Vol. 181, No. 7

Genetic Organization of the *Escherichia coli* K10 Capsule Gene Cluster: Identification and Characterization of Two Conserved Regions in Group III Capsule Gene Clusters Encoding Polysaccharide Transport Functions

BRADLEY R. CLARKE, ROWAN PEARCE, AND IAN S. ROBERTS*

School of Biological Sciences, University of Manchester, Manchester M13 9PT, United Kingdom

Analysis of the *Escherichia coli* K10 capsule gene cluster identified two regions, regions 1 and 3, conserved between different group III capsule gene clusters. Region 1 encodes homologues of KpsD, KpsM, KpsT, and KpsE proteins, and region 3 encodes homologues of the KpsC and KpsS proteins. An *rfaH* mutation abolished K10 capsule production, suggesting that expression of the K10 capsule was regulated by RfaH in a manner analogous to group II capsule gene clusters. An IS3 element and a φR73-like prophage, both of which may have played a role in the acquisition of group III capsule gene clusters, were detected flanking the K10 capsule genes.

Escherichia coli synthesizes at least 80 distinct capsular polysaccharides on the cell surface (22). These capsules have been classified into three groups based on biochemical and genetic criteria (17, 29). Group I capsules are heat-stable, high-molecular-weight polysaccharides with a low charge density. The reducing ends of some group I capsules are associated with lipid A-core of lipopolysaccharide, although this association is not essential for surface expression (20, 47). The genetic determinants for group I capsules map near *his* on the *E. coli* chromosome (25). Group II capsules are heat labile, have a high charge density, and have a lower molecular weight than those of group I (17). Phosphatidic acid is bound to the reducing ends of group II capsules and may anchor these polysaccharides to the cell surface (17, 38). The genetic locus for group II capsules has been mapped near *serA* at 64 min on the *E. coli* chromosome (24, 26, 45). Unlike group I capsules, those of group II are temperature regulated, with expression occurring only above 20°C (17) and with high levels of CMP-3-keto-3-deoxy-*manno*-octulosonate (CMP-KDO) synthetase activity at capsule-permissive temperatures (10, 11). This coincides with the detection of KDO at the reducing end of certain group II polysaccharides (18). Group III capsules (formerly group I/II) are also located near *serA* and have the same general characteristics as those of group II. However, in contrast to group II, group III strains express polysaccharide at all growth temperatures and do not have elevated levels of CMP-KDO synthetase (10, 23, 29).

Molecular cloning and analysis of gene clusters involved in biosynthesis of group II capsules has revealed a common segmental organization comprising three distinct regions (35, 36). Regions 1 and 3 are conserved among all group II strains analyzed. These regions contain the genes *kpsFEDUCS* and *kpsMT*, respectively, and encode functions involved in maturation and export of the polysaccharide (2, 27, 28, 41). Region 2, which is positioned between regions 1 and 3, is serotype

specific and encodes proteins involved in synthesis of the polymer (31).

The capsule gene clusters of group III strains, K10 and K54, have been cloned (29, 37). DNA probes derived from the cloned K10 genes hybridize to chromosomal DNA from group III strains, K3, K54, and K98, but not to DNA from group I or group II strains (29). In addition, within group III strains, the hybridization patterns indicate a segmental gene organization analogous to that found in group II capsule clusters in which a serotype-specific region is flanked by conserved genes (29). Although there appears to be no overall nucleotide homology between the group III and group II capsule clusters, complementation of mutations in genes encoding proteins for the export of group II capsules by cloned K10 and K54 genes reveals that there are functionally conserved steps in the export of group II and group III capsular polysaccharides (29). Partial DNA sequencing of the K54 capsule gene cluster revealed that homologues of some of the group II genes are included but that these genes are arranged differently to their counterparts in group II clusters (37).

We now report the nucleotide sequence of two regions (termed 1 and 3) of the *E. coli* K10 capsular gene cluster which are conserved amongst group III capsule gene clusters and which encode functions for the export of group III polysaccharides. Region 1 contains four genes encoding homologues of the group II region 1 and 3 proteins KpsD, KpsE, KpsM, and KpsT. Region 3 is composed of two genes which encode homologues of the group II region 1 proteins KpsC and KpsS. We demonstrate that the K10 gene cluster is regulated by RfaH and provide evidence suggesting that the group III capsule determinants may have been positioned under the control of the RfaH-regulated promoter through an IS110-mediated insertion into a group II capsule cluster. Finally, we identify the presence of IS3 and φR73-like prophage genes flanking the K10 capsule gene cluster, which may have played a role in the acquisition of capsule gene clusters at this chromosomal location.

Nucleotide sequence analysis of regions 1 and 3. The cosmid clone pRP1 (Fig. 1A) contains the entire K10 capsular gene cluster in addition to flanking DNA not involved in capsule synthesis (29). DNA sequencing was performed on subclones of pRP1 which harbor DNA flanking the serotype-specific re-

