

Molecular studies of *Escherichia coli* capsule
gene clusters

C. Rachel Drake

Thesis presented for the degree of Doctor of Philosophy

University of Leicester

1991

UMI Number: U041739

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U041739

Published by ProQuest LLC 2015. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



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



7501075934

X75194582X

Statement

The accompanying thesis submitted for the degree of Doctor of Philosophy entitled 'Molecular Studies of *Escherichia coli* capsule gene clusters' is based on work conducted by the author in the Department of Microbiology of the University of Leicester mainly during the period between January 1988 and September 1991.

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

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

signed.....*R. Hall*.....date.....10.10.91.....

Abstract

Molecular studies of *Escherichia coli* capsule gene clusters

C. Rachel Drake

Escherichia coli can produce a large number (over 70) of structurally distinct capsular polysaccharide (K antigens) which have previously been divided into two groups. Group II K antigens are encoded by *kps* near *serA*. The *kps* genes are homologous between group II capsule gene clusters which encode chemically distinct K antigens. Genes encoding the K4 antigen (group II), an unusual substituted polymer (a fructosylated chondroitin), were cloned and expressed in *E. coli* K-12 and shown to contain the group II *kps* determinants. By nucleotide sequence analysis it was shown that the 3' end of one group II capsule gene, *kpsS*, is different in the K1, K4 and K5 capsule gene clusters. It was demonstrated that strains expressing a group I capsule do not carry the group II *kps* determinants on the chromosome. The K3, K10, K11 and K54 capsules have characteristics of both groups I and II and were also shown to lack the group II capsule genes despite the K10 and K54 capsule genes having previously been mapped near *serA*. The existence of a third capsule group, group I/II, encoded by genes distinct from the group II capsule genes yet located in the same region of the chromosome is formally proposed. An unsuccessful attempt was made to clone the K9 antigen (group I) biosynthesis genes. DNA flanking the K1 and K4 antigen gene clusters was used as probes in Southern blot analysis of different *E. coli* isolates and was shown to be a common component of the *E. coli* chromosome. The K1, K4 and K5 capsule gene clusters appear to be located at the same chromosomal location. DNA adjacent to one end of the K4 capsule gene cluster was highly polymorphic and present in more than one copy in some strains. The organisation and variability of the *serA* region of the *E. coli* chromosome is discussed.

Acknowledgements

I sincerely thank Prof Graham Boulnois and Dr Ian Roberts for their supervision of this work and their meticulous reading of the manuscript. I also thank Prof Klaus and Dr Barbara Jann and the members of their group for providing antibodies and strains and for their interest and advice on my work. I am also grateful to Drs F. and I. Ørskov for the provision and serotyping of strains. I thank the other members of the capsule group, Annabel Smith, Carlo Pazzani, Sumita Ganguli and Rowan Pearce for their direct contribution to my work and for the all the colourful discussions we have had over the past few years. Thanks also go to the other people in my laboratory and those nearby for their help, particularly Drs Mark Coleman and Tim Wallis (and families) who have given me tremendous encouragement. I thank Dr Bill Grant and the members of his group for being so hospitable in providing me with bench and desk space. I thank Dr Tahir Yaqoob for leading me along the uphill struggle with the text handler. My thanks go to Mum, Dad, Sally and Lucy for their generosity over the years, for their encouragement and for their help with the typing of the introduction. Finally, my special thanks go to Terry McGenity for his patience in getting me through the difficult patches.

Contents

1 Introduction

1.1	Bacterial exopolysaccharide: general features	
1.2	The polysaccharide antigens of <i>E. coli</i>	3
1.3	The K antigens of <i>E. coli</i>	6
1.3.1	Function of the K antigen	6
1.3.2	<i>E. coli</i> capsules can be divided into at least two groups	7
1.3.3	Group II capsule expression	13
1.3.4	Group I capsule expression	25
1.4	The genetic diversity of <i>E. coli</i>	26
1.4.1	The variability of O:K serotype within the <i>E. coli</i> population	28
1.4.2	Generation of the group II K antigen gene cluster	29
1.5	The aims of this thesis	31

2 Materials and Methods 32

2.1	Bacterial Strains and Plasmids	32
2.1.1	Growth Conditions and Media	32
2.2	Transformation of Bacterial Cells	32
2.3	Procedures for DNA Extraction	34

2.3.1	Extraction of Chromosomal DNA	35
2.4	Techniques used in routine DNA manipulation	37
2.5	Cosmid Cloning	37
2.6	Transfer of DNA to Nylon Filters by Southern Blotting	39
2.7	Preparation of filters for colony hybridisation	39
2.8	DNA Hybridisation Procedures	40
2.9	DNA Sequencing	41
2.10	Polymerase Chain Reaction	42
2.11	Techniques used for Polysaccharide detection	43
2.12	Production of polyclonal antiserum against capsular polysaccharide	45
2.13	Haemagglutination assay	46
2.14	Detection of haemolytic activity	46
3	Cloning and Expression of the K4 antigen gene cluster	47
3.1	Introduction	47
3.2	Results	48
3.2.1	Southern Blot analysis of K4 chromosomal DNA	48
3.2.2	Isolation of the K4 antigen gene cluster	48
3.2.3	Analysis of the cloned K4 antigen gene cluster	52
3.3	Discussion	54
4	The variability of group II capsule gene clusters	58
4.1	Introduction	58
4.2	Results	61
4.2.1	PCR analysis of the region 1 - 2 and region 2 - 3 boundaries	61

4.2.2	Sequence analysis across the region 1 - 2 boundary of the K1 and K4 capsule gene clusters	62
4.2.3	KpsS has a variable C terminus	63
4.2.4	Sequence analysis beyond <i>kpsS</i> _{K1}	64
4.2.5	Sequence analysis beyond <i>kpsS</i> _{K4} and comparison with the K5 capsule genes	65
4.2.6	Comparison of the region 1 - 2 boundaries in the K1, K4 and K5 capsule gene clusters	65
4.3	Discussion	66
5	Southern blot analysis of the group II capsule gene locus and its flanking DNA	77
5.1	Introduction	77
5.2	Results	78
5.2.1	Analysis of group II capsule genes in strains not expressing a group II capsule	78
5.2.2	Analysis of sequences flanking Region 3 of the <i>kps</i> genes	81
5.2.3	Analysis of sequences adjacent to region 1 of the <i>kps</i> genes . . .	86
5.2.4	Analysis of K9 chromosomal DNA	90
5.2.5	Sequences flanking the K4 antigen gene cluster are not P-fimbrial adhesin genes	91
5.3	Discussion	94
6	Attempted cloning of the K9 biosynthesis genes	105
6.1	Introduction	105
6.2	Results	105
6.2.1	Production of a polyclonal antiserum against the K9 antigen .	106
6.2.2	Development of a K9 antigen detection system	106

6.2.3	Production of an <i>E. coli</i> 2667 cosmid library	107
6.2.4	Screening the cosmid library for K9 antigen expression	107
6.3	Discussion	108
7	Discussion	111
8	References	117

List of Abbreviations

BSA	bovine serum albumin
bp	basepair
CKS	CMP-KDO synthetase
EDTA	ethylenediamine-tetra acetic acid
EPS	exopolysaccharide
IPTG	isopropylthiogalactoside
kb	kilobase
kD	kilodalton
KDO	2-keto-3-deoxymanno-octonic acid
LPS	lipopolysaccharide
NANA	N-acetyl neuraminic acid
mins	minutes
MLEE	multilocus enzyme electrophoresis
MRHA	mannose resistant haemagglutination
ORF	open reading frame
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
RFLP	restriction fragment length polymorphism
SDS	sodium lauryl sulphate
SSC	saline sodium citrate (150mM NaCl, 15mM sodium citrate)
TEMED	N,N,N',N'-tetramethylethylenediamine
TE	Tris-EDTA (10mM Tris-HCl, 1mM EDTA pH8.0)
UTI	urinary tract infection
X-gal	5-bromo-chloro-3-indolyl-D-galactoside

Chapter 1

Introduction

The production of extracellular polysaccharide (exopolysaccharide) is an almost universal feature of bacteria. A wide variety of bacteria, both Gram positive and Gram negative from a wide variety of habitats, both clinical and natural, produce exopolysaccharide (EPS) and *Escherichia coli* is no exception. Not only is the number of bacterial species which produce EPS vast but also the array of chemically different polysaccharides displayed is enormous. The ability of some bacterial species to produce many polysaccharides presents an interesting genetic problem.

1.1 Bacterial exopolysaccharide: general features

Polysaccharides are linear macromolecules composed of repeating mono- or oligosaccharides which may be substituted with short side chains. The exact number of repeats in any one molecule is variable. Monosaccharides represented include neutral sugars, polyols, uronic acids and amino sugars. In addition, non-sugar substitutions including phosphate, formate, succinate, pyruvate and acetyl groups are common although sulphation has not been observed (Dudman, 1977; Sutherland, 1972, 1977). These macromolecules can exhibit greater diversity than any other by varying not only the components but also the linkage. Consequently, a wide range of structurally different bacterial polysaccharides have been identified (Table 1.1) (Kenne and Lindberg, 1983). Both homopolymers such as *E. coli* K1 and branched heteropolymers such as *E. coli* K30 (Table 1.1) which has a disaccharide backbone with a disaccharide side chain, can be found. Several cases are known in which different organisms synthesise chemically identical polysaccharides, for example *E. coli* K1 and *Neisseria meningitidis* group B capsular polysaccharides are chemically identical polymers of N-acetyl neuraminic acid (NANA) which is a sugar rarely found in bacteria (Table 1.1) (Kaspar *et al.*, 1973). Some species, including *E. coli*, *Klebsiella* spp and *Streptococcus pneumoniae*, synthesise a wide array of polysaccharides with chemically different repeat oligosaccharides.

Polysaccharide which exists beyond the surface membrane of a bacterial cell is termed exopolysaccharide (EPS). Polysaccharide itself comprises only a small proportion of the EPS bulk, upwards of 95% is reported to be water (Sutherland, 1972). EPS can be described as either slime or capsular polysaccharide although the distinction between the two is somewhat arbitrary and the terminology surrounding the subject confusing (Ørskov and Ørskov, 1990). EPS can exist as an amorphous slime layer which may encompass not only the producing cell, but also those nearby to create a microenvironment in which several cells exist. Slime polysaccharide is released into the growth medium and is easily extracted by washing. The alginate produced by *Pseudomonas*, the dextrans of *Agrobacterium* and xanthan gum synthesised by *Xanthomonas campestris* are all examples of this type of polysaccharide, however, the slime polysaccharide of *Erwinia amylovora* is often referred to as capsule (Sutherland, 1985; Whitfield, 1988). Alternatively, EPS may form a complete capsule, effectively a hydrated gel, around each individual cell. By virtue of its attachment to the cell surface, capsular polysaccharide requires a more vigorous extraction procedure than slime. Encapsulated bacteria do, however, release polysaccharide into the growth medium (Hancock and Cox, 1991). For instance, 20% of the capsular polysaccharide (K antigen) molecules of *Klebsiella aerogenes* are easily extracted by washing (Troy *et al.*, 1971). Identification of capsular polysaccharide as a precipitin halo surrounding a colony grown on serum agar is clear evidence that some molecules are not attached (Kaijser, 1977; Petrie, 1932). Capsular polysaccharides are the subject of this thesis and EPS which is clearly slime has largely been ignored.

The study of encapsulated bacteria has concentrated on those of medical importance including *S. pneumoniae*, *Haemophilus influenzae*, *Klebsiella* spp, *N. meningitidis* and *E. coli*. Bacterial capsules were originally identified as an exclusion zone around the cell visible by negative staining with Indian ink (Ørskov and Ørskov, 1990). Microscopic study of capsules is widely used but interpretation is complicated by the distorting effects of staining and fixing on the gelatinous structure (Bayer, 1990). Capsule thickness varies considerably (0.1-10 μ m) (Sutherland, 1977), some easily visible using Indian ink whilst others form a thin outer covering round the cell which is only visible using the electron microscope. It is not entirely clear how the capsule is maintained on the cell surface. Capsular polysaccharide is reported to be linked to peptidoglycan in a number of Gram positive bacteria (Sorensen *et al.*, 1990; Yeung and Mattingly, 1986). In Gram negative organisms it appears that capsular polysaccharide is attached in some cases to phospholipid presumed to reside in the outer membrane (Gotschlich *et al.* 1981; Kuo *et al.*, 1985; Schmidt and Jann, 1982) and in others to core-lipid A (Jann and Jann, 1990). Unattached molecules could be maintained in the capsule by ionic interaction with attached molecules (Jann and Jann, 1990).

The majority of bacteria isolated from natural habits including soil and water produce EPS (Costerton *et al.*, 1981). To account for this many theories have been put forward as to the selective advantage this confers, with varying degrees of experimental verification. These relate to both slime and capsular polysaccharide. EPS often constitutes the outermost layer of the cell exposed to the environment, therefore its properties affect the interactions of the bacterium. The physical nature of

EPS, regardless of exact composition, affects dehydration, ionic interaction, adhesion and colonisation of inert surfaces, infection by bacteriophage and virulence of the bacterium (Costerton *et al.*, 1981). The hydrophilic surface of the capsule may control the immediate environment of the cell by binding cations (both toxic and useful) including heavy metals (Dudman, 1977). Water is retained at the cell surface by the hydrated polysaccharide and this may delay desiccation and prolong bacterial survival in adverse conditions. The EPS produced by some bacteria can function as an adhesion factor, important in the colonisation of both inert and biological surfaces by the bacterium. This has practical consequences ranging from dental caries to the fouling of pipes (Costerton *et al.*, 1987). In aquatic environments it has been shown that adhesion to inert surfaces by *Acetobacter aceti* and others depends on the presence of a discrete capsule (Fletcher and Floodgate, 1983; Hermesse *et al.*, 1988). Evidence suggests that bacterial EPS may prevent predation by phagocytic protozoa and slime moulds (Dudman, 1977). Viral infection is an important cause of bacterial mortality and EPS can prevent adsorption and penetration of some bacteriophage (Costerton *et al.*, 1981) although for other bacteriophage EPS serves as the surface receptor (for example see Whitfield and Lam, 1986). In some cases specific polysaccharides are important in bacterial interactions, the most studied of which is *Rhizobium* spp in which EPS plays a key role in determining the specificity of *Rhizobium*-legume nodulation (Djordjevic *et al.*, 1987; Long, 1989).

Bacterial encapsulation has been recognised as a virulence determinant in bacteria pathogenic for either plants or animals. For instance, it has been postulated that the blockage of xylem vessels by the EPS of *E. amylovora* is important in the development of fireblight disease of apples and pears (Bennett and Billing, 1978). Encapsulated bacteria are responsible for some of the most serious invasive infections to which man is susceptible. In the classical experiment performed by Avery and Dubos (1931) capsular polysaccharide was enzymatically removed from *S. pneumoniae* and the lethal inoculum rose by a factor of 10^6 as a consequence; this clearly implicates the bacterial capsule as a virulence factor and many subsequent experiments have implied the same. Evidence suggests that encapsulation can confer on the bacterium a certain degree of resistance to phagocytosis and the bactericidal effects of serum (Cross, 1990; Horwitz and Silverstein, 1980; Joiner, 1988; Moxon and Kroll, 1990). Some capsules are more effective than others in conferring resistance to host defences and may be more important virulence factors (Kaijser *et al.*, 1977; Ørskov *et al.*, 1977). The multifactorial nature of bacterial virulence must not be forgotten as capsules are one of several potential virulence factors. Although most pathogenic bacteria are encapsulated, the majority of encapsulated bacteria are not pathogenic. Bacterial strains expressing certain capsule types are more often associated with infection than others (Joiner, 1988; Opal *et al.*, 1982; Stevens *et al.*, 1983).

Encapsulation is widespread, a feature of a large number of species yet many biological questions remain unanswered. Fundamental processes such as polysaccharide biosynthesis and transport across cell membranes are not clearly understood. In both medical and industrial spheres polysaccharides are important. Polysaccharides serve as lubricants and gelling agents; several bacterial polysaccharides are known to be useful and others could prove to be so. *E. coli* K5 polysaccharide is identi-

cal to an intermediate in heparin synthesis (Vann *et al.*, 1981) and could be used in the commercial production of heparin. Capsular antigens are vaccine candidates in protecting against infection caused by encapsulated organisms (Jennings, 1990). A pneumococcal vaccine is available including 23 of the known 83 capsular types (Robbins *et al.*, 1983). *E. coli* capsules are perhaps the best studied and offer the opportunity to explore all these avenues of interest.

1.2 The polysaccharide antigens of *E. coli*

E. coli can produce K and O antigens, both of which are polysaccharides attached to the cell surface. In general the capsular K antigen is acidic whilst the somatic O antigen is neutral (Ørskov *et al.*, 1977). In the 1940's *E. coli* capsular polysaccharides were formally acknowledged to be separate from the O antigen on the basis of different serological behaviour. The presence of a capsular antigen renders the bacteria inagglutinable in the homologous O antiserum (Ørskov *et al.*, 1977). This is because the capsule covers other surface components including lipopolysaccharide (LPS) and outer membrane protein; pili, fimbriae and flagella extend through the capsule. In *E. coli* K29 the capsule extends $1\mu\text{m}$ beyond the bacterium and the O antigen extends only about $0.02\mu\text{m}$ (Bayer, 1990). The term K antigen (for the German word Kapsel) was coined and the serological classification of the organism was expanded from the original O:H typing (LPS:flagella) to include the K antigen (Kauffman, 1966). Within the *E. coli* population many different O:K:H serotypes can be identified. Over 70 K antigens which differ in the structure of the repeat oligosaccharide have been described but a given strain expresses only one of these, switching between types does not occur (Ørskov *et al.*, 1977). Before discussing the K antigen in detail it is worthwhile considering briefly other features of the *E. coli* cell surface.

LPS is an integral part of the outer membrane, a common feature of the Enterobacteriaceae, composed of a lipid anchor (lipid A) and a polysaccharide moiety (Figure 1.1). The polysaccharide moiety has two components: the core oligosaccharide, the structure of which is highly conserved both within *E. coli* and between *E. coli* and *Salmonella typhimurium* spp, and the variable O antigen. Lipid A substituted with the inner core sugars is an essential component of the outer membrane (Luderitz *et al.*, 1973; Raeyz, 1988). The variable O antigen like the K antigen is composed of repeated sugar units and *E. coli* can express one of around 160 chemically different O antigens (Ørskov *et al.*, 1977). O antigen biosynthesis genes (*rfb*) are located near *his* at 44 mins on the *E. coli* chromosome (Bachmann, 1990). Various LPS mutants which are defective in sequential steps of LPS biosynthesis have been described (Rick, 1987). The wild type O antigen-carrying LPS is described as smooth. Rough mutants completely lack O antigen and semi-rough mutants produce core-lipid-A attached to only one O antigen repeat unit. *E. coli* K-12 is a rough *rfb* mutant (Schmidt, 1973).

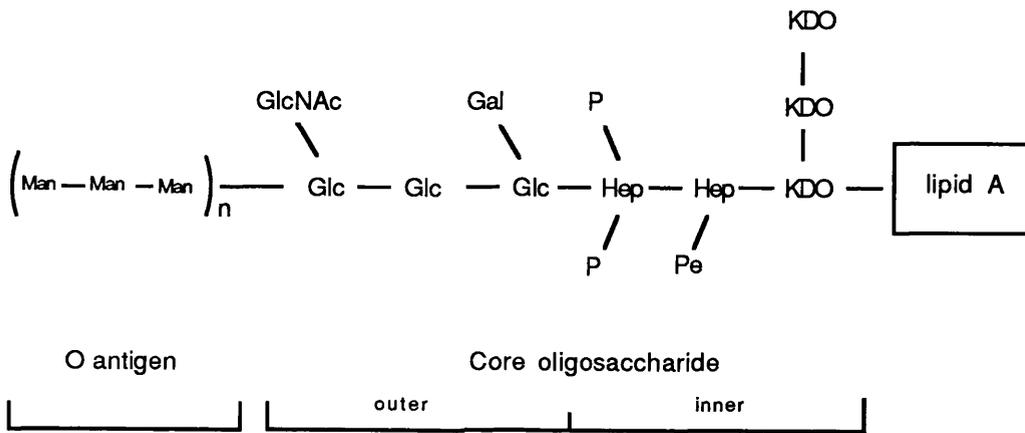
E. coli also has the ability to produce slime polysaccharides. This includes colanic acid (M antigen) and enterobacterial common antigen (ECA), neither of which are

Table 1.1: Repeating units of some *E. coli* K antigens

Antigen	Structure	Reference
K1	-8)- α -NANA-2	McGuire and Binckley (1964)
K2	4)- α -Gal-(1,2)-Gro-(3-P-	Jann <i>et al.</i> (1980)
K4	4)- β -GlcA-(1,3)- β -GalNAc-(1- \swarrow 3,2 \searrow β -Fru	Rodriguez <i>et al.</i> (1988)
K5	4)- β -GlcA-(1,4)- α -GlcNAc-(1-	Vann <i>et al.</i> (1981)
K3	2)- α -Rha-(1,3)- α -Rha-(1,3)- α -Rha-(1- \swarrow KHO	Dengler <i>et al.</i> (1988)
K10	3)- α -Rha-(1,3)- β -QuiNMal-(1-	Sieberth, V. and Jann, B. (unpublished)
K11	4)- β -Glc-(1,4)- α -Glc-(1-P- \swarrow 3,2 \searrow β -Fru	Rodriguez <i>et al.</i> (1989)
K19	3)- β -Rib-(1,4)- β -KDO-(2,8-OAc	Jann <i>et al.</i> (1988)
K54	3)- β -GlcA-(1,3)- α -Rha-(1- CO.NH threonine (serine)	Hofmann <i>et al.</i> (1985)
K9	3)- β -Gal-(1,3)- β -GalNAc-(1,4)- α -Gal-(1,4)- α -NeuNAc-(2,5-OAc	Dutton <i>et al.</i> (1987)
K29	2)-Man-(1,3)-Glc-(1,3)- β -GlcA-(1,3)- β -Gal-(1- \swarrow 4,1 Pyr ⁴ β -Glc-(1,2)- α -Man	Choy <i>et al.</i> (1975)
K30	20-Man-(1,3)-Gal-(1- \swarrow 3,1 \searrow β -GlcA-(1,3)-Gal	Chakraborty <i>et al.</i> (1980)

Abbreviations: Fru, fructose; Gal, galactose; Glc, glucose; GlcA, glucuronic acid; GlcNAC, N-acetyl glucuronic acid; Gro, glycerol; KHO, deoxy-hexulosonic acid; Man, mannose; NANA, N-acetyl neuraminic acid; Pyr, pyruvate; Rib, ribose; Rha, rhamnose; QuiNMal, 4-6-dideoxy-4-maolnylamidoglucose. The K antigens have been tabulated by group (from the top): group II, group I/II and group I.

Figure 1.1: Structure of *E. coli* LPS (Rick, 1987). Abbreviations: Gal, galactose; Glc, glucose; GlcNAc, N-acetyl glucuronic acid; Hep, heptose; KDO, keto-deoxy octulosonic acid; Man, mannose; P, phosphate; Pe, phosphorylethanolamine. The O antigen repeat unit shown is that of O8.



unique to *E. coli* but synthesised by many Enterobacteriaceae. Colanic acid is composed of glucose, galactose, fucose and glucuronic acid in a polysaccharide whose molecular weight is reported to be between 100 000 and 200 000 (Markovitz, 1977). It has been studied in *E. coli* K-12 where expression increases under conditions of environmental stress, particularly at lower growth temperatures (20°C) when the normally non-mucoid colonies give way to characteristically mucoid ones. The function of colanic acid has not been clearly demonstrated but it has been suggested that colanic acid, like most bacterial EPS has a protective role important in the survival of *E. coli* in adverse environmental conditions. Colanic acid is sometimes referred to a capsular polysaccharide but it is distinct from the capsular K antigens discussed in this thesis. Colanic acid biosynthesis is determined at the *cps* locus near *his* (Trisler and Gottesman, 1984).

1.3 The K antigens of *E. coli*

1.3.1 Function of the K antigen

E. coli is usually a member of the normal mammalian gut flora and is presumably adapted to this environment. Like other resident bacteria, including for example *Bacteroides*, *E. coli* is an opportunistic pathogen. The diverse gut flora of *E. coli* is not reflected in the range of serotypes of strains commonly causing disease. *E. coli* K1, K2, K3, K5, K13, K14 and K92 antigen expressing strains are more commonly associated with urinary tract infections (UTI) and pyelonephritis than others (Kaijser *et al.*, 1977; Ørskov *et al.*, 1982; Ørskov *et al.*, 1985). Similarly, *Klebsiella* spp are also opportunistic pathogens and can express over 80 chemically different K antigens, only some of which are commonly expressed by strains isolated from the site of infection (Ørskov and Ørskov, 1984). This does suggest that the capsule plays a role as a virulence factor, although the contribution made by other factors including adhesins and haemolysin are also important in determining the outcome of an infection (Johnson, 1991).

The K1 and K5 antigens have interesting properties. The K1 antigen (poly-NANA), like the *N. meningitidis* group B polysaccharide, is identical in structure to part of the oligosaccharide moiety of the neonatal neural cell adhesion molecule (N-CAM) and is therefore non-immunogenic (Finne *et al.*, 1983). This lack of immunogenicity may be responsible for the observation that 84% of *E. coli* strains causing neonatal meningitis in the U.S. expressed the K1 antigen (Robbins *et al.*, 1974). In addition, the K1 and group B polysaccharides are poor activators of the alternative complement pathway (Jarvis and Vedros, 1987; Plusche *et al.*, 1983) and are recognised as important virulence factors (Silver and Vimr, 1990). The K5 antigen is identical to an intermediate in heparin synthesis and is poorly immunogenic. K5 capsule expressing strains are frequently isolated from the blood and urinary tract (Kaijser and Jodal, 1984).

Whilst considerable interest has centred around clinical isolates and the role of the K antigen as a virulence factor, the role of the K antigen in its natural environment, the gut, is not clear. Polysaccharide could be important in colonisation or protection of the bacterium in the gut. Since any single *E. coli* strain is only a transient member of the gut flora of any one individual (complete turnover occurs approximately every two weeks) perhaps the capsule is important to the survival of the bacterium between excretion in the faeces and re-colonisation. Indeed, some capsules (Section 1.3.2.1) are expressed at low growth temperatures (18°C) not experienced in the intestine. However, other capsules are only expressed at 37°C and this would suggest that their role is not in survival outside the host. Colanic acid production may serve to engulf the cell in EPS when the bacterium is not in the mammalian gut.

1.3.2 *E. coli* capsules can be divided into at least two groups

Originally, K antigens were subdivided into A, B or L antigens (Ørskov *et al.*, 1977). Subsequently this classification was abandoned in favour of two groups of *E. coli* capsules, groups I and II (Jann and Jann, 1987, 1990). Most group II K antigens were formerly L antigens. The distinctions between group I and II capsular polysaccharides are summarised in Table 1.2. The main criteria for the division of K antigens into groups was their heat stability, electrophoretic mobility and composition. A number of other distinctions have also been identified such as temperature expression and genetic determination. The two groups also differ in their similarity, both biochemical and genetical, to capsular polysaccharides of other organisms (Jann and Jann, 1990). Those of group I resemble the K antigens of *Klebsiella* spp (Jann and Jann, 1983, 1987; Kenne and Lindberg, 1983). Those of group II are like those of *N. meningitidis* and *H. influenzae* (Jann and Jann, 1983, 1987). The K antigens associated with clinical isolates usually belong to group II and as a result group II capsules have been the subject of more intense study. Recently the notion of a third capsule group, group I/II was introduced (Finke *et al.*, 1990) and this will be discussed once groups I and II have been outlined.

1.3.2.1 Biochemical and structural distinctions between group I and II capsules

The majority of the oligosaccharide repeating units of group II K antigens are linear, comprising one or two different monomers, whilst those of group I tend to be more complex and may be substituted with single sugars or short side chains (Table 1.1). One distinction between capsule types is the relative heat stability of the capsular polysaccharide. Group II K antigens are heat-labile, at 100°C group II capsular polysaccharide is unstable and is released from the cell whilst those of group I remain intact. This is because the linkage between the group II polysaccharide and its anchor in the outer membrane is labile (Jann and Jann, 1983).

Table 1.2: A summary of the distinctions between the two groups of *E. coli* capsules (Jann and Jann, 1990)

Property	Group I	Group II
state at 100°C	stable	labile
electrophoretic mobility	low	high
acidic component	GlcA GalA pyruvate	GlcA NANA KDO ManNAcA
expressed below 20°C	yes	no
CKS activity at 37°C	low	high
co-expressed with	O8, O9, O20, O101	many O antigens
membrane anchor §	core-lipid A	KDO-phospholipid
genetic determination located near	<i>kst</i> <i>his (trp?)</i>	<i>kps</i> <i>serA</i>
intergenic relationship	<i>Klebsiella</i>	<i>H. influenzae</i> <i>N. meningitidis</i>

§The substitutions at the reducing end of neither groups I nor II polysaccharides have been verified in all cases. The molecules at the reducing end of the polymer are generally believed to form the membrane anchor of the polysaccharide.

The sugar components commonly found in group I and group II polymers differ (Table 1.1). Firstly, the majority of group I capsular polysaccharide lack amino sugars whilst all group II polysaccharides contain at least one (Jann and Jann, 1990). Secondly, although all K antigens are acidic, the acidic component varies between the two groups. Group II polysaccharides include a range of acidic components: glucuronic acid, NANA, KDO (keto-deoxy octulosonic acid) and N-acetylmannosaminuronic acid. Whereas group I polysaccharides are generally restricted to hexuronic acids (glucuronic acid, galacturonic acid) and pyruvate substitutions. In addition, phosphate substitution of the polysaccharide is more common in group II than group I K antigens and this confers additional negative charge on the molecule. As a result of these chemical differences, group II polysaccharides tend to be more acidic and have a higher charge density than those of group I.

Group I polymers are reported to have a higher molecular weight than those of group II, being greater than 100 000 and less than 50 000 respectively although exceptions to this rule can be found. Determination of polysaccharide molecular weight is not a reliable indication of capsule group. Corresponding difference in viscosity of aqueous solutions of group I and II polysaccharides has been noticed (Jann and Jann, 1983). On account of their high charge density and low molecular weight, group II polysaccharides have a greater electrophoretic mobility than group I enabling easy distinction of the two capsule groups. The capsules of group I appear thick and copious around the cell when viewed using the electron microscope whilst those of group II tend to be thin (Jann and Jann, 1983). This capsule morphology is reflected in the capsules of other species which those of *E. coli* resemble: *Klebsiella* spp capsules are particularly thick (Jann and Jann, 1977) in contrast to those of *H. influenzae* and *N. meningitidis*.

Group I K antigens are expressed in conjunction with only a few O antigens, namely O8 and O9 and occasionally O20 and O101 (Ørskov *et al.*, 1977). Group II K antigens can be co-expressed with all other O antigens but not with O8, O9, O20 or O101 (Ørskov *et al.*, 1977). A single group II polysaccharide is, however, restricted to a limited number of O serotypes by virtue of the clonal nature of the *E. coli* population (see Section 1.4), for instance the K1 antigen is expressed only with O6, O18 and O100 (Achtman *et al.*, 1983). Interestingly, O antigens can also be divided into two types. The four O antigens associated with group I capsules are 'rfe-dependent' whilst other O antigens are 'rfc-dependent'. This division is based solely on the mechanism of synthesis of the polymer which makes use of either rfe or rfc encoded polymerases (see Section 1.3.3.1) (Jann and Jann, 1987; Mäkelä and Stocker, 1984).

Another important distinction between the two groups of *E. coli* capsular polysaccharide is that group II capsule expression is regulated by temperature, capsular polysaccharide is produced at 37°C but not at 18°C (Ørskov *et al.*, 1984). In contrast group I capsules are expressed at all growth temperatures (Ørskov *et al.*, 1984). Interestingly, the expression of some other *E. coli* virulence factors such as P-fimbriae and haemolysin are similarly regulated by temperature. It has been suggested that co-ordinate regulation of virulence factors in response to environmental stimuli may occur (Dorman, 1991; Higgins *et al.*, 1990a).

Study of group II K antigens has revealed that some polymers which do not contain KDO in the repeat oligosaccharide have KDO at their reducing termini. This has been shown for K2, K4, K5, K54 and K62 (Schmidt and Jann, 1982; Finke *et al.*, 1989). In group II polysaccharides which do contain KDO in the repeat oligosaccharide, KDO is the end sugar at the reducing terminus. Recently it was shown that group II capsule expressing *E. coli* have CMP-KDO synthetase (CKS) activities some four or five times higher than group I and unencapsulated strains (Finke *et al.*, 1990). In addition, this elevated CKS activity was temperature regulated; at the capsule restrictive temperature CKS activities were as low as the group I capsule expressing strains. A high CKS activity at 37°C is therefore believed to be a feature of a group II K antigen expressing strain.

The way in which the capsular polysaccharide is attached to the cell surface is not entirely clear. Group II polysaccharides appear to be linked by a labile phosphodiester bridge to phosphatidic acid (Jann and Jann, 1990; Schmidt and Jann, 1982). The phospholipid has been postulated to serve as an anchor for the polysaccharide in the outer membrane. This has not been shown for all group II capsular polysaccharides but it is generally assumed to be so. There is evidence to suggest that the capsular polysaccharides of *N. meningitidis* and *H. influenzae* which resemble *E. coli* group II capsules are also linked to phospholipid (Gotschlich *et al.*, 1981; Kuo *et al.*, 1985). Attachment of group I K antigens to the cell surface is less clear. The group I capsular polysaccharides that have been studied, namely K9, K30 and K40, terminate in core-lipid A at the reducing end (Jann and Jann, 1990; Ørskov *et al.*, 1977). SDS polyacrylamide gel electrophoresis (PAGE) of K40 polysaccharide reveals a ladderlike pattern exactly as LPS would appear and low molecular weight fractions are linked to core-lipid A (Jann and Jann, 1990). Likewise, the K30 antigen exists as a low and high molecular weight species (Jann and Jann, 1987). The low molecular weight fraction contains core-lipid A with one sugar repeat unit attached and this is the configuration of LPS produced by a semi-rough mutant (Homonylo *et al.*, 1988). Core-lipid A has not been found attached to the high molecular weight fraction (Homonylo *et al.*, 1988; Whitfield *et al.*, 1989). In both capsule groups it has been reported that only 50% of molecules are attached to the putative anchor molecule, namely phospholipid or core-lipid A (Jann and Jann, 1990), though whether this reflects the state of the native polymer or the rigors of the extraction procedure is not clear.

It is apparent that group I K antigens and LPS are structurally similar consisting of core lipid A attached to a repeating oligosaccharide (K or O antigen). Cells expressing a group I capsule could be regarded as having two O antigens, one acidic and one neutral where the latter is probably O8 or O9. The acidic capsular K antigen covers the O antigen and these strains are characteristically inagglutinable in the homologous O antiserum (Ørskov *et al.*, 1977). Some polysaccharides have been isolated as a K antigen from some strains and as an acidic O-specific polysaccharide moiety of LPS from another strain. For instance, K87 when co-expressed with O8 is considered a K antigen but when expressed as an acidic cell wall antigen without O8 it is considered to be the O32 antigen (Jann and Jann, 1990). Other examples of this dual identity are K85 and O141, and K9 and O104 (Jann and Jann, 1990; Jann *et al.*, 1971; Ørskov *et al.*, 1977).

1.3.2.2 Genetic distinctions between group I and II K antigens

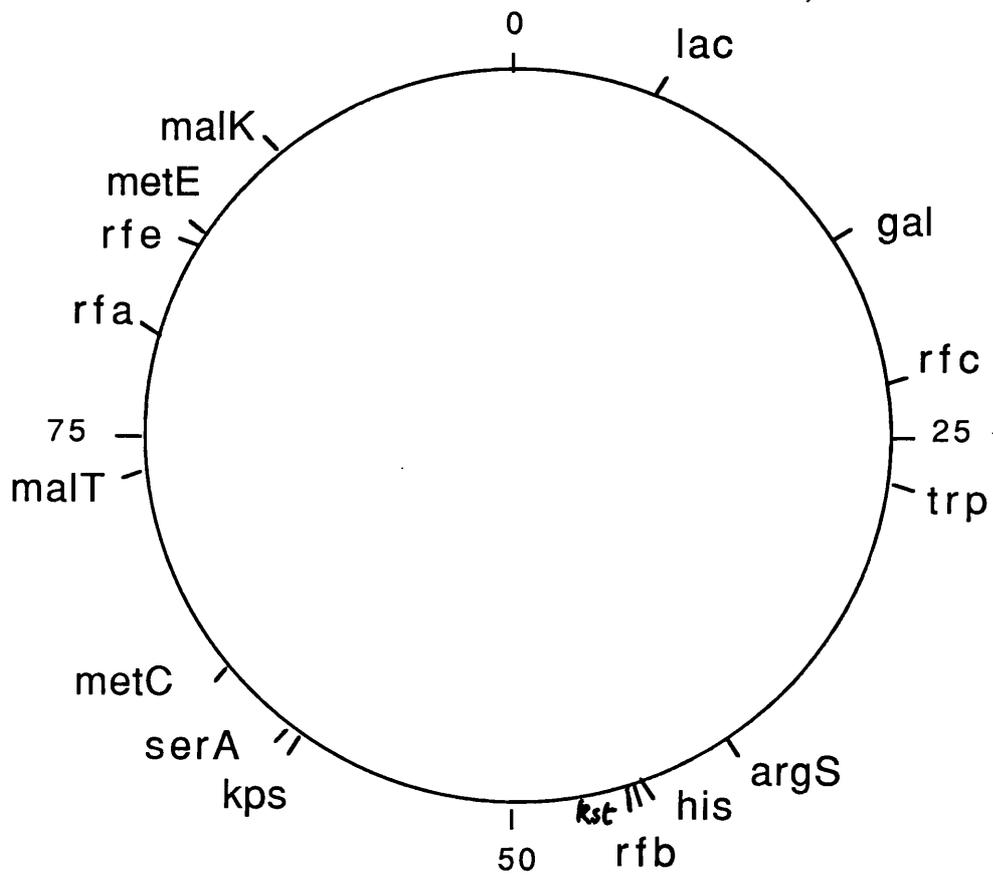
In the light of the biochemical and structural differences outlined above, it is not surprising that groups I and II capsules are encoded by different genes. Genes for the production of groups I and II antigens are not allelic. Expression of a group I, II or I/II capsule is mutually exclusive meaning that a single isolate produces only one of them despite the fact that group I capsule genes are located at a different site on the chromosome. In laboratory crosses using a K27 or K57 (group I) capsule expressing strain as the recipient and a K8 (group II) capsule expressing strains as the donor transconjugants expressing both capsular types were isolated (Ørskov *et al.*, 1976). Genetic determinants for the complete expression of K1 and K4 (group II) capsules have been mapped by classical genetic analysis to the *serA* region of the *E. coli* chromosome near 64 minutes (Figure 1.2) (Ørskov *et al.*, 1976). Capsule genes near *serA* have been termed *kps* (Silver *et al.*, 1984; Vimr *et al.*, 1989), formerly *kpsA* (Ørskov and Nyman, 1974). The group II *kps* genes have been cloned for several chemically different group II K antigens and have a common genetic organisation consisting of three functional regions (Figure 1.3). Two of these regions (regions 1 and 3) are homologous between different group II K antigen gene clusters. The group II capsule gene cluster will be considered in detail later (Section 1.3.3). A comparable organisation of capsule genes encoding similar proteins has been identified in *H. influenzae* and *N. meningitidis* (Frosch *et al.*, 1989, 1991; Kroll *et al.*, 1989, 1990).

In contrast, group I K antigen genes are located near *his* and *rfb* at around 44 minutes on the *E. coli* chromosome (Figure 1.2) (Mäkelä and Stocker, 1984). Those that have been mapped include K26 (Ørskov and Ørskov, 1962), K27 (Schmidt *et al.*, 1977), K30 (Laasko *et al.*, 1988), K8, K9, K17 and K57 (Ørskov *et al.*, 1977). There have, however, been conflicting reports about the involvement of an additional *trp*-linked locus in group I capsule biosynthesis (see Section 1.3.4). Interestingly, the genes for the *Klebsiella* K antigens are also located near *his* and *rfb* on the *Klebsiella* chromosome (Laasko *et al.*, 1988). The term *kst* (Kapsel heat STable) has been suggested for the group I capsular polysaccharide genes to distinguish them from those of group II and the O antigen biosynthesis genes, *rfb* (Whitfield *et al.*, 1989). The group I K antigen biosynthesis genes have not been studied in detail.

1.3.2.3 Group I/II capsules

Very recently the existence of another group of *E. coli* K antigens, group I/II has been suggested. Certain capsular polysaccharides do not fit comfortably into either group I or II. These include the K2, K3, K10, K11, K19, K54 and K98 antigens. On the basis of heat lability, charge density and composition these capsules were all originally classed as group II K antigens (Table 1.2) (Jann and Jann, 1987, 1990). In support of this classification, phospholipid has been detected at the reducing end of the K10, K11, K54 and K98 antigens. In addition, KDO was detected at the reducing end of the K10 antigen but this analysis was not extended to include the K11, K54 and K98 antigens (Jann, K., unpublished data). None of these capsular polysaccharides

Figure 1.2: Genetic map of the *E. coli* chromosome (Bachmann, 1990)



are expressed with the group I - associated 'rfe-dependent' O antigens. The K10 and K54 antigens have been mapped to the *serA* region of the *E. coli* chromosome (Ørskov and Nyman, 1974). It was as a result of this initial mapping that the term *kpsA* was originally proposed for the K antigen genes near *serA*.

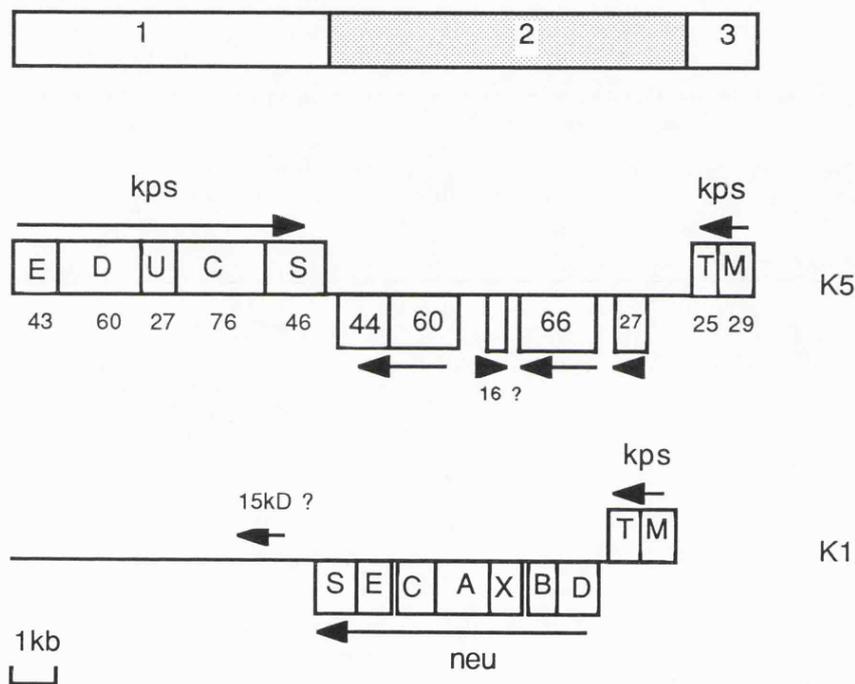
Group I/II capsules are similar to those of group I on two counts. Firstly K3, K10, K11, K54 and K98 capsules are expressed at 18°C (Ørskov and Ørskov, 1984). There is discrepancy in the case of K2 and K19 capsule expression: Ørskov and Ørskov (1984) report temperature-regulated expression but this is contradicted by Jann and Jann (personal communication) who have examined three different K2 antigen expressing strains. Secondly, the K2, K3, K10, K11, K19 and K54 capsule expressing strains do not have elevated CKS activity at 37°C (Finke *et al.*, 1990). It is unclear how significant these differences really are and the relationship between the group II and the group I/II capsule genes is examined in this thesis.

1.3.3 Group II capsule expression

Genes directing the production of several group II capsular polysaccharides have been cloned into *E. coli* K-12 laboratory strains. *E. coli* LE392 is not encapsulated and does not carry the group II capsule genes on its chromosome (Echarti *et al.*, 1983). The group II K antigens that have been cloned include K1 (Echarti *et al.*, 1983; Silver *et al.*, 1981), K5, K7, K12 and K92 (Roberts *et al.*, 1986). For details of the cloning strategies the reader is referred to a recent review (Boulnois and Roberts, 1990). Complete determinants for capsule expression are clustered and those for K1 occupy about 17 kilobases (kb) of DNA (Boulnois *et al.*, 1987). Analyses and comparisons of the cloned group II capsule gene clusters have been performed, the K1 and K5 antigen gene clusters being the subject of more intensive study. Very recently the nucleotide sequence of the entire K5 capsule gene cluster has been determined (Pazzani, C., Smith, A.N., Roberts, I.S. and Boulnois, G.J. unpublished data) and part of the nucleotide sequence of the K1 antigen gene cluster is also available (Pavelka *et al.*, 1991; Vann *et al.*, 1987; Vimr, E., unpublished data; Weisberger *et al.*, 1991).

