JOURNAL OF BACTERIOLOGY, Jan. 1998, p. 338–349 Vol. 180, No. 2
0021-9193/98/$04.00+0

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)

THOMAS A. RUSSO,[1,2,3]* SUZANNE WENDEROTH,[1]† ULRIKE B. CARLINO,[1,3]
JOSEPH M. MERRICK,[2,3] AND ALAN J. LESSE[1,2,4,5]

Department of Medicine,[1] Department of Microbiology,[2] The Center for Microbial Pathogenesis,[3] Department of Pharmacology and Toxicology,[4] and the Buffalo VA Medical Center,[5] SUNY at Buffalo, Buffalo, New York 14215

Received 12 June 1997/Accepted 8 November 1997

Group III capsular polysaccharides (e.g., K54) of extraintestinal isolates of *Escherichia coli*, similar to group II capsules (e.g., K1), are important virulence traits that confer resistance to selected host defense components in vitro and potentiate systemic infection in vivo. The genomic organization of group II capsule gene clusters has been established as a serotype-specific region 2 flanked by regions 1 and 3, which contain transport genes that are highly homologous between serotypes. In contrast, the organization of group III capsule gene clusters is not well understood. However, they are defined in part by an absence of genes with significant nucleotide homology to group II capsule transport genes in regions 1 and 3. Evaluation of isogenic, Tn*phoA*-generated, group III capsule-minus derivatives of a clinical blood isolate (CP9, O4/K54/H5) has led to the identification of homologs of the group II capsule transport genes *kpsDMTE*. These genes and their surrounding regions were sequenced and analyzed. The genomic organization of these genes is distinctly different from that of their group II counterparts. Although *kps*$_{K54}$*DMTE* are significantly divergent from their group II homologs at both the DNA and protein levels *phoA* fusions and computer-assisted analyses suggest that their structures and functions are similar. The putative proteins Kps$_{K54}$M and Kps$_{K54}$T appear to be the integral membrane component and the peripheral ATP-binding component of the ABC-2 transporter family, respectively. The putative Kps$_{K54}$E possesses features similar to those of the membrane fusion protein family that facilitates the passage of large molecules across the periplasm. At one boundary of the capsule gene cluster, a truncated *kpsM* (*kpsM*$_{truncated}$) and its 5′ noncoding regulatory sequence were identified. In contrast to the complete *kps*$_{K54}$*M*, this region was highly homologous to the group II *kpsM*. Fifty-three base pairs 3′ from the end of *kpsM*$_{truncated}$ was a sequence 75% homologous to the 39-bp inverted repeat in the IS*110* insertion element from *Streptomyces coelicolor*. Southern analysis established that two copies of this element are present in CP9. These findings are consistent with the hypothesis that CP9 previously possessed group II capsule genes and acquired group III capsule genes via IS*110*-mediated horizontal transfer.

Over 80 serologically and chemically unique capsular polysaccharides can be produced by *Escherichia coli* (22, 31). Initially, these polysaccharides were divided into group I and group II based on chemical, physical, genetic, and microbiological distinctions (20, 21). Subsequently, the division of the group II capsules into groups II and III (formerly I and II) has been proposed (36).

Group I capsules are chemically and physically characterized by a high molecular weight (>100,000), an acidic component usually consisting of hexuronic acid or pyruvate, a low charge density and electrophoretic mobility, and stability at pH 5 to 6 at 100°C. Group I capsules may protect against desiccation and may contribute to adherence in enteric disease-producing isolates of *E. coli* (17, 27, 30). However, a role in the pathogenesis of extraintestinal *E. coli* infection has not been demonstrated (45).

In contrast, epidemiologic and experimental evidence supports a role for group II and group III capsules as virulence factors for extraintestinal infection (10, 43, 44), and these capsules possess multiple similarities with the capsules of pathogenic strains of *Neisseria meningitidis* and *Haemophilus influenza* (26). The group II capsules are characterized by a molecular weight of <50,000; hexuronic acids; *N*-acetyl neuraminic acid, phosphate, or 2-keto-3-deoxyoctonic acid (KDO) as acidic components; a higher charge density and electrophoretic mobility; and a general lack of stability at pH 5 to 6 at 100°C. Several group II capsules are linked to KDO-phosphatidic acid, which may serve both as a recognition signal for transport across the cytoplasmic membrane and as a membrane anchor (6, 7). The genes that code for these capsules have been mapped near *serA* (32, 33, 53), and these capsules are coexpressed with a large number of O antigens. It was originally believed that a given *E. coli* strain possessed only genes for either a group I or a group II capsule. However, it has been subsequently shown that three group II or III capsule (K1, K5, and K54)-producing strains were also capable of producing the group I capsule colanic acid (25, 46). This finding suggests the possibility that many if not all strains of *E. coli* have the capability to produce a group I capsule, whereas only a subset can produce a group II or group III capsule. The gene clusters coding for the group II capsules K1, K4, K5, K7, K12, and K92 have been cloned and extensively studied (particularly K1 and K5) and have a common organization of three func-

* Corresponding author. Mailing address: Department of Medicine, Division of Infectious Diseases, 3435 Main St., Biomedical Research Building, Room 141, Buffalo, NY 14214. Phone: (716) 829-2674. Fax: (716) 829-2158. E-mail: trusso@acsu.buffalo.edu.

† Present address: Cornell University Medical School, New York, NY 10021.

tional regions (38–40, 48). Region 2, which is unique for a given capsular antigen, codes for genes whose products are responsible for the synthesis of the K-specific serotype. This region is flanked by regions 1 and 3, which are highly conserved among the group II capsule gene clusters evaluated to date. In fact, a DNA probe generated from region 1 in the K1 capsule gene cluster was used to identify the K4, K5, K7, K12, and K92 capsule gene clusters (13, 38). Region 1 contains six genes (*kpsFEDUCS*), and region 3 contains two genes (*kpsMT*); each region is organized in a single transcriptional unit and is temperature regulated. These gene products are needed for transport of the capsular polysaccharide across the cytoplasmic membrane and assembly onto the cell's surface (4).

