Molecular and Genetic Analysis of the *Escherichia coli* K5 Capsule Gene Cluster

C Pazzani

Thesis presented for the degree of Doctor of Philosophy

University of Leicester

1992

UMI Number: U555903

UMI

Dissertation Publishing

ProQuest

Abstract

Molecular and Genetic Analysis of the *Escherichia coli* K5 Capsule Gene Cluster

C Pazzani

Escherichia coli expressing group II capsules are associated with extra-intestinal diseases. *E.coli* expressing group II capsules have elevated CMP-KDO synthetase activity, with both enzyme activity and polysaccharide biosynthesis being temperature dependent. The locus (*kps*) necessary for biosynthesis of group II polysaccharides is organized into three functional regions, two of which (I and III) are conserved among different *kps* gene clusters. Determination of the nucleotide sequence of region I, of the K5 capsule gene cluster, revealed five genes (*kps*E, *kps*D, *kps*U, *kps*C and *kps*S) possibly organized into a single transcriptional unit. One of the genes, *kps*U, encoded for a functional CMP-KDO synthetase and was 65.5% identical to *kds*B of non-encapsulated *E.coli*. Whilst high level CMP-KDO synthetase activity was not essential for K5 polysaccharide synthesis, it was important for encapsulation. Disruption of either *kps*C or *kps*S resulted in cytoplasmic polysaccharide, suggesting a role for their products in an early stage of capsule biosynthesis. Whilst disruption of *kps*E and *kps*D generated periplasmic polysaccharide indicating a role for their products in the final stages of polysaccharide export. KpsE, KpsC and KpsS were homologous to proteins encoded by the capsule gene clusters of *Haemophilus influenzae* and *Neisseria meningitis*, suggesting a common functionality in the expression of capsules between these bacteria. Analysis of the region II-nucleotide sequence revealed four genes which are required for polysaccharide synthesis. This region II-sequence had a low GC content, atypical for the average GC ratio of *E.coli* DNA. The 3' end of *kps*S and *kps*T genes also had a low GC content. Since KpsS and KpsT encoded by different *kps* gene clusters had variable C-termini, it is possible that region II of the K5 capsule gene cluster might have been acquired by recombination events occured within *kps*S and *kps*T.

Statement

The accompanying thesis submitted for the degree of
Doctor of Philosophy entitled "Molecular and Genetic
Analysis of the *Escherichia coli* K5 Capsule Gene Cluster"
is based on work conducted by the author in the
Department of Microbiology of the University of Leicester
during the period between July 1988 and September 1991.

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

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

Signed............................

Date..............................

AKNOWLEDGEMENTS

INDEX

LIST OF ABBREVIATIONS

AP	Alternaternative pathway of complement
bp	basepair
CPS	capsular polysaccharide
CTP	Cytidine tri-phosphate
EDTA	Ethylenediamine-tetra acetic acid
GlcA	Glucuronic acid
GlcNAc	N-acetylglucosamine
IPTG	Isopropylthiogalactoside
Kb	Kilobase
KDa	Kilo Dalton
KDO	2-cheto-3-deoxymanno-octonic acid
LPS	lipopolysaccharide
ManNAc	N-acetylmannosaminuronic acid
min	Minute
MW	Molecular weight
NeuNAc	Sialic acid
orf	Open reading frame
PA	Phosphatidic acid
PEG	Polyethylene glycol
SD	Shine Dalgarno
SDS	Sodium lauryl sulphate
STC	Sialyltransferase complex
TE	Tris-EDTA (10mM Tris-HCl; 1mM EDTA pH8.0)
TEMED	N,N,N',N'-tetramethylethylenediamine
UDP	Uridine di-phopshate
X-Gal	5-Br-Cl-3-indolyl-D-galactoside

CHAPTER 1

INTRODUCTION

Encapsulated bacteria represent a broad range of medically important microbes causing invasive diseases such as septicaemia, meningitis, pneumonia, osteomyelitis, septic arthritis and urinary tract infections. The presence of a surface structure termed the capsule affects bacterial virulence, since removal of capsule by a variety of means results in bacteria which are no longer able to cause disease in experimental models of infection. Encapsulated micro-organisms are generally able to evade a number of different components of host defence, such as complement-mediated bacteriolysis, uptake and killing by phagocytes. For these reasons, the capsule is an important bacterial virulence factor and in recent years its study has greatly intrigued medical and industrial research, such as the production of vaccines. Also, capsule biosynthesis is a cellular process still poorly understood and its study represents an engaging challenge from a genetic, biochemical and structural point of view. In fact very little is known about how capsules are encoded, synthesized, exported and expressed on the bacterial cell surface. This thesis investigates mechanisms involved in the biosynthesis of the K5 capsule of *Escherichia coli*, which represents one of the main micro-organisms associated with childhood pyelonephritis and neonatal septicaemia.

1.1 BACTERIAL CAPSULES

1.1.1 GENERAL FEATURES

Capsules confer on an organism functional characteristics which play a key role in the pathogenesis of infectious disease of animals, plants and insects (Sutherland, 1977). The medical importance of encapsulated bacteria has focused studies on those micro-organisms responsible for human diseases. Encapsulated bacteria of medical importance are numerous and include *Streptococcus pneumoniae, Klebsiella spp., Haemophilus influenzae, Neisseria meningitidis* and *Escherichia coli*. For example, *N.meningitidis* is one of the major aetiological agents in septicaemia and meningitis. Meningococcal meningitis represents a worldwide problem with a mortality as high as 85% in untreated cases and 10% in those with appropriate antimicrobial therapy (Peltola, 1983; Gotschlich *et al.*, 1984). *N.meningitidis* can produce nine distinct capsular types of which those belonging to groups A, B and C are responsible for 90% of cases of meningococcal meningitis (Jennings, 1990). *H.influenzae* is also an important human pathogen and can be distinguished serologically into six different capsular types, termed a to f (Pittman, 1931). Strains expressing the type b capsule represent the commonest cause of non epidemic bacterial meningitis in children with a mortality of approximately 5%-10% (Turk, 1982; Jennings, 1990). In North America, infections caused by pneumococci are among the major causes of mortality. Pneumococci are mainly responsible for lower respiratory tract infections and in children they are the commonest cause of otitis media (Jennings, 1990). *E.coli* produces over 70 chemically distinct capsules, which are the basis of the K serotypes and have been classified into three groups (Orskov *et al.*, 1977; Orskov and Orskov, 1984;

2

Jann and Jann, 1990). Several studies have shown a correlation between specific K serotypes and certain human diseases. Generally, group I and II capsular polysaccharides are expressed by commensal *E.coli* of the gastero-intestinal tract. However group I capsules have been identified infrequently from *E.coli* associated with extra-intestinal diseases (Jann and Jann, 1983), whereas group II are characteristically found among *E.coli* strains isolated from extra-intestinal diseases (Orskov *et al.*, 1984). Additionally, serologic typing of clinical isolates, from all sites except blood, revealed that only a small number of K serotypes accounted for most of the extra-intestinal infections and that K1 and K5 were the predominant capsular serotypes (Cross, 1990).

Capsules are composed of either homo- or heteropolysaccharides, which can be linear or branched macromolecules that are located on the outer membrane of Gram-negative cells or the peptidoglycan layer of Gram-positive cells (Moxon and Kroll, 1990). Polysaccharides are composed of repeating units of neutral sugars, polyols, uronic acids and amino sugars. In addition, some polysaccharides contain non sugar substitutions including phosphate, formate, pyruvate and succinate (Sutherland 1977; Costerton *et al.*, 1981). This multitude of components can be linked to the broad range of different capsules expressed by *E.coli* (Kenne and Lindberg, 1983).

Capsules confer many properties on bacteria such as protection against dessication, the free passage of molecules and ions to the bacterial cell envelope and cytoplasmic membrane (Dudman, 1977), adherance to the surface of inanimate objects or living cells and the formation of biofilms and microcolonies (Costerton *et al.*, 1981). Additionally, capsules are thought to play a key

3

role in the invasion of the host by encapsulated bacteria (Robbins *et al.*, 1980).

1.1.2 ROLE IN VIRULENCE

In general, capsules protect bacteria against the action of phagocytic cells by impeding the binding of antibody directed toward cell surface components (Howard and Glynn, 1971; Horwitz and Silverstein, 1980). The cell surface of an encapsulated bacterium is hydrophilic with an overall negative charge. Such a feature contributes to general antiphagocytic properties (Moxon and Kroll, 1990). The hydrophilic surface is thought to act by reducing the surface tension at the interface between the phagocytic cells and the bacterium (van Oss and Gillman, 1973), resulting in an impairment of the phagocytic activity by mutual repulsion of charge (Moxon and Kroll, 1990). In addition to phagocytic cell evasion, capsules confer protection against bacteriolysis and opsonization mediated by activation of the alternative pathway of complement (AP), (Howard and Glynn, 1971; Joiner *et al.*, 1984).

It is known from studies in animals that antibodies, raised against the capsule, protect against disease caused by encapsulated bacteria. During the early stages of infection by encapsulated bacteria specific anti-capsule antibodies are often absent. In these circumstances, AP activation represents the major host defence mechanism (Moxon and Kroll, 1990). The alternative pathway initially involves the third component of complement (C3) and the deposition of the C3b fraction on the bacterial surface. This deposition promotes three potential immune functions (Moxon and Kroll, 1990). First, acting as a possible ligand for specific receptors on polymorphonuclear leucocytes or

4

macrophages. Second, by catalysing the complement cascade and generating membrane attack complexes (MACs). Such complexes, composed of C5b, C6, C7, C8 and multiple molecules of C9, result in membrane-pores that lead to the lysis of Gram-negative bacteria but, due to their thicker cell wall, not Gram-positive bacteria (Joiner *et al.*, 1984). Third, interaction of C3b with specific lymphocyte receptors promotes immunoglobulin secretion, lymphokine production and converts lymphocytes into memory cells.

In the initial stages of an infection C3b is first deposited on the bacterial surface. Then, it reacts with factor B to produce the complex C3bBb, which acts as a C3 convertase with the consequent amplification of C3b, a key step in AP activation. During this process a serum protein, termed factor H, acts as a regulator and competes with factor B for surface-bound C3b. Factor H, which has no enzymatic activity, is thought to modify the conformation of C3b. Such a modification facilitates the degradation of C3b by a second regulatory protein, called factor I. In this way both C3b amplification and deposition on the bacterial surface are reduced (Cross, 1990).

Some encapsulated bacteria are able to exploit the competition between factor B and H to their advantage. This has been demonstrated in the case of *S.agalactiae* type 3 and *E.coli* K1. Capsules produced by these micro-organisms contain sialic acid which increases the affinity of factor H for cell-bound C3b. Thus, although complement can bind to the capsule the C3 convertase complex (C3bBb), which should initiate the activation of the AP, is inhibited and the complex becomes detached from the cell surface (Cross, 1990). Group B meningococci also produce a sialic acid capsule and progressive enzymatic removal of meningoccocal surface sialic acid results in a proportional increase in

the deposition of factor B (Jarvis and Vedros, 1987). Group B meningococci from which 88% of the sialic acid had been removed showed susceptibility to bacteriolysis, which was comparable to that observed with naturally occurring sialic acid deficient mutants of meningococci (Jarvis and Vedros, 1987). Other organisms that contain sialic acid on their surface probably avoid opsonization in a similar manner (Hirsch et al., 1981). However, even capsules devoid of sialic acid, such as type 7 and 12 produced by *S.pneumoniae*, also appear to restrict AP activation by decreasing the binding affinity for factor B and increasing that for factor H. Thus, features of capsules other than chemical composition, must play an important role in bacterial evasion of complement-mediated bacteriolysis and opsonophagocytosis. Finally, not all encapsulated bacteria are able to avoid the activation of the AP. Encapsulated *E.coli* serotypes K6, K7, K27, K30, K42, K53, K57 and K75 are members of the normal human gut flora and activate efficiently the AP, whereas those isolated from extra-intestinal infections (serotypes K1, K3, K5, K12 and K92) are poor activators of this pathway (Stevens et al., 1983). Similar data have been produced from studies on different pneumococcal capsular types (Fine, 1975; Winkelstein et al., 1976), where the ability to avoid AP-mediated opsonophagocytosis was related to the ability to cause invasive disease. For example, *S.pneumoniae* type 1, which is the most common serotype isolated from adults with bacteremic pneumococcal pneumonia (Austrian and Gold, 1964), does not activate the AP. In contrast, *S.pneumoniae* type 25, which is infrequently isolated, efficiently activates this pathway (Fine, 1975).

The K1 and K5 capsules of *E.coli*, in addition to restricting AP activation, are also poorly immunogenic. Each of these capsular polysaccharides is similar to particular mammalian structures. The repeating unit of K1

is found in glycoproteins of mammalian cells (Finne *et al.*, 1983a; 1983b; Vimr *et al.*, 1984), whereas that of K5 is identical to the first polymeric intermediate in the biosynthesis of mammalian heparin (Prikas *et al.*, 1980). Thus, the poor immunogenicity of K1 and K5 capsules could be due to mimicry of certain host molecules and the consequent tolerance for the K1 and K5 capsule types. This tolerance might explain why K1 and K5 encapsulated *E.coli* are the most common serotypes found in extra-intestinal invasive diseases caused by *E.coli* (Vahlne, 1945; Cross, 1990). In summary, capsules undoubtedly are important virulence factors, however not all encapsulated bacteria are associated with invasive diseases. This suggests that the capsule, despite protecting bacteria against the host immune system, is not sufficient to confer virulence on encapsulated micro-organisms. Consequently, it is thought that other factors play a key role in the host-bacterium relationship during the course of an invasive disease (Robbins *et al.*, 1980).

1.2   THE K POLYSACCHARIDES OF
ESCHERICHIA COLI : CLASSIFICATION

E.coli produces more than 70 chemically and serologically distinct capsular polysaccharides (Orskov *et al.*, 1977). Individual isolates produce only one of these polymers, expression of which is stable and switching of capular type has not been documented (Boulnois and Roberts, 1990). The capsular polysaccharides of *E.coli* have been classified on the basis of their physical, chemical and biochemical features into three formations, termed group I, II and I/II (Jann and Jann, 1983; 1985; 1990; Finke *et al.*, 1990) (Table 1.1).

TABLE 1.1   CLASSIFICATION OF CAPSULAR POLYSACCHARIDES OF
E.coli

Property	Capsular polysaccharide group		
	I	II	I/II
Acid component	GlcA GalA Pyruvate	GlcA GalA KDO ManNAc Phosphate	GlcA GalA KDO ManNAc Phosphate
Expressed below 20°C	Yes	No	Yes
Lipid at the reducing end	Core-lipid A§	PA	PA
Chromosomal locus (close to)	(*his*) (*trp*)	*kps* (*serA*)	(*serA*)
High levels of CMP-KDO synthetase	No	Yes	No
Common features with	*Klebsiella* spp.	*H.influenzae* *N.meningitis*	NS
Co-expressed with	08,09,020	Many O-antigens	NS

TABLE 1.1 § this substitution has been verified for only a few polysaccharides. Abbreviations: GlcA, glucuronic acid; GalA, galacturonic acid; ManNAc, N-acetylmannosaminuronic acid; KDO, 2-cheto-3-deoxymanno-octonic-acid; PA, phosphatidic acid; NS, not stated.

1.2.1 GROUP I CAPSULES

The group I capsules expressed by *E.coli* are similar to those produced by *Klebsiella* spp. Group I capsular polysaccharides are composed mainly of glucuronic acid (GlcA), galacturonic acid (GalA) and their ketosidic-pyruvate substitutions. Such hexuronic acids are organized into macromolecules with repeating units ranging from tri- to hexasaccharides. Group I polysaccharides terminate at their reducing end directly in core-lipid A (Orskov *et al.*, 1977; Jann and Jann, 1990). Lipid A is also a component of lipopolysaccharide (LPS), which is an integral constituent of the outer membrane and represents a common feature among the Enterobacteriaceae. About 160 chemically different O-antigen side chains have been described (Orskov *et al.*, 1977). However, *E.coli* strains co-express group I polysaccharides only with certain O-antigens, namely O8, O9 and O20 (Jann and Jann, 1983; 1990).

A study of a large number of the different capsular polysaccharides, produced by *E.coli*, revealed that *E.coli* strains grown at 18°C were able to express only some of these polysaccharides, while when these strains were grown at 37°C all polysaccharides were expressed (Orskov *et al.*, 1984). *E.coli* strains expressing group I capsules do not show temperature-dependent polysaccharide biosynthesis (Jann and Jann, 1990). Two distinct genetic loci have been reported to be involved in group I polysaccharide biosynthesis (Schmidt *et al.*, 1977). The first is close to the *his* gene and is thought to represent the main genetic locus necessary for the biosynthesis of these polysaccharides. The second locus is close to the *trp* operon and its involvement in group I polysaccharide biosynthesis has been suggested to be connected in an unknown way with the glucuronic acid biosynthetic pathway

(Schmidt *et al.*, 1977). Glucuronic acid is a precursor of galacturonic acid, together with pyruvate they represent the major acid components of group I polysaccharides (Jann and Jann, 1990).

1.2.2 GROUP II CAPSULES

The group II capsules expressed by *E.coli* share similar features with those produced by *H.influenzae* and *N.meningitidis*.

A locus termed *kps*A, which maps near *ser*A, has been reported to be required for group II polysaccharide biosynthesis (Orskov and Nyman, 1974; Orskov *et al.*, 1977). Group II polysaccharides are mainly composed of N-acetylneuraminic acid (NeuNAc, also termed sialic acid), 2-cheto-3-deoxymanno-octonic acid (KDO), N-acetylmannosaminuronic acid (ManNAc), or phosphate, in repeating units of mono-, di- or trisaccharides (Jann and Jann, 1990). Group II polysaccharides terminate at their reducing end with the phospholipid 1,2-diacyl-*sn*-3-glycerol phosphate also termed phosphatidic acid (PA) (Gotschlich *et al.*, 1981; Schmidt and Jann, 1982; Jann and Jann, 1990). A similar substitition is found in *H.influenzae* type B and *N.meningitidis* group A, B and C (Kuo *et al.*, 1985). Phospholipids are normal membrane components of Gram-negative bacteria (Nikaido and Nakae, 1979). On the basis of these results, it has been suggested that PA found at the reducing end of *E.coli* group II and other capsules might act as the anchor which secures polysaccharide to the outer membrane (Gotschlich *et al.*, 1981; Boulnois and Jann, 1989; Jann and Jann, 1990). However, the labile chemical nature of the phosphatidiester bond between PA and polysaccharide (Schimdt and Jann, 1982; Jann and Jann,

1990) argues against this theory and, perhaps, other unknown factors may contribute to the formation and stabilization of capsules. In contrast to group I capsules, *E.coli* strains co-express group II capsules with a broad range of O-antigens and polysaccharide biosynthesis occurs only at growth temperatures higher than 20°C (Jann and Jann, 1990). Additionally, these strains contained a high level of CMP-KDO synthetase activity (Finke *et al.*, 1989). This enzyme, which is encoded by the *kds*B gene (Goldman and Kohlbrenner, 1985), is normally present at low level in non-encapsulated *E.coli* and is involved in the activation of KDO to CMP-KDO during LPS biosynthesis (Ghalambor and Heath, 1966). Interestingly, the high level of CMP-KDO synthetase, like group II polysaccharide biosynthesis itself, was evident only at bacterial growth temperatures above 20°C, whereas the enzyme activity was low when bacteria were grown at capsule restrictive temperatures (below 20°C) (Finke *et al.*, 1990). The high CMP-KDO synthetase activity and its temperature dependence suggest that this enzyme is a possible specific requirement for those polysaccharides with temperature-dependent biosynthesis. Consequently, the understanding of this high enzymatic activity has represented a key aim of this thesis. A recent study has shown that KDO is the reducing sugar of the K5 polysaccharide, despite not being a constituent of its repeating unit (Finke *et al.*, 1991). Taken together, these results have led to the working hypothesis that maybe all group II polysaccharides contain KDO at their reducing end. If this is true, KDO incorporation could represent the initial step in group II polysaccharide polymerization (Finke *et al.*, 1989), since polysaccharide polymerization is thought to occur from the reducing to the non-reducing end (Robbins *et al.*, 1967; Troy, 1979; Rohr and Troy, 1980). The high CMP-KDO synthetase activity could be interpreted as a specific requirement for group II polysaccharide biosynthesis. Additionally, the common temperature dependence of both

enzymatic activity and group II capsule biosynthesis
suggest that the high levels of this enzyme could represent
the rate-limiting step in KDO incorporation during
initiation of polymerization and by which group II
polysaccharides biosynthesis can be regulated. This
hypothesis is also strengthened by the assumption that CMP-
KDO synthetase activity represents the rate-limiting step
in KDO incorporation during LPS biosynthesis (Ray *et al.*,
1983). In part this thesis investigates why strains
expressing group II capsules contain high CMP-KDO
synthetase activity and if this activity represents a
regulatory step for group II capsule biosynthesis.

1.2.3 GROUP I/II CAPSULES

Recently the existence of a third group of capsules,
tentatively referred to as group I/II, has been postulated
by Finke *et al.*, (1990). These polysaccharides (K3, K10,
K11, K19 and K54) share common chemical and biochemical
features with those of both group I and II capsules. Group
I/II polysaccharides terminate at their reducing end with
PA, their biosynthesis is not temperature-dependent and
they are chemically similar to polysaccharides of group II.
Additionally, *E.coli* expressing group I/II capsules do not
contain high levels of CMP-KDO synthetase. Finally, like
group II polysaccharides, the genetic locus for group I/II
polysaccharide biosynthesis (such as for K10 and K54) maps
near *ser*A (Orskov and Nyman, 1974). The term *kps*A was
proposed for all K antigen genes mapping near *ser*A, raising
the possibility that some of these genes might be involved
in polysaccharide biosynthesis of both groups II and I/II.
However, recent studies (Drake *et al.*; Pearce *et al.*;
unpublished results) have shown that group I/II and II
capsule genes, despite sharing a common chromosomal
location, are not homologous. In this thesis, the term *kps*

12

will be used only for those gene clusters which encode
group II capsules.

CHAPTER 2

THE MOLECULAR GENETIC ANALYSIS OF CAPSULE PRODUCTION IN BACTERIA

2.1 INTRODUCTION

Epidemiological studies of *E.coli* producing group II polysaccharides revealed a clear relationship between these strains and serious human diseases. For example, *E.coli* producing the K1, K2, K3, K5, K12, K13, K20 and K51 group II capsules are the main isolates associated with childhood pyelonephritis and neonatal septicaemia (Kaijser *et al.*, 1977). Additionally, serological studies have shown that 80% of strains recovered from cases of *E.coli* neonatal meningitis express the K1 capsule (Robbins *et al.*, 1974). Group II capsules share similar features with those produced by *H.influenzae* and *N.meningitidis*. In addition, some group II capsules are serologically related to certain polysaccharides of *N.meningitidis* or *H.influenzae*. For example, antibodies raised against the group B capsule of *N.meningitidis* cross-react with K1 polysaccharide (Grados and Ewing, 1970). Similarly, antisera raised against capsule of *N.meningitidis* group A and group C, and *H.influenzae* type b cross-react with *E.coli* K93 (Guirguis *et al.*, 1985), K92 (Robbins *et al.*, 1973; 1975) and K100 (Schneerson *et al.*, 1972; Robbins *et al.*, 1973), respectively. This cross-reactivity is thought to be due to the possibility that identical or chemically related polysaccharides expose common epitopes. The immunological cross-reactivity between capsules produced by *N.meningitidis* group B and *E.coli* K1 was exploited in the

early 1980's for cloning of the K1 gene cluster (Silver *et al.*, 1981; Echarti *et al.*, 1983). At this time, information about the *kps* gene cluster was poor. The cloning of the K1 gene cluster represented the starting point of the genetic analysis of group II polysaccharide production and many studies carried out later, including those in this thesis, can be seen as the continuation of that initial work. Additionally, the inter-species cross-reactivity between capsules may be further exploited, for example some group II polysaccharides may represent key targets for the development of vaccines with a broad bacterial range.

2.2 ORGANIZATION OF *E.COLI* GROUP II CAPSULE GENE CLUSTERS

Laboratory strains of *E.coli* K12 are non-encapsulated and do not carry the group II capsule genes on their chromosome (Echarti *et al.*, 1983). These strains allowed the cloning and molecular analysis of several group II capsule gene clusters.

Cosmid libraries of chromosomal DNA from wild type strains identified contiguous DNA fragments which encode the biochemical functions necessary for biosynthesis and export of different group II capsules, such as K1 (Silver *et al.*, 1981; Echarti *et al.*, 1983), K5, K7, K12 and K92 (Roberts *et al.*, 1986), and K4 (Drake *et al.*, 1990). *E.coli* LE392 carrying the K1, K5, K7 and K92 gene clusters produced capsular polysaccharides which were immunologically indistinguishable from those produced by wild type strains. For example, the polysaccharide extracted from LE392 carrying the K1 gene cluster cross-reacted, like wild type K1, with antibodies raised against *N.meningitidis* group B (Silver *et al.*, 1981; Boulnois and Roberts, 1990). Similarly, polysaccharides extracted from LE392 carrying

the K7 or K92 gene clusters cross-reacted with antisera raised against the capsule of *S.pneumoniae* type III (Robbins *et al.*, 1972) and *N.meningitidis* group C (Glode *et al.*, 1977), respectively. These data suggested that the information necessary for polymerization and export of group II polysaccharides was encoded by contiguous DNA fragments and confirmed the hypothesis that a single genetic locus, previously termed *kps*A, was involved in group II capsule biosynthesis.

Capsule expression on laboratory strains harbouring different cosmids differed from that of clinical strains. *E.coli* LE392 carrying the K1 or K5 gene clusters were sensitive to K1 and K5 bacteriophages, respectively. However, plaques produced on wild type strains were clear, whereas those produced on recombinant strains were smaller and turbid (Roberts *et al.*, 1986). Additionally, immunoelectron microscopy showed that laboratory strains of *E.coli* carrying the K5 gene cluster expressed the K5 capsule only on 25% of cells, whereas all cells of the wild type strain produced capsules (Kroncke *et al*, 1990a). A similar phenomenon was observed for the cloned *N.meningitidis* group B capsule genes expressed in *E.coli*. *E.coli* LE392 harbouring the group B meningococci *cps* gene cluster (more details about the *cps* locus will be given later) expressed group B meningococcal capsule only on 50% of cells (Frosch *et al.*, 1989). The reduced level of capsule expression might explain the different plaque morphologies observed between wild type and laboratory strains carrying the respective gene cluster. Turbid plaques could result from a mixed population of encapsulated and non-encapsulated bacteria. Encapsulated bacteria can be recognized and lysed by the respective bacteriophage, whereas the remaining non-encapsulated cells grow normally to cause turbidity. The reason why not all cells of laboratory strains carrying different capsule gene

clusters express capsule on their surface is unclear. A possible explanation is the requirement of additional genetic information present in the wild type strains but missing in the various cosmid clones. However, in the absence of more exhaustive information one has to be very cautious in interpreting these data.

Molecular studies on the K1 gene cluster revealed that approximately 17 Kb of DNA, split into three functional regions, is involved in capsule biosynthesis (Echarti *et al.*, 1983; Silver *et al.*, 1984; Boulnois *et al.*, 1987) (Figure 2.1). A block of genes, termed region I, is required for polysaccharide translocation from the periplasm to the cell surface, since region I mutations resulted in intracellular polymer located in the periplasmic space. The second block of genes, termed region II, encodes the biochemical functions necessary for polysaccharide polymerization and mutations located in this region abolished polymer production. Additionally, some region II determinants, such as from K1, also encode proteins necessary for the biosynthesis of sugars which constitute the polysaccharide repeating unit. The third block, termed region III, is responsible for polysaccharide tansport from the cytoplasm to the periplasmic space, since region III mutations resulted in intracellular polymer with a cytoplasmic location. Subsequent studies showed that the organization of the K1 genes is shared by other group II gene clusters (Roberts *et al.*, 1986; 1988a; Roberts *et al.*; 1988b) (Figure 2.1). DNA probes from region I or III, of the K1 gene cluster, hybridized to DNA segments from other group II K antigen gene clusters, whereas region II DNA probes appeared to be capsule specific (Roberts *et al.* 1986; 1988a; Roberts *et al.*; 1988b). Additionally, LE392 strains carrying the K1 gene cluster deficient in either region I or III were able to express the K1 capsule when complemented with plasmids carrying the regions from other

K antigen gene clusters (Roberts *et al.*, 1986; 1988a).
These data revealed that regions I and III were conserved
DNA segments and functionally interchangeable between group
II capsule gene clusters. In particular, these results
demonstrated that polysaccharide polymerization is encoded
by a type-specific DNA region, whereas polysaccharide
export occurs regardless of the specific chemical
structure.

FIGURE 2.1   ORGANIZATION OF *E.coli* GROUP II CAPSULE GENE CLUSTERS
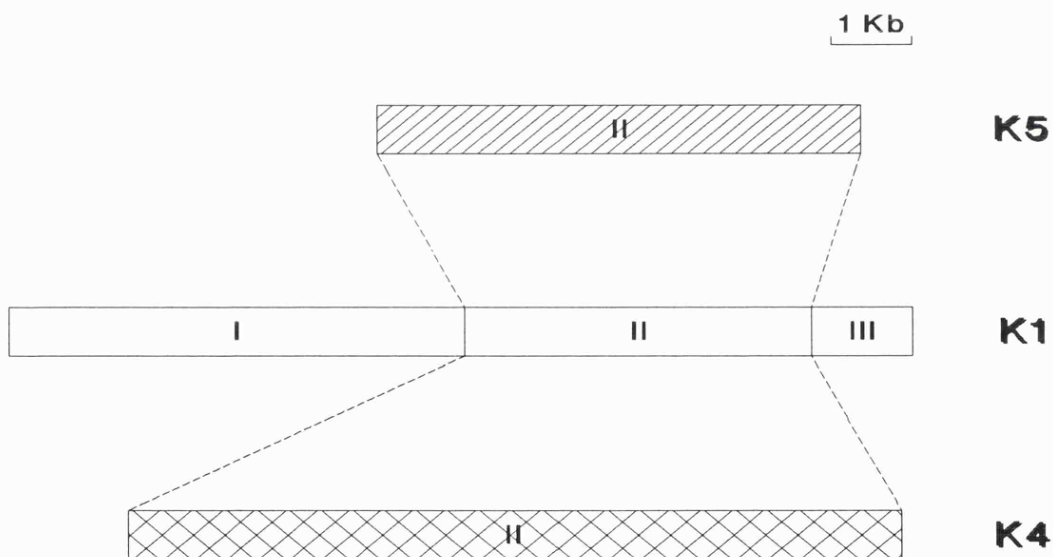
FIGURE 2.1   The organization of the K1, K4 and K5 capsular gene clusters is shown as boxes and the scale is shown in kilobase pairs. Boxes with diagonal and crossed lines refer to region II of the K5 and K4 capsular gene cluster, respectively. Broken lines highlight the difference in size of region II of the K4 and K5 capsular gene clusters.

On the basis of this common genetic organization Boulnois and Jann (1989) have proposed a "cassette-model" to explain the antigenic diversity of group II capsules. According to this theory, *E.coli* acquired during its early evolutionary history a *kps* gene cluster. Then, antigenic diversity might have occurred by acquisition and reorganization of a region II (also termed the region II cassette), perhaps through different mechanisms of recombination. For example, a resident region II could have been replaced by an incoming cassette through recombination events involving conserved DNA segments flanking the region II. Additionally, intra- and inter-cassette recombinations, point mutations and duplications could have produced hybrid cassettes that encoded novel polymers related to, but distinct from, the original.

2.3   THE MOLECULAR GENETIC ANALYSIS OF *HAEMOPHILUS* AND *NEISSERIA* CAPSULE GENE CLUSTERS

The *cap* locus of *H.influenzae* and the *cps* locus of *N.meningitidis*, which encode the biochemical functions necessary for capsule biosynthesis and export show a similar genetic organization with the *kps* locus of *E.coli* (Figure 2.2).

The *cap* locus of *H.influenzae* types a, b, c, and d, is composed of a central DNA fragment specific to each serotype (also termed region II), which is flanked by two common DNA segments (also termed regions I and III). As in the *E.coli kps* locus, the *cap* region II is thought to encode the biochemical functions necessary for polysaccharide polymerization, whereas regions I and III are thought to be required for polysaccharide export (Kroll *et al.*; 1989) (Figure 2.2).

FIGURE 2.2   ORGANIZATION OF THE CAPSULE GENE CLUSTERS OF
E.COLI, H.INFLUENZAE AND N.MENINGITIDIS
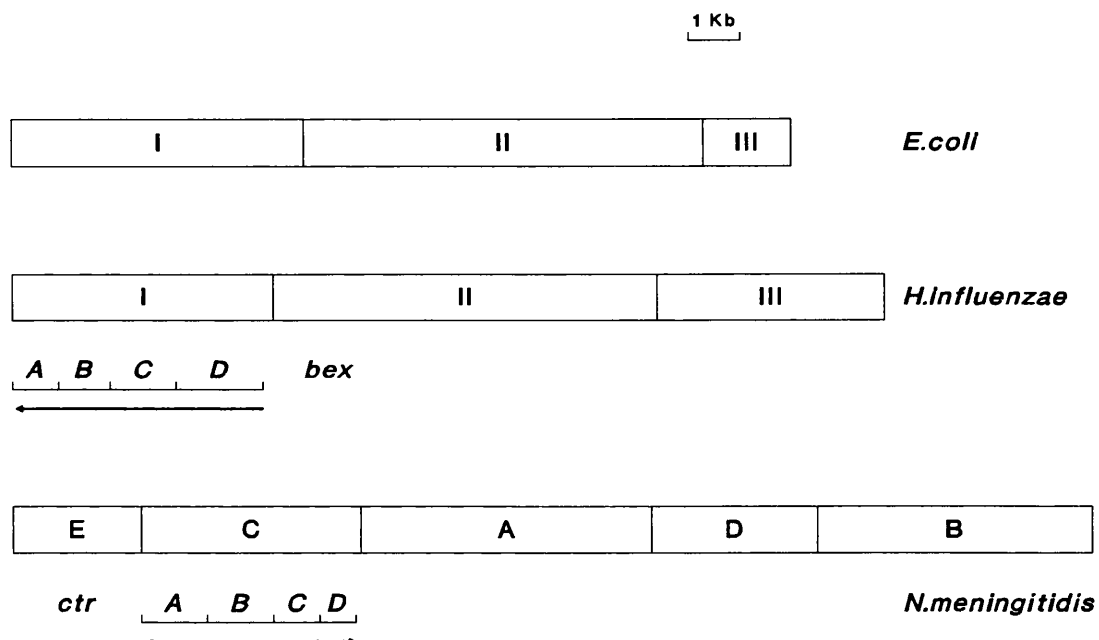
FIGURE 2.2 The organization of the *E.coli* K1, *H.influenzae* type b and *N.meningitidis* group B is shown as boxes and the scale is shown in kilobase pairs. The organization of region I of *H.influenzae* type b and region C of *N.meningitidis* group B is shown underneath boxes labelled I and C, respectively. The identified *orfs* are shown with horizontal lines and capital letters (A, B, C and D), with the side designation *bex* for *H.influenzae* and *ctr* for *N.meningitidis*, respectively. Arrows indicate the direction of transcription.

The region I of *H.influenzae* type b has been recently sequenced and contains four genes, termed *bex*D, *bex*C, *bex*B and *bex*A, possibly organized into a single transcriptional unit (Kroll *et al.*, 1990). This region is also thought to encode for an ATP-driven polysaccharide export system,

where BexA (24.7 kDa) and BexB (30.2 kDa) represent ATP-binding and inner membrane proteins, respectively. BexC (42.3 kDa) has been proposed to be a periplasmically orientated component of this export system, however the role of BexC and BexD (42 kDa) remains poorly defined.

The *cps* locus encodes the biochemical functions necessary for the biosynthesis of meningococcus group B capsule (Frosch et al., 1989) (Figure 2.2). This locus is composed of five regions, termed A to E. Region A is thought to encode proteins required for sugar biosynthesis and their polymerization. Regions B and C, are thought to be involved in polysaccharide export. *E.coli* strains harbouring a *cps* locus defective in region B resulted in polymer located in the cytoplasm, whereas strains harbouring a *cps* locus defective in region C resulted in a periplasmically located polymer. Deficiency in regions D and E caused higher levels of capsule polysaccharide production than wild type strains. The reason for this is as yet unclear, however it has been postulated that region D and E could play a regulatory role in the biosynthesis and/or transport of the polysaccharide itself (Frosch et al., 1989). Region C of *N.meningitidis* group B has been recently sequenced, and is composed of four genes, termed *ctrA*, *ctrB*, *ctrC* and *ctrD*, possibly organized into a single transcriptional unit (Frosch et al., 1991) (Figure 2.2). CtrA has been proposed to be an outer membrane protein with a predicted MW of 43.3 KDa and contains a putative cleavage site, characteristic of lipoproteins, which is recognized by signal peptidase II (Hussain et al., 1982). CtrA is thought to be involved in polysaccharide export across the outer membrane. CtrB and CtrC have been proposed to be cytoplasmic membrane proteins with calculated MWs of 41.9 and 30.1 KDa, respectively. CtrB and CtrC, like BexB and BexC, contain several hydrophobic regions which potentially span the inner membrane. CtrD has a calculated MW of 24.6 KDa and is

thought to be an ATP-binding protein. Computer analysis has revealed homology (higher than 60%) between proteins encoded by *N.meningitidis* region C and *H.influenzae* region I (Frosch *et al.*, 1991) (Table 2.1). Data described by this thesis and produced by A Smith in this laboratory, show that this protein homology is also shared with some of *E.coli* region III and I gene products, namely KpsM, KpsT and KpsE (for more details see section 2.4 and 4.2, respectively). The protein sequence homology, together with the similar genetic organization between the *kps*, *cps* and *cap* loci, has led to the hypothesis that encapsulation in *N.meningitidis*, *H.influenzae* and *E.coli*, could have a common evolutionary origin (Boulnois and Jann, 1989, Frosch *et al.*, 1991). Such a hypothesis is also supported by the observation of a low but detectable DNA homology, by southern blotting, between region C of the *cps* locus and DNA fragments belonging to *cap* locus of *Haemophilus spp.* and the *kps* locus of *E.coli* (Frosch *et al.*, 1991).

**TABLE 2.1   SEQUENCE SIMILARITY BETWEEN PROTEINS ENCODED BY
THE *CAP* LOCUS OF *H.INFLUENZAE* AND THE *CPS* LOCUS
OF *N.MENINGITIDIS*.**

N.meningitidis	*H.influenzae*	Sequence similarity	
		identity %	homology %
CtrA	BexD	50.8	66.8
CtrB	BexC	58.6	71.8
CtrC	BexB	68.7	82.2
CtrD	BexA	81.9	88.0

2.4   THE MOLECULAR GENETIC ANALYSIS OF *E.coli* GROUP II CAPSULE GENE CLUSTERS

Molecular studies on *E.coli* producing different group II capsules has led to the hypothesis that these polysaccharides are specifically polymerized in the cytoplasm and exported to the cell surface by a common transport system. In particular, this hypothesis implies the possibility to develop a new therapeutic approach by targeting this export system. A common export system, which functions regardless of chemical structure of polysaccharide, also represents a fascinating cellular

process. This thesis investigates the genetics of this export system and the biological functions of the encoded products. The three DNA regions which compose the *kps* locus will be described individually, starting from region III which was part of a study carried out by A Smith in this laboratory.

2.4.1 REGION III OF *E.coli* GROUP II CAPSULE GENE CLUSTERS

Region III is a DNA fragment of approximately 2 Kb, which encodes products required for surface expression of polysaccharide. These products are thought to function in a post polymerization process, since mutations in region III result in polymer either bound to the inner face of the cytoplasmic membrane (see section 6.1), or located in the cytoplasm (Boulnois and Jann, 1989). Taken together, these results suggest that region III-encoded products are likely to be involved in polysaccharide export across the cytoplasmic membrane. Analysis by immunoelectrophoresis and PAGE, has revealed that the inner membrane-bound polymer lacks phosphatidic acid and is shorter than wild type polysaccharide (Boulnois *et al.*, 1987; Kroncke *et al.*, 1990a). Whether KDO is present or not at the reducing end of these shorter polysaccharides remains unclear (Finke *et al.*, 1991).

Region III of the K1 and K5 gene clusters has been recently sequenced (Smith *et al.*, 1990; Pavelka *et al.*, 1991). It is composed of two genes, termed *kps*M and *kps*T, possibly arranged into a single transcriptional unit (Figure 2.3). *kps*M and *kps*T encode proteins with a predicted MW of 29 and 25 KDa respectively. KpsM is hydrophobic and computer analysis of its amino acid sequence has revealed the presence of at least six domains that could potentially

24

span the cytoplasmic membrane. Alkaline phosphatase positive Tn*pho*A insertions have identified, in KpsM, two periplasmic regions that are consistent with the hypothesis that KpsM may be a membrane spanning protein (Pavelka *et al.*, 1991). In contrast, KpsT is hydrophilic and contains sequences for a potential adenine nucleotide (ATP) binding domain.

FIGURE 2.3   REGION III OF THE K5 CAPSULAR GENE CLUSTER

I   II   III   K5   1 Kb

BG   BG

BG   kpsT   kpsM   BG   0.2 Kb
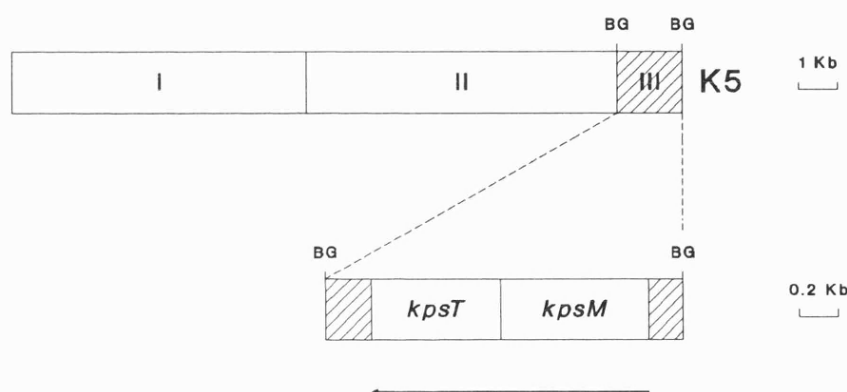
FIGURE 2.3 The organization of the K5 capsular gene cluster is shown as boxes and the scale is shown in kilobase pairs. Regions I and II are shown as empty boxes, whereas region III is shown as a box with diagonal lines. Region III is enlarged on scale 5:1 and broken lines highlight the enlargement. The organization of region III is shown in boxes labelled with *kps*M and *kps*T. Abbreviation : BG, *Bgl*II.