The gene clusters for all group II K antigens so far analysed have the same organisation consisting of three functional regions (1, 2, and 3) (Figure 1.3). These have been defined on the basis of mutation analysis (both transposon and deletion mutants), complementation analysis and DNA homology. Region 2, the central region, is unique to each K antigen gene cluster studied (Roberts, M. *et al.*, 1988) and encodes for the synthesis of the polysaccharide in question since disruption of region 2 abolishes polymer production (Boulnois *et al.*, 1987; Roberts *et al.*, 1986; Silver *et al.*, 1984). Silent copies of alternative region 2 elements have not been detected elsewhere on the chromosome of group II capsule expressing strains (Roberts, M. *et al.*, 1988). Regions 1 and 3 are distinct from each other yet each is common to all group II K antigen gene clusters studied: there is DNA homology (greater than 95%) between the equivalent regions 1 and 3 in the capsule gene clusters of chemically different group II K antigens (Echarti *et al.*, 1983; Roberts *et al.*, 1986, 1988). Consistent with the observed

Figure 1.3: The K1 and K5 capsule gene clusters showing the three regions and the genes which have been identified within them



The K5 and K1 capsule gene clusters showing the genes which have been identified within them. The genes within the K5 capsule gene cluster are drawn to scale whilst those in the K1 genes are drawn to the best approximation on the basis of the available information. Boxes above the lines represent region 1 and region 3 genes (*kps*) which are common components of the group II *kps* gene cluster. The boxes below the lines denote region 2 genes, termed *neu* in the K1 capsule gene cluster. The predicted sizes of the *kps* and gene products are indicated. The predicted sizes of the proteins encoded by genes within region 2 of the K5 capsule gene cluster are shown including the putative 16kD protein whose existence has not to date been verified.

DNA homology, protein analyses revealed similar sized proteins encoded by regions 1 and 3 of different capsule gene clusters (Roberts *et al.*, 1986; Silver *et al.*, 1984). In addition, the physical maps of different group II K antigen gene clusters can be aligned on the basis of common restriction endonuclease cleavage sites (Roberts *et al.*, 1986). If region 1 or 3 are mutated, polymer is produced but it is intracellular (see below) (Boulnois *et al.*, 1987; Kronke *et al.*, 1990a). Mutations which disrupt either region 1 or 3 can be complemented *in trans* by the equivalent region from the gene cluster of another group II K antigen (Roberts *et al.*, 1986). These observations suggest that the gene products of regions 1 and 3 of different group II capsule gene clusters are responsible for the export of polysaccharide to the cell surface and that these products perform the same function in the expression of chemically different group II capsules. Molecular analysis of the cluster has prompted the introduction of unified nomenclature such that shared loci (regions 1 and 3) retain the *kps* designation. Region 2 is named according to the polysaccharide for which it encodes, for instance K1 region 2 is termed neu for N-acetyl NEUraminic acid (Silver *et al.*, 1984; Vimr *et al.*, 1989).

Group II capsule biogenesis will be considered in temporal sequence, that is polymer synthesis followed by translocation to the cell surface. It is important to remember that capsule biogenesis is a dynamic process and the two steps may overlap. The mechanisms of bacterial polysaccharide synthesis were, for a long while, best understood in *E. coli* and *Salmonella* spp LPS. As a result, LPS biosynthesis has served as a model in the interpretation of data from other systems, including the synthesis of capsular polysaccharide. Before considering the synthesis of group II K antigens it is helpful to take a brief look at the biosynthesis of other bacterial polysaccharides.

1.3.3.1 The mechanisms of LPS biosynthesis

Bacterial polysaccharides are generally polymerised from the nucleotide derivative of their sugar components at the inner membrane and capsular polysaccharide synthesis is not exceptional (Troy, 1979). If the monomers are ubiquitous components of the bacteria they may be diverted to polysaccharide synthesis and this obviates the need for additional synthetases. This is the case in LPS synthesis where, for instance, the housekeeping gene *galU* is involved in supplying glucose, a component of the core oligosaccharide (Rick, 1987). In general, one of two basic mechanisms of polysaccharide polymerisation are employed and these are exemplified by *rfe*-dependent and *rfc*-dependent O antigen synthesis (Jann and Jann, 1984; Mäkelä and Stocker, 1984; Rick, 1987; Robbins and Wright, 1971). Both methods use *rfb* (O antigen biosynthesis) encoded transferases, sugar synthetases and nucleotide activators. *rfb* is located near *his* in *E. coli*, *Klebsiella* and *Salmonella* (Figure 1.2).

Synthesis of the majority of *E. coli* and *Salmonella* O antigens is *rfe*-independent and *rfc*-dependent. In this group of O antigens, repeat units are individually constructed by *rfb* gene products from sugar nucleotide donors on a carrier lipid (undecaprenol pyrophosphate) in the cytoplasmic membrane. The O antigen grows at the reducing

end by the transfer of the nascent O antigen to the non-reducing terminus of a newly synthesised repeat unit. This is mediated by polymerase(s) encoded at *rfc* near *trp* at 34 minutes on the chromosome (Figure 1.2) (Mäkelä and Stocker, 1984). Disruption of *rfb* results in a rough LPS and *rfc* mutants are semi-rough. This basic mechanism is common in the production of many bacterial macromolecules including peptidoglycan.

Synthesis of the *rfe*-dependent O antigens (*E. coli* O8, O9, O20, O101 and some Salmonella O antigens) and the LPS core oligosaccharide occurs by an altogether different mechanism. The involvement of *rfe* (Figure 1.2) and independence on *rfc* has recently been proven (Sugiyama *et al.*, 1991). Study of this class has centred around *E. coli* O8 and O9 both of which are mannans. Polymerisation occurs by the sequential addition of sugars from GDP-mannose directly on to the non-reducing end of the growing polysaccharide (Jann *et al.*, 1982). Lipid intermediates are not involved. This polymerisation is dependent on *rfe*-encoded products which may be involved in production or function of the endogenous mannose acceptor (Jann and Jann, 1985). Similarly, LPS core is synthesised directly onto lipid A by the sequential addition of monomers to the non-reducing terminus of the growing core from nucleotide sugar donors. Glycosyltransferases encoded by *rfa* (core biosynthesis) are located in the cytoplasmic membrane (Rick, 1987).

Subsequent steps are the same in both groups of O antigen synthesis. The O antigen is transferred from the carrier lipid to complete core lipid A. Finally, mature LPS is irreversibly exported to the outer membrane apparently at zones of adhesion between the inner and outer membrane, (Bayer junctions) (Bayer, 1979; Muhlradt *et al.*, 1973). Details of coupling synthesis and translocation, export across the outer membrane, and regulation of LPS biosynthesis are unknown.

1.3.3.2 Region 2 gene products are essential for polysaccharide synthesis

With this information in mind, return now to group II K antigen biosynthesis. Polymerisation is thought to occur at the inner face of the cytoplasmic membrane, as membrane fractions are found to have transferase activity and newly synthesised polymer is located at the inner face of the membrane (Finke *et al.*, 1991; Kronke *et al.*, 1990b; Rohr and Troy, 1980). In addition, polysaccharide is located in the cytoplasm of mutants defective in the polysaccharide transport mechanism (Boulnois *et al.*, 1987). Polymerisation of group II capsular polysaccharides is thought to take place at the non-reducing terminus (Rohr and Troy, 1980; Troy, 1979). The K1 and K5 antigens are the best studied examples and appear to be synthesised by different mechanisms, considered below. K1 polysaccharide is poly-NANA (formerly called polysialic acid) and the *N. meningitidis* group B polysaccharide has an identical structure. The K5 polysaccharide is a heteropolymer composed of glucuronic acid and N-acetyl glucosamine (Table 1.1).

Evidence suggests that K1 synthesis involves a carrier lipid on which short NANA

oligosaccharides are synthesised and subsequently transferred to a membrane-bound endogenous acceptor on which the polymer elongates, similar to the mechanism of synthesis of *rfe*-dependent O antigens (Finke *et al.*, 1991; Jann and Jann, 1990; Troy, 1979; Vijay and Troy, 1975). A similar mechanism is proposed for the production of the *N. meningitidis* group B polysaccharide in which lipid-linked oligosaccharide intermediates of around five NANA residues were found to participate in capsule biosynthesis (Masson and Holbein, 1985). In contrast, it seems that the mechanism of synthesis of the K5 polysaccharide is analogous to that of *rfe*-dependent O antigen synthesis in which the sequential addition of monomers to the growing polysaccharide takes place without the involvement of lipid intermediates (Finke *et al.*, 1991; Jann and Jann, 1990). It appears therefore, that although group II K antigens have many characteristics in common, synthetic mechanisms may differ.

Molecular analysis of region 2 of the group II capsule gene cluster is consistent with the notion that it is entirely concerned with polysaccharide synthesis. From the known features of the K1 biosynthesis pathway it is anticipated that enzymes necessary for the synthesis, activation and polymerisation of NANA would be encoded within region 2 of the K1 capsule gene cluster. Region 2 of the K1 capsule gene cluster is about 5kb (Boulnois *et al.*, 1987) and the most recent information suggests that it encodes seven proteins, the genes for which may form a single transcriptional unit (Figure 1.3) (Silver and Aaronson, 1988; Vimr, 1989; Vimr, E., personal communication). Four of the proteins have been assigned a function. The products of *neuB* and *neuC* are involved in the synthesis of NANA (Vann *et al.*, 1987; Vimr *et al.*, 1989): Neu B is probably NANA synthetase itself and NeuC may be required for precursor production (Silver *et al.*, 1984). NANA is probably activated by NeuA which is thought to be CMP-NANA synthetase (Silver and Aaronson, 1988; Vann *et al.*, 1987). It was reported that *neuS* encodes the only sialyl transferase and this finding conflicts with the proposed biochemical pathway (Steenbergen and Vimr, 1990). If polymerisation of the K1 polysaccharide does involve a lipid intermediate then at least two sialyltransferases would be predicted, one to transfer sugars to the growing oligosaccharides and one to transfer the oligosaccharides from the lipid carrier onto the growing polysaccharide. This discrepancy has not been resolved. The amino acid sequence of NeuS has recently been published and is similar to that of the sialyl transferase encoded within the *N. meningitidis* group B capsule genes, further evidence that these identical capsular polysaccharides are produced by similar mechanisms (Frosch *et al.*, 1991).

Region 2 of the K5 capsule gene cluster is about 7kb (Roberts *et al.*, 1988). Although K5 is a heteropolymer and K1 a homopolymer its constituents are ubiquitous bacterial sugars and the presence of sugar synthetase genes in region 2 of the K5 capsule gene cluster would not be anticipated. It is predicted from the nucleotide sequence of region 2 of the K5 gene cluster that four or five proteins are encoded (Figure 1.3). The 44, 66, 60 and 27 kD proteins have been visualised by protein analysis but the proposed 16kD product of an ORF encoded on the opposite strand to the others has not been observed and may not be expressed. One feature of region 2 of the K5 capsule gene cluster is the presence of large intragenic gaps and this makes it unlikely that the region 2 genes are transcribed as a single unit and are presumably

coordinated by another mechanism. Functions have not been assigned to any of the K5 region 2 products to date.

Since polymerisation of group II K antigens proceeds at the non-reducing end of the polymer and KDO has been found at the reducing end it has been postulated that transfer of KDO from CMP-KDO to an acceptor is the reaction initiating group II capsule synthesis (Finke *et al.*, 1989). Activated sugar components would then be transferred to KDO and the polymer elongated on the acceptor. A structural gene, *kpsU*, encoding CMP-KDO synthetase (CKS) has been identified within region 1 of the K5 capsule gene cluster (Pazzani, C., Roberts, I.S. and Boulnois, G.J., unpublished). Its position in region 1, a common region, implies a universal role for CKS in group II capsule expression. In addition, the synthesis of CMP-KDO by KpsU may be a means by which the synthesis and export of polysaccharide are co-ordinated since *kpsU* is present within an operon, the other products of which are involved in polysaccharide export. Temperature regulated expression of *kpsU* could, in turn, regulate polysaccharide synthesis. Deletion mutants in which region 1, including *kpsU*, is removed still synthesise polysaccharide and this may be explained by the provision of CMP-KDO from elsewhere. KDO, donated by CMP-KDO, is a critical constituent of the LPS core oligosaccharide CKS associated with LPS expression is encoded elsewhere on the chromosome by *kdsB* (Goldman and Kohlbrenner, 1985). CKS thus appears to be involved in both LPS and group II polysaccharide biosynthesis. It is conceivable that cross-talk between the LPS and capsule biosynthesis pathways occurs such that K antigen synthesis uses CMP-KDO generated by KdsB in the absence of the capsular KpsU. Alternatively, incorporation of KDO could be the last step in K antigen synthesis. The synthesised polymer could be transferred from the acceptor onto KDO either on another acceptor or on the phospholipid. This step could act as a signal for polysaccharide export. KDO at the reducing end of the variable K antigen may be important in the recognition of the polysaccharide by the common export mechanisms encoded by regions 1 and 3.

In summary, region 2 encoded products, in cooperation with the region 1 encoded KpsU and some housekeeping genes, are responsible for the synthesis of a particular K antigen at the inner face of the cytoplasmic membrane. The mechanism of synthesis by which each K antigen is produced may differ but there is every indication to suggest that thereafter group II capsule biogenesis utilises the products of regions 1 and 3 which function in a manner independent of polymer structure.

1.3.3.3 Polysaccharide handling and export by Region 1 and 3 encoded proteins

The mechanism by which the synthesis of different group II capsular polysaccharides may take place, mediated by unique region 2 gene products has been described above. The remaining steps in group II capsule biogenesis involve the transfer of the polysaccharide from the acceptor to its eventual phospholipid anchor molecule (although this may occur during, or as an integral part of, the synthesis), the translo-

cation of polysaccharide across the inner and outer membranes and the assembly of a complete capsule on the cell surface. It is generally accepted that these products perform the same function in the expression of different group II K antigens. It is possible, however, that the common regions 1 and 3 products interact with the unique region 2 encoded products. For ease of reference the region 1 and 3 encoded products are listed in Table 1.3 and illustrated in Figure 1.3.

Region 3 is 1.6kb and mutations in this region result in cytoplasmic polysaccharide which is of low molecular weight and is not linked to phospholipid, whether KDO is present is not known (Boulnois *et al.*, 1987; Kroncke *et al.*, 1990a). The absence of KDO would suggest that KDO is not involved in the reactions initiating polysaccharide synthesis. Disruption of membrane potential by the addition of carbonyl-cyanide-M-chlorophenol hydrazone (CCCP) prior to temperature upshift also results in cytoplasmic polysaccharide (Kroncke *et al.*, 1990b). This suggests that region 3 products are involved in the energy dependent translocation of the polysaccharide across the inner membrane (Boulnois and Jann, 1989). Region 3 from both the K1 and K5 capsule gene cluster has been sequenced and analysed (Pavelka *et al.*, 1991; Smith *et al.*, 1990). Both contain two genes, *kpsT* and *kpsM* present in one transcriptional unit (Figure 1.3). The nomenclature *kpsT*_{K1} and *kpsT*_{K5}, for example, will be used to distinguish products encoded by different capsule gene clusters. It is postulated that KpsM and KpsT proteins together belong to the family of both eukaryotic and prokaryotic proteins with import or export functions, the 'traffic ATP-Pases' or 'ABC (ATP-binding cassette) transporters' (Ames *et al.*, 1990; Higgins *et al.*, 1990b; Hyde *et al.*, 1990; Pavelka *et al.*, 1991; Smith *et al.*, 1990). The proteins exhibit the characteristics of this family: KpsT is extremely hydrophobic and postulated to be an integral membrane protein whilst KpsM is hydrophilic and has a consensus ATP-binding domain. It is therefore likely that KpsT is an integral component of the inner membrane whilst KpsM is cytoplasmic. On the basis of this it has been postulated that the region 3 gene products, KpsT and KpsM are involved in translocation of capsular polysaccharide across the inner membrane. It appears that lipidation occurs during or after this step. Why disruption of these two proteins should result in low molecular weight polysaccharide is not clear, perhaps this is due to a negative feedback mechanism or perhaps the synthesis and transport of polysaccharide is intimately linked. Synthesis and export may occur simultaneously or the region 3 encoded proteins may interact with the biosynthetic machinery perhaps as a scaffold at the inner membrane. Proteins similar to both KpsM and KpsT are encoded within the *H. influenzae* and *N. meningitidis* capsule gene clusters (Table 1.3).

Region 1 is about 7kb and encodes at least five proteins (Figure 1.3). The complete nucleotide sequence of region 1 of the K5 capsule gene cluster has been determined (Pazzani, C., Boulnois, G.J. and Roberts, I.S., in preparation). A set of similar proteins were identified in region 1 of K1, K5 and K7 capsule genes (Roberts *et al.*, 1986; Silver *et al.*, 1984). It is still not clear whether the proteins encoded by region 1 of different *kps* clusters are chemically identical or whether DNA hybridisation analyses have been insufficiently sensitive to detect small differences. The region 1 genes will be discussed using those in the K5 capsule gene cluster as the paradigm.

Table 1.3: Characteristics of the proteins encoded within regions 1 and 3 of the *E. coli* group II capsule gene cluster.

Protein	Size	Similar proteins
KpsE	43	BexC, CtrB
KpsD	60	-
KpsU	27	KdsB
KpsC	75	69kD §
KpsS	46	51kD §
KpsM	30	BexB, CtrC
KpsT	25	BexA, CtrD

The Kps proteins predicted from the nucleotide sequence of the K5 capsule gene cluster are listed. The first five proteins listed are encoded within region 1, the last two in region 3. The predicted size of each protein is given in kilodaltons. In the final column, proteins with similarity to the Kps proteins are given. The location of the genes encoding these proteins within the gene cluster are shown in Figure 1.3. Proteins designated Bex or Ctr are encoded by the *H. influenzae* or *N. meningitidis* capsule gene clusters respectively and KdsB is the CMP-KDO synthetase involved in LPS core biosynthesis. Proteins marked § are the predicted products of region 3 of the *H. influenzae* capsule gene cluster and are referred to by their predicted molecular weight.

The nucleotide sequence suggests that region 1 forms a single transcriptional unit which is transcribed towards region 2 (Figure 1.3). If region 1 of the K1 capsule gene cluster is deleted, full length polysaccharide which is linked to phosphatidic acid is located in the periplasm (Boulnois *et al.*, 1987; Roberts *et al.*, 1988) indicating that region 1 encoded proteins are, at least in part, responsible for translocation of capsular polysaccharide across the outer membrane.

The first gene to be transcribed in region 1 is *kpsE*. This encodes a 43kD protein which is postulated to be anchored in the membrane by means of a membrane spanning C terminal domain although the orientation of the protein is not clear. In temperature upshift experiments of wild type cells, extracellular capsular material was first visualised on the cell surface by electron microscopy as discrete tufts associated with specific membrane sites (Kronke *et al.*, 1990b). These sites are postulated to be zones of adhesion between the inner and outer membrane (Bayer junctions). This suggests that the capsular material is exported to the cell surface at specific sites and then distributed to form a complete capsule, as was suggested for LPS. Recombinants harbouring a deletion mutation in the site equivalent to *kpsE* in the K1 antigen gene cluster produced similar capsular tufts (cited in Boulnois and Roberts, 1990). It is postulated that KpsE may be involved in the assembly and distribution of capsule on the cell surface. Similar proteins are encoded by the *H. influenzae* and *N. meningitidis* capsule genes (Table 1.3) (Frosch *et al.*, 1991; Kroll *et al.*, 1990).

Downstream from *kpsE* in region 1 of the group II K antigen gene cluster is *kpsD* (Figure 1.3). This gene encodes a 60kD periplasmic protein which was first described in K1 and is likely to be homologous to that identified in the K5 capsule gene cluster (Silver *et al.*, 1987). This protein may be involved in translocation of the polysaccharide across the periplasm. Downstream of *kpsD* is *kpsU*. This gene encodes CKS and was described earlier in relation to polysaccharide synthesis (Section 1.3.3.2). KpsU is a 27kD protein which was identified as a product of region 1 of the K5 capsule gene cluster and although it has not been directly observed as a product of other group II capsule gene clusters, these are associated with elevated CKS activity (Finke *et al.*, 1989) and nucleotide sequence matching part of *kpsU*_{K5} was identified upstream of *kpsD*_{K1} (Vimr, unpublished). Group II capsule expressing strains appear to carry two structural genes for CKS, *kpsU* and *kdsB*, the former involved in K antigen production, the latter in LPS core biosynthesis. Comparison of the amino acid sequence of KpsU and KdsB reveals that these two proteins are 70% similar.

KpsC, a 76kD protein is believed to be a cytoplasmic protein on the basis of the proposed amino acid sequence and its function is not known. Finally, *kpsS*, the last gene in the region 1 operon of the K5 capsule gene cluster encodes a 46 kD protein which may have, as predicted from the nucleotide sequence, a cytoplasmic N terminus, a membrane spanning domain near the C terminus and a periplasmic, negatively charged C terminus. Mutants defective only in *kpsS* produce cytoplasmic polysaccharide which is not associated with the inner membrane. It is possible that KpsS is a scaffolding protein which may be responsible for linking the synthetic complex with a translocation complex formed by KpsT and KpsM. It has been found that proteins similar to KpsC and KpsS but not, to date, to KpsU are encoded by the

capsule gene clusters of other organisms (see Section 1.3.3.4 and Table 1.3). Although the majority of information relating to group II capsule gene clusters which direct the synthesis of different K antigens suggests that regions 1 and 3 are identical, this may not be entirely true. For instance, E. Vimr (unpublished data) has identified an open reading frame (ORF) on the lower strand within region 1 of the K1 capsule genes which encodes a putative 15 kD protein (Figure 1.3). This ORF was not found in region 1 of the K5 capsule gene cluster. The synthesis of K1 and K5 may also occur by different mechanisms. The group II capsules, although distinct may not be completely uniform.

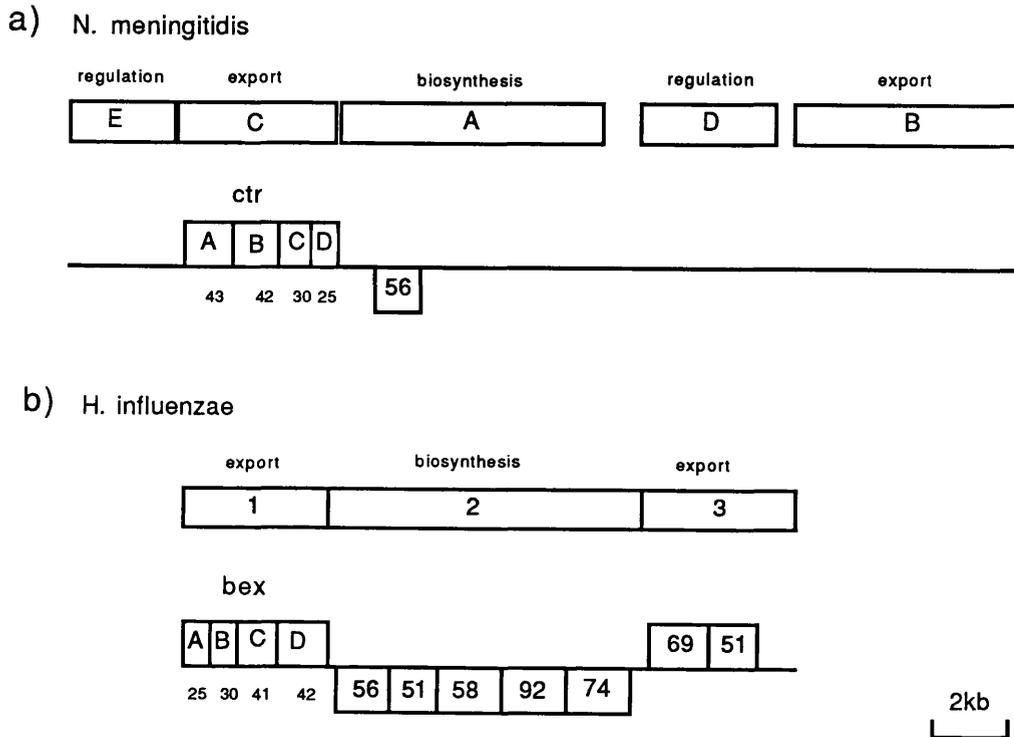
In summary, molecular analysis of the *kps* gene cluster has dissected the processes involved in group II capsule biogenesis and a speculative picture of group II capsule biogenesis can be constructed. Polysaccharide synthesis takes place from the non-reducing end of the molecule and is catalysed by K antigen-specific region 2 products. Synthesis occurs at the inner face of the cytoplasmic membrane. The mechanism of synthesis is not clear and may differ between different K antigens. Addition of KDO may be the initiating reaction in polymer synthesis, or may conclude it. Polysaccharide is synthesised and translocated to the cell surface by the cooperative actions of the region 1, 2 and 3 gene products. Polysaccharide is probably translocated across the inner membrane by KpsT and KpsM, possibly involving KpsS in the cytoplasm and KpsD in the periplasm and at some point at least some molecules are attached to phospholipid at the reducing terminus. Polysaccharide is translocated across the periplasm and outer membrane onto the cell surface where it is distributed as a capsule, perhaps involving the function of KpsE. This complete process takes place at 37°C.

1.3.3.4 *H. influenzae* and *N. meningitidis* capsules

In the discussion of *E. coli* group II K antigens there have been several references to the capsules of *H. influenzae* and *N. meningitidis*. The capsules produced by these organisms share some of the biochemical characteristics which are associated with the group II capsules of *E. coli*. In addition, genes encoding the capsules of *H. influenzae* and *N. meningitidis* are organised in a modular fashion (Figure 1.4). The similarity between the capsule gene clusters of these different Gram negative organisms is striking and it has been suggested that these genes share a common ancestor (Boulnois and Jann, 1989; Frosch *et al.*, 1991).

H. influenzae produces six different capsular types, types a through f (Pittman, 1931) type b being most commonly associated with disease causing isolates (Moxon and Kroll, 1984; Turk and May, 1967) and consequently the best studied. The capsule genes (*cap*) of types a, b, c, d and e have been analysed and shown to consist of three functional regions (Figure 1.4) (Falla *et al.*, 1991; Kroll *et al.*, 1989). Like the *E. coli* group II capsule gene cluster, the central region (region 2) is involved in polymer synthesis and the flanking regions (regions 1 and 3) in export of the polysaccharide to the cell surface. Regions 1 and 3 are conserved between different capsule gene clusters

Figure 1.4: The capsule gene clusters of *H. influenzae* and *N. meningitidis*



(a) The capsule gene cluster (*cps*) of *N. meningitidis* group B showing the five functional regions as boxes labelled A through E above. The possible functions of each region is written above each box. The nucleotide sequence of region C has been determined and the genes, *ctrA*, B, C and D are shown to scale with the sizes of the predicted proteins marked (Frosch *et al.*, 1991). Region A is serotype specific, the other regions are present in capsule gene clusters which encode serologically distinct polysaccharides (Frosch *et al.*, 1989).

(b) The capsule gene cluster *cap* of *H. influenzae* type b showing the three regions, 1, 2 and 3 as boxes labelled above. The proposed function of each region is written above. The genes present within region 1 are marked to scale, *bexA*, B, C and D and the sizes of the predicted proteins encoded by each gene is given below (Kroll *et al.*, 1990). The genes marked within regions 1 and 3 (notified by the size of the predicted product) have been identified in a preliminary analysis and are not marked to scale (Kroll, S.J., unpublished data).

(Kroll *et al.*, 1989) and the nucleotide sequence of regions 1 and 3 from the type b capsule gene cluster has been determined (Kroll *et al.*, 1990; Kroll, S.J. unpublished data). The organisation of regions 1 and 3 are not identical to those in *E. coli* group II *kps* but some of the gene products are similar (see below). Region 2 of the *H. influenzae* type b capsule gene cluster probably encodes five proteins of 56, 51, 58, 92 and 74kD (Figure 1.4) (Kroll, S.J., unpublished data). The functions of these individual proteins is not known. One striking difference between the *H. influenzae* capsule gene cluster and the capsule genes of *E. coli* and *N. meningitidis* is that the former exists as a tandem repeat in the majority of clinical isolates (Hosieth *et al.*, 1986).

The *N. meningitidis* capsule genes have also been analysed and study has centred around the medically important group B genes (*cps*), the polysaccharide of which is identical to that of the *E. coli* K1 antigen. Although DNA homology was not detected between the *E. coli* K1 antigen gene cluster and *N. meningitidis* group B (Echarti *et al.*, 1983), recent analysis has revealed some similarity in both the capsule biosynthesis and export proteins (Frosch *et al.*, 1991). The *cps* locus has been divided into five functional regions, regions A to E (Figure 1.4) (Frosch *et al.*, 1989). Region A, the central region, encodes enzymes for the synthesis of the polysaccharide in question and is therefore equivalent to region 2 in the *E. coli* and *H. influenzae* capsule genes. Region A of the group B capsule genes encodes a sialyl transferase, the predicted amino acid sequence of which has some similarity to that of NeuS encoded by the *E. coli* K1 capsule genes (Frosch *et al.*, 1991). Region C of the *cps* locus is involved in polysaccharide export and has been sequenced (Frosch *et al.*, 1991). Region B is also involved in polysaccharide export and regions D and E are thought to have regulatory functions (Frosch *et al.*, 1989), however, none of these regions have been subjected to a detailed analysis.

Region 1 of the *H. influenzae* *cap* locus and region C of the *N. meningitidis* *cps* locus strongly resemble one another. Nucleotide sequence analysis has revealed that both contain four genes, termed *bexA*, B, C and D in *H. influenzae* and *ctrD*, C, B and A in *N. meningitidis* (Figure 1.4), and that these probably form an operon (Frosch *et al.*, 1991; Kroll *et al.*, 1990). Comparison of the predicted amino acid sequence of these proteins reveals that *BexA*, B, C and D are analogous to *CtrD*, C, B and A respectively in terms of molecular weight, structure and proposed function (if known). In addition, some of these capsule export proteins are homologous to the *Kps* proteins encoded within regions 1 and 3 of the *E. coli* K5 group II capsule gene cluster. Like *KpsT* and *KpsM*, *BexA* and *BexB*, and *CtrD* and *CtrC* are postulated to be ABC transporters involved in the membrane translocation of capsular polysaccharide (Frosch *et al.*, 1990; Kroll *et al.*, 1991; Smith *et al.*, 1990). The predicted soluble components of these putative ABC transporters (*KpsT*, *BexA* and *CtrD*) show extensive amino acid sequence homology to each other and although the proposed integral membrane components (*KpsM*, *BexB* and *CtrC*) are less similar their hydropathy profiles are almost identical. Disruption of *bexA*, resulting in a failure to export capsular polysaccharide, has been associated with the observed high frequency loss of capsule expression in *H. influenzae* type b (Hosieth *et al.*, 1985; Kroll *et al.*, 1988). Further to the similarity between the putative ABC transporters,

KpsE is similar to BexC and CtrB but a Kps equivalent to BexD and CtrA has not been identified. Preliminary investigation of region 3 of the *H. influenzae* capsule genes indicates that two proteins are encoded with predicted sizes 69kD and 51kD and that these resemble KpsC and KpsS respectively. The prediction that analogous proteins are encoded by the capsule genes of *E. coli*, *H. influenzae* and *N. meningitidis* strongly suggests that the export of capsular polysaccharide in these three Gram-negative organisms is mediated, at least in part, by very similar mechanisms and that the capsule genes may share a common origin.

1.3.4 Group I capsule expression

A detailed study of *his*-linked group I capsule genes (*kst*) has not been made and it is not clear whether the genes for chemically different group I K antigens share a common organisation. Group I K antigens resemble LPS in containing core-lipid A, but it is uncertain whether LPS determinants are involved in group I capsule expression. Study of the structure of some *E. coli* group I capsular polysaccharides has provided evidence that an *rfc*-like polymerase may be involved in group I capsule biosynthesis. *rfc* LPS mutants are described as semi-rough meaning that core-lipid A is substituted with only one O antigen repeat oligosaccharide (Rick, 1987). An *rfc*-encoded polymerase is required to elongate the O antigen. *rfc* is located near *trp* at 34 mins on the *E. coli* chromosome (Figure 1.2) (Mäkelä and Stocker, 1984). A proportion of the wild type group I capsule has been found to have a structure equivalent to semi-rough LPS (Homonylo *et al.*, 1988; Ørskov *et al.*, 1977), in which a single K antigen repeat unit is attached to core-lipid A. Transconjugation and mutation experiments have also yielded 'semi-rough' K27 and K30 antigen forms and this phenotype has been termed Ki (K intermediate) (Schmidt *et al.*, 1977; Whitfield *et al.*, 1989). Hence, it seems that group I capsule synthesis makes use of an '*rfc*-like polymerase'. Consistent with this, preliminary biochemical evidence suggests that group I capsules are synthesised in an *rfc*-dependent fashion (Jann and Jann, 1990). In addition, *Klebsiella aerogenes* K antigens, which resemble those of *E. coli* group I, are synthesised by the polymerisation of oligosaccharide repeats on a carrier lipid, very much like the synthesis of *E. coli* and *Salmonella rfc*-dependent O antigens (Troy, 1979; Troy *et al.*, 1971, 1972).

Although an '*rfc*-like polymerase' has been implicated in both K27 and K30 capsule expression, its chromosomal location appears to differ between the two. Complete determinants for the expression of a K30 capsule appear to be located near *his* since mutants which either failed to express the K30 antigen or expressed only the Ki form of the antigen were restored by a *his* - linked locus (Whitfield *et al.*, 1989). This is supported by the results of crosses using *Klebsiella* K20 and *E. coli* K30 (Laasko *et al.*, 1988). The *Klebsiella* K20 and *E. coli* K30 antigens are biochemically indistinguishable although capsule morphology is different. In crosses using *E. coli* K-12 as the recipient it was shown that the determinants which dictate capsule morphology, in addition to those for the biosynthesis of the K30 antigen itself, are located near *his* (Laasko *et al.*, 1988). In contrast to K30 capsule expression, restoration

of *his*⁺ transconjugants which expressed the Ki form of the K27 antigen required a *trp* - linked locus (Schmidt *et al.*, 1988). The discrepancy between the K27 and K30 antigens may indicate that the group I capsules do not share a common genetic organisation as was found in the case of the group II.

Given the biochemical similarities of LPS and group I capsules it is feasible that LPS functions are involved in group I capsule expression. For example, some group I K antigens appear as an acidic LPS being attached to core-lipid A. In crosses involving *his*⁻ recipients and group I capsule expressing donors, hybrids which expressed the donor O antigen usually expressed the donor K antigen as well (Laasko *et al.*, 1988; Ørskov *et al.*, 1977; Schmidt *et al.*, 1977). This indicates that *rfb* and the group I capsule genes (for which the term *kst* has been proposed) are closely linked near *his*. Despite the genetic linkage and biochemical similarities between group I K antigens and LPS the number of shared loci must be small because the majority of K30⁻ mutants generated by Whitfield *et al.* (1989) expressed a normal O9 antigen and only a few were O9⁻:K30⁻.

1.4 The genetic diversity of *E. coli*

It has been suggested that the *E. coli* species comprises organisms of up to 15% sequence divergence with a G plus C content, ranging from 48% to 52% (Brenner *et al.*, 1972; Brenner and Falkow, 1971). Although some base changes do not affect the proteins expressed by the organism some do alter the amino acid sequence. Multi-locus enzyme electrophoresis (MLEE) analysis reveals a number of isoenzyme types, electrophoretic studies of some enzymes indicate that 94% of *E. coli* loci are polymorphic (Whittam *et al.*, 1983; Selander *et al.*, 1987). The majority of these changes are deemed to be selectively neutral (Selander *et al.*, 1987). A more striking variability within the *E. coli* population than isoenzyme types, is the phenotype of the cell surface. *E. coli* O and K antigens exhibit phenotypic variation in which a single isolate is restricted to the expression of one of a large number of alternative antigens, switching or conversion to another antigen type has not been observed (Ørskov and Ørskov, 1990). *E. coli* fimbriae exhibit phase variation; a single isolate comprises a mixture of cells expressing fimbriae with different antigenic, and often receptor, specificity (Smyth, 1986).

The *E. coli* population is not a continuous spectrum of individuals but is composed of a small number of broadly distributed, independently evolving lineages or clones (Ørskov and Ørskov, 1983; Selander and Levin, 1980; Whittam *et al.*, 1983). Despite the high incidence of variation, only a limited number of genotypes are apparent. This is observed as a non-random association of alleles (linkage disequilibrium), meaning that although individual loci vary, there is a strong tendency for certain combinations of alleles to be associated (Whittam *et al.*, 1983). These independent clones are stable, suggesting that genetic exchange, which would homogenise the population, is severely restricted and the entire chromosome maintains its clonal

identity (Ochman and Selander, 1984). Nucleotide sequence, MLEE and restriction fragment length polymorphism (RFLP) analyses have been employed to study the *E. coli* population and to assess clonal boundaries (Krawiec and Riley, 1990; Selander and Musser, 1990).

Despite the restricted genetic exchange, it is found that chromosomal regions have different ancestries and that horizontal gene transfer does occur (Arthur *et al.*, 1990; Maynard-Smith, 1991; Plos *et al.*, 1989). A number of mechanisms which generate genetic diversity have been proposed (for a review see Krawiec and Riley, 1990). These events take place superimposed on the background of the stable clonal chromosome and the majority of changes are deemed selectively neutral. The recombinational events described below occur at a very low rate, comparable to the spontaneous mutation rate, but they do have important long term consequences in the development of novel, adaptive phenotypes. The occurrence of random point mutations is one means by which genetic diversity is generated, although the majority are silent. Secondly, mutational events such as duplication, deletion or inversion of segments take place although large-scale alterations of the *E. coli* chromosome have not been observed, suggesting that the *E. coli* chromosome cannot tolerate such change. Analysis of individual housekeeping loci (*trp*, *phoA* and *gnd*) have revealed that nucleotide sequence polymorphisms are clustered and it is postulated that intragenic recombination involving small sequences of around 70-200 bp takes place resulting in, what is often referred to as 'mosaic genes' (Dubose *et al.*, 1988; Milkman and Crawford, 1983; Stoltzfus *et al.*, 1988). It has been proposed that horizontal gene transfer involving larger DNA segments (20-100 kb) also occur (Krawiec and Riley, 1990; Milkman and Bridges, 1990). Despite these rearrangements the chromosomal DNA between any two *E. coli* strains is highly homologous.

Selectively neutral changes are deemed to take place at a very low rate. However, for some genes selection pressure is strong and horizontal gene transfer is more likely to occur. This is the case for surface structures, such as adhesins where the ability to express the trait may confer the ability to colonise a new environment. Selectively advantageous genes are more likely to cross clonal boundaries, perhaps by conjugation, than selectively neutral 'housekeeping genes.' This was illustrated by Arthur *et al.* (1990) in which *rrn* (ribosomal RNA) RFLP profiles correlated with MLEE types whilst the frequency and restriction profiles of the F13 P-fimbrial adhesin genes were not related to clonal boundaries. Selectively neutral DNA which is located immediately adjacent to the advantageous allele may be transferred with it, a process referred to as 'hitchhiking' (Hedrick, 1982; Thomson, 1977). Some virulence factors are known to be linked and the idea that 'virulence gene blocks' are inherited or deleted from some strains has been suggested (Hacker, 1990; Hacker *et al.*, 1990).

1.4.1 The variability of O:K serotype within the *E. coli* population

Given the genetic division of the *E. coli* population, it is understandable that phenotypic variation, namely O and K antigen diversity, has arisen and been maintained in the population in the absence of unifying genetic exchange. Both O and K antigens are restricted to certain clones although serotyping is not a reliable indication of clonal identity because strains with the same O:K:H type may, on occasion, have very different chromosomal backgrounds (Achtman and Plusche, 1986; Caugant *et al.*, 1985; Ochman and Selander, 1984). K1 is restricted to six different clones and has been isolated in combination with O1, O7 and O18 (Achtman *et al.*, 1983). O8 and O9, and presumably group I K antigens are found in three distantly related pairs of clones (Ochman *et al.*, 1984). Despite the likely selective advantages of the K1 antigen gene cluster, it is more restricted than the P-fimbrial adhesin genes which are found over a more widespread selection of clones (Plos *et al.*, 1989; Selander *et al.*, 1986) at various positions on the *E. coli* chromosome associated with different restriction fragments (Arthur *et al.*, 1990).

LPS is a common feature of the Enterobacteriaceae and presumably has more ancient evolutionary origins than the K antigen. The ability to express K antigens has probably been acquired by *E. coli* since the two closely related species *E. coli* and *Salmonella* diverged. Riley and Krawiec (1987) identified several accretion domains throughout the *E. coli* (and *Salmonella*) chromosome by the comparison of the genetic linkage maps of *E. coli* and *S. typhimurium*. It was suggested that approximately 15% of the *E. coli* chromosome evolved by genomic accretion, that is, horizontal gene transfer. The *kps* genes have been placed in the accretion domain at 64 map units near *serA* (Riley and Krawiec, 1990; Vimr, 1991). The location of *kps* on such a domain suggests an independent origin for the *kps* genes and the domain to which they belong. Both *E. coli* and *Klebsiella* spp have acquired the ability to express over 70 different capsular K antigens but *E. coli* seems to have acquired more than one system with which to do so. The group I capsules resemble those of *Klebsiella* spp and both are encoded by *rfb* near *his*. The group II capsules resemble those of other Gram negative organisms: *H. influenzae* and *N. meningitidis*. Perhaps different capsule genes were acquired by different *E. coli* clones and because of the limited genetic exchange between clones have been fixed into different parts of the *E. coli* population.

O and K antigen expression is not restricted to *E. coli*. Expression of a wide array of chemically different polysaccharide antigens, particularly O antigens, is a common feature of the Enterobacteriaceae. It is not clear why so many structurally distinct variations of the O and K antigens are produced by a single organism and there are many unanswered questions. For instance, all *E. coli* K antigens, irrespective of precise chemical composition or capsule group provide an acidic, negatively charged, gelatinous layer around the outside of the cell. The natural reservoir of *E. coli* is the mammalian gut but the organism must also survive the outside environment in the transition between hosts. Although the presence of a capsule is likely to

be advantageous in both of these natural habitats the significance of the enormous variability of the K antigens is not clear. In pathogenicity a varied cell surface may be advantageous and some polysaccharide antigens, including for example the K1 antigen, may be more advantageous than others because of their capacity as a virulence factor. However, pathogenicity is the exception not the norm.

Despite the universal features of the capsule perhaps group I and group II capsules do not perform the same function. One indication that they may not, is that group II capsules are expressed in the mammalian body (37°C) and may not be expressed in the outside environment (18°C) (Table 1.2). Group I capsules are probably expressed in both habitats. It is possible that group II capsules and colanic acid together perform the basic function of the group I capsule, although this is entirely speculative. In group II capsule expressing strains at 37°C the group II capsule is expressed and colanic acid production is negligible. At 18°C the situation is reversed and colanic acid is predominant. It has been suggested that group I capsule expressing strains lack *cps*, the colanic acid biosynthesis genes (Whitfield, personal communication) which are tightly linked to *rfb*, near *his* as are the group I K antigen genes (Bachman, 1990). Colanic acid is not unique to *E. coli* and is produced by most Enterobacteriaceae, implying a universal function. If this is the case and group I capsules and colanic acid perform a similar role at 20°C why the group I K antigen should be structurally diverse is not clear. Even less is known about group I/II capsules. Theories about the acquisition and evolution of the group II K antigen gene cluster have been proposed.

1.4.2 Generation of the group II K antigen gene cluster

The group II capsule gene cluster consists of two common flanking regions (*kps*) separated by one of a number of K antigen-specific region 2 elements. This organisation has been maintained with apparent precision. A similar organisation exists within the *H. influenzae* and *N. meningitidis* capsule locus and all three encode homologous proteins (Section 1.3.3.2). It has been suggested that P-fimbrial adhesin genes are inherited by a transpositional event because terminal inverted repeats have been identified at either end of the cluster (Rhen, 1985). The *H. influenzae* capsule gene cluster also has the features associated with a transposon (Kroll *et al.*, 1991). Whether insertion sequences also flank the *E. coli* group II capsule gene cluster is not known. If these capsule gene clusters originated from a common ancestor, as has been suggested (Frosch *et al.*, 1991) it may have been by a transpositional event. Unlike *N. meningitidis* and *H. influenzae*, *E. coli* is not naturally transformable and this is therefore unlikely to be the method of acquisition of new genetic material by *E. coli*. Whatever the origin of the group II capsule gene cluster the precise organisation and DNA sequence of the *kps* genes is unique to *E. coli* and *kps* is highly conserved between different *E. coli* isolates which are spread over different clones. How the group II capsule gene cluster might have acquired so many alternative region 2 configurations is an interesting problem.