Group III capsules were originally categorized as group II capsular polysaccharides. Although these groups have similar biochemical and physical characteristics (31), map to the same location on the chromosome (32, 33, 53), confer resistance to selected host defense components in vitro, and potentiate systemic infection in vivo (10, 43, 44), differences exist between them. Group III serotypes K2, K3, K10, K11, K19, and K54 do not show temperature regulation of capsule expression, a characteristic which correlates with constitutive levels of CMP-KDO activity (14), whereas group II capsules have increased capsule expression and CMP-KDO activity at 37°C. Further, only the K2 capsule gene cluster, but not those from K3, K10, K11, K19, and K54, possesses DNA sequences homologous to group II capsule gene cluster regions 1 and 3 on the basis of Southern analysis (12, 36). In a recent study that described the cloning of a K10 and a different K54 (*E. coli* A12b) capsule gene cluster, Southern analysis and complementation studies were used to elucidate group III capsule gene organization (36). The preliminary results of these analyses suggested that a central serotype-specific region was flanked by two regions in which there was homology between the K10 and K54 gene clusters. Further, complementation studies demonstrated that group II $kps_{K5}D$ and $kps_{K5}E$ mutations, but not $kps_{K5}M$ or $kps_{K5}T$ mutations, were complemented by subclones from the K10 and K54 capsule gene clusters. Therefore, this finding suggested that, despite a lack of DNA homology, functional homology exists, at least in part, between proteins involved in the export of group II and group III capsular polysaccharides. The combination of these findings has resulted in the designation of serotypes K3, K10, K11, K19, and K54 (with or without K2) as group III capsules (36) and has suggested that these gene clusters are phylogenetically divergent from those of group II. In support of this concept, a clonal group of clinical *E. coli* isolates was recently identified from multiple geographic regions (23, 24). These strains were characterized in part by possession of the $papG_{J96}$ (class I) and $prsG_{J96}$ (class III) genes, the O4-specific antigen moiety of lipopolysaccharide, the H5 flagellar antigen, the F13 fimbrial antigen, and a group III capsule (K3, K10, and K54/96).

Researchers in our laboratory have been studying a clinical bacteremic isolate of *E. coli* (CP9, O4/K54/H5) as a model pathogen for extraintestinal infection (42). Its group III K54 capsular polysaccharide has been shown to be important for serum resistance in vitro (44) and systemic infection in vivo (43) but not for resistance to bactericidal permeability-increasing protein in vitro (45) or urinary tract infection in vivo (41). Previously, we reported the construction and initial characterization of Tn*phoA*-generated, isogenic K54-minus derivatives of CP9 that were used in these studies (42). In this study, we describe the genomic location and novel organization of a portion of the K54 capsule gene cluster, the DNA sequences of $kps_{K54}DMTE$, and an analysis of these genes.

MATERIALS AND METHODS

Strains. The strains used for this study are listed in Table 1. The wild-type strain (CP9, O4/K54/H5), a clinical blood isolate, and its K54 capsule-minus isogenic derivatives have been previously described in part (23, 42).

Construction of capsule gene subclones. Subclones of the K54 capsule gene locus 5′ to the Tn*phoA* insertions in CP9.29, CP9.108, CP9.137, CP9.171, CP9.C43, CP9.C54, and CP9.C56 were obtained by restricting whole-cell DNA with *Bam*HI, which recognizes a site located 3′ to the kanamycin resistance gene in Tn*phoA* with or without *Xba*I (CP9.108, 137, 171, C43, C54, and C56) or *Apa*I (CP9.29), neither of which restricts within Tn*phoA*. Ligations of these restrictions into pBSII SK−, electroporation into XL1 Blue (Stratagene, La Jolla, Calif.), and selection of ampicillin (100 μg/ml)- and kanamycin (40 μg/ml)-resistant transformants resulted in the identification of the subclones p29.1, p108.1, p137.1, p171.1, pC43.1, pC54.1, and pC56.1. To construct a second set of subclones, in which the active *phoA* fusion was in the opposite orientation, p108.1, p137.1, and p171.1 were restricted with *Bam*HI and *Xba*I, p29.1 was restricted with *Bam*HI and *Apa*I, and the inserts were purified by electroelution and ligated into pBSII KS−. This set of constructs has been designated p29.2, p108.2, p137.2, and p171.2. These plasmids are described in detail in Table 1.

Subclones of the K54 capsule gene locus 3′ to the Tn*phoA* insertions in CP9.108 and CP9.171 were obtained by restricting whole-cell DNA with *Cla*I, which recognizes a site 5′ to the kanamycin resistance gene in Tn*phoA*. Ligations of these restrictions into pBSII SK−, electroporation into XL1 Blue (Stratagene), and selection of ampicillin- and kanamycin-resistant transformants resulted in the identification of the subclones p108.3 and p171.3. Each contains the right 6.7 kb of Tn*phoA* and either 3.0 kb (p108.3) or 2.0 kb (p171.3) of chromosomal DNA 3′ to the respective Tn*phoA* insertions.