KpsT and KpsM are believed (Smith *et al.*, 1990) to function as dual components in a polysaccharide export system analogous to the periplasmic-binding-protein dependent

transport systems of certain Gram-negative bacteria (Higgins *et al.*, 1985). This kind of transport system consists of integral membrane proteins with a core structure of five or six membrane spanning α-helices separated by short hydrophilic regions (Higgins *et al.*, 1985). The hydrophilic regions which are periplasmically orientated, are thought to interact with substrate specific binding proteins located in the periplasm. Whereas, those domains cytoplasmically orientated are thought to interact with ATP-binding proteins which supply energy by coupling ATP-hydrolysis to the transport system (Higgins *et al.*, 1990a). These transport systems were generally considered only to play a role in the uptake of nutrients for cell growth. However, it is now becoming evident that many of these transporters have adapted to serve other functions. For example, evidence that such binding-protein-dependent transport systems could function to mediate export from cells, has already been reported for both prokaryotes and eukaryotes. The *E.coli* haemolysin (HlyA) is translocated directly from the cytoplasm to the outer membrane by a membrane bound system composed of two proteins (HlyB and HlyD), one of which (HlyB) is thought to bind ATP (Holland *et al.*, 1990). Similarly, the Mdr protein of mammalian cells is thought to function as an ATP-dependent transport system, expelling drugs from the cell (Gros *et al.*, 1986; Horio *et al.*, 1988). Thus, it has been proposed that KpsM might act as a "cytoplasmic membrane-pump" which, in conjunction with KpsT as an "energy-provider", mediates the transport of polysaccharide across the lipid bilayer from the cytoplasm to the periplasmic space (Smith *et al.*, 1990; Pavelka *et al.*, 1991). Additionally, site directed mutagenesis of the putative ATP-binding domain of KpsT resulted in *E.coli* unable to export capsular polysaccharide (Pavelka *et al.*, 1991). This latter result might reinforce, if KpsT binds ATP, the hypothesis that polysaccharide export occurs, in part, through an energy dependent system, where KpsT functions as an "energy-provider". However, this

proposed mechanism of export has still to be explored.

Polysaccharide export across the cytoplasmic membrane appears also to be dependent on membrane-potential, since the use of the membrane-potential decoupling agent carbonyl-cyanide-M-chlorophenyl hydrazone (CCCP) resulted in bacteria with cytoplasmic polymer (Kroncke et al., 1990b). In this experiment, CCCP was added to E.coli K1 and K5 prior to shifting growth conditions from the capsule restrictive temperature of 18°C to the permissive temperature of 37°C. Under these conditions, E.coli did not exhibit on its surface the K1 or K5 capsule. Such incubation resulted in a cytoplasmic polymer which lacked phosphatidic acid, was shorter than wild type and indistinguishable from that obtained in region III mutants. Taken together, these results suggest that polysaccharide export across the cytoplasmic membrane is a complex process, which might involve the co-participation of many factors such as the presence of membrane spanning proteins coupled with "energy-providers", and the requirement of an appropriate membrane-potential.

Computer analysis of the DNA sequence of K1 and K5 region III has confirmed previous experimental data (Roberts et al., 1986) that region III is a conserved DNA segment and a functionally interchangeable region among strains expressing different group II polysaccharides. $KpsM_{K1}$[1] and $KpsM_{K5}$[2] are both 258 amino acids in length with 97.6% identity over the entire protein. When conservative amino acid substitutions are taken into account, these proteins

---

[1] K1 = refers to the amino acid sequence encoded by the kpsM gene from the K1 gene cluster

[2] K5 = refers to the amino acid sequence encoded by the kpsM gene from the K5 gene cluster

are 99.2% homologous. The respective hydropathy plots displayed the same profile with the presence of six hydrophobic domains (Smith *et al.*, 1990; Pavelka *et al.*, 1991). $KpsT_{K1}$ and $KpsT_{K5}$ were 72% identical and 84% homologous. Additionally, the putative ATP binding sequences were perfectly conserved. Such data underline further the possibility that this sequence might represent an ATP binding domain of KpsT. Additionally, KpsT and KpsM proteins of *E.coli* showed significant homology with the BexA and BexB proteins of *H.influenzae*, respectively (Smith *et al.*, 1990; Kroll *et al.*, 1990; Pavelka *et al.*, 1991). KpsT is 45% homologous to BexA and in both proteins the putative ATP-binding consensus motif is conserved. KpsT showed only 26% homology with BexB, however the respective hydropathy plots exhibited a similar profile with six hydrophobic stretches (Kroll *et al.*, 1990).

In summary region III products might take part in polysaccharide export which is likely to be an energy-dependent process. Additionally, the extent of amino acid sequence homology identified between proteins encoded by capsule loci of *Escherichia*, *Haemophilus* and *Neisseria* has suggested that chemically different polysaccharides are exported through the cytoplasmic membrane via a common mechanism. Such a transport system seems to be a member of a superfamily, designated "ABC (ATP binding cassette) transporters" (Higgins *et al.*, 1990b; Hyde *et al.*, 1990), which consist of integral membrane proteins and ATP binding proteins.

2.4.2 REGION II OF *E.coli* GROUP II CAPSULE GENE CLUSTERS

Region II is capsule specific and a number of studies have suggested it encodes the biochemical functions necessary

for polysaccharide polymerization (Boulnois and Jann, 1989). Mutations located in region II of the K1 gene cluster result in strains which were defective in production of polysialic acid (Echarti *et al.*, 1983; Boulnois *et al.*, 1987). Region II has been reported to be variable in size, from as little as 5.8 Kb in the K1 gene cluster (Vimr *et al.*, 1989), and up to 14 Kb in the K4 gene cluster (Drake *et al.*, 1990). The size of region II correlates broadly with the chemical and structural complexity of the polysaccharide encoded by region II (Boulnois and Jann, 1989).

Most of the genetic and biochemical information about region II comes from studies carried out on the K1 capsular polysaccharide. The K1 polysaccharide is an α-2,8-linked homopolymer of NeuNAc (sialic acid) (Barry, 1959; DeWitt and Rowe, 1961) and its synthesis involves at least three steps (Silver *et al.*, 1984):

a) Biosynthesis of NeuNAc. This step has been proposed to occur through two sequential chemical reactions. Firstly, epimerization of N-acetylglucosamine (GlcNAc) or UDP-GlcNAc to N-acetylmannosamine (ManNAc). Secondly, condensation of ManNAc and pyruvate (or pyruvic acid) to NeuNAc

b) Activation of NeuNAc to CMP-NeuNAc by the enzyme CMP-NeuNAc synthetase.

c) Polymerization of NeuNAc by a sialyltransferase complex (STC)

FIGURE 2.4 REGION II OF THE K1 CAPSULAR GENE CLUSTER
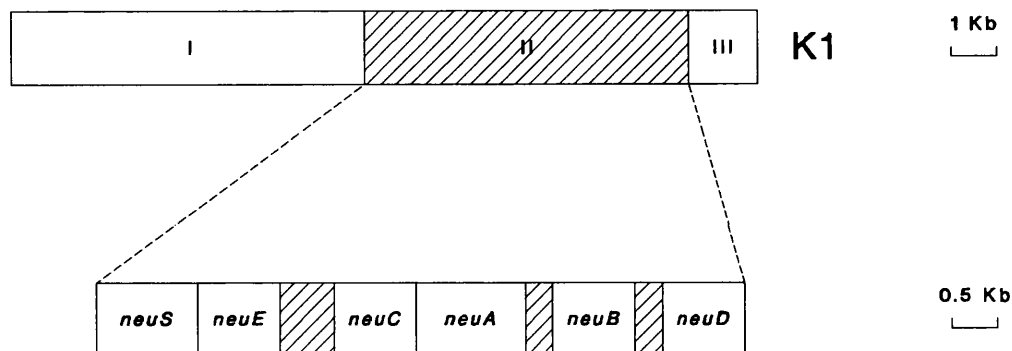
FIGURE 2.4 The organization of the K1 capsular gene cluster is shown as boxes and the scale is shown in kilobase pairs. Regions I and III are shown as empty boxes, whereas region II is shown as a box with diagonal lines. Region II is enlarged on scale 2:1 and broken lines highlight the enlargment. The organization of region II is shown as boxes labelled with *neu*A, B, C, D, E and S.

K1 region II contains at least six genes, termed *neu*A, B, C, D, E and S (Vimr *et al.*, 1992) (Figure 2.4). *neu*B and *neu*C encode proteins involved in the condensation reaction of ManNAc to NeuNAc (Vimr *et al.*, 1989). The possibility that more than one gene could take part in this condensation step has already been proposed by Vimr and Troy (1985). These authors have also suggested that another gene product is required for the epimerization of GlcNAc to ManNAc (Grosh and Roseman, 1965; Vimr and Troy, 1985), but to date this has not been identified. Mutation in *neu*C resulted in strains lacking the K1 polymer. However, these strains expressed the K1 capsule if complemented with exogenous NeuNAc, which demonstrated the direct involvement

of *neuC* in NeuNAc biosynthesis (Silver *et al.*, 1984). The fuction of the *neuD* encoded product is as yet unknown (Vimr *et al.*, 1989). *neuA*, encodes for the enzyme CMP-NeuNAc synthetase which catalyzes the activation of NeuNAc to CMP-NeuNAc. CMP-NeuNAc synthetase has been purified, and both MW (50 KDa) and NH_2-terminal sequence agree with the protein sequence predicted from the sequence of the *neuA* gene (Vann *et al.*, 1987). This enzyme was demonstrated to be reversibly inactivated at low temperatures (Merker and Troy, 1990). CMP-NeuNAc synthetase is synthesized but inactive at 15°C, whereas after temperature shifting to 37°C the enzyme activity is completely restored. These data have led to the hypothesis that the temperature inactivation of CMP-NeuNAc synthetase represents the main regulatory mechanism of the K1 capsule biosynthesis (Merker and Troy, 1990). However, NeuNAc is not present in all repeating units of group II polysaccharides, whereas the biosynthesis of this group is uniformly regulated by temperature. This suggests that other unknown factor(s) play a more common regulatory role in the biosynthesis of all group II capsules. *neuE* was very recently assigned to the K1 region II and it encodes a protein of at least 26 KDa (Steenbergen *et al.*, 1992). Computer analysis of the *neuE* sequence has revealed a putative membrane-spanning domain near the C-terminus which includes the presence of a sequence similar to the potential polyprenyl-binding motif (Steenbergen *et al.*, 1992). This motif has also been identified in three glycosyltransferases of *Saccharomices cerevisiae* which are involved in transferring of sugars from nucleotide sugars to dolichol phosphate derivatives [Dol-PP-$(GlcNAc)_2$-$(Man)_9$-$(Glc)_3$] (Albright *et al.*, 1989). A similar consensus motif has also been found in another yeast protein, termed SEC59, which is also thought to participate in assembly of the lipid-linked precursor oligosaccharide [Dol-PP-$(GlcNAc)_2$-$(Man)_9$-$(Glc)_3$] (Albright *et al.*, 1989). The polyprenyl-binding motif was identified in putative membrane-spanning domains and it has been

proposed to recognize specifically the isoprenoide region of dolichol phosphate. Additionally, this recognition is thought to occur in the membrane, since dolichol is very hydrophobic (Albright *et al.*, 1989). K1 polysaccharide biosynthesis requires the presence of a lipid intermediate (Troy *et al.*, 1975), which has been identified as undecaprenyl phosphate (Troy *et al.*, 1979). The presence of a sequence similar to the potential polyprenyl-binding motif and its location in a putative membrane-spanning domain has led to the hypothesis that NeuE might take part in initiation of K1 polysaccharide polymerization by transferring of one or more sialyl residues to undecaprenyl phosphate as a membrane lipid acceptor of the K1 polymer (Steenbergen *et al.*, 1992). *neuS*, encodes for the sialic acid polymerase (also termed sialyl transferase), which transfers K1 repeating units from the non reducing end to the growing polysaccharide chain. This enzyme is also thought to be a component of a membrane protein complex (STC) involved in K1 polysaccharide polymerization. STC might also include some region I gene products (KpsC and KpsS) in conjunction with KpsM and KpsT products of region III (Vimr *et al.*, 1989), since mutations located in *kps*T, *kps*C and *kps*S greatly diminished the endogenous sialyltransferase activity *in vitro*. This result suggests that products thought to take part in the process of polysaccharide export might also play an unknown role in polymer polymerization (Vimr *et al.*, 1989).

In a theoretical model for K1 polymerization (Troy and McCloskey, 1979) NeuNAc is added first to the lipid carrier undecaprenol phosphate (initiation) which is thought to function as a carrier of sialosyl residues in polymer assembly. Then, NeuNAc monomers are transferred sequentially from their CMP-NeuNAc activated form to the non-reducing end of the growing polysaccharide chain (elongation) (Troy *et al.*, 1975). Alternatively, the

undecaprenol phosphate could act as an intermediate "swinging arm" donor of sialosyl residues to another endogenous acceptor (Ortiz *et al.*, 1989). For example, Weisgerber and Troy (1990) have recently proposed that the K1 endogenous acceptor is a membrane protein of about 20 KDa. *E.coli* grown at 15°C did not express both polysialic acid and a membrane protein of 20 KDa (Troy and McCloskey, 1979; Whitfield and Troy, 1984), whereas a membrane protein of 20 KDa was expressed and found linked to polysialic acid at the capsule permissive temperature of 37°C. Additionally, this complex apparently does not contain PA or phospholipid (Weisgerber and Troy, 1990), suggesting that undecaprenyl phosphate does not represent the only membrane acceptor in polysialic acid biosynthesis. Polymerization of polysialic acid is thought to occur at the inner face of the cytoplasmic membrane (Janas and Troy, 1989) and requires protein synthesis (Whitfield and Troy, 1984; Whitfield *et al.*, 1984). Thus, it has been proposed that the 20 KDa protein might function as a membrane-located endogenous acceptor, to which monosialic acid residues or polysialic acid chains are transferred during polymerization, and promote translocation of the K1 polymer across the inner membrane (Weisgerber and Troy, 1990).

Despite the identification of a potential polyprenyl-binding motif, there is no experimental evidence that NeuE promotes the initiation of K1 polysaccharide polymerization. In the absence of more exhaustive data, it is still possible that NeuS might be multifunctional and catalize both initiaton and elongation reactions (Steenbergen and Vimr, 1990). This conclusion is based on the observation that NeuS will not elongate exogenous acceptors smaller then a NeuNAc-trimer (Vimr *et al.*,1986). Additionally, NeuS could also recognize a KDO acceptor for initiation and then elongate on the first sialosyl residue within the same catalytic site (Steenbergen and Vimr,

1990). This recognition might reflect the similar chemical structure of NeuNAc and KDO, and would also fit with the hypothesis that KDO incorporation represents the initial step in group II polysaccharide polymerization (Finke *et al.*, 1989). However, not all acidic sugars which constitute the group II polysaccharide repeating units are chemically related to KDO. Thus, if the model proposed by Finke *et al.* (1989) is true, it is possible that region I and III gene products in conjunction with inner membrane components, such as undecaprenol and/or exogenous acceptor(s), can co-partecipate in initiation, elongation and export of chemically different polysaccharides.

The K5 polysaccharide is a heteropolymer of glucuronic acid (GlcA) and N-acetyl glucuronic acid (GlcNAc) with a primary structure composed of 4)-ß-GlcA-(1,4)-α-GlcNAc-(1 (Vann *et al.*, 1981). Region II of the K5 capsule cluster is approximately 7 Kb (Roberts *et al.*, 1988a). However very little is known about the genetic organization of this region and the function of its encoded products involved in K5 polysaccharide biosynthesis. Recent studies on the biochemistry of the K5 polysaccharide biosynthesis have revealed that GlcA and GlcNAc are contained in equal amounts. Additionally, KDO, which is not a constituent of the repeating unit, is present at the reducing end of this polymer (Finke *et al.*, 1991). K5 polysaccharide polymerization is thought to occur at the inner face of the cytoplasmic membrane without participation of any lipid-linked oligosaccharides (Finke *et al.*, 1991). This result contrasts with the specific lipid involvement in K1 capsule biosynthesis. However, similar experimental data have already been reported in polysaccharide biosynthesis of *N.meningitidis*. For example, it has been demonstrated that undecaprenol phosphate participates in the polysaccharide polymerization of *N.meningitidis* group b (Masson and Holbein, 1985), whereas the presence of this lipid

intermediate seems not to be required in the polysaccharide polymerization of *N.meningitidis* group c (Vann *et al.*, 1979). The reason why a lipid intermediate is apparently involved in biosynthesis of only some capsule polysaccharides is as yet unclear. In a theoretical model for K5 polymerization (Finke *et al.*, 1991), KDO is transferred first from CMP-KDO to an inner membrane acceptor, the nature of which is unknown. However, despite the absence of a lipid intermediate, the authors are not in the position to rule out the possibility that undecaprenyl-phosphate may still represent a carrier for the K5 polymer. Then, initiation and elongation occurs by specific glycosyltransferases possibly associated with the cytoplasmic membrane. These enzymes would transfer alternately GlcA and GlcNAc from their UDP activated forms to the non-reducing end of the growing polysaccharide chain.

In summary, region II encodes functions directly involved in the specific polysaccharide biochemical pathway. Some of these products, such as NeuS, appear not to function autonomously but in a more complex membrane-associated system. It is possible to imagine a basic apparatus for secretion composed of a protein complex intimately associated with the inner membrane, to which specific region II gene products can be associated to constitute a polymer specific apparatus of biosynthesis. A more detailed knowledge about the genetic and biochemical information of regions II determinants from other capsule clusters, might reveal mechanisms involved in biosynthesis of group II polysaccharides biosynthesis. Thus, this thesis partly addresses the organization of region II of the K5 capsule cluster and the possible function of its encoded products.

2.4.3 REGION I OF *E.coli* GROUP II CAPSULE GENE CLUSTERS

Region I is thought to take part in the translocation of polysaccharide to the cell surface (Boulnois and Jann, 1989). Mutations in this region result in bacteria with periplasmic polymer indistinguishable from that isolated from the surface of wild type strains (Boulnois *et al.*, 1987; Kroncke *et al.*, 1990a). Thus, some region I products are thought to be involved in export of the mature polysaccharide across the outer membrane (Boulnois and Jann, 1989). The size of region I has been reported to vary in size, from 8 Kb (Boulnois and Jann, 1989) to 11.6 Kb (Silver *et al.*, 1988). In general , five proteins, ranging from 37 to 77 KDa, have been reported to be encoded by this region (Roberts *et al.*, 1986; Silver *et al.*,1987; Vimr *et al.* 1989; Boulnois and Jann 1989). Mutations located in two genes of region I of the K1 capsule cluster, termed *kps*C and *kps*S, diminished the endogenous sialyltransferase activity *in vitro* (Vimr *et al.*, 1989). Additionally, a mutation located in *kps*S of region I of the K5 capsule cluster (see section 6.1) resulted in strains with cytoplasmic polymer. Taken together, these results suggest that the presence of some region I products is required for polysaccharide export across the cytoplasmic membrane and might also be involved (either directly or indirectly) in the process of polymerization itself. Mutations located in another region I gene, termed *kps*D, resulted in bacteria with periplasmic polymer. *kps*D encodes a periplasmic protein with a calculated MW of 60 KDa, which is synthesized in the cytoplasm as a precursor protein of about 62 KDa (Silver *et al.*, 1987). The presence of a periplasmic gel between the inner and outer membrane has been described by Hobot *et al.* (1984). This gel is constituted by a polysaccharide frame filled with an aqueous solution of periplasmic proteins, oligosaccharides, monosaccharides and other small solutes. Diffusion of

substrates in this gel (Hobot *et al.*, 1984) would be
facilitated by the action of substrate-binding proteins
which direct the process of translocation. The periplasmic
polymer in strains carrying mutations in *kps*D and the
periplasmic location of KpsD, has led to the hypothesis
that KpsD might be involved in polysaccharide export
through the periplasm by protein-carbohydrate or protein-
protein interactions (Silver *et al.*, 1987). Finally
deletions which abolished the synthesis of the 38 KDa
protein, encoded by the region I of the K1 capsule cluster,
resulted in bacteria which were no longer sensitive to the
K1 bacteriophage infection (Boulnois and Jann, 1989).
However, immuno-electron microscopy has revealed that these
mutants expressed surface material in the form of small
"tufts" (Boulnois *et al.*, 1987). A similar phenotype is
also observed in temperature shift-up experiments from 18°C
to 37°C. The newly exported polysaccharide appears at first
as "tufts" on the bacterial surface in proximity to areas
in which the cytoplasmic and outer membranes came into
contact, termed Bayer junctions (Bayer, 1979). Then,
polysaccharide is rapidly dispersed around the outer
membrane with the formation of a classic capsule (Kroncke
et al., 1990b). Taken together, these results have led to
the hypothesis that polysaccharide might be exported
through the Bayer junctions and then organized into a
capsule by the 38 KDa protein (Boulnois and Jann, 1989).
Interestingly, laboratory strains of *E.coli* harbouring DNA
fragments which spanned region I, contain high levels of
CMP-KDO synthetase. Additionally, these levels were
detectable at growth temperature of 37°C, whereas at
temperatures below 20°C the enzyme activity decreased to
low levels (Finke *et al.*, 1989). These data indicated
region I contains the genetic information required for the
high level of CMP-KDO synthetase activity found in *E.coli*
expressing group II polysaccharides. This might involve
either the presence of an additional CMP-KDO synthetase
gene or a gene product which affected the expression of the

*kds*B gene present in all *E.coli* regardless of capsule status. Thus region I is a key region in the *kps* gene cluster. Its gene products are thought to be involved in different steps of polysaccharide production, ranging from co-participation in polysaccharide polymerization to polymer assembly into a capsule. Additionally, this region is also thought to be responsible for the high level of CMP-KDO synthetase which could represent the limiting step in group II polysaccharide biosynthesis. Thus, it is tempting to suggest that the regulation of expression of this region could constitute an important and general control mechanism for group II capsule production. This thesis focuses on region I of the K5 capsule cluster, its genetic organization and possible biological function of its encoded products.

CHAPTER 3

AIMS

Very little is known about how polysaccharides are
synthetized, exported across membranes and organized onto
the bacterial cell surface. The export system appears to be
a key target in the study of group II capsules, since such
a system is able to transport different polysaccharides
regardless of their chemical structure. This study
investigates the process of K5 polysaccharide biosynthesis
in *E.coli* K5 and two approaches were adopted. Firstly,
describe in detail the genes of region I, including those
involved in the high level of CMP-KDO synthetase activity.
Secondly, generate mutants defective in expression of the
K5 polysaccharide to produce information about the role
played by individual products involved in biosynthesis of
the K5 polysaccharide. This thesis has been carried out in
collaboration with the laboratory of Professor Klaus Jann
of the Max-Planck-Institute for Immunology of Freiburg
(Germany).

CHAPTER 4

MATERIAL AND METHODS

4.1 BACTERIAL STRAINS AND PLASMIDS

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

TABLE 2.1 BACTERIAL STRAINS

LE392	F$^-$, *hsd*R514 (r^-_K, m^+_K), *sup*E44 *sup*F58, *lac*Y1, *gal*K2, *gal*T22, *met*B1, *trp*R55, lambda$^-$	Maniatis *et al.*, 1982
JM101	F$^-$, *sup*E, *thi* , (*lac-pro*AB), *tra*D 361, *pro*AB, *lac*IqZM15	Yanisch-Perron *et al.*, 1985
CC118	*ara*D 139, (*ara*, *leu*) 7697, *lac*X 74, *pho*A 20, *gal*E, *gal*K, *thi*, *rps*E, *rpo*B, *arg*E am, *rec*A 1	Manoil and Beckwith, 1985
512	O1:K1:H7	K Jann
20026	O10:K5	K Jann
DS410	*min*A, *min*B, *ara*, *xyl*, *mtl*, *azi*, *thi*	Dougan and Sheratt, 1977

TABLE 2.2 PLASMIDS

pACYC184	Cm^R, Tc^R, cloning vector		Chang and Cohen, 1987
pUC18/19	Ap^R, cloning vector		Yanisch-Perron *et al.*, 1985
M13mp18/19	bacteriophage cloning/sequencing vector		Yanisch-Perron *et al.*, 1985
pGB110	Ap^R, $K5^+$		Roberts *et al.*, 1986
pGB118	Ap^R, $K5^+$		Roberts *et al.*, 1986
pGB118::1	Ap^R, $K5^-$		GB Boulnois
pKT274	Ap^R, $K1^+$		Echarti *et al.*, 1983
pGB14	Ap^R, $K1^-$		Echarti *et al.*, 1983
pGB15	Ap^R, $K1^-$		Echarti *et al.*, 1983

4.1.1 GROWTH CONDITIONS AND MEDIA

Bacteria were grown in Luria at 37°C with the addition of 1.5% agar (BBL) as required. D.O. medium was used as stated (per litre: 7g K_2HPO_4, 2g KH_2PO_4, 0.5g sodium citrate,

41

0.1g $MgSO_4$, 1g NH_4SO_4, 20g casamino acids, 5g yeast extract, 2g glucose to pH7.2 with NaOH). B-agar (0.1% peptone, 0.8% NaCl, 1.5% agar) was used where stated. Where necessary antibiotics were added to the growth medium at the following concentrations: ampicillin at 100μg/ml, chloramphenicol at 25μg/ml, kanamycin 25μl/ml and tetracycline at 20μg/ml. Antibiotics were obtained from Sigma Chemical Company Ltd. Bacterial cells were routinely harvested by centrifugation in a Sorval centrifuge (3300g, at 4°C for 10 mins) or in a bench top minifuge (13400g, at room temperature for 5 mins).

4.2 TRANSFORMATION OF BACTERIAL CELLS

4.2.1 PRODUCTION OF COMPETENT CELLS

100μl of an overnight culture were diluted 1:100 with 10ml of L-broth and grown to mid exponential phase (OD_{600} ~ 0.4). The cells were washed in 10mM NaCl and resuspended in 4ml ice-cold $CaCl_2$ (100mM). The cells were placed on ice for 30 mins and collected by gentle centrifugation (1800g) at 4°C for 5 mins. The cell pellet was resuspended in 1ml ice-cold $CaCl_2$ (100mM) and used immediately in transformation.

4.2.2 TRANSFORMATION WITH PLASMID DNA

Competent cells (100μl) were mixed with 5-20μl of DNA (in water) to be transformed and placed on ice for 1 hour. The cells were heat-shocked at 42°C for 3 mins. Immediately after heat-shocking 500μl of L-broth were added and the cells were incubated for one hour at 37°C. The transformed cells were plated onto L-agar plates (100μl per plate)

which contained the appropriate antibiotic(s). The plates were incubated overnight at 37°C.

4.2.3   TRANSFORMATION WITH BACTERIOPHAGE DNA

Competent cells of *E.coli* JM101 (100μl; see section 4.2.1), were mixed with the construct DNA, incubated on ice for one hour and heat-shocked at 42°C for 3 mins. The transformed cells were mixed with 100μl of JM101 (mid exponential phase), then 3 ml of molten B-agar (held at 45°C and containing 20μl 100mM IPTG, 50μl X-gal in dimethylformamide) were added. The suspension was immediately mixed and poured onto a B-agar plate, rocked for disperse and once set incubated at 37°C for overnight.

4.3   PROCEDURES FOR DNA EXTRACTION

DNA extraction protocols used the following solutions:

solution I   : 50mM glucose
                   25mM Tris-HCl pH 8.0
                   10mM EDTA
                   5mg/ml lysozyme

solution II   : 0.2M NaOH
                   1% SDS

solution III : 5M acetate (11.5ml glacial CH_3COOH)
                   3M potassium ions (60ml CH_3COOK)
                   H_2O (28.5ml)

43

4.3.1   SMALL SCALE EXTRACTION OF PLASMID DNA

Small scale preparation of plasmid DNA used 1.5ml of an overnight culture. Cells were suspended in 100μl of a freshly made solution I for 30 mins on ice. Solution II (200μl) was added and the tube was gently mixed and placed on ice for a further 10 mins. Solution III (150μl) was added, mixed gently and placed on ice for 15 mins. The supernatant was recovered after centrifugation avoiding the white pellet. Protein was removed by one phenol extraction followed by one chloroform extraction. The DNA was precipitated by adding two volume of ethanol but omitting CH_3COONa (see section 4.3.5). The DNA pellet was resuspended in 50μl of sterile distilled water.

4.3.2   LARGE SCALE EXTRACTION OF PLASMID DNA

Overnight cultures (400ml) were used for large scale preparation of plasmid (Birboim and Doly, 1979). The cells were collected by centrifugation, resuspended in 10ml of a freshly made solution I and left on ice for 30 mins. Solution II (20ml) was added and gently mixed, the whole was left on ice for another 15 mins. 15ml ice-cold solution III were added and mixed gently, then the preparation was placed on ice for a further 15 mins. Cell debribs was removed by centrifugation at 4°C for 30 mins at 36900g. Isopropyl alcohol (0.6 volumes) was added to the supernatant, mixed and left at room temperature for a minimum of 20 mins. The DNA was collected by centrifugation at 3500g for 30 mins at 20°C. The DNA pellet was dried *in vacuo* and resuspended in sterile water to a final volume of 17ml. CsCl was added to a final concentration of 1mg/ml and ethidium bromide to 50 μg/ml. Chromosomal and plasmid DNA was separated at 40000rpm using a Sorvall TV850 rotor in

Sorvall OTD 60 centrifuge for 20 hours at 20°C. The DNA was visualised under UV light and the lower band (plasmid DNA) extracted. Ethidium bromide was removed by several extraction with isopropanol saturated with CsCl. CsCl was removed by exhaustive dialysis against sterile distilled water at room temperature. Plasmid DNA was stored dissolved in sterile distilled water at -20°C.

4.3.3   EXTRACTION OF M13mp18/19 TEMPLATE DNA

The recombinant bacteriophage M13mp18/19 were transformed into JM101 (section 2.2.3) and white plaques were picked into 5ml L-broth containing 100μl of an overnight culture and incubated at 37°C for 5 hours with vigorous aeration. Replicative form DNA and template DNA were obtained from 1.5ml of a bacterial culture. The replicative form DNA was extracted from the cell pellet (see section 4.3.2) whereas the template DNA was isolated from the supernatant. The supernatant (1.2ml) was mixed with 300μl of a solution containing 2.5M NaCl and 20% PEG 6000. The mixed solution was left at room temperature for 30 mins. The phage pellet was recovered by two sequential centrifugations (the second centrifugation was performed to remove traces of PEG 6000), then resuspended in 120μl TES (10mM Tris-HCl pH 7.5; 1mM EDTA; 10mM NaCl) and extracted with an equal volume of phenol followed by one chloroform extraction. The template DNA was precipitated with ethanol (see section 4.3.5). The template DNA was collected by centrifugation at 13400g for 10 mins, dried *in vacuo* and resuspended in 20μl TE (10mM Tris-HCl pH 7.5; 1mM EDTA). 2μl of template DNA were visualised by agarose gel electrophoresis (see section 4.5), and 4-7μl were typically used in a sequencing reaction (see section 4.5).

4.3.4 PHENOL EXTRACTION AND ETHANOL PRECIPITATION OF DNA

Phenol extraction was performed using one volume of phenol containing 0.1% (w/v) of hydroxyquinoline and equilibrated with 100mM Tris-HCl pH 8.0. Chloroform extraction was performed using one volume of a mixture of chloroform:isoamyl alcohol (24:1 ; v/v). The aqueous phase (lower) was separated in a Sorval centrifuge at 20°C for 20 mins or in a bench top minifuge for 5 mins at 13400g, and collected avoiding the inter phase. Ethanol precipitation was performed with CH_3COONa to a final concentration of 300mM and 2 volumes of ethanol, at -20°C for a minimun of 30 mins. The DNA was collected by centrifugation either in a bench top minifuge for 5 mins at 13400g or in a sorval centrifuge for 30 mins at 3500g.

4.4 INFECTION OF E.COLI WITH BACTERIOPHAGE

E.coli was grown to mid exponential phase in the presence of 10mM $MgSO_4$ (if necessary the appropriate antibiotic was added to the medium). The cells were collected by centrifugation and resuspended in an equal volume of 10mM $MgSO_4$. The prepared cells (100μl) were mixed with bacteriophage (normally with a bacteriophage titre ranging from $1x10^2$ to $1x10^6$ pfu/ml). The suspension was left at room temperature for 15 mins, then 3ml of molten L-agar (held at 45°C) were added, mixed and poured onto L-agar plate (if necessary containing the appropriate antibiotic). Plates were rocked for disperse and once set incubated at 37°C for overnight (generally a preliminary check, for plaque formation, was performed after 4-6 hours of incubation).

phage dilution buffer: 10mM Tris-HCl pH 7.4

10mM MgSO$_4$

0.01% gelatin

4.5   TECHNIQUES USED IN ROUTINE DNA MANIPULATION

Restriction endonuclease cleavage of DNA was performed according to the manufactures recommandations, typically in 20μl reactions with one unit of enzyme per μg of DNA at 37°C. T4 DNA ligase was used at 14°C overnight. Restriction endonucleases and DNA modifying enzymes were purchased from Pharmacia Biochemicals Inc or Life Technologies Ltd (GIBCO/BRL). DNA fragments were separated by agarose gel electrophoresis using 0.7-1.2% Seakem agarose in TAE buffer (40mM Tris-acetate, 1mM EDTA) with 0.5μg/ml ethidium bromide and visualised using a longwave UV transilluminator. DNA samples were mixed with the appropriate volume of 6x gel-loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol, 15% w/v Ficoll) prior to loading. The DNA size markers used was 1 Kb ladder (BRL/GIBCO). For subcloning, the DNA fragment in question was first separated by agarose gel electrophoresis. Agarose containing the fragment was excised from the gel and the DNA recovered by elution as follows: the gel slice was placed in dialysis tubing containing 500μl TE pH 7.5 and subjected to 100V for 60 mins in TAE buffer, polarity was reversed for at least 30 sec. DNA was collected by ethanol precipitation (section 4.3.5) from the solution surrounding the slice.

4.6 DNA SEQUENCING

DNA was sequenced by the chain termination method described by Sanger *et al.*, (1977) in which DNA synthesis from deoxynucleotide triphosphates is terminated by the addition of dideoxynucleotide triphosphates. The M13 cloning vectors, M13mp18 and M13mp19 were used to generated single stranded DNA templates (section 2.3.4). Sequence reactions were performed using the Sequenase Version 2.0 kit produced by United States Biochemical Corporation. The protocol recommended by the manufacterers was followed using the universal (-40) primer. DNA fragments were radiolabelled by incorporating [α-^{35}S]dATP in the extention reactions. The radiolabelled DNA fragments were separated by gradient gel electrophoresis (Biggin *et al.*, 1983). Preparation of the gel used the following solutions:

gel solution 1: 7ml 5x TBE acrylamide/urea mix
45μ 10% ammonium persulphate
2.5μ TEMED

gel solution 2: 40ml 0.5%x TBE acrylamide/urea mix
180μ 10% ammonium persulphate
7.5μ TEMED

0.5x TBE acrylamide/urea mix
per litre: 430g urea
50ml 10x TBE
150ml 40% acrylamide

5x TBE acrylamide/urea mix
per litre: 430g urea
150ml 10x TBE
150ml 40% acrylamide
50g sucrose
50mg bromophenol blue

40% acrylamide

per litre: 380g acrylamide

   20g bisacrylamide

   deinised with 50g/l amberlite

Electrophoresis grade ammonium persulphate was purchased from BIO-RAD, TEMED from Sigma Chemical company Ltd, acrylamide and bisacrylamide (NN'-methylenebisacrylamide from BDH). 10x TBE is 0.089M Tris-borate, 0.089M boric acid, 0.002M EDTA. To prepare the gel, gel plate (20cm x 50cm) were taped together separated by 0.4mm spacers. 10ml gel solution 2 followed by 14ml gel solution 1 were drawn up into a 25ml pipette. Air bubbles were introduced to form a rough gradient. The liquid was run between the gel plates and the cavity filled with the remaining of gel solution 2. The comb was positioned and the plates clamped along each side. Gels were routinely freshly made. A vertical electrophoresis system was used. Running buffer in the top tank was 0.5x TBE and the lower 1x TBE in accordance with the gradient of the gel itself. The gel was clamped in position with aluminium sheets of a similar dimension as the gel plates on either side for even heat distribution. The gel was pre-run for 30 mins at a constant power of 40W and the wells rinsed with the running buffer prior to loading. Electrophoresis was performed at constant power of 40W for 3, 6 or 9 hours. After electrophoresis the gel plates were prised apart and the gel was soaked in fixing solution (10% methanol and 10% acetic acid), for 15 mins and then rinsed with distilled water. The gel was transferred to a pre-wetted filter paper, covered with Saran wrap and dried under vacum at 80°C. Autoradiography used Dupont Cronex film and took place at room temperature.

4.7 TECHNIQUES USED FOR POLYSACCHARIDE DETECTION

4.7.1 EXTRACTION OF POLYSACCHARIDE

Total polysaccharide was extracted from 10ml of an overnight culture (Jann and Jann, 1985). Cells were collected by centrifugation at 4°C for 10mins at 3000g and dehydrated by resuspending the cell pellet in 5ml ethanol, this was repeated twice. The preparation was again centrifuged, the pellet resuspended in 1ml acetone and transferred to a minifuge tube. The sample was centrifuged for 1min in a bench top minifuge and all traces of acetone removed from the final pellet by evaporation at room temperature for 20-30 mins. The final pellet was resuspended in 50μl 1M $MgCl_2$ and incubated at 37°C for 2 hours. Cell debris was removed by centrigugation for 5mins in a bench top minifuge and the supernatant, a crude polysaccharide extract, retained.

4.7.2 DOUBLE IMMUNODIFFUSION

Double immunodiffusion (Ouchterlony) analysis of polysaccharide extracts was performed using 1% Seakem HGT agarose (FMC Bioproducts) in barbitone buffer. To make barbitone buffer 12g 5'5' diethylbarbituric acid sodium salt was dissolved in 800ml distilled water. 4.4g 5'5 diethylbarbituric acid was dissolved at 95°C in 150ml distilled water. The two solutions were mixed together and adjusted to pH 8.2 with 5M NaOH. 4ml 1% agarose in barbitone were poured into a 35mm diameter petridish. Once set 3mm diameter holes were cut in the agarose using a punch. Typically, 10μl polysaccharide extract or serum was loaded per well.

4.8 MINICELL ANALYSIS

4.8.1 MINICELL PURIFICATION

Minicells were isolated using the procedure described by
Hallewell and Sherratt (1976). Minicell strains (*E.coli*
DS410) were grown to stationary phase in 400ml Brain Heart
Infusion (if necessary the appropriate antibiotic was added
to the nedium). The cells were separated from the culture
by centrifugation at 600g for 5 mins. The supernatants were
centrifuged at 8500g for 15 mins and pellets retained. The
pellets were resuspended in 6ml 1x M9 salts and minicells
were further purified by two successive sedimentations
through 20ml linear gradients of 5-20% (w/v) sucrose in 1x
M9 salts at 4650g for 20 mins (4°C). Purified minicells
were collected by centrifugation at 9500g for 10 mins and
resuspended in 1x M9 salts to a final OD_{650}=2.0. Minicells
were either immediately used for proteins labelling or
aliquoted (100μl) in 30% sterile glycerol and stored at -
20°C (for a period of time not exceeding 3-4 months).

10x M9 salts per litre:	60g	Na_2HPO_4	(337mM)
	30g	KH_2PO_4	(220mM)
	5g	NaCl	(85mM)
	10g	NH_4Cl	(187mM)

4.8.2 PROTEIN LABELLING

Proteins were labelled for 90 mins at 37°C in M9 minimal
medium containing ^{35}S-methionine (100μCi/ml ; 1Ci/nmol).
After a 10 mins chase with methionine-supplemented broth,
minicells were lysed by boiling in cracking buffer (62,5mM
Tris-HCl pH6.8 ; 3% SDS ; 10% glycerol ; 5% beta-

51

mercaptoethanol). About 5-10μl (from a final volume of 50μl) were run in a SDS polyacrylamide gel. The SDS polyacrylamide gel contained an acrylamide to bisacrylamide ratio of 37.5:1. The upper stacking gel contained 4.5% acrylamide in 0.125M Tris-HCl pH 6.8 and 0.1% SDS. The lower running gel containing 12.5% acrylamide in 0.375M Tris-HCl pH 8.8 and 0.1% SDS. The running buffer contained 0.02M Tris-HCl, 0.2M glycine, 0.1% SDS and 2.4mg/l sodium thioglycollate. The gel was run at a constant current of 25mA. After running the gel was soaked in fixing solution (10% acetic acid, 25% isopropanol) for 30 mins. The gel was dried under vacum at 80°C and autoradiographed at room temperature.

4.9   COMPUTER ANALYSIS

The DNA sequence produced by this thesis was analysed using the Wisconsin (Dereveux *et al.*, 1984) and Lipman-Pearson (Lipman and Pearson, 1985) molecular biology programs on the Vax VMS Cluster.

4.9.1   DATABASE SEARCHING

Amino acids can be classified on the basis of their chemical-physical properties, such as hydrophobicity, hydrophilicity, cross-link forming etc. (amino acid homology), and protein sequences can be aligned with respect to their amino acid homology and/or their amino acid identity. The homologies between protein sequences produced by this thesis and the NBRF-Protein sequence collection were identified by using of the computer program FastA. FastA uses the method of Lipman and Pearson to search for similarities between one sequence (the query)

52

and any group of sequences. FastA displays the best
similarities aligning horizontally tha query sequence with
the identified homologous sequences. Double vertical lines
identify identical amino acids, whereas double dots
indicate homologous amino acids. Horizontal lines (gaps)
within the protein sequence are automatically inserted to
maximize the alignment. FastA can also be used to search
for similarities between nucleotide sequences.

4.9.2 HYDROPATHY PROFILE

This computer program (Kyte and Doolittle, 1982) evaluates
the hydrophilicity and hydrophobicity of a protein along
its amino acid sequence. The hydropathy profile of a
protein is plotted on the basis of a hydropathy scale,
which takes into account the hydrophilic and hydrophobic
properties of each of the 20 amino acid side-chains. The
hydropathy scale is based on experimental observations
derived from the literature. The program uses a movement-
segment approach that continously determines the hydropathy
within a segment of predetermined lenght (window) as it
advances through the sequence. The consecutive scores are
plotted from the amino to the carboxy terminus. The upper
part of a hydropathy profile indicates protein
hydrophobicity, whereas the lower part indicates protein
hydrophilicity.