One theory, the so-called 'cassette theory', which has been proposed by Boulnois and Jann (1989) accounts for the generation of variable region 2 elements within the group II *kps* locus. In this theory region 2 undergoes genetic rearrangement *in situ* between regions 1 and 3. Deletion, duplication and insertion of sequences mediated by recombination and in conjunction with random point mutation result in different region 2 gene clusters which direct the expression of different polysaccharides. The oligosaccharide repeat unit of group II polysaccharides is variable but the range of sugars is not particularly wide, certain amino sugars and acidic sugar are common. Sugar synthetases (where necessary), activators and transferases may therefore be very similar. Genetic exchange in this model is thought to occur, in part, between different region 2 elements by homologous recombination resulting in the production of related region 2 elements which direct the synthesis of novel but not completely dissimilar polysaccharides. It is apparent that region 2 of the related K1 and K92 polysaccharides (both poly-NANA) may share some DNA sequences for some restriction endonuclease cleavage sites are common (Roberts *et al.*, 1986). In addition, polysaccharide biosynthesis genes may be donated from other sources, such as the LPS biosynthesis genes thus increasing the number of combinations of genes and the variability of the product. Perhaps the observed similarity between the LPS CKS gene, *kdsB* and the capsule equivalent, *kpsU*, indicates that LPS and group II capsule genes may be related. Group II K antigens containing NANA present an intriguing problem. As mentioned previously NANA is not a common bacterial sugar although it is a common component of mammalian polysaccharides. The ability to produce NANA does appear to be selectively advantageous for NANA-containing polysaccharides are typically poor activators of the alternative complement pathway.

The mechanism proposed by Boulnois and Jann (1989) is attractive but it is not the only possibility. An entire region 2 could be inherited from elsewhere and be inserted between the common regions 1 and 3 perhaps by a transpositional mechanism. The source of region 2 genes could be, for instance, *rfb*. Perhaps specific sequences exist at the junction of region 1 and 3 with region 2 which direct the shuffling and insertion of sequences.

These mechanisms relate purely to the group II capsule genes at *kps*, no reference is made to the capsule genes, *kst*, which encode for the production of group I K antigens. The genetic relationship between these two loci is far from clear. *kps* and *kst* may have a common evolutionary origin although this seems unlikely given the observed number of differences between capsules of the two groups. Alternatively, the two could be unrelated, running parallel evolutionary lines. Both loci could be related to *rfb*. Whether or not *kps* and *kst* are related, it would be interesting to know whether one capsule gene cluster is present on the chromosome of a strain expressing a capsule of the other type encoded by the other locus. A more informed speculation awaits the characterisation of the group I capsule genes. Similarly, it is impossible to fit the group I/II capsules into this scheme until more information is available.

1.5 The aims of this thesis

It is apparent that *E. coli* K antigens are variable, not only are they structurally diverse but also they can be divided into at least two groups. Thus, *E. coli* seems to have more than one means of producing what appears to be a variation of the same structure, the capsule. Within group II capsules, variability appears to be due to the presence of a K antigen-specific region 2 element and common products encoded by regions 1 and 3 seem able to handle and export structurally variable group II K antigens. Very little is known about the group I capsule genes. In addition, group I/II capsules have been separated from the main body of group II although it is far from clear how different they are or whether they are encoded by the group II capsule genes. It is the variability of the *E. coli* capsule genes, both within group II and between all three proposed groups, which is the subject of this thesis.

Chapter 2

Materials and Methods

2.1 Bacterial Strains and Plasmids

The bacterial strains and plasmids which were used in this study are listed in Table 2.1.

2.1.1 Growth Conditions and Media

Unless otherwise stated bacteria were grown in Luria broth at 37°C with the addition of 1.5% agar (BBL) as required. D.O. medium (Rodriguez *et al.*, 1988) was used as stated (per litre: 7g K₂HPO₄, 2g KH₂PO₄, 0.5g sodium citrate, 0.1g MgSO₄, 1g ammonium sulphate, 20g casamino acids, 5g yeast extract, 2g glucose to pH 7.2 with sodium hydroxide). 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 and tetracycline at 20µg/ml. Antibiotics were obtained from Sigma Chemical Company Ltd.

Bacterial cells were routinely harvested by centrifugation (3000g) at 4°C for 10 mins.

2.2 Transformation of Bacterial Cells

2.2.1 Production of Competent Cells

L-broth (10ml) was inoculated with 100µl of an overnight culture and grown to mid-log phase (OD₆₀₀ 0.5). The cells were washed in 10mM NaCl and resuspended in

Table 2.1: Bacterial strains and plasmids

LE392	F ⁻ <i>hsdR514</i> (r _K ⁻ , m _K ⁺), <i>supE44</i> , <i>supF58</i> , <i>lacY1</i> , <i>galK2</i> , <i>galT22</i> , <i>metB1</i> , <i>trpR55</i> , λ ⁻	Maniatis <i>et al.</i> , 1982
HB101	F ⁻ , <i>hsdS20</i> (r _B ⁻ , m _B ⁻), <i>recA13</i> , <i>ara-14</i> , <i>proA2</i> , <i>lacY1</i> , <i>galK2</i> , <i>rpsL20</i> (Sm ^r), <i>xyl-5</i> , <i>mtl-1</i> , <i>supE44</i> , λ ⁻	Maniatis <i>et al.</i> , 1982
JM101	F', <i>supE</i> , <i>thi</i> , (<i>lac-proAB</i>), <i>traD</i> ₃₆₁ , <i>proAB</i> , <i>lacIqZM15</i>	Yanisch-Perron <i>et al.</i> , 1985
512	O1:K1:H7	G.J. Boulnois
20022	O6:K2:H1	K. Jann †
20242	O4:K3:H4	K. Jann
U1-41	O5:K4:H4	K. Jann
2667	O9:K9:H12	K. Jann
21454	O11:K10:H ⁻	K. Jann
21455	O13:K11:H11	K. Jann
21456	O25:K19:H12	K. Jann
A12b	O6:K54:H10	F. Ørskov
B22 ‡	O9:K30:H12	C. Whitfield
2149	O9:K34:H ⁻	K. Jann
Plasmid	Characteristics	Reference
pACYC184	cloning vector	Chang and Cohen, 1978
pUC19	cloning vector	Yanisch-Perron <i>et al.</i> , 1985
M13mp18/19	bacteriophage cloning/sequencing vector	Yanisch-Perron <i>et al.</i> , 1985
pEMBLcos4	cosmid cloning vector	Hadfield, 1987
pGB110	carries the K5 antigen gene cluster	Roberts <i>et al.</i> , 1986
pKT172	carries the K1 antigen gene cluster	Echarti <i>et al.</i> , 1983
pKT274	carries the K1 antigen gene cluster, derived from pKT172	Echarti <i>et al.</i> , 1983
pGB792	carries the P-fimbrial (F14) adhesin genes	High <i>et al.</i> , 1988

†strains from K. Jann have a Freiburg Strain Collection number

‡derivative of the Ørskov K30 test strain, E69

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.

2.2.2 Transformation with Plasmid DNA

Competent cells (100 μl) and up to 25 μl of DNA (in water or TE) to be transformed were mixed and placed on ice for 1 hour. The cells were heat-shocked at 42°C for 3 mins. L-broth (500 μl) was added to the cells which were then incubated for a further one hour at 37°C and plated onto L-agar plates (100 μl per plate) which contained the appropriate antibiotics. The plates were incubated overnight at 37°.

2.2.3 Transformation with Bacteriophage DNA

Competent *E. coli* JM101 cells (the preparation is described in Section 2.2.1) were incubated on ice for 1 hour with the construct DNA and heat-shocked at 42°C for 3 mins. The transformed cells together with 200 μl mid log phase JM101 were added to 3ml molten B-agar (held at 42°C) containing 20 μl 100mM IPTG, and 50 μl X-gal in dimethylformamide. The suspension was mixed and poured immediately onto a B-agar plate, rocked to disperse and once set incubated at 37°C overnight.

2.3 Procedures for DNA Extraction

DNA extraction protocols used the following solutions:

Solution I

50mM glucose
25mM Tris-HCl pH 8.0
10mM EDTA

Solution II

0.2N NaOH
1% SDS

Solution III (100ml)

5M acetate (11.5ml glacial acetic acid)
3M potassium ions (60ml 5M potassium acetate)

2.3.1 Extraction of Chromosomal DNA

The method used to extract chromosomal DNA was based on that described by Saito and Muiira (1963). Bacterial cells from 5ml stationary phase cultures were washed in 10mM NaCl and resuspended in 5ml solution I for 30 mins on ice. SDS was added to a final concentration of 1% and EDTA to 50mM. The preparation was left at room temperature until the solution was clear (typically 20 mins). Sodium perchlorate to a final concentration of 1M and 2 volumes of a chloroform:isoamylalcohol mix (24:1) were added and thoroughly mixed by gentle inversion of the tube. The phases were separated by centrifugation at 20°C for 20 mins at 3000g. The aqueous phase was removed using a wide-bore pasteur pipette and a further 2 volumes of the chloroform:isoamylalcohol mix was added. The aqueous phase was retained as previously and extracted a further three times using phenol:chloroform (Section 2.3.5). Chromosomal DNA was finally retrieved by gently pouring 2 volumes of ethanol (chilled to -20°C) down the side of the tube onto the DNA-containing aqueous phase. DNA precipitated at the interface and was spooled out using the rounded end of a pasteur, resuspended in sterile distilled water and stored at -20°.

2.3.2 Small Scale Extraction of Plasmid DNA

Small scale preparation of plasmid DNA used 1.5ml stationary phase bacterial cultures. Cells were suspended in 100 μ l of solution I for 30 mins on ice. Solution II (200 μ l) was added and the tube was carefully mixed and placed on ice for a further 5 mins. Solution III (150 μ l) was added, mixed and placed on ice for between 10 and 60 mins. The preparation was spun in a bench top minifuge for 5mins and the supernatant retained. Protein was removed by phenol extraction and the plasmid DNA subsequently precipitated using ethanol (as described in Section 2.3.5 but omitting sodium acetate). The final DNA pellet was resuspended in 50 μ l TE containing 20 μ g RNase.

2.3.3 Large Scale Extraction of Plasmid DNA

Stationary phase bacterial cultures (400ml) were used for larger scale preparation of plasmid DNA (Birnboim and Doly, 1979). The cells were collected by centrifugation, resuspended in 10ml of solution I and left to stand on ice for 30 mins. Freshly made solution II (20ml) was mixed in and the whole left on ice for another 10 mins. After the addition of 15ml of ice-cold solution III, the preparation was placed on ice for a further 10 mins. Cell debris was removed from the plasmid preparation by centrifugation at 4°C for 20 mins at 35000g. Isopropyl alcohol (0.6 volumes) was added to the supernatant, mixed and left to stand at room temperature for a minimum of 15 mins. DNA was collected by centrifugation at 4000g for 30 mins at 20°. The DNA pellet was dried *in vacuo* and resuspended in sterile distilled water to a final volume of 17ml. Caesium chloride was added to a final concentration of 1mg/ml

and ethidium bromide to 50 $\mu\text{g}/\text{ml}$. Chromosomal and plasmid DNA was separated at 40000rpm using a Sorvall TV850 rotor in Sorvall OTD 60 centrifuge for 20 hours at 20°. DNA was visualised under UV light and the lower band of plasmid DNA extracted. Ethidium bromide was removed by equilibration with caesium chloride-saturated isopropanol. Caesium chloride was removed by exhaustive dialysis against distilled water at room temperature. Plasmid DNA was stored dissolved in sterile distilled water at -20°.

2.3.4 Extraction of M13mp18/19 Template DNA

The recombinant bacteriophage M13mp18/19 was transformed into JM101 (Section 2.2.3) and white plaques were picked into 5ml of L-broth containing 100 μl of an overnight culture and incubated at 37°C for 5 hours with vigorous aeration. An aliquot of the suspension was decanted into a 1.5ml tube and the cells collected by centrifugation in a minifuge for 1 min. The cell pellet was used to extract replicative form DNA to check the identity of the insert (for details of the plasmid DNA extraction procedure see Section 2.3.2). The supernatant was used to extract template DNA as follows: the supernatant (800 μl) was added to 200 μl of a freshly prepared solution containing 2.5M NaCl, 20% PEG 6000 and incubated at room temperature for 30 mins. The template DNA was collected by centrifugation in a minifuge for 5 mins. The supernatant was discarded, the sample was spun for a further 2 mins and all traces of the supernatant removed. The DNA pellet so produced was resuspended in 100 μl of 1.1M sodium acetate pH 7.0. Template DNA was purified by extraction with an equal volume of phenol:chloroform, followed by chloroform alone (Section 2.3.5). Finally, the DNA was precipitated with ethanol at -20°C (Section 2.3.5), collected by centrifugation and resuspended in 10 μl of TE. Template DNA in 2 μl was visualised by agarose gel electrophoresis (Section 2.4) and 7 μl were typically used in a sequencing reaction (Section 2.9).

2.3.5 Phenol extraction and ethanol precipitation of DNA

One volume of Phenol:chloroform (1:1 w/v) containing hydroxyquinoline and equilibrated in Tris-HCl pH7.5 was added to the DNA sample and mixed carefully (for chloroform extraction one volume of chloroform was added to the DNA sample). The aqueous phase was separated at 20°C for 20mins at 18000g or in a bench top minifuge for 5 mins and retained. For ethanol precipitation sodium acetate to a final concentration of 300mM and 2 volumes of ethanol at -20°C were added. The sample was placed at -20°C or -70°C for a minimum of 30 mins and the DNA collected by centrifugation either at in a benchtop minifuge for 5 mins or at 20°C for 30 mins at 4500g.

2.4 Techniques used in routine DNA manipulation

Restriction endonuclease cleavage of DNA was performed according to the manufacturers recommendations typically in 20 μ l reactions with one unit of enzyme 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% Seakem agarose in TAE buffer (40mM Tris-acetate, 1mM EDTA) with 0.5 μ g/ml ethidium bromide and visualised using a longwave UV transilluminator. DNA fragments over 40kb were separated on a 0.4% agarose gel and those less than 1kb on a 1% gel. Electrophoresis was performed in TAE containing 0.5 μ g/ml ethidium bromide at 140V for 2 hours or at 10V overnight. 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 were: bacteriophage λ DNA either uncut or digested with *Xho*I or 1kb ladder (BRL/GIBCO). λ DNA markers were incubated at 65°C in gel-loading buffer for 10 mins before use.

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 300 μ l TE and subjected to 100V for 30mins in TAE buffer, polarity was reversed for the last 30s. DNA was collected by ethanol precipitation (Section 2.3.5) from the solution surrounding the slice.

2.5 Cosmid Cloning

Genomic libraries were produced by the cosmid cloning method of Collins and Hohn (1978).

2.5.1 Preparation of chromosomal DNA for cloning

Chromosomal DNA was partially digested with the restriction endonuclease *Sau*3A for varying times. Cleavage was stopped by transfer of an aliquot of the pilot reaction mixture to an equal volume of loading buffer at 65°. After a 10 mins incubation at 65°C the DNA fragments were analysed by agarose gel electrophoresis. From this controlled digestion, conditions were optimised for the production of chromosomal fragments of 40-50kb. Chromosomal DNA was then cleaved under optimal conditions with *Sau*3A, the reaction was stopped with phenol:chloroform and the chromosomal

DNA recovered by ethanol precipitation (Section 2.3.5).

2.5.2 Preparation of the cosmid vector for cloning

The cosmid vector, pEMBLcos4, was cleaved with *Pvu*II and dephosphorylated using calf intestinal alkaline phosphatase. The enzyme was inactivated and removed by phenol extraction, the phenol was removed by thoroughly mixing the aqueous phase with an equal volume of ether and retaining the lower phase. That the phosphatase had been successful was confirmed by the failure of the de-phosphorylated, *Pvu*II-cleaved vector DNA to ligate to itself. The vector DNA was then cleaved with the restriction endonuclease *Bam*HI to generate the two arms.

2.5.3 Ligation and packaging of the recombinant DNA

The prepared chromosomal DNA was ethanol precipitated with the cosmid arms in the molar ratio 1:6 of ligatable ends. This was estimated by differential ethidium bromide staining taking the fragment lengths into account. The chromosomal and cosmid DNA were ligated together in a 5 μ l reaction at an approximate DNA concentration of 1 μ g/ μ l. An aliquot of the ligation mixture (1 μ l) was analysed by agarose gel electrophoresis and the remaining DNA was packaged into bacteriophage λ heads by the method of Hohn (1979) using the λ DNA *in vitro* packaging kit produced by Amersham International PLC as recommended in the instructions.

2.5.4 Infection of *E. coli* with the cosmid library

E. coli was grown to mid log phase in the presence of 0.2% maltose, 10mM MgSO₄. The cells were collected by centrifugation and resuspended in an equal volume of 10mM MgSO₄. The prepared cells (100 μ l) were mixed with 100 μ l of the packaged cosmid DNA and allowed to adsorb at 37°C for 20 mins. For a control, prepared LE392 cells were mixed with 100 μ l phage dilution buffer (10mM Tris-HCl pH 7.4, 10mM MgSO₄, 0.01% gelatin) and treated in the same fashion. L-broth (100 μ l) was added and the samples incubated at 37°C for 1 hour to allow expression of the ampicillin resistance gene. The library was then plated onto L-agar (100 μ l per plate) containing ampicillin and incubated at 37°C overnight.

2.6 Transfer of DNA to Nylon Filters by Southern Blotting

DNA was transferred to filters as described by Southern (1975). DNA samples were separated by agarose gel electrophoresis as described above and the gel photographed along side a linear rule. The DNA was de-purinated by soaking the gel in 0.25M HCl for 7 mins. The gel was rinsed briefly in distilled water and placed in denaturing solution (0.5M NaOH, 1.5M NaCl) for 30 mins with occasional shaking. The gel was again rinsed in distilled water and placed in neutralising solution (0.5M Tris-HCl pH7.5, 3M NaCl) for another 30 mins with occasional shaking as before. The gel was rinsed again and placed on six sheets of pre-wetted (20xSSC) Whatman paper (3mm) without trapping any air bubbles. A pre-wetted (10xSSC) sheet of nylon membrane (Hybond-N, Amersham International PLC) was placed on the gel with a pre-wetted sheet (10xSSC) Whatman paper ontop, again taking care to avoid air bubbles. Four sheets of dry Whatman paper were placed above this with a stack of paper towels. Finally a glass plate and a 500g weight were placed on the top. The lower sheets of Whatman paper were regularly soaked with 20xSSC and the paper towels changed. The apparatus was left overnight for the DNA to transfer and then dismantled. The nylon filter was air dried, wrapped in Saran wrap and exposed to UV light from a long wave transilluminator for 5 mins to fix the DNA to the filter. Filters were stored at room temperature in the dark until required for DNA hybridisation (Section 2.8).

2.7 Preparation of filters for colony hybridisation

Bacteria were grown overnight at 37°C on a nylon filter (Hybond-N) which had been placed on the surface of an L-agar plate containing the appropriate antibiotics. Whatman paper (3mm) was placed in a shallow tray and soaked in denaturing solution. The nylon filter was removed from the L-agar plate and placed on the soaked Whatman paper for 5 mins. The filter was then transferred to Whatman paper, this time soaked in neutralising solution for a further 5 mins and air dried. For details of the composition of the solutions see Section 2.6. DNA was fixed to the filter by exposing to longwave UV light from a transilluminator for 5 mins. Cell debris was removed from the surface of the filters by gentle scrubbing in 5x SSC using polymer wool and the filters left to air dry in preparation for DNA hybridisation (Section 2.8).

2.8 DNA Hybridisation Procedures

2.8.1 Production of a radiolabelled DNA probe

Plasmid DNA was cleaved with the appropriate restriction endonucleases and the fragments separated by agarose gel electrophoresis on a 1% low melting point agarose gel (BRL). The required DNA fragment was excised from the gel and added to sterile water (1.5ml water per gram agarose). The sample was placed in a boiling water bath for 7 mins then stored at -20°. Prior to use the sample was boiled for an additional 3 mins and incubated at 37°C for 10-60 mins.

DNA was radiolabelled using random hexanucleotide primers exactly as described by Feinberg and Vogelstein (1983). Nucleotides and hexanucleotides were obtained from Pharmacia and [α -³²P]dCTP from Amersham International PLC.

2.8.2 Hybridisation of DNA immobilised on filters with the probe

Southern blot or colony hybridisation filters were shaken at 65°C in 100ml of pre-hybridisation solution (see below) for 1 hour. This solution was then discarded and replaced by 20ml hybridisation solution (see below) containing the radiolabelled probe DNA which had been boiled for 5 mins before adding. The filter was shaken overnight (or a minimum of 4 hours) at 65°. Hybridisation solution is 3x SSC, 2x Denhardtts (50x is 1% each Ficoll, BSA, polyvinolpyrrolindine), 200 μ g/ml salmon sperm DNA, 0.1% SDS, 6% PEG 6000. Prehybridisation solution is the same except with 5x Denhardtts. Both solutions were stored at -20°C without the salmon sperm DNA. Salmon sperm DNA was sheared by forcing through a narrow gauge syringe needle and denatured by boiling prior to use.

After the hybridisation period the filters were washed twice by shaking in 250ml 2x SSC, 0.1% SDS at 65°C for 15 mins and twice in the final wash solution of choice under the same conditions. The degree of base pair mismatch was estimated using the equation of Howley *et al.* (1979) and SSC concentrations in the final wash were varied to control the degree of hybridisation accordingly: 2x SSC for approximately 70% homology, 0.5x SSC for approximately 85% homology and 0.1x SSC for approximately 95% DNA homology. SDS concentration (0.1%) and temperature (65°C) were not varied. The filters were air dried partially if re-use was planned and completely dried if not.

The filters were wrapped in Saran wrap for autoradiography and placed in a cassette carrying intensifying screens. Kodak X-Omat AR film was exposed to the filters at -70°. Films were developed in an Agfa-Geveart automatic film processing machine.

2.8.3 Removal of bound probe from a filter

The probe was stripped from some filters following autoradiography and the filter was re-used with another probe. The probe was removed according to the manufacturer's instructions: the filters were shaken in 0.4M NaOH at 45°C for 30mins and then transferred to pre-warmed 0.1x SSc, 0.1% SDS, 0.2M Tris-HCl pH 7.5 and shaken at 45°C for a further 30 mins. Successful removal of the probe was gauged using a Guiger counter (full scale deflection was reduced to background).

2.9 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 generate 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 manufacturers was followed using either the universal (-40) primer or oligonucleotide primers synthesised for this purpose. DNA fragments were radiolabelled by incorporating [α -³⁵S]dATP in the extension 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
45 μ l 10% ammonium persulphate
2.5 μ l TEMED

0.5xTBE acrylamide/urea mix

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

40% acrylamide (per litre):

380g acrylamide
20g bisacrylamide
deionised with 50g amberlite per litre

gel solution 2

40ml 0.5x TBE acrylamide/urea mix
180 μ l 10% ammonium persulphate
7.5 μ l TEMED

5xTBE acrylamide/urea mix

per litre:
430g urea
150 ml 10x TBE
150ml40% acrylamide
50g sucrose
50mg bromophenol blue

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, clean gel plates (20cm x 50cm) were taped together separated by 0.4mm spacers. 10ml gel solution 2 followed by all of gel solution 1 were drawn up into a 25ml pipette. Four air bubbles were introduced to form a rough gradient. The liquid was run between the gel plates and the cavity filled with the remainder of gel solution 2. The comb was positioned and the plates clamped along each side. Gels were routinely left to set overnight.

A vertical electrophoresis system was used. Running buffer in the top tank was 0.5x TBE and in 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 at a constant power of 40W and the wells rinsed with running buffer prior to loading. Electrophoresis was performed at 40W for 2 hours 15 mins to visualise the smallest DNA fragments and for increasingly longer periods to separate larger fragments, typically 4 hours, 6 hours or 8 hours.

After electrophoresis the gel plates were prised apart and the gel was soaked in fixing solution (10% methanol, 10% acetic acid) for 10 mins and then rinsed briefly in distilled water. It was transferred to pre-wetted filter paper, covered with Saran wrap and dried under vacuum at 80°. Autoradiography used Dupont Cronex film and took place at room temperature.

Nucleotide sequences were analysed using the University of Wisconsin Genetics Computer Group sequence analysis software (Devereux *et al.*, 1984) on the Leicester University VAX/VMS cluster. The programme FASTA, which uses the algorithm of Pearson and Lipman (1988), was used for sequence comparisons.

2.10 Polymerase Chain Reaction

The polymerase chain reaction (PCR) was performed as described by Saiko *et al.* (1988). Amplification reactions contained 1x amplification buffer, 0.5-5.0 mM MgCl₂, 200µM each dNTP (dATP, dTTP, dCTP, dGTP), 1ng template DNA, 1.0µM of each primer and 2.5 units *Taq* DNA polymerase (Sigma Chemical Company Ltd) in a total volume of 100µl. MgCl₂ concentrations were empirically optimised in the reaction by varying the amounts of a 50mM stock solution added; 2.0mM MgCl₂ was then selected for routine use. 10x amplification buffer contains 500mM KCl, 100mM Tris-HCl pH 8.3, 0.1% gelatin and was stored at -20°. Ultrapure dNTPs were purchased from Pharmacia Biochemicals Inc. Reaction mixtures were briefly vortexed, collected by centrifugation and overlain with 50µl sterile mineral oil prior to amplification. DNA amplification was performed in a Perkin-Elmer Cetus thermal cycler. 30 cycles of

the following conditions were performed:

denaturing step	94°C	1 min
annealing step	55°C	1 min
extension step	72°C	2 min

Following amplification 50 μ l chloroform was added to the reaction tube and mixed by vortexing. The aqueous phase was separated by centrifugation and retained. The PCR product (5 μ l) was analysed by agarose gel electrophoresis.

2.11 Techniques used for Polysaccharide detection

2.11.1 Extraction of polysaccharide

Total polysaccharide was extracted from 10ml stationary phase bacterial cultures (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 1 min in a benchtop minifuge and all traces of acetone removed from the final pellet *in vacuo*. The final pellet was resuspended in 50 μ l 1M MgCl₂ and incubated at 37°C for 2 hours. Cell debris was removed by centrifugation for 5 mins in a benchtop minifuge and the supernatant, a crude polysaccharide extract, retained.

2.11.2 Double immunodiffusion analysis

Double immunodiffusion (Ouchterlony) analysis of polysaccharide extracts was performed using 1% Seakem HGT agarose (FMC Bioproducts) in barbital buffer. To make barbital 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 barbital were poured into a 35mm diameter petri-dish. Once set 3mm diameter holes were cut in the agarose using a punch. Typically, 10 μ l polysaccharide extract or serum was loaded per well.

2.11.3 Serum agar plates

The presence of cell surface polysaccharide was investigated using serum agar plates. Bacteria were grown on L-agar plates containing the appropriate antibiotics and supplemented with 10% serum. A precipitin halo around the colony indicates the presence of extracellular antigen; a negative control was not surrounded by a halo.

2.11.4 Immunochemical detection of polysaccharide

Polysaccharide from whole cells (colony immunoblotting) or polysaccharide extracts (polysaccharide immunoblotting) was detected immunochemically.

2.1.4.1 Preparation of filters for colony immunoblotting

For immuno-blot analysis of whole cells nylon filters (Hybond-N) were placed over fresh bacterial colonies grown on agar plates and incubated for 1 hour at 37°. Alternatively, bacteria were grown directly on nylon filters placed on L-agar. The filters were carefully peeled off the plates and baked for 4 hours at 80°C and then treated for immunochemical detection of polysaccharide. To lyse cells, nylon filters carrying bacteria were exposed to chloroform vapour in a sealed chamber for 30 mins.

2.11.4.2 Preparation of filters for polysaccharide immunoblotting

Polysaccharide extracts (Section 2.11.1) were spotted directly onto nylon filters (Hybond-N) in multiple 1 μ l aliquots (up to 10 μ l in total) which were allowed to dry between applications.

Subsequently, both colony and polysaccharide filters were treated as described below.

Immunoblot filters were washed four times in TN (50mM Tris-HCl, 0.9% NaCl, pH7.5) and then incubated at room temperature for 1 hour in blocking solution (TN containing 3% skimmed milk powder (Beta Lab)). A second incubation in the blocking solution, this time containing primary antibody at the appropriate dilution (often 1:2000), was performed either for 2 hours at room temperature or overnight at 4°. The filters were washed four times in TN and incubated for a final 2 hours at room temperature in blocking solution containing donkey anti-rabbit immunoglobulin conjugated with horseradish peroxidase (Amersham International plc) at 1:500 dilution. The filters were washed four times in TN and the substrate added. To prepare the substrate, 3mg 4-chloro-1-naphthol were dissolved in 10ml methanol and added to 50ml TN with 20 μ l hydrogen peroxide. Once a positive reaction (purple)

had been observed with the controls (approximately 10 mins at room temperature) the filters were briefly rinsed in distilled water and air dried.

2.12 Production of polyclonal antiserum against capsular polysaccharide

2.12.1 Preparation of heat inactivated cells

E. coli 2667 was grown to mid-log phase in L-broth, collected by centrifugation and resuspended in an equal volume of sterile 0.04% glucose. An aliquot was plated onto L-agar at appropriate dilutions to determine the viable counts of the culture. The remaining cells were inactivated at 90°C for 45 mins and checked by plating 100 μ l onto L-agar, failure to grow after an overnight incubation at 37°C was indicative of complete heat inactivation. With reference to the cell count performed immediately before heat inactivation the cells were collected and resuspended in saline (10mM) to a final concentration of 5x10⁸ cells per 500 μ l.

2.12.2 Immunization Procedure

Serum was raised in a 100 day old rabbit. Injection and bleeding of the animal was performed behind closed doors by trained staff. For details of the procedure see Catty and Raykundalia (1988).

A preliminary test bleed (1-2ml) was taken from the rear marginal ear vein prior to immunization. 5x10⁸ heat inactivated bacterial cells in 500 μ l 0.04% glucose were injected with an equal volume of Freund's Complete Adjuvant over three subcutaneous sites. After a period of one month a test bleed was taken and the animal injected with a further 5x10⁸ cells this time with Freund's Incomplete Adjuvant over three sites as before. Blood (20-50ml) was taken after one week and subsequently at one month intervals.

2.12.3 Harvesting serum

Blood was allowed to clot for 1 hour at 37°. The clot was freed from the walls of the container and left to stand overnight at 4°C to retract. Serum was removed from around the clot with a pasteur pipette and any red cells were removed by centrifugation (1800g) at 4°C for 20 mins. Serum was tested for the presence of anti-capsular antibodies by double immunodiffusion (Section 2.11.2).

2.12.4 Absorption of antiserum to remove cross-reactive antibodies

Cells from a mid-log phase culture (5ml) were resuspended in 1ml serum and the suspension rotated either for 1 hour at room temperature or overnight at 4°. The bacteria were then removed from the serum by repeated centrifugation in a minifuge. Sodium azide was added to the serum at a final concentration of 0.1% (w/v) and the serum stored at -20°C in 100µl aliquots for routine use and at -70°C for long term storage.

2.13 Haemagglutination assay

Erythrocytes from whole blood were washed three times in phosphate buffered saline (PBS) and suspended to a final concentration of 3% (v/v). To detect haemagglutination agar-grown bacteria were suspended in 10µl PBS on a glass slide and mixed with an equal volume of the prepared erythrocytes. The slide was gently rocked to mix and agglutination visualised as a granular precipitate not present on the control plates (PBS alone). To test for mannose resistant haemagglutination, bacteria were suspended in PBS supplemented with 50mM mannose.

2.14 Detection of haemolytic activity

Production of haemolysin was detected by plating the bacteria onto L-agar containing 5% PBS-washed sheep red blood cells. A positive reaction was seen as a zone of clearing around a colony after overnight incubation at 37°.

Chapter 3

Cloning and Expression of the K4 antigen gene cluster

3.1 Introduction

Genetic analysis of group II K antigens has been restricted to linear homo- and heteropolymers including K1, K5, K7, K12 and K92. It has been demonstrated that the Kps proteins encoded by regions 1 and 3 of the group II capsule gene cluster function in the expression of these chemically different group II capsules. The means by which the same proteins can handle different polysaccharides is not clear. Recognition of chemically different polysaccharides by the Kps transport proteins could be mediated by common components, for instance, a terminal KDO residue or phosphatidic acid. Alternatively, polymer size and conformation may be important. The majority of group II polymers for which the structure has been determined are linear molecules (Jann and Jann, 1990) and in this respect those that have been subjected to a genetical analysis are typical. However, two substituted capsular polysaccharide have been described (K4 and K52) which have been placed in group II (Hofmann *et al.*, 1985b; Rodriguez *et al.*, 1988). Despite the apparent versatility of the export proteins, it is not known whether these functions are capable of handling substituted polysaccharides or whether an alternative system operates in the expression of these K antigens.

The K4 capsular polysaccharide comprises a backbone of β -D-glucuronic acid - (1,4)- β -N-acetylgalactosamine - (1- which is substituted at glucuronic with fructose (Rodriguez *et al.*, 1988). The K4 backbone is identical with the sugar repeat structure of chondroitin, which in its sulphated form is a major component of the extracellular matrix of mammalian tissue. The fructose substituent is the immunodominant residue of the K4 antigen (Rodriguez *et al.*, 1988). At low pH (pH 5) up to 50% of the fructose substituents are lost from the polymer and its reaction with antiserum raised against the K4 polysaccharide is reduced (Rodriguez *et al.*, 1988). Extreme conditions are necessary for complete de-fructosylation yielding the chondroitin-like

linear backbone and this could probably not occur under physiological conditions. Therefore, although the K4 and chondroitin sugar backbones are identical, the K4 antigen itself is immunogenic. In contrast, the K1 and K5 polysaccharides which are identical with mammalian oligosaccharides are poorly immunogenic. Partly as a consequence of this, strains expressing the K1 and K5 capsule types are often associated with disease.

Genetic analysis of K4 antigen expression might reveal whether the group II capsule determinants are capable of handling substituted polysaccharides and may shed light on the mechanisms involved in the biosynthesis of branched polymers.

3.2 Results

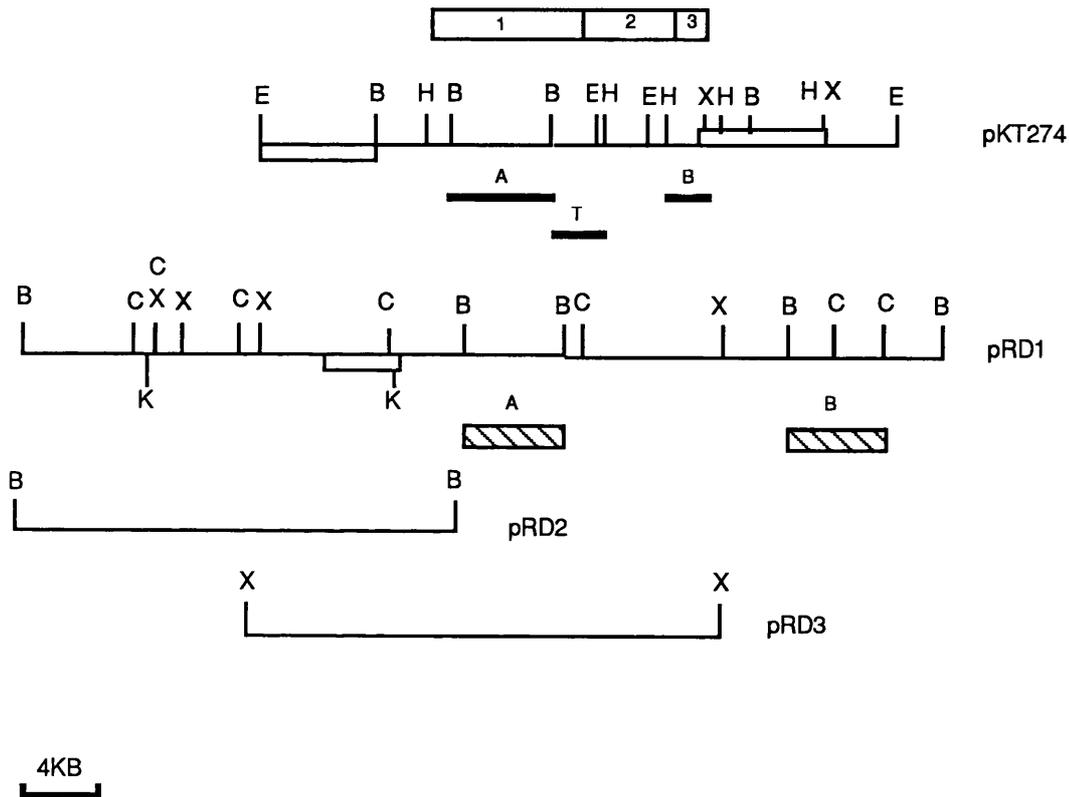
3.2.1 Southern Blot analysis of K4 chromosomal DNA

Southern blot analysis was used in the first instance to determine whether the group II capsule determinants (regions 1 and 3) were present on the chromosome of an *E. coli* strain expressing the K4 antigen. *E. coli* U1-41 (O5:K4:H4) was selected for study as this was the strain used to determine the structure of the polysaccharide and to produce the K4 antiserum (Rodriguez *et al.*, 1988). DNA probes were taken from pKT274 (Figure 3.1) which carries the cloned K1 antigen gene cluster (Echarti *et al.*, 1983). Probe A is a 5.2 kb *Bam*HI fragment which is present within region 1 of the K1 genes and a homologous fragment of the same size exists in region 1 of the K5, K92 and K12 antigen gene clusters (Roberts *et al.*, 1986). Probe B is a 2.7kb *Hind*III-*Xho*I fragment which contains region 3 of the K1 gene cluster and the end of Tn5 (Figure 3.1). Neither probe is homologous to sequences in the unencapsulated laboratory strain LE392 (Roberts *et al.*, 1986). Probes A and B were used in Southern blot analyses of chromosomal DNA extracted from *E. coli* U1-41 and hybridised to fragments under stringent conditions which require a minimum of 95% DNA homology (0.1x SSC). Probe A hybridised to a 5.2kb *Bam*HI chromosomal DNA fragment and probe B to an 8kb *Bam*HI fragment in the *E. coli* K4 genome. The results of these experiments are not shown as similar blots are illustrated in Chapter 5 (Figure 5.2bc). Thus the common group II capsule genes are present on the chromosome of a K4 antigen expressing strain and it seems likely that they are involved in capsule expression.

3.2.2 Isolation of the K4 antigen gene cluster

The demonstration of *kps* regions 1 and 3 on the chromosome of a K4 antigen expressing strain does not prove their involvement in K4 capsule expression, nor does it show that K4 capsule genes have the same organisation as others. It was decided to clone the K4 antigen biosynthesis genes to resolve these problems.

Figure 3.1: Physical maps of the pKT274 and pRD1



Physical maps of the recombinant plasmids pKT274 and pRD1 carrying the biosynthesis genes for the K1 and K4 antigens respectively. Beneath pRD1 are the subclones derived from it. The labelled boxes 1 through 3 above pKT274 refer to functional gene blocks involved in the production of the K1 antigen. The open boxes above and below the line refer to vector and Tn5 sequences respectively. The three DNA probes taken from pKT274 are represented by the thick lines labelled A, B and T. The hashed boxes labelled A and B refer to the smallest restriction fragments in pRD1 homologous to probes A and B respectively. Enzyme target sites: B, *Bam*HI; C, *Cla*I; E, *Eco*RI; H, *Hind*III; K *Kpn*I; X, *Xho*I.

3.2.2.1 Production of a K4 cosmid library

The K1 antigen gene cluster spans 17kb. To maximise the chances of cloning the K4 antigen gene cluster, which may be larger than that of K1, it was decided to use a cosmid vector which allows insert sizes of 34-48kb and is compatible with the highly efficient bacteriophage λ packaging system for library construction. The vector of choice was pEMBLcos4 as it has multiple *cos* sites and packaging is dependent on ligatable cosmid arms such that the production of cosmid concatemers, a problem often associated with cosmid cloning, is reduced.

High molecular weight chromosomal DNA from *E. coli* U1-41 was cleaved with the restriction endonuclease *Sau3A* to generate fragments of 35-45kb which were analysed by agarose gel electrophoresis (Figure 3.2a). This confirmed that the chromosomal DNA fragments were within the size range suitable for cosmid cloning. The cleaved chromosomal DNA and prepared cosmid DNA were ligated together and an aliquot of the ligation mix analysed by agarose gel electrophoresis (Figure 3.2b). The cosmid arms had clearly ligated together in each of the three possible combinations and presumably also ligated with the chromosomal DNA. The ligated chromosomal and cosmid DNA was packaged *in vitro* into bacteriophage λ heads and used to infect *E. coli* LE392. This generated 1300 recombinants.

3.2.2.2 Screening the library by colony hybridisation

The U1-41 cosmid library was screened by colony hybridisation with the same probes A and B as used above (Section 3.2.1). This approach was chosen in favour of immunodetection due to a limited supply of K4 antiserum. It was based on two assumptions: that the K4 capsule genes are clustered and that they include both regions 1 and 3, such that a recombinant plasmid with homology to both probes would also carry a K4-specific region 2 element and direct the expression of the K4 antigen in LE392. Firstly, colony hybridisation with probe A identified thirteen recombinants containing sequences homologous to region 1 of the K1 capsule gene cluster (Figure 3.2c). These thirteen were screened again by colony hybridisation this time using probe B. One recombinant was identified with homology to both probes A and B. This recombinant was selected for further study and the plasmid present named pRD1.

3.2.2.3 Immunological analysis of LE392 (pRD1)

To determine whether pRD1 carries the K4 antigen biosynthesis genes, polysaccharide was extracted from *E. coli* LE392 (pRD1), LE392 alone and from the K4 wild type strain, U1-41 grown in L-broth. The polysaccharide preparations were used in double immunodiffusion analysis (Ouchterlony) using the polyclonal antiserum raised against the K4 antigen (Rodriguez *et al.*, 1988) (kindly donated by Prof Jann). A

Figure 3.2: Production and screening of the U1-41 cosmid library

(a) Chromosomal DNA from *E. coli* U1-41 was cleaved with 0.1 units of the restriction enzyme *Sau3A* at 37°C for 60s. The gel was loaded as follows: lane 1 molecular weight markers of 12, 11, 10, 9kb; lane 2, uncut lambda; lanes 3 and 6, lambda cut with *XhoI*; lane 4, uncut U1-41 chromosomal DNA; lane 5, U1-41 chromosomal DNA cut with *Sau3A*. Uncut lambda is 49.7kb and the lambda *XhoI* fragments are 34.2 and 15.5 kb.

(b) An ethidium bromide stained gel showing the ligated chromosomal and cosmid DNA. The gel was run at high voltage and the separation is poor. The bands correspond to (from the bottom): two small cosmid arms ligated, small and large cosmid arms ligated, two large cosmid arms ligated, chromosomal and cosmid DNA ligated. Molecular weight markers (kb): 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1.6, 1.

(c) Colony hybridisation analysis of part of the U1-41 library using radiolabelled probe A. The spots mark the site of LE392(pKT274) colonies which were used as a positive control and the lines to LE392 harbouring recombinant cosmids with homology to the probe.

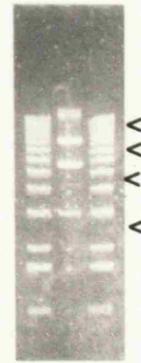
(d) A sketch illustrating a double immunodiffusion analysis of the K4 wild type strain (U1-41), LE392 and LE392 harbouring pRD1. The cells were grown overnight in different medium as indicated and polysaccharide extracted from the cell culture. Polyclonal antiserum (10 μ l) against the K4 polysaccharide was placed in the centre well and 10 μ l polysaccharide extracts from each of the cultures in the outer wells as follows: well 1, LE392 L-broth; 2, U1-41 L-broth; 3, LE392(pRD1) L-broth; 4, LE392 D.O. medium; 5, U1-41 D.O. medium; 6, LE392(pRD1) D.O. medium.

Figure 3.2

a)



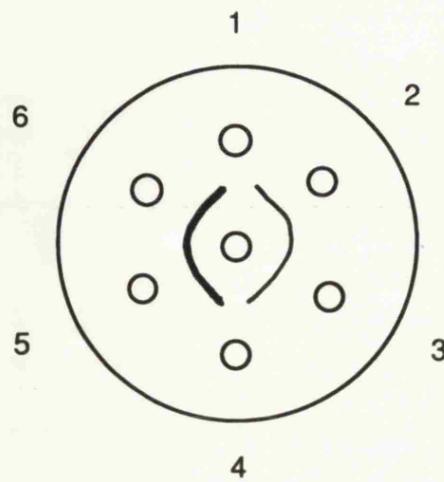
b)



c)



d)



weak but discernable precipitin line of identity was seen between both the wild type and LE392(pRD1) polysaccharide extracts and the antiserum but not with LE392 alone. Since the fructose substitution is the major antigenic determinant of the K4 polysaccharide this reaction indicates that LE392(pRD1) produces the K4 polysaccharide. Bacterial growth in L-broth is associated with a decrease in pH (Boyd, 1988). Thus the weak reaction of the wild type and LE392(pRD1) polysaccharide extracts with the antiserum raised against the K4 antigen could be because the K4 polysaccharide has lost up to 50% of its fructose substitutions in the culture and has a correspondingly reduced antigenicity. D.O. medium is buffered and de-fructosylation should be reduced. It was found that the intensity of the precipitin line was increased when cells were grown in D.O. medium. This indicates that the immunoreactive material produced by LE392(pRD1) is the substituted K4 polysaccharide. The precipitin lines did not photograph well and are sketched in Figure 3.2d.

Extraction of polysaccharide with $MgCl_2$ from bacterial cells does not reveal the cellular location of the polymer. Therefore, LE392(pRD1) was grown on serum agar plates containing polyclonal K4 antiserum. Single colonies were surrounded by a precipitin halo on the serum agar plate as were those of U1-41 but not LE392. This is consistent with the presence of extracellular immunoreactive material. It was concluded that pRD1 confers the ability to express cell surface K4 antigen on *E. coli* LE392.

