Ammonium persulfate
10µl TEMED

2x Sample loading buffer
50mM Tris-HCl (pH 6.8)
4% (w/v) SDS
1% (v/v) 2-β-mercaptoethanol
20% (w/v) Sucrose
0.001% (w/v) bromophenol blue

A Bio-Rad mini-PROTEAN II apparatus was used to cast and run gels. The
resolving gel was poured immediately following the addition of TEMED and overlaid
with n-butanol to ensure a straight top edge. After 30 minutes the n-butanol was
rinsed away with distilled water and the cavity dried with blotting paper before adding
the stacking gel solution. A comb was inserted and the gel allowed to polymerise for at
least 1 hour.

For electrophoresis, the gel was submerged in running buffer (25mM Tris-HCl,
192mM glycine, 0.1% (w/v) SDS). Samples were mixed with an equal volume of 2x
sample loading buffer and placed in a boiling water bath for 10 minutes prior to
loading. Broad range prestained protein standards (Bio-Rad) were used as molecular
weight markers. Gels were run at a constant voltage of 150V for approximately 1 hour.

Total cell protein for SDS-PAGE was prepared from 10ml late logarithmic phase
cultures. The bacteria were chilled on ice and an aliquot removed to determine the
protein concentration. The bacteria were then collected by centrifugation, adjusted to
the same concentration with sample loading buffer and boiled before loading. Proteins
were detected by staining for at least 30 minutes with 0.5% (w/v) Coomassie blue R-
250 (Bio-Rad) in 40% (v/v) methanol and 10% (v/v) glacial acetic acid. Stain not associated with protein was removed by placing the gel in 40% (v/v) methanol, 10% (v/v) glacial acetic acid for at least 2 hours.

2.16 Western blotting

Proteins were transferred to nylon membranes following SDS-polyacrylamide gel electrophoresis by the method of Towbin et al., (1979). The SDS-PAGE gel and a sheet of nylon membrane (Immobilon™ PVDF, Millipore) were equilibrated for 15 minutes in transfer buffer (25mM Tris-HCl, 192mM glycine, 20% (v/v) methanol). The gel was placed on a sheet of Whatman 3mm paper, pre-soaked with transfer buffer, then the nylon membrane placed on the gel and overlaid with a second sheet of Whatman 3mm paper. The assembly was submerged in transfer buffer in a mini-Transblot apparatus (Bio-Rad) and proteins transferred at 250mA for 2 hours at 4°C.

After transfer, the nylon membrane was blocked in 3% (w/v) skimmed milk in TN buffer pH 7.4 (10mM Tris-HCl, 0.15M NaCl) for 2 hours at ambient temperature. Primary antibody (rabbit) was then added in 3% (w/v) skimmed milk in TN buffer pH 7.4 and incubation continued for 2 hours with gentle rocking. The membrane was washed four times with TN buffer pH 7.4 for 5 minutes each wash before adding a 1/1000 dilution of secondary antibody (donkey anti-rabbit IgG-horseradish peroxidase conjugate, Amersham). After incubation at ambient temperature for 2 hours with rocking, the membrane was washed as before with TN buffer pH 7.4, then placed in developing solution. The membrane was then rinsed with distilled water, dried on Whatman 3mm paper and photographed promptly.

*Developing solution (30ml)*

15mg 4-Chloro-1-naphthol in 5ml methanol
15µl 6% (v/v) Hydrogen peroxide

TN buffer pH 7.4 to 30ml
Chapter 3

Involvement of known regulatory genes in expression of the E. coli K5 antigen

3.1 Introduction

For several reasons, the role of known regulatory genes in expression of the E. coli K5 antigen is best studied in a strain with a single copy of the capsule genes in their normal chromosomal location rather than on multicopy plasmids. First, the superhelicity of chromosomal and plasmid DNA is different (Drlica, 1992), and this may affect the action of regulators which act by altering DNA topology (section 1.7.1.1). In addition, DNA supercoiling regulates the expression of some genes in response to temperature (section 1.7.4), so plasmids carrying the capsule genes may not exhibit the same thermoregulation as the chromosomal genes. Secondly, the presence of multiple copies of the capsule genes may titrate out the activity of regulatory proteins which may be present in small amounts. Wild-type E. coli expressing the K5 antigen are not amenable to genetic analysis, being poorly transformable and difficult to transduce with bacteriophage P1. It was therefore decided to transfer the K5 capsule genes onto the chromosome of a laboratory strain of E. coli by Hfr mating. This would generate a K5/K-12 hybrid strain into which mutations in known regulatory genes could easily be introduced.

In a population of cells containing the F factor, F may integrate into the chromosome at low frequency forming an Hfr strain. Integrated F retains its ability to transfer and may therefore mobilise chromosomal genes to another strain (reviewed in Reimmann and Haas, 1993). Hfr is an acronym for high frequency of recombinants and refers the ability of cells containing integrated F to transfer chromosomal genes at about 1000 fold higher frequency than F+ cells. The integration of F into the E. coli chromosome occurs by homologous recombination via insertion sequences such as IS2, IS3 or Tn1000 (Reimmann and Haas, 1993). In the transfer process, the chromosome of the donor is replicated from the site of insertion of F and the newly synthesised strand is led into the recipient starting from the 5' end. On entry into the recipient, the donor DNA may replace the endogenous sequences by homologous recombination. Selection for transfer of the K5 capsule genes was provided by the close linkage of kps to the serA gene, which is required for the biosynthesis of serine (Vimr, 1991).
3.2 Results

3.2.1 Construction of an *E. coli* K5/K-12 hybrid

To select for transfer of the F factor between strains, transposon Tn10, which confers resistance to tetracycline, was inserted into F. This was achieved by infecting *E. coli* strain BW96F with bacteriophage λ840::Tn10 and selecting for tetracycline resistant colonies (section 2.5). Tn10 insertions in F which did not impair transfer were isolated by pooling the tetracycline resistant cells and mating them with *E. coli* strain LE392, with selection on L-agar containing tetracycline but lacking thymine to prevent growth of the donor. Subsequently, F::Tn10 was transferred by filter mating from strain LE392 to the prototrophic *E. coli* K5 wild-type strain Bi8337-41, with selection on minimal medium M9 supplemented with glucose and tetracycline. Strain LE392 cannot grow on this medium as it requires methionine and tryptophan. The frequency of transfer of F into the wild-type encapsulated *E. coli* was approximately 10 fold lower than into *E. coli* K-12 strains (data not shown). This may be because the capsule or O antigen layer of Bi8337-41 inhibits the formation of mating pairs.

Strain Bi8337-41 F::Tn10 was conjugated with *E. coli* K-12 PA360, a streptomycin resistant *serA* mutant. Selection was on minimal medium M9 containing streptomycin and all the supplements needed for the growth of PA360 except serine. Streptomycin resistant transconjugants not requiring serine for growth were then screened for the ability to produce a K5 capsule by sensitivity to a K5-specific bacteriophage. Approximately 70% of *ser* transconjugants produced a K5 capsule, giving clear plaques of 2-3mm in diameter identical to those of Bi8337-41. The *E. coli* K5/K-12 hybrid strain is referred to as MS101. Immunelectron microscopy revealed that MS101 produced a capsule that was morphologically identical to that of the wild-type (P. Hänfling, personal communication). In addition, MS101 did not produce a capsule at 18°C and was found to have elevated levels of CMP-KDO synthetase at 37°C (data not shown). This is typical for *E. coli* strains expressing group II capsules (Jann and Jann, 1990).

3.2.2 The effect of mutations in known regulatory genes on expression of the K5 antigen

Mutations in known regulatory genes were introduced into MS101 by transduction using bacteriophage P1vir. Production of a capsule was then assayed by sensitivity to K5-specific phage. Table 3.1 lists the mutations studied and their effect on the K5 phage sensitivity of MS101 at 37°C and 18°C. Where possible, the transfer of the regulatory mutations was confirmed as described in Table 3.2. The regulators
studied include the histone-like proteins H-NS, integration host factor, and the leucine-responsive regulatory protein, which affect gene expression by altering DNA topology, as well as mutations in topoisomerase I and protein HU, which have a more general effect on DNA supercoiling. The role of RfaH, which regulates LPS core biosynthesis, and the rcs system, which controls the expression of *E. coli* group I K antigens and other polysaccharides in enteric bacteria, was also studied. The effect of a mutation in the *tcp* locus, which regulates the expression of *E. coli* Pap pilin in response to temperature, was also investigated.

Table 3.1. Sensitivity of regulatory mutants of MS101 to K5-specific bacteriophage at 37°C and 18°C.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Mutation</th>
<th>Donor strain</th>
<th>Sensitive (S)/Resistant (R)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>37°C</td>
</tr>
<tr>
<td>MS101</td>
<td>hns::Tn10</td>
<td>GM230</td>
<td>S</td>
</tr>
<tr>
<td>MS102</td>
<td>topA::Tn10</td>
<td>RED31</td>
<td>S</td>
</tr>
<tr>
<td>MS103</td>
<td>himA::Cm^R</td>
<td>HN1491</td>
<td>S</td>
</tr>
<tr>
<td>MS104</td>
<td>himD::Cm^R</td>
<td>CH1569</td>
<td>S</td>
</tr>
<tr>
<td>MS105</td>
<td>rfaH::Tn5</td>
<td>1.8</td>
<td>R</td>
</tr>
<tr>
<td>MS106</td>
<td>lrp::Tn10</td>
<td>CV1008</td>
<td>S</td>
</tr>
<tr>
<td>MS107</td>
<td>rcsA::Km^R</td>
<td>VS20184</td>
<td>S</td>
</tr>
<tr>
<td>MS108</td>
<td>rcsB::Km^R</td>
<td>VS20185</td>
<td>S</td>
</tr>
<tr>
<td>MS111</td>
<td>hupA::Cm^R</td>
<td>OHP109</td>
<td>S</td>
</tr>
<tr>
<td>MS112</td>
<td>hupB::Km^R</td>
<td>OHP96</td>
<td>S</td>
</tr>
<tr>
<td>MS113</td>
<td>tcp::mTn10</td>
<td>DL652</td>
<td>S</td>
</tr>
</tbody>
</table>

For several other regulators, mutations caused by the insertion of a selectable marker could not be obtained. In these instances, the mutant was transformed with the cloned K5 capsule genes (pPC6). These included mutants affected in the alternative σ factors, σ^/i^ (KY1429), σ^/i^ (ET8045), and σ^/i^ (UM56-64), and the cAMP-cAMP receptor protein system (CA8445). In each case, the mutations did not affect capsule production, as determined by sensitivity to K5 phage (data not shown). Further, MS101 grown at 37°C in minimal medium M9 with glucose as the sole carbon source was sensitive to K5 phage, indicating that group II capsule production is not subject to catabolite repression.
Table 3.2. Confirmation of the regulatory mutations studied.

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Confirmation</th>
</tr>
</thead>
<tbody>
<tr>
<td>hns::Tnl0</td>
<td>Inability to PCR with primers flanking hns</td>
</tr>
<tr>
<td>himA::Cm(^#)</td>
<td>Colony hybridisation with a 339 bp Dral fragment from the pACYC184 chloramphenicol acetyl transferase gene (Cm(^#) gene probe). Poor growth in minimal medium in the absence of alanine</td>
</tr>
<tr>
<td>himD::Cm(^#)</td>
<td>Non-haemolytic when transformed with the cloned (\alpha)-haemolysin (hly) genes (pEK50)</td>
</tr>
<tr>
<td>rfaH::Tn5</td>
<td>Elevated l-serine deaminase activity(^\dagger)</td>
</tr>
<tr>
<td>brp::Tnl0</td>
<td>Colony hybridisation with a 402 bp PvuI-Xhol fragment from the pKOK(^#) kanamycin resistance gene (Km(^#) gene probe)</td>
</tr>
<tr>
<td>rcsA::Km(^#)</td>
<td>Colony hybridisation with the Cm(^#) gene probe</td>
</tr>
<tr>
<td>rcsB::Km(^#)</td>
<td>Colony hybridisation with the Km(^#) gene probe</td>
</tr>
<tr>
<td>hupA::Cm(^#)</td>
<td>Inability to grow at 42°C</td>
</tr>
<tr>
<td>hupB::Km(^#)</td>
<td>Inability to grow using arginine as the sole nitrogen source</td>
</tr>
<tr>
<td>rpoH</td>
<td>Reduced catalase activity</td>
</tr>
</tbody>
</table>

\(^\dagger\) Assayed by the method of Tuan et al., (1990) (data not shown).

Of the regulatory mutations studied only one had any detectable effect. This was a mutation in the \(rfaH\) gene, and it abolished the expression of any detectable capsule at 37°C. None of the mutations induced capsule expression at 18°C, indicating that the temperature-dependent expression of group II capsules does not require any of the regulators studied here.

To quantitate the effect of some of the regulatory mutations, the biosynthesis of the K5 polysaccharide was assayed \textit{in vitro}. Membranes were isolated from bacteria grown at 37°C and incubated with the nucleotide sugar components of the polysaccharide. One of the nucleotide sugars, UDP-glucuronic acid, was labelled with \(^{14}\)C so that polysaccharide biosynthesis could be quantitated using a scintillation counter (section 2.13.4).
The *in vitro* UDP-glucuronic acid transferase activity of the MS101 *rfaH* mutant (MS106) was less than 10% that of MS101 and was comparable to that of PA360, which lacks the capsule genes (Table 3.3). Mutations affecting H-NS and integration host factor, also reduced the *in vitro* UDP-glucuronic transferase activity, but to a lesser extent than the *rfaH* mutation. The data do not support roles for the leucine-responsive regulatory protein or the *rcs* system in expression of the K5 antigen.

### Table 3.3. UDP-Glucuronic acid transferase activities of regulatory mutants of MS101.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Mutation</th>
<th>μmol ^14*C UDP-GlcA incorporated per mg protein (in 30 min. @ 37°C)</th>
<th>Standard error of the mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA360</td>
<td>kps^-</td>
<td>0.49</td>
<td>0.23</td>
</tr>
<tr>
<td>MS101</td>
<td>kps^+</td>
<td>7.23</td>
<td>0.57</td>
</tr>
<tr>
<td>MS102</td>
<td>hns::Tn10</td>
<td>3.17</td>
<td>0.31</td>
</tr>
<tr>
<td>MS105</td>
<td>himD::Cm^8</td>
<td>2.23</td>
<td>0.20</td>
</tr>
<tr>
<td>MS106</td>
<td>rfaH::Tn5</td>
<td>0.68</td>
<td>0.11</td>
</tr>
<tr>
<td>MS107</td>
<td>lrp::Tn10</td>
<td>5.87</td>
<td>0.79</td>
</tr>
<tr>
<td>MS109</td>
<td>rcsA::Km^R</td>
<td>7.63</td>
<td>0.45</td>
</tr>
<tr>
<td>MS110</td>
<td>rcsB::Km^R</td>
<td>6.67</td>
<td>0.35</td>
</tr>
</tbody>
</table>

† Mean of 6 measurements using membranes from 2 separate cultures.

By immunoelectron microscopy it was found that MS106 does not accumulate intracellular polysaccharide (Figure 3.1). This indicates that RfaH is required for the biosynthesis, and not export, of the K5 polysaccharide. The isogenic parent of MS101 (PA360) did not bind the gold-conjugated anti-K5 antibody (data not shown), indicating that the labelling was specific.

#### 3.2.3 Expression of the K5 antigen is directly regulated by RfaH

To preclude the possibility that MS106 contains a secondary mutation that confers an acapsular phenotype, *E. coli* strain 1.8, which carries an *rfaH* mutation, was transformed with plasmids pPC6 and pKT274, carrying the cloned K5 and K1 capsule genes respectively. Both strains were found to be resistant to capsule-specific bacteriophage, indicating that RfaH has a general role in the expression of group II
Figure 3.1. Immunoelectron micrographs of thin sections from *E. coli* MS101 (A) and MS106 (*rfaH*) (B) grown at 37°C. Bar = 1μm.
capsules. This was confirmed by transformation of *E. coli* strain ED3869, which has a different *rfaH* allele (*sfrBl1*), caused by a point mutation. This result also indicates that the Tn5 insertion in *rfaH* in strains 1.8 and MS106 does not have a polar effect on the expression of downstream genes that may be required for capsule expression.

The notion that MS106 does not contain secondary mutations is supported by the observation that it can be complemented by a plasmid carrying a functional copy of the *rfaH* gene (pKZ17). Transformation of MS106 with pKZ17 conferred sensitivity to K5 phage and restored the *in vitro* transferase activity to a level comparable to that of MS101 (Table 3.4).

### Table 3.4. Complementation of MS106 by the *rfaH* subclone pKZ17.

<table>
<thead>
<tr>
<th>Strain</th>
<th>µmol ¹⁴C UDP-GlcA incorporated per mg protein (in 30 min. @ 37°C)†</th>
<th>Standard error of the mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA360</td>
<td>0.28</td>
<td>0.06</td>
</tr>
<tr>
<td>MS101</td>
<td>6.60</td>
<td>0.48</td>
</tr>
<tr>
<td>MS106</td>
<td>0.29</td>
<td>0.07</td>
</tr>
<tr>
<td>MS106 (pKZ17)</td>
<td>6.54</td>
<td>0.29</td>
</tr>
</tbody>
</table>

† Mean of 6 measurements using membranes from 2 separate cultures.

RfaH mutants produce rough LPS with a heterologous core structure (Lindberg and Hellerqvist, 1980). Mutations which affect LPS core biosynthesis, such as *galU* and *rfaG* defects, have been reported to alter the expression of several surface structures in *E. coli*, causing the loss of pili and flagella and inducing colanic acid biosynthesis (Komeda *et al.*, 1977; Parker *et al.*, 1992). The loss of flagella in *E. coli* *galU* mutants appears to be due to a decrease in flagellin mRNA (Komeda *et al.*, 1977), and it has been suggested that the expression of flagella may be regulated by a system which is capable of sensing the degree of completion of the LPS core (Schnaitman and Klena, 1993).

To determine if capsule loss in an *rfaH* mutant was the result of LPS truncation, capsule production was assayed in an *E. coli* *galU* mutant, which is unable to make UDP-glucose and has a truncated LPS core similar to that of an *rfaH* mutant (Schnaitman and Klena, 1993). The K5 polysaccharide cannot be made in *galU* mutants since it contains glucuronic acid, for which UDP-glucose is the immediate
precursor. Instead, production of the K1 antigen was assayed as this does not require UDP-glucose. E. coli CAIO (galU) was transformed with pKT274 and found to be sensitive to K1-specific phage, suggesting that LPS truncation does not affect the expression of group II capsules.