CHAPTER 5

REGION I OF THE K5 CAPSULAR GENE CLUSTER

Region I of the *kps* gene cluster is thought to play a key role in the biosynthesis of group II capsules. Products of this region are thought be involved in different steps of polysaccharide biosynthesis, ranging from the possible direct co-participation in polysaccharide polymerization to polymer assembly on the cell surface. At least one region I gene is also thought to be responsible for the high levels of CMP-KDO synthetase activity characteristically found in strains expressing group II capsules. Additionally, levels of this enzyme are high only when bacteria are grown at capsule permissive temperatures, indicating that this enzyme might represent a limiting step by which group II polysaccharide biosynthesis is regulated. The aims of this thesis were, in part, to descibe the organization of region I.

The initial step was to define more accurately the extent of region I, which has been reported to range in size from 8 Kb (Boulnois and Jann, 1989) to 11.6 Kb (Silver *et al.*, 1988). Several region I mutants had been described, which were deficient in polysaccharide export (Figure 5.1). pGB14 was derived from pKT274, which encodes the biochemical functions necessary for biosynthesis and export of the K1 capsular polysaccharide, and contains a 4 Kb deletion (Echarti *et al.*, 1983) which removes the beginning of region I. LE392(pGB14) has surface polymer organized in the form of small "tufts" (Boulnois and Jann, 1989). pGB15 derived from pKT274 and carries a 5.3 Kb deletion thought to remove the central part of region I (Echarti *et al.*,

1983). LE393(pGB15) results in intracellular polymer in the periplasm (Boulnois *et al.*, 1987). pGB118::1 is a Tn1000 transposon insertion in pGB118. This insertion is located at the end of region I, close to region II, and LE392(pGB118::1) results in intracellular polymer in the cytoplasm (D Bronner and K Jann, personal communication). pGB118 derives from pGB110, which encodes the biochemical functions necessary for biosynthesis and export of the K5 capsular polysaccharide. pGB118 contains a 20 Kb *BamHI* deletion (Roberts *et al.*, 1986) (Figure 5.1). LE392(pGB110) and LE392(pGB118) express on their cell surface a K5 polysaccharide immunologically indistinguishable from that of K5 wild type strains. LE392(pGB110) and LE392(pGB118) are sensitive to K5 bacteriophage infection, whereas LE392(pGB118::1) is resistant (Table 5.1). Likewise, LE392(pKT274) is sensitive to K1 bacteriophage infection, whereas LE392(pGB14) and LE392(pGB15) are resistant (Table 5.1).

Identification of K-specific polysaccharides on the bacterial surface is readily shown using capsule-specific bacteriophage infection. This technique was used to identify the shortest DNA fragment of the K5 capsular gene cluster able to complement strains carrying pGB14, pGB15 or pGB118::1. A 10.4 Kb *HindIII* fragment, which spanned region I was subcloned from pGB110 into pUC18 generating pH18 (Figure 5.1). Verification of the structure of pH18 was obtained by digestion of the plasmid with several restriction enzymes and their sites were mapped. Subsequently, a 10.4 Kb *HindIII* fragment and a 7.7 Kb *HindIII-EcoRV* fragment, were subcloned from pH18 into pACYC184 generating pH54 and pHEV1, respectively (Figure 5.1). pKT274 and pGB118::1 have the pMB1 replicon (Hershfield *et al.*,1974), and are therefore compatible with the p15A replicon of pACYC184 (Selzer *et al.*, 1983) allowing complementation experiments to be performed.

55

LE392(pGB14) or LE392(pGB15) also carrying either pH54 or pHEV1 were sensitive to K1 bacteriophage infection (Table 5.1). Likewise, LE392(pGB118::1) carrying pH54 or pHEV1 were sensitive to K5 bacteriophage infection (Table 5.1). The presence of the above mentioned plasmids, in each strain, was verified by small scale extraction of plasmid DNA and their digestion with appropriate restriction enzymes. Since pHEV1 was the smallest plasmid able to complement a representative collection of region I mutations, the *Hind*III-*EcoR*V fragment (7.7 Kb) better defined the region I of the K5 capsular gene cluster. This was in good agreement with the size of region I of 8 Kb proposed by Boulnois and Jann (1989).

FIGURE 5.1 COMPLEMENTATION OF REGION I
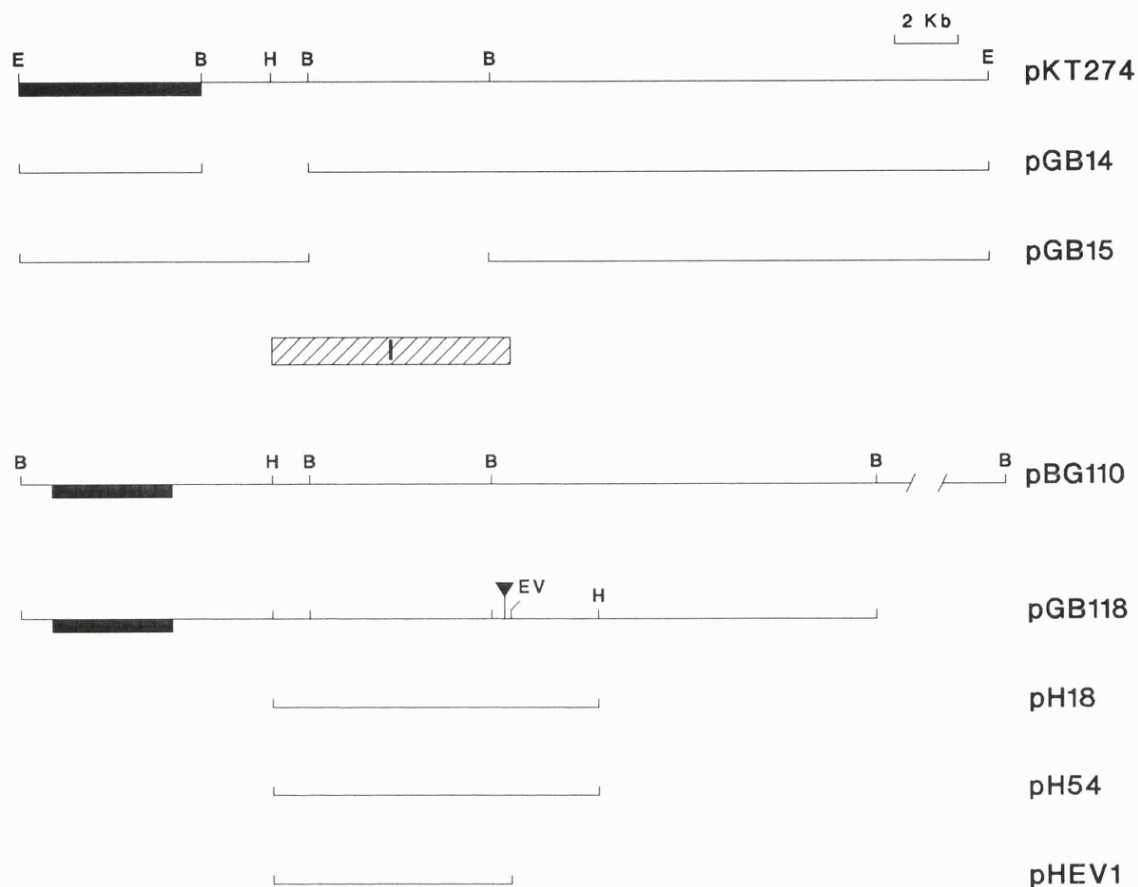
FIGURE 5.1 The linear map of various constructs is shown with horizontal lines and the scale is shown in kilobase pairs. Filled boxes indicate vector sequence, vertical and oblique lines indicate restriction sites (capital letters) or Tn1000 insertion (filled triangle; pGB118::1). The space within horizontal lines interrupted by dashed lines indicate sequence present in the original construct but not shown in figure. The space within horizontal lines interrupted by vertical lines indicates the extent of deletions. The box with diagonal lines aligns region I along the various constructs. Abbreviations: E, *EcoRI*; B, *BamHI*; H, *HindIII*; EV, *EcoRV*.

TABLE 5.1   COMPLEMENTATION EXPERIMENTS OF REGION I:
BACTERIOPHAGE SENSITIVITY

Strain	K1ϕ	K5ϕ
LE392	R	R
LE392(pGB110)	ND	S
LE392(pGB118)	ND	S
20026	ND	S
512	S	ND
LE392(pKT274)	S	ND
LE392(pGB14)	R	ND
LE392(pGB15)	R	ND
LE392(pGB118::1)	ND	R
LE392(pGB14, pH54)	S	ND
LE392(pGB15, pH54)	S	ND
LE392(pGB15, pH54)	S	ND
LE392(pGB15, pHEV1)	S	ND
LE392(pGB118::1, pH54)	ND	S
LE392(pGB118::1, pHEV1)	ND	S

TABLE 4.1 Abbreviation: ϕ, bacteriophage; R, resistant; S, sensitive;
ND, not determined.

5.1 CMP-KDO SYNTHETASE ACTIVITY

LE392(pH18) had CMP-KDO synthetase levels comparable to those produce by LE392(pGB110) (Table 5.2) (for discussion about the levels of CMP-KDO synthetase see section 7.2). Additionally, these levels of enzyme were apparent only when bacteria were grown at the capsule permissive temperature of 37°C. At the capsule restrictive temperature of 18°C the levels of CMP-KDO synthetase in LE392(pH18), LE392(pGB110) or K5 wild type strains were comparable with those produced by non encapsulated bacteria, such as LE392 (A Finke, personal communication). These data agreed with a previous study, which had demonstrated that DNA fragments carrying region I of the K5 capsular gene cluster were responsible for the high levels of CMP-KDO synthetase activity in *E.coli* strains expressing the K5 capsular polysaccharide (Finke *et al.*, 1989). To identify the smallest DNA region responsible for these high levels of CMP-KDO synthetase activity, two different experimental approaches were adopted. Firstly, several DNA fragments were subcloned from pH18 into vectors pACYC184 and pUC18 generating a series of plasmids termed pHSP1, pPS1, pHPS1, pBA6 and pBGSM2 (Figure 5.2). Then, the CMP-KDO synthetase activity of LE392 carrying the above mentioned plasmids was determined (Table 5.2). Secondly, pH18 was subjected to a Tn5*pho*A transposon mutagenesis. Tn5*pho*A was selected to provide information about the cellular location of some region I encoded products and for possible investigations of gene expression at different bacterial growth temperatures.

LE392(pH18) was infected with lambda::Tn5*pho*A and several insertions were identified and mapped by appropriate restriction enzyme digestions (data not shown). However, none of these mutants had alkaline phosphatase activity.

Transposon insertions spread throughout pH18 were chosen:
pH18::5, pH18::16, pH18::18, pH18::22 and pH18::24 (Figure
5.2). CMP-KDO synthetase activity was determined for LE392
carrying either pH18 or the above mentioned insertions
(Table 5.2) (N.B. the CMP-KDO synthetase assays were
carried out by A Finke). Only pBA6, which contains the 5.3
Kb *BamHI* fragment from the middle of region I, encoded for
high levels of CMP-KDO synthetase activity at 37°C. Three
insertions in pH18 affected the high levels of CMP-KDO
synthetase activity in LE392 (pH18). Mutants pH18::5 and
pH18::18, resulted in levels of CMP-KDO synthetase activity
comparable to those of LE392, whilst pH18::16 had only 33%
of the activity seen in LE392(pH18). Each of these
insertions map in the 5.3 Kb *BamHI* fragment. The insertion
pH18::22, which maps outside the 5.3 Kb *BamHI* fragment,
affected the levels of CMP-KDO synthetase activity only
when bacteria were grown at 18°C. At this temperature the
levels of enzyme activity were four fold higher than those
in LE392 or in K5 wild type strains. The reason for this is
unknown. Interestingly, LE392(pBA6) produced high levels of
CMP-KDO synthetase activity at bacterial growth
temperatures of 18°C and 37°C. This datum suggests that DNA
regions flanking the 5.3 *BamHI* fragment, which are present
in pH18 but missing in pBA6, could be responsible for
production of low and high levels of CMP-KDO synthetase
activity at 18°C and 37°C, respectively. pBA6 was
constructed by cloning the 5.3 Kb *BamHI* fragment of pH18
into the *BamHI* site of pACYC184. This cloning site maps in
the *tet* gene. It is possible that expression of the 5.3 Kb
BamHI fragment in pACYC184, is under the control of the *tet*
gene promoter (Figure 5.2). These results might suggest
that high and low levels of CMP-KDO synthetase activity
might be due to different amounts of mRNA transcribed at
37°C and 18°C, respectively.

FIGURE 5.2   IDENTIFICATION OF THE SMALLEST DNA REGION
RESPONSIBLE FOR HIGH LEVELS OF CMP-KDO SYNTHETASE
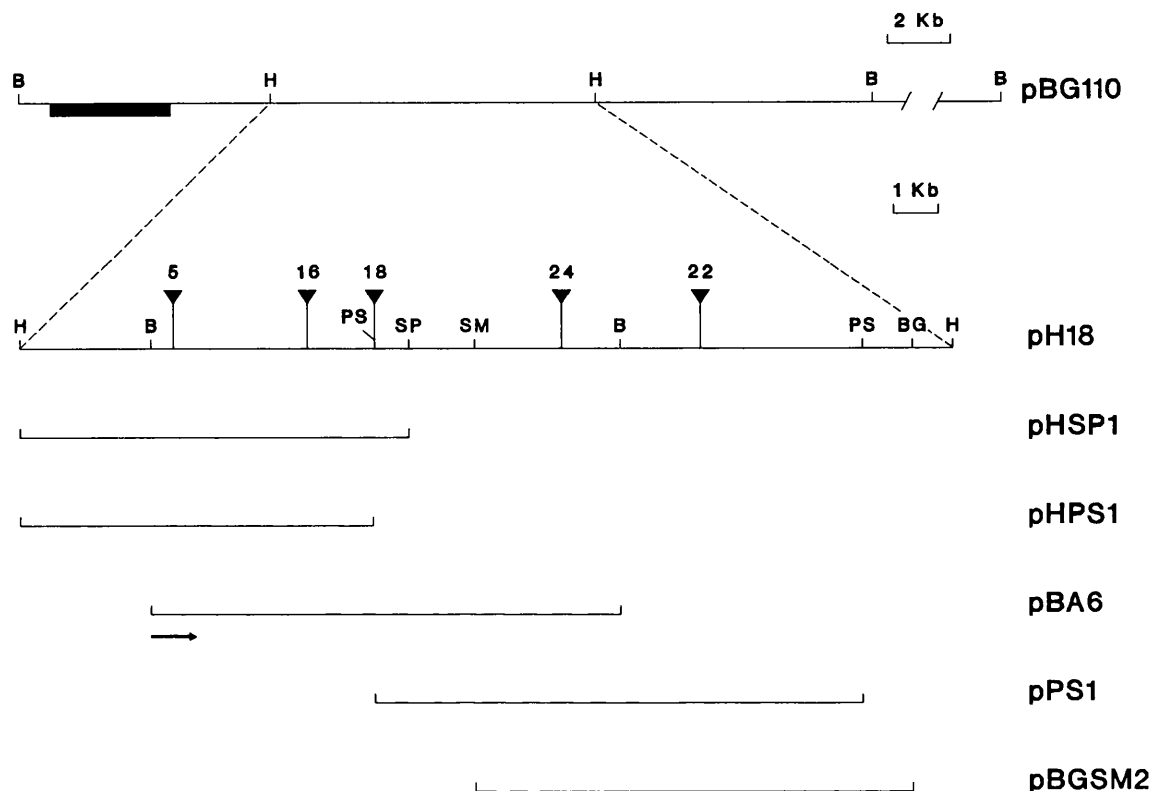
FIGURE 5.2 The linear map of various constructs is shown with horizontal lines and the scale is shown in kilobase pairs. The filled box indicates vector sequence. Vertical and oblique lines indicate restriction sites (capital letters) or Tn5*pho*A insertions (filled triangles, numbers refer to the specific insertion reported in text). The space within horizontal lines interrupted by dashed lines indicates sequence present in the original construct but not shown in figure. Broken lines highlight enlargement on scale (3:1). The arrow indicates the direction of transcription from the *tet* promoter of pACYC184. Abbreviations: B, *Bam*HI; H, *Hind*III; PS, *Pst*I; SP, *Sph*I; SM, *Sma*I; BG, *Bgl*II.

TABLE 5.2 CMP-KDO SYNTHETASE ACTIVITY OF THE VARIOUS MUTANTS AND SUBCLONES OF REGION I

Strain	CMP-KDO synthetase activity (nmol of CMP-KDO/mg of protein in 15 min)	
	18°C	37°C
LE392	55.8	35.1
20026	52.3	391.8
LE392(pGB110)	58.3	835.8
LE392(pGB118)	66.3	847.2
LE392(pH18)	50.0	1025.0
LE392(HSP1)	<50.0	<50.0
LE392(pPS1)	<50.0	<50.0
LE392(pHPS1)	<50.0	<50.0
LE392(pBA6)	1287.6	1300.4
LE392(pBGSM2)	<50.0	<50.0
LE392(pH18::5)	50.0	54.0
LE392(pH18::16)	347.0	53.8
LE392(pH18::18)	52.0	17.0
LE392(pH18::22)	266.1	900.0
LE392(pH18::24)	57.0	984.0

5.2 DNA SEQUENCE AND COMPUTER ANALYSIS OF REGION I OF THE K5 CAPSULE GENE CLUSTER

The pH18 plasmid was shown to encode for high levels of CMP-KDO synthetase activity at 37°C but not 18°C. In addition, pH18 contains the 7.7 Kb *Hind*III-*EcoR*V DNA fragment defined as carrying region I of the K5 capsular gene cluster. For these reasons, pH18 was used for sequencing region I. A multitude of DNA fragments were generated by appropriate restriction enzymes and subcloned into appropriately cleaved M13mp18 and m13mp19. These fragments were chosen to obtain the entire DNA sequence of the 10.4 Kb *Hind*III fragment on both strands (for DNA sequencing strategy and sequencing gel see figures 6.2 and 6.3, respectively). The DNA sequence of region I was analyzed using the Wisconsin (Devereux *et al.*, 1984) and Lipman-Pearson (Lipman and Pearson, 1985) molecular biology programs on the Vax VMS Cluster (for details about computer programs see section 5.9).

The *Hind*III-*EcoR*V fragment, which includes region I of the K5 gene cluster, consists of 7721 bp and comprises several open reading frames (*orf*) (Figure 5.3 and 5.13). These *orfs* are defined by utilization of the AUG codon as translation start, since other start codons, such as GUG and UUC, are used only rarely as translation starts (McCarthy and Gualerzi, 1990). *orf*1 starts 5' to the *Hind*III site and terminates 71 bp upstream of the second *orf* (*orf*2). *orf*2 (coordinates 782 to 1939 bp) is 1149 bp and terminates 23 bp upstream of the third *orf* (*orf*3). *orf*3 (coordinates 1954 to 3630 bp) is 1677 bp and terminates 9 bp upstream of the fourth *orf* (*orf*4). *orf*4 (coordinates 3640 to 4380 bp) is 742 bp and its UGA stop codon overlaps the AUG start codon of the fifth *orf* (*orf*5). *orf*5 (coordinates 4377 to 6404 bp) is 2037 bp and terminates 34 bp upstream of the sixth *orf*

(*orf6*). *orf6* (coordinates 6439 to 7608 bp) is 1161 bp and terminates 155 bp before the *EcoRV* restriction site. Other small *orfs* were identified on the opposite strand. These small *orfs* encode proteins with predicted MWs smaller than 11 KDa and the minicell analysis carried out here (see section 5.3) and in previous studies (Roberts *et al.*, 1986; Silver *et al.*, 1987) revealed proteins with calculated MWs ranging from 27 to 77 KDa. In the absence of data, which might give evidence that some of these minor *orfs* are expressed only *orf1* to *orf6* will be discussed. *orf2* to *orf6* have, in addition to the first AUG start codon, other AUGs which might be potential translation starts. To assign to each of these ORFs the most likely translational start, two criteria were used. Firstly, the presence of a ribosome binding site also referred as the Shine Dalgarno (SD) sequence, a conserved stretch of polypurines (5')-AGGAGG-(3') (Shine and Dalgarno, 1974) usually located within the region -15 to -2 bp upstream of a tranlation start (McCarthy and Gualerzi, 1990). Secondly, the calculated MW of products encoded by region I and visualized with the minicell analysis (see section 5.3) and in previous studies (Roberts *et al.*, 1986; Silver *et al.*, 1987). Additionally, during the course of this study, the N-terminal amino acid sequence of the product of ORF4 became available. This allowed the accurate assignment of translation start for *orf4*.

FIGURE 5.3 *orfs* IDENTIFIED IN pHEV1 DEFINED AS CARRYING REGION I OF THE K5 GENE CLUSTER
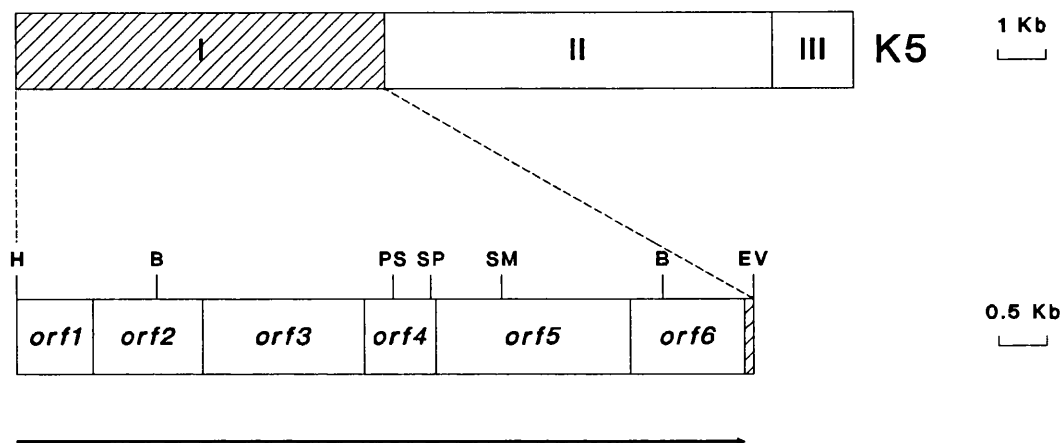
FIGURE 5.3 The organization of the K5 capsular gene cluster and *orfs* identified in pHEV1 are shown as boxes. The scale is shown in kilobase pairs. Region I is enlarged on scale 2:1 and broken lines highlight the enlargement. Vertical lines indicate restriction sites (capital letters). The arrow indicates the direction of transcription. Abbreviation: H, *Hind*III, B, *Bam*HI; PS, *Pst*I; SP, *Sph*I; SM, *Sma*I; EV, *Eco*RV.

*orf*1 extends beyond the *Hind*III restriction site, however the sequence 5' to the *Hind*III restriction site has not been determined. The first AUG start codon of *orf*1, 3' to the *Hind*III restriction site, is located at coordinate 24 bp (Figure 5.13). Translation from this AUG results in a predicted protein of 228 amino acids with a MW of 24.8 KDa, however this AUG is not preceded by a putative SD sequence. Assuming that this AUG is the authentic translation start of *orf*1, the product encoded by this *orf* does not affect the K5 polysaccharide biosynthesis, since strains harbouring the K5 capsular gene cluster with a DNA deletion

located in *orf1* are sensitive to K5 bacteriophage infection (see section 7.1).

Only the first four AUG codons of *orf2* might account for the 38.5 KDa protein visualized with the minicell analysis and identified as the product of *orf2* (see section 5.3 and figure 5.5). Translation from these codons (coordinates 782, 815, 890 and 938 bp) (Figure 5.13) yields proteins with predicted MWs of 43, 41.7, 39 and 37 KDa, respectively. However, none of these AUG codons is preceded by an obvious SD sequence. On the basis of this preliminary analysis it was not possible to assign the translation start codon for *orf2*. Computer analysis of the product of *orf2* was carried out using the first AUG (coordinate 782 bp) as the translational start. The hydropathy profile (Kyte and Doolittle, 1982) of ORF2 (Figure 5.6) revealed two hydrophobic regions, which consist of stretches longer than 20 amino acids which could represent potential membrane spanning segments (Engelman *et al.*, 1986). The first hydrophobic region (coordinates 21 to 52 amino acids) is preceded by a stretch of twenty amino acids with an overall positive charge. The second hydrophobic region is located towards the C-terminus and is followed by a hydrophilic and charged stretch of five amino acids. Oligotopic membrane proteins span the cytoplasmic membrane twice with amino and carboxy hydrophobic regions (Dalbey, 1990) and ORF2 shares with these proteins several features. Firstly, a stretch of charged amino acids precedes the N- and follows the C-terminal hydrophobic regions, respectively. Computer analysis using the amino acid sequence of ORF2 identified several homologous sequences. ORF2 is 22.1% identical and 73.2% homologous (Figure 5.4), over 359 amino acids, with the BexC protein encoded by region I of the *cap* locus of *H.influenzae* type b (Kroll *et al.*, 1990). BexC is thought to be composed of 377 amino acids with a MW of 42.1 KDa. *phoA* fusions in

frame with *bex*C have suggested that BexC contains regions periplasmically orientated (Kroll *et al.*, 1990). Additionally, BexC has been proposed to be involved in polysaccharide export from the cytoplasm to the bacterial cell surface (Kroll *et al.*, 1990). ORF2 is also 26.8% identical and 73.2% homologous (Figure 5.5), over 355 amino acids, with the CtrB protein encoded by the *cps* locus of *N.meningitidis* group B (Frosch *et al.*, 1991). CtrB is thought to be composed of 387 amino acids with a MW of 41.9 KDa. *pho*A fusions in frame with *ctr*B have suggested that CtrB contains regions periplasmically orientated and Western-blot analysis has identified these fusion products associated with the inner membrane (Frosch *et al.*, 1991). Additionally, CtrB has been proposed to be involved in polysaccharide export. CtrB is also 59.9% identical and 91.6% homologous, over 379 amino acids, with BexC (Frosch *et al.*, 1991).

FIGURE 5.4   PROTEIN HOMOLOGY BETWEEN KpsE AND BexC

```
              10        20        30        40        50        60
KpsE  MLIKVKSAVSWMRARLSAISLADIQKHLAKIIILAPMAVLLIYLAIFSQPRYMSESKVAI
                      |:|    :::|  |  |    :|:::|:    |:|||: ::
BexC       MTTENAAIPTKKKKSFWKKMKPLFGLTVLIPTAFSAVYFGLFASDIYVSESSFVV
                 10        20        30        40        50


              70        80        90       100       110
KpsE  KRSDDLNSGSLNFGLLLGASNPSSAEDALYLKEYINSPDMLAALDKQLNFREAFSHSG--
      ::: : :| |   :|| ::: ::|::|:  ::||:::|:: |:||:: | :|: :|::|
BexC  RSPRSQSSLSGVGALLQSTGFSRSQDDTYSVQEYMRSRTALSALEQGLPLRTFYSEKGDL
         60        70        80        90       100       110


          120       130       140       150       160       170
KpsE  LDFLNHLSKDETAEGFLKYYKDRINVSYDDKTGLLNIQTQGFSPEFALKFNQTVLKESER
      |: :|  :: ::|:|:| :|:|:|::|: |: :|: ::::::|::| : ::|: :|||:|:
BexC  LSRFNGFGLNDTQEAFYRYFKERLSVDVDSISGIATLRVHAFDAEEGYQINERLLKEGES
         120       130       140       150       160       170


          180       190       200       210       220       230
KpsE  FINEMSHRIARDQLAFAETEMEKARQRLDASKAELLSYQDNNNVLDPQAQAQAASTLVNT
      :||  :::| :| ::||| :::::| :::::: ::| |: :|:::| :||: :: :|:::
BexC  LINRLNERARKDTIEFAEQAVKDAEKNVNETAQALSQYRIKNKIFDLPAQSGVQLSLISS
         180       190       200       210       220       230


          240       250       260       270       280       290
KpsE  LMGQKIQMEADLRNLLTYLREDAPQVVSARNAIQSLQAQIDEEKSKITAPQGDKLNRMAV
      |::: |::|::| :|:: :  |:||| : :    :||: :|||::::::::: :|:: :  ::
BexC  LKSELIRVETQLAQLVS-ITPDNPQVPALQMRQKSLKKEIDEQTRQLSG-NGNSAATQTA
         240       250       260       270       280       290


          300       310       320       330       340       350
KpsE  DFEEIKSKVEFNTELYKLTLTSIEKTRVEAARKLKVLSVISSPQLPQESSFPNIPYLIAC
      |:: ::  |:: :   ::||:::||  ||:|:  |:|||  |  |: :  |:  || |
BexC  DYQRLMLANELAQQQLAAAMTSLQNTRGEADRQQLYLEVISQPSKPDWALQPSRIYNIIA
         300       310       320       330       340       350


          360       370       380
KpsE  WLLVCCLLFGTLKLLLAVIEDHRD
      :::   :|:|:|:|:||:|  | :|::
BexC  TFIIGLMLYGVLNLLIASIREHKN
         360       370
```

FIGURE 5.4 KpsE refers to the product of orf2 subsequently designated as kpsE. Numbers indicate amino acid coordinates of the respective proteins. Double vertical lines align identical amino acids. Double dots align homologous amino acids. Horizontal lines represent gaps automatically inserted to maximize the alignment.

FIGURE 5.5 PROTEIN HOMOLOGY BETWEEN KpsE AND CtrB

```
                10        20        30        40        50
KpsE       MLIKVKSAVSWMRARLSAISLADIQKHLAKIIILAPMAVLLIYLAIFSQPRYMSES
                              ::::  |  ::  |:|:::|:   |:  |:|
CtrB  MSEQLPVAVATETKAERKKPKKKSWIKKLSPLFWVTVIIPTVISLVYFGFFASDRFTSQS
          10        20        30        40        50        60


          60        70        80        90       100       110
KpsE  KVAIKRSDDLNSGSLNFGLLLGASNPSSAEDALY-LKEYINSPDMLAALDKQLNFREAFS
      :  ::  ||  :  :|:   ::|  :|  :::  ::|:|::|  :  ||::|::  |::|  |  |   ||  ::
CtrB  SFVV-RSPKSQSSLNGLGAILQGTGFARAQDDIYTVGEYMRSRSSLDELRKILPVREFYE
            70        80        90       100       110


      120       130       140       150       160       170
KpsE  HSGLDF--LNHLSKDETAEGFLKYYKDRINVSYDDKTGLLNIQTQGFSPEFALKFNQTVL
      :|  :|   :|  ::   :::|:|  :|||:::  :::|:  :|:  ::::  :|::    :  |:|:::|
CtrB  TKGDAFSRFNGFGFRGEEEAFYQYYKNQVMINFDTVSGISTLNVTSFDALESKKINEALL
      120       130       140       150       160       170


      180       190       200       210       220       230
KpsE  KESERFINEMSHRIARDQLAFAETEMEKARQRLDASKAELLSYQDNNNVLDPQAQAQAAS
      |::|  :||:::::|   |  :  :||:  :::|  :|:::::::::|  :|:  :|:|:|  :||::::
CtrB  KQGEALINQLNDRARADTVRYAEEVVKTAAERVKEASQNLTDYRIANGVFDLKAQSEVQM
      180       190       200       210       220       230


      240       250       260       270       280       290
KpsE  TLVNTLMGQKIQMEADLRNLLTYLREDAPQVVSARNAIQSLQAQIDEEKSKITAPQGDKL
      :||::|  ::  |  :::::|  ::  :  :  ::||:  :  ::   |||:  :||::  :  |::     ::|
CtrB  GLVSKLQDELIVIQTQLDQVKA-VTPENPQIPGLQAREQSLRKEIDQQLRAISGGGHSSL
      240       250       260       270       280       290


      300       310       320       330       340       350
KpsE  NRMAVDFEEIKSKVEFNTELYKLTLTSIEKTRVEAARKLKVLSVISSPQLPQESSFPNIP
      ::  |:::::  :   :  :::::        ::||:|:::|||:|:   |:|||  |  ||:  :   |:
CtrB  SNQAAEYQRVYLENQLAEQQLAAAMTSLESAKVEADRQQLYLEVISQPSLPDLAHEPKRL
      300       310       320       330       340       350


      360       370       380
KpsE  YLIACWLLVCCLLFGTLKLLLAVIEDHRD
      |  |:   |::   :::|:|:|:||  |  |  :|::
CtrB  YNIVATLIIGLMVYGILSLLTASIREHKN
      360       370       380
```

FIGURE 5.5 KpsE refers to the product of *orf2* subsequently designated as *kps*E. Numbers indicate the amino acid coordinates of the respective proteins. Double vertical lines align identical amino acids. Double dots align homologous amino acids. Horizontal lines represent gaps automatically inserted to maximize the alignment.

Hydropathy profiles of BexC, CtrB and ORF2 show common features (Figure 5.6). Firstly, they contain two hydrophobic regions located at the N- and C-termini of sufficient length to span the membrane. In addition, they all have hydrophilic and charged amino acids preceding the N-terminal and following the C-terminal hydrophobic regions, respectively. It is possible that ORF2, BexC and CtrB may be examples of oligotopic membrane proteins with their hydrophilic central region periplasmically orientated and with their N- and C-charged termini protruding in the cytoplasm. In addition to BexC and CtrB, ORF2 is 16% identical and 65.5% homologous, over 275 amino acids, with the FanD protein (Roosendaal and deGraat, 1989) of enterotoxigenic *Escherichia coli*. This protein is thought to be involved in export of the K99 fimbriae to the bacterial cell surface. The homology is restricted to the hydrophilic cental region of ORF2. ORF2 is also 16.1% identical and 64.8% homologous over 236 amino acids with the phospho-B-glucosidase (BglB) of *Escherichia coli* (Schnetz *et al.*, 1987), and 18.2% identical and 66.9% homologous, over 148 amino acids, with the PlsB protein of *Escherichia coli* (Lightner *et al.*, 1983). PlsB is a membrane-bound *sn*-glycerol-3-phosphate acyltransferase and catalyzes condensation of *sn*-glycero-3-phosphate and fatty acids thioesters to yield lysophosphatic acid. Homologies with BglB and PlsB are also restricted to the hydrophilic central domain of ORF2.

FIGURE 5.6 HYDROPATHY PROFILES OF KpsE, BexC AND CtrB
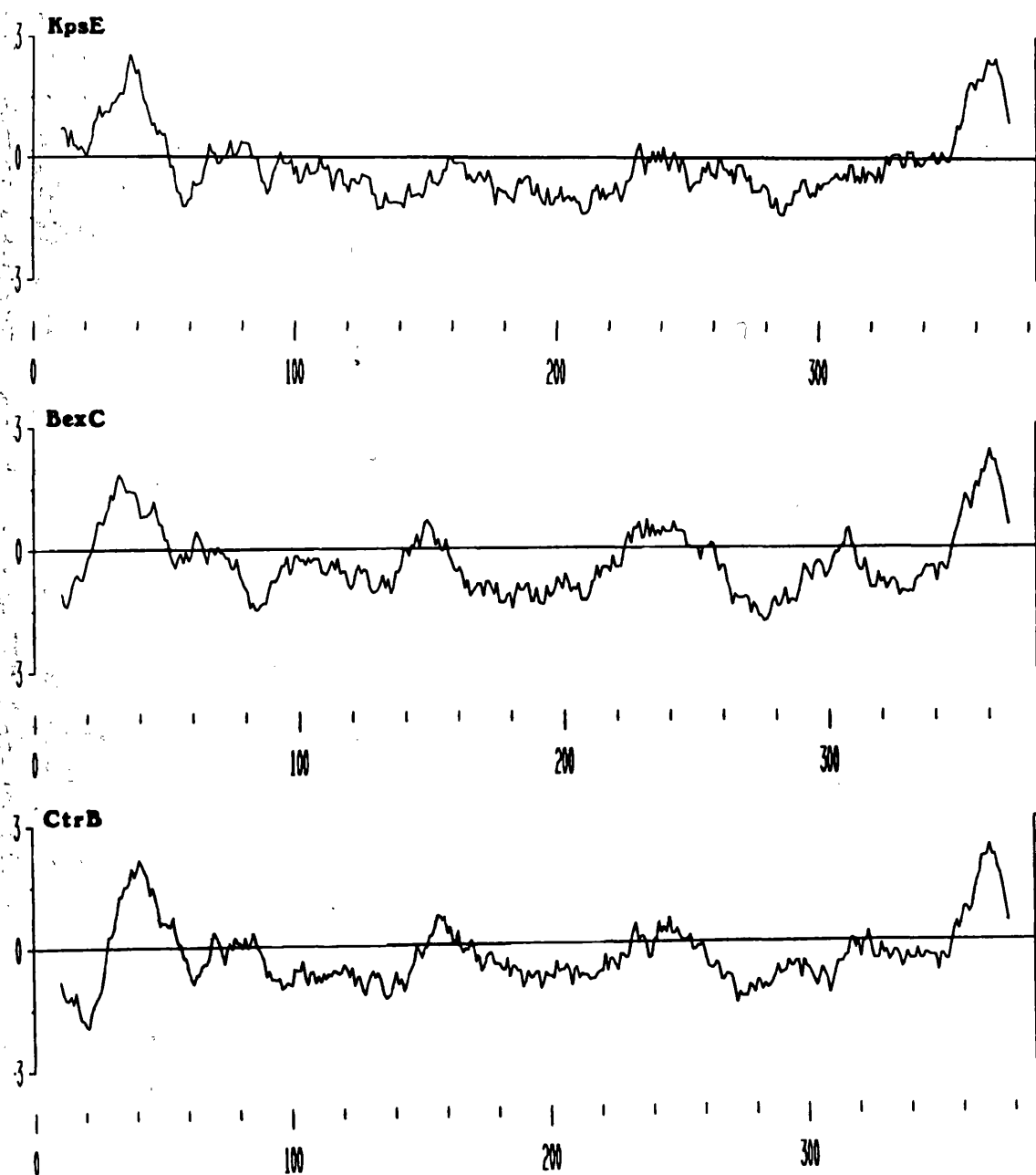
FIGURE 5.6 <u>KpsE</u> refers to the product of *orf2* (*kpsE*). Hydropathy profiles were plotted with a window 20.

orf3, in addition to the first AUG start codon that is preceded by a putative SD sequence, contains other AUGs which are also preceded by a putative SD sequence and may represent potential translation starts. Minicell analysis identified a protein of 59 KDa as the product of orf3 (see section 5.3 and figure 5.5). Additionally a 60 KDa protein has also been reported to be encoded by region I and is proposed to have a periplasmic location (Vimr et al., 1987). Translation from the first AUG of orf3 yields a protein with a predicted MW of 60414 Da. Analysis of the amino acid sequence of ORF3 identified a typical N-terminal signal peptide (von Heijne, 1983; 1985), which is composed of two positively charged residues (lysine) followed by a hydrophobic stretch of 12 amino acids and a putative processing site (Figure 5.7). On this basis it seems likely that the first AUG of orf3 is the most probable translation start. Computer analysis using the amino acid sequence of ORF3 identified 25% identity and 60.3% homology, over 68 amino acids, with the phosphatydyl glycerophosphate B phosphatase (PgpB) of Escherichia coli (Icho, 1988). PgpB contains a putative N-terminal signal peptide and is thought to be both outer and inner membrane bound. PgpB hydrolyzes three different substrates namely phosphatidyl glycerophosphate (PGP), phosphatidic acid (PA) and lysophosphatidic acid (LPA). The PgpB enzyme activity is also dependent on its cellular location. For example, it has a high PGP phosphatase activity in the cytoplasmic membrane and a high PA and LPA phosphatase activity in the outer membrane. The homology between ORF3 and PgpB is restricted to a short stretch of amino acids, however this homology might be significant since it is located in the hydrophilic central domain of PgpB which is thought to interact with the hydrophilic parts of PGB, PA and PLA. ORF3 also shows extensive homology, 60.2% over 226 amino acids, with the Malk protein, which is involved in the Maltodextrin and Maltose binding protein-dependent transport system of Escherichia coli (Gilson et al., 1982).

Despite the exact function of MalK not being known it contains a putative ATP binding site and this protein might have an energizing role by coupling ATP hydrolysis to the Maltodextrin and Maltose transport system (Higgins *et al.*,1985). However, ORF3 does not contain any obvious potential ATP binding site.

FIGURE 5.7 PUTATIVE N-TERMINAL SIGNAL SEQUENCE IDENTIFIED IN KpsD

N-TERMINAL SIGNAL SEQUENCE   KpsD


```
N

H

C    -3 : Ala;Gly;Ser;Leu;Ile;Val

     -1 : Ala;Gly;Ser

            cleavage site
```

```
Met
Lys  ⊕
Leu
Phe
Lys  ⊕

Ser
Ile
Leu
Leu
Ile
Ala
Ala
Cys
His
Ala
Ala
Gln

Ala
Ser
Ala

Thr
Ile
Asp  ⊖
Ile
Met
```
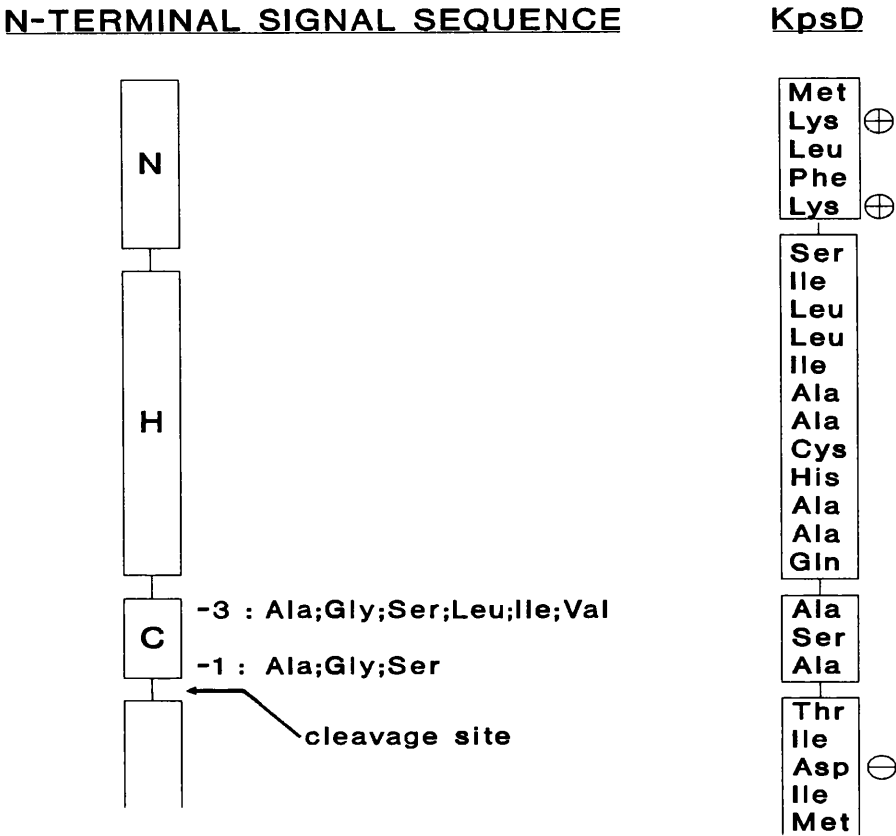
FIGURE 5.7 KpsD refers to the product of *orf3* (*kps*D, see later in the text). Prokaryotic signal peptides, like eukaryotic ones, have three distinct regions, a positive charged amino-terminus (N-region), a hydrophobic core with a minimal length of 7-8 amino acids (H-region), and a hydrophilic carboxy-terminus (C-region). The signal peptidase recognizes or prefers certain amino acids for their steric configuration which determines the position for proteolytic cleavage. In prokaryotes the position -1 contributes the carboxyl group to the peptide bond that is cleaved and is restricted to small uncharged residues (Ala, Gly or Ser). The position -3 is generally restricted to Ala, Gly, Ser, Leu, Ile or Val. Downstream of the cleavage site a negative charge in the form of Asp or Glu is preferred. Abbreviations: N, N-region; H, H-region; C, C-region.

