The expression and transcriptional organisation of region 1 of the *Escherichia coli* K5 capsule gene cluster

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

by

David A. C. Simpson BSc (Leicester)
Department of Microbiology and Immunology
University of Leicester

September 1996

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

The expression and transcriptional organisation of region 1 of the *Escherichia coli* K5 capsule gene cluster

David A. C. Simpson

E. coli strains associated with extra-intestinal infections are generally encapsulated. The gene clusters required for expression of group II capsules comprise three functional regions. The expression of region 1, which is conserved between the different serotypes was investigated. The K5 *kps* region 1 was transcribed as an 8.0 kb polycistronic mRNA which was processed to form a 1.3 kb transcript encoding the most promoter-distal gene *kpsS*. Transcription initiated at a single position and the nucleotide sequence of the promoter revealed two AT-rich sequences also present at the equivalent position in the region 3 promoter. These may represent sites for the co-ordinated regulation of the conserved regions 1 and 3. Two potential IHF binding sites were also identified near the region 1 promoter. The region 1 transcript included an additional gene, *kpsF*, previously thought to be outside the capsule cluster. This encoded a 35.6 kDa KpsF protein with a predicted amino acid sequence 95% identical to KpsF from *E. coli* K1 and also homologous with GutQ, ORF328 and an hypothetical open reading frame from *Haemophilus influenzae*. KpsF is not essential for capsule production and no regulatory role could be established. An intragenic, Rho-dependent transcriptional terminator was discovered within *kpsF* and may link transcription of region 1 with the physiological status of the cell. Group II capsules are expressed at 37°C but not 18°C and region 1 was regulated at the level of transcription, with transcripts barely detectable at 18°C by Northern blotting or RNase protection. The temperature-regulation of region 1 was unaffected by mutations in regulatory genes known to control the expression of other temperature-dependent virulence factors. A mutation in *rfaH*, which affects group II capsule expression also had no effect on region 1 expression.

ACKNOWLEDGEMENTS

I would like to thank Professor Ian Roberts for his supervision of this work. I am grateful to all the members of the Department of Microbiology and Immunology who have given me advice and practical assistance. In particular I have appreciated the help and support of the members of Lab 228, among them Mark Stevens, Rowan Pearce, Julie Eastgate, Chantal Petit, Anna Ostrowski, Brad Clarke and Gary Griffiths. I am indebted to Gordon Rigg for teaching me about the computer analysis of sequence data and his help with the innumerable problems encountered when using computers! Of the many friends I have made during the course of this project I would particularly like to thank Fred Esumeh, with whom I have been able to share the trials and tribulations of completing a PhD thesis. A special thanks also to Andrew Wallace for reminding me never to take things too seriously! Greatest thanks to my wife, Fionnuala, who has gone out of her way to support me throughout. Finally a mention of my daughter Ella, to whom I dedicate this thesis.

CONTENTS

ABBREVIATIONS

ABC	ATP-binding cassette
AMV	Avian myeloblastosis virus
BLAST	basic local alignment search tool
bp	base pair
CCCP	*m*-chlorophenylhydrazone
CKS	CMP-KDO synthetase
CMP-KDO	cytidine monophosphate-KDO
CRP	cyclic AMP receptor protein
DEPC	diethylpyrocarbonate
DTT	dithiothreitol
ECA	enterobacterial common antigen
EDTA	ethylenediamine-tetra acetic acid
EPS	extracellular polysaccharide
FUR	ferric uptake regulator
GalA	galacturonic acid
Glc	glucose
GlcA	glucuronic acid
GlcNAc	*N*-acetylglucosamine
GST	glutathione S-transferase
IHF	integration host factor
IPTG	isopropylthiogalactoside
IS	insertion sequence
KDO	2-keto-3-deoxymanno-octonic acid
kb	kilobase
kDa	kilo Dalton
LB	Luria Bertani
LPS	lipopolysaccharide
LRP	leucine-responsive regulatory protein
MAC	membrane attack complex
MOPS	3-[N-Morpholino]-propane-sulphonic acid
MW	molecular weight
NCBI	National Center for Biotechnology Information
NeuNAc	*N*-acetylneuraminic acid
ONPG	*o*-nitrophenyl-ß-D-galactoside
ORF	open reading frame
PBS-T	phosphate buffered saline/Tween 20
PEG	polyethylene glycol
PGP	phosphatidyl glycerophosphate
PMNL	polymorphonuclear leukocytes
R-LPS	rough LPS
REP	repetitive extragenic palindromic (sequence)
RT-PCR	reverse transcriptase polymerase chain reaction
S-LPS	smooth LPS
SDS-PAGE	sodium lauryl sulphate-polyacrylamide gel electrophoresis
SD	Shine-Dalgarno
und-PP	undecaprenolpyrophosphate

CHAPTER 1

Introduction

1.1 The biology of polysaccharide capsules

1.1.1 General features

Bacteria produce a diverse range of polysaccharides. They are comprised of repeating monosaccharide units linked by glycosidic bonds formed at one of several hydroxy groups on each monosaccharide. The variety of configurations this enables and the different monosaccharides available are responsible for the great structural diversity. Further diversity is achieved by formation of branched chains and substitution with organic or inorganic molecules.

Most bacterial polysaccharides are found on the outer surface of the cell. Extracellular polysaccharides with little or no cell association are termed slime polysaccharides. Capsular polysaccharides (K antigens) form a cohesive layer or capsule and are often covalently attached to the cell surface via a hydrophobic anchor. Lipopolysaccharide (LPS) (O antigen) is anchored in the outer leaflet of the outer membrane by lipid A. LPS forms a layer extending 20-40 nm from the cell surface while capsular polysaccharide can extend up to several micrometres, masking the O antigen (Whitfield and Valvano, 1993). The various surface polysaccharides of a typical Gram-negative bacterium, illustrated by *Escherichia coli,* are depicted in Fig. 1.1 (Whitfield *et al.,* 1994).

Polysaccharide capsules are highly hydrated (95-99% H_2O) (Moxon and Kroll, 1990) and hydrophilic (anionic). Their location on the cell surface means they often mediate interactions with the environment. As such they are involved in the pathogenesis of animal and plant infections (Moxon and Kroll, 1990)(Cross, 1990), adherence, biofilm formation (Costerton *et al.,* 1987) and prevention of desiccation (Ophir and Gutnick, 1994). These functions are described in the following sections.

1.1.2 Role in adhesion

Bacteria interact with inanimate surfaces, other bacteria and host tissues. Extracellular polysaccharides often initiate adhesion. Cell division and recruitment of bacteria from the planktonic phase leads to formation of a biofilm (Costerton *et al.,* 1987). Biofilms reduce the efficiency of industrial systems and

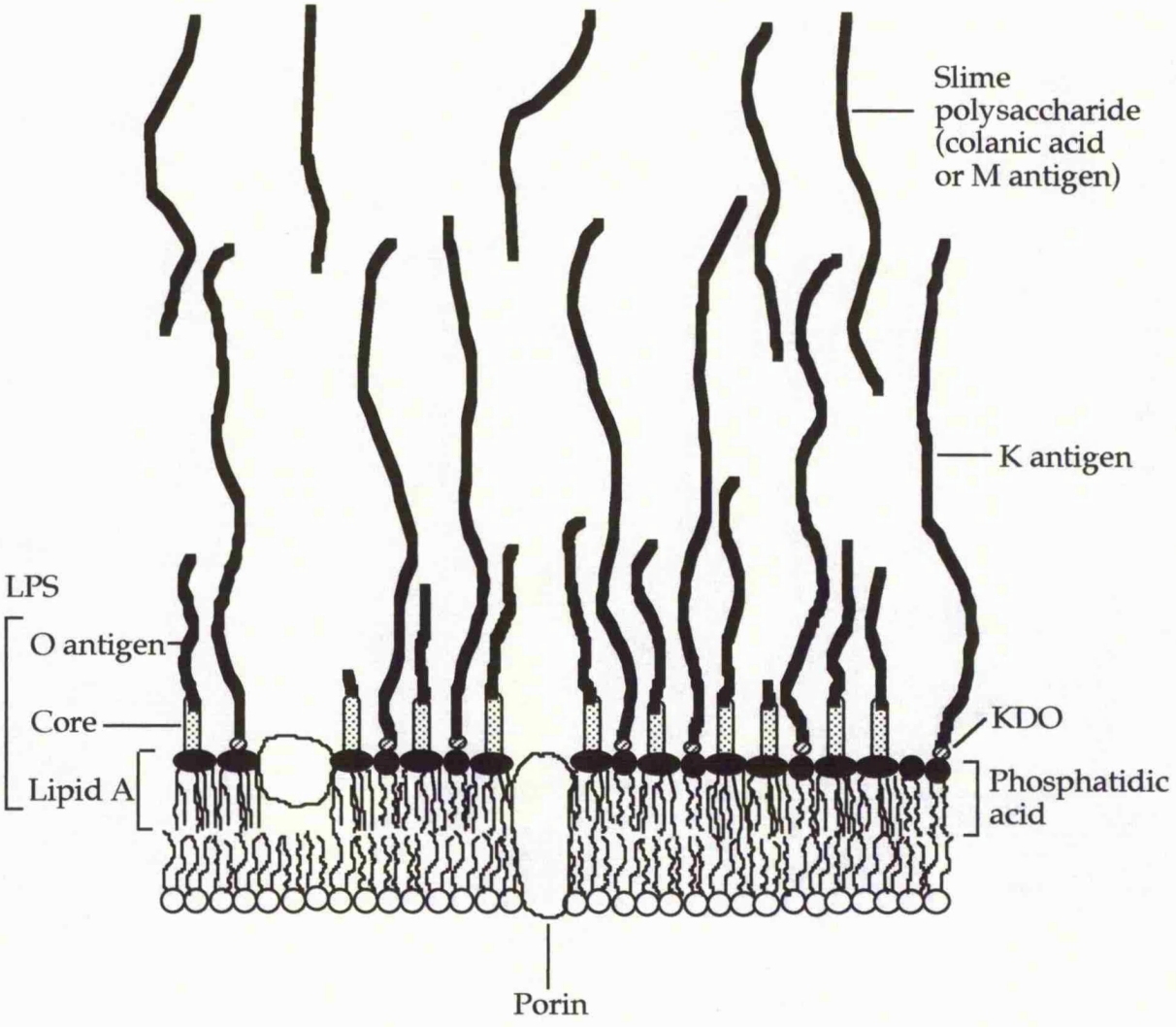


Figure 1.1. The surface polysaccharides of *E. coli*. The diagram illustrates the attachment of LPS and a typical group II K antigen to the periplasmic membrane.

are costly to remove. When they form on medical devices and prostheses biofilms become a centre for infection, especially in immunocompromised patients.

Biofilms may form on host tissues. While pili and other recognition proteins are involved in specific interactions with host receptors exopolysaccharides mediate less specific adhesion mechanisms. *E. coli* can coexpress adhesive proteins and polysaccharide capsules (Kröncke *et al.*, 1990). Biofilms may comprise a structured consortium of different bacteria with the initial colonisers providing a suitable environment for the establishment of later species (Costerton *et al.*, 1987). The cells are bound within a predominantly polysaccharide matrix. This matrix traps nutrients, while limiting antibacterial agents such as antibiotics and host defences. During prolonged infection of the lung of cystic fibrosis patients *Pseudomonas aeruginosa* shifts to a mucoid form producing large amounts of the exopolysaccharide alginate, which may provide a barrier to antibiotics (Deretic *et al.*, 1994).

1.1.3 Prevention of desiccation

It has been suggested that exopolysaccharides may enhance survival and promote transmission between hosts by preventing desiccation. Such a function is likely to have provided the selective advantage which lead to the fixation of the genes required for capsule expression rather than the promotion of invasive infections leading to the decimation of the host (Moxon and Kroll, 1990). The large amounts of water absorbed by capsular polysaccharides may slow the drying rate and maintain a humid environment around the cell. It has been demonstrated that mucoid strains of *E. coli, Acinetobacter calcoaceticus* and *Erwinia stewarti* are more resistant to desiccation than their corresponding isogenic non-mucoid mutants (Ophir and Gutnick, 1994). In the case of *E. coli* desiccation was shown to induce expression of the colanic acid biosynthetic genes (*cps*).

1.1.4 Role in virulence

The LPS and capsular polysaccharides produced by many of the bacteria associated with invasive disease, including *E. coli, Salmonella typhi, Haemophilus influenzae, Neisseria meningitidis, Klebsiella* spp and *Streptococcus pneumoniae* help them to evade host defences (Cross, 1990). However, only a few serotypes or combinations of O and K antigens are associated with extra-intestinal infections. *E. coli* serotypes K1 and K5 are

frequently isolated from cases of bacteraemia and urinary tract infections and *E. coli* K1 comprises 80% of isolates from neonatal meningitis (Ørskov and Ørskov, 1992). Certain polysaccharide capsules inhibit phagocytosis and increase resistance to complement-mediated serum killing in the pre-immune host. In the early phase of infection complement is activated by the alternative pathway in which deposition of C3b is amplified by binding of factor B to form a C3 convertase C3bBb (Cross, 1990). Opsonisation with C3b performs three functions: to act as a ligand for specific receptors on polymorphonuclear leukocytes (PMNL) or macrophages, promoting phagocytosis; to trigger the cascade of the terminal complement sequence C5-C9, resulting in formation of the membrane attack complex (MAC) and subsequent bacteriolysis; and to attach to lymphocyte receptors, promoting immunoglobulin secretion and lymphokine production.

Polysaccharide capsules can interfere with the amplification of C3b. The complement regulatory protein factor H binds to C3b facilitating its degradation by factor I, thereby terminating amplification (Cross, 1990). Capsules containing sialic acid (for example *E. coli* K1 and *Streptococcus* group B, type III) increase the affinity of factor H for C3b and inhibit formation of the C3 convertase (Stevens *et al.*, 1978). Other capsules (pneumococcal types 7 and 12) have a low affinity for factor B and therefore increased relative affinity for factor H.

Certain *E. coli* capsules, in the absence of specific capsular antibodies, have a steric effect on complement fixation, masking structures on the bacterial surface which would otherwise activate complement (Horwitz and Silverstein, 1980). However, masking of, for example, LPS is not complete because antibodies against O antigen can in some cases induce phagocytosis and killing by PMNL, suggesting that sufficient LPS is exposed to allow specific opsonisation (Horwitz and Silverstein, 1980). Rather than preventing complement activation the capsules of *Staphylococcus aureus* and *S. pneumoniae* mask bound C3b from the C3b receptors on the surface of PMNL (Roberts *et al.*, 1989). The presence of a thick capsule may result in formation of the MAC too far from the cell surface to be effective. Shedding of capsule with attached host factors may reduce the effects of complement (Moxon and Kroll, 1990). Capsular polysaccharides form a hydrophilic, negatively charged (Allen *et al.*, 1987) layer on the bacterial cell surface. This may inhibit phagocytosis by mutual repulsion of the phagocyte or by altering surface tension.

Although polysaccharide capsules play an important role in pathogenesis it is difficult to establish the precise role of individual structures. Comparison of isogenic capsular transformants has shown that the *H. influenzae* type b capsule

is a determinant of virulence (Zwahlen *et al.*, 1989). However, although the type b capsule is necessary for *H. influenzae* to survive and replicate in blood it is not sufficient alone (Moxon and Kroll, 1990). Three strains of *E. coli* expressing K1 capsule were shown to have increased *in vitro* serum resistance and virulence in newborn rats than their isogenic parents (Kim *et al.*, 1986). However, expression of the cloned *E. coli* K1 genes in a laboratory strain of *E. coli* only slightly increased resistance to the bactericidal effects of complement (Allen *et al.*, 1987)(Roberts *et al.*, 1989), therefore indicating that other cell surface structures such as LPS play a role in serum resistance. The anti-phagocytic role of capsules is dependent upon the specific combination of cell surface molecules that are expressed: *E. coli* K1 isolates are associated with a limited number of O antigens (Ørskov and Ørskov, 1992), which may reflect a functional relationship (Whitfield *et al.*, 1994). The polysialic acid capsule of *N. meningitidis* was thought to be the main factor preventing activation of the alternative complement pathway, however sialylation of lipo-oligosaccharide has now been shown to be more important in preventing complement-mediated killing (Hammerschmidt *et al.*, 1994).

The protective effect of the capsule is usually lost after the development of an immune response. The immunogenicity of certain capsular polysaccharides depends upon the age of the host and many are not immunogenic in young infants (Jennings, 1990). Certain polysaccharides are similar to host components and because of the mechanisms which prevent autoimmune responses, are also exceptionally poor immunogens in adults (Jann and Jann, 1992). The K1 poly-α-2,8-sialic acid capsule is identical to the carbohydrate terminus of the embryonic neural cell adhesion molecule n-CAM. The K5 polysaccharide (-4)-ß-GlcA-(1,4)-αGlcNAc-(1-, Vann *et al.*, 1981) is identical with the first polymeric intermediate of heparin biosynthesis. Bacteria expressing these capsules are therefore afforded some resistance against the host's specific immune response, which may explain the prevalence of serotypes K1 and K5 among *E. coli* isolates from certain infections.

1.2 Surface polysaccharides of *Escherichia coli*

1.2.1 K antigens

Classification

The capsules or K antigens (kapselantigene) produced by *E. coli* are acidic polysaccharides (Jann and Jann, 1990). A given isolate produces one of over 80 serologically distinct types (Ørskov and Ørskov, 1992). The K antigens can be divided into at least three groups according to chemical, genetic and physical

criteria (Table 1.1)(Jann and Jann, 1990)(Pearce and Roberts, 1995). Group I polysaccharides have a high molecular weight and low electrophoretic mobility. They are charged because they contain glucuronic acid (GlcA) or galacturonic acid (GalA) or are substituted with pyruvate. Group II and III polysaccharides have a lower molecular weight and higher charge density, with a variety of acidic components such as *N*-acetylneuraminic acid (NeuNAc), 2-keto-3-deoxymanno-octonic acid (KDO) or *N*-acetylmannosaminuronic acid (ManNAc) and greater phosphate substitution. The repeating units of some representative K antigens are shown in Table 1.2: The shorter mono-, di- or trisaccharide repeats of group II and III polysaccharides results in higher charge densities.

Group I capsules are expressed at all growth temperatures while group II are not expressed below 20°C. Strains expressing group II capsules have elevated levels of CMP-KDO synthetase activity at capsule-permissive temperatures. In serotype K5 the enzyme responsible for this activity, KpsU, has been shown to be encoded within a region common among group II capsule gene clusters (sections 1.3.2 and 1.3.4)(Pazzani *et al.*, 1993). Group III polysaccharides are structurally similar to group II but resemble group I in that they are not temperature regulated and lack elevated levels of CMP-KDO synthetase (Finke *et al.*, 1990). Group I capsules are co-expressed only with O antigens O8, O9 and O20, while group II and III capsules may be co-expressed with one of many O antigens, other than those found with group I.

The group I capsules can be further subdivided into those that lack amino sugars in the repeating unit and resemble the capsules of *Klebsiella* spp. (group 1A) and those that contain amino sugars (group IB) (Jann and Jann, 1992). This division is strengthened by variations in their association with the slime polysaccharide colanic acid and mode of regulation. Group IA capsules are not coexpressed with colanic acid and are regulated by the *rcs* system while strains expressing group IB capsules are able to express colanic acid (Jayaratne *et al.*, 1993). The *rcs* system was shown previously to regulate expression of colanic acid in *E. coli* K-12 (Gottesman and Stout, 1991). It comprises the RcsC and RcsB sensor and effector proteins respectively, the positive regulators RcsA and RcsF and the negative regulator the Lon protease (section 1.5.2). The introduction of multiple copies of *rcsB* into *E. coli* group IB strains activated expression of colanic acid but in group IA strains such as K30 no colanic acid was produced and the amount of K antigen produced was increased (Jayaratne *et al.*, 1993). Based on genetic evidence from *E. coli* K30 it has been suggested that the genes for the biosynthesis of group IA capsular polysaccharide and colanic acid are allelic (Keenleyside *et al.*, 1992). Colanic acid can therefore be classified as a group IA capsular

Table 1.1. Classification of K antigens.

Property	Group IA	Group IB	Group II	Group III
Molecular weight	> 100 kDa	> 100 kDa	< 50 kDa	< 50 kDa
Acidic component	GlcA, GalA, Pyruvate	GlcA, NeuNAc	GlcA, NeuNAc, KDO, ManNAcA, Phosphate	GlcA, NeuNAc, KDO, ManNAcA, Phosphate
Amino sugars	No	Yes	Yes/No	Yes/No?
Expressed below 20°C	Yes	Yes	No	Yes
Lipid at reducing end	Lipid A, ?	Lipid A	Phosphatidic acid	Phosphatidic acid
Removal of lipid at pH5-6, 100°C	No	No	Yes	Yes
Chromosomal locus (near to)	? (*his, rfb*)	*cps* (*his, rfb*), ?(*trp*)	*kps* (*serA*)	? (*serA*)
Co-expression with O antigens	O8, O9, O20, O101	O8, O9, O20, O101	Many O-antigens (other than those with group I)	Many O-antigens (other than those with group I)
Co-expression with colanic acid	No	Yes	Yes	Yes?
Elevated CMP-KDO synthetase activity	No	No	Yes	No
Intergeneric relationships	*Klebsiella* spp	-	*H. influenzae*, *N. meningitidis*	?

Table 1.2. Repeating units of typical K antigens.

K antigen	Repeat unit
Group IA	
K30	2)-Man-(1,3)-Gal-(1- 3 | β-GlcA-(1,3)-Gal
K31	2)-α-Glc(1,3)-β-Gal-(1,3)-β-GlcA-(1,2)-β-Rha-(1,2)-β-Rha-(1,
Group IB	
K9	3)-β-Gal-(1,3)-β-GalNAc-(1,4)-α-Gal-(1,4)-α-NeuNAc-(2, 5-OAc
Group II	
K4	4)-β-GlcA-(1,3)-β-GalNAc-(1, 3 2 β-Fru
K5	4)-β-GlcA-(1,4)-α-GlcNAc-(1,
K97	2)-β-Rib-(1,5)-β-KDO-(2,
K1	8)-α-NeuNAc-(2,
K92	8)-α-NeuNAc-(2,9)-α-NeuNAc-(2,
Group III	
K10	3)-α-Rha-(1,3)-β-QuiNMal-(1,
K54	3)-β-GlcA-(1,3)-α-Rha-(1, | CO.NH threonine (serine)

Abbreviations: Fru, fructose; Gal, galactose; GalNAc, *N*-acetylgalactosamine; Glc, glucose; GlcA, glucuronic acid; GlcNAc, *N*-acetylglucosamine; KDO, 2-keto-3-deoxymanno-octonic acid; Man, mannose; NeuNAc, *N*-acetylneuraminic acid; QuiNMal, 4,6-dideoxy-4-malonylamidoglucose; Rha, rhamnose; Rib, ribose. For original references see (Jann and Jann, 1990)

polysaccharide. Introduction of multicopy *rcsA* or *rcsB* into *E. coli* group II strains induced synthesis of colanic acid at 37°C but did not affect expression of the K antigens (Keenleyside *et al.*, 1992)(Keenleyside *et al.*, 1993).

Most group I capsular polysaccharides are linked to core-lipid A (section 1.2.2), but this is not a universal feature. The attachment of *E. coli* K30 antigen was studied as a typical example of a group IA capsule (MacLachlan *et al.*, 1993). Low molecular weight K30 antigen with a single oligosaccharide repeat unit was attached to core-lipid A but the high molecular weight polysaccharide was expressed on the surface by an independent pathway which was unaffected by mutations in core biosynthesis genes. There is therefore presumably an alternative membrane anchor other than core-lipid A for the attachment of group IA polysaccharides. In contrast, a high molecular weight fraction linked to core-lipid A was isolated from the *E. coli* K40, group IB strain and following SDS-PAGE it exhibited a ladder-like pattern characteristic of LPS (Jann *et al.*, 1992). Group IB capsules also resemble LPS in that strains expressing them have the *rol* gene on their chromosome, unlike group IA strains in which the *rol* gene has not been detected (Dodgson *et al.*, 1996). The product of the *rol* gene controls O-polysaccharide chain length (Bastin *et al.*, 1993)(section 1.2.2) and its presence in multicopy had a similar effect on group IB polysaccharides, reducing their length (Dodgson *et al.*, 1996). Certain group IB polysaccharides have been described as acidic LPS. The K87 and O32 or K9 and O104 antigens are examples of polysaccharides with identical structures that have been classified as group IB K antigens when expressed with an O antigen (O8 or O20) and as O antigens when expressed alone (Jann and Jann, 1990).

Group II polysaccharides are attached to phosphatidic acid via a phosphodiester linkage (extremely labile at pH < 6) to their reducing terminus which is almost always KDO, irrespective of whether it is in the repeating unit (Schmidt and Jann, 1982). The lipid substitution is assumed to act as a hydrophobic anchor attaching the polysaccharide to the cell surface. This function has been demonstrated for the terminal lipid moieties of the related capsular polysaccharides of *N. meningitidis* group C and *H. influenzae* type b which were shown to promote association with outer membrane vesicles (Arakere *et al.*, 1994) (this association was eliminated by removal of lipid with phospholipase). However, only 20-50% of the *E. coli* group II polysaccharide chains are lipid-substituted (Jann and Jann, 1990) and ionic interactions may hold the remainder at the cell surface.

The group IA K antigen biosynthesis genes (*cps*) are located near *rfb* and *his* on the *E. coli* chromosome (section 1.3.1) with an additional *trp*-linked locus involved with synthesis of serotype K27 (Whitfield *et al.*, 1994). The group IB antigen biosynthesis genes are also assumed to map near *his*. The genes required for the production of group II polysaccharides are located near *serA* and those from many different serotypes have been cloned (section 1.3.2). The K10 and K54 group III capsule gene clusters have been cloned and also map near *serA* (Pearce and Roberts, 1995). There is some genetic evidence to support the inclusion of several serotypes in a fourth group (see section 1.3.3.).

Biosynthesis and export

Capsular polysaccharides are synthesised from activated precursors available in the cytoplasm, which are usually nucleotide sugars and may or may not be capsule-specific. Synthesis of the polysaccharide or oligosaccharide intermediates occurs at a complex located on the cytoplasmic face of the cytoplasmic membrane (Roberts, 1996). This is suggested by the retention of *in vitro* polymerisation activity in membrane fractions and the need to use cytoplasmic precursors. The polymerisation of K5 polysaccharide was shown to occur on the inside of the cytoplasmic membrane by the much greater *in vitro* activity of inside-out vesicles compared with right-side-out vesicles (Finke *et al.*, 1991). Involvement of lipid-linked intermediates in the biosynthesis of many polysaccharides implies membrane association. The known glycosyl transferases are basic proteins likely to associate with the membrane through ionic interactions (Steenbergen *et al.*, 1992) and generally lack membrane-spanning domains (Liu *et al*, 1993).

The initiation of polysaccharide biosynthesis is poorly understood, but generally requires a specific enzyme distinct from the glycosyl transferases involved in chain elongation. The polymerisation and export of the group I and II polysaccharides follows two different pathways (Fig. 1.2). Group II capsular polysaccharides grow at their non-reducing end by the sequential addition of single residues from nucleotide sugar precursors. Much of this understanding has been derived from studies of the *E. coli* K1 polysialic acid capsule (Table 1.2). The growing polysaccharide is thought to be attached to undecaprenol and sialic acid residues attached to undecaprenol monophosphate (und-P) have been isolated from membrane preparations of cells synthesising K1 antigen (Troy *et al.* 1975). (The presence of und-P differentiates this mechanism from that of O antigen biosynthesis which involves undecaprenolpyrophosphate (und-PP)-linked intermediates (section 1.2.2)). Alternatively undecaprenol could be acting as a carrier for sialic acid residues between CMP-sialic acid and the growing polymer attached to a protein acceptor (Whitfield and Valvano, 1993). The demonstration that a single gene encoded the K1 polymerase limited the

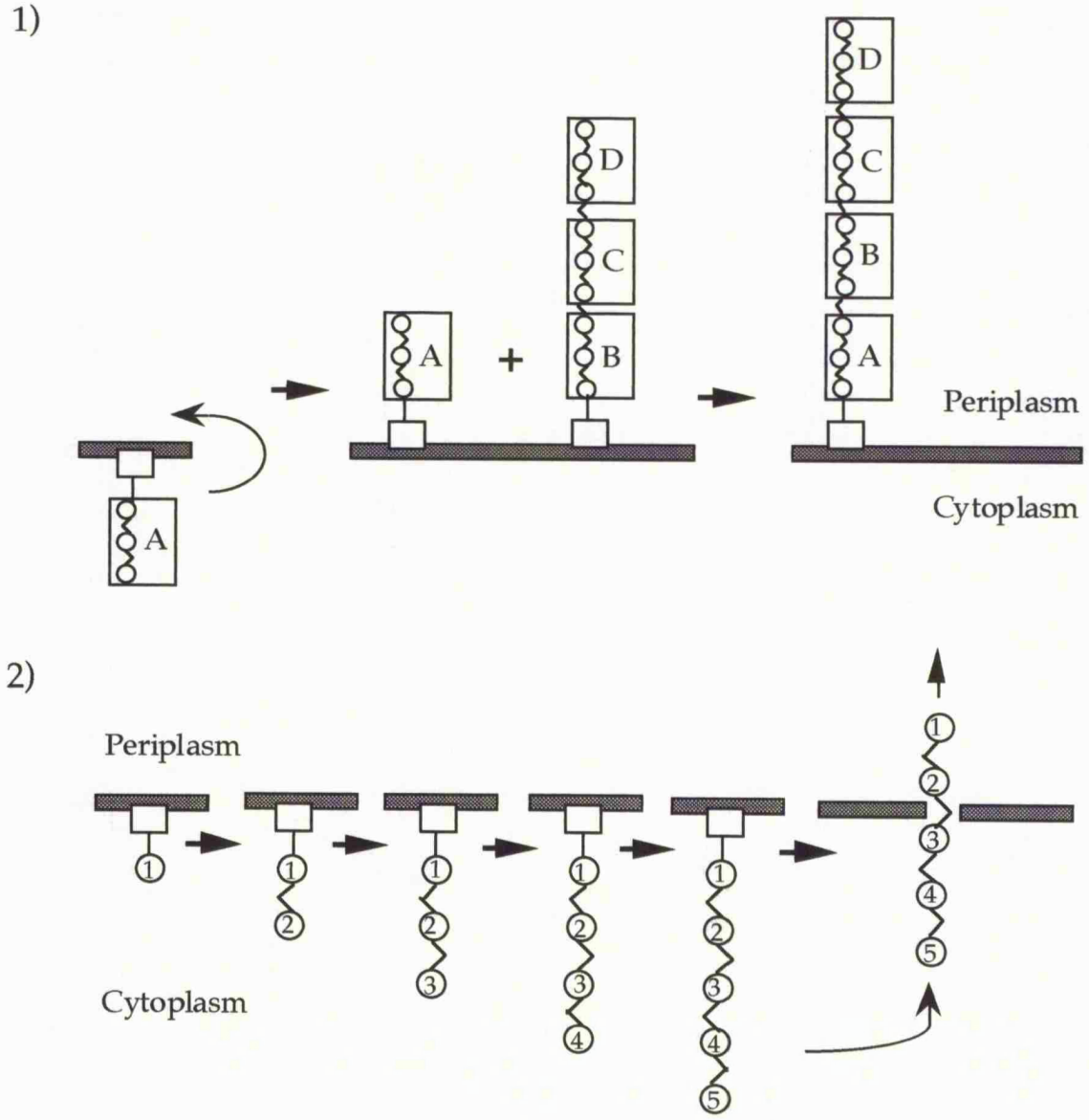


Figure 1.2. Two models for polysaccharide biosynthesis (see Whitfield, 1993 and references therein). 1) Model based on polymerisation of *Salmonella enterica* O-polysaccharide. Repeat units are assembled on the carrier lipid undecaprenol (□) at the cytoplasmic face of the inner membrane before translocation to the periplasmic face. The nascent polysaccharide (B-C-D) grows at the reducing end by transfer to a lipid-linked repeat unit (A). The biosynthesis of group I K antigens is thought to resemble this process. 2) Processive polymerisation model based on *E coli* O-polysaccharides O8 and O9. The polysaccharide grows at the non-reducing terminus by sequential addition of individual residues. The completed polysaccharide is then exported across the cytoplasmic membrane. The biosynthesis of group II K antigens resembles this model, although the involvement of lipid is uncertain.

number of potential assembly mechanisms (Steenbergen and Vimr, 1990). Membrane preparations from *E. coli* K1 incorporated sialosyl residues from radiolabelled CMP-sialic acid into acceptors present in the membrane (endogenous acceptors).

When a purified sialic acid pentamer was provided as an exogenous acceptor single sialosyl residues were incorporated, suggesting that chain elongation proceeded by sequential transfer of monomeric units (Steenbergen and Vimr, 1990)(Steenbergen *et al.*, 1992). Complementation of a mutation in the K1 polymerase gene by that from *E. coli* K92 demonstrated that this similar polysaccharide (Table 1.2) was synthesised by the same processive mechanism.

In vitro studies with another group II polysaccharide, K5, have not identified any undecaprenol-linked derivatives (Finke *et al.*, 1991)(Jann and Jann, 1990). Also, biosynthesis *in vitro* is neither inhibited by bacitracin, which prevents lipid recycling, nor stimulated by exogenous polyprenol lipid, suggesting that undecaprenol-linked intermediates are not involved (Finke *et al.*, 1991). However, chain elongation does occur by sequential addition of single residues: A single glycosyl transferase enzyme, KfiC (section 1.3.4), that transfers both *N*-acetylglucosamine (GlcNAc) and glucuronic acid (GlcA) from the respective UDP-sugars has been identified. When incubated with either of the radiolabelled precursors alone this enzyme can incorporate the radiolabelled sugar into exogenous K5 polysaccharide acceptor (Petit *et al.*, 1995) or partially N-sulphated acceptors (Lidholt *et al.*, 1994). The apparent differences in biosynthesis of the K1 and K5 polysaccharides indicate that although group II polysaccharide gene clusters share a common genetic organisation (section 1.3.2) this may not necessarily imply that the polysaccharides are polymerised by exactly the same mechanism. The processive polymerisation mechanism of group II K antigens resembles that of *E. coli* O polysaccharides O8 and O9 (section 1.2.2) but differs, at least in K7, in that it is *rfe*-independent (Meier-Dieter *et al.*, 1990).

Analysis of the biosynthesis of the group II-like capsules of *N. meningitidis* and *H. influenzae* supports a processive model for the biosynthesis of this class of polysaccharides. When cells of *H. influenzae* group B, which synthesises a polysaccharide with a ribose–ribitol-5-phosphate repeat unit were permeabilised and incubated with radiolabelled CDP-ribitol, a single ribitol phosphate residue was incorporated into endogenous polysaccharide (Eldere *et al.*, 1995). The fact that ribitol phosphate is added to endogenous polysaccharide in the absence of ribose, the donor for which is unknown, implies that the sugars are added in a

stepwise process rather than being assembled as a ribitol phosphate-ribose repeat unit on a lipid carrier and then transferred to the growing polymer.

The polymerisation of group II polysaccharides is apparently completed prior to export from the cytoplasm. Mutations in certain genes involved in the export of *E. coli* K1 and K5 polysaccharides result in full length polymer located in the cytoplasm (section 1.3.4). This suggests that polymerisation is completed at the cytoplasmic face of the membrane and is consistent with the direction of growth at the non-reducing terminus, distal from the lipid carrier. This contrasts with polymerisation of group I polysaccharides at the reducing end of the polymer which enables the growing terminus to remain near the near the cytoplasmic membrane during polymerisation at the periplasmic face.

The intracellular polysaccharide from *E. coli* K1 and K5 mutants with defects in transport functions (section 1.3.4) is not substituted with phospholipid, suggesting that this is a prerequisite for export. Likewise, lipid substitution of the group II-like polysaccharide of *N. meningitidis* was shown to be a requirement for translocation (Frosch and Müller, 1993). Group II polysaccharides are exported across the cytoplasmic membrane by an ATP-binding cassette (ABC) transport system (Fath and Kolter, 1993)(Higgins *et al.*, 1990)(Reizer *et al.*, 1992). This comprises an ATP-binding protein which couples hydrolysis of ATP with translocation of polysaccharide through a pore formed by a dimer of integral membrane proteins (Fig. 1.4)(Pavelka *et al.*, 1991; Smith *et al.*, 1990). The ATP-binding and integral membrane components (KpsT and KpsM, respectively) from *E. coli* K1 and K5 are described in section 1.3.4 and Table 1.3. Homologous systems are involved in the export of various polysaccharides including the group II-like capsules of *H. influenzae* (Kroll *et al.*, 1990) and *N. meningitidis* (Frosch and Müller, 1993) (Fig. 1.5) and the *E. coli* serotype O9 LPS (Kido *et al.*, 1995).

Translocation of polysaccharide across the periplasm and into the outer leaflet of the periplasmic membrane is poorly understood. Analysis of the phenotype of mutants defective in specific cloned group II capsule genes has indicated proteins involved in the export process (sections 1.3.2 and 1.3.4). It has been suggested that capsular polysaccharides may be exported at membrane adhesion zones (Bayer junctions)(Bayer, 1968)(Bayer, 1990) where the inner and outer membranes come close together. In temperature upshift experiments newly synthesised group II polysaccharide of serotypes K1, K5 and K12 was observed by immunoelectronmicroscopy as tufts above membrane adhesion zones (Kröncke *et al.*, 1990).

Proteins may be involved in polysaccharide assembly in the outer membrane. Whitfield *et al* identified five proteins in the outer membrane of *E. coli* which correlated with the expression of the K1 antigen (Whitfield *et al.*, 1985). One of these was the porin protein K which had previously been shown to be present in encapsulated bacteria. However, the expression of K1 polysaccharide from recombinant plasmids in *E. coli* K-12, which lacks protein K, led to the suggestion that it may be the preferred porin, while others can perform a similar function (Whitfield and Valvano, 1993). There is however no direct evidence linking protein K or other specific outer membrane proteins with expression of extracellular polysaccharide.

The polymerisation and export of the group I capsular polysaccharides is poorly understood. The mechanism is thought to differ from that of group II polysaccharides in that oligosaccharide repeat units are synthesised individually and then assembled into polysaccharide (Fig. 1.2). This was suggested by studies of the biosynthesis of the group I-like polysaccharide of *Klebsiella aerogenes* in which tetrasaccharide repeat units are synthesised on an undecaprenol carrier (Troy *et al.*, 1971). The und-P-linked polysaccharide grows at the reducing end by transfer of a repeat unit from und-P. If the sugar, rather than sugar phosphate, is transferred this leaves und-PP which must be dephosphorylated to und-P before it can act as a carrier for another repeat unit. This dephosphorylation reaction is inhibited by bacitracin. The inhibition of *K. aerogenes* polysaccharide biosynthesis by bacitracin therefore demonstrates that it involves cycling of undecaprenol phosphate. This mechanism is similar to the well characterised pathway for the synthesis of the O polysaccharides of *Salmonella enterica* (section 1.2.2). Initial studies suggest that colanic acid synthesis in *E. coli* is similar to that of *Klebsiella* capsular polysaccharide (Whitfield *et al.*, 1994). Given the close relationships between group IA capsules and colanic acid and between group IB capsules and LPS it is likely that group I capsules are assembled via a blockwise mechanism.

1.2.2 Lipopolysaccharide (LPS)

LPS consists of lipid A, core oligosaccharide and O polysaccharide (O antigen). The hydrophobic lipid A forms the outer leaflet of the outer membrane and a minimal LPS (Re LPS) consisting of lipid A and two KDO moieties is essential for bacterial growth (Raetz, 1993). Lipid A is synthesised from UDP-GlcNAc which is acylated at positions 2 and 3 forming UDP-2,3-diacylglucosamine. Cleavage at the pyrophosphate bond generates 2,3-diacylglucosamine-1-phosphate (lipid X). UDP-2,3-diacylglucosamine and lipid X are condensed to form the disaccharide backbone, UDP is removed and two KDO residues are added prior to the final

acylation reactions. Lipid A therefore comprises a diglucosamine backbone phosphorylated at positions 4' and 1 and attached to six fatty acid residues (ß-hydroxymyristol) (Raetz, 1993). Lipid A (endotoxin) is responsible for activating eukaryotic signal transduction pathways in macrophages and stimulating excess cytokine production which can lead to endotoxin-induced shock (Raetz, 1993).

Lipid A is attached to a non-repeating core oligosaccharide, which is divided into an inner conserved and an outer variable region. The inner core contains KDO and L-*glycero*-D-*manno*-heptose. KDO is transferred from the precursor CMP-KDO, which is synthesised by the KdsA (KDO 8-phosphate synthetase) and KdsB (CMP-KDO synthetase) enzymes (Strohmaier *et al.*, 1995). Five outer core types have been defined in *E. coli* (Whitfield *et al.*, 1994). Most of the genes involved in core synthesis are located in the *rfa* cluster (Schnaitman and Klena, 1993). LPS which terminates with the core oligosaccharide is described as rough LPS (R-LPS). Attachment of O polysaccharide to the outer core oligosaccharide forms smooth LPS (S-LPS). O polysaccharides consist of oligosaccharide subunits which may be repeated over 50 times. Most extra-intestinal *E. coli* isolates are smooth, reflecting the role of O-specific side chains in resisting complement-mediated serum killing (Valvano, 1992)(section 1.14). However, the resistance afforded to complement depends on the chemical composition of the specific O chain and only a few of the 173 designated O serotypes (Ørskov and Ørskov, 1992) are associated with invasive infections (Jann and Jann, 1992).

Most of the genes required for O antigen biosynthesis are located in the *rfb* cluster at 40 min between *his* and the *cps* genes for colanic acid biosynthesis (Schnaitman and Klena, 1993). O polysaccharides are synthesised on a lipid carrier from nucleotide sugars prior to ligation onto core-lipid A to form LPS. However, there are two distinct pathways which couple different polymerisation processes with different translocation mechanisms (Fig. 1.2)(Whitfield, 1995). Most *E coli* O polysaccharides are synthesised by the blockwise assembly mechanism (Fig. 1.2(1)) exemplified by the O polysaccharides of *S. enterica*. RfbP is required to transfer the first galactose-1-phosphate from UDP-Gal onto und-P. Oligosaccharide repeat units are assembled on und-PP-gal at the cytoplasmic face of the inner membrane. Sugars are transferred sequentially from their sugar nucleotide precursors by glycosyl transferases (Liu *et al.*, 1993). Polymerisation occurs at the periplasmic face of the inner membrane and the lipid-linked repeat units are translocated or 'flipped' across the membrane by a process probably involving RfbX (Whitfield, 1995). The polysaccharide grows at the reducing terminus by transfer of the partially polymerised polysaccharide to a single repeat unit linked to und-P. This polymerisation process requires the Rfc protein – *rfc*

mutants have only a single repeat unit of O antigen attached to lipid A core, giving a semi-rough (SR) phenotype (Mäkelä and Stocker, 1984).

LPS biosynthesis is terminated by the transfer of the polymerised polysaccharide (or monomeric oligosaccharide) from undecaprenol to core-lipid A in a ligation reaction involving the RfaL protein. The length of certain O polysaccharides has been correlated with serum resistance (Valvano, 1992). Analysis of O polysaccharides by SDS-PAGE demonstrates that chain length has a modal distribution. A protein involved in controlling chain length has been identified as the regulator of O antigen chain length (Rol) (Batchelor *et al.*, 1992) or O chain length determinator (Cld) (Bastin *et al.*, 1993). Rol may allow polymerisation by Rfc to continue for a defined period and then promote ligation of the O antigen to core-lipid A by RfaL. Several Rol proteins have been studied and a given protein can impose a specific modal length on heterologous polysaccharides, although the cognate components are required for optimum activity.

The second pathway for biosynthesis of LPS in *E. coli* involves a processive mechanism exemplified by the O8 and O9 O antigens of *E coli* K-12. This pathway requires a functional *rfe* gene, which encodes UDP-GlcNAc: und-P GlcNAc-1-phosphate transferase (Meier-Dieter *et al.*, 1992). Rfe is required to produce und-PP-GlcNAc which has been shown to act as an acceptor of mannose residues in the synthesis of O8 antigen (Rick *et al.*, 1994). A single GlcNAc residue was identified at the reducing terminus of lipid-carrier-linked O8 side chains. Chain elongation may therefore progress by successive addition of mannosyl residues from GDP-mannose to the non-reducing terminus of the growing polymer. That the residues are added processively is supported by the lack of a mutation analogous to *rfc* resulting in a single repeat unit attached to core-lipid A and the absence of inhibition *in vitro* by bacitracin (which prevents recycling of undecaprenol) (Whitfield and Valvano, 1993). The O8 and O9 polysaccharides are both homopolymers of mannose with tri- and pentasaccharide repeat units respectively and it is not clear how the polymerisation mechanism incorporates these repeat structures. The completed polymer may be transferred with the terminal GlcNAc residue to core-lipid A creating mature LPS.

It was demonstrated previously that und-PP-Glc was an acceptor for at least 30 mannosyl residues in the biosynthesis of O9 polysaccharide (Jann and Jann, 1984). The R1 LPS core structure of *rfe* mutants of *E. coli* O9 lacks a nonreducing terminal glucose (Glc) residue and it was therefore proposed that transfer of the O9 side chain with a terminal glucose completed the core structure. However

und-PP-Glc is not required for synthesis of lipid-linked O8 chains because these intermediates accumulated in a phosphoglucose isomerase (*pgi*) mutant in the absence of exogenously supplied glucose (Rick *et al.*, 1994). The apparent ability of both und-PP-Glc and und-PP-GlcNAc to act as acceptors for mannose residues while both processes required *rfe* was explained by the recent demonstration that Rfe can transfer either Glc-1-phosphate or GlcNAc-1-phosphate to und-P (Kido *et al.*, 1995). In the case of O9 polysaccharide biosynthesis the MtfC enzyme, which catalyses transfer of the first mannose residue, prefers the GlcNAc derivative (Kido *et al.*, 1995). However, the specific structure of the core oligosaccharide may influence which sugar is transferred.

To date only homopolymeric O antigens have been shown to be synthesised by the processive mechanism (the *E. coli* O8 and O9 mannans and the *Klebsiella pneumoniae* O1 D-galactan (Bronner *et al.*, 1994)). However, group II K antigens appear to be synthesised by a similar mechanism (section 1.2.1) and include heteropolymers such as the K5 polysaccharide (Table 1.2). The bifunctional transferase may be responsible for maintaining the fidelity of the repeat structure of the K5 polysaccharide (section 1.3.4).

Polymerisation is completed at the cytoplasmic face of the inner membrane and consequently this system lacks a 'flippase', no Rfc protein is required for polymerisation of subunits and chain length is not influenced by Rol. Instead an ABC transporter is required for export of the polysaccharide (although its precise role has not been demonstrated). This consists of an ATP-binding protein which is believed to couple ATP hydrolysis with polysaccharide export through a pore formed by an integral membrane component. Such systems have been found in the *E. coli* O9 (Kido *et al.*, 1995) and *K. pneumoniae* O1 (Bronner *et al.*, 1994) *rfb* clusters. ABC transporters are also required for the export of group II K antigens and are associated with the processive polymerisation mechanism and absence of an Rfc homologue.

While biosynthesis by the processive mechanism is *rfe*-dependent this requirement is not diagnostic for this pathway because biosynthesis of *E. coli* GlcNAc-containing O polysaccharides, including O7, O18, O75 and O111 also requires Rfe (Alexander and Valvano, 1994). The blockwise synthesis of ECA is also *rfe*-dependent (section 1.2.3)(Meier-Dieter *et al.*, 1992). How O antigen, synthesised by either pathway, is translocated across the periplasm and onto the outer membrane is unknown.

1.2.3 Enterobacterial Common Antigen (ECA)

ECA, which has no apparent role in virulence is produced by almost all members of the Enterobacteriaceae, including *E. coli*. It is a heteropolymer with a trisaccharide repeat unit containing the amino sugars GlcNAc, *N*-acetyl-D-mannosaminuronic acid (ManNAcA) and 4-acetamido-4,6-dideoxy-D-galactose (Fuc4NAc) (Meier-Dieter *et al.*, 1990). ECA is assembled on und-P and all the lipid-linked intermediates involved in biosynthesis of the repeat unit have been isolated (Barr *et al.*, 1989). As for certain O antigens (section 1.2.2) biosynthesis of ECA is dependent on the product of the *rfe* gene to catalyse the transfer of GlcNAc-1-phosphate from UDP-GlcNAc to und-P (Meier-Dieter *et al.*, 1992). The proteins required for the biosynthesis of ECA-specific precursors and the relevant glycosyl transferases are encoded by the adjacent *rff* locus (Meier-Dieter *et al.*, 1990). ECA polymerisation is apparently blockwise and while the precise mechanism has not been established it is likely to resemble that of O antigens because the ECA gene cluster encodes an open reading frame (ORF) homologous with the Rol (Cld) protein (Meier-Dieter *et al.*, 1992)(Bastin *et al.*, 1993). ECA is generally attached to the cell surface via a phosphodiester linkage with phosphatidic acid but can be found attached to core-lipid A in certain R-LPS strains. The similarity with O antigen biosynthesis is emphasised by the dependence of the attachment of ECA to core-lipid A on RfaL (Whitfield *et al.*, 1994).

1.2.4 Colanic acid (M antigen)

Many *E. coli* strains and related members of the Enterobacteriaceae express colanic acid (M antigen) giving their colonies a mucoid appearance. Colanic acid has a six-residue repeating unit containing glucose, galactose, glucuronic acid and fucose and has recently been recognised as a group IA K antigen (section 1.2.1) because of the similar chemical structure and the observation that it is not co-expressed with other group IA K antigens (Jayaratne *et al.*, 1993). The *E. coli* K-12 *cps* cluster which contains genes involved in colanic acid synthesis has been cloned (section 1.3.1)(Aoyama *et al.*, 1994).

1.3 Genetics of *E. coli* capsules

1.3.1 Group I capsule gene clusters

Genes involved in the synthesis of group IA capsules form the *cps* cluster near *his* and adjacent to the *rfb* O antigen genes at 45 min on the *E. coli* chromosome

(Whitfield *et al.*, 1994). In *E. coli* O9:K30 proteins encoded within the *rfb* cluster provide the precursor GDP-mannose which is required for both O and K antigen expression, demonstrating co-operation between polysaccharide biosynthetic pathways (Jayaratne *et al.*, 1994). An additional *trp*-linked locus may be involved in the synthesis of serotype K27 (Schmidt *et al.*, 1977).

The *cps* cluster for colanic acid, recognised recently as a group IA capsule, has been cloned from *E. coli* K-12 and part of the nucleotide sequence determined (Aoyama *et al.*, 1994). The DNA encodes five genes, two of which are involved in the synthesis of GDP-fucose, and has a G+C content above the average for *E. coli* suggesting that it has undergone lateral transfer. The homologous genes from *S. enterica* LT2 have an even higher G+C content and analysis of the effect of random genetic drift on codon base 3 suggests that the two *cps* clusters were acquired separately from a high G+C strain after the divergence of the two species. The discovery that the *cps* locus is polymorphic is consistent with these independent acquisitions. The widespread expression of colanic acid by *E. coli* strains appears to be due to a selective advantage of this particular group IA capsule rather than the presence of the colanic acid *cps* genes in an ancestral strain as was previously suggested. In the absence of direct data on the location of the group IB antigen biosynthesis genes they are assumed to also map near *his*. In strains expressing group IB K antigens the *rol* gene, which is involved in determining the polysaccharide chain length, is present on the chromosome (Dodgson *et al.*, 1996).

1.3.2 Group II capsule gene clusters

Group II capsule gene clusters map near *serA* (63 min) (Boulnois *et al.*, 1992) and sequence data has recently confirmed that the group II K1 and K5 *kps* clusters are both at the same site at 64 min on the *E. coli* chromosome, adjacent to *speC* (see chapter 3). The genetic organisation of the group II capsules is better characterised than that of the other groups. The genes required for expression of serotypes K1 (Silver *et al.*, 1981)(Echarti *et al.*, 1983), K4 (Drake *et al.*, 1990), K5, K7, K12 and K92 (Roberts *et al.*, 1986) have been cloned. Three functional regions have been identified within approximately 17 kb of DNA (Fig. 1.3)(Boulnois *et al.*, 1987). Hybridisation studies have shown that regions 1 and 3 are largely homologous between serotypes, while region 2 is serotype-specific (Roberts *et al.*, 1986)(Roberts *et al.*, 1988)(Silver *et al.*, 1987).

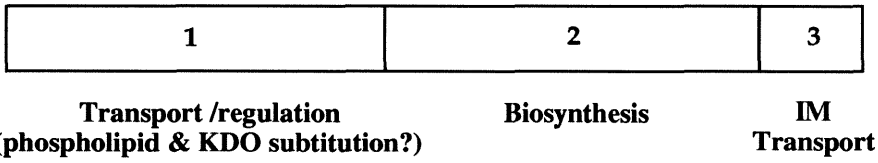


Figure 1.3. The regional organisation of group II capsule gene clusters. A central serotype-specific region is flanked by regions involved primarily in polysaccharide export. Some genes in region 1 appear to be involved with later stages of biosynthesis which are associated with export.

Mutations in region 1 result in intracellular polysaccharide (Bronner *et al.*, 1993). Such mutations in one K antigen cluster can be complemented by the equivalent region from a different cluster (Roberts *et al.*, 1986)(Roberts *et al.*, 1988)(Boulnois and Roberts, 1990). The manner in which the proteins encoded by region 1 interact with polysaccharide during the process of translocation from the cytoplasm to the cell surface is therefore independent of the structure of the repeat unit. The products of certain region 1 genes are functionally conserved with proteins encoded by group III capsule clusters (section 1.3.3): Mutations in *kpsE*, *kpsD* and *kpsC* in K5 (section 1.3.4) have been complemented by subclones of the K10 and K54 group III gene clusters (Pearce and Roberts, 1995). Group II polysaccharides have KDO at their reducing end irrespective of whether KDO is present in the repeating unit. This correlates with high levels of CMP-KDO synthetase in strains expressing group II capsules (section 1.2.1) (Finke *et al.*, 1990) and suggests that CMP-KDO is involved in an essential reaction in group II polysaccharide biosynthesis. In K5 the gene encoding the enzyme responsible for the elevated CMP-KDO synthetase activity, *kpsU*, is located in region 1 (Pazzani *et al.*, 1993; Rosenow *et al.*, 1995). This is distinct from the *kdsB* gene which encodes the CMP-KDO synthetase involved in LPS core biosynthesis (section 1.2.2). The substitution of group II polysaccharides with KDO and phosphatidic acid by region 1 proteins (see section 1.3.4 and (Frosch and Müller, 1993)) may provide the common signal for export of heterologous polysaccharides.

Region 2 is serotype-specific and encodes the functions required for capsule biosynthesis (Boulnois and Jann, 1989). Mutants in region 2 of K1 and K5 are defective in polysaccharide production (Echarti *et al.*, 1983)(Drake *et al.*, 1990)(Boulnois *et al.*, 1987)(Kröncke *et al.*, 1990). The size of region 2 varies between serotypes from approximately 6 kb to 14 kb, broadly correlating with the complexity of the polysaccharide encoded.

Region 3 is approximately 2 kb of DNA encoding products required for polysaccharide export across the inner membrane. Mutations in this region

result in polysaccharide either in the cytoplasm or bound to the cytoplasmic face of the inner membrane (Echarti *et al.*, 1983)(Kröncke *et al.*, 1990). Analysis of region 3 in K1 and K5 (Pavelka *et al.*, 1991)(Smith *et al.*, 1990) has revealed two genes, *kpsM* and *kpsT* which resemble ABC transporters (Reizer *et al.*, 1992)(Fath and Kolter, 1993) and are described in section 1.3.4.

1.3.3 Group III capsule gene clusters

Group III capsule gene clusters map near *serA* (63 min), the same location as group II clusters. They do however form a distinct set of gene clusters because hybridisation studies have shown that group III strains lack the group II cluster (Boulnois *et al.*, 1992)(Drake *et al.*, 1993). The K10 and K54 group III capsule gene clusters have been cloned and appear to have a segmental organisation analogous to group II clusters (Pearce and Roberts, 1995). The K10 gene cluster encodes proteins which complement mutations in the *kpsE, D* and *C* genes of K5 (Pearce and Roberts, 1995). The nucleotide sequence of sections of the K10 gene cluster was determined recently and revealed homologues of these K5 genes (Pearce, R., personal communication). A gene with nucleotide sequence similar to the K5 *kpsM* gene was also detected, but has not been shown to complement a mutation in the K5 gene. The K11 and K19 capsule gene clusters have group III features but no DNA similarity with the K10 capsule genes was detected by Southern blot analysis (Pearce and Roberts, 1995). The K11 gene cluster is located near *his*, the location of group I capsule genes (Pearce, R., personal communication). K11 and K19 have therefore been proposed to belong to a fourth group of K antigens.

1.3.4 The K5 capsule gene cluster

The K5 capsule gene cluster has been cloned (Roberts *et al.*, 1986) and an 18 kb DNA fragment shown to be sufficient to direct synthesis of the K5 polysaccharide (Bronner *et al.*, 1993). The nucleotide sequence of the entire cluster has been determined (Pazzani *et al.*, 1993)(Petit *et al.*, 1995)(Smith *et al.*, 1990) and shown to encode 12 potential ORFs (Table 1.3 and Fig. 1.4). All of these proteins except a 16 kDa ORF in region 2 have been observed in minicells or *in vitro* transcription-translation systems. By analysis of the effect of mutations in these genes, comparison of the predicted amino acid sequences with proteins in databases and biochemical analyses it has been possible to assign functions to many of the encoded proteins (Table 1.3 and Fig. 1.4).

Table 1.3. The proteins encoded by the K5 capsule gene cluster.

Protein	Homologue(s) (%identity, %similarity /number amino acids)	Features
Region 1		
KpsE	BexC (28.1%, 73.2%/359) CtrB (26.8%, 73.2%/355)	43 kDa, N- and C-terminal cytoplasmic membrane-spanning domains and large central periplasmic region (Rosenow *et al.*, 1995)
KpsD	GumB (25.3%, 68.8%/170) ExoF (29.3%, 67.7%/99) PgpB (25.0%, 60.3%/68) MalK (17.6%, 65.7%/102)	60.5 kDa periplasmic protein with typical signal sequence
KpsU	KdsB (44.3%, 70.7%/246) NeuA (21.1%, 65%/183)	27.1 kDa CMP-KDO synthetase (Rosenow *et al.*, 1995)
KpsC	LipA (48.9%, 82.2%/393) LpsZ (38.1%, 75.9%/312)	75.6 kDa
KpsS	LipB (37.4%, 76.0%/396)	30-46 kDa
Region 2*		
KfiD	Udg (72.2%, 91.2%/388) HasB (53.8%, 85.8%/400) AlgD (27%, 69.3%/359)	44 kDa UDP-Glc Dehydrogenase (Sieberth *et al.*, 1995)
KfiC	ExoO (23.9%, 67.4%/230) NodC (17.6%, 60.7%/267) HasA (23.1%, 73.1%/52)	60 kDa UDP-GlcA and UDP-GlcNAc transferase (Petit *et al.*, 1995)
KfiB	No homologues	66 kDa
KfiA	No homologues	27 kDa
Region 3		
KpsT	BexA (45.2%, 89.4%/220) CtrD (43.8%, 87.6%/217) NodI (24.4%, 68.8%/205)	25 kDa ATP-binding component of putative ABC transport system
KpsM	BexB (26.0%, 69.0%/258) CtrC (24.8%, 71.7%/246) NodJ (14.8%, 60.4%/81)	29 kDa integral membrane protein of putative ABC transport system

*Region 2 also encodes a predicted 16 kDa ORF which has not been observed and has no detectable phenotype when disrupted (Petit *et al.*, 1995). The %similarity figures include conservative amino acid changes. Only several examples of the many proteins similar to KfiC, KfiD, KpsT and KpsM are shown.