3.2.3 Analysis of the cloned K4 antigen gene cluster

3.2.3.1 Restriction endonuclease analysis of pRD1

In an attempt to isolate a smaller recombinant plasmid which still contained the K4 biosynthesis genes, pRD1 was cleaved with several different restriction endonucleases. It was noted that cleavage with either *Bam*HI or *Xho*I yielded large DNA fragments and that these might contain the genes of interest. pRD1 was therefore cleaved with these enzymes individually and ligated to itself. In this way the subclones pRD2 and pRD3 which contain the cosmid vector were generated (Figure 3.1). Neither LE392(pRD2) nor LE392(pRD3) produced K4 polysaccharide detectable by double immunodiffusion analysis. These plasmids did, however, prove useful in the construction of a restriction endonuclease cleavage map of pRD1.

pRD1 was cleaved with a variety of restriction endonucleases and those with infrequent target sites chosen for map construction. These restriction endonucleases were used singly and in combinations to cleave pRD1, pRD2 and pRD3. By summing the sizes of the restriction fragments the total size of pRD1 was calculated to be 48.6kb. The restriction enzyme cleavage data was used to generate a physical map of pRD1 (Figure 3.3, Table 3.1) Since the vector must lie within the overlapping sequences of pRD2 and pRD3 its exact position to the left of region 1 was easily determined on the basis of the single *Cla*I and *Kpn*I site in both the vector and the overlap (Figure 3.1).

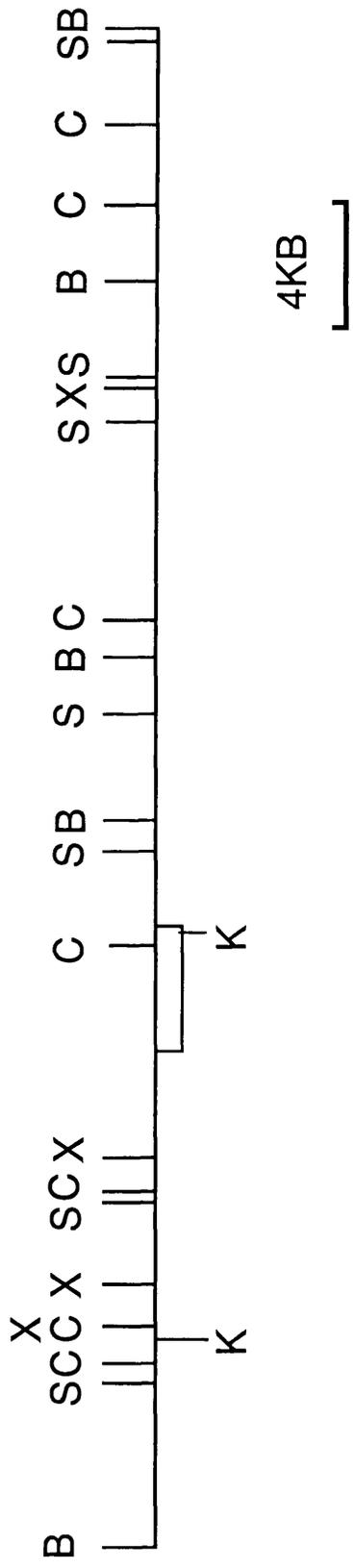
Figure 3.3: Physical map of pRD1

Physical map of the recombinant plasmid pRD1 which contains the K4 antigen gene cluster. The open box below the line refers to the vector sequence (pEMBLcos4). Restriction enzyme target sites: B, *Bam*HI ; C, *Cla*I ; K *Kpn*I; S, *Sma*I; X, *Xho*I.

Table 3.1: DNA fragment sizes in kilobases generated by cleavage of pRD1 with the given restriction endonucleases

<i>Bam</i> HI	<i>Cla</i> I	<i>Kpn</i> I	<i>Sma</i> I	<i>Xho</i> I
23.3	13.3	34.7	11.3	24.9
12.1	10.6	13.9	10.8	17.3
8.0	7.9		9.8	5.0
5.2	7.7		6.8	1.4
	5.5		4.4	
	2.4		4.2	

plasmid pRD1 - encoding the K4 capsule gene cluster



3.2.3.2 Localisation of the K4 antigen genes on pRD1 by Southern blot analysis

The approximate positions of regions 1 and 3 in pRD1 were already apparent as probes A and B hybridised to 5 and 8kb *Bam*HI fragments in the K4 chromosomal DNA respectively (Section 3.2.1). To determine more precisely the extent of the *kps* genes on pRD1 and to confirm the restriction map, Southern blot analysis of pRD1 was performed. Three identical Southern blot filters were prepared in which pRD1 was cleaved with each of the five mapped enzymes (Figure 3.4a). Two filters were hybridised with probes A and B respectively (Figure 3.1). An additional DNA probe (termed probe T) was taken from the K1 antigen gene cluster. This was a 2.4kb *Bam*HI-*Eco*RI fragment which carries the 3' end of the last region 1 gene (*kps*S) and extends into region 2 of the K1 genes (Figure 3.1). The filters were washed with 0.1% SSC. The results with all three radiolabelled DNA probes are shown in Figure 3.4. Fragment sizes were determined with reference to the physical map of pRD1 (Figure 3.3). The region 1 probe (probe A) hybridised to DNA fragments of 5.2 kb (*Bam*HI), 10.6kb (*Cla*I), 34.7kb (*Kpn*I), 4.2kb and 9.8 kb (*Sma*I), and 24.9kb (*Xho*I). The adjacent probe (probe T) hybridised to DNA fragments of 12.1kb (*Bam*HI), 10.6kb (*Cla*I), 34.7kb (*Kpn*I), 9.8kb (*Sma*I) and 24.9kb (*Xho*I). The region 3 probe (probe B) hybridised to DNA restriction fragments of 8.0kb (*Bam*HI), 13.3kb and 2.4kb (*Cla*I), 34.7kb (*Kpn*I), 11.3kb (*Sma*I) and 17.3kb (*Xho*I). By comparing the restriction fragments which hybridise to each of the probes with the restriction endonuclease cleavage map of pRD1 the extent and position of regions 1 and 3 within pRD1 were deduced. Region 1 spans the 5.2kb *Bam*HI fragment and extends into the 1.6kb *Bam*HI-*Cla*I fragment to the right. It can only be assumed that region 1 is the same size as in other group II capsule gene clusters also and extends to the right of the 5.2kb *Bam*HI fragment accordingly. Region 3 is contained within the 8.0kb *Bam*HI fragment but its exact position is unclear. Probe B hybridised to the 2.4kb and the 13.3kb *Cla*I fragments. Hence, region 3 spans the left-most *Cla*I site in the 8.0kb *Bam*HI fragment (Figure 3.1). There is a *Cla*I site in region 3 of the K5 antigen gene cluster (Figure 5.5) and if this is a conserved site then only one third of region 3 of the K4 antigen gene cluster would lie within the 2.4kb *Cla*I fragment of pRD1. The proposed positions of regions 1 and 3 on pRD1 are marked in Figure 4.1 and this gives the gene cluster a similar appearance to others encoding group II capsule genes. It is reasonable to suppose that the K4 capsule gene cluster is not exceptional and that the stretch of DNA which lies between regions 1 and 3 in pRD1 corresponds to region 2.

3.3 Discussion

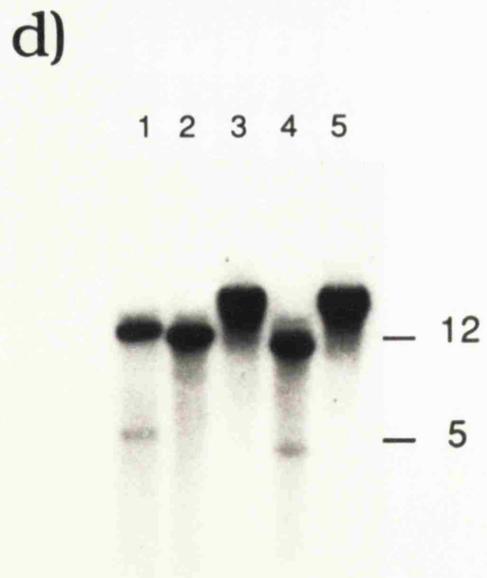
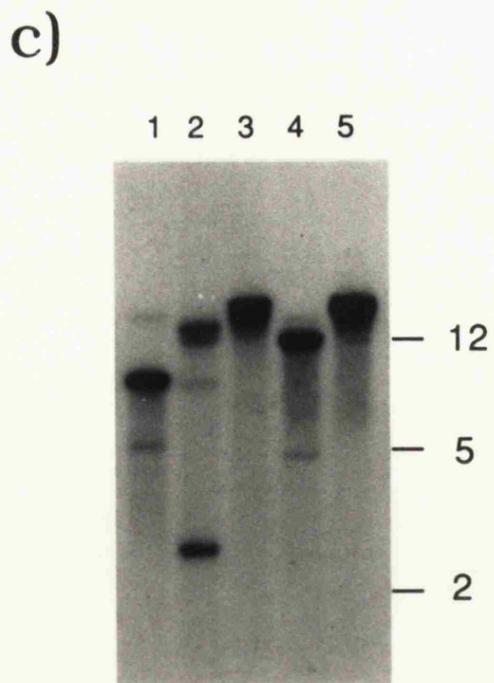
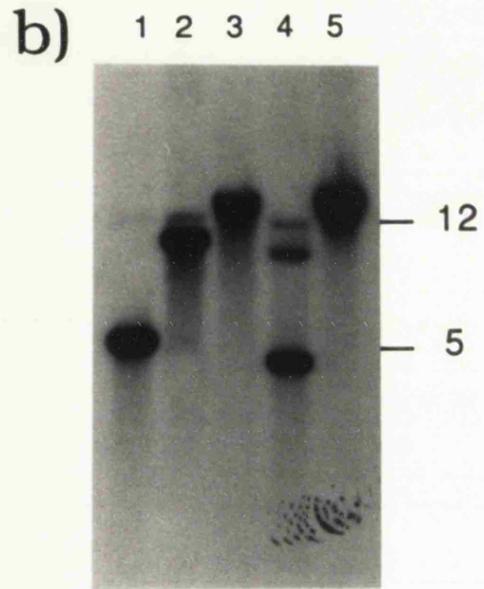
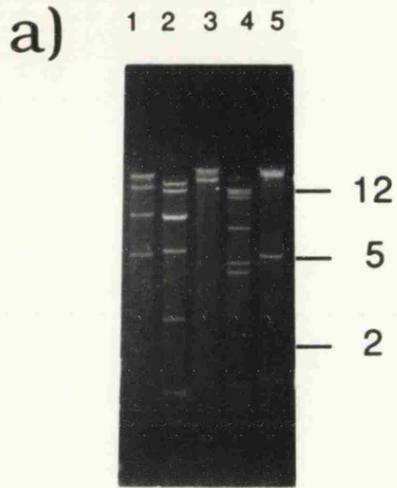
The conserved group II capsule genes, *kps*, were shown to be present on the chromosome of a K4 capsule producing strain (*E. coli* U1-41) and presumed to function in K4 antigen biogenesis. A cosmid library was produced and from it a 48.6kb recombinant plasmid containing U1-41 genomic DNA was isolated by virtue of its homology to

Figure 3.4: Restriction endonuclease and Southern blot analysis of pRD1

(a) Ethidium bromide-stained agarose gel showing pRD1 cleaved with five different restriction enzymes. Three gels identical to the one illustrated were used in the Southern blot analyses shown in (b), (c) and (d). Plasmid DNA was cleaved with the following enzymes: lane 1, *Bam*HI; 2, *Cla*I; 3, *Kpn*I; 4, *Sma*I; 5, *Xho*I. Fragment sizes in kilobases are marked on the left hand side.

(b-d) Three identical agarose gels (Figure 3.4a) were used for Southern blot analysis of pRD1 with (b) Probe A, (c) Probe B and (d) Probe T. Faint high molecular weight bands correspond to uncut plasmid DNA. Fragment sizes in kilobases are marked on the left hand side. pRD1 was cleaved with the following restriction endonucleases: lane 1, *Bam*HI; 2, *Cla*I; 3, *Kpn*I; 4, *Sma*I; 5, *Xho*I.

Figure 3.4



regions 1 and 3 of the K1 capsule gene cluster and was termed pRD1. pRD1 confers the ability to produce extracellular K4 polysaccharide on LE392. This was demonstrated using a polyclonal antiserum raised against K4 polysaccharide in serum agar plates and in double immunodiffusion analysis of cell extracts; in both tests a precipitin line was seen with LE392(pRD1) and U1-41 but not with LE392 alone. Since the furanofructose is the immunodominant residue of the K4 polysaccharide and the antiserum reacts only very weakly with the non-fructosylated backbone (Rodriguez *et al.*, 1988), LE392(pRD1) probably produces fructose substituted K4 polysaccharide. A more intense precipitin reaction was observed in double immunodiffusion analysis of polysaccharide extracts from cells that were grown in buffered medium and this is further evidence that the polysaccharide produced by the recombinant contains the fructose substituent. This can be accounted for by the increased stability of the pH labile, immunodominant fructose residue, if cross reaction had been due to the backbone alone this increase would not be predicted. A complete chemical analysis could confirm the identity of the antigen produced by LE392(pRD1) as the substituted K4 polysaccharide, such an analysis was beyond the scope of this study. A precipitin halo around LE392(pRD1) grown on serum agar plates is evidence that the K4 polysaccharide produced by LE392(pRD1) is extracellular although it is possible that this reaction was with intracellular material released by cell lysis. In the absence of additional information provided by, for example, a K4-specific bacteriophage or immuno-electron microscopy it can only be assumed that this K4 material forms a discrete capsule around the cell as in the wild type strain *E. coli* U1-41.

It is likely that pRD1 itself encodes all the functions necessary for the expression of the fructosylated K4 capsule although the possibility that some capsule functions are encoded on the chromosome of LE392 and complement those on pRD1 cannot be excluded. However, fructose is a rare component of bacterial polysaccharides and it therefore seems likely that enzymes for fructosylation of the K4 chondroitin backbone are located within pRD1. The organisation of the K4 genes carried by pRD1 was determined by restriction mapping and Southern blot analysis using probes from regions 1 and 3 of the K1 antigen gene cluster. It appears that the K4 capsule genes are organised in a similar manner to that demonstrated for the *E. coli* K1, K5, K7, K12 and K92 antigens, with a central fragment flanked by the common regions 1 and 3 (Roberts *et al.*, 1986). The presence and relative positions of region 1 and 3 on pRD1 infer that the 14kb region between them is the K antigen-specific region 2 which encodes for synthesis of the substituted K4 polymer. In the other group II K antigen gene clusters analysed capsule determinants are confined by regions 1 and 3 and this is probably the case for the K4 antigen genes. The unlikely possibility that synthesis enzymes are encoded in the 18kb of DNA flanking region 3 of the K4 capsule genes on pRD1 cannot be rigorously excluded. Further analysis of pRD1 may exclude these possibilities and at the same time may shed light on the mechanism of synthesis of the substituted heteropolymer, K4.

In both chondroitin sulphate and the fructose substituted *H. influenzae* type e capsular polysaccharide it is likely that the linear backbone is synthesized first and subsequently modified by sulphatisation or fructosylation respectively (Branefors-Helander *et al.*, 1981; Delfert and Conrad, 1985). If synthesis of K4 polysaccharide also pro-

ceeds by this mechanism then it is reasonable to suppose that a mutant defective in the fructosylation process would synthesise the linear non-substituted backbone. The linear non-substituted backbone acts as a substrate for chondroitinase (Rodriguez *et al.*, 1988) and polysaccharide, regardless of structure can be detected by CTAB precipitation (Jann and Jann, 1983). Alternatively, K4 biosynthesis may proceed by the sequential addition of each sugar component either onto the growing chain or into repeat units which are subsequently polymerised. In this case, a defect in fructosylation, or indeed any of the synthetic steps, may completely abolish polysaccharide production. These are just two alternatives and the situation may not be this simple. For instance, *E. coli* O9 polysaccharide biosynthesis proceeds by the sequential addition of monomers to the growing chain (*rfe*-dependent) (Sugiyama *et al.*, 1991). A deletion in the biosynthesis determinants (*rfb*) did not abolish polysaccharide synthesis but resulted in the production of an O antigen antigenically distinct from O9 (Kido *et al.*, 1989). An analysis of region 2 of the K4 antigen gene cluster in pRD1 may prove useful in deciphering the synthetic mechanism of this substituted polysaccharide. Such a study, which is likely to be time consuming was not attempted.

The group II capsule gene-encoded Kps proteins appear to export the substituted K4 polysaccharide and not to be restricted in their function to linear molecules. This suggests that chemically diverse polysaccharides can be exported onto the cell surface by common products encoded by regions 1 and 3. Mechanisms involved in group II capsule export must be capable of recognising the different K1 and K4 polysaccharides. The results of this study suggest that the presence or absence of substituents is not important for export of the capsular antigen onto the cell surface. However, it is interesting to note that the heavily substituted complex capsular polysaccharides produced by *E. coli* such as K30 (a disaccharide repeat unit with a disaccharide substitution) are restricted to group I. Whether K antigen-specific proteins are produced which bind to the polysaccharide and are in turn bound by the common transport proteins is not known. Alternative motifs for the recognition of chemically different group II polymers are the KDO and phosphatidic acid residues found at the reducing end of all group II polysaccharides so far studied.

Those group II capsule gene clusters which have been studied, ranging from the simple K1 capsule to the substituted K4 capsule share a common genetic organisation. A comparison of the group II capsule gene clusters which have been cloned shows that the size of region 2 is variable and that this size is broadly a reflection of polysaccharide complexity. For instance, K1 is a linear homopolymer and has a region 2 of only 5kb; K7 is a linear heteropolymer and its region 2 is 7kb. K4 is a substituted heteropolymer with three sugar components and is the most complex group II polysaccharide for which the genes have been cloned. At 14kb region 2 of the K4 capsule gene cluster is the largest identified to date. Nucleotide sequence analysis, however, indicates that K1 region 2 encodes more proteins than that of K5. The K1, K4 and K5 capsule gene clusters are aligned in Figure 4.1 (Chapter 4) and the relationship between region 2 elements is clear. This alignment is consistent with the concept of the cassette theory in which region 1 and 3 are preserved and flank one of a selection of very different region 2 cassettes. This idea is developed in the following chapter.

Chapter 4

The variability of group II capsule gene clusters

4.1 Introduction

A large number of capsular polysaccharides which differ in the composition of the oligosaccharide repeat unit are produced by *E. coli*. The group II capsule gene clusters which have been studied to date appear to be remarkably similar. Regions 1 and 3 are homologous between different group II capsule gene clusters. Although the overall organisation of the cluster is the same, the central region (region 2) is very variable and appears as a 'gene cassette' slotted in between the common regions 1 and 3 (Boulnois and Jann, 1989). Region 2 encodes for the synthesis of the polysaccharide itself and structural variability of group II capsules is thought to be entirely dependent on the variability of the region 2 cassette within the gene cluster. How the very similar group II capsule gene clusters with variable central regions may have arisen is an intriguing problem.

Two basic mechanisms may have been involved in the generation of the variable group II capsule gene cluster but the two are not mutually exclusive. First, region 2 could have been inherited as a complete functional block into the *kps* locus from another source, such as the LPS biosynthesis genes. Secondly, a series of insertions, deletions and mutations within the region 2 of the gene cluster may have created unique sets of genes encoding for the production of novel polysaccharides. In the former case, the region 2 cassette would probably have inserted at a specific point between regions 1 and 3. The DNA at the junctions between the common flanking sequences and the unique region 2 may give clues as to the nature of any such event. For instance, if region 2 was inherited in a transpositional event, as was suggested for the entire *H. influenzae* capsule gene cluster (Kroll *et al.*, 1991), then insertion sequences may now flank region 2. In the latter case the junctions between the regions and sequences within region 2 itself may have been the site of recombination between resident and incoming DNA. Whatever the mechanism for the generation of the group II capsule

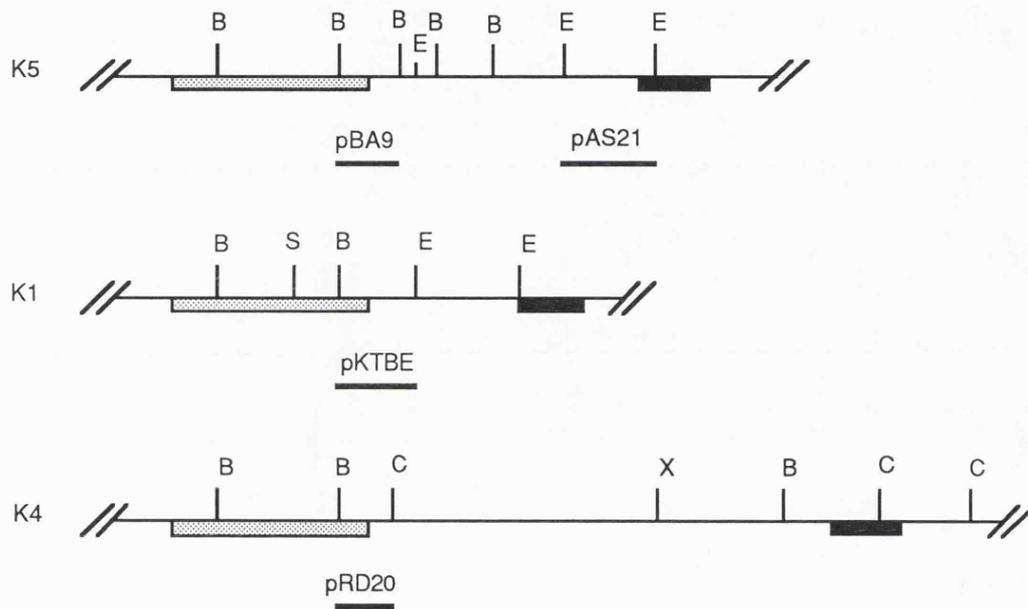
gene cluster the junctions between region 2 and the regions 1 and 3 may have been important.

The definitions of the three functional regions of the group II capsule gene cluster were based on DNA homology between different clusters, functional homology (assessed by complementation studies using entire gene blocks) and the phenotype of mutants (Roberts *et al.*, 1988). There is DNA homology in regions 1 and 3 between different group II capsule gene clusters and the proteins encoded appear to be similar. Whether the gene products of regions 1 and 3 from different capsule gene clusters are indeed identical is not clear. Region 2 is unique for each K antigen gene cluster its products are necessary for polysaccharide synthesis. The precise extent of each region has not been defined and as a consequence the junctions between regions are difficult to describe.

The complete nucleotide sequence of the entire K5 capsule gene cluster has been now determined (Pazzani, C., Smith, A.N., Boulnois, G.J. and Roberts, I.S., unpublished data) and allows the junction between regions to be analysed in detail. The *kpsS* gene is at the 3' end of region 1, adjacent to region 2 (Figure 1.2). Next to *kpsS*, in region 2 of the K5 capsule gene cluster, is a gene which encodes a 44kD protein (Figure 1.2). These two genes are transcribed towards each other and the stop codons of the two genes are 339bp apart. Between regions 2 and 3 of the K5 capsule gene cluster a 987bp, non-coding, sequence separates the stop codon of *kpsT* from the initiation codon of the region 2 gene which encodes a 27kD protein. These non-coding sequences between the regions could be important in rearrangements of region 2 but analysis of these sequences to date has not been informative. For the purpose of this study, the term 'boundary region' or 'boundary' will refer to the non-coding sequences which lie at either end of region 2 between region 2, and region 1 and 3 genes. The sequence of the boundary regions in group II capsule gene clusters other than K5 is not known.

The restriction endonuclease cleavage maps of the group II K antigen gene clusters so far analysed can be aligned on the basis of conserved sites (Roberts *et al.*, 1988). In the K1, K4 and K5 capsule gene clusters a 5.2kb *Bam*HI fragment lies within region 1 (Figure 4.1). Adjacent to this in the K5 capsule gene cluster (carried on pGB110) is a 1.6kb *Bam*HI fragment which has been cloned using the vector pUC19 and the resulting plasmid termed pBA9 (C. Pazzani, unpublished). pBA9 contains the 3' end of both *kpsS* (region 1) and the region 2 gene which encodes the 44kD protein with the 339bp boundary region between the stop codons (Figure 4.3). Region 1 is homologous between different group II capsule gene clusters (Roberts *et al.*, 1988) and is thought to encode homologous proteins, it is therefore reasonable to suppose that DNA fragments to the right of the 5.2kb *Bam*HI fragment in region 1 of the K1 and K4 capsule genes will also contain the 3' end of *kpsS* and possibly a boundary region and part of a respective region 2 gene. In the K1 antigen gene cluster (carried on pKT274) a 2.1kb *Bam*HI-*Eco*RI fragment lies adjacent to the 5.2kb *Bam*HI fragment and in the K4 capsule gene cluster (carried on pRD1) a 1.8kb *Bam*HI-*Cla*I fragment occupies this position (Figure 4.1). In region 2 of the K1 antigen gene cluster *neuS* (sialyl transferase) is adjacent to, and transcribed towards region 1 (Figure 1.2). The

Figure 4.1: K5, K1 and K4 Capsule Gene Clusters



Linear maps of the capsule gene clusters of K5 (pGB110) K1 (pKT172) and K4 (pRD1) aligned on the basis of common region 1 restriction enzyme fragments. The dotted boxes represent region 1 and the black boxes region 3 in each of the capsule gene clusters shown. The position of region 3 in the K4 capsule gene cluster is estimated (see Section 3.2.3.2). The lines labelled pBA9, pKTBE and pRD20 refer to the equivalent fragments in each cluster which span from region 1 into region 2. These fragments form the inserts of the plasmids with the same names. The line labelled pAS21 refers to the DNA fragment from the K5 capsule gene cluster which spans from region 2 into region 3 which forms the insert of a plasmid with the same name (Smith, A.N., unpublished). Restriction enzyme target sites: B, *Bam*HI; C, *Cla*I; E, *Eco*RI; S, *Sal*I; X, *Xho*I.

arrangement of region 2 of the K4 capsule genes is not known.

If the hypothesis that region 2 rearrangement, by whatever mechanism, involves the boundary regions between the variable region 2 and the common regions 1 and 3 is true, then the boundaries between regions may exhibit common features either within or between capsule gene clusters. In this study the K1, K4 and K5 capsule gene clusters were analysed and compared in the light of this hypothesis and in doing so revealed variability within the region 1 gene *kpsS*.

4.2 Results

4.2.1 PCR analysis of the region 1 - 2 and region 2 - 3 boundaries

As a preliminary examination of the boundary regions at both ends of region 2 from different capsule gene clusters it was decided to use the polymerase chain reaction (PCR). When these experiments were performed the nucleotide sequence of the K5 capsule gene cluster around the region 2 - 3 boundary was not available. Using the known DNA sequence of the region 1 - 2 boundary in the K5 antigen genes PCR primers were chosen. If there were DNA homology these primers could be used to amplify similar DNA sequences from the region 1 - 2 boundary of the capsule genes in pRD1, pKT172 and pGB110 itself or from any other site within the cluster or in the flanking DNA around it. Sequence analysis and comparison of the PCR product could then be performed.

The DNA sequence of the 1.6kb *Bam*HI fragment cloned in pBA9 was analysed. Beyond the stop codons of the two genes imperfect inverted repeats which may serve as transcription termination loops were identified (Figure 4.2) using the FOLD and STEMLOOP programmes. These were not good sites to choose primers and were avoided. Two DNA primers were chosen which would bind beyond the potential loops and serve to amplify a 213bp junction fragment (Figure 4.2). Primers were checked for the formation of internal stemloops and other binding sites elsewhere on pBA9. Restriction endonuclease target sites were incorporated into the 5' end of each primer to facilitate cloning the PCR product should this become necessary. A *Bam*HI and *Sal*I restriction endonuclease target sites were introduced into one primer (PR1 below) and *Pst*I and *Bam*HI into the second (PR2 below). The sequence of the two primers, excluding the restriction endonuclease sites were as follows:

PR1 5'- CTTAGATATCTTTGG -3'

PR2 5'- ACTATTAAGAGAAGG -3'

The primers were initially used to amplify DNA from the K5 capsule gene cluster using three different templates: pBA9 linearised with *Bam*HI, pGB110 linearised with *Sma*I and pAS21 linearised with *Sal*I. pAS21 contains an *Eco*RI fragment from pGB110 which spans the region 2 - 3 boundary of the K5 capsule genes and was produced by A.N. Smith (Figure 4.1). A single fragment of the predicted size was synthesised using PCR from the pBA9 and pGB110 templates but not from pAS21 (Figure 4.2). The unlikely possibility that product of PCR using the pGB110 template was a dimer cannot be excluded. The annealing temperature was lowered from 55°C to 37°C to decrease the specificity of primer binding but no further fragments were synthesised. It was concluded that amplification using these two primers from the K5 capsule gene cluster, including the region 2 - 3 boundary and the DNA flanking the cluster on pGB110 is not possible except from the site at which the primers were actually chosen in the region 1 - 2 boundary. During the course of this work DNA sequence of the entire K5 capsule gene cluster became available. Sequence comparisons using FIND revealed that no further binding sites for PR1 or PR2 within the gene cluster including the region 2 - 3 interface.

PCR was also attempted from the K1 and K4 capsule genes using the following templates: pRD1 linearised with *Xho*I and pKT172 linearised with *Sal*I. pKT172 (from which pKT274 was derived) carries the K1 capsule gene cluster and about 17kb flanking DNA (Echarti *et al.*, 1983). DNA amplification using pRD1 or pKT172 as a template was unsuccessful despite a change in annealing temperature as before (see above).

The failure to amplify DNA using primers PR1 and PR2 from all templates except those containing the region 1 - 2 boundary of the K5 capsule gene cluster from which the primers originate, could be due to one of two reasons. First, DNA homology between the template and the primers could be insufficient to promote PCR indicating that at least at the site of the primer sequences, the region 1 - 2 sequence of the K5 capsule gene cluster is unique. Secondly, sequences homologous to the primers may be too widely spaced to allow DNA amplification. An alternative approach to investigate the boundary regions, namely DNA sequencing, was adopted.

4.2.2 Sequence analysis across the region 1 - 2 boundary of the K1 and K4 capsule gene clusters

To investigate the boundary between regions 1 and 2 in the K1 and K4 capsule gene clusters DNA sequence was determined from the fragments which span the junction as follows. First, a restriction endonuclease cleavage map of the appropriate fragments (described earlier) was constructed as follows. The *Bam*HI - *Cla*I fragment which lies to the right hand end of region 1 of the K4 capsule genes (Figure 4.1) was subcloned from pRD1 into the *Bam*HI-*Acc*I sites of pUC19 and the resulting plasmid termed pRD20. pRD20 was cleaved with a range of restriction enzymes. *Dra*I, *Eco*RV and *Hinc*II cut only once. Single, double and triple digests with *Bam*HI and *Hind*III and each of these enzymes in turn were performed and the products analysed by

agarose gel electrophoresis. A restriction enzyme map of the fragment was thus constructed (Figure 4.3) and checked by double and triple digests with the test enzymes themselves. In preparation for sequencing three constructs were produced. The insert in pRD20 was cloned into both M13mp18 and M13mp19 using the *Bam*HI and *Hind*III sites in the polylinker of the vector. These recombinants thus carried the *Acc*I-*Hind*III portion of the pUC polylinker in addition to the pRD20 insert. The 0.7kb *Bam*HI-*Dra*I fragment of pRD20 was cloned into the *Bam*HI-*Sma*I sites of M13mp19.

Similarly, the *Bam*HI - *Eco*RI fragment which occupies a similar position in the K1 antigen gene cluster (Figure 4.1) and is homologous to the insert of pRD20 (Chapter 3) was cloned from pKT274 into pUC19 and a restriction map of the resulting plasmid, pKTBE was produced (Figure 4.3). The entire insert fragment was cloned into the *Bam*HI-*Eco*RI sites of M13mp19 and a 1.2kb *Bam*HI-*Dra*I fragment (the result of partial cleavage with *Dra*I) cloned into M13mp19 cleaved with *Bam*HI and *Sma*I.

These M13 constructs of pRD20 and pKTBE were used in sequencing reactions (Figure 4.3). Two oligonucleotides (below) were synthesised once some sequence data was available and used to prime reactions in sequence extension beyond the *Dra*I sites in pKTBE and pRD20 (Figure 4.3). The two oligonucleotides, oligo 1 and oligo 4 bound to pKTBE and pRD20 at base 1064 and 712 respectively (Figure 4.4).

oligo 1 5'- TTTTATCGTGCAAAGAG -3'

oligo 4 5'- ATGCCCTGTACGACATC -3'

The DNA sequence generated from the K1 and K4 capsule genes cloned in pRD20 and pKTBE was analysed using READ to search for open reading frames. The sequences were aligned with the known sequence from the K5 capsule gene cluster (Figure 4.4). The sequence extending from the *Cla*I site towards region 1 in pRD20 is separated by 0.4kb of unsequenced DNA from the rest of the pRD20 sequence and is presented in Figure 4.7.

4.2.3 KpsS has a variable C terminus

Region 1 of the group II capsule gene cluster exhibits DNA homology of over 95% between different gene clusters. The DNA fragments from which sequence has been determined in this study, namely the 1.8kb *Bam*HI - *Cla*I fragment from the K4 antigen gene cluster (in pRD20) and the 2.1kb *Bam*HI - *Eco*RI fragment from the K1 antigen gene cluster (in pKTBE) both contain the 3' end of region 1 (Figure 4.1). It would therefore be expected that the nucleotide sequence of part of the fragments would be identical or very similar to the equivalent sequence from the K5 capsule gene cluster which includes the *kpsS* gene. Sequence analysis revealed that, as predicted, these DNA fragments do contain the 3' end of *kpsS*. To distinguish between *kpsS* in

the different capsule gene clusters, *kpsS* encoded by the K1 antigen gene cluster will be referred to as *kpsS_{K1}* and so on (*kpsS_{K4}*, *kpsS_{K5}*).

The predicted KpsS_{K5} peptide is 389 amino acids in length. *kpsS_{K1}* and *kpsS_{K4}* were sequenced from the conserved *Bam*HI site which corresponds to the position of amino acid 105 in KpsS_{K5}. *kpsS_{K5}* extends 855 nucleotides from the conserved *Bam*HI site (Figure 4.4). The nucleotide sequence of *kpsS* from the K1, K4 and K5 capsule gene clusters is near identical up to base 855 (Figure 4.4). DNA sequence homology is typically 96.5% between any two *kpsS* sequences. This homology is reflected in the deduced amino acid sequences (Figure 4.5). KpsS_{K1} and KpsS_{K4} have 99.6% identical amino acids, KpsS_{K5} and KpsS_{K4} 97.7% and KpsS_{K1} and KpsS_{K5} 98.6% identical. Accounting for conservative amino acid changes 100% homology is seen between KpsS_{K1} and KpsS_{K5}. KpsS_{K1} shows a non-conservative substitution of alanine for tryptophan at position 282, although both are non-polar residues. However, neither *kpsS_{K1}* nor *kpsS_{K4}* terminate at base 855 (Figure 4.4). *kpsS_{K1}* extends to base 891 and the predicted protein is 12 amino acids longer at the C terminus than KpsS_{K5} (Figure 4.5). *kpsS_{K4}* terminates at base 897 and the predicted protein is 14 amino acids longer than KpsS_{K5} (Figure 4.5). Beyond the first residue, glycine, there is no similarity between the C termini of KpsS_{K1} and KpsS_{K4} (Figure 4.5). The C terminus of KpsS encoded by the K1, K4 and K5 gene clusters is hydrophilic as predicted from the deduced amino acid sequence and the C terminus of KpsS_{K4} is exceptionally basic. The nucleotide sequence of *kpsS_{K1}* has also been determined in the laboratory of Dr E. Vimr (personal communication). The DNA sequence and proposed peptide sequence, including the C terminus, found in the two studies are identical (Figure 4.5). The C terminus of KpsS encoded by the K1, K4 and K5 capsule gene clusters is therefore variable; KpsS_{K1} and KpsS_{K4} extend 12 or 14 amino acids respectively beyond the C terminus of KpsS_{K5} although up to this point the predicted proteins are almost identical.

4.2.4 Sequence analysis beyond *kpsS_{K1}*

An open reading frame (ORF) of 429bp was identified beyond *kpsS_{K1}* on the lower strand of the sequence determined from pKTBE. The ORF terminates with a double stop codon at 1046 (Figure 4.4), 155bp from the *kpsS_{K1}* stop codon. The start codon of the ORF precedes the sequence available. Translation of this ORF (Figure 4.6) revealed an amino acid sequence identical with the C terminal 237 amino acid sequence of the published NeuS (Frosch *et al.*, 1991). This incomplete ORF is therefore positively identified as *neuS*. Organisation of the K1 and K5 capsule genes around the region 1 - 2 boundary is similar. The adjacent region 1 and region 2 genes are transcribed towards each other leaving 155bp or 339bp respectively between stop codons.

4.2.5 Sequence analysis beyond $kpsS_{K4}$ and comparison with the K5 capsule genes

A similar, though less precisely defined situation exists in the case of the K4 capsule genes. Only 203 bases of sequence were determined beyond the stop codon of $kpsS_{K4}$ but 458 bases of nucleotide sequence extending from the *Cla*I site in region 2 towards region 1 was also obtained (Figure 4.3). This sequence lies within region 2 of the K4 capsule gene cluster. An ORF, termed ORF K4, extends through this sequence with neither start nor stop codons in the opposite direction to $kpsS$. The sequence and translation of ORF K4 are displayed in Figure 4.7. A comparison of ORF K4 with the nucleotide sequence of the entire K5 capsule gene cluster (Pazzani and Smith, unpublished) identified a region of 56% DNA homology over 310bp. The coordinates of this homology corresponded to those of the gene in region 2 of the K5 capsule genes which encodes the 60kD protein. The predicted amino acid sequence encoded by ORF K4 and that of the 60kD protein encoded by the K5 genes were compared (Figure 4.8). The predicted 60kD protein is 520 amino acids in length and between 240 and 470 lies a region with 35% identity over a 143 amino acid overlap with the deduced 152 amino acid sequence encoded by ORF K4. Considering conservative amino acid changes this homology rises to 77%. A stop codon which may terminate ORF K4 was not identified within the 203 bases of sequence beyond $kpsS_{K4}$ which was determined. If ORF K4 is a real gene it presumably stops within the 0.4kb of DNA which was not sequenced. Database searches (NBRF-PIR, SWISS-PROT Protein Sequence Database) with either the 60kD protein encoded by region 2 of the K5 capsule genes (search was performed by C. Pazzani) or the predicted protein encoded by ORF K4 did not identify any proteins with significant homology. The function of the 60kD encoded within region 2 of the K5 capsule genes is not known.

4.2.6 Comparison of the region 1 - 2 boundaries in the K1, K4 and K5 capsule gene clusters

A boundary region comprising 155bp of non-coding DNA was found between the stop codons of $kpsS_{K1}$ and *neuS* (region 2). At least 203bp of apparently non-coding DNA lies beyond the stop codon of $kpsS_{K4}$. These sequences constitute the region 1 - 2 boundaries of the K1 and K4 capsule gene clusters and are comparable with 339bp between region 1 and region 2 genes in the K5 capsule gene cluster (Pazzani and Smith unpublished). The boundary region sequences were compared and analysed using BESTFIT, REPEAT and FIND. DNA homology was not identified. Database searches (EMBL, GenBANK) did not identify any similar DNA sequences.

4.3 Discussion

DNA sequence analysis from the common *Bam*HI site at the end of region 1 into the beginning of region 2 of the K1 and K4 antigen gene cluster was performed and a comparison made with the equivalent sequence from the K5 capsule gene cluster (Pazzani *et al.*, in preparation). The experiments which were performed were designed to answer questions about the region 1 - 2 boundary in different capsule gene clusters but actually revealed some interesting results about region 1 itself. In the K5 capsule gene cluster this common *Bam*HI site bisects *kpsS* with the terminal (3') 855bp of the gene lying beyond the *Bam*HI site towards region 2 (Figures 4.3, 4.4). Region 1 of the group II capsule genes is defined, in part, as a region of DNA homology (over 95%) implying that the proteins encoded by region 1 of different group II K antigen gene clusters are identical. Therefore, it was no surprise to discover that beyond the common *Bam*HI site in the K1 and K4 capsule genes was the *kpsS* gene (*kpsS*_{K1} and *kpsS*_{K4}). Comparison of *kpsS* (from the *Bam*HI site) revealed highly homologous (typically 96.5%) DNA sequence between the *kpsS*_{K1}, *kpsS*_{K4} and *kpsS*_{K5} genes up to but not including the stop codon of *kpsS*_{K5} at base 855 (Figure 4.4). The predicted amino acid sequences of KpsS_{K4} and KpsS_{K5} were identical (accounting for conservative amino acid changes) up to and including the C terminal tryptophan residues in KpsS_{K5} (Figure 4.5). However, the C terminus of KpsS encoded by different capsule gene clusters was not the same (Figure 4.5). KpsS_{K1} is 12 amino acids longer than KpsS_{K5} at the C terminus and KpsS_{K4} is 14 residues longer than KpsS_{K5}. There was no amino acid sequence homology between the C termini of KpsS_{K1} and KpsS_{K4} (Figure 4.5). This is an unexpected result for it was assumed that the gene products of region 1 from different K antigen gene clusters were the same. It is now apparent that at least in the case of KpsS this is not the case.

A comparison of a given gene from different *E. coli* strains would reveal minor differences in nucleotide sequence some of which do not alter the amino acid sequence of the gene product and some result in conservative changes. This was the case for most of the *kpsS* gene but the 3' end of the gene was variable and clearly each capsule gene cluster encodes a slightly different version of the KpsS protein. That KpsS has a variable C terminus raises several questions. *kpsS* no longer conforms to the definition of region 1. The gene product is involved in polysaccharide transport but the gene is not completely homologous and the 3' end is unique for each K antigen; this is a characteristic of a region 2 gene. There is a precise point in *kpsS* at which DNA homology between different capsule gene clusters breaks down and the sequence becomes K serotype-specific. The significance of the variable C terminus of KpsS is not clear.

It is possible that the unique C terminus of KpsS is involved in the expression of a specific K antigen. Perhaps the C terminus of KpsS recognises the polysaccharide or even the polysaccharide-specific region 2 proteins and the rest of the protein is somehow concerned with polysaccharide transport. If this was the case KpsS could act as a link between polysaccharide synthesis (a unique process) and polysaccha-

ride export (a common process). There are, however, several arguments against this. First, consider the likely structure and function of KpsS. The observation that a TnPhoA insertion in the 3' end of the gene (in the K5 capsule genes) results in a PhoA positive fusion protein and the identification of a putative membrane spanning domain in the 40-80 amino acids before the C terminus is good evidence that the C terminus of KpsS is periplasmic and the N terminus cytoplasmic (Pazzani *et al.*, in preparation). In a mutant which lacked KpsS_{K5} alone polysaccharide was located in the cytoplasm, not associated with the inner membrane (Bronner, D., unpublished data). It has been postulated that the larger N terminal domain of KpsS is cytoplasmic and may constitute the active site of the protein, and that the smaller C terminal domain, the very end of which extends into the periplasm, serves to anchor the protein to the cytoplasmic membrane. If this is the case, the exact nature of the variable C terminus of KpsS (which in all cases is hydrophilic) may not be relevant to the function of the protein, although of course, this theory is only speculative. Another indication that the C terminus of KpsS is not important is that KpsS_{K5} does not appear to have a unique amino acid sequence. The expression of the K5 antigen would therefore appear to proceed without a specific protein-protein or protein-polysaccharide recognition facility. However, the amino acid sequence of KpsS_{K5} may not be unique but the C terminus itself may be. In the absence of information about KpsS encoded by other group II capsule gene clusters it is impossible to speculate further. Further evidence that a unique C terminus of KpsS is not essential for K antigen export is provided by complementation analyses. Mutants defective in region 1, lacking one or several proteins, can be complemented *in trans* by region 1 from a different capsule gene cluster. This implies that region 1 proteins, including KpsS, can handle chemically different group II K antigens without the need for any K antigen-specific domains. However, in these experiments a complete region 1 from a different capsule gene cluster was used to complement mutants. The possibility cannot be excluded that the region 1-encoded proteins are specific to each other and only function if the complete set of proteins encoded by any single region 1 is present. In conclusion, given that region 1 gene products can handle different polysaccharides and that KpsS_{K5} does not have a unique amino acid sequence it seems most likely that the variable C terminus of KpsS is redundant and that, as predicted by the very definition of the region, this region 1 encoded protein is not capsule-specific. Perhaps KpsS_{K5} represents the minimum functional unit. It would be interesting to extend this study to include *kpsS* from other capsule gene clusters. It has recently been shown that *kpsT* also differs between the K1 and K5 capsule gene clusters and KpsT_{K5} is five amino acids longer at the C terminus than KpsT_{K1} (Pavelka *et al.*, 1991; Smith *et al.*, 1990).