The product of *orf*4 has been recently purified and the determined N-terminal protein sequence is identical to that predicted by the use of the first AUG start codon of *orf*4 (Rosenou and Jann, unpublished results). The purified product of *orf*4 has also been demonstrated to have CMP-KDO synthetase activity. Computer analysis using the amino acid sequence of ORF4 identified 44.3% identity and 70.7% homology (Figure 5.8), over 246 amino acids, with CTP:CMP-3-deoxy-*manno*-octusolonate cytidyltransferase (CMP-KDO synthetase) of *Escherichia coli* (Goldman and Kohlbrenner, 1985). CMP-KDO synthetase is encoded by the *kds*B gene, has a calculated MW of 27488 Da and its function is to activate KDO to CMP-KDO during lipopolysaccharide biosynthesis (Ray and Benedict, 1982). Interestingly, both ORF4 and KdsB contain two cysteines, however their position is different. Cysteine residues can be involved in formation of either disulfide "bridges" or catalytic sites. The finding that, two cysteines are present in ORF4 and KdsB but differently located, raises the following questions. May the different position of these cysteines couple the enzyme to the biosynthesis apparatus in two different ways depending on their role in CPS and LPS production? Does the position of these cysteines affect the kinetics of reaction of ORF4 and KdsB? Experiments are now in progress to extabilish the kinetics of reaction for both ORF4 and KdsB (Rosenou and Jann, personal communication).

*orf*4 is 63.5% identical to *kds*B over a nucleotide sequence of 540 bp. This suggests that *orf*4 and *kds*B might represent products of a gene duplication event. Additionally, all these data imply that the high level of CMP-KDO synthetase activity found in *E.coli* expressing group II capsules is due to the presence of an additional CMP-KDO synthetase encoded by a gene located in the middle of region I.

ORF4 is also 20.8% identical and 65% homologous over 183 amino acids with CMP-NeuNAc synthetase (NeuA) of *Escherichia coli* (Figure 5.9). NeuA, KdsB and ORF4 carry out similar biochemical reactions involving CTP and the chemically related substrates NeuNAc and KDO. NeuNAc and KDO are synthesized in bacteria through similar reactions which involve the common substrate phosphoenolpyruvic acid (Jann and Jann, 1983). Interestingly, the homology between NeuA, KdsB and ORF4 is exclusively located at their N-termini. However, a multiple alignment of ORF4, KdsB and NeuA did not reveal any conserved sequence which, on the basis of their biochemical function, might be expected.

FIGURE 5.6 PROTEIN HOMOLOGY BETWEEN KpsU AND KdsB

```
          10        20        30        40        50        60
KpsU MSKAVIVIPARYGSSRLPGKPLLDIVGKPMIQHVYERALQVAGVAEVWVATDDPRVEQAV
     |::|||||||:|:|||||||||:||  ||||| || |||  :  :|::  :  ||||:   |::||
KdsB  SFVVIIPARYASTRLPGKPLVDINGKPMIVHVLERA-RESGAERIIVATDHEDVARAV
          10        20        30        40        50


          70        80        90        100       110
KpsU QAFGGKAIMTRNDHESGTDRLVEVMHK---VEADIYINLQGDEPMIRPRDVETLLQGMRD
     :|  ||:: |||:||:|||:||:||::|     ::::  :|:|||||||||:: :   :  :::  :
KdsB EAAGGEVCMTRADHQSGTERLAEVVEKCAFSDDTVIVNVQGDEPMIPATIIRQVADNL-A
       60        70        80        90        100       110


     120       130       140       150       160
KpsU DPALPVATLCHAISAAEAA-EPSTVKVVVNTRQDALYFSRSPIPYPRN-------AEKAR
     ::::  :|||   :|  :|:|:|  :|::|||||:::    ||||||::||:  |:         :   ::
KdsB QRQVGMATLAVPIHNAEEAFNPNAVKVVLDAEGYALYFSRATIPWDRDRFAEGLETVGDN
     120       130       140       150       160       170


     170       180       190       200       210       220
KpsU YLKHVGIYAYRRDVLQNYSQLPESMPEQAESLEQLRLMNAGINIR-TFEVAATGPGVDTP
     :|:|:|||:||  :  :::|  :  :  |   |: |  ||||||::   |  :|:  :   :  :::|:|||||
KdsB FLRHLGIYGYRAGFIRRYVNWQPSPLEHIEMLEQLRVLWYGEKIHVAVAQEVPGTGVDTP
     180       190       200       210       220       230


     230       240
KpsU ACLEKVRALMAQELAENA
     :  ||:|||  |
KdsB EDLERVRAEMR
     240
```

FIGURE 5.8

KpsU refers to the product of *orf*4 (subsequently designated *kps*U). Numbers refer to the amino acid coordinates of the respective proteins. Double vertical lines align identical amino acids. Double dots align homologous amino acids. Horizontal lines represent gaps automatically inserted to maximize the alignment.

FIGURE 5.9   PROTEIN HOMOLOGY BETWEEN KpsU AND NeuA

```
             10        20        30        40        50
KpsU  MSKAVIVIPARYGSSRLPGKPLLDIVGKPMIQHVYERALQVAGVAEVWVATDDPRVEQA
      :|  :  :||||  ||: |::|   |  :::||:::::  |  |||  :   ::|  |:||:  :  ::
NeuA  MRTKIIAIIPARSGSKGLRNKNALMLIDKPLLAYTIEAALQSEMFEKVIVTTDSEQYGAI
             10        20        30        40        50        60


  60        70        80        90        100       110
KpsU  VQAFGGKAIMTRNDHESGTDRLVE-VMHKVE-ADIYINLQGDEPMIRPRDVETLLQGMRD
      ::::|::  ::  :::  ::::::  |  ::|  ::   :  |  :::   :|   :  ||  :  ::::::
NeuA  AESYGADFLLRPEELATDKASSFEFIKHALSIYTDYESFALLQPTSPFRDSTHIIEAVKL
             70        80        90        100       110       120


  120       130       140       150       160       170
KpsU  DPALPVATLCHAISAAEAAEPSTVKVVVNTRQDALYFSRSPIPYPRNAEKARYLKHVGIY
      ::|      |  ::|::  :::|| :    :::   :   :|: :   |  ||:  :  |   :  :|:
NeuA  YQTLE-KYQC-VVSVTRSNKPSQIIRPLDDYSTLSFFDLDYSKYNRNS-IVEYHPNGAIF
             130       140       150       160       170


  180       190       200       210       220       230
KpsU  AYRRDVLQNYSQLPESMPEQAESLEQLRLMNAGINIRTFEVAATGPGVDTPACLEKVRAL
         :::
NeuA  IANKQHYLHTKHFFGRYSLAYIMDKESSLDIDDRMDFELAITIQQKKNRQKIDLYQNIHN
      180       190       200       210       220       230


  240
KpsU  MAQELAENA

NeuA  RINEKRNEFDSVSDITLIGHSLFDYWDVKKINDIEVNNLGIAGINSKEYYEYIIEKELIV
      240       250       260       270       280       290

NeuA  NFGEFVFIFFGTNDIVVSDWKKEDTLWYLKKTCQYIKKKNAASKIYLLSVPPVFGRIDRD
      300       310       320       330       340       350

NeuA  NRIINDLNSYLRENVDFAKFISLDHVLKDSYGNLNKMYTYDGLHFNSNGYTVLENEIAEI
      360       370       380       390       400       410

      VK
```

FIGURE 5.9 KpsU refers to the product of orf4 (kpsU). Numbers refer to the amino acid coordinates of the respective proteins. Double vertical lines align identical amino acids. Double dots align homologous amino acids. Horizontal lines represent gaps automatically inserted to maximize the alignment.

For *orf5* the first AUG start codon is preceded by a putative SD sequence. However, other in frame AUGs are also preceded by a putative SD sequence and therefore may represent alternative potential translation starts. A region I encoded protein has been reported with a calculated MW ranging in size from 72 to 77 KDa (Roberts et al., 1986; Silver et al., 1988; Boulnois and Jann, 1989). *orf5* is the only *orf*, in region I, which can encode a protein with this MW range. Translation from the first AUG of *orf4* yields a protein with a predicted MW of 75685 Da, whereas the other AUGs yield proteins with predicted MWs smaller than 62 KDa. On the basis of these considerations it is likely that the first AUG is the translation start. Computer analysis using the amino acid sequence of ORF5 identified 46.3% identity and 77.9% homology, over 512 amino acids, with the 58 KDa protein encoded by region III of the capsule gene cluster of *H.influenzae* type b (Brophy et al., unpublished results). ORF5 is also homologous, 61.5% over 135 amino acids, with the UgpC protein of the *sn*-glycerol-3-phosphate (G3P) uptake system of *Escherichia coli* (Overduin et al., 1988). UgpC contains a potential ATP binding site and has been proposed to function by coupling ATP-hydrolysis to the G3P uptake system. However, ORF5 does not contain any obvious potential ATP binding site.

Only the second AUG codon of *orf6* is preceded by a putative SD sequence which yields a protein with a predicted MW of 44 KDa. ORF6 is 38.1% identical and 75.3% homologous, over 373 amino acids, with the 44 KDa protein *H.influenzae* type b (Brophy et al., unpublished results). The exact function of the 44 KDa protein is not known. However, the 44 and 58 KDa proteins are encoded by a DNA region which is thought to be required for polysaccharide export in *H.influenzae* (Kroll et al., 1989). This, in conjunction with the knowledge that chemically different polysaccharides can be exported by a common system, suggests that the protein

homology between region III encoded products of *H.influenzae* with ORF5 and ORF6, might be due to similar roles played by these proteins in polysaccharide export. The homology between ORF6 and the 44 KDa protein encoded by region III of the capsule gene cluster of *H.influenzae* type b started before the second AUG of ORF6 and computer analysis was performed using the amino acid sequence of ORF6 taking into account the first AUG codon. ORF6 is 61.1% homologous over 239 amino acids with the UDP-glucuronosyltransferase of rat (Iyanagi *et al.*, 1986), and 56.8% homology over 162 amino acids with the UgpA protein of the *sn*-glycerol-3-phosphate (G3P) uptake system of *Escherichia coli* (Overduin *et al.*, 1988). UgpA has been proposed to be a membrane spanning protein involved, with an unknown function, in the uptake of G3P. However, hydropathy profile of ORF6 did not reveal any domain which could potentially span the membrane.

Thus, it has been possible to determine the AUG start codon for *orf*4 and assign the first AUG of *orf*3 and *orf*5 as the most likely translation start. More indeterminate is the possible translation start of *orf*2 and *orf*6. Data from computer analyses and protein data base searchs indicate the first AUG of *orf*2 as the most likely translation start codon. Whereas, the putative SD sequence present before the second AUG of *orf*6, idicates this codon as the most likely translation start of *orf*6. However, in the absence of the N-terminal sequence of purified proteins one has to be cautious in assigning potential translation start codon for *orf*2, *orf*3, *orf*5 and *orf*6, respectively. Taking into account the putative translation starts and the short gap between the various *orf*s, region I of the K5 capsular gene cluster might be organized into a single transcritional unit. However, more data are required before defining accurately the transcriptional organization of this region. Silver *et al.* (1988) have proposed for genes of region I,

of the K1 capsular gene cluster, encoding products of 40, 60 and 77 KDa the designation of *kps*E, *kps*D and *kps*C, respectively. Additionally, another gene, termed *kps*S, has been identified in region I and mapped 3' of *kps*C (Vimr *et al.*, 1989). In keeping with this nomenclature it is possible to assign *orf*2 to *kps*E, *orf*3 to *kps*D, *orf*5 to *kps*C and *orf*6 to *kps*S. For *orf*4 I propose the designation *kps*U (Figure 5.10).

FIGURE 5.10   REGION I OF THE K5 CAPSULAR GENE CLUSTER
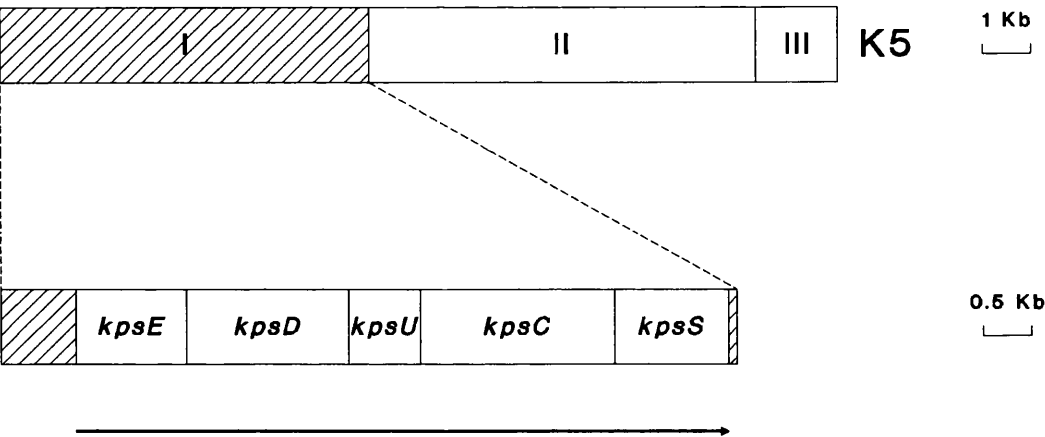
FIGURE 5.10 The organization of the K5 capsular gene cluster and the proposed organization of region I are shown as boxes. The scale is shown in kilobase pairs. Region I is enlarged on a scale 2:1 and broken lines highlight the enlargement. The arrow indicates the likely direction of transcription.

DNA 5' to *kps*E has been sequenced for the K7 capsular gene cluster (Smith *et al.*, unpublished results) and this region is 99% identical to that of the K5 capsular gene cluster. Computer analysis using the amino acid sequence of ORF1

81

identified 42.3% identity and 83.9% homology, over 137 amino acids, with the GutQ protein of *Escherichia coli* (Yamada *et al.*, 1990). GutQ has a biological function as yet not defined and is encoded by the last gene of the glucitol operon, which maps at 58 min in the *E.coli* chromosome (Yamada and Saier, 1987). The *kps* gene cluster for PSA synthysis in *E.coli* K1 was mapped near the unit 64 on the chromosome of *E.coli* K-12 (Vimr, 1991). *orf1* is 55.5% identical over 695 bp with *gulQ*. However, these levels of DNA homology are not sufficient to define *gulQ* as the chromosomal location of the *kps* gene cluster.

E.coli promoters contain two conserved motifs termed -35 (5')-TTGACA-(3') and -10 (5')-TATAAT-(3') consensus sequences, which are recognized by *E.coli* RNA polymerase for transcription initiation (Robenger and Court, 1979; Siebenlist *et al.*, 1980; Hawley and McClure, 1983;). An average *E.coli* promoter conserves 3.9 bp of the -35 consensus and 4.2 bp of the -10 consensus, with an inter-consensus spacing ranging from 15 to 21 bp (O'Neil, 1989). Individual bases within these consensus sequences are conserved to differing levels depending on many factors, for example the inter-consensus spacing (spacing class). In general, most of the available data take into account the three major spacing classes of 16, 17 and 18 bp (O'Neil, 1989). At this level of degeneracy, promoters would occur in one or the other direction approximately every 200 bp in a random sequence (Dykes *et al.*, 1975). Therefore it is difficult to identify only on the basis of the -35 and -10 consensus sequences, the position of an authentic promoter. However, the region 5' to *kps*E might contain some promoter activity since LE392(pH18), which retains this region, exhibits high levels of CMP-KDO synthetase only at 37°C whereas LE392(pBA6), which lacks this region, exhibits high levels of CMP-KDO synthetase regardless of the bacterial growth temperatures. Computer analysis identified four DNA

sequences, two of which retain 4 bp of the -35 consensus and the other two which retain 5 bp of the -10 consensus, 5' to *kps*E at coordinates 112, 138, 187 and 213 bp. These sequences have an inter-consensus spacing of 21 bp (Figure 5.13). It has not been possible, due to the absence of data for this spacing class, to compare the position of individual bases of these motifs with the percentage found in other -35 and -10 consensus sequences of *E.coli* promoters. Consequently, these motifs have to be interpreted as the best DNA sequences 5' to *kps*E which are most homologous to the -35 and -10 consensus sequences. No predicted stem and loop DNA structures, which could function in a rho-independent or rho-dependent transcription termination (Robenberg and Court, 1979), were identified 3' to *kps*S.

The proposed "cassette-model" to explain the antigenic diversity of group II capsules implies different mechanisms of recombination, which also included the possible presence of conserved DNA segments flanking region II (Boulnois and Jann, 1989). On the basis of this possibility the DNA sequence between region I and II for the K1 and K4 capsular gene clusters has been recently sequenced (Drake *et al.*, unpublished results). Computer analysis of these DNA sequences, together with that available for the K5 capsular gene cluster, did not reveal conserved DNA regions (Drake *et al.*, unpublished results). This result would tend to exclude recombinations via conserved DNA fragments flanking the region II, as a possible mechanism by which the antigenic diversity might have occurred. This result is apparently in contrast with the presence of "islands" of DNA homology identified in the junction between regions I and II of the capsule gene clusters of *H.influenzae* type b and c (Kroll *et al.*, 1990). Interestingly KpsS, as expected, is extremely conserved, over 98% identity, among the K1, K4 and K5 capsular gene clusters. However, KpsS

encoded by these capsular gene clusters contain a variable C-terminus. $KpsS_{K1}$ is twelve amino acids longer at the C-terminus than $KpsS_{K5}$, whereas $KpsS_{K4}$ is fourteen amino acids longer at the C-terminus than $KpsS_{K5}$ (Figure 5.11) (Drake *et al.*, unpublished results). Additionally, the twelve extra C-terminal amino acids of $KpsS_{K1}$ are not homologous to the fourteen extra C-terminal amino acids of $KpsS_{K4}$. Likewise, KpsT encoded by the K1 and K5 capsular gene clusters contain also a variable C-terminus. $KpsT_{K5}$ is six amino acids longer at the C-terminus than $KpsT_{K1}$ (Figure 5.11). The identification of variable C-termini in KpsS and KpsT, and the absence of homologous DNA segments between regions I and II from different K gene clusters, raises the possibility that recombination between an incoming and outcoming "region II-cassette" might occur by recombination events within *kps*S and *kps*T. It would be interesting to see if KpsS and KpsT from other capsular gene clusters contain also variable C-termini.

FIGURE 5.11   VARIABILITY OF THE C-TERMINI OF KpsS AND KpsT
IN DIFFERENT CAPSULE GENE CLUSTERS
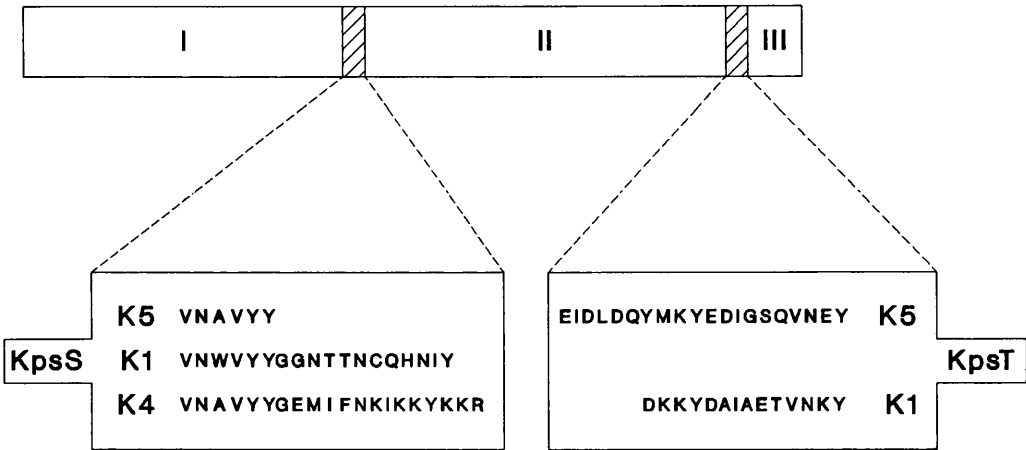
FIGURE 5.11 The organization of the *kps* locus of *Escherichia coli* is shown as boxes labelled I, II and III. Boxes with diagonal lines refer to the C-termini of KpsS and KpsT, respectively. The C-terminal sequences of KpsS and KpsT are shown in enlarged boxes. Broken lines highlight the enlargement.

5.3 MINICELL ANALYSIS

Minicells contain only plasmid DNA and newly synthesized proteins are mainly encoded by this DNA. DS410 carrying pH18, pBA6, pHEV1 or pHSP1 (see figure 5.2) was used to produce minicells harbouring these plasmids and ^{35}S-methionine was used to label proteins synthesized by these minicells (Figure 5.12b). Minicells from DS410(pUC18) and strain DS410(pACYC184) were used to distinguish betwen vector and non vector encoded products (figure 5.12a).

pH18 encoded four non vector proteins with relative MWs of 28, 38.5, 43 and 59 KDa. pHEV1 encoded three non vector proteins with relative MWs of 28, 38.5 and 59 kDa. Since pHEV1 contained the shortest DNA fragment which functionally defined the K5 region I, it was assumed that the extra protein of 43 KDa encoded by pH18 did not represent a region I product. This protein was subsequently shown to be encoded by a region II gene included in pH18 but not present in pHEV1 (see chapter 6). pHSP1, which includes kpsE and kpsD, encoded non vector polypeptides with relative MWs of 38.5 and 59 KDa. Whereas pBA6, which includes kpsD, kpsU and kpsC, encoded non vector polypetides with relative MWs of 28 KDa and 59 KDa.

Taking into account the predicted MWs of region I encoded products, the identified polypeptides encoded by pH18, pHEV1, pHSP1 and pBA6, and the organization of these plasmids it was possible to assign the 38.5 KDa polypeptide to kpsE, the 59 KDa to kpsD and the 28 KDa to kpsU. No protein which might have represented the kpsC and kpsS product has been identified in this analysis. A region I encoded protein with MW close to that predicted by kpsC has already been reported in previous studies (Roberts et al.,

1986; Silver *et al.*, 1988; Boulnois and Jann, 1989), whereas no data are available for the predicted *kps*S encoded product. Why minicells carrying pH18 or pHEV1 did not synthesize the *kps*C and *kps*S encoded products still remains unclear.
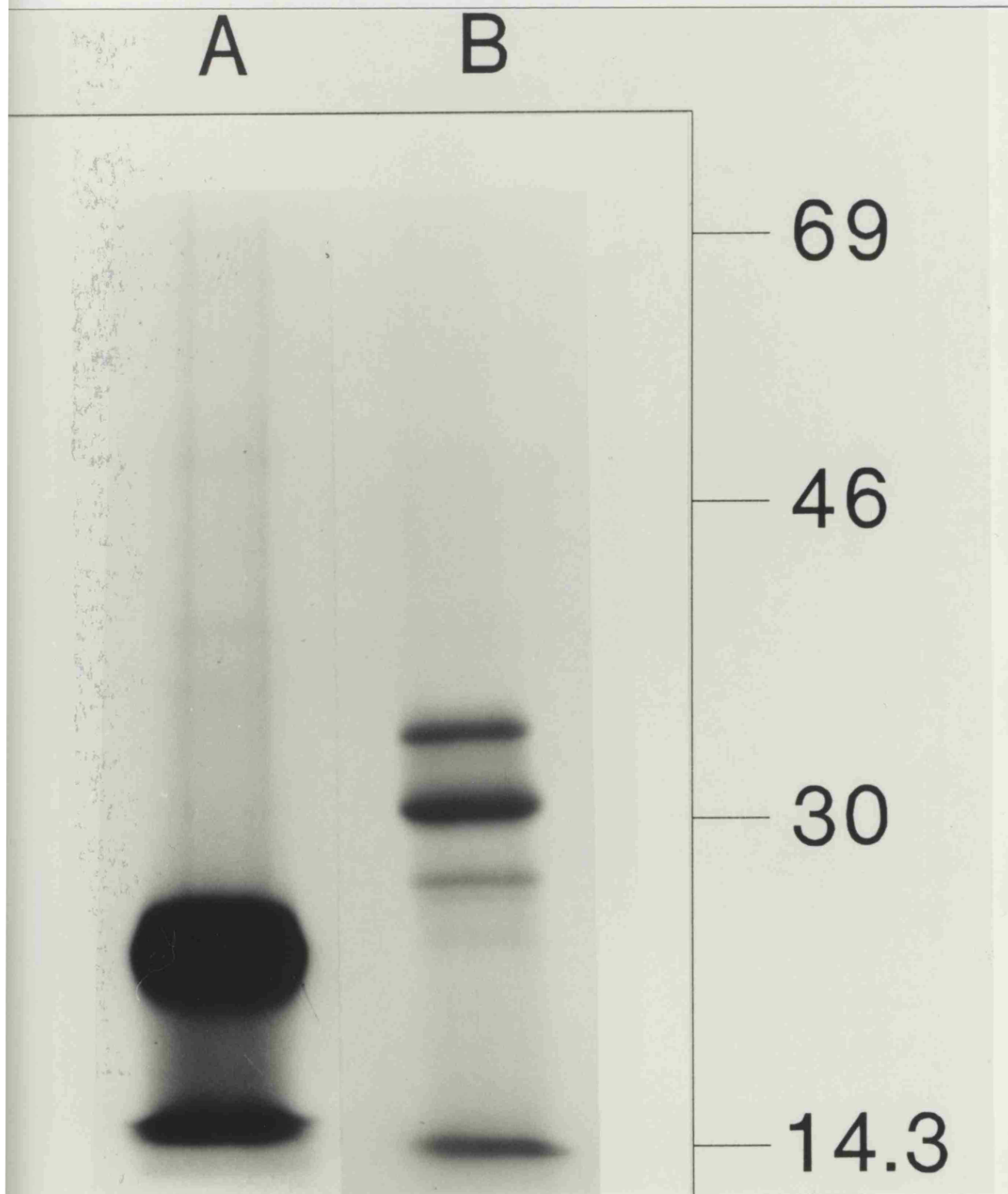
FIGURE 5.12a Lines A and B indicate protein patterns produced by minicells harbouring pACYC184 and pUC18, respectively. Numbers on the right side indicate MWs of protein markers.

FIGURE 5.12b PROTEIN PRODUCED IN MINICELLS HARBOURING pBA6, pHEV1, pH18 OR pHSP1.
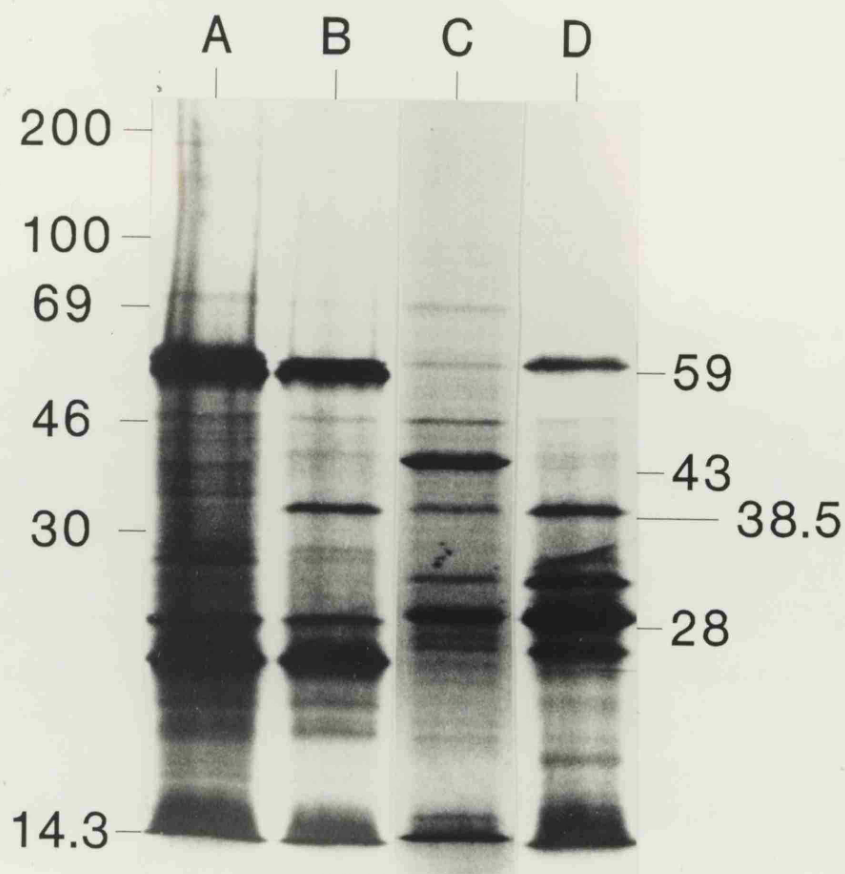
FIGURE 5.12b Lines A, B, C and D indicate protein patterns produced by minicells harbouring pBA6, pHEV1, pH18 or pHSP1, respectively. Numbers on the left side indicate MWs of protein markers, whereas numbers on the right side indicate MWs of the identified non vector proteins.

5.4 DISCUSSION

The region I of the K5 capsule gene cluster is 7.7 Kb and is composed of, at least, five genes possibly organized into a single transcriptional unit (Figure 5.13). Analysis of the protein pattern produced by minicells carrying different plasmids identified three of the five predicted proteins encoded by region I, namely kpsE, KpsD and KpsU. It was not possible to identify the other two predicted KpsC and KpsS proteins and the reason for this is as yet not known.

E.coli expressing group II capsules contains characteristically high level of CMP-KDO synthetase activity and region I was shown to be responsible for this high level. Data presented in this thesis demonstrated that the high level of this enzyme activity, produced by E.coli expressing the K5 capsule, is due to the presence of an additional CMP-KDO synthetase gene termed kpsU and located in the middle of region I. kpsU is 63.5% identical to kdsB, the CMP-KDO synthetase gene present in the chromosome of non-encapsulated E.coli. This homology strongly suggests a possible common origin of evolution for kpsU and kdsB. The presence of an additional CMP-KDO synthetase gene in the K5 capsular gene cluster will be debated later (see section 7.2). The high levels of CMP-KDO synthetase activity are only produced at the capsule permissive temperature of 37°C. Interestingly, pBA6 which lacks part of kpsE and the DNA region upstream of this gene, encoded high levels of CMP-KDO synthetase activity at the capsule restrictive and permissive temperatures of 18°C and 37°C, respectively. This result implies that DNA 5' to kpsE might play a role in determining the production of different levels of CMP-KDO synthetase activity at 18°C and 37°C. If this occurs at the level of transcription, it is likely that the DNA 5'

to *kps*E might be involved in regulation of *kps*U and probably of the entire region I. The region upstream of *kps*E is not necessary for K5 polysaccharide biosynthesis but appears to have some effect on the percentage of cells which are encapsulated (see section 7.2). Additionally, this region is highly conserved between the K5 and K7 gene clusters. These results suggest that the region 5' to *kps*E might play a regulatory role in group II polysaccharide biosynthesis. This possible regulatory role will be descussed later.

Computer analysis has identified homology between KpsE, KpsC and KpsS of region I with BexC, the 58 and 44 KDa proteins encoded by the capsule gene cluster of *H.influenzae* type b, respectively. These results, in conjunction with those obtained from studies on region III of the K5 and K1 capsular gene clusters (Smith *et al.*, 1990; Pavelka *et al.*, 1991), highlight two important aspects. Firstly, they re-enforce further the theory of a common evolutionary origin of encapsulation between *E.coli*, *H.influenzae* and *N.meningitidis*. The *kps* locus of *E.coli* encodes proteins which are homologous to those encoded by *cap* and *cps* loci of *H.influenzae* and *N.meningitidis*, respectively. There is the possibility that a common export system, which may function regardless of the polysaccharide chemical structure, may also be functionally interchangeable among these genera. Secondly, these results reveal for the first time evidence of differences between *kps*, *cap* and *cps* loci. KpsD and KpsU appear to be products uniquely encoded by the *kps* locus, since homology between these proteins and those encoded by *cap* and *cps* loci is either restricted to short amino acid stretches or longer amino acid stretches with a low percentage of amino acid identity. Also, the organization within specific regions of these loci appears different. The equivalent of region III

plus *kps*E and possibly *kps*D, of *E.coli*, is organized in *H.influenzae* and *N.meningitidis* into a single unit, termed region I and C, respectively. Whereas *kps*C and *kps*S, which are part of a larger transcriptional unit in the *kps* locus of *E.coli*, represent region I of the *cap* locus of *H.influenzae*.

In summary, it appears that *kps*, *cap* and *cps* loci may have derived from a common origin but, various selective pressures must have played a different role during the course of their subsequent evolution and generated remarkable diversities. The presence of specific proteins, such as KpsD or its equivalent, raises the question if these proteins carry out similar or genera-specific polysaccharide export steps. Similar questions are raised by the presence of the high levels of CMP-KDO synthetase activity. Are these levels of enzyme activity necessary for group II polysaccharide biosynthesis? Do they represent a specific feature of the *kps* gene cluster?

Region I products have homology with components belonging to binding-protein-dependent transport systems, which are closely related in terms of sequence, organization, structure, mechanism and probably evolutionary origin (Higgins *et al.*, 1990b). These results, together with data from region III, suggest that group II polysaccharide export might occur by a binding-protein-dependent transport system. Alternatively, or additionally, the homology between region I products and several proteins might reflect a direct relationship with the specific function carried out by these homologous proteins. For example, KpsE, KpsD, KpsC and KpsS show homology with proteins associated with different reactions involving PA or *sn*-glycerol-3-phosphate, which is a precursor of PA. Group II polysaccharides are linked to PA and are exported

regardless of their chemical structure. Thus, it is possible that these region I products might function in polysaccharide export via recognition of a common component such as PA.

Finally, very little is known about the sub-cellular location of region I products. However, on the basis of computer analyses and homologies with Bex and Ctr proteins, it is possible to hypothesise the most probable cellular location for region I products. KpsE appears to be a protein which spans the cytoplasmic membrane, with its hydrophilic central region orientated towards the periplasm and a membrane topology which is consistent with that of the oligotopic membrane proteins (Dalbey, 1990). KpsD is located in the periplasm. KpsU, KpsC and KpsS appear to be soluble proteins with a cytoplasmic location, nevertheless it is not possible to rule out that some of these proteins might be either directly bound to the cytoplasmic membrane or part of a protein complex associated with the inner membrane.

FIGURE 5.13 DNA SEQUENCE OF REGION I OF THE K5 CAPSULAR GENE CLUSTER

```
             10                    30                      50
Hind III
AAGCTT TCCATGGCGATCTGGGCATGATTACGCCTTACGATCTTCTGATCCTTATTTCTG
   A    F    H    G    D    L    G    M    I    T    P    Y    D    L    L    I    L    I    S    A


             70                    90                     110 (-35)
CCAGCGGTGAAACGGATGAAATCCTCAAGCTAGTTCCTTCACTGAAAAATTTCGGCAACC
   S    G    E    T    D    E    I    L    K    L    V    P    S    L    K    N    F    G    N    R


            130          (-10)      150                    170
GAATTATCGCCATTACCAATAATGGAAATTCCACGCTGGCGAAAAATGCTGATGCCGTGC
   I    I    A    I    T    N    N    G    N    S    T    L    A    K    N    A    D    A    V    L


            190                   210                     230
           (-35)                        (-10)
TGGAACTCCACATGGCGAATGAAACCTGCCCGAATAATCTTGCACCAACAACGTCTACCA
   E    L    H    M    A    N    E    T    C    P    N    N    L    A    P    T    T    S    T    T


            250                   270                     290
CGCTGACGATGGCGATCGGCGATGCGCTGGCGATTGCCATGATCCACCAACGCAAATTTA
   L    T    M    A    I    G    D    A    L    A    I    A    M    I    H    Q    R    K    F    M


            310                   330                     350
                                  Sma I
TGCCGAATGATTTTGCGCGCTATCACCCGGGCGGTTCATTAGGTCGTCGCCTGCTGACCC
   P    N    D    F    A    R    Y    H    P    G    G    S    L    G    R    R    L    L    T    R


            370                   390                     410
GCGTTGCTGATGTCATGCAGCATGATGTTCCTGCGGTACAGCTGGATGCGTCATTTAAAA
   V    A    D    V    M    Q    H    D    V    P    A    V    Q    L    D    A    S    F    K    T


            430                   450                     470
CCGTGATTCAACGTATCACCAGCGGATGCCAGGGAATGGTGATGGTAGAAGACGCAGAAG
   V    I    Q    R    I    T    S    G    C    Q    G    M    V    M    V    E    D    A    E    G


            490                   510                     530
GTGGGCTAGCGGGCATTATCACCGACGGTGACCTGCGTCGCTTTATGGAAAAAGAGGATT
   G    L    A    G    I    I    T    D    G    D    L    R    R    F    M    E    K    E    D    S


            550                   570                     590
CTCTGACATCCGCCACGGCTGCGCAGATGATGACACGTGAACCGCTGACGCTACCGGAAG
   L    T    S    A    T    A    A    Q    M    M    T    R    E    P    L    T    L    P    E    D


            610                   630                     650
ACACCATGATCATTGAAGCGGAAGAAAAAATGCAAAAGCACCGCGTCTCAACGTTATTGG
   T    M    I    I    E    A    E    E    K    M    Q    K    H    R    V    S    T    L    L    V


            670                   690                     710
TGACCAACAAGGCAAATAAAGTCACTGGCCTTGTGCGCATTTTCGACTAATTAAGCAACG
   T    N    K    A    N    K    V    T    G    L    V    R    I    F    D    *
```

94

```
          730                    750                    770
GGGTGCGAGAGGTTAGTCTCTCATTGAGTTTGCGCCCCACTGAAACTTTGATAATCGTTA

kpsE      790                    810                    830
CATGTTGATAAAAGTGAAGTCTGCCGTATCCTGGATGCGTGCTCGTCTGTCTGCCATCTC
 M   L   I   K   V   K   S   A   V   S   W   M   R   A   R   L   S   A   I   S

          850                    870                    890
ACTGGCAGATATCCAAAAACACCTGGCGAAAATCATCATTCTGGCACCGATGGCGGTGCT
 L   A   D   I   Q   K   H   L   A   K   I   I   I   L   A   P   M   A   V   L

          910                    930                    950
GCTGATCTATCTGGCTATCTTCAGCCAGCCTCGCTATATGAGCGAGTCGAAAGTCGCCAT
 L   I   Y   L   A   I   F   S   Q   P   R   Y   M   S   E   S   K   V   A   I

          970                    990                    1010
TAAACGCTCGGATGATTTAAACAGCGGCAGCCTGAATTTTGGTCTGCTTCTGGGTGCCTC
 K   R   S   D   D   L   N   S   G   S   L   N   F   G   L   L   L   G   A   S

          1030                   1050                   1070
TAACCCCAGTTCCGCAGAAGATGCGTTGTATCTGAAAGAGTACATCAACTCGCCGGATAT
 N   P   S   S   A   E   D   A   L   Y   L   K   E   Y   I   N   S   P   D   M

          1090                   1110                   1130
GCTGGCGGCGCTGGATAAGCAACTAAATTTTCGTGAAGCGTTTAGCCACAGCGGGCTCGA
 L   A   A   L   D   K   Q   L   N   F   R   E   A   F   S   H   S   G   L   D

          1150                   1170                   1190
TTTTCTTAATCATCTTAGCAAGGATGAAACCGCAGAAGGCTTCCTGAAGTACTACAAGGA
 F   L   N   H   L   S   K   D   E   T   A   E   G   F   L   K   Y   Y   K   D

          1210                   1230                   1250
CCGTATCAACGTCTCGTATGACGATAAAACCGGATTACTGAATATTCAGACGCAGGGCTT
 R   I   N   V   S   Y   D   D   K   T   G   L   L   N   I   Q   T   Q   G   F

          1270                   1290                   1310
TAGCCCGGAGTTTGCGCTTAAGTTTAACCAAACCGTGCTGAAAGAGTCAGAGCGCTTTAT
 S   P   E   F   A   L   K   F   N   Q   T   V   L   K   E   S   E   R   F   I

          1330                   1350                   1370
CAATGAGATGTCACATCGCATCGCGCGTGACCAGCTTGCCTTTGCAGAAACGGAGATGGA
 N   E   M   S   H   R   I   A   R   D   Q   L   A   F   A   E   T   E   M   E

          1390                   1410                   1430
AAAGGCACGCCAGCGTCTGGACGCCAGCAAAGCGGAATTGCTCTCTTATCAGGACAACAA
 K   A   R   Q   R   L   D   A   S   K   A   E   L   L   S   Y   Q   D   N   N

          1450BamH I            1470                   1490
CAACGTTCTGGATCCACAGGCACAGGCACAGGCGGCGAGCACGTTAGTGAATACGCTGAT
 N   V   L   D   P   Q   A   Q   A   Q   A   A   S   T   L   V   N   T   L   M

          1510                   1530                   1550
GGGCCAGAAGATCCAGATGGAAGCGGACCTGCGGAACTTGCTGACGTATCTGCGTGAGGA
 G   Q   K   I   Q   M   E   A   D   L   R   N   L   L   T   Y   L   R   E   D
```