* Corresponding author. Mailing address: School of Biological Sciences, 1.800 Stopford Building, University of Manchester, Oxford Rd., Manchester M13 9PT, United Kingdom. Phone: 0044-(0)161-275-5601. Fax: 0044-(0)161-275-5656. E-mail: ISROBERT@fs1.scg.man.ac.uk.
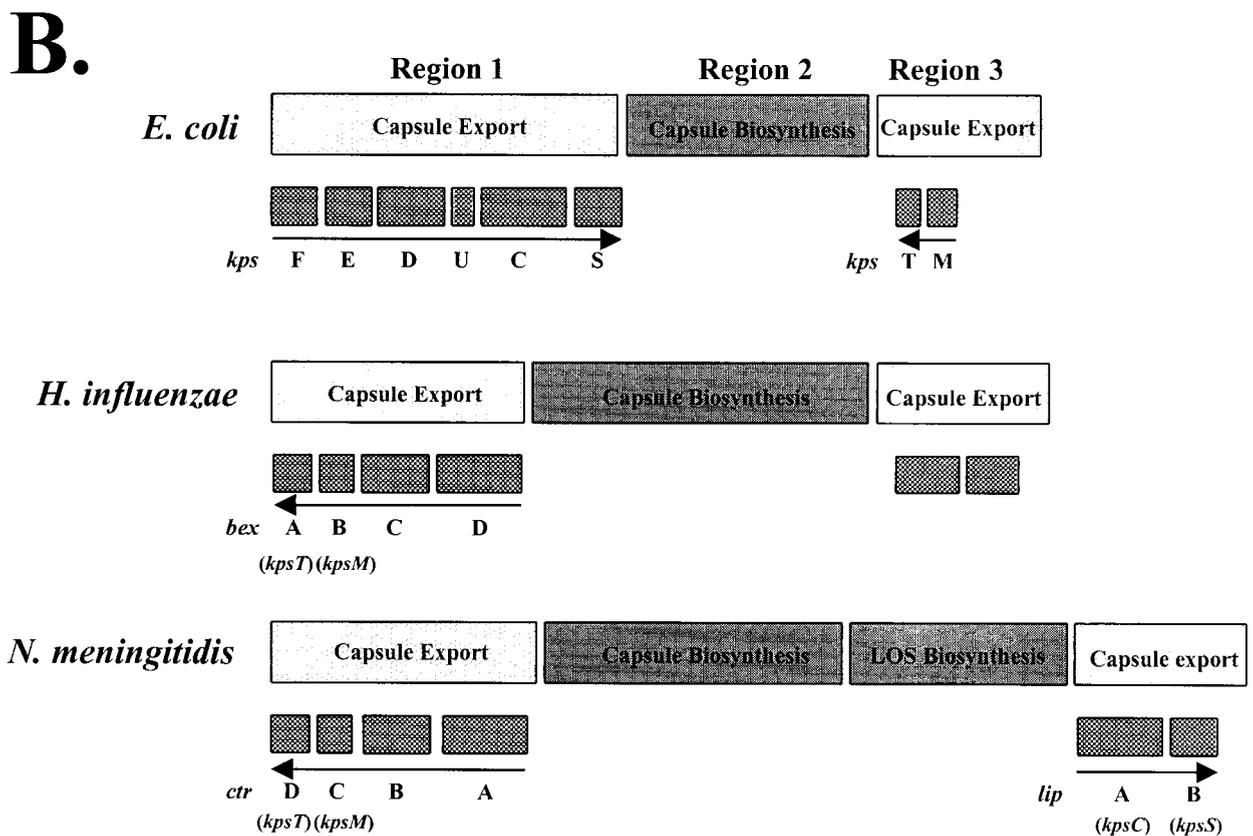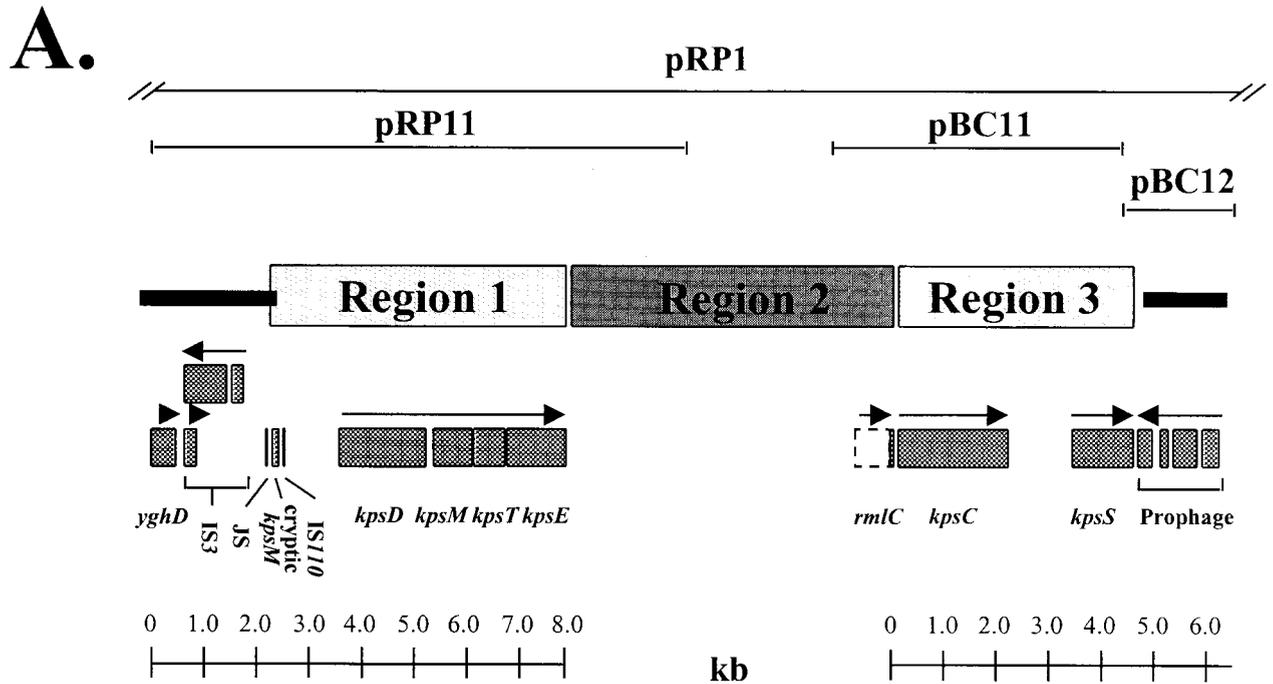
FIG. 1. (A) Scale diagram of the *E. coli* K10 capsule genes and flanking sequences. Recombinant plasmids used in this study are depicted as thin lines at the top. Genes within the K10 capsule cluster lie below regions 1, 2, and 3, whereas genes flanking the cluster lie below the thick lines. Arrows show direction of transcription. A scale is shown at the bottom. JS, JUMPstart sequence. (B) Diagram (not to scale) comparing the general organization of the capsule gene clusters of *E. coli* K5, *H. influenzae*, and *N. meningitidis*. *E. coli kps* homologues to *H. influenzae* and *N. meningitidis* are shown in parentheses under the corresponding genes.

TABLE 1. Degree of amino acid homology between the putative proteins of the *E. coli* K10 capsule cluster and their homologues
in *E. coli* K5, *H. influenzae*, and *N. meningitidis*

E. coli K10 protein	Function[a]	Homologue and homology (% identity)		
		E. coli K5	*H. influenzae*	*N. meningitidis*
KpsD	Periplasmic protein	KpsD, 34.6		
KpsM	ABC-transporter (membrane component)	KpsM, 39.7	BexB, 28.2	CtrC, 27.5
KpsT	ABC transporter (ATP-binding component)	KpsT, 45.7	BexA, 46.5	CtrD, 45.8
KpsE	Polysaccharide export protein	KpsE, 30.0		
KpsC	Polysaccharide modification protein	KpsC, 34.3		LipA, 39.9[b]
KpsS	Polysaccharide modification protein	KpsS, 87.2		LipB, 38.5

[a] ABC, ATP-binding cassette.
[b] LipA is homologous to amino acids 197 to 593 of $KpsC_{K10}$.