LPS truncation has been reported to affect the protein composition of the outer membrane (Ames et al., 1974), and it has been postulated that LPS may act as a ligand to promote the correct folding and insertion of membrane proteins (Schnaitman and Klena, 1993). It has been shown that LPS is involved in the assembly of outer membrane porins in E. coli (Ried et al., 1990), and that LPS interacts with OmpF directly in vitro, facilitating its renaturation, trimerisation and membrane incorporation (Sen and Nikaido, 1990). The outer membrane protein TolC, which is required for the secretion of α-haemolysin (Wandersman and Délepelaire, 1990), is also believed to require LPS for folding and insertion into the outer membrane (Wandersman and Létoffé, 1993). Mutations in rfaH reduce the secretion of α-haemolysin, and it is believed that this is due, in part, to the effect of LPS truncation on TolC function (Wandersman and Létoffé, 1993). RfaH mutations also reduce transcription of the α-haemolysin (hly) operon (Bailey et al., 1992).

To confirm that the effect of an rfaH mutation on capsule expression was not the result of altered TolC function, E. coli A586, which has a mutation in tolC, was transformed with the cloned K1 and K5 capsule genes. In both cases, the transformants were sensitive to capsule-specific phage. It is unlikely that the outer membrane porin OmpF is required for K5 capsule expression, because almost no OmpF is made in tolC mutants (Morona et al., 1983).

3.3 Discussion

In order to study the regulation of group II capsule expression, a laboratory strain of E. coli carrying the K5 capsule genes in their normal chromosomal location was constructed. This was accomplished by Hfr mating making use of the close linkage of kps to serA. The K5/K-12 hybrid (MS101) produced a capsule identical to that of the wild-type when examined by immunoelectron microscopy and exhibited temperature-dependent capsule production and CMP-KDO synthetase activity. Mutations in known regulatory genes were introduced into MS101 by P1 transduction and their effect on capsule production assayed by sensitivity to a K5-specific phage and by measuring the in vitro UDP-glucuronic acid transferase activity.

None of the mutations studied induced expression of the K5 antigen at 18°C, as determined by sensitivity to K5 phage. Thus, the hns and tcp genes, which repress the expression of E. coli Pap pili at low temperatures (Göransson et al., 1990; White-Ziegler et al., 1992), do not appear to be involved in the mechanism by which K5 capsule
production is repressed at 18°C. However, in the absence of in vitro transferase data for the MS101 regulatory mutants at 18°C, it is not possible to preclude a role for any of the regulators studied here in the thermoregulation of group II capsule production. It may be that a combination of environmental cues is needed for the activation of group II capsule expression by temperature. This is true of the *Shigella flexneri* invasion genes, which are only induced by temperature if the external osmolarity is high (Porter and Dorman, 1994).

Mutations affecting the histone-like proteins H-NS and integration host factor reduced the in vitro transferase activity of MS101 by about 2-3 fold. This may indicate that the expression of genes involved in biosynthesis of the K5 antigen is sensitive to DNA supercoiling. Topoisomerase I and protein HU mutations, which have a more general effect on DNA supercoiling, may be expected to have a similar effect on biosynthesis of the K5 polysaccharide in vitro. It is believed that most promoters are sensitive to DNA topology to some extent (Pérez-Martín et al., 1994). Topoisomerase I mutants rapidly acquire compensatory mutations, for instance in the DNA gyrase (gyrAB) genes (Dorman et al., 1989), so their effect is unreliable and difficult to interpret. The topA mutants RED31 and MS103 were grown on low osmolarity media as this reduces the selection for compensatory mutations (C. Dorman, personal communication).

Mutations affecting H-NS and IHF are highly pleiotropic (section 1.7.1.1), and it is possible that they have an indirect effect on capsule expression, perhaps by affecting the expression of other factors needed for capsule production. Regulatory hierarchies of this kind are common in the control of gene expression (Dorman and Ni Bhriain, 1992). The nucleotide sequence of the *E. coli* K5 capsule gene cluster was searched for the 13 bp IHF consensus binding sequence (Goodrich et al., 1990) using the FINDPATTERNS program (Devereux et al., 1984). A putative IHF binding site was identified 40 bp from the translation initiation codon for kfiA. Transcripts which span all of region 2 have been shown to initiate 68 bp from the start of kfiA (Petit et al., 1995), therefore the binding of IHF at this site may affect the expression of region 2 genes. No close matches to the LRP or CRP consensus binding sites were identified.

H-NS negatively regulates the expression of colanic acid in *E. coli* by repressing the transcription of rcsA (section 1.8.1) (Sledjeski and Gottesman, 1995). Homologues of RcsA regulate the production of diverse polysaccharides in enteric bacteria, therefore H-NS may have a general role in the expression of polysaccharides that are regulated by the *rcs* system. The involvement of H-NS and DNA supercoiling in these systems may mean that environmental cues, such as osmolarity and temperature, can directly influence capsule production through their effect on DNA topology.

The *rcs* system does not appear to regulate the expression of the K5 antigen, since the in vitro transferase activities of *rcsA* and *rcsB* mutants of MS101 do not differ
significantly from that of MS101. In support of this data, it has been reported that the introduction of multicopy \textit{rcsA} or \textit{rcsB} into \textit{E. coli} K1 and K5 does not affect the biosynthesis of these polysaccharides \textit{in vitro}, whereas colanic acid production in these strains is greatly increased by these plasmids (Keenleyside \textit{et al}., 1992, 1993).

\textit{RcsA} has been reported to negatively regulate expression of the \textit{E. coli} K54 antigen (Russo and Singh, 1993). This effect requires \textit{RcsB} and varies with the growth temperature, being maximal at 28°C. The authors conclusions were based on the increased serum resistance of an \textit{E. coli} K54 \textit{rcsA} mutant and on the activity of alkaline phosphatase fusions to capsule genes in \textit{rcsA} and \textit{lon} mutants. Neither of these parameters necessarily correlate with increased capsule biosynthesis. The K54 antigen was described by Russo and Singh (1993), as a group II capsule, however it is expressed at all growth temperatures and genetic data suggest that it belongs to group III (Pearce and Roberts, 1995). The role of \textit{RcsA} as a negative regulator of group III capsule expression is in contrast to its role as a positive regulator in all other systems (section 1.8.1) (Stout \textit{et al}., 1991; McCallum and Whitfield, 1991; Keenleyside \textit{et al}., 1992).

For regulatory mutations that could not be transduced into MS101, including mutations in alternative \textit{\sigma} factors and the cAMP-cAMP receptor protein system, the mutants were transformed with the cloned K5 capsule genes (pPC6). Although these strains were sensitive to K5 phage, roles for these regulators cannot be dismissed for several reasons. These include the limitations of using plasmids to study regulation addressed earlier and the fact that pPC6 does not contain the entire K5 capsule gene cluster. Although pPC6 gives wild-type plaques with the K5 phage and exhibits temperature-dependent capsule production and CMP-KDO synthetase activity, it does not contain the region 1 promoter or a functional copy of \textit{kpsF}. It follows that the control sequences that may be required for the action of these regulators may be absent. The transcription of region 1 on pPC6 is believed to result from a vector promoter (D. Simpson, unpublished results), and this may compensate for the loss of these sequences. By restriction endonuclease mapping, the region 3 promoter (section 5.2.5) was found to be present on pPC6.

Only one of the regulatory mutations studied had a detectable effect on K5 capsule expression at 37°C. This was a mutation in the \textit{rfaH} gene and it conferred resistance to K5-specific phage at 37°C. The \textit{in vitro} UDP-glucuronic acid transferase activity of the \textit{rfaH} mutant (MS106) was comparable to that of a strain which lacks the capsule genes. This suggests that RfaH is needed for the biosynthesis of the K5 polysaccharide. The lack of intracellular polysaccharide detected in an \textit{rfaH} mutant by immunoelectron microscopy supports this notion. The phenotype of MS106 is the same as that of a region 2 mutant (Boulnois \textit{et al}., 1987; Roberts \textit{et al}., 1988). This does not automatically mean that RfaH regulates the expression of region 2 genes, as
proteins from other regions of the capsule gene cluster are also needed for efficient polysaccharide biosynthesis (section 1.6) (Vimr et al., 1989; Bronner et al., 1993a, 1993b).

It was shown that indirect effects of the rfaH mutation on LPS structure do not affect capsule production. This notion is supported by the observation that deep rough strains of E. coli are capable of producing K5 capsules that are morphologically the same as the wild-type (Finke et al., 1991). Indeed, MS101 produces rough LPS (it has an rfbD mutation) and yet still produces a typical capsule. The possibility that the rfaH::Tn5 mutation was polar or that MS106 contains secondary mutations was also precluded. Taken together, these results indicate that RfaH directly regulates the expression of the K5 antigen. The finding that RfaH is required for production of the K1 antigen, suggests that it has a general role in the expression of group II capsules.

RfaH regulates the expression of several virulence and fertility factors in E. coli. It is required for biosynthesis of core lipopolysaccharide (section 1.8.3) (Pradel and Schnaitman, 1991), synthesis and secretion of α-haemolysin (Bailey et al., 1992), and production of the F factor sex pilus (Beutin and Achtman, 1979). The rfaH (sfrB) gene is located at 87 minutes on the E. coli chromosome (Bachmann, 1990), and encodes a protein of 18.3 kDa that has been visualised in minicells (Rehemtulla et al., 1986; Bailey et al., 1992). The isolation of amber mutations in rfaH (Creeger et al., 1984), implies that rfaH function is mediated by its protein product. Hybridisation and gene fusion studies have indicated that RfaH is a transcriptional activator and may act by causing antitermination (Beutin et al., 1981; Gaffney et al., 1983; Farewell et al., 1991; Bailey et al., 1992). RfaH lacks known DNA binding motifs and exhibits no significant homology to histone-like proteins which activate transcription by altering DNA topology (Bailey et al., 1992).

The transcriptional regulation of the F factor transfer region (tra) is complex and may involve several internal promoters and terminators (Figure 3.2) (reviewed in Frost et al., 1994). The polarity of insertions in the tra operon suggests that a single transcript extends from traY to traI (Frost et al., 1994). Beutin et al., (1981) analysed transcription of the traY-I operon in an rfaH mutant by hybridising DNA fragments of different lengths from the tra operon to total RNA from an rfaH mutant containing F. RfaH was reported to reduce the premature termination of transcripts at two sites within the traY-I operon, between traC and traG and between traG and traD. The activity of lacZ transcriptional fusions along the tra operon in an rfaH mutant supports this notion (Gaffney et al., 1983). However these data are not conclusive since there was no assessment of the activity of internal promoters or the stability of transcripts in wild-type and rfaH mutant strains. There is evidence that the tra operon contains multiple Rho-dependent terminators (Penfold et al., 1994), therefore RfaH may be needed for complete transcription of the tra operon. The finding that rho mutations can suppress the effect of an rfaH mutation on F transfer (Beutin et al.,
lends support to the hypothesis that RfaH regulates transcription of the *tra* operon by antitermination. In addition, suppressors of the effect of *rfaH* mutations on core LPS biosynthesis map in the genes for Rho and subunits of RNA polymerase (section 1.8.3) (Farewell *et al*., 1991). Thus, RfaH may interfere with the ability of RNA polymerase to recognise Rho-dependent termination signals.

**Figure 3.2. Transcriptional organisation of *tra* genes in the F transfer region.** Boxes represent the *tra* open reading frames. The figure does not show the location of *trb* genes, or genes that are transcribed counter to the major *tra* promoter (P₉) (*artA* and *finP*). Minor promoters which transcribe the *tra* genes in the same direction as P₉ have been identified upstream of *traH, traS, traT* and *traD* (Ham *et al*., 1989). Two putative Rho-dependent transcription terminators have been identified in *traK* (Penfold *et al*., 1994).

Production of the *E. coli* α-haemolysin toxin is determined by the *hly* operon, which comprises four contiguous genes, *hlyCABD*, that may be located on the chromosome or on conjugative plasmids (Figure 3.3) (Welch and Pellett, 1988). HlyA is a non-toxic prohaemolysin which is activated by acylation by HlyC, and then exported by HlyB and HlyD (Welch, 1994).

**Figure 3.3. Transcriptional organisation of plasmid borne *E. coli* α-haemolysin (*hly*) operons.** Open boxes represent the open reading frames. The *hlyR* regulatory region is shown as a filled box. Arrows represent the two *hly* transcripts described in the text. The promoter for the *hly* operon is designated P₉. Chromosomal *hly* operons have a homologous organisation but the upstream sequences, including *hlyR*, are less well characterised (Cross *et al*., 1990).
Two *hly* transcripts have been detected; a major 4 kb *hlyCA* transcript and a minor 8 kb *hlyCABD* transcript, both of which initiate at the same site (Figure 3.3) (Welch and Pellett, 1988). The minor transcript is believed to result from antitermination at a Rho-independent terminator located between *hlyA* and *hlyB* (Koronakis et al., 1988). Transcription of the entire *hly* operon is activated by RfaH (Bailey et al., 1992). It is not yet known if RfaH regulates antitermination within the *hly* operon.

Plasmid borne α-haemolysin determinants contain a c. 600 bp sequence (*hlyR*) about 1.5 kb upstream of the *hly* promoter, which enhances α-haemolysin production (Figure 3.3) (Vogel et al., 1988; Koronakis et al., 1988, 1989). It appears that *hlyR* is required for antitermination within the *hly* operon (Koronakis et al., 1989). Thus, in the absence of *hlyR*, fewer transcripts continue into the *hlyB* and *hlyD* genes. Removal of *hlyR* also reduces transcription of the entire *hly* operon (Jubete et al., 1995). The sequences upstream of the chromosomal *hly* genes are different from those upstream of the plasmid *hly* determinant, but are also required for α-haemolysin production (Cross et al., 1990). Expression of the *hly* operon is also regulated by differential mRNA stability, with the *hlyCA* transcript being more stable (Welch and Pellet, 1988).

Transcription of the *hly* operon is negatively regulated by *hha* (Godessart et al., 1988; Nieto et al., 1991). The *hha* gene encodes a protein of 8.6 kDa that is highly homologous to the YmoA protein of *Yersinia enterocolitica* which modulates the expression of several virulence genes in response to temperature (De la Cruz et al., 1992). It is believed that Hha interacts with a sequence (*hlyM*), in the coding region of *hlyC* and suppresses transcription from the *hly* promoter (Jubete et al., 1995), perhaps by altering DNA topology (Carmona et al., 1993). The presence of *hlyR* suppresses this effect (Jubete et al., 1995).

It is not clear if *hha* regulates the expression of other RfaH-regulated operons. A *hha* mutant of the K5/K-12 hybrid strain was found to be sensitive to K5 phage (D. Simpson, unpublished results). However, if *hha* is a negative regulator it would be expected that capsule production would be increased. In the absence of *in vitro* transferase data for the MS101 *hha* mutant, a role for *hha* in the regulation of group II capsule production cannot be precluded. No homology between the *hlyM* sequence, which is needed for the action of Hha, and the sequence of other RfaH-regulated operons was detected (data not shown).

Because RfaH appears to be a transcriptional activator, it was decided to compare the pattern of transcription of genes from each of the three regions of the K5 capsule gene cluster in wild-type and *rfaH* mutant strains. This is the subject of the next chapter.
Chapter 4

RfaH regulates transcription of the K5 capsule genes

4.1 Introduction

To investigate the regulation of the K5 capsule gene cluster, bacterial luciferase (luxAB) transcriptional fusions were made in genes from each of the three regions. The luxAB genes used lack their native promoter, therefore the production of luciferase is a measure of transcription of the capsule genes. In the presence of its aldehyde substrate n-decanal, bacterial luciferase generates a blue-green fluorescence that can be sensitively detected in a luminometer (Meighen, 1991). For the reasons addressed in section 3.1, expression of the capsule genes is best studied in a strain with a single copy of the capsule genes in their normal chromosomal location. Thus, luxAB fusions to capsule genes were transferred to the chromosome of the K5/K-12 hybrid strain by allelic exchange using the positive-selection suicide vector pCVD442 (Donnenberg and Kaper, 1991).

Plasmid pCVD442 contains the origin of replication from plasmid R6K, the mob region from plasmid RP4, the Bacillus subtilis sacB gene and a blaM gene conferring resistance to ampicillin (Donnenberg and Kaper, 1991). Replication at the origin of plasmid R6K requires π protein (Kolter et al., 1978). Plasmid R6K carries a gene (pir) which encodes the π protein, however pCVD442 does not have this gene and so can only replicate in strains which supply the π protein in trans. Derivatives of pCVD442 carrying luxAB fusions in capsule genes are unable to replicate in MS101 as it does not contain the pir gene. However, if selection for ampicillin resistance is applied, merodiploid strains in which the plasmid has integrated into the chromosome can be isolated (Figure 4.1). Integration occurs by homologous recombination and should therefore disrupt the resident capsule gene. It is possible to select for a second recombination event in which the reporter gene fusion replaces the wild-type copy of the capsule gene and the suicide vector is lost. This selection involves the sacB gene, which encodes the enzyme levan sucrase, that is toxic for Gram negative organisms when grown in the presence of sucrose (Blomfield et al., 1991). It follows that if the merodiploid is cultured in the absence of selection for ampicillin resistance and plated onto medium containing sucrose, only colonies which have lost the suicide vector will form. In some cases, the wild-type copy of the gene will be replaced by the reporter gene fusion (Figure 4.1).
Figure 4.1. Recombination events required for the transfer of a plasmid borne allele onto the chromosome. Integration can occur by recombination on either side of the plasmid allele (a or b). For excision to lead to successful allelic exchange, recombination must occur on the opposite side of the mutation to which integration occurred. Adapted from Blomfield et al., (1991).

4.2 Results

4.2.1 Construction of a luxAB fusion in kfiA in MS101

A transcriptional fusion between kfiA and the luxAB genes was made by ligating a 2.4 kb Smal-PvuII fragment from pHV100 into pAS21 partially digested with XmnI (Figure 4.2). The products of this ligation were used to transform E. coli strain JM101 selecting for ampicillin resistance. Recombinants were identified by the ability to produce light (section 2.14.1), and the orientation of the insertion was verified by digestion with AccI (Figure 4.2). This plasmid (pMS1) was then digested with Smal.
Figure 4.2. Construction of plasmids pMS1 and pMS2. Capsule genes are represented by black boxes and the luxAB genes by clear boxes. Restriction endonuclease cleavage sites are abbreviated as follows; A, AccI; Pv, PvuII; Sm, SmaI; St, StuI; X, XmnI. Plasmids are not drawn to scale.
and SfuI and a 4.2 kb fragment carrying the kfiA::luxAB fusion ligated with SmaI digested pCVD442, generating pMS2 (Figure 4.2). E. coli SY327λpir was used for the isolation of recombinants as it may be transformed more efficiently than strain SM10λpir (L. Zhao, unpublished results). Plasmid pMS2 was then transformed into SM10λpir and introduced into MS101 by conjugation with selection on L-agar containing ampicillin and streptomycin. Strain SY327λpir cannot be used to mobilise derivatives of pCVD442 as it lacks the plasmid RP4 transfer genes.