95

```
          1570                 1590                 1610
CGCCCCGCAAGTTGTGAGTGCGCGTAATGCGATTCAGTCATTGCAGGCACAAATTGACGA
 A   P   Q   V   V   S   A   R   N   A   I   Q   S   L   Q   A   Q   I   D   E

          1630                 1650                 1670
AGAAAAAAGCAAAATCACTGCGCCACAGGGTGACAAGCTAAACCGTATGGCAGTGGATTT
 E   K   S   K   I   T   A   P   Q   G   D   K   L   N   R   M   A   V   D   F

          1690                 1710                 1730
TGAAGAAATCAAATCAAAAGTAGAGTTCAACACCGAGCTGTACAAACTGACCCTGACCTC
 E   E   I   K   S   K   V   E   F   N   T   E   L   Y   K   L   T   L   T   S

          1750                 1770                 1790
CATTGAAAAGACCCGTGTAGAAGCGGCTCGTAAGCTCAAGGTGCTGTCAGTGATCAGTTC
 I   E   K   T   R   V   E   A   A   R   K   L   K   V   L   S   V   I   S   S

          1810                 1830                 1850
GCCACAGTTGCCGCAGGAATCGTCTTTTCCAAATATCCCTTATTTGATCGCCTGCTGGTT
 P   Q   L   P   Q   E   S   S   F   P   N   I   P   Y   L   I   A   C   W   L

          1870                 1890                 1910
ACTGGTGTGCTGCCTGCTGTTCGGCACCCTGAAACTGTTGCTGGCTGTTATTGAAGATCA
 L   V   C   L   L   F   G   T   L   K   L   L   L   A   V   I   E   D   H

          1930                 1950                 1970
                            SD            kpsD
CCGAGACTAACGCTGTCGCTGAATGAGTTTGTGATGAAATTATTTAAATCAATTTTACTG
 R   D   *                                M   K   L   F   K   S   I   L   L

          1990                 2010                 2030
ATTGCCGCCTGTCACGCGGCGCAGGCCAGCGCGACCATTGATATTAACGCTGACCCAAAC
 I   A   A   C   H   A   A   Q   A   S   A   T   I   D   I   N   A   D   P   N

          2050                 2070                 2090
CTGACAGGAGCCGCGCCGCTTACCGGTATTCTGAACGGACAGAAGTCGGATACGCAAAAC
 L   T   G   A   A   P   L   T   G   I   L   N   G   Q   K   S   D   T   Q   N

          2110                 2130                 2150
ATGAGCGGCTTCGACAATACCCCGCCGCCCGCACCGCCGGTGGTAATGAGCCGTATGTTT
 M   S   G   F   D   N   T   P   P   P   A   P   P   V   V   M   S   R   M   F

          2170                 2190                 2210
GGTGCTCAACTTTTCAACGGCACCAGCGCGGATAGCGGTGCGACGGTAGGATTCAACCCT
 G   A   Q   L   F   N   G   T   S   A   D   S   G   A   T   V   G   F   N   P

          2230                 2250                 2270
GACTATATTCTGAATCCGGGTGATAGCATTCAGGTTCGCTTGTGGGGTGCGTTCACCTTT
 D   Y   I   L   N   P   G   D   S   I   Q   V   R   L   W   G   A   F   T   F

          2290                 2310                 2330
GATGGTGCGTTACAGGTTGATCCCAAAGGTAATATTTTCCTGCCGAACGTTGGTCCGGTG
 D   G   A   L   Q   V   D   P   K   G   N   I   F   L   P   N   V   G   P   V

          2350                 2370                 2390
AAAATTGCTGGCGTCAGTAATAGTCAGCTAAATGCCTTGGTCACATCCAAAGTGAAGGAA
 K   I   A   G   V   S   N   S   Q   L   N   A   L   V   T   S   K   V   K   E
```

96

```
         2410                    2430                    2450
GTATACCAGTCCAACGTCAACGTCTACGCCTCCTTATTACAGGCGCAGCCAGTAAAAGTG
V   Y   Q   S   N   V   N   V   Y   A   S   L   L   Q   A   Q   P   V   K   V

         2470                    2490                    2510
TACGTGACCGGATTTGTGCGTAATCCTGGTCTGTATGGCGGTGTGACGTCTGATTCGTTA
Y   V   T   G   F   V   R   N   P   G   L   Y   G   G   V   T   S   D   S   L

         2530                    2550                    2570
CTCAATTATCTGATCAAGGCTGGCGGCGTTGATCCAGAGCGCGGAAGTTACGTTGATATT
L   N   Y   L   I   K   A   G   G   V   D   P   E   R   G   S   Y   V   D   I

         2590                    2610                    2630
GTGGTCAAGCGCGGTAACCGCGTGCGCTCCAACGTCAACCTGTACGACTTCCTGCTGAAC
V   V   K   R   G   N   R   V   R   S   N   V   N   L   Y   D   F   L   L   N

         2650                    2670                    2690
GGCAAACTGGGACTTTCGCAGTTCGCCGATGGTGACACCATCATCGTCGGGCCACGTCAG
G   K   L   G   L   S   Q   F   A   D   G   D   T   I   I   V   G   P   R   Q

         2710                    2730                    2750
CATACTTTCAGCGTTCAGGGCGATGTCTTTAACAGCTACGACTTTGAGTTCCGCGAAAGC
H   T   F   S   V   Q   G   D   V   F   N   S   Y   D   F   E   F   R   E   S

         2770                    2790                    2810
AGCATTCCCGTAACGGAAGCGTTGAGCTGGGCGCGCCCTAAGCCTGGCGCGACTCACATT
S   I   P   V   T   E   A   L   S   W   A   R   P   K   P   G   A   T   H   I

         2830                    2850                    2870
ACGATTATGCGTAAACAGGGGCTGCAAAAACGCAGCGAATACTATCCGATCAGTTCTGCG
T   I   M   R   K   Q   G   L   Q   K   R   S   E   Y   Y   P   I   S   S   A

         2890                    2910                    2930
CCAGGCCGTATGTTGCAAAATGGCGATACCTTAATCGTGAGCACTGACCGCTATGCCGGC
P   G   R   M   L   Q   N   G   D   T   L   I   V   S   T   D   R   Y   A   G

         2950                    2970                    2990
ACCATTCAGGTGCGGGTTGAAGGCGCACACTCCGGTGAACATGCCATGGTATTGCCTTAT
T   I   Q   V   R   V   E   G   A   H   S   G   E   H   A   M   V   L   P   Y

         3010                    3030                    3050
GGTTCCACTATGCGTGCGGTTCTGGAAAAAGTCCGCCCGAACAGCATGTCGCAGATGAAC
G   S   T   M   R   A   V   L   E   K   V   R   P   N   S   M   S   Q   M   N

         3070                    3090                    3110
GCAGTTCAGCTTTATCGCCCATCAGTGGCTCAGCGTCAGAAAGAGATGCTGAATCTCTCG
A   V   Q   L   Y   R   P   S   V   A   Q   R   Q   K   E   M   L   N   L   S

         3130                    3150                    3170
CTGCAAAAGCTGGAGGAAGCCTCACTTTCTGCCCAGTCATCCACCAAAGAAGAAGCCAGC
L   Q   K   L   E   E   A   S   L   S   A   Q   S   S   T   K   E   E   A   S

         3190                    3210                    3230
CTGCGAATGCAGGAAGCACAACTGATCAGCCGCTTTGTGGCGAAAGCACGCACCGTGGTT
L   R   M   Q   E   A   Q   L   I   S   R   F   V   A   K   A   R   T   V   V
```

```
          3250                 3270                 3290
CCGAAAGGTGAAGTGATCCTCAACGAATCCAATATTGATTCTGTTCTGCTTGAAGATGGC
P   K   G   E   V   I   L   N   E   S   N   I   D   S   V   L   L   E   D   G

          3310                 3330                 3350
GACGTCATCAATATTCCGGAGAAAACATCGCTGGTTATGGTTCATGGCGAGGTGCTGTTC
D   V   I   N   I   P   E   K   T   S   L   V   M   V   H   G   E   V   L   F

          3370                 3390                 3410
CCGAACGCGGTGAGCTGGCAGAAAGGGATGACCACCGAGGATTACATCGAGAAATGCGGT
P   N   A   V   S   W   Q   K   G   M   T   T   E   D   Y   I   E   K   C   G

          3430                 3450                 3470
GGCCTGACGCAAAAATCGGGTAACGCCAGAATTATCGTCATTCGTCAGAACGGTGCGCGA
G   L   T   Q   K   S   G   N   A   R   I   I   V   I   R   Q   N   G   A   R

          3490                 3510                 3530
GTCAACGCTGAAGATGTAGATTCACTCAAGCCTGGCGATGAGATTATGGTTCTGCCGAAA
V   N   A   E   D   V   D   S   L   K   P   G   D   E   I   M   V   L   P   K

          3550                 3570                 3590
TATGAATCGAAAAACATTGAAGTTACCCGTGGTATTTCCACCATCCTCTATCAGCTGGCG
Y   E   S   K   N   I   E   V   T   R   G   I   S   T   I   L   Y   Q   L   A

          3610                 3630                 3650
                             SD            KpsU
GTGGGTGCAAAAGTGATTCTGTCTTTGTAAGGGGTTGAAATGAGCAAAGCAGTTATTGTC
V   G   A   K   V   I   L   S   L   *                 M   S   K   A   V   I   V

          3670                 3690                 3710
ATTCCGGCTCGCTATGGCTCCTCGCGCCTGCCGGGTAAGCCACTGCTCGATATTGTTGGT
I   P   A   R   Y   G   S   S   R   L   P   G   K   P   L   L   D   I   V   G

          3730                 3750                 3770
AAACCGATGATCCAGCATGTTTACGAGCGTGCGTTACAGGTGGCGGGCGTTGCGGAAGTT
K   P   M   I   Q   H   V   Y   E   R   A   L   Q   V   A   G   V   A   E   V

          3790                 3810                 3830
TGGGTGGCAACAGACGATCCGCGTGTTGAACAGGCCGTACAGGCGTTTGGCGGGAAAGCC
W   V   A   T   D   D   P   R   V   E   Q   A   V   Q   A   F   G   G   K   A

          3850                 3870                 3890
ATCATGACGCGCAACGATCATGAATCCGGCACCGATCGGCTGGTCGAGGTGATGCATAAA
I   M   T   R   N   D   H   E   S   G   T   D   R   L   V   E   V   M   H   K

          3910                 3930                 3950
                        Pst I
GTCGAGGCAGATATCTACATTAACCTGCAGGGCGACGAACCAATGATTCGCCCGCGGGAT
V   E   A   D   I   Y   I   N   L   Q   G   D   E   P   M   I   R   P   R   D

          3970                 3990                 4010
GTAGAAACGCTGCTACAAGGAATGCGTGACGATCCCGCGTTGCCGGTGGCAACGCTATGC
V   E   T   L   L   Q   G   M   R   D   D   P   A   L   P   V   A   T   L   C
```

98

```
         4030                  4050                  4070
CACGCGATTTCTGCCGCAGAAGCGGCTGAGCCAAGCACGGTAAAAGTGGTTGTGAATACC
H   A   I   S   A   A   E   A   A   E   P   S   T   V   K   V   V   V   N   T

         4090                  4110                  4130
CGCCAGGATGCGCTTTATTTCAGCCGCTCACCGATTCCGTATCCGCGTAATGCTGAAAAA
R   Q   D   A   L   Y   F   S   R   S   P   I   P   Y   P   R   N   A   E   K

         4150                  4170                  4190
GCGCGCTACCTGAAACACGTCGGTATTTACGCTTATCGTCGCGATGTGCTGCAAAACTAC
A   R   Y   L   K   H   V   G   I   Y   A   Y   R   R   D   V   L   Q   N   Y

         4210                  4230                  4250
AGCCAGTTACCGGAGTCCATGCCGGAGCAGGCGGAATCACTGGAGCAGCTAAGGTTGATG
S   Q   L   P   E   S   M   P   E   Q   A   E   S   L   E   Q   L   R   L   M

         4270                  4290                  4310
AACGCGGGGATCAACATCCGCACATTTGAGGTTGCCGCAACCGGTCCGGGCGTCGACACC
N   A   G   I   N   I   R   T   F   E   V   A   A   T   G   P   G   V   D   T

         4330                  4350                  4370
         Sph I                            SD                    KpcC
CCAGCATGCCTGGAAAAAGTGCGCGCCCTGATGGCACAGGAACTGGCTGAAAACGCATGA
P   A   C   L   E   K   V   R   A   L   M   A   Q   E   L   A   E   N   A   *
                                                              M   I

         4390                  4410                  4430
TTGGCATTTACTCGCCTGGCATCTGGCGTATTCCGCATCTGGAGAAATTTCTGGCGCAAC
   G   I   Y   S   P   G   I   W   R   I   P   H   L   E   K   F   L   A   Q   P

         4450                  4470                  4490
CGTGCCAGAAACTTTCTCTGCTGCGCCCTGTTCCACAAGAAGTTGATGCTATCGCCGTGT
   C   Q   K   L   S   L   L   R   P   V   P   Q   E   V   D   A   I   A   V   W

         4510                  4530                  4550
GGGGACATCGTCCCAGCGCGGCGAAACCAGTCGCCATCGCCAAAGCAGCGGGAAAACCCG
   G   H   R   P   S   A   A   K   P   V   A   I   A   K   A   A   G   K   P   V

         4570                  4590                  4610
TCATTCGTCTGGAAGATGGATTTGTGCGTTCGCTGGATCTTGGCGTCAATGGCGAGCCGC
   I   R   L   E   D   G   F   V   R   S   L   D   L   G   V   N   G   E   P   P

         4630                  4650                  4670
CGCTTTCTCTGGTGGTGGATGATTATTGCATTTACTACGATGCCAGCAAGCCTTCAGCAC
   L   S   L   V   V   D   D   Y   C   I   Y   Y   D   A   S   K   P   S   A   L

         4690                  4710                  4730
TGGAGAAACTGGTACAGGATAAAGCCGGAAATACAGCTCTGATAAGCCAGGCCAGAGAAG
   E   K   L   V   Q   D   K   A   G   N   T   A   L   I   S   Q   A   R   E   A

         4750                  4770                  4790
CGATGCACACCATCGTGACCGGGGATTTGTCGAAATATAACCTGGCACCTGCGTTTGTGG
   M   H   T   I   V   T   G   D   L   S   K   Y   N   L   A   P   A   F   V   A
```

```
        4810                    4830                    4850
CTGATGAGTCAGAACGTTCAGACATCGTTCTGGTTGTCGATCAGACATTTAATGATATGT
  D   E   S   E   R   S   D   I   V   L   V   V   D   Q   T   F   N   D   M   S

        4870                    4890                    4910
CAGTGACGTATGGCAATGCTGGCCCGCATGAGTTTGCTGCCATGCTGGAAGCCGCGATGG
  V   T   Y   G   N   A   G   P   H   E   F   A   A   M   L   E   A   A   M   A

        4930                    4950                    4970
CGGAAAATCCTCAAGCCGAAATTTGGGTGAAGGTGCATCCGGATGTCCTGGAAGGAAAGA
  E   N   P   Q   A   E   I   W   V   K   V   H   P   D   V   L   E   G   K   K

        4990                    5010                    5030
AAACAGGTTATTTCGCCGCTCTGCGCGCCACGCAACGAGTACGTTTAATTGCCGAGAATG
  T   G   Y   F   A   A   L   R   A   T   Q   R   V   R   L   I   A   E   N   V

        5050                    5070 Sma I                5090
TCAGCCCGCAGTCGCTGTTGCGACACGTTTCCCGGGTTTACGTCGTGACCTCCCAGTACG
  S   P   Q   S   L   L   R   H   V   S   R   V   Y   V   V   T   S   Q   Y   G

        5110                    5130                    5150
GCTTTGAAGCCTTGCTGGCAGGAAAACCGGTGACATGCTTCGGCCAGCCCTGGTATGCAG
  F   E   A   L   L   A   G   K   P   V   T   C   F   G   Q   P   W   Y   A   G

        5170                    5190                    5210
GCTGGGGCTTAACCGACGATCGCCATCCGCAGTCCGCTTTGTTATCTGCCCGACGCGGTT
  W   G   L   T   D   D   R   H   P   Q   S   A   L   L   S   A   R   R   G   S

        5230                    5250                    5270
CTGCCACGCTGGAGGAACTTTTTGCCGCTGCATACCTGCGTTACTGTCGCTATATCGATC
  A   T   L   E   E   L   F   A   A   A   Y   L   R   Y   C   R   Y   I   D   P

        5290                    5310                    5330
CGCAAACGGGAGAAGTAAGCAATCTATTTACCGTGCTGCAATGGCTGCAATTACAACGTC
  Q   T   G   E   V   S   N   L   F   T   V   L   Q   W   L   Q   L   Q   R   R

        5350                    5370                    5390
GACATCTGCAACAGCGTAATGGTTATTTATGGGCGCCAGGCTTAACGCTGTGGAAGTCAG
  H   L   Q   Q   R   N   G   Y   L   W   A   P   G   L   T   L   W   K   S   A

        5410                    5430                    5450
CGATCCTGAAACCTTTCTTGCAAACGGCAACAAACCGGCTGAGTTTTTCACGTCGCTGTA
  I   L   K   P   F   L   Q   T   A   T   N   R   L   S   F   S   R   R   C   T

        5470                    5490                    5510
CTGCGGCGAGCGCCTGCGTGGTATGGGGGTGTAAAGGGAGAACAGCAATGGCGAGCCGAAG
  A   A   S   A   C   V   V   W   G   V   K   G   E   Q   Q   W   R   A   E   A

        5530                    5550                    5570
CGCAGCGAAAATCACTGCCGTTATGGCGAATGGAAGATGGTTTTCTACGTTCATCCGGAC
  Q   R   K   S   L   P   L   W   R   M   E   D   G   F   L   R   S   S   G   L

        5590                    5610                    5630
TTGGCTCTGACCTTTTGCCGCCGCTATCGTTGGTCCTGGATAAACGCGGGATCTACTATG
  G   S   D   L   L   P   P   L   S   L   V   L   D   K   R   G   I   Y   Y   D
```

100

```
            5650                5670                5690
ACGCCACGCGCCCCAGCGACCTGGAAGTGCTGCTTAATCACAGCCAGTTAACACTGGCGC
 A   T   R   P   S   D   L   E   V   L   L   N   H   S   Q   L   T   L   A   Q

            5710                5730                5750
AGCAGATGCGAGCTGAAAAAATTACGCCAGCGACTGGTTGAAAGCAAACTGAGCAAATACA
 Q   M   R   A   E   K   L   R   Q   R   L   V   E   S   K   L   S   K   Y   N

            5770                5790                5810
ACCTGGGGGCCGATTTCTCTCTGCCTGCCGAGGCCAAAGATAAAAAAATCATCCTGGTGC
 L   G   A   D   F   S   L   P   A   E   A   K   D   K   K   I   I   L   V   P

            5830                5850                5870
CGGGTCAGGTAGAAGACGATGCCTCAATTAAAACAGGCACTGTGTCGATTAAGAGCAACC
 G   Q   V   E   D   D   A   S   I   K   T   G   T   V   S   I   K   S   N   L

            5890                5910                5930
TTGAGTTATTACGCACAGTACGCGAGCGCAATCCGCACGCTTACATTATTTATAAACCGC
 E   L   L   R   T   V   R   E   R   N   P   H   A   Y   I   I   Y   K   P   H

            5950                5970                5990
ACCCGGATGTACTGGTGGGGAATCGCAAGGGCAATATTCCGACAGAACTAATTGCTGAAC
 P   D   V   L   V   G   N   R   K   G   N   I   P   T   E   L   I   A   E   L

            6010                6030                6050
TCGCTGATTATCAGGCACTGGACGCAGATATTATTCAATGCATTCAGCGCGCAGATGAAG
 A   D   Y   Q   A   L   D   A   D   I   I   Q   C   I   Q   R   A   D   E   V

            6070                6090                6110
TGCACACCATGACATCATTGTCCGGGTTTGAAGCGTTATTACATGGCAAGCAAGTTCATT
 H   T   M   T   S   L   S   G   F   E   A   L   L   H   G   K   Q   V   H   C

            6130                6150                6170
GTTACGGCCTGCCCTTCTATGCCGGTTGGGGTTTAACCGTTGATGAACATCACTGCCCGC
 Y   G   L   P   F   Y   A   G   W   G   L   T   V   D   E   H   H   C   P   R

            6190                6210                6230
GTCGCGAGCAAAAATTAACGATAGCAGATTTGATCTATCAGACGTTGATTGTTTATCCAA
 R   E   Q   K   L   T   I   A   D   L   I   Y   Q   T   L   I   V   Y   P   T

            6250                6270                6290
CCTATATCCACCCAACACTGCTACAACCTATATCGGTTGAAGAGGCGGCGGAATATTTGA
 Y   I   H   P   T   L   L   Q   P   I   S   V   E   E   A   A   E   Y   L   I

            6310                6330                6350
TCCAGACGCCGCGCAAGCCGATGTTTATTACCCGAAAAAAAGCGGGACGGGTAATACGCT
 Q   T   P   R   K   P   M   F   I   T   R   K   K   A   G   R   V   I   R   Y

            6370                6390                6410
ATTACCGCAAATTAATTATGTTCTGCAAGGTCAGATTTGGCTAAACAATTTCACAGTTGA
 Y   R   K   L   I   M   F   C   K   V   R   F   G   *

            6430      Kps8      6450                6470
CTATTACACAAATTATTTATGCAAGGTAATGCACTAACCGTTTTATTATCCGGTAAAAAA
                      M   Q   G   N   A   L   T   V   L   L   S   G   K   K
```

101

```
        6490   SD                  6510                    6530
TATCTGCTATTGCAGGGGCCAATGGGACCCTTTTTCAGTGATGTTGCCGAGTGGCTAGAG
Y   L   L   L   Q   G   P   M   G   P   F   F   S   D   V   A   E   W   L   E

        6550                    6570                    6590
TCATTAGGTCGTAACGCTGTGAATGTTGTATTCAACGGTGGGGATCGTTTTTACTGCCGC
S   L   G   R   N   A   V   N   V   V   F   N   G   G   D   R   F   Y   C   R

        6610                    6630                    6650
CATCGACAATACCTAGCTTACTACCAGACACCGAAAGAGTTTCCCGGATGGTTACGGGAT
H   R   Q   Y   L   A   Y   Y   Q   T   P   K   E   F   P   G   W   L   R   D

        6670                    6690                    6710
CTCCACCGGCAATATGACTTTGACACAATCCTCTGCTTTGGCGACTGCCGCCCATTGCAT
L   H   R   Q   Y   D   F   D   T   I   L   C   F   G   D   C   R   P   L   H

        6730                    6750   BamH I              6770
AAAGAAGCAAAACGCTGGGCAAAGGCGAAAGGGATCCGCTTCCTGGCATTTGAAGAAGGA
K   E   A   K   R   W   A   K   A   K   G   I   R   F   L   A   F   E   E   G

        6790                    6810                    6830
TATTTACGCCCGCAATTTATTACCGTTGAAGAAGGCGGAGTGAACGCATATTCATCGCTA
Y   L   R   P   Q   F   I   T   V   E   E   G   G   V   N   A   Y   S   S   L

        6850                    6870                    6890
CCGCGCGATCCGGATTTTTATCGTAAGTTACCAGATATGCCTACGCCGCACGTTGAGAAC
P   R   D   P   D   F   Y   R   K   L   P   D   M   P   T   P   H   V   E   N

        6910                    6930                    6950
TTAAAACCTTCAACGATGAAACGTATAGGCCATGCTATGTGGTATTACCTGATGGGCTGG
L   K   P   S   T   M   K   R   I   G   H   A   M   W   Y   Y   L   M   G   W

        6970                    6990                    7010
CATTACCGTCATGAGTTTCCTCGCTACCGCCACCACAAATCATTTTCCCCCTGGTATGAA
H   Y   R   H   E   F   P   R   Y   R   H   H   K   S   F   S   P   W   Y   E

        7030                    7050                    7070
GCACGTTGCTGGGTTCGTGCATACTGGCGCAAGCAACTTTACAAGGTAACACAGCGTAAG
A   R   C   W   V   R   A   Y   W   R   K   Q   L   Y   K   V   T   Q   R   K

        7090                    7110                    7130
GTATTACCGAGGTTAATGAACGAACTGGACCAGCGTTATTATCTTGCTGTTTTGCAGGTG
V   L   P   R   L   M   N   E   L   D   Q   R   Y   Y   L   A   V   L   Q   V

        7150                    7170                    7190
TATAACGATAGCCAGATTCGTAACCACAGCAGTTATAACGATGTGCGTGACTATATTAAT
Y   N   D   S   Q   I   R   N   H   S   S   Y   N   D   V   R   D   Y   I   N

        7210                    7230                    7250
GAAGTCATGTACTCATTTTCGCGTAAAGCGCCGAAAGAAAGTTATTTGGTGATCAAACAT
E   V   M   Y   S   F   S   R   K   A   P   K   E   S   Y   L   V   I   K   H

        7270                    7290                    7310
CATCCGATGGATCGTGGTCACAGACTCTATCGACCATTAATTAAACGGTTGAGTAAGGAA
H   P   M   D   R   G   H   R   L   Y   R   P   L   I   K   R   L   S   K   E
```

```
       7330                   7350                   7370
TATGGCTTAGGTGAGCGAATCCTTTATGTGCACGATCTCCCGATGCCGGAATTATTACGC
Y  G  L  G  E  R  I  L  Y  V  H  D  L  P  M  P  E  L  L  R

       7390                   7410                   7430
CATGCAAAAGCGGTGGTGACGATTAACAGTACGGCGGGGATCTCTGCGCTGATTCATAAC
H  A  K  A  V  V  T  I  N  S  T  A  G  I  S  A  L  I  H  N

       7450                   7470                   7490
AAACCACTCAAAGTGATGGGCAATGCCCTGTACGACATCAAGGGCTTGACGTATCAAGGG
K  P  L  K  V  M  G  N  A  L  Y  D  I  K  G  L  T  Y  Q  G

       7510                   7530                   7550
CATTTGCACCAGTTCTGGCAGGCTGATTTTAAACCAGATATGAAACTGTTTAAGAAGTTT
H  L  H  Q  F  W  Q  A  D  F  K  P  D  M  K  L  F  K  K  F

       7570                   7590                   7610
CGTGGGTATTTATTGGTGAAGACGCAGGTTAATGCGGTTTATTATTAACATCTTATCAAC
R  G  Y  L  L  V  K  T  Q  V  N  A  V  Y  Y  *

       7630                   7650                   7670
AATCGACATCTCCATTATATCAATCATAACAGATGGATGATACCTTTATTAGATGTAAAA

       7690                   7710        EcoR V
AGATAAAGTATTTTAGATATTACTAATATATATTAGATATC
```

FIGURE 5.13 The DNA sequence of the 7.7 Kb *Hind*III-*Eco*RV fragment is shown with capital letters. Numbers indicate nucleotide coordinates. Amino acid translation for each *orf* is shown underneath the nucleotide sequence in the single letter code. Restriction sites are shown, above the nucleotide sequence identified by empty characters. The DNA sequences, identified 5' to *kps*E, which are most homologous to the -35 and -10 consensus sequences, are shown with bold characters. The proposed designation for region I genes is shown highlighted over the first ATG of the respective *orf*. Potential translational start codons and the encoded methionine (M) are shown with highlighted and bold characters. Potential Shine Dalgarno sequences (SD) are highlighted with double lines. Stars indicate translational stop codons.

CHAPTER 6

REGION II OF THE K5 CAPSULAR GENE CLUSTER

Region II of the *kps* locus is capsule specific and is thought to encode functions necessary for group II polysaccharide synthesis, since mutations located in this region abolish the production of the encoded polysaccharide. Polysaccharide synthesis is still a poorly defined process. For example, very little is known about the number of proteins required for synthesis of a specific capsular polysaccharide. Some polysaccharides, such as K1 and K92, are homopolymers of the same repeating unit (NeuA). However, the linkage between these repeating units varies, K1 has a sialyl $\alpha2,8$-linkage (Barry, 1959; Dewitt and Rowe, 1961) whereas K92 has a sialyl $\alpha2,9$-linkage (Glode *et al.*, 1977). The region II determinants of both the K1 and K92 capsular gene clusters are very similar (Vimr *et al.* 1992), suggesting involvement of similar proteins in the biosynthesis of these polymers (Vimr *et al.*, 1992). For example, the sialyltransferases encoded by the K1 and K92 capsular gene clusters are 83% identical and 92% homologous over 409 amino acids (Vimr *et al.*, 1992). This raises the question of how highly homologous transferases can perform similar reactions but involving different chemical bonds. Most capsular polysaccharides are heteropolymers, implying the involvement of more than one transferase. How can these transferases insert accurately sugars in a growing polysaccharide chain? Additionally, polysaccharide biosynthesis is thought to occur at the inner phase of the cytoplasmic membrane (Janas and Troy,

1989) and preliminary results have suggested that proteins involved in polysaccharide export might also play a role albeit indirectly in the process of polymerization (Vimr *et al.*, 1989). This suggests that polysaccharide polymerization and export might be coupled, although the mechanism is not known. Finally, very little is known about initiation of polymerization. Does this process, in the synthesis of different capsular polysaccharides, require specific or common functions?

A molecular study of region II of the K5 capsular gene cluster would provide useful data about the number of its encoded products and preliminary information about their possible function. Additionally, information provided by this study might also be further exploited in generating specific mutants. For example, by appropriate deletions or transposon insertions, it is possible to generate mutants which lack a single product. Study of these mutants might firstly provide information about the role played by specific proteins in polysaccharide biosynthesis and might also be informative about the integration of the biochemical pathway for K5 biosynthesis.

6.1 DNA SEQUENCE AND COMPUTER ANALYSIS OF REGION II OF THE K5 CAPSULAR GENE CLUSTER

Tn5*phoA* mutagenesis of pGB110 identified two insertions in region II which affect the K5 synthesis. LE392 harbouring these mutations are resistant to K5 bacteriophage infection (see section 7.1). One of these insertions, termed pGB110::9 (Figure 6.1), was used to identify a DNA segment of region II which complemented the pGB110::9 insertion. A 3.6 Kb *Cla*I-*Nsi*I fragment,

which spans the pGB110::9 insertion, was subcloned from pGB110 into pACYC184 generating pXSP1 (Figure 6.1). LE392(pGB110::9, pXSP1) is sensitive to K5 bacteriophage infection and the presence of the above mentioned plasmids was verified by small scale extraction of plasmid DNA and digestion with appropriate restriction enzymes. pXSP1 was used to further extend the DNA sequence of region I (see chapter 5) towards region II. The DNA segment from *EcoRV* (3' to *kpsS*) to *Hind*III had already been sequenced on both strands during the sequence analysis of pH18 (Figure 6.1). This was further extended, about 1.2 Kb, to complete the DNA sequence of the 3.6 Kb *ClaI-NsiI* fragment (Figure 6.1). Several DNA fragments of pXSP1 were generated by appropriate restriction enzymes and subcloned into appropriately cleaved M13mp18 and M13mp19 (for a representative example of both the DNA sequencing strategy adopted in this study and sequencing gel see Figure 6.2 and 6.3, respectively).

FIGURE 6.1 COMPLEMENTATION OF LE392(pGB110::9)
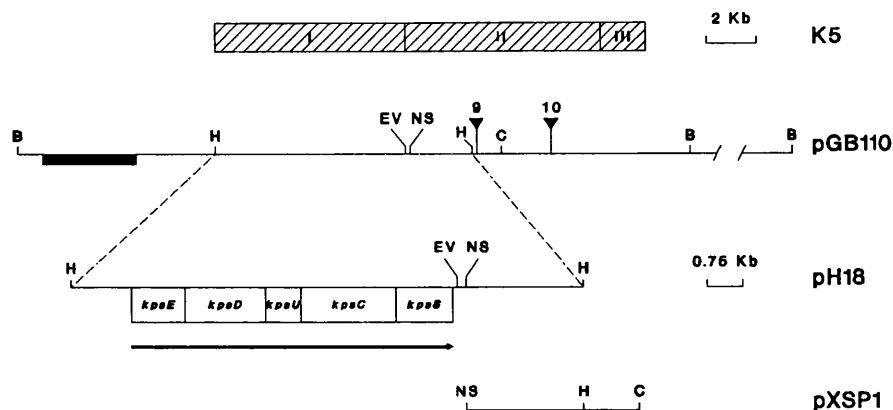
FIGURE 6.1 The linear map of constructs used for complementation of LE392(pGB110::9) is shown with horizontal lines and the scale is shown in kilobase pairs. The space within horizontal lines interrupted by dashed lines indicates sequence present in the original construct but not shown in figure. The filled box indicates vector sequence. Vertical and oblique lines indicate restriction sites (capital letters) or Tn5phoA insertions (filled triangles; numbers refer to specific insertions reported in the text). Boxes with diagonal lines and labelled I, II and III, define the organization of the K5 capsular gene cluster along pGB110. The orfs identified in region I are shown in boxes labelled kpsE, kpsD, kpsU, kpsC and kpsS and the arrow indicates the direction of their transcription. Broken lines highlight the enlargement on scale. Abbreviations: B, BamHI; EV, EcoRV; NS, NsiI; C, ClaI. H, HindIII.
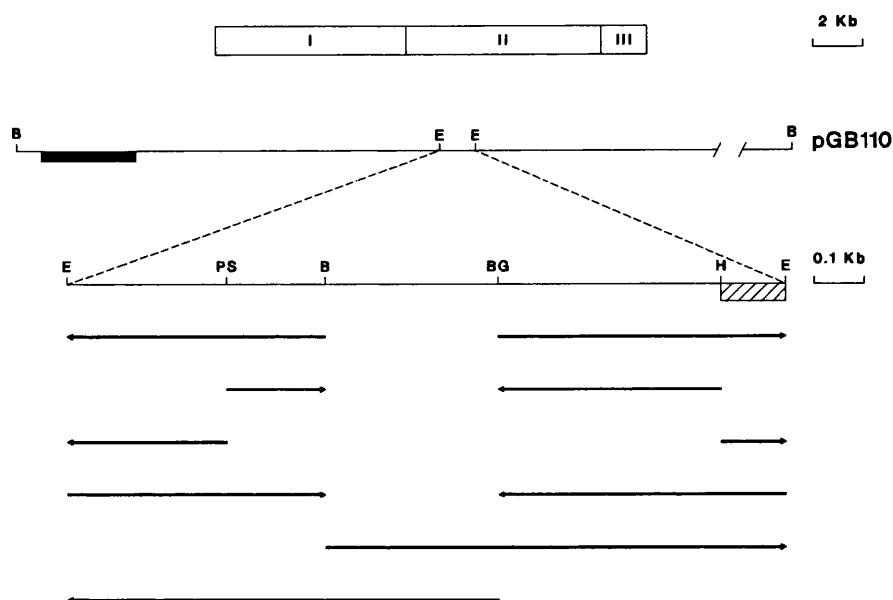
FIGURE 6.2 The linear map of pGB110 is shown with horizontal lines and the scale is shown in kilobase pairs. The filled box refers to vector sequence. The space within horizontal lines interrupted by dashed lines indicates sequence present in the original construct but not shown in figure. Boxes labelled I, II and III show the organization of the K5 capsular gene cluster in pBG110. Vertical lines indicate restriction sites (capital letters). Broken lines highlight the enlargement on scale (20:1) of the 1.5 Kb *EcoRI* fragment. The 1.5 Kb *EcoRI* fragment has been chosen as a representative example of the sequencing strategy adopted in this study. The box with diagonal lines refers to the DNA sequence showed on figure 5.3 (sequencing gel). Arrows indicate DNA fragments subcloned from pH18 and pXSP1 into appropriately cleaved M13mp18 and M13mp19. Abbreviations: E, *EcoRI*; B, *BamHI*; PS, *PstI*; BG, *BglII*; H, *HindIII*.

108

FIGURE 6.3   SEQUENCING GEL

A T C G

C(3')
T
T   *Eco*RI
A
A
G(5')

T(3')
T
C   *Hind*III
G
A
A(5')

The sequence of the 1.2 Kb *Hind*III-*Cla*I fragment and that overlapping the *Hind*III restriction site was determined on both strands. The DNA region extending from *Eco*RV (3' to *kps*S) to *Cla*I consists of 3659 bp and comprises several *orfs* (Figure 6.4).


FIGURE 6.4 *orfs* IDENTIFIED IN pXSP1
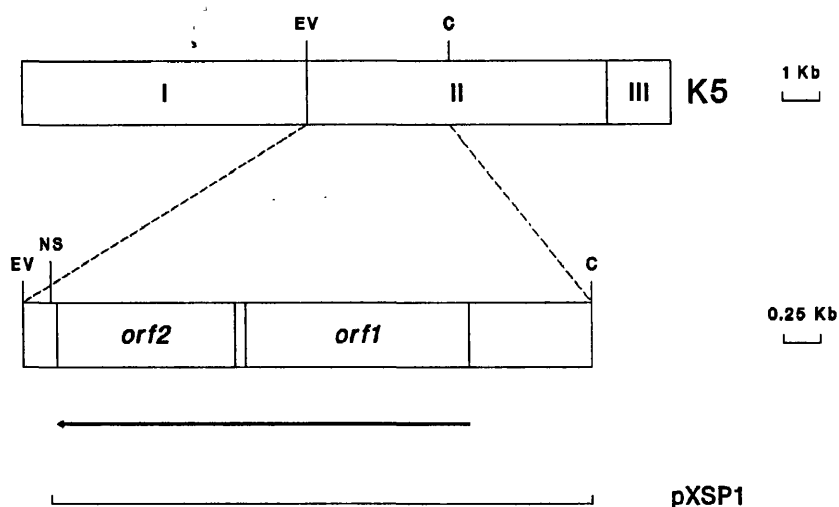
FIGURE 6.4 The organization of the K5 capsular gene cluster is shown as boxes labelled with I, II and III. The scale is shown in kilobase pairs. *orfs* identified in pXSP1 and discussed in the text are shown in enlarged boxes (scale 4:1). Broken lines highlight the enlargement. Vertical lines indicate restriction sites (capital letters). The arrow indicates the direction of transcription of *orf*1 and *orf*2. The linear map of pXSP1 is shown with a horizontal line. Abbreviations: EV, *Eco*RV; C, *Cla*I; NS, *Nsi*I.

The first *orf* (*orf*1) starts 3' to *Cla*I (coordinates 834 to 2396; figure 6.11) and is 1563 bp and terminates 41 bp

110

before the second *orf* (*orf2*). *orf2* (coordinates 2438 to 3616; figure 6.11) is 1179 bp and terminates 232 bp before the *EcoRV* restriction site. Other small *orfs* were identified either 5' to *orf1* or on the opposite strand.

Translation from the first three AUGs of *orf1* (coordinates 834, 1101 and 1425 bp, Figure 6.11) results in proteins with predicted MWs of 59.9, 49.7 and 37.2 KDa, respectively. None of these AUG codons is preceded by a putative SD sequence. However, only use of the first AUG may account for the 60 KDa protein identified as the product of this region by minicell analysis (A Smith, personal communication). On this basis it seems likely that the first AUG is the translation start of *orf1*. Computer analysis using the amino acid sequence of ORF1 revealed 21.4% identity and 64.1% homology, over 192 amino acids, with the RodC protein of *Bacillus subtilis* (Honeyman and Stewart, 1989). *rodC* lies within an operon which encodes proteins involved in the synthesis of poly(glycero-phosphate) teichoic acid of *B.subtilis*. The synthesis of poly(glycero-phosphate), [poly(groP)], involves two main steps (Pooley *et al.*, 1992). Firstly, the formation of a linkage unit: Prenol-PP-GlcNAc-ManNAc-(P-Gro)$_{1-3}$, and secondly, polymerization of the main polyol-phosphate chain:

$$\text{Prenol-PP-GlcNAc-ManNAc-(P-Gro)}_{1-3}\text{-}\begin{bmatrix} \text{P-Gro} \\ | \\ \text{Glc} \end{bmatrix}_n$$
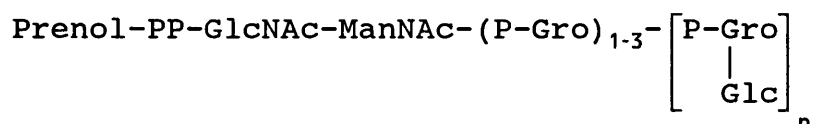
Very recently *tagF* (also termed *rodC*) has been demonstrated to encode for a CDP-glycerol:poly(glycero-phosphate) glycerophosphotransferase (CGPTase) (Pooley *et al.*, 1992). CGPTase catalyses the extension of the main poly(groP) backbone through sequential transfer of

glycero-phosphate units from CDP-glycerol (CDP-gro) to an acceptor in the form of growing polyglycerol-phosphate chains attached to the linkage unit lipid. Additionally, TagF has also been proposed to be part of a multienzyme complex which is associated with the cytoplasmic membrane. ORF1 is also 21% identical and 66% homologous, over 100 amino acids, with the rat liver bilirubin UDP-glucuronosyl transferase (Sato et al., 1990).

Only the first two AUG codons of orf2 may account for the 43 KDa protein identified as the gene product of orf2 (see section 5.3 and figure 5.5). Translation from these AUG codons results in proteins with predicted MWs of 44 and 41.6 KDa, respectively. The second AUG is preceded by a putative SD sequence, whereas no obvious SD sequence was identified 5' to the first AUG. This suggests that the second AUG is the most likely translation start of orf2. However, it is not possible to exclude the first AUG as a translation start codon. It is important to underline that a SD sequence is not absolutely essential for translational initiation, but does enhance the efficiency of translation (Calagero et al., 1988; McCarthy and Gualerzi, 1990). With this in mind computer analysis was performed by taking into account the amino acid sequence derived by translation from the first AUG. ORF2 is 27% identical and 69.1% homologous, over 359 amino acids, with the GDP-mannose dehydrogenase (AlgD) of Pseudomonas aeruginosa (Figure 6.6) (Deretic et al., 1987). AlgD is a protein with a predicted MW of 48 KDa and is a key enzyme in alginate biosynthesis (Roychoudhury et al., 1989). Alginate is a linear copolymer consisting of ß-1,4-linked D-mannuronic acid and variable amounts of the C-5 epimer L-guluronic acid (Evans and Linker, 1973). This polymer plays a crucial role in chronic pulmonary infection caused by alginate-producing P.aeruginosa, which is responsible for high

mortality among cystic fibrosis patients (Govan and Harris, 1986). AlgD converts GDP-mannose to GDP-mannuronic acid as follows:

$$\text{GDP-mannose} + 2\text{NAD}^+ + \text{H}_2\text{O} \longrightarrow \text{GDP-mannuronic acid} + 2\text{NADH} + 2\text{H}^+$$
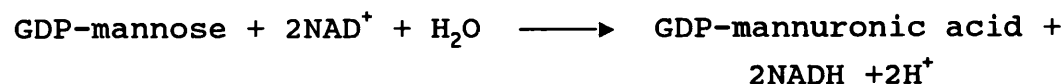
AlgD exhibits a number of characteristics common to other four-electron transfer dehydrogenases, such as the UDP-D-Glucose dehydrogenase from bovine liver (UdhG) (Feingold and Frazen, 1981), and L-histidinol dehydrogenase from *E.coli* (Ordman and Kirkwood, 1977). In particular AlgD shares with these dehydrogenases their putative catalytic site which includes a conserved cysteine residue. ORF2 shares with AlgD and UdhG the conserved cysteine residue and its surrounding amino acid sequence (Figure 6.5).