gion identified previously (29). Sequencing of the 5′ end of the cluster (pRP11) identified four open reading frames (ORFs) with homology to the *E. coli* K5 genes *kpsD*, *kpsM*, *kpsT*, and *kpsE* (Fig. 1A). These ORFs are arranged on the K10 chromosome in the same order as their homologues recently identified in the *E. coli* group III serotype K54 (37). This organization is clearly different from that in group II strains, where *kpsD* and *kpsE* are present in region 1 within the operon *kpsFEDUCS* and where *kpsM* and *kpsT* together constitute region 3 (Fig. 1B) (34). The organization of the group III capsule cluster resembles those of *Haemophilus influenzae* and *Neisseria meningitidis*, in which the *kpsT*, *kpsM*, and *kpsE* homologues (*bexA/ctrD*, *bexB/ctrC*, and *bexC/ctrB*) (13–15, 19) are arranged together in a single transcriptional unit (Fig. 1B). However, in these cases the order of genes is different and there is no *kpsD* homologue (Fig. 1B). The levels of amino acid homology between the encoded K10 polypeptides and their homologues in *E. coli* K5, *H. influenzae*, and *N. meningitidis* are presented in Table 1. $KpsD_{K10}$, $KpsM_{K10}$, $KpsT_{K10}$, and $KpsE_{K10}$ are very similar to their K54 counterparts, with only 3 amino acid substitutions each in KpsD and KpsT, 1 substitution in KpsM, and 18 substitutions in KpsE (data not shown). The DNA sequence of the K10 and K54 capsule gene clusters is 99% identical over the *kpsDMTE* coding region, which suggests that region 1 may have been acquired as a block into *E. coli* K10 and K54 from a common ancestor. The average G+C content of the *kpsDMTE* coding sequences is 39%, which is significantly lower than the average of 51% for the *E. coli* K-12 genome (1) and supports the probability of lateral gene transfer. It is notable that the G+C content of region 1 of the group III capsule genes more closely resembles that of the *H. influenzae* genome (37%) than that of regions 1 and 3 of the group II capsule gene clusters (46%) (28). This suggests that region 1 of the group III capsule gene clusters may have been acquired by a different route from that of the group II capsule export genes.

Although there is a high degree of nucleotide homology between $kpsE_{K10}$ and $kpsE_{K54}$, this homology ceases immediately 3′ of the coding sequence. Analysis of this 3′ sequence failed to identify any ORFs with homology to those in the available databases. Likewise, DNA sequence data within 1.3 kb 3′ of $kpsE_{K54}$ also failed to identify any ORFs that encode capsule export or synthesis functions (37). It is likely that this long intergenic region defines the junction between the conserved group III capsule region 1 genes and the variable region 2, which is involved in polymer synthesis. Interestingly, this DNA also has a low G+C (25%) content, typical of genes, present within capsule gene clusters, that encode enzymes for polysaccharide biosynthesis (33).

Nucleotide sequence analysis of subclones, pBC11 and pBC12, containing DNA 3′ to the serotype-specific region (Fig.

1A) identified two ORFs encoding polypeptides with homology to region 1 proteins of *E. coli* K5. An ORF of 2,147 bp encodes a polypeptide with limited homology to $KpsC_{K5}$ (Fig. 1A; Table 1). Plasmid pBC11 can restore K5 capsule expression when supplied in *trans* in a $kpsC_{K5}$ mutant (results not shown). This functional and amino acid homology suggests that this ORF encodes the K10 equivalent of $KpsC_{K5}$. A second ORF of 1,419 bp encodes a polypeptide with 87.2% identity to $KpsS_{K5}$ over 389 amino acids (Fig. 1; Table 1). However, compared with $KpsS_{K5}$, $KpsS_{K10}$ contains an additional 17 amino acids at its C terminus. The $kpsC_{K10}$ and $kpsS_{K10}$ genes contain 43.8 and 41.3% G+C, respectively. The nucleotide sequence of the region immediately 3′ to $kpsS_{K10}$ (1,707 bp) contains an average of 51% G+C, typical for the *E. coli* K-12 genome (1). The low G+C content of $kpsC_{K10}$ and $kpsS_{K10}$ indicates that they may have been acquired by lateral gene transfer from another organism with low G+C content. However, the G+C content of $kpsC_{K10}$ and $kpsS_{K10}$ is higher than that of the K10 region 1 genes and indicates that region 1 and region 3 may have been acquired separately. The $kpsC_{K10}$ and $kpsS_{K10}$ genes are separated by an intergenic region of 1,170 bp. In contrast, the distance between *kpsC* and *kpsS* is 134 bp in *E. coli* K5 (28). The relatively large distance between $kpsC_{K10}$ and $kpsS_{K10}$ might reflect the acquisition of each of these genes separately. Alternatively, this *kpsC-kpsS* intergenic region may contain elements which function in maintenance or regulation of the K10 capsule in vivo. It is notable that this intergenic sequence has an extremely low G+C content of 28.1%, the significance of which is unclear. The arrangement of *kpsC* and *kpsS* genes in a separate region in the group III capsule cluster is the same as that for their homologues, *lipA* and *lipB*, respectively, in *N. meningitidis* (Fig. 1B).

The DNA sequence analysis performed in this study failed to identify an ORF corresponding to *kpsU* in either region 1 or 3 of the K10 capsule gene cluster. The absence of *kpsU* is consistent with the lack of elevated levels of CMP-KDO synthetase found in group III-expressing strains (18). The identification of homologues to the KpsC and KpsS proteins is in keeping with the observation that phosphatidyl-KDO has been detected at the reducing end of the K10 polysaccharide (39) and would be in agreement with the suggested role of the KpsC and KpsS proteins in the addition of phosphatidyl-KDO to the reducing end of polysaccharide (34). The lack of a capsule-specific CMP-KDO synthetase enzyme would suggest that the CMP-KDO for group III capsule expression is supplied by the KdsB enzyme (11, 28). The KpsC protein of *E. coli* K5 contains two repeated domains, and it has been suggested that these arose from a gene duplication resulting in the fusion of two homologous proteins (32). It is speculated that these domains either represent two active sites acting on similar substrates or represent a single active site involving two domains. Analysis of

the amino acid sequence of KpsC$_{K10}$ reveals the presence of these repeated domains, where the N-terminal domain (amino acids 65 to 365) exhibits 30% homology to the C-terminal domain (amino acids 428 to 682). Therefore, although the overall amino acid homology of 35% suggests considerable divergence between KpsC$_{K10}$ and KpsC$_{K5}$, the retention of these repeated domains indicates a common KpsC progenitor and perhaps an important functional requirement for these domains. In addition to the absence of KpsU, the K10 capsule cluster does not appear to encode a homologue of the group II region 1 protein, KpsF. This protein is not essential for group II capsule biosynthesis (6, 40), and the significance of its absence from the K10 cluster is not apparent.