Identification of a cosmid clone containing capsule genes. Whole-cell DNA was purified from CP9 as described previously (42), and DNA fragments (30 to 50 kb) were ligated into the unique *Bam*HI site of the 8.8-kb cosmid cloning vector pWE15 (Clontech Laboratories, Palo Alto, Calif.). The ligation mix was packaged into lambda phage in vitro and transduced into *E. coli* NM554, and the resultant CP9-derived DNA library was amplified once. The amplified library was screened for clones containing capsule genes via colony filter hybridization as described previously (16). The probe used for detection was generated by digesting p171.1 with *Pvu*I/*Xba*I, purifying the 1.3-kb restriction product via electroelution, and subsequent radioactive labelling with [α-^{32}P]dCTP by random oligonucleotide priming. Approximately 1,000 colonies of NM554 containing the CP9 DNA library were screened, and a cosmid clone (cos9a) was detected. Cos9a was confirmed to contain capsule gene DNA via Southern analysis (42), with the p171.1 *Pvu*I/*Xba*I 1.3-kb fragment as the probe.

DNA sequencing, determination of Tn*phoA* insertion sites, and analysis of capsule genes. DNA sequence was determined by the dideoxy chain termination method of Sanger et al. (47) with the capsule gene subclones (p29.1, p108.1, p108.3, p137.1, p171.1, p171.3, and pC56.1) and cos9a as the DNA templates. DNA sequencing of the capsule gene subclones p29.1, p108.1, p137.1, p171.1, and pC56.1 was initially with a Tn*phoA′* fusion joint primer (5′ AATATCGCC CTGAGC 3′), which established the location for a given Tn*phoA* insertion. Sequencing of capsule gene subclones p108.3 and p171.3 was initially with the Tn*phoA* primer (5′ CATGTTAGGAGGTCACAT 3′). Subsequent DNA sequence was determined with primers derived from the deduced sequences of the capsule gene subclones or the cosmid cos9a. A consensus sequence was generated by assembling and editing the DNA sequence obtained from 76 overlapping but independent sequencing reactions with AssemblyLIGN 1.0.2 (Oxford Molecular Group, Beaverton, Oreg.). Both strands of the capsule gene sequence submitted in this report were sequenced. The organization of the assembled subclone sequences and that of the cosmid cos9a sequence were in agreement. Sequence analysis, comparisons, and CLUSTAL alignments were performed, in part with MacVector (version 6.0; Oxford Molecular Group). Comparisons were also performed via BLAST analysis of the nonredundant GenBank, EMBL, DDBJ, and PDB sequences. Percentages of similarity and identity were determined by the GAP program of the Wisconsin Sequence Analysis Packages (Genetics Computer Group, Madison, Wis.). The PROSITE database was used for motif searches (2). SignalP V1.1 was used for identification of signal sequences (28). A terminator sequence search was performed by the method of Brendel and Trifonov adapted for the Wisconsin Sequence Analysis Packages (TERMINATOR) (5).

Capsule loci based on *Xba*I DNA fragments. We previously reported, based on Southern analysis of pulsed-field gel electrophoretically separated DNA from CP9 and Tn*phoA*-generated isogenic capsule-minus derivatives, that the K54 capsule genes were located on at least three different *Xba*I DNA fragments (capsule loci [*cl*] 1, 2, and 3). These fragments were linked within the transducing range of bacteriophage T4 (100 to 150 kb) (42). *cl1* was estimated to be a 10.3-kb fragment, and strains CP9.29, CP9.108, CP9.137, CP9.171, and CP9.C56 had insertions within this locus. Sequence analysis from the present study confirmed this finding and indicated that *cl1* was approximately 6.0 kb (Fig. 1). The fragment identified as *cl3* and containing the Tn*phoA* insertion in CP9.C54 was proven by sequence analysis to also be *cl1*. CP9.C54 has subsequently been shown to contain a truncated form of Tn*phoA*, a finding which led to the incorrect interpretation that *cl1* and *cl3* were separate loci. *cl2* was an estimated 18.5 kb, and strain CP9.C43 had a Tn*phoA* insertion within this locus. Sequence

TABLE 1. *E. coli* strains and plasmids used in this study

Strain or plasmid	Genotype or other relevant characteristics	Source or reference
Strains		
CP9	O4/K54/H5 clinical blood isolate, serum resistant	42
CP9.29	*cl1.29*::Tn*phoA* active Tn*phoA* fusion, serum sensitive	42
CP9.108	$kps_{K54}M$::Tn*phoA* (formerly *cl1.108*:Tn*phoA*) active Tn*phoA* fusion, serum sensitive	42
CP9.137	$kps_{K54}D$::Tn*phoA* (formerly *cl1.137*:Tn*phoA*) active Tn*phoA* fusion, serum sensitive	42
CP9.171	$kps_{K54}E$::Tn*phoA* (formerly *cl1.171*:Tn*phoA*) active Tn*phoA* fusion, serum sensitive	42
CP9.C56	*cl1*::Tn*phoA* inactive Tn*phoA* fusion, serum sensitive	42
CP9.C54	*cl1*::Tn*phoA* inactive Tn*phoA* fusion, serum sensitive	42
CP9.C43	*cl2*::Tn*phoA* inactive Tn*phoA* fusion, serum sensitive	42
NM554	*recA13* F^- *araD139* Δ(*ara-leu*)*7696 galE15 galK16* Δ(*lac*)*X74 rpsL*(Str^r) *hsdR2* (r_K^- m_K^+) *mcrA mcrB1*	Lab strain
XL1 Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* (F^+ *proAB* $lacI^q$ *ZΔM15* Tn*10*)	Stratagene
Plasmids		
p29.1	14.0-kb *Bam*HI/*Apa*I fragment from CP9.29 containing the leftward 5.0 kb of Tn*phoA* (active fusion) and capsule genes cloned into pBS II SK (−); capsule gene transcription is in the opposite orientation to the pBS II SK *lac* promoter	This study
p29.2	The above 14.0-kb fragment cloned into pBS II KS (−); capsule gene transcription is in the same orientation as the pBS II SK *lac* promoter	This study
p108.1	8.5-kb *Bam*HI/*Xba*I fragment from CP9.108 containing the leftward 5.0 kb of Tn*phoA* (active fusion) and capsule genes cloned into pBS II SK (−); capsule gene transcription is in the same orientation as the pBS II SK *lac* promoter	This study
p108.2	The above 8.5-kb fragment cloned into pBS II KS (−); capsule gene transcription is in the opposite orientation to the pBS II SK *lac* promoter	This study
p108.3	9.7-kb *Cla*I fragment from CP9.108 containing the rightward 6.7 kb of Tn*phoA* (nonfusion) and capsule genes cloned into pBS II SK (−)	This study
p137.1	7.5-kb *Bam*HI/*Xba*I fragment from CP9.137 containing the leftward 5.0 kb of Tn*phoA* (active fusion) and capsule genes cloned into pBS II SK (−); capsule gene transcription is in the same orientation as the pBS II SK *lac* promoter	This study
p137.2	The above 7.5-kb fragment cloned into pBS II KS (−); capsule gene transcription is in the opposite orientation to the pBS II SK *lac* promoter	This study
p171.1	6.4-kb *Bam*HI fragment from CP9.171 containing the leftward 5.0 kb of Tn*phoA* (active fusion) and capsule genes cloned into pBS II SK (−); capsule gene transcription is in the same orientation as the pBS II SK *lac* promoter	This study
p171.2	The above 6.4-kb fragment cloned into pBS II SK (−); capsule gene transcription is in the opposite orientation to the pBS II SK *lac* promoter	This study
p171.3	8.7-kb *Cla*I fragment from CP9.171 containing the rightward 6.7 kb of Tn*phoA* (nonfusion) and capsule genes cloned into pBS II SK (−)	This study
pC56.1	5.25 kb *Bam*HI/*Xba*I fragment from CP9.C56 containing the leftward 5.0 kb of Tn*phoA* (inactive fusion) and capsule genes cloned into pBS II SK (−)	This study
pcos9a	An approximately 40-kb fragment that contains capsule genes cloned into the *Bam*HI site of the cosmid pWE15	This study