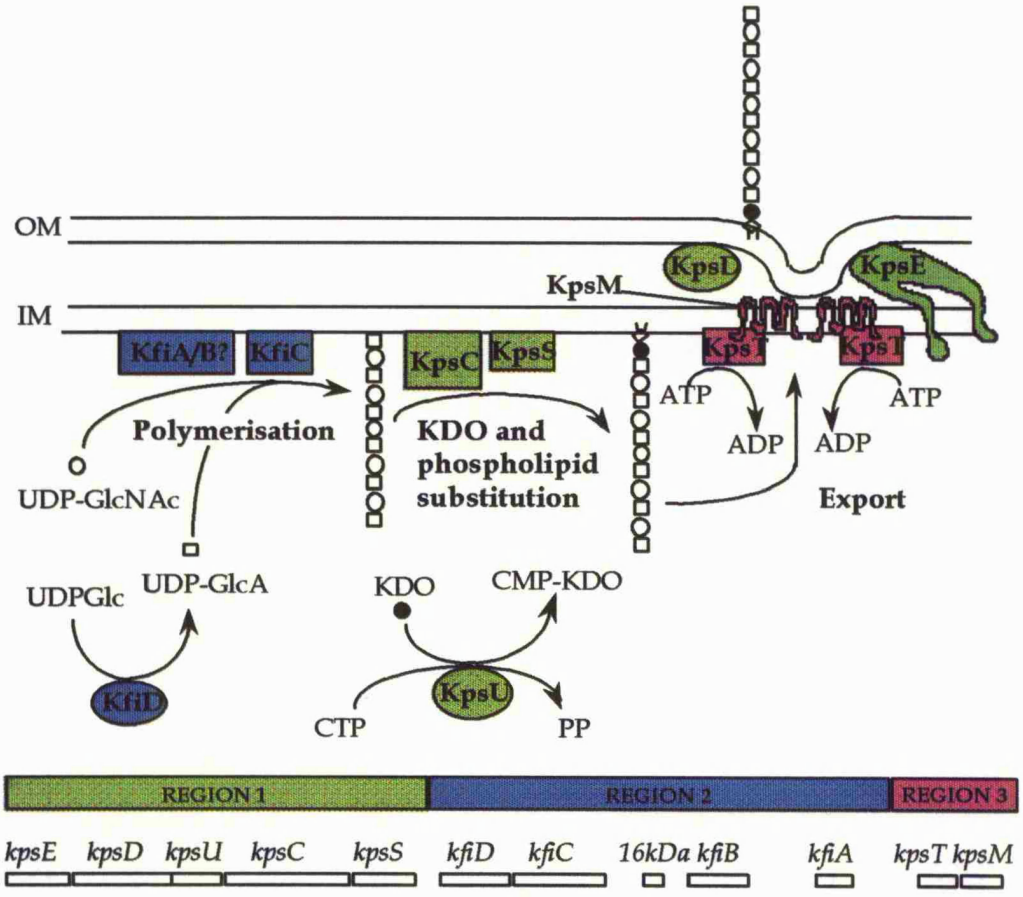


Figure 1.4. The *E. coli* K5 capsule gene cluster and the production of K5 capsular polysaccharide. The schematic diagram shows the main stages of polysaccharide biosynthesis and export and the proteins believed to be involved in each process. The proteins associated with the cytoplasmic membrane (IM) are likely to form a complex and have been drawn individually for clarity. The depression in the outer membrane (OM) represents a membrane adhesion zone (Bayer junction). The colours of the proteins correspond with their genetic organisation which is depicted in the lower section of the figure.

Region 1 encodes five proteins on the same strand and is potentially organised as a single transcriptional unit (Figure 1.4). Upstream of the first gene is an ORF, beginning beyond the known sequence, with homology to the GutQ protein from the glucitol operon (Yamada *et al.*, 1990). Constructs lacking most of this ORF continue to direct synthesis of K5 polysaccharide (Bronner *et al.*, 1993; Pazzani, 1992) indicating that it is non-essential for capsule production. Work in this thesis and with serotype K1 (Cieslewicz *et al.*, 1993) suggests that this gene is in fact part of the capsule gene cluster and it has been designated *kpsF* .

Immunoelectron-microscopy showed that the amount of polysaccharide produced by strains lacking the first two genes of region 1, *kpsE* and *kpsD*, was greatly reduced relative to the wild type (Bronner *et al.*, 1993)(Bronner *et al.*, 1993). The polysaccharide was located in the periplasm and was substituted with KDO and phospholipid. A similar phenotype was observed when *kpsE* was disrupted by insertion of an oligonucleotide linker, although polar effects could have been involved (Pazzani, 1992). The hydropathic profile of KpsE generated from the predicted amino acid sequence indicates two hydrophobic membrane-spanning helices, near the ends of the protein. It was predicted that KpsE is anchored in the cytoplasmic membrane with a large central, relatively hydrophilic periplasmic domain. This topology was confirmed by measuring the activity of Tn*phoA* and ß-lactamase fusions constructed at different positions within KpsE (Rosenow *et al.*, 1995). Anti-KpsE antibodies confirmed the location of the protein in the cytoplasmic membrane, while treatment of membranes with proteinase K and cytofluorometric analysis of spheroplasts demonstrated the exposed periplasmic domain (Rosenow *et al.*, 1995). KpsE is homologous with BexC from *H. influenzae* (Kroll *et al.*, 1990) and CtrB from *N. meningitidis* (Frosch *et al.*, 1991)(Table 1.3), both of which are involved in polysaccharide export. The phenotype of *kpsE* mutants and the protein topology and homologies are consistent with a role for KpsE in translocation of polysaccharide through the periplasm. KpsE, BexC and CtrB are homologous with the N-terminal domain of ExoP from *Rhizobium meliloti* (Becker *et al.*, 1995) and with other proteins involved in polysaccharide chain length determination (Bastin *et al.*, 1993). It is however unlikely that KpsE is involved in determining the length of K5 polysaccharide because although less polysaccharide is produced in *kpsE* mutants the chain length is unaltered (Bronner *et al.*, 1993).

KpsD is a 60 kDa protein with a typical signal sequence, suggesting that it is exported across the cytoplasmic membrane (Pazzani *et al.*, 1993). The periplasmic location of KpsD in K1 has been confirmed (Silver *et al.*, 1987) and processing of a precursor by leader peptidase demonstrated. A deletion within *kpsD* resulted in

polysaccharide within the periplasm, but polar effects may have been involved (Pazzani, 1992). Disruption of *kpsD* in K1 also resulted in periplasmic polysaccharide (Silver *et al.*, 1987; Wunder *et al.*, 1994). KpsD is therefore involved in translocation of the polysaccharide to the cell surface. KpsD is homologous over a short region with phosphatidyl glycerophosphate (PGP) B phosphatase (PgpB) which hydrolyses PGP, phosphatidic acid and lysophosphatidic acid (Icho, 1988). The homology is in the central hydrophilic region of PgpB thought to interact with the substrate and it is tempting to speculate that this reflects an interaction of KpsD with the phosphatidic acid-linked K5 polysaccharide. Unusually PgpB is located in the inner and outer membranes, with greater phosphatidic acid phosphatase activity in the outer membrane.

KpsU is a cytoplasmic CMP-KDO synthetase (CKS) enzyme, designated K-CKS and is responsible for the elevated levels of CKS activity at capsule-permissive temperatures (Finke *et al.*, 1990; Finke *et al.*, 1989). The protein has been purified and the N-terminal sequence is in agreement with that predicted from the nucleotide sequence (Rosenow *et al.*, 1995). KpsU is homologous with the product of the *kdsB* gene, a CMP-KDO synthetase (L-CKS) involved in LPS biosynthesis (section 1.2.2) and responsible for the low basal CKS activity observed at all growth temperatures (Strohmaier *et al.*, 1995). The activity of both enzymes is dependent on magnesium ions. At 30°C and with an optimal pH of 9.5 K-CKS has apparent K_m values for KDO and CTP of $2.0\pm0.4\times10^{-3}$ M and $2.5\pm0.5\times10^{-3}$ M, respectively and V_{max} values of $2.8\pm0.56\times10^{-6}$ $mol^{-1}min^{-1}mg$ and $3.0\pm0.6\times10^{-6}$ $mol^{-1}min^{-1}mg$, respectively. These kinetic parameters vary from those of L-CKS; the K-CKS K_m values are 10 times higher than those of L-CKS and the V_{max} values 100 times lower (Rosenow *et al.*, 1995). These differences may reflect the different functions of L-CKS which is required at all temperatures for the synthesis of LPS, an essential component of the cell wall and K-CKS which is involved in the synthesis of capsular polysaccharides which contribute to virulence and are only produced under certain conditions. Even when CTP and KDO are present at low concentrations L-CKS can catalyse the production of CMP-KDO for LPS but only when the substrates are abundant are they directed into the non-essential capsular polysaccharide synthesis by K-CKS. KpsU is also homologous with CMP-*N*-acetylneuraminic acid synthetase (NeuA) from *E. coli* K1 (Zapata *et al.*, 1989) and *N. meningitidis* group B (Edwards and Frosch, 1992; Ganguli *et al.*, 1994), which catalyses formation of the CMP derivative of a similar sugar.

KpsC is a 75 kDa hydrophilic protein, probably located in the cytoplasm. Mutations in *kpsC* result in full length intracellular polysaccharide not linked with lipid or KDO and located in areas of low electron density in the cytoplasm (Bronner *et al.*, 1993). Mutants lacking KpsS have the same phenotype (Bronner *et al.*, 1993). The mature wild-type K5 polysaccharide has phosphatidyl-KDO at its reducing end, although KDO is not in the repeat unit. The presence of KDO at the reducing terminus is a feature of group II polysaccharides. The location of the *kpsC* and *kpsS* genes in a region conserved among group II capsule gene clusters is consistent with them encoding proteins involved in the substitution with KDO and phosphatidic acid. The presence of full length polysaccharide in *kpsC* and *kpsS* mutants suggests that substitution with phosphatidyl-KDO is a requirement for translocation rather than initiation of biosynthesis as previously suggested (Boulnois and Jann, 1989).

KpsC and KpsS are homologous with the LipA and LipB proteins respectively of *N. meningitidis* group B (Frosch and Müller, 1993). Mutants lacking LipA or LipB accumulate capsular polysaccharide which lacks a lipid moiety in the cytoplasm. Frosch and Müller (Frosch and Müller, 1993) conclude that LipA and LipB are involved in the cytoplasmic phospholipid substitution of polysaccharide prior to export, supporting the role for KpsC and KpsS suggested above. KpsC is homologous with the LpsZ protein of *R. meliloti,* mutations in which affect the molecular weight range of a group II-like surface polysaccharide required for the Fix$^+$ phenotype of *R. meliloti* (Reuhs *et al.*, 1995). This polysaccharide contains KDO in the repeating unit and is a K antigen, not LPS as previously reported (prompting the redesignation of *lpsZ* as *rkpZ*). The phenotype of an *lpsZ* mutant differs from that of a *kpsC* mutant in that the same amount of polysaccharide is synthesised in the normal location and only the size is altered. KpsC is also homologous (77.9% similarity over 512 amino acids) with a 58 kDa protein from region 3 of the *H. influenzae* type b capsule gene cluster (Brophy, L., unpublished). KpsS is homologous (75.3% similarity over 373 aminio acids) with a 44 kDa protein from the same region (Brophy, L., unpublished).

Region 2 encodes the enzymes required for biosynthesis of the K5 polysaccharide (Table 1.2) and consists of four genes, KfiA-D and one 16 kDa ORF the putative product of which has not been observed and when disrupted results in no detectable phenotype (Petit *et al.*, 1995). Analysis of the amino acid sequence of KfiD (Petit *et al.*, 1995) revealed a nicotinamide adenine dinucleotide (NAD)-binding motif in the amino-terminal region and similarity with a number of dehydrogenase enzymes (Table 1.3), including UDP-Glc dehydrogenases from

Shigella flexneri (Udg) and *Streptococcus pyogenes* (HasB) and AlgD, a GDP-mannose dehydrogenase essential for alginate biosynthesis by *P. aeruginosa* (Deretic *et al.*, 1994). KfiD was subsequently shown to be an NAD-dependent UDP-Glc dehydrogenase, required for synthesis of UDP-Glucuronic acid (UDP-GlcA) (Sieberth *et al.*, 1995). UDP-GlcA is an essential substrate for the synthesis of K5 polysaccharide but is not otherwise required by *E. coli*; hence the enzyme required for synthesis of UDPGlcA is encoded within the capsule cluster. The other substrate, UDP-GlcNAc, is provided by housekeeping genes not specifically associated with capsule biosynthesis.

The amino acid sequence of KfiC is homologous with various glycosyl transferases (Table 1.3)(Petit *et al.*, 1995), including several Exo proteins involved in the biosynthesis of succinoglycan by *R. meliloti*, NodC involved in synthesis of D-glucosamine-containing lipooligosaccharide nodulation factors by *Rhizobium* spp (section 1.5.7) and HasA from group A streptococci which transfers glucuronic acid and glucosamine to the growing chain of hyaluronic acid (which is structurally very similar to the K5 polysaccharide (Vann *et al.*, 1981)). Incorporation of radiolabelled glucuronic acid and *N*-acetylglucosamine from their UDP-activated precursors (in the presence of exogenous K5 polysaccharide) by membranes from strains which over-expressed the cloned *kfiC* gene demonstrated that the KfiC protein has K5 transferase activity (Petit *et al.*, 1995). The *kfiC* and *kfiD* genes are co-transcribed (Petit *et al.*, 1995) as are the streptococcal *hasA* and *hasB* genes and probably the equivalent genes in heparin biosynthesis, leading to the suggestion that genetic linkage and co-regulation of a transferase with enzymes synthesising its precursors may be a general phenomenon (Sieberth *et al.*, 1995). KfiC is unable to initiate K5 polysaccharide biosynthesis. Likewise, the biosynthesis of heparan sulphate, which is similar to K5 polysaccharide, requires two GlcNAc transferases: One for initiation and another for polymerisation (Fritz *et al.*, 1994). The initiation of K5 polysaccharide synthesis is not understood and a separate enzyme, possibly encoded by *kfiA* or *kfiB* may be required to transfer the first sugar.

Western blot analysis has shown that KfiA, C and D localise to the inner membrane (Rigg, G. P. and I. S. Roberts, unpublished), supporting suggestions that a biosynthetic complex is located on the inner face of the cytoplasmic membrane (Fig. 1.4) (Roberts, 1996). This complex is thought to be closely linked with the export process and proteins encoded by regions 1 and 3 are required for the membrane targeting of the region 2 proteins. This interaction is also reflected by the observation that region 2 alone is not sufficient to direct polysaccharide synthesis (Bronner *et al.*, 1993). K5 transferase activity is substantially reduced in

the absence of KpsC or KpsS (Bronner *et al.*, 1993). The KpsS protein is required for transferase activity with acceptors present within the membrane (endogenous acceptors), but activity is at least partially restored when K5 polysaccharide is added to act as an exogenous acceptor. Endogenous activity is low in region 3 mutants in K1 but is also restored with exogenous acceptor (Pavelka *et al.*, 1994) A block in export apparently feeds back on the synthesis pathway reducing the number of acceptors available within the membrane.

Kröncke reported that mutations in K5 region 3 result in polysaccharide located in the cytoplasm which is shorter than surface polymer and lacks phosphatidic acid substitution (Kröncke *et al.*, 1990). This differs from the phenotype of mutants in the equivalent region C of the *N. meningitidis cps* cluster (see below and Fig. 1.5) in which polysaccharide was phospholipid-substituted and located at the inner face of the cytoplasmic membrane. This phenotype is consistent with the proposed role of LipA/LipB and KpsC/KpsS in phosphatidic acid substitution. The lack of substitution in K5 region 3 mutants may indicate a functional association between the region 3 products and KpsC and KpsS. In K1 region 3 mutants polysaccharide was located against the inner membrane, but was not tested for phosphatidic acid substitution (Pavelka *et al.*, 1994).

The presence of polysaccharide in the cytoplasm in region 3 mutants suggests a defect in the translocation of polysaccharide across the cytoplasmic membrane from its site of synthesis on the inner face of the membrane (Boulnois and Jann, 1989). A similar phenotype is observed when the membrane-potential decoupling agent carbonyl cyanide *m*-chlorophenylhydrazone (CCCP) is added shortly after a temperature upshift (Kröncke *et al.*, 1990). The polysaccharide export process therefore apparently requires an appropriate membrane potential. Nucleotide sequence analysis of region 3 revealed two genes, *kpsM* and *kpsT* (Smith *et al.*, 1990) which are homologous with the equivalent genes from K1 (Table 1.3) (Pavelka *et al.*, 1991) and probably form a single transcriptional unit. Comparison of the amino acid sequences of KpsM and KpsT with those of proteins in the database showed that they belong to a large family of proteins called the ATP-binding cassette (ABC) transporters (Higgins *et al.*, 1990)(Fath and Kolter, 1993). KpsT includes a highly conserved region comprising the characteristic ATP-binding cassette. ABC transporters couple the energy provided by ATP hydrolysis with translocation of substrate across membranes. The family can be divided into those systems involved in import into the bacterial cell (periplasmic permeases), eukaryotic transporters and bacterial exporters (Fath and Kolter, 1993). The basic ABC exporter system consists of a dimer of ABCs and a dimer of membrane-spanning domains (MSD) which may be on the same or

separate polypeptides. The hydropathy plot (Kyte and Doolittle, 1982) of KpsM indicates an integral membrane protein with six membrane-spanning domains. This topology has been confirmed for $KpsM_{K1}$ which has also been shown to form homodimers in the cytoplasmic membrane (Pigeon and Silver, 1994). It has been shown that $KpsT_{K1}$ binds ATP and that this function is required for transport of polysaccharide across the inner membrane (Pavelka *et al.*, 1994). KpsM and KpsT almost certainly comprise an ABC-transporter system for translocation of polysaccharide across the cytoplasmic membrane.

Amino acid sequence comparisons suggest that among the ABC transporters the KpsTM system is most closely related to the capsular polysaccharide export systems from *H. influenzae* (BexAB) (Kroll *et al.*, 1990), *N. meningitidis* (CtrDC) (Frosch *et al.*, 1991) and a proposed lipo-oligosaccharide export system in *Rhizobium leguminosarum* (NodIJ) (Vázquez *et al.*, 1993). Based on quantitative sequence comparisons these systems, together with the drug resistance transporter of *Streptomyces peucetius* (DrrAB), have been proposed to comprise an 'ABC-2' subfamily of ABC transporters (Reizer *et al.*, 1992).

It has been suggested that several of the bacterial ABC transport systems have a third component; the BexC, CtrD and KpsE proteins in the *H. influenzae, N. meningitidis* and *E. coli* systems, respectively (Reizer *et al.*, 1992). The degree of similarity between these proteins indicates that they are almost certainly homologous and therefore, especially considering their very similar topology, likely to perform similar functions. Mutations in *kpsE* result in polysaccharide located in the periplasm (Bronner *et al.*, 1993) and it has been suggested that KpsE accepts the polysaccharide on the outer face of the cytoplasmic membrane (Cieslewicz *et al.*, 1993)(Rosenow *et al.*, 1995). The genetic organisation of *bexC* and *ctrD* (Fig. 1.5) suggests that they are co-transcribed with their respective ABC transporter genes and are therefore likely to encode proteins with a closely related function. *kpsE* is in a different region from *kpsM* and *kpsT* and if their products do interact in polysaccharide export this poses the question of how the genes are co-ordinately regulated. Whether BexC, CtrD and KpsE form an integral part of the ABC transporters or are simply involved in the next step in polysaccharide export is unknown. It has also been suggested that the periplasmic protein KpsD, encoded by region 1 may be a periplasmic binding component of this transport system (Pavelka *et al.*, 1991). The K5 polysaccharide may be exported onto the cell surface at adhesion sites between the inner and outer membrane (Bayer and Bayer, 1994). Following a shift from the capsule-restrictive to capsule-permissive temperature (i.e. 18°C to 37°C) newly synthesised polymer

appears as tufts on the cell surface before eventually covering the entire surface (Jann and Jann, 1990).

An understanding of the functions of the genes in region 1 will be necessary to fully interpret the observed patterns of expression. Conversely, a detailed model of gene regulation may provide clues to gene functions.

1.3.5 Evolution of the *kps* locus

Common organisation

To investigate the evolution of the *E. coli kps* gene cluster (Fig. 1.3) it is helpful to examine the gene clusters required for expression of the group II-like capsules of *H. influenzae* and *N. meningitidis* which have a similar organisation. The central regions 2, II and A, of the *kps, cap* and *cps* loci respectively (Fig. 1.5) encode serotype-specific biosynthetic functions. Regions 3, I and C encode ABC transport systems involved in export of polysaccharide across the cytoplasmic membrane (section 1.3.4). The *E. coli kpsE* gene encodes a protein involved in export of the polysaccharide across the periplasm and unlike its homologues in the *cap* and *cps* clusters is separated from the genes for the integral membrane and ATP-binding ABC-transporter proteins (Fig. 1.5). Genes involved in phospholipid substitution are also conserved between the three clusters. Region E in *N. meningitidis* is regulatory while region D is involved in lipo-oligosaccharide production and was probably acquired late in the phylogeny. The presence of conserved genes and the common organisation of these three clusters suggests that they may have a common origin (Boulnois and Jann, 1989).

"Cassette" model for acquisition of biosynthetic genes

A model has been proposed in which a new biosynthetic region may occasionally be acquired *en bloc* from a different micro-organism (Boulnois and Jann, 1989). The idea of each region as a "cassette" was suggested by the low GC content of region 2 (33.4%) of the K5 cluster compared with region 1 (50.6%), region 3 (42.3%) (Pazzani, 1992) and the average GC content for *E. coli* (48-52%). Analysis of the *kpsS* and *kpsT* genes flanking region 2 in the *kps* cluster (Fig. 1.4) revealed that their 3' termini have a low GC content (Pazzani, 1992) and vary between serotypes (Drake, 1991)(Pavelka *et al.*, 1991)(Smith *et al.*, 1990). This suggested that different region 2 cassettes may have been acquired by homologous recombination between the ends of the conserved regions 1 and 3.

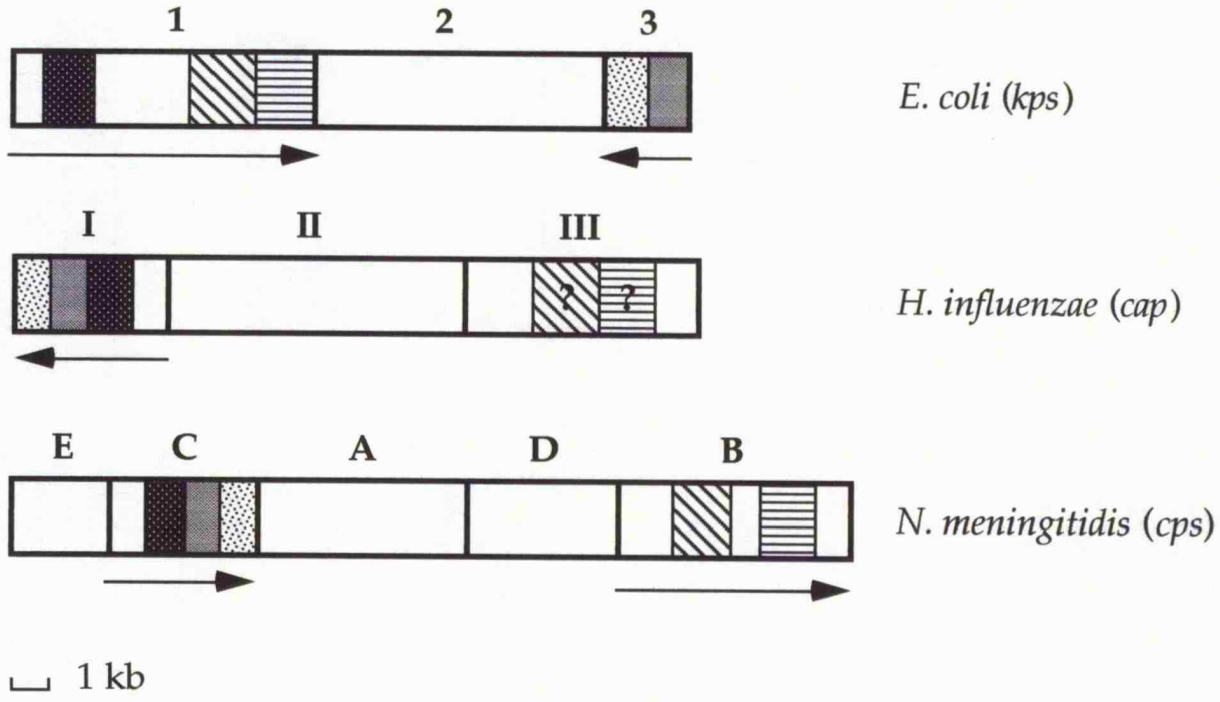


Figure 1.5. The conserved organisation of group II-like capsule gene clusters, illustrated by *E. coli* K5, *H. influenzae* type b and *N. meningitidis* group C. Genes with the same shading are homologous and encode, in *E. coli*, *H. influenzae* and *N. meningitidis*, respectively: KpsE, BexC and CtrB (black); KpsM, BexB and CtrC (medium stippling); KpsT, BexA and CtrD (light stippling); KpsC, the 58 kDa protein and LipA (sloping lines) and KpsS, the 44 kDa protein and LipB (horizontal lines)(the nucleotide sequence of region III of the *H. influenzae* cluster is unpublished (L. Brophy *et al*) and the details and exact location of the genes are unclear). Arrows indicate the direction of transcription. The functions of the regions and specific genes are described in the text.

Further evidence for the horizontal transfer of a biosynthetic region is provided by the *cap* locus of *H. influenzae.* Like the *kps* region 2, *cap* region II has a low GC content of 32% (Eldere *et al.*, 1995), lower than regions I (39%) and III (40%) which are comparable with the average for *H. influenzae* (38%) (Fleischmann *et al.*, 1995). The *H. influenzae* population has a clonal structure with two major phylogenetic groups (Musser *et al.*, 1988; Musser *et al.*, 1988). Type b strains are predominantly in division I, however a few are found in division II. The organisation of the type b *cap* locus is the same in each division (Kroll and Moxon, 1990). Comparison of an homologous pair of *cap* loci showed that the nucleotide sequence of region 1 had diverged by approximately 12% between divisions while serotype-specific region 2 was identical. This suggested that a cassette of serotype-specific DNA had been transferred within the bacterial population since the creation of the *cap* locus and too recently to have acquired sequence variation by genetic drift (Kroll and Moxon, 1990).

Chromosomal location

All *E. coli* group II gene clusters have the same chromosomal location near *serA* but this differs from the determinants of group I and the putative fourth group which are located near *his* (section 1.2.1 Classification). If these clusters are genetically related insertion sequence (IS) elements may have been involved in their dispersion. Again studies of the *H. influenzae cap* locus provide a possible model for how this could have occurred. The *cap* loci of strains in *H. influenzae* phylogenetic division I are flanked by direct repeats of IS element, IS*1016* (Kroll *et al.*, 1991). Those versions of IS*1016* for which the nucleotide sequence has been determined no longer encode a functional transposase. However, it is suggested that in the past IS*1016* transposed to both sides of the *cap* locus forming a compound transposon (Kroll *et al.*, 1991). This enabled the cluster to move to its current location in division I strains, which differs from that in division II strains (Kroll and Moxon, 1990). The *cap* loci of all division I strains examined have the same flanking DNA (Kroll *et al.*, 1989) suggesting that transposition has been very rare and that serotype-specific genes were acquired after this event. Homologous recombination between IS*1016* repeats can result in capsule-deficient mutants or amplification of the *cap* locus. However, spontaneous *E. coli* capsule-deficient mutants are not observed, as would be predicted if an *Haemophilus*-type system existed. Determination of the nucleotide sequence flanking the *kps* gene cluster is required to investigate the presence of IS elements. IS elements may have mediated transfer of single genes or regions, indeed the relic of an IS2 element is present in region 2 of the K5 cluster (Rigg, G. P., unpublished observation).

1.4 Bacterial gene expression and regulation

1.4.1 Introduction

Expression of group II capsules, including the K5 antigen is regulated by temperature (Ørskov *et al.*, 1984) and probably also by other environmental factors. The levels of the proteins required for capsule expression must be maintained at appropriate levels under capsule-permissive conditions and reduced under capsule-restrictive conditions. This chapter outlines the parameters which affect gene expression and the levels at which regulation can occur. Each protein must be maintained at a specific cellular concentration. The intrinsic values of a number of parameters determine this concentration:

1) Gene copy number. This may be increased by amplification. For example, duplication of the *cap* locus of *H. influenzae* approximately doubles capsular polysaccharide production (Kroll and Moxon, 1988). Location of a gene near the origin of replication increases its effective copy number in rapidly dividing cells (Dorman, 1994).

2) Concentration of mRNA. The concentration of viable mRNA (i.e. that containing intact coding regions) depends on the rate of transcription initiation and termination (before the end of the gene)(section 1.4.2) and the rate of degradation (section 1.4.3).

3) Translation efficiency (section 1.4.4). Initiation is generally the rate-limiting step, although codon usage can influence translational elongation (McCarthy and Gualerzi, 1990). The Shine-Dalgarno (SD) sequence is the most important feature determining the efficiency of translation initiation, but other primary sequence elements may be involved (McCarthy and Gualerzi, 1990).

4) Protein stability and regulation of activity. Most bacterial proteins are stable with half-lives longer than one cell generation. The level of functional protein may be altered by protein modification or by other interactions (section 1.4.5).

In the absence of regulation the above parameters will determine the constitutive level of the protein, but because they mostly result from the association and interaction of diffusing molecules there will be stochastic variations. These are particularly significant at the level of mRNA for which the renewal time is similar to the mean interval between initiation events (Axe and Bailey, 1993). Regulatory mechanisms are therefore required to prevent

unacceptable deviations from the mean protein concentration. These may be at the level of transcription (sections 1.4.2 and 1.4.3), translation (section 1.4.4) and/or post-translation (section 1.4.5).

The importance of regulatory mechanisms in maintaining stable steady state protein concentrations was demonstrated by a model, based on experimental results, for the expression of *rpoB* which encodes the ß subunit of RNA polymerase (Axe and Bailey, 1993). Levels of RpoB were unstable in the absence of regulation but the size of deviations from the mean were reduced and the rate at which levels returned to the mean after a displacement were improved by an autogenous transcriptional regulatory mechanism (i.e. reduced levels of RpoB increased transcription of *rpoB* mRNA and vice-versa)(Steward and Linn, 1992). The level of RpoB maintained in the model was further stabilised by the addition of a simple translational regulatory mechanism whereby binding of RNA polymerase to *rpoB* mRNA prevented translation.

In addition to maintaining constitutive protein concentrations regulation is required to alter gene expression in response to changes in growth conditions and the external environment. *E. coli* K5 begins expressing capsular polysaccharide when moved from 18°C to 37°C. The cell must sense the change in environment and respond by altering the expression of specific genes. This regulation may be at several levels but a mechanism preventing mRNA production at the restrictive temperature would be most energetically favourable for the cell. A new environment may require a co-ordinated change in the expression of various genes, for example the activation of virulence genes upon entering the host (section 1.4.6). Shifting conditions may also require more general or global changes in gene expression (section 1.4.7).

1.4.2 Transcriptional regulation

The average half-life of an mRNA molecule in *E. coli* is 2-4 minutes (Belasco, 1993). This rapid turn-over enables a rapid response to environmental change. Transcriptional regulatory mechanisms are reviewed by Dorman, C. J. (Dorman, 1994) and may affect the rate of initiation or termination (at sites before the end of the full length mRNA). The concentration of mRNA is equally dependent on the rate of degradation (section 1.4.3).

Initiation of transcription involves recognition of the promoter region by RNA polymerase holoenzyme (α_2ßß'σ) and the reversible formation of a closed complex (Hippel *et al.*, 1982). In an irreversible step involving local DNA

melting this becomes an open complex. Following repeated rounds of abortive initiation the sigma (σ) factor is released, the polymerase clears the promoter region and the core enzyme begins processive elongation of RNA. The forward rate constants for the formation of the closed and open complexes indicate the strength of the promoter. These are determined largely by the sequence of the promoter region. The consensus sequence required for recognition by σ^{70}, the most common σ factor has been determined (Hippel *et al.*, 1982). This includes the -10 (TATAAT) and -35 (TTGACA) boxes. In general, the closer the sequence of a promoter is to this consensus the greater the rate of initiation it will direct. The spacing of recognition elements is critical and deviation from the consensus 17 bp reduces promoter activity (O'Neill, 1989). Alternative σ factors enable RNA polymerase to recognise different promoter sequences. By controlling production of an alternative σ factor the cell can regulate a group of genes which have the appropriate promoter sequences (section 1.4.6).

Transcription is both positively and negatively regulated by factors which bind in the promoter region. The classic example is the *lac* operon in which the LacI repressor bound to an operator site overlapping the promoter interacts with RNA polymerase preventing it from leaving the initial transcribing complex. Binding of allo-lactose to the repressor causes it to dissociate, enabling induction of ß-galactosidase in the presence of lactose. There are many similar examples of repressors binding to the promoter region and precluding RNA polymerase. Regulation is often enhanced by cooperative interactions between binding proteins. Transcription from several promoters, including *lac*, is positively regulated by the cyclic AMP receptor protein (CRP), which binds upstream of the promoter region inducing a DNA bend (Pérez-Martin *et al.*, 1993). Protein interactions between CRP and RNA polymerase activate transcription initiation and the DNA bending may aid formation of the closed complex, which itself involves bending of the DNA by RNA polymerase. In some cases CRP-induced bending allows upstream DNA sequences (or activators) to contact the back of the enzyme, promoting open complex formation. The importance of DNA bending in CRP activation has been demonstrated by the functional replacement of CRP-binding sites by statically curved DNA. Other regulatory factors which bend DNA include integration host factor (IHF) and the leucine-responsive regulatory protein (LRP) (Dorman, 1994).

DNA may be intrinsically bent and A tracts in phase with the DNA helical repeat are the main determinant (see chapter 3). Curved DNA is often located upstream of bacterial promoters and may favour the wrapping of DNA around RNA polymerase required during initiation (Pérez-Martin *et al.*, 1993). The presence of

an upstream bend has been shown to correlate with strong promoter activity (Plaskon and Wartell, 1987) and is dependent on orientation relative to the RNA polymerase binding site as insertions of non-integral helical turns produce mutant promoters. DNA loop formation allows proteins bound at distant *cis*-acting enhancer sites to interact with RNA polymerase (Matthews, 1992).

DNA supercoiling can enhance transcriptional initiation by assisting in the unwinding of DNA during open complex formation (Ehrlich *et al.*, 1985). Changes in DNA twist may enable -10 and -35 sequences with spacing varying from the standard 17 bp to align on the same side of the helix and interact with RNA polymerase. This feature may be exploited in the regulation of such promoters. Supercoiling-sensitive promoters can be affected by the transcription of neighbouring genes, which alters local supercoiling levels (Chen *et al.*, 1992). The effect of decreased supercoiling has been studied *in vivo* by inactivation of topoisomerases and shown to have diverse effects on many promoters. DNA methylation can potentially affect protein-DNA interactions and the Dam methylation pattern at certain GATC sites can regulate gene expression (Hale *et al.*, 1994). Transcription of Pap pili and F1845 fimbrial adhesin depends on the methylation state of GATC sites in their promoter regions (Woude *et al.*, 1992)(Bilge *et al.*, 1993).

Transcription termination has been reviewed by Holmes, W. M. (Holmes *et al.*, 1994) and more recently by Richardson, J. P. (Richardson, 1993). The efficiency of transcription termination at sites within a polycistronic operon can control the expression of downstream genes. Termination involves pausing of the RNA polymerase core enzyme elongation complex to form an unstable termination complex which dissociates from the DNA template and releases the RNA transcript (Hippel and Yager, 1992). This may occur either at sites that depend only on the nucleic acid sequence (Rho-independent terminators)(Reynolds *et al.*, 1992) or at Rho-dependent terminators which require the Rho protein to release the paused transcription complex from the template. Rho-independent terminators consist of a GC rich stem-loop followed by a run of U residues while Rho-dependent terminators have less well defined characteristics which are discussed in chapter 5.

In the process of attenuation the interaction between ribosomes and nascent RNA is used to control the efficiency of a terminator, thereby regulating the expression of downstream genes. For example, when ribosomes are able to translate the leader sequence in the tryptophan operon the terminator works efficiently, preventing expression of the tryptophan biosynthetic genes. In the

absence of tryptophan the ribosome stalls at two *trp* codons in the leader RNA, enabling the RNA to adopt an alternative structure which prevents formation of the terminator stem-loop and reduces termination efficiency. Latent intragenic Rho-dependent terminators provide another mechanism to link transcription with translation. Under conditions of stress, when translation is blocked, the nascent RNA is free of ribosomes enabling the binding of Rho and subsequent termination (this model also explains the polarity of nonsense mutations (Adhya and Gottesman, 1978)). This mechanism saves energy by preventing the synthesis of transcripts which are not being translated.

Rho-dependent terminators provide a site for regulation by antiterminator proteins. For example the BglG protein of the *bgl* operon, which is involved in the catabolism of aromatic ß-glucosides, prevents termination by binding to the RNA at two terminator sites within the operon. This antitermination function is linked with transport of ß-glucosides through Enzyme II Bgl (BglF) which in the absence of ß-glucosides inactivates BglG by phosphorylation.

Regulation of termination is important in the development of bacteriophage λ and involves proteins N and Q (Richardson, 1993). Protein Q only directs antitermination of transcripts from a specific promoter and is enhanced by the host factor NusA. Protein N only recognises transcripts with a *nut* sequence and antitermination involves formation of a complex containing various host factors including NusA (Li *et al.*, 1993). NusA may be involved in loading N and Q onto RNA polymerase, preventing Rho-mediated termination.

These examples demonstrate that regulation of termination is an important but complex and poorly understood mechanism of regulating polycistronic operons.

1.4.3 mRNA stability

Bacterial mRNA half-lives vary from approximately 30 sec to 50 min (Belasco, 1993) and control of degradation is equally important as initiation in determining mRNA levels. The half-lives of transcripts do not depend on their length, indicating that degradation is not due simply to random cleavage. Four main ribonucleases have been identified: Two 3' to 5' exoribonucleases, RNase II and Polynucleotide phosphorylase; and two endoribonucleases, RNase E and RNase III. The cleavage site specificities of the endoribonucleases are poorly defined but RNase E cleaves in single-stranded AU-rich regions (Melefors *et al.*, 1993) while RNase III cleaves within certain double helical regions, including stem-loops (Court, 1993). Other features of the mRNA affect the efficiency of

cleavage. A 3' stem-loop is required to prevent immediate cleavage by 3' exonucleases and additional intercistronic stem-loops protect upstream cistrons from exonuclease digestion following removal of the 3' terminal stem-loop by endonucleolytic cleavage. Stem-loops, such as that provided by the common repetitive extragenic palindromic (REP) sequence (Newbury *et al.*, 1987) efficiently impede 3' exonucleases and mRNA degradation is often initiated by endonucleolytic cleavage, which is usually the rate-limiting step (Axe and Bailey, 1993).

Several 5' mRNA stabilisers have been defined by their ability to confer stability to heterologous sequences that are fused downstream (Bechhofer, 1993). Stabilisation of the long-lived *ompA* mRNA involves formation of a secondary structure in the 5' untranslated region which includes base pairing at the extreme RNA 5' terminus (Hansen *et al.*, 1994). No 5' to 3' exoribonuclease has been identified in *E. coli* and it is likely that 5' terminal base pairing increases mRNA stability by impeding access of RNase E to downstream cleavage sites, as has been demonstrated for RNA I (Bouvet and Belasco, 1992). The efficiency of translation can affect mRNA stability, but this process is poorly understood (Petersen, 1993). The presence of ribosomes appears to protect mRNA from degradation.

Individual cistrons within a polycistronic mRNA may have different half-lives as demonstrated for the *puf* mRNA of *Rhodobacter capsulatus* (Klug, 1993). This differential stability results in differential expression of the encoded products, components of the photosynthetic apparatus, and contributes to their stoichiometry. The difference in stability of individual segments depends on the distribution of decay promoting and impeding elements, including an RNase E-like cleavage site (Fritsch *et al.*, 1995). Control of mRNA stability is also involved in the oxygen-dependent regulation of these genes (Klug, 1993).

The pilin structural gene *papA* and the regulatory gene *papB* are encoded on the same mRNA, but the more rapid decay of the *papB* region results in differential expression of the two proteins (Baga *et al.*, 1988). The F1845 fimbrial adhesin operon provides a good example of mRNA processing. The *daaE* gene encoding the structural subunit is located downstream of four genes, but is encoded on a stable 1.3 kb transcript produced by endonucleolytic cleavage of a longer polycistronic transcript (Bilge *et al.*, 1993). An antisense RNA spanning the processing site is required to promote cleavage (Cantwell *et al.*, 1994).

1.4.4 Translational regulation

Translation involves initiation, elongation and termination. Although the rate of elongation is influenced by codon usage the most important point for regulation is initiation. The efficiency of initiation is determined by interactions between the 30S ribosomal subunit and the SD sequence, AUG codon and additional recognition elements reviewed by Mc Carthy, J. E. G. and C. Gualerzi (McCarthy and Gualerzi, 1990). The secondary structure of the translational initiation region may affect the efficiency of initiation by forming recognition motifs or by altering the accessibility of recognition sequences. The variety of regulatory mechanisms involving inhibition of initiation are indicated by the examples below.

Binding of a regulatory protein may prevent access of the ribosome to the SD sequence. For example, binding of RpoB to a promoter-like sequence spanning the SD in its own mRNA provides a mechanism for autogenous translational regulation (Axe and Bailey, 1993). Ribosomal protein S4 binds to a pseudoknot which encompasses the SD sequence of the α operon mRNA. Rather than blocking translation by precluding ribosome binding S4 traps the mRNA in a conformation able to bind 30S subunits but unable to form an initiation complex (Spedding and Draper, 1993). Translation of the *cob* operon, which encodes enzymes for biosynthesis of vitamin B_{12}, is repressed by sequestration of the SD and AUG codon in an RNA hairpin (Richter-Dahlfors *et al.*, 1994)(Richter-Dahlfors and Andersson, 1992). In the absence of B_{12} an upstream translational enhancer is thought to unfold the hairpin and permit translation. The mechanism whereby translational control of the first gene mediates repression of the whole operon is discussed in chapter 6. When pyrimidines are in excess translation of the *pyr*C gene, which is involved in pyrimidine biosynthesis, is also blocked by formation of a hairpin overlapping the SD sequence (Liu and Turnbough, 1994). However when pyrimidines are limiting and CTP levels are low transcription initiates 3 bases downstream, after two C residues. The shorter transcript no longer forms the inhibitory hairpin and high levels of PyrC are synthesised.

LcrF and VirF are closely related homologues which activate transcription of virulence genes in response to temperature in different *Yersinia* spp (Hoe and Goguen, 1993). The LcrF protein is present at higher levels at 37°C than 26°C although it is stable at both temperatures and the *lcrF* gene is transcribed at similar levels. It has been proposed that temperature regulation is achieved by a stem-loop structure which sequesters the translation initiation region of the *lcrF*

mRNA at 27°C but at 37°C is sufficiently destabilized to enable efficient translation initiation (Hoe and Goguen, 1993).

Antisense mRNAs complementary to the translational initiation region are common in plasmid incompatibility systems (Siemering *et al.*, 1993) and in the regulation of related chromosomal genes (Poulsen *et al.*, 1991). The antisense RNA binds to the cognate mRNA and prevents ribosome binding and translation of a toxic product, which is however synthesised in cells which have not received the plasmid at segregation and no longer produce the short-lived inhibitory antisense RNA. The *ompC* gene is regulated by an antisense RNA, *micF* which is encoded upstream and transcribed in the opposite orientation (Green and Inouye, 1986).

The translational efficiencies of genes in a polycistronic operon, such as that predicted for region 1 of the K5 capsule cluster, are often not independently determined. The efficiency of translation initiation may be coupled with translation of an upstream gene (McCarthy and Gualerzi, 1990). The *atpHA* genes of the *atp* operon, which encodes subunits of the proton-translocating ATP synthase, are translationally coupled, with *atpA* being translated at least three times more efficiently than *atpH* (Rex *et al.*, 1994). The mechanism involves formation of an mRNA conformation which sequesters the *atpA* translational initiation region, inhibiting *de novo* initiation. However, ribosomes translating *atpH* induce this inhibitory structure to refold into an open conformation which promotes both *de novo* and re-initiation of translation. A similar mechanism is thought to underlie translational coupling of ribosomal subunit genes.

1.4.5 Post-translational regulation

Various processes control protein activity and therefore alter the effective concentration. These include phosphorylation, protein-protein and other interactions, proteolytic cleavage and changes in location. Although most proteins are relatively stable there is a subset of proteolytically unstable, short-lived proteins, often involved in regulatory circuits (Gottesman and Maurizi, 1992). For example, RcsA which is rapidly degraded by the Lon protease, positively regulates colanic acid synthesis (section 1.5.2). The activity of many alternative σ factors is controlled by the binding of specific anti-sigma factors which prevent unwanted transcription of the cognate genes (Brown and Hughes, 1995). Of particular relevance to the regulation of capsular polysaccharide

synthesis is the inactivation of sialic acid synthase below 20°C (section 1.5.3) (Merker and Troy, 1990).

1.4.6 Environmental regulation

Regulatory networks

For a bacterium to adapt to a change in environment a new pattern of gene expression is required, involving the activation of some genes and the repression of others. Regulatory networks have evolved to coordinate these changes. A group of genes and operons controlled by the same regulatory protein is called a regulon (Neidhardt, 1987)(Miller *et al.*, 1989). Although each regulon is controlled by separate regulatory circuits there is evidence for cross-talk between signal-transducing systems (Gross, 1993). Individual genes or operons may belong to more than one regulon and be controlled by interacting regulatory systems in response to various stimuli. Regulatory networks also provide a way for individual systems that recognise single signals to combine and control expression of a particular gene(s) such that it is expressed only when various environmental stimuli occur simultaneously. For example, the production of EPS I by *Pseudomonas solanacearum* requires activation by at least three separate signal transduction systems (Huang *et al.*, 1995).

Regulation of virulence genes

Invasion of the host involves induction of a number of virulence genes organised in regulons and the expression of K5 polysaccharide is presumably coordinated with that of other virulence genes. Identification of sequences involved in regulation of region 1 of the K5 capsule cluster may identify similarities with known regulatory systems. Conversely analysis of region 1 expression in known regulatory mutants may reveal which networks are involved with capsular polysaccharide expression. Some of the best characterised regulatory systems displaying mechanisms likely to be involved in regulation of capsular polysaccharide are outlined below.

Regulons are usually controlled at the level of transcription (section 1.4.2). Alternative σ factors enable RNA polymerase to recognise groups of promoters which share a specific element not recognised by σ^{70}. Regulation of a specific σ factor couples transcription of a regulon with a particular environmental parameter, for example nitrogen availability (σ^{54} or RpoN), temperature (heat shock) (σ^{32} or RpoH), growth phase/starvation (σ^{38} or RpoS). While most alternative σ factors resemble σ^{70} the structure and mode of action of σ^{54} is different (Merrick, 1993). The σ^{54} recognition boxes (TGGCAC-N5-TTGCa/t) are unusually close at -12 and -24. To activate transcription σ^{54} requires specific

activator proteins which bind at enhancer-like elements 100-200 bp upstream of the promoter and contact the σ^{54} holoenzyme by DNA loop formation (Morett and Segovia, 1993). The importance of alternative σ factors in the control of bacterial gene expression is indicated by the continuing discovery of many new putative σ factors.

At each step of infection different combinations of virulence factors are required. With the exception of iron the cues for expression are general environmental factors such as temperature, osmolarity, pH, O_2 and CO_2 (Mekalanos, 1992). The change in environment is related to gene expression by a signal transduction pathway, often a two-component system (Gross, 1993). A sensory protein spanning the cell membrane responds to a specific stimulus by autophosphorylating a histidine residue in its conserved cytoplasmic domain. The phosphate is then transferred to a cytoplasmic response regulator protein, typically enabling it to activate gene expression. Examples involved in regulation of virulence factors include the *bvgA* and *bvgS* genes of *Bordetella pertussis* (Miller *et al.*, 1989) which respond to temperature and PhoP/Q in *S. typhimurium* which regulate genes involved in survival inside macrophages, probably in response to low pH or starvation (Gross, 1993). Two-component systems in *E. coli* include the EnvZ/OmpR proteins which control outer membrane protein expression in response to osmolarity, and RcsB/RcsC which regulate colanic acid synthesis (section 1.5.2). The regulation of several virulence factors in *Vibrio cholerae* is controlled by the ToxR protein in response to temperature, osmolarity, pH and the presence of amino acids (DiRita, 1992)(Gross, 1993). The sensory and regulatory functions are combined in the ToxR protein which can sense a stimulus in the periplasm (with the aid of another protein, ToxS) and dimerise, enabling it to bind DNA and activate transcription.

Host tissues are iron-restricted and under these conditions bacteria must express various iron-sequestering systems. Expression of genes in the *E. coli* iron regulon is controlled by the Fur protein in response to intracellular levels of iron. When complexed with Fe(II) the Fur protein can bind DNA at an 'iron box' operator sequence upstream of iron-regulated genes, repressing transcription. When iron is restricted Fur becomes inactive and the genes are expressed. Fur therefore both senses the change in environment and in response regulates directly gene expression.

The AraC family comprises a large group of proteins which regulate virulence genes or carbohydrate metabolism (Dorman, 1994). They include CfaD from *E.*

coli and virF from *Shigella* (Dorman, 1992). These proteins have a conserved C-terminal helix-turn-helix DNA-binding domain, which in the case of AraC has been shown to bind to an operator site and repress transcription. The N-terminal domains are less conserved, presumably reflecting the ability to respond to different environmental signals.

1.4.7 Global regulation

The changes in gene expression which occur throughout the genome following a shift in conditions have been demonstrated at the protein and mRNA level (Chuang *et al.*, 1993). The various regulons which respond to a particular stimulus comprise a stimulon (Neidhardt, 1987). In addition to regulation of specific genes in response to a particular stimulus there is evidence for more general changes in gene expression. Control of mRNA degradation may be used to match overall RNA turnover to growth conditions. In *E. coli* during anaerobiosis bulk mRNA stability increases and RNA synthesis decreases, maintaining mRNA levels while saving energy (Georgellis *et al.*, 1993). The rate of transcription and translation also alters with growth rate (Vogel and Jensen, 1994)(Jacques *et al.*, 1992).

The mechanisms of global regulation are unclear. Changes in the environment, such as temperature, alter levels of DNA supercoiling which have a very broad effect on transcription (Dorman, 1994)(Dorman, 1995). Proteins that affect DNA topology, such as the histone-like protein H-NS have very pleiotropic effects on gene expression. H-NS often represses temperature-regulated genes which are activated by AraC-like regulators.

1.5 Regulation of extracellular polysaccharide expression

1.5.1 Introduction

The regulatory mechanisms controlling expression of extracellular polysaccharides are poorly understood. However various studies, in particular of the production of colanic acid by *E. coli* and alginate by *P. aeruginosa,* are revealing the complex interactions between various regulatory systems which determine the level of polysaccharide expression in response to environmental requirements. The principles and mechanisms involved are likely to be similar to those regulating expression of the *E. coli* K5 capsule.

1.5.2 Group I capsules

Although far from complete, understanding of the regulation of colanic acid expression is more advanced than that of other extracellular polysaccharides (Gottesman and Stout, 1991). Colanic acid (section 1.2.4) is a group IA capsular polysaccharide (section 1.2.1). It is expressed from the *cps* gene cluster by most *E. coli* strains at 20°C, but only weakly at 37°C. The regulatory system is complex and demonstrates how multiple, interacting mechanisms can affect expression of a particular set of genes (Fig. 1.6).

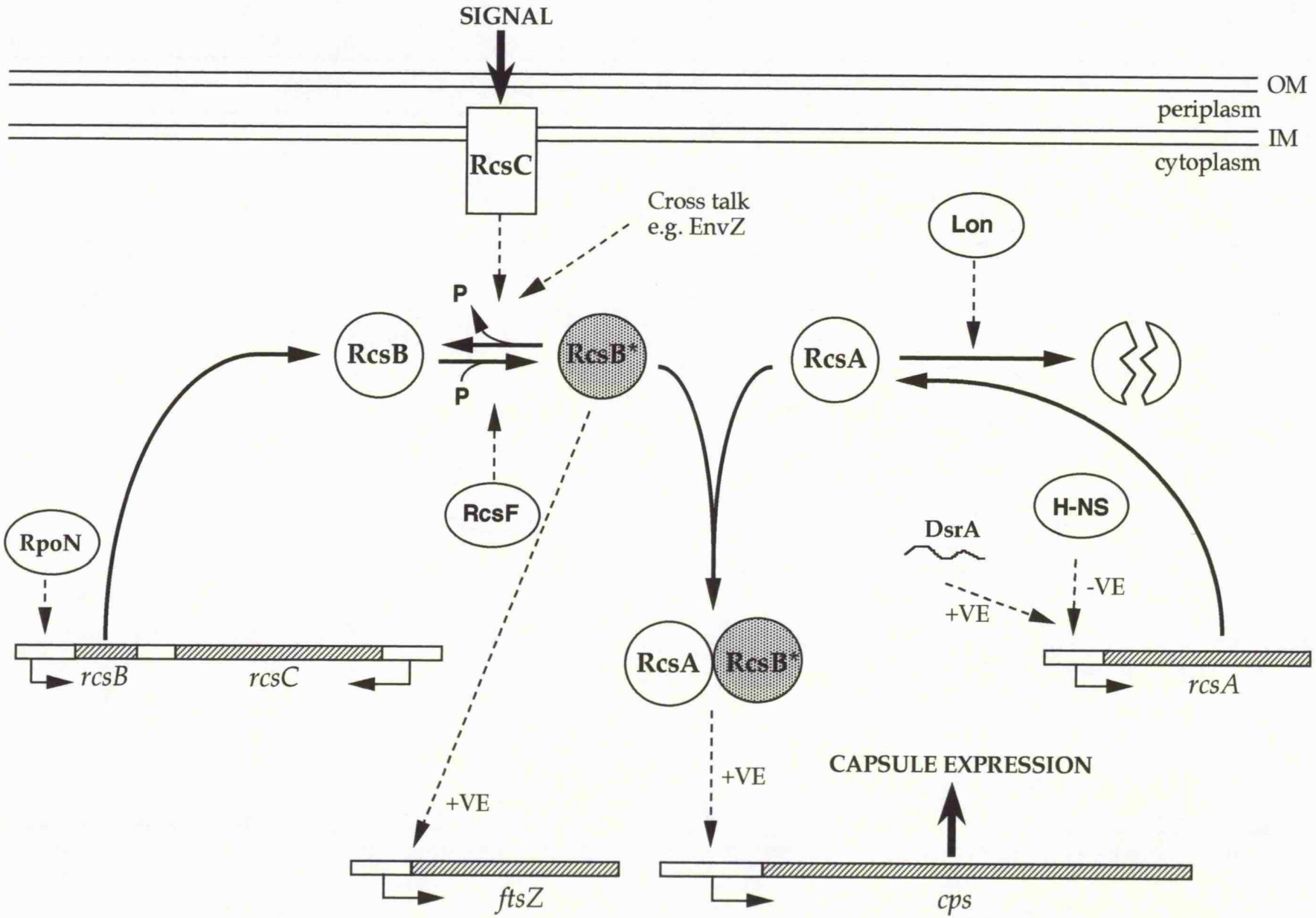
Transcription of *cps* is regulated by a two-component system (section 1.4.6) comprising the regulation of capsule synthesis RcsC sensor and RcsB effector proteins (Fig. 1.6)(Stout and Gottesman, 1990). RcsC is a transmembrane protein with a periplasmic domain which senses an undetermined environmental signal (perhaps temperature). This signal is transferred to a characteristic cytoplasmic protein kinase domain homologous with the sensors from other two-component systems including EnvZ and PhoM (Stout and Gottesman, 1990). The N-terminal domain of RcsB is homologous with other effectors including OmpR and NtrC. By analogy with other two-component systems it was proposed that, in response to the appropriate signal, RcsC activated RcsB by phosphorylation, although this has not been demonstrated directly (Gottesman and Stout, 1991). However, the activation of RcsB by effectors from other systems (eg. EnvZ) was reduced by multicopy *rcsC* suggesting that RcsC is primarily a phosphatase (Gervais *et al.*, 1992). A gene, *rcsF*, has been identified which stimulates colanic acid biosynthesis in the wild type but not in an *rcsB* mutant and it has been proposed that RcsF is a kinase which promotes phosphorylation of RcsB (Gervais and Drapeau, 1992). The MucZ protein stimulates RcsB/C-dependent colanic acid expression and may be involved in activation of RcsB (Zuber *et al.*, 1995).

Another positive regulatory protein, RcsA, is involved in colanic acid expression (Stout *et al.*, 1991). RcsA is rapidly degraded by the Lon protease (Fig. 1.6) and high levels accumulate in a *lon* mutant causing overexpression of colanic acid (Gottesman and Maurizi, 1992). Transcription of *rcsA* is silenced by H-NS but this effect is counteracted by a small antisense RNA (DsrA) (Fig. 1.6) which is transcribed from a region downstream of *rcsA* (Sledjeski and Gottesman, 1995). DsrA interferes with the ability of H-NS to silence, but the mechanism is unclear as it does not act on the *rcsA* transcript, is not homologous with the promoter and does not affect H-NS levels. Controlling the level of RcsA and its interactions with RcsB could contribute to the regulation of *cps* expression.

Homologues of RcsA have been cloned from *Klebsiella aerogenes* (Allen *et al.*, 1987) and *Erwinia amylovora* (Coleman *et al.*, 1990) and shown to stimulate expression of their extracellular polysaccharides (Bugert and Geider, 1995), which resemble group I *E. coli* capsules (Jann and Jann, 1990). These *rcsA* genes enhance colanic acid biosynthesis when transferred to *E. coli* K-12 (Allen *et al.*, 1987)(Coleman *et al.*, 1990). This indicates that the *rcs* system for regulation of polysaccharide expression is conserved among these bacteria. Another protein from *Klebsiella pneumoniae*, RmpA2, enhances colanic acid expression in *E coli* K-12 and serotype-specific capsular polysaccharides in *K. pneumoniae* (Wacharotayankun *et al.*, 1993). RmpA2 may be an additional transcriptional regulator to RcsA and RcsB because while the termini resemble RcsA the central domain is similar to NtrC.

Expression of *cps* is dependent on RcsB while RcsA is an accessory regulatory protein. RcsA and RcsB both have helix-turn-helix motifs at their carboxy termini and are members of the LuxR family of regulatory proteins (Stout *et al.*, 1991). The observation that RcsA was stabilised by RcsB led to the accepted model whereby the two proteins interact to activate transcription of *cps*. In a Lon^+ host with low levels of RcsA, activation of *cps* transcription depends on the proportion of RcsB which is activated (phosphorylated). In a *lon* mutant host non-phosphorylated RcsB can interact with the abundant RcsA to direct high level *cps* transcription. When RcsA is limiting and RcsB is not phosphorylated the cps genes are not transcribed (Gottesman and Stout, 1991).