To confirm that plasmid pMS2 had integrated at the correct site, chromosomal DNA was prepared from ampicillin resistant transconjugants, digested with BglII and Southern blots performed using a 2.4 kb EcoRI fragment from region 2 (Figure 1.2) as a probe. Plasmid pMS2 (10.4 kb) does not contain any BglII sites, therefore the native 5.7 kb BglII region 2 fragment (Figure 1.2) was found to shift in size to 16.1 kb on the integration of pMS2 (Figure 4.3). As a control, chromosomal DNA was also prepared from MS101 strains into which plasmid pCVD442 alone had been introduced. The integration of pCVD442 occured outside region 2, indicating that the integration of pMS2 was specific (Figure 4.3). Strains in which pMS2 had integrated correctly were found to be resistant to K5-specific bacteriophage. The merodiploid strains still contain functional copies of all the capsule genes (Figure 4.1), however it is possible that the insertion had a polar effect on the expression of downstream genes.

A merodiploid of MS101 containing pMS2 integrated in region 2 was grown to late logarithmic phase (Aoo ~ 1) in L-broth in the absence of ampicillin. Dilutions of the culture (10⁰ through 10⁴) were then plated onto L-agar (lacking sodium chloride) containing 6% (w/v) sucrose and streptomycin, with overnight incubation at 30°C. These conditions were reported by Blomfield et al., (1991), to be optimal for the selection of double recombinants. Sucrose resistant colonies were first replica plated onto media containing streptomycin and ampicillin and media containing streptomycin alone to check that they had lost the suicide vector. The second recombination often results in the loss of the entire integrated plasmid (Figure 4.1) (Blomfield et al., 1991), therefore ampicillin sensitive, sucrose resistant isolates were screened for bioluminescence and sensitivity to K5 phage. Chromosomal DNA was prepared from bioluminescent and K5 phage resistant isolates on the assum patents that they had retained the kfiA::luxAB fusion. This DNA was digested with BglII and a Southern blot performed using the 2.4 kb EcoRI fragment from region 2 as a probe. These isolates were found to contain an 8.1 kb BglII fragment, indicating that the native copy of kfiA has been replaced by the kfiA::luxAB fusion (Figure 4.4). The replacement of kfiA with a kfiA::luxAB fusion was confirmed by PCR using primers 5'-ATGATTGTTGCAAATATGTCATCATAC-3' and 5'-TTAGTTAATATGTGGAAGGGTAAAAAG-3', which amplify the kfiA gene (Figure 4.5). A PCR product of 0.6 kb was amplified using MS101 chromosomal DNA, whereas a product of 3.0 kb was amplified using DNA
Figure 4.3. Southern blot analysis of BgIII digested chromosomal DNA from MS101 (track 1), MS101 merodiploids containing pMS2 (tracks 2 to 5), and MS101 merodiploids containing pCVD442 (tracks 6 to 9). A 2.4 kb EcoRI fragment from pAS21 was used as probe. Markers (kb) are HindIII fragments of λ DNA.
Figure 4.4. Southern blot analysis of BglII digested chromosomal DNA from MS101 (track 1), and MS101 derivatives containing a kfiA::luxAB fusion (MS108) (tracks 2 to 5). A 2.4 kb EcoRI fragment from pAS21 was used as probe. Markers (kb) are HindIII fragments of λ DNA.

Figure 4.5. Verification of the kfiA::luxAB fusion in strain MS108 by PCR. PCR reactions were performed with primers flanking kfiA using chromosomal DNA from MS101 (track 1), an MS101 merodiploid containing pMS2 integrated in region 2 (track 2) and MS108 (track 3). 1 kb DNA ladder was used for size markers. 
from the strain containing the \textit{kfiA::luxAB} fusion, owing to insertion of the 2.4 kb \textit{Smal-Pvul} fragment from pHV100. The strain carrying the \textit{kfiA::luxAB} fusion is referred to as MS108.

4.2.2 Construction of a \textit{luxAB} fusion in \textit{kpsT} in MS101

A \textit{luxAB} fusion in \textit{kpsT} was generated by cloning the 2.4 kb \textit{Smal-Pvul} fragment from pHV100 into a unique \textit{EcoRV} site in pAS1, generating pMS3 (Figure 4.6). A 4.3 kb \textit{SacI-SphI} fragment of pMS3 was then cloned into pCVD442 digested with \textit{SacI} and \textit{SphI}, generating pMS4 (Figure 4.6). Plasmid pMS4 was introduced into MS101 from SM10\textit{pir} as described in section 4.2.1. Chromosomal DNA was prepared from recipients of pMS4 that were resistant to K5-specific phage on the assumption that the plasmid had integrated at the correct site in these strains. This DNA was digested with \textit{BglII} and Southern blots performed using a 1.8 kb \textit{SacI-SphI} fragment from pAS1 as a probe. Plasmid pMS4 (10.5 kb) does not contain any \textit{BglII} sites, therefore the native 1.8 kb \textit{BglII} region 3 fragment (Figure 1.2) was found to shift in size to 12.3 kb on the integration of pMS4 (data not shown). A double recombinant was selected by growing this merodiploid in the absence of ampicillin on medium containing sucrose as before. Chromosomal DNA was prepared from sucrose resistant isolates that were ampicillin sensitive, bioluminescent and resistant to K5 phage. This DNA was digested with \textit{BglII} and a Southern blot performed using the 1.8 kb \textit{SacI-SphI} fragment from pAS1 as a probe. The region 3 probe hybridised to a 4.2 kb \textit{BglII} fragment in these strains, indicating that the native copy of \textit{kpsT} had been replaced by the \textit{kpsT::luxAB} fusion (data not shown). The strain carrying the \textit{kpsT::luxAB} fusion is referred to as MS120.

4.2.3 Construction of a \textit{luxAB} fusion in \textit{kpsE} in MS101

A \textit{kpsE::luxAB} fusion was generated by cloning the 2.4 kb \textit{Smal-PvuII} fragment of pHV100 into a unique \textit{EcoRV} site in pCR6, generating pMS5 (Figure 4.7). A 4.5 kb \textit{Smal-Bst1107I} fragment of pMS5 was then cloned into \textit{Smal} digested pCVD442 (Figure 4.7), and this plasmid (pMS6) introduced into MS101. Chromosomal DNA was prepared from K5 phage resistant recipients of pMS6, digested with \textit{HindIII} and a Southern blot performed using a \textit{kpsE} PCR product as a probe. On the integration of pMS6, the native 10.4 kb \textit{HindIII} region 1 fragment (Figure 1.2) was replaced by fragments of c. 3 and 12 kb, which were predicted from the location of \textit{HindIII} sites within pMS6 (data not shown). Following sucrose selection, chromosomal DNA was
Figure 4.6. Construction of plasmids pMS3 and pMS4. Capsule genes are represented by black boxes and the luxAB genes by clear boxes. Restriction endonuclease cleavage sites are abbreviated as follows; Ev, EcoRV; Pv, PvuII; Sc, SacI; Sm, SmaI; SphI. Plasmids are not drawn to scale.
Figure 4.7. Construction of plasmids pMS5 and pMS6. Capsule genes are represented by black boxes and the luxAB genes by clear boxes. Restriction endonuclease cleavage sites are abbreviated as follows; B, Bst11071; Ev, EcoRV; Pv, PvuII; Sm, SmaI. Plasmids are not drawn to scale.
prepared from ampicillin sensitive, bioluminescent colonies. This DNA was digested with *SmaI* and a Southern blot performed using the *kpsE* PCR product as a probe. A double recombinant in which the native 4.7 kb *SmaI* fragment (Figure 1.2) was replaced by a 7.1 kb fragment was isolated (data not shown).

The strain carrying the *kpsE::luxAB* fusion (MS130) was found to be sensitive to K5 phage. The insertion of the *luxAB* genes into *kpsE* was found to have created a translational fusion in which the N-terminal 23 amino acids of KpsE were replaced by 56 residues from the N-terminal end of LuxE, which are encoded at the end of the *luxAB* insert (Figure 4.8). LuxE is a cytoplasmic protein that is involved in the biosynthesis of aldehydes for the luciferase reaction (Meighen, 1991). It was predicted that this fusion would increase the molecular weight of KpsE by about 3.6 kDa. By Western blotting using KpsE-specific antibody, MS130 was found to produce a larger form of KpsE than MS101 (Figure 4.9). The N-terminal residues of KpsE which were replaced in this fusion precede a predicted membrane spanning domain and are believed to reside in the cytoplasm (Rosenow *et al.*, 1995a). The ability of MS130 to make a capsule indicates that these N-terminal residues are not important for the activity of KpsE in polysaccharide transport.

### 4.2.4 The effect of growth temperature on transcription of *kpsE, kfiA and kpsT*

To determine if the transcription of *kpsE, kfiA* and *kpsT* is regulated by temperature, strains MS130, MS108 and MS120 were grown to mid-logarithmic phase at 18°C and 37°C and light production assayed as described in section 2.14.1 (Figure 4.10).

Transcription of *kpsE* was found to be reduced by about half at 18°C. This is not in agreement with the results of Northern blot and slot blot hybridisations in which region 1 transcripts have not been detected at 18°C (D. Simpson, unpublished results). Further, no KpsE protein has been detected in whole cell lysates of MS101 grown at 18°C by Western blotting (D. Simpson, unpublished results). This indicates that the luciferase fusion in *kpsE* may not be providing accurate information about the regulation of region 1 transcription.

Transcription of *kfiA* was also reduced by about half at 18°C. The KfiA protein has not been detected at 18°C by Western blotting (C. Petit, unpublished results). This suggests that the expression of *kfiA* may be regulated at the translational level. The transcription of the *kfiA::luxAB* fusion at 18°C and 37°C was not affected by the introduction of an integration host factor mutation. At 18°C the mean relative light units (rlu)/μg total protein expressed by a *himA* mutant of MS108 (MS117) was 5.9 ± 1.6 compared to 6.8 ± 1.4 for MS108. At 37°C the values were 12.2 ± 2.8 for MS117 and 14.7 ± 2.1 for MS108. Thus, despite the presence of an IHF consensus binding sequence
Figure 4.8. Strain MS130 contains a translational fusion between kpsE and the *Vibrio harveyii* luxE gene. The lux genes were cloned into the EcoRV site in kpsE on a 2.4 kb Smal-PvuII fragment from pHV100. Boxes represent the open reading frames. The nucleotide sequence of the positive sense strand 12 bp either side of the site of insertion is shown. The protein translation is given below the nucleotide sequence in single letter code.

Figure 4.9. Western blot analysis of KpsE produced by *E. coli* MS101 and MS130 grown at 37°C. Total protein (50μg) from PA360 (track 1), MS101 (track 2) and MS130 (kpsE::luxAB) (track 3), was probed with rabbit anti-KpsE polyclonal antibody using a goat anti-rabbit IgG-horseradish peroxidase conjugate as secondary antibody. Markers (kDa) are Bio-Rad broad range protein standards.
Figure 4.10. Transcription of *kpsE*, *kfiA* and *kpsT* luciferase fusions at 37°C and 18°C. Transcription is reported in terms of relative light units (per second) per μg total protein. Values are the mean of at least three separate assays using duplicate cultures. Error bars represent the standard error of the mean.
in the promoter region for \( kfiA \), IHF does not appear to regulate the expression of this gene.

The transcription of \( kpsT \) was reduced by about 16 fold at 18°C. To determine if the regulation of region 3 transcription by temperature involves the \( hns \) or \( tcp \) genes, which repress the transcription of the \( E. \ coli \ pap \) operon at low temperature (section 1.7.4), \( hns \) and \( tcp \) mutations were introduced into strain MS120. These mutations did not significantly affect \( kpsT \) transcription at 18°C (the mean rlu/μg total protein were; MS120, 0.4 ± 0.1; MS122 (MS120 \( hns::Tn10 \)), 0.3 ± 0.1; MS124 (MS120 \( tcp::mTn10 \)), 0.1 ± 0.1). It follows that neither H-NS nor the product of the \( tcp \) gene repress \( kpsT \) transcription at the capsule-restrictive temperature.

### 4.2.5 The effect of RfaH on transcription of \( kpsE, kfiA \) and \( kpsT \)

To determine the effect of RfaH on the transcription of \( kpsE, kfiA \) and \( kpsT \), the \( rfaH::Tn5 \) mutation from \( E. \ coli \) strain 1.8 was transduced into strains MS130, MS108 and MS120, respectively. The transductants (MS131 (\( kpsE::luxAB \)), MS115 (\( kfiA::luxAB \)) and MS121 (\( kpsT::luxAB \))), were shown to be non-haemolytic when transformed with pEK50, indicating that they carry the \( rfaH \) mutation. These strains were assayed for light production at 37°C (Figure 4.11).

The transcription of \( kpsE \) was not significantly affected by the \( rfaH \) mutation. Because the \( kpsE::luxAB \) fusion appeared to provide inaccurate information about the transcription of \( kpsE \) at 18°C, the effect of the \( rfaH \) mutation was confirmed by measuring KpsE production in strain MS106 by Western blotting. Strains MS101 and MS106 were found to produce equal amounts of KpsE at 37°C (Figure 4.12).

RfaH was found to positively regulate the transcription of \( kfiA \), with strain MS115 producing about 70 fold less light than MS108. The transcription of \( kpsT \) was reduced by a third by the \( rfaH \) mutation. The introduction of the \( rfaH \) subclone pKZ17 into MS115 and MS121 restored transcription of the \( kfiA \) and \( kpsT \) genes at 37°C to their wild-type levels (the mean rlu/μg total protein were; MS108, 14.7 ± 2.1; MS115 (\( rfaH \)), 0.2 ± 0.03; MS115 (pKZ17), 13.0 ± 1.4; MS120, 6.5 ± 0.7; MS121 (\( rfaH \)) 4.1 ± 0.4; MS121 (pKZ17), 6.6 ± 0.4).

Several problems associated with the use of luciferase as a reporter gene necessitate the confirmation of these effects. It has been reported that the \( luxAB \) genes can affect the activity of nearby promoters (Forsberg et al., 1994). This has been attributed to local bending of the DNA caused by an A+T rich region of curved DNA in the 5' coding sequence of \( luxA \). This may not be a general effect on all promoters since transcription from the \( gyrB \) promoter, which is sensitive to DNA topology, was unaffected by insertion of the \( luxAB \) genes downstream of the promoter (Forsberg et al., 1994). Secondly, superoxide and other reactive oxygen intermediates are generated.
Figure 4.11. Transcription of *kpsE, kfiA* and *kpsT* luciferase fusions in an *rfaH* mutant. Transcription is reported in terms of relative light units (per second) per µg total protein. Values are the mean of at least three separate assays using duplicate cultures. Error bars represent the standard error of the mean.
Figure 4.12. Western blot analysis of KpsE produced by *E. coli* MS101 and MS106 at 37°C. Total protein (50μg) from MS101 (track 1), MS106 (track 2), and PA360 (track 3), was probed with rabbit anti-KpsE polyclonal antibody using a goat anti-rabbit IgG-horseradish peroxidase conjugate as secondary antibody. Markers (kDa) are Bio-Rad broad range protein standards.
as by-products of the luciferase reaction and may indirectly affect gene expression by causing oxidative stress (González-Flecha and Demple, 1994). Thirdly, the activity of luciferase may vary with the growth phase, owing to the availability of reduced flavine mononucleotide, which is required for the luciferase reaction (Meighen, 1991).

In order to confirm that kfiA transcription is activated by RfaH, an attempt was made to generate a β-galactosidase (lacZ) transcriptional fusion in kfiA in MS101. β-galactosidase reporter fusions do not exhibit the problems associated with luciferase fusions (Forsberg et al., 1994). A 4.7 kb PstI fragment from pKOK6 carrying a promoterless copy of lacZ and a kanamycin resistance gene was cloned into a unique NsiI site in kfiA in pAS21, generating pMS7 (Figure 4.13). The kfiA::lacZ fusion was then subcloned from pMS7 on a 5.9 kb XmnI-Sall fragment into SmaI-Sall digested pCVD442, generating pMS8 (Figure 4.13). This plasmid was introduced into MS101 from SM10pir as described in section 4.2.1. Ampicillin resistant transconjugants were screened for the integration of pMS8 in region 2 by Southern blotting of BglII digested chromosomal DNA using the 2.4 kb EcoRI fragment of pAS21 as a probe. However, after screening over 50 transconjugants, no merodiploids containing single insertions of pMS8 in region 2 were identified. It is possible that the integration of pMS8 occurred via the lacZ gene because MS101 is lac+. However, a lac mutant of MS101, constructed by transducing MS101 with a P1vir lysate prepared on JM101 (Δlac-proAB), also showed aberrant integration of pMS8.

Because a kfiA::lacZ fusion in MS101 could not be isolated, kfiA transcription in an rfaH mutant was monitored by direct detection of the RNA in slot blot hybridisations. Total RNA from MS101 and MS106 was immobilised on a nylon filter as described in section 2.8.2 and hybridised to a 32P-labelled kfiA PCR product. The relative amounts of RNA bound by the probe were quantitated using a PhosphorImager. The reduction in kfiA transcription in MS106 is estimated to be about 20 fold (Figure 4.14). To confirm that the same amount of RNA was loaded from each strain, the membrane was stripped and re-hybridised using a kpsT PCR product as a probe. The transcription of kpsT was not affected by the rfaH mutation and signals of approximately equal intensity were obtained with RNA from MS101 and MS106 (Figure 4.14). The probe did not hybridise to RNA from PA360, which lacks the capsule genes. RNA samples treated with RNase were also loaded on the membrane but did not hybridise to the probe (data not shown), indicating that the signals were not due to contaminating DNA.
Figure 4.13. Construction of plasmids pMS7 and pMS8. Capsule genes are represented by black boxes and the *lacZ* and kanamycin resistance (KmR) genes by clear boxes. The *lacZ::KmR* cassette contains a bi-directional transcription terminator from phage *fd* (*tfd*) 3' of *lacZ*. Restriction endonuclease cleavage sites are abbreviated as follows; N, *NsiI*; Ps, *PstI*; Sa, *SalI*; Sm, *Smal*; X, *XmnI*. Plasmids are not drawn to scale.
Figure 4.14. Direct detection of kfa transcripts in MS101 and MS106 by hybridisation with a $^{32}$P-labelled DNA probe. Total RNA (50µg) from MS101 (slot 1), MS106 (slot 2) and PA360 (slot 3) was hybridised to a $^{32}$P-labelled kfa PCR product (panel A). To check the amount of RNA loaded from each strain, the membrane was stripped and re-hybridised with a kpsT PCR product (panel B).
The result of the slot blot hybridisation using a radiolabelled \( kpsT \) DNA probe and RNA from MS101 and MS106 is not in agreement with the finding that transcription of the \( kpsT: luxAB \) fusion is reduced by a third in an \( rfaH \) mutant. It is possible that the luciferase fusion in \( kpsT \) provides a more sensitive measure of region 3 transcription than can be obtained in slot blot hybridisations.