analysis has demonstrated that the Tn*phoA* insertion, responsible for the $K54^-$ phenotype in CP9.C43, is in a novel DNA sequence which has no identifiable homology with any known capsule gene. This sequence is not part of the K54 capsule gene cluster reported here (Fig. 1). This data, in conjunction with the information described below, suggests that the Tn*phoA* insertion in CP9.C43 and *cl2* are located 3′ to the end of $kps_{K54}E$ (bp 6132).

Southern analysis. Whole-cell DNA was prepared as described previously (42) and restricted with *Acc*I as suggested by the manufacturer (New England Biolabs, Beverly, Mass.). Southern hybridization was performed as described previously (42) with the following modifications. A Robbins Scientific model 1000 hybridization oven was used. Salmon sperm DNA (180 μl of 150 μg/ml stock) was added to 10 ml of prehybridization solution, which was removed and replaced with 10 ml of hybridization solution (5× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate]-1% sodium dodecyl sulfate). An oligonucleotide was made from the IS_{CP9}*110* sequence (bp 581 to 626), labelled with [γ-^{32}P]dATP with T4 polynucleotide kinase according to the manufacturer's instructions (Gibco BRL, Gaithersburg, Md.), and used as a probe. After hybridization at 65°C for 18 h, the blot was washed once at 65°C with 1× SSC-0.1% sodium dodecyl sulfate for 3 min, followed by five washes at 25°C with 6× SSC-1% Sarkosyl for 5 min.

Alkaline phosphatase assays. Alkaline phosphatase assays were performed as previously described except that a Beckman DU 640B spectrophotometer was used to record the hydrolysis rates of *p*-nitrophenyl phosphate (46). The baseline activity of CP9 is negligible and therefore was not accounted for in this calculation. PhoA activity from each of the measured constructs represents the mean of five independent evaluations.

Nucleotide sequence accession number. The accession no. of the nucleotide sequence shown in Fig. 2 is AF007777.

RESULTS AND DISCUSSION

Location, organization, and analysis of the K54 capsule gene locus. K54-deficient strains generated by Tn*phoA* insertion mutagenesis (42) were used to identify clones carrying the group III capsule genes (see Materials and Methods), which were in turn used to identify a cosmid (pcos9a) carrying the wild-type genes. Sequencing of the region and identification of the sites of Tn*phoA* insertion were carried out. The K54 group III capsule gene locus is depicted in Fig. 1. As expected, since group II or III capsule gene sequences have not been detected in *E. coli* K-12, a search of GenBank did not identify any DNA or protein homology of this capsule gene locus with the deposited *E. coli* K-12 sequence. Novel loci of unique DNA not present in laboratory strains of *E. coli* have been termed "pathogenicity islands," and this sequence likely represents a portion of such a locus. One of the two boundaries of this novel CP9 DNA sequence with *E. coli* sequence from the K-12 genome was established (Fig. 1). The boundary was contiguous with the third base (bp 6222, section 269, accession no. AE000379) of a 178-amino-acid open reading frame (ORF) (bp 6220 to 6756) of unknown identity from the complete *E. coli* K-12 genome. Interestingly, this novel CP9 sequence is 150