Figure 1.6. The proteins involved in the regulation of colanic acid expression in *E. coli* and their proposed interactions (adapted from (Whitfield and Valvano, 1993)). An environmental signal is sensed by the inner membrane (IM) protein RcsC which is thought to respond by dephosphorylating activated RcsB* to the less active RcsB form (P represents a phosphate group). The positive regulatory protein RcsF catalyses the opposite reaction. RcsB may also be phosphorylated with low efficiency by the sensor proteins from other two-component systems (cross talk). RcsB* interacts with RcsA forming a heterodimer which activates transcription of the *cps* genes (hashed box). RcsB* also activates transcription of the *ftsZ* gene which encodes a protein required for cell division. The levels of the unstable RcsA protein are determined by the proteolytic activity of the Lon protease. Interaction between different regulatory systems is demonstrated by the various factors determining the level of transcription of the *rcs* genes. The H-NS protein represses expression of *rcsA* and is counteracted by the DsrA RNA. The *rcsB* and *rcsC* genes are transcribed towards each other and one of the several promoters upstream of *rcsB* is recognised by RpoN.


SIGNAL
OM
periplasm
IM
cytoplasm
RcsC
Cross talk
e.g. EnvZ
Lon
P
P
RcsB
RcsB*
RcsA
RcsF
RpoN
H-NS
DsrA
-VE
+VE
rcsB
rcsC
RcsA
RcsB*
rcsA
+VE
+VE
CAPSULE EXPRESSION
ftsZ
cps

The regulatory mechanisms involved in colanic acid expression are likely to apply broadly to group IA capsular polysaccharides whose expression has already been shown to be increased by multicopy *rcsB* (Jayaratne *et al.*, 1993). However, RcsA and RcsB are not essential for group I capsule expression but are required for high levels of synthesis (Jayaratne *et al.*, 1993). In strains expressing group IB capsules colanic acid expression was elevated in response to *rcsB*, but it was not determined whether levels of the group IB polysaccharides were also elevated (Jayaratne *et al.*, 1993). The length of group IB but not IA polysaccharides is regulated by the Rol protein which also affects O-antigen chain length (section 1.5.4) (Dodgson *et al.*, 1996).

1.5.3 Group II capsules

Production of specific precursors is an efficient point for regulation of capsule expression. This is often achieved by feedback inhibition of the precursor upon the first committed step of the biosynthetic pathway, as occurs in LPS biosynthesis (Whitfield and Valvano, 1993). *E. coli* K1 expresses a group II polysialic acid capsule (Table 1.2) but synthesis of NeuNAc (sialic acid) by *E. coli* K1 is not inhibited by CMP-NeuNAc (Vimr and Troy, 1985). However, intracellular levels of sialic acid are regulated by an *N*-acylneuraminate pyruvate-lyase (Vimr and Troy, 1985)(Vimr and Troy, 1985). The *N. meningitidis* group B capsule is chemically identical to that of *E. coli* K1. However availability of precursor may be controlled by a CMP-NeuNAc hydrolase (Masson and Holbein, 1983).

E. coli group II capsular polysaccharides are expressed at 37°C but not at 18°C (section 1.2.1). CMP-KDO synthetase expression is elevated at capsule-permissive temperatures (Finke *et al.*, 1990) and the *kpsU* gene which encodes the K-CKS enzyme has been located in region 1 of the *E. coli* K5 *kps* cluster (section 1.3.4) (Pazzani *et al.*, 1993). K-CKS was active *in vitro* at 37°C and 18°C and the lack of activity at capsule-restrictive temperatures is not therefore due to cold sensitivity of the enzyme (Rosenow *et al.*, 1995). However it is not known whether expression of *kpsU* is regulated at the level of transcription or translation. The transcriptional organisation of region 2 of the *E. coli* K5 *kps* cluster has been determined recently (Petit *et al.*, 1995). The level at which the expression of the biosynthetic enzymes were regulated was investigated (Petit, C. P., unpublished). While some genes were transcriptionally regulated the products of others continued to be synthesised at 18°C (see chapter 6). Expression of the K1 capsule is regulated by reversible cold inactivation of the sialic acid synthase at 15°C: sialic acid synthase activity was detected only at 37°C, not 15°C and its expression after

upshift from 15°C did not require transcription or translation (Merker and Troy, 1990). CMP-NeuNAc synthetase and polysialyl polymerase also appeared to be cold inactivated but transcriptional regulation of genes involved in polysialic acid biosynthesis was not excluded.

While the work for this PhD was being performed the product of the *rfaH* (*sfrB*, *hlyT*) gene was shown to be required for K5 capsule expression (Stevens *et al.*, 1994). RfaH is also a positive regulator of *rfa* genes involved in biosynthesis of LPS core oligosaccharide (Pradel and Schnaitmann, 1991)(section 1.5.4), the *tra* operon of plasmid F (Beutin *et al.*, 1981) and the haemolysin (*hly*) operon (Bailey *et al.*, 1992). It is an 18 kDa basic protein with no similarity to known regulatory proteins (Bailey *et al.*, 1992). Evidence from various systems suggested that RfaH acts by preventing transcription termination (see chapter. 6). The model incorporating antitermination mediated by RfaH which has been proposed to regulate the expression of K5 and other group II K antigens (Stevens, M. P., unpublished) is discussed in chapter 6.

1.5.4 LPS

Lipid A-core is an essential component of the cell membrane of enteric bacteria (section 1.2.2) and its expression is therefore not restricted by environmental conditions. However the amount of R-LPS capped with O-polysaccharide and the O-chain length can affect serum resistance (Valvano, 1992)(Whitfield and Valvano, 1993). An increase in temperature from 30°C to 42°C reduces the median size of *E. coli* K-12 O-polysaccharide by 2 to 3 units while the number of R-LPS molecules carrying O units decreases (Schnaitman and Klena, 1993). The Rol (Cld) protein (Bastin *et al.*, 1993)(Batchelor *et al.*, 1992) determines the median O-chain length in *E. coli* and *Salmonella* spp (section 1.2.2) and may mediate the effects of temperature (Schnaitman and Klena, 1993).

Lipid A and core oligosaccharide are linked through two or three KDO molecules. KDO is incorporated into lipid A from the activated precursor CMP-KDO, the biosynthesis of which involves KDO-8-phosphate phosphatase (encoded by *kdsA*) and CMP-KDO synthetase (encoded by *kdsB*). Expression of both *kdsA* and *kdsB*, which are at separate loci, is growth-phase regulated at the level of transcription and is maximal at the onset of bacterial growth, when the need for LPS is probably greatest (Strohmaier *et al.*, 1995).

Mutations in the *rfaH* gene (section 1.5.3 & chapter 6) result in R-LPS with a heterogeneous core structure (Mäkelä and Stocker, 1984). Genes involved in

biosynthesis of LPS core oligosaccharide (section 1.2.2) are encoded within the *rfa* operon. In the absence of functional RfaH transcription terminates at Rho-dependent sites within the *rfa* operon; the reduced expression, particularly of the most promoter distal genes results in synthesis of an incomplete core oligosaccharide (Pradel and Schnaitmann, 1991)(Schnaitman and Klena, 1993). The role of RfaH in regulating the *rfa* operon and other genes involved in polysaccharide biosynthesis is discussed in chapter 6. Increasing temperature changed the gel profiles of core-LPS, possibly by differential regulation of the *rfa* genes involved in core biosynthesis (Pradel and Schnaitmann, 1991). The *rfaD* gene is unique among polysaccharide biosynthetic genes in being a member of the heat shock regulon and may play an important role in the regulation of core biosynthesis (Schnaitman and Klena, 1993).

1.5.5 Vi polysaccharide

The Vi polysaccharide is a homopolymer of *N*-acetylgalactosamine uronic acid and although it resembles group II K antigens is regulated by the *rcs* system (section 1.5.2)(Whitfield and Valvano, 1993). It is synthesised by *Citrobacter freundii* and some *Salmonella* spp. The proteins required for the synthesis and translocation of Vi polysaccharide are encoded by the *viaB* locus which includes genes with moderate similarity to the group II ABC transporters (section 1.2.1)(Hashimoto *et al.*, 1993). Another locus, *viaA*, is also required for Vi antigen expression and was shown to correspond to *rcsB* (Houng *et al.*, 1992). The EnvZ/OmpR two-component system which regulates the OmpC and OmpF porins also affects Vi antigen expression: In an *ompR* mutant the Vi antigen is not expressed (Pickard *et al.*, 1994). The signal recognised by OmpR may be osmolarity because Vi antigen expression is down-regulated at high NaCl concentrations (> 0.3M). These conditions are equivalent to those encountered in the gut and the reduction in Vi polysaccharide may facilitate interactions with epithelial cells (Pickard *et al.*, 1994). Vi antigen is a virulence factor of *S. typhi* and its expression is important for survival in the blood which has a lower osmolarity equivalent to 0.15M NaCl.

Cross talk between, in the case of Vi antigen expression, the RcsB/C and EnvZ/OmpR two-component regulatory systems enables the bacterium to regulate polysaccharide synthesis in response to the many different environments encountered within the host. Cross talk also occurs in the regulation of alginate biosynthesis in *P. aeruginosa* (section 1.5.6), with both OmpR and AlgR capable of activating transcription of *algD*.

Expression of Vi polysaccharide is affected by DNA rearrangements. *C. freundii* switches between a Vi^+ and Vi^- phenotype due to instability of the *viaB* locus (Whitfield and Valvano, 1993).

1.5.6 Regulation of alginate expression by *P. aeruginosa*

The overproduction of the exopolysaccharide alginate by *P. aeruginosa* results in a mucoid phenotype and is an important pathogenic determinant in infection of the lungs of cystic fibrosis patients (section 1.1.2). Alginate comprises randomly arranged, ß(1-4)-linked D-mannuronic acid and its C5 epimer L-guluronic acid (Deretic *et al.*, 1994). Most of the genes required for alginate biosynthesis are located in a cluster and are all transcribed in the same direction. Genetic evidence suggests that they form an operon, although a large transcript of the predicted size (~18 kb) has not been detected, suggesting that processing to form smaller transcripts may occur (Chitnis and Ohman, 1993). All the genes are co-expressed from a promoter upstream of the first gene, *algD* (encoding GDP-mannose dehydrogenase which catalyses the final step in precursor biosynthesis and channels sugar intermediates into alginate biosynthesis). Regulation of alginate production centres on this promoter (Fig. 1.7) (Deretic *et al.*, 1994).

Transcription is directed by the AlgU alternative σ factor (section 1.4.6), which is a member of a group of σ^{70} factors which are involved in the regulation of extracytoplasmic functions (the ECF subfamily) and are regulated by extracytoplasmic conditions (Lonetto *et al.*, 1994). Other members include *E. coli* σ^E which regulates genes involved in protein processing in the extracytoplasmic compartment and *Pseudomonas syringae* HrpL which controls transcription of a secreted plant virulence factor (Lonetto *et al.*, 1994). Transcription of *algD* is positively regulated by two proteins, AlgR and AlgB, which are homologous to the response regulators (effectors) of two-component regulatory systems (Goldberg and Dahnke, 1992). AlgR binds to three sites within the *algD* promoter (Fig. 1.7) and may be phosphorylated by the AlgQ (AlgR2) protein, which is similar to PfrA, a positive regulator of siderophore biosynthesis (Deretic *et al.*, 1994). The OmpR effector protein is also capable of activating *algD* transcription (Berry *et al.*, 1989), demonstrating cross-talk between regulatory systems. AlgR activates *algD* transcription at high osmolarity (Berry *et al.*, 1989), one of the stress conditions which may trigger enhanced alginate expression in the dehydrated cystic fibrosis lung. Constitutive expression of *algD* and therefore mucoidy can be caused by mutations, for example in the *muc* genes (Fig. 1.7) whose products counteract the effect of AlgU possibly by acting as anti-sigma factors (Brown and Hughes, 1995). Reversion to non-mucoidy can occur through

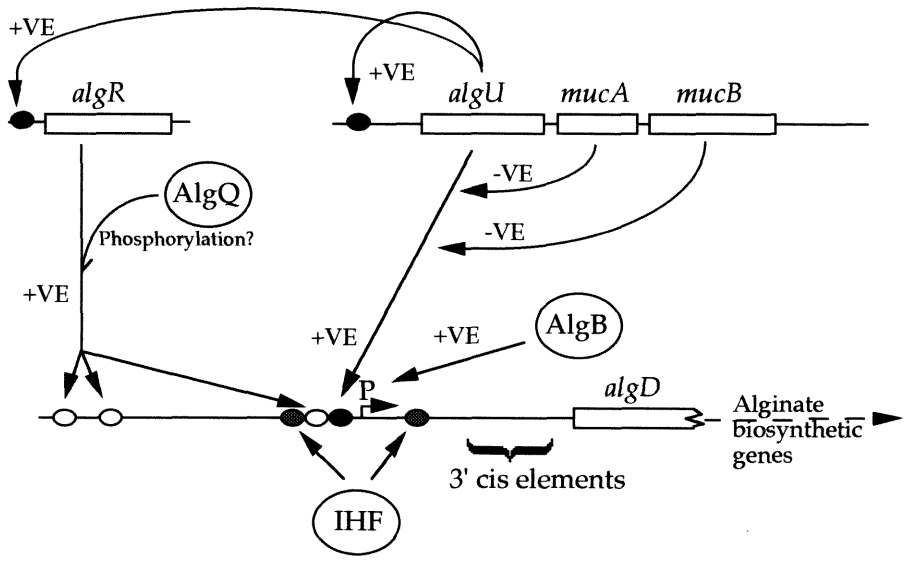


Figure 1.7. The regulation of alginate expression in *P. aeruginosa.* Arrows represent the interactions between the genes which regulate transcription from the promoter (P) upstream of *algD,* the first gene of an operon encoding most of the genes required for alginate biosynthesis. Where shown gene products are circled. The *algU* gene encodes an alternative sigma factor which directs transcription from promoters upstream of *algD, algR* and *algU* (solid ovals). AlgR has two high affinity binding sites 380 and 460 bp and one low affinity site 40 bp upstream of the *algD* promoter (open ovals). IHF binding sites are indicated by shaded ovals. The MucA and MucB proteins counteract the positive effect of AlgU on *algD* transcription.

mutations in *algU* or other genes (Schurr *et al.*, 1994). The histone-like IHF and AlgP proteins are involved in *algD* transcription (Wozniak, 1994). Two IHF binding sites are required for maximal expression, including a high affinity site 90 bp downstream of the transcription start site (Fig. 1.7). Given the role of IHF in promoting loop formation this suggested that sequences further downstream of the promoter might affect *algD* transcription; removal of sequences in this region did reduce transcription indicating the presence of 3' *cis* elements (Fig. 1.7).

1.5.7 Regulation of the expression of other bacterial extracellular polysaccharides

EPS I is an extracellular polysaccharide virulence factor of the phytopathogen *P. solanacearum* and causes wilting of infected plants. Expression of EPS I is positively regulated by a network of interacting signal transduction pathways consisting of the PhcA protein, which is a member of the LysR family of transcriptional regulators and by analogy may respond to a small signal molecule and two two-component systems comprising the VsrA/VsrD and VsrB/VsrC sensor and response regulator proteins respectively (Huang *et al.*, 1995). The two two-component systems are separate because they share low amino acid and structural similarities and while inactivation of either reduces EPS I expression the effect on extracellular protein expression and *in planta* growth varies (Schell *et al.*, 1994). A novel feature of the network is the basic XpsR protein, which has no known homologues and interconnects the PhcA and VsrA/D pathways. Both pathways are required for maximum *xpsR* transcription and PhcA has been shown to bind the *xpsR* promoter. Most of the genes required for EPS I expression form a single *eps* operon and are transcribed from a single promoter (Huang and Schell, 1995). XpsR probably mediates VsrA/D and PhcA control by binding within the 140 bp region upstream of the *eps* transcription start site shown to be sufficient for normal regulation (Huang and Schell, 1995). The VsrB/C system does not act through XpsR but both XpsR and VsrB/C are required simultaneously for activation of *eps* transcription.

Rhizobium spp produce various polysaccharides which are involved in the formation of nitrogen-fixing root nodules in leguminous plants (Johnston, 1989). These include extracellular polysaccharides (EPS), LPS, lipo-oligosaccharides (Nod factors) and KDO-containing polysaccharides analogous to the group II K antigens of *E. coli* (Reuhs *et al.*, 1993). The *exo* genes involved in the biosynthesis of EPS I (succinoglycan) in *R. meliloti* form a cluster on a large indigenous plasmid and are negatively regulated by the product of the *exoR* gene (Reed *et al.*, 1991). Genetic studies suggested that expression of EPS I was

also regulated post-transcriptionally (Müller *et al.*, 1993) by the products of the *exoX* and *exoY* genes (whose effects are similar to the *psiA* and *pssA* genes respectively of *R. leguminosarum* biovar *phaseoli*) (Reed *et al.*, 1991)(Mimmack *et al.*, 1994). Mutations in *exoX* increase EPS I expression while multicopy *exoX* results in an EPS^- phenotype which can be suppressed by multicopy *exoY* (Reed *et al.*, 1991). ExoY is a membrane-associated protein which is homologous with the RfbP protein of *Salmonella typhimurium* and was therefore proposed to be a galactosyl transferase (Müller *et al.*, 1993). ExoY and other proteins essential for EPS I expression form a biosynthetic complex associated with the cytoplasmic membrane. ExoX is thought to regulate this complex via protein:protein interactions with ExoY.

The ratio of low molecular weight to high molecular weight EPS I produced by *R. meliloti* is regulated by the ExoP protein (Becker *et al.*, 1995). ExoP comprises a C-terminal cytoplasmic and an N-terminal periplasmic domain. A mutated ExoP protein lacking the C-terminal domain resulted in an increased ratio of low to high molecular weight EPS I (Becker *et al.*, 1995). The N-terminal domain has limited sequence homology and similar hydropathic profile to the Rol (Cld) proteins of *S. enterica* and *E. coli* which determine O-antigen chain length (section 1.5.4) and proteins involved in group II capsular polysaccharide and ECA biosynthesis (Becker *et al.*, 1995). These include the KpsE protein which is involved in the export of K5 polysaccharide but has not been shown to affect its length (section 1.3.4). The LpsZ (RkpZ) protein affects the size of rhizobial K polysaccharides (Reuhs *et al.*, 1995) and is homologous with the KpsC protein of *E. coli* K5 (section 1.3.4) (Pazzani *et al.*, 1993). However, unlike a mutation in *lpsZ* which results in the export of smaller polysaccharides, a mutation in *kpsC* results in intracellular accumulation of a greatly reduced amount of polysaccharide (Bronner *et al.*, 1993).

The Nod factors secreted by *Rhizobium* spp are modified oligomers of ß-1,4-linked GlcNAc and are essential for root nodule formation. Their synthesis requires the products of the nodulation (*nod*) genes. Expression of the *nod* genes is induced by flavonoid molecules present in root exudates of host plants. These inducers activate the NodD protein which binds to a regulatory sequence (*nod*-box) preceding the *nod* operons and activates transcription (Johnston, 1989). The *nodD* gene forms a single transcriptional unit preceded by a *nod*-box and is transcribed constitutively: In *R. leguminosarum,* but not *R. meliloti* NodD has an autoregulatory function, repressing its own transcription.

DNA rearrangements affect the expression of several polysaccharides, including the Vi antigen (section 1.5.5). Production of EPS by the marine bacterium *Pseudomonas atlantica* is reversibly inactivated by site-specific insertion of the novel IS element IS*492* into the *eps* locus (Bartlett and Silverman, 1989). EPS production is restored by the precise excision of IS*492*. The *cap* locus of division I *H. influenzae* strains of serotypes a-d, which directs the synthesis of group II-like capsular polysaccharides, is flanked by copies of the IS-like element IS*1016* (Kroll *et al.*, 1989)(Kroll *et al.*, 1991). Amplification of *cap* can occur through IS*1016*-mediated homologous recombination events. The amount of polysaccharide produced by a strain with a duplicated type b *cap* locus was greater than that from one with a single copy but whether this confers a selective advantage is unclear as strains with a single copy *cap* locus are pathogenic (Kroll and Moxon, 1988). The reduction in polysaccharide synthesis due to the instability of the *cap* locus may provide an advantage at specific stages of infection.

1.6 Aims

Region 1 is common to all group II capsule gene clusters and information determined from *E. coli* K5 is likely to apply to other group II strains. The proteins encoded by region 1 are required for polysaccharide export. The processes involved in the expression of region 1 are therefore potential targets for therapeutic intervention to prevent capsule expression by bacteria causing infections. The aim of this thesis is to describe the pattern of region 1 expression at the molecular level. First, the transcriptional organisation of region 1 will be investigated. By analysis of the promoter region it is hoped to determine the sequences involved in transcriptional regulation and to investigate the involvement of known or novel regulatory proteins.

Expression of group II capsules is regulated by temperature. Region 1 encodes the CMP-KDO synthetase responsible for the elevated CMP-KDO synthetase activity observed at capsule-permissive temperatures. This suggests that region 1 is temperature-regulated. The level at which this temperature regulation occurs and the mechanisms responsible will be investigated.

CHAPTER 2

Materials and Methods

2.1 Chemicals

Unless stated otherwise the chemicals used in this study were purchased from Sigma Chemical Company Ltd or Fisons Scientific Equipment.

2.2 Bacterial strains and plasmids

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

2.3 Bacterial growth and manipulation

2.3.1 Media

Bacteria were routinely grown in Luria Bertani medium (LB broth: 0.5% [w/v] NaCl, 1% [w/v] trypticase peptone and 0.5% [w/v] yeast extract) with the addition of 1.5% [w/v] agar (BBL) as required, to make LB agar. The 2x YT medium for M13KO7 infection (section 2.6.2) contained more trypticase peptone (1.6% w/v) and yeast extract (1% w/v). *E. coli* JM101 was maintained on B agar (0.8% [w/v] NaCl, 1% [w/v] trypticase peptone). Soft top agar for use in bacteriophage assays (section 2.4) contained 0.8% [w/v] NaCl, 1% [w/v] trypticase peptone and only 0.8% agar. Blood agar plates were made with 4% [w/v] blood agar base (Oxoid) and 7% [v/v] defibrinated horse blood. Minimal medium consisted of M9 salts (per litre: 12.8 g Na_2HPO_4, 3.0 g KH_2PO_4, 0.5 g NaCl, 1.0 g NH_4Cl, adjusted to pH 7.4 with NaOH) supplemented with 0.1 mM $CaCl_2$, 2 mM $MgSO_4$, Thiamine (10 μg/ml), glucose (4 mg/ml) and appropriate amino acids (50 μg/ml), with the addition of 1.5% [w/v] agarose (Seakem) as required. MacConkey agar consisted of 4% [w/v] MacConkey agar base (Oxoid) supplemented with 0.5% [w/v] glucose or galactose as required. Colonies grown on this medium and able to ferment the sugar turned purple due to a change in pH.

When necessary the above growth media were supplemented with the following antibiotics: ampicillin (100 μg/ml), chloramphenicol (25 μg/ml), kanamycin (25 μg/ml), streptomycin (25 μg/ml) and tetracycline (20 μg/ml). All antibiotics were purchased from Sigma Chemical Company Ltd.

Table 2.1. Bacterial strains.

Strain	Relevant properties	Source/reference
LE392	F⁻*hsdR514 supE44 supF58 lacY1 galK2 galT2 metB1 trpR55*	(Murray *et al.*, 1977)
LE392.23	LE392 Δ(*argF–lac*)U169	T. J. Silhavy
JM101	*thi supE* Δ(*lac proAB*) F'[*traD36 proAB* ⁺ *lacZ*ΔM15 *lacI* ᵠ]	(Yannisch-Perron *et al.*, 1985)
DH5α		(Woodcock *et al.*, 1989)
SURE™		(Greener, 1990)
20026	O10:K5	K. Jann
MS101	PA360 *serA*⁺, *K5*⁺	(Stevens *et al.*, 1994)
MS102	MS101 *hns*::*Tn*10	(Stevens *et al.*, 1994)
MS106	MS101 *rfaH*::*Tn*5	(Stevens *et al.*, 1994)
MS113	MS101 *tcp1*::*Tn*5	M. P. Stevens
MS114	MS101 *tcp2*::*Tn*5	M. P. Stevens
MS140	MS101 *hha*::*Tn*5	This study
DS410	*minA minB ara xyl mtl azi thi*	(Dougan and Sherratt, 1977)
W3110	F⁻IN(*rrnD-rrnE*)1 λ⁻	B. Bachmann
2055 (CGSC#5072)	(W3110) *trpE*9851 *leu-277 rho-4*	(Morse and Guertin, 1972)

Table 2.2. Plasmids.

Plasmid	Description	Source/reference
pACYC184	cloning vector (Cm^r, Tc^r)	(Chang and Cohen, 1978)
pTZ18/19R	phagemid cloning/expression vector	(Mead *et al.*, 1986)
pGEM5Z	phagemid cloning/expression vector	Promega
pGEX-2T	glutathione S-transferase gene fusion vector	(Smith and Johnson, 1988)
pCB192	*lacZ* promoter-probe vector	(Schneider and Beck, 1986)
pHV100	*luxAB* vector for the study of transcription termination	(Peabody *et al.*, 1989)
pKK223-3	cloning/expression vector	(Brosius and Holy, 1984)
pGB110	K5⁺ cosmid	(Roberts *et al.*, 1986)
pPC6	K5 gene cluster cloned in pACYC184	C. Pazzani
pH18	*Hin*dIII fragment encoding region 1 except *kpsF* cloned in pUC18	C. Pazzani
pDS200	11 kb EcoRI fragment encompassing region 1, cloned in pKK223-3	This study

2.3.2 Growth conditions and harvesting

Bacteria were grown with shaking at 200 rpm at 37°C or 18°C, as stated. The bacterial cells were routinely harvested by centrifugation in a Minifuge-T centrifuge (Heraeus) at 2740 *g*, 4°C for 10 min or in a bench top minifuge at 13400 *g*, room temperature for 2 min.

2.3.3 Transformation with plasmid DNA

Bacteria were transformed with plasmid DNA by either the $CaCl_2$ method (Mandel and Higa, 1970) or electroporation (Dower *et al.*, 1988). In the $CaCl_2$ method competent cells were prepared by inoculating a 10 ml culture with 100 µl of an overnight culture. At an OD_{600} of approximately 0.4 the cells were harvested and resuspended in 4 ml of ice-cold 10 mM NaCl. They were immediately collected again by centrifugation at 1931 *g*, 4°C for 5 min and resuspended in 4 ml of ice-cold 100 mM $CaCl_2$. After 30 min on ice the cells were collected gently by centrifugation at 1320 *g*, 4°C for 5 min, resuspended in ≤1 ml of ice-cold 100 mM $CaCl_2$ and stored on ice until required. Approximately 0.1-1.0 µg of plasmid DNA (in 2-20 µl) was added to a 100 µl aliquot of competent cells. The mixture was kept on ice for 1 hr, heat shocked at 42°C for 3 min and after the addition of 500 µl of LB broth incubated at 37°C for 1 hr. Aliquots of 100 µl were then spread on LB agar plates containing appropriate antibiotics and incubated at 37°C overnight.

Electrocompetent cells were also prepared from a 10 ml mid-logarithmic phase culture, which was chilled on ice for 15 min. The cells were then harvested by centrifugation at 4°C and resuspended in 10 ml of ice-cold sterile water. This wash was repeated twice with water and finally with ice-cold 10% glycerol. The cells were resuspended in a final volume of 80 µl of 10% glycerol and 1-3 µl of DNA added to 40 µl aliquots. The mixture was transferred to an ice-cold Gene Pulser™ 2 mm cuvette (Biorad). This was immediately placed between the electrodes of a Biorad Gene Pulser set with the following parameters: Voltage 2.4 kV, capacitance 25 µF and resistance 200 Ω in parallel with the sample chamber (field strength 12 kV/cm). After delivery of a high voltage electric pulse the time constant was recorded (values of >4.8 indicate a high transformation efficiency) and 1 ml of ice-cold SOC recovery medium added to the cuvette (SOC: 2% [w/v] trypticase peptone, 0.5% [w/v] yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose). Following incubation for 1 hr at 37°C the mixture was plated out as above.

2.3.4 Transformation with bacteriophage DNA

E. coli JM101 was transformed with the replicative form of the bacteriophage DNA by the $CaCl_2$ method described above (section 2.3.3). However, following the heat shock there was no recovery period and the cells were mixed with 200 µl of a mid-logarithmic culture of JM101. To this mixture was added 3 ml of LB soft top kept molten at 45°C and containing 20 µl 100 mM IPTG and 50 µl 2% X-gal in dimethylformamide. The suspension was immediately spread on a B agar plate and incubated at 37°C overnight. Plaques were visible as a small zone of slow-growing infected cells in a lawn of more turbid bacterial growth.

2.3.5 Long term storage of cultures

For long term storage of cultures 800 µl of logarithmic phase culture (OD_{600} = 0.6) was added to 200 µl of sterile glycerol in a 1.5 ml eppendorf tube, mixed and stored at -70°C.

2.4 Assay for K5 bacteriophage sensitivity

2.4.1 Preparation of lysate

To prepare a high titre K5 bacteriophage lysate a $K5^+$ strain was grown overnight and 100 µl aliquoted into a 10 ml universal. A K5 bacteriophage plaque was picked with a sterile toothpick and placed in the universal which was vortexed and incubated at room temperature for 10 min before adding 2 ml of LB broth containing 10 mM $MgSO_4$. After incubation at 37°C for 6 hr the culture was transferred to two 1.5 ml eppendorf tubes and 100 µl of chloroform added. The tubes were vortexed, incubated at 37°C for 10 min, centrifuged at 13400 *g* for 2 min and the supernatants containing the bacteriophage collected. To prevent any cell growth during storage at 4°C 20 µl of chloroform was added to each bacteriophage stock, which should have a titre of approximately $1x10^9$ pfu/ml.

2.4.2 Assay for sensitivity

The culture to be assayed was grown to mid logarithmic phase (OD_{600} = 0.4-0.6) and resuspended in 5 ml of 10 mM $MgSO_4$. A serial dilution of the bacteriophage stock was prepared in bacteriophage dilution buffer (10 mM Tris-HCl pH 7.4, 10 mM $MgSO_4$ and 0.01% [w/v] gelatin). 100 µl of bacteriophage (at appropriate dilutions) were mixed with 100 µl of cells and left to adsorb for 20 min at room temperature. After this period 3ml of soft top agar, kept molten at 45°C, was

added to each tube and the mixture poured over an LB agar plate. Plates were examined for the formation of plaques following incubation at 37°C for 4-6 hr or if necessary overnight.

2.5 P1 transduction

2.5.1 Preparation of lysate

The donor culture carrying the marker of interest was grown overnight and 50 µl used to inoculate 5 ml of LB broth containing 0.2% [w/v] glucose and 5 mM $CaCl_2$. This was incubated with shaking at 37°C for 30 min and then 100 µl of a $P1_{vir}$ lysate (ca. $5x10^9$ pfu/ml) was added. The culture was incubated at 37°C for a further 2-3 hr until clearing occurred due to cell lysis. The culture was vortexed with 100 µl of chloroform, transferred to 1.5 ml eppendorf tubes and centrifuged at 4500 *g* for 10 min to pellet the cell debris. The supernatant was transferred carefully to a sterile eppendorf and vortexed with 20 µl of chloroform before storage at 4°C.

2.5.2 Transduction

The recipient strain to which the marker was to be moved was grown to an OD_{600} > 0.8 in LB broth containing 2.5 mM $CaCl_2$ and 1 ml dispensed to a sterile tube. To this was added 100 µl of P1 bacteriophage stock prepared on a strain carrying the marker to be moved. (Different volumes of P1 bacteriophage may be added to separate aliquots of recipient culture to discover the optimum ratio of bacteriophage to cells.) The mixture was incubated stationary at 30°C for 30 min then 4 ml of LB broth containing 5 mM $Na_3C_6H_5O_7$ was added and incubation continued with shaking for 1 hr at 30°C. The cells were collected by centrifugation and resuspended in 100µl LB broth with 5 mM $CaCl_2$ and plated on LB agar containing 5 mM $Na_3C_6H_5O_7$ and appropriate antibiotics to select for the desired transductants.

2.6 Techniques used in DNA analysis

2.6.1 Extraction of plasmid DNA

Plasmid DNA was extracted by alkaline lysis, essentially as described by Birnboim and Doly (Birnboim and Doly, 1979) using the following solutions:

Solution I	2% [w/v] glucose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA
Solution II	0.2M NaOH, 1% [w/v] SDS
Solution III	Potassium acetate buffer, pH4.8: 5M acetate, 3M potassium (In 100 ml: 11.5 ml glacial CH_3COOH, 60 ml CH_3COOK)

For large scale extractions a 400 ml culture was incubated overnight at 37°C. The cells were harvested by centrifugation in a Sorvall centrifuge (Dupont) at 3330 *g*, 4°C for 10 min using a Beckman GSA rotor, resuspended in 10 ml of ice-cold solution I and kept on ice for 30 min. Solution II was added (20 ml) with gentle mixing and the tubes kept on ice for 10 min. Solution III was then added (15 ml) with gentle mixing and after 10 min on ice the cell debris was removed by centrifugation at 36900 *g*, 4°C for 30 min in a Sorvall centrifuge using a Beckman SS-34 rotor. The supernatant was transferred in 18 ml aliquots to 30 ml Corex tubes and the DNA precipitated by the addition of 12 ml of isopropanol with mixing and incubation at room temperature for 15 min. The DNA was recovered by centrifugation at 3500 *g*, 20°C for 20 min, the supernatant discarded and the pellets allowed to air dry. The pellets were then resuspended in a combined volume of exactly 17 ml nanopure water. The plasmid DNA was purified by caesium chloride-ethidium bromide density gradient centrifugation. Exactly 17 g of ultrapure caesium chloride (ICN Biomedicals Inc.) was dissolved in the DNA solution which was then transferred to a 30 ml Sorval fixed angle rotor centrifuge tube containing 1 ml of a 10 mg/ml ethidium bromide solution. The tube was filled with paraffin oil and sealed before centrifugation in a Sorval TV850 fixed angle rotor at 128700 *g*, 20°C for 20 hr using a Sorvall OTD 60 Ultracentrifuge. The DNA was visualised under UV light and the lower band consisting of closed circular plasmid DNA collected. The ethidium bromide was extracted repeatedly with caesium chloride-saturated isopropanol until no pink colour remained in the lower aqueous phase. The CsCl was removed by dialysis against 10 l of distilled water, which was changed at least once. The plasmid DNA was stored in aliquots at -20°C.

For small scale extractions 1.5 ml of a 10 ml overnight culture was resuspended in 100 µl of solution I in a 1.5 ml eppendorf tube. After 30 min on ice 200 µl of solution II was added and the tube inverted 5 times, kept on ice for 5 min and 150 µl of ice-cold solution III added. The tube was vortexed gently and after 5 min on ice centrifuged at 13400 *g* for 5 min. The supernatant was recovered and mixed with an equal volume of phenol:chloroform (1:1). After centrifugation for 5 min the upper aqueous layer was recovered and extracted with an equal volume of chloroform. The DNA was then precipitated by addition of 1 ml of ethanol with mixing and after 5 min at room temperature was collected by

13400 *g* for 5 min. The pellet was dried *in vacuo*, resuspended in 25 µl of nanopure water and stored at -20°C.

2.6.2 Preparation of M13mp18/19 and phagemid single-stranded DNA template

E. coli JM101 was grown overnight and 100 µl used to inoculate 5 ml of LB broth. A single plaque from a transformation of JM101 with a recombinant M13mp18/19 bacteriophage (section 2.3.4) was picked into this culture and incubated at 37°C with shaking for 5 hr. Two aliquots of 1.5 ml were then centrifuged at 13400 *g* in a bench top minifuge for 5 min. The replicative form was extracted from the pellet (section 2.6.1) and the template DNA from the supernatant, which was transferred to fresh tubes and centrifuged again. Two 800 µl aliquots of supernatant were transferred to fresh tubes and 200 µl of 2.5 M NaCl, 20% [w/v] polyethylene glycol (PEG) 6000 was added to each. After 30 min at room temperature the tubes were centrifuged for 5 min and the supernatant removed. Any remaining supernatant was discarded after a second centrifugation. The pellets were resuspended in a combined volume of 200 µl 1.1M sodium acetate pH 7.0 and extracted with an equal volume of phenol:chloroform (1:1) and then with chloroform. After the chloroform extraction 150 µl of the aqueous layer was removed and 750 µl of ethanol added. The template DNA was collected after 30 min at -20°C by centrifugation at 13400 *g* for 10 min. The pellet was dried *in vacuo* and resuspended in 15 µl TE (10 mM Tris-HCl, 1 mM EDTA pH 8.0). The concentration of the template was estimated by agarose gel electrophoresis (section 2.6.5) and 5-7 µl was typically used in a sequencing reaction (section 2.6.8).

Template DNA was also produced using the phagemids pTZ18/19 by the system originally devised by Vieira and Messing (Vieira and Messing, 1987). pTZ18/19 does not encode the proteins required to generate progeny single-stranded DNA and these must be provided by the helper bacteriophage M13KO7 which is itself poorly replicated. A fresh bacterial colony containing pTZ18/19 with the DNA of interest was used to inoculate 3 ml of 2x YT broth. M13KO7 was added to a final concentration of approximately $2x10^7$ pfu/ml and incubated at 37°C with strong agitation for 1-1.5 hr until the culture was slightly turbid. Kanamycin was added to a final concentration of 70 µg/ml, to select for maintenance of M13KO7, and incubation continued at 37°C overnight. Following centrifugation single-stranded bacteriophage DNA was prepared from the supernatant as described above.

2.6.3 Extraction of genomic DNA

The pellet from 1.5 ml of an overnight culture was resuspended in 600 µl of TE buffer (10 mM Tris, 1 mM EDTA pH 8.0) containing 0.5% [w/v] SDS and 0.1 mg/ml proteinase K. After incubation at 37°C for 30 min 100 µl of 5 M NaCl and 80 µl of a 10% [w/v] solution of cetlytrimethylammonium bromide (CTAB) in 0.7 M NaCl were added. Following incubation at 65°C for 10 min two extractions were performed with chloroform:isoamyl alcohol (24:1). The aqueous phase was then extracted with phenol:chloroform (1:1) and precipitated with 0.6 volumes of isopropanol. The DNA was collected by centrifugation at 13400 *g* for 5 min, the pellet washed with 70% [v/v] ethanol, dried *in vacuo* and resuspended in 30 µl TE. Contaminating RNA was removed by incubation at 37°C for 30 min with 5 units of ribonuclease (RNace-it™, Stratagene).

2.6.4 DNA purification and concentration

DNA was routinely purified by phenol extraction. The DNA solution was mixed with an equal volume of phenol equilibrated to pH > 7.6 by repeated extraction with Tris buffer. Following centrifugation in a bench top minifuge at 13400 *g* for 5 min the upper aqueous layer was carefully transferred to a fresh tube and an equal volume of chloroform:isoamyl alcohol (24:1, v/v) added. The mixture was vortexed and centrifuged as above and the upper aqueous layer transferred to a fresh tube taking care to avoid the interface. If required the DNA solution was concentrated by precipitation and resuspension in a smaller volume. Sodium acetate, pH 5.2 was added to a final concentration of 300 mM followed by 2.5 volumes of ethanol. After ≥ 30 min at -20°C the DNA was collected by centrifugation at 13400 *g* in a bench top minifuge for 5 min, the pellet dried *in vacuo* and resuspended in an appropriate volume of nanopure water.

2.6.5 Routine DNA manipulations

Restriction endonucleases and DNA modifying enzymes were purchased from Gibco BRL. Typically 0.5-1.0 µg of DNA was digested with one unit of endonuclease in a volume of 20 µl at 37°C (or other recommended temperature) for 1 hr in the buffers provided. If the DNA was to be ligated the restriction enzyme was heat inactivated or removed by phenol extraction. Ligations were performed with T4 DNA ligase at 14°C overnight. The insert and vector DNA were mixed in a ratio of approximately 3:1, co-precipitated, resuspended in 15 µl of the buffer provided and mixed with 0.5 ul of enzyme. To prevent self ligation of vector DNA with cohesive ends the 5' terminal phosphates were removed

(dephosphorylation) with Calf Intestinal Alkaline Phosphatase (CIP). The DNA was incubated with 1 µl of enzyme in CIP buffer (250 mM Tris-HCl pH 9.0, 0.1 mM $MgCl_2$, 0.01 mM $ZnCl_2$) at 37°C for 30 min. DNA fragments synthesised by PCR (section 2.6.7) using Taq DNA polymerase (NBS Biologicals) have a protruding 3' adenosine residue. To clone such fragments into blunt ended vectors 3' terminal thymidine residues were added by incubating ≤ 0.5 µg of vector DNA in 100 µl of PCR buffer (20 mM Tris-HCl pH 8.0, 25 mM KCl, 0.05% Tween 20, 1.5 mM $MgCl_2$) with 2 mM dTTP and 1 µl of Taq polymerase (5 units) at 70°C for 2 hr.

DNA fragments were separated by agarose gel electrophoresis in 0.7-1.2% Seakem agarose in ELFO buffer (33 mM Tris, 1 mM EDTA pH 7.7) with 0.5 µg/ml ethidium bromide. Samples were generally in a volume of 20 µl to which 2 µl of 10x DNA loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol, 25% Ficoll) was added prior to loading. Following electrophoresis at approximately 5.0 V/cm for 1-2 hr the fragments were visualised under UV. The size of fragments was estimated by comparison with 1 kb standards (Gibco BRL). If required DNA was purified from an agarose slice cut from the gel using a Sephaglas™BandPrep Kit (Pharmacia Biotech) according to the manufacturer's instructions.

The concentration of DNA solutions was determined by measuring the optical density at 260 nm and 280 nm, given that an OD_{260} = 1.0 corresponds to a 40 µg/ml solution of double-stranded DNA and an approximately 20 µg/ml solution of oligonucleotides. The ratio of the OD_{260} to OD_{280} values indicates the purity of the solution, the ideal value for double-stranded DNA being 1.8.

2.6.6 Detection of recombinant plasmids

Certain vectors including pUC18/19, pTZ18/19 and M13mp18/19 enable 'blue/white screening' for recombinants. In these vectors the multiple cloning site is located in the *lacZ'* gene which encodes an amino terminal fragment of ß-galactosidase. JM101 carries an F' episome encoding enzymatically inactive ß-galactosidase lacking residues 11-41 (*lacZ*ΔM15). The ability of the vector *lacZ'* gene to complement this mutation (α-complementation) is usually abolished when *lacZ'* is disrupted by the insertion of cloned DNA. The products of a ligation (section 2.6.5) were transformed (section 2.3.3) into JM101 and recombinant clones selected by their inability to metabolise X-gal and the fact that they were therefore white rather than blue when plated on LB agar containing 0.04% X-gal and 0.1 mM IPTG. The putative recombinant clones were confirmed

by extraction of plasmid DNA (section 2.6.1) and restriction analysis (section 2.6.5).

2.6.7 The polymerase chain reaction

The polymerase chain reaction (PCR) was performed with 10 ng of template DNA, oligonucleotide primers (each 0.25 µM) and dNTPs (each 0.2 mM) in 100 µl PCR buffer (20 mM Tris-HCl pH8.3, 25 mM KCl, 0.05% Tween 20, 0.1 mg/ml bovine serum albumin (BSA) and 2.0 mM $MgCl_2$). The template DNA was added after the other components had been UV irradiated on an old UV transilluminator for 15 min to crosslink any contaminating DNA and prevent it acting as a template. The reaction mixture was overlaid with 20 µl of molecular biology grade mineral oil and placed in a Perkin Elmer Cetus thermal cycler. After 5 min at 95°C 0.5 µl of Taq DNA polymerase (NBS Biologicals) was added and 30 cycles of the following program performed. The primers and template were denatured at 95°C for 1 min, allowed to anneal at 55°C for 1 min and the primers then extended at 72°C for 2 min. The products of the reaction were analysed by agarose gel electrophoresis (section 2.6.5)

2.6.8 DNA sequencing

Sequencing was performed by the dideoxy chain termination method of Sanger *et al.* (Sanger *et al.*, 1977) using the Sequenase® Version 2.0 kit (United States Biochemical Corporation) according to the manufacturer's protocol. For the annealing reaction 0.5 pmoles of primer DNA were added to 7 µl of template DNA (section 2.6.2) and 2 µl of the 5x reaction buffer supplied. This mixture was heated to 65°C for 2 min and then cooled slowly to < 35°C. Extension reactions were performed separately with four dNTP mixes each containing a different dideoxynucleotide triphosphate and the products labelled by incorporation of [α-^{35}S]dATP. The radiolabelled fragments were separated by buffer gradient gel electrophoresis (Biggin *et al.*, 1983). The gels were prepared with the following solutions:

5x Tris/borate/EDTA (TBE)	44.5 mM Tris-HCl 44.5 mM boric acid 2.0 mM EDTA
0.5x TBE acrylamide/urea mix (per 500 ml)	215 g urea 50 ml 5x TBE 75 ml 40% [w/v] acrylamide

5X TBE acrylamide/urea mix (per 200 ml)	86 g urea 100 ml 5X TBE 30 ml 40% [w/v] acrylamide 10 g sucrose 10 mg bromophenol blue
Gel solution 1:	65 ml 0.5X TBE acrylamide/urea mix 300 µl 10% [w/v] ammonium persulphate 70 µl TEMED
Gel solution 2:	12 ml 5X TBE acrylamide/urea mix 60 µl 10% [w/v] ammonium persulphate 20 µl TEMED

The Accugel 40% [w/v] acrylamide solution (National Diagnostics) contained acrylamide:bisacrylamide (19:1). Electrophoresis grade ammonium persulphate was obtained from Biorad. The glass gel plates (20 cm X 50 cm) were cleaned with Teepol detergent and then ethanol. One plate was siliconised with a thin layer of dimethyldichlorosilane, allowed to air dry and the two plates taped together separated by 0.4 mm spacers. Gel solutions 1 and 2 were prepared and 10 ml of solution 1 followed by 15 ml of solution 2 drawn into the same 25 ml plastic pipette and poured between the plates. The remaining space was filled with solution 1 and a comb inserted to form a single large well. The gel was left to polymerise for at least 1 hr before removing the tape and clamping to a vertical electrophoresis system. Aluminium plates were attached to either side to dissipate heat, the top reservoir was filled with 0.5X TBE and the bottom with 5X TBE. The gel was pre-run at 40 W for 30 min and then the well rinsed before inserting a "shark's tooth" comb with the teeth just contacting the acrylamide. The samples were denatured by heating to 80°C for 3 min immediately before loading between the teeth of the comb. Electrophoresis was performed at a constant power of 40 W for 3 or 6 hr. The apparatus was dismantled leaving the gel attached to the non-siliconised plate where it was covered with cold fixing solution (10% acetic acid, 10% methanol) for 15 min and then rinsed with distilled water. The gel was transferred to damp filter paper (Whatman 3 MM), covered with Saran wrap and dried under vacuum for 45 min at 80°C. The fragments were visualised by autoradiography by placing film (Cronex, Dupont) directly on top of the dried gel.

2.6.9 Transfer of DNA to filters

DNA was transferred to a filter from either an agarose gel (Southern blotting (Southern, 1975)) or directly from bacterial colonies. In the former genomic DNA (section 2.6.3) was restricted with appropriate endonucleases (section 2.6.5) and the fragments separated by electrophoresis in 0.7% agarose gels (section 2.6.5). The gel was placed in 0.25 M HCl for 7 min, rinsed in distilled water and placed in denaturation buffer (0.5 N NaOH, 1.5 M NaCl) with gentle shaking for 30 min. After rinsing in distilled water the gel was placed in neutralising solution (0.5 M Tris-HCl pH 7.5, 3 M NaCl) for 30 min. A capillary blot was set up by placing the gel on 4 sheets of filter paper (Whatman 3 MM) cut to the size of the gel, soaked in 20x SSC (3 M NaCl, 0.3 M sodium citrate) and linked to a reservoir of this buffer. A sheet of Hybond-N nylon membrane (Amersham) the size of the gel was placed on top of the gel followed by several sheets of filter paper cut just smaller than the membrane (care being taken to avoid entrapping bubbles). Paper towels were placed on top of the filter paper and the sandwich compressed by placing a 0.4 kg weight on a plate on top of the towels. Transfer was performed overnight and the towels replaced periodically when damp. The apparatus was dismantled, the membrane rinsed in 2x SSC and allowed to air dry before wrapping in Saran wrap. The DNA was crosslinked to the membrane by exposure to UV light from an old transilluminator for 4 min.

Bacterial colonies were transferred from an agar plate to Hybond-N by placing the membrane carefully on top of the agar and then lifting it off vertically to avoid smearing. The membrane was placed colony side up on filter paper (Whatman 3 MM) soaked in denaturing solution (0.5 M NaOH, 1.5 M NaCl). After 5 min the membrane was transferred to filter paper soaked in neutralising solution (0.5 M Tris-HCl pH 7.5, 1.5 M NaCl) for 5 min and air dried for 30 min on fresh filter paper before baking at 80°C for 2 hr to fix the DNA. Bacterial cell debris was removed by gently scrubbing with polyallomer wool moistened in 5x SSC.

2.6.10 Preparation of radiolabelled DNA probes

DNA fragments generated by restriction endonuclease digestion or PCR (section 2.6.7) were purified following agarose gel electrophoresis (section 2.6.5) and radioactively labelled by random hexanucleotide primer extension (Feinberg and Vogelstein, 1983). Oligo-labelling buffer (OLB) was prepared by mixing solutions A, B and C in the ratio 10:25:15 (Table 2.3) and stored in aliquots at -20°C. The DNA was denatured by boiling 10 ng in a volume of 10 µl for 3 min. The tube was cooled on ice and 3 µl of OLB, 1 µl of [α^{32}P]dCTP and 1 µl of the large

fragment of DNA polymerase I (Gibco BRL) added. The labelling reaction was incubated at room temperature for ≥ 5hr. If required the labelled probe was separated from unincorporated dNTPs by ethanol precipitation.

Table 2.3 The components of oligo-labelling buffer (OLB)

Solution A	1 ml Soln O (1.25 MTris-HCl, 0.125 M $MgCl_2$ pH 8.0 18 µl ß-mercaptoethanol 5 µl each of dATP, dTTP, dGTP (100 mM) (Pharmacia)
Solution B	HEPES (2 M, titrated to pH 6.6 with NaOH)
Solution C	Hexanucleotides (90 OD units/ml) (Pharmacia)

2.6.11 DNA:DNA hybridisation and detection

The filters with immobilised DNA (section 2.6.9) were incubated with 25 ml of hybridisation buffer in glass hybridisation bottles rotating in a Hybaid hybridisation oven heated to 65°C. Hybridisation buffer comprised:

3x SSC (section 2.6.9)
0.1% [w/v] SDS
200 µg/ml herring sperm DNA (see below)
5x Denhardt's solution (Ficoll (Pharmacia), BSA and polyvinylpyrollidone (pvp) each 0.1% [w/v])

Herring sperm DNA was dissolved to make a 5 mg/ml stock and the DNA sheared to produce smaller fragments by repeated boiling and passage through progressively finer gauge needles. The filters were pre-hybridised for ≥ 2hr before the addition of radiolabelled DNA probe (section 2.6.10), which was boiled for 2 min immediately prior to adding to the hybridisation mixture. After hybridisation unbound probe and dNTPs were removed by four 15 min washes at 65°C with a pre-heated solution containing 0.1x SSC and 0.1% [w/v] SDS. This high stringency wash was used because the probes were identical with the DNA to be detected. The filters were dried briefly and wrapped in Saran wrap before the radiolabelled DNA was visualised by autoradiography. If required the radiolabelled probe was stripped from the filter by pouring on boiling 0.1% SDS and agitating periodically as it cooled to room temperature. The filter could then be hybridised with a different probe.

2.7 Techniques used in RNA analysis

2.7.1 Preparation of cellular RNA

The cells from 1.5 ml of mid-logarithmic phase culture (OD_{600} = 0.4-0.6) were collected by centrifugation and resuspended in 0.5 ml phenol preheated to 60°C and equilibrated to pH 5.2 by repeated extraction with 0.2 M sodium acetate pH 5.2. This suspension was immediately added to 0.8 ml of lysis buffer (0.02 M sodium acetate pH 5.2, 0.5% [w/v] SDS, 1 mM EDTA) and incubated at 60°C for 5 min with periodic mixing. After 10 min on ice and centrifugation at 13400 *g* for 5 min the top aqueous layer was transferred to a fresh tube and the extraction with phenol pH 5.2 at 60°C was repeated. A chloroform extraction was then performed and the RNA precipitated by adding two volumes of ethanol and incubating at -20°C for ≥ 1 hr. The RNA was collected by centrifugation at 13400 *g* for 5 min, dried briefly *in vacuo* and resuspended in nanopure water treated with diethylpyrocarbonate (DEPC). DEPC was added to all solutions used in RNA analysis (0.05% v/v) to inhibit RNase activity. The solutions were then incubated at 37°C overnight and autoclaved at 15 lb/sq inch for 15 min to destroy any remaining DEPC before use. If required RNA preparations were treated with RNase-free DNaseI (Pharmacia) to ensure that they contained no DNA.

2.7.2 Northern blotting

RNA was electrophoresed under denaturing conditions in 1% agarose gels containing 6.3% formaldehyde (17 ml of 37% [v/v] formaldehyde solution per 100 ml) in MOPS buffer (20 mM 3-[N-Morpholino]-propane-sulphonic acid). The sample, typically 30 µg, was resuspended in 6 µl nanopure water and 12.5 µl deionised formamide, 2.5 µl 10x MOPS and 4 µl of formaldehyde (37% v/v) added. After incubation at 65°C for 10 min the sample was chilled on ice and 2.5 µl 50% [v/v] glycerol containing 2.5 mg/ml bromophenol blue was added prior to loading. Electrophoresis was performed at a constant 20 V overnight, with the buffer recirculating. A capillary blot was set up to transfer the RNA to a Hybond N membrane (Amersham) as described in section 2.6.9.

Alternatively, RNA was transferred directly to Hybond N using a slot blot apparatus (Biorad) which was first washed in 0.5 M H_2SO_4 for 30 min. The apparatus was assembled according to the manufacturers instructions. Samples were prepared as for electrophoresis (see above) until after incubation at 65°C when they were kept on ice and an equal volume of cold 20x SSC added. 100 µl of nanopure water was placed in each well and the vacuum applied until the water

was sucked through the membrane. The samples were then loaded, gently sucked through and the apparatus dismantled.

The membranes were wrapped in Saran wrap and the RNA fixed by UV crosslinking (section 2.6.9) RNA size markers (Gibco BRL) were visualised by detaching the section of the membrane to which they had transferred and staining with methylene blue.

Specific RNA was detected by hybridisation with radiolabelled DNA probes (section 2.6.10). Conditions were essentially as described in section 2.6.11 with 3x SSC replaced by 5x SSPE (20x SSPE stock is 3.6 M NaCl, 0.2 M NaH_2PO_4, 0.02 M EDTA) and a hybridisation temperature of 60°C. In certain cases sensitivity was increased by replacing the buffer following prehybridisation with fresh buffer lacking herring sperm DNA but containing 0.1 g/ml dextran sulphate (Molecular weight 500 000). Following hybridisation washing was performed at 60°C for 15 min periods with buffers of increasing stringency (from 1x SSPE, 0.1% SDS to 0.1x SSPE, 0.1% SDS) as required.

2.7.3 Transcript mapping

Single-stranded radiolabelled RNA probes were used to analyse *in vivo* transcripts. Firstly the DNA in the region of interest was cloned downstream of the T7 promoter in the vector pTZ18/19 (Table 2.2). The recombinant plasmid was linearised with an appropriate restriction enzyme that does not generate a protruding 3' terminus. The DNA was separated from any undigested plasmid by agarose gel electrophoresis and subsequent purification (sections 2.6.4 and 2.6.5). Radiolabelled RNA probes were synthesised using the Riboprobe™ *in vitro* transcription system (Promega). The transcription reaction was performed in 100 µl of the buffer provided (40 mM Tris-HCl pH 7.5, 6 mM $MgCl_2$, 2 mM spermidine and 10 mM NaCl) with the following components: 10 mM dithiothreitol (DTT), 20U RNasin® ribonuclease inhibitor, 0.5 mM each rATP, rGTP, rCTP, 12 µM rUTP, 1.0 µg template DNA, 50 µCi [α-^{32}P]rUTP and 15-20 U T7 RNA polymerase. The reaction was incubated at 37°C for 1 hr and then the DNA template was removed by adding 7 U of RNase-free DNase I (Pharmacia) and incubating for a further 15 min. The sample was purified by phenol extraction and ethanol precipitation.

The probe was then gel purified on a DNA sequencing gel (section 2.6.8) to remove shorter transcripts. The full length probe was detected by autoradiography and eluted from a gel slice by incubation in elution buffer (2 M

ammonium acetate, 1% [w/v] SDS, 25 μg/ml tRNA) at 37°C for ≥ 4 hr. The gel fragments were removed by centrifugation and the RNA precipitated from the supernatant by addition of 1ml ethanol, incubation on ice for 15 mins and centrifugation at 13400 *g* for 10 min. The probe was then hybridised with 20-100 μg of the RNA being studied in 50 μl 80% formamide hybridisation buffer (80% [v/v] deionised formamide, 40 mM PIPES, pH 6.7, 0.4 M NaCl, 1 mM EDTA). The mixture was heated to 85°C for 5 min and then incubated at 40°C overnight. Unhybridised single-stranded ^{32}P-RNA probe was digested by treatment with RNase ONE. 450 μl of RNase digestion buffer (10 mM Tris-HCl pH 7.5, 5 mM EDTA, 200 mM sodium acetate) and 40 U of RNase ONE were added to the hybridisation mixture and incubated at 30°C for 45 min (In some cases the temperature was lowered to prevent nicking of single-stranded regions formed due to breathing of the RNA:RNA duplex in regions with high AU content). The RNase digest was terminated by the addition of 20 μl SDS (10% w/v) and 10μl proteinase K (10 mg/ml) and incubation at 37°C for 20 min. Following phenol:chloroform extraction and ethanol precipitation the products were resuspended in formamide loading buffer (80% formamide [v/v], 10 mM EDTA, 1 mg/ml xylene cyanol, 1 mg/ml bromophenol blue) and analysed by electrophoresis and autoradiography.

2.7.4 Primer extension

Primer extension analysis was performed using Promega's AMV Reverse Transcriptase Primer Extension System, essentially according to the manufacturer's instructions. An oligonucleotide complementary to the RNA of interest was radiolabelled by transferring the terminal [γ-^{32}P] phosphate from dATP to its 5'-hydroxyl terminus. 3 μl of [γ-^{32}P]dATP (3000Ci/mMole @ 10 mCi/ml) and 10 pmoles of oligonucleotide (calculated assuming the molecular weight of a nucleotide to be 350) were incubated with 10 U T4 polynucleotide kinase in 10 μl 'forward exchange buffer' (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 5 mM DTT and 0.1 mM spermidine) at 37°C for 10 min. The T4 polynucleotide kinase was inactivated by heating to 90°C for 2 min. If necessary the radiolabelled primer was diluted before use in the primer extension reaction. Up to 100 μg of cellular RNA was annealed with 1 μl of radiolabelled primer in AMV Primer Extension Buffer (50 mM Tris-HCl, pH 8.3, 50 mM KCl, 10 mM MgCl2, 10 mM DTT, 1 mM each dNTP, 0.1 mM spermidine) at 58°C for for 20 min. The elongation reaction was performed with 1U of AMV reverse transcriptase at 42°C for 30 min using the same buffer with the addition of 3.7 mM sodium pyrophosphate. Samples with 100 μg of RNA were treated with DNase-free RNase (RNace-it™, Stratagene) before being precipitated and redissolved in

loading dye (80% formamide, 10 mM EDTA, 0.1% xylene cyanol and 0.1% bromophenol blue). The products were analysed by denaturing polyacrylamide gel electrophoresis (section 2.6.8) and visualised by autoradiography.