4.2.6 RfaH regulates the expression of all region 2 genes

To determine if RfaH regulates the transcription of the other region 2 genes, slot blot hybridisations were performed using total RNA from MS101 and MS106 and radiolabelled \( kfiB \) and \( kfiC \) PCR products as probes. Transcription of \( kfiB \) and \( kfiC \) was found to be reduced in the \( rfaH \) mutant (Figure 4.15). This reduction was quantitated using a PhosphorImager and was of a similar magnitude to that for \( kfiA \) (c. 20 fold). The images of the \( kpsT \) control hybridisations for these experiments are not shown as they are of a poor quality. In addition to these observations, no KfiD protein could be detected in an \( rfaH \) mutant by Western blotting (Figure 4.16). It was not considered necessary to quantitate \( kfiD \) transcription in MS106 as \( kfiD \) is in the same transcriptional unit as \( kfiC \). One explanation for these results is that KfiA positively regulates the expression of the other region 2 genes. Thus in an \( rfaH \) mutant, the expression of \( kfiB, kfiC \) and \( kfiD \) would be reduced because \( kfiA \) is transcribed at a lower level (section 4.2.5). However, this is unlikely as strain MS108, which has a \( luxAB \) insertion in \( kfiA \), was found to produce KfiD (Figure 4.16).

4.2.7 Plasmid pAS30 restores capsule production in MS106

Because the expression of region 2 genes is reduced in an \( rfaH \) mutant, an attempt was made to restore capsule production in MS106 by the introduction of multiple copies of the region 2 genes on a plasmid. Transformation of MS106 with pAS30 conferred sensitivity to K5 phage. Plasmid pAS21, which carries only \( kfiA \), was unable to restore capsule expression in the \( rfaH \) mutant. This is consistent with the observation that RfaH regulates the expression of all the region 2 genes, not just \( kfiA \). In addition, this provides further evidence that the reduction in the expression of region 2 genes in an \( rfaH \) mutant is not the result of less KfiA being made.
Figure 4.15. Direct detection of kfiB and kfiC transcripts in MS101 and MS106 by hybridisation with $^{32}$P-labelled DNA probes. Total RNA (50 μg) from MS101 (slot 1), MS106 (slot 2) and PA360 (slot 3) was probed with a $^{32}$P-labelled kfiB PCR product (panel A) and a kfiC PCR product (panel B). RNase treated samples did not hybridise to the probe (data not shown). Both membranes were stripped and re-hybridised to a kpsT PCR product and the amount of RNA loaded from MS101 and MS106 found to be roughly equal (data not shown).

Figure 4.16. Western blot analysis of KfiD produced by E. coli MS101, MS106 (rfaH) and MS108 (kfiA::luxAB) grown at 37°C. Total protein (50 μg) from PA360 (track 1), MS101 (track 2), MS106 (track 3) and MS108 (track 4), was probed with rabbit anti-KfiD polyclonal antibody using a goat anti-rabbit IgG-horseradish peroxidase conjugate as secondary antibody. Markers (kDa) are Bio-Rad broad range protein standards.
4.3 Discussion

To investigate the effect of temperature and RfaH on transcription of the K5 capsule genes, bacterial luciferase (luxAB) transcriptional fusions were made in each of the three regions. These fusions were transferred to the chromosome of MS101 using a positive-selection suicide vector so that expression of the genes could be studied in their natural context.

By measuring the bioluminescence of a strain carrying a kpsE::luxAB fusion at 37°C and 18°C, the transcription of kpsE was found to be reduced by almost a half at the capsule-restrictive temperature. However, region 1 transcripts have not been detected at 18°C in Northern blots and slot blot hybridisations (D. Simpson, unpublished results). In addition, neither KpsE nor KpsD can be detected in whole cell lysates of MS101 grown at 18°C (D. Simpson, unpublished results). The observation that the CMP-KDO synthetase activity of E. coli K5 is comparable to that of an acapsular strain at 18°C (Finke et al., 1990), suggests that the production of KpsU is also reduced at this temperature. Taken together, these data indicate that the expression of region 1 is repressed at low temperature. Sequences between kpsF and kpsE may be important in the thermoregulation of region 1 transcription, since deletion of these sequences prevents the temperature-dependent expression of CMP-KDO synthetase (Pazzani, 1992). It is possible that the activity of these sequences is altered by the insertion of the luxAB genes in kpsE. This may be attributed to the ability of the luxAB genes to alter DNA topology (Forsberg et al., 1994). Evidence that the luxAB insertion in kpsE uncouples the normal temperature regulation of region 1 could be obtained by Western blotting of total protein from strain MS130 grown at 18°C using antibody against KpsE. Because transcription of the kpsE::luxAB fusion was reduced by only a half at 18°C, it would be expected that the KpsE::LuxF fusion protein will still be made. The kpsE::luxAB fusion may therefore be of little use for investigating the role of RfaH in the regulation of region 1 transcription.

The transcription of a kfiA::luxAB fusion was found to be reduced by about a half at 18°C. The activity of the kfiA promoter cloned in a cat promoter probe vector has been found to be approximately the same at 18°C and 37°C (C. Petit, unpublished results). However, no KfiA protein has been detected at 18°C by Western blotting, implying that its expression may be subject to translational control (section 1.7.2). This notion is supported by the finding that a chromosomally encoded LacZ translational fusion to KfiA is active at 37°C but not 18°C (C. Petit, unpublished results). It is possible that the translational control of KfiA production involves thermo-labile secondary structures in the mRNA which obscure the Shine-Dalgarno sequence or start codon, as with the Yersinia pestis icrF mRNA (section 1.7.4) (Hoe and Goguen, 1993).
The transcription of the region 3 gene *kpsT* was significantly reduced at 18°C, being about 16 fold lower than at 37°C. The repression of *kpsT* transcription at 18°C was shown not to involve the histone-like protein H-NS, which controls the transcription of several virulence genes at low temperatures (Göransson *et al.*, 1990; Dorman *et al.*, 1990), or the tcp locus, which represses transcription of the *E. coli* pap operon below 25°C (White-Ziegler *et al.*, 1992). It has also been found that hns and tcp mutations do not induce production of the region 1 protein KpsD in MS101 at 18°C (D. Simpson, unpublished results). An attempt has been made to isolate the gene(s) required for the thermoregulation of region 3 transcription. This involved transposon mutagenesis of MS120, with selection for colonies that make more light at 18°C or less light at 37°C. So far, no regulatory genes have been isolated (C. Petit, unpublished results).

By transducing the strains carrying *luxAB* fusions in *kpsE*, *kfiA* and *kpsT* with an *rfaH* mutation, it was found that RfaH activates transcription of the region 2 gene *kfiA*. The *rfaH* mutant carrying the *kfiA::luxAB* fusion produced about 70 fold less light at 37°C than the wild-type strain. Transcription of *kpsT::luxAB* fusion was reduced by a third in an *rfaH* mutant, whereas the transcription of *kpsE* was not affected. Because the *kpsE::luxAB* fusion does not appear to provide accurate information about region 1 transcription at 18°C, the effect of RfaH on the expression of KpsE was determined by Western blotting. Strain MS106 was found to produce approximately the same amount of KpsE as MS101, confirming that the expression of *kpsE* does not require RfaH.

Because of the potential problems associated with the use of luciferase as a reporter gene, it was necessary to confirm the effect of an *rfaH* mutation on *kfiA* transcription. An attempt was made to generate a chromosomal β-galactosidase transcriptional fusion in *kfiA*. However, integration of the plasmid containing the *kfiA::lacZ* fusion failed to occur in region 2. Transcription of *kfiA::lacZ* fusion was therefore monitored by hybridisation of total RNA from strains MS101 and MS106 with a radiolabelled *kfiA* DNA probe. Transcription of *kfiA* was estimated to be reduced by about 20 fold in MS106. It was shown that the reduced signal was not due to degradation or unequal loading of the RNA by stripping the membrane and re-hybridising with a *kpsT* probe. RNA samples treated with RNase did not hybridise to the probe, indicating that the signals were not due to contaminating chromosomal DNA. Further, the probe did not bind to RNA from PA360, which lacks the capsule genes. Therefore the *kfiA::luxAB* fusion would appear to provide a faithful measure of the transcription of *kfiA* in an *rfaH* mutant.

The finding that *kpsT* transcripts detected by hybridisation with a radiolabelled DNA probe are made in equal amounts in wild-type and *rfaH* mutant strains is not in
accord with the data obtained with the kpsT::luxAB fusion. A reduction in region 3 transcription in an rfaH mutant has also been demonstrated by RNase protection analysis (section 5.2.4) and primer extension (5.2.5). These methods are more sensitive for the detection of specific mRNAs than slot blot hybridisations (Sambrook et al., 1989).

By direct detection of RNA in slot blot hybridisations, RfaH was also found to activate the transcription of the kfiB and kfiC genes. Further, KfiD was not detected in an rfaH mutant by Western blotting. The finding that RfaH regulates the transcription of the serotype-specific region 2 genes is consistent with the phenotype of an rfaH mutant determined by immunoelectron microscopy and measurement of the in vitro UDP-glucuronic acid transferase activity (section 3.2.2). The data also support previous observations that RfaH is a transcriptional activator (Beutin et al., 1981; Gaffney et al., 1983; Pradel and Schnaitman, 1991; Brazas et al., 1991; Bailey et al., 1992). The observation that capsule production in strain MS106 can be restored by the introduction of multiple copies of the region 2 genes on a plasmid, but not by kfiA alone, supports the finding that RfaH is required for the expression of all region 2 genes. It also indicates that the reduction in the expression of the region 2 genes in an rfaH mutant is not the result of less KfiA being made. The fact that capsule expression in MS106 can be restored by pAS30 also implies that there is no absolute requirement for RfaH for the activity of the region 2 promoters. However, it is possible that the region 2 genes on pAS30 are also transcribed from vector promoters.

It has been reported that transcription of the E. coli hly operon, which is regulated by RfaH, varies with growth phase (Koronakis et al., 1989), with antitermination being prominent throughout logarithmic growth but absent in stationary phase. Thus, transcription of the luciferase fusions should be analysed through the growth cycle in wild-type and rfaH mutant strains. This may be complicated by the fact that the luciferase reaction requires reduced flavine mononucleotide, the levels of which decline in stationary phase. Thus, transcription of the fusions may appear to decrease as the cells enter stationary phase, even if luciferase continues to be made at the same rate. In addition, the luciferase enzyme may accumulate during the growth cycle. For this sort of analysis, it may be better to express bioluminescence as a proportion of the viable count, as specific activities based on the total protein concentration take account of dead or non-viable cells which do not express the gene.

From the observations made so far in this study and considering how RfaH regulates other operons (sections 1.8.3 and 3.3), a model for the regulation of the K5 capsule gene cluster by RfaH has been devised. This is the subject of the next chapter.
5.1 Introduction

Studies on the regulation of the rfa, hly and tra operons indicate that RfaH may act by causing transcription antitermination (sections 1.8.3 and 3.3) (Beutin et al., 1981; Gaffney et al., 1983; Pradel and Schnaitman, 1991; Brazas et al., 1991; Bailey et al., 1992). Assuming this to be the case, it is proposed that RfaH activates the transcription of region 2 of the K5 capsule gene cluster by preventing the termination of region 3 transcripts (Figure 5.1). This model accounts for the observation that in an rfaH mutant, transcription of kfiA is reduced but transcription of region 3 is less affected (section 4.2.5). The reduction in region 2 transcription in an rfaH mutant explains the low in vitro transferase activity of MS106 and the absence of intracellular polysaccharide. An implicit requirement of this model is that the promoters for the region 2 genes are so weak that, in the absence of readthrough transcripts, insufficient amounts of the region 2 proteins are made for polysaccharide biosynthesis to occur.

Figure 5.1. Model for the regulation of region 2 transcription in E. coli K5 by RfaH. Boxes represent the open reading frames. Arrows indicate the direction of transcription. The extent of the readthrough transcript (broken line) is not known.
5.2 Results

5.2.1 Transcription from the kfiA promoter is not regulated by RfaH

To determine if transcription from the kfiA promoter is regulated by RfaH, the β-galactosidase activity of wild-type and rfaH mutant strains carrying pMS7 was determined. Plasmid pMS7 contains a lacZ transcriptional fusion in kfiA in pAS21 and sequences to the 3' end of kpsT, including the kfiA promoter identified by Petit et al., (1995) (section 4.2.5; Figure 4.13). The direction of kfiA transcription on pMS7 is counter to that from the lac promoter. The rfaH mutant 1.8 could not be used for measuring β-galactosidase activity as it is lac+, therefore the rfaH::Tn5 mutation was transduced into JM101 using bacteriophage P1vir, generating strain MS135. The β-galactosidase activity of MS135 carrying pMS7 did not differ significantly from that of JM101 (pMS7) at 37°C (Table 5.1), indicating that transcription from the kfiA promoter is not regulated by RfaH. A similar result was obtained by measuring light production in JM101 and 1.8 carrying pMS1, which has a luxAB fusion in kfiA in pAS21 (section 4.2.1; Figure 4.3) (mean rlu/μg total protein for 2 assays were JM101 (pMS1), 11.3 and 1.8 (pMS1), 17.6). Because the transcription of kfiA on the chromosome requires RfaH (section 4.2.5), these data support the model that the expression of kfiA occurs by readthrough transcription from region 3 and that RfaH is needed for this to occur.

Table 5.1. Transcription of a kfiA::lacZ fusion on pMS7 in an rfaH mutant.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Relevant genotype</th>
<th>Units of β-galactosidase†</th>
</tr>
</thead>
<tbody>
<tr>
<td>JM101 (pMS7)</td>
<td></td>
<td>1026 ± 39</td>
</tr>
<tr>
<td>MS135 (pMS7)</td>
<td>JM101 rfaH</td>
<td>975 ± 35</td>
</tr>
</tbody>
</table>

† Mean of at least three separate measurements ± the standard error of the mean.
To test the hypothesis that the kpsT-kfiA intergenic region is transcribed, reverse transcriptase-PCR was performed (Figure 5.2). Primer RT1 (5'-AACTCTTTTTTTCCAGGGTGGG-3') was annealed to transcripts entering kfiA using total RNA from MS101 and extended with AMV reverse transcriptase. Primer RT2 (5'-GGTTATTATGAGAATGTTGAAAGTGG-3') was then annealed to the cDNA at the 3' end of kpsT and second strand synthesis performed using Taq polymerase. The product was amplified over 35 cycles of PCR. An RT-PCR product of c. 1.1 kb was generated (Figure 5.3). A control reaction to which no AMV reverse transcriptase was added did not give a product, indicating that the RT-PCR product did not arise from contaminating MS101 chromosomal DNA. No RT-PCR product was obtained using RNA from PA360, which lacks the capsule genes. The RT-PCR product generated fragments of the expected size when digested with BglII and Stul (Figure 5.3).

Figure 5.2. Strategy for the detection of transcripts spanning the kpsT-kfiA intergenic region by reverse transcriptase-PCR. Arrows with broken lines indicate the direction of DNA synthesis. The location of restriction endonuclease cleavage sites used to verify the RT-PCR product is shown (B, BglII; S, Stul).

Given the possible importance of readthrough transcription in the expression of region 2 genes, the nucleotide sequence upstream of region 3 was determined and the site of initiation of region 3 transcripts mapped in wild-type and rfaH mutant strains.
Figure 5.3. Detection of transcripts spanning the *kpsT-kfiA* intergenic region by reverse transcriptase-PCR. Tracks 1 and 3 contain the products of RT-PCR reactions using primers RT1 and RT2 with RNA from MS101 and PA360 grown at 37°C, respectively. Control reactions using RNA from MS101 and PA360 but without reverse transcriptase were run in tracks 2 and 4 respectively. The RT-PCR product from MS101 digested with *BgIII* (track 5) and *StuI* (track 6), gave fragments of c. 0.3 and 0.8 kb with *BgIII* and c. 0.5 and 0.6 kb with *StuI*, as expected. Reactions containing template or primers only did not give RT-PCR products (data not shown). 1 kb DNA ladder was used for size markers.
5.2.3 Determination of the nucleotide sequence upstream of region 3

To determine the nucleotide sequence upstream of region 3 a 2.5 kb ClaI fragment including kpsM and about 2 kb of upstream sequence was subcloned from pCP1 into the AccI site of M13mp19, generating pMS17. The nucleotide sequence of the insert in pMS17 was determined and confirmed by sequencing the reverse strand using double stranded plasmid DNA (Figure 5.4).

The sequence upstream of region 3 in E. coli K5 has a %G+C content of 35.9% and is 97.3% identical to that described for E. coli K1 (accession number U05251, Silver et al., unpublished results). No open reading frames were detected upstream of region 3 in E. coli K1 and K5 using the FRAMES program (Devereux et al., 1984). The consensus binding sequences recognised by IHF, CRP, LRP and alternative σ factors were not identified upstream of region 3 using the FINDPATTERNS program and no prominent secondary structures were predicted for the DNA or mRNA in this region.

A search of all nucleotide sequences in the database with the sequence upstream of region 3 in E. coli K5 revealed 97% identity in a 175 bp overlap to DNA in the 65 minute region of the E. coli K-12 chromosome. This homology begins 1058 bp upstream of the translation initiation codon for kpsM and marks the end of the K5 capsule gene cluster. The E. coli K1 kps locus was found to end at exactly the same position. The nucleotide sequence upstream of region 1 was not homologous to DNA in the 65 minute region, indicating that further capsule-specific sequences may be present at this end of the capsule gene cluster. The complete nucleotide sequence of the 65 to 68 minute region of the E. coli K-12 chromosome is known (accession number U28377). The region 3 end of kps is 10.85 kb from the glcB gene, which encodes malate synthetase, and 41.56 kb from metC, which is required for methionine biosynthesis. This is in good agreement with the map position of the E. coli K1 capsule genes predicted by Vimr (1991).

5.2.4 Mapping of the region 3 promoter by RNase protection analysis

To locate the region 3 promoter an RNase protection assay was performed. This involved the generation of radiolabelled RNAs of the opposite sense to the region 3 transcript. The antisense RNAs were synthesised in vitro using T7 RNA polymerase and [α-32P] rUTP (section 2.11.1) and hybridised to total RNA from MS101 grown at 37°C. RNA hybrids were then digested with a ribonuclease which only cleaves single stranded RNA and the products analysed by polyacrylamide gel electrophoresis (section 2.11.2). The size of the protected RNA hybrid reflects the extent of the transcript which is recognised by the probe.
Figure 5.4. Nucleotide sequence upstream of region 3 of the *E. coli* K5 *kps* locus.
The nucleotide sequence from 1 to 191 was also reported by Smith et al. (1990). Restriction endonuclease cleavage sites are shown above the sequence and protein translation is shown below in single letter code. The site of initiation of region 3 transcripts is marked with an arrow and the putative -10 sequence is underlined. Residues 1074 to 1248 are 97% identical to DNA in the 65 minute region of the E. coli K-12 chromosome (accession number U28377). The end of the capsule gene cluster is marked with an asterisk.
To generate antisense RNA probes for RNase protection assays, a 2.5 kb ClaI fragment from pCP1 including kpsM and about 2 kb of upstream sequence (Figure 1.2) was first cloned into AccI digested M13mp19, generating pMS18. A 2.3 kb FspI-SmaI fragment from pMS18, including 120 bp of kpsM, was then subcloned into HincII cut pTZ18R, generating pMS23 (Figure 5.5). Recombinants were selected by blue/white screening on medium containing X-gal and IPTG. Transcription from the T7 promoter of pMS23 generates an RNA of the opposite sense to the region 3 transcript.