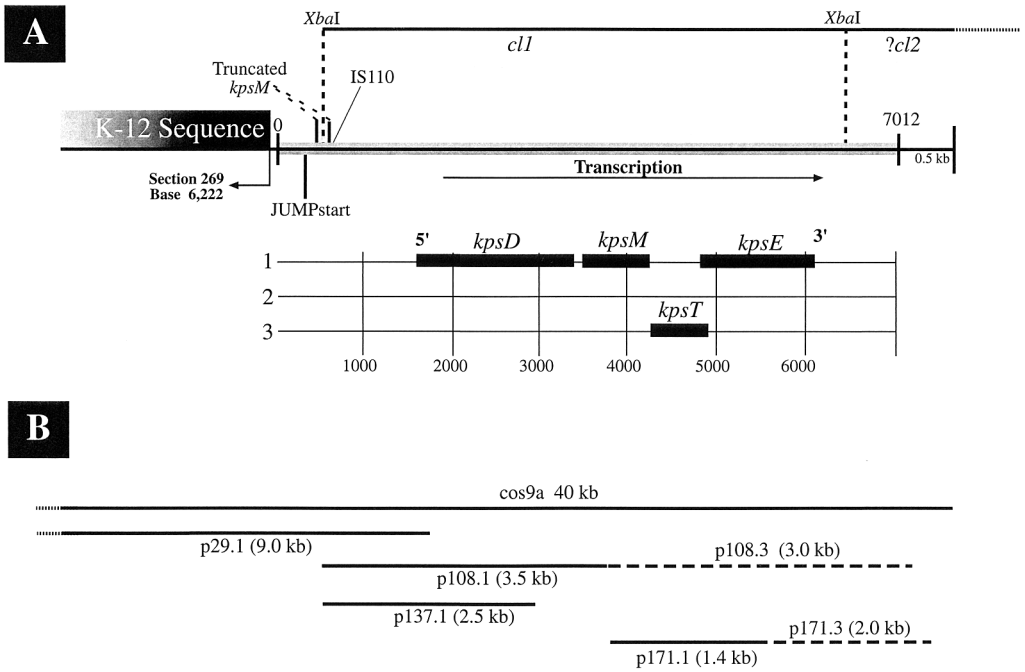

FIG. 1. (A) Schematic diagram of the K54 group III capsule gene sequence described in this study. From left to right are (i) the sequence homologous with the K-12 genome and its intersection (section 269, bp 6222) with a sequence unique to CP9 (90° arrow); (ii) bp 0 to 345, which are 85 to 90% homologous to the 5′ noncoding region of $kps_{K1,5}M$ (including the JUMPstart site as marked); (iii) a truncated *kpsM* (bp 346 to 476 and 501 to 526 are designated $kpsM_{truncated}$) that is 85 to 90% homologous to the corresponding region of $kps_{K1,5}M$; (iv) an IS*110* element (bp 581 to 626) 53 bp from the 3′ end of $kpsM_{truncated}$; and (v) the shaded region marked from 0 to 7012, representing the capsule gene sequence submitted in this report. The region from bp 627 to 1644 is unidentified but is probably capsule gene sequence. This region is followed by $kps_{K54}DMTE$, with their respective ORFs and reading frames depicted below. The 0.9 kb 3′ to $kps_{K54}E$ (bp 6133 to 7012) plus 0.5 kb is unidentified K54 capsule gene sequence. Prior to sequence analysis, the K54 *cl* were defined as *Xba*I fragments (44). The defined location of *cl1* (6.0 kb) and the presumed location of *cl2* are marked above. (B) The lines represent various inserts of subclones used for sequence analysis and promoter localization. Insert sizes are as marked. Length is proportional, and location corresponds to the schematic diagram above. The insert in p29 consists of the first 154 bp of $kps_{K54}D$ and an 8.9-kb region 5′ to the start of $kps_{K54}D$. The insert in p108 contains the first half of $kps_{K54}M$, all of $kps_{K54}D$, and 1.2 kb 5′ to $kps_{K54}D$ (bp 452 to 3878). The insert in p137 consists of two-thirds of $kps_{K54}D$ and the 1.2 kb 5′ to it (bp 452 to 2801). The insert in p171 (bp 3894 to 5288) covers the first half of $kps_{K54}E$, all of $kps_{K54}T$, and the last half of $kps_{K54}M$. The dotted lines at the leftward boundaries of cos9a and p29.1 represent extension into K-12 homologous sequence beyond what is depicted above.

bp 3′ to the phenylalanine tRNA (bp 5996 to 6071). The points of insertion of several pathogenicity islands are within various tRNAs (15). This location was consistent with the genetic linkage of the K54 capsule genes with *serA* (32), and its point of insertion is identical with that of the K5 capsule gene cluster.

Four homologs of group II capsule transport genes were recognized in the K54 capsule gene locus described in this study (Fig. 1). $kps_{K54}D$ (bp 1645 to 3387, Fig. 2), $kps_{K54}M$ (bp 3457 to 4254, Fig. 2), $kps_{K54}T$ (bp 4269 to 4916, Fig. 2), and $kps_{K54}E$ (bp 4888 to 6132, Fig. 2) were identified. However, while in group II capsule gene loci, *kpsD* and *kpsE* are in region 1 along with four other genes (*kpsFEDUCS*), and *kpsMT* are in region 3, in the CP9 (K54) capsule gene locus, these four genes are grouped together (*kpsDMTE*). These findings demonstrate that the organization of the K54 capsule transport genes in CP9 is unequivocally different from that of the corresponding regions in strains with group II capsule genes (39). Further, this data confirms the prediction, from complementation studies, that functional homologs of $kps_{K5}D$ and $kps_{K5}E$ existed in the K10 and K54 capsule gene clusters (36).

The putative molecular weights, estimated pIs, guanosine-plus-cytosine content, and presence or absence of an identifiable Shine-Dalgarno or signal sequence of $kps_{K54}DMTE$ and their comparison with $kps_{K1,5}DMTE$ are summarized in Table 2 (9, 34, 35, 49, 54). The guanosine-plus-cytosine content of $kps_{K54}DMTE$ ranged from 37 to 43%, compared to the 51% observed for *E. coli* K-12, and suggested that these genes were acquired by horizontal transfer from an unknown species. A nucleic acid subsequence analysis program and a manual search failed to identify any highly conserved Shine-Dalgarno sequences. The reason for this is unknown. However, since these genes are in essence "foreign DNA," perhaps their mRNAs possess sequence elements with complementarity to parts of the 16S rRNA that are distinct from those recognized by Shine-Dalgarno sequences, which in turn serve as translational enhancers.