A partial ORF encoding 37 amino acids with 38 to 48% identity to the C terminus of RmlC homologues from numerous bacteria (data not shown) was identified 50 bp from the start of kpsC$_{k10}$ (Fig. 1A). The rmlC gene encodes dTDP-4-keto-6-deoxy-D-glucose-3,5-epimerase, which is required for synthesis of an intermediate in the dTDP-rhamnose pathway (47). The E. coli K10 capsule contains rhamnose as one of its component sugars (39), and so this RmlC homologue is probably involved in the synthesis of dTDP-rhamnose required for biosynthesis of the K10 polymer. Therefore, rmlC$_{K10}$ defines the boundary between the serotype-specific region 2 and a conserved region 3 involved in capsule export.

Analysis of the nucleotide sequence 5′ of region 1 and the regulation of the K10 capsule gene cluster by RfaH. DNA sequence obtained 5′ to kpsD$_{K10}$ (Fig. 2) revealed the presence of a JUMPstart motif located 1,326 bp from kpsD$_{K10}$. The JUMPstart motif is found 5′ of numerous virulence-associated operons and appears to be required for RfaH-dependent transcriptional antitermination (16, 21, 42). For group II capsule gene clusters, it has been suggested that RfaH and the JUMPstart sequence together allow transcription from the region 3 promoter to override rho-dependent terminators and permit transcription to run through into region 2 (42). A mutation in rfaH results in termination of the transcript upstream of region 2, thus preventing capsule synthesis. The transcriptional start site 5′ of region 3 in E. coli K5 has been mapped to 741 bp 5′ to the ATG of kpsM (42). While the transcription start site 5′ to region 1 in the K10 capsule gene cluster has not been identified, the sequence surrounding the transcriptional start site in K5 is highly homologous to the equivalent sequence in K10 (Fig. 2), suggesting that transcription may be initiated at the same point in both group II and group III capsule gene clusters. Although the K10 region 1 promoter region is highly homologous to the K5 region 3 promoter, the K10 sequence contains a deletion relative to the K5 sequence (Fig. 2). The location of this deletion positions the JUMPstart sequence 531 bp closer to the putative transcriptional start site relative to that in K5 (Fig. 2). The significance of this deletion in the regulation of K10 capsule gene expression is not clear. However, to determine if the K10 capsule cluster is RfaH regulated, plasmid pRP1 (Fig. 1A) was introduced into the E. coli K-12 strain JM101 and its rfaH mutant, MS135 (42), and K10 capsule production was monitored by measuring the sensitivity to K10-specific bacteriophage. Strain JM101(pRP1) expressed a K10 capsule and was sensitive to bacteriophage, whereas strain MS135(pRP1) was bacteriophage resistant, indicating a lack of K10 capsule. This demonstrates that the K10 capsule cluster is RfaH regulated. It can be inferred, by analogy to E. coli K5, that transcription from the K10 region 1 promoter proceeds through region 1 into region 2. The presence of a JUMPstart sequence 5′ to K54 region 1 implies that the K54 capsule is also RfaH regulated (37). However a putative rho-independent terminator, which could possibly prevent transcription from re-

gion 1 running into region 2 even with a functional RfaH, has been identified immediately 3′ to kpsE$_{K54}$ (37). This putative terminator is not present in the K10 sequence.

The region of nucleotide homology between the K5 and K10 region 1 promoter regions extends approximately 190 bp into the coding sequence of the K5 kpsM gene (Fig. 2). This DNA sequence is identical to the corresponding region in K54 (37) but does not show significant homology to the functional kpsM$_{K10}$ and kpsM$_{K54}$ genes located 3′ to kpsD (Fig. 1A). The fact that the cryptic kpsM gene shows homology to kpsM$_{K5}$ but not to the group III kpsM genes suggests that the group II and III promoter regions were acquired from a common ancestor expressing a group II capsule. Russo et al. (37) have suggested that the group III capsule cluster was derived from a group II cluster by insertion of foreign capsule genes 3′ of the start of the kpsM gene. As a consequence, the group III region 1 genes may be transcribed from the group II region 3 promoter. However, one must be cautious, since it is known that the region 3 promoter of the K5 capsule gene cluster is thermoregulated (36a) whereas group III capsules are expressed at all growth temperatures (18). As such, it is possible that the group III region 1 promoter is located between the cryptic kpsM gene and the kpsD$_{k10}$ gene (Fig. 1) or that the deletions in the group III promoter region abolish temperature regulation. The regulation of both group II and III capsule gene clusters by RfaH may imply that RfaH regulation of capsule expression is an evolutionary advantage to pathogenic E. coli strains.

The locations of any region 3 promoters remain to be elucidated. The close proximity of rfbC$_{K10}$ and kpsC$_{K10}$ implies that either promoter-transcribing region 3 lies within region 2 or expression of group III capsule genes occurs from a single promoter 5′ to region 1 which generates a single large transcript spanning the capsule gene cluster. Such a transcriptional organization would be consistent with that seen for many other RfaH-regulated operons (21). It is possible, however, that transcription of kpsS$_{K10}$ is initiated from a promoter within the kpsC-to-kpsS intergenic region.

Detection of IS110 and IS3 5′ to the K10 capsule gene cluster. It has been postulated that mobilization of the group III determinants into the progenitor group II strain was mediated by IS110 (37). A remnant of IS110 from Streptomyces coelicolor (4) is present 53 bp 3′ to the cryptic kpsM gene in K54 (37) and in K10 (Fig. 2). In addition to the IS110 sequence, a region with 99% homology to IS3 (44) was identified 5′ to the K10 JUMPstart sequence (Fig. 2). Three ORFs were identified within the IS3 sequence (Fig. 1A). This IS element was not identified in the reported K54 sequence and would appear to be specific to the K10 capsule gene cluster. IS elements have been implicated in the duplication of genes in the group I capsule locus of E. coli K30 (8) and have been found near the capsule genes of Klebsiella pneumoniae (46). In addition, remnants of IS600 and IS630 elements may have been involved in lateral transfer of a pathogenicity island into enteropathogenic and enterohemorrhagic E. coli (7, 30). Conceivably, a block of capsule genes could be mobilized through transposition if they were flanked by IS elements. Although numerous IS3 elements are present on the E. coli chromosome (1), we have not identified a second IS3 element within or flanking the K10 capsule gene cluster. However, a second flanking IS3 element could have been lost through subsequent recombination events. Alternatively, capsule genes could be transferred by homologous recombination between IS elements located in E. coli and in DNA from another organism.