Now consider the organisation of the group II capsule gene clusters around the junction between regions 1 and 2. The organisation of the K1, K4 and K5 capsule gene clusters in the region studied is similar and can be summarised as follows (Figure 4.9). A comparison of the region 1 - 2 boundaries in the K1, K4 and K5 capsule genes did not shed any light on a possible mechanism involved in the generation of the group II capsule gene cluster as was hoped. There was no significant DNA homology between the different gene clusters in this region and therefore this is not a site of homologous recombination. In contrast, examination of the DNA downstream

of region 1 of the *H. influenzae* capsule genes (downstream of *bezD*), towards region 2 of the capsule gene cluster (Figure 1.4, Section 1.3.3.4) identified three islands of DNA homology between the type b and type c capsule gene clusters (Kroll *et al.*, 1990). This suggests that homologous recombination between gene cluster encoding for the production of chemically different capsules could take place in the region 1 - 2 boundary of the *H. influenzae* gene cluster but the same is not true for *E. coli*. These regions may, however, adopt a configuration or contain a sequence which facilitates genetic exchange either at this site or elsewhere. Whether these boundary regions are merely gaps between operons or are functionally significant is still not clear. With respect to the cassette theory proposed by Boulnois and Jann (1989), homologous recombination may not occur at the boundary region (in *E. coli*) as defined in this study but region 2 is still flanked by several kilobases of homologous DNA, that is regions 1 and 3 themselves. Recombination could, for instance, take place within *kpsS* (upstream of the *kpsS*_{K5} stop codon). It is possible that the variable 3' end of *kpsS* arose by in-frame fusion between the resident *kpsS* and an incoming region 2 element (or part thereof) during the generation of the various group II capsule gene clusters. Alternatively, the 3' end of *kpsS*_{K5} may have been lost by a similar process. The same may be true for *kpsT*. Perhaps *kpsS* and *kpsT* are unique members of the common regions 1 and 3 by virtue of their variability. This could be because of their proximity to the variable region 2 cassette. A clearer picture awaits the comparison of other *kps* genes between different gene clusters.

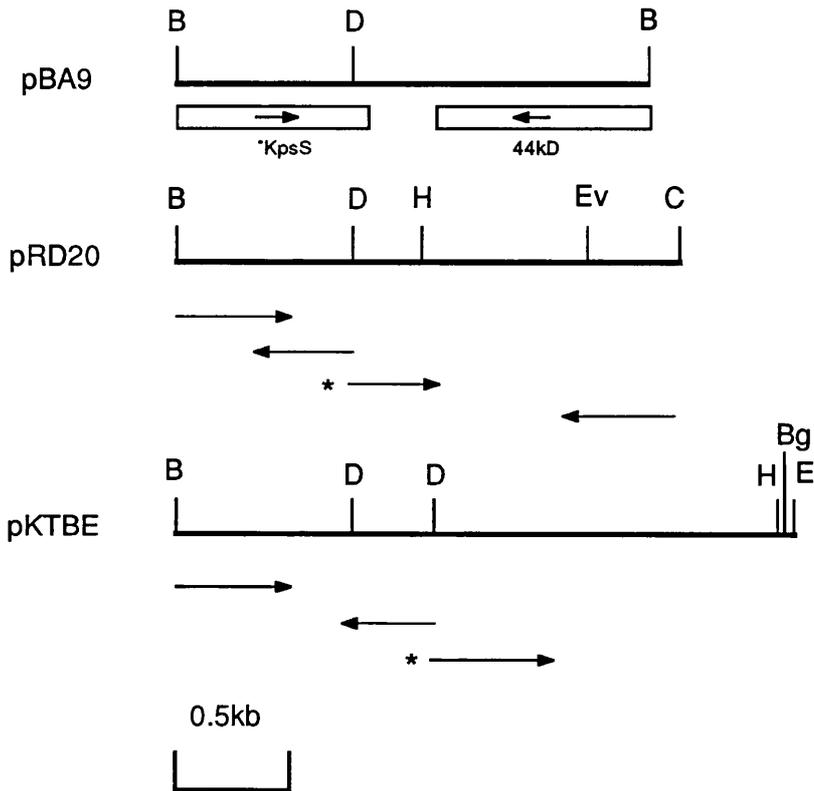
An open reading frame was identified in region 2 of the K4 capsule gene cluster and termed ORF K4. On the assumption that ORF K4 is a gene and that its product has a function, the protein translation of ORF K4 was analysed. The predicted peptide was found to have 35% identity over 143 amino acids (77% similarity) with the predicted amino acid sequence of the 60kD protein encoded within region 2 of the K5 antigen gene cluster. The K4 and K5 polysaccharides are similar in structure, both contain glucuronic acid with either N-acetyl galactosamine or N-acetyl glucosamine respectively; the K4 antigen also contains fructose. It therefore seems likely that the synthetic machinery for the production of these two K antigens may be similar. The identification of two similar but not identical genes in region 2 of the respective capsule determinants is not altogether surprising. The degree of DNA homology (56%) and amino acid sequence similarity between the two would suggest that the two genes have a common ancestor and have not arisen by convergent evolution.

Figure 4.2: PCR strategy

(a) DNA sequence of part of pBA9 (subcloned from pGB110) as determined by Pazzani *et al.* (unpublished). The sequence shown includes the 3' terminus of *kpsS* (region 1), the region 1 - 2 boundary and the 3' end of the region 2 gene which encodes a 44kD protein as marked. The unlabelled arrows represent the site of potential transcription termination loops and the single lines represent the binding site of the two PCR primers (PR1 and PR2) as indicated. The sequence is numbered from the *Bam*HI site at the region 1 end of the insert in pBA9 (Figure 4.1).

(b) Different plasmid templates were used for PCR (each linearised as stated in the text) with primers PR1 and PR2 and the product separated by agarose gel electrophoresis in a 1% gel, stained with ethidium bromide and visualised under UV. The template in each track was as follows: Lane 2, pKT172; 3, pRD1; 4, pAS21; 5, pBA9; /, pGB110. Molecular weight markers in the left hand track (bp) were as follows: 1635, 1018, 516, 394, 344, 298 and 211.

Figure 4.3: Sequencing Strategy



Restriction enzyme cleavage maps of the fragments which span the region 1 - 2 boundaries in the K5 (pBA9), K4 (pRD20) and K1 (pKTBE) capsule gene clusters (Figure 4.1). Additional *DraI* sites are present in pKTBE but their position was not determined. The arrowed boxes indicate the position and direction of the open reading frames in regions 1 and 2 of the K5 capsule genes as marked. The plain arrows illustrate the strategy used to sequence K1 and K4. The two reactions extending to the right of the *DraI* site were primed with oligonucleotides (marked with a star). Restriction enzyme target sites: B, *Bam*HI; Bg, *Bgl*II; C, *Cla*I; D, *Dra*I; E, *Eco*RI; Ev, *Eco*RV; H, *Hinc*II.

Figure 4.4: The nucleotide sequence of pBA9 aligned with the incomplete sequences of pRD20 and pKTBE

The complete nucleotide sequence of the *Bam*HI fragment (cloned in pBA9) which spans the region 1 - 2 boundary in the K5 capsule gene cluster (Pazzani *et al.*, unpublished data) is displayed in the top line marked K5. The sequence is numbered from the first *Bam*HI site as was the sequence shown in Figure 4.2. The incomplete sequence of the comparable fragments from the K4 and K1 capsule gene clusters has been aligned below. The sequence marked K4 was determined from pRD20 and that marked K1 from pKTBE. The binding sites of the oligonucleotides oligo 1 and 4 on pKTBE and pRD20 respectively are marked underneath the sequence. The K1 sequence has been adjusted by the insertion of dots on the basis of DNA homology with K4 and K5 to allow for the absence of data in the middle of the fragment. The 3' end of the K1 and K5 region 2 genes (*neuS* and that encoding the 44kD protein respectively) are on the lower strand and terminate as shown. Restriction endonuclease cleavage sites are clearly marked with boxes and labelled above the sequence. Stop codons are boxed and labelled below the sequence. Open reading frames are marked above the sequence in question, in black in the case of the K5 sequence, in blue in the K4 sequence and in green in the K1 sequence. The sequences begin with the 3' end of *kpsS* (region 1) in all three cases, each terminating at a different place.

kpsS →

BamHI

1

50

K5 GGATCCGCTT CCTGGCATTT GAAGAAGGAT ATTTACGCC GCAATTTATT
 K4 GGATCCGCTT TCTGGCATTT GAGGAAGGAT ATTTACGCC GCAATTTATT
 K1 GGATCCGCTT TCTGGCATTT GAGGAAGGAT ATTTACGCC GCAATTTATT

51

100

K5 ACCGTTGAAG AAGGCGGAGT GAACGCATAT TCATCGCTAC CGCGCGATCC
 K4 ACCGTTGAAG AAGGCGGAGT AAACGCATAT TCATCGCTAC CGCGCGATCC
 K1 ACCGTTGAAG AAGACGGAGT GAACGCATAT TCATCGCTAC CGCGCGATCC

101

150

K5 GGATTTTTAT CGTAAGTTAC CAGATATGCC TACGCCGCAC GTTGAGAACT
 K4 GGATTTTTAT CGTAAGTTAC CAGATATGCC TACGCCGCAC GTTGAGAACT
 K1 GGATTTTTAT CGTAAGTTAC CAGATATGCC TACGCCGCAC GTTGAGAACT

151

200

K5 TAAAACCTTC AACGATGAAA CGTATAGGCC ATGCTATGTG GTATTACCTG
 K4 TAAAACCTTC AACGATGAAA CGTATAGGTC ATGCGATGTG GTATTACCTG
 K1 TAAAACCTTC AACGATGAAA CGTATAGGTC ATGCGATGTG GTATTACCTG

201

250

K5 ATGGGCTGGC ATTACCGTCA TGAGTTCCT CGCTACCGCC ACCACAAATC
 K4 ATGGGCTGGC ATTACCGCCA TGAGTTCCT CGCTACCGCC ACCATAAATC
 K1 ATGGGCTGGC ATTACCGCCA TGAGTTCCT CGCTACCGCC ACCATAAATC

251

300

K5 ATTTTCCCC TGGTATGAAG CACGTTGCTG GGTTTCGTGCA TACTGGCGCA
 K4 GTTTTCCCC TGGTATGAGG CTCGTTGCTG GGTTTCGTGCA TACTGGCGCA
 K1 GTTTTCCCC TGGTATGAAG CACGTTGCTG GGTTTCGTGCA TACTGGCGCA

301

350

K5 AGCAACTTTA CAAGTAACA CAGCGTAAGG TATTACCGAG GTTAATGAAC
 K4 AGCAACTTTA CAAGTAACA CAGCGTAAGG TATTACCGAG GTTAATGAAT
 K1 AGCAACTTTA CAAGTAACA CAGCGTAAGG TATTACCGAG GTTAATGAAC

351

400

K5 GAACTGGACC AGCGTTATTA TCTTGCTGTT TTGCAGGTGT ATAACGATAG
 K4 GAGCTGGACC AGCGTTATTA TCTTGCCGTT TTGCAGGTGT ATAACGATAG
 K1 GAGTTGGACC AGCGTTATTA TCTTGCCGTT TTGCAGGTAT ATAACGATAG

401

450

K5 CCAGATTCGT AACCACAGCA GTTATAACGA TGTGCGTGAC TATATTAATG
 K4 CCAGATTCGT AACCACAGCA ATTATAACGA TGTGCGTGAC TATATTAATG
 K1 CCAGATTCGT AACCACAGCA ATTATAACGA TGTGCGTGAC TATATTAATG

kpsS →

451 500
 K5 AAGTCATGTA CTCATTTTCG CGTAAAGCGC CGAAAGAAAAG TTATTTGGTG
 K4 AAGTCATGTA CTCATTTTCA CGTAAAGCAC CGAAAGAAAAG TTATTTGGTG
 K1 AAGTCATGTA CTCATTTTCA CGTAAAGCGC CGAAAGAAAAG TTATTTGGTG

501 550
 K5 ATCAAAACATC ATCCGATGGA TCGTGGTCAC AGACTCTATC GACCATTAAT
 K4 ATCAAGCACC ATCCAATGGA TCGTGGTCAC AGACTCTATC GACCATTAAT
 K1 ATCAAAACATC ATCCG.....

551 600
 K5 TAAACGGTTG AGTAAGGAAT ATGGCTTAGG TGAGCGAATC CTTTATGTGC
 K4 TAAGCGATTG AGTAAGGAAT ATGGCTTAGA TGAGCGAGTC ATTTATGTGC
 K1

601 650
 K5 ACGATCTCCC GATGCCGGAA TTATTACGCC ATGCAAAAAGC GGTGGTGACG
 K4 ACGATCTCCC GATGCCGGAA TTATTACGCC ATGCAAAAAGC GGTAGTGACG
 K1

651 700
 K5 ATTAACAGTA CGGCGGGGAT CTCTGCGCTG ATTCATAACA AACCACTCAA
 K4 ATTAACAGTA CGGCGGGGAT CTCTGCGCTG ATTCATAACA AACCACTCAA
 K1

701 750
 K5 AGTGATGGGC AATGCCCTGT ACGACATCAA GGGCTTGACG TATCAAGGGC
 K4 AGTGATGGGC AATGCCCTGT ACGACATCAA AGGCTTGACG TATCAAGGGC
 K1TCAA GGGCTTGACG TATCAAGGGC

oligo 4

Dral

751 800
 K5 ATTTGCACCA GTTCTGGCAG GCTGATTTTA AACCCAGATAT GAAACTGTTT
 K4 ATTTGCACCA GTTCTGGCAG GCCGATTTTA AACCCGAATAT GAAACTGTTT
 K1 ATTTGCACCA GTTCTGGCAG GCCGATTTTA AACCCGATAT GAAACTGTTT

801 850
 K5 AAGAAGTTTC GTGGGTATTT ATTGGTGAAG ACGCAGGTTA ATGCGGTTTA
 K4 AAGAAGTTTC GAGGGTATTT ATTGGTGAAG ACGCAGGTTA ATGCGGTTTA
 K1 AAGAAGTTTC GGGGGTATTT ATTGATGAAG ACGCAGGTTA ATTGGGTTTA

851 900
 K5 TTATTAACAT CTTATCAACA ATCGACATCT CCATTATATC AATCATAACA
 K4 TTATGGGGAG ATGATTTTAA ACAAGATAAA AAAATATAAA AAAAGATAAA
 K1 TTATGGGGGG AACACAACAA ACTGCCAACA TAATATATAT TAAATTTCAA

kpsS STOP

901 950
K5 GATGGATGAT ACCTTTATTA GATGTAAAAA GATAAAGTAT TTTAGATATT
K4 AAAGATTGTA TTATGCTATC AAAAAAACAT CACAATGAAT GGCATTAACA
K1 GTCAATATCT TTTGAATTTT AAGTAGCCAA AAAATCATT CCATCCTCTT

951 1000
K5 ACTAATATAT ATTAGATATC TTTGGATATA GCACATAATA TACATTTCCA
K4 TAACCATTCA TAAACTCCCA TTCATTAAAT AAATGATTAA CATACTATAA
K1 TAACATAAAG GTAATAAAAC ATACTACTTT AAGATTTAAT

1001 1050
K5 ACGCACAAAGC CAAGTATTCT AATTAACATT AGAGTTATGA AAAATTCGAT
K4 TTAGAAAACA CAATAAATTA TAAAAATAAA TATTATAGAT TATACATTTG
K1 TTTACGACTG G TACTGTAAT AGAATATAAA ATGTAGCTTT TATTACTCCC
neuS STOP

1051 HincII 1100
K5 ATAAACACAA CAGCAGGGCA AGAAATTTAA CCAACATCTT ATTTTACGAT
K4 TGATAAGTTA TATACTGGTA AAGTTAACAT GAGCACTTTG TAATGCGTTT
K1 CCAAGAAAAT CCTTTTATCG TGCAAAGAGG GAGATGTTAT ATCATCCGAT

oligo 1
1101 DraI 1150
K5 CGAGGATTAA TAAAGAACAT ACCAACGAGG CATATTTTTA ACGGATTA
K4 TGATGGCATT AGTATCTACT AATATTGAGT CGGGCGCAGA AATGGTCGAT
K1 AATATTTTAA CATTATCAAA ATTCTTTAAA ATATCGAAAT GGAGACGCAG

1151 1200
K5 TAATGCATT CCTTCTCTTA ATAGTTAATA GTATATAACA ATACATTAGT
K4 TTTTATTAGT GAAATCAGAA TTTCTATGT AAAGACTTTC TTTATTCTT
K1 CGTATTAATC CCTTTTTGGG ATTTATCATT ATCACACAAC TTTATCATAA
44kD STOP

1201 1250
K5 CACATTTAAA CAAATCGCGA CTATAGACTT TGTCGACCAC ATCGTTCAAC
K4 AA.....
K1 GAGGCGCTAT TGAATAAGAC TGACATCTTT TAGAAACTAA AGGTGCATAT

1251 1300
K5 TCCTCTGACA TTCGGTTAGT GATAATAATA TCAGCTTTCC CTTTAAAGAT
K4
K1 ACCAAAGAAG ATGATGTTAA TCCAATTAAG TGTTTTATTT TTGCTTTTTT

1301 1350
K5 CGCCAGCTCC CGTTCCAAAG GTGAGTAAA GAATGTATCT CCAGAAATAA
K4
K1 TATTACCGGC TCTATTAATAA AGTCTGGCTC TGTTATAAGA ATTATGTCAC

44kD



neuS

1351 1400
K5 GCGGCTCATA AATAATTACT TTCACGCCTT TTTTCTTGAT ACGCTTTATA
K4
K1 GACATTTTGC CTTTTTATT GCCTTTTGAA TTGCATTAAT ATTTTTTTA

1401 1450
K5 ATACCAAGAA TAGAAGAATC TCTAAAATTA TCTGAACCAC TTTTCATAAT
K4
K1 GGCTCTTTAG GGTGAGGTTT TATAAATATT CTTGCATATC AGGCTTATCT

1451 1500
K5 CAAACGATAA ACCCCACAAA CTTGTGGTCT ATGTTTCAA ATAACATTAG
K4
K1 ATTCTAAGTA AGATAGAAAT TAACGAATCC GCAAATAATG TATGTTCAAT

1501 1550
K5 TGATAAAGTC CTTACGTGTA CGGTTAGCAT CAACAATTGC AGATATAAGT
K4
K1 CAAATACTTT TGATTTCAT ATATTATATC ATCTCTAGT.

1551 1600
K5 TTATTCGGAA CAGACTGATA GTTCGCTAAT AATTGCTTGG TATCTTTGG
K4
K1

1601 1650
K5 CAAACAGTAG CCACCATAAC CAAAAGAAGG ATTATTGTAG TAATTACCAA
K4
K1

BamHI

1651 1661

K5 TGCGCGGATC C
K4
K1

Figure 4.5: Predicted C termini of KpsS encoded by the K1, K4 and K5 antigen gene clusters

1	60
K4	IRFLAFEEGYLRPQFITVEEGGVNAYSSLPRDPDFYRKLPDMPTPHVENLKPSTMKRIGH						
K1	IRFLAFEEGYLRPQFITVEEDGVNAYSSLPRDPDFYRKLPDMPTPHVENLKPSTMKRIGH						
K1VMPTPHVENLKPSTMKRIGH						
K5	IRFLAFEEGYLRPQFITVEEGGVNAYSSLPRDPDFYRKLPDMPTPHVENLKPSTMKRIGH						
		+					
61	120
K4	AMWYYLMGWYHRHEFPRYRHHKSFSPWYEARCWVRA YWRKQLYKVTQRKVLPRLMNELDQ						
K1	AMWYYLMGWYHRHEFPRYRHHKSFSPWYEARCWVRA YWRKQLYKVTQRKVLPRLMNELDQ						
K1V	AMWYYLMGWYHRHEFPRYRHHKSFSPWYEARCWVRA YWRKQLYKVTQRKVLPRLMNELDQ						
K5	AMWYYLMGWYHRHEFPRYRHHKSFSPWYEARCWVRA YWRKQLYKVTQRKVLPRLMNELDQ						
121	180
K4	RYYLAVLQVYND SQIRNHSNYNDVRDYINEVMYSFSRKAPKESYLVIKHHMPDRGHRLYR						
K1	RYYLAVLQVYND SQIRNHSNYNDVRDYINEVMYSFSRKAPKESYLVIKHHP.....						
K1V	RYYLAVLQVYND SQIRNHSNYNDVRDYINEVMYSFSRKAPKESYLVIKHHMPDRGHRLYR						
K5	RYYLAVLQVYND SQIRNHSSYNDVRDYINEVMYSFSRKAPKESYLVIKHHMPDRGHRLYR						
		+					
181	240
K4	PLIKRLSKEYGLDERVIYVHDLPMPELLRHAKAVVTINSTAGISALIHNKPLKVMGNALY						
K1						
K1V	PLIKRLSKEYGLSERVIYVHDLPMPELLRHAKAVVTINSTAGISALIHNKPLKVMGNALY						
K5	PLIKRLSKEYGLGERILYVHDLPMPELLRHAKAVVTINSTAGISALIHNKPLKVMGNALY						
		+ ++					
241	298
K4	DIKGLTYQGHLHQFWQADFKPDMKLFKKFRGYLLVKTQVNAVYYGEMIFNKIKKYKKR						
K1	.XKGLTYQGHLHQFWQADFKPDMKLFKKFRGYLLMKTQVNWVYYGGNTTNCQHNIY						
K1V	DIKGLTYQGHLHQFWQADFKPDMKLFKKFRGYLLMKTQVNWVYYGGNTTNCQHNIY						
K5	DIKGLTYQGHLHQFWQADFKPDMKLFKKFRGYLLVKTQVNAVYY						
		+	+	*	****	*****	

Alignment of the C terminal region of *kpsS*_{K4}, *KpsS*_{K1} and *KpsS*_{K5} as marked. K1V is the *KpsS*_{K1} predicted amino acid sequence as determined by Dr E. Vimr. + denotes a conservative amino acid change between any two proteins and a star denotes a non-conservative substitution. The amino acid numbered 1 is the 105th residue in the complete *KpsS*_{K5}.

Figure 4.6: Incomplete sequence and translation of *neuS*

```

ACTAGAGATGATATAATATATGCAAATCAAAAGTATTTGATTGAACATACATTATTTGCG
1  -----+-----+-----+-----+-----+-----+-----+ 60
T R D D I I Y A N Q K Y L I E H T L F A

GATTCGTTAATTTCTATCTTACTTAGAATAGATAAGCCTGATAATGCAAGAATATTTATA
61  -----+-----+-----+-----+-----+-----+-----+ 120
D S L I S I L L R I D K P D N A R I F I

AAACCTCACCTAAAGAGCCTAAAAAAAATATTAATGCAATTCAAAAGGCAATAAAAAAG
121  -----+-----+-----+-----+-----+-----+-----+ 180
K P H P K E P K K N I N A I Q K A I K K

GCAAAATGTCGTGACATAATTCTTATAACAGAGCCAGACTTTTAAATAGAGCCGGTAATA
181  -----+-----+-----+-----+-----+-----+-----+ 240
A K C R D I I L I T E P D F L I E P V I

AAAAAAGCAAAAATAAAACACTTAATTGGATTAACATCATCTTCTTTGGTATATGCACCT
241  -----+-----+-----+-----+-----+-----+-----+ 300
K K A K I K H L I G L T S S S L V Y A P

TTAGTTTCTAAAAGATGTCAGTCTTATTCAATAGCGCCTCTTATGATAAAGTTGTGTGAT
301  -----+-----+-----+-----+-----+-----+-----+ 360
L V S K R C Q S Y S I A P L M I K L C D

AATGATAAATCCCAAAAAGGGATTAATACGCTGCGTCTCCATTCGATATTTTAAAGAAT
361  -----+-----+-----+-----+-----+-----+-----+ 420
N D K S Q K G I N T L R L H F D I L K N

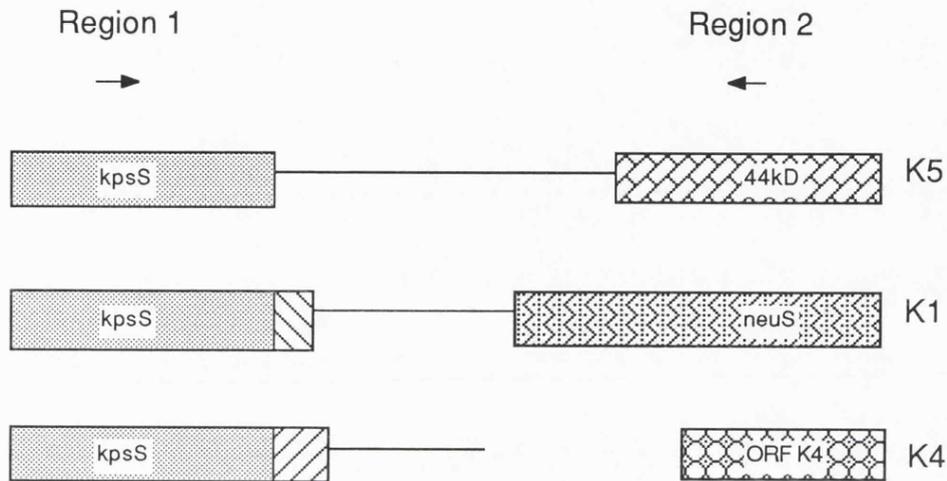
TTTGATAATGTTAAAATATTATCGGATGATATAACATCTCCCTCTTGCACGATAAAAGG
421  -----+-----+-----+-----+-----+-----+-----+ 480
F D N V K I L S D D I T S P S L H D K R

ATTTTCTTGGGGGAGTAA
481  -----+----- 498
I F L G E *

```

The ORF and its translation identified in pKTBE beyond *kpsS*_{K1} in the K1 gene cluster. The translation is identical to the C terminal region of NeuS (Frosch *et al.*, 1991). The sequence shown corresponds to bases 1132 - 535 of the K1 sequence shown in Figure 4.4 and the protein therefore reads N to C terminus.

Figure 4.9: Schematic representation of different capsule gene clusters around the junction of regions 1 and 2



A diagram showing the 3' end of *kpsS*, the non-coding boundary region and the 3' end of the region 2 in the K5, K1 and K4 capsule gene clusters. DNA homology between the different clusters is striking up to the 3' end of the *kpsS*_{K5} gene (dotted boxes), beyond this point there is no DNA homology. *kpsS* in the K1 and K4 gene clusters extend further and are not homologous (hashed boxes) and encode dissimilar C termini. In each case the first gene in region 2 is transcribed in the opposite direction to *kpsS* (arrows). The 3' end of the region 1 and region 2 genes are separated by a non-coding gap of a few hundred bases (plain lines). In the K1 antigen gene cluster this region 2 gene was confirmed as *neuS* and in the K5 cluster it encodes a 44kD protein (patterned boxes). ORF K4 may be the gene in region 2 of the K4 capsule gene cluster which is closest to region 1. The exact organisation of the K4 capsule genes is not known hence the unlinked box tentatively marked ORF K4.

Chapter 5

Southern blot analysis of the group II capsule gene locus and its flanking DNA

5.1 Introduction

Genes encoding both group II (K1 and K4) and group I/II (K10 and K54) capsules map near *serA* on the *E. coli* chromosome and have been given the designation *kps* (Ørskov and Nyman, 1974; Ørskov *et al.*, 1976). Implicit in this was the assumption that K antigens whose genes map near *serA* are encoded by similar if not identical genes. Indeed, analysis of group II capsule gene clusters has revealed a common genetic organisation (Roberts *et al.*, 1988) in which conserved determinants have been termed *kps* (Silver *et al.*, 1984; Vimr *et al.*, 1989). These common genes within a given capsule gene cluster lie in regions 1 and 3 and DNA homology of over 95% is apparent between the different capsule gene clusters analysed to date (Boulnois *et al.*, 1987; Roberts *et al.*, 1988). Recently, some intriguing problems have arisen because group II was divided on biochemical grounds into two: group II and group I/II (Finke *et al.*, 1990). The K2, K3, K10, K11, K19, K54 and K98 capsules are biochemically distinct from group II, being expressed at all growth temperatures (Ørskov *et al.*, 1984) and having a low CMP-KDO synthetase (CKS) activity at 37° (Finke *et al.*, 1990), yet the K10 and K54 genes map to the same chromosomal region as the group II genes and both are encompassed in the term *kps* (Ørskov and Nyman, 1974). It is feasible that the group I/II K antigens are encoded by the same *kps* genes near *serA* that have been described for group II but in which regulation of capsule expression is unusual hence the observed differences in temperature regulation and CKS activity. Alternatively, group I/II capsules may be encoded by genes which are distinct from those described for group II capsules. The nature of the group I/II capsule genes and their relationship to group II *kps* is not clear.

Group I capsule genes are located near *his*, and the biochemical and expressional

properties of group I capsules are very different to those of groups II and I/II; in fact group I K antigens in many ways resemble LPS. It therefore seems unlikely that group I capsule expression involves the group II capsule genes although the demonstration that group II *kps* genes can handle structurally diverse polymers such as K4 may be an indication that this is not necessarily so. In this study Southern blot analysis was used to resolve some of these problems.

5.2 Results

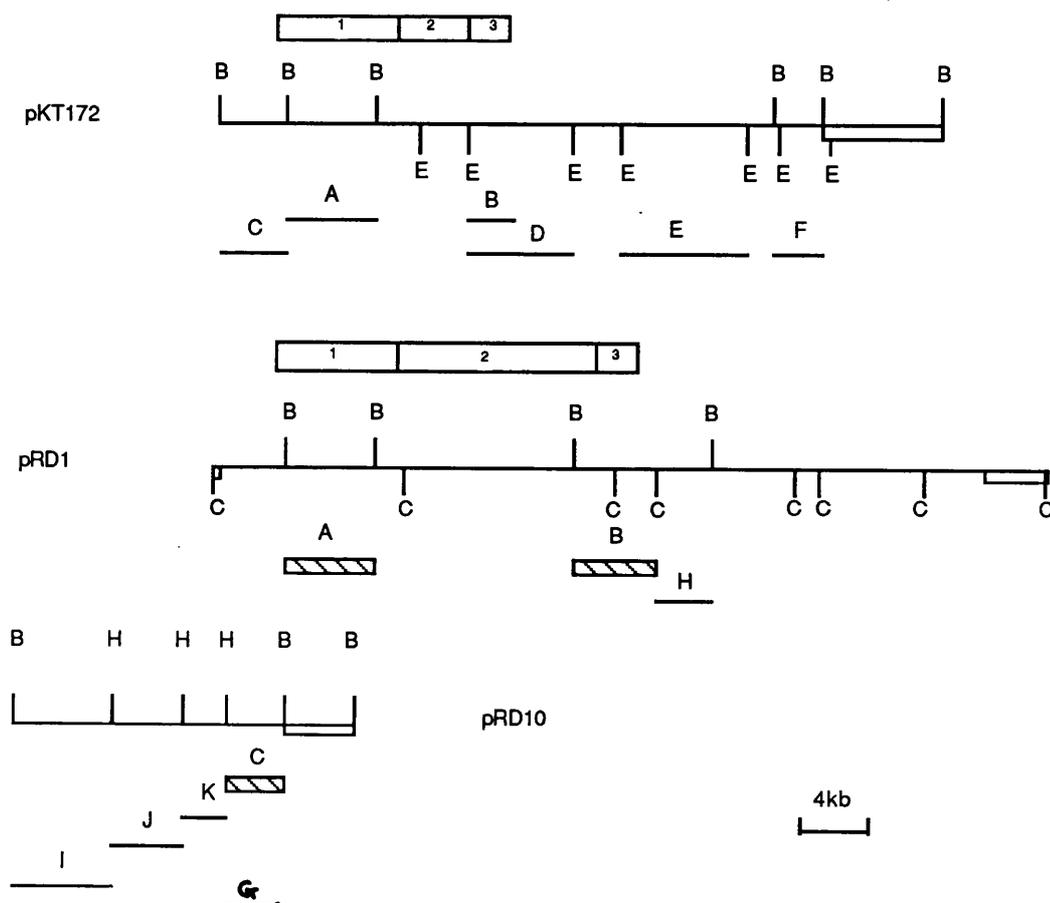
Chromosomal DNA was extracted from *E. coli* strains expressing capsules of different groups (Table 2.1). K9, K29 and K30 capsule expressing strains were chosen as representative of group I and K1, K4 and K5 as representative of group II. Strains expressing the K2, K3, K10, K11, K19 and K54 antigens, which have all been placed in group I/II, were also studied. In each of the Southern hybridisation experiments described below a similar protocol was followed. Chromosomal DNA was extracted from the test strains and cleaved with the restriction endonuclease *Bam*HI or *Hpa*I as indicated. The cleaved DNA was separated by agarose gel electrophoresis (Figure 5.2a) and transferred by Southern blotting to a nylon filter. The filters were made in duplicate and hybridised together with the same probe. One of the pair was washed in 0.5xSSC and the other in 2xSSC. The filters were subjected to autoradiography for 24hours and then again for a longer period. Unless otherwise stated the autoradiographs displayed result from the higher stringency wash and a 24hour exposure. Fragment sizes up to 12kb were calculated from multiple autoradiographs to the nearest 1kb (0.5kb for smaller fragments) with reference to gel photographs. Fragment sizes over 12kb were put into one size bracket (>12). The majority of autoradiographs generated are shown and all results are summarised in tables. All strains were sero-typed in the laboratory of Drs F. and I. Ørskov in Denmark.

5.2.1 Analysis of group II capsule genes in strains not expressing a group II capsule

5.2.1.1 Analysis of group I capsule expressing *E. coli*

Southern blot analysis was used to determine whether strains which express a group I capsule also carry the group II K antigen genes in their chromosome. Two radiolabelled DNA probes, A and B (Figure 5.1) from regions 1 and 3 of the K1 antigen gene cluster in pKT274 were used in Southern blot analysis of chromosomal DNA from group I capsule producing *E. coli*. Probes A and B have been described in Chapter 3. Probe A, a 5.2kb *Bam*HI fragment, encodes the majority of the region 1 genes. Probe B, a 2.7kb *Hind*III-*Xho*I fragment, encodes the region 3 genes (Figure 5.1). Group I (K9, K29, K30) and group II (K1, K4, K5) capsule expressing strains were included in the study. Chromosomal DNA was cleaved with the restriction endonu-

Figure 5.1: Source of DNA probes



Physical map of the recombinant plasmid pKT172 (carrying the biosynthesis genes for the K1 antigen, Echarti *et al.*, 1983), pRD1 (cloned K4 biosynthesis genes, Chapter 3) and pRD10 (see text). The open boxes below the lines refer to vector sequences. The boxes labelled 1 through 3 refer to the gene regions involved in the production of the K1 and K4 antigens although the location of region 3 of the K4 capsule gene cluster is estimated (Chapter 3). The DNA probes are represented by the lines labelled A through K. Probes A and B were taken from pKT274 but align with pKT172 as shown (Figure 3.1). Probes C through F were taken from pKT172. Probe H was taken from pRD1 and probes G, I, J and K from pRD10 as shown. Hashed boxes denote regions of homology with the appropriate K1 antigen gene derived-probe. Restriction enzyme target sites: B, *Bam*HI; C, *Cla*I; E, *Eco*R1; H, *Hpa*I.

Figure 5.2: Chromosomal DNA digestion and hybridisation with region 1 and 3 probes

(a) Ethidium bromide stained agarose gel showing chromosomal DNA cut with the restriction endonuclease *Bam*HI prior to Southern blotting. Similar gels were used in the Southern blot analyses described in this chapter. The gel was photographed alongside a linear rule to aid the calculation of fragment sizes. Molecular weight markers (kb): 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1.6, 1.0, 0.5.

(b,c) Southern blot analysis of *Bam*HI-cleaved chromosomal DNA from eleven *E. coli* strains each expressing a different K antigen as indicated by each track. Identical filters were hybridised with (b) Probe A and (c) Probe B. The source of the probes is shown in Figure 5.1.

(d,e) Southern blot analysis of chromosomal DNA from the K4 and K54 capsule expressing strains as indicated by each track. Identical filters were probed with (d) probe A and (e) probe B.

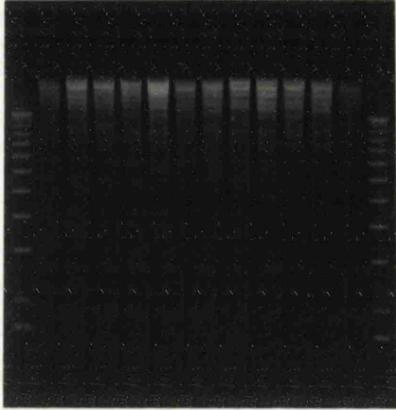
Table 5.1: Results with probes A and B

group	I			II			I/II					
K antigen	K9	K29	K30	K1	K4	K5	K2	K3	K10	K11	K19	K54
probe A	-	-	-	5	5	5	5	-	-	-	-	-
probe B	-	-	-	>12	8	12	>12	-	-	-	-	-

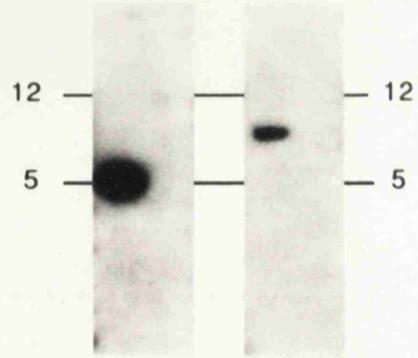
Summary of a series of Southern blot analyses of twelve different *E. coli* strains using radiolabelled DNA probes A and B (Figure 5.1). The K antigen expressed by each strain and the capsule group to which it belongs are listed. Note that K2 was later placed in group II. The size of chromosomal DNA fragments homologous to each probe are listed in kilobases and a dash indicates that no fragments bound the probe. Autoradiographs are shown in Figure 5.2.

Figure 5.2

a) 1 4 5 2 3 10 11 19 9 29 30



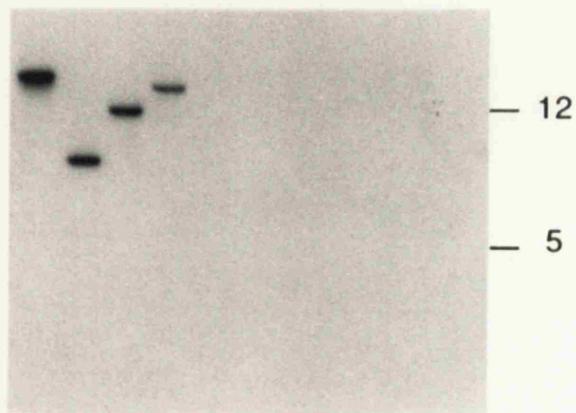
d) 4 54 e) 4 54



b) 1 4 5 2 3 10 11 19 9 29 30



c) 1 4 5 2 3 10 11 19 9 29 30



cleave *Bam*HI. Probe A hybridised to the same sized (5.2kb) *Bam*HI fragment in the chromosomal DNA of the K1, K4 and K5 capsule expressing strains (Figure 5.2b, Table 5.1), exactly as predicted from the known restriction enzyme cleavage maps of these group II K antigen gene clusters. Probe B hybridised to different *Bam*HI fragments in the chromosomal DNA of the group II strains as predicted from the maps (Figure 5.2c, Table 5.1). Neither probe A nor B hybridised to the chromosomal DNA of the group I capsule expressing strains (Figure 5.2ab, Table 5.1). The same results were obtained at both high and low washing conditions. Therefore, strains elaborating a group I capsule do not have the chromosomal determinants necessary to produce a group II capsule and group I capsule biogenesis must be independent of group II capsule gene products. It is feasible, however, that sequences more than 30% divergent from group II *kps* are present on the chromosome.

5.2.1.2 Analysis of group I/II capsule producing *E. coli*

Southern blot analysis was used to determine whether group I/II capsule encoding genes are similar to or distinct from, the well defined group II capsular polysaccharide genes which, like the group I/II capsules, are encoded near *serA*. Included in the Southern hybridisation experiments described above was the *Bam*HI cleaved chromosomal DNA of the K2, K3, K10, K11 and K19 (group I/II) capsule expressing strains. At the time of these experiments a K54 capsule expressing strain was unavailable for study but similar experiments were subsequently performed and chromosomal DNA from a K54 capsule expressing strain was analysed with probes A and B (Figure 5.2de). Neither probe A nor probe B hybridised to the chromosomal DNA of K3, K10, K11, K19 or K54 capsule expressing strains even at low stringency (Table 5.1, Figure 5.2bc). Interestingly, both probes A and B hybridised to the K2 capsule producing *E. coli* at high stringency and probe A hybridised to a 5kb *Bam*HI fragment found in region 1 of several group II capsule gene clusters. Thus the K2 capsule expressing strain carries group II *kps* genes on the chromosome. Although there is no evidence to suggest that these genes are actually involved in K2 capsule expression, given the obvious restriction enzyme site similarity and DNA homology between these sequences and those known to function in group II capsule producing strains, it would seem probable. The conclusion from the Southern blot experiments with the other group I/II strains is clear: homologues of regions 1 and 3 of the group II capsule gene cluster are not present on the chromosomes of the group I/II capsule expressing *E. coli* strains (with the exception of K2).

5.2.2 Analysis of sequences flanking Region 3 of the *kps* genes

The chromosomes of group I and I/II capsule expressing strains lack the group II capsule genes. This implies that at some point group II capsule genes, or the ancestral equivalent, were acquired by the chromosome of some *E. coli* isolates and not others. This may not have been confined to the capsule genes and there could be other

differences in the *serA* region of the chromosome between group II capsule expressing *E. coli* and other isolates. Group I and group I/II capsule producing *E. coli* strains might also, therefore, lack sequences found immediately adjacent to the group II capsule genes.

pKT172 is the primary cosmid which contains the K1 antigen gene cluster from which pKT274 is derived (Echarti *et al.*, 1983). Plasmid pKT172 carries 17kb of DNA adjacent to region 3 of the K1 antigen gene cluster but not involved in capsule expression. To determine the extent to which this DNA is found in *E. coli* strains expressing group I, II and I/II capsules three DNA probes were used in Southern blot analysis. Probes D, E and F (Figure 5.1) span the sequences flanking region 3 of the K1 antigen gene cluster in pKT172. They represent the following restriction endonuclease fragments: probe D (6.0kb *EcoRI*), probe E (7.7kb *EcoRI*) and probe F (2.5kb *BamHI*). Probe D overlaps by 3kb with the K1 antigen gene cluster but it has already been shown that region 3, in the form of probe B, is not present on the chromosome of group I and I/II capsule expressing strains (Section 5.2.1.). Probe D therefore contains sequences which are immediately adjacent to *kpsM* in region 3, and probes E and F cover sequences progressively further from the K1 capsule gene cluster. Each probe was used in Southern blot analysis of *BamHI*-cleaved chromosomal DNA from the K1, K3, K10, K11, K9, K29 and K30 capsule expressing *E. coli* strains. The results are illustrated in Figure 5.3(a,b,c) and summarised in Table 5.2. The three probes hybridised to single *BamHI* fragments in the DNA from all group I, group II and group I/II capsule expressing strains at both high and low stringency with one exception: sequences homologous to probes E and F were not detected in the chromosomal DNA of the K11 expressing isolate (Table 5.2). Interestingly, probe D hybridised to a chromosomal fragment of the same size in the K9, K30 and K3 capsule expressing strains (Figure 5.3a). Probe D was also used in Southern blot analysis of chromosomal DNA of the laboratory strain *E. coli* LE392 (with a K4 capsule expressing strain as a control) and hybridised to a single *BamHI* fragment of greater than 20kb (Figure 5.4ab). It appears from this that the sequences flanking region 3 of the K1 capsule gene cluster are common components of the *E. coli* chromosome, not restricted to chromosomes which carry the group II *kps* genes.

In order to determine whether the sequences adjacent to region 3 of the K1 antigen gene cluster are unusual in their widespread occurrence a probe was also taken from the sequences flanking region 3 of the K4 antigen gene cluster. This comprised a 3kb *ClaI*-*BamHI* fragment from pRD1 which was termed probe H (Figure 5.1). This probe is equivalent in position to probe D but does not contain any region 3 sequences. The exact position of region 3 in pRD1 has not been determined but probe H probably lacks the sequence which is immediately adjacent to region 3 of the K4 capsule gene cluster. Probe H was used in Southern blot analysis of the K1, K3, K10, K11, K9, K29, K30, K2, K4, K5 and K19 capsule expressing strains. Chromosomal DNA was cleaved with either *BamHI* or *HpaI* and the results are shown in Figure 5.3(d,e) and summarised in Table 5.2. Probe H hybridised to chromosomal DNA from all strains tested except the K11 or K19 capsule expressing strains. Probes D and H hybridised to the same sized set of *BamHI* cleavage fragments in the chromosomal DNA of the

Figure 5.3: Southern blot analysis using probes flanking region 3 of the group II capsule genes

(a-d) Southern blot analysis of *Bam*HI cleaved chromosomal DNA from different K antigen expressing strains (the K antigen expressed by each strain is indicated above each track) using (a) probe D, (b) probe E, (c) probe F and (d) probe H. The origins of the probes are shown in Figure 5.1.

(e) Southern blot analysis of *Hpa*I cleaved chromosomal DNA from different K antigen expressing strains with probe H. The K antigen expressed by each strain is indicated above each track.