FIGURE 6.5 ALIGNMENT BETWEEN PUTATIVE CATALYTIC SITES OF UdhG AND AlgD WITH ORF2

```
UdhG  ASVGFGGSCFZZGK
      : |  |:||  :    :
ORF2  PSFGYGGYCLPKDT
      |:|::||  |||| :
AlgD  PGFAFGGSCLPKDV
```

FIGURE 6.5 Double vertical lines align identical amino acids. Double dots align homologous amino acids. Conserved cysteines are shown with bold characters highlighted with small dots. Z refers to either Glutamine (Q) or to Glutamic acid (E).

ORF2 is also 72.9% homologous, over 70 amino acid residues, with the ecdysteroid UDP-glucosyl transferase

113

of the insect baculovirus *Autographa californica* (O'Reilly and Miller, 1989). This enzyme transfers glucose from UDP-glucose to ecdysteroids and also shares homology with several mammalian UDP-glucuronosyl transferases.
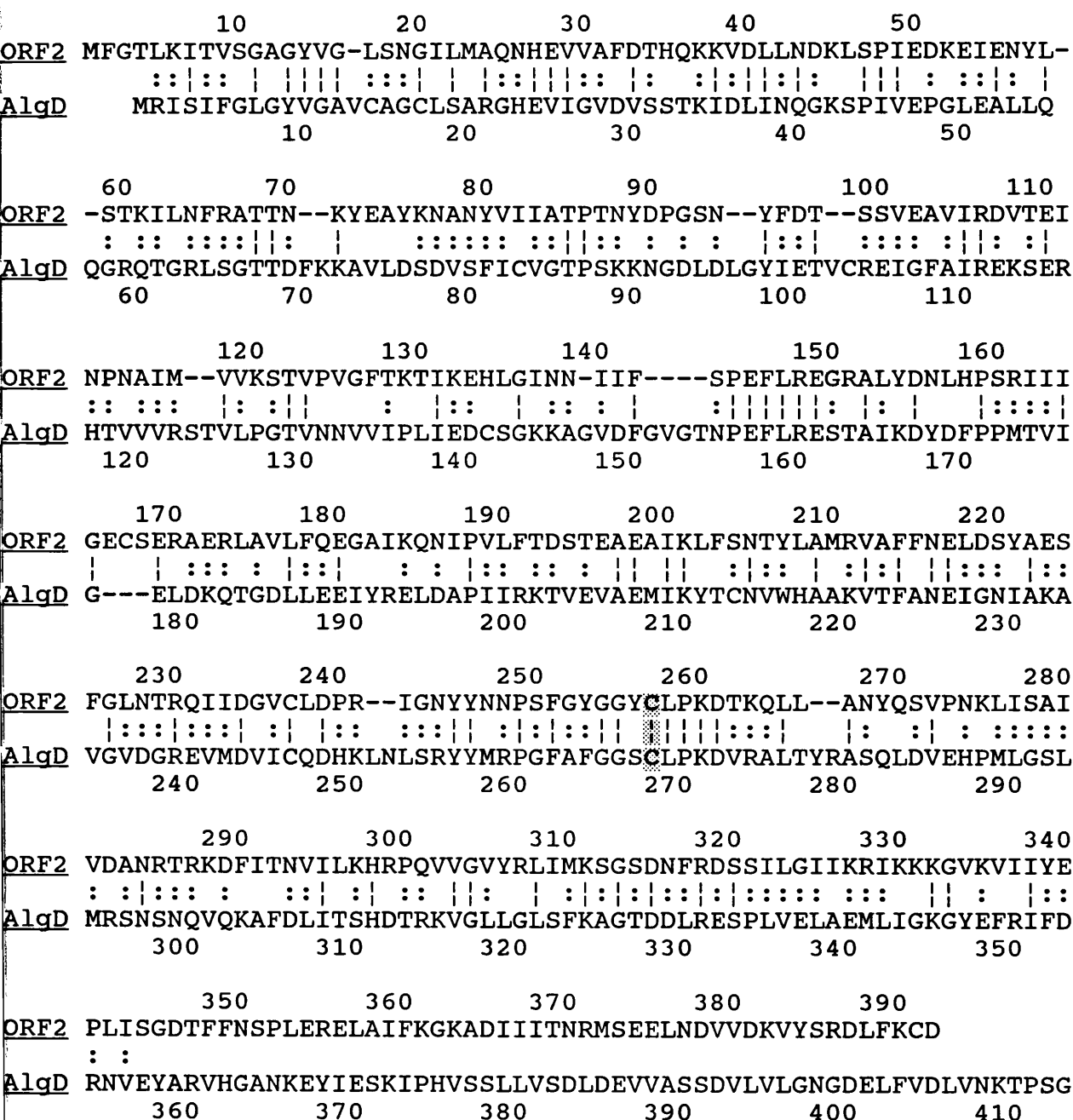
FIGURE 6.6 HOMOLOGY BETWEEN ORF2 AND AlgD

```
           10        20        30        40        50
ORF2 MFGTLKITVSGAGYVG-LSNGILMAQNHEVVAFDTHQKKVDLLNDKLSPIEDKEIENYL-
      ::|:: | |||| :::| | |::|||:: |: :|:|:|:| ||| : ::|: |
AlgD    MRISIFGLGYVGAVCAGCLSARGHEVIGVDVSSTKIDLINQGKSPIVEPGLEALLQ
            10        20        30        40        50


      60        70        80        90        100       110
ORF2 -STKILNFRATTN--KYEAYKNANYVIIATPTNYDPGSN--YFDT--SSVEAVIRDVTEI
      : :: ::::||: |    ::::::: ::||:: : : : |::| :::: :||: :|
AlgD QGRQTGRLSGTTDFKKAVLDSDVSFICVGTPSKKNGDLDLGYIETVCREIGFAIREKSER
       60        70        80        90        100       110


           120       130       140       150       160
ORF2 NPNAIM--VVKSTVPVGFTKTIKEHLGINN-IIF----SPEFLREGRALYDNLHPSRIII
      :: ::: |: :|| : |:: | :: : | :||||||: |: | |::::|
AlgD HTVVVRSTVLPGTVNNVVIPLIEDCSGKKAGVDFGVGTNPEFLRESTAIKDYDFPPMTVI
      120       130       140       150       160       170


      170       180       190       200       210       220
ORF2 GECSERAERLAVLFQEGAIKQNIPVLFTDSTEAEAIKLFSNTYLAMRVAFFNELDSYAES
      |    | ::: : |::|    : : |:: :: : || || :|:: | :|:| ||:::  |::
AlgD G---ELDKQTGDLLEEIYRELDAPIIRKTVEVAEMIKYTCNVWHAAKVTFANEIGNIAKA
         180       190       200       210       220       230


      230       240       250       260       270       280
ORF2 FGLNTRQIIDGVCLDPR--IGNYYNNPSFGYGGYCLPKDTKQLL--ANYQSVPNKLISAI
      |:::|:::: | :| |:: ::::|| :|:|:|:|| ::||||:::| |: :| : :::::
AlgD VGVDGREVMDVICQDHKLNLSRYYMRPGFAFGGSCLPKDVRALTYRASQLDVEHPMLGSL
      240       250       260       270       280       290


      290       300       310       320       330       340
ORF2 VDANRTRKDFITNVILKHRPQVVGVYRLIMKSGSDNFRDSSILGIIKRIKKKGVKVIIYE
      : :|:::  :   ::| :| :: ||: | :|:|:|::|:|::::: ::: || : |::
AlgD MRSNSNQVQKAFDLITSHDTRKVGLLGLSFKAGTDDLRESPLVELAEMLIGKGYEFRIFD
         300       310       320       330       340       350


           350       360       370       380       390
ORF2 PLISGDTFFNSPLERELAIFKGKADIIITNRMSEELNDVVDKVYSRDLFKCD
      : :
AlgD RNVEYARVHGANKEYIESKIPHVSSLLVSDLDEVVASSDVLVLGNGDELFVDLVNKTPSG
         360       370       380       390       400       410
```

FIGURE 6.6 Numbers refer to amino acid coordinates of the respective proteins. Double vertical lines align identical amino acids. Double dots align homologous amino acids. Horizontal lines represent gaps automatically inserted to maximize the alignment. Conserved cysteines are shown with bold characters highlighted with small dots.

Computer analysis identified 5' to *orf1*, at coordinates 544 and 568 bp (figure 6.11), two DNA motifs which share homology with the -35 and -10 consensus sequences of *E.coli* promoters. The first motif (TTGATC) retains 5 bp of the -35 consensus, the second (TATATT) retains 5 bp of the -10 consensus and the spacing between these motifs is 18 bp. The -35 consensus sequence for promoters of this spacing class is TTGACY (Y = T or C) and each base is over 50% conserved. Whereas, the -10 consensus sequence is TATGWT (W = A or T) and each base is also over 50% conserved. An average *E.coli* promoter of this spacing class retains 3.4 bp of the -35 and 3.8 bp of the -10 consensus sequence, respectively (O'Neil, 1989). The two DNA motifs identified 5' to *orf1* are in good agreement with the average features for *E.coli* promoters of the 18 bp spacing class. These motifs represent good candidates for a possible promoter 5' to *orf1*. No predicted rho-independent or rho-dependent termination sequence was identified 3' to *orf1*. Whereas, a stem and loop structure is located 60 bp after the UAA stop codon of *orf2* and, potentially, it might take part in rho-dependent transcriptional termination (Figure 6.11).

In summary, the DNA region extending from *EcoRV* to *ClaI* is composed of at least two genes (*orf1* and *orf2*) (Figure 6.4). Analysis of the protein pattern produced by minicells harbouring plasmids carrying either region II or region I plus *orf2* (pH18), identified proteins with calculated MWs close to those predicted by translation from the first AUG of *orf1* and *orf2*. In the case of *orf2* a second AUG might also account for the observed 43 KDa protein. On the basis of these preliminary results, it is difficult to distinguish which of these two AUGs represents the most likely translation start of *orf2*. Whereas, there is more confidence in assigning the first AUG of *orf1* as its most probable translation start.

Taking into account the first AUG and the short gap between these *orf*s, it is possible that *orf*1 and *orf*2 are organized into a single transcriptional unit. However, more data are required before defining accurately the transcriptional organization of this region.

The remaining part of region II, a 4 Kb DNA fragment from the *Cla*I to the *Bgl*II restriction site 3' to *kps*T (Figure 6.7), was sequenced by A Smith (personal communication). This DNA fragment comprises two *orf*s, *orf*3 and *orf*4, with a direction of transcription from region III towards region I. *orf*3 starts 986 bp after the stop codon of *kps*T. *orf*3 is 717 bp and terminates 431 bp before *orf*4. *orf*4 is 1692 bp and terminates 1291 bp before the first AUG of *orf*1. Translation initiation using only the first two AUGs of *orf*3 would account for the 27 KDa protein observed in minicells analysis (A Smith, personal communication). Translation from these AUG codons results in proteins with predicted MWs of 27.3 and 26.7 KDa, respectively. However, none of these AUG codons is preceded by an obvious SD. The first AUG of *orf*4 is preceded by a putative SD and translation from this start codon results in a protein with a predicted MW of 65732 Da. Only the use of this AUG codon would account for the 66 KDa protein seen in minicells (A Smith, personal communication). Computer analysis was performed using the amino acid sequence obtained from translation of the first AUG of *orf*3 and *orf*4. Homologies shared by ORF3 and ORF4 with the NBRF-Protein sequence collection are restricted to short amino acid stretches (A Smith, personal communication).

FIGURE 6.7   ANALYSIS OF REGION II OF THE K5 CAPSULAR GENE CLUSTER
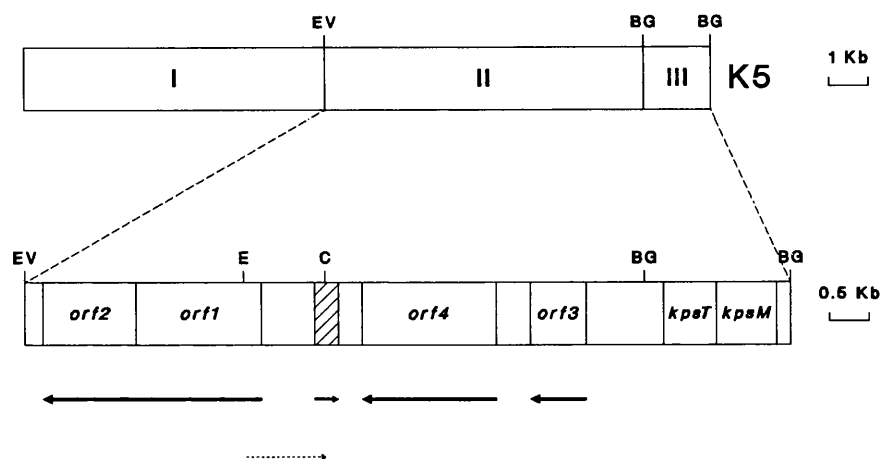
FIGURE 6.7 The organization of the K5 capsular gene cluster is shown in boxes labelled I, II and III. The scale is shown in kilobase pairs. Regions II and III are enlarged on scale (2:1) and broken lines highlight the enlargement.The *orfs* identified in region II and III, are shown in boxes labelled *orf* or *kps*, respectively. The box with oblique lines indicates an *orf* which might be included in the organization of region II (see text). Horizontal arrows with unbroken lines indicate the direction of transcription. The horizontal arrow with broken lines indicates the direction of transcription due to a promoter activity identified in the 1 Kb *EcoRI-ClaI* fragment (see text). Vertical lines indicate restriction sites (capital letters). Abbreviations: EV, *EcoRV*; BG, *BglII*; E, *EcoRI*; C, *ClaI*.

Promoter activity has been identified with a direction of transcription from *orf*1 to *ClaI* (Figure 6.7) (A Smith, personal communication), by subcloning the 1 Kb *EcoRI-ClaI* DNA fragments into the vector pKK232-8, which contains a promoterless chloramphenicol acetyltransferase

gene (*cat*) (Brosius, 1984). A small *orf* starts 138 bp
before and terminates 162 bp after the *Cla*I restriction
site (Figure 6.7). The first AUG of this *orf* is preceded
by a putative SD sequence and translation from this AUG
results in a protein with a predicted MW of 11.5 KDa.
However, some factors impose caution in assigning this
orf as a gene of region II of the K5 capsular gene
cluster. Firstly, analysis of the protein pattern
produced by minicells harbouring plasmids carrying region
II did not reveal proteins with a calculated MW smaller
than 27 KDa (A Smith, personal communication). This
analysis was performed using experimental conditions
which would have allowed the identification of proteins
smaller than 10 KDa.

The average GC ratio in *E.coli* DNA is about 48-52 mol%
(Orskov, 1984). Computer analysis of the DNA sequence of
the K5 capsular gene cluster revealed that the G+C
content is 50.6% in region I, 33.4% in region II and
42.3% in region III (Figure 6.8). The G+C content of
50.2% is uniformly distributed within region I with the
lowest value of 45.6% in *kps*S. Likewise, the G+C content
of 33.4 is uniformly distributed within region II, with
the highest value of 40.4% in the region between *orf*4 and
*orf*1. In region III, the G+C content is 44.8% in *kps*M and
39.4% in *kps*T. A more detailed computer analysis of *kps*T
and *kps*S has revealed a variable G+C content within these
genes. An average G+C content of 43.2% is uniformly
distributed in the first 507 nucleotides (5') of *kps*T,
whereas the G+C content drops to 28% in the last 168
nucleotides (3') of *kps*T. Likewise, an average G+C
content of 47.2% is uniformly distributed within the
first 1097 nucleotides (5') of *kps*S, whereas the G+C
content drops to 32.9% in the last 73 nucleotides (3') of
*kps*S. The variable G+C content within *kps*T and *kps*S seems
to match, approximately, with the variable C-termini of

KpsT and KpsS encoded by different capsule gene clusters (see section 5.2). Additionally, the low G+C content of the 3' ends of *kps*T and *kps*S appears consistent with the low G+C content of region II. On this basis, it is tempting to suggest that the K5 region II might have been acquired through recombination events within *kps*T and *kps*S. Additionally, the low G+C content of region II implies the possibility that this region might have been acquired from a micro-organism different from *E.coli*.

FIGURE 6.8   GC CONTENT OF THE K5 CAPSULAR GENE CLUSTER
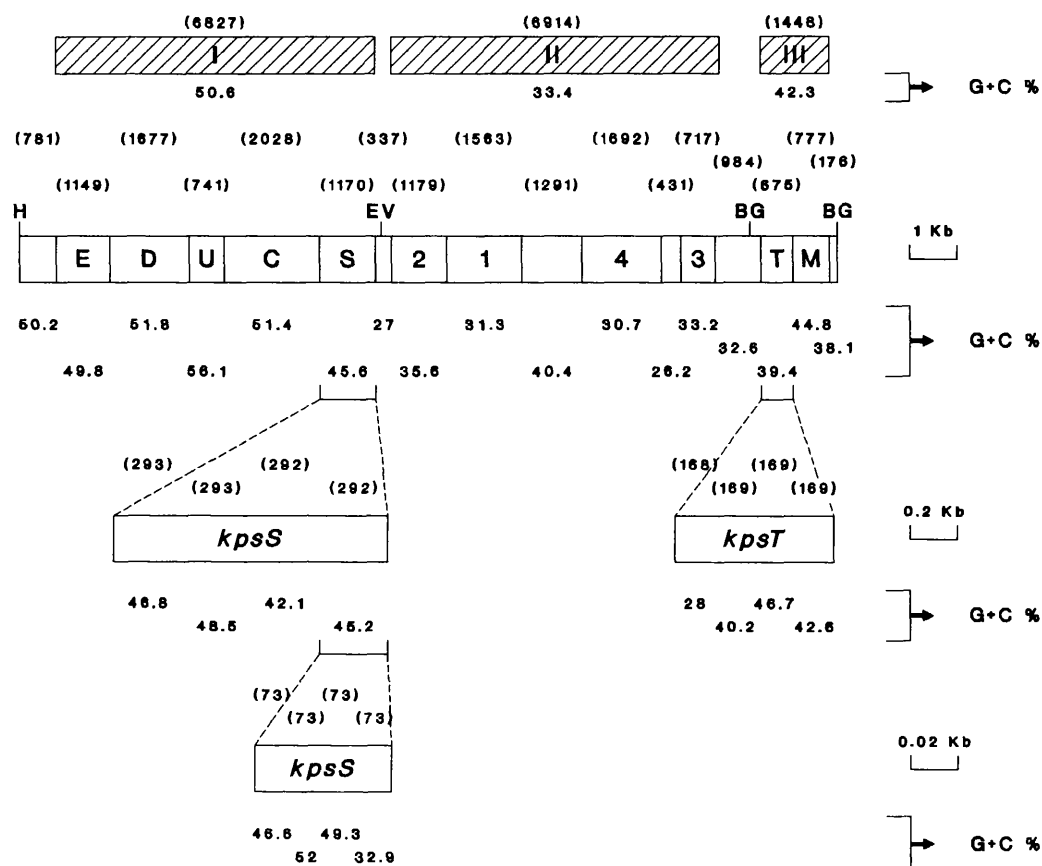
FIGURE 6.8 Organization of the K5 capsular gene cluster is shown as boxes and the scale is shown in kilobase pairs. Boxes labelled with capital letters refer to *orfs* identified in region I and III, whereas boxes labelled with numbers refer to *orfs* identified in region II. Numbers in paretheses refer to length of genes in nucleotides. Numbers below boxes refer to their average G+C content. *kpsS* and *kpsT* are enlarged to show the G+C content within these genes. Vertical lines indicate restriction sites. Abbreviations: H, *Hind*III; EV, *EcoRV*; BG, *Bgl*II.

6.2 DISCUSSION

Region II of the K5 capsular gene cluster is approximately 7.9 Kb and composed of at least four genes (*orf*1 to *orf*4) possibly organized into three transcriptional units (Figure 6.9). Additionally, data from preliminary experiments (A Smith, personal communication) indicate another *orf* might be located between *orf*1 and *orf*4 (Figure 6.7). However, these data need to be investigated further before including additional *orfs* to the proposed organization of region II of the K5 capsular gene cluster. For example, gene fusions located in this *orf* would undoubtedly provide more convincing evidence of its existence.

FIGURE 6.9   ORGANIZATION OF REGION II OF THE K5 CAPSULAR GENE CLUSTER
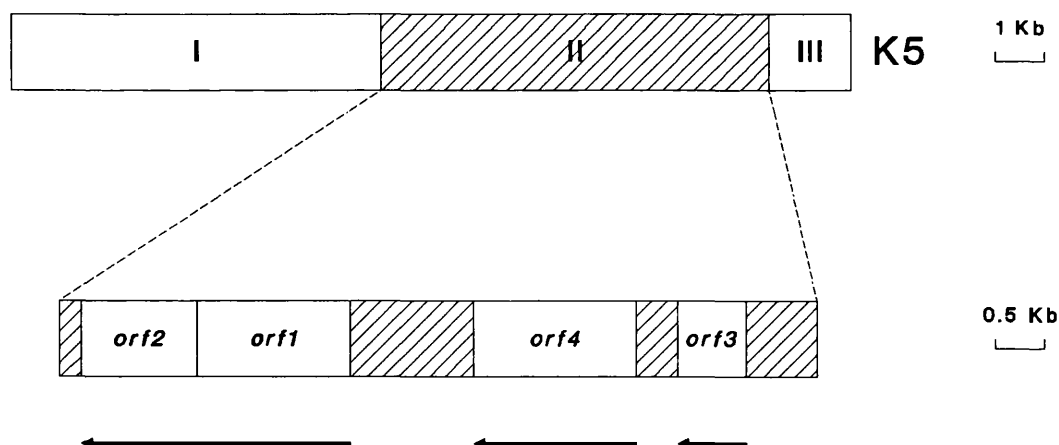
FIGURE 6.9 The organization of the K5 capsular gene cluster is shown as boxes labelled I, II and III. The scale is shown in kilobase pairs. Boxes with oblique lines highlight both region II and the enlargement on scale (2:1) of this region. The identified *orfs* of region II are shown in boxes. Arrows indicate the direction of transcription.

The K5 polysaccharide repeating unit is composed of GlcA and GlcNAc. GlcNAc is also found in other Gram-negative structures such as peptidoglycan and LPS, whereas GlcA is found in capsular and capsular-like structures, such as the M-antigens. Monosaccharides can be transferred from their activated forms (UDP-derivatives) to a growing polysaccharide chain, by specific transferases. However, very little was known about the number of function required or the role they play in synthesis of the K5 polysaccharide.

The product of ORF2 might be responsible for GlcA synthesis, for example by catalysing the dehydrogenation of UDP-Glc to UDP-GlcA. This hypothesis is based on the extensive homology between ORF2 product and the AlgD dehydrogenase (Figure 6.5) and the presence of an amino acid sequence highly conserved among different dehydrogenases and thought to constitute their catalytic site (Figure 6.6). Recent studies have indicated that ORF2 might function as a GlcNAc transferase (Sieberth and Jann, unpublished results). This hypothesis is based on results obtained by supplementing LE392(pGB110::9) with exogenous UDP-GlcA and UDP-GlcNAc. The insertion in pGB110::9 is located in *orf*1 and LE392(pGB110::9) does not produce K5 polysaccharide. Additionally, LE392(pGB110::9) grown in medium containing exogenous sugars (UDP-GlcA*[1] and UDP-GlcNAc) did not incorporate radioactive material into products associated with the cytoplasmic membrane, suggesting that exogenous sugars cannot bypass functions missing in this mutant. pHSA2 and pBGSA1 were generated by subcloning, from pH18, the 2.4 Kb *Hind*III-*Sal*I and the 2 Kb *Bgl*II-*Sal*I fragments into pACYC184, respectively (Figure 6.10). LE392(pGB110::9; pHSA2) or LE392(pGB110::9; pBGSA1) grown in medium containing only exogenous UDP-GlcNAc*

---

[1]* indicates radioactive sugars

membrane fraction, whereas when grown in medium containing only exogenous UDP-GlcA* no radioactive material was incorporated in this fraction. This suggests that both pHSA2 and pBGSA1 encode function(s) able to promote the transfer of GlcNAc, but not of GlcA, to inner membrane-associated material. On this basis, it has been hypothesised that the ORF2 product might function as a GlcNAc transferase. However, there are some objections to this hypothesis. For example, pHSA2 and pBGSA1 carry only part of *orf*1 and *orf*2 (Figure 6.10). The *Hind*III and *Bgl*II restriction sites map 240 and 790 bp 3' to the first AUG of *orf*1, respectively (Figure 6.11). Whereas the *Sal*I restriction site, present in both constructs, maps 40 bp 5' to the UAA stop codon of *orf*2 (Figure 6.11). Nevertheless, it is possible that the C-terminus of ORF2 might not play a crucial role in determining the function of this protein. The variable 3'ends of *kps*S and *kps*T among different *kps* gene clusters, suggests that proteins with a variable C-terminus are able to perform identical functions. Thus, there is the possibility that the truncated ORF2 product encoded by pHSA2 or pBGSA1, which lacks eleven amino acid C-terminal (Figure 6.11), might still be functional and promote the transfer of GlcNAc. ORF2 might also be bifunctional, promoting the transfer of GlcNAc to a growing polysaccharide chain and catalysing the dehydrogenation of UDP-Glc to UDP-GlcA. ORF1 shows homology with two transferases, the glycerolphosphotransferase of *B.subtilis* and the UDP-glucuronosyl transferase of rat. pGB110::9 is located in *orf*1 and experiments involving complementation by exogenous UDP-GlcA and UDP-GlcNAc have suggested that ORF1 is involved in step(s) of polysaccharide polymerization which might be independent from synthesis of sugars constituting the repeating unit of the K5 polysaccharide. Thus, if ORF2 functions as a GlcNAc transferase and maybe also as a UDP-Glc dehydrogenase, there is the possibility that ORF1 might act as a GlcA transferase.

FIGURE 6.10   COMPLEMENTATION OF pGB110::9 BY EXOGENOUS UDP-GlcA AND UDP-GlcNAc
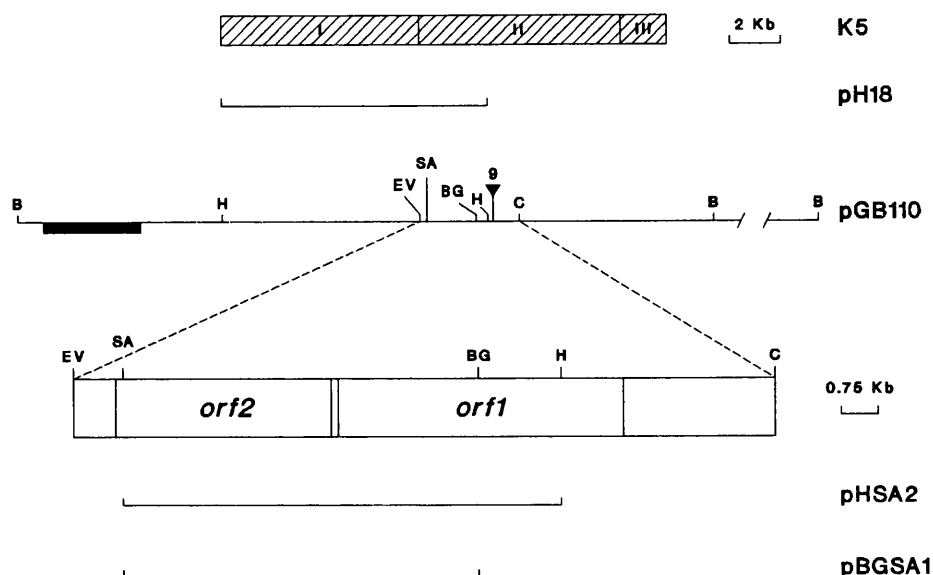
FIGURE 6.10 The linear map of various constructs is shown with horizontal lines and the scale is shown in kilobase pairs. The filled box refers to vector sequence. The space within horizontal lines interrupted by dashed lines indicates sequence present in the original construct but not shown in figure. Boxes with oblique lines and labelled I, II and III align the organization of the K5 capsular gene cluster along pGB110. Vertical and oblique lines indicate restriction sites (capital letters) or Tn5phoA insertion (filled triangle; pBG110::9). The organization of the 3.6 Kb EcoRV-ClaI fragment in enlarged on scale and shown with boxes. Broken lines highlight the enlargement. Abbreviations: B, BamHI; H, HindIII; EV, EcoRV; BG, BglII; C, ClaI; SA, SalI.

It is more difficult to speculate about ORF3 and ORF4. Homologies between ORF3 and ORF4 of the K5 capsular gene cluster and the NBRF-Protein sequence collection are

restricted to short amino acid stretches, and these homologies do not provide any useful indication about the possible role played by ORF3 and ORF4 in synthesis of the K5 polysaccharide.

Finally, the low G+C content of both region II and 3' ends of *kps*T and *kps*S raises the possibility that region II might have been acquired through recombination events within these genes. Additionally, the low G+C content also suggests that region II might have been acquired from a different micro-organism.

FIGURE 6.11   DNA SEQUENCE OF THE 3.8 Kb *ClaI-EcoRV* FRAGMENT
OF REGION II OF THE K5 CAPSULAR GENE CLUSTER

Cla I 10   30   50
*ATCGA*TAGGGAATAGCCGGTCTGCACGGCAAAAGAAGAGTTTGATGATATCACAGTTAAG

    70   90   110
TGACGGCACGGCCTGTGCCAGAGCGTAGTTGCTGTGTTCGTCGATCATGGTAATGACATA

   130   150   170
ACAGCGCAGCTCGCCCATTCTGAGCTCAATAGCATCTATACCAACGAGCTCACTGCTCTT

   190   210   230
TACCGGGCGATAGCGTTTTGATCTGCGGGTTGGAGGATTGTATTTTTATGAACAGTGTT

   250   270   290
TTTCCTTGGGGAGGCAGTCACATCGGTATCAGTCGCATTTTATCGTGTGCGGCCGTGAAC

   310   330   350
ATCCTGCCGATGGTTGAAATGCTCGGACAGACCAGGCGGTGCTGTTCACTCCACGGCTTC

   370   390   410
AAGCAAACAAAAATCTGTTCTTCTAGGTTGGGAAGCTCTATTCTCAGTCGACGTATTTCT

   430   450   470
TACAGAATTACGAGATGCCATTGCTTTGTGCGATGTACTAGCGGAGCTTTGCTACGCGAA

   490   510   530
ATAGGTGCCTCTGGGCCATAATACAGCGTTCGTGTGGATACAGCAAAAACTTCCGCAACT

   550   570   590
  (-35)   (-10)
GTATTGATCTCATGTTTCTCCCAGAAGTATATTTTTTCATCCTTAATTTTGTAATCTCAG

   610   630   650
GTATAACAAAGTGTTTCATCACATAGATGTTGGCATGGTAATGCCTCAAATATCCGCCGC

   670   690   710
AGATACGTTGCATCAACTTAGCATTTCCCTCGCTTGTCCGGAGATAATTGCAATATCTCT

   730   750   770
GTGAGCTTACACTGTGACATTCGTTGAGTTTTAGTGATGTTTTTAAAGATTTATATTTAT

   790   810   830
                   SD SD orfI
AATATTTAGTAAATGCAGTTTTATTCTCATTTTATTTATCATTAAGTGAATGTATGAACG
                            M N A
   850   870   890
CAGAATATATAAATTTAGTTGAACGTAAAAAGAAATTAGGGACAAATATTGGTGCTCTTG
 E Y I N L V E R K K K L G T N I G A L D

```
           910                    930                    950
ATTTTTTATTATCAATTCATAAGGAGAAAGTTGATCTTCAACATAAAAACTCGCCTTTAA
 F   L   L   S   I   H   K   E   K   V   D   L   Q   H   K   N   S   P   L   K

           970                    990                   1010
AAGGTAACGATAACCTTATTCACAAAAGAATAAACGAATACGACAATGTACTTGAACTAT
 G   N   D   N   L   I   H   K   R   I   N   E   Y   D   N   V   L   E   L   S

          1030           EcoR I   1050                   1070
CTAAGAATGTATCAGCTCAGAATTCTGGCAATGAGTTTTCTTATTTATTGGGATATGCAG
 K   N   V   S   A   Q   N   S   G   N   E   F   S   Y   L   L   G   Y   A   D

          1090                   1110                   1130
ATTCTCTTAGAAAAGTTGGTATGTTGGATACTTATATTAAAATTGTTTGTTATCTAACAA
 S   L   R   K   V   G   M   L   D   T   Y   I   K   I   V   C   Y   L   T   I

          1150                   1170         Hind III   1190
TTCAATCTCGTTATTTTAAAAATGGCGAACGAGTTAAGCTTTTTGAACATATAAGTAACG
 Q   S   R   Y   F   K   N   G   E   R   V   K   L   F   E   H   I   S   N   A

          1210                   1230                   1250
CTCTACGGTATTCAAGGAGTGATTTTCTCATTAATCTTATTTTTGAACGATATATCGAAT
 L   R   Y   S   R   S   D   F   L   I   N   L   I   F   E   R   Y   I   E   Y

          1270                   1290                   1310
ATATAAACCATCTAAAATTGTCGCCCAAACAAAAGATTTTTATTTTTGTACGAAGTTTT
 I   N   H   L   K   L   S   P   K   Q   K   D   F   Y   F   C   T   K   F   S

          1330                   1350                   1370
CAAAATTTCATGATTATACTAAAAATGGATATAAATATTTAGCATTTGATAATCAAGCCG
 K   F   H   D   Y   T   K   N   G   Y   K   Y   L   A   F   D   N   Q   A   D

          1390                   1410                   1430
ATGCAGGGTATGGCCTGACTTTATTATTAAATGCAAACGATGATATGCAAGATAGTTATA
 A   G   Y   G   L   T   L   L   L   N   A   N   D   D   M   Q   D   S   Y   N

          1450                   1470                   1490
ATCTACTCCCTGAGCAAGAACTTTTTATTTGTAATGCTGTAATAGATAATATGAATATTT
 L   L   P   E   Q   E   L   F   I   C   N   A   V   I   D   N   M   N   I   Y

          1510                   1530                   1550
ATAGGAGTCAATTTAACAAATGTCTACGAAAATACGATTTATCAGAAATAACTGATATAT
 R   S   Q   F   N   K   C   L   R   K   Y   D   L   S   E   I   T   D   I   Y

          1570                   1590                   1610
ACCCAAATAAAATTATATTGCAAGGAATTAAGTTCGATAAGAAAAAAAATGTTTATGGAA
 P   N   K   I   I   L   Q   G   I   K   F   D   K   K   K   N   V   Y   G   K

  Bgl II  1630                   1650                   1670
AAGATCTTGTTAGTATAATAATGTCAGTATTCAATTCAGAAGATACTATTGCATACTCAT
 D   L   V   S   I   I   M   S   V   F   N   S   E   D   T   I   A   Y   S   L

          1690                   1710                   1730
TACATTCATTGTTGAATCAAACATATGAAAATATTGAAATTCTCGTGTGCGATGATTGTT
 H   S   L   L   N   Q   T   Y   E   N   I   E   I   L   V   C   D   D   C   S
```

```
              1750                   1770                   1790
CATCGGACAAAAGCCTTGAAATAATTAAGAGCATAGCTTATTCTAGTTCAAGAGTGAAAG
  S   D   K   S   L   E   I   I   K   S   I   A   Y   S   S   S   R   V   K   V

              1810                   1830                   1850
TATATAGCTCACGAAAAAACCAAGGCCCTTATAATATAAGAAATGAGCTAATAAAAAAAG
  Y   S   S   R   K   N   Q   G   P   Y   N   I   R   N   E   L   I   K   K   A

              1870                   1890                   1910
CACACGGTAATTTCATCACCTTTCAAGATGCAGATGATCTTTCTCATCCGGAGAGAATAC
  H   G   N   F   I   T   F   Q   D   A   D   D   L   S   H   P   E   R   I   Q

              1930                   1950                   1970
AAAGACAAGTTGAGGTTCTTCGCAATAATAAGGCTGTAATCTGTATGGCTAACTGGATCC
  R   Q   V   E   V   L   R   N   N   K   A   V   I   C   M   A   N   W   I   R

              1990                   2010                   2030
GTGTTGCGTCAAATGGAAAAATTCAATTCTTCTATGATGATAAAGCCACAAGAATGTCTG
  V   A   S   N   G   K   I   Q   F   F   Y   D   D   K   A   T   R   M   S   V

              2050                   2070                   2090
TTGTATCGTCAATGATAAAAAAAGATATTTTTGCGACAGTTGGTGGCTATAGACAATCTT
  V   S   S   M   I   K   K   D   I   F   A   T   V   G   G   Y   R   Q   S   L

              2110                   2130                   2150
TAATTGGTGCAGATACGGAGTTTTATGAAACAGTAATAATGCGTTATGGGCGAGAAAGTA
  I   G   A   D   T   E   F   Y   E   T   V   I   M   R   Y   G   R   E   S   I

              2170                   2190                   2210
TTGTAAGATTACTGCAGCCATTGATATTGGGGTTATGGGGAGACTCCGGACTTACCAGGA
  V   R   L   L   Q   P   L   I   L   G   L   W   G   D   S   G   L   T   R   N

              2230                   2250                   2270
ATAAAGGAACAGAAGCTCTACCTGATGGATATATATCACAATCTCGAAGAGAATATAGTG
  K   G   T   E   A   L   P   D   G   Y   I   S   Q   S   R   R   E   Y   S   D

              2290                   2310                   2330
ATATCGCGGCAAGACAACGAGTGTTAGGGAAAAGTATCGTAAGTGATAAAGATGTACGTG
  I   A   A   R   Q   R   V   L   G   K   S   I   V   S   D   K   D   V   R   G

              2350                   2370                   2390
GTTTATTATCTCGCTATGGTTTGTTTAAAGATGTATCAGGAATAATTGAACAATAGTTTG
  L   L   S   R   Y   G   L   F   K   D   V   S   G   I   I   E   Q   *

              2410                   2430        orf2       2450
TTATTCTATATATATTAAATTTTTGGGGCTATATAAAATGTTCGGAACACTAAAAATAAC
                                          M   F   G   T   L   K   I   T

              2470                   2490       SD           2510
TGTTTCAGGCGCTGGTTACGTTGGGCTTTCAAATGGAATTCTAATGGCTCAAAATCATGA
  V   S   G   A   G   Y   V   G   L   S   N   G   I   L   M   A   Q   N   H   E

              2530                   2550                   2570
AGTGGTTGCATTTGATACCCATCAAAAAAAAGTTGACTTACTTAATGATAAACTCTCTCC
  V   V   A   F   D   T   H   Q   K   K   V   D   L   L   N   D   K   L   S   P
```

129

```
        2590                    2610                    2630
TATAGAGGATAAGGAAATTGAAAATTATCTTTCAACTAAAATACTTAATTTTCGCGCAAC
 I   E   D   K   E   I   E   N   Y   L   S   T   K   I   L   N   F   R   A   T

        2650                    2670                    2690
TACTAACAAATATGAAGCCTATAAAAATGCCAATTACGTTATTATTGCTACACCAACGAA
 T   N   K   Y   E   A   Y   K   N   A   N   Y   V   I   I   A   T   P   T   N

        2710                    2730                    2750
TTATGACCCAGGTTCAAATTACTTTGATACATCAAGCGTTGAAGCTGTCATTCGTGACGT
 Y   D   P   G   S   N   Y   F   D   T   S   S   V   E   A   V   I   R   D   V

        2770                    2790                    2810
AACGGAAATCAACCCAAACGCAATTATGGTGGTTAAATCTACGGTCCCAGTAGGTTTCAC
 T   E   I   N   P   N   A   I   M   V   V   K   S   T   V   P   V   G   F   T

        2830                    2850                    2870
AAAAACAATTAAAGAACATTTAGGTATTAATAATATTATCTTCTCTCCAGAATTTTTACG
 K   T   I   K   E   H   L   G   I   N   N   I   I   F   S   P   E   F   L   R

        2890                    2910                    2930
AGAAGGAAGAGCCCTATACGATAATCTCCATCCATCTCGCATTATTATCGGTGAATGTTC
 E   G   R   A   L   Y   D   N   L   H   P   S   R   I   I   I   G   E   C   S

        2950                    2970                    2990
TGAACGGGCAGAACGTTTGGCAGTGTTATTTCAGGAAGGAGCGATTAAACAAAATATACC
 E   R   A   E   R   L   A   V   L   F   Q   E   G   A   I   K   Q   N   I   P

        3010                    3030                    3050
CGTTTTATTTACAGATTCTACGGAAGCGGAAGCGATTAAGTTATTTTCAAATACTTATTT
 V   L   F   T   D   S   T   E   A   E   A   I   K   L   F   S   N   T   Y   L

        3070                    3090                    3110
GGCTATGCGAGTTGCATTTTTTAATGAATTGGATAGTTACGCAGAAAGTTTTGGTCTGAA
 A   M   R   V   A   F   F   N   E   L   D   S   Y   A   E   S   F   G   L   N

        3130                    3150                    3170
TACGCGTCAGATTATTGACGGTGTTTGTTTGGATCCGCGCATTGGTAATTACTACAATAA
 T   R   Q   I   I   D   G   V   C   L   D   P   R   I   G   N   Y   Y   N   N

        3190                    3210                    3230
TCCTTCTTTTGGTTATGGTGGCTACTGTTTGCCAAAAGATACCAAGCAATTATTAGCCAA
 P   S   F   G   Y   G   G   Y   C   L   P   K   D   T   K   Q   L   L   A   N

        3250                    3270                    3290
CTATCAGTCTGTTCCGAATAAACTTATATCTGCAATTGTTGATGCTAACCGTACACGTAA
 Y   Q   S   V   P   N   K   L   I   S   A   I   V   D   A   N   R   T   R   K

        3310                    3330                    3350
GGACTTTATCACTAATGTTATTTTGAAACATAGACCACAAGTTGTGGGGGTTTATCGTTT
 D   F   I   T   N   V   I   L   K   H   R   P   Q   V   V   G   V   Y   R   L

        3370                    3390                    3410
GATTATGAAAAGTGGTTCAGATAATTTTAGAGATTCTTCTATTCTTGGTATTATAAAGCG
 I   M   K   S   G   S   D   N   F   R   D   S   S   I   L   G   I   I   K   R
```

130

```
        3430                3450                3470
TATCAAGAAAAAAGGCGTGAAAGTAATTATTTATGAGCCGCTTATTTCTGGAGATACATT
 I  K  K  K  G  V  K  V  I  I  Y  E  P  L  I  S  G  D  T  F

        3490                3510                3530
CTTTAACTCACCTTTGGAACGGGAGCTGGCGATCTTTAAAGGGAAAGCTGATATTATTAT
 F  N  S  P  L  E  R  E  L  A  I  F  K  G  K  A  D  I  I  I

        3550                3570   Sal I        3590
CACTAACCGAATGTCAGAGGAGTTGAACGATGTGGTCGACAAAGTCTATAGTCGCGATTT
 T  N  R  M  S  E  E  L  N  D  V  V  D  K  V  Y  S  R  D  L

        3610                3630                3650   Nsi I
GTTTAAATGTGACTAATGTATTGTTATATACTATTAACTATTAAGAGAAGGAAATGCATT
 F  K  C  D  *

        3670                3690                3710
ATTTAATCCGTTAAAAATATGCCTCGTTGGTATGTTCTTTATTAATCCTCGATCGTAAAA
          ————————▶      ◀————————
        3730                3750                3770
TAAGATGTTGGTTAAATTTCTTGCCCTGCTGTTGTGTTTATATCGAATTTTTCATAACTC

        3790                3810                3830
TAATGTTAATTAGAATACTTGGCTTGTGCGTTGGAAATGTATATTATGTGCTATATCCAA

EcoR V
AGATATC
```

FIGURE 6.11 The DNA sequence of the 3.8 Kb *Cla*I–*Bgl*II fragment is shown. Numbers indicate nucleotide coordinates. Amino acid translation for each *orf* is shown underneath the nucleotide sequence with capital letters. Restriction sites are shown, over the respective nucleotide sequence, with empty characters. The DNA motifs, identified 5' to *orf1*, which are most homologous to the −35 and −10 consensus sequences, are shown with bold characters. Potential translational start codons and the encoded methionine (M) are shown with bold characters highlighted with double lines. Stars indicate translational stop codon. Arrows indicate a stem and loop structure (sequence is shown with bold characters).