2.7.5 Reverse transcriptase polymerase chain reaction

The reverse transcriptase polymerase chain reaction (RT-PCR) was used to detect specific transcripts. Between 0.2-5.0 µg of cellular RNA were mixed with 2.5 µl of 25 µM 3'oligonucleotide (complementary to the transcript of interest) in 50 µl of PCR buffer (50 mM KCl, 10 mM Tris-HCl, 1.68 mM $MgCl_2$, 200 µM each dNTPs) and denatured at 90°C for 2 min. 10 U Stratascript reverse transcriptase (Stratagene) and 10 U Human Placental RNase inhibitor (Amersham) were added and the reverse transcription reaction performed at 37°C for 1 hr. The reverse transcriptase was denatured by heating to 92°C for 2 min. A 20 µl solution of PCR buffer containing 2.5 µl 5'primer (25 µM) and 1 U BIOTAQ™ Polymerase (Bioline UK) was added and the mixture overlaid with mineral oil. 35-40 cycles (92°C, 2 min; 55°C, 1 min; 72°C, 2 min) of PCR were performed using a GeneE thermal cycler (Techne). The products were analysed by agarose gel electrophoresis (section 2.6.5).

2.8 Techniques used in protein analysis

2.8.1 Preparation of whole cell lysates

Cultures were grown to logarithmic phase and the OD_{600} determined. The cells from 4 ml of culture were collected and resuspended in a volume of loading buffer (Tris-HCl 50 mM pH 6.8, 0.4% [w/v] SDS, 4% [v/v] glycerol, 1% [v/v] 2-mercaptoethanopl, 0.02% [w/v] bromophenol blue) proportional to the OD_{600} of the culture (4 ml from a culture with OD_{600} = 0.7 was resuspended in 400 µl). The samples were denatured by boiling for 5 min.

Alternatively cultures were resuspended in sonication buffer (20 mM Tris-HCl, 5 mM EDTA pH 7.5) and lysed by sonication. The cells were kept on ice while a small probe (Labsonic) was used at high power for four 15 sec periods with 15 sec intervals. The sonicate was separated into a supernatant and pellet fraction by centrifugation at 13400 *g* for 30 min, 4°C. The pellet was resuspended in sonication buffer and the protein concentration of each fraction determined using a Coomassie blue protein estimation kit (Biorad). A 1/10 volume of 10x loading buffer was added to each sample and a volume inversely proportional to the protein concentration analysed by SDS-PAGE (section 2.8.2).

2.8.2 SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE (Laemmli, 1970) was performed using a Mini-PROTEAN™ II system (Biorad). The following buffers were prepared: Buffer A, 0.75 M Tris-HCl pH 8.8, 0.2% [w/v] SDS; and Buffer B, 0.25 M Tris-HCl pH 6.8, 0.2% [w/v] SDS. The Protogel acrylamide solution (30% [w/v] acrylamide, 0.8% [w/v] bisacrylamide) was purchased from National Diagnostics. The gels contained:

Resolving gel	4.95 ml acrylamide (30%)
(15% acrylamide)	4.91 ml Buffer A
	348 µl Ammonium persulphate (1% w/v)
	40 µl TEMED
Stacking gel	1.04 ml acrylamide (30%)
(5% acrylamide)	3.0 ml Buffer B
	2.12 ml nanopure water
	150 µl Ammonium persulphate (1% w/v)
	20 µl TEMED

Immediately after pouring the lower, resolving gel was overlaid with butanol and allowed to set. The butanol was removed and the stacking gel composed and poured. Samples were boiled for 5 min in loading buffer (Tris-HCl 50 mM pH 6.8, 0.4% [w/v] SDS, 4% [v/v] glycerol, 1% [v/v] 2-mercaptoethanopl, 0.02% [w/v] bromophenol blue) and 10-20 µl loaded on the gel. Electrophoresis was performed in running buffer (0.025 M Tris-HCl, 0.192 M glycine, 0.1% [w/v] SDS) at 100 V until the samples entered the resolving gel and 150 V thereafter. Proteins were visualised by staining the gel in Coomassie blue followed by destain solution (40% [v/v] methanol and 10% [v/v] acetic acid). Protein sizes were estimated by comparison with molecular weight markers (Sigma).

2.8.3 Protein overexpression

Fusion proteins encoded by genes cloned in the vector pGEX-2T (Smith and Johnson, 1988) were overexpressed by transforming the recombinant plasmid into *E. coli* SURE (section 2.3.3). The culture was grown to OD_{600} = 0.2 and the *tac* promoter then induced by addition of isopropyl ß-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. Incubation was continued at 37°C for a further 3 hrs and cell lysates prepared (section 2.8.1).

2.8.4 Pre-adsorption of polyclonal sera with acetone powders

Contaminating antibodies were removed from antibody solutions by treatment with a saturating amount of competitor protein. A suspension of proteins was prepared from an *E. coli* strain not expressing the protein to which the antibody was raised (JM109 (DE3)) by acetone precipitation. The cells from a 400 ml overnight culture were resuspended in 10 ml saline (0.9% [w/v] NaCl) and 40 ml acetone added with vigorous mixing. Following incubation on ice for 30 min with repeated vigorous mixing the precipitate was collected by centrifugation at 10000 *g* for 10 min. The pellet was resuspended in acetone (-20°C) with vigorous mixing and kept on ice for 10 min. The pellet was collected, spread on filter paper and dispersed into a fine powder. When dry the powder was stored in an airtight container. Approximately 200 µl of acetone powders were mixed with 1.0 ml of antibody solution overnight at 4°C. The powder was removed by centrifugation at 13400 *g* for 10 min. The supernatant containing the antibody was transferred to a fresh tube and the centrifugation repeated. The supernatant was removed and stored at 4°C.

2.8.5 Western analysis

The proteins were transferred from SDS-PAGE gels to a Hybond-C membrane (Amersham) by Western blotting (Towbin *et al.*, 1979). The gel was placed on two sheets of filter paper (Whatman 3 MM) dampened with transfer buffer (20% [v/v] methanol, 0.1% [w/v] SDS, 20 mM Tris-HCl, glycine). The membrane was soaked in methanol, rinsed in water and then placed on top of the gel, avoiding the introduction of bubbles. Four sheets of filter paper were placed on top of the membrane and the whole sandwich inserted between plastic slats in a blotting tank with the membrane nearest the positive terminal. Transfer was performed in transfer buffer at 70 mA for 2 hrs. Molecular weights were estimated by comparison with pre-stained markers (Biorad) which remained visibe on the transfer membrane. After transfer the membrane was rinsed in phosphate buffered saline/Tween 20 (PBS-T - Per litre: 8 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4, 0.24 g KH_2PO_4, 0.1% [v/v] Tween 20).

Non-specific sites on the membrane were blocked by incubation with gentle rocking overnight in 5% [w/v] skimmed milk in PBS-T. The membrane was rinsed in PBS-T before incubation for > 1 hr with the primary antibody diluted appropriately in 5% [w/v] skimmed milk in PBS-T. Following four washes of at least 15 min each with PBS-T the membrane was incubated for > 1 hr with secondary antibody (Horse radish peroxidase conjugated Affinipure goat anti

rabbit IgG Fc fragment specific, Jackson Immuno Research) diluted 1/2000 in 5% [w/v] skimmed milk in PBS-T. The membrane was washed as before in PBS-T and the bound antibody visualised using the Amersham ECL detection system. The two detection reagents were mixed together (1:1) and spread over the membrane ($\approx$0.125ml/cm^{-2}). After 1 min the excess was drained and the membrane wrapped in Saran wrap prior to autoradiography (exposures ranged from 10 sec - 20 min).

2.8.6 Minicell analysis

Minicells were prepared from *E. coli* strain DS410 (Dougan and Sherratt, 1977) following transformation (section 2.3.3) with the plasmid encoding the genes of interest. A 400 ml culture was grown overnight in Brain Heart Infusion medium (BHI: 37g/l BHI base (Oxoid). The culture was centrifuged at 1200 *g*, 4°C for 5 min (section 2.6.1) and the supernatant containing the minicells carefully transferred to fresh pots. These were centrifuged at 8540 *g* for 15 min to sediment the minicells and the pellets resuspended in 3 ml ice-cold 1x M9 salts (section 2.3.1). The minicells were loaded onto a 20% sucrose gradient (prepared by freezing a 20% [w/v] sucrose solution in 1x M9 salts and thawing overnight at 4°C) and centrifuged at 2740 *g* for 20 min at 4°C. Contaminating vegetative cells were sedimented to the bottom of the tube while minicells formed a band in the central region. The upper 2/3 of this band was collected and transferred to a fresh tube. The minicells were collected by centrifugation at 9220 *g*, 4°C for 10 min and the sucrose centrifugation repeated. The volume and OD_{600} of the minicells collected from the second sucrose gradient was determined. The minicells were then collected as before and the pellet resuspended in a volume of 70% 1x M9, 30% glycerol calculated to create a suspension with OD_{600} = 2.0. Aliquots of 100 µl were stored at -20°C until required

A 50 µl aliquot of minicells was washed with 100 µl of 1x M9 and then resuspended in 100 µl of 1x M9 containing 0.4% [w/v] glucose and 1.5 µl of methionine assay medium (MAA, Difco). Cycloserine was added to a final concentration of 200 µg/ml and the mixture incubated at 37°C for 90 min to kill any remaining vegetative cells. The cells were collected by centrifugation and resuspended in 100 µl of 1x M9, 0.4% [w/v] glucose, 0.105 g/ml MAA solution preheated to 37°C. Approximately 12 µCi of [^{35}S]Methionine were added and the cells incubated at 37°C for 45 min. The culture was then resuspended in 100 µl 1x M9, 0.4% [w/v] glucose containing 200 µg/ml non-radiolabelled methionine and incubated for 15 min at 37°C. The cells were collected, resuspended in 20 µl SDS-PAGE loading buffer (section 2.8.2) and boiled for 5 min prior to analysis by SDS-

PAGE and autoradiography. The size of radiolabelled proteins was estimated by comparison with molecular weight markers radiolabelled with ^{14}C (Biorad). The signal was enhanced using Amplify (Amersham) to convert the weak ß-emissions to light.

2.9 Assay for ß-galactosidase activity

The assay for ß-galactosidase activity was performed essentially as described by Miller (Miller, 1972). ß-galactosidase hydrolyses *o*--nitrophenyl-ß-D–galactoside (ONPG) to galactose and yellow *o*-nitrophenol. When ONPG is in excess the production of *o*-nitrophenol (measured by absorption at 420 nm) is proportional to the amount of ß-galactosidase present and the time of reaction. Cultures were grown to mid-logarithmic phase and the OD_{600} recorded. Four 100 µl aliquots were removed and added to 900 µl Z buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM 2-mercaptoethanol) containing 20 µl 0.1% [w/v] SDS and 40 µl chloroform. Each sample was vortexed for 10 sec to lyse the cells and equilibrated to 37°C. 200 µl of ONPG (4 mg/ml in Z buffer) was added and the time recorded for production of a faint yellow colour. The reaction was stopped by addition of 0.5 ml 1M Na_2CO_3 and 1 ml transferred to a cuvette and the OD_{550} and OD_{420} determined. The units of ß-galactosidase were calculated from the equation:

$$\text{Activity (Miller units)} = 1000 \times \frac{OD_{420} - (1.75 \times OD_{550})}{t \times v \times OD_{600}}$$

OD_{420} and OD_{550} refer to reaction mixture
OD_{600} refers to the culture immediately before the assay
t = time of reaction (min)
v = volume of culture in assay (ml)(In these experiments v = 0.1)

The reading at OD_{550} compensates for the contribution of light scattering by cells to the OD_{420} allowing calculation of the true absorbance of the *o*-nitrophenol.

2.10 Assay for luciferase activity

Bacterial growth and expression of luciferase were measured as described by Peabody *et al* (Peabody *et al.*, 1989). Cultures were grown overnight in L-broth and 0.1 ml used to inoculate 15ml of minimal media (M9) containing ampicillin (100 µg/ml) thiamine (5 µg/ml), 0.02% glucose and appropriate amino acids. At OD_{600} = 0.1 IPTG was added to a final concentration of 100 µM. Optical density was measured at defined intervals thereafter and culture samples diluted to

OD_{600} = 0.1 with SSC buffer. The light emitted by 100 µl of the diluted sample upon addition of 10µl of the luciferase substrate, n-decanal (1% in ethanol, Sigma) was measured over 10 sec using a luminoskan RS luminometer (Labsystems).

2.11 Computer analysis

Nucleotide sequence was analysed using the programs of the Wisconsin Genetics Computer Group (GCG) (Devereux *et al.*, 1984) running on a Silicon Graphics Crimson mainframe (SERC Daresbury Seqnet facility, UK) under a UNIX platform.

2.11.1 Nucleic acid sequences

The program Map was used to map restriction enzyme recognition sites and predict open reading frames and Mapsort to predict the products of restriction digests. Inverted repeats and potential stem-loop structures were detected with IREP and Stemloop. RNA secondary structure was predicted by MFold using a free energy minimisation method (Jaeger *et al.*, 1989) and the output displayed by PlotFold. Sequences comprising Rho-independent transcriptional terminators were predicted using Terminator (Brendel and Trifonov, 1984). The DNA sequence of open reading frames was converted to the predicted amino acid sequence by the program Translate.

2.11.2 Amino acid sequences

The hydropathy plots (Kyte and Doolittle, 1982) of predicted amino acid sequences were calculated using PepPlot. Accurate secondary structure predictions were obtained from an alignment of the query protein with homologous proteins using the PredictProtein server (Rost and Sander, 1993) at the European Molecular Biology Laboratory, Heidelberg.

2.11.3 Database searching

Similarities between predicted amino acid sequences and those in the National Center for Biotechnology Information (NCBI) non-redundant nucleotide sequence database translated in all reading frames were determined using the Blastx program (Altschul *et al.*, 1990). This database comprises the Brookhaven Protein Data Bank, April 1994 Release, GenBank(R) Release 90, EMBL Data Library, Release 43.0 and the daily updates. Nucleotide sequences were compared

with other nucleotide sequences using the Blastn (Altschul *et al.*, 1990) or Fasta (Pearson and Lipman, 1988) programs. Amino acid sequences were analysed for known motifs using the Motifs program to search the PROSITE database of protein sites and patterns compiled by Dr. Amos Bairoch of the University of Geneva.

2.11.4 Multiple alignments

Multiple sequences were aligned to show conserved residues using the ClustalW program.

CHAPTER 3

The transcriptional organisation of region 1

3.1 The region 1 promoter

3.1.1 Introduction

The expression of group II capsular polysaccharides is temperature regulated (Ørskov *et al.*, 1984). Region 1 is conserved between group II capsule gene clusters and encodes proteins involved in precursor synthesis and polysaccharide modification and export (Boulnois and Jann, 1989)(Roberts, 1996). The CMP-KDO synthetase encoded by region 1 of the K5 capsule gene cluster is expressed at 37°C but not at 18°C, indicating that expression of region 1 is temperature regulated. Initiation of transcription is likely to be the main point of regulation because this prevents the synthesis of unused mRNA and is therefore the most energetically favourable level for regulation. To investigate this possibility it was necessary to identify the promoter(s) responsible for expression of region 1. Analysis of the nucleotide sequence of region 1 (Pazzani *et al.*, 1993) reveals that all the genes are transcribed in the same direction with no large intergenic gaps (Fig. 3.1) and it was therefore proposed that region 1 comprises a single polycistronic operon. This model is supported by the polar effects of mutations in this region (Bronner *et al.*, 1993)(Pazzani, 1992).

The number and sizes of transcripts spanning the specific region of DNA which is used as a probe may be determined by Northern blot analysis. This information can then be combined with genetic and sequence data to compile a model for the transcriptional organisation of the region which includes the approximate location of the promoters. To determine the precise location of a promoter it is necessary to map the 5' end of the transcript. The approach adopted with region 1 was to localise the proposed promoter to a fragment of DNA by its ability to direct expression of a reporter gene cloned downstream. The exact transcription start point was then determined by primer extension from a primer located an appropriate distance downstream of the region containing the promoter. The predicted mRNA 5' end was confirmed by an RNase protection assay. Analysis of a promoter sequence may show whether it is recognised by σ^{70} or an alternative σ factor. The more similar the sequence is to the consensus binding motif the more efficiently transcription is likely to be initiated (Hippel *et al.*, 1982). Features of the surrounding DNA and binding sites for known regulatory proteins provide further information.

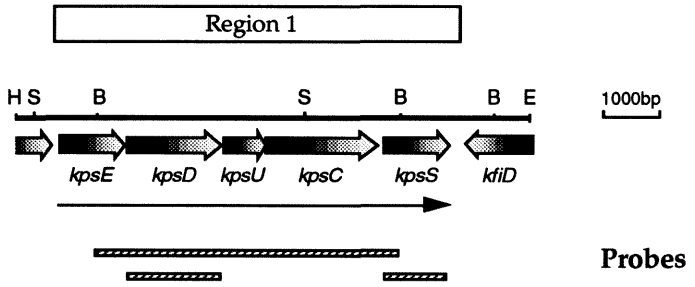


Figure 3.1. Region 1 of the K5 capsule gene cluster (*kps*). The five genes of region 1 (*kpsE-S*) are depicted by shaded arrows indicating the direction of transcription. They potentially comprise a polycistronic operon represented by the lower arrow. The end of an upstream ORF similar to *gutQ* and the first gene of region 2 are also shown. The hatched boxes represent the 5.3 kb *Bam*HI restriction fragment and the *kpsD* and *kpsS* PCR products used as probes in Northern blot analysis. Restriction endonuclease abbreviations: B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; S, *Sma*I.

3.1.2 Northern blots

There is an open reading frame (ORF) which ends 71 bp upstream of *kpsE* and is homologous to *gutQ*, a gene of uncertain function from the glucitol operon (Yamada *et al.*, 1990). However, clones of the *kps* cluster beginning at the *Hin*dIII or *Sma*I sites within this ORF (Fig. 3.1) and lacking the 5' end of the gene can direct capsule production (Bronner *et al.*, 1993)(Pazzani, 1992). This ORF is therefore not required for polysaccharide production and *kpsE* was assumed to be the first gene of the capsule cluster with the promoter located between it and the upstream ORF. Transcription presumably terminates downstream of *kpsS*, the last gene of region 1, because region 2 is transcribed in the opposite direction (Fig. 3.1). No transcriptional terminators, comprising a GC-rich stem-loop followed by a run of U residues, are detected in this region by the Terminator program (Brendel and Trifonov, 1984). However, there are several potential stem-loops which could perform the same function. Region 1 was therefore predicted to be expressed as a single mRNA of approximately 7.0-7.5 kb.

To test this hypothesis RNA was extracted from strain LE392(pGB110) which was shown to be expressing a K5 capsule by sensitivity to K5-specific bacteriophage. In Northern analysis of this RNA a 5.3 kb DNA probe spanning region 1 (Fig. 3.1) hybridised with a transcript of approximately 8.0 kb (Fig. 3.2). A smear of transcripts consisting of intermediates in synthesis and degradation was also detected and produced artefactual bands at the positions were large amounts of rRNA also migrated (Fig. 3.2, open arrows). This phenomenon is commonly

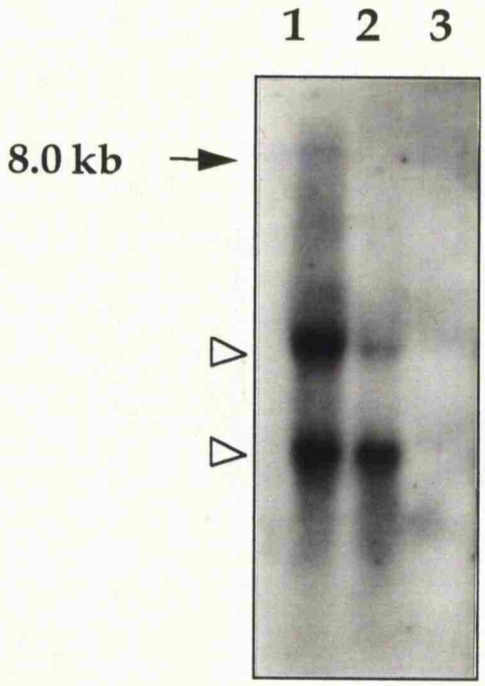


Figure 3.2. Northern analysis of region 1. A 5.3 kb probe spanning region 1 (Fig. 3.1), was hybridised with RNA extracted from LE392(pGB110) grown at either 37°C (Lane 1) or 18°C (Lane 2) and from LE392 at 37°C (Lane 3). An approximately 8.0 kb transcript (solid arrow) and a smear of smaller transcripts was detected in RNA from the capsule-expressing culture (Lane 1). This smear produces artefactual bands at the positions where large amounts of rRNA also migrate (open arrows).

observed in blots with prokaryotic RNA (Newbury *et al.*, 1987). No region 1 mRNA could be detected in the K5 wild type, suggesting a low level of expression from the single chromosomal copy of the *kps* cluster (data not shown). In many repeated analyses with RNA from LE392(pGB110) or other constructs encoding region 1 a high molecular weight smear was observed, suggesting that the transcript is degraded rapidly (Fig. 3.15 and chapter 4, Fig. 4.2). Assuming the top of the smear to represent full-length transcripts, a length of 7.0-8.0 kb was confirmed. Given the uncertainty of the location of the 3' end of the mRNA and the error margin in determination of the length of the transcript these observations are consistent with the promoter being located upstream of *kpsE* as suggested above.

3.1.3 Promoter-probe analysis

Constructs beginning from the *Hin*dIII site upstream of region 1 (Fig. 3.1) can direct capsule expression and it was therefore assumed that the promoter lay between this site and *kpsE*. To confirm this a *Hin*dIII-*Bam*HI fragment including 781 bp upstream of *kpsE* was cloned into the promoter-probe vector pCB192 (Schneider and Beck, 1986)(Fig. 3.3), creating pCB192HB (Fig. 3.4). The K5 DNA

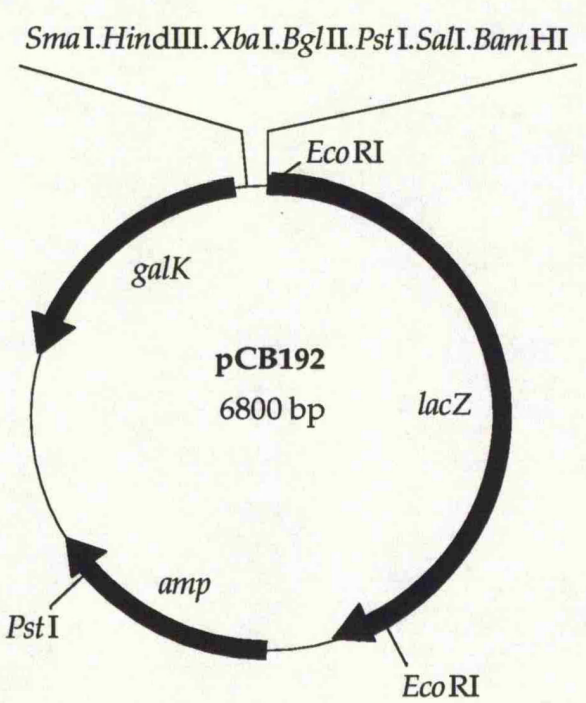
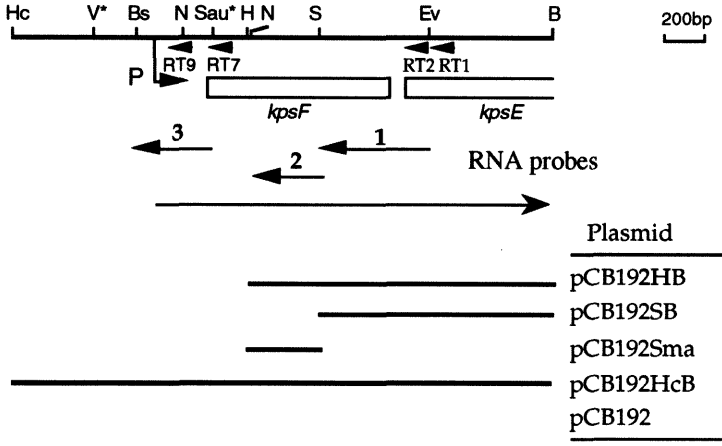


Figure 3.3. Promoter probe vector pCB192 (Schneider and Beck, 1986). The fragment of interest is inserted in the polylinker and promoter activity detected by measuring ß-galactosidase activity expressed from the *lacZ* gene. Promoter activity in the opposite direction can be measured by assaying *galK* activity.

was inserted at the polylinker cloning site upstream of the promoterless *lacZ* gene. ß-galactosidase activity is expressed only if a promoter is present within the inserted fragment and no significant ß-galactosidase activity was detected with pCB192HB or two subclones pCB192SB and pCB192Sma (Fig. 3.4). This suggested that the promoter was further upstream and therefore a larger *Hinc*II-*Bam*HI fragment was cloned into pCB192, creating pCB192HcB (Fig. 3.4). β-galactosidase activity was expressed from pCB192HcB, indicating the presence of a promoter (Fig. 3.4).



Plasmid	β-gal activity (Miller units)
pCB192HB	1.3 (0.6)
pCB192SB	0.56 (0.4)
pCB192Sma	0.0 (0)
pCB192HcB	77.4 (1.9)
pCB192	0.76 (1.0)

Figure 3.4. Detection of the region 1 promoter. The DNA fragments cloned into pCB192 and the resulting constructs with their respective ß-galactosidase activities are shown. The assays were performed in *E. coli* SURE™ and the values are the average of at least three readings, with the standard deviation in brackets. The map of the 5' end of region 1 shows the relevant restriction endonuclease sites and the positions of *kpsE* and the upstream ORF (*kpsF*). The oligonucleotides used in primer extension are indicated by arrowheads and the antisense RNA probes 1, 2 and 3 by arrows. The position determined for the promoter (P) and the resulting transcript are indicated by arrows. Restriction endonuclease abbreviations: B, *Bam*HI; Bs, *Bsa*BI; EV, *Eco*RV; Hc, *Hinc*II; H, *Hin*dIII; N, *Nco*I; Sau, *Sau*3A;V, *Vsp*I (*not unique site).

To facilitate further analysis of the promoter the restriction endonuclease map and then the nucleotide sequence of approximately 1.5 kb further upstream were determined (section 3.1.4). The activity of pCB192HcB and lack of activity of pCB192HB indicated that the promoter must lie within the *Hinc*II-*Hin*dIII fragment (Fig. 3.4). The most likely location was upstream of the ORF adjacent to *kpsE*, now designated *kpsF* (chapter 5).

3.1.4 Determination of nucleotide sequence upstream of *kpsE*

When it was demonstrated that the promoter was not immediately upstream of *kpsE* it became necessary to sequence further upstream. A DNA fragment spanning from an *Eco*RI site in the cosmid vector used to clone the capsule gene cluster to the *Bam*HI site in *kpsE* (Fig. 3.4) was cloned in pTZ19 (Fig. 3.6). A restriction map was generated and smaller fragments subcloned into M13mp18/19 or pTZ18/19 for the preparation of single-stranded templates. The templates were sequenced by the Sanger dideoxy chain termination method (Sanger *et al.*, 1977) using the universal primers or oligonucleotides synthesised specifically for this purpose. A series of overlapping sequences covering both strands were generated (Fig. 3.5). These were compiled to form the final sequence shown in Figure 3.11.

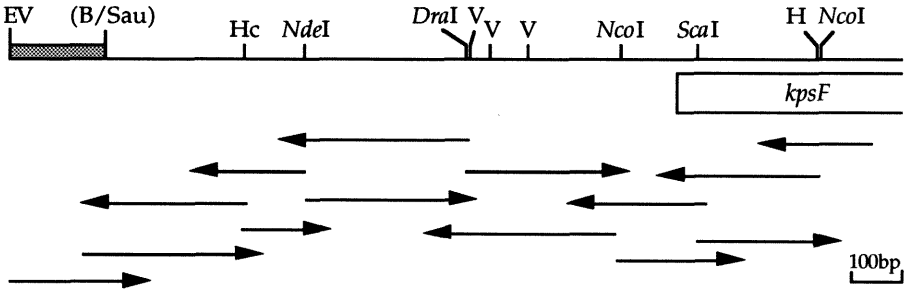


Figure 3.5. Sequencing strategy. The arrows show the overlapping sequences obtained from each reaction, which were compiled to give the entire sequence of both DNA strands upstream of region 1. The top line shows the restriction endonuclease sites: Abbreviations; EV, *Eco*RV; Bs, *Bsa*BI; Hc, *Hin*cII; H, *Hin*dIII; V, *Vsp*I; (B/Sau), *Bam*HI vector site destroyed by insertion of the *Sau*3A-cleaved K5 DNA. The shaded box represents the vector pcos4. The position of the upstream ORF, now designated *kpsF* is indicated.

3.1.5 RNase protection and primer extension analysis

RNase protection is a sensitive method for detecting specific transcripts. A radiolabelled antisense RNA probe is hybridised with total cellular RNA and the region of the probe for which the complementary sense strand exists is protected during subsequent digestion with a single-strand specific ribonuclease. Antisense

RNA probes were generated using the multifunctional phagemid pTZ19R (Fig. 3.6)(Mead *et al.*, 1986). The appropriate DNA fragment was inserted at the multiple cloning site (MCS) and the specific RNA synthesised from the adjacent T7 promoter. It was also possible to produce single-stranded DNA from the f1 origin of replication with the help of bacteriophage M13KO7, enabling the nucleotide sequence of the inserted DNA to be determined using the reverse primer (Rev).

To preclude the possibility of a weak promoter immediately upstream of *kpsE* antisense RNA probe 1 (Figs. 3.4 and 3.6) was generated. Firstly, recombinant plasmid pDS111 was constructed by cloning an *Eco*RI-*EcoRV* DNA fragment which included sequences upstream of region 1 (the shaded box represents the original vector, pcos4) into the MCS of pTZ19R (Fig. 3.6). pDS111 was then linearised with *Sma*I and RNA probe 1 (Fig. 3.4) transcribed from the T7 promoter. When RNA probe 1 was hybridised with total cellular RNA from the K5 wild type and LE392 carrying constructs expressing region 1 the full length probe was protected (Fig. 3.7), confirming that the promoter was not adjacent to *kpsE*, but lay further upstream. Antisense RNA probe 2, which includes sequences further upstream, was therefore generated. pDS110 (Fig. 3.6) carries an *Eco*RI-*Sma*I fragment and was linearised with *Nco*I to enable the synthesis of RNA probe 2 (Fig. 3.4). RNA probe 2 was also fully protected in RNase protection assays (Fig. 3.7), indicating that the promoter was located not within, but 5' to the ORF upstream of *kpsE* (*kpsF*).

Primer extension enables determination of the exact start point of transcription. Oligonucleotides complementary to the mRNA downstream of the proposed promoter were hybridised with RNA from a capsule-expressing strain. They were used to prime the synthesis of DNA complementary to the 5' end of the region 1 mRNA by reverse transcriptase. Comparison of the cDNA with a DNA sequence ladder generated with the same oligonucleotide indicates which nucleotide is the first to be transcribed.

When primer extension was performed with oligonucleotide RT1 (5'-AGCCAGATAGATCAGCAGCACCGC-3') or RT2 (5'-CATCCAGGATACGG CAGACTTCAC-3') complementary to the 5' end of *kpsE* (Fig. 3.4) a band was observed which confirmed a transcript start point far upstream of *kpsE* (data not shown). Primer RT7 (5'-GCGAACAGAGGTAATTAGATATGG-3') located nearer to the promoter (Fig. 3.4) was chosen to define the transcription start point more accurately. Primer extension with RT7 revealed three bands potentially representing the 5' ends of transcripts approximately 40, 90 and 220 bp

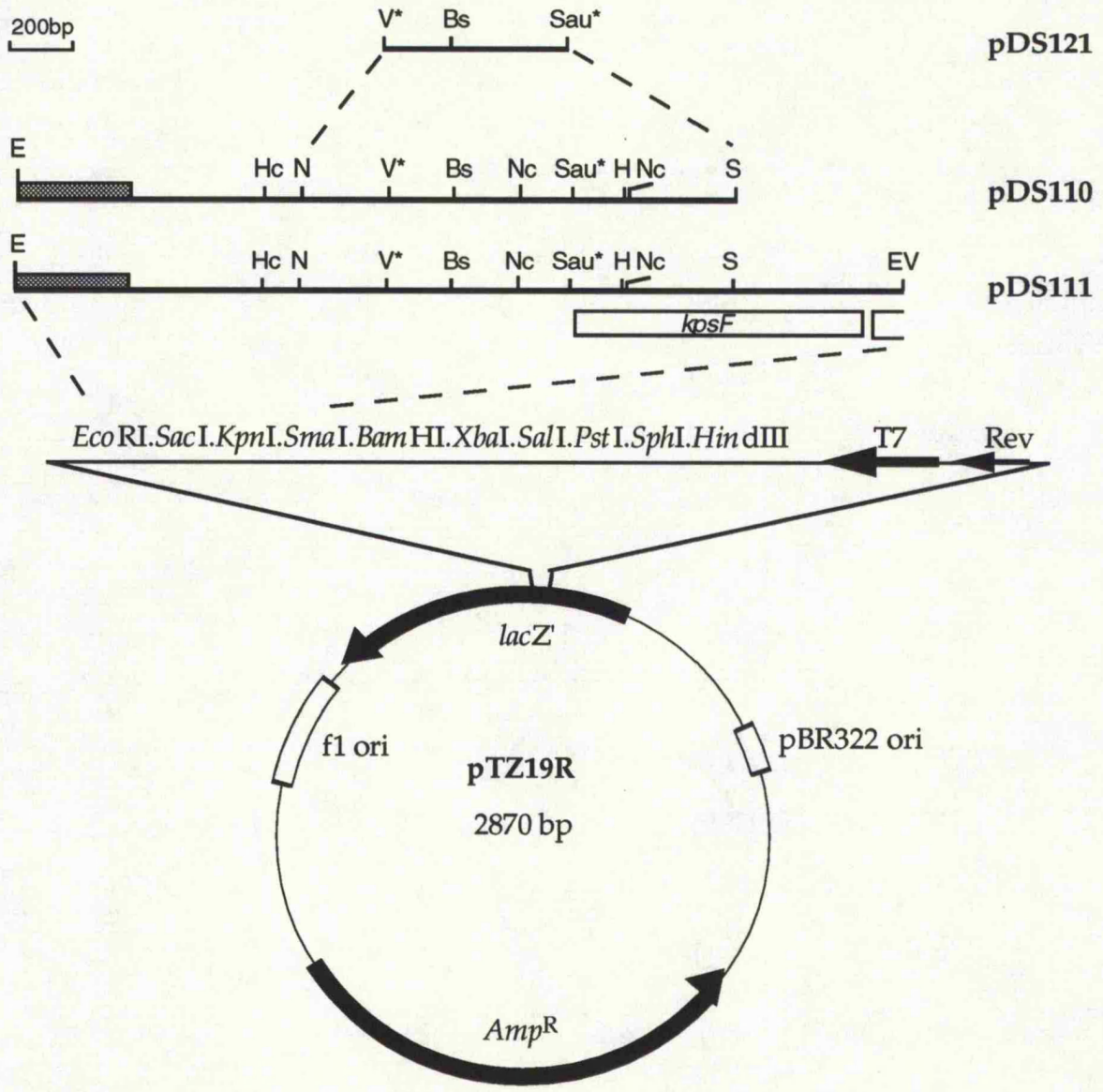


Figure 3.6. Phagemid pTZ19R (Mead *et al.*, 1986) and constructs for the synthesis of antisense RNA probes. The positions of the T7 promoter and reverse DNA sequencing primer are indicated with arrows. Horizontal lines represent the DNA fragments from upstream of region 1 which were cloned to form recombinant plasmids pDS111, pDS110 and pDS121 and generate antisense RNA probes 1, 2 and 3, respectively (the shaded boxes represent the original vector, pcos4). The position of *kpsF*, the ORF upstream of *kpsE*, is indicated with a box. The restriction endonuclease sites present in the pTZ19R multiple cloning site (MCS) are shown. Abbreviations: B, *Bam*HI; Bs, *Bsa*BI; E, *Eco*RI; EV, *Eco*RV; Hc, *Hin*cII; H, *Hin*dIII; Nc, *Nco*I; N, *Nde*I; S, *Sma*I; Sau, *Sau*3A;V, *Vsp*I (*not unique site).

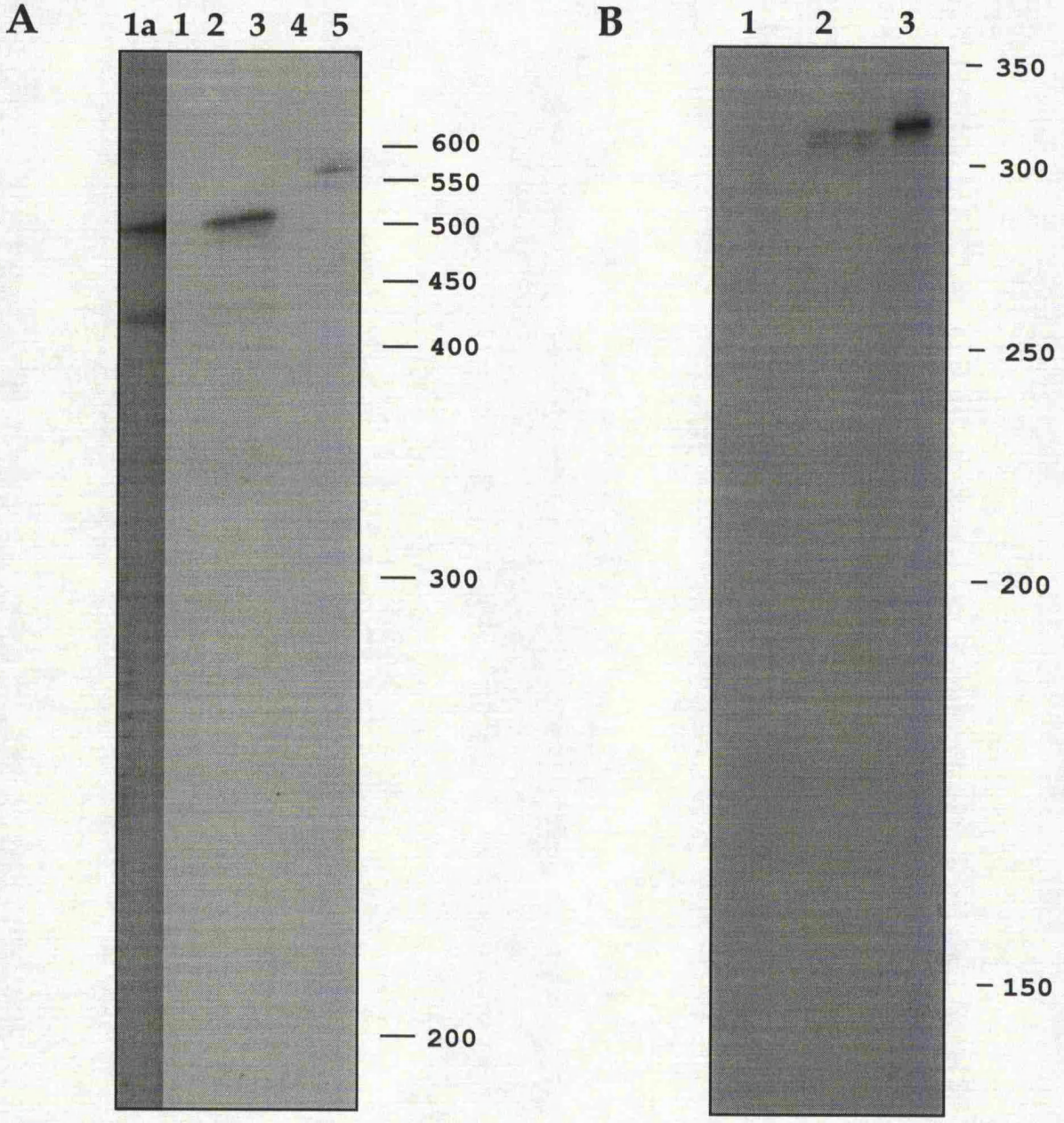


Figure 3.7. RNase protection. A. Radiolabelled RNA probe 1 (Fig. 3.4) was synthesised from pDS111 (Fig. 3.6), hybridised with total RNA, digested with RNase One and the protected fragments analysed by gel electrophoresis. Lane 1; K5 wild type (100 μg). Lane 1a; longer exposure of lane 1. Lane 2; LE392/pDS200 (50μg). Lane 3; LE392/pH18 (80 μg). Lane 4; LE392/pUC18 (80 μg). Lane 5; full length undigested probe. Sizes are indicated in nucleotides. B. Radiolabelled RNA probe 2 was hybridised with RNA from: LE392 (Lane 1), K5 (Lane 2), and LE392/pDS200 (Lane 3). The length of probe complementary to K5 sequences is 324 bases.

upstream of *kpsF* (Fig. 3.8, Lane 3). These findings are in general agreement with results presented for the region 1 promoter of the K1 cluster (Cieslewicz and Vimr, 1994). However, reverse transcriptase may pause and dissociate from the mRNA at certain points, perhaps in regions of secondary structure. This results in cDNAs which are shorter than those extending to the 5' end of the mRNA. To distinguish whether the bands observed with RT7 represented three transcription start points or were in part due to reverse transcriptase dissociation an RNase protection assay was performed.

An antisense RNA probe spanning the 5' ends of the potential transcripts was required. The only available restriction endonuclease site downstream of the potential transcript starts was *Sau*3A (Fig. 3.4). Plasmid pDS110 comprises an *Eco*RI-*Sma*I fragment spanning the region upstream of *kpsF*, cloned in pTZ19R (Fig. 3.6). A *Vsp*I-*Sau*3A fragment was ligated to the compatible ends generated by cleavage of pDS110 with *Nde*I and *Bam*HI to create pDS121 (Fig. 3.6). pDS121 was linearised at a *Bsa*BI site upstream of the first potential transcript 5' end and radiolabelled run-off transcripts synthesised from the T7 promoter in pTZ19 (antisense RNA probe 3, Fig. 3.6).

Knowing the position of the *Sau*3A site it was possible to calculate the lengths of fragments which would be protected by transcripts initiating at each of the potential start points. When the RNase One digestion was performed at 37°C strong bands of 250 bases and weak bands of 260 bases were protected by RNA from LE392/pDS200 (Fig. 3.9, Lane 4) and the K5 wild type (Fig. 3.9, Lane 2). The smaller product was probably due to unwinding at the end of the RNA duplex resulting in digestion of approximately ten ribonucleotides. This explanation was supported when the RNase One digestion was repeated at 25°C; the main protected band was at 260 bases with a range of products down to 250 bases, probably due to reduced unwinding. A protected fragment of 260 bases is in agreement with transcription initiation occurring at the furthest upstream site suggested by primer extension (Fig. 3.8). The protection of the 260 bp transcript by mRNA from the K5 wild type confirmed that this is the main promoter *in vivo*. The presence of only one major band corresponding to the first initiation point (Fig. 3.9), suggested that the other 2 bands observed in primer extension do not represent separate transcripts.

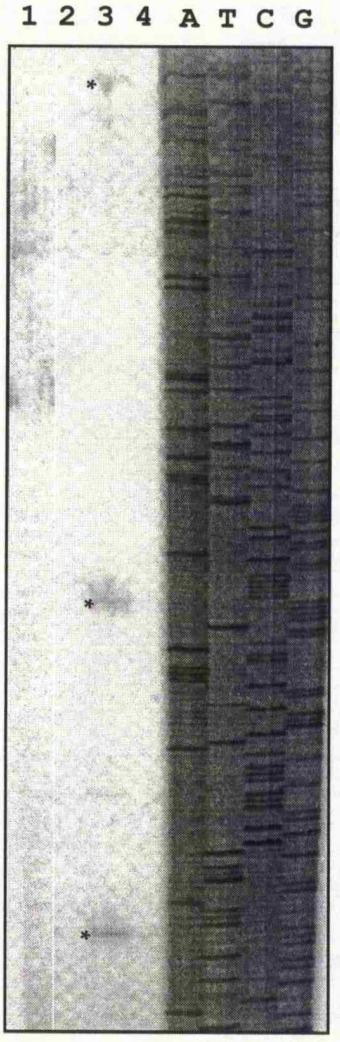


Figure 3.8. Primer extension with oligonucleotide RT7. Extension reactions were performed with oligonucleotide RT7 (Figs. 3.4 and 3.11) annealed with RNA extracted from: LE392/pKK223-3 incubated at 37°C (Lane 1), LE392/pDS200 at 18°C (Lane 2) and 37°C (Lane 3) and K5 at 37°C (Lane 4). The three bands corresponding to the potential transcription start sites are indicated (*).

When the RNase ONE digestion to remove single-stranded RNA was performed at 37°C two fragments of approximately 110 ribonucleotides were observed (Fig. 3.9) which did not correspond to any of the transcripts predicted by primer extension. However, when the temperature was reduced to 25°C the bands disappeared suggesting that they were caused by nicking of the RNA in AU-rich regions where the duplex was unstable and likely to be temporarily single-stranded at the higher temperature. The lack of protection from RNA extracted from the K5 wild type incubated at 18°C (Fig. 3.9. Lane 3) suggests that transcription is regulated by temperature (see chapter 4).

To determine the nucleotide at which transcription initiates primer extension was performed with oligonucleotide RT9 (5'-GTCAGTCCATGCTTATAT GCAGG-3'), designed to anneal 100 bp downstream from the estimated start point (Figs. 3.4 and 3.11). A single band was identified (Fig. 3.10) indicating that transcription starts at a C residue on the sense strand of the DNA template, 225 bp upstream of *kpsF* (Fig. 3.11).

Figure 3.9. RNase protection with radiolabelled antisense RNA probe 3 (Figs. 3.4 and 3.5). The probe was hybridised with total cellular RNA from the K5 wild type, 37°C (Lane 2) and 18°C (Lane 3), LE392/pDS200, 37°C (Lane 4), and LE392/pKK223-3, 37°C (Lane 5). Digestion with RNase One was performed at 37°C (left panel) or 25°C (right panel). The protected fragments were analysed by gel electrophoresis and their approximate sizes determined by comparison with a DNA sequence ladder. The full length undigested probe is 355 bases (Lane 1), in agreement with its predicted length.

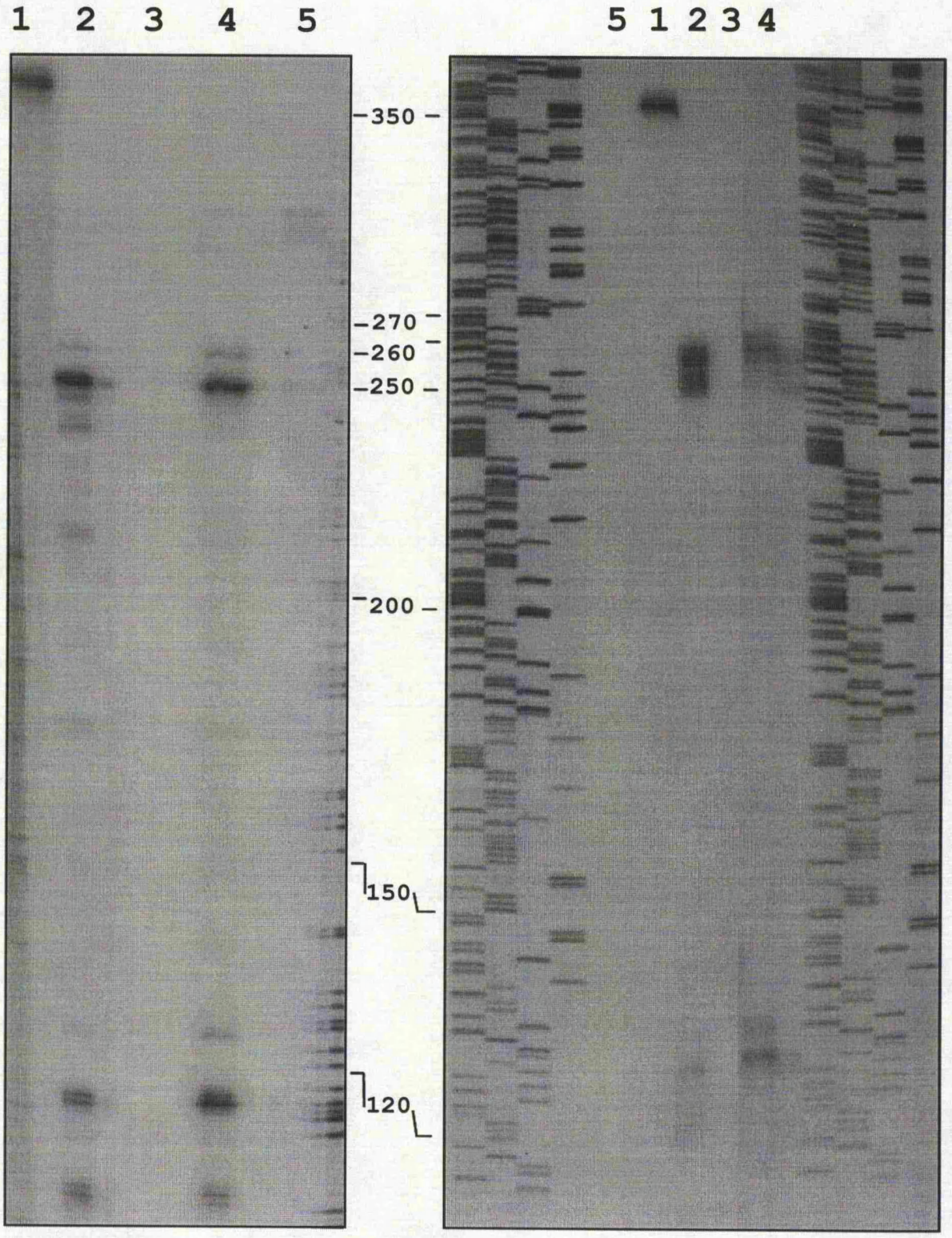


Figure 3.9

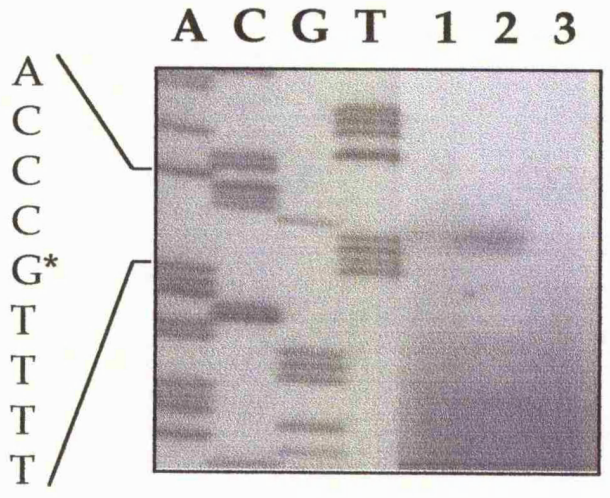


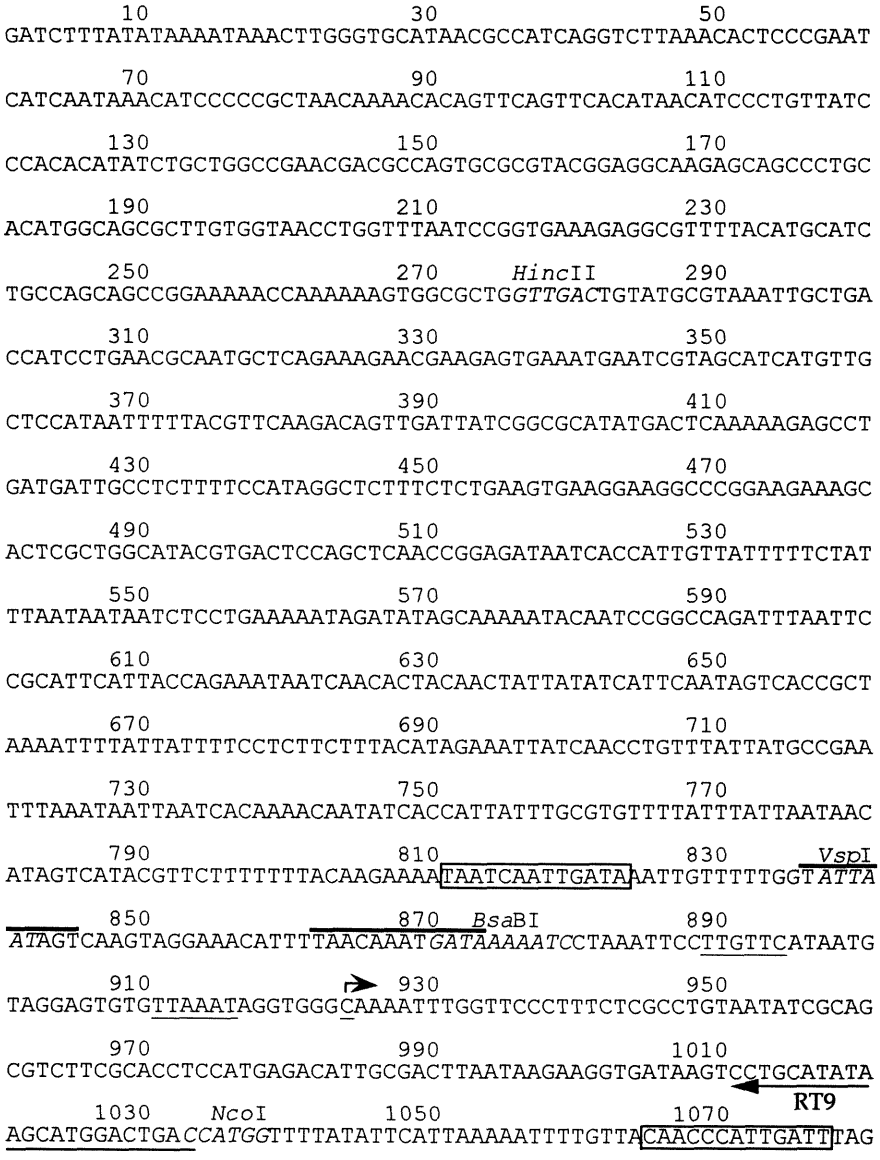
Figure 3.10. Primer extension with oligonucleotide RT9. Extension reactions were performed with oligonucleotide RT9 (Figs. 3.4 and 3.11) annealed with RNA extracted from the K5 wild type (Lane 1) LE392/pDS200 (Lane 2) and LE392/pKK223-3 (Lane 3), all incubated at 37°C. The products were analysed by gel electrophoresis adjacent to a DNA sequencing ladder generated using the same primer. The bands observed with LE392/pDS200 and K5 end at a G residue (starred). Transcription therefore starts at cytosine residue 924 in the sense strand (marked with an arrow in Fig. 3.11).

3.1.6 Analysis of DNA sequence upstream of *kpsE*

Analysis of the nucleotide sequence (Fig. 3.11) revealed that the region 1 promoter is located upstream of an ORF, the 3' end of which had previously been observed upstream of *kpsE* (Fig. 3.1). This ORF is now called *kpsF* (Figs. 3.4, 3.5 and 3.11) and is described in detail in chapter 5. The transcription start point is 125 nucleotides upstream of *kpsF* and upstream of it are potential σ^{70} -35 and -10 regions (Fig. 3.11). The likelihood that transcription is directed by σ^{70} is reinforced by the absence of an alternative σ factor binding site.

The only regulatory protein known to be required for the expression of K5 capsular polysaccharide is RfaH (Stevens *et al.*, 1994). RfaH is thought to mediate transcriptional antitermination and downstream of the promoters of the genes which it regulates is a highly conserved sequence called the JUMPStart (Hobbs and Reeves, 1994), which probably acts in the mRNA. This sequence is present upstream of region 3 but is not found upstream of region 1. This suggested that RfaH is not involved in the temperature regulation of region 1 and this was confirmed by Western blotting (chapter 4). The role of RfaH in regulation of the K5 gene cluster is discussed in chapter 6.

Many regulatory proteins which affect the activity of specific promoters have been identified in *E. coli*. In some cases the basis of specificity is unclear while in others the protein acts at a conserved DNA sequence. The regulation of capsular polysaccharide expression is likely to be co-ordinated with that of other virulence factors (Miller *et al.*, 1989) and the DNA upstream of *kpsF* was therefore searched for the consensus sequences associated with proteins which regulate virulence factors. These included the cyclic AMP receptor protein (CRP), which requires cAMP to enable binding at the consensus sequence (5'-TGTGAnntngnTCACA-3'). CRP regulates a diverse set of genes, indirectly in response to the carbon source which affects cAMP levels. CRP is important in *Salmonella* virulence and regulates enterotoxin and flagellar gene expression in *E. coli* (see (Dorman, 1994) and references therein). Integration host factor (IHF) binds DNA with the sequence: 5'-(C/T)AANNNNTTGAT(A/T)-3' and affects the activity of a large number of operons (Freundlich *et al.*, 1992). IHF bends the DNA by up to 140°C and enables distant sites, often bound by regulatory proteins, to come together and for example activate RNA polymerase. The leucine-responsive regulatory protein (LRP) controls a regulon of genes, including several virulence factors such as the Pap and other pili operons (Woude *et al.*, 1992). The ferric uptake regulator (FUR) acts with Fe^{2+} to repress genes of the iron assimilation pathways. FUR binds to an inverted repeat sequence or 'iron box'


10
30
50
GATCTTTATATAAAATAAACTTGGGTGCATAACGCCATCAGGTCTTAAACACTCCCGAAT
70
90
110
CATCAATAAACATCCCCCCGCTAACAAAACACAGTTCAGTTCACATAACATCCCTGTTATC
130
150
170
CCACACATATCTGCTGGCCGAACGACGCCAGTGCGCGTACGGAGGCAAGAGCAGCCCTGC
190
210
230
ACATGGCAGCGCTTGTGGTAACCTGGTTTAATCCGGTGAAAGAGGCGTTTTACATGCATC
250
270
HincII
290
TGCCAGCAGCCGGAAAAACCAAAAAAGTGGCGCTGGTTGACTGTATGCGTAAATTGCTGA
310
330
350
CCATCCTGAACGCAATGCTCAGAAAGAACGAAGAGTGAAATGAATCGTAGCATCATGTTG
370
390
410
CTCCATAATTTTTACGTTCAAGACAGTTGATTATCGGCGCATATGACTCAAAAAGAGCCT
430
450
470
GATGATTGCCTCTTTTCCATAGGCTCTTTCTCTGAAGTGAAGGAAGGCCCGGAAGAAAGC
490
510
530
ACTCGCTGGCATACGTGACTCCAGCTCAACCGGAGATAATCACCATTGTTATTTTTCTAT
550
570
590
TTAATAATAATCTCCTGAAAAATAGATATAGCAAAAATACAATCCGGCCAGATTTAATTC
610
630
650
CGCATTCATTACCAGAAATAATCAACACTACAACTATTATATCATTCAATAGTCACCGCT
670
690
710
AAAATTTTATTATTTTCCTCTTCTTTACATAGAAATTATCAACCTGTTTATTATGCCGAA
730
750
770
TTTAAATAATTAATCACAAAACAATATCACCATTATTTGCGTGTTTTATTTATTAATAAC
790
810
830
VspI
ATAGTCATACGTTCTTTTTTTACAAGAAAATAATCAATTGATAAATTGTTTTTGGTATTA
850
870
BsaBI
890
ATAGTCAAGTAGGAAACATTTTAACAAATGATAAAAATCCTAAATTCCTTGTTCATAATG
910
930
950
TAGGAGTGTGTTAAATAGGTGGGCAAAATTTGGTTCCCTTTCTCGCCTGTAATATCGCAG
970
990
1010
CGTCTTCGCACCTCCATGAGACATTGCGACTTAATAAGAAGGTGATAAGTCCTGCATATA
RT9
1030
NcoI
1050
1070
AGCATGGACTGACCATGGTTTTATATTCATTAAAAATTTTGTTACAACCCATTGATTTAG

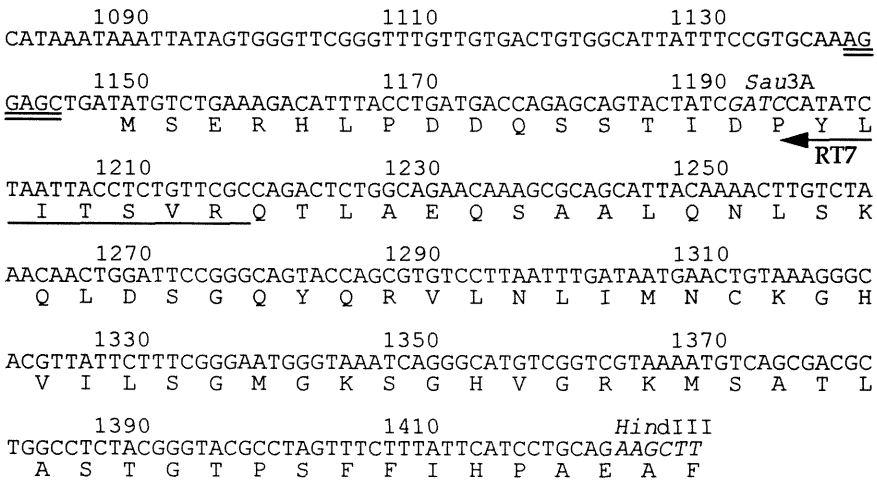


Figure 3.11. Nucleotide sequence upstream of region 1 of the K5 *kps* locus. The protein translation of the ORF now designated KpsF is shown in single letter code below the sequence and the putative Shine-Dalgarno ribosome-binding site is doubly underlined. Selected restriction endonuclease cleavage sites are indicated in italics. Primers RT7 and RT9 used to determine the transcript start point are indicated by arrows. The site of transcription initiation is underlined and marked with an arrow. Potential -10 and -35 promoter sequences are underlined. The two sequences matching the IHF binding site consensus are boxed. The two AT-rich regions conserved with the region 3 promoter are overlined.