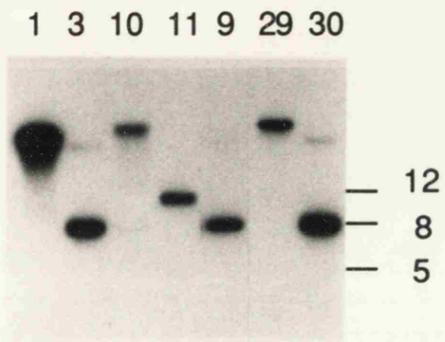
Table 5.2: Results with probes D, E, F and H

Group	I			II				I/II			
	K9	K29	K30	K1	K2	K4	K5	K3	K10	K11	K19
probe D	8	>12	8	>12	nd	nd	nd	8	>12	11	nd
probe H _B	8	>12	8	>12	>12	8	>12	8	>12	-	-
probe H _H	3.5 3	3 2	3.5 3	3.5	10	7	7	>12	>12	-	-
probe E	>12	>12	>12	>12	nd	nd	nd	>12	8	-	nd
probe F	>12	>12	>12	3	nd	nd	nd	>12	12	-	nd

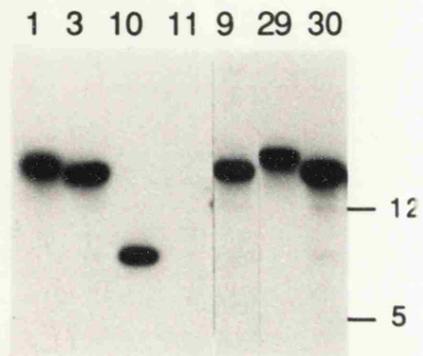
Summary of Southern blot analyses of eleven different *E. coli* strains using probes D, E, F and H (Figure 5.1). The K antigen expressed by each strain and the capsule group to which it belongs are listed. The size of chromosomal DNA fragments homologous to each probe are listed in kilobases and a dash indicates that no fragments bound the probe. In the case of probe H, subscript H or B indicates that chromosomal DNA was cleaved with either *Hpa*I or *Bam*HI respectively. The notation nd indicates that the experiment was not done. The autoradiographs themselves appear in Figure 5.3.

Figure 5.3

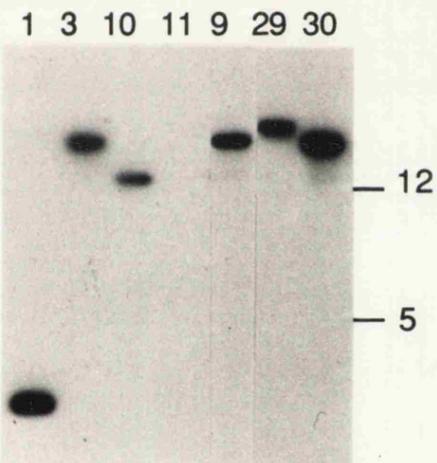
a)



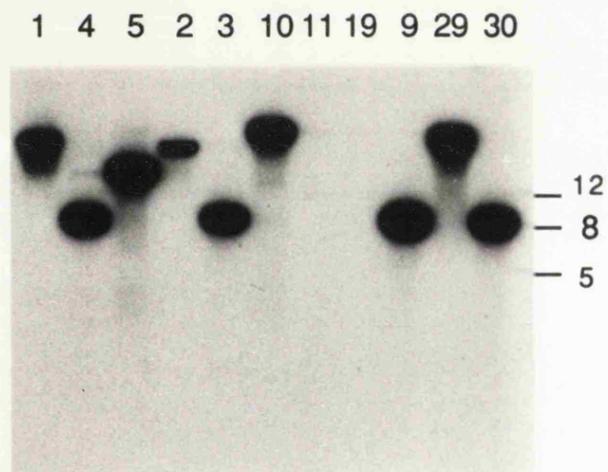
b)



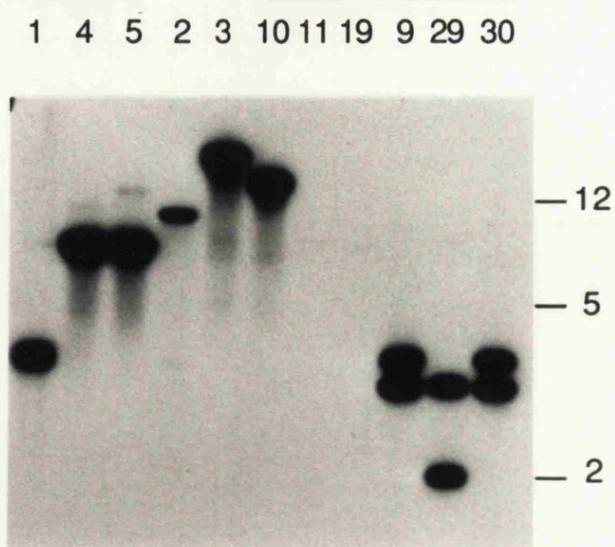
c)



d)



e)



test strains. (Figure 5.3ad). This shared hybridisation pattern could be evidence that the K1 and K4 capsule gene clusters are flanked by the same chromosomal sequences. In the case of K3, K9 and K30 probes D and H hybridised to the same sized 8kb *Bam*HI fragments and probe H identified *Hpa*I fragments of the same size in the chromosomal DNA of the group I capsule expressing strains (Figure 5.3a, d and e). The K11 and K19 capsule expressing *E. coli* showed a different pattern of hybridisation with probes D, E, F and H to the other strains, a similar observation was made in subsequent analysis (Section 5.2.3).

The results with probes D, E, F and H have several implications. Unlike group II capsule genes, sequences adjacent to region 3 of the K1 and K4 group II capsule gene clusters are present on the chromosome of most, but not all, *E. coli* isolates regardless of capsule status. Since probe D overlaps region 3 of the K1 gene cluster by 3kb, somewhere within the remaining 4kb presumably lies a point at which group II capsule gene-associated sequences meet common *E. coli* chromosomal DNA. Chromosomal DNA from the K11 antigen producing strain was homologous to probe D which overlaps with the K1 capsule gene cluster but not to probes which extend further from the capsule genes (probes E and F) nor indeed to the cluster itself (probes A and B). Failure to hybridise with probe H implies that that homology between probe D and the chromosome of the K11 producing strain is restricted to the sequence immediately adjacent to region 3 of the K1 capsule genes which is not present on probe H. The *E. coli* K11 and K19 strains used in this study are exceptional in lacking a common, though obviously not essential, portion of the *E. coli* chromosome.

5.5.5.1 Homologous sequences flank region 3 of the K1, K4 and K5 antigen genes clusters

Since a probe which lies close to region 3 of the K4 capsule gene cluster (probe H) and one which overlaps with region 3 of the K1 capsule gene cluster (probe D) hybridise to the same set of fragments at high stringency (Table 5.2) it is likely that the probes are homologous. This implies that the sequences adjacent to region 3 of different group II capsule gene clusters are the same. If this is so then sequences homologous to probes E and F (which lie beyond probe D in the sequence adjacent to region 3 of the K1 antigen genes) would be expected in the DNA which flanks region 3 of the K4 and K5 capsule gene clusters. The sequences flanking region 3 of the K4 and K5 capsule genes are present in pRD1 and pGB110 respectively. Plasmids pKT172, pGB110 and pRD1 were subjected to Southern blot analysis using radiolabelled probes E and F and washed under conditions requiring 95% DNA homology. The results are shown in Figure 5.4cd and summarised in Figure 5.5. Probes E and F hybridised to the predicted fragments in pKT172. Probe E hybridised to the 7.6kb *Eco*RI fragment in pGB110 which lies close to region 3 (Figure 5.5). Probe F hybridised to the 12.7kb *Eco*RI fragment in pGB110 which is the next *Eco*RI away from region 3 of the capsule genes (Figure 5.5). The results with pRD1 were more difficult to interpret because it is difficult to distinguish between the 7.7kb and 7.9kb *Cla*I fragments in Southern blot analysis. Both probes E and F hybridised to the position on the filter

Figure 5.4: Analysis of LE392 with probes D, I, J and K, and the cloned capsule genes with probes E and F. The origin of all the probes is shown in Figure 5.1.

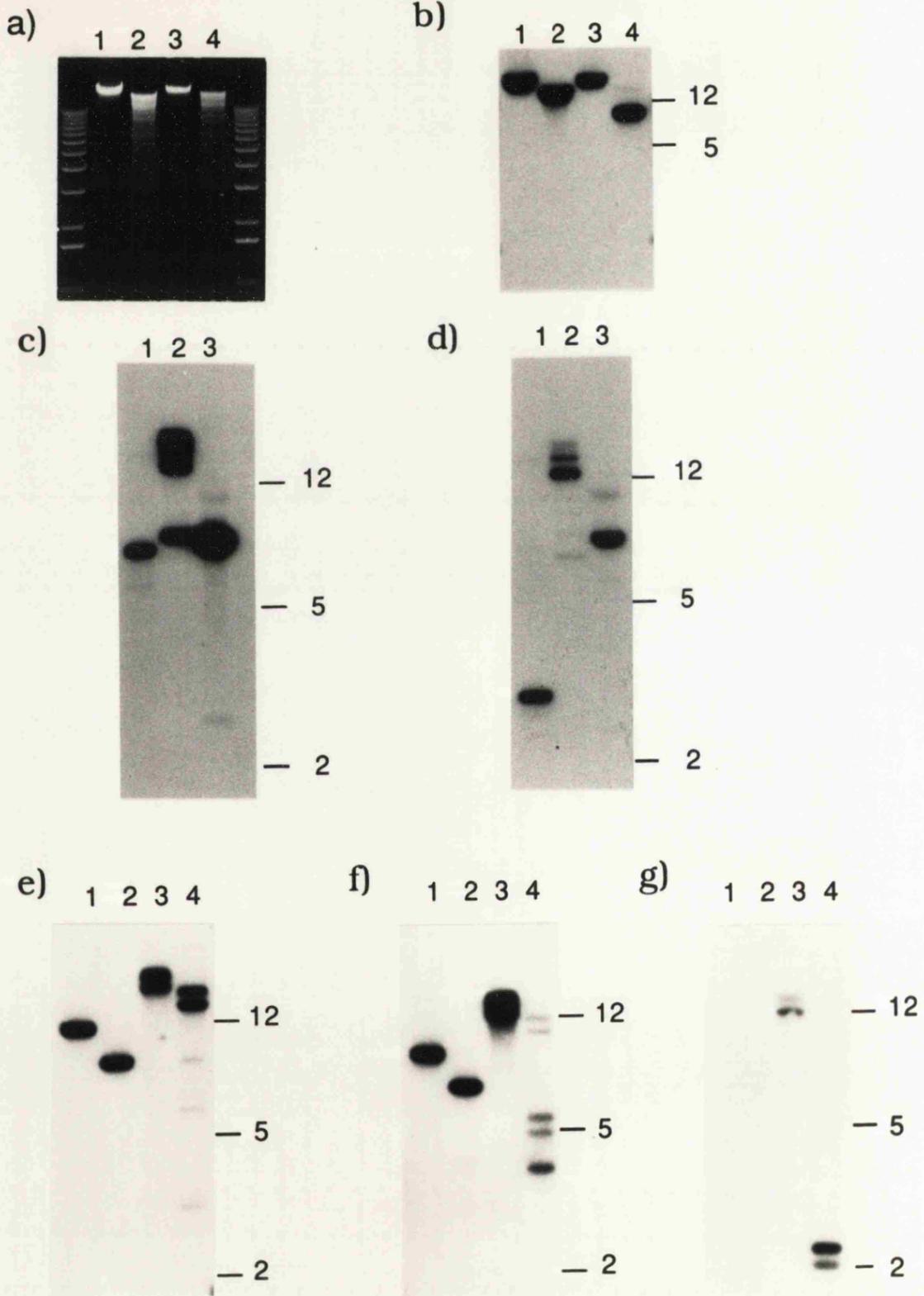
(a) Chromosomal DNA from *E. coli* LE392 and the K4 capsule expressing strain was either cleaved with *Bam*HI or untreated and analysed by agarose gel electrophoresis. DNA was loaded in the following order: lane 1, LE392 uncut; 2, LE392x*Bam*HI; 3, K4 uncut; 4, K4x*Bam*HI. Molecular weight markers (kb): 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1.6, 1.0, 0.5.

(b) Southern blot analysis of the above agarose gel with probe D.

(c,d) pKT172 was cleaved with *Eco*RI, pGB110 also with *Eco*RI and pRD1 was cleaved with *Cla*I and all were subjected to Southern blot analysis with (c) probe E or (d) probe F. The lanes were loaded as follows: Lane 1, pKT172; 2, pGB110; 3, pRD1. Relatively more pRD1 DNA was loaded onto the gels.

(e,f,g) Southern blot analysis of chromosomal DNA extracted from *E. coli* LE392 and the K4 capsule expressing strain. DNA was cleaved with either *Bam*HI or *Hpa*I and subjected to Southern blot analysis with (e) probe I, (f) probe J or (g) probe K. The DNA was loaded as follows: Lane 1, LE392 x *Bam*HI; 2, LE392 x *Hpa*I; 3, K4 x *Bam*HI; 4, K4 x *Hpa*I.

Figure 5.4



corresponding to the 7.7kb and 7.9kb *Cla*I fragments in pRD1. These two fragments are both located in the DNA adjacent to region 3 of the K4 capsule genes. Since this Southern blot analysis was performed with plasmid and not genomic DNA it is clear that the sequences flanking the K1, K4 and K5 capsule gene clusters are homologous. In addition, the organisation of the DNA flanking region 3 of the K1 and K5 capsule genes clusters appears to be the same (Figure 5.5). Comparison of the restriction enzyme cleavage maps of pGB110 and pRD1 in the 3.5kb beyond region 3 reveals shared *Sma*I, *Bam*HI and *Cla*I sites (Figure 5.5). These observations suggest that the K1, K4 and K5 capsule gene clusters are positioned at the same chromosomal location near *serA* and this may be the case for all group II K antigen gene clusters.

5.2.3 Analysis of sequences adjacent to region 1 of the *kps* genes

DNA sequences adjacent to region 3 of the group II *kps* genes are present in many *E. coli* isolates (see above). Similar studies were performed to see whether this is also true for those sequences flanking region 1 of the group II capsule genes. Probe C, a 4.0kb *Bam*HI which overlaps by 0.5kb with region 1 of the K1 genes in pKT172 (Figure 5.1) was used in Southern blot analysis of the genomic DNA of various encapsulated *E. coli*. Sequence homologous to probe C was detected in the group II K1, K2, K4 and K5 capsule expressing strains but not in those producing K3, K10, K11, K19, K9, K29 or K30 capsules (Table 5.3). This suggests that at least 3.5kb of DNA adjacent to region 1 of the K1 antigen gene cluster is only present on the chromosome of strains which carry the group II capsule determinants, although there is no evidence as yet to suggest this is essential or involved in group II capsule biogenesis. Sequences common to the *E. coli* chromosome may lie further than 3kb from region 1 of the K1 capsule gene cluster .

5.2.3.1 Isolation of a DNA fragment adjacent to K4 region 1

In all of the group II capsule gene clusters that have been cloned, the vector sequence is less than 4kb from the end of region 1 and consequently there is no source of probes extending beyond this end of the capsule gene cluster. The *E. coli* U1-41 cosmid library from which the K4 antigen gene cluster was isolated (Chapter 3) proved useful. Screening the library with probe A (region 1) identified thirteen recombinants only one of which actually carried the complete K4 capsule gene cluster (Section 3.2.2.3). The remaining twelve cosmids must therefore contain sequences which extend into the region 1 flanking DNA. Plasmid DNA was extracted from ten of these recombinants, including pRD1 and cleaved with the restriction endonuclease *Bam*HI (Figure 5.6a). Two of the ten proved to be identical, seven of the ten contained the 5.2kb *Bam*HI fragment from region 1 of the K4 capsule genes. A *Bam*HI fragment greater than 12kb was present in seven plasmids but not in pRD1, this was a likely contender for the genomic *Bam*HI fragment which encodes the 5' end of region 1. Probe C, which

contains 0.5kb of region 1 of the K1 capsule genes was used in Southern blot analysis of the plasmids. Probe C hybridised to the greater than 12kb *Bam*HI fragment which was present in seven cosmids and to the 23kb *Bam*HI fragment in pRD1 which contains the vector (Figure 5.6b). The greater than 12kb *Bam*HI fragment was subcloned from one of the plasmids into the *Bam*HI site of the vector pACYC184 and the resulting plasmid termed pRD10. The entire fragment was used in a preliminary Southern blot analysis of the K1, K3, K10, K11, K9, K29 and K30 encapsulated *E. coli* and hybridised to all strains tested except that expressing the K29 antigen (data not shown). A restriction enzyme cleavage map of pRD10 was generated with the following enzymes: *Bam*HI, *Cla*I, *Hind*III, *Hpa*I and *Sst*I (Figure 5.7). On the basis of the restriction map the *E. coli* U1-41 genomic DNA inserted into pRD10 was calculated to be 15.5kb. To determine whether the region of homology between pRD10 and probe C was restricted to the 5' end of region 1, pRD10 was cleaved with the restriction enzymes that have been mapped, singly and in combinations (Figure 5.6c) and subjected to Southern blot analysis with probe C. A 3.4kb *Bam*HI-*Hpa*I fragment and 17.8kb and 1.7kb *Bam*HI-*Hind*III fragments but not the 0.5kb *Hpa*I-*Sst*I fragment were homologous to probe C (Figures 5.6d and 5.7). Since region 1 extends only 0.5kb into probe C it is concluded that the sequences adjacent to region 1, like those adjacent to region 3, of the K1 and K4 capsule gene clusters are homologous.

5.2.3.2 Southern blot analysis with probes flanking K4 region 1

pRD10 was used as a source of DNA probes beyond region 1 of the K4 antigen gene cluster. Firstly, the 3.4kb *Bam*HI - *Hpa*I fragment (probe G, Figure 5.7) with homology to probe C was used in Southern blot analysis with the different encapsulated strains and as expected, gave the same results as probe C (Table 5.3). A further three probes were chosen which together spanned the entire 15.5kb of pRD10. Probe K is the 2.5kb *Hpa*I fragment next to probe G, probe J is the next 4.0kb *Hpa*I fragment and probe I, the furthest from region 1, is a 5.8kb *Hpa*I - *Bam*HI fragment (Figure 5.7). Southern blot analyses with these probes were performed only at the higher stringency (0.5x SSC) because in the previous experiments there was no detectable difference between results obtained at the two wash conditions. Chromosomal DNA was cleaved with *Bam*HI or *Hpa*I and hybridised with each probe. K1, K2, K3, K4, K5, K3, K10, K11, K19, K9, K29 and K30 capsule expressing strains were analysed. Chromosomal DNA of the laboratory strain *E. coli* LE392 was also included in the analysis (Figure 5.4efg). The three probes hybridised in a complex manner to chromosomal DNA of different capsule expressing strains and the intensity of the bands varied. The results are shown in Figure 5.8 and summarised in Table 5.3.

Consider the results with each probe in turn. Probe K did not hybridise to the chromosomal DNA of the group I capsule expressing strains or *E. coli* LE392 but hybridised to all group II and group I/II encapsulated strains. Interestingly, probe K detected multiple *Bam*HI and *Hpa*I fragments in the chromosome of all the group II

Table 5.3: Hybridisations with probes adjacent to region 3

group	I			II				I/II			
	K9	K29	K30	K1	K2	K4	K5	K3	K10	K11	K19
probe C	-	-	-	7	12	>12	>12	-	-	-	-
probe G	-	-	-	7	12	>12	>12	-	-	-	-
probe K _B	-	-	-	>12 >12	>12 12	>12 >12	12 8	>12 >12 10 7	>12 >12	>12 11	>12
probe K _H	-	-	-	>12 >12 8	>12 3	3 2.5	>12 3	>12 7 4 2.5	2.5 2	10 3	12
probe J _B	>12	-	>12	>12 >12	>12 >12 >12	>12 >12	>12 >12 9	>12 >12 12 8	>12 >12 >12	>12	>12
probe J _H	3.5	-	3.5	12 10 5.5 3.5	12 8 4 4	>12 12 6 5 4	>12 >12 5 4.5 3.5	>12 11 8 7 3.5	6.5 5.5 5 3.5	7	10
probe I _B	4	-	4	>12 4	>12 12 8 6 4	>12 >12	>12 >12 12	>12 >12 9 8 7 4	>12 >12 >12	>12 >12	-
probe I _H	>12	-	>12	6.5	>12 >12 9 6.5 5.5	>12 12 9 6 3.5	>12 >12 9 6 2	>12 9 6.5 3	9 6 3	8	-

Summary of a series of Southern blot analyses of chromosomal DNA of different encapsulated *E. coli* strains with various probes (Figure 5.1). The capsular group and the K antigen produced by each strain tested is indicated across the top and the size of fragments bound by each probe is given in kilobases in the table. All fragments over 12 kilobases have been put into the same size bracket (>12) and a dash indicates no homology. Subscript H indicates that the chromosomal DNA was cleaved with *HpaI* and subscript B that it was cleaved with *BamHI*.

and group I/II capsule expressing strains, except K19 in which only single fragments hybridised (Figure 5.8ab). Probe J hybridised to chromosomal DNA from *E. coli* LE392 and all the group I, II and I/II capsule producing strains except *E. coli* K29 (Figure 5.8cd). Hybridisation of probe J with *E. coli* K11 and K19 was extremely weak. Hybridisation of probe J to LE392, the K9, K30 (group I) and the K11 and K19 (group I/II) capsule expressing strains was restricted to single fragments. Multiple *Bam*HI and *Hpa*I fragments were homologous to probe J in the chromosomal DNA of the group II and two of the four group I/II (K3, K10) capsule expressing strains. The pattern with probe J was even more elaborate than was seen with probe K. In the case of K4, probe J hybridised to the fragment used as a probe and an additional four *Hpa*I fragments (Figure 5.8c). Finally, probe I hybridised to all strains except the K19 and K29 encapsulated strains (Figure 5.8). With probe I hybridisation to the K11 antigen expressing strain was strong. Single *Bam*HI and *Hpa*I fragments in the chromosomal DNA of *E. coli* LE392, the group I and the K11 capsule expressing strains were homologous to probe I. Multiple bands were homologous to probe I in the chromosomal DNA of the group II and two of the four group I/II (K3 and K10) capsule expressing strains. However, probe I did not identify a very complex pattern of homology in the K1 antigen expressing isolate.

Consideration of the results with all three probes as a whole reveals some trends. The hybridisation patterns of the K9 and K30 capsule producing *E. coli* were the same with probes I, J and K (Figure 5.8) as they had been with probe D (Section 2.2). None of the probes from pRD10 hybridised to the K29 antigen producing strain although the significance of this is not clear. Probe K did not hybridise to chromosomal DNA of group I capsule producing strains and probes J and I hybridised only to single fragments. This same hybridisation profile was exhibited by the unencapsulated laboratory strain *E. coli* LE392. Probes more than 3kb beyond region 1 of the K4 capsule gene cluster (I, J and K) hybridised to multiple *Bam*HI and *Hpa*I fragments (with the exception of probe J hybridising to K1) in the chromosomal DNA of the group II capsule expressing strains and two (K3, K10) of the four group I/II strains. Sequences flanking region 1 of the K4 antigen biosynthesis genes are clearly present in multiple copy on the chromosome of some capsule expressing strains. Consideration of the sizes of the *Bam*HI and *Hpa*I fragments in which the repeat sequences lie indicates that these repeats are not all immediately adjacent to each other. For instance, if the *Hpa*I homologues of some probes were arranged in tandem they could not span the *Bam*HI homologues of the same probe. Probes I, J and K hybridised to single chromosomal fragments in the DNA from the other two group I/II strains (K11, K19) although hybridisation with probe J was weak (Figure 5.8cd) and probe I failed to hybridise with *E. coli* K19. These two group I/II capsule producing strains also differed from the other group I/II strains in hybridisation with probes D, E and F (Table 5.2). Probes J and K produced a similar *Bam*HI restriction fragment length profile in each of the encapsulated *E. coli* indicating that the homologues of probes J and K maybe contiguous as J and K themselves are in pRD10 (Figure 5.8ac).

5.2.4 Analysis of K9 chromosomal DNA

Group I and I/II encapsulated *E. coli* and LE392 do not carry sequences homologous either to the group II capsule gene cluster or a certain amount of DNA adjacent to region 1. The organisation of the *E. coli* chromosome therefore differs at the *serA* region. What, if anything, is present at this site in *E. coli* chromosomes which lack group II *kps* is far from clear. At least in the K10 and K54 group I/II capsule expressing strains, capsule genes distinct from group II genes must be present in this region of the chromosome though not necessarily at the same site as group II *kps*. In group I capsule producing strains there are no clues. If capsule genes were inherited near *serA* it might be predicted that in clones not inheriting such genes the sequences which lie either side of the group II capsule genes are contiguous. The probe closest to region 1 of the group II capsule gene cluster which hybridised to LE392, and the K9 and K30 (group I) capsule expressing strains was probe J. The probe closest to region 3 of the K1 capsule gene cluster which hybridised to these strains was probe D. Probe H was the closest probe to region 3 of the K4 capsule genes that was used. To determine whether sequences represented by probes D and J are in fact contiguous on the chromosome of strains which lack capsule genes at *serA*, the K9 (group I) capsule expressing strain was chosen for Southern blot analysis. The hypothesis was that if sequences homologous to probes D and J are adjoining then probes D and J should hybridise to a very similar set of restriction endonuclease cleavage fragments in the chromosomal DNA of the strain in question. The data already collected are inconclusive, for probes D and J hybridise to different sized *Bam*HI fragments whilst probes H (equivalent to D) and J both detected a similar sized *Hpa*I fragment (Tables 5.2 and 5.3).

5.2.4.1 Sequences flanking the Group II capsule genes are not contiguous in *E. coli* K9

Chromosomal DNA from the K9 capsule producing strain was cleaved with a variety of restriction endonucleases with a 6 basepair recognition sites, in both single and double digests. Two Southern blot filters were prepared and hybridised with radio-labelled probe D (Figure 5.9a). Bound probe D was then removed from the nylon filters which were re-hybridised with probe J (Figure 5.9b). A comparison of the patterns of hybridisation with the two probes clearly showed that the probes D and J do not hybridise to any fragments of the same size (Figure 5.9). Probe D hybridised to a 3kb *Hpa*I fragment, probe J to a 3.5kb fragment (Table 4). Probe D hybridised to relatively large *Kpn*I, *Eco*RV and *Cla*I fragments whilst probe J hybridised to much smaller fragments. This consistent failure to hybridise to the same restriction enzyme fragments indicates that on the chromosome of *E. coli* K9 the sequences homologous to probes D and J are not contiguous. It is highly unlikely that all the restriction enzymes used in the analysis cleave at the site between the sequences homologous to the probes. Thus, in the K9 capsule expressing strain the sequences which flank the group II *kps* determinants in other strains are not contiguous. If sequences which flank the group II capsule genes are not contiguous when group II genes are absent,

then DNA, which is less than 70% homologous to the group II *kps* genes, must lie between them.

There was a difference in the hybridisation pattern of probes D and H. In the previous analysis probe H which is equivalent to probe D hybridised to *Hpa*I fragments of 3kb and 3.5kb in the DNA from the K9 antigen producing strain (Table 5.2). From this it is concluded that probe H extends further from the region 3 of the K4 capsule genes than does probe D from the K1 capsule genes and therefore hybridises to an additional 3.5kb *Hpa*I fragment in the K9 chromosomal DNA. The 3.5kb fragment identified by probes H and J are therefore not the same (Tables 5.2 and 5.3). Probes D and H also hybridised differently to the DNA from the K11 antigen expressing strain when probe D hybridised but probe H did not (Table 5.2). Although probes D and H originate from comparable positions in pKt172 and pRD1 respectively (upstream from region 3) their DNA sequence is not entirely the same.

5.2.5 Sequences flanking the K4 antigen gene cluster are not P-fimbrial adhesin genes

A most striking observation is that sequences which flank region 1 of the K4 antigen gene cluster are present in multiple copy on the chromosome of some strains, namely those expressing the K1, K2, K3, K4, K5 and K10 antigens (Table 5.3). These repeats are not tandem repeats but may be at different chromosomal locations. The function(s) of the product(s) (if any) of this repeated DNA is unclear. Whatever the role of these sequences they do appear to be restricted to isolates which express group II capsular antigens, these are more frequently associated with extra-intestinal disease than others although the origin of the test strains is unfortunately not known.

It is interesting to note that some isolates of *E. coli* produce multiple fimbrial adhesins many of which are considered to be important virulence factors. These are proteinaceous hair-like structures which extend through the capsule. Fimbrial adhesins are composed of an antigenically distinct subunit with a receptor at the tip (Hinson and Williams, 1989). The receptors are specific to the oligosaccharide moiety of mammalian molecules and allow the bacterium to bind to a mammalian cell, a vital step in colonisation (Ørskov and Ørskov, 1990b). *E. coli* has the ability to express many different fimbrial adhesins which fall into several categories including type 1, P-fimbriae and S-fimbriae (Jann and Hoschützky, 1990; Ørskov and Ørskov, 1990b). Adhesins are important in colonisation of epithelia by *E. coli* and are believed to be important virulence factors (Tennent *et al.*, 1990). Unlike the O and K antigens, fimbrial adhesins exhibit phase variation (Smyth, 1986). An individual strain may have the ability to express several different adhesins but an individual cell may express only one or two of these adhesins at any one time. The appropriate genes are switched on and off such that within the population all of the adhesins are expressed.

Adhesin genes can be spread over a number of clones and P- and type I fimbrial

adhesin genes are particularly widespread in their occurrence (Selander *et al.*, 1986). Adhesin genes can be found either linked or dispersed at various loci on the *E. coli* chromosome (Hacker, 1990; High *et al.*, 1988; Krallmann-Wenzel *et al.*, 1989). There is DNA homology between gene clusters which encode serologically distinct fimbriae of the same type (Hacker, 1990). Therefore, Southern hybridisation with fimbrial gene probes might reveal multiple bands as did probes I, J and K. In addition, P-fimbrial adhesin genes have been found at the *serA* region of the chromosome (Hacker *et al.*, 1990; Hull, *et al.*, 1986) and this is the origin of the fragment which has been cloned in pRD10. *E. coli* strains with three P-fimbrial adhesin gene clusters have been identified (Plos *et al.*, 1989; Hull *et al.*, 1988; Authur *et al.*, 1990) an observation entirely consistent with the hybridisation patterns of the probes I, J and K in which strains with no, one or several homologues of pRD10 were found. It is therefore possible that the sequences adjacent to the K4 antigen gene cluster, contained on pRD10 encode adhesins, possibly P-fimbriae.

Several P-fimbrial adhesin gene clusters have been cloned and analysed to date (for a review see Tennent *et al.*, 1990). Original genetic analysis centred around the F13 gene cluster (*pap*) and homologous gene clusters encoding different P-fimbriae (with different designations) have since been described. The genes encoding serologically distinct P-fimbriae share a common genetic organisation. A cluster of 9-13kb encodes nine proteins which include the major subunit (PapA), the adhesin proteins (PapE, F and G), regulatory proteins (PapB and I) and proteins involved in export and anchorage of the pilus (PapC, D and H). DNA sequence and restriction enzyme site homology is apparent between gene clusters encoding serologically distinct P-fimbriae even in the *papA* equivalent which is the most variable domain; serological classification of the fimbriae is dependent on the antigenicity of the major subunit. In addition, the regulatory genes (*papB* and I) are similar to those in the S-fimbrial adhesin gene cluster (Göransson *et al.*, 1988). Different P-fimbrial adhesin gene clusters can be aligned on the basis of shared restriction enzyme sites (van Die *et al.*, 1986). pRD10 does not, however, contain the highly conserved *Cla*I, *Hind*III and *Bgl*II (no sites in pRD10) cleavage sites present within several different clusters. Preliminary experiments were performed to test the hypothesis that pRD10 contains adhesin genes, possibly P-fimbrial adhesin genes.

5.2.5.1 Haemagglutination and haemolysis assays

Haemagglutination is the standard *in vitro* assay for adhesins (Jann and Hoschützky, 1990), different fimbriae recognise different erythrocytes and agglutination can be mannose resistant (MRHA) or mannose sensitive (type I fimbriae). To test the hypothesis that sequences adjacent to region 1 of the K4 capsule gene cluster are adhesin genes a haemagglutination assay was performed. *E. coli* U1-41, LE392, and LE392 harbouring pRD1, pRD10 and the cosmids from the U1-41 library which were homologous to probe A but not to probe B were tested (Figure 5.6). These cosmids carry more U1-41 genomic DNA than pRD10 and may also encode adhesins. Neither the recombinant strains nor LE392 alone promoted agglutination of human or sheep

erythrocytes whilst U1-41 exhibited MRHA of both cell types, consistent with the expression of adhesins. None of the recombinants encode these functions but this does not exclude the possibility that adhesins which agglutinate erythrocytes other than sheep or human are expressed.

P-fimbrial genes can be linked to α -haemolysin genes (*hly*) and multiple copies of *hly* have been identified in strains with multiple P-fimbrial adhesin genes (Hull *et al.*, 1988). The repeated sequences on pRD10 could therefore encode haemolysin. The same set of test strains were grown on blood agar plates but the wild type U1-41 was not haemolytic and neither were the recombinants. *E. coli* 20025, which is known to produce α -haemolysin (High *et al.*, 1988) was used as a positive control in this analysis. This repeated DNA does not encode α -haemolysin.

5.2.5.2 Southern blot analysis with F14 adhesin gene probes *pap*

Some P-fimbrial adhesin genes have been cloned previously in this laboratory and were therefore available for study. The F14 gene cluster is contained on a 12.5kb *Bam*HI fragment within the plasmid pGB792 (High *et al.*, 1988; High, N.J. PhD thesis 1988, University of Leicester) (Figure 5.10). To determine whether pRD10 and the U1-41 cosmids share any DNA homology with the F14 fimbrial adhesin genes Southern blot analysis was used. pGB792, pRD1, and the cosmids flanking the K4 capsule genes were cleaved with *Bam*HI, pRD10 was cleaved with *Hpa*I and all were probed with the entire 12.5kb *Bam*HI fragment from pGB792 (Figure 5.10). The filter was washed in 0.5xSSC exposed to film for 48 hours. The probe bound very strongly to the 12.5kb *Bam*HI fragment in pGB792 and exceedingly weakly to pRD10 and the cosmids. The probe bound to the 15.5kb *Bam*HI fragment cloned in pRD10 present in some of the recombinants, but not to the 5.2kb region 1 *Bam*HI fragment. Hybridisation was far more pronounced to pGB792 than to the U1-41 DNA in the recombinants (Figure 5.10).

To compare the degree of hybridisation of the F14 gene probe to the U1-41 recombinant DNA with its hybridisation to another P-fimbrial adhesin gene cluster another probe was taken from pGB792. This was a 3.5kb *Kpn*I-*Bam*HI fragment (Figure 5.10) which encodes the F14 adhesin genes (PapE, F and G equivalents). The adhesin genes themselves are largely conserved between different P-fimbrial adhesin clusters although the *pap*G equivalent may vary slightly with receptor specificity (Tennent *et al.*, 1990). pGB792 was again included in the Southern blot analysis of the recombinants and this time a plasmid containing *pap*E, F and G from the F13 gene cluster was also included. The probe hybridised to the same set of fragments in the U1-41 recombinants as the entire F14 gene cluster had done. Hybridisation to pGB792 was again pronounced as was the hybridisation to the equivalent genes from the F13 gene cluster (data not shown).

Hybridisation of the F14 *pap* probes with the recombinants carrying sequences flanking the K4 antigen gene cluster was consistently weak. Clearly, the sequences con-

tained on pRD10 and the related plasmids do not share the same degree of homology with the F14 genes as do the F13 genes and given the results of the haemagglutination assays it is unlikely that the sequences adjacent to the K4 capsule gene cluster direct the expression of P-fimbrial adhesins.

5.3 Discussion

DNA probes from regions 1 and 3 of the K1 antigen gene cluster detect highly homologous sequences in the DNA of group II capsule expressing strains (Roberts, M. *et al.*, 1988) (Figure 5.2) but did not hybridise to chromosomal DNA of group I capsule expressing *E. coli* (Figure 5.2, Table 5.1). This was the case under conditions of low stringency requiring 70% DNA homology for hybridisation. Group I capsule producing strains have therefore not acquired (or retained) the ability to express a group II capsule. Although the transport mechanisms encoded by regions 1 and 3 are capable of handling a diversity of group II polysaccharides, ranging from the homopolymer K1 to the substituted heteropolymer K4 (Chapter 3), these results suggest that group I capsule expression is independent of these mechanisms. This is not surprising given the observed number of biochemical differences between group I and group II capsular polysaccharides and the non-allelic nature of their genetic determinants. The failure of the group II *kps* gene probes to hybridise to the chromosomal DNA of the group I capsule expressing strains is more likely to be due to an absence of group II genes than to extensive sequence divergence. A parallel can be drawn with P-fimbriae and haemolysin: strains which do not express these traits do not carry the appropriate genes (Hull *et al.*, 1988; Plos *et al.*, 1989). The unencapsulated laboratory strain of *E. coli*, LE392, also lacks the group II capsule gene cluster regions 1 and 3 as assessed by Southern blot analysis (Echarti *et al.*, 1983).

Similarly, sequences with over 70% DNA homology to the group II capsule gene cluster were not detected in the K3, K10, K11, K19 or K54 (group I/II) capsule expressing strains (Table 5.1 and Figure 5.2) even though both group II and group I/II capsule genes have been mapped to the same region of the chromosome. The genes (*kps*) involved in the expression of chemically different group II K antigens are highly homologous to each other (less than 5% sequence divergence) but this degree of homology does not extend to include the group I/II capsule genes which must be either completely different or over 70% divergent from those of group II. Hence, expression of group I/II capsules (except K2) appears not to utilise the group II capsule gene products nor are these genes present on the chromosome. This is a surprising result given the observed biochemical similarities between the group II and group I/II capsules. Consequently, group I/II polysaccharides must be encoded by genes distinct from the group II *kps* genes. Studies of CMP-KDO synthetase activity lead Finke *et al.* (1990) to postulate the existence of a third group of *E. coli* capsular polysaccharides. With the additional genetic information presented here the existence of such a group, encoded by genes distinct from the group II capsule determinants (*kps*), probably located at the same region of the chromosome near

serA, is formally proposed.

The existence of another set of capsule genes in *E. coli* is intriguing. The group I/II and group II capsules are biochemically similar but their genetic determinants are different. The *H. influenzae* and *N. meningitidis* capsule genes are also over 30% divergent from those of *E. coli* group II but these three sets of genes (*bez*, *ctr*, *kps*) are considered to be related and capsule biogenesis in these organisms share common features (Section 1.3.3.4). It is feasible, therefore, that the *E. coli* group I/II K antigens belong to the same family of capsules as those of group II; the respective genes may be related despite the observed failure of the group II capsule gene probes to hybridise to the chromosomal DNA of group I/II capsule expressing *E. coli*. The extent to which the products of the group I/II capsule genes are related to group II capsule gene-encoded products awaits analysis of the group I/II capsule genes which is currently underway. Of course, the possibility that group I/II K antigens (or indeed those of groups I and II) can be subdivided cannot be overlooked. Only the K10 and K54 capsule genes have been mapped to the *serA* region of the chromosome and others could be located elsewhere. The only proven similarity between the genes encoding chemically different group I/II K antigens is a lack of homology with the group II *kps* genes. Evidence for an identical chromosomal location of the K1, K4, K5 and K10 genes is presented later.

The term *kps* does not now seem appropriate to encompass all capsule genes near *serA* for this term now includes more than one set of genes: the well defined group II genes and at least the K10 and K54 group I/II capsule genes. Perhaps *kps* should be used only to describe the common group II capsule genes present within regions 1 and 3 of the cluster and an alternative designation used to refer to those of group I/II.

The K2 antigen was placed in group I/II by Finke *et al.* (1990) but a strain expressing the K2 antigen was shown in this study to carry the group II *kps* genes on its chromosome. Both the K2 and K19 antigens were found to be expressed at low growth temperatures (18°) (Jann and Jann, personal communication) and to have a low CKS activity at 37°C (Finke *et al.*, 1990): characteristics of group I/II capsules. However, Ørskov *et al.* (1984) found that both K2 and K19 capsules were not expressed at 18°: a characteristic of group II capsules. This discrepancy has not been resolved. The results of the Southern blot analysis described here suggest that K19 is a group I/II antigen (see above) and K2 a group II antigen since sequences homologous to regions 1 and 3 of group II *kps* were found on the chromosome of a K2 antigen expressing strain (Figure 5.2b,c, Table 5.1). In addition, a *Bam*HI fragment of the same size and with homology to the region 1 probe was identified in the chromosomal DNA of the K2 capsule expressing strain tested suggesting that region 1 is highly homologous. Whether the homologues of regions 1 and 3 participate in expression of the K2 capsule is not clear. Since a central role has been implicated for CKS in group II capsule expression, perhaps in supplying CMP-KDO for the initiation of polymer synthesis, it is surprising that a group II K antigen expressing strain has a low CKS activity. Interestingly, the K10 (group I/II) antigen is also substituted with KDO at the reducing end but also has a low CKS activity. CKS

is encoded by *kpsU* which lies within the 5.2kb *Bam*HI fragment in region 1 of the K5 capsule gene cluster (Pazzani *et al.*, in preparation) and therefore *kpsU* may be present in the equivalent *Bam*HI fragment in *E. coli* K2 but for some reason an elevated CKS activity is not detected. Capsule expression in this K2 antigen expressing strain could make use of the *kdsB*-encoded CKS which produces CMP-KDO for LPS core biosynthesis. It is possible that CMP-KDO does not play an important role in group II capsule biogenesis. Alternatively, K2 capsule expression in this strain may not conform with the other group II capsules. It is also possible that the *kps* genes are silent and that capsule expression makes use of alternative genes. If the K2 capsule genes were cloned some of these problems could be resolved.

DNA probes from the sequences either side of the K1 and K4 capsule gene clusters were also used as probes in Southern blot analysis. The number of strains analysed with these probes was very low and the pattern of hybridisation very variable, therefore, the conclusions that have been drawn from the results are generalisations and must be viewed as such. Not surprisingly, given the stability of the *E. coli* chromosome, sequences flanking the group II K antigen gene cluster are common components of the *E. coli* chromosome, similar enough to permit detection by cross-hybridisation. Although common, these sequences are not universal: the sequence flanking region 3 of the K1 and K4 capsule genes were largely absent from the chromosome of the K11 and K19 encapsulated strains (Table 5.2) and the sequence adjacent to region 1 absent from the K29 capsule expressing strain (Table 5.3). Sequences which are adjacent to region 3 of the K1 and K4 capsule genes (probes D, E, F and H, Figure 5.1) hybridised to single fragments in the chromosomal DNA of most strains regardless of capsule status (Figure 5.3, Table 5.2). Sequences adjacent to region 3 hybridised in a more complex fashion and are considered below.

Adjacent to region 1 of group II *kps* about 3.5kb (probes C and G) is only found when the group II *kps* genes are present on the chromosome (Figure 5.1; Table 5.3). A limited amount of nucleotide sequence has been determined in this region from pGB110 and pGB108 which encodes the K7 capsule gene cluster (Roberts *et al.*, 1986). In the DNA upstream of *kpsE*_{K5} an ORF was detected on the same strand as *kpsE*. The ORF which is at least 707bp starts before the available sequence information and terminates 71bp upstream of *kpsE* (Pazzani, C., unpublished data). An homologous ORF was detected upstream of *kpsE*_{K7} (Smith, A.N., unpublished data). There is no evidence to suggest that this ORF (either in K5 or K7) is expressed. Neither preliminary deletion analysis using pGB110 nor transposon mutagenesis using pKT274 (Boulnois *et al.*, 1987) suggest that the DNA upstream of *kpsE* is necessary for capsule expression although this possibility cannot be precluded. The high degree of similarity in this sequence upstream of region 1 of the K1, K2, K4, K5 and K7 capsule genes (as shown by both nucleotide sequence analysis and the Southern blot data presented here) and its apparent restriction to group II capsule expressing strains is intriguing.

Further upstream from probes C and G which flank region 1 of the K4 capsule gene cluster, another 2.5kb (Probe K, Figure 5.1; Table 5.3) is only present on the chromosome of strains expressing a group II or group I/II capsule. Beyond this 2.5kb

of DNA the next 10kb of DNA (Probes I and J, Figure 5.1; Table 5.3) is present in all strains tested except the K29 antigen expressing strain. The hybridisation patterns of probes derived from this DNA (probes I, J and K, Figure 5.1; Table 5.3) is consistent with it being repeated at varying locations on the chromosome of all the group II and two of the four group I/II capsule expressing strains. *E. coli* LE392, the group I capsule strains and the K11 and K19 capsule strains have only one copy of this sequence and *E. coli* K29 has none. Despite the widespread and variable occurrence of adhesin genes in the *E. coli* population, the repeated sequence does not seem to direct the expression of adhesins because recombinants do not cause haemagglutination (although the test was limited to only two types of erythrocytes) nor is there any sequence with DNA homology to a P-fimbrial F14 gene probe (Figure 5.10). The possibility that adhesins other than P-type, that do not recognise sheep or human erythrocytes are encoded on this repeat DNA cannot be excluded. The role of this repeated DNA remains elusive. Although the group II K antigen expressing strains used in this study are consistent in their possession of multiple fragments homologous to probes upstream of region 1 (probes I, J and K, Figure 5.1), within *E. coli* strains expressing either groups I or I/II capsules two patterns of homology were observed. The polymorphism associated with this repeat sequence may be independent of capsule type.