The ORF (ORF1) from bp 1645 to 3387 encoded $kps_{K54}D$. It is 50% homologous at the nucleotide level and has 33% identity and 54% similarity at the predicted protein level with the group II genes $kps_{K1,5}D$. The limited homology with GumB (16% identity, 32% similarity) and OtnA (17% identity, 32% similarity) may represent common functional regions involved with transport (4). The identification of a putative signal sequence (Fig. 3), the presence of active Tn*phoA* fusions within $kps_{K54}D$, a hydrophilic hydropathy profile, and secondary-structure predictions similar to those of $Kps_{K1,5}D$ (data not shown) suggest that KpsD has a periplasmic location and a function similar to that of its group II counterpart, despite sequence divergence.

ORF2 (bp 3457 to 4254) encoded $kps_{K54}M$. No significant homology was detected at the nucleotide level; however, it has 39% identity and 52% similarity at the predicted protein level with the group II genes $kps_{K1,5}M$. Other homologs in *Actinobacillus pleuropneumoniae* (CpxB), *N. meningitidis* (CtrC), *H. influenzae* (BexB), and *Salmonella typhi* (VexB) revealed amino

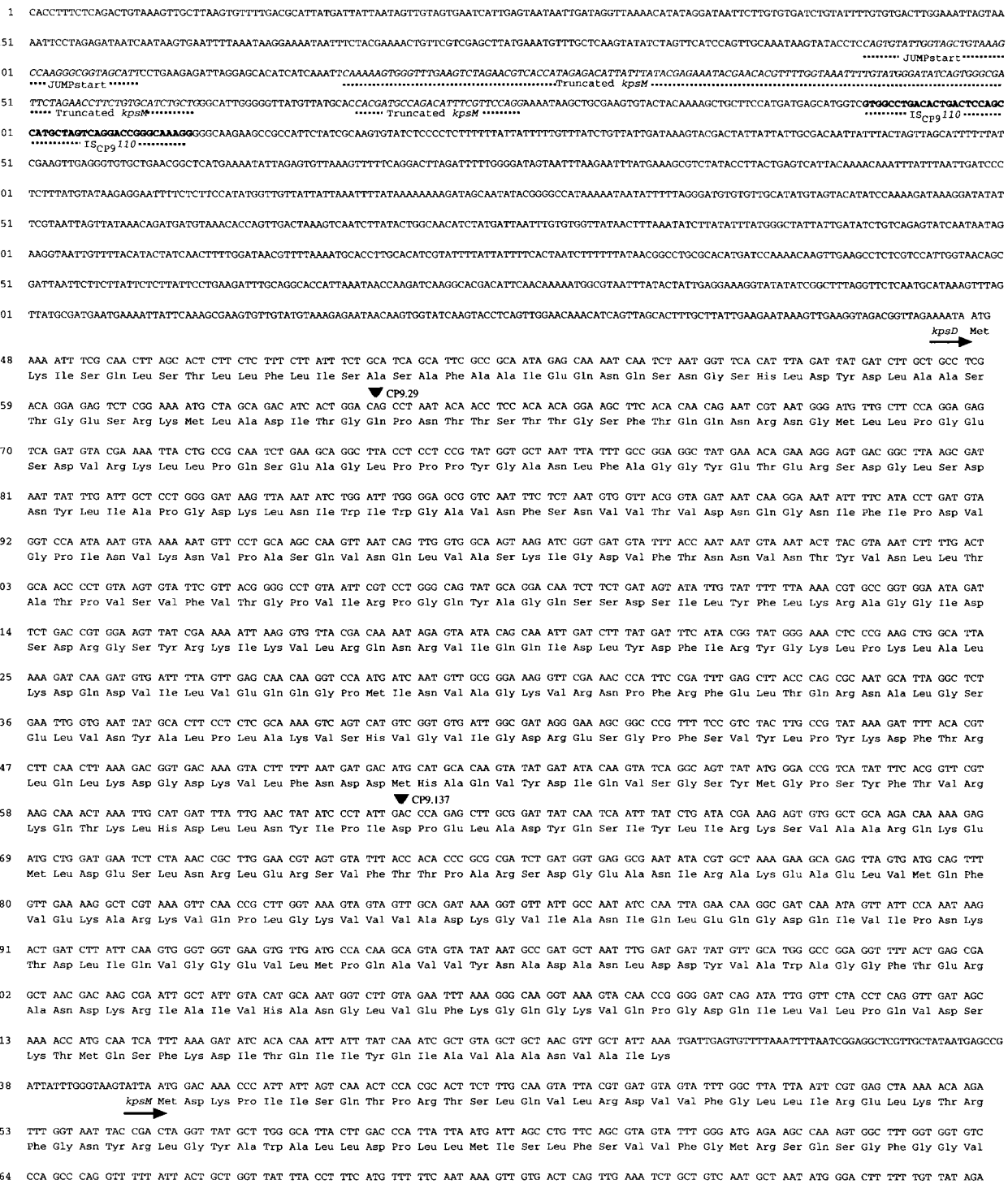

FIG. 2. Nucleotide sequence and deduced amino acid sequence of $kps_{K54}D$, $kps_{K54}M$, $kps_{K54}T$, and $kps_{K54}E$. Arrows identify putative transcriptional start sites, solid triangles identify the insertion site of active Tn*phoA* fusions, the open triangle identifies the insertion site of an inactive Tn*phoA* fusion, and the underlined regions identify the inverted repeats of a strong theoretical rho-independent RNA polymerase terminator. The JUMPstart site, the truncated *kpsM* ($kpsM_{truncated}$), and IS_{CP9}*110* are marked and identified by the dotted lines.