CHAPTER 7

MUTAGENESIS OF THE K5 CAPSULAR GENE CLUSTER

K5 polysaccharide biosynthesis is a complex process which involves several steps not yet well characterized. The aims of this work were to detail the organization of the K5 capsular gene cluster. The first step was to sequence the *kps* locus of the K5 capsular gene cluster and to assign the identified *orfs* to proteins visualized by minicells analysis. Additionally, computer analysis of the K5 capsular gene cluster-encoded products provided preliminary information about the possible function played by some of products in K5 polysaccharide synthesis and export. The second step was to generate mutants, the analysis of which might have provided useful data about the role played by products missing in these mutants. For example, immunoelectron microscopy of mutants defective in polysaccharide export can be used to identify the cellular location of polymer and make it possible to relate steps of polysaccharide export with products missing in these mutants. Additionally, purification and biochemical analysis of intracellular polymers can provide information about the presence of PA and/or KDO, and make it possible to identify those gene products which might promote the linkage between polysaccharides and PA and/or KDO.

An initial attempt to mutagenize pGB110 was performed using lambda::Tn5*pho*A. Tn5*pho*A was selected to provide information about the cellular location of the *kps*

encoded products and for possible investigation of gene expression at different growth temperatures.

7.1 Tn5*pho*A MUTAGENESIS OF pGB110

LE392(pGB110) was infected with lambda::Tn5*pho*A bacteriophage. Several (ApKm)^R colonies were investigated by small scale extraction of plasmid DNA and digestion with appropriate restriction enzymes. Only five contained a Tn5*pho*A insertion in pGB110, namely pGB110::9, pGB110::10, pGB110::21, pGB110::28 and pGB110::34 (Figure 7.1). LE392 harbouring pGB110 with these insertions was analyzed for K5 bacteriophage sensitivity and by immunoelectron microscopy (NB, the immunoelectron microscopy was performed by D Bronner, Freiburg). Additionally, polysaccharide was extracted from each of these mutants and examined by double immunodiffusion against K5 monoclonal antibody (Table 7.1). K5 bacteriophage sensitivity provided information about polysaccharide expression on the cell surface, double immunodiffusion expanded about synthesis of the K5 polysaccharide itself, and immunoelectron microscopy provided data about the cellular location of intracellular polymer.

FIGURE 7.1   Tn5phoA MUTAGENESIS OF pGB110
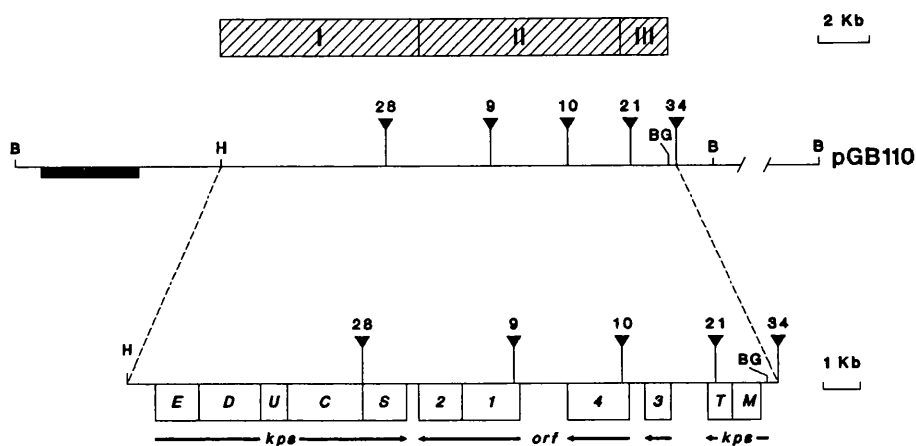
FIGURE 7.1 The linear map of pGB110 is shown with horizontal lines and the scale is shown in kilobase pairs. The filled box refers to vector sequence. The space between horizontal lines interrupted by dashed lines indicates sequence present in the original construct but not shown in figure. Filled boxes with vertical lines and labelled I, II and III, identify the regions of the K5 capsular gene cluster on pGB110. Vertical and oblique lines indicate restriction sites (capital letters) and Tn5phoA insertions (filled triangles; numbers refer to specific insertions discussed in text). Broken lines highlight the enlargement on scale of the organization of the K5 capsular gene cluster. Boxes labelled with capital letters or numbers refer to the identified *orfs* of the K5 capsular gene cluster. Arrows indicate the direction of transcription (*kps* refers to the identified *orfs* in region I and III, whereas *orf* refers to those identified in region II). Abbreviations: B, *Bam*HI; H, *Hind*III; BG, *Bgl*II.

TABLE 7.1   K5 BACTERIOPHAGE SENSITIVITY AND DOUBLE IMMUNODIFFUSION OF LE392(pGB110) MUTANTS

Strain	Polymer detected by immunodiffusion	K5 ϕ sensitivity
LE392(pGB110)	+	S
LE392(pGB110::9)	–	R
LE392(pGB110::10)	–	R
LE392(pGB110::21)	+	R
LE392(pGB110::28)	+	R
LE392(pGB110::34)	+	S

TABLE 6.1 Abbreviations: R, resistant; S, sensitive

Insertion pGB110::28 maps in region I either at the end of *kps*C or at the beginning of *kps*S (Figure 7.1). LE392(pGB110::28) is resistant to K5 bacteriophage infection but produces intracellular polysaccharide which reacts with K5 monoclonal antibody (Table 7.1). Immunoelectron microscopy of LE392(pGB110::28) identified intracellular polysaccharide in the cytoplasm. In LE392(pGB110::28) the K5 monoclonal antibody reacted over large areas of the cytoplasm which corresponded to areas of low electron density (this phenotype will be referred in this thesis as "holes"). Insertions pGB110::9 and pGB110::10 map in region II within *orf*1 and *orf*4, respectively (Figure 7.1). LE392(pGB110::9) and LE392(pGB110::10) are resistant to K5 bacteriophage infection and do not produce intracellular polysaccharide which reacts with K5 monoclonal antibody (Table 7.1). Immunoelectron microscopy of LE392(pGB110::9) did not

reveal any intracellular polysaccharide, whereas that of LE392(pGB110::10) identified intracellular polysaccharide associated with the inner face of the cytoplasmic membrane. The reason why intracellular polysaccharide was not detected by the double immunodiffusion analysis of LE392(pGB110::10) is unknown. Insertion pGB110::21 maps in region III, within *kps*T (Figure 7.1). LE392(pGB110::21) is resistant to K5 bacteriophage infection and produces intracellular polysaccharide which reacts with K5 monoclonal antibody (Table 7.1). Immunoelectron microscopy of LE392(pGB110::21) identified polysaccharide in the cytoplasm. Insertion pGB110::34 maps about 400 bp 5' to the first AUG of *kps*M (Figure 7.1). LE392(pGB110::34) is sensitive to K5 bacteriophage infection yielding plaques indistinguishable from those formed by the same bacteriophage on LE392(pGB110) (Table 7.1). LE392(pGB110::34) produces polysaccharide which reacts with K5 monoclonal antibody and immunoelectron microscopy of this mutant showed that LE392(pGB110::34) is encapsulated. None of these insertions abolishes the high levels of CMP-KDO synthetase typical of LE392(pGB110) (D Bronner, personal communication). However a variable level of enzyme activity was seen with these mutants (Table 7.2). The reason for these different levels of CMP-KDO synthetase activity is not known.

TABLE 7.2   CMP-KDO SYNTHETASE ACTIVITY IN LE392(pGB110) MUTANTS

Strain	CMP-KDO synthetase activity at 37°C (nmol of CMP-KDO/mg of protein in 15 min)
LE392	35.1
LE392(pGB110)	835.7
LE392(pGB110::9)	380.0
LE392(pGB110::10)	590.0
LE392(pGB110::21)	1290.0
LE392(pGB392::28)	350.0
LE392(pGB110::34)	480.0

In summary it appears that ORF1 is essential for K5 polysaccharide production. However the absence of polysaccharide in LE392(pGB110::9) might also be due to possible transcriptional polar effects, which would result in bacteria lacking both ORF1 and ORF2. Complementation experiments of LE392(pGB110::9) with exogenous UDP-GlcNAc and UDP-GlcA suggest that ORF1 and ORF2 are involved in steps of K5 synthesis which might be independent from the availability of UDP-GlcA and UDP-GlcNAc (see section 6.2). There is the possibility that ORF1 and ORF2 might function as GlcNAc and GlcA transferases, respectively. This might explain why LE392(pGB110::9) does not produce K5 polymer. orf4 is likely to constitute a single transcriptional unit (see sections 6.1 and 6.2) and the insertion in pGB110::10 maps in this orf. This mutant should, theoretically, produce all products required for K5 polysaccharide

biosynthesis except from ORF4. LE392(pGB110::10) contains intracellular polymer associated with the inner face of the cytoplasmic membrane. This suggests that ORF4 is involved in step(s) of K5 synthesis after initiation of polymerization.

Region III is composed of two genes, *kps*M and *kps*T, possibly organized into a single transcriptional unit (Smith *et al.*, 1990; Pavelka *et al.*, 1991). pGB110::21 maps in *kps*T and LE392(pGB110::21) should, theoretically, produce all products necessary for K5 polysaccharide biosynthesis except from KpsT. LE392(pGB110::21) results in intracellular polymer in the cytoplasm. Interestingly, LE392 harbouring the K5 capsular gene cluster with a deleted region III results in intracellular polymer associated with the inner face of the cytoplasmic membrane (D Bronner, personal communication). Taken together, these results suggest firstly that the K5 polymer is, during its production, associated with the cytoplasmic membrane. KpsM is thought to be a membrane-spanning protein involved in polysaccharide export across the inner membrane, whereas KpsT has been proposed to function as an "energy provider" coupling ATP-hydrolysis to the export system itself (Smith *et al.*, 1990; Pavelka *et al.*, 1991). Polysaccharide polymerization has been proposed to occur at the inner face of the cytoplasmic membrane (Janas and Troy, 1989) and membrane acceptors, such as undecaprenyl phosphate, are thought to be required for polymerization (Troy *et al.*, 1975; 1979). It is possible that either during or after polymerization, polysaccharide is exported across the cytoplasmic membrane by interacting with proteins involved in polysaccharide export. At this stage it is also possible that polysaccharide is detached from the cytoplasmic membrane and may become associated with polysaccharide export proteins. The absence of both KpsM and KpsT might

decouple polysaccharide synthesis from export. In this case polymer would remain associated with the site of polymerization, the inner face of the cytoplasmic membrane. Polysaccharide export is also thought to be an energy dependent process (Smith *et al.*, 1990; Pavelka *et al.*, 1991). Absence of only KpsT might block the export system from an energetic point of view. Taken together, these considerations suggest firstly that KpsM might play a key role in coupling polysaccharide synthesis to the export system. Secondly, the hypothesised association between polysaccharide and export proteins might require KpsT, the absence of which would de-couple polysaccharide from the export system, possibly leading to release of polymer into the cytoplasm.

Region I is composed of at least five genes (*kps*E, *kps*D, *kps*U, *kps*C and *kps*S), possibly organized into a single transcriptional unit (see sections 5.2 and 5.4). pGB110::28 maps either at the end of *kps*C or at the beginning of *kps*S and LE392(pGB110::28) results in intracellular polymer in the form of "holes". pGB118::1 is a Tn1000 insertion in pGB118 and maps within *kps*S and, interestingly, immunoelectron microscopy of LE392(pGB118::1) showed an intracellular polymer also in the form of "holes" (D Bronner, personal communication). However, very little is known about the basis of this particular phenotype. For example whether it is simply due to an accumulation of polymer in the cytoplasm or to the presence of a polysaccharide-protein(s) complex. KpsC and KpsS are encoded by region I, which is required for polysaccharide export (Boulnois and Jann, 1989). Additionally, KpsC and KpsS have also been hypothesised to play an unknown role in K1 polymerization, since mutations located in *kps*C or *kps*S affect sialyltransferase activity *in vitro* (Vimr *et al.*, 1989). On the basis of results obtained from analysis of LE392

harbouring pGB110::28 or pGB118::1 it seems likely that KpsS and maybe KpsC are required for polysaccharide export through the cytoplasmic membrane. However, these results provide no support for the hypothesised involvement of KpsS and KpsC in polysaccharide polymerization.

Finally, LE392(pGB110::34) is encapsulated suggesting that the DNA region where insertion pGB110::34 maps does not play an obvious role in K5 polysaccharide biosynthesis. pGB110::34 maps about 400 bp 5' to the first AUG of *kps*M and this insertion might define the boundary, 5' to region III, of the *kps* locus.

7.2 MUTAGENESIS OF REGION I OF THE K5 CAPSULAR GENE CLUSTER

The aims of this work are mainly focused on the role played by region I products in K5 polysaccharide biosynthesis. Data produced by previous studies (Boulnois *et al.*, 1987; Silver *et al.*, 1987; Vimr *et al.*, 1989; Finke *et al.*, 1989) and described in this thesis (see chapter 5) have highlighted the importance of this region. KpsU might play a role in initiation of polymerization (Finke *et al.*, 1989). KpsS and maybe KpsC might promote, in conjunction with KpsM and KpsT, polysaccharide transfer across the cytoplasmic membrane (see section 6.1). Additionally, KpsS and KpsC might also play an undefined role in polysaccharide polymerization (Vimr *et al.*, 1989). Whereas, KpsE and KpsD might be required for polysaccharide export from the cytoplasmic to the outer membrane (Silver *et al.*, 1987; Boulnois and Jann, 1989). Region I is composed of five genes possibly

organized into a single transcriptional unit. Mutagenesis of this region by transposon insertions result in transcriptional polar effects, which complicate the interpretation of results obtained from the study of mutants. Thus, it was important to choose strategies of mutagenesis which did not result, at least theoretically, in possible polar effects. For example, appropriate DNA deletions within a gene should not produce transcriptional polar effects 3' to the site of DNA deletion. Additionally, cloning of appropriate synthetic oligonucleotides which contain a translational stop within and in frame with a gene, would have interrupted the gene's translation without affecting the transcription 3' to the site of cloning.

The first step was to subclone the smallest DNA fragment which still encoded K5 polysaccharide synthesis and export to reduce the number of restriction sites and hence to facilitate mutagenesis. The insertion pGB110::34 was thought to define the boundary, 5' to region III, of the K5 capsular gene cluster. A 13.3 Kb *XhoI-SmaI* fragment was subcloned from pGB110::34 into pACYC184 generating pPC1 (Figure 7.2). pPC1 lacks *kpsE*, *kpsD*, *kpsU* and part of *kpsC*. LE392(pPC1) is resistant to K5 bacteriophage infection and produces levels of CMP-KDO synthetase activity comparable to LE392 (Table 7.3). Immunoelectron microscopy of LE392(pPC1) showed intracellular polymer in the cytoplasm in the form of "holes" (NB, immunoelectron microscopy and CMP-KDO synthetase assays of all pPC constructs were performed by D Bronner). Subsequently, a 8.4 Kb *XbaI-MluI* fragment was subcloned from pH18 into pPC1, cleaved with *XbaI* and *MluI*, generating pPC2 (Figure 7.2). pPC2 contains regions I, II and III. LE292(pPC2) is sensitive to K5 bacteriophage infection yielding plaques similar to those formed by the same bacteriophage on LE392(pGB110).

LE392(pPC2) produces high levels of CMP-KDO synthetase activity only at capsule permissive temperatures (Table 7.3). Additionally, immunoelectron microscopy of LE392(pPC2) showed that only 20% of cells are encapsulated. pPC2 demonstrates that regions I, II and III of the K5 capsular gene cluster are sufficient for synthesis and export of this polysaccharide. pPC1 and pPC2 allowed targeted mutagenesis of region I without, theoretically, affecting transcription 3' to the site of mutagenesis.

The region 5' to *kps*E is thought to play an unknown role in production of levels of CMP-KDO synthetase activity at different growth temperatures (see sections 5.1 and 5.4). A DNA deletion in this region might have provided insight into its possible function. A 4 Kb *Sma*I-*Sph*I was subcloned from pH18 into pUC19, cleaved with *Hinc*II and *Sph*I, generating pSMSP2 (Figure 7.3). Then, the 4 Kb *Xba*I-*Sph*I fragment was subcloned from pSMSP2 into pPC3, cleaved with *Xba*I and *Sph*I, generating pPC6 (Figure 7.2). pPC6 lacks the 0.3 Kb *Hind*III-*Sma*I fragment 5' to *kps*E. LE392(pPC6) is sensitive to K5 bacteriophage infection yielding clear plaques similar to those formed by the same bacteriophage on K5 wild type strains. LE393(pPC6) contains high levels of CMP-KDO synthetase activity only at capsule permissive temperatures (Table 7.3). Interestingly, immunoelectron microscopy of LE392(pPC6) showed that 80% of cells are encapsulated. A 9.6 Kb *Xba*I-*Bgl*II fragment was subcloned from pPC6 into pACYC184, cleaved with *Xba*I and *Bam*HI, generating pC1 (Figure 7.3). Then, a 0.6 Kb *Sma*I-*Hpa*I deletion of pC1 generated pC2 (Figure 7.3). The 7.5 Kb *Xba*I-*Mlu*I fragment, which contains the 0.6 Kb *Sma*I-*Hpa*I deletion, was subcloned from pC2 into pPC1, cleaved with *Xba*I and *Mlu*I, generating pPC7 (Figure 7.2). pPC7 lacks, in addition to the 0.3 Kb *Hind*III-*Sma*I fragment 5' to *kps*E, the central

142

region of *kps*C. LE392(pPC7) is resistant to K5 bacteriophage infection and produces levels of CMP-KDO synthetase activity comparable to LE392 (Table 7.3). Immunoelectron microscopy of LE392(pPC7) showed intracellular polymer in the form of "holes". pPC2 contains two unique *Aat*II restriction sites in *kps*D. pPC4 was generated by digestion of pPC2 with *Aat*II and selfligation. pPC4 contains a 800 bp *Aat*II deletion in *kps*D (Figure 7.2), and LE392(pPC4) is resistant to K5 bacteriophage infection and produces levels of CMP-KDO synthetase comparable to LE392 (Table 7.3). Immunoelectron microscopy of LE392(pPC4) showed intracellular polymer in the periplasm.

FIGURE 7.2 MUTAGENESIS OF REGION I OF THE K5 CAPSULAR GENE CLUSTER
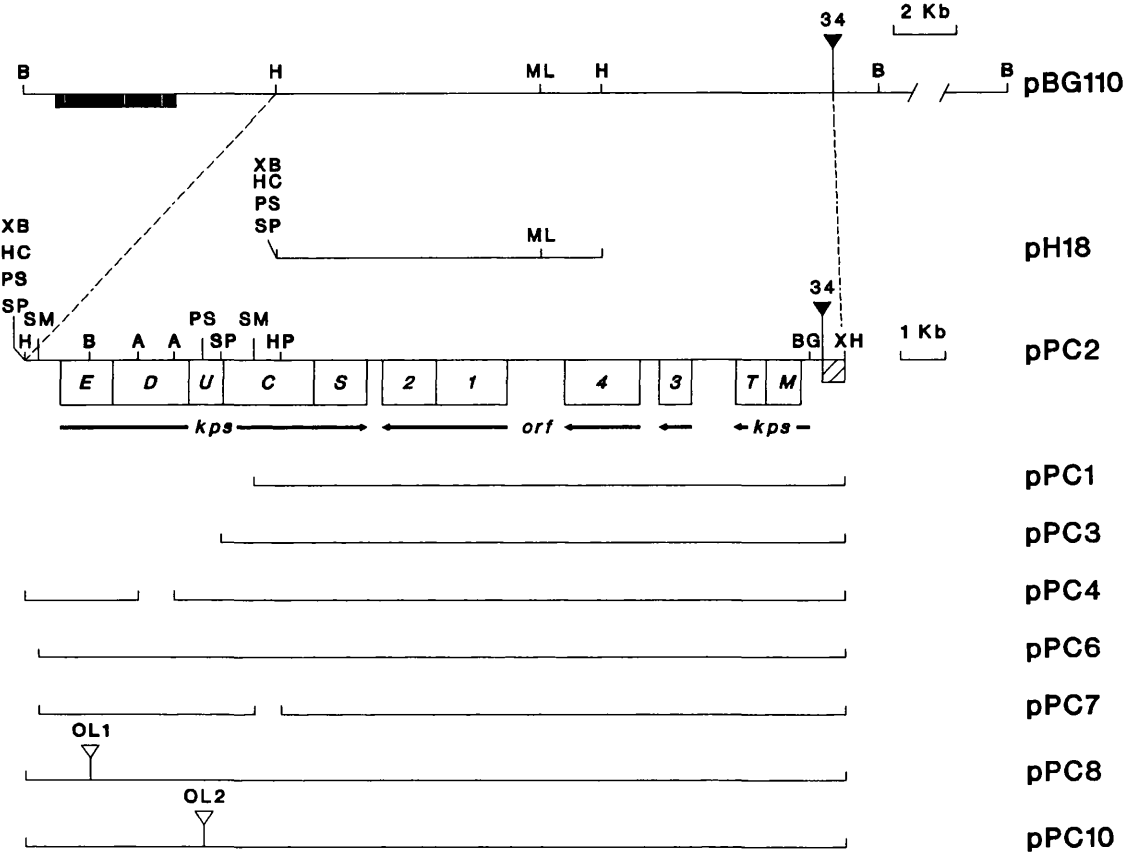
FIGURE 7.2 The linear map of various constructs is shown with horizontal lines and the scale is shown in kilobase pairs. The filled box indicates vector sequence whereas the box with oblique lines refers to sequence of Tn5*pho*A. The space between horizontal lines interrupted by dashed lines indicate sequence present in the original construct but not shown in figure. The space between horizontal lines interrupted by vertical lines indicates DNA deletions. Vertical and oblique lines indicate restriction sites (capital letters) or Tn5*pho*A insertion (filled triangle) or double stranded synthetic oligonucleotides (empty triangles, labelled OL1 or OL2; see text). Broken lines highlight the enlargement on scale of various pPC

144

constructs. Boxes labelled with capital letters or numbers refer to the identified *orfs* of the K5 capsular gene cluster. Arrows indicate the direction of transcription (*kps* refers to the identified *orfs* in region I and III, whereas *orf* refers to those identified in region II). Abbreviations: B, *BamHI*; H, *HindIII*; ML, *MluI*; XB, *XbaI*; HC, *HincII*; PS, *PstI*; SP, *SphI*; SM, *SmaI*; A, *AatII*; HP, *HpaI*; BG, *BglII*; XH, *XhoI*.

TABLE 7.3   K5 BACTERIOPHAGE SENSITIVITY AND CMP-KDO SYNTHETASE ACTIVITY IN LE392 HARBOURING VARIOUS pPC CONSTRUCTS

Strain	K5ϕ sensitivity	CMP-KDO synthetase activity (nmol of CMP-KDO/mg of protein 15 min)	
		18°C	37°C
LE392	R	55.8	35.1
LE392(pPC1)	R	<50.0	55.0
LE392(pPC2)	S	<50.0	176.0
LE392(pPC3)	R	<50.0	25.0
LE392(pPC4)	R	<50.0	30.0
LE392(pPC6)	S	<50.0	190.0
LE392(pPC7)	R	<50.0	120.0
LE392(pPC8)	R	<50.0	40.0
LE392(pPC10)	S	<50.0	20.0

TABLE 7.3 Abbreviations: ϕ, bacteriophage; R, resistant; S, sensitive

FIGURE 7.3   CONSTRUCTS GENERATED FOR MUTAGENESIS OF REGION I OF THE K5 CAPSULAR GENE CLUSTER
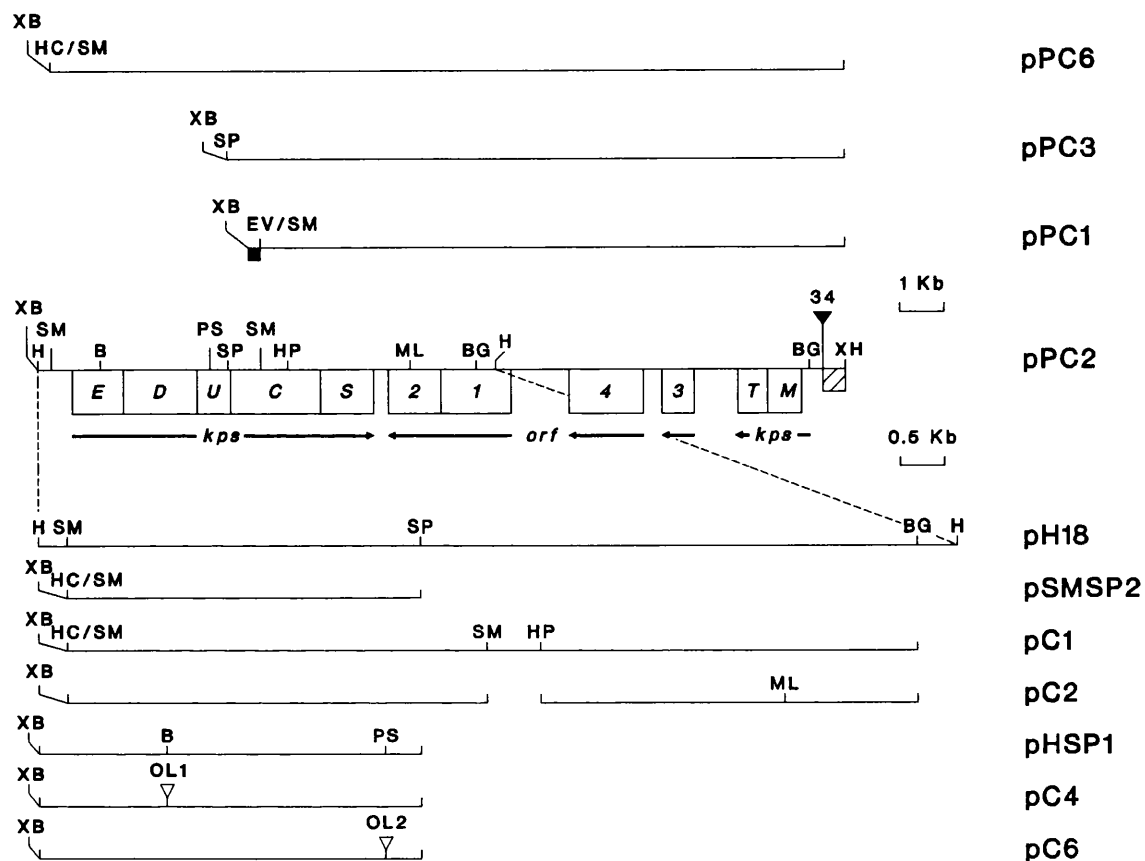
FIGURE 7.3 The linear map of various constructs is shown with horizontal lines and the scale is in kilobase pairs. Boxes labelled with capital letters or numbers refer to the identified *orfs* of the K5 capsular gene cluster. Arrows indicate the direction of transcription (*kps* refers to the identified *orfs* in region I and III, whereas *orf* refers to those identified in region II). The box with oblique lines refers to sequence of Tn5*pho*A, whereas the filled box refers to sequence of pACYC184. Horizontal lines interrupted by vertical lines indicate DNA deletions. Vertical and oblique lines indicate restriction sites (capital letters), or Tn5*pho*A insertion

(filled triangle) or double stranded synthetic oligonucleotides (empty triangles and labelled with OL1 or OL2). Broken lines highlight the enlargement on scale (2:1). Abbreviations: XB, *XbaI*; HC, *HincII*; SM, *SmaI*; SP, *SphI*; EV, *EcoRV*; H, *HindIII*; B, *BamHI*; PS, *PstI*; HP, *HpaI*; ML, *MluI*; BG, *BglII*; XH, *XhoI*.

The absence of obvious suitable restriction sites within *kps*E and *kps*U, did not allow mutagenesis by DNA deletion within these genes. However, the presence of single restriction sites allowed a mutagenesis of these genes by cloning of appropriate synthetic oligonucleotides, which contained translational stop codon in frame with *kps*E or *kps*U. A 4.3 Kb *HindIII-SphI* fragment was subcloned from pH18 into pACYC184 cleaved with *HindIII* and *SphI*, generating pHSP1 (Figure 7.3). pHSP1 contains an unique *BamHI* site in *kps*E and a unique *PstI* site in *kps*U. Two complementary oligonucleotides were used to generate the double stranded oligonucleotide OL1 (Figure 7.4). OL1 contains ends compatible with *BamHI* and a recognition sequence for *NotI*. OL1, regardless of the possible orientation of cloning, also contains a translational stop in frame with *kps*E. Additionally, cloning of OL1 into a *BamHI* site generates a sequence which no longer retains the *BamHI* site because the base 3' to the *BamHI* ends of OL1 is different from that of the *BamHI* site. OL1 was subcloned into pHSP1, cleaved with *BamHI*, generating pC4 (Figure 7.3). Cloning of OL1 into pHSP1 was verified by small scale extraction of plasmid DNA and digestion with appropriate restriction enzymes. pC4 lacks the unique *BamHI* site of pHSP1 and contains a *NotI* site absent in pHSP1. The 4.4 Kb *XbaI-SphI* fragment, which contains OL1, was subcloned from pC4 into pPC3, cleaved with *XbaI* and *SphI*, generating pPC8 (Figure 7.2). Verification of pPC8 was performed by appropriate restriction digestions (Figure 7.5). LE392(pPC8) is resistant to K5 bacteriophage infection and produces

levels of CMP-KDO synthetase activity comparable to LE392 (Table 7.3). Immunoelectron microscopy of LE392(pPC8) showed intracellular polymer in the periplasm. Two other complementary synthetic single stranded oligonucleotides were used to generate the double stranded oligonucleotide OL2 (Figure 7.4). OL2 contains ends compatible with *Pst*I and a recognition sequence for *Not*I. OL2, regardless of the possible orientation of cloning, also contains a translational stop in frame with *kps*U. Additionally, cloning of OL2 into a *Pst*I site generates a sequence which no longer retains the *Pst*I site. The base 5' to the compatible *Pst*I ends of OL2 is different from that of the *Pst*I site. OL2 was subcloned into pHSP1, cleaved with *Pst*I, generating pC6 (Figure 7.3). Cloning of OL2 into pHSP1 was verified by small scale extraction of plasmid DNA and digestion with appropriate restriction enzymes. pC6 lacks the unique *Pst*I site of pHSP1 and contains a *Not*I site absent in pHSP1. The 4.4 Kb *Xba*I-*Sph*I fragment, which contains OL2, was subcloned from pC6 into pPC3, cleaved with *Xba*I and *Sph*I, generating pPC10 (Figure 7.2). Verification of pPC10 was performed by appropriate restriction digestions (Figure 7.5). LE392(pPC10) is sensitive to K5 bacteriophage infection and produces levels of CMP-KDO synthetase activity comparable to LE392 (Table 7.3). However, immunoelectron microscopy of LE392(pPC10) showed a mixed phenotype. Some cells are partially encapsulated, with the K5 polysaccharide not uniformly distributed around the cell surface. Whereas other cells contain intracellular polymer in the form of "holes".

alLeu   AspHisOchre   LeuVal

TTCTG³'   ⁵'GATCATTAATAGCGGCCGCTCATCAAT³'   ⁵'GATCCAGAAC
AAGACCTAG₅'   ₃'TAATTATCGCCGGCGAGTAGTTACTAG₅'   ₃'GTCTTG

UmberHisAsp

cohesive   _Not_I   cohesive
_Bam_HI   _Bam_HI

| _kps_E | OL1 | _kps_E |

snLeu   HisUmber   LeuAsn

ACCTGCA³'   ⁵'TTGATGAGCGGCCGCTATTAATGCA³'   ⁵'GGTT
TGG₅'   ₃'ACGTAACTACTCGCCGGCGATAATT₅'   ₃'ACGTCCAA

OchreHisLeu

cohesive   _Not_I   cohesive
_Pst_I   _Pst_I

| _kps_U | OL2 | _kps_U |
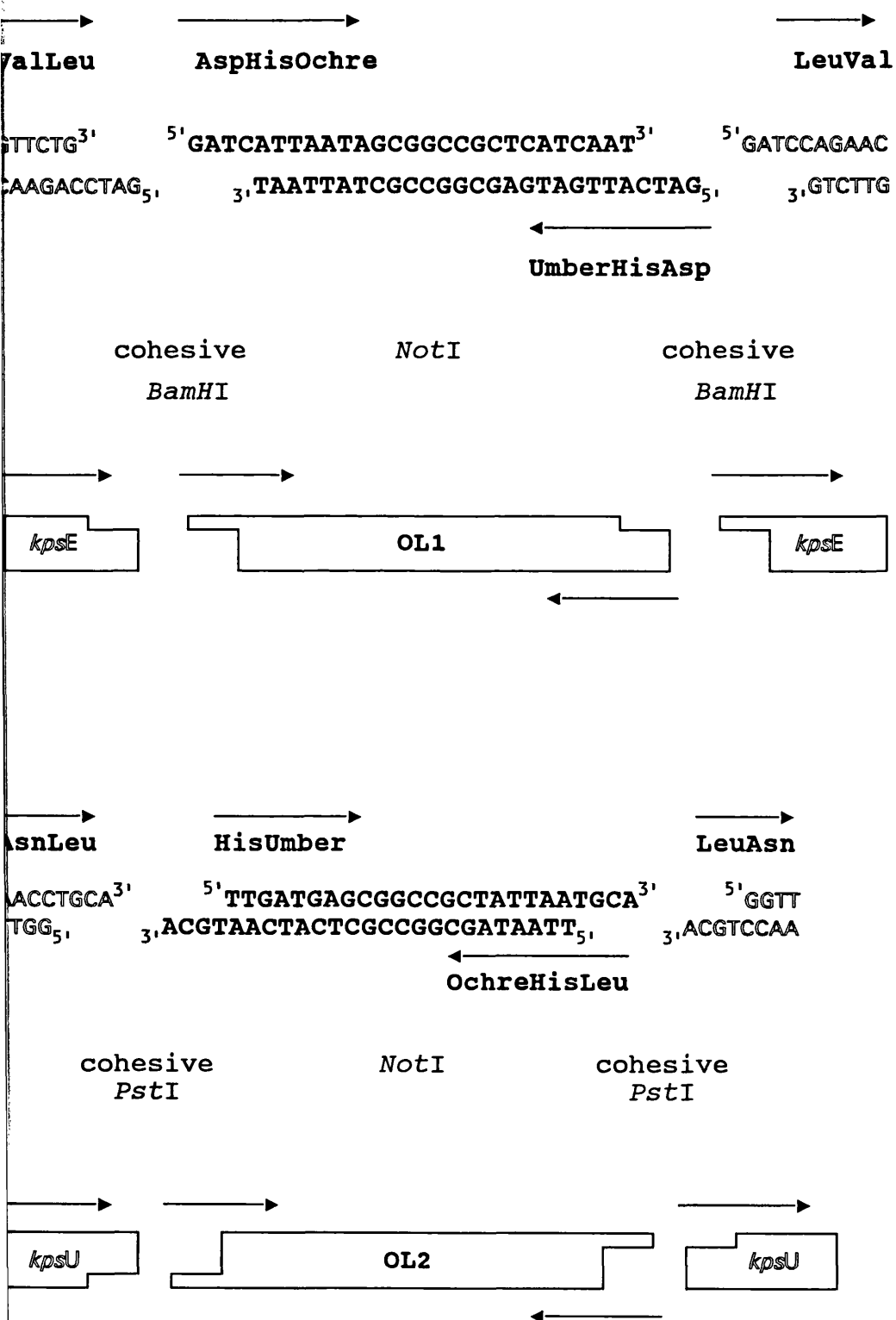
FIGURE 7.4 Letters with empty characters refer to the nucleotide sequence of *kps*E and *kps*U. Letters with bold characters refer to either synthetic oligonucleotides or to the translational frame of *kps*E, *kps*U, OL1 and OL2. Arrows indicate the direction of translation. Boxes labelled with empty characters (*kps*E or *kps*U) or bold characters (OL1 or OL2) align genes and synthetic oligonucleotides to the respective sequences. Ochre (TAA) and Umber (TGA) refer to translational stops in frame with *kps*E or *kps*U.

FIGURE 7.5   VERIFICATION OF pPC8 AND pPC10
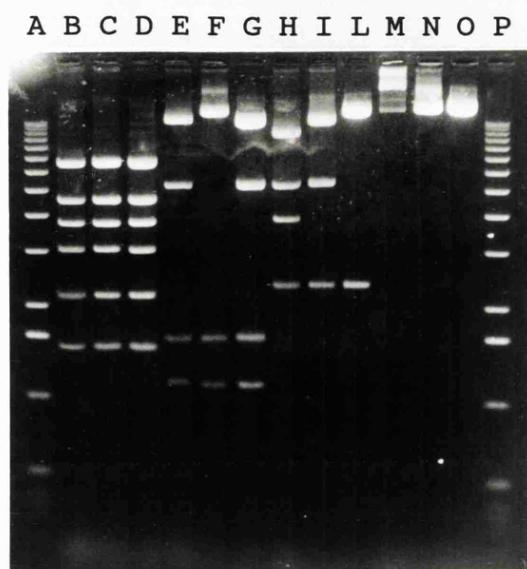
FIGURE 7.5   Lanes A and P contain molecular weight markers (Kb): 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1.6, 1, 0.5. Lanes B, E, H and M contain pPC2 cleaved with *EcoRV*, *BamHI*, *PstI* and *NotI*, respectively. Lanes C, F, I and N contain pPC8 cleaved with *EcoRV*, *BamHI*, *PstI* and *NotI*, respectively. Lanes D, G, L and O contain pPC10 cleaved with *EcoRV*, *BamHI*, *PstI* and *NotI*, respectively.