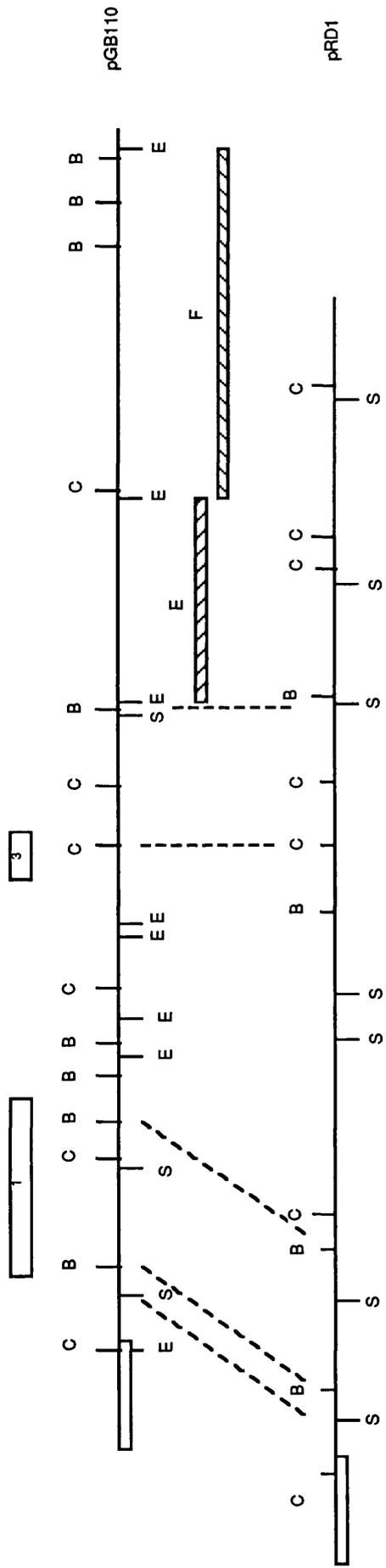
Analysis of the sequence adjacent to region 3 of the cloned group II capsule genes gave some interesting results. The DNA upstream of region 3 of the K1 (probes E and F, Figure 5.1), K4 and K5 capsule genes is homologous and in the case of K1 and K5 at least appears to be organised similarly (Figures 5.4c,d and 5.5). This strongly suggests that in K1, K4 and K5 capsule expressing strains their respective determinants are located at the same site on the chromosome. K1 *kps* has been mapped to 64 map units on the physical map of the *E. coli* chromosome with region 1 orientated towards *serA* (Vimr, 1991). This was achieved using the DNA flanking the K1 capsule gene cluster to probe the set of overlapping fragments generated by Kohara *et al.* (1987) which together represent the entire *E. coli* genome. The probes hybridised to a particular recombinant whose origin was known and the position of the K1 capsule gene clusters on the physical map of the chromosome calculated. Since the DNA flanking the K4 and K5 capsule genes is homologous to that flanking the K1 antigen genes it is predicted that the flanking sequences would hybridise to the same recombinant. Therefore, the K1 and K5 capsule genes would map to the same position on the physical map as do the K1 capsule genes. This may be the case for all group II capsule gene clusters. The K10 (group I/II) biosynthesis genes have very recently been cloned but not yet characterised in detail (Pearce, R., Roberts, I.S. and Boulnois G.J., personal communication). It has been demonstrated that a cosmid which directs surface expression of the K10 antigen in *E. coli* LE392 has DNA homology with probes flanking both ends of the K4 antigen gene cluster (probes K and H respectively, Figure 5.1). The region of DNA homology with the two probes is separated by about 20kb in which the K10 antigen biosynthesis genes presumably lie. Thus, sequences which flank the group II capsule genes (K1, K4 and K5) and the group I/II capsule genes (K10) are homologous and by the argument above these different *E. coli* capsule genes probably map to the same position on the chromosome. Further analysis of the cloned K10 genes may reveal exactly how different these gene clusters really are.

Presumably, the 8kb *Bam*HI fragment detected in the chromosomal DNA of the K9 and K30 capsule expressing strains which has homology to the sequence upstream of region 3 (probe D, Figure 5.1) could also be mapped to 64 map units. Since this fragment has not been cloned perhaps it could be orientated on the physical map of the *E. coli* chromosome by virtue of its restriction enzyme sites (Médigue *et al.*, 1990).

Although it was not the original intention of the study, the probes taken from either side of the group II capsule gene cluster have been used to perform a restriction fragment length polymorphism (RFLP) analysis of the test strains. The probes revealed distinctive hybridisation profiles in the different strains. In more comprehensive studies, RFLP profiles have tended to fall within clonal boundaries as assessed by MLEE and have given a measure of sequence diversity (Arthur *et al.*, 1990; Harshman and Riley, 1980; Krawiec and Riley, 1990; Milkman and Bridges, 1990). The RFLP study reported here is insufficient to expose clonal boundaries although the striking similarity in RFLP profiles of the K9 and K30 capsule expressing strains with all the probes may indicate a common clonal descent. It has previously been suggested that *E. coli* K-12 is related to *rfe*-dependent O antigen producing strains (Kido *et al.*, 1989) and this is consistent with the observation made here in which the profile of *E. coli* LE392 was more like that of the group I strains than the other encapsulated *E. coli*.

Figure 5.5: Alignment of pGB110 and pRD1

Linear maps of the pGB110 and pRD1 which carry the K4 and K5 capsule gene clusters respectively. The open boxes denote the vector sequences. The numbered boxes above the lines refer to the capsule gene regions. The hashed boxes below pGB110 refers to the *EcoRI* fragments with homology to probes E and F which were derived from pKT172 as shown in Figure 5.1. The restriction enzyme site homology between pRD1 and pGB110 is highlighted by the dotted lines. Restriction enzyme target sites: B, *Bam*HI; C, *Cla*I; E, *Eco*RI; S, *Sma*I.



Alignment of pRD1 and pGB110

Figure 5.6: Construction and analysis of pRD10

(a) Ten recombinant plasmids from the *E. coli* U1-41 genomic library containing sequences homologous to probe A were cleaved with *Bam*HI and analysed by agarose gel electrophoresis. The cosmids have been labelled A through J above each track, where J corresponds to pRD1. The greater than 12kb *Bam*HI fragment common to seven of them is marked with a single arrow and the region 1-derived 5.2kb fragment with a double arrow (this band is barely visible in track J).

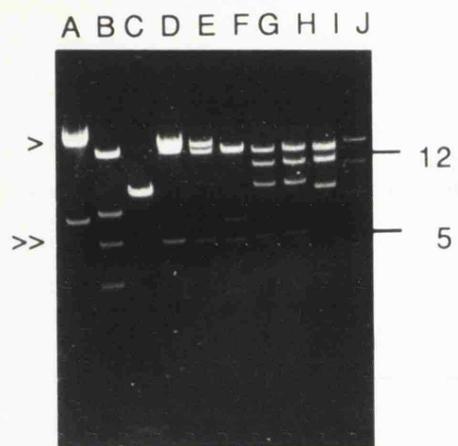
(b) Southern blot analysis of the gel shown in (a) with radiolabelled probe C (Figure 5.1).

(c) pRD10 was cleaved with a variety of restriction enzymes and analysed by agarose gel electrophoresis. Plasmid DNA was cut with the following enzyme(s) in each track: Lane 1, *Bam*HI ; 2 *Hind*III; 3 *Hpa*I ; 4 *Cla*I ; 5 *Hpa*I + *Sst*I; 6 *Hpa*I + *Hind*III; 7 *Cla*I + *Sst*I; 8 *Bam*HI + *Sst*I; 9 *Cla*I + *Hind*III; 10 *Hpa*I + *Bam*HI . Molecular weight markers (kb): 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1.6, 1.0, 0.5.

(d) Southern blot analysis of the gel displayed in (c) with radiolabelled probe C (Figure 5.1).

Figure 5.6

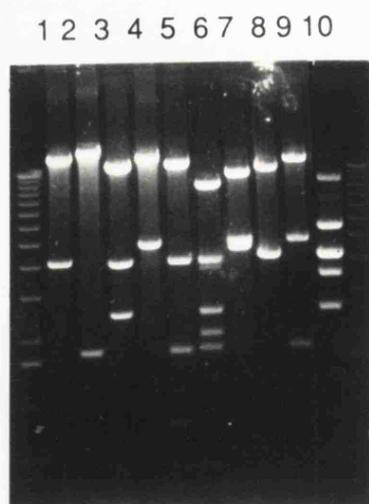
a)



b)



c)



d)

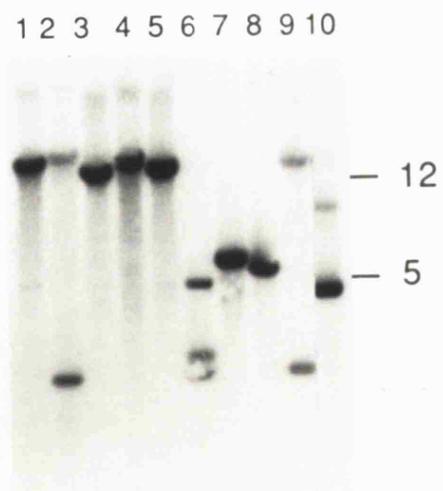
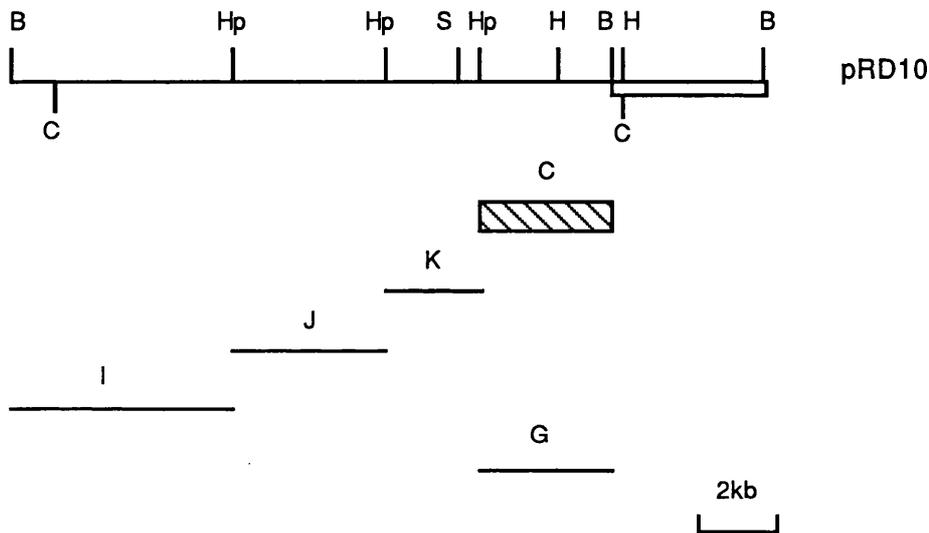


Figure 5.7: Linear map of plasmid pRD10



Linear map of plasmid pRD10 which contains a 15.5kb genomic *Bam*HI fragment from *E. coli* U1-41 which lies adjacent to the 5.2kb *Bam*HI fragment in region 1 of the K4 capsule gene cluster on the chromosome (see text). The open box denotes the vector sequence pACYC184. The lines below the plasmid represent the probe fragments chosen for study and the hashed box marks the region of homology with probe C. Restriction enzyme target sites: B, *Bam*HI; C, *Cla*I; H, *Hind*III; Hp, *Hpa*I; S, *Sst*I.

Figure 5.8: Hybridisations with probes taken from pRD10

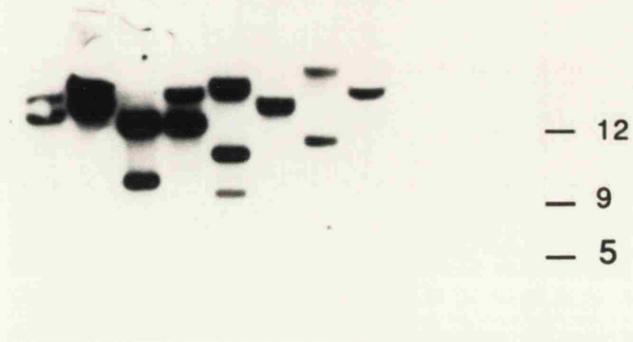
Southern blot analysis of chromosomal DNA from strains expressing different capsular polysaccharides, as indicated above each track. Chromosomal DNA was cleaved with either *Bam*HI or *Hpa*I and subjected to Southern blot analysis with one of the three probes I, J or K. The bands corresponding to the hybridisation of probe J to chromosomal DNA from the K11 and K19 capsule expressing strains have not photographed well. The chromosomal digest and the probe with which each filter was analysed were as follows:

- (a) Probe K, *Bam*HI
- (b) Probe K, *Hpa*I
- (c) Probe J, *Bam*HI
- (d) Probe J, *Hpa*I
- (e) Probe I, *Bam*HI
- (f) Probe I, *Hpa*I

Figure 5.8

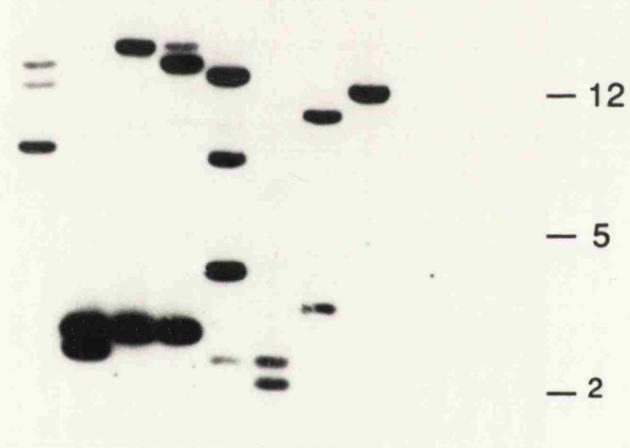
a) Probe K, BamHI

1 4 5 2 3 10 11 19 9 29 30



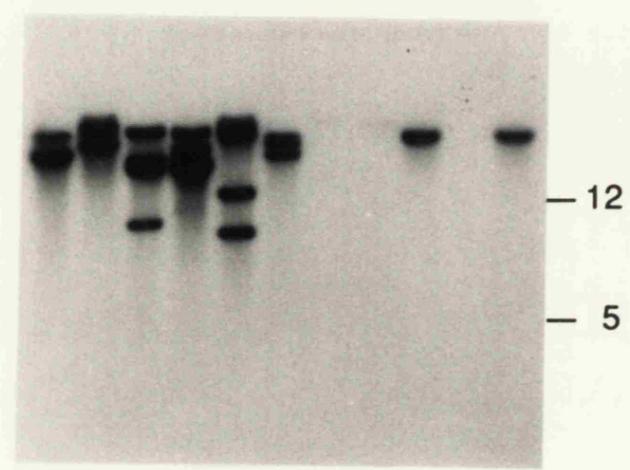
b) Probe K, HpaI

1 4 5 2 3 10 11 19 9 29 30



c) Probe J, BamHI

1 4 5 2 3 10 11 19 9 29 30

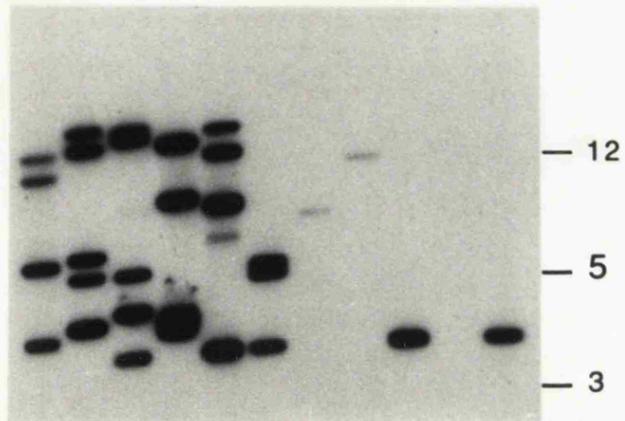


continued
overleaf...

Figure 5.8 continued...

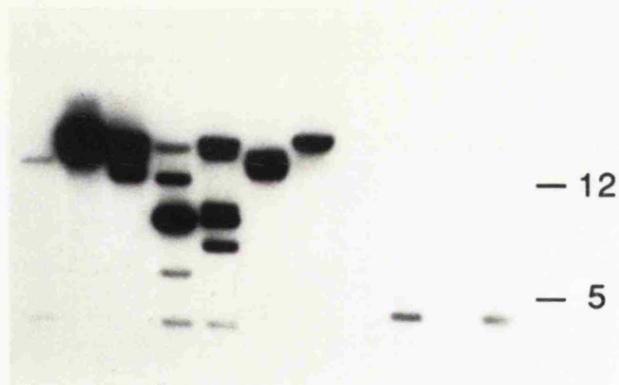
d) Probe J, HpaI

1 4 5 2 3 10 11 19 9 29 30



e) Probe I, BamHI

1 4 5 2 3 10 11 19 9 29 30



f) Probe I, HpaI

1 4 5 2 3 10 11 19 9 29 30

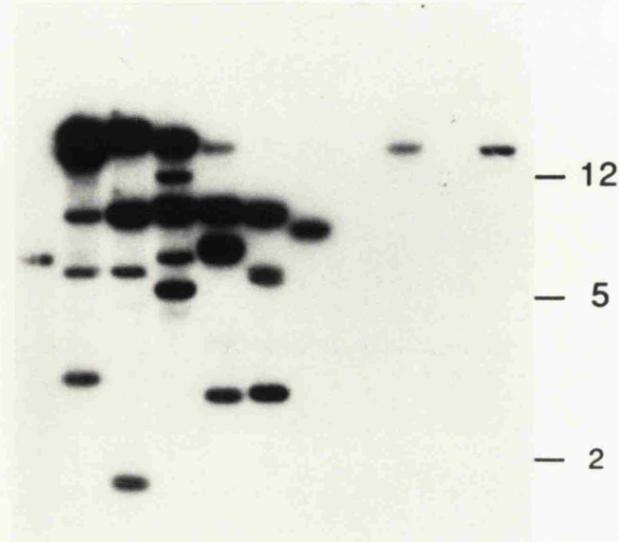


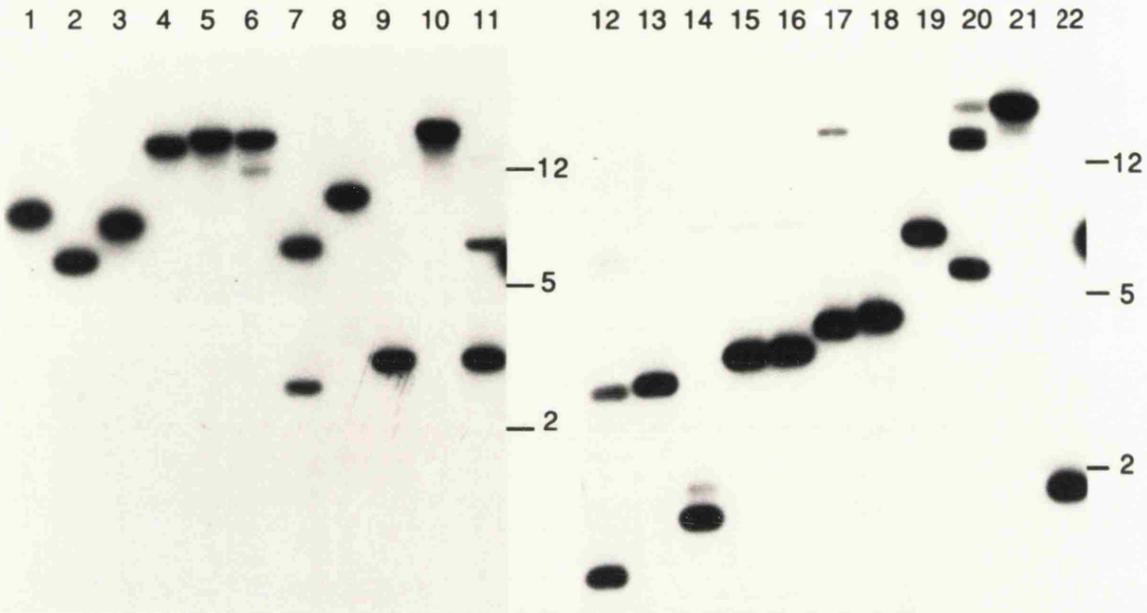
Figure 5.9: Southern blot analysis of *E. coli* K9 chromosomal DNA

Chromosomal DNA from the K9 capsule expressing DNA was cleaved with the restriction endonuclease(s) indicated below and hybridised first with (a) probe D and then (b) probe J. The same two Southern blot filters were hybridised with each probe.

Lane:	1	<i>Bam</i> HI
	2	<i>Eco</i> RI
	3	<i>Pst</i> I
	4	<i>Hind</i> III
	5	<i>Kpn</i> I
	6	<i>Cla</i> I
	7	<i>Pvu</i> II
	8	<i>Dra</i> I
	9	<i>Hinc</i> II
	10	<i>Eco</i> RV
	11	<i>Hpa</i> I
	12	<i>Cla</i> I + <i>Pvu</i> II
	13	<i>Hinc</i> II + <i>Pvu</i> II
	14	<i>Cla</i> I + <i>Hinc</i> II
	15	<i>Kpn</i> I + <i>Hinc</i> II
	16	<i>Hinc</i> II + <i>Hind</i> III
	17	<i>Cla</i> I + <i>Hind</i> III
	18	<i>Hind</i> III + <i>Eco</i> RV
	19	<i>Cla</i> I + <i>Eco</i> RV
	20	<i>Cla</i> I + <i>Kpn</i> I
	21	<i>Kpn</i> I + <i>Hind</i> III
	22	<i>Hinc</i> II + <i>Eco</i> RV

Figure 5.9

a)



b)

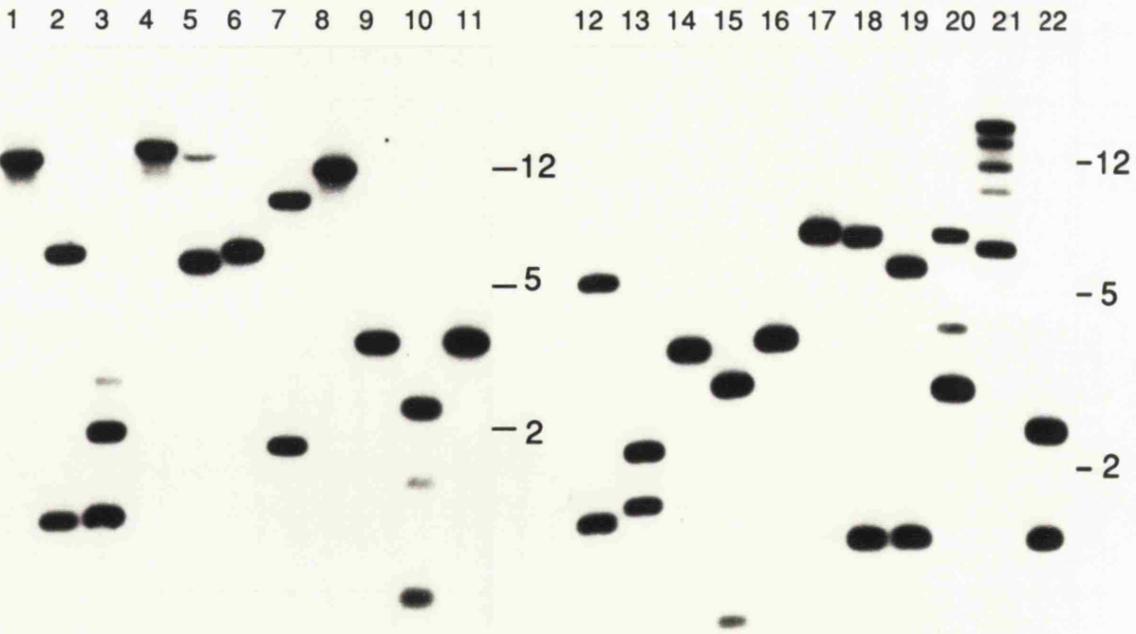


Figure 5.10: Investigation with the F14 gene cluster

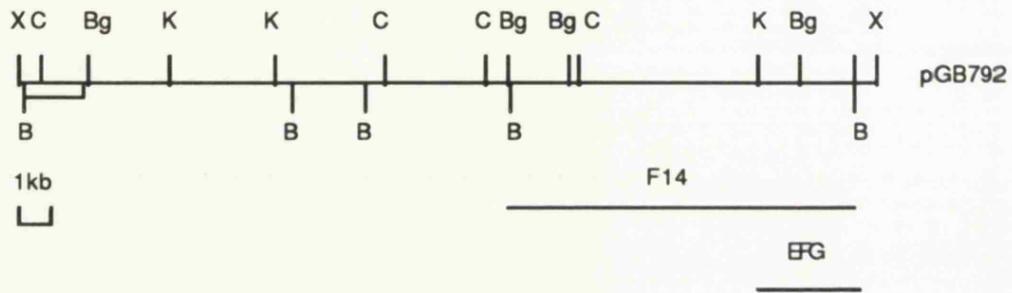
(a) Physical map of the recombinant plasmid pGB792 which encodes the entire F14 P-fimbrial adhesin gene cluster. The line marked F14 denotes the 12.5kb *Bam*HI fragment used as the F14 gene probe and that marked EFG denotes the smaller probe fragment which carries the *pap*E,F and G equivalents. The open box denotes the vector sequence. Restriction enzyme target sites: B, *Bam*HI; Bg, *Bgl*II; C, *Cla*I; K, *Kpn*I; X, *Xho*I.

(b) Agarose gel showing six of the cosmid clones isolated from the U1-41 library (Chapter 3) which have homology with region 1 of the group II capsule genes. The cosmids have been labelled according to the scheme adopted in Figure 5.6(a). Lane 7 contains pRD10 cleaved with *Hpa*I and lane 8 contains pGB792 cleaved with *Bam*HI. Molecular weight markers (kb): 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1.6, 1.0, 0.5.

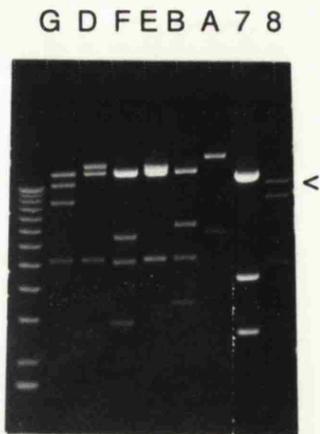
(c) Southern blot analysis of the above gel using the F14 gene probe after a 48 hour exposure to film.

Figure 5.10

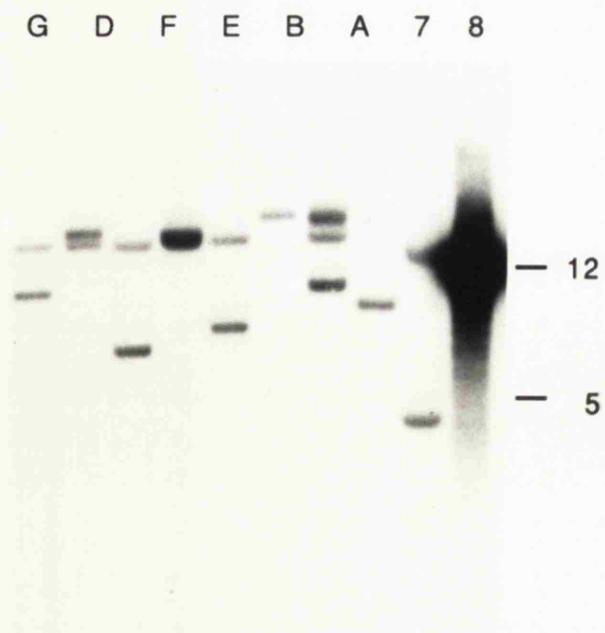
a)



b)



c)



Chapter 6

Attempted cloning of the K9 biosynthesis genes

6.1 Introduction

In order to understand the genetic basis of capsule variability in *E. coli* it is essential to investigate all the systems involved in capsule production. It is not sufficient to restrict a study to the capsule genes located near *serA*. The group II capsule genes, particularly those of K1 and K5 have been studied in detail and an analysis of the K10 antigen (group I/II) has recently been initiated. If the group I capsule genes were also cloned, fundamental questions concerning capsule gene diversity in *E. coli* could be addressed. Although there is no detectable DNA homology between the group II capsule genes and any others (Chapter 5), the capsule genes of *E. coli* could all be variants of a common theme with a common ancestry. Another possibility is that the group I capsule genes originated as a duplication of the LPS biosynthesis genes and are unrelated to group II. Certainly some relationship between LPS and group I K antigens seems likely in the light of biochemical evidence. Comparisons between the group I capsule genes and those of other surface polysaccharides may reveal why the K antigens of *E. coli* fall into more than one distinct group. The mechanisms of group I polysaccharide synthesis and export are not understood. Cloning the group I capsule biosynthesis genes would allow an independent analysis of their function. In this chapter experiments designed to clone the K9 capsule biosynthesis genes are described.

6.2 Results

The group I K9 antigen was chosen for study using *E. coli* 2667 (09:K9:H12). The strategy adopted for cloning the K9 biosynthesis genes was as follows: to generate

a polyclonal antiserum against the K9 antigen and using this antiserum to screen a cosmid library for the expression of the K9 polysaccharide.

6.2.1 Production of a polyclonal antiserum against the K9 antigen

Polyclonal antiserum was raised against *E. coli* 2667 heat-inactivated whole cells. Antiserum against capsular antigens is routinely produced using inactivated cells because the immune response mounted against a purified polysaccharide alone is often weak. The cell acts as a natural adjuvant and promotes an increased immune response to the polysaccharide antigen. Pre-immune serum collected from the animal prior to the challenge procedure was checked for pre-existing antibodies against the K9 antigen by double immunodiffusion analysis using a polysaccharide extract of *E. coli* 2667. No detectable antibodies against *E. coli* 2667 were found.

The anti-serum which was finally collected had been raised against whole *E. coli* 2667 cells and would therefore contain antibodies against bacterial components in addition to the K9 antigen. The specificity of the antiserum was therefore improved by absorption with unencapsulated *E. coli*: LE392 and a K⁻ strain which expressed the homologous O antigen (*E. coli* 2388 09:K29⁻). To assess the presence of K9 antibodies in the serum, polysaccharide was extracted from *E. coli* 2667 and 2388 (K⁻); LE392 was treated in the same way. The cell extracts were analysed by double immunodiffusion with the absorbed serum. A single precipitin line was produced between the serum and the extract from the K9 capsule expressing strain only (Figure 6.1a). The fact that this reaction was unaffected by proteinase treatment indicates that the reactive antigen is not a protein. In this simple test the serum reacted with a single antigen which is expressed by 2667 and not by the unencapsulated smooth strain or the K-12 strain. It is reasonable to suppose therefore that the serum contains antibodies against the K9 antigen. Although the antiserum could be used in the first instance to screen a library for the expression of K9, an immunoreactive recombinant would have to be tested by other means to confirm its identity. Such tests could involve chemical analysis or the use of a K9-specific bacteriophage.

6.2.2 Development of a K9 antigen detection system

It was proposed to use the antiserum against the K9 antigen to screen a cosmid library for the expression of the K9 antigen, the logical technique of choice being by colony immunoblotting. First, the system had to be optimised. *E. coli* 2667 and LE392 (which was to be the host strain for the library) were grown either in mixed culture or separately. Bacteria were either grown directly on nylon filters or transferred to them after overnight growth on agar plates. Nylon was chosen rather than nitrocellulose (usual for protein immunoblotting) because its properties are more suitable for binding negatively charged molecules, more usually DNA but in this case

capsular polysaccharide. The filters were probed with the K9 antiserum using a range of primary antibody concentrations. A final dilution of 1:2000 was chosen for easy distinction between *E. coli* LE392 and the K9 capsule expressing strain and economy of antiserum. At this antiserum concentration a strong positive reaction (purple) was evident with K9 capsule producing colonies and an exceedingly weak reaction with the negative control. Some reaction, however feeble with the negative would be useful in locating any positive recombinants on a background of negatives. Since it was possible that the K9 antigen may not be expressed on the cell surface of a recombinant, for instance if the transport functions were not cloned, the immunoblot was also performed using chloroform-lysed cells and the same results were obtained.

6.2.3 Production of an *E. coli* 2667 cosmid library

The number and nature of the K9 biosynthesis genes is not known and to maximise the chance of cloning the intact K9 biosynthesis genes it was decided to clone large chromosomal DNA fragments using a cosmid vector. *E. coli* 2667 chromosomal DNA was extracted and treated exactly as described in Chapter 3 for the generation of the *E. coli* U1-41 cosmid library from which the K4 antigen gene cluster was isolated. Briefly, chromosomal DNA was partially digested with the restriction endonuclease *Sau*3A generating fragments of 35-45kb which were ligated into the prepared arms of the cosmid vector, pEMBLcos4. The DNA was packaged into lambda heads *in vitro* and infected into *E. coli* LE392. This generated 1700 recombinants. Plasmid DNA was extracted from 24 recombinants and each one was shown to contain insert DNA in addition to the vector itself. Assuming the *E. coli* genome is 4.2×10^6 bp and the average insert size is 45kb then by the equation of Clarke and Carbon (1976) to have a 99% probability of cloning any particular sequence 420 recombinant DNA molecules must be recovered.

6.2.4 Screening the cosmid library for K9 antigen expression

6.2.4.1 *E. coli* LE392 as the host strain

The 1700 recombinants containing *E. coli* 2667 chromosomal DNA inserts were transferred to nylon filters and screened for K9 antigen production using the colony immunoblot procedure. A separate filter carrying *E. coli* 2667 colonies was included in the immunoreactions as a positive control. Screening chloroform-lysed cells was chosen in the first instance as this would ensure detection of all recombinants producing the K9 antigen whether on the cell surface as a complete capsule or not. No obvious positives were detected but 24 recombinants showed very weak colour reactions which although stronger than the majority of colonies on the plate was still considerably paler than the reaction with the wild type (Figure 6.1b). It was decided to study these 24 in more detail.

The weak reaction exhibited by some of the recombinants could have been because the K9 antigen was produced in very low amounts. The immunodetection procedure was therefore modified to concentrate the antigen on the filter. Polysaccharide was extracted from *E. coli* 2667 and LE392 was treated in the same way. Varying amounts of the extracts were spotted onto the filter and the K9 antigen detected as in the colony immunoblots. This technique gave a marked difference in the intensity of the colour reaction between 2667 and LE392, the degree of which increased with the addition of more extract. Polysaccharide was extracted from the 24 weakly reactive recombinants and treated in the same way. Although reaction intensities varied none were stronger than the reaction of the LE392 negative control and none approached the intensity of reaction with the wild type (Figure 6.1c). The colony immunoblot of the entire library was repeated using chloroform lysed cells. Again all recombinants were negative and this time the 24 which previously gave a weak reaction failed to react. Therefore, a recombinant expressing the K9 antigen was not isolated.

6.2.4.2 *E. coli* HB101 as the host strain

At this time Homonylo and Whitfield (1988) reported in an abstract the cloning of the K30 biosynthesis genes in *E. coli* HB101 by cosmid cloning and immunodetection which was the same approach that was adopted in this study. The cloned genes were not sufficient for complete surface expression of a K30 capsule but the K30 antigen was produced, possibly in the cytoplasmic membrane. It was therefore decided to introduce the *E. coli* 2667 cosmid library into the same host strain, *E. coli* HB101. 1500 recombinants were screened by colony immunoblotting using chloroform lysed cells but recombinants expressing the K9 antigen were not detected. This attempt to clone the K9 antigen biosynthesis genes had failed for reasons that are unclear.

6.3 Discussion

An attempt was made to clone the K9 antigen biosynthesis genes using a cosmid vector and a polyclonal anti-K9 serum generated for the specific purpose of screening the library. Such an antiserum was produced by the inoculation of a rabbit with heat-inactivated *E. coli* 2667 (09:K9:H12) followed by absorption of the serum with a K-12 strain and an 09:K⁻ strain. An antibody-based system for the detection of the K9 antigen was developed and judged sufficiently sensitive to screen a library. An *E. coli* 2667 cosmid library was generated in two host strains, LE392 and HB101 but a K9 antigen-expressing recombinant was not isolated.

This is not the only attempt to clone group I capsule genes that has failed. The cosmid clone which was reported to direct the incomplete surface expression of the K30 antigen in HB101 proved to be unstable thus precluding further analysis (Whitfield, C., personal communication). One explanation for the failure to clone the K9 antigen genes is that the genes may not be clustered and cannot be contained on a 45kb

plasmid. There is already doubt as to the number of loci involved in group I capsule expression. The possibility that more than one locus is necessary is supported by the preliminary studies conducted with the K27 antigen. In this case two unlinked loci (the *his*-linked biosynthesis genes and the *trp*-linked 'rfc-like polymerase' genes) were necessary for complete capsule expression (Schmidt *et al.*, 1977). It is not known whether an 'rfc-like polymerase' is necessary for K9 biosynthesis or whether such a function is encoded by the laboratory strains used as hosts for the cosmid library. In addition to K27 capsule expression, expression of the cloned *Klebsiella* K2 biosynthesis genes in HB101 was dependent on the presence of an additional locus which was thought to have a regulatory function (Arakawa *et al.*, 1991). It is quite possible therefore that the K9 capsule biosynthesis genes comprise more than one locus. A wild type strain rather than *E. coli* HB101 was found more useful as a host for the cloned *Shigella sonnei* form I antigen (LPS) (Yoshida *et al.*, 1991). This may also be the case for the *E. coli* group I capsule genes.

Figure 6.1: Immunodetection of the K9 antigen

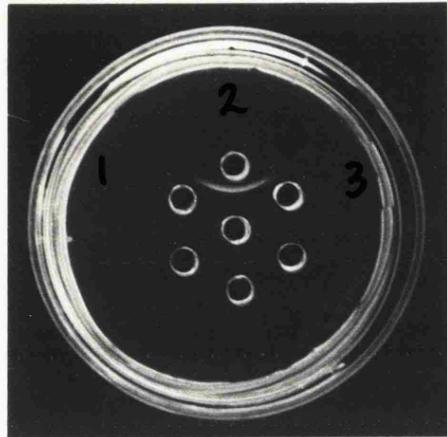
(a) Double immunodiffusion of polysaccharide extracts against the polyclonal antiserum raised against whole *E. coli* 2667 cells after the cross reactive antibodies had been adsorbed with the unencapsulated strains. Antiserum (10 μ l) was placed in the central well and 10 μ l of polysaccharide extracts were loaded as follows: well 1 *E. coli* LE392; 2, *E. coli* 2667 (09:K9:H12); 3, *E. coli* 2388 (09:K29⁻).

(b) A colony immunoblot filter produced in the screening of the *E. coli* 2667 cosmid library in LE392.

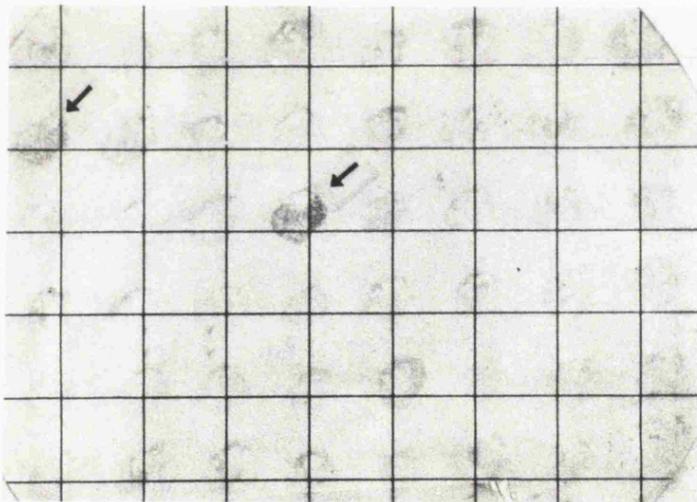
(c) Immuno-detection of the K9 antigen in polysaccharide extracted from *E. coli* LE392 and 2667 spotted onto a nylon filter. The volume of extract applied (μ l) is given above the spots.

Figure 6.1

a)



b)



positive control



c)

		K9	LB			
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30		K9	LE	X	

Chapter 7

Discussion

E. coli can express a large number of capsular polysaccharides which differ in the structure of the repeat oligosaccharide but have previously been grouped on the basis of shared biochemical characteristics. At the outset of this study there appeared to be two mutually exclusive genetic systems for the production of a polysaccharide capsule: the group I capsule genes near *his* and the group II capsule genes near *serA*. In this thesis it was demonstrated that there are three, the third being the group I/II capsule genes which are apparently located at the same chromosomal site (near *serA*) as the group II capsule genes. The possibility that additional genetic systems for capsule production exist in *E. coli* cannot be excluded. In the process of investigating the variability of capsule genes in the *E. coli* population, it has become apparent that the *serA* region of the chromosome is highly variable. A picture of this chromosomal region in different *E. coli* strains can be constructed using the group II capsule genes as a base (Figure 7.1). Only a limited number of strains expressing different groups of capsule have been studied but trends in the organisation of the chromosome around 64 map units, where group II *kps* is located, have been observed.

In group II encapsulated strains the group II capsule gene cluster, with its three genetic regions, is located near *serA* with region 1 orientated towards *serA* (Vimr, 1991). Genetic analysis of the group II capsule gene cluster has prompted the definition of three discrete functional regions, regions 1, 2 and 3, (Boulnois *et al.*, 1987; Roberts *et al.*, 1988) and it has been proposed that region 2 forms a gene 'cassette' which varies between the common regions 1 and 3 (Boulnois and Jann, 1989). However, nucleotide sequence analysis of different capsule gene clusters has revealed that the *kps* genes can be variable. It was shown in this study that the region 1 gene, *kpsS*, is variable at the 3' end. For example, the C terminus of KpsS_{K4} is fourteen amino acids longer than that of KpsS_{K5} (Chapter 5). Up to the 3' end, *kpsS* was highly homologous between different gene clusters such that the encoded proteins are largely identical and it was suggested that the variable C terminus of KpsS may be functionally insignificant. *kpsT* is also variable between different clusters (Pavelka *et al.*, 1991; Smith *et al.*, 1990). Not only is the C terminus of KpsT_{K5} five amino acids longer than that of KpsT_{K5} but also KpsT can be split into two regions on

the basis of DNA homology between *kpsT*_{K1} and *kpsT*_{K5}. The 5' two-thirds of the two genes are highly homologous to each other but the 3' one third has only 70% DNA homology. The characteristics of KpsT as an integral membrane protein are preserved but the significance of this variability on capsule expression is not clear. Despite these differences in Kps proteins encoded by different group II capsule gene clusters the proteins do appear to be functionally homologous. *kpsS* and *kpsT* are the two genes which are positioned immediately adjacent to region 2 and the variable 3' end of the genes are orientated towards region 2. At the level of nucleotide sequence the variable 3' ends of these two genes fit the previous criteria of region 2 sequences, not those associated with the common region 1 or 3 genes. This questions the idea that each of the regions in the group II capsule gene cluster is a discrete entity (see below). It is not known whether other *kps* genes are variable between clusters or whether the two genes adjacent to region 2 are exceptional. One possibility is that *kpsS* and *kpsT* are mosaic genes, the result of intragenic recombination or sequence divergence which may affect any gene.

The results of the analyses with *kpsS* and *kpsT* suggest that there are no distinct boundaries between the regions as was previously supposed. Due to the location of the *kpsS* and *kpsT* it is tempting to speculate that the variable 3' end of these genes arose as a consequence of the generation of various region 2 elements within the group II capsule gene cluster. However, the mechanisms by which the variability of region 2 was generated is still not clear. In the cassette theory, homologous recombination is postulated to one mechanism involved in the generation of diverse region 2 elements (Boulnois and Jann, 1989). If homologous recombination took place between clusters in either the 3' end of *kpsS* or *kpsT* and at some site within region 2 this would generate new combinations of genes but it would not generate novel sequence. This cannot therefore account for the novel 3' sequence of *kpsS* or *kpsT*. This implies that region 2 DNA sequences must, at least in part, have originated from outside the *kps* gene cluster. Sequence comparison of *kpsS* and *kpsT* from additional group II capsule gene clusters must be made to determine whether the 3' sequences of these genes are indeed unique.

It is possible that the common *kps* genes and the unique region 2 genes have different origins and have combined to produce the characteristic group II capsule gene cluster with its three regions. The G+C content of region 2 of the K5 and K1 capsule gene clusters is considerably lower than that of regions 1 and 3. The common region 1 and 3 comprise 50% G+C, typical of the *E. coli* genome, but region 2 of the K5 capsule genes is 33% G+C (Pazzani and Smith, unpublished data) and that of K1 is reported to be AT rich (Vimr, unpublished data). This difference in G+C content of the regions implies that regions 1 and 3, and region 2 do have alternative origins or that region 2 of the K1 and K5 capsule genes may have been inherited by *E. coli* more recently than the *kps* genes themselves. It suggests that each group II capsule gene cluster was not inherited as a complete ready-made block from elsewhere. What is more, the 3' one third of *kpsT*_{K5} is composed of 28% G+C (although this is a very short sequence for statistical analysis) and this suggests that the 3' end of *kpsT* may not share the same origin as the rest of region 3. It is interesting to note that the *Salmonella rfb* genes also have a low G+C content and it has been postulated that

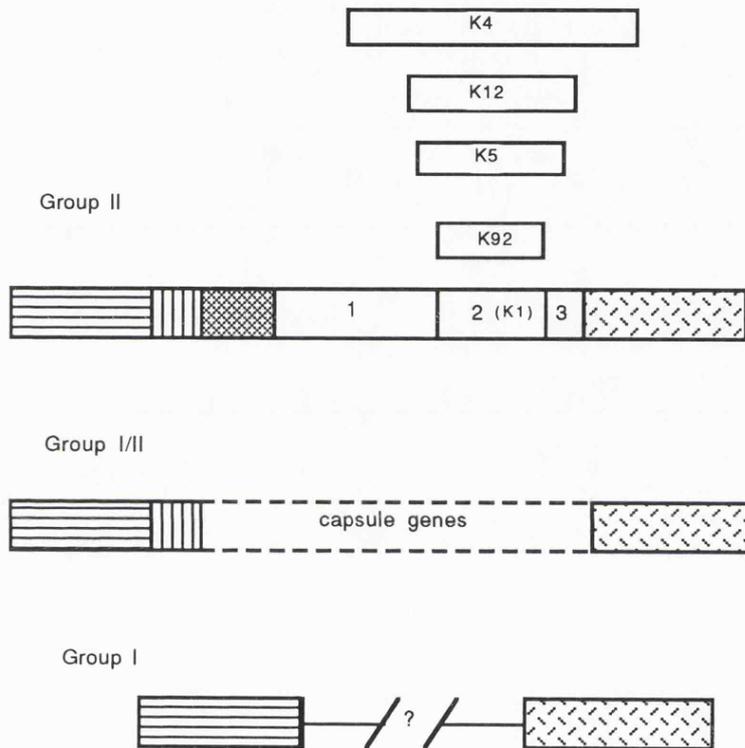
the *rfb* genes do not originate from the Enterobacteriaceae (Liu *et al.*, 1991; Verma and Reeves, 1989). Further speculation about the generation and organisation of the group II capsule gene cluster might be possible if the nucleotide sequence from a greater number of different clusters were available.