![Figure 5.5. Map of restriction endonuclease cleavage sites in pMS23. Arrows indicate the direction of transcription. The bold line denotes the extent of cloned DNA. Plasmid pMS26 is identical to pMS23 except for a deletion between the HincII and SmaI sites. Restriction endonuclease cleavage sites are abbreviated as follows; Hc, HincII; Sm, SmaI; V, VspI; X, XmnI.](image)

Plasmid pMS23 was digested with HincII and the linear template used for in vitro synthesis of a run-off 467 nucleotide antisense RNA which extends 306 nucleotides 5' of the translation initiation codon for kpsM. After hybridisation of this probe to total RNA from MSlOl and RNase digestion, polyacrylamide gel electrophoresis revealed that the entire complementary region of the probe was protected (data not shown). The probe was reduced in length by 38 nucleotides because polylinker sequences (between the T7 promoter and the region 3 insert) are transcribed from pMS23 but are not present on region 3 transcripts. This indicates that the region 3 promoter must be at least 306 bp from the start of kpsM. It was therefore decided to generate an RNA probe complementary to sequences further upstream of kpsM.

To do this, a deletion was made in pMS23 between the HincII site 5' of region 3 and a SmaI site in the polylinker, generating pMS26 (Figure 5.5). This deletion brings
the sequence upstream of the HincII site next to the T7 promoter. Plasmid pMS26 was digested with XmnI and a 2.2 kb fragment used for the in vitro synthesis of a 382 nucleotide antisense RNA. Accounting for non-complementary sequences encoded by the polylinker, the entire probe was protected (data not shown). This indicates that the promoter for region 3 must be at least 668 bp upstream of the start of kpsM.

A 538 nucleotide antisense RNA, extending a further 156 bp upstream of the XmnI site, was generated using a 0.75 kb VspI fragment of pMS26. Total RNA from MS101 protected a fragment approximately 70 nucleotides smaller than the 538 nucleotide probe (Figure 5.6). Because the probe contains 21 nucleotides not complementary to region 3, this indicates that the 5' end of the region 3 transcripts must be approximately 50 bp from the VspI site, or about 770 bp upstream of the translation initiation codon for kpsM. The smaller protected fragments may be due to breathing at the ends of the RNA hybrid and/or to degradation of the probe or target RNA. The exact size of the largest protected fragment cannot be accurately determined as the DNA sequence was poorly resolved and because RNA:RNA hybrids do not have the same electrophoretic mobility as single stranded DNA molecules of the same size. However, this approach allows the design of appropriate oligonucleotides to precisely locate the site of transcription initiation by primer extension.

5.2.5 Determination of the 5' end of region 3 transcripts by primer extension

To precisely locate the site of initiation of region 3 transcripts primer extension analysis was used. A 32P-end labelled primer, MS17 (5'-GTCGCACCAATAATCCCG-3'), was annealed to region 3 transcripts approximately 230 bp from the start site predicted by RNase protection analysis and extended using AMV reverse transcriptase (section 2.12). Repeated attempts to map the region 3 transcript using RNA from MS101 and MS106 were unsuccessful, therefore total RNA (50μg) from LE392 (pCP1) and 1.8 (rfaH) (pCP1) was used. The products were analysed by polyacrylamide gel electrophoresis against a DNA sequence ladder obtained with the same primer and single stranded DNA from plasmid pMS17 (section 5.2.3) (Figure 5.7).

The 5' end of region 3 transcripts is 741 bp upstream of the translation initiation codon for kpsM. Region 3 transcripts initiate from the same site in an rfaH mutant, indicating that alternative promoters are not involved in the regulation of region 2/3 transcription by RfaH. No primer extension products were obtained using RNA from LE392 or 1.8 containing pTZ19R, which lacks the region 3 insert (data not shown). The discrepancy in the site of initiation of region 3 transcripts predicted by RNase protection analysis (c. 770 bp 5' of kpsM) and primer extension (741 bp 5' of kpsM) is probably due to the difficulty in sizing the protected RNA hybrid for the reasons indicated in section 5.2.4.
Figure 5.6. RNase protection analysis of region 3 transcripts. The full length probe (track 5) was generated using a 0.75 kb VspI fragment of pMS26. Protected hybrids with RNA from MS101 and MS106 were run in tracks 1 and 2, respectively. The probe did not protect RNA from PA360, which lacks the capsule genes (track 3), and was completely degraded by RNase in the absence of RNA (track 4). The sequencing ladder was generated using pMS17 single stranded DNA (section 5.2.3) and primer MS13. The ratio of ddNTPs to dNTPs in the termination reaction was reduced 2 fold to improve the sequence distal to the primer.
Figure 5.7. Mapping of the site of initiation of region 3 transcripts by primer extension. Tracks 1 and 2 contain the products of extension reactions with primer MS17 and RNA from LE392 (pCP1) and 1.8 (rfaH) (pCP1), respectively. Tracks A, T, C and G show the result of sequencing reactions using the same primer and single stranded pMS17 DNA (section 5.2.3). The 5' end of the transcript is marked with an asterisk.
In *E. coli* K1, the 5' end of region 3 transcripts has been mapped to 743 bp upstream of the translation initiation codon for *kpsM* (Q. Zhao and R. Silver, unpublished results). There are two additional nucleotides in the region between *kpsM* and the promoter in *E. coli* K1 that are not present in *E. coli* K5. It follows that the region 3 transcripts initiate at precisely the same location in these bacteria. In *E. coli* K1 the predicted σ^70^-10 sequence for the region 3 promoter is identical to that in *E. coli* K5, but the predicted first nucleotide of the transcript is cytosine and not uracil.

5.3 Discussion

The notion that RfaH causes transcription antitermination can be used to explain the pattern of transcription of the capsule genes in an *rfaH* mutant. It is proposed that RfaH acts to allow transcripts that would normally terminate after region 3 to continue into region 2. Thus, in the absence of RfaH, region 2 transcription is reduced, whereas transcription of region 3 is less affected. A number of observations made in this section support this model. Firstly, the RfaH-dependent transcription of *kfiA* does not involve the *kfiA* promoter since, a *kfiA::lacZ* fusion on pMS7 is transcribed at the same level in wild-type and *rfaH* mutant strains. Thus, the expression of *kfiA* may rely on readthrough transcription from region 3, where RfaH is needed for this to occur. Each of the region 2 promoters has been cloned in a cat promoter probe vector (Petit et al., 1995), and it would be interesting to compare the activity of these promoters in wild-type and *rfaH* mutant strains. If it can be shown that the region 2 promoters are active in an *rfaH* mutant, this may suggest that all of the region 2 genes are expressed by readthrough transcription.

The existence of a readthrough transcript was demonstrated by reverse transcriptase-PCR using RNA from MS101. A control reaction to which no reverse transcriptase was added did not give a product, indicating that contaminating MS101 chromosomal DNA was not present. It was found that the region between *kpsT* and *kfiA* is also transcribed in an *rfaH* mutant (MS106), however the assay used is not quantitative, and this result does not imply that readthrough transcription occurs at the same frequency in an *rfaH* mutant. To quantitate the abundance of the readthrough transcripts in MS101 and MS106, an RT-PCR assay using radioactively labelled primers or dNTPs could be used (Wang et al., 1989).

It has been proposed that RfaH may act by modifying RNA polymerase such that it fails to recognise downstream termination signals (Farewell et al., 1991). The isolation of extragenic suppressors of *rfaH* phenotypes in the genes for termination factor Rho and subunits of RNA polymerase provides indirect support for this notion (Beutin et al., 1981; Farewell et al., 1991). RfaH may therefore act in an analogous way
to the bacteriophage \( \lambda \) N and Q proteins (section 1.7.1.5) (Greenblatt et al., 1993; Roberts, 1993; Friedman and Court, 1995). Antitermination by N requires a number of host proteins (NusA, NusB, NusG and ribosomal protein S10), as well as a short sequence \( nut \), in the nascent RNA (Mason et al., 1992; DeVito and Das, 1994). It is believed that N interacts with RNA polymerase as it transcribes the \( nut \) site and catalyses the formation of a termination resistant elongation complex which is stabilised by the Nus factors (Das, 1992; DeVito and Das, 1994). Modification of RNA polymerase by the Q protein does not require host factors, although it is enhanced by NusA (Grayhack et al., 1985), and occurs while the polymerase is paused at a short DNA sequence \( qut \) that overlaps the late gene promoter (Yarnell and Roberts, 1992).

It has been proposed that both N and Q act by increasing the elongation rate of RNA polymerase (Jin et al., 1992). This may facilitate antitermination at Rho-dependent terminators by causing RNA polymerase to pass through the terminator so quickly that it does not interact with Rho or other termination factors (Jin et al., 1992).

A requirement of the antitermination model is that the region 2 promoters are so weak that, in the absence of readthrough transcription, insufficient region 2 proteins are made for polysaccharide biosynthesis to occur. Promoter probe data for the region 2 promoters appears to support this notion (Petit et al., 1995). The observation that KfiD is not made in an rfaH mutant (section 4.2.6), even though \( kfiA-D \) transcripts are still made (section 5.2.1) also indicates that the \( kfiA \) promoter is weak. However, the observation that the \( kfiA::luxAB \) fusion is transcribed at approximately double the level of the \( kpsT::luxAB \) fusion at 37°C (section 4.2.4) indicates that \( kfiA \) promoter is at least as strong as the region 3 promoter. The \( kfiA::luxAB \) fusion therefore measures total \( kfiA \) transcription, arising from both readthrough transcription and its own promoter.

The fact that the region 2 genes have promoters of their own makes interpretation of the action of RfaH difficult. The situation is analogous in the regulation of the F factor \( tra \) operon by RfaH. Hybridisation and gene fusion studies have indicated that the \( traS, traT \) and \( traD \) genes rely on RfaH for their expression (section 3.3) (Beutin et al., 1981; Gaffney et al., 1983), and yet each of these genes is preceded by a promoter (Ham et al., 1989). It has been suggested that RfaH serves to enhance the expression of the \( traS, traT \) and \( traD \) genes by preventing the termination of transcripts from the major \( tra \) promoter (\( P_\gamma \)) at sites upstream of these genes (Frost et al., 1994). Evidence for this comes from the observation that deletion of the \( P_\gamma \) promoter significantly decreases production of the TraD protein (Maneevanakul et al., 1992). It may be that RfaH has a similar role in regulation of the K5 capsule gene cluster, serving to boost the expression of region 2 genes to a level at which efficient polysaccharide biosynthesis can occur. The importance of readthrough transcription in the expression of region 2 of the K5 capsule gene cluster could be determined by
deleting the region 3 promoter in MS101 and assaying the effect this has on the transcription of region 2 genes.

Although RfaH may activate transcription of the region 2 genes by preventing the premature termination of region 3 transcripts, it also activates the transcription of kpsT (section 4.2.5). The situation is similar in the tra, rfa and hly operons, where RfaH increases the transcription of genes which precede putative terminators (Gaffney et al., 1983; Brazas et al., 1991; Pradel and Schnaitman, 1991; Bailey et al., 1992). It is difficult to see how RfaH affects transcription initiation as it lacks known DNA binding motifs and is not homologous to histone-like proteins which activate transcription by altering DNA topology (Bailey et al., 1992). In the rfa operon, it has been shown that the extent to which genes rely on RfaH for their expression, increases with their distance from the promoter (Brazas et al., 1991; Pradel and Schnaitman, 1991). Thus, RfaH may in some way stabilise the transcription complex so that it can continue to the end of the operon (Farewell et al., 1991). It is possible that RfaH is required to suppress pausing of RNA polymerase at sites between the region 3 promoter and kpsT. Therefore, in the absence of RfaH fewer transcripts reach kpsT than in the wild-type.

It is arguable that RfaH will not regulate the expression of all group II K antigens to the same extent. This is because the region 2 genes from different serotypes may be transcribed in the opposite direction to region 3 or may have strong promoters of their own. If this were the case, the region 2 genes would not rely on antitermination for their expression. It has been shown that multiple copies of the region 2 genes can restore capsule production in an rfaH mutant (section 4.2.7), suggesting that if the region 2 genes are expressed at a high enough level, the need for RfaH is diminished. It is striking however, that in both E. coli K1 and K5, all of the region 2 genes are transcribed in the same direction as region 3 (Silver, 1994). The capsule biosynthesis genes are also transcribed in the same direction in N. meningitidis (Edwards et al., 1994), and H. influenzae (Van Eldere et al., 1995). It would be interesting to determine if the regulation of K1 capsule expression by RfaH is also mediated by antitermination and readthrough transcription of the region 2 genes. This could be done by comparing the hybridisation of region 2 and 3 specific probes to total RNA from wild-type and rfaH mutant strains containing pKT274.

It may also be predicted that for RfaH to regulate region 2 transcription in different group II strains the terminator at the end of region 3 must be conserved. The TERMINATOR program (Devereux et al., 1984) was used to search the region between kpsT and kfiA for Rho-independent terminators. In general, Rho-independent terminators comprise a G+C-rich inverted repeat, capable of forming a stem-loop structure in the RNA, followed by a run of thymidine residues on the non-template strand (Holmes et al., 1983). Two putative Rho-independent terminators were
identified 154 and 181 nucleotides 3’ of kpsT in E. coli K5, however a large number of other stem-loops followed by a thymidine-rich region were also identified and it is difficult to assess the importance of these given that the region between kpsT and kfiA is so A+T-rich (%A+T= 67.3). The putative Rho-independent terminators 3’ of kpsT in E. coli K5 are not present at the end of region 3 in E. coli K1 (Pavelka et al., 1991). In fact there is only 50 bp between kpsT and the first gene of region 2 (neuD) (Figure 1.3) (Anunziato et al., 1995), and there are no obvious stem-loop structures in this region that may correspond to a Rho-independent terminator. In addition, there is a marked divergence in the nucleotide sequence at the 3’ end of kpsT in E. coli K1 and K5 (Drake, 1991).

It is possible that the terminator at the end of region 3 is Rho-dependent. A conserved feature common to all Rho-dependent terminators in E. coli so far studied has recently been identified. This comprises a G+C-rich region in the RNA of at least 78 nucleotides in which the content of cytosine is higher than that of guanosine (a C>G-rich bubble) (Alifano et al., 1991). It has been shown that this consensus motif can cause Rho-dependent termination (in the absence of translation) if placed within a gene (Rivellini et al., 1991). Computer programs have been developed to search nucleotide sequences for potential Rho-dependent terminators (Alifano et al., 1991; Penfold et al., 1994) and could be used to search the region between kpsT and kfiA.

Because of the possible importance of antitermination in the regulation of region 2 transcription, the site of initiation of transcripts upstream of region 3 was determined in wild-type and rfaH mutant strains. RNase protection analysis was used to locate the promoter and the 5’ end of region 3 transcripts mapped to 741 bp upstream of the translation initiation codon for kpsM by primer extension. The transcripts were found to initiate at the same site in an rfaH mutant, indicating that alternative promoters are not involved in the regulation of region 2/3 transcription by RfaH. Region 3 transcripts also initiate at this site in E. coli K1, accounting for two nucleotides that are not present between kpsM and the region 3 promoter in E. coli K5. To confirm that the cDNA generated using primer MS17 reflects the genuine end of the transcript and is not due to stalling of the reverse transcriptase at secondary structures in the mRNA, a reaction should be performed with a different primer. This would also rule out any ambiguity in the exact 5’ end of the transcript due to gel migration artefacts.

Homology to the E. coli ς70 -10 consensus sequence was identified upstream of the 5’ end of region 3 transcripts in E. coli K5, but the -35 motif was absent. It is known that promoters with only a -10 sequence can still function successfully as sites for transcription initiation (Lisser and Margalit, 1993). To determine if the promoters for regions 1 and 3 of the E. coli K5 capsule gene cluster possess common motifs that may be important for temperature regulation, the nucleotide sequence upstream of regions
1 and 3 was compared using the FASTA program. No significant homology was detected.

In the process of mapping the region 3 promoter approximately 1.2 kb of sequence upstream of *kpsM* was determined. Sequences homologous to DNA in the 65 minute region of the *E. coli* K-12 chromosome were detected about 1 kb upstream of *kpsM*, marking the end of the K5 capsule gene cluster. The %G+C for the sequence upstream of region 3 to the end of *kps* is 35.9% whereas the %G+C of the 175 nucleotides in the 65 minute region is 49.1%, which is typical for *E. coli* (Ørskov, 1984). Low G+C ratios have been reported for the capsule genes of other Gram negative bacteria (Frosch *et al.*, 1991; Arakawa *et al.*, 1995; Van Eldere *et al.*, 1995), and it is possible that this reflects a common ancestry (section 1.5.7). The finding that the K1 and K5 capsule gene clusters end at the same nucleotide, lends support to notion that they have a common evolutionary origin and diverged by internal variation (section 1.5.7) (Boulnois and Jann, 1989; Frosch *et al.*, 1991). The sequence upstream of region 3 in *E. coli* K1 and K5 contains no open reading frames and is not significantly homologous to any other sequences in the database.

The data accrued so far do not provide direct evidence for antitermination. This could be obtained by Northern blotting of total RNA from MS101 and MS106 using a region 3 probe. It would be predicted that in MS101 two transcripts would hybridise to a region 3 probe; a readthrough transcript and a region 3 only transcript. Because the location of the region 3 promoter is known, the size of the larger transcript would indicate how far the readthrough transcript extends across region 2. In MS106, it would be predicted that the region 3 only transcript would predominate, with only a small proportion of transcripts, if any, continuing into region 2. The large size of the readthrough transcript (c. 10 kb if the transcript ends 3' of *kfd*), may render its detection by Northern blotting difficult.

Because the capsule genes appear to be weakly transcribed, evidence for antitermination could also be obtained in an RNase protection assay, which is more sensitive than a Northern blot for the detection of mRNA (Sambrook *et al.*, 1989). An antisense RNA probe spanning the terminator would be predicted to give two protected fragments with RNA from MS101; a large fragment, protected by the readthrough transcript, and a smaller fragment, protected by the terminated RNA. Using RNA from MS106, the major protected species would be predicted to be the smaller fragment, owing to the termination of region 3 transcripts in the absence of RfaH. Although this approach may define the precise location of the terminator, the extent of the readthrough transcript cannot be determined.