Identification of a cryptic prophage 3′ to region 3. DNA sequence analysis of the region 3′ to region 3 (Fig. 1A) revealed four contiguous ORFs with homology to those which

```
K10   ACGCGATAAATTTATTCACTATTTTCAGCAATGGCGTGAACGCCAGTTAAGCCGTGGCGAACACTGGCTGACACAACACCTGGCGGGGCGTTCTCCGCGT   100
YghD  R  D  K  F  I  H  Y  F  Q  Q  W  R  E  R  Q  L  S  R  G  E  H  W  L  T  Q  H  L  A  G  R  S  P  R
K10   GAAAAGGGCATGTTACTGGCAGCGGTGGTGTTCCTGTTTAGCGTCGGATATTACGTCCTCATCTGGCAGCCGTTGAGCGAACGGATAGAACAACAGGAGA   200
      E  K  G  M  L  L  A  A  V  V  F  L  F  S  V  G  Y  Y  V  L  I  W  Q  P  L  S  E  R  I  E  Q  Q  E  T
K10   CGATGTTGCAGCAGCTCGTGGCGATGAACGCGCGACTGAAGAGCGCCGGCGGGATATTATTACGTCCGCGCGAAAATCCGCCACAACAACGCCTGCGCAGGT   300
        M  L  Q  Q  L  V  A  M  N  A  R  L  K  S  A  A  P  D  I  I  A  A  R  K  S  A  T  T  T  P  A  Q  V
K10   ATCGCGGGTCATCAGCGACAGTGCTTCCGCGCACTCGGTGGTCATTAAGCGGATAGCCGAGCGCGGGGAGAATATCCAGGTCTGGATAGATCCTGTGGTA   400
        S  R  V  I  S  D  S  A  S  A  H  S  V  V  I  K  R  I  A  E  R  G  E  N  I  Q  V  W  I  D  P  V  V
K10   TTTAATGATCTGCTGAACTGGTTAAACGCGCTGGATGAAAAGTATGCGCTGCGGGTGACACAAATTGATGTCAGTGCAGATGAAAAGGCTGGAATGGTGA   500
        F  N  D  L  L  N  W  L  N  A  L  D  E  K  Y  A  L  R  V  T  Q  I  D  V  S  A  D  E  K  A  G  M  V  N
K10   ACGTGCAGCGGCTGGAGTTTGGATGGGGGTAAGTATTGCCATCCCCTTAAACCCCACCTGATAACC-TTAGCCCAAAGGTGAGTGCGGAACTGTCAAACC   599
   D  V  Q  R  L  E  F  G  W  G   **  ** ********  ****** **********  ******* ********** ********  ********************
K5                                  TAGGTATTGCCATTTCCTTAA-CCCCACCTGATAACCCTTAGCCCAATGGTGAGTGCGGAACTGTCAAACC   243

K10   CGCTTACCGGGCTGAGTTCCGATGTAGCAAGGGTTATCTGATCCTACCCACGTAATATGGACACAGGCCTAAGCGAGGTTCTTGTTTTCAAATTGTTCCG   699
      *** *********************** ******     IS3
K5    AGCTCACCGGGCTGAGTTCCGATGTAGCAAGATTTATCT                                                              282

K10   GACTGAGGCCGCCACACCAACTGTGCCGCCGCCACCGATTGTAATCACATTCGATATAATTAAACACCGTTGCCCGCATTATTTCCCGGCTGATAAAGTG   799
K10   TTCTCCATGGATACATTCCACTTTCAGCGAATGAAAGAAGCTTTCACGCAGGCATTATCGTAGCAGCAACCTTTTGCGCTCATACTTCCACGCAGATTA   899
K10   TGCCGCTTCAGTTGCGCCTGATAATCTGCTGAACAGTACTGGCCTCCACGGTCCGTGAACGATAACGTTCCGGGGCCTCTTACGCCACACAGCGCCA    999
K10   TGTGCAGGGCATCGCAGGCCAGTTGCGCCGTCATGCGTGGCGACATTGACCAGCCAATAACGGCACGTGACCACAGGTCAATGACCACTGCCAGATACAG   1099
K10   CCAGCCTTCATCTGTACGTAAGTACGTGATGTCTCCTGCCCACTTCTGGTTCGGGCCACTGGCGTAAAAATCCTGCTCCAACAGATTTTCTGACACAGGC   1199
K10   AGGCCGTGTGCGCGGTAGCTGACCGGGCTGAACTTCCGGGAGGCGTTTGCCCTCAGTCCCTGACGGCGCAGGCTTGCCGCCACGGTTTTTACGTTAAAGG   1299
K10   GGTAACCCTGAGCACGCAGTTCATCCGTCAGGCGTGGGGCACCGTAACGCTGTTTTGACCGGGTAAAAGCCGCGAGGACAACGCTGTCGCAGTGTTGGCG   1399
K10   GAACTGCTGACGCGTGCTTATCCTTGTCCGCCGCTGACACCACGTATACCAGCCGCTGCGGGCCACCCGGAGCACGCGGCACATTGCTTTGATGCTGAAC   1499
K10   TCAGCCTGATGTTTTTCAATAAAGACATACTTCATTTCAGGCGCTTCGCGAAGTATGTCGCGGCCTTTTGGAGGATAGCCAGCTCGTCTTCATCCGTTCTGC   1599
K10   CAGCTGGCGTTTGAGACGTGCAATCTCGGTAGACATCTCCAGTTCACGTTCAGAAGACGTCTGCTGATTTTGCTGTTTACTGCGCCAGTTGTAGAGCTGT   1699
K10   GATTCATACAGGCTGAGTTCACGGGCTGCGGCAGTAACACCGATGCGTTCAGCAAGCTTCAGGGCTTCACTGCGAAATTCAGGCGAATGCTGTTTACGGG   1799
K10   GTTTTTTACTGGTTGATACTGTTTTTTGTCATGTGAGTCACCTCTGACTGAGAGTTTTACTCACTTAGCCGCGTGTCCACTATTGCTGGGTAAGATCAATCT   1899

K10   CCCTCCAGACCACCACGGAATGTTCGATAATACTGTAGGAATTCGTTTTTCACTATGCAGCCCTTGTGAATCGACACCTTCCTCAGACTGTAAAGTTGCT   1999
      ********************* *** **** *****   ******** *** **** * ** *********** ** ********** ****