151

7.3 DISCUSSION

A DNA fragment of about 18 Kb has been subcloned from pGB110::34 into pACYC184 generating pPC2. pPC2 contains only regions I, II and III, and LE392(pPC2) is sensitive to K5 bacteriophage infection. Consequently this demonstrates for the first time that regions I, II and III are sufficient for K5 polysaccharide synthesis and export. However, immunoelectron microscopy of LE392(pPC2) showed that only 20% of cells are encapsulated. Similar data were obtained by immunoelectron microscopy of LE392(pGB110) (Kroncke *et al.*, 1990a). This raises the question of why surface expression of the K5 polysaccharide varies between K5 wild type strains and laboratory strains harbouring plasmids with the genetic information sufficient for K5 polysaccharide biosynthesis. The 0.3 Kb *Hind*III-*Sma*I deletion 5' to *kps*E, which generated pPC6, results in increased encapsulation from 20% to 80%. Plaques formed on LE392(pPC6) were clear and similar to those formed by the same bacteriophage on a K5 wild type strains (Kroncke *et al.*, 1990a). In contrast, plaques formed by the same bacteriophage on LE392(pPC2) or LE392(pGB110) were turbid. Thus, it is firstly possible to relate, approximately, plaque morphology with extent of encapsulation. Secondly, the 0.3 Kb *Hind*III-*Sma*I fragment might play an undefined role in determining the extent of encapsulation. The region 5' to *kps*E is conserved among the K5 and K7 capsular gene clusters (A Smith, personal communication). Additionally, by Southern blot experiments, the region immediately 5' to *kps*E has been shown to be conserved among different group II capsular gene clusters and also to be specific for this group of capsular polysaccharides (R Drake *et al.*, unpublished results). The different extent of encapsulation between LE392(pPC2) and LE392(pPC6) might imply for this region a

152

possible regulatory function. Additionally, the different
levels of CMP-KDO synthetase activity between K5 wild
type strains, LE392(pGB110) and LE392(pPC2) might be
explained by different levels of DNA supercoiling, in
which the region 5' to *kps*E might be involved. For
example, two motifs similar to the -35 and -10 consensus
sequences map in this region (see section 5.2) and might
account for a possible promoter activity, the expression
of which might be sensitive to changes in DNA
supercoiling. Different levels of DNA supercoiling have
already been shown to play a regulatory function in
expression of virulence genes. For example, in *Salmonella
typhimurium* (Galan and Curtis III, 1990) and *Shigella
flexneri* (Dorman *et al.*, 1990), the expression of
invasion genes has been shown to be sensitive to changes
in supercoiling. However, LE392(pPC2) and LE392(pPC6)
contain comparable levels of CMP-KDO synthetase activity
at different growth temperatures (Table 7.3), implying
that the 0.3 Kb *Hind*III-*Sma*I fragment 5' to *kps*E does not
affect the expression of *kps*U. Thus, the region 5' to
*kps*E might be a regulatory element in group II
polysaccharide biosynthesis, however the mechanism of
regulation is unknown.

pPC7 lacks the central region of *kps*C and immunoelectron
microscopy of LE392(pPC7) showed intracellular polymer in
the form of "holes". A similar phenotype was obtained by
immunoelectron microscopy of LE392(pGB110::28) and
LE392(pGB118::1) (see section 7.1). pPC3 lacks *kps*E, *kps*D
and *kps*U and, interestingly, immunoelectron microscopy of
LE392(pPC3) showed intracellular polymer in the
periplasm. Taken together, these results suggest that
KpsC and KpsS, in conjunction with KpsM and KpsT, are
required for polysaccharide export across the cytoplasmic
membrane. Purification and biochemical analysis of the
intracellular polymer of LE392(pGB118::1) has revealed

that this intracellular polymer is devoid of both KDO and PA (V Sieberth, personal communication). This firstly implies that KDO incorporation does not represent the initial step in group II polysaccharide polymerization, as proposed by Finke *et al.*, 1989. Secondly, KDO and PA are likely to be linked to polysaccharide after or during its polymerization. The assumption that KDO is not involved in initiation of polysaccharide polymerization also necessitates a re-examination on the meaning of high levels of CMP-KDO synthetase activity in group II polysaccharide biosynthesis. The previous hypothesised regulatory role of high levels of CMP-KDO synthetase activity (see section 2.6) might be re-interpreted as a possible functional requirement for polysaccharide export rather than for polysaccharide synthesis. For example, KDO represents the linkage between polysaccharide and PA, and PA is thought to anchor polymer to the outer membrane (Gotschlish *et al.*, 1981; Boulnois and Jann, 1989; Jann and Jann, 1990). Thus, it is possible that low levels of CMP-KDO synthetase activity or the absence of KpsU might affect the linkage between polysaccharide and PA and, consequently, represent a limiting factor in polysaccharide export.

pPC10 contains a mutation in *kps*U and LE392(pPC10) produces levels of CMP-KDO synthetase activity comparable to LE392. Immunoelectron microscopy of LE392(pPC10) showed a mixed phenotype. Some cells are partially encapsulated whereas others contain intracellular polymer in the form of "holes". This implies a series of possible hypotheses. Firstly, high levels of CMP-KDO synthetase activity appear essential for full encapsulation. Secondly, *kds*B cannot completely substitute for *kps*U. However, it is not understood if this lack of complementation reflects production of different levels of CMP-KDO synthetase activity or it is due to a

different role played by KdsB and KpsU in LPS and CPS production, respectively. Thirdly, intracellular polymer in the form of "holes" suggests the possibility of polar effects, since mutants in *kps*C and/or *kps*S result in intracellular polymer in the form of "holes". The first AUG of *kps*C overlaps the UGA translational stop of *kps*U (see section 5.2). It is possible that *kps*C is translationally coupled to *kps*U. In this case a premature translational termination 5' to *kps*C might affect expression of *kps*C and maybe *kps*S. Examples of translationally coupled genes have already been reported for tryptophan (Oppenheim and Yanofsky, 1980) and galactose (Schumperli *et al.*, 1982) operons, where the expression of some genes can be drastically reduced by a premature translational termination upstream of these genes. However, despite translational polar effects representing the most likely explanation, the possibility of transcriptional polar effects cannot be ruled out.

pPC4 and pPC8 contain a mutation in *kps*D and *kps*E, respectively. Immunoelectron microscopy of LE392(pPC4) and LE392(pPC8) showed intracellular polysaccharide in the periplasm, suggesting for KpsE and KpsD a role in polysaccharide export from the cytoplasmic to the outer membrane. However, LE392(pPC4) and LE392(pPC8) contain levels of CMP-KDO synthetase activity comparable to LE392 (Table 7.3). This suggests the possibility of polar effects which might affect expression of *kps*U and imposes caution in drawing conclusions about the possible function of KpsE and KpsD. Nevertheless, KpsD has been proposed to be a periplasmic protein (Silver *et al.*, 1987), whereas KpsE might be a protein which spans the cytoplasmic membrane with its hydrophilic central region orientated towards the periplasm (see sections 5.2 and 5.4). This implies a role for KpsE and KpsD in polysaccharide export after translocation across the

155

cytoplasmic membrane. Additionally, the intracellular location of polymer in LE392(pPC4) and LE392(pPC8) seems to exclude polar effects on expression of *kps*C and *kps*S. Thus, despite data obtained from LE392(pPC4) and LE392(pPC8) having to be interpreted with caution, it seems likely that KpsE and KpsD might promote polysaccharide export from the cytoplasmic membrane to the outer membrane.

CHAPTER 8

DISCUSSION

Biosynthesis of group II polysaccharide is a complex process which involves several steps not yet well characterized. It is possible to divide this process into polymerization and export phases. The phase of polymerization involves proteins which are encoded by a capsule specific DNA region, whereas the process of polysaccharide export occurs through a common mechanism which is independent of the chemical structure of the polysaccharide. Additionally, other features, such as the high levels of CMP-KDO synthetase activity and the temperature dependence of both enzyme activity and polysaccharide biosynthesis, make the study of group II capsules intriguing.

The K5 polysaccharide is a heteropolymer with a repeating unit composed of GlcA and GlcNAc. GlcNAc is also found in other structures of Gram-negative bacteria such as peptidoglycan and LPS, whereas GlcA is found in capsular and capsular-like structures such as the M-antigens. KDO, which is not part of the K5 polysaccharide repeating unit, has been found at the reducing end of the K5 polysaccharide. The presence of KDO at the K5 polysaccharide reducing end, the high level of CMP-KDO synthetase activity and the common temperature dependence of both enzyme activity and polysaccharide biosynthesis firstly led to the hypothesis that transfer of KDO might represent the initial step in polymerization. Secondly, that high level of CMP-KDO synthetase may reflect a specific functional requirement and maybe a regulatory

157

element in group II polysaccharide biosynthesis.

Results described in this thesis have firstly demonstrated that high level of CMP-KDO synthetase, produced by *E.coli* K5, is due to the presence of an additional CMP-KDO synthetase gene (*kps*U) located in the middle of region I. Secondly, this gene is 63.5% identical to *kds*B and both are likely to represent products of a gene duplication event. Thirdly, the high level of CMP-KDO synthetase is probably not essential for K5 polysaccharide synthesis but appears necessary for full encapsulation. The hypothesis that KDO is the first sugar which initiates K5 polysaccharide synthesis, and maybe all group II polysaccharides, seems very unlikely. Consequently, the presence of high level of CMP-KDO synthetase activity does not appear to be related with a mechanism by which the whole process of group II polysaccharide biosynthesis might be regulated. It is possible that high levels of CMP-KDO synthetase activity might represent a functional requirement for polysaccharide export. KDO is the linkage between polysaccharide and PA, and the latter is thought to anchor group II polysaccharides in the outer membrane.

Polysaccharide export is a process which occurs regardless of the chemical structure of polymers. The K5 capsular gene cluster has been sequenced and, at least, six proteins are involved in this export process. Four proteins, KpsM, KpsT, KpsC and KpsS are required for polysaccharide export across the cytoplasmic membrane. It is possible that polysaccharide is translocated across the cytoplasmic membrane via a protein-dependent process. KpsM, which is thought to be a cytoplasmic membrane-spanning protein, might promote polysaccharide export by interacting with KpsC and/or KpsS. Polysaccharide export

is also thought to be energy-dependent and KpsT might provide the required energy by coupling ATP-hydrolysis to the export system itself. KpsC and/or KpsS might take part in export by promoting protein-polysaccharide interactions. For example, KpsC is positively charged (it contains 80 positively charged amino acids versus 67 negatively charged), whereas group II polysaccharides are negatively charged. It is possible that KpsC interacts directly with group II polysaccharides, regardless of their chemical structures, by chemical bonds which involve different electrochemical charges.

Two other proteins, KpsE and KpsD, are likely to be involved in polysaccharide export after polymer translocation across the cytoplasmic membrane. KpsE, which might be an oligotopic protein with its hydrophilic central domain orientated towards the periplasm, shows homology with the membrane-bound PlsB *sn*-glycerol-3-phosphate acyltransferase of *E.coli*. PlsB catalyzes condensation of *sn*-glycerol-3-phosphate and fatty acids thioesters to yield lysophosphatidic acid (LPA) a precursor of PA. Whereas KpsD, which is thought to be a periplasmic protein, shares with the periplasmic PgpB protein of *E.coli* an homologous, hydrophilic amino acid stretch that might interact with PA. This suggests that KpsE might promote polysaccharide transfer to PA. Whereas KpsD might act as a chaperone for polysaccharide through the periplasmic space. In this model, KDO might simply represent the linkage between the reducing end of a group II polysaccharide and PA. Once the polysaccharide is transferred to PA, KDO would be found at the reducing terminus without having taken part in any aspect of the initiation of polymerization.

Production of high levels of CMP-KDO synthetase activity and group II polysaccharide biosynthesis are temperature dependent. This common feature has focussed attention on CMP-KDO synthetase activity as a possible regulatory element for group II polysaccharide biosynthesis. *kps*U is located in the middle of region I, the products of which are required for polysaccharide export. From an energetic point of view, synthesis, at temperatures below 20°C, of proteins involved in polysaccharide export would represent a waste of energy. Thus, the common temperature dependence of both high level of CMP-KDO synthetase activity and polysaccharide biosynthesis might simply reflect the location of *kps*U in a DNA region under the control of a regulation system for group II polysaccharide biosynthesis. The DNA sequence 5' to *kps*E has been found to be conserved between the K5 and K7 capsular gene clusters, and demonstrated to be specific to group II capsules. This region also appears to play a role in determining both production of high and low level of CMP-KDO synthetase activity at different growth temperatures and the extent of encapsulation. However, very little is known about mechanism(s) by which this DNA region might affect both CMP-KDO synthetase activity and encapsulation. The region 5' to *kps*E shows homology (55.5% identity over 695 bp) with *gul*Q of the glucitol operon of *E.coli*. This operon maps at 58 min of the *E.coli* chromosome, whereas *kps* maps at 64 min. Additionally, the function of GulQ in the D-glucitol transport system is still unknown. Thus, despite the region 5' to *kps*E remaining of key interest, it is not possible to speculate about the specific function played by this region in group II polysaccharide biosynthesis.

Proteins encoded by *kps* of *E.coli* show homology with those encoded by *cap* of *H.influenzae* and *cps* of *N.meningitidis*, respectively. This datum has further re-

enforced the hypothesis of a common evolutionary origin of encapsulation between these three genera. Additionally, data presented here have provided the first evidence of differences between some functions encoded by the *kps* locus and those encoded by *cap* and *cps* loci. It would be interesting to see if these differences reflect genera-specific requirements or are merely due to acquisition of different but functionally related genes.

Finally, region II of the K5 capsular gene cluster encodes at least four proteins which are required for synthesis of the K5 polysaccharide. Two of these products might function as GlcA and GlcNAc transferases. Whereas, very little is known about the possible function played by the other two products. Interestingly the G+C content of region II (33.4%) is low compared to the average GC ratio (48-52 mol%) of *E.coli* DNA. This atypical G+C content is also found at the 3' ends of *kps*S and *kps*T of the K5 capsular gene cluster. Additionally KpsS and KpsT, encoded by different capsular gene clusters, contain variable C-termini. Taken together, these results suggest firstly that recombination between an incoming and outgoing "region II-cassette" might occur by recombination events within *kps*S and *kps*T. Secondly, that region II of the K5 capsular gene cluster might have been acquired from another micro-organism.

The study of group II polysaccharide biosynthesis represents an engaging challange from many viewpoints such as gene regulation, polysaccharide polymerization and polysaccharide export across membranes. Data provided by this thesis have contributed to our understanding of the detailed genetic organization of the K5 capsular gene cluster and produced preliminary information about the function played by some of its products. It is also hoped

that these data will be useful for future experiments, which could detail more accurately the process of polysaccharide biosynthesis. At this stage, it is very difficult and premature to describe in detail the process of group II polysaccharide biosynthesis. Whereas, it may be more constructive to speculate about this complex mechanism of biosynthesis. The purpose of this final chapter is essentially to enlarge the debate on this mechanism and to propose new lines of investigation, the exploring of which would further gratify the efforts of this thesis.

REFERENCES

Albright CF, P Orlean and PW Robbins (1989). A 13-amino acid peptide in three yeast glycosyltransferases may be involved in dolichol recognition. PNAS USA 86:7366-7369

Austrian R and J Gold (1964). Pneunococcal bacteremia with reference to bacteremic pneumococcal pneumoniae. Ann Intern Med 60:759-776

Barry GT (1959). Colominic acid, a polymer of N-acetylneuraminic acid. J Exp Med 107:507-521

Bayer ME (1979). The fusion sites between outer and inner membrane of bacteria:their role im membrane assembly and virus infection. In:Inouye M (ed) Bacterial outer membrane. Wiley, New York.

Biggin MD, TJ Gibson and GF Hong (1983). Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. PNAS USA 80:3963-3965

Birnboim HC and J Doly (1979). A rapid alkaline extraction procedure for screening reconbinant plasmid DNA. Nucl Acid Res 1:1513-1523

Boulnois GJ, IS Roberts, R Hodge, KR Hardy, KB Jann, and KN 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

Boulnois GJ and K Jann (1989). Bacterial polysaccharide capsule synthesis, export and evolution of structural diversity. Molecular Microbiology 3(12):1819-1823

Boulnois GJ and IS Roberts (1990). Genetics of capsular polysaccharide production in bacteria. Curr Top Microbiol Immunol 150:1-18

Brosious J (1984). Plasmid vectors for the selection of promoters. Gene 27:151-160

Calogero RA, CL Pon, MA Canonaco and CO Gualerzi (1988). Selection of the mRNA translation initiation region by *Escherichia coli* ribosomes. Proc Natl Acad Sci USA 85:6427-6431

Chang ACY and SN Cohen (1978). Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the p15A cryptic miniplasmid. J of Bacteriology 134:1141-1156

Corsterton JN, Irwin RT and K-J Cheng (1981). The bacterial glycocalyx in nature and disease. Annu Rev Microbiol 35:299-324

Cross AS (1990). The Biologic Significance of Bacterial Encapsulation. Curr Top Microbiol Immunol 150:87-95

Dalbey RE (1990). Positively charged residues are important determinants of membrane protein topology. TIBS 15(6):253-257

Deretic V, JF Gill and AM Chakrabarty (1987). *Pseudomonas aeruginosa* infection in cystic fibrosis: nucleotide sequence and transcriptional regulation of the *alg*D gene. Nucleic Acids Res 15(11):4567-4581

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

De Witt CW, and JA Rowe (1961). Sialic acid (N, 7-O-diacetylneuraminic acid and N-acetylneuraminic acid) in *Escherichia coli*. Isolation and identification. J of Bacteriology 82:838-848

Dorman CJ, NI Bhriain and CF Higgins (1990). DNA supercoiling and environmental regulation of virulence gene expression in *Shigella flexneri*. Nature (London) 344:789-792

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

Drake CR, IS Roberts, B Jann, K Jann and GJ 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

Dudman WF (1977). The role of surface polysaccharides in natural environments. In:Surface carbohydrates of the prokaryotic cell. Sutherland IW (ed) Academic, New York, pp 357-414

Dykes G, R Bambara, K Marians and R Wu (1975). On the statistical significance of primary structural features found in DNA-protein interaction sites. Nucleic Acids Res 2:327-345

Echarti C, B Hirschel, GJ Boulnois, JM Valey, F Waldvogel, and KN Timmis (1983). Cloning and Analysis of the K1 Capsule Biosynthesis Genes of *Escherichia coli*: Lack of Homology with *Neisseria meningitidis* Group B DNA Sequences. Infection and Immunity 41(1):54-60

Engelman DM, TA Steitz and A Goldman (1986). Identifying non polar transbilayer helices in amino acid sequences of membrane proteins. Ann Rev Biophys Chem 15:321-353

Evans LR and A Linker (1973). Production and characterization of the slime polysaccharide of *Pseudomonas aeruginosa*. J of Bacteriology 116(2):915-924

Falkow S, IR Ryman and O Washington (1962). Deoxyribonucleic acid base composition of *Proteus* and *Providence* organisms. J of Bacteriology 83:1318-1321

Feingold DS and JS Frazen (1981). Pyridine nucleotide-linked four-electron transfer dehydrogenases. Trends Biochem Sci 6:103-105

Fine DP (1975). Pneumococal type-associated variability in alternated complement pathway activation. Infection and Immunity 12:772-778

Finke A, J Roberts, G Boulnois, C Pazzani, and K Jann (1989). Activity of CMP-2-Keto-3-Deoxyoctulosonic Acid Synthetase in *Escherichia coli* Strains Expressing the K5 Polysaccharide: Implication For K5 Polysaccharide Biosynthesis. J of Bacteriology 171(6):3074-3079

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

Finke A, D Bronner, AV Nikolaev, B Jann and K Jann (1991). Biosynthesis of the *Escherichia coli* K5 Polysaccharide, a Representative of Group II Polysaccharides: Polymerization *In Vitro* and Characterization of the Product. J of Bacteriology 173(13):4088-4094

Finne J, Finne V, Dlagostini-Bazin H and C Goridis (1983a). Occurence of α-2-8-linked polysialosyl units in a neural cell adhesion molecule. Biochim Biophys Res Commun 112:482-487

Finne J, M Leinomen, and PH Makela (1983b). Antigenic similarities between brain components and bacteria causing meningitis. Lancet ii: 355-357

Frosch M, C Weisgerber, and TF Meyer (1989). Molecular characterization and expression in *Escherichia coli* of the gene complex encoding the polysaccharide capsule of *Neisseria meningitidis* group B. Proc Natl Acad Sci USA 84:1669-1673

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. Molecular Microbiology 5(5):1251-1263

Galan JE and R Curtis III (1990). Expression of *Salmonella typhimurium* genes required for invasion is regulated by changes in DNA supercoiling. Infection and Immunity 58:1879-1885

Ghalambor MA and EC Heath (1966). The biosynthesis of cell wall lipopolysaccharide in *Escherichia coli*. IV. Purification and properties of cytidine monophosphate 3-deoxy-D-manno-octulosonate synthetase. J biol Chem 241:3216-3221

Gilson E, H Nikaido and M Hofnung (1982). Sequence of the *malK* gene in *E.coli* K12. Nucleic Acids Res 10(22):7449-7458

Glode MP, JB Robbins, T-Y Lin, EC Gotschlich, I Orskov and F Orskov (1977). Cross-antigenicity and immunogenicity between capsular polysaccharides of group C *Neisseria meningitidis* and *Escherichia coli* K92. J Infect Dis 135:94-102

Goldman RC and WE Kohlbrenner (1985). Molecular cloning of the structural gene coding for CTP:CMP-3-deoxy-manno-octulosonate Cytidyltransferase from *Escherichia coli* K-12. J of Bacteriology 163(1):256-261

Gotschlich EC, BA Fraser, O Nishimura, JB Robbins and T-Y Liu (1981). Lipid on capsular Polysaccharides of Gram-negative Bacteria. J Biol Chem 256(17):8915-8921

Gotschlich EC (1984). Meningococcal meningitis: Germanier E (ed) Bacterial vaccines. Academic, New York, pp 237-255

Govan JRW and GS Harris (1986). *Pseudomonas aeruginosa* and cystic fibrosis: unusual bacterial adaptation and pathogenesis. Microbiol Sci 3 (10):302-308

Grados O and WH Ewing (1970). Antigenic relantionships between *Escherichia coli* and *Neisseria meningitidis* group B. J Infect Dis 122:100-103

Gros P, J Croop and D Housman (1986). Mammalian multidrug resistance gene: complete cDNA sequence indicates strong homology to bacterial transport protein. Cell 47:371-380

Grosh S and S Roseman (1965). The sialic acids. IV. N-Acyl-D-Glucosamine-6-phosphate 2-epimerase. J Biol Chem 240:1525-1530

Guirguis NR, R Schneerson R, A Bax, W Egan, JB Robbins, J Shiloach, I Orskov, F Orskov and AE Kholly (1985). *Escherichia coli* K51 and K93 capsular polysaccharides are cross-reactive with group A capsular polysaccharide of *Neisseria meningitidis*. J Exp Med 162:1837-1851

Hallewell RA and DJ Sherratt (1976). Isolation and characterization of ColE2 plasmid mutants unable to kill Colicin-sensitive cells. Mol Gen Genet 146:239-245
Harley CB and RP Reynolds (1987). Analysis of *E.coli* promoters. Nucliec Acids Res 15:2343-2361

Hawley DK and WR McClure (1983). Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res 11:2237-2255

Hershfield V, HW Boyer, C Yanofsky, MA Lovett and DR Helinski (1974). Plasmid ColE1 as a molecular vehicle for cloning and amplification of DNA. Proc Natl Acad Sci USA 71:3455-3459

Higgins CF, Hiles ID, Whalley K and Jamieson DJ (1985). Nucleotide binding by membrane components of bacterial periplasmic binding protein-dependent transport systems. EMBO J 4:1033-1040

Higgins CF, MP Gallagher, SC Hyde, ML Mimmack and SR Pearce (1990a). Periplasmic binding protein-dependent transport systems: the membrane-associated components. Phil Trans R Soc Lond B 326:353-365

Higgins CF, Hyde SC, Mimmack MM, Gileadi U, Gill DR, and Gallagher MP (1990b). Binding protein-dependent transport systems. J Bioenerget Biomembr 22:571-592

Hirsch RL, DE Griffin and JA Winkelsteiln (1981). Host modification of Sindbis virus sialic acid content influences alternative complement pathway activation and virus clearance. J Immunol 127:1740-1743

Hobot JA, E Carlemalm, W Villiger and E Kellenberger (1984). Periplasmic gel: new concept resulting from the reinvestigation of bacterial cell envelope ultrastructure by new methodhs. J of Bacteriology 160(1):143-152

Holland IB, B Kenny and M Blight (1990). Haemolysin secretion from *E.coli*. Biochimie 72:131-141

Honeyman AL and GC Stewart (1989). The nucleotide sequence of the *rodC* operon of *Bacillus subtilis*. Molecular Microbiology 3(9):1257-1268

Horio M, MM Gottesman and I Pastan (1988). ATP-dependent transport of vinblastine in vescicles from human multidrug resistant cells. Proc Natl Acad Sci USA 85:3580-3584

Howard CJ and AA Glynn (1971). the virulence for mice of strains of *Escherichia coli* related to the effects of K antigens on their resistance to phagocythosis and killing by complement. Immunology 20:767-777

Horwitz MA and SC Silverstein (1980). influence of the *Escherichia coli* capsule on complement fixation and on phagocytosis and killing by human phagocytes. J Clin Invest 65:82-94

Hussain M, S Ichihara and S Mizushima (1982). Mechanism of signal peptide cleavage in the biosynthesis of the major lipoprotein of the *Escherichia coli* outer membrane. J Biol Chem 257(9):5177-5182

Hyde SC, Emsley P, Hartshorn MJ, Mimmack MM, Gileadi U, Pearce SR, Gallagher MP, Gill DR, Hubbard RF, and Higgins CF (1990). Structural model of ATP-binding proteins associated with cystic fibrosis multidrug resistance and bacterial transport. Nature 346:362-365

Icho T (1988). Membrane-bound phosphatase in *Escherichia coli*:sequence of the *pgp*B gene and dual subcellular localization of the *pgp*B product. J of Bacteriology 170(11):5117-5124

Iyanagi T, M Haniu, K Sogawa, Y Fujii-Kuriyama, S Watanabe, JE Shively and KF Anan (1986). Cloning and characterization of cDNA encoding 3-methylcholantrene inducibile rat mRNA for UDP-glucuronosyltransferase. J Biol Chem 261(33):15607-15914

Janas T and FA Troy (1989). in Proceedings of the Xth International Symposium on Glycoconjugates, Jerusalem, Israel (N Sharon, H Lis, D Duksin and I Kahane, eds) pp207-208. Organizing Committee X[th] International Symposium on Glycoconjugates, Jerusalem, Israel.

Jann K and B Jann (1983). The K antigens of *Escherichia coli*. Prog Allergy 33:53-79

Jann K, and B Jann (1985). Cell surface components and virulance: *Escherichia coli* O and K antigens in relation to virulence and pathogenicity, pp 156-176. In M.Sussman (ed), the virulance of Escherichia coli. Academic Press, Inc., New York.

Jann B and K Jann (1990). Structure and Biosynthesis of the Capsular Antigens of *Escherichia coli*. Curr top Microbiol Immunol 150:19-42

Jarvis GA and NA Vedros (1987). Sialic acid of group B *Neisseria meningitidis* regulates alternative complement pathway activation. Infection and Immunity 55(1):174-180

Jennings HJ (1990) Capsular Polysaccharides as Vaccine Candidates. Curr Top Microbiol Immunol 150:97-127

Joiner KA, EJ Brown and MM Frank (1984). Complement and bacteria: chemistry and biology in the host defence. Annu Rev Immunol 2:461-492

Kaijser B, Hanson LA, Jodal U, Lidin-Janson G, Robbins JB (1977) Frequency of *Escherichia coli* K antigens in urinary tract infections in children. Lancet i:663-664

Kenne L and Lindberg B (1983) Bacterial polysaccharides in Aspinall, G.O. (ed) The polysaccharides. Academic Press, New York pp. 287-360

Kroll IS, S Zamze, B Loynds, and ER Moxon. (1989) Common Organization of Chromosomal Loci for Production of Different Capsular Polysaccharides in *Haemophilus influenzae*. J of Bacteriology 171(6):3343-3347

Kroll JS, B Loynds, LN Brophy and ER Moxon. (1990) The *bex* locus in encapsulated *Haemophilus influenzae*: a chromosomal region involved in capsule polysaccharide export. Molecular Microbiology 4(11):1853-1856

Kroncke KD, G Boulnois, J Roberts, D Bitter-Suermann, JR Golecki, B Jann. (1990a) Expression of the *Escherichia coli* K5 Capsular Antigen: Immunoelectron microscopic and Biochemical Studies with Recombinant E.coli. J of Bacteriology 172(2):1085-1091

Kroncke KD, JR Golecki, and K Jann. (1990b) Further electron microscopic studies on the expression of *Escherichia coli* group II capsules. J of Bacteriology 172:3469-3472

Kuo JS-C, VW Doelling, JF Graveline and DW McCoy. (1985) Evidence for Covalent Attachment of Phospholipid to the Capsular Polysaccharide of *Haemophilus influenzae* Type b. J of Bacteriology 163(2):769-773

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

Lightner VA, RM Bell and P Modrich (1983). The DNA sequence encoding *plsB* and *dsk* loci of *Escherichia coli*. J Biol Chem 258(18):10856-10861

Lipman DJ and WR Pearson (1985). Rapid and sensitive protein similarity searches. Science 227:1435-1441

Maniatis T, EF Fritsch and J Sambrook (1982). Molecular cloning, a laboratory manual. Cold Spring Harbor, USA

Manoil C and J Beckwith (1985). Tn*phoA*: a transposon probe for protein export signals. Proc Natl Acad Sci USA 82:8129-8133

Masson L and BE Holbein (1985). Role of lipid intermediate(s) in the synthesis of serogroup B *Neisseria meningitidis* capsular polysaccharide. J of Bacteriology 161:861-867

McCarthy JEG and C Gualerzi (1990). Translation control of prokaryotic gene expression. TIG 6(3):78-85

Merker RI and FA Troy (1990). Biosynthesis of the polysialic acid capsule in *Escherichia coli* K1. Cold inactivation of sialic acid synthetase regulates capsule expression below 20°C. Glycobiology 1(1):93-100

Moxon ER and JS Kroll. (1990) The role of bacterial polysaccharide capsules as virulence factors. Curr Top Microbiol Immunol 150:65-85

Nikaido H and T Nakae (1979). The outer membrane of gram-negative bacteria. Adv Microb Physiol 20:163-250

O'Neil MC (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

Oppenheim DS and C Yanofsky (1980). Translational coupling during expression of the tryptophan operon of *Escherichia coli*. Genetics 95:785-795

Ordman AB and S Kirkwood (1977). Mechanism of action of uridine diphosphoglucose dehydrogenase. Evidence for an essential lysine residue at the active site. J Biol Chem 252:1320-1326

O'Reilly DR and LK Miller (1988). A Baculovirus blocks insect molting by producing ecdysteroid UDP-Glucosyl transferase. Science 245:1110-1112

Orskov I and K Nyman. (1974) Genetic Mapping of the Antigenic determinants of Two Polysaccharide K Antigens, K10 and K54, in *Escherichia coli*. J of Bacteriology 120(1):43-51

Orskov I, F Orskov, B Jann and K Jann. (1977) Serology, Chemistry, and Genetics of O and K Antigens of *Escherichia coli*. Bacteriological Reviews 41(3):667-710

Orskov I and F Orskov (1984). Serotyping of *Klebsiella*. Methods Microb 14:143-164

Orskov F (1984). The genus *Escherichia*, pp 420-423 In: Krieg NR and JG Holt (eds), Bergey's manual of systematic bacteriology, vol 1. Williams and Wilkins, Baltimore

Orskov F, V Sharma and I Orskov. (1984) Influence of Growth Temperature on the Development of *Escherichia coli* Polysaccharide K Antigens. J Gen Microbiol 130:2681-2684

Ortiz AI, A Reglero, CB Roderiguez-Aparicio, and J Luengo. (1989) *In vitro* synthesis of colominic acid by membrane-bound sialyltransferase of *Escherichia coli* K-235. Eur J Biochem 178:741-749

Overduin P, W Boos and J Tommassen (1988). Nucleotide sequence of the *ugp* genes of *Escherichia coli* K-12:homology to the maltose system. Molecular Microbiology 2(6):767-775

Owen RJ, AU Ahmed and CA Dawson (1987). Guanosine-plus-cytosine contents of type strains of the genus *Providencia*. Int J Syst Bacteriol 37:449-450

Pavelka MS, JR Lori, LF Wright, and RP Silver. (1991) Identification of Two Genes, *kpsM* and *kpsT*, in Region 3 of the Polysialic Acid Gene Cluster of *Escherichia coli* K1. J of Bacteriology 173(15):4603-4610

Peltola H. (1983) Meningococcal disease: still with us. Rev Infect Dis 5:71-91

Pittman M. (1931) Variation and type specific in the bacterial species *Haemophilus influenzae*. J Exp Med 53:471-492

Pooley HM, F-X Abellan and D Karamata (1992). CDP-Glycerol:Poly(glycerophosphate) Glycerophosphotransferase, which is involved in the synthesis of the major wall teichoic acid in *Bacillus subtilis* 168, is encoded by *tagF* (*rodC*). J of Bacteriology 174(2):646-649

Prikas HS, P Campbell, DS Feingold, I Jacobson, JW Jensen, U Lindahl, and L Roden. (1980) Biosynthesis of Heparin. Hydrogen Exchange at Carbon 5 of The Glycuronosyl residues. Biochemistry 19;495-500

Ray PH and CD Benedict (1982). CTP:CMP-3-deoxy-D-manno octulosonate cytidyl transferase (CMP-KDO synthetase). Methods Enzymol 83:535-540

Ray PH, JE Kelsey, EC Bigham, CD Benedict and T Miller (1983). Synthetis and use of 3-deoxy-D-manno-2-octulosonate (KDO) in *Escherichia coli*:potential sites of inhibition. In L Anderson and F Unger (eds), Bacterial lipopolisaccharides. American Chemical Society, Washington, D.C., p.141-149

Robbins JB, RL Myerowitz, JK Whisnant, M Argaman, R Schneerson, ZT Handzel and EC Gotschlich (1972). Enteric bacteria crossreactive with *Neisseria meningitidis* group A and group C and *Diplococcus pneumoniae* types I and III. Infection and Immunity 6:651-656

Robbins JB, EC Gotschlich, T-Y Liu, R Schneerson, ZT Handzel, M Argaman, JC Jr Parke and RL Myerowith (1973). Bacterial antigens cross-reactive with the capsular polysaccharides of *Haemophilus influenzae* type b, *Neisseria meningitidis* group A and C and *Diplococcus pneumoniae* types I and III. In:Robbins JB, Horton RE, Krause RM (eds) New approaches for including natural immunity to pyogenic organisms. Department of Health, Education, and Welfare, Washington, pp 45-56 (DHEW publication n° (NIH) 74-553)

Robbins JB, GH McCracken, EC Gotschlich, F Orskov, and LA Hanson. (1974) *Escherichia coli* K1 capsular polysaccharide associated with neonatal meningitis. The New England Journal of medicine 290:1216-1220

Robbins JB, R Schneerson, T-Y Liu, M-S Schiffer, G Schiffman, RL Myerowitz, GH McGracken *et al*. (1975). Cross-reacting bacterial antigens and immunity to disease caused by encapsulated bacteria. In:Neter E, Milgram F (eds). the immune system and infectious diseases. 4th Internatioanl Convocation on Immunology, Buffalo, NY Karger, New York, pp 218-241

Robbins JB, R Schneerson, WB Egan, W Vann, and DT Liu. (1980) Capsular Polysaccharides-Unanswered Questions. In The Molecular Basis of Microbial Pathogenicity. H Smith, JJ Skehel, and MJ Turner (eds) Weinhein:Verlag Ghemie Gmbh pp. 115-132

Robbins PW, D Bray, M Dankert and A Wright (1967). Direction of chain growth in polysaccharide synthesis. Work on a bacterial polysaccharide suggests that elongation can occur at the "reducing" end of growing chains. Science 158:1536-1542

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

Roberts JS, R Mounford, R Hodge, KB Jann, and GJ Boulnois. (1988a) Common Organization of Gene Clusters for Production of Different Capsular Polysaccharides (K Antigens) In *Escherichia coli*. J of Bacteriology 170(3):1305-1310

Roberts M, I Roberts, TK Korhonen, K Jann, D Bitter-Suermann, and PH Williams. (1988b)   DNA Probes for K-Antigen (Capsule) Typing of *Escherichia coli*. J Clin Microbiol 26(2):385-387

Rohr TE and FA Troy (1980). Structure and biosynthesis of surface polymers containing polysialic acid in *Escherichia coli*. J Biol Chem 255:2332-2342

Roosendaal B and FK deGraaf (1989). The nucleotide sequence of the *fanD* gene encoding the large outer membrane protein involved in the biosynthesis of K99 fimbriae. Nucleic Acids Res 117(3):1263

Robenberg M and D Court (1979). Regulatory sequences involved in the promoter and terminator of RNA transcription. Annu Rev Genet 13:319-353

Roychoudhury S, TB May, JF Gill, SK Singh, DS Feingold and AM Chakrabarty (1989). Purification and characterization of Guanosine diphospho-D-mannose dehydrogenase. J Biol Chem 264(16):9380-9385

Sanger F, S Nicklen and AR Coulson (1977). DNA sequencing with chain termination inihibitors. Proc Natl Acad Sci USA 74:5463-5467

Sato H, O Koiwai, K Tanabe and S Kashiwamata (1990). Isolation and sequencing of rat liver bilirubin UDP-Glucuronosyltransferase cDNA: possible alternate splicing of a common primary transcript. Biochem Biophys Res Common 169(1):260-264

Schmidt G, B Jann and K Jann, I Orskov and F Orskov. (1977) Genetic Determinants of the Synthesis Of the Polysaccharide Capsular Antigen K27 (a) of *Escherichia coli*. J Gen Microbiol 100:335-361

Schmidt MA and K Jann. (1982) Phospholipid substitution of capsular (K) polysaccahride antigens from *Escherichia coli* causing extraintestinal infections. FEMS Microbiology Letters 14:69-74

Schneerson R, M Bradshaw, JK Whisnant, RL Myerowitz, JE Parke and JB Robbins (1972). An *Escherichia coli* antigen cross-reactive with capsular polysaccharide of *Haemophilus influenzae* type b: occurence among known serotypes and immunochemical and biologic properties of *E.coli* antisera toward *H.influenzae* type b. J Immunol 108:1551-1562

Schnetz K, C Toloczyki and B Rak (1987). Glucoside (*bgl*) operon of *Escherichia coli* K-12:nucleotide sequence, genetic organization, and possible evolutionary relantionship to regulatory components of two *Bacillus subtilis* genes. J of Bacteriology 169(6):2579-2590

Schumperli D, K McKenney, DA Sobieski and M Rosenberg (1982). Translational coupling at an intercistronic boundary of the *Escherichia coli* galactose operon. Cell 30:865-871

Selzer G, T Som, T Itoh and J Tomizawa (1983). The origin of replication of plasmid p15A and comparative studies on the nucleotide sequences around the origin of related plasmids. Cell 32:119-129

Shine J and L Dalgarno (1974). The 3' terminal sequence of *Escherichia coli* 16S ribosomal RNA:complementary to non-sense triplets and ribosome binding sites. Proc Natl Acad Sci USA 71:1342-1346

Siebenlist U, RB Simpson and W Gilbert (1980). *E.coli* RNA polymerase interacts homologously with two different promoters. Cell 20:269-281

Silver RP, CW Finn, WF Vann, W Aaronson, R Schnnerson, PJ Kretschmer and CF Garon (1981). Molecular cloning of the K1 capsular polysaccharide genes of *E.coli*. Nature 289:696-698

Silver RP, WF Wann, and W Aaronson. (1984) Genetic and Molecular Analysis of *Escherichia coli* K1 Antigen Genes. J of Bacteriology 164(2):854-860

Silver RP, W Aaronson and WF Vann (1987). Translocation of capsular polysaccharides in pathogenic strains of *Escherichia coli* requires a 60-kilodalton periplasmic protein. J of Bacteriology 169(12):5489-5495

Silver RF, W Aaronson and WF Vann (1988). The K1 capsular polysaccharide of *Escherichia coli*. Rev Infect Dis 10:282-286

Smith AN, GJ Boulnois and IS Roberts. (1990) Molecular analysis of the *Escherichia coli* K5 *Kps* locus: identification and characterization of an inner-membrane capsular polysaccharide transport system. Molecular Microbiology 4(11):1863-1869

Steenbergen SM and ER Vimr. (1990) Mechanism of polysialic acid elongation in *Escherichia coli* K1. Molecular Microbiology 4(4):603-611

Steenbergen SM, TJ Wrona and ER Vimr (1992). Functional analysis of the sialyltransferase complexes in *Escherichia coli* K1 and K92. J of Bacteriology 174(4):1099-1108

Stevens P, LS Young and S Adamu (1983). Opsonization of various capsular (K) *E.coli* by the alternative complement pathway. Immunology 12:772-778

Sutherland IW (1977) Bacterial exopolysaccharides their nature and production. In: Sutherland IW (ed) Surface Carbohydrates of the Procaryotic Cell Academic Press Inc (London) Ltd. pp.27-96

Troy FA, IK Vijak, and N Tesche. (1975) Role of Undecaprenyl Phosphate in Synthesis of Polymers Containing Sialic Acid in *Escherichia coli*. J Biol Chem 250(1):156-163

Troy FA. (1979) The chemistry and biosynthesis of selected bacterial capsular polymers. Annu Rev Microbiol 33:519-560

Troy FA and MA McCloskey (1979). Role of a membranous sialyltransferase complex in the syntheis of surface polymers containing polysialic acid in *Escherichia coli*. temperature-induced alteration in the assembly process. J Biol Chem 254(15):7377-7387

Turk DC (1982). Clinical importance of *Haemiphilus influenzae*. In *Haemophilus influenzae*:epidemiology, immunology and prevention of diseases, SH Sell and PF Wright (eds.), New York:Elsevier Science Publishing Co.,pp.3-9

Vahlne G (1945). Serological typing of the colon bacteria with special reference to the occurrence of *E.coli* in man under normal and pathological conditions, particularly in appendicitis. Acta Pathol Microbiol Scand [Suppl] 62:1-127

van Oss CJ and CF Gillman (1973). Phagocytosis as a surface phenomenon: influence of C1423 on the contact angle and on the phagocytosis of sensitized encapsulated bacteria. Immun Commun 2:415

Vann WF, TY Lin, and JB Robbins (1979). Cell-free biosynthesis of the O-acetylated N-acetylneuraminic acid capsular polysaccharide of group C meningococci. J of Bacteriology 133:1300-1306

Vann WF, MA 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 desulfo heparin. Eur J Biochem 116:359-364

Vann WF, RP Silver, C Abeijon, K Chang, W Aaronson, A Sutton, CW Finn, W Linder and M Kotsatos. (1987). Purification, properties, and genetic location of *Escherichia coli* cytidine 5'-monophosphate N-acetyneuraminic acid synthetase. J Biol Chem 262(36):17556-17562

Vimr ER, RD McCoy, HF Vollger, NC Wilkison, and FA Troy (1984). Use of prokaryotic derived probes to identify poly (sialic acid) in neonatal neural membranes. Proc Natl Acad Sci USA 81:1971-1975

Vimr ER and FA Troy. (1985) Regulation of sialic acid metabolism in *Escherichia coli*: role of N-acylneuraminate pyruvate-lysate. J of Bacteriology 164(2):854-860

Vimr ER, BL Bassler and FA troy (1986). Abst Ann Meet Sm Soc Microbiol, pp154, K-152

Vimr ER, W Aaronson, and RP Silver. (1989) Genetic analysis of chromosomal mutations in polysialic acid gene cluster of *Escherichia coli* K1. J of Bacteriology 171(2):1106-1117

Vimr ER (1991). Map position and genome organization of the *kps* cluster for polysialic acid biosynthesis in *Escherichia coli* K1. J of Bacteriology 173:1335-1338

Vimr ER, R Bergstrom, SM Steenbergen, G Boulnois and I Roberts (1992). Homology among *Escherichia coli* K1 and K92 polysialyltransferases. J of Bacteriology 174(15):5127-5131

vonHeijne G (1983). Patterns of amino acids near the signal-sequence cleavage sites. Eur J Biochem 133:17-21

vonHeijne G (1985). Signal sequences. The limits of variation. J Mol Biol 184:99-105

Weisgerber C and FA Troy (1990). Biosynthesis of the polysialic acid capsule of *Escherichia coli* K1. The endogenous acceptor of polysialic acid is a membrane protein of 20 KDa. J Biol Chem 265(3):1578-1587

Whitfield C and FA Troy (1984). Biosynthesis and assembly of the polysialic acid capsule in *Escherichia coli* K1. Activation of sialyl polymer synthesis in inactive sialyltransferase complexes requires protein synthesis. J Biol Chem 259(20):12776-12780

Whitfield C, ER Vimr, JW Costerton and FA Troy (1984). Protein synthesis is required for *in vivo* activation of polysialic acid capsule synthesis in *Escherichia coli* K1. J of Bacteriology 159(1):321-328

Winkelstein JA, JA Bocchini Jr and G Schiffman (1976). The role of capsular polysaccharide in the activation of the alternative pathway by the *pneumococcus*. J Immunol 116:367-370

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

Yamada M, Y Yamadas and MH Saier, Jr (1990). Nucleotide sequence and expression of the *gutQ* gene within the glucitol operon of *Escherichia coli*. J DNA Sequencing and Mapping 1:141-145

Yanisch-Perron C, J Viera and J Messing (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of M13mp18 and pUC19 vectors. Gene 33:103-119