(5'-GATAATGATAATCATTATC-3') located in the promoter regions of the genes which it regulates.

Two sequences at -101 to -113 and +142 to +154 relative to the transcript start were identified which match the IHF consensus (Fig. 3.11). No other matches were detected. The position of the binding sites relative to the promoter are consistent with those described in other systems. It has been suggested that high AT content may be involved in the bending process mediated by IHF (Wozniak, 1994). The DNA upstream of the transcript start has a high AT content (Fig. 3.12) with polyA and polyT tracts. Periodic phasing of such tracts with the helix repeat has been correlated with experimentally observed intrinsically curved DNA. Curved DNA upstream of various promoters has been shown to activate transcription (Collis *et al.*, 1989)(Ohyama *et al.*, 1992). However, analysis of the DNA upstream of the region 1 promoter using the algorithm devised by Plaskon and Wartell (Plaskon and Wartell, 1987) suggests that it is not curved in the absence of IHF.

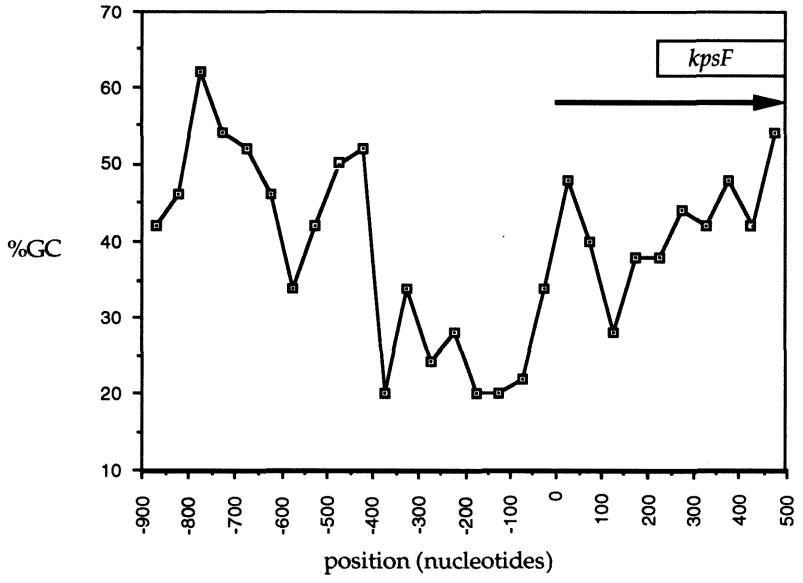


Figure 3.12. GC content of the promoter region. The %GC content was calculated over a window of 50 nucleotides, with the first nucleotide to be transcribed designated +1. The position of the transcript (arrow) and *kpsF* are indicated. There is a region of very low GC content, compared to the average of approximately 50% for *E coli* DNA, spanning from 50 to 400 nucleotides upstream of the transcription start point.

In an attempt to reveal any uncharacterised but conserved regulatory sequences the promoter region was compared with those of other genes involved in polysaccharide production but no similarities were observed. However, alignment of the transcript start point with that upstream of region 3 of the K5 *kps* cluster (Stevens, M. P., unpublished) revealed two conserved sequences at similar positions within each promoter region (Fig. 3.13). Both sequences are extremely AT rich: Sequence A (5'-TATTAATAGT-3') is totally conserved while sequence B (5'-TAACAA(A/T)TGATA-3') varies at one residue. The conserved B region resembles but does not match the consensus for an IHF binding site. The conserved sequences upstream of region 3 in K5 are identical to those in the K1 *kps* cluster (it should be noted that these sequences are almost identical throughout their length). A search of the NCBI non-redundant database for these consensus sequences revealed no perfect matches, the closest having three mismatches. The similar sequences were generally not located in potential promoter regions and are therefore unlikely to be related. The conserved sequences appear to be specific to group II capsule gene clusters and are likely to be involved in the regulation of the conserved regions 1 and 3, probably by binding of a protein which affects transcription initiation by RNA polymerase.

Translation of the sequence upstream of *kpsF* in the first forward reading frame reveals a short polypeptide with similarity to several transposase proteins encoded by insertion sequence (IS) elements (Fig. 3.14). The polypeptide begins 109 bp into the known K5 sequence and is similar with the C-termini of the other proteins. The similarity does not continue upstream in another reading frame, therefore if the K5 sequence is derived from a gene within an IS element the reading frame has since been disrupted several times or a DNA rearrangement has occurred. The similarities range from 48% identity over 56 amino acids with IS*492* from *Pseudomonas atlantica* to 32% over 50 amino acids with IS*900* from *Mycobacterium paratuberculosis*. The lengths of these similarities are too short to confirm that the sequences share a common ancestor (Doolittle *et al.*, 1986). However, the likelihood that the K5 sequence did encode a transposase is increased because it is reasonable to expect IS elements flanking the *kps* locus, as occurs in the *cap* locus (Kroll *et al.*, 1991)(see discussion, section 3.4).

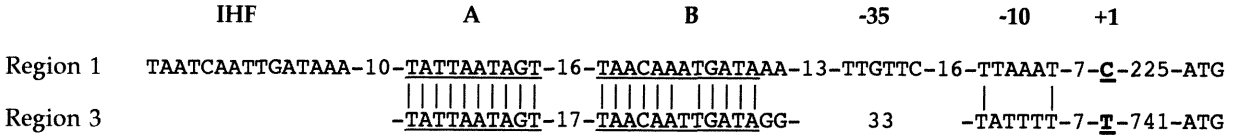


Figure 3.13. Comparison of the K5 region 1 and region 3 promoter sequences. Both promoters have a region resembling the σ^{70} -10 consensus (TATAAT), but while the region 1 sequence contains three of the conserved nucleotides comprising the -35 consensus sequence (TTGACA) no -35 box is present in region 3. At a similar distance upstream of both transcription start points are two conserved sequences labelled A and B. A putative IHF-binding site matching the consensus (C/T)AANNNNTTGAT(A/T) (Freundlich et al., 1992) is present further upstream of region 1 but not in region 3.

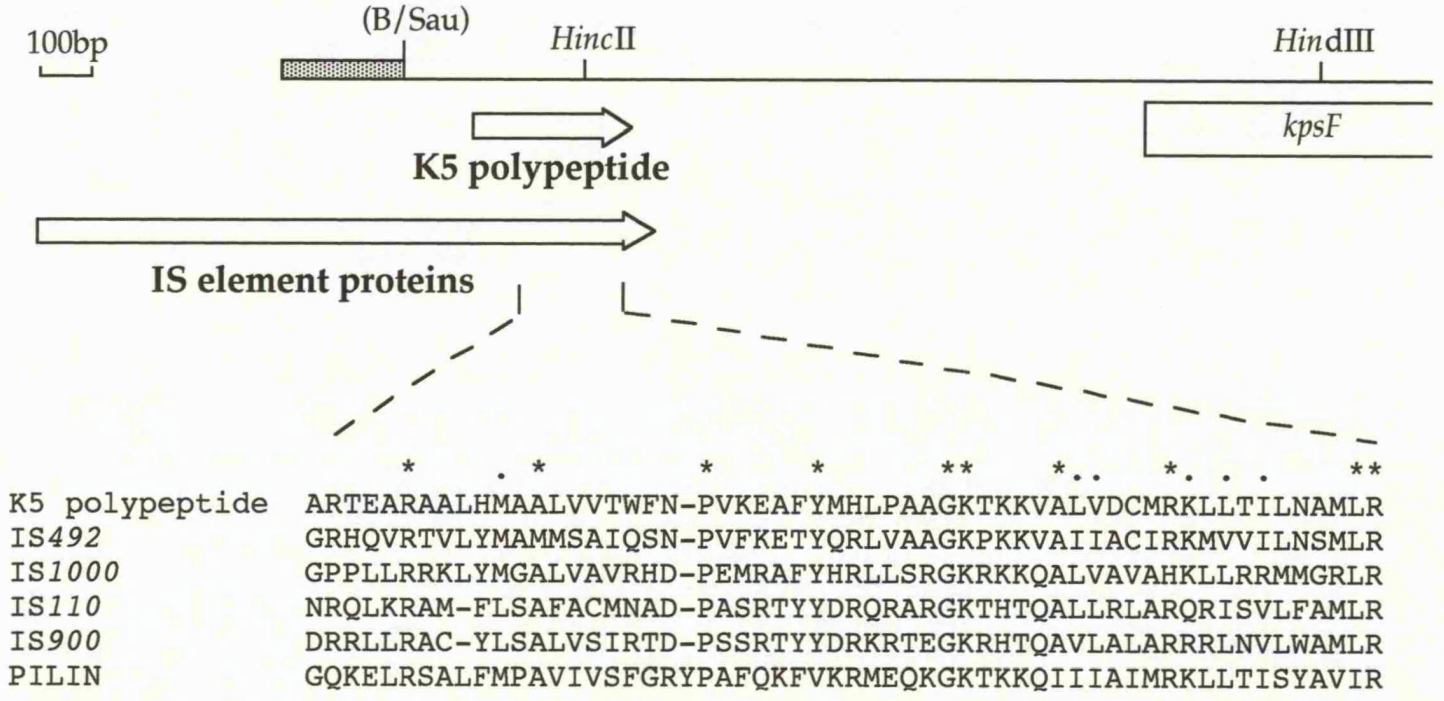


Figure 3.14. Similarity of K5 polypeptide with IS element proteins. A putative polypeptide (open arrow) encoded upstream of *kpsF* (box) is similar with the C-terminal regions of several proteins encoded by IS elements (large open arrow). These are: IS*492* hypothetical 33 kDa protein from *Pseudomonas atlantica* (Accession no. A32816); IS*1000* hypothetical protein 1A from *Thermus aquaticus* (Accession no. S27739); IS*110* hypothetical 43.6 KDa protein from *Streptomyces coelicolor* (Accession no. P19780); IS*900* hypothetical 42 kDa protein from *Mycobacterium paratuberculosis* (Accession no. P14322) and the pilin gene inverting protein from *Moraxella lacunata* (Accession no. A37759). The sequences are aligned over the region of greatest similarity. Amino acids which the K5 polypeptide shares with at least four of the other proteins or at which conservative changes have occurred are indicated by a star (*) or dot (.), respectively. The shaded box represents the vector pcos4 and (B/Sau) indicates the junction with the inserted K5 DNA.

3.2 *kpsS*-specific transcript

3.2.1 Introduction

The results presented in section 3.1 suggest that region 1 is transcribed primarily as a large polycistronic operon. However, smaller stable transcripts produced by RNA processing or transcription initiation at a second promoter may exist. Previous studies have suggested that there may be a promoter upstream of *kpsS* in *E. coli* K1 (Cieslewicz and Vimr, 1995). A *kpsS*-specific transcript would be approximately 1.3 kb and could be obscured in Northern blot analysis by the effects of rRNA (Newbury *et al.*, 1987). The first potential ATG codon of *kpsS* is only 34 bp from the end of *kps*C, however this start lacks a SD sequence and translation may initiate at the second ATG codon which is preceded by a SD sequence. In this case the gap of 100 bp between the genes would be sufficient to encode a promoter. Region 1 is functionally conserved between group II capsule gene clusters and the transcriptional organisation is likely to be retained. The available sequence of *kpsS* from K1 is almost identical with that from K5 except at the 3' end near the junction with region 2 (Drake, 1991). Vimr E. R., *et al* (Vimr *et al.*, 1989) reported that in K1 a plasmid carrying a fragment spanning from within region 2 to the *Bam*HI site equivalent to that marked within the putative K5 *kpsS* gene (Fig. 3.16) can complement *kpsS* mutations. This suggests that *kpsS* may be transcribed separately (Cieslewicz and Vimr, 1995). Steenbergen, S. M., *et al* suggested that translation of *kpsS* starts downstream of the *Bam*HI site and have demonstrated a KpsS'-LacZ fusion protein with a size consistent with this hypothesis (Steenbergen *et al.*, 1992). However, this start seems unlikely, at least in K5 in which the ORF continues for 165 amino acids upstream. In addition a 44 kDa protein, consistent with translational initiation at the first or second ATG codon, is expressed from a plasmid containing a 1.5 kb *Hin*cII fragment spanning the K5 *kpsS* region (Pazzani *et al.*, 1993). The presence of a promoter upstream of *kpsS* and other potential promoter regions and RNA processing sites within region 1 are investigated.

Derivatives of phage Mu can be used to insert a promoterless *lacZ* gene, which also lacks signals for translation initiation, at random into plasmid DNA (Berg *et al.*, 1989). ß-galactosidase activity is expressed only when a translational fusion is formed with a gene which is being transcribed and translated. Such a fusion was isolated within a clone of the *E. coli* K1 *kps* cluster upstream of the *Bam*HI site in *kpsS* and was expressed in the opposite orientation to *kpsS* (Steenbergen and Vimr, 1990)(Steenbergen *et al.*, 1992). An active Tn*phoA* insertion has been isolated just beyond the 3' end of the K5 *kpsS* gene but also in the opposite

orientation (Esumeh, F., unpublished). In the absence of any significant ORFs encoded on the opposite strand to *kpsS* the transcriptional activity suggested by these observations might reflect a regulatory antisense RNA, which would be detected by Northern blot analysis. DNA fragments within and downstream of *kpsS* were therefore tested for promoter activity directed against the gene.

3.2.2 Northern blot, promoter-probe and RT-PCR analysis

The KpsS protein may be encoded by a specific transcript resulting from RNA processing or transcription initiation at a second promoter. A Northern blot was hybridised with a *kpsS*-specific probe (Fig. 3.15A) then stripped and re-probed with a *kpsD*-specific probe (Fig. 3.15B)(both probes were prepared by PCR, Fig. 3.1). The two probes both detected a high molecular weight smear, presumably the degradation products of the 8.0 kb region 1 transcript initiating upstream of *kpsF*. However, the *kpsS* probe hybridised strongly to transcripts of approximately 1.3 kb (Fig. 3.15A, Lane 1), just below the lower ribosomal band, which were not detected by the *kpsD* probe which hybridised predominantly to higher molecular weight transcripts (Fig. 3.15B, Lane 1). This suggests that a small transcript spans at least part of the *kpsS* gene. This transcript could be generated from a promoter directing transcription of this region in either direction. To determine whether such a promoter existed various DNA fragments (Fig. 3.16) were cloned into pCB192 (Fig. 3.3) and assayed for the ability to direct transcription of the *lacZ* gene. No promoter activity was detected from a 5.3 kb *Bam*HI fragment spanning from *kpsE* to the beginning of *kpsS* (pCB192S1). This suggested that *kpsS* is not expressed as a separate transcript from a promoter within region 1. None of the clones designed to look for promoter activity in the opposite direction (pCB192S2, 3 and 4), which would be required to generate an antisense RNA within *kpsS*, expressed significant ß-galactosidase activity.

If, as suggested by promoter-probe analysis, there is no *kpsS*-specific promoter *kpsS* must be transcribed as part of the main region 1 transcript. In this case the mRNA would span from *kpsC* into *kpsS*. Alternatively the lack of activity observed by promoter-probe analysis might be due to the disruption of promoters or the absence of cis-acting sequences on the fragments chosen. To investigate these possibilities the pattern of transcription from pDS200, which encodes all of region 1 and part of region 2 (chapter 4, Fig. 4.1), was investigated using reverse transcriptase (RT) PCR. This is a sensitive technique because the PCR amplification step enables the detection of low levels of a specific mRNA following its reverse transcription into cDNA. The primers used and the

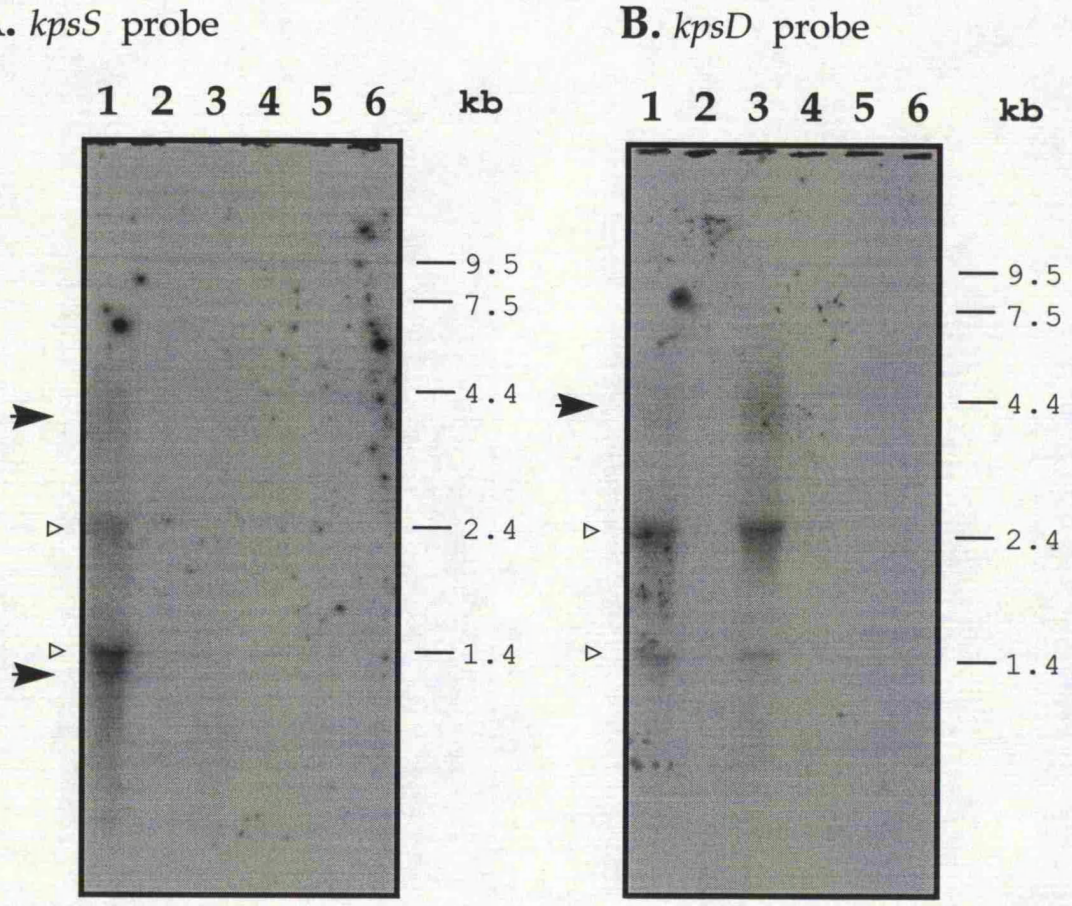


Figure 3.15. Northern blot probed with a radiolabelled PCR product encoding *kpsS* (Panel A), stripped and reprobed with a PCR product encoding *kpsD* (Panel B). RNA was extracted from the following cultures: LE392/pDS200, 37°C (Lane 1); LE392/pDS200, 18°C (Lane 2); LE392/pDS200T, 37°C (Lane 3); LE392/pDS200T, 18°C (Lane 4);. LE392/pDS201, 37°C (Lane 5); LE392/pKK223-3, 37°C (Lane 6). The *kpsS* probe hybridises mainly to small transcripts (lower arrow), although larger transcripts are detected (top arrow). The artifactual bands often observed at the position of the rRNA are indicated by open arrows. The size of the RNA markers is given in kb.

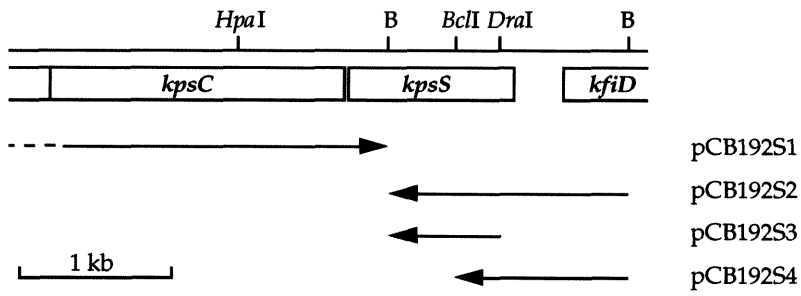


Figure 3.16. *kpsS* promoter-probe constructs. The arrows show the DNA fragments cloned into pCB192 such that the promoterless *lacZ* gene is downstream of the arrowhead. In pCB192S1 the insert spans from a *Bam*HI site in *kpsE* (Fig. 3.1) to the end shown in *kpsS*. None of the recombinant plasmids expressed ß-galactosidase activity above background levels. Abbreviation: B, *Bam*HI.

predicted products are shown in Fig. 3.17A. The amplification of a product with primers C1 and RT8 following reverse transcription with RT8 confirmed that transcription continues from *kpsC* into *kpsS* (Fig. 3.17B, Lane 5). As predicted a product spanning *kpsS* is produced when the sense strand is used as the template sequence in the reverse transcription reaction (Fig. 3.17B, Lane 1), but not from the other strand (Fig. 3.17B, Lane 3). Therefore if an antisense RNA is present in this region it does not span the gene. The absence of products amplified by reactions performed without the reverse transcription step (Fig. 3.17B, Lanes 2, 4, 6 and 8) demonstrated that there was no contaminating DNA in the RNA preparation.

To confirm that *kpsS* is expressed from the same promoter as the other region 1 genes a transcriptional terminator (*rrnB*T_1T_2, (Brosius *et al.*, 1981)) was inserted into the *Hpa*I site in the *kpsC* gene of pDS200 (Fig. 3.16), creating pDS200T. In a second construct the *kpsF* promoter was removed from pDS200 by deletion of an *Sph*I fragment spanning from *kpsU* (chapter 4, Fig. 4.1) to within the vector, creating pDS201. RNA from LE392 carrying the above plasmids and grown at 37°C was subjected to Northern analysis. A *kpsD*-specific probe was predicted to detect an approximately 8.0 kb transcript from pDS200, a 6.2 kb transcript from pDS200T and no transcript from pDS201 in which *kpsD* is deleted. However, no full length transcripts were detected from pDS200 (Fig. 3.15B, Lane 1) or pDS201 (Fig. 3.15B, Lane 3), but a range of smaller transcripts, presumably degradation products, were observed. A *kpsS*-specific probe detected transcripts only from pDS200 (Fig. 3.15A, Lane 1) and not from pDS200T (Fig. 3.15A, Lane 3) or pDS201

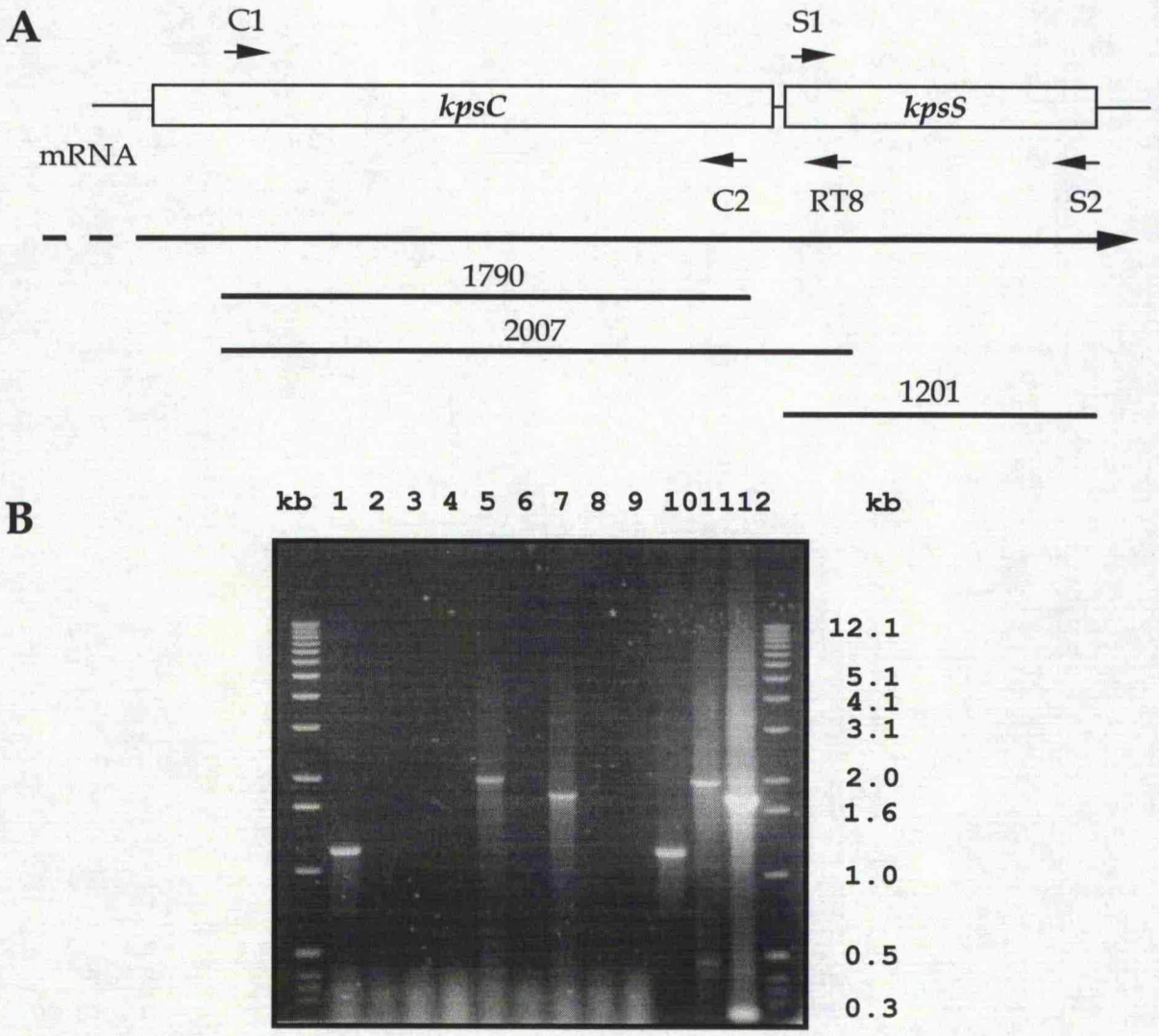


Figure 3.17. Reverse transcriptase PCR (RT-PCR) across the *kps*C-*kps*S intergenic region. A. The names and positions of the primers (arrowheads) used in RT-PCR are shown relative to the *kps*C and *kps*S genes. The long arrow indicates the predicted mRNA and horizontal lines represent the amplification products (with their lengths shown in bp). B. The following reactions were performed with 5 µg of RNA extracted from LE392/pDS200 grown at 37°C. The first primer listed was used in the reverse transcription step and the second added at the amplification step: Lane (1) S2 + S1; (3) S1 + S2; (5) RT8 + C1; (7) C2 + C1. Similar reactions were performed, omitting the reverse transcription step to demonstrate that there was no contaminating DNA in the RNA preparation: Lane (2) S2 + S1; (4) S1 + S2; (6) RT8 + C1; (8) C2 + C1. Lane (9) contained RNA only. PCR products were amplified from a DNA template using: Lane (10) S1 + S2; (11) RT8 + C1; (12) C2 + C1.

(Fig. 3.15, Lane 5). Transcription of *kpsS* is blocked by the terminator and therefore depends on a promoter upstream of *kpsC*, presumably the *kpsF* promoter. The lack of a *kpsS* transcript synthesised from pDS201 provides additional evidence that the gene does not have its own promoter.

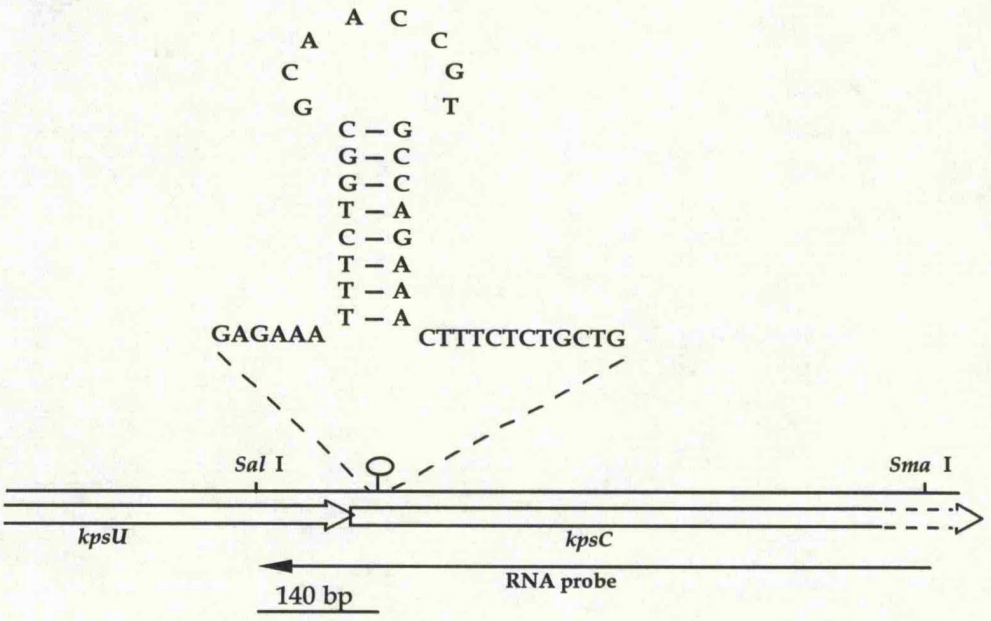
The absence of a *kpsS* promoter suggests that the 1.3 kb *kpsS*-specific transcript observed in Fig. 3.15A, Lane 1 is created by RNA processing of the larger region 1 transcript. The cleavage site specificity of the major endoribonucleases, RNase III (Court, 1993) and RNase E (Melefors *et al.*, 1993) is poorly defined. RNase III is specific for duplex RNA, but can cleave stems with unpaired bases. Although a functional site may contain as few as half of the conserved bases a 26 bp palindromic consensus sequence has been proposed (the position of the staggered cut site is indicated, N=A,U,G or C; W=A or U):

WNAGWGNNCWUNNN^NAWGNNCWCUNW

RNase E cleaves single-stranded AU-rich RNA and RAUUW (R=G or A; W=U or A) has been suggested as a consensus RNase E cleavage site (Melefors *et al.*, 1993). The region between the termination codon of *kpsC* and the second potential ATG codon of *kpsS* was searched for the above consensus sequences. The best match with the RNase III consensus contained six mismatches while two positions matched the RNase E consensus sequence. To further investigate these potential cleavage sites the predicted secondary structure of the RNA spanning the *kpsC-kpsS* intergenic region was investigated using the program MFold, which employs a free energy minimisation method (Jaeger *et al.*, 1989). No stable secondary structure was predicted which, together with the high AU content of 67%, is consistent with cleavage by RNase E (Melefors *et al.*, 1993).

3.3 *kpsC* transcriptional terminator

Near the beginning of *kpsC* is a putative stem-loop structure (ΔG=-24 kJ) which may act as a transcriptional terminator, reducing transcription of the downstream genes and thus explaining the low level of expression suggested by the difficulty encountered observing the KpsC and KpsS proteins (Pazzani *et al.*, 1993). Termination at this site could explain the transcripts of approximately 4-5 kb observed in Northern analysis (which might otherwise be formed by degradation of a longer transcript)(Fig. 3.15 and chapter 4, Fig. 4.2). If transcripts terminate at this site it is likely that the stem-loop would protect them from 3' exoribonucleases. In this case the 3' end of the truncated transcripts should be detected in an RNase protection assay. A *Sal*I-*Sma*I fragment spanning the



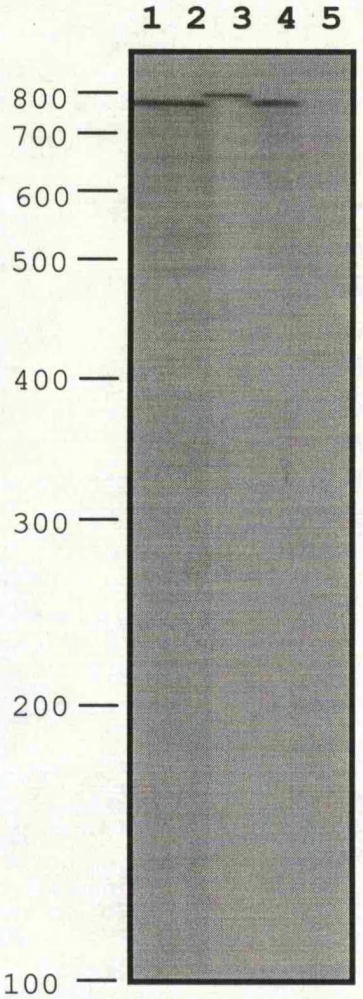


Figure 3.18. RNase protection. The potential stem-loop structure downstream of *kpsU* is shown above. *kpsU* and *kpsC* are indicated by open arrows. An 800 base radiolabelled antisense RNA probe spanning this region (arrow) was hybridised with total RNA (left) from the K5 wild type (Lane 1), LE392/pDS200 (Lane 2), LE392/pH18 (Lane 4) and LE392/pUC18 (Lane 5). Following digestion with RNase One the protected fragments and the undigested probe (Lane 3) were analysed by gel electrophoresis and their approximate sizes determined by comparison with a DNA sequence ladder.

putative stem-loop structure (Fig. 3.18) was cloned into pTZ18R. The recombinant plasmid was linearised with *Sal*I and used as a template for the synthesis of a radiolabelled antisense RNA probe. If a proportion of transcripts terminate at the stem-loop structure only 140 bases of antisense RNA molecules which hybridise to these transcripts would be protected from digestion by single-strand specific ribonuclease. However, the full length of the probe complementary to *kps* sequences (the 40 nucleotides of vector sequence at the 5' end of the probe do not hybridise and are therefore digested) was protected by RNA extracted from LE392 carrying either of several plasmids containing region 1 and from the K5 wild type (Fig. 3.18). No band of 140 bases was observed.

3.4 Discussion

Before this study region 1 of the K5 capsule gene cluster was believed to be transcribed from a promoter located upstream of *kpsE*. The results presented in this chapter demonstrate that transcription initiates from a single promoter upstream of an additional gene previously thought to be outside the *kps* cluster. The potential role of this new gene, termed *kpsF*, is analysed in chapter 5. The inclusion of *kpsF* in the region 1 transcript accounts for the observed length of 8.0 kb, approximately 1.0 kb longer than that estimated for a transcript initiating upstream of *kpsE*. The difficulty encountered in observing the region 1 transcript suggests that the full length mRNA has a short half-life. Interpretation of Northern blot results was complicated by specific hybridisation at the position of the rRNA bands (Fig. 3.2), a phenomenon noted previously (Newbury *et al.*, 1987). The high concentration of rRNA can cause masking of mRNA of the same size and cause a continuous smear of degradation products to appear as separate bands. However, by comparison of the hybridisation patterns with gene-specific probes a *kpsS* transcript was identified. The origin of this transcript is discussed below.

Region 1 is transcribed weakly, with transcripts undetectable in the wild type by Northern hybridisation. The region 1 promoter shares only three out of the six conserved nucleotides which comprise the -35 box and four out of six in the -10 box. This limited similarity with the σ^{70} consensus sequence is consistent with a low level of activity (the region 3 promoter lacks a -35 box altogether). Regulatory proteins may be required to compensate for the poor consensus and enhance transcription initiation. The conserved sequences present at the same positions upstream of the region 1 and region 3 transcript start sites (Fig. 3.13) may be binding sites for regulatory proteins. The similarity of the sequences to the consensus for IHF binding might reflect recognition by IHF itself or by a similar

protein which could bend the DNA. This could potentially provide a mechanism for the co-ordinated regulation of the conserved regions 1 and 3, possibly in response to temperature. These conserved sequences provide a target for future investigations.

From analysis of the region 1 promoter sequence the only potential known regulatory protein identified was IHF. IHF is involved in determining DNA structure and is a general regulator affecting the expression, positively or negatively, of 15 to 20% of the proteins in *E. coli* (Freundlich *et al.*, 1992)(Chuang *et al.*, 1993). The expression of various virulence factors is altered by IHF, including type 1 and F1845 fimbriae (Bilge *et al.*, 1993) and the outer membrane proteins OmpC and OmpF (Dorman and Higgins, 1987). IHF binding sites have been identified 75 bp upstream and 90 bp downstream of the promoter for *algD* the first gene in an operon required for production of the exopolysaccharide alginate by *Pseudomonas aeruginosa* (Wozniak, 1994). Mutations in each element reduced transcription. IHF binding bends DNA and is thought to activate transcription by enabling regulatory factors bound at distant sites to contact RNA polymerase by DNA loop formation. IHF mediates DNA bending upstream of σ^{54} promoters allowing upstream activators to contact bound σ^{54} (Merrick, 1993)(Morett and Segovia, 1993). When a fragment downstream of the 3' IHF site in *algD* was removed transcription was greatly reduced, suggesting that in this case IHF mediates the formation of a higher order loop structure involving a 3' enhancer element (Wozniak, 1994).

Given the putative IHF binding sites either side of the region 1 promoter one might expect IHF to affect capsular polysaccharide expression. It has been shown recently that mutations in either of the *himA* or *himD* genes, which encode the two IHF subunits, cause a five-fold reduction in expression of the KpsE protein (Simpson *et al.*, In press). This suggests strongly that IHF is required at these sites for efficient transcription of region 1. Binding of IHF could be confirmed by gel shift experiments.

IHF is however neither essential for region 1 expression; K5 capsular polysaccharide is produced in IHF mutants (of either subunit) incubated at 37°C, nor temperature regulation; no capsule is produced in IHF mutants at 18°C (Stevens *et al.*, 1994). The lack of a strong phenotype in IHF mutants is common and the protein is believed to be functionally redundant (Freundlich *et al.*, 1992). The glycosyl transferase activity encoded by *kfiC* (region 2) is reduced in an IHF mutant (Stevens, M. P., unpublished results). There are no IHF binding sites near the *kfiC* promoter but there are two overlapping sites on opposite strands of the

DNA immediately downstream of the promoter reported for *kfiA* (Petit *et al.*, 1995). If as suggested by Petit *et al* transcripts initiating at the *KfiA* promoter span *KfiC* this would enable IHF acting at the *kfiA* promoter to affect transferase activity. Alternatively a reduction in region 1 expression could have a feedback effect on transferase activity.

Upstream of the region 1 promoter the K5 sequence encodes a short polypeptide similar to the transposase proteins encoded by genes from several IS elements. Two of these IS elements are involved directly in gene expression. Insertion of IS*492* causes inactivation of extracellular polysaccharide production in *Pseudomonas atlantica* which is restored following excision (Bartlett and Silverman, 1989). The pilin invertase gene product (Piv) of *Moraxella bovis* mediates the inversion of a 2.1 kb region which brings the constant N-terminus of the pilin structural gene in frame with either the Q or I pilin remainder of the gene (Fulks *et al.*, 1990). The length of alignments between the putative K5 polypeptide and the proteins encoded by the IS elements are too short to conclude from the sequence data alone that they are homologous (share a common ancestor) (Doolittle *et al.*, 1986). However, the chance that this is the case is increased by the expectation of IS elements flanking the capsule cluster as has been demonstrated for the *H. influenzae cap* locus (Kroll *et al.*, 1991). It has been proposed that the *kps* locus was inserted into the chromosome of non-encapsulated *E. coli* by a transposition-like event (Vimr, 1991) which would require an IS element.

The *kps* locus has been mapped previously near *serA* (Ørskov *et al.*, 1976) and Vimr E. R. established its orientation and point of insertion into the *E. coli* K-12 genome by hybridisation of DNA probes flanking the K1 *kps* cluster with the Kohara library (Vimr, 1991). The sequence of the K-12 chromosome spanning this region (near 64 min) has been determined recently as part of the *E. coli* genome project (Accession no. U28377). The sequences available beyond region 3 of the cloned K1 (Accession no. U05251) and K5 (Stevens, M. P., unpublished) *kps* clusters both extend into the common K-12 sequence. This has revealed the precise site of insertion into the K-12 chromosome of the *kps* locus upstream of region 3 (Fig. 3.19). The junction is the same in K1 and K5 and probably also in K4 because probes flanking region 3 of K1 and K4 show a similar pattern of hybridisation (Drake, 1991). The various region 2 determinants have presumably been acquired since the insertion of the capsule genes at this site. The insertion is therefore very stable and one would predict that the IS element responsible has been inactivated, which is the case with the putative K5 IS element. The *kps*-specific sequences flanking region 1 are not required for capsule expression,

which is consistent with a role for this region in the acquisition of the *kps* locus rather than the regulation of its expression.

Southern blot analysis showed that the DNA for at least 3 kb upstream of *KpsE*, and therefore 2 kb upstream of *kpsF* was specific to strains which express group II capsules, including K5 (Drake, 1991). As predicted by this data the nucleotide sequence of the 1150 bp upstream of *kpsF* (Fig. 3.11) does not match that of the *E. coli* K-12 chromosome. However, although the K5 *kps* clone (pGB110) does not include K-12 sequences upstream of region 1 the K1 clone pSR23 (Silver *et al.*, 1984) extends further upstream and includes sequences which hybridised with the K-12 chromosome (Vimr, 1991). DNA probes flanking either region 1 or region 3 of the K1 *kps* cluster hybridised with the same, approximately 12 kb Kohara clone (Vimr, 1991). This indicates that if the insertion of the *kps* cluster into the chromosome resulted in the loss of chromosomal sequences this amounted to no more than several kb. Therefore the site of insertion of the region 1 end of the *kps* cluster into the K-12 chromosome must be near that of the region 3 end. To determine the extent of *kps*-specific sequences upstream of region 1 the restriction map of pSR23 (Silver *et al.*, 1988) was compared with that of the K-12 chromosome adjacent to the site of insertion of region 3 (Fig. 3.19). The relative positions of the *Bam*HI, *Hin*dIII and *Pst*I restriction sites 5' to region 1 in the K1 clone (marked with an asterisk in Fig. 3.19) correspond reasonably well with the sites predicted from the nucleotide sequence of the K-12 chromosome near the junction with region 3. The proposition that these are the same sequence is supported by the observation that this region of pSR23 directs expression of an 80 kDa protein (KpsB) in minicells (Silver *et al.*, 1984) which could correspond to SpeC (ornithine decarboxylase) which has predicted MW = 81.5 kDa (Fig. 3.19). A more detailed restriction map of the K1 clone is required to confirm whether these regions are the same and that KpsB is SpeC.

If these regions are equivalent it is possible to predict where the recombination event which introduced the region 1 end of the *kps* locus into the K-12 chromosome occurred: Probe C is specific to group II strains (Drake, 1991) and recombination must therefore have occurred just downstream of the shared *Pst*I site (Fig. 3.19, P*). This event would remove approximately 1 kb of chromosomal DNA, consistent with the flanking regions not being contiguous, as suggested by Drake, C. R. (Drake, 1991).

Analysis of the chromosome adjacent to the site of insertion upstream of region 3 revealed ORFs 178 and 286 which are homologous with proteins of the General secretory pathway (Fig. 3.19). If the suggestion above is correct, region 1 is inserted

adjacent to *speC* which encodes ornithine decarboxylase. No obvious repeated sequences were detected at the region 3 and putative region 1 boundaries. The chromosomal region between the insertion sites contains the *pheV* gene which encodes phenylalanyl-tRNA and ORF 325 which has no known homologues.

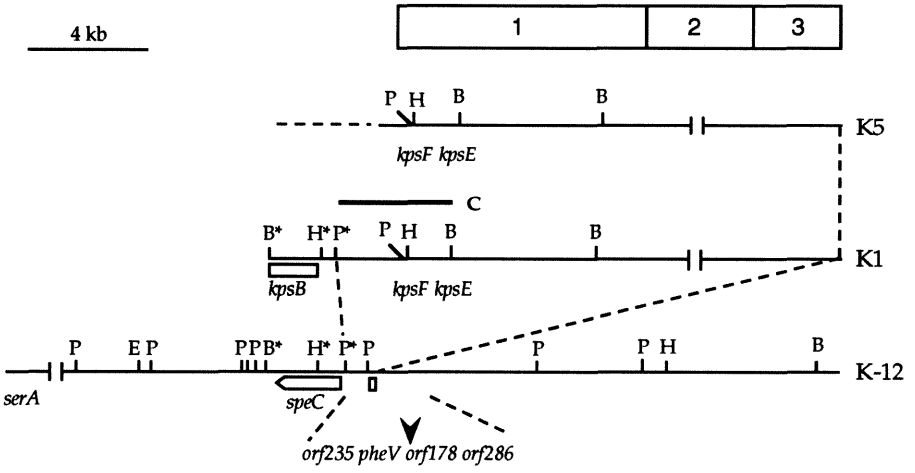


Figure 3.19. Analysis of the chromosomal location of the *kps* locus. The restriction maps of region 1 and flanking sequences of cosmid clones encoding the K1 and K5 *kps* loci are shown below boxes depicting the three functional regions of group II capsule gene clusters. Probe C from the K1 cluster is indicated by a thick horizontal line. The site of insertion into the chromosome of the region 3 ends of the K1 and K5 *kps* loci is indicated with an arrowhead. The letters thought to represent the same restriction endonuclease sites in the K-12 chromosome and the K1 clone are marked with an asterisk. Abbreviations: B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; P, *Pst*I.

The detection of a 1.3 kb *kpsS*-specific transcript suggested that factors other than transcription initiation at the promoter upstream of *kpsF* affect expression of specific region 1 genes. Given the reports of genes inserted near *kpsS* which were expressed in the opposite direction (Steenbergen and Vimr, 1990)(Esumeh, F., unpublished) an antisense regulatory RNA which interfered with transcription or created an endoribonuclease site seemed a likely explanation. Such a system exists in the F1845 fimbrial adhesin operon in which an antisense RNA anneals to the mRNA creating an endoribonucleolytic site upstream of the last gene. Cleavage at this site results in a stable mRNA encoding the fimbrial subunit gene *daaE* (Bilge *et al.*, 1993; Cantwell *et al.*, 1994). However, promoter-probe analysis precluded this possibility for the K5 *kpsS* transcript, unless the promoter was

located between the start of the *kpsS* gene and the *Bam*HI site or after the *Bcl*I site (Fig. 3.16) – this seemed unlikely as the resulting transcript would have a very short overlap with the *kpsS*-specific probe. Insertion of a transcriptional terminator in *kpsC* demonstrated that the *kpsS*-specific transcript was dependent on the upstream promoter. This suggested that the initial region 1 transcripts included *kpsS* and that a proportion were cleaved to create smaller *kpsS*-specific transcripts. There are numerous examples of bacterial polycistronic mRNAs which are processed. This is a way of controlling the stoichiometry of the gene products. For example, the operon encoding the bacterial photosynthesis genes in *Rhodobacter capsulatus* is expressed as a single transcript which is processed into segments with varying stabilities (Klug, 1993). If KpsS is translated from a monocistronic mRNA one would expect the gene to have a ribosome binding site resembling the consensus AGGAGG (McCarthy and Gualerzi, 1990) because translation could not be linked with that of upstream genes and de novo translation initiation would be required. The first potential ATG codon of *kpsS* is not preceded by a SD sequence and it is therefore likely that translation initiates at the second potential site 63 bp downstream, which does have a potential SD sequence.

The basis for the cleavage specificity of the two main endoribonucleases in *E. coli*, RNase E and RNase III, is poorly defined (Court, 1993)(Melefors *et al.*, 1993). The RNase E consensus sequence is not sufficient to explain the specificity of cleavage. Sequences upstream of the RNase E cleavage site are important for processing of 9S rRNA to yield 5S rRNA but not for the cleavage of *papBA* mRNA to yield stable *papA* mRNA (encoding the major pilin, PapA) (Nilsson *et al.*, 1996). mRNA secondary structure surrounding the cleavage site can effect the efficiency of cleavage (Melefors *et al.*, 1993). However, the presence of the tentative consensus sequence for an RNase E cleavage site in the *kpsC-kpsS* intergenic region, the lack of stable predicted RNA secondary structure and the relatively high AT content suggest that cleavage by RNase E is likely. If an antisense RNA probe spanning this region was only partially protected in an RNase protection assay this would demonstrate that processing occurred and indicate the cleavage site. Analysis of whether cleavage occurred in an *rne* (*ams*) mutant which lacked RNase E would demonstrate whether this enzyme is responsible for this event.

An antisense RNA probe spanning a putative stem-loop structure near the beginning of *kpsC* was fully protected in an RNase protection assay with RNA from the K5 wild type and LE392 carrying plasmids encoding region 1. The smaller fragment which would also have been protected if a proportion of

transcripts terminated at the stem-loop was not observed. This demonstrated that even if the putative stem-loop structure exists *in vivo* it does not cause transcriptional termination. The absence of other protected fragments indicated that there are no promoters or stable processed transcripts in this region.

The nucleotide sequence flanking region 1 has been shown previously to be conserved among group II capsule gene clusters (Drake, 1991). This study has shown that region 1 extends into this 'group II-associated DNA' to include *kpsF*. Transcription of region 1 is initiated 125 bp upstream of *kpsF* and a fragment containing a further 650 bp has promoter activity and is likely to contain regulatory elements (see chapter 4). Region 1 is transcribed from this single promoter and the initial transcript is processed to form smaller fragments.

CHAPTER 4

Regulation of the expression of region 1

4.1 Introduction

The expression of group II capsular polysaccharides is temperature regulated, with a capsule expressed at 37°C but not at 18°C (Ørskov *et al.*, 1984). This is typical of many virulence factors which show environmental regulation (Mekalanos, 1992). The expression of a capsule is correlated with high levels of CMP-KDO synthetase activity (Finke *et al.*, 1990). A structural gene, *kpsU*, for a CMP-KDO synthetase enzyme has been located in region 1 of the K5 capsule gene cluster (Pazzani *et al.*, 1993). Although not all the genes of the *kps* cluster need be temperature regulated to prevent capsule production at 18°C it is likely that region 1 is regulated, given the observations with KpsU. As a first step to investigate how this regulation is achieved it was necessary to determine whether it occurred at the level of transcription or translation. The most likely and energetically favourable point for regulation is transcription initiation; regulation of this step prevents synthesis of unnecessary mRNA. Cloning of the region 1 promoter in a promoter-probe vector (chapter 3) enabled comparison of promoter strength at capsule permissive and restrictive temperatures.

The activity of an isolated promoter on a multicopy plasmid may differ from the wild-type situation. To support results from such a system it is necessary to measure mRNA levels directly in strains carrying the cloned genes of interest or where possible in the wild type. In this study several methods involving hybridisation of specific probes with total RNA preparations were employed. If expression is regulated at the level of transcription and mRNA is not produced at the capsule-restrictive temperature one would expect no proteins to be produced under these conditions. If expression is regulated at the level of translation no proteins may be produced despite the presence of mRNA at the capsule-restrictive temperature. A situation between these extremes is possible and certain proteins may be post-translationally regulated, being present at 18°C but inactive. Specific antibodies can be used to investigate the presence of proteins in extracts from cultures grown at each temperature and confirm or add to conclusions drawn from RNA studies. To this end antibodies were produced against the KpsD protein (in collaboration with Peter Hänfling, Max-Planck-Institut für Immunobiologie, Freiberg, Germany) and KpsE-specific antibodies became available during the course of the project.

Virulence genes are organised in regulatory networks, enabling their co-ordinated expression in response to environmental stimuli (Mekalanos, 1992). It is therefore likely that some of the genes which control expression of other virulence genes may also regulate capsule expression. The high temperature encountered in the body compared with ambient temperature activates the expression of various virulence factors including group II capsular polysaccharides. The nucleoid protein H-NS is required for temperature regulation of genes involved in cell invasion by *Shigella flexneri* and Pap pilus expression in *E. coli* (Dorman, 1994). Regulation is usually achieved at the level of transcription and H-NS causes transcriptional silencing. The effect on the expression of region 1 of mutations in *hns* and other genes implicated in the temperature regulation of virulence factors was investigated.

4.2 Temperature regulation

4.2.1 Promoter-probe analysis

A DNA fragment including the region 1 promoter has been cloned in pCB192 upstream of a promoterless *lacZ* gene creating pCB192HcB (chapter 3, section 3.1.3). ß-galactosidase activity expressed from pCB192 and pCB192HcB was assayed at 37°C and 18°C. A *lacZ*$^-$ derivative of MS101 was used to provide any regulatory factors encoded specifically by the *kps* gene cluster. The results, shown in table 4.1, indicated that the region 1 promoter was temperature regulated, with the activity at 18°C reduced to 45% of that at 37°C.

Table 4.1. Promoter activity at 37°C and 18°C.

Plasmid	ß-galactosidase activity (Miller units) (SD)	
	37°C	18°C
pCB192HcB	166.9 (8.0)	75.3 (1.8)
pCB192	16.3 (1.9)	12.9 (2.0)

4.2.2 RNA analysis

All the specific functions required for K5 capsule production are encoded by pGB110, a large cosmid clone carrying the whole *kps* cluster (Roberts *et al.*, 1986). Cultures of LE392(pGB110) were grown at 37°C and 18°C and the presence of a capsule at 37°C demonstrated by sensitivity to K5-specific bacteriophage. In dot blots a radiolabelled 5.3 kb *Bam*HI fragment specific for region 1 (Fig. 4.1)

hybridised with RNA extracted from LE392(pGB110) grown at 37°C but not with that from 18°C (Fig. 4.2A). In Northern blots an approximately 8.0 kb transcript and a range of degradation products were detected in extracts from 37°C, while almost no region 1 mRNA was detected at 18°C (chapter 3, Fig. 3.2). This supports the conclusion from promoter-probe data that region 1 is temperature regulated primarily at the level of transcription. Surprisingly the production of region 1 mRNA from two constructs lacking the *kpsF* promoter, pH18 and pPC6 (Pazzani, 1992)(Fig. 4.1) was also temperature regulated (Fig. 4.2B and C).

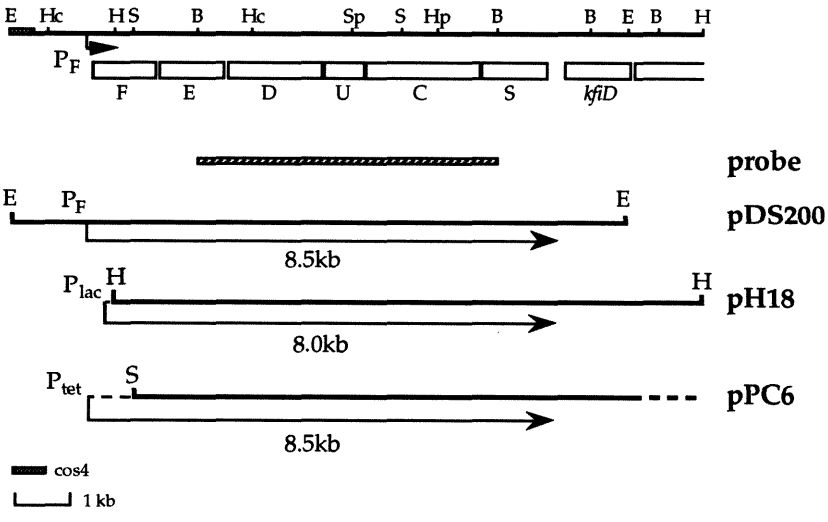


Figure 4.1. Region 1 transcripts predicted from various recombinant plasmids. The genes of region 1, *kpsF, E, D, U, C* and *S* are indicated below a simplified restriction map. The hatched box indicates the DNA probe used in Northern analysis. Horizontal lines represent the fragments cloned into pKK223-3 to make pDS200 (this study) or into pUC18 to make pH18 (Pazzani, 1992). pPC6 comprises a DNA fragment spanning from a *Sma*I site in *kpsF* to a site beyond region 3 cloned in pACYC184 (Pazzani, 1992). The promoter now known to be located upstream of *kpsF* is marked (P_F) and the predicted transcript indicated with an arrow. When pH18 and pPC6 were used to analyse transcription it was believed that they contained the region 1 promoter. This is not the case and the probable vector promoters responsible for their expression are marked. Abbreviations: B, *Bam*HI; E, *Eco*RI; Hc, *Hinc*II; H, *Hin*dIII; Hp, *Hpa*I (site of insertion of *rrn*T_1T_2 in pDS200T); S, *Sma*I; Sp, *Sph*I (sequences upstream of this site deleted in pDS201).