Another explanation for the reduction in region 2 transcription in an *rfaH* mutant is that RfaH is required for the stability of region 2 transcripts. To address this possibility, the stability of region 2 transcripts should be compared in wild-type and
rfaH mutant strains. This could be achieved by hybridising a radiolabelled region 2 probe to total RNA from MS101 and MS106 extracted at intervals after arresting transcription with rifampicin. The rate of decay of region 2 transcripts following the addition of rifampicin will be a function of their stability and could be determined by densitometry. This analysis is complicated by the existence of three overlapping region 2 transcripts and a readthrough transcript from region 3. There is no evidence that RfaH is required for the stability of transcripts in the tra, rfa and hly operons.
Chapter 6

Involvement of \textit{cis}-acting sequences in regulation of the K5 capsule gene cluster

6.1 Introduction

Recently, a conserved DNA sequence has been identified in the non-coding region upstream of several gene clusters in enteric bacteria for the biosynthesis of polysaccharides (Hobbs and Reeves, 1994; Arakawa et al., 1995). This sequence (JUMPstart, Just Upstream of Many Polysaccharide-associated gene \textit{starts}), is present in the gene clusters for the production of \textit{E. coli} group II capsules (\textit{kps}), the Vi antigen of \textit{Salmonella} (\textit{viaB}), and the outer core (\textit{rfa}) and O antigen (\textit{rfb}) components of LPS. The JUMPstart sequence is 39 bp long, G+C-rich, and contains a 6 bp direct repeat. In \textit{E. coli} K1 and K5 it is located upstream of region 3. The high degree of conservation of the JUMPstart sequence in the non-coding regions of otherwise quite dissimilar genes implies that it has an important function. As the JUMPstart sequence is often located in the promoter regions of polysaccharide biosynthesis genes, it was postulated by Hobbs and Reeves (1994), that it may be involved in the regulation of transcription. It was also suggested that the JUMPstart sequence may be the recognition sequence of a site-specific recombinase, allowing the exchange of genes for polysaccharide biosynthesis and export between different bacteria. In this section, the role of the JUMPstart sequence in expression of the K5 antigen is addressed.

6.2 Results

6.2.1 The JUMPstart sequence is conserved upstream of RfaH-regulated operons

A search of all nucleotide sequences in the database using a 12 bp motif that is almost absolutely conserved among JUMPstart sequences (GGCGGTTAGCGT), was performed using the FINDPATTERNS program (Devereux \textit{et al.}, 1984), allowing one mismatch. This search identified homology to sequences that are found in the upstream non-coding regions of operons that are regulated by RfaH, that is, \textit{hly}, \textit{tra}, \textit{rfa} and \textit{kps}. An alignment of these sequences, generated using the CLUSTAL-V program (Higgins \textit{et al.}, 1992), is shown in Figure 6.1.
Figure 6.1. Alignment of sequences upstream of RfaH-regulated operons. The numbers on the right hand side refer to the distance, in base pairs, between the last residue shown and the site of translation initiation. Sequences were extracted from GenBank/EMBL entries with accession numbers X53819, U05251, M80599, X07565, and U01159 respectively. Gaps (.) have been introduced to improve the alignment. The consensus sequence is shown in bold. Conserved residues are shown in upper case, residues in lower case represent conservation in four out of the five aligned sequences. Part of the JUMPstart sequence derived by Hobbs and Reeves (1994), is shown for comparison (W = A or T; R = A or G; N = any nucleotide).

The FINDPATTERNS search also identified a second JUMPstart sequence in the K5 capsule gene cluster. This sequence is located 620 bp upstream of the translation initiation codon for kfiC in region 2, but is only weakly homologous to the consensus sequence obtained by aligning the JUMPstart sequences from RfaH-regulated operons on either side of the 12 bp GGGCGGTAAGGTT motif (Figure 6.2). The orientation of this sequence relative to the downstream genes is the same as the JUMPstart sequence upstream of region 3.

Figure 6.2. Alignment of the JUMPstart sequence located 5' of kfiC with the consensus obtained by aligning the sequences 5' of RfaH-regulated operons.

The finding that the JUMPstart sequence is conserved upstream of operons that are regulated by RfaH suggests that it may be involved in the mechanism by which RfaH acts. In support of this notion, the JUMPstart sequence is located in a region upstream of the rfa operon that is needed for RfaH to activate rfa transcription (Brazas et al., 1991). This region is highly conserved between the rfa operons of E. coli and S. typhimurium (Klena et al., 1993). In addition, the JUMPstart sequence in plasmid borne hly operons is located in hlyR, a 600 bp sequence that activates transcription of the entire hly operon and is required for antitermination (section 3.3) (Vogel et al., 1988; Koronakis et al., 1989; Jubete et al., 1995). The alignment of the sequences found
upstream of RfaH-regulated operons redefines the JUMPstart sequence and identifies nucleotides that may be involved in the regulation of the downstream genes by RfaH. The implication is that genes preceded by the JUMPstart sequence may share a common mode of regulation by RfaH.

The orientation of the JUMPstart sequence with respect to the downstream genes is the same in all instances. In all but the hly operon, the JUMPstart sequence is located downstream of the promoter. In the cases where the JUMPstart sequence is transcribed, it is a variable distance from the promoter, for example it is about 700 bp 3' of the region 3 promoter in kps, 85 bp 3' of the promoter in rfa and about 5.5 kb 3' of the major tra promoter P₁ (between traV and traR) (Figure 3.3). To determine if the JUMPstart sequence is important for the regulation of K5 capsule expression by RfaH, mutations were made in the JUMPstart sequence upstream of region 3 and transferred to the chromosome of MS101.

6.2.2 Insertion of an oligonucleotide linker in the JUMPstart sequence in MS101

In E. coli K5, the JUMPstart sequence upstream of region 3 contains a SgI site which is unique in region 3. A linker (30-mer) generated by annealing the complementary oligonucleotides MS9 (5'-CAAGTGATCGACAAGCTTGCTACGAGTCGA-3') and MS10 (5'-GGTGTGGAGTGGTAGGAAGGTTGTGGATGA-3') (section 2.6.4), was inserted into the unique SgI site in the JUMPstart sequence of pAS1, generating pMS11 (Figure 6.3). The linker has cohesive ends compatible with those of SgI and destroys the site on insertion. Thus, after ligation of the linker with SgI cut pAS1, the T4 DNA ligase was heat inactivated and the products digested with SgI to enrich for recombinants. The linker also contains a HindIII site and this was used to verify the insertion. The oligonucleotides used lacked 5' phosphate groups, therefore only single insertions should be possible. To confirm this, the nucleotide sequence of the insert in pMS11 was determined (Figure 6.4).

A 1.9 kb SacI-SphI fragment of pMS11 carrying region 3 and the linker insertion was subcloned into SacI-SphI digested pCVD442, generating plasmid pMS12 (Figure 6.3). Plasmid pMS12 was then introduced into MS101 from SM10kpir as described in section 4.2.1. To confirm integration of the plasmid at the correct site, chromosomal DNA from ampicillin resistant transconjugants was digested with BglII and Southern blots performed using the 1.8 kb SacI-SphI fragment from pAS1 as a probe. Plasmid pMS12 (8.1 kb) does not contain any BglII sites, therefore the native 1.8 kb BglII region 3 fragment (Figure 1.2) was found to shift in size to c. 9.9 kb on integration of the plasmid (Figure 6.5).
Figure 6.3. Construction of plasmids pMS11 and pMS12. The JUMPstart sequence is represented by a black box and the linker by a clear box. Restriction endonuclease cleavage sites are abbreviated as follows; H, HindIII; Sc, SacI; SphI; St, Styl. Plasmids are not drawn to scale.
Figure 6.4. Verification of the JUMPstart mutations by nucleotide sequencing. Nucleotide sequences were derived from pMS11 (linker insertion) and pMS14 (deletion) double stranded plasmid templates using primer MS13 (5'-GCCTCTTCGTATTTCTCG-3') which anneals in kpsM about 80 bp from the JUMPstart sequence. The initiation codon and putative Shine-Dalgarno sequence for kpsM are shown in bold (Smith et al., 1990). Residues of the JUMPstart sequence that are conserved between RfaH-regulated operons are underlined. Restriction endonuclease cleavage sites used to construct the mutations are also shown.

A double recombinant was selected by growing an MS101 merodiploid containing pMS12 correctly integrated in region 3 in the absence of ampicillin on medium containing sucrose. Isolates which retained the linker insertion were identified by colony PCR using an oligonucleotide from the linker (MS9) and primer kpsT(c) (5'-TAAATTACATTCAAATCTAAATCTGATACATTATACATC-3') which anneals at the 3' end of kpsT. Of 50 isolates screened, 3 gave a product of the expected size (c. 1.5 kb) (Figure 6.6).

The linker insertion in MS101 was confirmed by Southern blotting of HindIII digested chromosomal DNA using the 1.8 kb SacI-SphI fragment from pAS1 as a probe. The 7.5 kb HindIII fragment which spans region 3 in MS101 (Figure 1.2) was found to be replaced by fragments of 3 and 4.5 kb in isolates containing the linker insertion (MS136) (Figure 6.7). The 3 kb fragment hybridised weakly to the probe as it is only homologous over about 150 bp.
Figure 6.5. Southern blot analysis of BgIII digested chromosomal DNA from MS101 merodiploids containing pMS12 integrated in region 3 (tracks 1 to 4) and MS101 (track 5). A 1.8 kb SacI-SphI fragment from pAS1 was used as probe. Markers (kb) are HindIII fragments of λ DNA.
Figure 6.6. Identification of an MS101 mutant containing a linker insertion in the JUMPstart sequence. Colony PCR reactions with primers MS9 and kpsT(c) were performed using MS101 (track 1), a merodiploid of MS101 containing pMS12 integrated in region 3 (track 2), and MS101 containing a linker insertion in the JUMPstart sequence (MS136) (track 3). 1 kb DNA ladder was used for size markers.

Figure 6.7. Southern blot analysis of HindIII digested chromosomal DNA from MS101 (track 1) and MS101 derivatives containing a linker insertion in the JUMPstart sequence (MS136) (tracks 2 to 4). A 1.8 kb SacI-SphI fragment from pAS1 was used as probe. Markers (kb) are HindIII fragments of λ DNA.
6.2.3 Deletion of the JUMPstart sequence in MSlOl

To delete the JUMPstart sequence upstream of region 3 on the chromosome of MSlOl, a 2.2 kb XmnI-StuI region 3 fragment from pCP1 was first subcloned into HincII cut pUC19, generating pMS13. Recombinants were selected by blue/white screening on L-agar containing X-gal and IPTG. Plasmid pMS13 contains unique BsaBl and BstXI sites, which flank the JUMPstart sequence and are separated by 49 bp (Figure 6.8). Plasmid pMS13 was digested with BsaBl (at 60°C), then phenol:chloroform extracted and digested with BstXI (at 55°C). After a second phenol:chloroform extraction, the 3' protruding end left by BstXI was removed using bacteriophage T4 DNA polymerase (section 2.6.3). After heat inactivation of the enzyme, the blunt ends were ligated with T4 DNA ligase. The deletion between the BsaBl and BstXI sites removes the Styl site, which is unique in pMS13. Following ligation therefore, the T4 DNA ligase was heat inactivated and the products digested with StyI to enrich for plasmids which had lost the JUMPstart sequence. A derivative of pMS13 containing the JUMPstart deletion was identified (pMS14), and the deletion confirmed by sequencing (Figure 6.4).

A 1.8 kb BglII fragment containing the JUMPstart deletion from pMS14 was then cloned into BamHI cut pTZ18R to provide suitable flanking sites for cloning the insert into pCVD442 (Figure 6.8). This plasmid (pMS24) was digested with SacI and SalI and the 1.8 kb region 3 fragment cloned into SacI-SalI digested pCVD442, generating pMS25 (Figure 6.8). Plasmid pMS25 was then introduced into MSlOl from SM100pir by conjugation. As pMS25 (8.0 kb) does not contain any BglII sites, the native 1.8 kb BglII region 3 fragment (Figure 1.2) was predicted to shift in size to c. 9.8 kb on integration of the plasmid. This was confirmed by Southern blotting of BglII digested chromosomal DNA from ampicillin resistant transconjugants using the 1.8 kb SacI-Spal fragment from pAS1 as a probe (Figure 6.9).

Following sucrose selection with an appropriate merodiploid, ampicillin sensitive isolates which retained the JUMPstart deletion were identified by colony PCR using primer kpsT(c) which anneals at the 3' end of kpsT and a primer which anneals c. 200 bp 5' of the JUMPstart sequence (MS14, 5'-GGCCACTTTTCTGAAATATTACCCAG-3'). To facilitate the identification of isolates containing the JUMPstart deletion, each product was digested with StyI. For isolates containing the JUMPstart deletion, the 1.7 kb PCR product did not cut with StyI, whereas isolates in which the entire pMS25 insertion was lost, gave products of 0.2 and 1.5 kb. Of 50 isolates screened, only 1 contained the JUMPstart deletion (MS137) (Figure 6.10). The low frequency of isolation of JUMPstart mutants following sucrose selection may be because there is
Figure 6.8. Construction of plasmids pMS14, pMS24 and pMS25. The JUMPstart sequence is represented by a black box. The deletion between the BsaBI-BstXI sites of the JUMPstart sequence is shown as Δ. Restriction endonuclease cleavage sites are abbreviated as follows; Bg, BgIII; Bs, BsaBI; Bx, BstXI; Sa, Sall; Sc, SacI; St, StyI. Plasmids are not drawn to scale.
Figure 6.9. Southern blot analysis of *BgIII* digested chromosomal DNA from MS101 (track 1), and MS101 merodiploids containing pMS25 integrated in region 3 (tracks 2 to 5). A 1.8 kb *Sacl-SphI* fragment from pAS1 was used as probe. Markers (kb) are *HindIII* fragments of λ DNA.
Figure 6.10. Identification of an MS101 mutant containing a JUMPstart deletion. Colony PCR products generated with primers MS14 and kpsT(c) and digested with StyI were obtained from MS101 (track 1), a merodiploid of MS101 containing pMS25 integrated in region 3 (track 2), and MS101 containing a JUMPstart deletion (MS137) (track 3). 1 kb DNA ladder was used for size markers.
only about 200 bp on the 5' side of the mutation over which the second recombination event needed for successful allelic exchange can take place (Figure 4.1).

The JUMPstart deletion in MS101 was confirmed by Southern blotting of Styl digested chromosomal DNA using the 1.8 kb SacI-SphI region 3 fragment from pAS1 as a probe. In MS101 this probe hybridised to Styl fragments of 4.8 and 6 kb whereas for MS137, a 10.8 kb Styl fragment was detected (Figure 6.1). The BglII, ClaI and HindIII fragments spanning region 3 in MS137 appeared to be the same as those of MS101, indicating that no gross rearrangements of the DNA in this region had occurred.

6.2.4 JUMPstart mutations affect biosynthesis of the K5 antigen

The linker insertion in the JUMPstart sequence did not affect the sensitivity of MS101 to K5 phage, with MS136 giving plaques of the same morphology as MS101. MS137 however, gave very faint turbid plaques with the K5 phage. Membranes were prepared from these mutants and the \textit{in vitro} UDP-glucuronic acid transferase activity assayed. The linker insertion in the JUMPstart sequence reduced the \textit{in vitro} membrane transferase activity of MS101 by about 4.5 fold (Table 6.1). Thus, biosynthesis of the K5 antigen is impaired by the linker insertion, even though the sensitivity of MS136 to K5 phage is not affected. The JUMPstart deletion reduced the \textit{in vitro} transferase activity of MS101 to a level comparable to that of the acapsular strain PA360.

**Table 6.1. UDP-Glucuronic acid transferase activities of JUMPstart mutants of MS101.**

<table>
<thead>
<tr>
<th>Strain</th>
<th>JUMPstart mutation</th>
<th>( \mu \text{mol} \ ^{14}\text{C} \text{UDP-GlcA} ) incorporated per mg protein (in 30 min. @ 37°C)(^ \dagger )</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA360</td>
<td></td>
<td>0.28 ± 0.06</td>
</tr>
<tr>
<td>MS101</td>
<td></td>
<td>5.96 ± 0.39</td>
</tr>
<tr>
<td>MS136</td>
<td>Linker insertion</td>
<td>1.31 ± 0.16</td>
</tr>
<tr>
<td>MS137</td>
<td>Deletion</td>
<td>0.16 ± 0.04</td>
</tr>
</tbody>
</table>

\( \dagger \) Mean (± standard error of the mean) of 6 measurements using membranes from 2 separate cultures.
Figure 6.11. Southern blot analysis of chromosomal DNA from MS101 digested with BglII, ClaI, HindIII and Styl (tracks 1, 2, 3 and 4, respectively), and chromosomal DNA from MS101 containing a JUMPstart deletion (MS137) digested with BglII, ClaI, HindIII and Styl (tracks 5, 6, 7 and 8, respectively). A 1.8 kb SacI-SphI fragment from pAS1 was used as probe. Markers (kb) are HindIII fragments of λ DNA.
6.2.5 Transcription of kfiA requires the JUMPstart sequence

To test the hypothesis that the JUMPstart sequence is required for the action of RfaH, the transcription of kfiA was analysed in the JUMPstart mutants of MS101. Total RNA from MS136 and MS137 was immobilised on a nylon membrane and hybridised to a 32P-labelled kfiA PCR product. RNA from MS101, MS106 (rfaH) and PA360 was also loaded on the membrane for comparison. The relative amounts of RNA bound by the probe were quantitated using a PhosphorImager. Transcription of kfiA was found to be reduced by about 3 fold by the linker insertion. No signals were obtained with RNA from MS106 or MS137 (Figure 6.12). This effect was confirmed by Western blotting using KfiA-specific antibody (data not shown). To check that the same amount of RNA was loaded from each strain, the membrane was stripped and re-hybridised using a kpsT PCR product as a probe. Signals of approximately equal intensity were obtained from all strains, indicating that the JUMPstart mutations do not affect the transcription of region 3 (Figure 6.12).