K5    CCCTCCAGACCACCACGGAATGTTCGGTAACACTGAAGGAACCTGTTTTTCACCATGTAGCCTTCGTTAATCGACACCTTTCTAAGACTGTAAAGATGCT   382

K10   TAAGTGTTTTGACGCATTATGATTATTAATAGCTGTAATGAATCATTGAGTAATAATTGATAGGTTAAAACATATAGGATAATTCTTGTGTGATCTGTAT   2099
      ************ ********************* ****** ************** ********** * ****************** **************
K5    TAAGTGTTTTGACACATTATGATTATTAATAGTTGTAATAAATCATTGAGTAACAATTGATAGGTCAAAACATATAGGATAATTCTTGTGTGATCTGTAT   482

K10   TTTGTGTGACTTGGAAATTAGTAAAATTCCTAGAGATAATCAATAAGTGAATTTTAAATAAGGAAAATAATTTCTACGAAAACTGTTCGTCGAGCTTATG   2199
      **************************** ********* ****** *** ************** * ********* *  ***** ******** *** *
K5    TTTGTGTAGCTTGGAAATTAGTAAAATTCCTGGAGATAATCAGAAAGGGAATTTCAAATAAGCATAATAAATTTCCAGTGAAACTATTCGTCGATTTTAAG   582
                                                                                            JUMPstart
K10   AAAATGTTTGCTCAAGTATATCTAGT    TCATCCAGTTGCAGATAAGTAT-ACCTCCAGTGTATTGGTAGCTGTAAAGCCAAGGGCGGTAGCATTCCTGA   2295
      ****** ****** ********* *      * ******** ** ***** ** ***** ********************* ****************** *****
K5    AAAATGTCTGCTCAGGTATATCTACT Δ TTATCCAGTTACAAATAAGCATTACCTCCAGTGTATTGGTAGCTGTTAAGCCAAGGGCGGTAGCGTACCTGA   1110
             Cryptic kpsM
K10   AGAGATTAGGAGCACATCATCAAATT-CAAAAGTGGGTTTGAAGTCTAGAACGTCACCATAGAGACATTATTTATACGAGAAATACGAACACGTTTTGG    2394
      ********** ********** **   *** **** ***** **** **** ***** ***** *********** ****** ***********  *****
K5    AGAGATTAGGATCACATCATCAAATGGCAAGAAGTGGATTTGAAGTCCAGAAAGTCACCGTAGAGGCATTATTTCTACGAGAAATACGAACACGCTTTGG   1210
       Cryptic kpsM cont.
K10   TAAATTTTGTATGGGATATCAGTGGGCGATTCTAGAACCTTCTGTGCATCTGCTGGGCATTGGGGGTTATGTTATGCACCACGATGCCAGACATTTCGTT   2494
      *** ** ** ******** *********** **** *** ***** ***** ** ****** * ****** ** ***** ***  ** ** **  ***
K5    TAAGTTCCGTCTGGGATATTTGTGGGCGATTCTCGAACCCTCTGCGCATTTGCTGATACTGTTGGGGCATTTTTGGTTACATT-ATGCACCGCACGATGCC   1309

K10   CCAGGAAAATAAGCTGCGAAGTGTTCTACAAAAGCTGCTTCCATGATGAGCATGGTCGTGGCCTGACACTGACTCCAGCCATGCTAGTCAGGACCGGGCA   2594

K10   AAGGGGGGCAAGAAGCCGCCATTCTATCGCAAGTGTATCTCCCCTCTTTTTTATTATTTTTGTTTATCTGTTATTGATAAAGTACGACTATTATTATTGC   2694
K10   GACAATTATTTACTAGTTAGCATTTTTTATCGAAGTTGAGGGTGTGCTGAACGGCTCATGAAAATATTAGAGTGTTAAAGTTTTTCAGGACTTAGATTTT   2794
K10   TGGGGATAGTAATTTAAGAATTTATGAAAACGTCTATGCCTTACTGAGTCATTACAAAACAAATTTATTTAATTGTTCCCTCTTTATGTATAAGAGGAAT   2894
K10   TTTCTCTTCCATATGGTTGTTATTATTAAATTTTATAAAAAAGATAGCAATATACGGGGCCATAAAAATAATATTTTTAGGGATGTGTGTTGCATATGTA   2994
K10   GTACATATCCAAAAGATAAAGGATATATTCGTAATTAGTTATAGACAGATGATGTAAACACCAGTTGACTAAAGTCAATCTTATACTGACAACATCTATG   3094
K10   ATTAAATTTGCGTGGTTATAATTTTAAATATCTTATATTTATGGGCTATTTTTGATATCTGTCAGAGTATCAATAACAGAAGGTAATTGTCCGACATACTA   3194
K10   TCAACTTCGGGATAACGTTTTAAAATGCACCTTACACATCGTATTTTATTATTTCCACTAATCTTTTTTATAACGGCCTGCGCACATGATCCAAAACAAG   3294
K10   TTGAAGCCTCTCGTCCATTGGTAACAGCGATTAATTCTTCTTATTCTCTTATTCCTGAAGATTTGCAGGCACCATTAAATAACCAAGATCAAGGCACGAC   3394
K10   ATTCAACAAAAATGGCGTAATTTATACTATTGAGGAAAGGTATATATCGGCTTTAGGTTCTCAATGCATAAAGTTAAGTTATGCGATGAATAAAAATTAT   3494
K10   TCAAAGCGAAGTGTTGTTTGTAAAGAGAATAACAAGTGGTATCAAGTACCTCAGTTGGAACAAACATCAGTTAGCACTTTGCTTATTGAAGAATAAAGTG   3594

K10   GAGGGTAGACGGTTAGAAAATAATGAAAATTTCGCAACTTAGCACT                                                         3640
KpsD                                                 M  K  I  S  Q  L  S  T
```