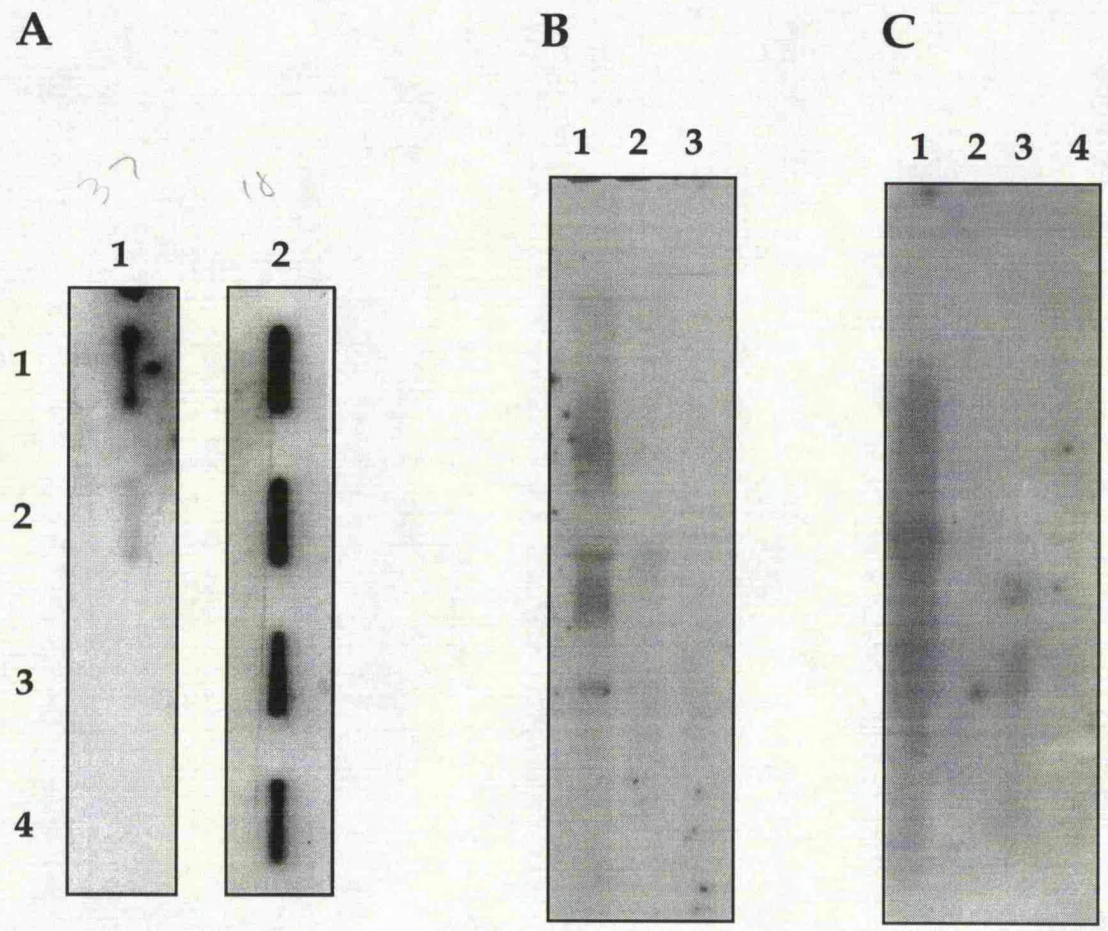


Figure 4.2. A. Dot blot. RNA samples (40 μg) extracted from LE392/pGB110 incubated at 37°C (row 1) or 18°C (row 2) and LE392 at 37°C (row 3) or 18°C (row 4) were fixed to a nylon filter. The filter was hybridised with a radiolabelled 5.3 kb *Bam*HI DNA fragment specific for region 1 (Fig. 4.1)(column 1), stripped and reprobed with a rRNA-specific probe (column 2), to confirm equal loading. B. Northern blot. RNA samples extracted from LE392/pH18 (Fig. 4.1) incubated at 37°C (Lane 1) or 18°C (Lane 2) and LE392/pUC18 at 37°C (Lane 3) were subjected to Northern analysis with the region 1 probe described above. C. Northern blot. RNA samples extracted from LE392/pPC6 (Fig. 4.1) incubated at 37°C (Lane 1) or 18°C (Lane 2) and LE392/pACYC184 at 37°C (Lane 3) or 18°C (Lane 4) were subjected to Northern analysis with a probe specific for *kpsD*.

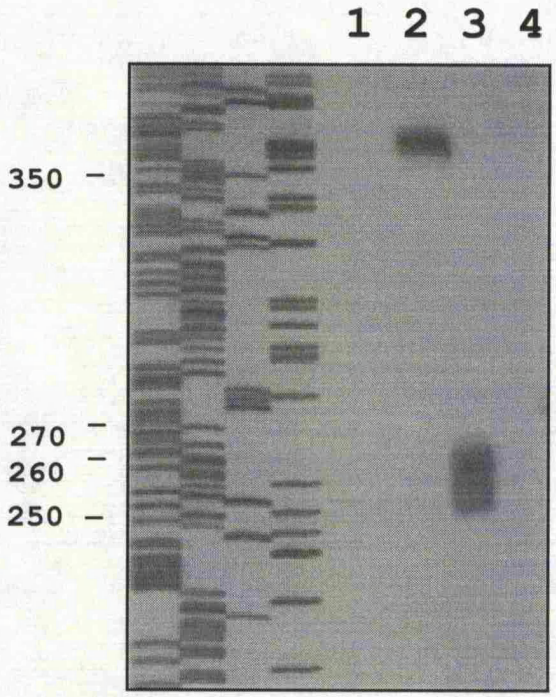


Figure 4.3. RNase protection. Radiolabelled antisense RNA probe 3 hybridised with RNA extracted from LE392 (pKK223-3) (Lane 1) and the K5 wild type grown at 37°C (Lane 3) and 18°C (Lane 4). The full length undigested probe is in Lane 2. The length of the DNA sequencing ladder is indicated in nucleotides.

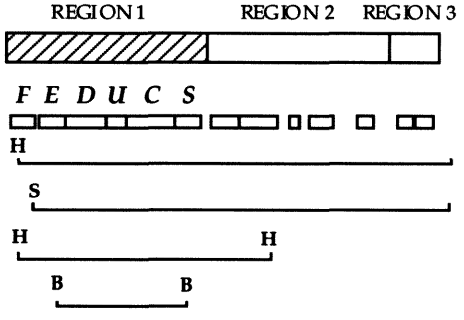
The levels of mRNA produced by the region 1 promoter at 37°C and 18°C were also investigated by primer extension (chapter 3, Fig. 3.8) and RNase protection (Fig. 4.3). Both methods confirmed the greatly reduced level of mRNA observed at 18°C by Northern blot analysis. The RNase protection method is particularly sensitive because the probe is radiolabelled throughout its length and being single-stranded its availability is not reduced by re-annealing with a complementary strand. It was therefore possible to detect the low levels of mRNA synthesised by the K5 wild type. Transcripts were detected at 37°C (Fig. 4.3 Lane 3) but not at 18°C (Lane 4), demonstrating that transcriptional regulation occurs *in vivo* .

4.2.3 Regulation of region 1 in *kps* subclones

Subclones of the *kps* cluster, for example pPC2 and pPC6 (Figs. 4.1 and 4.4) which continue to direct capsule production despite lacking the region 1 promoter have been generated previously (Pazzani, 1992). The expression of surface polysaccharide by pPC6 is temperature regulated and analysis of mRNA (Fig. 4.2) and protein (section 4.2.5) levels showed that region 1 is temperature regulated. Similarly, expression of region 1 from pH18, which carries a *Hin*dIII fragment spanning from within *kpsF* to region 2 (Figs. 4.1 and 4.4), is temperature regulated according to CMP-KDO synthetase activity (Pazzani, 1992) and mRNA levels (Fig. 4.2). This could be due to elements which prevent transcription or destabilise the mRNA at 18°C. Expression of CMP-KDO synthetase activity from pBA6 (Fig. 4.4), which lacks more of the 5' end of region 1 and possibly these elements, is not temperature regulated (Pazzani, 1992). In all these constructs region 1 is expressed from vector promoters (Fig. 4.1) and the difference in regulation patterns may therefore be due to the use of different vectors and the insertion of region 1 at varying sites, rather than a property of the specific DNA inserted. To test whether vector promoter activity is temperature regulated a *lacZ*-kanamycin-resistance cassette was excised from pKOK6 (Kokotek and Lotz, 1989) and inserted into the appropriate vector at the same site as the *kps* DNA. In pPC6 the region 1 end of the insert is at an *Xba*I site in pACYC184 and no appropriate site is available at the end of the pKOK6 *lacZ* cassette. The vector pCB192 (chapter 3, Fig. 3.3) was therefore used as a source of a promoterless *lacZ* gene – it was linearised with *XbaI* and cloned into pACYC184, also cleaved with *Xba*I, such that the *lacZ* gene was located in the same position as region 1 in pPC6.

The K5 subclones and their equivalent *lacZ* reporter gene constructs, plac2-4, are shown in Fig. 4.4. The constructs were transformed into the *lacZ* $^-$ derivative of

LE392, LE392.23 and cultures grown at 37°C and 18°C. The ß-galactosidase activity produced from plac4 was much higher at 37°C than at 18°C (Fig. 4.5), suggesting that the temperature regulation of region 1 expression observed in pH18 could be due to regulation of the vector promoter. However, the ß-galactosidase activity of plac2 was the same at both temperatures (Fig. 4.5), suggesting that the absence of temperature regulation in pBA6 is not due to the lack of specific *kps* sequences but because the activity of the vector promoter directing expression of region 1 is not regulated. No ß-galactosidase activity was observed from plac3.



Construct	Temp. regulated	Vector	*lacZ* construct
pPC2	Yes	pACYC	—
pPC6	Yes	pACYC	plac3
pH18	Yes	pUC18	plac4
pBA6	No	pACYC	plac2

Figure 4.4. Analysis of vector transcription patterns. A: The three regions of the K5 *kps* cluster are indicated and the region 1 *kpsF, E, D, U, C* and *S* genes labelled. Horizontal lines represent the DNA fragments carried by pPC2, pPC6, pH18 and pBA6, and the table shows the vector into which they are cloned, whether their expression is temperature regulated and the corresponding *lacZ* reporter constructs. Abbreviations: B, *Bam*HI; H, *Hin*dIII; S, *Sma*I.

4.2.4 Production of KpsD antibody

To confirm the prediction from RNA analyses that the region 1 proteins are not synthesised at 18°C a KpsD-specific antibody was produced. The Glutathione S-transferase (GST) gene fusion system was chosen for the expression and purification of KpsD. In this system the gene of interest is cloned into the vector pGEX-2T (Pharmacia Biotech, Fig. 4.6) to create a translational fusion with the C-terminus of the GST protein. Expression of the fusion protein is under the control of P_{tac} and may be induced with IPTG. The *lacI*^{q} gene represses expression of the fusion protein from P_{tac} in the absence of IPTG. The over-expressed fusion protein can be purified using its affinity for immobilised glutathione (Smith and Johnson, 1988). A thrombin cleavage site between the GST section and the

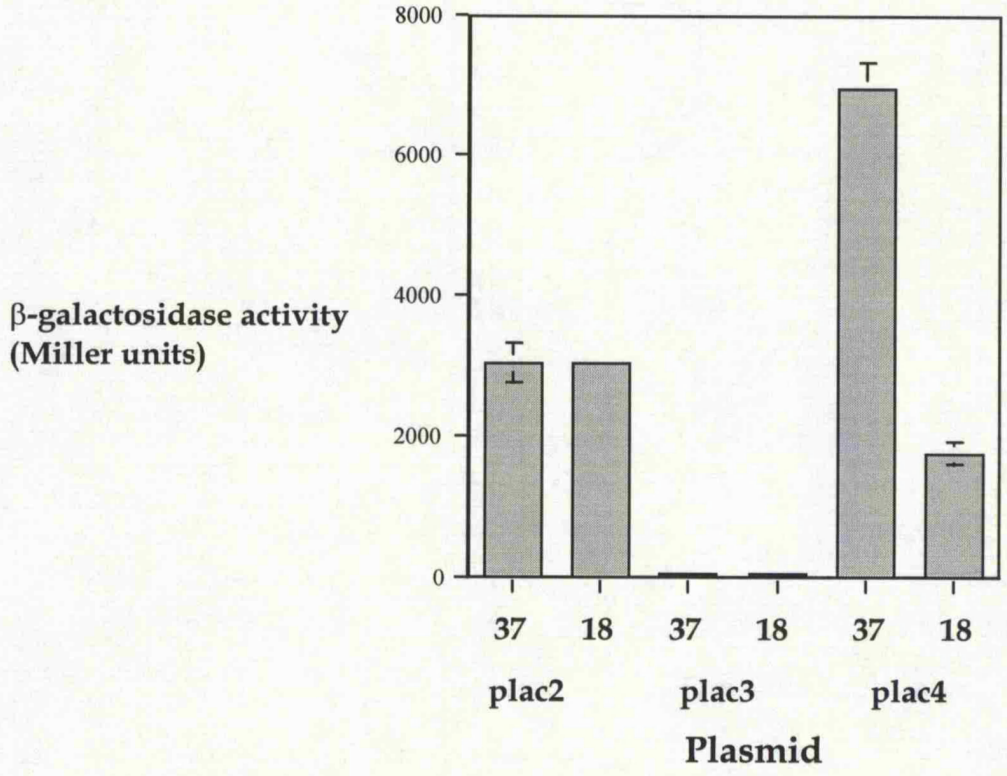


Figure 4.5. Vector transcriptional activity. The ß-galactosidase activity of the plac constructs described above (Fig. 4.4) was measured in strain LE392.23 grown at 37°C and 18°C. The readings for each construct at 37°C (left) and 18°C (right) are indicated. Each value is the average of four readings and the error bars show the standard deviation.

remainder of the fusion protein enables removal of GST by digestion with thrombin (Fig. 4.8).

A 1617 bp fragment encoding KpsD, but lacking the N-terminal signal sequence was amplified by PCR using the primers: 5'-ACCATTGATATTAACGCTGAC-3' and 5'-TTACAAAGACAGAATCACTTTTG-3'. pGEX-2T was cleaved with *Sma*I and single thymidine residues added to the 3' ends. This facilitated cloning of the PCR product which had 3' terminal adenosine residues added by *Taq* polymerase during PCR. The orientation of insertion was determined by restriction enzyme digests and clones chosen which generated an in frame fusion between the C-terminus of GST and the KpsD protein (Fig. 4.6). The recombinant clones were transformed into *E. coli* SURE™ and production of the fusion protein induced by addition of IPTG. Whole cell lysates were prepared and analysed by SDS-PAGE. Induction of pGEX-2T with IPTG resulted in production of the 26 kDa GST protein (Fig. 4.7, Lane 1) which was not observed in uninduced cultures (Fig. 4.7, Lane 2). In the recombinant clone pDS2 a protein of approximately 87 kDa, the size predicted for the GST-KpsD fusion protein, was greatly overexpressed following IPTG induction and GST was absent (Fig. 4.7, Lane 3).

The KpsD protein was purified and antibodies prepared by Peter Hänfling (Max-Planck-Institut für Immunobiologie, Freiberg, Germany). The fusion protein was affinity purified with glutathione, cleaved with thrombin and the GST removed (Fig. 4.8). The purified KpsD protein was used for immunisation and the resulting serum ProteinA-Sepharose purified.

4.2.5 Expression of the KpsD protein at 37°C and 18°C

The low level of region 1 mRNA observed at 18°C (section 4.2.2) suggested that region 1 proteins were not made at this temperature. To confirm this LE392 carrying pDS200 (Fig. 4.1), which directs temperature-regulated transcription of region 1, was grown at 37°C and 18°C and lysates prepared. Western blots with anti-KpsD antibody confirmed that KpsD was synthesised at 37°C (Fig. 4.9A, Lane 2) but not at 18°C (Fig. 4.9A, Lane 3). The level of KpsD produced by pDS200 (Fig. 4.9B, Lane 1) was higher than that from pGB110 (Fig. 4.9B, Lane 2) which was more than in the K5 wild type (Fig. 4.9B, Lane 4). No KpsD was detected from the K5 wild type grown at 18°C (Fig. 4.9B, Lane 5), supporting the results obtained from the plasmid models. The amount of KpsD produced by pPC6 was low (Fig. 4.9B, Lane 6), comparable to that in the wild type, but was also detectable at 18°C (Fig. 4.9B, Lane 7). Even less KpsD was produced by LE392(pPC2) and was barely detectable in the western blot (Fig. 4.9B, Lane 8).

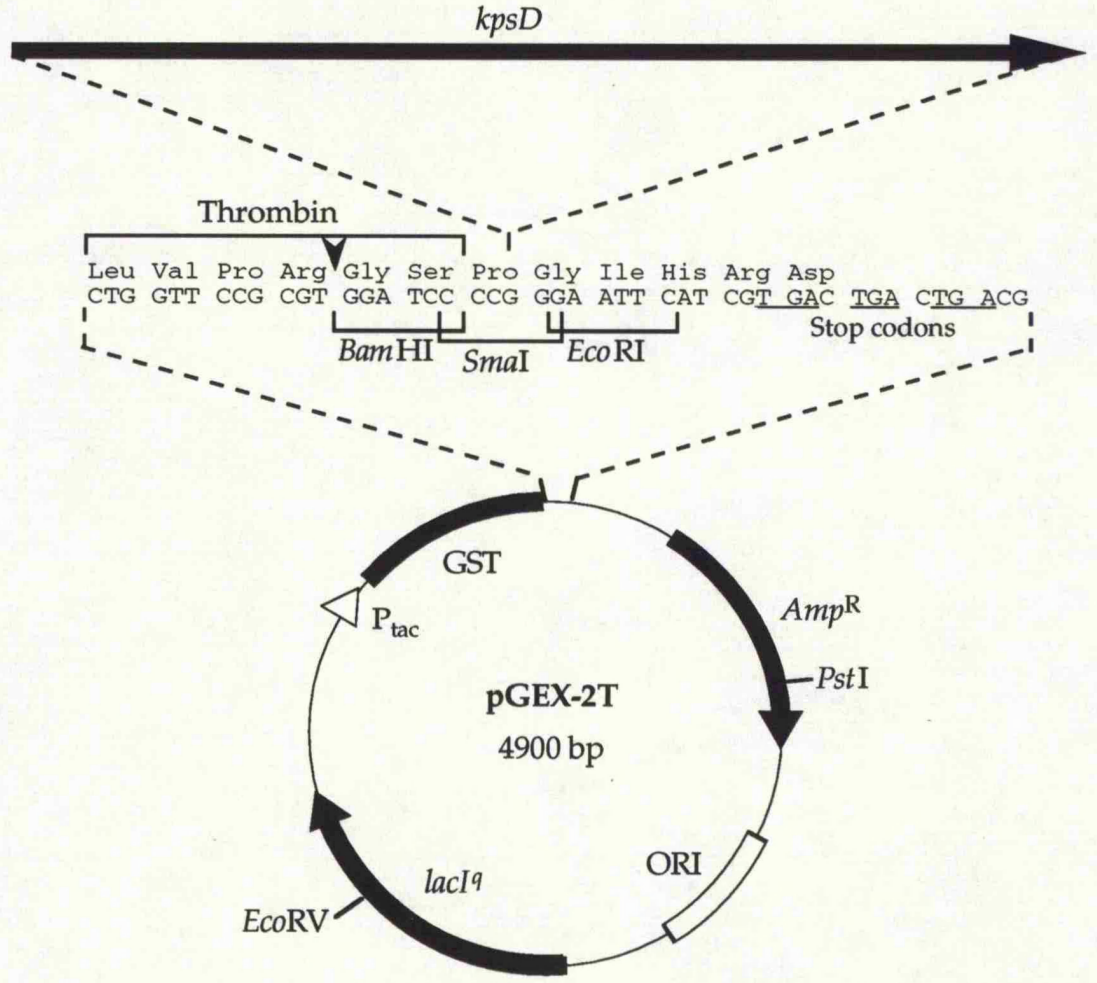


Figure 4.6. The Glutathione S-transferase (GST) gene fusion system. The horizontal arrow represents the PCR fragment encoding KpsD which was cloned into the *Sma*I site of pGEX-2T (Pharmacia Biotech) to generate pDS2. The nucleotide sequence surrounding the cloning site is shown and the position of the thrombin cleavage site in the amino acid sequence encoded by this region indicated by an arrowhead.

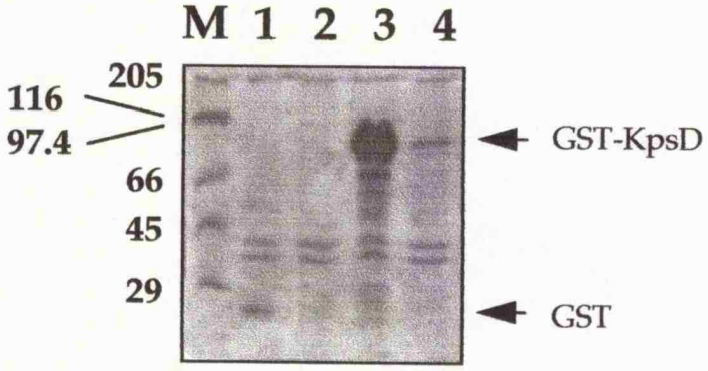


Figure 4.7. SDS-PAGE gel demonstrating over-expression of the GST-KpsD fusion protein. The 26 kDa C-terminal GST polypeptide is over-expressed by SURE™/ pGEX-2T after (Lane 1) but not before IPTG induction (Lane 2). Clone pDS2 (Fig. 4.6) expresses an approximately 87 kDa fusion protein weakly without IPTG induction (Lane 4) and very strongly after induction (Lane 3). The molecular weight of the markers (M) is shown in kDa.

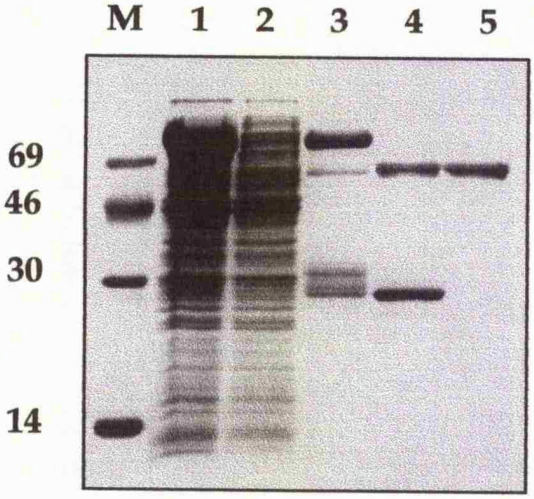


Figure 4.8. Purification of the KpsD protein. SURE™/pDS2 after (Lane 1) and before (Lane 2) IPTG induction. The fusion protein was affinity purified (Lane 3) and cleaved with thrombin (Lane 4). The purified KpsD protein (Lane 5) was used for immunisation. The size of molecular weight markers is indicated in kDa.

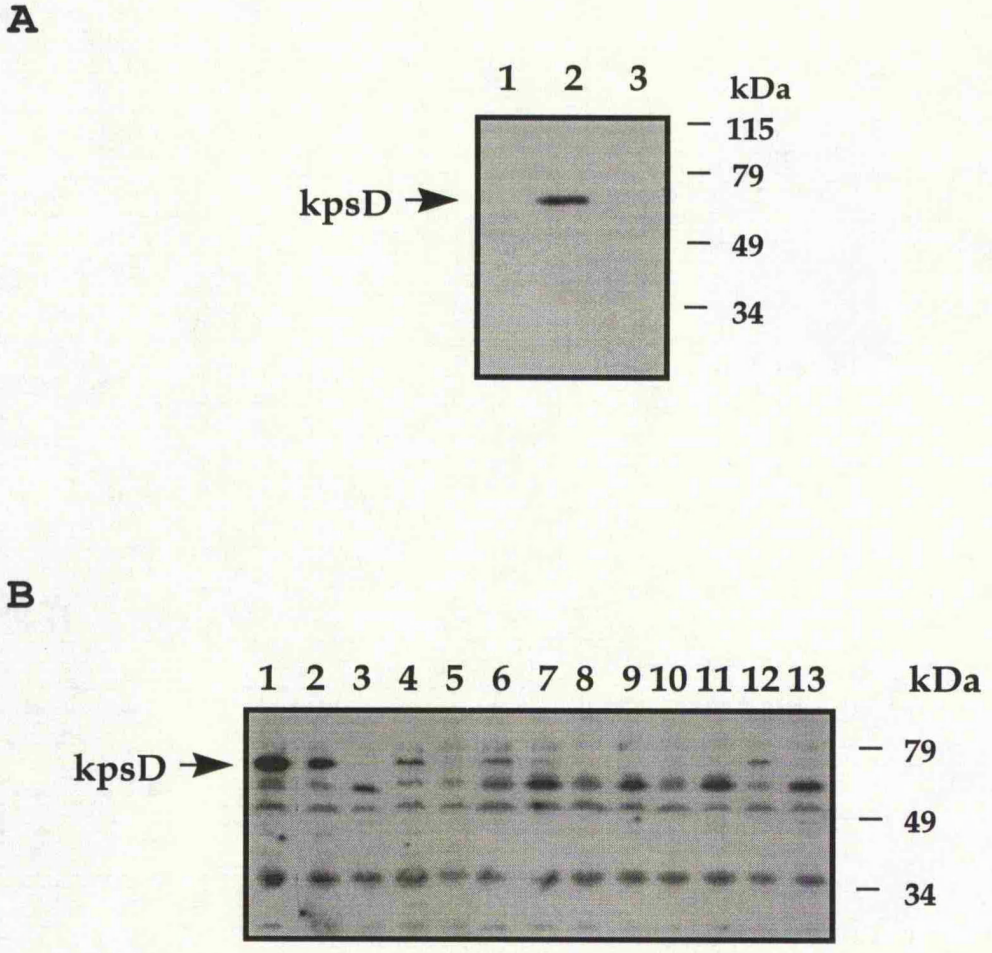


Figure 4.9. Western blots with anti-KpsD antibody. A. Whole cell lysates from the following cultures were probed with antibody against KpsD: LE392/pKK223-3, 37°C (Lane 1); LE392/pDS200, 37°C (Lane 2) and LE392/pDS200, 18°C (Lane 3). Equal loading was confirmed by Coomassie blue staining of an identical SDS-PAGE gel. The molecular mass of markers is indicated in kDa. B. As above, with: LE392/pDS200, 37C° (Lane 1); LE392/pGB110, 37°C (Lane 2) and 18°C (Lane 3); K5 wild type, 37°C (Lane 4) and 18°C (Lane 5); LE392/pPC6, 37°C (Lane 6) and 18°C (Lane 7); LE392/pPC2, 37°C (Lane 8) and 18°C (Lane 9); and LE392/pACYC, 37°C (Lane 10) and 18°C (Lane 11); L32706/pPC6, 37°C (Lane 12); L32706/pPC6, 18°C (Lane 13).

4.3 Involvement of known regulatory proteins

4.3.1 Construction of MS101*hha*::Tn*5*

The Hha protein of *E. coli* is a small 8.6 kDa polypeptide which is 82% identical with the YmoA protein of *Yersinia enterocolitica*. YmoA influences DNA topology and regulates virulence genes in response to temperature, for example derepressing transcription of *yadA* and *yop* genes at low temperatures (De la Cruz *et al.*, 1992). Hha and YmoA are functionally interchangeable and are members of a new class of histone-like regulatory proteins (Mikulskis and Cornelis, 1994). The *hha* gene was identified by a mutation that increased the production of plasmid-encoded haemolysin in *Escherichia coli* (Godessart *et al.*, 1988)(Nieto *et al.*, 1991). In the absence of the upstream enhancer element *hlyR* the Hha protein interacts with *hlyM* to silence the haemolysin promoter (Jubete *et al.*, 1995). The recombinant plasmid pANN202-312 encodes the haemolysin genes but lacks *hlyR* and therefore shows poor haemolysin expression. However, in *hha* cells pANN202-312 directs high level haemolysin expression comparable to that from plasmids containing *hlyR* (Carmona *et al.*, 1993). Haemolysin expression can therefore be used to assay for the presence of a *hha* mutation. The potential role of Hha in temperature regulation of virulence factors and its involvement in the expression of haemolysin, a system regulated by the RfaH protein which is also required for K5 capsular polysaccharide production (Stevens *et al.*, 1994), suggest a possible role in regulation of the *kps* cluster.

The K5 *kps* cluster has been introduced onto the chromosome of K-12 strain PA360 to create the $K5^+$ strain MS101, which is more amenable to genetic manipulation than the K5 wild type (Stevens *et al.*, 1994). To study the effect of Hha on expression of region 1 a *hha* mutation was introduced into MS101. Strain Hha3 has a Tn*5* insertion in the *hha* gene and expresses large amounts of haemolysin when carrying plasmid pANN202-312, as demonstrated by formation of a large haemolysis halo around colonies grown on blood agar (Godessart *et al.*, 1988). While various Tn*5* insertions in the *hha* gene exhibit a high frequency of excision and secondary transposition Hha3 is a stable mutant (Badenas *et al.*, 1994). The *hha*::Tn*5* allele was transferred from Hha3 to MS101 (carrying pANN202-312) by P1 transduction and selection for the kanamycin resistance encoded by Tn*5* (Fig. 4.10). Clones in which the wild-type chromosomal *hha* allele was replaced by *hha*::Tn*5* by homologous recombination were detected by their increased zone of haemolysis on blood agar. The genotype of several putative MS101*hha*::Tn*5* clones was confirmed by

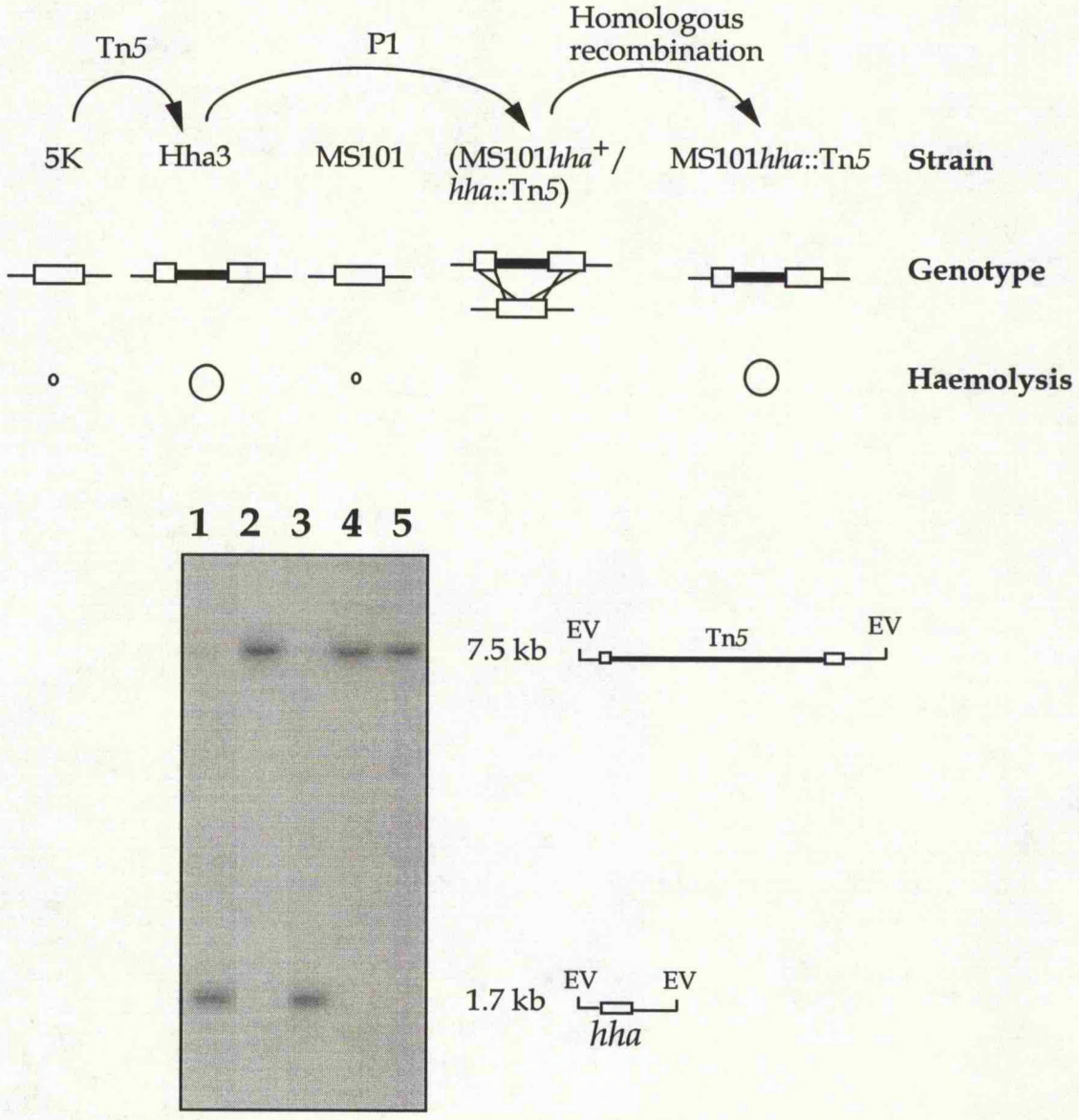


Figure 4.10. Construction of MS101*hha*::Tn*5*. The strains and procedures involved are shown at the top of the figure. Open boxes depict the *hha* gene and thick black lines Tn*5*. A large circle indicates a large zone of haemolysis when carrying pANN202-312 (a characteristic Hha$^-$ phenotype) and a small circle a narrow zone. In the Southern blot a *hha*-specific probe detects either a 1.7 kb *Eco*RV fragment spanning the *hha* gene or a 7.5 kb fragment containing the 5.8 kb Tn*5* insertion. The wild-type *hha* fragment is detected in 5K (Lane 1) and MS101 (Lane 3) while the longer, interrupted fragment is present in Hha3 (Lane 2) and two separate MS101*hha*::Tn*5* transductants (Lanes 4 and 5). Abbreviation: *Eco*RV, EV.

Southern blotting (Fig. 4.10). Chromosomal DNA was digested with *Eco*RV and hybridised with a radiolabelled 1.4 kb *hha*-specific *Bgl*II-*Eco*RV DNA probe prepared from pUBM22 (Nieto *et al.*, 1991). The predicted 1.7 kb fragment spanning the *hha* gene was observed in 5K and MS101, while this increased to 7.5 kb with the insertion of Tn*5* in Hha3 and MS101*hha*::Tn*5* (Fig. 4.10).

4.3.2 Analysis of mutant phenotypes

Mutations in genes which regulate the expression of virulence factors have previously been introduced into MS101 and their effect on capsule production studied (Stevens *et al.*, 1994). MS102 has a mutation in the *hns* gene which encodes the histone-like protein H-NS, a pleiotropic regulator which affects DNA topology (Hulton *et al.*, 1990) and has been implicated in the temperature regulation of virulence gene expression (Porter and Dorman, 1994). MS106 has a mutation in the *rfaH* gene which encodes a regulator of several virulence genes (Pradel and Schnaitmann, 1991)(Bailey *et al.*, 1992). MS113 has a mutation in the gene encoding RimJ, which is involved in thermoregulation of the *pap* operon (White-Ziegler *et al.*, 1990). The *pap* pilin gene is normally transcribed at 37°C and not at 23°C but in the absence of RimJ, the N-terminal acetylase of ribosomal protein S5, transcription at 23°C is restored to levels similar to those at 37°C (White-Ziegler and Low, 1992).

The mutant strains MS102 (MS101*hns*::Tn*10*) and MS113 (MS101*tcp1*::mTn*10*) were shown to be sensitive to K5-specific bacteriophage when grown at 37°C and resistant at 18°C, indicating temperature regulation of capsule expression (Stevens, M. P., unpublished) (Stevens *et al.*, 1994). MS101*hha*::Tn*5* has the same pattern of K5 bacteriophage susceptibility showing that the Hha protein is neither required for expression at 37°C nor derepresses all the necessary genes at 18°C.

Although capsule production is temperature regulated in these mutants, implying that region 1 products are present at 37°C, their concentrations may be altered and expression at 18°C induced. Antibodies were used to determine the levels of region 1 proteins in lysates prepared from cultures of the mutant strains grown at 37°C and 18°C. The anti-KpsD antibody detected KpsD only weakly in lysates of MS101 from 37°C. This suggests that less region 1 proteins are synthesised in MS101 than in strains carrying multicopy plasmids such as pDS200. Although KpsD appeared to be temperature regulated in MS101 and the regulatory mutants the positive signal at 37°C was very weak (data not shown).

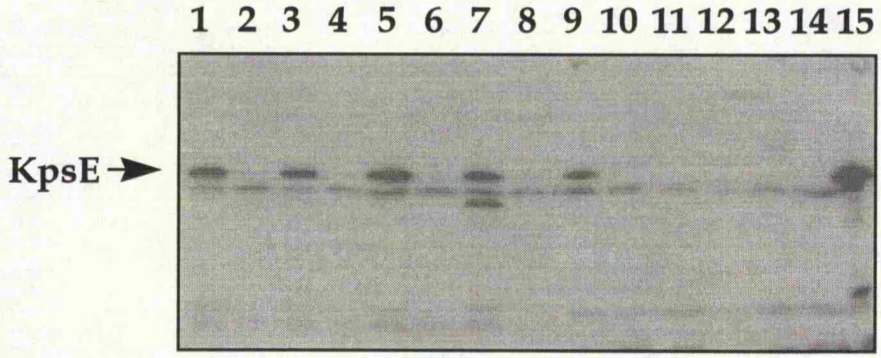


Figure 4.11. Western blot with anti-KpsE antibody. Whole cell lysates from the following cultures were probed with antibody against KpsE: MS101, 37°C (Lane 1) and 18°C (Lane 2); MS106, 37°C (Lane 3) and 18°C (Lane 4); MS102, 37°C (Lane 5) and 18°C (Lane 6); MS113, 37°C (Lane 7) and 18°C (Lane 8); MS101*hha*::Tn5, 37°C (Lane 9) and 18°C (Lane 10); PA360, 37°C (Lane 11) and 18°C (Lane 12); Hha3, 37°C (Lane 13) and 18°C (Lane 14); LE392/pDS200, 37°C (Lane 15).

The western blots were therefore probed with the more specific and higher titre anti-KpsE antibody. KpsE is present in the regulatory mutants at 37°C at about the same level as in MS101 and is absent at 18°C (Fig. 4.11). This suggests that none of these proteins are involved in the regulation of region 1.

The resistance of MS106 (MS101*rfaH*::Tn*5*) to K5-specific bacteriophage at both 37°C and 18°C demonstrated that RfaH is required for K5 capsule production (Stevens *et al.*, 1994). However, KpsE was produced by MS106 at 37°C at a level comparable to that in MS101 (Fig. 4.11, Lane 3) suggesting that RfaH does not affect the expression of region 1.

4.5 Discussion

The results presented in this chapter demonstrate that the temperature regulation of region 1 of the K5 capsule gene cluster is achieved by controlling the level of mRNA. By Northern blot analysis, RNase protection assays and primer extension region 1 mRNA was shown to be absent at 18°C, or at least greatly reduced compared with 37°C. This reduction could be due to reduced promoter activity or mRNA stability. The expression of all the genes in region 1 is dependent upon a single promoter upstream of the first gene, *kpsF* (chapter 3). Transcription initiation would therefore provide an efficient point for regulation and prevent the synthesis of unnecessary transcripts. Promoter-probe analysis suggested that the activity of this promoter is temperature regulated. However, the activity at 18°C remained just under half that at 37°C in contrast with the extremely low levels of region 1 mRNA at 18°C relative to 37°C detected by RNA analysis. The relatively high activity of the isolated promoter at 18°C may be an effect of being in a multicopy plasmid, with the many copies of an operator site perhaps titrating out a repressor. Alternatively the DNA fragment analysed may lack regulatory elements.

Promoter-probe analysis suggests that transcription initiation is the main point of regulation, but mRNA stability may also affect expression of region 1. For example, the mRNA might be destabilized at 18°C by sequences which assume conformations susceptible to endoribonuclease cleavage. A transcriptional terminator active at 18°C might result in incomplete transcripts which were rapidly degraded. In the extreme case, if sequences at the 5' end of the region 1 mRNA destabilized the whole mRNA at 18°C, these could be present in pCB192HcB and be responsible for conferring temperature regulation upon expression of ß-galactosidase activity rather than changes in transcription

initiation. Elements within region 1 which modulate the mRNA stability in response to temperature could explain how the expression of region 1 from recombinant plasmids pPC6 and pH18 (Fig. 4.2) is temperature-regulated despite their lacking the region 1 promoter. The 71 bp between *kpsF* and *kpsE* provide a possible regulatory site; expression of region 1 from pBA6 which lacks this region is not temperature regulated. An alternative explanation is that the vector promoters responsible for transcription of region 1 in pPC6 and pH18 (but not in pBA6) are themselves temperature regulated. Measurement of transcriptional activity in the vectors at 37°C and 18°C, at the point were the K5 DNA was inserted, using *lacZ* as a reporter gene showed that this was the case, at least with pH18. Transcription initiation therefore remains the most likely point for regulation of region 1.

The conclusions from the study of vector promoter activity must however be drawn cautiously. Insertion of pCB192 into the *Bam*HI site of pACYC184 such that the *lacZ* gene lay downstream of the vector promoter resulted in very low ß-galactosidase activity (data not shown) compared with the high level observed when the *lacZ* gene alone was inserted creating plac2. Insertion of a whole plasmid into pACYC may disrupt transcriptional patterns or result in an unstable hybrid plasmid. The activity expressed from a single inserted gene is more likely to reflect the original transcriptional pattern at the point of insertion. Results obtained using the whole of pCB192 as a source of a reporter gene should therefore be treated with extreme caution. The lack of ß-galactosidase activity expressed by plac3 suggested that there was no transcription through the *Xba*I site in pACYC184 even at 37°C. Region 1 is inserted at this site in pPC6 and although the level of region 1 proteins expressed from pPC6 is low (Fig. 4.9B) one would expect to detect some transcription. The insertion of pCB192 to generate plac3 may have disrupted the promoter in pACYC184 or the insertion of the *kps* gene cluster to generate pPC6 may have induced a promoter not otherwise active in pACYC184. Construction of a *lacZ* gene flanked by the appropriate restriction enzyme sites and its insertion into the *Xba*I site of pACYC might provide a more accurate measure of transcription at this point and demonstrate whether temperature regulation of region 1 expression in pPC6 is due to modulation of the vector promoter.

Western analysis of the expression of the region 1 proteins KpsE and KpsD confirmed the prediction from RNA analysis that are not expressed at 18°C. Regulatory proteins are presumably required to either repress expression at 18°C or activate expression at 37°C. However, the introduction of mutations in the genes encoding various regulatory proteins had no effect on the levels of KpsE or

KpsD. The RfaH protein, which is required for capsule production, was shown not to affect expression of region 1 and must therefore exert its effect solely on regions 2 and 3. To identify genes whose products do affect the expression of region 1 it should be possible to isolate mutations which alter the expression of a reporter gene inserted in this region (see chapter 6). The effect of the mutations could be confirmed by western blot analysis to measure the levels of KpsE and KpsD expressed in the mutant strains.

CHAPTER 5

Analysis of KpsF

5.1 Introduction

The nucleotide sequence of region 1 of the K5 antigen gene cluster reported previously encoded the C-terminus of an ORF upstream of *kpsE* (Pazzani *et al.*, 1993). Nucleotide sequence determined in this study (chapter 3, Fig. 3.11) revealed the entire ORF which is almost identical with a gene located in the same position in the K1 *kps* cluster (Cieslewicz *et al.*, 1993). In keeping with the nomenclature of Silver *et al* (Silver *et al.*, 1988) the K1 protein has been designated KpsF and will be referred to here as $KpsF_{K1}$ to distinguish it from the K5 protein, $KpsF_{K5}$.

Region 1 is transcribed as a polycistronic operon with *kpsF* as the first gene (chapter 3). The KpsF protein would therefore be expected to be involved in capsule production. However, K5 *kps* clones such as pPC6 (chapter 4, Fig. 4.1)(Pazzani, 1992) which lack *kpsF* can still direct capsule synthesis. These observations have several possible interpretations. The KpsF protein may not be involved in capsule production, but the promoter of the *kpsF* gene is required for expression of region 1. In this case the KpsF protein might not be expressed or may perform an unrelated function. Alternatively KpsF may have a non-essential, perhaps regulatory role and although affecting the process of capsular polysaccharide production sufficient surface polymer is produced in its absence to maintain sensitivity to K5-specific bacteriophage. Another possibility is that KpsF is functionally redundant and when *kpsF* is deleted another protein is capable of performing the same function in capsule production.

The first step in assigning a function to an unknown protein is to analyse the predicted amino acid sequence. This can suggest the likely cellular location and secondary structure of the protein. Comparison of the sequence with those in databases may reveal similarities with known proteins. Proteins with a low degree of similarity may be derived from a common ancestor by divergent evolution and therefore be homologous, but the similarity could alternatively have arisen by convergent evolution. The greater the similarity between the amino acid sequences of two proteins the greater the probability that they are homologous. This probability is also increased if the proteins perform a common function, are of the same length and have similar topology (Doolittle *et al.*, 1986)(Saier, 1994). Once a group of homologous proteins are optimally aligned a

phylogenetic tree can be constructed to show their relatedness. If the function of the homologous protein or family of proteins is known this may provide clues to the role of the new protein. Proteins have a modular organisation and with programs such as Motifs (Devereux *et al.*, 1984) it is possible to identify known functional domains.

Detection of a protein without the use of an over-expression system suggests that it is synthesised *in vivo*. Demonstration that the KpsF protein is produced and therefore almost certainly has a function is particularly important given the lack of a direct role in the biosynthesis of the K5 capsular polysaccharide, which can be made in its absence. A potential role for KpsF is in regulation of capsule expression and its effect on the region 1 promoter is investigated. It has been suggested for the K1 *kps* locus that at 37°C KpsF may activate a promoter in the *kpsF-kpsE* intergenic region (Cieslewicz and Vimr, 1995) and this hypothesis is tested for the K5 region 1. Alternatively, regulation might be achieved by transcriptional termination in this region, possibly involving KpsF.

5.2 Analysis of the predicted amino acid sequence of KpsF

5.2.1 Primary structure and hydropathicity of KpsF

The predicted amino acid sequence of KpsF (chapter 3, Fig. 3.11) indicates a relatively hydrophilic, 35.6 kDa protein. The hydropathicity plot suggests a cytoplasmic location with no hydrophobic stretches long enough to be likely to comprise a membrane-spanning domain (Fig. 5.1).

5.2.2 Database searches

The Basic Local Alignment Search Tool (BLAST) algorithm (Altschul *et al.*, 1990) was used to search the non-redundant database maintained at the National Center for Biotechnology Information (NCBI) in Bethesda, Maryland, USA to find sequences that are similar to the predicted amino acid sequence of KpsF. The BLASTX version compares the amino acid query sequence to the nucleotide database translated in all six reading frames.

KpsF homologues

$KpsF_{K5}$ is 94.6% identical to $KpsF_{K1}$ over 317 amino acids, with the variation mostly at the C-terminus of the shorter $KpsF_{K1}$. However the similarity continues if the K1 DNA sequence is translated in the next reading frame,

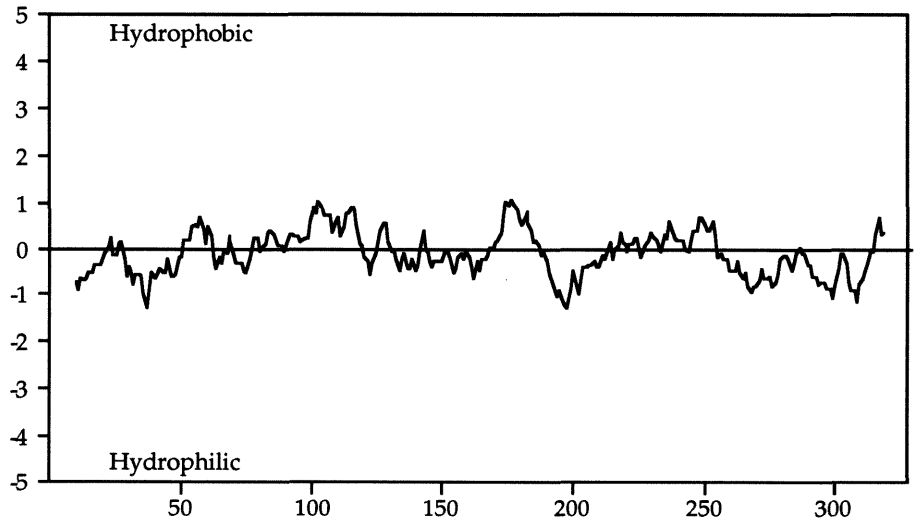


Figure 5.1. Hydropathicity plot of KpsF, calculated with a window of 20 amino acids using the algorithm of Kyte, J. and R. F. Doolittle (Kyte and Doolittle, 1982).

indicating a frame-shift mutation or a sequencing error. This extended translation would generate a $KpsF_{K1}$ of the same length as $KpsF_{K5}$ (328 amino acids) and would share 98.2% identity. The conclusion that $KpsF_{K5}$ represents the correct sequence is supported by the similarity between its C-terminus and those of the homologous proteins described below. The $kpsF_{K5}$ and $kpsF_{K1}$ genes are 98.8% identical at the nucleotide level and this degree of identity continues for the 130 bp of sequence available upstream of $kpsF_{K1}$. The 3' end of a *kpsF* gene has also been identified in *E. coli* K7 (Smith, A. N., unpublished).

$KpsF_{K5}$ is similar to three other proteins (Fig. 5.2); GutQ (Yamada *et al.*, 1990) and ORF 328 from *E. coli*, and KpsF from *H. influenzae* (Fleischmann *et al.*, 1995), which I shall refer to as $KpsF_H$. When the degree of similarity between proteins is >25% over a stretch of at least 100 amino acids it is highly likely that they are homologous (Doolittle *et al.*, 1986). Given the length of the alignments and degree of identity between $KpsF_{K5}$ and GutQ, ORF328 and $KpsF_H$ these proteins may be described as homologous. This assumption is supported by the similar length of the proteins. KpsF, GutQ, ORF328 and $KpsF_H$ were aligned using the ClustalW program (Thompson *et al.*, 1994) (Fig. 5.2).

The GutQ protein is a member of the glucitol operon which is responsible for glucitol catabolism. The operon comprises the *gutO, P, A, B, D, M, R* and *Q* genes (Fig. 5.3) and is located at 58 min on the *E. coli* chromosome. *gutA* and *gutB* encode the glucitol-specific enzyme II and enzyme III, respectively of the phosphoenolpyruvate-dependent sugar phosphotransferase system (PTS) and *gutD* encodes glucitol 6-phosphate dehydrogenase (Yamada and Saier, 1987). GutM is a transcriptional activator and GutR a repressor of the *gut* operon (Yamada and Saier, 1988). GutQ was originally reported as a 23.5 kDa protein and Yamada *et al* observed a 21 kDa protein in maxicells presumed to be GutQ (Yamada *et al.*, 1990). However a recent sequencing project spanning this region predicts that gutQ is a larger 32.7 kDa ORF. Yamada *et al* also reported a 34 kDa band which they assumed to be a dimer, but now seems likely to be the intact GutQ protein (with the smaller polypeptide perhaps a degradation product). The extended ORF is nearer the size of the homologous proteins and similarity continues into the new C-terminal end. The revised GutQ predicted amino acid sequence therefore appears to be correct and is used for analyses in this study. The function of GutQ is unclear and although its over-expression had a small effect on expression of *gutA, B* and *D* there was no effect on glucitol fermentation or the activity of the glucitol PTS system *in vitro* (Yamada *et al.*, 1990). An insertion in *gutQ* had no effect on the induced state, but in the uninduced state the level of the dehydrogenase (GutD) was increased three-fold (Saier, M. H., personal communication). GutQ may therefore be a co-repressor with GutR. GutQ has a predicted nucleotide-binding site, reported to resemble those of ATP-binding proteins (but see below).

ORF328 is a hypothetical protein predicted from nucleotide sequence (Accession no. U18997) at approximately 70 min on the *E. coli* chromosome. It is located near to, but probably not transcriptionally linked with *rpoN,* the gene encoding σ^{54} (Fig. 5.3). Nucleotide similarity between *kps* and U18997 is restricted to the regions encoding KpsF and ORF328 respectively. $KpsF_H$ is also a hypothetical protein, designated purely by its similarity with $KpsF_{K1}$. Certain strains of *Haemophilus influenzae* produce group II-like capsular polysaccharides and it is possible that $KpsF_H$ could perform the same function in capsule expression as KpsF in *E. coli.* The gene encoding $KpsF_H$ was sequenced as part of the *H. influenzae* Rd genome (Fleischmann *et al.*, 1995). This is a capsule-deficient mutant in which the serotype d capsulation locus (*cap*) has been lost by recombination between flanking IS*1016* elements (Kroll *et al.*, 1989)(Kroll *et al.*, 1991). The residual *capd* locus comprises a single copy of IS*1016* (Accession no. X58173) which is located at 1,081,965 nucleotides on the *H. influenzae* Rd genome, far from $kpsF_H$ (1,746,301). The chromosomal location of *cap* loci from

```
                10                  30                  50
KpsF        -MSERHLPDDQSSTIDPY----------LITSVRQTLAEQSAALQNLSKQLDSGQYQRVL
GutQ        ------------------------------------MLELQEASRLPERLG-DDFVRAA
KpsFH       MPNFSFVFFYDSAKITPISTALLGRRMNYLKIAQDSLSVESNALLQLSQRLG-DDFNQVI
ORF328      ---------MSHVELQPG--------FDFQQAGKEVLAIERECLAELDQYIN-QNFTLAC
                                                .     *   .    .

                70                  90                  110
KpsF        NLIMNCKGHVILSGMGKSGHVGRKMSATLASTGTPSFFIHPAEAFHGDLGMITPYDLLIL
GutQ        NIILHCEGKVVVSGIGKSGHIGKKIAATLASTGTPAFFVHPAEALHGDLGMIESRDVMLF
KpsFH       DLILACEGRLVIGGIGKSGLIGKKMVATFASTGTPSFFLHPTEAFHGDLGMLKPIDIVML
ORF328      EKMFWCKGKVVVMGMGKSGHIGRKMAATFASTGTPSFFVHPGEAAHGDLGMVTPQDVVIA
              .  * *.... *.**** .*.*. ** ******.**.** ** ******.   *...

                130                 150                 170
KpsF        ISASGETDEILKLVPSLKNFGNRIIAITNNGNSTLAKNADAVLELHMANETCPNNLAPTT
GutQ        ISYSGGAKELDLIIPRLEDKSIALLAMTGKPTSPLGLAAKAVLDISVEREACPMHLAPTS
KpsFH       ISYSGETDDVNKLIPSLKNFGNKIIAVTSNKNSTLARHADYVLDITVEREVCPNNLAPTT
ORF328      ISNSGESSEITALIPVLKRLHVPLICITGRPESSMARAADVHLCVKVAKEACPLGLAPTS
            ** ** . ..  ..* *      ....*    * .   *   * . .  * **  ****.

                190                 210                 240
KpsF        STTLTMAIGDALAIAMIHQRKFMPNDFARYHPGGSLGRRLLTRVADVMQHD--VPAVQLD
GutQ        STVNTLMMGDALAMAVMQARGFNEEDFARSHPAGALGARLLNKVHHLMRRDDAIPQVALT
KpsFH       SALVTLALGDALAVSLITARNFQPADFAKFHPGGSLGRRLLCKVKDQMQTR--LPTILPT
ORF328      STTATLVMGDALAVALLKARGFTAEDFALSHPGGALGRKLLLRVNDIMHTGDEIPHVKKT
            *.  *. .*****....  * *   ***  ** *.** .** .*   *.    .* .

                260                 280                 300
KpsF        ASFKTVIQRITSGCQGMVMVEDAEGGLAGIITDGDLRRFMEKEDS-LTSATAAQMMTREP
GutQ        ASVMDAMLELSRTGLGLVAVCDAQQQVQGVFTDGDLRR-WLVGGG-ALTTPVNEAMTVGG
KpsFH       TNFTDCLTVMNEGRMGVALVMENEQ-LKGIITDGDIRRALTANGAGTLNKTAKDFMTSSP
ORF328      ASLRDALLEVTRKNLGMTVICDDNMMIEGIFTDGDLRRVFDMGVD-VRQLSIADVMTPGG
            .     .  .     *.  . .    . *. ****.**                  **

                320                 340
KpsF        LTLPEDTMIIEAEEKMQKHRVSTLLVTNKANKVTGLVRIFD-------
GutQ        TTLQSQSRAIDAKEILMKRKITAAPVVDENGKLTGAINLQDFYQAGII
KpsFH       KTIHQDEFLSKAEDFMKAKKIHSLVVVNDENHVVGLVEFSS-------
ORF328      IRVRPGILAVEALNLMQSRHITSVMVADG-DHLLGVLHMHDLLRAGVV
              .       *   .  ... .  *     .. * .
```

Protein	Amino acids: % identity with KpsF	DNA similarity	% GC content	Accession number
KpsF	-	-	49.1	X95264
$KpsF_H$	49.7% in 308aa	56.5% in 863 bp	40.7	U32841
ORF328	41.4% in 307aa	57.7% in 690 bp	51.9	U18997
GutQ	41.4% in 292aa	60.1% in 544 bp	57.6	X51361(P17115) U29579

Figure 5.2. Multiple alignment of KpsF with gutQ, $KpsF_H$ and ORF 328 generated by the program ClustalW (Thompson *et al.*, 1994). Wholly conserved residues are marked with an asterisk (*) and those where conservative amino acid changes have occurred with a dot (.). The putative nucleotide binding domain is shown in bold. The table gives details of the similarity between KpsF and the other proteins at the amino acid and nucleotide level. The overall GC content and database accession number of each gene is shown.

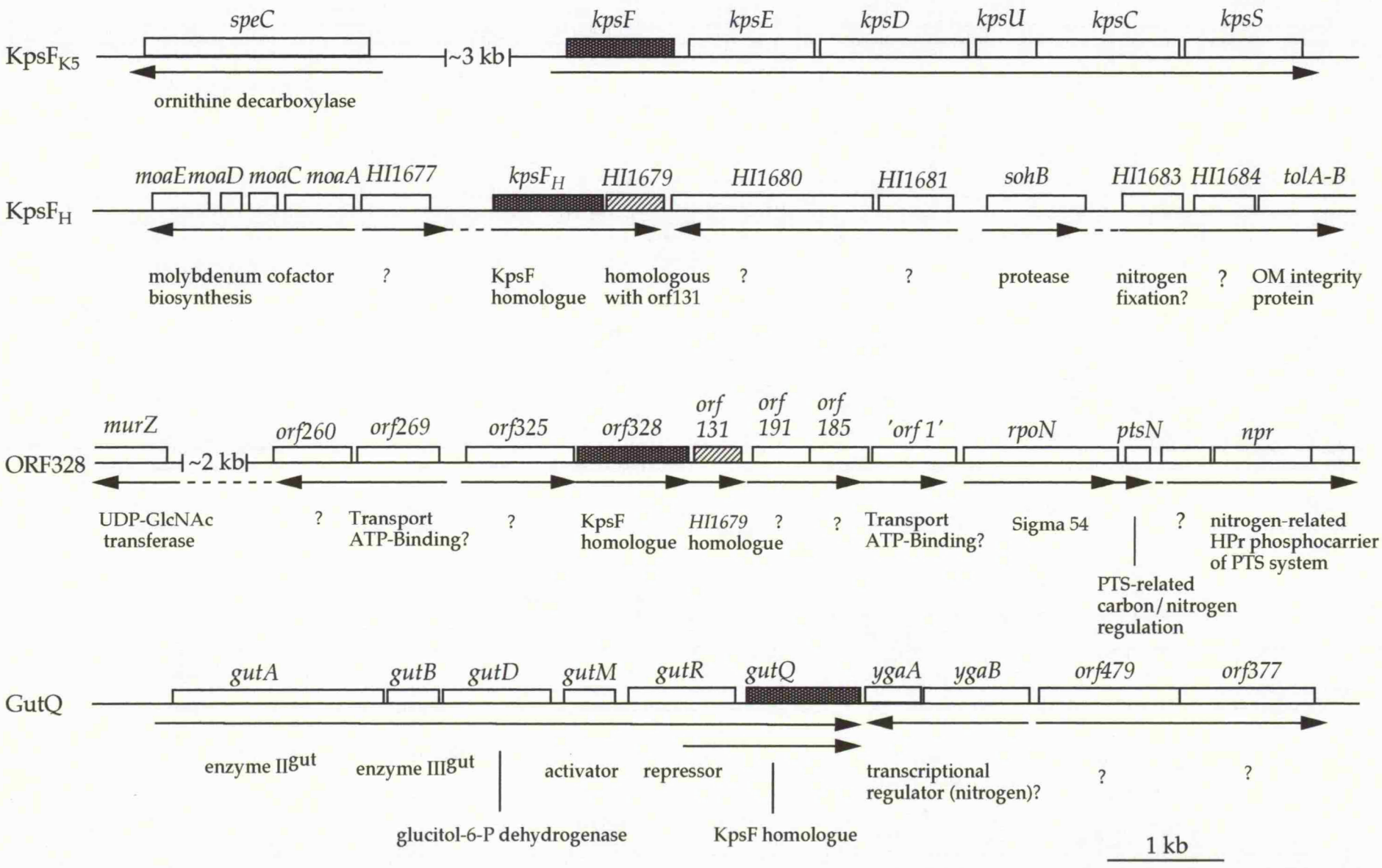
different serotypes is the same, at least within phylogenetic division I strains (Kroll *et al.*, 1989)(Kroll and Moxon, 1990). The $kpsF_H$ gene is therefore not part of the *cap* locus, unlike *kpsF* in the *E. coli kps* locus.