To confirm that the JUMPstart sequence is required for kfiA transcription, the effect of the JUMPstart mutations on the transcription of a plasmid borne β-galactosidase transcriptional fusion in kfiA was determined. A kfiA::lacZ fusion was made in plasmid pCP1, which contains kfiA, region 3 and about 2 kb of upstream sequence. This was achieved by cloning the 4.7 kb PstI fragment from pKOK6 into a unique NsiI site in pCP1, generating pMS9 (Figure 6.13). Plasmid pMS9 contains a unique StyI site in the JUMPstart sequence and this was used for insertion of the linker formed from oligonucleotides MS9 and MS10 as in section 6.2.1, generating pMS10 (Figure 6.13). The JUMPstart deletion was transferred to pMS9 by replacing the 1.8 kb region 3 BglII fragment of pMS9 with the corresponding BglII fragment from pMS14 (section 6.2.3; Figure 6.8), generating pMS15 (Figure 6.13). These plasmids were transformed into E. coli JM101 and the effect of the JUMPstart mutations on kfiA transcription determined by measuring β-galactosidase activity. For comparison, the effect of an rfaH mutation on kfiA transcription was determined by transforming pMS9 into MS135 (JM101 rfaH).
Figure 6.12. Direct detection of kfiA transcripts in MS101 JUMPstart mutants by hybridisation with a $^{32}$P-labelled DNA probe. Total RNA (50µg) from MS101 (slot 1), MS106 (slot 2), MS136 (slot 3), MS137 (slot 4), and PA360 (slot 5) was hybridised to a $^{32}$P-labelled kfiA PCR product (panel A). To check the amount of RNA loaded from each strain, the membrane was stripped and re-hybridised with a kpsT PCR product (panel B). RNase treated samples did not hybridise to the probe (data not shown), indicating that the signals are not due to contaminating DNA.
Figure 6.13. Construction of plasmids pMS9, pMS10 and pMS15. The JUMPstart sequence is represented by a black box and the lacZ and kanamycin resistance (KmR) genes by clear boxes. Restriction endonuclease cleavage sites are abbreviated as follows; B, BglII; N, NsiI; Ps, PstI; S, Styl. Plasmids are not drawn to scale.
Transcription of the $kfiA::lacZ$ fusion on pMS9 was reduced by only 2.5 fold in an $rfaH$ mutant (Table 6.2). This compares to a reduction of about 70 fold for the chromosomal $kfiA::luxAB$ fusion in an $rfaH$ mutant (section 4.2.5). Plasmid pMS15, which is the same as pMS9 but lacks the JUMPstart sequence, showed a 2.2 fold reduction in transcription of the $kfiA::lacZ$ fusion (Table 6.2). The transcription of the $kfiA::lacZ$ fusion on plasmid pMS10, which contains a linker insertion in the JUMPstart sequence, was reduced by 1.7 fold. The linker insertion therefore has a less pronounced effect on $kfiA$ transcription than the JUMPstart deletion. Although the plasmid borne $kfiA$ gene exhibits weaker regulation by RfaH than the chromosomal gene, these data support the results obtained by direct detection of $kfiA$ transcripts in the MS101 JUMPstart mutants.

Table 6.2. Transcription of $kfiA::lacZ$ reporter fusions on plasmids with JUMPstart mutations.

<table>
<thead>
<tr>
<th>Strain</th>
<th>JUMPstart mutation</th>
<th>Units of $\beta$-galactosidase$^\dagger$</th>
</tr>
</thead>
<tbody>
<tr>
<td>JM101 (pMS9)</td>
<td></td>
<td>1583 ± 72</td>
</tr>
<tr>
<td>JM101 (pMS10)</td>
<td>Linker insertion</td>
<td>930 ± 60</td>
</tr>
<tr>
<td>JM101 (pMS15)</td>
<td>Deletion</td>
<td>709 ± 27</td>
</tr>
<tr>
<td>MS135 (pMS9)</td>
<td>JM101 $rfaH$</td>
<td>601 ± 84</td>
</tr>
</tbody>
</table>

$^\dagger$ Mean of at least three separate measurements (± the standard error of the mean).

6.2.6. Plasmid pAS30 restores capsule production in MS137

Because the transcription of $kfiA$ is reduced by mutations in the JUMPstart sequence, an attempt was made to restore capsule production MS137 by the introduction of the region 2 genes on a plasmid. Transformation of MS137 with pAS30 conferred sensitivity to K5-specific phage. Plasmid pAS21, which carries only $kfiA$, was unable to restore production of the K5 antigen in MS137. This indicates that the JUMPstart sequence is required for the expression of region 2 genes other than $kfiA$. The observation that capsule expression in MS137 can be restored by the region 2 genes on a plasmid also indicates that the JUMPstart deletion has not significantly affected the expression of the region 3 genes.
To determine if the JUMPstart sequence can be supplied in trans, an attempt was made to complement MS137 by introducing functional copies of the JUMPstart sequence on a plasmid. Neither plasmid pAS1 nor pCP1 were able to complement MS137, indicating that the JUMPstart sequence only acts in cis. The hlyR sequence of plasmid borne hly operons, which contains the JUMPstart sequence, also acts in cis (Vogel et al., 1988).

6.3 Discussion

A short sequence (JUMPstart), that is found upstream of many polysaccharide biosynthesis genes in enteric bacteria, was found to be conserved upstream of all known RfaH-regulated operons. This suggested that the JUMPstart sequence may be required for the action of RfaH. In support of this notion, the JUMPstart sequence is located in a region upstream of the E. coli and S. typhimurium rfa operons that is required for RfaH to activate rfa transcription (Brazas et al., 1991). In addition, in plasmid borne hly operons the JUMPstart sequence is located in a region (hlyR) that activates transcription of the hly operon and which is also required for antitermination (section 3.3) (Vogel et al., 1988; Koronakis et al., 1989; Jubete et al., 1995). The hlyR element contains several other motifs, including direct and inverted repeats and a putative IHF binding site, and it is possible that these are also required for its activity. An alignment of the sequences found upstream of RfaH-regulated operons redefines the JUMPstart sequence and identifies highly conserved nucleotides that may be important for its activity. A second JUMPstart sequence was identified in region 2 of the K5 capsule gene cluster, 620bp upstream of the translation initiation codon for kfc.

It has been proposed that RfaH may act in an analogous way to the N and Q proteins of bacteriophage λ (Beutin et al., 1981; Gaffney et al., 1983; Bailey et al., 1992; Schnaitman and Klena, 1993). Antitermination by N and Q requires cis-acting sequences (nut and qut, respectively) (Roberts, 1993; Greenblatt et al., 1993; Friedman and Court, 1995), and it is possible that the JUMPstart sequence has an analogous role in the action of RfaH. Thus, RfaH may modify RNA polymerase while it associated with the JUMPstart sequence such that it does not recognise downstream terminators. Assuming this to be the case, it would be unlikely that the JUMPstart sequence acts as a signal in the nascent RNA as with nut, as it is not transcribed in the hly operon. In the cases where the JUMPstart sequence is transcribed, it is a variable distance from the promoter. This contrasts with the qut site, which overlaps the late gene promoter (Yarnell and Roberts, 1992).
Sequences similar to the phage λ nut site are found in several E. coli operons. The leader regions of E. coli ribosomal RNA operons contain sequences similar to boxA of the nut site and it has been shown that these are required for antitermination (section 1.7.1.5) (Li et al., 1984; Berg et al., 1989; Albrechtsen et al., 1990; Heinrich et al., 1995). BoxA-like sequences have also been identified in the spacer regions of rrr operons (section 1.7.1.5) (Berg et al., 1989), and it has been suggested that these sequences may serve to ‘recharge’ the termination resistant state of the elongation complex so that it may continue to the end of the operon (Heinrich et al., 1995). It is possible that the second JUMPstart sequence located in region 2 upstream of kfiC has an analogous function, allowing readthrough transcripts to extend to the end of region 2.

A 14 bp consensus for the boxA sequences found in the nut sites of lambdoid phages and leader and spacer regions of E. coli rrr operons has been derived (Morgan, 1986). This consensus sequence was found to match the consensus JUMPstart sequence described by Hobbs and Reeves (1994) at 12 of the 14 positions, allowing for a gap of 4 bp (Figure 6.14). A similar number of mismatches occurs between the boxA elements of nut sites of some lambdoid phages (Campbell, 1994). No homology was detected between the JUMPstart sequence and boxB or the qut site.

| BoxA consensus | G C G A G A . . . A T T G T G | BoxA consensus
| JUMPstart consensus | G C G G G G A | A C C G A A N T G T C G |

Figure 6.14. Alignment of the consensus boxA and JUMPstart sequences. Only part of the consensus JUMPstart sequence is shown. A 4 bp gap (•) has been introduced into the boxA consensus sequence to improve the alignment. N= any nucleotide.

Whilst homology over such a short stretch cannot normally be considered as significant, it may be important given that these sequences have been implicated in antitermination. BoxA-like sequences are also required for antitermination in the E. coli tryptophanase (tna) operon (Stewart and Yanofsky, 1985), and have been identified in the leader regions of several attenuators (Friedman and Gottesman, 1983). The antitermination systems in E. coli and lambdoid phages may therefore have a common evolutionary origin.

To examine the role of the JUMPstart sequence upstream of region 3 in the regulation of K5 capsule expression, mutations were made in this sequence and transferred to the chromosome of MS101 using a positive-selection suicide vector. The insertion of an oligonucleotide linker into the JUMPstart sequence did not affect the
K5 phage sensitivity of MS101, but was found to reduce the in vitro UDP-glucuronic acid transferase activity by about 4.5 fold. A mutant in which the JUMPstart sequence was deleted gave very faint turbid plaques with the K5 phage and exhibited an in vitro transferase activity comparable to that of an acapsular strain. It was not possible to complement the MS101 mutant with the JUMPstart deletion by the introduction of a functional JUMPstart sequence on a plasmid. This indicates that the JUMPstart sequence is only active in cis.

By direct detection of RNA in slot blot hybridisations it was found that deletion of the JUMPstart sequence dramatically reduced the transcription of kfiA, but had little or no effect on the transcription of kpsT. This phenotype is the same as that of an rfaH mutation (section 4.2.5). The linker insertion in the JUMPstart sequence had a less pronounced effect on kfiA transcription. To confirm the effect of the JUMPstart mutations on kfiA transcription, the linker insertion and deletion mutations were transferred to a plasmid carrying a kfiA::lacZ transcriptional fusion. The JUMPstart mutations were found to reduce the transcription of kfiA to a similar level to that observed in an rfaH mutant. These data indicate that the JUMPstart sequence may be required for the action of RfaH. It would be interesting to determine if the JUMPstart sequence is required to prevent the termination of region 3 transcripts. This could be achieved by Northern blotting of RNA from MS136 and MS137 with a region 3 probe as described in section 5.3.

Transcription of the plasmid borne kfiA::lacZ fusion did not show as pronounced regulation by RfaH as the chromosomal gene. In an rfaH mutant transcription of the kfiA::lacZ fusion on pMS9 was reduced by 2.5 fold, whereas transcription of a chromosomal kfiA::luxAB was reduced by 70 fold (section 4.2.5). The disparity in the regulation of kfiA transcription by RfaH on plasmids and the chromosome may be explained in several ways. It is possible that the activity of the kfiA promoter is higher on plasmids than on the chromosome. This may be due to differences in the superhelicity of plasmid and chromosomal DNA. The transcription of kfiA on plasmids may also be boosted by vector promoters. In addition, the frequency of transcription from the region 3 promoter may be different on plasmids. Transcription frequency is known to affect readthrough transcription in other systems, such as the rplL-rpoB intergenic region (Steward and Linn, 1992). It is also possible that a factor exists which antagonises the effect of RfaH, (i. e., negatively regulates readthrough transcription) which is titrated out by multiple copies of the capsule genes. A plasmid borne reporter gene fusion in the S. typhimurium rfa operon has also been reported to exhibit weaker regulation by RfaH than the corresponding chromosomal gene (Brazas et al., 1991).

The alignment of JUMPstart sequences located upstream of RfaH-regulated operons (Figure 6.1) and polysaccharide biosynthesis genes (Hobbs and Reeves, 1994;
Arawaka et al., 1995), identified 12 contiguous residues (GGCGGTAGCGT) that are almost absolutely conserved. This motif may therefore be important for the activity of the JUMPstart sequence. The linker insertion, although roughly central to the JUMPstart sequence, is located to one side the 12 bp motif (Figure 6.4) and this may explain why its effect on expression of the K5 antigen and kfiA transcription was not as pronounced as the JUMPstart deletion. The linker insertion separates the 6 bp direct repeats of the JUMPstart sequence and is central to the boxA-like sequences, suggesting that these motifs, or their spatial organisation, may not be important for its activity.

To better define the nucleotides of the JUMPstart sequence required for its activity, a smaller deletion (25 bp) which removes the 12 bp motif and some of the unconserved region between the the JUMPstart sequence and the start of kpsM could be made by digesting pMS13 (or pMS24) with BsaBI and StyI, followed by end-repair of the StyI end with Klenow and ligation of the blunt ends. This deletion could be transferred to MS101 in the same way as the BsaBI-BstXI deletion (section 6.2.3). It would be also be interesting to change individual nucleotides of the 12 bp motif. This could be achieved by site-directed mutagenesis or by replacing the region between the the BsaBI and StyI sites with oligonucleotide linkers with defined mutations and ends for StyI and BsaBI. The effect of these mutations on kfiA transcription could be determined by cloning the mutations on the 1.8 kb BglII region 3 fragment into pMS9, which has a kfiA::lacZ fusion. To determine the effect of these mutations in MS101 however, there must be some way of selecting recombinants following sucrose selection. This may be possible if the nucleotide changes result in the introduction or removal of a restriction enzyme site, as double recombinants retaining the mutation could be identified by PCR and digestion of the product. These mutations could also be introduced into MS108 (kfiA::luxAB), to determine their effect on transcription of kfiA on the chromosome.

It is possible that the effect of the JUMPstart mutations on production of the K5 antigen may be due, in part, to altered translation of region 3 transcripts. The mutations are very close to the Shine-Dalgarno sequence and start codon for kpsM and it is possible that the mutations affect the accessibility of these sequences by altering the secondary structure of the mRNA. Further, the JUMPstart deletion removed 1 nucleotide from the 5' end of the putative Shine-Dalgarno sequence for kpsM described by Smith et al., (1990) (Figure 6.4). It is known that region 3 proteins are required for efficient polysaccharide biosynthesis in vitro, perhaps forming part of the transferase complex (Vimr et al., 1989; Bronner et al., 1993b). The observation that capsule production in strain MS137 can be restored by the region 2 genes on a plasmid suggests that the JUMPstart deletion has not significantly affected the production of the region 3 proteins. This should be confirmed by assaying the production of region 3 proteins in strains containing the JUMPstart mutations by Western blotting. In this
laboratory, antibody is only available against KpsT, however \textit{kpsM} and \textit{kpsT} are translationally coupled (Smith \textit{et al.}, 1990; Pavelka \textit{et al.}, 1991), therefore if the initiation of \textit{kpsM} translation is impaired, \textit{kpsT} translation will also be affected. Attempts to detect KpsT in MS101 and the JUMPstart mutants by Western blotting have been unsuccessful. This may be due to poor reactivity of the KpsT-specific antibody and/or to the low level of KpsT synthesis in these strains. It may be easier to compare KpsT production in strains carrying the JUMPstart mutations on a plasmid.
Chapter 7

Discussion

The regulation of group II capsule expression in *E. coli* has until now received little attention compared to group I K antigens, lipopolysaccharides and other bacterial polysaccharides, such as alginate. In this study, the role of known regulatory genes in production of the group II K5 antigen was investigated. An *E. coli* K5/K-12 hybrid strain was constructed so that expression of a single copy of the capsule genes could be studied in their normal chromosomal location. This was achieved by Hfr mating, making use of the close linkage of *kps* to *serA*, and obviates the problems associated with using plasmids to study gene regulation. The K5/K-12 hybrid (MS101) exhibited all the properties of wild-type *E. coli* K5, but could be transduced at high frequency.

Mutations in regulatory genes were introduced into MS101 using bacteriophage P1vir and their effect on capsule expression assayed by sensitivity to a K5-specific phage and by measuring the *in vitro* membrane transferase activity. Only one mutation had a detectable effect on capsule production. This was mutation in the *rfaH* gene, and it conferred resistance to K5-phage at 37°C. The *in vitro* UDP-glucuronic acid transferase activity of the *rfaH* mutant (MS106) was less than 10% that of MS101 and was comparable to that of an acapsular strain of *E. coli*. This indicated that polysaccharide biosynthesis is impaired by the *rfaH* mutation. The lack of intracellular polysaccharide detected in MS106 by immunoelectron microscopy supports this notion. Indirect effects of the *rfaH* mutation, for example on LPS structure, do not account for this phenotype. RfaH was also shown to be required for production of the K1 antigen, suggesting that it has a general role in the expression of group II capsules.

The *rcs* system, which regulates the expression of colanic acid, group I K antigens and diverse polysaccharides in other bacteria, was not required for production of the K5 antigen. This is in accord with the findings of Keenleyside *et al.*, (1992, 1993) who reported that the *in vitro* biosynthesis of the K1 and K5 polysaccharides is not affected by multicopy *rcsA* or *rcsB*. It has been reported that RcsA negatively regulates expression of the K54 antigen (Russo and Singh, 1993). The K54 antigen was claimed by the authors to be a group II capsule, however it is made at all growth temperatures and genetic data indicate that it belongs to group III (Pearce and Roberts, 1995). The role of RcsA as a negative regulator of group III capsule expression is in contrast to its role as a positive regulator in all other systems (Stout *et al.*, 1991; McCallum and Whitfield, 1991; Keenleyside *et al.*, 1992). It would be interesting to determine if RfaH is required for the expression of group III capsules.
The genes for production of the group III K10 and K54 antigens have been cloned in this laboratory (Pearce and Roberts, 1995), and could be transformed into an rfaH mutant and capsule production assayed with a capsule-specific phage or by immunoblotting using capsule-specific antibody.

RfaH is required for the expression of several virulence and fertility factors in E. coli. It is needed for biosynthesis of the LPS core oligosaccharide (Pradel and Schnaitman, 1991), synthesis and secretion of α-haemolysin (Bailey et al., 1992), and production of the F factor sex pilus (Beutin and Achtman, 1979). Hybridisation and gene fusion studies have indicated that in each of these cases, RfaH acts as a transcriptional activator (Beutin et al., 1981; Gaffney et al., 1983; Brazas et al., 1991; Bailey et al., 1992). RfaH reduces the termination of transcripts in the F tra operon (Beutin et al., 1981; Gaffney et al., 1983) and the rplL-rpoB intergenic region (Railing and Linn, 1987), and may therefore act as a transcription antiterminator.

In order to study regulation of the K5 capsule gene cluster, bacterial luciferase transcriptional fusions were made in genes from each of the three regions of the cluster (kpsE, kfiA and kpsT). These fusions were transferred to the chromosome of the K5/K-12 hybrid strain using a positive-selection suicide vector. Into strains containing these fusions an rfaH mutation was introduced and the effect determined by assaying bioluminescence. Transcription of the serotype-specific region 2 gene, kfiA, was found to require RfaH. Because of potential problems associated with the use of luciferase as a reporter gene, this was confirmed by the direct detection of RNA in slot blot hybridisations. The transcription of the other region 2 genes was also found to require RfaH.

Transcription of the kpsT::luxAB fusion was reduced by about a third in an rfaH mutant. The reduction in kpsT transcription in an rfaH mutant was confirmed by RNase protection analysis (section 5.2.4) and primer extension (section 5.2.5), but was not observed in slot blot hybridisations. It follows that the control of hybridising a kpsT probe to RNA from MS101 and MS106 to check the relative amounts of RNA loaded is inappropriate. It would have been better to use a probe against transcripts that would be present at equal levels in both strains, for example ribosomal RNAs.