acid identities from 23 to 24% and similarities from 37 to 34%. All of these homologs have been implicated as the integral membrane component of the ABC-2 transporters of capsular polysaccharide across the cytoplasmic membrane (1, 18, 37). In $kps_{K54}M$, the identification of the ABC-2 transporter system integral membrane protein signature (Fig. 4), a similar hydropathy profile (hydrophobic protein with six transmembrane regions), and secondary-structure predictions similar to those of

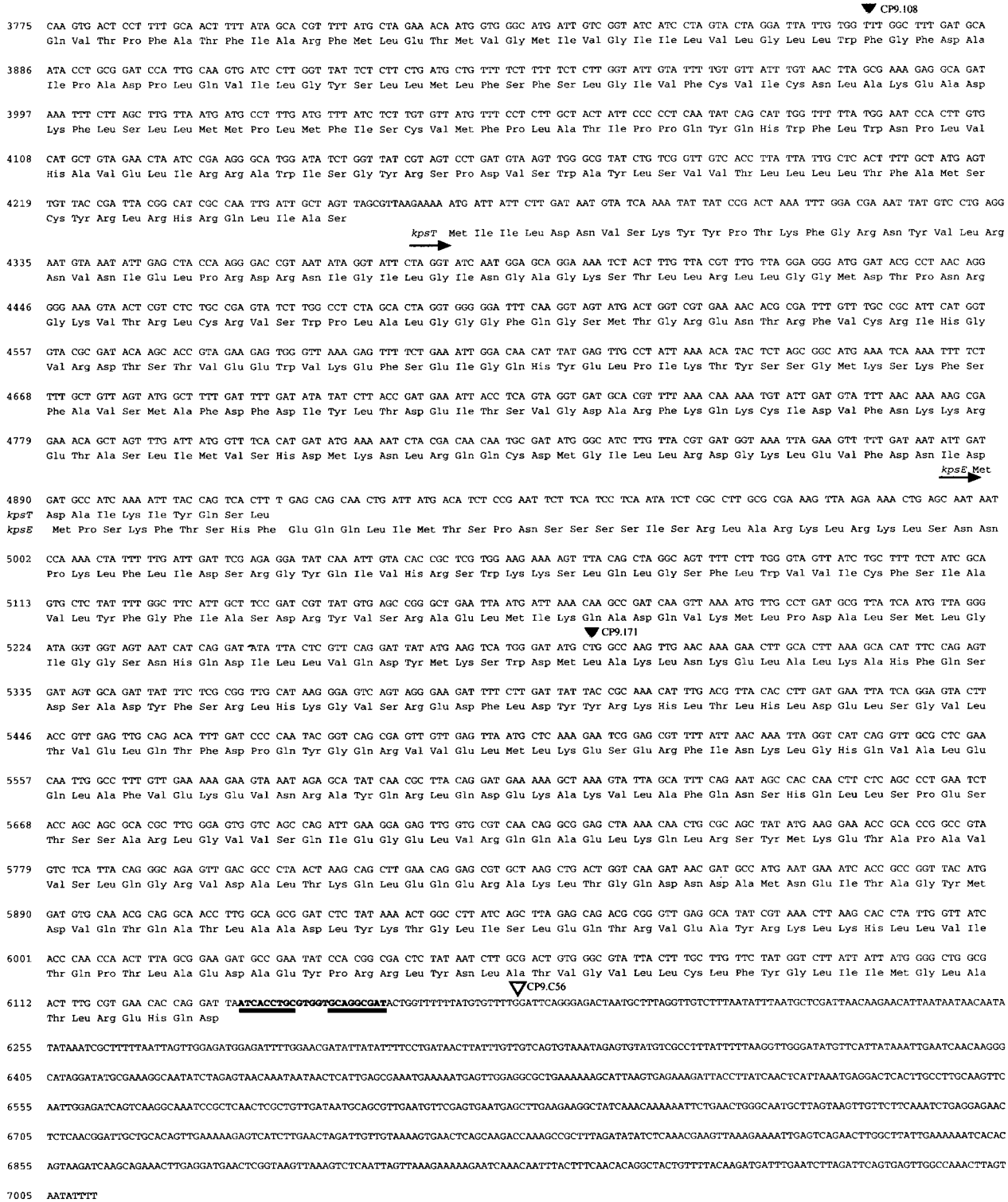

FIG. 2—*Continued.*

$Kps_{K1,5}M$ (data not shown) support the notion that $Kps_{K54}M$ is also a member of this family.

ORF3 (bp 4269 to 4916) encoded $kps_{K54}T$. It is 65 to 62% homologous at the nucleotide level and has 51 to 45% identity and 66 to 62% similarity at the predicted protein level with the group II genes $kps_{K1,5}T$. Other homologs in *A. pleuropneumoniae* (CpxA), *N. meningitidis* (CtrD), *H. influenzae* (BexA), and *S. typhi* (VexC) showed amino acid identities from 47 to 24% and similarities of 59 to 36%. These proteins are the peripheral ATP-binding components of the ABC-2 transporter protein

TABLE 2. Comparisons of putative molecular weight, pI, and GC content and the presence or absence of Shine-Dalgarno and signal sequences between *kps*$_{K54}$*DMTE* and *kps*$_{K1,5}$*DMTE*