Genetic organisation of the KpsF homologues

Analysis of the genes surrounding *gutQ*, *orf328* and $kpsF_H$ may provide clues to the functions of the KpsF homologues. From the orientation and relative positions of genes and predicted ORFs it is possible to predict the likely transcriptional organisation (Fig. 5.3). The KpsF homologues do not share a common genetic organisation. $kpsF_{K5}$ has been shown to be the first gene of a long operon and while *orf328* appears to be the second gene, *gutQ* is at the end of an operon (Fig. 5.3). The $kpsF_H$ gene is potentially co-transcribed with only one gene. If the KpsF homologues perform a similar function, perhaps regulating the expression of linked genes, the observation that they are not all located as the first gene of an operon may have implications for the models investigated below (sections 5.4 and 5.5).

HI1679 and ORF131 are encoded by genes downstream of $kpsF_H$ and *orf328*, respectively. They share 42.1% identity over 95 amino acids and are therefore likely to be homologous. Among the *E. coli* KpsF homologues $KpsF_H$ is most similar to ORF328. These observations suggest that $KpsF_H$ and ORF328 perform the same function.

Figure 5.3. The genetic organisation of the KpsF homologues. The genes encoding KpsF and its homologues $KpsF_H$, ORF328 and GutQ are represented by shaded boxes. Transcripts demonstrated experimentally or predicted from the nucleotide sequence are indicated by arrows. Gene names and known functions are given and those functions suggested by similarity of the predicted protein to known proteins are indicated with a question mark (the functions of the *kps* proteins are described in chapter 1). HI1679 from *H. influenzae* and ORF 131 from *E. coli*, which are 42.1% identical over 95 amino acids, are represented by hatched boxes.


KpsF$_{K5}$
speC
~3 kb
kpsF
kpsE
kpsD
kpsU
kpsC
kpsS
ornithine decarboxylase
KpsF$_H$
moaE moaD moaC moaA HI1677
kpsF$_H$
HI1679
HI1680
HI1681
sohB
HI1683 HI1684
tolA-B
molybdenum cofactor biosynthesis
?
KpsF homologue
homologous with orf131
?
?
protease
nitrogen fixation?
?
OM integrity protein
ORF328
murZ
~2 kb
orf260
orf269
orf325
orf328
orf 131
orf 191
orf 185
'orf 1'
rpoN
ptsN
npr
UDP-GlcNAc transferase
?
Transport ATP-Binding?
?
KpsF homologue
HI1679 homologue
?
?
Transport ATP-Binding?
Sigma 54
?
nitrogen-related HPr phosphocarrier of PTS system
PTS-related carbon/nitrogen regulation
GutQ
gutA
gutB
gutD
gutM
gutR
gutQ
ygaA
ygaB
orf479
orf377
enzyme II^{gut}
enzyme III^{gut}
activator
repressor
transcriptional regulator (nitrogen)?
?
?
glucitol-6-P dehydrogenase
KpsF homologue
1 kb

Similarity of KpsF with glucosamine-6-P synthases

The predicted amino acid sequence of KpsF is similar over a short region (Fig. 5.4) with a number of glucosamine-6-P synthase (GlmS, L-glutamine:D-fructose-6-P amidotransferase, EC 2.6.1.16) enzymes. The transfer of an amino group from glutamine to fructose-6-P to form glucosamine-6-P is the first step in the hexosamine biosynthetic pathway which is required for the synthesis of extracytoplasmic polysaccharides (Winterburn and Phelps, 1972). The glucosamine-6-P synthases are divided into two functional domains, the first 200 amino acids containing the glutamine amide transfer domain and the subsequent 400 amino acids the sugar-binding domain (Denisot *et al.*, 1991). KpsF is very unlikely to have glucosamine-6-P synthase activity because it is much smaller than the approximately 65 kDa GlmS enzymes, contains only one of multiple conserved regions and lacks conserved cysteine and lysine residues involved in catalysis (Fernández-Herrero *et al.*, 1995). However, the similarity with KpsF is at approximately 350 amino acids from the N-terminus of the GlmS enzymes and suggests that KpsF might also interact with a sugar.

```
                  * ****      +   *  *       *   *      *
KpsF_K5     DLLILISASGETDEILKLVPSLKNFGNRIIAITNNGNSTLAKNADAVLEL
Bacillus    PLFIFLSQSGETADSRAVLVQVKALGHKALTITNVPGSTLSREADYTLLL
Rhizobium   SAALFISQSGETADTLASLRYCKAHGLRIGAVVNTRESTMAREADAIFPI
Myco.       TLVVAISQSGETADTLEAVRHAKEQKAKVLAICNTNGSQIPRECDAVLYT
Mus         DVCFFISQSGETADTLMGLRYCKERGALTVGITNTVGSSISRETDCGVHI
Thermus     TLALAISQSGETIDTLEGLREAKRKGARSLGVINAKGSTLTREVEDVLYI
Esch.       SLMITLSQSGETADTLAGLRLSKELGYLGSLAICNVPGSSLVRESDLALM
Yeast       DVCVFVSQSGETADTMLALNYCIERGALTVGIVNSVGSSISRVTHCGVHI
```

Figure 5.4 Region of homology between $KpsF_{K5}$ and aminotransferases from: *Bacillus subtilis* (GlmS, P39754); *Rhizobium leguminosarum,* (NodM, P08633, (Surin and Downie, 1988) which is almost identical with NodM from *R. meliloti*); *Mycobacterium leprae* (GlmS, U00020); *Mus musculus* (U00932, which is identical with the human protein); *Thermus thermophilus* (GlmS, U17352); *E. coli* K-12 (GlmS, P17169); *Saccharomyces cerevisiae* (Gfa1, P14742). Protein names and database accession numbers are shown in parentheses.

5.2.3 Secondary structure

It is possible to predict secondary structures such as α-helices and ß-sheets from the primary amino acid sequence. The accuracy of such predictions is increased if one or more homologous proteins are available. The PredictProtein program

accessible via the EMBL-Heidelberg server (Rost and Sander, 1993) was used to predict the secondary structure of KpsF from an alignment with GutQ, ORF328 and $KpsF_H$ (Fig. 5.5).

```
........10....,...20....,...30....,...40....,...50....,...60
AA       |MSERHLPDDQSSTIDPYLITSVRQTLAEQSAALQNLSKQLDSGQYQRVLNLIMNCKGHVI|
PHD sec  |                 HHHHHHHHHHHHHHHHHHHHHHH    HHHHHHHHHH  EEE|
Rel sec  |988889975767798326578999999999999999845198725899999897157389|

          ....,...70....,...80....,...90....,..100....,..110....,..120
AA       |LSGMGKSGHVGRKMSATLASTGTPSFFIHPAEAFHGDLGMITPYDLLILISASGETDEIL|
PHD sec  |EEE   HHHHHHHHEEEE      EEEE            EE   EEEEEE    HHHHH|
Rel sec  |997375434433542575149975142445411158753121455699985499469999|

          ....,..130....,..140....,..150....,..160....,..170....,..180
AA       |KLVPSLKNFGNRIIAITNNGNSTLAKNADAVLELHMANETCPNNLAPTTSTTLTMAIGDA|
PHD sec  |HHHHHHHHH   EEEEE    HHHHHHHHHHEEEEE                EEHHHHH|
Rel sec  |999999982993999819998577799875312564112686688999865222236699|

          ....,..190....,..200....,..210....,..220....,..230....,..240
AA       |LAIAMIHQRKFMPNDFARYHPGGSLGRRLLTRVADVMQHDVPAVQLDASFKTVIQRITSG|
PHD sec  |HHHHHHHHH    HHHHH    HHHHHHHHHHHHHHH     EEE HHHHHHHHHHH |
Rel sec  |999999994179167664168778979999999999927899512231148999999955|

          ....,..250....,..260....,..270....,..280....,..290....,..300
AA       |CQGMVMVEDAEGGLAGIITDGDLRRFMEKEDSLTSATAAQ MMTREPLTLPEDTMIIEAEE|
PHD sec  |   EEEEE      EEEE  HHHHHHHHH         HHH         HHHHHHHHH|
Rel sec  |982799751555523752884889989416986444522541689975563203899999|

          ....,..310....,..320....,..
AA       |KMQKHRVSTLLVTNKANKVTGLVRIFD|
PHD sec  |HHHHH  EEEEEE     EEEEEEE  |
Rel sec  |999536838999659995479898419|
```

Figure 5.5 Secondary structure prediction for KpsF. A multiple sequence alignment of KpsF with GutQ, $KpsF_H$ and ORF328 was used to create a secondary structure prediction using the profile neural network method (Rost and Sander, 1993). The three states, helix (H), Strand (E) and loop (blank) are predicted and an index of the reliability of the prediction (0-9) provided.

5.2.4 Protein motifs

Proteins contain motifs which are involved in catalytic or ligand binding activity and the Prosite database (A. Bairoch) is a 'dictionary of the known motifs'. The presence of such a motif in a new protein can suggest putative functions. No perfect matches were detected between KpsF and any of the consensus sequences in Prosite. However, there is a region conserved among the KpsF homologues and shown in bold in Fig. 5.2 with the following consensus:

GX(V,L)(V,I)(V,L,I)XGXGKSG

This sequence is similar to the Walker A motif (one mismatch) for an ATP- or GTP-binding site (GXXXXGKS)(Traut, 1994). On the basis of this similarity it has been suggested previously that GutQ (Yamada *et al.*, 1990) and $KpsF_{K1}$ (Cieslewicz *et al.*, 1993) are ATP-binding proteins. However, although the KpsF motif contains the invariant lysine, essential for interaction with bound phosphate, it lacks the highly conserved first glycine. Also, the motif does not match any of the refined phosphate-binding loop (P-loop) consensus sequences from different ATP- or GTP-binding protein families (Saraste *et al.*, 1990). The motif does satisfy the principal requirements for a nucleotide binding site, forming a short, glycine-rich loop between a ß-sheet and an α-helix (see secondary structure predictions, Fig. 5.5)(Branden and Tooze, 1991). The spacing of the hydrophobic patch and conserved glycines suggest that it may bind dinucleotides rather than ATP, although the charged lysine is a feature of mononucleotide binding sites. The position of the glycines match the dinucleotide binding motif, GXGXXG, with the conserved glycines facilitating a sharp bend in the loop, while the upstream hydrophobic side chains stabilise the ß-sheet interactions (Möller and Amons, 1985). A hydrophobic residue (such as the leucine found in $KpsF_H$) would be expected after the third glycine rather than the histidine present in KpsF and the other homologues.

5.3 Expression of KpsF in minicells

An *Eco*RV DNA fragment spanning *kpsF* was cloned into pTZ19 in the opposite orientation to the *lac* promoter such that the gene would be transcribed by its own, rather than the *lac* promoter (pDS111, chapter 3, Fig. 3.6). A *Hin*dIII fragment encoding a truncated KpsF protein was also cloned into pTZ19 (pDS110) and the proteins synthesised from the recombinant plasmids in minicells of strain DS410 (Dougan and Sherratt, 1977) examined. Plasmid pDS111 encoded a non-vector protein of approximately the molecular mass predicted for KpsF (35.6 kDa, Fig. 5.6, Lane 1). The expression from pDS110 of a protein with the molecular mass predicted for the truncated KpsF (22.1 kDa, Fig. 5.6, Lane 2) confirmed that *kpsF* was the gene being expressed. The demonstration that *kpsF* is expressed from its own promoter and ribosome binding site, albeit on a plasmid, suggests strongly that it is expressed when in its wild-type chromosomal location.

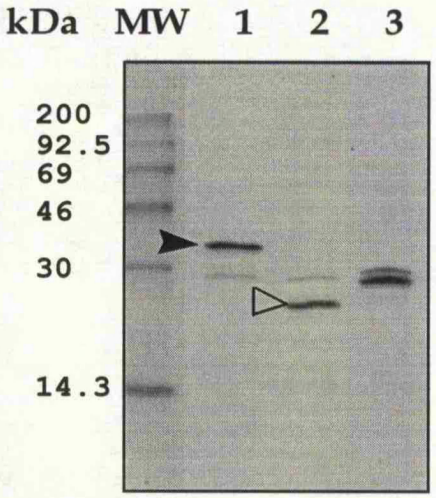


Figure 5.6. Expression of KpsF in minicells. Proteins expressed by minicells harbouring the following plasmids. Lane 1; pDS110, encoding KpsF (solid arrow) predicted MW 35.6 kDa. Lane 2; pDS111, encoding a truncated KpsF (open arrow) with predicted MW of 22.1 kDa. Lane 3; pTZ19. The sizes of MW markers are indicated in kDa.

5.4 *kpsF-kpsE* intergenic region and *kpsF* intragenic transcriptional terminator

5.4.1 Investigation of KpsF-dependent promoter upstream of *kpsE*

A model has been proposed for the K1 *kps* locus in which KpsF is required to activate a promoter upstream of *kpsE* (Cieslewicz and Vimr, 1995). At 37°C KpsF is produced, but at 18°C transcription from the *kpsF* promoter is blocked preventing synthesis of KpsF and therefore expression of region 1 from the *kpsE* promoter. Given the near identity of K1 and K5 region 1 sequences (section 5.2.2) such a regulatory mechanism would be expected to apply to both systems. A *Hin*dIII-*Bam*HI fragment spanning the K5 *kpsF-kpsE* intergenic region has been cloned previously upstream of *lacZ* creating plasmid pCB192HB (chapter 3, Fig. 3.4). No significant activity was detected in the strain SURE™ which lacks *kpsF* (chapter 3, Fig. 3.4). To investigate whether KpsF activates a promoter in this region pCB192HB was introduced into MS101 (which carries a chromosomal copy of the *kps* locus, including *kpsF*) but the level of ß-galactosidase activity produced (0.010 ± 0.003) was equivalent to that from pCB192 (0.012 ± 0.004)(values are in Miller units ± SD). Plasmid pDSF was constructed to investigate whether a higher concentration of KpsF was required to activate the putative promoter. This comprised the *Eco*RV fragment which was demonstrated to direct synthesis of KpsF in minicells (Fig. 5.6) cloned in pACYC184, which has a p15A origin of replication. pDSF is compatible with the pCB192 constructs which have a pMB1 replicon. When pDSF was introduced into MS101 carrying pCB192HB, the presence of extra KpsF had no significant effect on ß-galactosidase activity (0.017 ± 0.007).

5.4.2 Assay for transcriptional termination

An alternative regulatory model could involve transcriptional termination in the *kpsF-kpsE* intergenic region. In keeping with the apparent role of its homologue GutQ as a repressor, KpsF might mediate termination at 18°C. To study transcriptional termination a *Hin*dIII-*Bam*HI fragment spanning from within *kpsF* to within *kpsE* was cloned into the vector pHV100 (Peabody *et al.*, 1989) creating pDS300 (Fig. 5.7A). The DNA was inserted at a polylinker located between the *lac* promoter and the *luxAB* genes, which encode the luciferase enzyme of the bioluminescent marine bacterium *Vibrio harveyi*. Translation termination codons in all three frames prevent formation of luciferase fusion proteins and luciferase expression therefore represents the level of transcriptional readthrough. Bioluminescence was measured at intervals

following the induction of the *lac* promoter by the addition of IPTG. Reduction in bioluminescence compared with the vector indicates transcriptional termination occurring within the inserted fragment.

The DNA inserted in pDS300 was shown to cause significant transcriptional termination at both 37°C (Fig. 5.7B) and 18°C (Fig. 5.8). Subclones and deletion derivatives of pDS300 (pDS301, pDS302 and pDS304, Fig. 5.7A) were constructed to define more precisely the location of the terminator. Transcriptional termination activity was retained by a *Hin*dIII-*Dra*I fragment (pDS304, Fig. 5.7C) but lost in a construct starting at a *Sma*I site 90 bp upstream of the *Dra* I site (pDS301, Fig. 5.7B). The transcriptional terminator is therefore located in a 330 bp *Hin*dIII-*Sma*I fragment within *kpsF*.

Figure 5.7. Transcriptional termination. A. Fragments cloned into pHV100 to study transcriptional termination. Restriction endonuclease abbreviations: B, *Bam*HI; D, *Dra*I; EV, *Eco*RV; H, *Hin*dIII; Hc, *Hin*cII; S, *Sma*I. B. Assays for transcriptional termination at 37°C. The solid symbols refer to the OD_{600} values on the left hand scale and the open symbols to the luminescence values on the right hand scale: pHV100 (Δ), pDS300 (□), pDS301 (◊) pDS302 (O). IPTG was added at Time=0. C. Assay for transcriptional termination with pHV100 (Δ), pDS300 (□) and pDS304 (+).

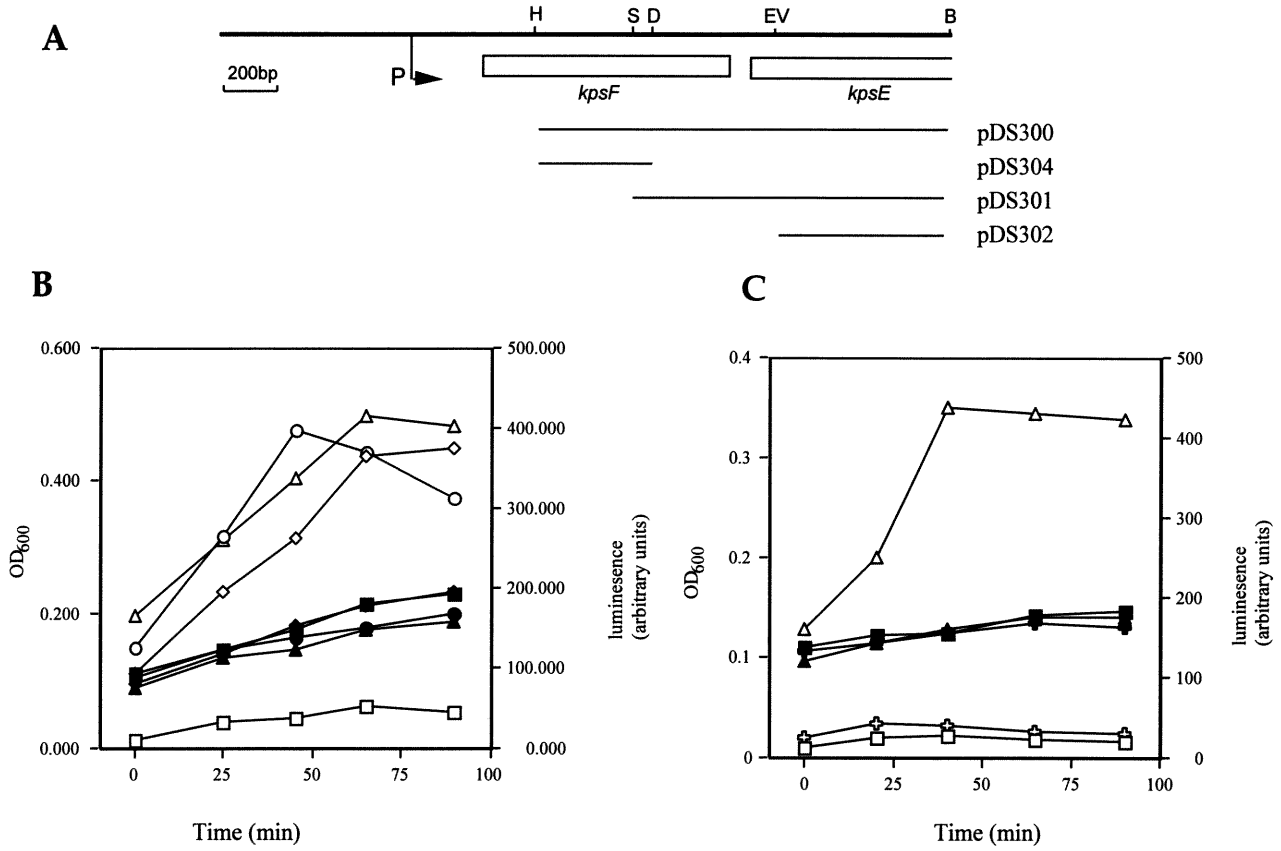


Figure 5.7. Assays for transcriptional termination.

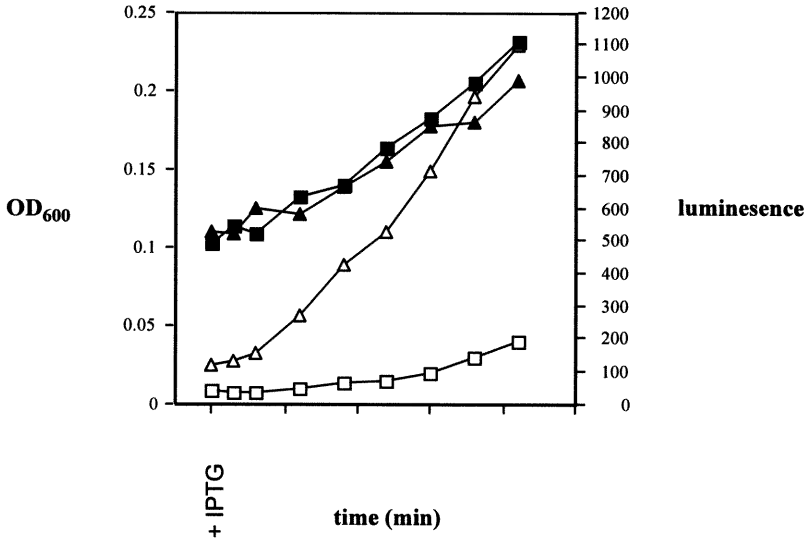


Figure. 5.8. Transcriptional termination at 18°C. The solid symbols refer to the OD_{600} values on the left hand scale and the open symbols to the luminesence values on the right hand scale: pHV100 (Δ), pDS300 (□).

5.4.3 Promoter-probe analysis

The promoter-probe construct used previously to demonstrate the presence of the region 1 promoter (pCB192HcB, Fig. 5.9A and chapter 3, Fig. 3.4) contained the transcriptional terminator region described above (section 5.4.2). One would expect a *lacZ* gene fused to the promoter upstream of the terminator to be more highly expressed. To demonstrate this, and thereby confirm the presence of the proposed terminator, plasmid pCB192HcH which carries the promoter on a fragment ending upstream of the terminator, was constructed (Fig. 5.9A). Cultures carrying pCB192HcH expressed much higher ß-galactosidase activity than those with pCB192HcB (Fig. 5.9B). To demonstrate that the higher activity of pCB192HcH was due to the absence of the specific terminator and not to the shorter length of DNA between the promoter and the *lacZ* gene, pCB192HcS was constructed (Fig. 5.9A). pCB192HcS includes 330 bp spanning the putative terminator which are not present in pCB192HcH but the ß-galactosidase activity expressed (Fig. 5.9B) was comparable with that from pCB192HcB, despite the insert being over 1 kb shorter. The reduction in transcription is therefore due to termination within the region identified using pHV100 (section 5.4.2). The ß-galactosidase activity expressed from all the constructs was reduced at 18°C.

5.4.4 Sequence analysis of terminator region

Transcriptional termination may be either Rho-dependent or Rho-independent (chapter 1, section 1.4.2). Rho-independent terminators comprise a GC-rich stem-loop followed by a run of U residues (Richardson, 1993)(Holmes *et al.*, 1994). No such sequence is present in the 330 bp *Hin*dIII-*Sma*I fragment responsible for termination activity (sections 5.4.2 and 5.4.3). The features of Rho-dependent termination signals are less well defined. It has been suggested that a cytosine-over guanosine-rich sequence is required in the RNA to activate Rho-dependent transcriptional termination (Rivellini *et al.*, 1991). By plotting the percentage of G and C residues potential Rho-dependent transcriptional terminator regions may be identified (Alifano *et al.*, 1991). The G and C contents of *kpsF* were plotted with a window of 78 residues (Fig. 5.10). A cytosine-rich region corresponds with the fragment containing the transcriptional terminator, increasing the possibility that termination is dependent on Rho. Another feature of RNA bound by Rho is a weak stem-loop, usually of six bases with the sequence CAA in the loop and preceded by a CCCCA consensus. This specificity was discovered by selective enrichment for sequences which bound Rho (Schneider *et al.*, 1993). A five base stem loop with a CAA loop is present within the sequences directing termination (110 bp downstream of the *Hin*dIII site marked in Fig. 5.7), but is not preceded by the CCCCA consensus.

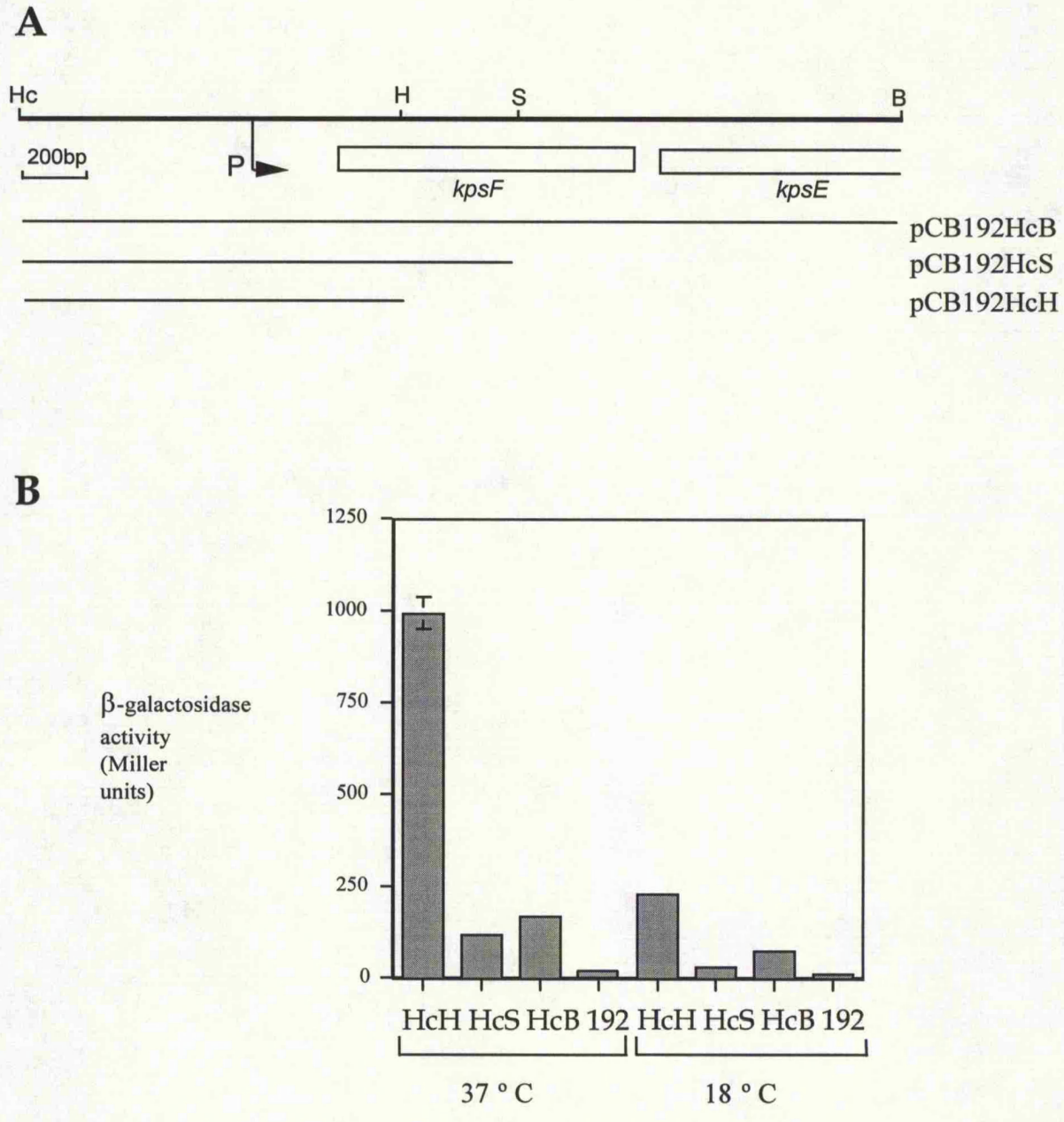


Figure 5.9. Promoter probe analysis. A. DNA fragments of different lengths containing the region 1 promoter cloned into promoter probe vector pCB192. The position of the promoter (P) is indicated with an arrow and *kpsF* and *kpsE* with boxes. Restriction endonuclease abbreviations: B, *Bam*HI; H, *Hin*dIII; Hc, *Hinc*II; S, *Sma*I. B. ß-galactosidase activity (Miller units) expressed from the recombinant plasmids described above. Assays were performed in MS101 and where visible the error bars show the standard deviation of the readings.

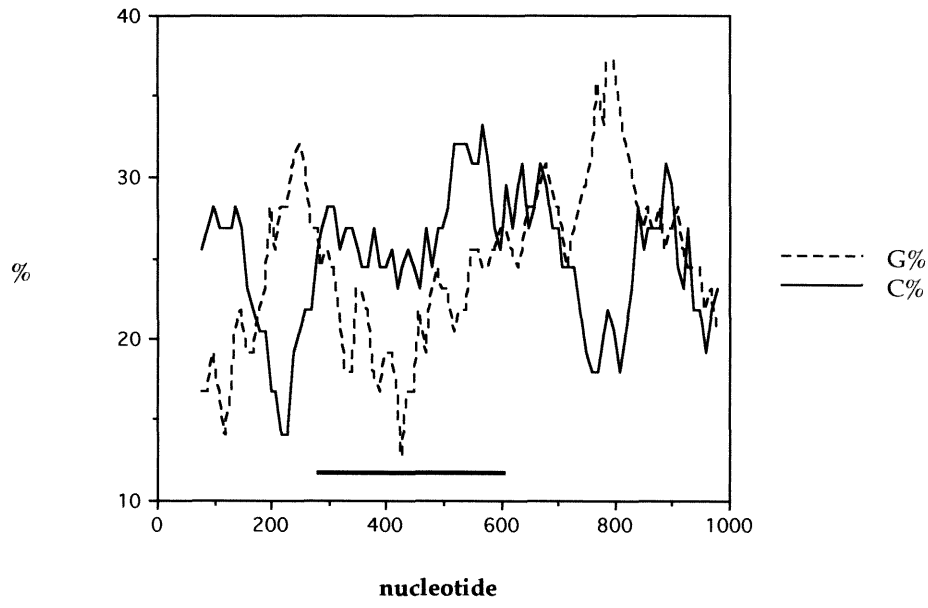


Figure 5.10. Percentage of C and G residues in *kpsF*. The percentage of C (solid line) and G residues (broken line) was calculated over a window of 78. The bar indicates the region of the gene implicated in transcriptional termination and corresponds with a cytosine- over guanosine-rich region, a feature of Rho-dependent terminators.

5.4.5 Termination activity in a Rho⁻ background

To investigate whether termination was Rho-dependent the activities of the promoter-probe constructs when in a Rho⁻ mutant host (2055)(Morse and Guertin, 1972) were compared with those in the Rho⁺ parent (W3110). In the Rho⁺ strain the ratio of the activity of pCB192HcH to that of pCB192HcB or pCB192HcS was approximately 3:1 while in the isogenic strain lacking functional Rho it was approximately 1:1. This demonstrates that sequences present in pCB192HcB and pCB192HcS comprise a Rho-dependent terminator which is responsible for the reduced transcriptional readthrough compared with pCB192HcH observed in Rho⁺ backgrounds.

5.5 Autoregulation of KpsF expression

Many proteins control their own expression by acting as a negative regulator of their own promoter. By controlling the activity of the region 1 promoter by such an autoregulatory mechanism KpsF could maintain the appropriate level of the other region 1 proteins which are expressed from the same polycistronic mRNA. If this system existed increasing the concentration of KpsF artificially would be expected to reduce promoter activity. To test this hypothesis KpsF was expressed *in trans* from the plasmid pDSF and the effect on region 1 promoter activity assessed by measuring the ß-galactosidase activity of the *kpsF-lacZ* transcriptional fusions in pCB192HcH and pCB192HcB. The introduction of pDSF, when compared with pACYC184, had no significant effect on ß-galactosidase activity (Fig. 5.11). Expression of KpsF therefore appears not to be autoregulated. The assays were performed in MS101 which has a chromosomal copy of *kpsF*. Introduction of pDSF was also shown to have no significant effect on the ß-galactosidase activity of the promoter fusions in SURE™ which lacks *kpsF* (data not shown).

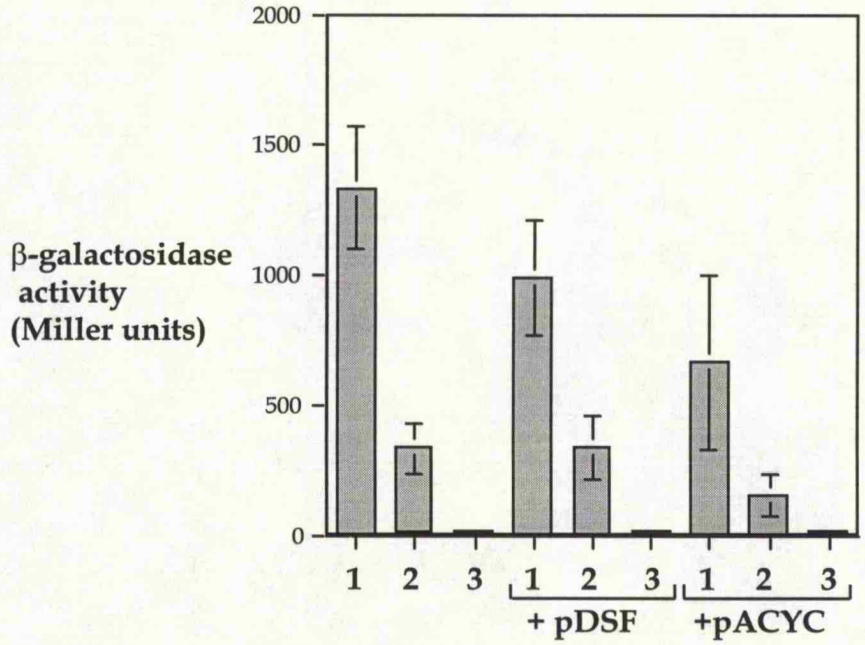


Figure 5.11. Effect of KpsF on region 1 promoter activity. The β-galactosidase activity produced by MS101 carrying pCB192HcH (Lane 1), pCB192HcB (Lane 2) and pCB192 (Lane 3) was compared with the activity from the same plasmids following the introduction of pDSF, from which additional KpsF protein was synthesised, or the vector pACYC184.

5.6 Discussion

The transcriptional organisation of region 1 suggests that *kpsF* is part of the K5 capsule gene cluster. KpsF is not required for capsule expression but it seems likely that it performs some function. It was shown that *kpsF* could be expressed in minicells using its own promoter and translational initiation region. The conservation of the *kpsF* nucleotide and predicted amino acid sequences between serotypes K5 and K1 (and the known sequence from K7) suggests that the KpsF protein has a conserved function. The cytoplasmic location predicted from the amino acid sequence is in keeping with a regulatory role rather than a direct involvement in the export of the polysaccharide across the cell membranes. The identification of three proteins homologous to KpsF (GutQ, ORF328 and $KpsF_H$), none of which appear to be directly linked with expression of capsular polysaccharide, point to a general function. The observation that GutQ affects expression of the glucitol operon (Yamada *et al.*, 1990) suggests that these proteins may represent a new family of regulatory proteins. They may be involved in the regulation of transport functions because *kpsF* is in an operon with genes involved in export of polysaccharide, *gutQ* in an operon concerned with glucitol uptake and *orf328* potentially linked with *orf1* which encodes an ORF homologous with proteins involved in transport. Analysis of the predicted amino acid sequences of KpsF and its homologues suggests that their action may involve binding of a dinucleotide.

However, there is no direct evidence that KpsF affects expression of region 1. Expression of KpsF *in trans* neither affected *kpsF* promoter activity nor activated a promoter upstream of *kpsE*. This is in keeping with an RNase protection assay with the K5 wild type in which a probe spanning the *kpsF-kpsE* intergenic region was fully protected (chapter 3, Fig. 3.7), suggesting that there is no promoter within this region. The absence of KpsF may affect the amount of capsular polysaccharide expressed or alter the structure of the K5 polysaccharide in a manner which does not alter susceptibility to K5-specific bacteriophage. Alternatively a requirement for KpsF might be masked by complementation by another protein. This is not the case with GutQ because plasmid pPC6, in which the *kpsF* gene is deleted, can direct synthesis of a capsule in *E. coli* L32706 in which the *gutQ* gene is inactivated. Although this precludes complementation by GutQ for a structural role for KpsF a regulatory role is still possible because pPC6 lacks the native promoter, the likely point of regulation. Also, this experiment does not preclude complementation by the more recently discovered ORF328.

The higher ß-galactosidase activity expressed from pCB192HcH than pCB192HcB demonstrated that a greater number of region 1 transcripts extend to the *Hin*dIII site within the *kpsF* gene than to the *Bam*HI site within *kpsE*. A possible explanation was that the functional KpsF expressed from pCB192HcB might be responsible for the reduced level of transcription. This is not the case because the ß-galactosidase activity of pCB192HcS, which like pCB192HcH carries a truncated *kpsF* gene, is comparable to that of pCB192HcB.

As well as reducing the level of transcription directed from the *kpsF* promoter, a DNA fragment within *kpsF* was shown to reduce the level of readthrough transcription from the *lac* promoter. One possible explanation for this could be a promoter in the opposite orientation within the fragment. This might interfere with transcription or destabilize the mRNA. It is possible to measure transcriptional activity in both orientations from fragments cloned in pCB192 (chapter 3, Fig. 3.3) by assaying β-galactosidase or galactokinase activity. pCB192SB (chapter 3, Fig. 3.4) carries the fragment in question but no galactokinase activity was detected from LE392 (pCB192SB) as determined by development of a purple colour due to fermentation of galactose in McConkey agar (data not shown). This fragment is therefore unlikely to contain a promoter directed against that upstream of *kpsF* and the reduction in transcription is likely to be caused by transcriptional termination. This conclusion is supported by the presence of features consistent with a Rho-dependent terminator and the dependence of the reduction in transcription on a functional *rho* gene. The observations in the *rho* mutant strain must however be treated cautiously as a mutation in *rho* has pleiotropic effects. The possible biological significance of a Rho-dependent terminator in this region is discussed in chapter 6.

While sequences within *kpsF* reduce transcription at 37°C they are not required for temperature regulation. The transcriptional activity measured from promoter-probe constructs lacking this region was reduced at 18°C.

The plasmids pPC2 and pPC6 (chapter 4, Figs. 4.1 and 4.3) carry the entire K5 capsule gene cluster except the 5' end of *kpsF*. They differ only in that pPC2 includes the *Hin*dIII-*Sma*I fragment implicated in transcriptional termination and pPC6 does not. K5 bacteriophage form very turbid plaques on a lawn of LE392 (pPC2), whilst those on LE392 (pPC6) are clear, presumably reflecting a difference in the expression of the K5 polymer on the cell surface. Although the level of the region 1 protein KpsD detected in Western blots of lysates of LE392 (pPC6) was low it was almost undetectable in LE392 (pPC2) (Fig. 4.9). These observations

further demonstrate the importance of the region within *kpsF* in determining the level of region 1 expression.

CHAPTER 6

Discussion

Certain *E. coli* group II capsular polysaccharides are associated with invasive infections of man (Ørskov and Ørskov, 1992). Their expression is regulated by temperature such that they are expressed at 37°C, the temperature encountered in the host, but not below 20°C (Ørskov *et al.*, 1984). The expression of these polysaccharides has been correlated with elevated levels of CMP-KDO synthetase activity (Finke *et al.*, 1990). The K5 *kps* locus has been studied as a paradigm for group II capsule genes and the gene encoding the CMP-KDO synthetase enzyme responsible for the high activity at 37°C has been identified in region 1 (Pazzani *et al.*, 1993). Region 1 is common to all group II capsule gene clusters and is involved in polysaccharide export. It is hoped that an understanding of how this region is regulated will lead to new methods of treating infections caused by encapsulated bacteria by preventing capsule expression. This information may also have biotechnological applications: It may be possible to modify the biosynthesis of K5 polysaccharide to produce hyaluronic acid (by changing one linkage from an α to ß configuration) or heparin (Lidholt *et al.*, 1994). An understanding of the genetic and metabolic regulation of capsule production will be required to realise the potential of such systems.

This study has described the transcriptional organisation of region 1 of the K5 *kps* locus and demonstrated that its expression is regulated by temperature at the level of transcription. An investigation of certain regulatory proteins has not revealed those involved in this process. The only protein known to regulate expression of K5 capsular polysaccharide is RfaH (section 1.5.3). A mutation in *rfaH* prevented both the production of a capsule at 37°C and the synthesis of intracellular polysaccharide (Stevens *et al.*, 1994). RfaH is required for expression of complete LPS (Pradel and Schnaitmann, 1991), synthesis and secretion of α-haemolysin (Bailey *et al.*, 1992) and regulates genes involved in fertility (Beutin *et al.*, 1981). RfaH is also required for expression of the K1 capsule and presumably the other group II capsules. The effect on capsule expression is not an indirect effect of LPS truncation because a *galU* mutant which also has a truncated LPS core was shown to produce a K1 capsule (Stevens *et al.*, 1994). RfaH is a transcriptional activator but has no similarity to known regulatory proteins and lacks a DNA-binding motif (Bailey *et al.*, 1992). It is thought to mediate transcription antitermination in a manner analogous to the N and Q proteins of λ (Beutin *et al.*, 1981). This role was suggested by premature transcription termination at Rho-dependent sites within the *tra* operon (Beutin *et al.*, 1981)

and an increase in termination frequency at the *rplL-rpoB* intercistronic region (Ralling and Linn, 1994) in *rfaH* mutants. The observation that an *rfaH* mutation had the greatest effect on the activity of the most promoter-distal reporter fusions within the *rfa* operon was consistent with the RfaH protein acting as an antiterminator (Pradel and Schnaitmann, 1991). The conclusion that RfaH acts by preventing Rho-dependent termination was supported by the isolation of *rfaH* suppressor mutations in the *rho* gene (Farewell *et al.*, 1991).

To study the effect of a mutation in *rfaH* on the expression of the three regions of the K5 capsule gene cluster transcriptional fusions were made with *kfiA* and *kpsT* (Stevens, M. P. and Roberts, I. S., unpublished). Transcription of *kpsT* was slightly reduced (by one third) in an *rfaH* mutant, whereas that of *kfiA* was greatly reduced (approximately 80-fold). These results were confirmed by dot-blot analysis of RNA levels (Stevens, M. P. and Roberts, I. S., unpublished). Western blots showed that the *rfaH* mutation had no effect on the expression of the region 1 protein KpsE (chapter 4, Fig. 4.11). The *rfaH* mutant was restored to K5 bacteriophage sensitivity by introduction of the region 2 genes on a separate plasmid, indicating that the reduction in expression of region 2 was the critical effect. The model was proposed that RfaH mediates an antitermination process which allows transcripts that would otherwise terminate at the end of region 3 to continue into region 2. Transcripts have been demonstrated which span from *kpsT* into *kfiA* (Stevens, M. P. and Roberts, I. S., unpublished) and could potentially continue across region 2. It may be pertinent that all the region 2 genes in K5 and K1 and the one analysed from the K4 capsule gene cluster (Drake, 1991) are transcribed in the same direction as region 3. This model might represent a common mechanism among group II capsule gene clusters for the regulation of heterologous region 2 genes.

However, the identification of promoters upstream of *kfiA*, *kfiB* and *kfiC* and three transcripts within region 2 of the K5 cluster (Petit *et al.*, 1995) suggested that the pattern of expression of this region is complex and may vary between serotypes. Recent experiments have shown that while *kfiC* and *kfiD* are not transcribed at 18°C *kfiA* and *kfiB* are transcribed at similar levels at 37°C and 18° (Roberts, I. R. and Petit, C., unpublished). None of the region 2 proteins were detected at 18°C and it was proposed that *kfiA* and *B* are regulated post-transcriptionally.

A short conserved sequence termed JUMPStart (after its location) was identified Just Upstream of Many Polysaccharide-associated gene clusters in enteric bacteria, including the *rfa* (LPS core oligosaccharide), *rfb* (LPS O antigen), *viaB* (Vi

antigen) and *kps* (group II K antigen) loci (Hobbs and Reeves, 1994). This sequence has now been found upstream of the operons regulated by RfaH (*rfa, hly, tra* and *kps*) (Stevens *et al.*, 1994). Transcriptional antitermination is directed at RNA polymerase molecules that have initiated transcription at a specific promoter or transcribed a specific sequence element (Richardson, 1993). For example the λ N protein recognises the *nut* sequence and forms a complex with other proteins which modifies the elongating RNA polymerase and causes it to ignore downstream termination signals. In all cases in which the location of the promoter is known the JUMPStart sequence is located in the same orientation between this and the first gene (with the exception of the *hly* operon in which it is over 1 kb upstream of the promoter – it may be significant that RfaH (HlyT) has been proposed to affect transcription initiation rather than termination in the *hly* operon (Bailey *et al.*, 1992)). It is possible that the JUMPStart sequence is a recognition element involved in regulation by RfaH, which may act as an antiterminator by modifying RNA polymerase as it transcribes this sequence. In this case deletion of the JUMPStart sequence would be predicted to prevent antitermination by RfaH and therefore, according to the model proposed above, reduce expression of region 2. RNA dot blots showed that transcription of region 2 was greatly reduced upon deletion of the JUMPStart sequence while region 3 was unaffected (Stevens, M. P. and Roberts, I. S., unpublished). Deletion of the JUMPStart also resulted in an acapsular phenotype and K5 transferase activity was reduced to background levels. These data support the antitermination model proposed above.

The significant reduction in expression of region 2 upon deletion of the JUMPStart sequence upstream of region 3 suggests that transcription of region 2 is largely dependent on the region 3 promoter. Region 3 was shown to be thermoregulated (expressed at 37°C, but not at 18°C) by assaying the luminescence expressed from a *luxAB* fusion in *kpsT* (Stevens, M. P. and Roberts, I. S., unpublished). It is difficult to reconcile the role of the region 3 promoter in the expression of region 2 given the apparently temperature-independent transcription of *kfiA* and *kfiB*. It has been suggested that transcription from region 3 boosts the expression of region 2 gene products to levels sufficient for polysaccharide biosynthesis. The RfaH protein is required for normal expression of region 2 at 37°C. The condition to which RfaH may respond and regulate expression of region 2 is unknown, but is unlikely to be temperature because it exerts its action via transcripts initiating from the region 3 promoter which is independently regulated by temperature (Roberts, I. S. and Stevens, M. P., unpublished). Region 1 is expressed independently of RfaH and

lacks the JUMPStart sequence. This is consistent with a link between the JUMPStart sequence and the action of RfaH.

All the regions, 1 (this study), 2 (Roberts, I. S. and Petit, C., unpublished) and 3 (Roberts, I. S. and Stevens, M. P., unpublished) are expressed at 37°C but not at 18°C. The mechanism of temperature regulation is likely to be the same, at least for the common group II regions. The two conserved sequences identified at similar positions upstream of the region 1 and 3 promoters (chapter 3, Fig. 3.13) may be binding sites for a protein involved in the co-ordinated regulation of the rate of transcription from these two promoters. This protein(s) is unknown but was shown not to include Hha, H-NS or RimJ.

Potential IHF binding sites were identified from the nucleotide sequence either side of the promoter region. Electrophoresis of a DNA fragment containing these sites in the presence of purified IHF would demonstrate whether they are bound by this protein; if so the progress of the DNA would be retarded. This seems likely because mutations in *himA* or *himD*, the genes which encode IHF, reduced expression of region 1 five-fold (Simpson *et al.*, In press). However, IHF is not required for normal temperature regulation of K5 capsule expression (Stevens *et al.*, 1994). A physical change, such as the curvature of the DNA, which decreases with increasing temperature (Ussery, 1995) might be responsible for altering gene expression in response to changes in temperature. This could, for example, result in two regulatory proteins being brought into contact at 37°C but not at 18°C. The degree of superhelicity does not however appear to have a major effect on the activity of the region 1 promoter as mutations in the *hns* gene, which encodes the nucleoid protein H-NS had no detectable effect. Such mutations have been correlated with changes in DNA supercoiling (Higgins *et al.*, 1988).

This work has established that region 1 is regulated by temperature at the level of transcription. Control of transcriptional initiation is the most energetically favourable point for regulation and would enable co-ordinated regulation of the region 1 genes which are all transcribed from the *kpsF* promoter. The activity of the promoter is also likely to be regulated by other environmental parameters, such as osmolarity, and to be linked with the expression of other virulence factors. An approach to identify the genes whose products are involved in this regulation would be to construct a *lacZ* transcriptional fusion on the chromosome within the *kpsF* gene. Random transposon mutagenesis would include insertions in genes encoding proteins which regulate the *kpsF* promoter. These insertions could be identified by their effect on β-galactosidase activity

expressed from the *lacZ* fusion. Similar preliminary experiments with plasmid promoter fusions were unsuccessful because of insertions into the plasmid and the potential disruption of regulatory mechanisms in a multicopy plasmid system (results not shown). A chromosomal *kpsF-lacZ* fusion would provide a simple quantitative assay for the effect of different environmental conditions on the expression of region 1. For example, growth on a low osmolarity medium (L-agar without NaCl) did not affect temperature regulation of K5 capsular polysaccharide as measured by sensitivity to K5-specific bacteriophage (data not shown). This may have hidden changes in region 1 expression which would be detected by assaying β-galactosidase activity. The role of promoter elements such as the AT-rich sequences conserved between regions 1 and 3 could be investigated by measuring the effect of deletions or site-directed mutagenesis on β-galactosidase activity. The chromosomal fusion could be used to confirm results obtained in a plasmid model. The use of a *kpsF-lacZ* fusion would however not detect effects dependent on KpsF or the products of other region 1 genes whose expression might be disrupted.

The discovery of a rho-dependent terminator within *kpsF* may represent another level at which expression of region 1 is adapted to the current environment. If the nascent mRNA encoding such a terminator is not being actively co-translated it is free to bind Rho factor which could translocate unimpeded along the mRNA and when it encountered a stalled RNA polymerase catalyse release of the RNA and dissociation of the elongation complex (Platt, 1994). Under conditions of physiological stress when the mRNA is not being translated efficiently the level of transcriptional termination would increase. This provides a mechanism for preventing the unnecessary synthesis of transcripts which would not be translated (Adhya and Gottesman, 1978)(Richardson, 1991).

Determination of the transcriptional organisation of region 1 has shown that it contains an additional gene, *kpsF*. The KpsF protein is not required for capsule production. However, the degree of conservation with the equivalent K1 protein and the demonstration that it is expressed suggest that it performs some function. Comparison with the homologous protein GutQ suggested that its role might be regulatory, but it has not been possible to demonstrate this function. To investigate thoroughly the KpsF$^-$ phenotype it would be necessary to construct a chromosomal non-polar mutation in *kpsF* which retained the promoter region. To achieve this an in-frame deletion would first be created in *kpsF* carried in a suicide vector such as pCVD442 (Donnenberg and Kaper, 1991). In this case replication of the vector requires the product of the *pir* gene. Introduction of the

recombinant plasmid into a target strain lacking the *pir* gene and selection for plasmid-encoded antibiotic resistance would enable isolation of strains in which the plasmid had integrated into the chromosomal *kpsF* gene by homologous recombination. Growth on sucrose would select against the *sacB* gene encoded by the vector. Strains could therefore be selected that had lost the vector sequence by a second recombination event leaving only the mutated version of *kpsF* on the chromosome. The genotype would be confirmed by Southern blotting. It is essential that the mutation be non-polar: Insertion of a kanamycin cassette into the chromosomal $kpsF_{K1}$ gene was reported to reduce expression of capsular polysaccharide as shown by the formation of minute plaques by K1-specific bacteriophage (Cieslewicz and Vimr, 1994). However this observation was probably due to a polar effect on the expression of downstream genes in region 1. KpsF may have a subtle effect on the efficiency of capsule expression or cause an undetectable alteration in the structure of the polysaccharide. Assays of the serum resistance, K5 transferase activity and intracellular polysaccharide levels in the *kpsF* mutant might detect the effect of the loss of such functions.

An alternative approach to investigate the role of KpsF would be to clone the *kpsF* gene in an expression vector, replacing the native promoter with one inducible by for example IPTG. The effect of induction of high levels of the KpsF protein on for example the activity of the region 1 promoter could be measured. This would avoid the complications in interpretation which arise from the presence of the native promoter region in the plasmid pDSF used to express KpsF *in trans* in this study.

The *kpsS* gene is apparently expressed as a separate 1.3 kb transcript. This transcriptional organisation might enable the specific regulation of *kpsS* expression, as for example oxygen tension affects the stability of the segment of *puf* mRNA which encodes the reaction centre in *R. capsulatus* (Klug, 1993). A mutation in the *kpsS* gene drastically reduced K5 polymerisation activity (Bronner *et al.*, 1993). It has been proposed that KpsS may be involved in the formation of phosphatidyl-KDO and its attachment to the K5 polysaccharide (Roberts, 1996). This may represent the common element recognised by the machinery that exports group II polysaccharides. Genes are often organised in the same order as their products act in a particular pathway. This is the case with the genes for the biosynthesis of O antigen in group B *Salmonella,* with the last gene to be transcribed the first to act (Schnaitman and Klena, 1993). If KpsS acts in the export of polysaccharide before the other proteins encoded by region 1 and also has an effect on biosynthesis, regulation of *kpsS* expression might provide a method of fine-tuning capsular polysaccharide expression.

The determination of the sequence upstream of region 1 adds to our understanding of the genetic organisation of group II capsule gene clusters. Approximately 3.5 kb of group II-associated DNA was identified upstream of *kpsE* by Southern blotting (Drake *et al.*, 1993). Of this, 1.5 kb comprises *kpsF* and the region 1 promoter, but the function of the remainder is unclear. 1 kb upstream of *kpsF* is the relic of an IS element (chapter 3, Fig. 3.14) which could have been involved in the acquisition of the *kps* cluster. The K5 cosmid clone does not extend beyond this but comparison of the restriction map and proteins encoded by the K1 clone with the K-12 chromosomal DNA sequence adjacent to the junction with region 3 of the *kps* cluster suggest that the *kps*-specific DNA joins the K-12 chromosome approximately 1 kb beyond the putative IS element.

This study has indicated the complexity of the regulation of gene expression and shown how several mechanisms can combine to achieve the appropriate level of expression. It has provided an understanding of the transcriptional organisation of region 1 of the K5 capsular polysaccharide gene cluster and shown that thermoregulation occurs at the level of transcription. It is hoped that this will aid future work investigating the regulatory mechanisms which control capsule expression.

References

Adhya, S. and M. Gottesman (1978). Control of transcription termination. Annu. Rev. Biochem. **47**: 967-996.

Alexander, D. C. and M. A. Valvano (1994). Role of the *rfe* gene in the biosynthesis of the *Escherichia coli* O7-specific lipopolysaccharide and other O-specific polysaccharides containing *N*-acetylglucosamine. J. Bact. **176**(22): 7079-7084.

Alifano, P., F. Rivellini, D. Limauro, C. B. Bruni and M. S. Carlomagno (1991). A consensus motif common to all rho-dependent prokaryotic transcription terminators. Cell **64**: 553-563.

Allen, P. M., I. S. Roberts, G. J. Boulnois, J. R. Saunders and C. A. Hart (1987). Contribution of capsular polysaccharide and surface properties to virulence of *Escherichia coli* K1. Infect. Immun. **55**: 2662-2668.

Allen, P., C. A. Hart and J. R. Saunders (1987). Isolation from *Klebsiella* and characterization of two *rcs* genes that activate colanic acid capsular biosynthesis in *Escherichia coli*. J. Gen. Microbiol. **133**: 331-340.

Altschul, S. F., W. Gish, W. Miller and E. W. Myers (1990). Basic local alignment search tool. J. Mol. Biol. **215**: 403-410.

Aoyama, K., A. M. Haase and P. R. Reeves (1994). Evidence for effect of random genetic drift on G+C content after lateral transfer of fucose pathway genes to *Escherichia coli* K-12. Molecular Biology and Evolution **11**(6): 829-838.

Arakere, A. D., A. L. Lee and F. C.E. (1994). Involvement of phospholipid end groups of group C *Neisseria meningitidis* and *Haemophilus influenzae* type b polysaccharides in association with isolated outer membranes and in immunoasssays. J Bact **176**(3): 691-695.

Axe, D. D. and J. E. Bailey (1993). Modeling the regulation of bacterial genes producing proteins that strongly influence growth. Biotech. Bioeng. **43**: 242-257.

Badenas, C., C. Madrid and A. Juárez (1994). Excision and transposition of Tn*5* upon insertion in the *hha* gene of *Escherichia coli*. Can. J. Microbiol. **40**: 597-601.

Baga, M., M. Goransson, S. Normark and B. E. Uhlin (1988). Processed mRNA with differential stability in the regulation of *E. coli* pilin gene expression. Cell **52**: 197-206.

Bailey, M. J. A., V. Koronakis, T. Schmoll and C. Hughes (1992). *Escherichia coli* HlyT protein, a transcriptional activator of haemolysin synthesis and secretion, is encoded by the *rfaH* (*sfrB*) locus required for expression of sex factor and lipopolysaccharide genes. Mol. Micro. **6**(8): 1003-1012.

Barr, K., P. Nunes-Edwards and P. D. Rick (1989). In vitro synthesis of a lipid-linked trisaccharide involved in synthesis of enterobacterial common antigen. J. Bact. **171**(3): 1326-1332.

Bartlett, D. H. and M. Silverman (1989). Nucleotide sequence of IS*492*, a novel insertion sequence causing variation in extracellular polysaccharide production in the marine bacterium *Pseudomonas atlantica*. J. Bact. **171**(3): 1763-1766.

Bastin, D. A., G. Stevenson, P. K. Brown, A. Haase and R. Reeves (1993). Repeat unit polysaccharides of bacteria: A model for polymerization resembling that of ribosomes and fatty acid synthetase, with a novel mechanism for determining chain length. Mol. Micro. 7(5): 725-734.

Batchelor, R. A., P. Alifano, E. Biffali, S. I. Hull and R. A. Hull (1992). Nucleotide sequences of the genes regulating O-polysaccharide chain length (*rol*) from *Escherichia coli* and *Salmonella typhimurium*: Protein homology and functional complementation. J. Bact. **174**(16): 5228-5236.

Bayer, M. E. (1968). Areas of adhesion between wall and membrane of *E. coli*. J. Gen. Microb. **53**: 395-404.