FIG. 2. DNA sequence 5′ to region 1 of the *E. coli* K10 capsule gene cluster. The sequence homologous to IS*3* and the JUMPstart sequence are indicated by shaded boxes. The K5 sequence is displayed under the K10 sequence where the two are homologous (boldface). The numbering of the K10 DNA starts at the beginning of the known sequence within the *yghD* gene. The K5 sequence is numbered from the stop codon of the *yghD* gene. The IS*3* element interrupts regions of homology between K5 and K10. The cryptic *kpsM* gene is indicated by unshaded boxes. The remnant of IS*110* is marked by a solid line. Δ indicates the region of K10 sequence deleted (531 bp) relative to K5. An arrow marks the transcriptional start site in the K5 sequence.

are part of a putative prophage inserted at one end of the LEE pathogenicity island of enterohemorrhagic *E. coli* EDL933 (30). This prophage is related to retronphage φR73 and other CP4-like cryptic prophages found in *E. coli* K-12 (1, 43). In strain EDL933, the putative prophage contains 13 ORFs flanked by directly repeated attachment (*att*) sequences (30). Numerous virulence determinants have been associated with lysogenic bacteriophages (5). These genes are carried on the bacteriophage genome and are found inserted into the host chromosome between two *att* sites. The prophage in strain EDL933 was not implicated in transduction of the LEE pathogenicity island, because both *att* sites were at only one end of the LEE gene cluster. We were unable to identify the *att* direct repeats within the K10 sequence due to absence of DNA sequence data for both ends of the prophage. In the φR73 family of prophages, the bacteriophage integrase gene is located ad-

jacent to one of the *att* sites. An integrase homologue was not identified among the four ORFs sequenced. However, one may be present within the unsequenced region distal to the K10 cluster. Therefore, transduction of the K10 locus cannot be ruled out, due to the possibility of unidentified *att* repeats flanking the capsule genes.

Localization of the K10 capsule gene cluster on the *E. coli* chromosome. The K10 sequence from bp 1 to 530 encodes a polypeptide with 93% amino acid identity (95% similarity) to the YghD protein of *E. coli* K-12 (1) (Fig. 1A). This protein is presumed to be an M-type component of a general protein secretion pathway. The *pheV* gene, coding for phenylalanine-tRNA, is positioned 150 bp 3′ to *yghD* in *E. coli* K-12. DNA homologous to the *pheV* gene was not found in the region sequenced in this study. Nucleotide homology between the K-12 and K10 sequences ceases immediately 3′ of the *yghD* stop codon, indicating the region of insertion of the K10 capsule cluster. The position of the K10 capsule genes between *yghD* and *pheV* on the *E. coli* K-12 chromosome is consistent with the map position near *serA* previously determined for the *E. coli* K10 capsule gene cluster (24). This is the same insertion site reported for the K54 and K5 capsule gene clusters (37). Presumably, *pheV* is located at the distal end of the ϕR73-like prophage identified flanking the K10 capsule gene cluster. ϕR73 and other related bacteriophages integrate at the 3′ end of tRNA genes (5, 43). The observation that numerous pathogenicity islands are found adjacent to prophages (5) implicates these prophages in the evolution of virulence gene clusters inserted in the bacterial chromosome near tRNA genes. It is notable that the K5 capsule gene cluster is inserted at the same chromosomal location as that of K10 although prophage genes have not been identified adjacent to the K5 cluster. Therefore, the ϕR73-like prophage may have played a role in the acquisition of the primordial group III capsule gene cluster mediating the integration into a group II capsule gene cluster at this site. The remnant of this group II cluster is identified by the promoter, JUMPstart, and group II *kpsM* sequences 5′ to the group III region 1 genes, with the remainder of this group II cluster having been lost during the subsequent evolution of the group III capsule gene cluster. Identification of the remnants of an IS*110* sequence 5′ to region 1 in both the K10 and K54 capsule gene clusters would suggest that IS*110* may have played a role in the acquisition of the primordial group III capsule gene cluster. Detection of IS*3* 5′ to region 1 of the K10 capsule gene cluster, but not in K54, would suggest that this IS element has been involved in events specific to the evolution of the K10 capsule gene cluster.

Therefore using the *E. coli* K10 capsule as a paradigm, we have described the genetic organization of and identified the proteins conserved among different group III clusters. The similarity of the group III proteins to those in group II and the ability to complement mutations in group II polysaccharide export indicate a conserved functionality among group II and group III capsule clusters. We have also shown that the group III capsule genes are under RfaH-mediated transcriptional regulation. Finally, we have identified IS*3* and IS*110* elements and a ϕR73-like prophage which may have contributed to the evolution of the group III capsule clusters from a group II progenitor.