The notion that RfaH acts by causing transcription antitermination can be used to explain the pattern of transcription of the capsule genes in an rfaH mutant. It is proposed that RfaH acts to allow transcripts that would normally end after region 3 to continue into region 2. The finding that transcription from the kfiA promoter is not regulated by RfaH, even though kfiA transcription on the chromosome requires RfaH, supports the hypothesis that kfiA is expressed by readthrough transcription and that RfaH is needed for this to occur. By reverse transcriptase-PCR it was shown that the region between kpsT and kfiA is transcribed, providing further support for the antitermination model.
The model does not explain how RfaH activates the transcription of region 3. It has been reported that RfaH activates the transcription of genes in the tra, rfa and hly operons that precede putative transcription terminators (Gaffney et al., 1983; Brazas et al., 1991; Pradel and Schnaitman, 1991; Bailey et al., 1992). The extent to which the transcription of genes in the rfa operon relies on RfaH increases with their distance from the promoter (Brazas et al., 1991; Pradel and Schnaitman, 1991). It is therefore possible that RNA polymerase acts by suppressing pausing of the polymerase so that it may continue to the end of the operon. RfaH may therefore act to ensure that transcripts from the region 3 promoter reach region 3. RfaH lacks known DNA binding motifs and is not thought to affect transcription initiation (Bailey et al., 1992). This could be confirmed by cloning the region 3 promoter in a promoter probe vector and measuring its activity in wild-type and rfaH mutant strains.

Because of the possible importance of readthrough transcription in the expression of region 2 genes, the site of initiation of region 3 transcripts was mapped and compared in wild-type and rfaH mutant strains. By primer extension analysis, region 3 transcripts were found to initiate 741 bp 5' of the translation initiation codon for kpsM. The same site of initiation was used in an rfaH mutant strain, indicating that RfaH does not invoke alternative promoters to regulate the capsule genes. Region 3 transcripts also initiate at this site in E. coli K1, accounting for 2 bp that are not present in the E. coli K5 sequence.

The data accrued in this study do not provide direct evidence that RfaH causes antitermination. This could be obtained by Northern blotting of total RNA from MS101 and MS106 using a region 3 probe. It would be predicted that a region 3 probe would hybridise to two transcripts in MS101, a readthrough transcript and a region 3 only transcript. Because the location of the region 3 promoter is known, the extent of the readthrough transcript and the approximate location of the terminator can also be determined. In the rfaH mutant it would be expected that little, if any, readthrough transcription occurs, so the region 3 only transcript would predominate.

It proved particularly difficult to detect region 2 and 3 transcripts in total RNA from MS101 using DNA probes, suggesting that these regions are weakly transcribed. It may therefore be difficult to detect the readthrough transcript by Northern blotting, especially if it large (c. 10 kb if it ends 3' of kfdD). The detection of readthrough transcripts may be improved by using total RNA from a strain carrying multiple copies of the capsule genes on a plasmid, such as pPC6. However, a plasmid borne lacZ transcriptional fusion in kfiA was found to exhibit less pronounced regulation by RfaH than the corresponding chromosomal gene. This effect also occurs with a reporter gene fusion in the S. typhimurium rfaGBIJ operon on a plasmid (Brazas et al., 1991). The disparity in the regulation of plasmid and chromosomal genes by RfaH
underlines the importance of studying the expression of a single copy of the capsule genes in their normal chromosomal location.

Evidence for antitermination could also be obtained in an RNase protection assay, which is more sensitive than a Northern blot for the detection of mRNA (Sambrook et al., 1989). This approach requires the location of the terminator at the end of region 3 to be known. An antisense RNA probe spanning the terminator would be expected to protect two fragments with RNA from MS101; a large fragment, protected by the readthrough transcript, and a smaller fragment, protected by the terminated RNA. Using RNA from MS106, it would be expected that the smaller fragment would predominate, owing to the termination of region 3 transcripts in the absence of RfaH. This approach will not reveal how far the readthrough transcript extends across region 2.

A requirement of the antitermination model is that the promoters for the region 2 genes are so weak that, in the absence of readthrough transcription, insufficient region 2 proteins are made for polysaccharide biosynthesis to occur. Promoter probe data for the region 2 promoters would appear to support this notion (Petit et al., 1995). By preventing the termination of region 3 transcripts, RfaH may serve to boost the expression of the region 2 genes to a level at which capsule expression can occur. The situation is analogous in the regulation of the F factor traS, traD and traT genes by RfaH. Each of these genes is preceded by a promoter (Ham et al., 1989), and yet they are poorly expressed in an rfaH mutant (Beutin et al., 1981; Gaffney et al., 1983). It has been suggested that RfaH enhances the transcription of traS, traD and traT by preventing the termination of transcripts from the major tra promoter at sites upstream of these genes (Frost et al., 1994). The importance of readthrough transcription in the expression of region 2 could be confirmed by deleting the region 3 promoter in MS101 and assaying the effect this has on the transcription of region 2 genes.

The control of termination of region 3 transcripts offers a strategy for regulating the relative amounts of region 2 and 3 proteins. It follows that RfaH may have an important role in maintaining the correct stoichiometry of proteins in the multi-enzyme complex that mediates polysaccharide biosynthesis and export. It is arguable that such a system would not be conserved among *E. coli* expressing other group II capsules. There are several possible reasons for this. First, it may be expected that if the region 2 genes in different serotypes are transcribed from strong promoters of their own, or are transcribed in the opposite direction to region 3, they would not rely on readthrough transcription for their expression. It was shown that multiple copies of the region 2 genes can restore capsule production in an rfaH mutant (section 4.2.7), suggesting that if the region 2 proteins are made at a high enough level, the need for RfaH is diminished. The requirement for region 2 genes to be transcribed weakly and
in the same direction as region 3 may impose restrictions on how different region 2 cassettes evolve in group II strains. At the least, it would be expected that capsule expression in different group II strains will rely on RfaH to different extents.

The second requirement of the antitermination model is that the terminator at the end of region 3 is conserved in group II strains. However, in E. coli K1 there is only 50 bp between kpsT and the first gene of region 2 (neuD) and there is no obvious stem-loop structure that may correspond to a Rho-independent terminator in this region (Annunziato et al., 1995). In addition, there is a marked divergence in the nucleotide sequence at the 3' end of kpsT in different group II strains (K1, K4 and K5) (Drake, 1991). It is possible that the terminator is located within kpsT. However, if this were the case it would be expected that little (if any) functional KpsT would be made in an rfaH mutant. The finding that capsule production in strain MS106 can be restored by plasmid pAS30, which contains region 2 but lacks kpsT, suggests that the terminator is not located in kpsT. An attempt is currently being made to map the terminator between kpsT and kfiA in E. coli K5 by RNase protection analysis.

To determine if RfaH regulates the expression of other group II capsule gene clusters in the same way as in E. coli K5, RNA from wild-type and rfaH mutant strains carrying the cloned genes for different group II capsules could be hybridised with region 2 and 3 specific probes. The genes for the K1, K4, K7, K12 and K92 antigens have been cloned in this laboratory (Echarti et al., 1983; Roberts et al., 1988; Drake et al., 1990) and could be used for this study.

It is not clear how RfaH may cause transcription antitermination. A comparison of the predicted amino acid sequence of RfaH with that of the known antitermination factors, BglG, SacT and the λ N and Q proteins revealed no significant homology. However, a search of the protein sequence database using the FASTA program (Devereux et al., 1984), revealed that RfaH is 25% identical and 68% similar to the NusG protein from Thermus thermophilus. RfaH is also 20% identical and 65% similar (over 129 amino acids) to NusG from E. coli, and shares similar homology to the NusG proteins from Bacillus subtilis, Staphylococcus carnosus and Streptomyces coelicolor. In E. coli, NusG is a 20 kDa protein required for both transcription termination and antitermination events (Sullivan and Gottesman, 1992; Roberts, 1993). NusG is required for antitermination by the phage λ N protein (Li et al., 1992) and ribosomal RNA (rrn) operons (Squires et al., 1993). In common with RfaH, NusG also increases readthrough transcription in the rplL-rpoB intergenic region (Linn and Greenblatt, 1992). It is believed that NusG regulates transcription termination and antitermination by interacting with the termination factor Rho (Li et al., 1993; Burova and Gottesman, 1995).

It is proposed that Rho causes transcription termination in the following way (reviewed in Platt, 1994). Firstly, Rho binds as a hexamer to a cytidine-rich
unstructured region in the nascent mRNA. It then undergoes a conformational change that allows it to move towards the 3' end of the transcript using the energy derived from ATP hydrolysis. Rho then uses an ATP-dependent RNA-DNA helicase activity to release the nascent RNA when RNA polymerase pauses at a Rho-dependent terminator. In *E. coli*, NusG is required for the action of Rho at some terminators (Sullivan and Gottesman, 1992). NusG binds weakly to both Rho and RNA polymerase, and may facilitate termination by recruiting Rho to the elongation complex (Li *et al.*, 1992, 1993). Overexpression of NusG has been found to inhibit Rho-dependent termination in *E. coli*, perhaps by sequestering soluble Rho (Burova and Greenblatt, 1995). It follows that if the intracellular ratio of NusG to Rho is high enough, antitermination can occur at Rho-dependent terminators. Antitermination might also occur if NusG is prevented from interacting with RNA polymerase, because Rho may not be recruited to the elongation complex so well (Lin *et al.*, 1993; Burova and Greenblatt, 1995).

Recently, NusG has been reported to increase the rate of elongation of RNA polymerase by suppressing pausing (Burova *et al.*, 1995; Burns and Richardson, 1995). It is possible that this facilitates antitermination at Rho-dependent terminators by preventing Rho from 'catching up' with the elongation complex as it passes through the terminator. The phage λ N and Q proteins may also act in this way (Jin *et al.*, 1992). To determine if RfaH acts by increasing the elongation rate of RNA polymerase, the rate of elongation of region 3 transcripts could be measured in wild-type and *rfaH* mutant strains. This could be achieved by following the kinetics of induction of luciferase in strains carrying the *kpsT::luxAB* fusion (MS120 and MS121 (*rfaH*)) following a shift from 18°C to 37°C. The time taken for luciferase to appear will be a function of the elongation rate (Jin *et al.*, 1992). A similar method has been used to estimate RNA polymerase elongation rates in other systems (Burova *et al.*, 1995).

The finding that suppressors of *rfaH* phenotypes map in Rho and subunits of RNA polymerase (Beutin *et al.*, 1981; Farewell *et al.*, 1991), suggests that RfaH, like NusG, may interact with both Rho and the elongation complex. These interactions could be confirmed by affinity chromatography using purified RfaH as described by Li *et al.*, (1992, 1993). Purified RfaH would be covalently linked to an inert support, packed into a column and crude cell extracts from *E. coli* applied. Proteins would then be eluted from the column using salt solutions of increasing concentration and the eluates analysed by SDS-PAGE and Western blotting using antibodies specific for Rho and RNA polymerase.

It will also be important to determine if host factors, such as the Nus proteins, are needed for the action of RfaH. The host factor requirements for antitermination by the N and Q proteins are different (Friedman and Court, 1995). In the case of N, at least four host proteins are required for antitermination (NusA, NusE, NusG and
ribsosomal protein S10) (Mason et al., 1992; DeVito and Das, 1994), whereas Q has fewer host factor requirements, although NusA is essential for antitermination at some terminators (Grayhack et al., 1985). If possible the effect of mutations in nus genes should be studied in MS101 because the capsule genes exhibit less pronounced regulation by RfaH when carried on plasmids.

Superimposed on the regulation of the K5 capsule gene cluster by RfaH is the effect of temperature on expression of the capsule genes. None of the regulatory mutations studied induced expression of the K5 antigen at 18°C, suggesting that the temperature regulation of group II capsule production is not mediated by any of the regulators studied here. Transcription of regions 1 and 3, which are conserved among group II capsule gene clusters, is reduced at 18°C. The repression of kpsT transcription at 18°C was shown not to involve the hns or tcp genes, which repress transcription of the E. coli pap operon at low temperature. Transposon mutagenesis of the strain carrying the kpsT::luxAB fusion (MS120) has been performed, with selection for mutants which make more light at 18°C, or less light at 37°C. It is hoped that this approach will identify the protein(s) involved in the regulation of region 3 transcription by temperature.

It would be interesting to investigate the effect of other environmental cues, such as osmolarity, on the expression of group II K antigens. Osmolarity regulates the biosynthesis of alginate in P. aeruginosa (Berry et al., 1989), and the S. typhi Vi antigen (Pickard et al., 1994). In addition, desiccation activates transcription of the E. coli cps genes for the biosynthesis of colanic acid (Ophir and Gutnick, 1994), and it may be expected that this is also sensed as an increase in external osmolarity. Biosynthesis of the S. typhi Vi antigen is regulated by the OmpR-EnvZ two component system, which is capable of sensing osmolarity (Pickard et al., 1994). It would be interesting to investigate the role of this system in the biosynthesis of group II K antigens and other polysaccharides in enteric bacteria.

A short sequence (JUMPstart), that is found upstream of many polysaccharide biosynthesis genes in enteric bacteria, was found to be conserved in the non-coding regions 5' of RfaH-regulated operons. Deletion of this sequence from the K5 capsule gene cluster reduced the transcription of kfiA without affecting region 3 transcription. This phenotype is the same as that of an rfaH mutant and suggests that the JUMPstart sequence may be required for the action of RfaH. Removal of the JUMPstart sequence also reduced the in vitro UDP-glucuronic acid transferase activity of MS101 to a level comparable to that of an acapsular strain. It is possible that this phenotype is due in part to reduced translation of region 3 transcripts, since the JUMPstart deletion removed one nucleotide from the Shine-Dalgarno sequence for kpsM. Mutations in the JUMPstart sequence may also affect the accessibility of the Shine-Dalgarno sequence.
or start codon for kpsM by altering the secondary structure of the mRNA. However, the finding that the region 2 genes on a plasmid can restore capsule production in strain MS137 suggests that the JUMPstart deletion has not significantly affected the expression of the region 3 genes. This should be confirmed by comparing the production of region 3 proteins in a wild-type strain and strains carrying the JUMPstart mutations.

The role of the JUMPstart sequence in antitermination by RfaH may be analogical to that of the phage λ nut and qut sites in the action of N and Q (Roberts, 1993; Greenblatt et al., 1993; Friedman and Court, 1995). Thus, RfaH may interact with the elongation complex while it is associated with the JUMPstart sequence, and modify it in such a way that it fails to stop at downstream terminators. Assuming this to be the case, it is unlikely that the JUMPstart sequence would act as a signal in the nascent RNA, as with nut, since it is not transcribed in the hly operon (Vogel et al., 1988). In cases where the JUMPstart sequence is transcribed, the distance between the promoter and the JUMPstart sequence is variable and in no instance does it overlap the promoter. This is in contrast to the qut site, which overlaps the λ late gene promoter (Yarnell and Roberts, 1992). It may be that the JUMPstart sequence, or factors associated with it, are able to contact RNA polymerase paused in the promoter region or at another site before the terminator. In this respect, the JUMPstart sequence may act as a kind of enhancer (Collado-Vides et al., 1991). This would seem unlikely however, as MS137 could not be complemented by supplying a functional copy of the JUMPstart sequence in trans. The hlyR element, which contains a JUMPstart sequence and is required for antitermination in the hly operon, also cannot be supplied in trans (Vogel et al., 1988). The orientation of the JUMPstart sequence relative to the direction of transcription is the same in all instances and it would be interesting to determine if this is important for its activity. It is known that hlyR only functions in one orientation (Vogel et al., 1988).

The JUMPstart sequence described by Hobbs and Reeves, (1994) shares some homology to boxA of the nut sites of lambdoid phages. BoxA-like sequences have been implicated in antitermination in several E. coli operons, including ribosomal RNA operons (Li et al., 1984; Berg et al., 1989; Albrechtsen et al., 1990; Heinrich et al., 1995) and the tryptophanase (tna) operon (Stewart and Yanofsky, 1985). Sequences similar to boxA are also found in several attenuators (Friedman and Gottesman, 1983). This may suggest that the antitermination systems of E. coli and lambdoid phages have a common ancestry.

Further mutagenesis of the JUMPstart sequence may help to define the role of this sequence in the regulation of the capsule genes by RfaH. Defined deletions or single base changes could be made in the boxA-like sequences or in the 12 contiguous residues that are most conserved among RfaH-regulated operons. Because of the
proximity of the JUMPstart sequence to the Shine-Dalgarno sequence for \textit{kpsM}, the
effect of these mutations on expression of the region 3 proteins should be determined.
Where possible the effect of JUMPstart mutations should be assayed in MS101 or
MS108, which contains a \textit{kfaA::luxAB} fusion. For the transfer of these mutations to
the chromosome of MS101 or MS108 to be possible, the mutations must introduce or
remove a restriction enzyme site so that double recombinants can be identified by PCR
and digestion of the product.

To summarise, RfaH may activate transcription of region 2 of the K5 capsule
gene cluster by preventing the termination of region 3 transcripts. It is appears that the
JUMPstart sequence is required for this process. The implication is that genes preceded
by the JUMPstart sequence may share a common mode of regulation by RfaH. Thus,
RfaH may coordinately regulate the expression of several cell surface polysaccharides
and other virulence factors in enteric bacteria.
References


Arrecubieta, C., López, R. and García, E. (1994) Molecular characterisation of *cap3a*, a gene from the operon required for the synthesis of the capsule of *Streptococcus*
pneumoniae type 3: sequencing of mutations responsible for the unencapsulated phenotype and localisation of the capsular cluster on the pneumococcal chromosome. *Journal of Bacteriology* 176, 6375-6383.


Goldberg, J. and Dahneke, T. (1992) Pseudomonas aeruginosa AlgB, which modulates the expression of alginate, is a member of the NtrC subclass of prokaryotic regulators. Molecular Microbiology 6, 59-66.


Merrick, M. (1993) In a class of its own- the RNA polymerase sigma factor $\sigma^{54} (\sigma^5)$. *Molecular Microbiology* 10, 903-909.


Morgan, E. (1980) Insertions of Tn10 into an *E. coli* ribosomal RNA operon are incompletely polar. *Cell* 21, 257-265.


Russo, T. and Singh, G. (1993) An extraintestinal, pathogenic isolate of *Escherichia coli* (O4:K54:H5) can produce a group 1 capsule which is divergently regulated from its constitutively expressed group 2, K54 capsular polysaccharide. *Journal of Bacteriology* 175, 7617-7623.


Troy, F. and McCloskey, M. (1979) Role of a membranous sialyltransferase complex in the synthesis of surface polysaccharides containing polysialic acid in *Escherichia coli*.
Temperature-induced alteration in the assembly process. *Journal of Biological Chemistry* 254, 7377-7387.