Bayer, M. E. (1990). Visualization of the bacterial polysaccharide capsule. Curr. Top. Microb. Immun. **150**: 129-157.

Bayer, M. E. and M. H. Bayer (1994). Biophysical and structural aspects of bacterial capsule. ASM News **60**(4): 192-198.

Bechhofer, D. (1993). 5' mRNA stabilizers. In G. Brawerman and J.G. Belasco (Eds.), Control of mRNA stability (pp. 31-52). New York: Academic Press.

Becker, A., K. Neihaus and A. Puhler (1995). Low-molecular-weight succinoglycan is predominantly produced by *Rhizobium meliloti* strains carrying a mutated ExoP protein characterized by a periplasmic N-terminal domain and a missing C-terminal domain. Mol. Micro. **16**(2): 191-203.

Belasco, J. G. (1993). mRNA degradation in prokaryotic cells: An overview. In G. Brawerman and J.G. Belasco (Eds.), Control of mRNA stability (pp. 3-12). New York: Academic Press.

Berg, C. M., D. E. Berg and E. A. Groisman (1989). Transposable elements and the genetic engineering of bacteria. In D. E. Berg and M. M Howe (Eds), Mobile DNA (pp. 879-920). Washington D. C.: American Society for Microbiology.

Berry, A., J. D. DeVault and A. M. Chakrabarty (1989). High osmolarity is a signal for enhanced *algD* transcription in mucoid and nonmucoid *Pseudomonas aeruginosa* strains. J. Bact. **171**(5): 2312-2317.

Beutin, L., P. A. Manning, M. Achtman and N. Willetts (1981). *sfrA* and *sfrB* products of *Escherichia coli* K-12 are transcriptional control factors. J. Bact. **145**(2): 840-844.

Biggin, M. D., T. J. Gibson and G. F. Hong (1983). Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. (USA) **80:** 3963-3965.

Bilge, S. S., J. Apostol J.M., K. J. Fullner and S. L. Moseley (1993). Transcriptional organization of the F1845 fimbrial adhesin determinant of *Escherichia coli*. Mol. Micro. 7(6): 993-1006.

Birnboim, H. C. and J. Doly (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7(6): 1513-1523.

Boulnois, G. J., R. Drake, R. Pearce and I. S. Roberts (1992). Genome diversity at the *serA*-linked capsule locus in *Escherichia coli*. FEMS Microbiology Letters **100:** 121-124.

Boulnois, G. J. and K. Jann (1989). Bacterial polysaccharide capsule synthesis, export and evolution of structural diversity. Mol. Micro. 3(12): 1810-1823.

Boulnois, G. J. and I. S. Roberts (1990). Genetics of capsular polysaccharide production in bacteria. Curr. Top. Microb. Immun. **150:** 1-18.

Boulnois, G. J., I. S. Roberts, R. Hodge, K. R. Hardy, K. B. Jann and K. N. Timmis (1987). Analysis of the K1 capsule biosynthesis genes of *Escherichia coli*: Definition of three functional regions for capsule production. Mol. Gen. Genet. **208:** 242-246.

Bouvet, P. and J. G. Belasco (1992). Control of RNase E-mediated RNA degradation by 5'-terminal base pairing in *E. coli*. Nature **360:** 488-491.

Branden, C. and J. Tooze (1991). Introduction to protein structure. New York: Garland Publishing, Inc.

Brendel, V. and E. N. Trifonov (1984). A computer algorithm for testing potential prokaryotic terminators. Nucleic Acids Res. 12(10): 4411.

Bronner, D., B. R. Clarke and C. Whitfield (1994). Identification of an ATP-binding cassette transport system required for translocation of lipopolysaccharide O-antigen side-chains across the cytoplasmic membrane of *Klebsiella pneumoniae* serotype O1. Mol. Micro. 14(3): 505-519.

Bronner, D., V. Sieberth, C. Pazzani, I. S. Roberts, G. J. Boulnois, B. Jann and K. Jann (1993). Expression of the capsular K5 polysaccharide of *Escherichia coli*: Biochemical and electron microscopic analyses of mutants with defects in region 1 of the K5 gene cluster. J. Bact. 175(18): 5984-5992.

Bronner, D., V. Sieberth, C. Pazzani, A. Smith, G. Boulnois, I. Roberts, B. Jann and K. Jann (1993). Synthesis of the K5 (groupII) capsular polysaccharide in transport-deficient recombinant *Escherichia coli*. FEMS Microbiology Letters **113:** 279-284.

Brosius, J., T. J. Dull, D. D. Sleeter and H. F. Noller (1981). Gene organization and primary structure of a ribosomal RNA operon from *Escherichia coli*. J. Mol. Biol. **148**: 107-127.

Brosius, J. and A. Holy (1984). Regulation of ribosomal RNA promoters with a synthetic *lac* operator. Proc. Natl. Acad. Sci. (USA) **81**: 6929-6933.

Brown, K. L. and K. T. Hughes (1995). The role of anti-sigma factors in gene regulation. Mol. Micro. **16**(3): 397-404.

Bugert, P. and K. Geider (1995). Molecular analysis of the *ams* operon required for exopolysaccharide synthesis of *Erwinia amylovora*. Mol. Micro. **15**(5): 917-933.

Cantwell, B. J., S. S. Bilge, J. A. Apostol, W. P. Loomis and S. L. Moseley (1994). Mutations in the F1845 fimbrial adhesin determinant affecting mRNA processing and stability, abstr. D130, p. 119. In: Abstracts of the 94th General Meeting of the American Society for Microbiology 1994. American Society for Microbiology, Washington, D. C.

Carmona, M., C. Balsalobre, F. Muñoa, M. Mouriño, Y. Jubete, F. De la Cruz and A. Juárez (1993). *Escherichia coli hha* mutants, DNA supercoiling and expression of the haemolysin genes from the recombinant plasmid pANN202-312. Mol. Micro. **9**(5): 1011-1018.

Chang, A. C. Y. and S. N. Cohen (1978). Construction and characterisation of amplifiable multicopy DNA cloning vehicles from the p15A cryptic miniplasmid. J. Bact. **134**: 1141-1156.

Chen, D., R. Bowater, C. J. Dorman and D. M. J. Lilley (1992). Activity of a plasmid-borne *leu-500* promoter depends on the transcription and translation of an adjacent gene. Proc. Natl. Acad. Sci. (USA) **89**: 8784-8788.

Chitnis, C. E. and D. E. Ohman (1993). Genetic analysis of the alginate biosynthetic gene cluster of *Pseudomonas aeruginosa* shows evidence of an operonic structure. Mol. Micro. **8**(3): 583-590.

Chuang, S., D. L. Daniels and B. F.R. (1993). Global regulation of gene expression in *Escherichia coli*. J. Bact. **175**(7): 2026-2036.

Cieslewicz, M. J., S. M. Steenbergen and E. R. Vimr (1993). Cloning, sequencing, expression, and complementation analysis of the *Escherichia coli* K1 *kps* region 1 gene, *kpsE*, and identification of an upstream open reading frame encoding a protein with homology to GutQ. J. Bact. **175**(24): 8018-8023.

Cieslewicz, M. J. and E. R. Vimr (1995). Genetic and biochemical analysis of region 1 mutants of the *kps* cluster of *Escherichia coli* K1, abstr. D-105, p. 267. In: Abstracts of the 95th General Meeting of the American Society for Microbiology 1995. American Society for Microbiology, Washington, D. C.

Cieslewicz, M. J. and E. R. Vimr (1994). Identification of *kpsF* as the first gene of region 1 in the *kps* cluster for polysialic acid capsule synthesis in *Escherichia coli* K1, abstr. D-232. In Abstracts of the 94th General Meeting of the American Society

for Microbiology, (pp. p137). Las Vegas: American Society for Microbiology, Washington, D. C.

Coleman, M., R. Pearce, E. Hitchin, F. Busfield, J. W. Mansfield and I. S. Roberts (1990). Molecular cloning, expression and nucleotide sequence of the *rcsA* gene of *Erwinia amylovora,* encoding a positive regulator of capsule expression: Evidence for a family of related capsule activator proteins. J. Gen. Microb. **136:** 1799-1806.

Collis, C. M., P. L. Molloy, G. W. Both and H. R. Drew (1989). Influence of the sequence-dependent flexure of DNA on transcription in *E. coli.* Nucleic Acids Res. 17(22): 9447-9468.

Costerton, J. W., K. Cheng, G. G. Geesy, T. I. Ladd, J. C. Nickel, M. Dasgupta and T. J. Marrie (1987). Bacterial biofilms in nature and disease. Ann. Rev. Microbiol. **41:** 435-464.

Court, D. (1993). RNA processing and degradation by RNase III. In G. Brawerman and J.G. Belasco (Eds.), Control of mRNA stability (pp. 71-115). New York: Academic Press.

Cross, A. S. (1990). The biologic significance of bacterial encapsulation. Curr. Top. Microb. Immun. **150:** 87-95.

De la Cruz, F., M. Carmona and A. Juárez (1992). The Hha protein from *Escherichia coli* is highly homologous to the YmoA protein from *Yersinia enterocolitica.* Mol. Micro. 6(22): 3451-3452.

Denisot, M., F. Le Goffic and B. Badet (1991). Glucosamine-6-phosphate synthase from *Escherichia coli* yields two proteins upon limited proteolysis: Identification of the glutamine amidohydrolase and 2R ketose/aldose isomerase-bearing domains based on their biochemcal properties. Arch. Biochem. Biophys. 288(1): 225-230.

Deretic, V., M. J. Schurr, J. C. Boucher and D. W. Martin (1994). Conversion of *Pseudomonas aeruginosa* to mucoidy in cystic fibrosis: Environmental stress and regulation of bacterial virulence by alternative sigma factors. J. Bact. 176(10): 2773-2780.

Devereux, J., P. Haeberli and O. Smithies (1984). A comprehensive set of sequence analysis programs for the Vax. Nucleic Acids Res. 12(1): 387-395.

DiRita, V. J. (1992). Co-ordinate expression of virulence genes by ToxR in *Vibrio cholerae.* Mol. Micro. 6(4): 451-458.

Dodgson, C., P. Amor and C. Whitfield (1996). Distribution of the *rol* gene encoding the regulator of lipopolysaccharide O-chain length in *E. coli* and its influence on the expression of group I capsular K antigens. J. Bact. **In Press.**

Donnenberg, M. S. and J. B. Kaper (1991). Construction of an *eae* deletion mutant of enteropathogenic *E. coli* by using a positive-selection suicide vector. Infect. Immun. 59(12): 4310-4317.

Doolittle, R. F., D. F. Feng, M. S. Johnson and M. A. McClure (1986). Relationships of human protein sequences to those of other organisms. Cold Spring Harb. Symp. Quant. Biol. **51**: 447-455.

Dorman, C. J. (1995). DNA topology and the global control of bacterial gene expression: implications for the regulation of virulence gene expression. Microbiology **141**: 1271-1280.

Dorman, C. J. (1994). Genetics of bacterial virulence. Oxford: Blackwell Scientific Publications.

Dorman, C. J. (1992). The VirF protein from *Shigella flexneri* is a member of the AraC transcription factor superfamily and is highly homologous to Rns, a positive regulator of virulence genes in enterotoxigenic *Escherichia coli*. Mol. Micro. **6**(11): 1575-1576.

Dorman, C. J. and C. F. Higgins (1987). Fimbrial phase variation in *Escherichia coli*: Dependence on integration host factor and homologies with other site-specific recombinases. J. Bact. **169**(8): 3840-3843.

Dougan, G. and D. J. Sherratt (1977). The use of transposon I as a probe of the structure and function of plasmid ColE1. Mol. Gen. Genet. **151**: 151-160.

Dower, J. D., J. F. Miller and C. W. Ragsdale (1988). High efficiency transformation of *E. coli* by high voltage electroporation. Nucleic Acids Res. **16**(13): 6127-6143.

Drake, C. R. (1991). Molecular studies of *Escherichia coli* capsule gene clusters. PhD, University of Leicester.

Drake, C. R., G. J. Boulnois and I. S. Roberts (1993). The *Escherichia coli serA*-linked capsule locus and its flanking sequences are polymorphic, genetic evidence for the existence of more than two groups of capsules. J. Gen. Microb. **139**.

Drake, C. R., I. S. Roberts, B. Jann, K. Jann and G. J. Boulnois (1990). Molecular cloning and expression of the genes encoding the *Escherichia coli* K4 capsular polysaccharide, a fructose-substituted chondroitin. FEMS Microbiology Letters **66**: 227-230.

Echarti, C., B. Hirschel, G. J. Boulnois, J. M. Varley, F. Waldvogel and K. N. Timmis (1983). Cloning and analysis of the K1 capsule biosynthesis genes of *Escherichia coli*: Lack of homology with *Neisseria meningitidis* group B DNA sequences. Infect. Immun. **41**(1): 54-60.

Edwards, U. and M. Frosch (1992). Sequence and functional analysis of the cloned *Neisseria meningitidis* CMP-NeuNAc synthetase. FEMS Microbiology Letters **96**: 161-166.

Ehrlich, R., A. Larousse, M.-A. Jacquet, M. Marin and C. Reiss (1985). *In vitro* transcription initiation from three different *Escherichia coli* promoters: Effect of supercoiling. Eur. J. Biochem. **148**: 293-298.

Eldere, J. V., L. Brophy, B. Loynds, P. Celis, I. Hancock, S. Carman, J. S. Kroll and E. R. Moxon (1995). Region II of the *Haemophilus influenzae* Type b capsulation locus is involved in serotype-specific polysaccharide synthesis. Mol. Micro. **15**(1): 107-118.

Farewell, A., R. Brazas, E. Davie, J. Mason and L. I. Rothfield (1991). Suppresion of the abnormal phenotype of *Salmonella typhimurim rfaH* mutants by mutations in the gene for transcription termination factor rho. J. Bact **173**(16).

Fath, M. J. and R. Kolter (1993). ABC transporters: Bacterial exporters. Microbiological Reveiws **57**(4): 995-1017.

Feinberg, A. P. and B. Vogelstein (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Analytical Biochemistry **132**: 6-13.

Fernández-Herrero, L. A., M.-A. Badet-Denisot, B. B. and J. Berenguer (1995). *glmS* of *Thermus thermophilus* HB8: an essential gene for cell-wall synthesis identified immediately upstream of the S-layer gene. Mol. Micro. **17**(1): 1-12.

Finke, A., D. Bronner, A. V. Nikolaev, B. Jann and K. Jann (1991). Biosynthesis of the *Escherichia coli* K5 polysaccharide, a representative of group II capsular polysaccharides: Polymerisation in vitro and characterisation of the product. J. Bact. **173**(13): 4088-4094.

Finke, A., B. Jann and K. Jann (1990). CMP-KDO-synthetase activity in *Escherichia coli* expressing the capsular polysaccharides. FEMS Microbiology Letters **69**: 129-134.

Finke, A., I. Roberts, G. Boulnois, C. Pazzani and K. Jann (1989). Activity of CMP-2-keto-3-deoxyoctulosonic acid synthetase in *Escherichia coli* strains expressing the capsular K5 polysaccharide: Implication for K5 polysaccharide biosynthesis. J. Bact. **171**(6): 3074-3079.

Fleischmann, R. D. *et al.* (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science **269**: 496-512.

Freundlich, M., N. Ramani, E. Mathew, A. Sirko and P. Tsui (1992). The role of integration host factor in gene expression in *Escherichia coli*. Mol. Micro. **6**(18): 2557-2563.

Fritsch, J., R. Rothfuchs, R. Rauhut and G. Klug (1995). Identification of an mRNA element promoting rate-limiting cleavage of the polycistronic *puf* mRNA in *Rhodobacter capsulatus* by an enzyme similar to RNase E. Mol. Micro. **15**(6): 1017-1029.

Fritz, T. A., M. M. Gabb, G. Wei and J. D. Esko (1994). Two N-acetylglucosaminyltransferases catalyze the biosynthesis of heparan sulfate. J. Biol. Chem. 269(46): 28809-28814.

Frosch, M., U. Edwards, K. Bousset, B. Kraube and C. Weisgerber (1991). Evidence for a common molecular origin of the capsule gene loci in Gram-negative bacteria expressing group II capsular polysaccharides. Mol. Micro. 5(5): 1251-1263.

Frosch, M. and A. Müller (1993). Phospholipid substitution of capsular polysaccharides and mechanisms of capsule formation in *Neisseria meningitidis*. Mol. Micro. 8(3): 483-493.

Fulks, K. A., C. F. Marrs, S. P. Stevens and M. R. Green (1990). Sequence analysis of the inversion region containing the pilin genes of *Moraxella bovis*. J. Bact. 172(1): 310-316.

Ganguli, S., G. Zapata, T. Wallis, C. Reid, G. Boulnois, W. F. Vann and I. S. Roberts (1994). Molecular cloning and analysis of genes for sialic acid synthesis in *Neisseria meningitidis* Group B and purification of the meningococcal CMP-NeuNAc synthetase enzyme. J. Bact. 176(15): 4583-4589.

Georgellis, D., T. Barlow, S. Arvidson and v. A. Gabain (1993). Retarded RNA turnover in *Escherichia coli* : A means of maintaining gene expression during anaerobosis. Mol. Micro. 9(2): 375-381.

Gervais, F. G. and G. R. Drapeau (1992). Identification, cloning and characterisation of *rcsF*, a new regulator gene for exopolysaccharide synthesis that suppresses the division mutation *ftsZ84* in *E. coli* K-12. J. Bact. 174(24): 8016-8022.

Gervais, F. G., P. Phoenix and G. R. Drapeau (1992). The *rcsB* gene, a positive regulator of colanic acid biosynthesis in *E. coli* is also an activator of *ftsZ* expression. J. Bact. 174: 3964-3971.

Godessart, N., F. J. Muñoa, M. Regue and A. Juárez (1988). Chromosomal mutations that increase the production of a plasmid-encoded haemolysin in *Escherichia coli*. J. Gen. Microb. 134: 2779-2787.

Goldberg, J. B. and T. Dahnke (1992). *Pseudomonas aeruginosa* AlgB, which modulates the expression of alginate, is a member of the NtrC subclass of prokaryotic regulators. Mol. Micro. 6(1): 59-66.

Gottesman, S. and M. R. Maurizi (1992). Regulation by proteolysis: Energy-dependent proteases and their targets. Microbiological Reveiws 56(4): 592-621.

Gottesman, S. and V. Stout (1991). Regulation of capsular polysaccharide synthesis in *Escherichia coli* K12. Mol. Micro. 5(7): 1599-1606.

Green, P. J. and M. Inouye (1986). Posttranscriptional regulation of gene expression. In M. Inouye (Eds.), Bacterial outer membranes as model systems (pp. 17-35). New York: John Wiley & Sons.

Greener, A. (1990) Strategies 3: 5-6.

Gross, R. (1993). Signal transduction and virulence regulation in human and animal pathogens. FEMS Microb. Rev. **104**: 301-326.

Group, G. C. (1994). Program Manual for the Wisconsin Package, Version 8. In 575 Science Drive, Madison, Wisconsin, USA 53711:

Hale, W. B., M. W. Van Der Woude and D. A. Low (1994). Analysis of nonmethylated GATC sites in the *Escherichia coli* chromosome and identification of sites that are differentially methylated in response to environmental stimuli. J. Bact. 176(11): 3438-3441.

Hammerschmidt, S., C. Birkholz, U. Zähringer, B. D. Robertson, J. van Putten, O. Ebeling and M. Frosch (1994). Contributions of genes from the capsule gene complex (cps) to lipooligosaccharide biosynthesis and serum resistance in *Neisseria meningitidis*. Mol. Micro. 11(5): 885-896.

Hansen, M. J., L.-H. Chen, M. L. S. Fejzo and J. G. Belasco (1994). The *ompA* 5' untranslated region impedes a major pathway for mRNA degradation in *Escherichia coli*. Mol. Micro. 12(5): 707-716.

Hashimoto, Y., N. Li, H. Yokoyama and T. Ezaki (1993). Complete nucleotide sequence and molecular characterisation of ViaB region encoding Vi antigen in *Salmonella typhi*. J. Bact. **175**: 4456-4465.

Higgins, C. F., C. J. Dorman, D. A. Stirling, L. Waddell, I. R. Booth, G. May and E. Bremer (1988). A physiological role for DNA supercoiling in the osmotic regulation of gene expression in *S. typhimurium* and *E. coli*. Cell **52**: 569-584.

Higgins, C. F., S. C. Hyde, M. M. Mimmack, U. Gileadi, D. R. Gill and M. P. Gallagher (1990). Binding protein-dependent transport systems. J. Bioenerget. Biomem. 22(4): 571-592.

Hippel, v. P. H., D. G. Bear, R. B. Winter and O. G. Berg (1982). Molecular aspects of promoter function: An overview. In R.L.a.M.J.C. Rodriguez (Eds.), Promoters: Structure and Function (pp. 3-33). Praeger.

Hippel, v. P. H. and T. D. Yager (1992). The elongation-termination decision in transcription. Science **255**: 809-812.

Hobbs, M. and P. R. Reeves (1994). The JUMPstart sequence: a 39 bp element common to several polysaccharide gene clusters. Mol. Micro. 12(5): 855-856.

Hoe, N. P. and J. D. Goguen (1993). Temperature sensing in *Yersinia pestis*: Translation of the LcrF activator protein is thermally regulated. J. Bact. 175(24): 7901-7909.

Holmes, W. M., T. Platt and M. Rosenberg (1994). Termination of transcription in *E. coli*. Cell **32**: 1029-1032.

Horwitz, M. A., and S. C. Silverstein (1980). Influence of the *E. coli* capsule on complement fixation and on phagocytosis and killing by human phagocytes. J. Clin. Invest. **65:** 82-94

Houng, H. H., K. F. Noon, J. T. Ou and L. S. Baron (1992). Expression of Vi antigen in *E. coli* K-12: Characterization of ViaB from *Citrobacter freundii* and identity of ViaA with RcsB. J. Bact. **174:** 5910-5915.

Huang, J., B. F. Carney, T. P. Denny, A. K. Weissinger and M. A. Schell (1995). A complex network regulates expression of *eps* and other virulence genes of *Pseudomonas solanacearum*. J. Bact. **177**(5): 1259-1267.

Huang, J. and M. Schell (1995). Molecular characterization of the *eps* gene cluster of *Pseudomonas solanacearum* and its transcriptional regulation at a single promoter. Mol. Micro. **16**(5): 977-989.

Hulton, C. S. J., A. Seirafi, J. C. D. Hinton, J. M. Sidebotham, L. Waddell, G. D. Pavitt, T. Owen-Hughes, A. Spassky, H. Buc and C. F. Higgins (1990). Histone-like protein H1 (H-NS), DNA supercoiling, and gene expression in bacteria. Cell **63:** 631-642.

Icho, T. (1988). Membrane-bound phosphatases in *Escherichia coli*: Sequence of the *pgpB* gene and dual subcellular localization of the *pgpB* product. J. Bact. **170**(11): 5117-5124.

Jacques, N., J. Guillerez and M. Dreyfus (1992). Culture conditions differentially affect the translation of individual *Escherichia coli* mRNAs. J. Mol. Biol. **226:** 597-608.

Jaeger, J. A., D. H. Turner and M. Zuker (1989). Improved predictions of secondary structures for RNA. Proc. Natl. Acad. Sci. (USA) **86:** 7706-7710.

Jann , K. and B. Jann (1984). Structure and biosynthesis of O-antigens. In E.T. Reitschel (Eds.), Handbook of Endotoxin (pp. 138-186). Amsterdam: Elsevier Science Publishers.

Jann, B. and K. Jann (1990). Structure and biosynthesis of the capsular antigens of *Escherichia coli*. Curr. Top. Microb. Immun. **150:** 19-42.

Jann, K., T. Dengler and B. Jann (1992). Core-lipid A on the K40 polysaccharide of *Escherichia coli* O8:K40:H9, a representative of group I capsular polysaccharides. Zbl. Bakt. **276:** 196-204.

Jann, K. and B. Jann (1992). Capsules of *Escherichia coli*, expression and biological significance. Can. J. Microbiol. **38:** 705-710.

Jayaratne, P., D. Bronner, R. MacLachlan, C. Dodgson, N. Kido and C. Whitfield (1994). Cloning and analysis of duplicated *rfbk* and *rfbm* genes involved in formation of GDP-mannose in *E. coli* O9:K30 and participation of *rfb* genes in the synthesis of the group I K30 capsular polysaccharide. J. Bact. **176:** 3126-3139.

Jayaratne, P., W. J. Keenleyside and P. R. Maclachlan (1993). Characterization of *rcsB* and *rcs*C from *Escherichia coli* O9:K30:H12 and examination of the role of the *rcs* regulatory system in expression of Group I capsular polysaccharides. J. Bact. 175(17): 5384-5394.

Jennings, H. J. (1990). Capsular polysaccharides as vaccine candidates. Curr. Top. Microb. Immun. 150: 97-127.

Johnston, A. W. B. (1989). The symbiosis between *Rhizobium* and legumes. In D.A. Hopwood and K.F. Chater (Eds.), Genetics of bacterial diversity (pp. 393-414). London: Academic Press.

Jubete, Y., J. C. Zabala, A. Juárez and F. De la Cruz (1995). *hlyM,* a transcriptional silencer downstream of the promoter in the *hly* operon of *Escherichia coli.* J. Bact. 177(1): 242-246.

Keenleyside, W. J., D. Bronner, K. Jann, B. Jann and C. Whitfield (1993). Coexpression of colanic acid and serotype-specific capsular polysaccharides in *Escherichia coli* strains with Group II K antigens. J. Bact. 175(20): 6725-6730.

Keenleyside, W. J., P. Jayaratne, P. R. Maclachlan and C. Whitfield (1992). The *rcsA* gene of *E. coli* O9:K30:H12 is involved in the expression of serotype-specific group I K (capsular) polysaccharide. J. Bact. 174: 8-16.

Kido, N., V. I. Torgov, T. Sugiyama, K. Uchiya, H. Sugihara, T. Komatsu, N. Kato and K. Jann (1995). Expression of the O9 polysaccharide of *Escherichia coli*: Sequencing of the *E. coli* O9 *rfb* gene cluster, characterization of mannosyl transferases and evidence for an ATP-binding cassette transport system. J. Bact. 177(8): 2178-2187.

Kim, K. S., J. H. Kang and A. S. Cross (1986). The role of capsular antigens in serum resistance and *in vivo* virulence of *Escherichia coli.* FEMS Microbiology Letters 35: 275-278.

Kirby, J. E., J. E. Trempy and S. Gottesman (1994). Excision of a P4-like cryptic prophage leads to Alp protease expression in *Escherichia coli.* J. Bact. 176(7): 2068-2081.

Klug, G. (1993). The role of mRNA degradation in the regulated expression of bacterial photosynthesis genes. Mol. Micro. 9(1): 1-7.

Kokotek, W. and W. Lotz (1989). Construction of a *lacZ*-kanamycin-resistance cassette useful for site-directed mutagenesis and as a promoter probe. Gene 84: 467-471.

Kroll, J. S., B. Loynds, L. N. Brophy and E. R. Moxon (1990). The *bex* locus in encapsulated *Haemophilus influenzae*: a chromosomal region involved in capsule polysaccharide export. Mol. Micro. 4(11): 1853-1862.

Kroll, J. S., B. M. Loynds and E. R. Moxon (1991). The *Haemophilus influenzae* capsulation gene cluster: a compound transposon. Mol. Micro. 5(6): 1549-1560.

Kroll, J. S. and E. R. Moxon (1988). Capsulation and gene copy number at the *cap* locus of *Haemophilus influenzae* type b. J. Bact. 172(2): 859-864.

Kroll, J. S. and E. R. Moxon (1990). Capsulation in distantly related strains of *Haemophilus influenzae* type b: Genetic drift and gene transfer at the capsulation locus. J. Bact. 172(3): 1374-1379.

Kroll, J. S., S. Zamze, B. Loynds and E. R. Moxon (1989). Common organization of chromosomal loci for production of different capsular polysaccharides in *Haemophilus influenzae*. J. Bact. 171(6): 3343-3347.

Kröncke, K., G. J. Boulnois, I. S. Roberts, D. Bitter-Suermann, J. R. Golecki, B. Jann and K. Jann (1990). Expression of the *Escherichia coli* K5 capsular antigen: Immunoelectron microscopic and biochemical studies with recombinant *E. coli*. J. Bact. 172(2): 1085-1091.

Kröncke, K. D., J. R. Golecki and K. Jann (1990). Further electron microscopic studies on the expression of *Escherichia coli* group II capsules. J. Bact. 172(6): 3469-3472.

Kröncke, K. D., I. Ørskov, F. Ørskov, B. Jann and K. Jann (1990). Electron microscopic study of coexpression of adhesive protein capsules in *Escherichia coli*. Infect. Immun. 58(8): 2710-2714.

Kyte, J. and R. F. Doolittle (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157: 105-132.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680-685.

Li, J., S. W. Mason and J. Greenblatt (1993). Elongation factor NusG interacts with termination factor ρ to regulate termination and antitermination of transcription. Genes and Development 7: 161-172.

Lidholt, K., M. Fjelstad, K. Jann and U. Lindahl (1994). Substrate specificities of glycosyltransferases involved in formation of heparin precursor and E. coli K5 capsular polysaccharides. Carbohydrate Research 255: 87-101.

Liu, D., A. M. Haase, L. Lindqvist, A. A. Lindberg and P. R. Reeves (1993). Glycosyl transferases of O-antigen biosynthesis in *Salmonella enterica*: Identification and characterisation of transferase genes of groups B, C2, and E1. J. Bact. 175(11): 3408-3413.

Liu, J. and C. L. Turnbough Jr. (1994). Identification of the Shine-Dalgarno sequence required for expression and translational control of the *pyrC* gene in *Escherichia coli*. J. Bact. 176(9): 2513-2516.

Lonetto, M. A., K. L. Brown, K. E. Rudd and M. J. Buttner (1994). Analysis of the *Streptomyces coelicolor sigE* gene reveals the existence of a subfamily of eubacterial RNA polymerase σ factors involved in the regulation of extracytoplasmic functions. Proc. Natl. Acad. Sci. (USA) 91: 7573-7577.

MacLachlan, P. R., W. J. Keenleyside, C. Dodgson and C. Whitfield (1993). Formation of the K30 (Group I) capsule in *Escherichia coli* O9:K30 does not require attachment to lipopolysaccharide lipid A core. J. Bact. 175(23): 7515-7522.

Mäkelä, P. H. and B. A. D. Stocker (1984). Genetics of lipopolysaccharide. In E.T. Rietschel (Eds.), Handbook of endotoxin, Vol 1. Chemistry of endotoxin (pp. pp59-137). Amsterdam: Elsevier Biomedical Press.

Mandel, M. and A. Higa (1970). Calcium dependent bacteriophage DNA infection. J. Mol. Biol. 53: 154.

Masson, L. and B. E. Holbein (1983). Physiology of sialic acid capsular polysaccharide synthesis in serogroup B *Neisseria meningitidis*. J. Bact. 154: 728-736.

Matthews, K. S. (1992). DNA looping. Microbiological Reveiws 56(1): 123-136.

McCarthy, J. E. G. and C. Gualerzi (1990). Translational control of prokaryotic gene expression. T. I. G. 6(3): 78-85.

Mead, D. A., E. Szczesna-Skorupa and B. Kemper (1986). Single-stranded DNA "blue" T7 promoter plasmids: a versatile tandem promoter system for cloning and protein engineering. Protein Engineering 1(1): 67-74.

Meier-Dieter, U., K. Barr, R. Starman, L. Hatch and P. D. Rick (1992). Nucleotide sequence of the *E. coli rfe* gene involved in the synthesis of enterobacterial common antigen. J. Biol. Chem. 267(2): 746-753.

Meier-Dieter, U., R. Starman, K. Barr, H. Mayer and P. D. Rick (1990). Biosynthesis of enterobacterial common antigen in *Escherichia coli*. Biochemical characterisation of Tn*10* insertion mutants defective in enterobacterial common antigen synthesis. J. Biol. Chem. 265: 13490-13497.

Mekalanos, J. J. (1992). Environmental signals controlling expression of virulence determinants in bacteria. J. Bact. 174(1): 1-7.

Melefors, O., U. Lundberg and v. A. Gabain (1993). RNA processing and degradation by RNase K and RNase E. In G. Brawerman and J.G. Belasco (Eds.), Control of mRNA stability (pp. 53-70). New York: Academic Press.

Merker, R. I. and F. A. Troy (1990). Biosynthesis of the polysialic acid capsule in *Escherichia coli* K1. Cold inactivation of sialic acid synthase regulates capsule expression below 20°C. Glycobiology 1(1): 93-100.

Merrick, M. J. (1993). In a class of its own - the RNA polymerase sigma factor σ^{54} (σ^{N}). Mol. Micro. 10(5): 903-909.

Mikulskis, A. V. and G. R. Cornelis (1994). A new class of proteins regulating gene expression in enterobacteria. Mol. Micro. 11(1): 77-86.

Miller, J. F., J. J. Mekalanos and S. Falkow (1989). Coordinate regulation and sensory transduction in the control of bacterial virulence. Science **243**: 916-922.

Miller, J. H. (1972). Assay of β-Galactosidase. In Experiments in molecular genetics (pp. 352-355). Cold Spring Harbor, N. Y.: Cold Spring Harbor Laboratory Press.

Mimmack, M. L., D. Borthakur, M. A. Jones, J. A. Downie and A. W. B. Johnston (1994). The *psi* operon of *Rhizobium leguminosarum* biovar *phaseoli*: Identification of two genes whose products are located at the bacterial cell surface. Microbiology **140**(1223): 1229.

Möller, W. and R. Amons (1985). Phosphate-binding sequences in nucleotide-binding proteins. FEBS **186**(1): 1-7.

Morett, E. and L. Segovia (1993). The σ^{54} bacterial enhancer-binding protein family: Mechanism of action and phylogenetic relationship of their functional domains. J. Bact. **175**(19): 6067-6074.

Morse, D. E. and M. Guertin (1972). Amber *suA* mutations which relieve polarity. J. Mol. Biol. **63**: 605-608.

Moxon, E. R. and J. S. Kroll (1990). The role of bacterial polysaccharide capsules as virulence factors. Curr. Top. Microb. Immun. **150**: 65-85.

Müller, P., M. Keller, W. M. Weng, J. Quandt, W. Arnold and A. Pühler (1993). Genetic analysis of the *Rhizobium meliloti exoYFQ* operon: ExoY is homologous to sugar transferases and ExoQ represents a transmembrane protein. Mol. Plant-Microbe Interact. **6**(1): 55-65.

Murray, N. E., W. J. Brammer and K. Murray (1977). Lamboid phages that simplify the recovery of *in vitro* recombinants. Mol. Gen. Genet. **150**: 53.

Musser, J. M., J. S. Kroll, E. R. Moxon and R. K. Selander (1988). Clonal population structure of encapsulated *Haemophilus influenzae*. Infect. Immun. **56**(8): 1837-1845.

Musser, J. M., J. S. Kroll, E. R. Moxon and R. K. Selander (1988). Evolutionary genetics of the encapsulated strains of *Haemophilus influenzae*. Proc. Natl. Acad. Sci. USA **85**: 7758-7762.

Neidhardt, F. C. (1987). Multigene systems and regulons. In F.C. Neidhardt (Eds.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology (pp. 1313-1317). Washington, D.C.: American Society for Microbiology.

Newbury, S. F., N. H. Smith, E. C. Robinson, I. D. Hiles and C. F. Higgins (1987). Stabilization of translationally active mRNA by prokaryotic REP sequences. Cell **48**: 297-310.

Nieto, J. M., M. Carmona, S. Bolland, Y. Jubete, F. De la Cruz and A. Juárez (1991). The *hha* gene modulates haemolysin expression in *Escherichia coli.* Mol. Micro. 5(5): 1285-1293.

Nilsson, P., S. Naureckiene and B. E. Uhlin (1996). Mutations affecting mRNA processing and fimbrial biogenesis in the *E. coli pap* operon. J. Bact. 178(3): 683-690.

O'Neill, M. C. (1989). *Escherichia coli* promoters: Consensus as it relates to spacing class, specificity, repeat substructure and three-dimensional organization. J. Biol. Chem. 264(10): 5522-5530.

Ohyama, T., N. Nagumo, Y. Hirota and S. Sakuma (1992). Alteration of the curved helical structure located in the upstream region of the β–lactamase promoter of plasmid pUC19 and its effect on transcription. Nucleic Acids Res. 20(7): 1617-1622.

Ophir, T. and D. L. Gutnick (1994). A role for exopolysaccharides in the protection of microorganisms from desiccation. Appl. Environ. Microb. 60(2): 740-745.

Ørskov, F. and I. Ørskov (1992). *Escherichia coli* serotyping and disease in man and animals. Can. J. Microbiol. 38: 699-704.

Ørskov, F., V. Sharma and I. Ørskov (1984). Influence of growth temperature on the development of *Escherichia coli* polysaccharide K antigens. J. Gen. Microb. 130: 2681-2684.

Ørskov, I., V. Sharma and F. Ørskov (1976). Genetic mapping of the K1 and K4 antigens (L) of *Escherichia coli.* Non-allelism of K(L) antigens with K antigens of O8:K27(A), O8:K8(L) and O9:K57(B). Acta. Pathol. Microbiol. Scand. Sect. B 84: 125-131.

Pavelka, M. S., Jr., S. F. Hayes and R. P. Silver (1994). Characterization of KpsT, the ATP-binding component of the ABC-transporter involved with the export of capsular polysialic acid in *Escherichia coli* K1. J. Biol. Chem. 269(31): 20149-20158.

Pavelka, M. S., Jr., L. F. Wright and R. P. Silver (1991). Identification of two genes, *kpsM* and *kpsT,* in region 3 of the polysialic acid gene cluster of *Escherichia coli* K1. J. Bact. 173(15): 4603-4610.

Pazzani, C. (1992). Molecular and genetic analysis of the *Escherichia coli* K5 capsule gene cluster. PhD Thesis, University of Leicester.

Pazzani, C., C. Rosenow, G. J. Boulnois, D. Bronner, K. Jann and I. S. Roberts (1993). Molecular analysis of region 1 of the *Escherichia coli* K5 antigen gene cluster: A region encoding proteins involved in cell surface expression of capsular polysaccharide. J. Bact. 175(18): 5978-5983.

Peabody, D. S., C. L. Andrews, K. W. Escudero, J. H. Devine, T. O. Baldwin and D. G. Bear (1989). A plasmid vector and quantitative techniques for the study of transcription termination in *Escherichia coli.* Gene 75: 289-296.

Pearce, R. and I. S. Roberts (1995). Cloning and analysis of gene clusters for production of the *Escherichia coli* K10 and K54 antigens: Identification of a new group of *serA*-linked capsule gene clusters. J. Bact. 177(14): 3992-3997.

Pearson, W. R. and D. J. Lipman (1988). Improved tools for biological sequence analysis. Proc. Natl. Acad. Sci. (USA) 85: 2444-2448.

Pérez-Martin, J., F. Rojo and V. de Lorenzo (1993). Promoters responsive to DNA bending: a common theme in prokaryotic gene expression. Microbiological Reveiws 58(2): 268-290.

Petersen, C. (1993). Translation and mRNA stability in bacteria: A complex relationship. In G. Brawerman and J.G. Belasco (Eds.), Control of mRNA stability (pp. 117-145). New York: Academic Press.

Petit, C., G. P. Rigg, C. Pazzani, A. Smith, V. Sieberth, M. Stevens, G. Boulnois, K. Jann and I. S. Roberts (1995). Region 2 of the *Escherichia coli* K5 capsule gene cluster encoding proteins for the biosynthesis of the K5 polysaccharide. Mol. Micro. 17(4): 611-620.

Pickard, D., J. Li, M. Roberts, D. Maskell, D. Hone, M. Levine, G. Dougan and S. Chatfield (1994). Characterisation of defined *ompR* mutants of *Salmonella typhi*: *ompR* is involved in the regulation of Vi polysaccharide expression. Infect. Immun. 62(9): 3984-3993.

Pigeon, R. P. and R. P. Silver (1994). Topological and mutational analysis of KpsM, the hydrophobic component of the ABC-transporter involved in the export of polysialic acid in *Escherichia coli* K1. Mol. Micro. 14(5): 871-881.

Plaskon, R. P. and R. M. Wartell (1987). Sequence distributions associated with DNA curvature are found upstream of strong *E. coli* promoters. Nucleic Acids Res. 15(2): 785-796.

Platt, T. (1994). Rho and RNA: models for recognition and response. Mol. Micro. 11(6): 983-990.

Porter, M. E. and C. J. Dorman (1994). A role for H-NS in thermo-osmotic regulation of virulence gene expression in *Shigella flexneri*. J. Bact. 176(13): 4187-4191.

Poulsen, L. K., A. Refn, S. Molin and P. Andersson (1991). The *gef* gene from *Escherichia coli* is regulated at the level of translation. Mol. Micro. 5(7): 1639-1648.

Pradel, E. and C. A. Schnaitmann (1991). Effect of *rfaH* (*sfrB*) and temperature on expression of *rfa* genes of *Escherichia coli* K-12. J. Bact. 173(20): 6428-6431.

Raetz, C. R. H. (1993). Bacterial endotoxins: Extraordinary lipids that activate eucaryotic signal transduction. J. Bact. 175(18): 5745-5753.

Ralling, G. and T. Linn (1994). Evidence that Rho and NusA are involved in termination in the *rplL-rpoB* intercistronic region. J. Bact. **169**(5): 2277-2280.

Reed, J. W., M. Capage and G. C. Walker (1991). *Rhizobium meliloti exoG* and *exoJ* mutations affect the ExoX-ExoY system for modulation of exopolysaccharide production. J. Bact. **173**: 3776-3788.

Reed, J. W., J. Glazebrook and G. C. Walker (1991). The *exoR* gene of *Rhizobium meliloti* affects RNA levels of other *exo* genes but lacks homology to known transcriptional regulators. J. Bact. **173**: 3789-3794.

Reizer, J., A. Reizer and M. H. Saier Jr. (1992). A new subfamily of bacterial ABC-type transport systems catalyzing export of drugs and carbohydrates. Protein Science **1**: 1326-1332.

Reuhs, B. L., R. W. Carlson and J. S. Kim (1993). *Rhizobium fredii* and *Rhizobium meliloti* produce 3-deoxy-D-*manno*-2-octulosonic acid-containing polysaccharides that are structurally analagous to group II K antigens (capsular polysaccharidides) found in *Escherichia coli*. J. Bact. **175**(11): 3570-3580.

Reuhs, B. L., M. N. V. Williams, J. S. Kim, R. W. Carlson and F. Côté (1995). Suppression of the Fix^- phenotype of *Rhizobium meliloti exoB* mutants by *lpsZ* is correlated to a modified expression of the K polysaccharide. J. Bact. **177**(15): 4289-4296.

Rex, G., B. Surin, G. Besse, B. Schneppe and J. E. G. McCarthy (1994). The mechanism of translational coupling in *Eschericia coli*. J. Biol. Chem. **269**(27): 18118-18127.

Reynolds, R., R. M. Bermudez-Cruz and M. J. Chamberlin (1992). Parameters affecting transcription termination by *Escherichia coli* RNA polymerase. J. Mol. Biol. **224**: 31-63.

Richardson, J. P. (1991). Preventing the synthesis of unused transcripts by Rho factor. Cell **64**: 1047-1049.

Richardson, J. P. (1993). Transcription termination. Crit. Rev. Biochem. Mol. Biol. **28**(1): 1-30.

Richter-Dahlfors, A. A. and D. I. Andersson (1992). Cobalamin (vitamin B_{12}) repression of the Cob operon in *Salmonella typhimurium* requires sequences within the leader and the first translated open reading frame. Mol. Micro. **6**(6): 743-749.

Richter-Dahlfors, A. A., S. Ravnum and D. I. Andersson (1994). Vitamin B_{12} repression of the *cob* operon in *Salmonella typhimurium* : translational control of the *cbi* gene. Mol. Micro. **13**(3): 541-553.

Rick, P. D., G. L. Hubbard and K. Barr (1994). Role of the *rfe* gene in the synthesis of the O8 antigen in *E. coli*. J. Bact. **176**(10): 2877-2884.

Rivellini, F., P. Alifano, C. Piscitelli, V. Blasi, C. B. Bruni and M. S. Carlomagno (1991). A cytosine- over guanosine-rich sequence in RNA activates Rho-dependent transcription termination. Mol. Micro. 5(12): 3049-3054.

Roberts, I. S. (1996). The biochemistry and genetics of capsular polysaccharide production in bacteria. In Press.

Roberts, I. S., R. Mountford, N. High, D. Bitter-Suermann, K. Jann, K. Timmis and G. J. Boulnois (1986). Molecular cloning and analysis of genes for production of K5, K7, K12, and K92 capsular polysaccharides in *Escherichia coli*. J. Bact. 168(3): 1228-1233.

Roberts, I. S., R. Mountford, R. Hodge, K. Jann and G. J. Boulnois (1988). Common organization for gene clusters for production of different capsular polysaccharides (K antigens) in *Escherichia coli*. J. Bact. 170: 1305-1310.

Roberts, I. S., F. K. Saunders and G. J. Boulnois (1989). Bacterial capsules and interactions with complement and phagocytes. Biochemical Society Transactions 17: 462-464.

Rosenow, C., F. Esumeh, I. S. Roberts and K. Jann (1995). Characterization and localisation of the KpsE protein of *Escherichia coli* K5, which is involved in polysaccharide export. J. Bact. 177(5): 1137-1143.

Rosenow, C., I. S. Roberts and K. Jann (1995). Isolation from recombinant *Escherichia coli* and characterisation of CMP-Kdo synthetase, involved in the expression of the capsular K5 polysaccharide (K-CKS). FEMS Microbiology Letters 125: 159-164.

Rost, B. and C. Sander (1993). Improved prediction of protein secondary structure by use of sequence profiles and neural networks. Proc. Natl. Acad. Sci. (USA) 90: 7558-7562.

Saier, M. H., Jr. (1994). Computer-aided analyses of transport protein sequences: Gleaning evidence concerning function, structure, biogenesis, and evolution. Microbiological Reveiws 58(1): 71-93.

Sanger, F., S. Nicklen and A. R. Coulson (1977). DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. (USA) 74: 5463-5467.

Saraste, M., P. R. Sibbald and A. Wittinghofer (1990). The P-loop - A common motif in ATP- and GTP-binding proteins. TIBS 15: 430-434.

Schell, M., T. Denny and J. Huang (1994). VsrA, a second two-component sensor regulating virulence genes of *Pseudomonas solanacearum*. Mol. Micro. 11(3): 489-500.

Schmidt, G., B. Jann, K. Jann, I. Ørskov and F. Ørskov (1977). Genetic determinants of the synthesis of the polysaccharide capsular antigen K27(A) of *E. coli*. J. Gen. Microb. 100: 355-361.

Schmidt, M. A. and K. Jann (1982). Phospholipid substitution of capsular (K) polysaccharide antigens from *Escherichia coli* causing extra-intestinal infections. FEMS Microbiology Letters **14**: 69-74.

Schnaitman, C. A. and J. D. Klena (1993). Genetics of lipopolysaccharide biosynthesis in enteric bacteria. Microbiological Reveiws **57**(3): 655-682.

Schneider, D., L. Gold and T. Platt (1993). Selective enrichment of RNA species for tight binding to *Escherichia coli* rho factor. FASEB J **7**: 201-207.

Schneider, K. and C. F. Beck (1986). Promoter-probe vectors for the analysis of divergently arranged promoters. Gene **42**: 37-48.

Schurr, M. J., D. W. Martin, M. H. Mudd and V. Deretic (1994). Gene cluster controlling conversion to alginate-overproducing phenotype in *Pseudomonas aeruginosa*: Functional analysis in a heterologous host and role in the instability of mucoidy. J. Bact. **176**(11): 3375-3382.

Sieberth, W., G. P. Rigg, I. S. Roberts and K. Jann (1995). Expression and characterization of UDPGlc Dehydrogenase (KfiD), which is encoded in the type-specific region 2 of the *Escherichia coli* K5 capsule genes. J. Bact. **177**(15): 4562-4565.

Siemering, K. R., J. Praszkier and A. J. Pittard (1993). Interaction between the antisense and target RNAs involved in the regulation of IncB plasmid replication. J. Bact. **175**(10): 2895-2906.

Silver, R. P., W. Aaronson and W. F. Vann (1988). The K1 capsular polysaccharide of *Escherichia coli*. Reviews of Infectious Diseases **10**(2): S282-S286.

Silver, R. P., W. Aaronson and W. F. Vann (1987). Translocation of capsular polysaccharides in pathogenic strains of *Escherichia coli* requires a 60-Kilodalton periplasmic protein. J. Bact. **169**(12): 5489-5495.

Silver, R. P., C. W. Finn, W. F. Vann, W. Aaronson, R. Schneerson, P. J. Kretschmer and C. F. Garon (1981). Molecular cloning of the K1 capsular polysaccharide genes of *E. coli*. Nature **289**: 696-698.

Silver, R. P., W. F. Vann and W. Aaronson (1984). Genetic and molecular analyses of *Escherichia coli* K1 antigen genes. J. Bact. **157**(2): 568-575.

Simpson, D. A., T. Hammarton and I. S. Roberts (1996). The transcriptional organization and regulation of region 1 of the *Escherichia coli* K5 capsule gene cluster. J. Bact. **In Press**.

Sledjeski, D. and S. Gottesman (1995). A small RNA acts as an antisilencer of the H-NS-silenced *rcsA* gene of *Escherichia coli*. Proc. Natl. Acad. Sci. (USA) **92**: 2003-2007.

Smith, A. N., G. J. Boulnois and I. S. Roberts (1990). Molecular analysis of the *Escherichia coli* K5 *kps* locus: identification and characterization of an inner-membrane capsular polysaccharide transport system. Mol. Micro. **4**(11): 1863-1869.

Smith, D. B. and K. S. Johnson (1988). Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione *S*-transferase. Gene **67:** 31-40.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:** 503-517.

Spedding, G. and D. E. Draper (1993). Allosteric mechanism for translational repression in the *Escherichia coli* α operon. Proc. Natl. Acad. Sci. (USA) **90:** 4399-4403.

Steenbergen, S. M. and E. R. Vimr (1990). Mechanism of polysialic acid chain elongation in *Escherichia coli*. Mol. Micro. 4(4): 603-611.

Steenbergen, S. M., T. J. Wrona and E. R. Vimr (1992). Functional analysis of the sialyltransferase complexes in *Escherichia coli* K1 and K92. J. Bact. 174(4): 1099-1108.

Stevens, M. P., P. Hanfling, B. Jann, K. Jann and I. S. Roberts (1994). Regulation of *Escherichia coli* K5 capsular polysaccharide expression: Evidence for involvement of RfaH in the expresion of group II capsules. FEMS Microbiology Letters **124:** 93-98.

Stevens, P., S. N. H. Huang, W. D. Welch and L. S. Young (1978). Restricted complement activation by *Escherichia coli* with the K1 capsular serotype: a possible role in pathogenicity. J. Immunol. 121: 2174-2180.

Steward, K. L. and T. Linn (1992). Transcription frequency modulates the efficiency of an attenuator preceding the *rpoBC* RNA polymerase genes of *Escherichia coli* : possible autogenous control. Nucleic Acids Res. 20(18): 4773-4779.

Stout, V. and S. Gottesman (1990). RcsB and RcsC: a two-component regulator of capsule synthesis in *Escherichia coli*. J. Bact. 172(2): 659-669.

Stout, V., A. Torres-Cabassa, M. R. Maurizi, D. Gutnick and S. Gottesman (1991). RcsA, an unstable positive regulator of capsular polysaccharide synthesis. J. Bact. 173(5): 1738-1747.

Strohmaier, H., P. Remler, W. Renner and G. Högenauer (1995). Expression of genes *kdsA* and *kdsB* involved in 3-deoxy-D-*manno*-octulosonic acid metabolism and biosynthesis of enterobacterial lipopolysaccharide is growth phase regulated primarily at the transcription level. J. Bact. 177(15): 4488-4500.

Surin, B. P. and J. A. Downie (1988). Characterization of the *Rhizobium leguminosarum* genes *nodLMN* involved in efficient host-specific nodulation. Mol. Micro. 2(2): 173-183.

Thompson, J. D., D. G. Higgins and T. J. Gibson (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res. 22(22): 4673-4680.

Towbin, H., T. Staehelin and J. Gordon (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. Proc. Natl. Acad. Sci. (USA) **76:** 4350-4354.

Traut, T. W. (1994). The functions and consensus motifs of nine types of peptide segments that form different types of nucleotide-binding sites. Eur. J. Biochem. **222:** 9-19.

Trempy, J. E., J. E. Kirby and S. Gottesman (1994). Alp suppression of Lon: Dependence on the *slpA gene*. J. Bact. **176**(7): 2061-2067.

Troy, F. A., I. K. Vijay and N. Tesche (1975). Role of undecaprenyl phosphate in synthesis of polymers containing sialic acid in *E. coli*. J. Biol. Chem. **250:** 156-163.

Troy, F. A., F. E. Frerman and E. C. Heath (1971). The biosynthesis of capsular polysaccharide in *Klebsiella aerogenes*. J. Biol. Chem. **246:** 118-133.

Ussery, D. (1995). A review of curved DNA. Biochimie **In Press.**

Valvano, M. A. (1992). Pathogenicity and molecular genetics of O-specific side-chain lipopolysaccharides of *Escherichia coli*. Can. J. Microbiol. **38:** 711-719.

Vann, W. F., M. A. Schmidt, B. Jann and K. Jann (1981). The structure of the capsular polysaccharide (K5 antigen) of urinary tract infective *Escherichia coli* O10:K5:H4. A polymer similar to desulfoheparin. Eur. J. Biochem. **116:** 359-364.

Vázquez, M., O. Santana and C. Quinto (1993). The NodI and NodJ proteins from *Rhizobium* and *Bradyrhizobium* strains are similar to capsular polysaccharide secretion proteins from Gram-negative bacteria. Mol. Micro. 8(1): 369-377.

Vieira, J. and J. Messing (1987). Production of single-stranded plasmid DNA. Methods Enzymol. **153:** 3.

Vimr, E. R. (1991). Map position and genomic organization of the *kps* cluster for polysialic acid synthesis in *Escherichia coli* K1. J. Bact. **173**(3): 1335-1338.

Vimr, E. R., W. Aaronson and R. P. Silver (1989). Genetic analysis of chromosomal mutations in the polysialic acid gene cluster of *Escherichia coli* K1. J. Bact. **171**(2): 1106-1117.

Vimr, E. R. and F. A. Troy (1985). Identification of an inducible catabolic system for sialic acids (*nan*) in *Escherichia coli*. J. Bact. **164**(2): 845-853.

Vimr, E. R. and F. A. Troy (1985). Regulation of sialic acid metabolism in *Escherichia coli*: Role of *N*-acylneuraminate pyruvate-lyase. J. Bact. **164**(2): 854-860.

Vogel, U. and K. F. Jensen (1994). The RNA chain elongation rate in *Escherichia coli* depends on growth rate. J. Bact. **176**(10): 2807-2813.

Wacharotayankun, R., Y. Arakawa, M. Ohta, K. Tanaka, T. Akashi, M. Mori and N. Kato (1993). Enhancement of extracapsular polysaccharide synthesis in *Klebsiella pneumoniae* by RmpA2, which shows homology to NtrC and FixJ. Infect. Immun. 61(8): 3164-3174.

White-Ziegler, C. A., L. B. Blyn, B. A. Braaten and D. A. Low (1990). Identification of an *Escherichia coli* locus involved in thermoregulation of the *pap* operon. J. Bact. 172(4): 1775-1782.

White-Ziegler, C. A. and D. A. Low (1992). Thermoregulation of the *pap* operon: Evidence for the involvement of RimJ, the N-terminal acetylase of ribosomal protein S5. J. Bact. 174(21): 7003-7012.

Whitfield, C. (1995). Biosynthesis of lipopolysaccharide O antigens. T. I. M. 3(5): 178-185.

Whitfield, C., W. J. Keenleyside and B. R. Clarke (1994). Structure, function and synthesis of surface polysaccharides in *Escherichia coli*. In C.L. Gyles (Eds.), *Escherichia coli* in domestic animals and humans (pp. 437-494). Wallingford, Oxon, UK: CAB International.

Whitfield, C. and M. A. Valvano (1993). Biosynthesis and expression of cell-surface polysaccharides in Gram-negative bacteria. Advances in Microbial Physiology 35: 135-246.

Whitfield, C., E. R. Vimr, J. W. Costerton and F. A. Troy (1985). Membrane proteins correlated with expression of the polysialic acid capsule in *Escherichia coli* K1. J. Bact. 161(2): 743-749.

Winterburn, P. J. and C. F. Phelps (1972). The influence of substrates and modifiers on L-glutamine D-Fructose 6-phosphate amidotransferase. In S. Prusiner and E.R. Stadtman (Eds.), The enzymes of glutamine metabolism (pp. 343-360). New York: Academic Press.

Woodcock, D. M., P. J. Growther, J. Doherty, S. Jefferson, E. Decruz, M. Noyerweidner, S. S. Smith, M. Z. Michael and M. W. Graham (1989). Quantitative evaluation of *Escherichia coli* host strains for tolerance to cytosine methylation in plasmid-phage recombinants. Nucleic Acids Res. 17: 3469-3478.

Woude, v. d., M. W., B. A. Braaten and D. A. Low (1992). Evidence for a global regulatory control of pilus expression in *Escherichia coli* by Lrp and DNA methylation: model building based on analysis of *pap*. Mol. Micro. 6: 2429-2435.

Wozniak, D. J. (1994). Integration Host Factor and sequences downstream of the *Pseudomonas aeruginosa algD* transcription start site are required for expression. J. Bact. 176(16): 5068-5076.

Wunder, D. E., W. Aaronson, S. F. Hayes, J. M. Bliss and R. P. Silver (1994). Nucleotide sequence and mutational analysis of the gene encoding KpsD, a periplasmic protein involved in transport of polysialic acid in *Escherichia coli* K1. J. Bact. 176(13): 4025-4033.

Yamada, M. and M. H. Saier Jr. (1987). Physical and genetic characterization of the glucitol operon in *Escherichia coli*. J. Bact. **169**(7): 2990-2994.

Yamada, M. and M. H. Saier Jr. (1988). Positive and negative regulators for glucitol (*gut*) operon expression in *Escherichia coli*. J. Mol. Biol. **203**(569-583).

Yamada, M., Y. Yamada and M. H. Saier Jr. (1990). Nucleotide sequence and expression of the *gutQ* gene within the glucitol operon. J. DNA Seq. Map. **1**: 141-145.

Yannisch-Perron, C., J. Vieira and J. Messing (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13mp18 and pUC19 vectors. Gene **33**: 103-119.

Zapata, G., W. F. Vann, W. Aaronson, M. S. Lewis and M. Moos (1989). Sequence of the cloned *Escherichia coli* K1 CMP-*N*-acetyl neuraminic acid synthetase gene. J. Biol. Chem. **264**: 14769-14774.

Zuber, M., T. A. Hoover and D. L. Court (1995). Analysis of a *Coxiella burnetii* gene product that activates capsule synthesis in *Escherichi coli*: Requirement for the heat shock chaperone DnaK and the two-component regulator RcsC. J. Bact. **177**(15): 4238-4244.

Zwahlen, A., J. S. Kroll, L. G. Rubin and E. R. Moxon (1989). The molecular basis of pathogenicity in *Haemophilus influenzae*: comparative virulence of genetically related capsular transformants and correlation with changes at the capsulation locus *cap*. Microb. Pathogen. **7**: 225-235.