REFERENCES

1. **Blattner, F. R., G. Plunkett III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao.** 1997. The complete genome sequence of *Escherichia coli* K-12. Science **277:**1453–1462.
2. **Bliss, J. M., and R. P. Silver.** 1996. Coating the surface: a model for expression of capsular polysialic acid in *Escherichia coli* K1. Mol. Microbiol. **21:**221–231.
3. **Bronner, D., V. Sieberth, C. Pazzani, I. S. Roberts, G. J. Boulnois, B. Jann, and K. Jann.** 1993. Expression of the capsular K5 polysaccharide of *Escherichia coli*: biochemical and electron microscopical analyses of mutants with defects in region 1 of the K5 gene cluster. J. Bacteriol. **175:**5984–5992.
4. **Bruton, C. J., and K. F. Chater.** 1987. Nucleotide sequence of IS*110*, an insertion sequence of *Streptomyces coelicolor* A3(2). Nucleic Acids Res. **15:**7053–7065.
5. **Cheetham, B. F., and M. E. Katz.** 1995. A role for bacteriophages in the evolution and transfer of bacterial virulence determinants. Mol. Microbiol. **18:**201–208.
6. **Cieslewicz, M. J., and E. R. Vimr.** 1997. Reduced sialic acid capsule expression in *Escherichia coli* K1 mutants with defects in *kpsF*. Mol. Microbiol. **26:**237–249.
7. **Donnenberg, M. S., L.-C. Lai, and K. A. Taylor.** 1997. The locus of enterocyte effacement pathogenicity island of enteropathogenic *Escherichia coli* encodes secretion functions and remnants of transposons at its extreme right end. Gene **184:**107–114.
8. **Drummelsmith, J., P. A. Amor, and C. Whitfield.** 1997. Polymorphism, duplication, and IS*1*-mediated rearrangement in the chromosomal *his-rfb-gnd* region of *Escherichia coli* strains with group IA capsular K antigens. J. Bacteriol. **179:**3232–3238.
9. **Finke, A., D. Bronner, A. V. Nikolaev, B. Jann, and K. Jann.** 1991. Biosynthesis of the *Escherichia coli* K5 polysaccharide, a representative of group II capsular polysaccharides: polymerization in vitro and characterization of the product. J. Bacteriol. **173:**4088–4094.
10. **Finke, A., B. Jann, and K. Jann.** 1990. CMP-KDO synthetase activity in *Escherichia coli* expressing different capsular polysaccharides. FEMS Microbiol. Lett. **69:**129–134.
11. **Finke, A., I. S. Roberts, G. J. Boulnois, C. Pazzani, and K. Jann.** 1989. Activity of CMP-KDO synthetase in *Escherichia coli* strains expressing the capsular K5 polysaccharide: implication for polysaccharide biosynthesis. J. Bacteriol. **171:**3074–3079.
12. **Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J.-F. Tomb, B. A. Dougherty, J. M. Merrick, K. McKenney, G. G. Sutton, W. FitzHugh, C. A. Fields, J. D. Gocayne, J. D. Scott, R. Shirley, L. I. Liu, A. Glodek, J. M. Kelley, J. F. Weidman, C. A. Phillips, T. Spriggs, E. Hedblom, M. D. Cotton, T. Utterback, M. C. Hanna, D. T. Nguyen, D. M. Saudek, R. C. Brandon, L. D. Fine, J. L. Fritchman, J. L. Fuhrman, N. S. Geoghagen, C. L. Gnehm, L. A. McDonald, K. V. Small, C. M. Fraser, H. O. Smith, and J. C. Venter.** 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science **269:**496–512.
13. **Frosch, M., U. Edwards, K. Bousset, B. Kausse, and C. Weisgerber.** 1991. Evidence for a common molecular origin of the capsule gene loci in Gram-negative bacteria expressing group II capsular polysaccharides. Mol. Microbiol. **5:**1251–1263.
14. **Frosch, M., and A. Müller.** 1993. Phospholipid substitution of capsular polysaccharides and mechanisms of capsule formation in *Neisseria meningitidis*. Mol. Microbiol. **8:**483–493.
15. **Frosch, M., C. Weisgerber, and T. F. Meyer.** 1989. Molecular characterization and expression in *Escherichia coli* of the gene complex encoding the polysaccharide capsule of *Neisseria meningitidis* group B. Proc. Natl. Acad. Sci. USA **86:**1669–1673.
16. **Hobbs, M. and P. R. Reeves.** 1994. The JUMPstart sequence: a 39 bp element common to several polysaccharide gene clusters. Mol. Microbiol. **12:**855–856.
17. **Jann, B., and K. Jann.** 1990. Structure and biosynthesis of the capsular antigens of *Escherichia coli*. Curr. Top. Microbiol. Immunol. **150:**19–42.
18. **Jann, K., and B. Jann.** 1992. Capsules of *Escherichia coli*, expression and biological significance. Can. J. Microbiol. **38:**705–710.
19. **Kroll, J. S., B. Loynds, L. N. Brophy, and E. R. Moxon.** 1990. The *bex* locus in encapsulated *Haemophilis influenzae*: a chromosomal region involved in capsule polysaccharide export. Mol. Microbiol. **4:**1853–1862.
20. **MacLachlan, P. R., W. J. Keenleyside, C. Dodgson, and C. Whitfield.** 1993. Formation of the K30 (group I) capsule in *Escherichia coli* O9:K30 does not require attachment to lipopolysaccharide lipid A-core. J. Bacteriol. **175:**7515–7522.
21. **Neito, J. M., M. J. A. Bailey, C. Hughs, and V. Koronakis.** 1996. Suppression of transcription polarity in the *Escherichia coli* haemolysin operon by a short upstream element shared by polysaccharide and transfer determinants. Mol. Microbiol. **19:**705–713.
22. **Ørskov, F., and I. Ørskov.** 1992. *Escherichia coli* serotyping and disease in man and animals. Can. J. Microbiol. **38:**699–704.

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

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

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

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

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

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

29. **Pearce, R. and I. S. Roberts.** 1995. Cloning and analysis of gene clusters for production of the *Escherichia coli* K10 and K54 antigens: identification of a new group of *serA*-linked capsule gene clusters. J. Bacteriol. **177:**3992–3997.

30. **Perna, N. T., G. F. Mayhew, G. Pósfal, S. Elliott, M. S. Donnenberg, J. B. Kaper, and F. R. Blattner.** 1998. Molecular evolution of a pathogenicity island from enterohemorrhagic *Escherichia coli* O157:H7. Infect. Immun. **66:** 3810–3817.

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

32. **Rigg, G. P., B. Barrett, and I. S. Roberts.** 1998. The localization of KpsC, S and T, and KfiA, C and D proteins involved in the biosynthesis of the *Escherichia coli* K5 capsular polysaccharide: evidence for a membrane-bound complex. Microbiology **144:**2905–2914.

33. **Roberts, I. S.** 1995. Bacterial capsules in sickness and in health. Microbiology **141:**2023–2031.

34. **Roberts, I. S.** 1996. Biochemistry and genetics of capsular polysaccharide production in bacteria. Annu. Rev. Microbiol. **50:**285–315.

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

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

36a.**Rowe, S., and I. S. Roberts.** Unpublished results.

37. **Russo, T. A., S. Wenderoth, U. C. Carlino, J. M. Merrick, and A. J. Lesse.** 1998. Identification, genomic organization, and analysis of the group III capsular polysaccharide genes *kpsD*, *kpsM*, *kpsT*, and *kpsE* from an extraintestinal isolate of *Escherichia coli* (CP9, O4/K54/H5). J. Bacteriol. **180:**338–349.

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

39. **Sieberth, V., B. Jann, and K. Jann.** 1993. Structure of the K10 capsular antigen from *Escherichia coli* O11:K10:H10, a polysaccharide containing 4,6-dideoxy-4-malonylamino-D-glucose. Carbohydr. Res. **246:**219–228.

40. **Simpson, D. A. T. C. Hammerton, and I. S. Roberts.** 1996. Transcriptional organization and regulation of expression of region 1 of the *Escherichia coli* K5 capsule gene cluster. J. Bacteriol. **178:**6466–6474.

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

42. **Stevens, M. P., B. R. Clarke, and I. S. Roberts.** 1997. Regulation of the *Escherichia coli* K5 capsule gene cluster by transcription antitermination. Mol. Microbiol. **24:**1001–1012.

43. **Sun, J., M. Inouye, and S. Inouye.** 1991. Association of a retroelement with a P4-like cryptic prophage (retronphage ϕR73) integrated into the selenocytostyl tRNA gene of *Escherichia coli*. J. Bacteriol. **173:**4171–4181.

44. **Timmerman, K. P., and C. P. D. Tu.** 1985. Complete sequence of IS3. Nucleic Acids Res. **13:**2127–2139.

45. **Vimr, E. R.** 1991. Map position and genome organization of the *kps* cluster for polysialic acid biosynthesis in *Escherichia coli* K1. J. Bacteriol. **173:**1335–1338.

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

47. **Whitfield, C., and M. A. Valvano.** 1993. Biosynthesis and expression of cell-surface polysaccharides in Gram-negative bacteria. Adv. Microb. Physiol. **35:** 135–246.