DNA adjacent to region 3 of the group II capsule gene cluster was shown to be a common component of the *E. coli* chromosome (Figure 7.1), present in most isolates in single copy and this is consistent with the findings of Vimr (1991). At the other end of the capsule gene cluster, adjacent to region 1, the situation is more complicated and the DNA can be divided into three sections (Figure 7.1). The sequence immediately adjacent to region 1 is only present in strains which carry the group II capsule genes although this DNA has not been implicated in group II capsule expression. Beyond this group II capsule gene-associated DNA about 3kb was found to be present only on the chromosome of strains expressing either a group II or a group I/II capsule, that is, in strains in which capsule genes are located near *serA*. Further still from region 1 of the group II capsule gene cluster the DNA was shown to be a common component of the *E. coli* chromosome. In conclusion, in group II capsule expressing strains the group II *kps* genes plus a certain amount of flanking DNA are surrounded by the common *E. coli* chromosome.

In group I/II capsule expressing strains the organisation of the *serA* region of the chromosome is similar to that in group II capsule expressing strains. It was shown by Southern blot analysis that the group II capsule genes are absent from the chromosome and are probably replaced by those of group I/II. The K10 antigen genes at least appear to be located at the same chromosomal position as the group II capsule gene clusters (Pearce, R., unpublished) and this may be the case for other group I/II capsule gene clusters. To account for the expression of chemically different group I/II polysaccharides the group I/II capsule genes, like the group II capsule genes, are presumably internally variable. Adjacent to one end of the group I/II capsule gene cluster are common components of the *E. coli* chromosome, that is the sequence found adjacent to region 3 of the group II *kps* genes which is present in most *E. coli* isolates (Figure 7.1). Adjacent to the other end of the group I/II capsule gene cluster is the DNA sequence which was shown to be present only in isolates which express a group II or a group I/II capsule. Presumably, the conserved *E. coli* chromosome lies beyond this capsule gene-associated DNA (Figure 7.1).

Finally, consider the *serA* region of the chromosome in group I capsule expressing strains and *E. coli* LE392. Like group I/II encapsulated strains, the chromosome of group I capsule expressing strains and *E. coli* LE392 do not carry the group II *kps* genes but there is no evidence to suggest that capsule genes of any type are present near *serA*. The common components of the *E. coli* chromosome which lie either side of the group II and I/II capsule genes are present on the chromosome, presumably at this site (Chapter 5). However, there is a discrepancy. In the K9 capsule expressing strain examined in this thesis, a single copy of the sequences which flank the group II capsule gene cluster were present on the chromosome and evidence suggested that the two sequences are not contiguous. Therefore, DNA of unknown coding capacity must lie in this position at 64 map units on the chromosome of this strain. In contrast,

Figure 7.1: The proposed organisation of the *serA* region of the *E. coli* chromosome in strains expressing different groups of capsule



Schematic representation of the *serA* region of the chromosome in group II (top), group I/II (middle) and group I (bottom) capsule expressing strains. The picture is based mainly on the data presented in Chapter 5. In the top picture the three boxes labelled 1 to 3 represent the three functional units of the group II *kps* genes. The region 2 in the diagram is that of K1 and other region 2 cassettes encoding different capsular polysaccharides as marked are pictured above. The patterned box to the left of region 1 refers to the group II capsule gene-associated DNA and the adjacent striped box refers to the DNA which is only found in group II and group I/II capsule expressing strains. The common components of the chromosome on either side of the capsule gene cluster are depicted as patterned boxes. In the group I/II capsule expressing strains capsule genes are present at this locus but their character is not clear and they are therefore depicted with dotted lines. The capsule gene associated DNA and the common components of the chromosome are present as above. In the group I capsule expressing strains the common *E. coli* chromosome is present and DNA of unknown length and unknown function is located at this site. Stripes (vertical or horizontal) are used to denote sequences which may be repeated on the chromosome of some strains.

Vimr (1991) found that DNA probes taken from either side of the K1 antigen gene cluster hybridised to the same DNA fragment cloned from *E. coli* W3110 but the probe taken from the DNA adjacent to region 3 also hybridised to a DNA fragment from the 62 minute region of the *E. coli* W3110 chromosome. This result suggests that the region 3-flanking sequence is repeated on the W3110 chromosome but the result could also be explained by a cloning artefact. The result could indicate that DNA flanking the group II capsule gene cluster in W3110 is contiguous in contrast to that in *E. coli* K9. A repeat of this sequence, on the chromosome of the group I capsule expressing strains and the K-12 strain tested in this thesis, should have been detected in the Southern blot analysis performed. The sequence flanking region 1 of the group II capsule gene cluster were found to be repeated on the chromosome of some group II and group I/II capsule expressing strains (see below). The model of the *serA* region of the chromosome in different strains depicted in Figure 7.1 is based on the analysis of only a small number of strains. Therefore, it is quite possible that the genetic organisation of the *serA* region of the chromosome in *E. coli* W3110 shares what may wrongly have been categorised as a characteristic of a group II or group I/II capsule expressing strain. PCR could be employed, using chromosomal DNA and primers taken from the DNA flanking the group II capsule gene cluster, to resolve these problems.

At this stage it is impossible to state how the various organisations of the *serA* region of the chromosome, apparent in different *E. coli* strains, may have arisen. Vimr (1991) suggested that the group II capsule gene cluster was inherited by a transpositional event to the 64 minute region of the *E. coli* chromosome. Insertion sequences may flank the group II and group I/II capsule gene clusters and may also include the capsule-gene associated DNA which has been identified in either case. Nucleotide sequence analysis of the DNA flanking the capsule gene clusters may therefore reveal whether transposition is likely to have been an important mechanism in the inheritance of the capsule genes near *serA*. Clearly, a single genetic rearrangement near *serA* cannot be responsible for the observed variability. First, a hypothesis that ancestral capsule genes inserted near *serA* in some strains does not explain the observation that the sequence at this putative insertion site may not be contiguous in others. Such a hypothesis could explain why the group II and group I/II capsule genes appear to be at the same chromosomal site. Since the group II and group I/II capsule genes are not homologous it is not easy to explain the observation that both are associated with the same DNA sequence which is not found in other strains (Figure 7.1). Sequence analysis of the group I/II capsule genes may reveal whether the group II and group I/II capsule genes share common evolutionary origins as those of *H. influenzae* and *N. meningitidis* appear to have (Frosch *et al.*, 1991).

There a further variable feature of the *serA* region of the *E. coli* chromosome. Sequences adjacent to region 1 of the group II capsule gene cluster (K4) were found to be polymorphic and were repeated on the chromosome of some strains but not others. This sequence includes both the sequence found only in group II and group I/II capsule expressing strains and the DNA further upstream from region 1 which was a common *E. coli* sequence. These sequences were present in multiple copy, possibly at various chromosomal locations not just near *serA* on the chromosome of the group

II and some of the group I/II capsule producing strains but in single copy if at all in group I encapsulated strains and in single copy in *E. coli* K-12. The function of this DNA is not known and although the restriction profile matches that expected for adhesin genes it is unlikely that functional adhesins are encoded by this DNA. It is unclear whether the abundance of this DNA is related to capsule type or whether the apparent trends between strains expressing different groups of capsules is a reflection of the low sample number.

Several differences between the *serA* region of the chromosome in different capsule expressing strains has become apparent and these differences extend beyond those associated with the expression of different capsules. Not only are the group II and group I/II capsule genes different but also the sequences around the capsule genes are variable in occurrence, some even being repeated on the chromosome of some strains but not others. Superimposed on this variability between groups is the internal variability of the group II capsule gene cluster itself. In an organism in which the chromosome is fairly stable this variability is somewhat surprising and the processes involved in its development are not clear.

References

Achtman, M. and Pluschke, G. (1986) Clonal analysis of descent and virulence among selected *Escherichia coli*. *Ann. Rev. Microbiol.* **40**:185-210.

Achtman, M., Mercer, A., Kusecek, B., Pohl, A., Heuzenroeder, M., Aaronson, W., Sutton, A. and Silver, R.P. (1983) Six widespread clones among *Escherichia coli* isolates. *Infect. Immun.* **39**:315-335.

Ames, G.F-L., Mimura, C.S. and Shymala, V. (1990) Bacterial periplasmic permeases belong to a family of transport proteins operating from *Escherichia coli* to humans: traffic ATPases. *FEMS Microbiol. Rev.* **75**:429-446.

Arakawa, Y., Ohta, M., Wacharotayankun, R., Morei, M., Kido, N., Komatsu, T., Sugiyama, T. and Kato, N. (1991) Biosynthesis of *Klebsiella* K2 capsular polysaccharide in *Escherichia coli* HB101 requires the functions of *rmpA* and the chromosomal *cps* gene cluster of the virulent strain *Klebsiella pneumoniae* Chedid (01:K2). *Infect. Immun.* **59**:2043-2050.

Arthur, M., Arbeit, R.D., Kim, C., Beltran, P., Crowe, H., Steinbach, S., Campanelli, G. and Wilson, R.A. (1990) Restriction fragment length polymorphisms among uropathogenic *Escherichia coli* isolates: *pap*-related sequences compared with *rrn* operons. *Infect. Immun.* **58**:471-479.

Avery, O.T. and Dubose, R. (1931) The protective action of a specific enzyme against type III pneumococcus infection in mice. *J. Experimental Medicine* **54**:73-89.

Bachmann, B. (1990) Linkage map of *Escherichia coli* K-12, edition 8. *Microbiol. Rev.* **54**:130-197.

Bayer, M.E. (1979) The fusion sites between the outer membrane and cytoplasmic membrane; their role in membrane assembly and lines infection. in Inouye, M. (ed) *Bacterial outer membranes: biogenesis and functions*. John Wiley and Sons, Inc., New York. pp167-202.

Bayer, M.E. (1990) Visualisation of the bacterial polysaccharide capsule. *Curr. Top. Microbiol. Immunol.* **150**:129-157.

Bayer, M.E. and Thurow, H. (1977) Polysaccharide capsules of *Escherichia coli*. microscopic study of its size, structure and sites of synthesis. *J. Bacteriol.* **130**:911-936.

Bennett, R.A. and Billing, E. (1978) Capsulation and virulence in *Erwinia amylovora*. *Annal. Appl. Biol.* **89**:41-45.

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

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

Boulnois, G.J., Roberts, I.S., Hodge, R., Hardy, K.R., Jann, K.B. and Timmis, K.N. (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, G.J. and Jann, K. (1990) Bacterial polysaccharide capsule synthesis, export and evolution of diversity. *Mol. Microbiol.* **3**:1819-1823.

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

Boyd, R.F. (1988) General microbiology. Times Mirror/Mosby College Publications, Missouri, USA.

Branefors-Hellander, P., Kenne, L., Lindberg, B., Petersson, K. and Unger, P. (1981) Structural studies of two capsular polysaccharides elaborated by different strains of *Haemophilus influenzae* type e. *Carbohydr. Res.* **88**:77-84.

Brenner, D.J. and Falkow, S. (1971) Molecular relationships among members of the Enterobacteriaceae. *Advan. Genet.* **16**:81-118.

Brenner, D.J., Fanning, G.R., Steigerwalt, A.G., Ørskov, I. and Ørskov, F. (1972) Polynucleotide relatedness among three groups of pathogenic *Escherichia coli* strains. *Infect. Immun.* **6**:308-315.

Catty, D. and Raykundalia, C. (1988) Production and quality control of polyclonal antibodies. In *Antibodies Volume I a practical approach*, Catty, D. (ed) p19-79. IRL Press Limited, Oxford.

Caugant, D.A., Levin, B.R., Ørskov, I., Ørskov, F., Svanborg-Eden, C. and Selander, R.K. (1985) Genetic diversity in relation to serotype in *Escherichia coli*. *Infect. Immun.* **49**:407-413.

Chang, A.C.Y. and Cohen, S.N. (1978) Construction and characterisation of ampli-

fiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J. Bacteriol.* **134**:1141-1156.

Chakraborty, A.K., Friebolin, H. and Stirm, S. (1980) Primary structure of the *Escherichia coli* serotype K30 capsular polysaccharide. *J. Bacteriol.* **141**:971-972.

Choy, Y.M., Fehmel, F., Frank, N. and Stirm, S. (1975) *Escherichia coli* capsule bacteriophages. VI. Primary structure of the bacteriophage 29 receptor, the *Escherichia coli* serotype 29 capsular polysaccharide. *J. Virol.* **16**:581-590.

Clark, I. and Carbon, J. (1976) A colony bank containing synthetic Col E1 hybrid plasmids representative of the entire *Escherichia coli* genome. *Cell* **9**:91-99.

Collins, J. and Hohn, B. (1978) Cosmids: a type of plasmid gene-cloning vector that is packageable *in vitro* in bacteriophage λ heads. *Proc. Natl. Acad. Sci. USA* **75**:4242-4246.

Costerton, J.W., Irvin, R.T. and Cheng, K-J. (1981) The bacterial glycocalyx in nature and disease. *Ann. Rev. Microbiol.* **35**:299-324.

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

Cross, A.S. (1990) The biological significance of bacterial encapsulation. *Curr. Top. Microbiol. Immunol.* **150**:87-96.

Delfert, D.M. and Conrad, H.E. (1985) Sulfation of chondroitin oligosaccharides *in vitro*. *J. Biol. Chem.* **260**:14446-14451.

Dengler, T., Himmelsbach, K., Jann, B. and Jann, K. (1988) Structure of the capsular K3 antigen of *Escherichia coli* O4:K3:H4 a polysaccharide containing a 4-deoxy-2-hexulosonic acid. *Carb. Res.* **178**:191-201.

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

Djordjevic, M.A., Gabriel, D.W. and Rolfe, B.G. (1987) *Rhizobium*-the refined parasite of legumes. *Ann. Rev. Phytopath.* **25**:145-168.

Dorman, C.J. (1991) DNA supercoiling and environmental regulation of gene expression in pathogenic bacteria. *Infect. Immun.* **59**:745-749.

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

- Dutton, G.G.S., Parolis, H. and Parolis, L.A.S. (1987) The structure of the neuraminic acid containing capsular polysaccharide of *Escherichia coli* serotype K9. *Carbohydr. Res.* **170**:193-206.
- DuBose, R.F., Dykhuizen, D.E. and Hartl, D.L. (1988) Genetic exchange among natural isolates of bacteria: recombination with the *phoA* gene of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **85**:7036-7040.
- Echarti, C.E., Hirschel, B., Boulnois, G.J., Varley, J.M., Waldvogel, F. and Timmis, K.N. (1983) Cloning and analysis of the K1 capsule biosynthesis genes of *Escherichia coli*: lack of homology with *Neisseria meningitidis* group B DNA sequences. *Infect. Immun.* **41**:54-60.
- Feinberg, A.P. and Vogelstein, B. (1983) A technique for radiolabelling restriction endonuclease fragments to a high specific activity. *Anal. Biochem.* **132**:6-13.
- Finke, A., Roberts, I.S., Boulnois, G.J., Pazzani, C. and Jann, K. (1989) Activity of CMP-KDO synthetase in *Escherichia coli* strains expressing the capsular K5 polysaccharide: implication for polysaccharide biosynthesis. *J. Bacteriol.* **171**:3074-3075.
- Finke, A., Jann, B. and Jann, K. (1990) CMP KDO synthetase activity in *Escherichia coli* expressing different capsular polysaccharides. *FEMS Microbiol. Lett.* **69**:129-134.
- Finke, A., Bronner, D., Nikolaev, A.V., Jann, B. and Jann, K. (1991) Biosynthesis of the *Escherichia coli* K5 polysaccharide, a representative of group II capsular polysaccharides: polymerisation *in vitro* and characterisation of the product. *J. Bacteriol.* **173**:4088-4094.
- Finne, J., Leinonen, M. and Mäkelä, P.H. (1983) Antigenic similarities between brain components and bacteria causing meningitis, implications for vaccine development. *Lancet* **2**:355-357.
- Fletcher, M. and Floodgate, G.D. (1973) An electron-microscopic demonstration of an acidic polysaccharide involved in the adhesion of a marine bacterium to solid surfaces. *J. Gen. Microbiol.* **74**:325-324.
- Frosch, M., Weisberger, C. and Meyer, T.F. (1989) Molecular characterisation and expression in *Escherichia coli* of the gene complex encoding the polysaccharide capsule of *Neisseria meningitidis* group B. *Proc. Natl. Acad. Sci. USA* **86**:1669-1673.
- Frosch, M., Edwards, U., Bousset, K., Krauß, B. and Weisberger, C. (1991) Evidence for a common molecular origin of the capsule gene loci in Gram-negative bacteria expressing group II capsular polysaccharides. *Mol. Microbiol.* **5**:1251-1263.
- Göransson, M., Forsman, K. and Uhlin, B.E. (1988) Functional and structural homology among regulatory cistrons of pili-adhesin determinants in *Escherichia coli*.

Mol Gen. Genet. **212**:412-417.

Gotschlich, E.C., Fraser, B.A., Nishmura, O., Robbins, J.B. and Liu, T-Y. (1981) Lipid on capsular polysaccharides of gram-negative bacteria. *J. Biol. Chem.* **256**:8915-8921.

Hacker, J. (1990) Determinants coding for fimbriae and adhesins of extra-intestinal *Escherichia coli*. *Curr. Top. Microbiol. Immunol.* **151**:1-27.

Hacker, J., Bender, L., Ott, M., Wingender, J., Lund, B., Marre, R. and Werner, G. (1990) Deletions of chromosomal regions coding for fimbriae and hemolysins occur *in vitro* and *in vivo* in various extra-intestinal *Escherichia coli* isolates. *Microbiol. Path.* **8**:213-225.

Hadfield, C. (1987) Genomic library cloning - with plasmid, λ and cosmid vectors. In *Gene cloning and analysis a laboratory guide*. p61-105, Boulnois, G.J. (ed). Blackwell Scientific Publications, Oxford.

Harshman, L. and Riley, M. (1980) Conservation and variation of nucleotide sequences in *Escherichia coli* strains isolated from nature. *Infect. Immun.* **50**:620-629.

Hedrick, P.W. (1982) Genetic hitchhiking: a new factor in evolution? *Bioscience* **32**:845-853.

Hermesse, M.P., Dereppe, C., Batholome, Y. and Rouxhet, G. (1988) Immobilisation of *Acetobacter aceti* by adhesion. *Can. J. Microbiol.* **34**:638-644.

Hancock, I.C. and Cox, C.M. (1991) Turnover of cell surface-bound capsular polysaccharide in *Staphylococcus aureus*. *FEMS Microbiol Lett.* **77**:25-30.

Higgins, C.F., Hinton, J.C.D., Hulton, C.S.J., Owen-Hughes, T., Pavitt, G.D. and Seirafi, A. (1990a) Protein H1: a role for chromatin structure in the regulation of bacterial gene expression and virulence? *Mol. Microbiol.* **4**:2007-2012.

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

High, N.J., Hales, B.A., Jann, K. and Boulnois G.J. (1988) A block of urovirulence genes encoding multiple fimbriae and hemolysin in *Escherichia coli* O4:K12:H⁻. *Infect. Immun.* **56**:513-517.

Hinson, G. and Williams, P.H. (1989) Adhesins of pathogenic *Escherichia coli*. *Genetics of bacterial diversity*, Hopwood, D.A. and Chater, K.E. (eds). Academic Press Ltd, London. pp287-307.

Hofmann, P., Jann, B. and Jann, K. (1985) Structure of the amino-acid containing

capsular polysaccharide K54 antigen from *Escherichia coli* O6:K54:H10. Carbohydr. Res. **139**:261-272.

Hofmann, P., Jann, B. and Jann, K. (1985b) Structure of the fructose-containing capsular polysaccharide of uropathogenic *Escherichia coli* O4:K52:H⁻. Eur. J. Biochem. **147**:601-609.

Hohn, B. (1979) *in vitro* packaging of λ and cosmid cloning. Meths. Enzymol. **68**:299-309.

Homonylo, M.K. and Whitfield, C. (1988) Cloning of the capsular polysaccharide (K antigen) biosynthetic genes from *Escherichia coli* K30. Abstr. Annu. Meet. Am. Soc. Microbiol. Miami Beach, Florida USA May 8-13 1988. **88**:83.

Homonylo, M.K., Wilmot, S.J., Lam, J.S., MacDonald, L.A. and Whitfield, C. (1988) Monoclonal antibodies against the capsular K antigen of *Escherichia coli* O9:K30(A):H12 characterisation and use in analysis of K antigen organisation on the cell surface, Can. J. Microbiol. **34**:1159-1165.

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

Hosieth, S.K., Connelly, C.J. and Moxon, E.R. (1985) Genetics of high-frequency loss of b capsule expression in *Haemophilus influenzae*. Infect. Immun. **49**:389-395.

Hosieth, S.K., Moxon, E.R. and Silver, R.P. (1986) Genes involved in *Haemophilus influenzae* type b capsule expression are part of an 18kb tandem duplication. Proc. Natl. Acad. Sci. USA **83**:1106-1110.

Hull, R., Bieler, S., Falkow, S. and Hull, S. (1986) Chromosomal map positions of genes encoding P-adhesins in uropathogenic *Escherichia coli*. Infect. Immun. **51**:693-695.

Hull, R., Bieler, S., and Hull, S. (1988) Restriction fragment length polymorphism and multiple copies of DNA sequences homologous with probes for P-fimbriae and hemolysin genes among uropathogenic *Escherichia coli*. Can. J. Microbiol. **34**:301-311.

Howley, P.M., Israel, M.F., Law, M-F. and Martin, M.A. (1979) A rariip method for detecting and mapping homology between heterologous DNA. J. Biol. Chem. **254**:4876-4883.

Hyde, S.C., Emsley, P., Hartshorn, M.J., Mimmack, M.M., Gileadi, U., Pearce, S.R., Gallagher, M.P., Gill, D.R., Hubbard, R.F. and Higgins, C.F. (1990) Structural model of ATP-binding proteins associated with cystic fibrosis, multidrug resistance and bacterial transport. Nature **346**:362-365.

- Jann, B. and Jann, K. (1977) Bacterial polysaccharide antigens. in Surface carbohydrates of the prokaryotic cell, Sutherland, I.W. (ed) Academic Press, New York. pp 247-287.
- Jann, B., and Jann, K. (1990) Structure and biosynthesis of the capsular antigens of *Escherichia coli*. Curr. Top. Microbiol. Immunol. **150**:19-42.
- Jann, B., Jann, K., Schmidt, G., Ørskov, F. and Ørskov, I. (1971) Comparative immunochemical studies of the surface antigens of *Escherichia coli* strains O8:K87(B):H19 and (O32):K87(B):H45. Eur. J. Biochem. **23**:515-522.
- Jann, B., Ahrens, R., Dengler, T. and Jann, K. (1988) Structure of the capsular polysaccharide K19 from uropathogenic *Escherichia coli* O25:K19:H12. Carb. Res **177**:273-277.
- Jann, K. and Jann, B. (1983) The K antigens of *Escherichia coli*. Prog. Allergy **33**:53-79.
- Jann, K. and Jann, B. (1984) Structure and biosynthesis of O-antigens. p138-186 in Reitschel, E.T. (ed), Handbook of endotoxins, vol. 1: chemistry of endotoxin. Elsevier Scientific Publishers B.V., Amsterdam.
- Jann, K. and Jann, B. (1985) Cell surface components and virulence: *Escherichia coli* O and K antigens in relation to virulence and pathogenicity. in The Virulence of *Escherichia coli* (M. Sussman, Ed), pp 156-176. Academic Press Inc., New York.
- Jann, K. and Jann, B. (1987) Polysaccharide antigens of *Escherichia coli*. Rev. Inf. Dis. (suppl. 5)**9**:S517-S526.
- Jann, K. and Hoshützky, H. (1990) Nature and organisation of the adhesins. Curr. Top. Microbiol. Immunol. **151**:55-70.
- Jann, K., Jann, B., Schmidt, M.A. and Vann, W.F. (1980) Structure of the *Escherichia coli* K2 capsular antigen, a teichoic acid like polymer. J. Bacteriol. **143**:1102-1115.
- Jann, K., Goldmann, G., Weisberger, C., Wolf-Ullisch, C. and Kanegasaki, S. (1982) Biosynthesis of the O9 antigen of *Escherichia coli*. Initial reaction and overall mechanism. Eur. J. Biochem. **127**:157-164.
- Jann, K., Pillat, M., Weisberger, C., Shibaev, V.N. and Torgov, V.I. (1985) Biosynthesis of the O9 antigen of *Escherichia coli*. Synthetic glycosyldiphosphomorphamides as probes for the requirement of mannose acceptors. Eur. J. Biochem. **151**:393-397.
- Jarvis, G.A. and Vedros, N.A. (1987) Sialic acid of Group B *Neisseria meningitidis* regulates alternative complement pathway activation. Inf. Immun. **55**:174-180.

- Jennings, H.J. (1990) Capsular polysaccharides as vaccine candidates. *Curr.Top. Microbiol. Immunol.* **150**:97-127.
- Johnson, J.R. (1991) Virulence factors in *Escherichia coli* urinary tract infection. *Clin. Microbiol. Rev.* **4**:80-128.
- Joiner, K.A. (1988) Complement and evasion by bacteria and parasites. *Ann. Rev. Microbiol.* **42**:201-230.
- Kaijser, B. (1977) A simple method for typing acidic polysaccharide K antigens of *Escherichia coli*. *FEMS Microbiol. Lett.* **1**:285-288.
- Kaijser, B. and Jodal, U. (1985) *Escherichia coli* K5 antigen in relation to various infections in healthy individuals. *J. Clin. Microbiol.* **19**:264-266.
- Kaijser, B., Hanson, L.A., Jodal, U., Lidin-Janson, G. and Robbins, J.B. (1977) Frequency of *Escherichia coli* K antigens in urinary tract infections in children. *Lancet* **i**, 663-664.
- Kaspar, D.L. Winkelhake, J.L., Zollinger, W.D., Brandt, B.L. and Artenstein, M.S. (1973) Immunochemical similarity between polysaccharide antigens of *Escherichia coli* O7:K1L:NM and group B *Neisseria meningitidis*. *J. Immunol.* **110**:262-268.
- Kauffman, F. (1966) The bacteriology of Enterobacteriaceae. E. Munksgaard, Copenhagen.
- Kenne, L. and Lindberg, B. (1983) Bacterial polysaccharides. in Aspinall, G.O. (ed) The polysaccharides. Academic Press, New York. pp287-360.
- Kido, N., Ohta, M., Iida, K., Hasegawa, T., Ito, H., Arakawa, Y., Komatsu, T. and Kato, N. (1989) Partial deletion of the cloned *rfb* genes of *Escherichia coli* O9 results in synthesis of new O-antigenic lipopolysaccharide. *J. Bacteriol.* **171**:3629-3633.
- Kohara, Y., Akiyama, K. and Isono, K. (1987) The physical map of the whole *Escherichia coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell* **50**:495-508.
- Krawiec, S. and Riley, M. (1990) Organisation of the bacterial chromosome. *Microbiol. Rev.* **54**:502-539.
- Krallmann-Wenzel, U., Ott, M., Hacher, J., and Schmidt, G. (1989) Chromosomal mapping of genes encoding mannose-sensitive (type 1) and mannose-resistant F8 (P) fimbriae of *Escherichia coli* O18:K5:H5. *FEMS. Microbiol. Lett.* **58**:315-322.
- Kroll, J.S., Hopkins, I. and Moxon, E.R. (1988) Capsule loss in *Haemophilus influenzae* type b occurs by recombination-mediated disruption of a gene essential for polysaccharide export. *Cell* **53**:347-356.

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

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

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

Kröncke, K.D., Boulnois, G.J., Roberts, I.S., Bitter-Suerman, D., Golecki, J., Jann, B., and Jann, K. (1990a) Expression of the *Escherichia coli* K5 capsular antigen: immunoelectron microscopic and biochemical studies with recombinant DNA. *J. Bacteriol.* **172**:1085-1091.

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

Kuo, J.S.-C., Doelling, V.W., Graveline, J.F. and McCoy, D.W. (1985) Evidence for covalent attachment of phospholipid to the capsular polysaccharide of *Haemophilus influenzae*. *J. Bacteriol.* **163** 769-773.

Laasko, D.M., Homonylo, M.K., Wilmot, S.J. and Whitfield, C. (1988) Transfer and expression of the genetic determinants for O and K antigen synthesis in *Escherichia coli* O9:K(A)30 and *Klebsiella* spp O1:K20, in *E. coli* K12. *Can. J. Microbiol.* **34**:987-992.

Liu, D., Verma, N.K., Romana, L.K. and Reeves, R.P. (1991) Relationships among the *rfb* regions of *Salmonella* serovars A, B and D. *J. Bacteriol.* **173**:4814-4819.

Long S.R. (1989) Rhizobium-legume nodulation: life together in the underground. *Cell* **56**:203-214.

Luderitz, O., Galanos, C., Lehmann, V., Nurminen, E., Reitschel, E., Rosenfelder, G., Simon, M. and Westphal, O. (1973) Lipid A chemical structure and biological activity. *J. Inf. Dis.* **128** (Suppl.):9-21.

Mäkelä, P.M. and Stocker, B.A.D. (1984) Genetics of lipopolysaccharide . in Handbook of endotoxin, vol. 1. Chemistry of endotoxin. Reitschel, E.T. (ed), Elsevier Science Publishers, Amsterdam. pp59-137.

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

Markovitz, A. (1977) Genetics and regulation of bacterial capsular polysaccharide

biosynthesis and radiation sensitivity. in Sutherland, I.W. (ed) Surface carbohydrates of the prokaryotic cell. Academic Press Inc. (London) Ltd. pp415-462.

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

Maynard Smith, J., Dowson, C.G. and Spratt, B.G. (1991) Localised sex in bacteria. Nature **349**:29-31.

McGuire, E.J. and Binckley, S.B. (1964) The structure and chemistry of colominic acid. Biochemistry **3**:247-251.

Médigue, C., Bouché, J.B., Hénaut, A. and Danchin, A. (1990) Mapping of sequences genes (700kbp) in the restriction map of the *Escherichia coli* genome. Mol. Microbiol. **4**:169-187.

Milkman, R. and Crawford, I.P. (1983) Clustered third base substitutions among three wild strains of *Escherichia coli*. Science **221**:378-380.

Milkman, R. and Bridges, M.M. (1990) Molecular evolution of the *Escherichia coli* chromosome. III Clonal fragments. Genetics **126**:505-517.

Moxon, R.E. and Kroll, J.S. (1984) Type b capsular polysaccharide as a virulence factor of *Haemophilus influenzae*. Vaccine **6**:113-115.

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

Muhlradt, P.F., Menzel, J., Golecki, J.R. and Speth, V. (1973) Outer membrane of Salmonella. Site of export of newly synthesised lipopolysaccharide on the surface. Eur. J. Biochem. **35**:471-481.

Ochman, H. and Selander, R.K. (1984) Evidence for a clonal population structure in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **81**:198-201.

Ochman, H., Whittam, T.S., Caugant, D.A. and Selander, R.K. (1983) Enzyme polymorphism and genetic population structure in *Escherichia coli* and *Shigella*. J. Gen. Microbiol. **129**:2715-2726.

Ochman, H., Wilson, R.A., Whittam, T.S., and Selander, R.K. (1983) Genetic diversity within serotypes of *Escherichia coli*. Proceedings of the Fourth International Symposium on Neonatal Diarrhea. 202-224.

Opal, S., Cross, A. and Gemski, P. (1982) K antigen and serum sensitivity of rough *Escherichia coli*. Inf. Immun. **37**:956-960.

Ørskov, F. and Ørskov, I. (1962) Behaviour of *Escherichia coli* antigens in sexual recombination. *Acta Pathol. Microbiol. Scand.* —bf 55:99-109.

Ørskov, F. and Ørskov, I. (1983) Summary of a workshop on the clone concept in the epidemiology, taxonomy, and evolution of the Enterobacteriaceae and other bacteria. *J. Inf. Dis.* **148**:346-357.

Ørskov, F. and Ørskov, I. (1990) The serology of capsular antigens. *Curr. Top. Microbiol. Immunol.* **150**:44-63.

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

Ørskov, I., and Nyman, K. (1974) Genetic mapping of the antigenic determinants of two polysaccharide K antigens, K10 and K54, in *Escherichia coli*. *J. Bacteriol.* **120**:43-51.

Ørskov, I. and Ørskov, F. (1984) Serotyping of *Klebsiella*. *Methods Microbiol.* **14**:143-164.

Ørskov, I. and Ørskov, F. (1985) *Escherichia coli* in extra-intestinal infections. *J. Hyg.* **95**:551-575.

Ørskov, I. and Ørskov, F. (1990) Serological classification of the adhesins. *Curr. Top. Microbiol. Immunol.* **151**:71-90.

Ørskov, I., Ørskov, F., Birch-Andersen, A., Knamori, M. and Eden, S. (1982) O, K, H and fimbrial antigens in *Escherichia coli* serotypes associated with pyelonephritis. *Scand. J. Inf. Dis. (Suppl.)* **33**:18-25.

Ørskov, I., Ørskov, F., Jann, B. and Jann, K. (1977) Serology, chemistry and genetics of O and K antigens of *Escherichia coli*. *Bacteriol. Rev.* **41**:667-710.

Ørskov, I., Sharma, V. and Ørskov, F. (1976) Genetic mapping of the K1 and K4 antigens (L) of *Escherichia coli*. *Acta Pathol. Microbiol. Scand. Sect. B.* **84**:125-131.

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

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

Petrie, G.F. (1932) A specific precipitin reaction associated with the growth on agar plates of meningococcus, pneumococcus and *B-dysenteriae* (Shiga). *Br. J. Exp.*

Pathol. **13**:993-996.

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

Plos, K., Hull, I.S., Hull, R.A., Levin, B.R., Ørskov, I., Ørskov, F. and Svanborg-Eden, C. (1989) Distribution of the P-associated-Pilus (*pap*) region among *Escherichia coli* from natural sources: evidence for horizontal gene transfer. Infect. Immun. **57**:1604-1611.

Pluschke, G., Mayden, J., Achtman, M., and Levine, R.P. (1983) Role of the capsule and the O antigen in the resistance of O14:K1 *Escherichia coli* to complement-mediated killing. Inf. Immun. **42**:903-913.

Raetz, C.R.H. (1988) Biosynthesis and pharmacological properties of *Escherichia coli* lipid A. in Inouye, M. (ed) Bacterial outer membranes as model systems. pp229-245. John Wiley & Sons, Inc., New York.

Rick, P.D. (1987) Lipopolysaccharide biosynthesis in Neidhardt, F.C., Ingraham, J.L., Low, K.B., Magasanik, B, Schaechter, M. and Umbarger, H.E. (eds) *Escherichia coli* and *Salmonella typhimurium* cellular and molecular biology. vol 1 pp648-662. American Society for Microbiology, Washington D.C.

Riley, M., and Krawiec, S. (1987) Genome organisation. in Neidhardt, F.C. Ingraham, J.L., Low, K.B., Magasanik, B, Schaechter, M. and Umbarger, H.E. (eds), *Escherichia coli* and *Salmonella typhimurium* cellular and molecular biology. pp 967-981. American Society for Microbiology, Washington D.C.

Robbins J.B., McCracken, G.H., Gotschlich, E.C., Ørskov, I. and Hanson, L.A. (1974) *Escherichia coli* K1 capsular polysaccharide associated with neonatal meningitis. New England J. Med. **290**:267-271.

Robbins, P.W. and Wright, A. (1971) Biosynthesis of O antigens. in Microbial toxins, vol. 4. Weinbaum, G., Kadis, S. and Ajl, S.J. (eds) New York Academic Press pp351-368.

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

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

Roberts, M., Roberts, I.S., Korhonen, T.K., Jann, K., Bitter-Suermann, D., Boulnois, G.J., and Williams, P.H. (1988) DNA probes for K antigen (Capsule) typing of

Escherichia coli. J. Clin. Microbiol. **26**:385-387.

Rodriguez, M-L., Jann, B. and Jann, K. (1988) Structure and serological characteristics of the capsular K4 antigen of *Escherichia coli* O5:K4:H4, a fructose-containing polysaccharide with a chondroitin backbone. Eur. J. Biochem. **177**:117-124.

Rodriguez, M-L., Jann, B. and Jann, K. (1989) Structure and serological properties of the capsular K11 antigen of *Escherichia coli* O13:K11:H11, a fructose containing polysaccharide with a diglucosyl phosphate backbone. Carb. Res.

Rohr, T.E. and Troy, F.A. (1980) Structure and biosynthesis of surface polymers containing polysialic acid in *Escherichia coli*. J. Biol. Chem. **255**:2332-2342.

Saiki, R.K., Gyllensten, U.B. and Erlich, H. (1988) The polymerase chain reaction. in Genome analysis, Davies, K.E. (ed). IRL Press, Oxford.

Saito, H., and Miura, K.I. (1963) Preparation of transforming deoxyribonucleic acid by phenol treatment. Biochim. Biophys. Acta. **72**:619-629.

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

Schmidt, G. (1973) Genetical studies on the lipopolysaccharide of *Escherichia coli* K-12. J. Gen. Microbiol. **77**:151-160.

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

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

Selander, R.K. and Levin, B.R. (1980) Genetic diversity and structure in *Escherichia coli* populations. Science **210**:545-547.

Selander, R.K. and Musser, J.M. (1990) Population genetics of bacterial pathogenesis. p11-36, in Molecular basis of bacterial pathogenesis. Iglewski, B.H. and Clark, V.L. (eds) Academic Press Inc., London.

Selander, R.K., Korhoenen, T.K., Vaisanen-Rhen, V., Williams, P.H., Pattison, P.E. and Caugant, D.A. (1986) Genetic relationships and clonal structure of strains of *Escherichia coli* causing neonatal septicemia and meningitis. Infect. Immun. **52**:213-222.

Selander, R.K., Caugant, D.A. and Whittam, T.S. (1987) Genetic structure and variation in natural populations of *Escherichia coli* in Neidhardt, F.C., Ingraham,

J.L., Low, K.B., Magasanik, B, Schaechter, M. and Umbarger, H.E. (eds) *Escherichia coli* and *Salmonella typhimurium* cellular and molecular biology. vol. 2 pp1625-1648. American Society for Microbiology, Washington D.C.

Silver, R.P. and Aaronson, W. (1988) The K1 capsular polysaccharide of *Escherichia coli*. Rev. Infect. Dis. vol. 10 suppl. 2:S282-S286.

Silver, R.P. and Vimr, E.R. (1990) Polysialic acid capsule of *Escherichia coli* K1. p39-60, in Molecular basis of bacterial pathogenesis. Iglewski, B.H. and Clark, V.L. (eds) Academic Press Inc., London.

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

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

Silver, R., Aaronson, W. and Vann, W.F. (1987) Translocation of capsular polysaccharides in pathogenic strains of E requires a 60kDa periplasmic protein. J. Bacteriol. **169**:5489-5495.

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

Smyth, C.J. (1986) Fimbrial variation in *Escherichia coli*. in Birkbeck, T.H. and Penn, C.W. (eds), Antigenic variation in infectious diseases. pp95-125. IRL Press, Oxford.

Sorensen, U.B.S., Henrichsen, J. and Chen, H-C. (1990) Covalent linkage between the capsular polysaccharide and the cell wall peptidoglycan of *Streptococcus pneumoniae* revealed by immunochemical methods. Microbial Path. **8**:325-334.

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **89**:503-517.

Steenbergen, S.M. and Vimr, E.R. (1990) Mechanisms of polysialic acid chain elongation in *Escherichia coli* K1. Mol. Microbiol. **4**:603-611.

Stevens, P, Young, L.S. and Adamu, S. (1983) Opsonisation of various capsular (K) *Escherichia coli* by the alternative complement pathway. Immunol. **50**:497-502.

Stoltzfus, A., Leslie, J.F. and Milkman, R. (1988) Molecular evolution of the *Escherichia coli* chromosome. 1. Analysis of structure and natural variation in a previously uncharacterised region between *trp* and *tonB*. Genetics **120**:345-358.

Sugiyama, T., Kido, N., Komatsu, T., Phta, M. and Kato, N. (1991) Expression of the cloned *Escherichia coli* O9 *rfb* gene in various mutant strains of *Salmonella typhimurium*. J. Bacteriol. **173**:55-58.

Sutherland, I.W. (1972) Microbial exopolysaccharides. Advan. Microbial Phys. **8**:143-213.

Sutherland, I.W. (1977) Bacterial exopolysaccharides - their nature and production. in Sutherland, I.W. (ed) Surface carbohydrates of the prokaryotic cell. Academic Press Inc. (London) Ltd. pp. 27-96.

Sutherland, I.W. (1985) Biosynthesis and composition of Gram negative bacterial extracellular and wall polysaccharides. Ann. Rev. Microbiol. **39**:243-270.

Taylor, P.W. (1983) Bactericidal and bacteriolytic activity of serum against gram-negative bacteria. Microbiol. Rev. **47**:46-83.

Tennent, J.M., Hultgren, S., Marklund, B-I., Forsman, K., Göransson, M., Uhlin, B.E. and Normark, S. (1990) Genetics of adhesin expression in *Escherichia coli*. in Iglewski, B.H. and Clark U.L. (eds) Molecular basis of bacterial Pathogenicity. Academic Press Inc, London. p79-110.

Thomson, G. (1977) The effect of a selected locus on linked neutral loci. Genetics **85**:753-788.

Troy, F.A. (1979) The chemistry and biosynthesis of selected bacterial capsular polymers. Ann. Rev. Microbiol. **33**:19-60.

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

Troy, F.A., Frerman, F.E. and Heath, E.C. (1972) Synthesis of capsular polysaccharide of bacteria. Methods Enzymol. **28**:602-624.

Turk, D.C. and May, J.R. (1967) *Haemophilus influenzae*: its clinical importance. English University Press, London.

van Die, I., van Megen, I., Zuidweg, E., Hoekstra, W., de Ree, H., van den Bosch, H. and Bergmans, H. (1986) Functional relationship among the gene clusters encoding F7₁, F7₂, F9 and F11 fimbriae of human uropathogenic *Escherichia coli*. J. Bacteriol. **167**:407-410.

Vann, W.F., Schmidt, M.A., Jann, B. and Jann, K. (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, W.F., Silver, R.P., Abeijon, C., Chang, K., Aaronson, W., Sutton, A., Finn,

C.W., Linder, W. and Katsatos, M. (1987) Purification, properties and genetic determination of *Escherichia coli* 5'-monophosphate N-acetylneuraminic acid synthetase. *J. Biol. Chem.* **262**:17556-17562.

Verma, N.K. and Reeves, P.R. (1989) Identification and sequence of *rfbS* and *rfbE*, which determine antigenic specificity of group A and group D salmonellae. *J. Bacteriol.* **171**:5694-5701.

Vijay, I.K. and Troy, F.A. (1975) Properties of membrane-associated sialyltransferase of *Escherichia coli*. *J. Biol. Chem.* **250**:164-170.

Vimr, E. Map position and genomic organisation of the *kps* cluster for polysialic acid synthesis in *Escherichia coli* K1. *J. Bacteriol.* **173**:1335-1338.

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

Whitfield, C. (1988) Bacterial extracellular polysaccharides. *Can. J. Microbiol.* **34**:415-420.

Whitfield, C. and Lam, M. (1986) Characterisation of coliphage K30, a bacteriophage specific for *Escherichia coli* capsular serotype K30. *FEMS Microbiol. Lett.* **37**:351-355.

Whitfield, C., Schoenhals, G. and Graham, L. (1989) Mutants of *Escherichia coli* O9:K30 with altered synthesis and expression of the capsular K30 antigen. *J. Gen. Microbiol.* **135**:2589-2599.

Whittam, T.S., Ochman, H. and Selander, R.K. (1983) Multilocus enzyme electrophoresis in natural populations of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **80**:1751-1755.

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

Yeung, M.K. and Mattingly, S.J. (1986) Covalent linkage of the type- and group-specific antigens to the peptide moiety of the peptidoglycan of serotype III group B *Streptococcus*. *Curr. Microbiol.* **14**:205-211.

Yoshida, Y., Okamura, N., Kato, J. and Watanabe, H. (1991) Molecular characterisation of form I antigen genes of *Shigella sonnei*. *J. Gen. Microbiol.* **137**:867-874.