Gene and region	Mol wt (in thousands)	Estimated pI	% GC	Presence or absence of sequence[a]	
				Shine-Dalgarno	Signal
kps$_{K54}$*D*	63.9	6.29	40	−	+
kps$_{K1}$*D*	60	7.33	51	+	+
kps$_{K5}$*D*	60	8.37	52	+	+
kps$_{K54}$*M*	30.2	10.04	38	−	−
kps$_{K1}$*M*	29.6	9.22	43	+	−
kps$_{K5}$*M*	29.5	9.06	45	+	−
kps$_{K54}$*T*	24.5	9.36	37	−	−
kps$_{K1}$*T*	24.9	9.29	35	+	−
kps$_{K5}$*T*	25.5	8.73	39	+	−
kps$_{K54}$*E*	47.3	8.64	43	−	−
kps$_{K1}$*E*	39.0	5.0	50	+	−
kps$_{K5}$*E*	43.0	6.02	50	+	−

[a] +, present; −, absent. The presence or absence of Shine-Dalgarno and signal sequences was determined by MacVector version 6.0 and PSORT, respectively.

family. The identification of Walker motifs (Fig. 5), an ABC transporter signature sequence, a similar hydrophilic hydropathy profile, and secondary-structure predictions similar to those of Kps$_{K1,5}$T support the notion that Kps$_{K54}$T is also a member of the ABC-2 transporter family, and its structure seems conserved in comparison with that of Kps$_{K1,5}$T.

ORF4 (bp 4888 to 6132) encoded *kps*$_{K54}$*E*. No significant homology was detected at the nucleotide level; however, it has 31% identity and 46% similarity at the predicted protein level with the group II genes *kps*$_{K1,5}$*E*. Other homologs in *A. pleuropneumoniae* (CpxC), *N. meningitidis* (CtrB), *H. influenzae* (BexC), and *S. typhi* (VexD) showed amino acid identities from 27 to 20% and similarities from 40 to 32%. Analysis of the putative Kps$_{K54}$E protein via hydropathy profiles and secondary-structure predictions suggests that this protein is similar to the membrane fusion protein family (4, 11). These proteins are believed to interact with ABC-type transport proteins (and others) and perhaps outer membrane proteins to facilitate substrate transport of large molecules. Kps$_{K54}$E has a number of features of this family, including (i) a hydrophilic amino terminus located in the cytoplasm (amino acids 1 to 60 of Kps$_{K54}$E with an excess of basic over acidic residues of a net +11), (ii) a hydrophobic amino terminus region that may both span and anchor the protein in the cytoplasmic membrane, (iii) a hydrophilic, largely alpha-helical periplasmic region (supported by the presence of active Tn*phoA* fusions at amino acid residue 136), and (iv) a hydrophobic carboxy terminus. However, the Kps$_{K54}$E hydrophobic carboxy terminus is significantly smaller than in the membrane fusion protein family, and the conservation of residues in this region is absent. Kps$_{K1,5}$E possess a similar predicted structure (4) except that Kps$_{K1}$E has a deletion of amino acid residues 1 to 71 and Kps$_{K5}$E has a deletion of residues 37 to 71 (Fig. 6).

Based on DNA and protein homologies, the sequence from bp 1645 to 6132 clearly comprised genes involved in K54 transport. Located 5′ to the end of this cluster is 1.0 kb (bp 626 to 1644) of sequence which has no identifiable DNA or protein homology. No Tn*phoA* insertions have been mapped to this region. A 369-bp ORF was identified from bp 1247 to 1615. Whether this ORF or another sequence in this 1.0-kb region codes for products involved with capsule transport or synthesis is unclear.

However, 5′ to this region, a sequence homologous to the initial 20% of *kpsM* and its entire 5′ noncoding regulatory region (0.35 kb) was identified. The sequence homologous to the 5′ coding region of *kpsM* (bp 346 to 476 and 501 to 526) has been designated *kpsM*$_{truncated}$ (Fig. 1 and 2). Over the entire *kpsM*$_{truncated}$ sequence and its 5′ noncoding regulatory region, an 85 to 90% DNA sequence homology to its K1 and K5 counterparts was observed. Previous investigators, using Southern analysis, did not identify any regions of homology with a variety of strains that contained group III capsule gene clusters (including a different K54 serotype) when probes containing *kps*$_{K5}$*M* were used (12, 36). The reasons for this discrepancy are unclear. However, the degree of homology of *kpsM*$_{truncated}$ with *kps*$_{K1,5}$*M* was notable, since the complete *kps*$_{K54}$*M* possessed no significant homology at the nucleotide level to *kps*$_{K1,5}$*M*. This finding suggested the possibility that CP9 originally possessed a group II capsule gene locus and that the K54 capsule genes were subsequently acquired by horizontal transfer. Therefore, the DNA sequence immediately 3′ to *kpsM*$_{truncated}$ was analyzed for potential insight into the evolution of this group III capsule gene locus. A 51-bp region (bp 581 to 626) identified 53 bp 3′ to the end of *kpsM*$_{truncated}$ (Fig. 1 and 2) was 75% homologous with the IS*110* insertion sequence identified from *Streptomyces coelicolor*. The area of homology coincided with a 39-bp inverted repeat that has been identified within this element (8) and the 12 bp 5′ to the repeat (bp 1073 to 1122, accession no. Y00434). Runs of cytosines appear to be the target site for this element, and the 11-bp region 5′ to this element contained two runs of cytosines (5′ CCCGTTTCCCCC 3′). These findings were consistent with the hypothesis that CP9 previously possessed group II capsule genes and acquired group III capsule genes via IS*110*-mediated horizontal transfer. A second prediction of this model would be the presence of a second IS$_{CP9}$*110* element at the 3′ end of the transferred segment. Southern analysis of *Acc*I-restricted chromosomal DNA isolated from CP9 and the closely related strain J96 (23) was performed under high-stringency conditions, with the IS$_{CP9}$*110* element used as an oligonucleotide probe. Two copies of the IS$_{CP9}$*110* element were detected in the chromosomes of CP9 and J96 on identically sized restriction fragments (Fig. 7). One of the fragments was the size predicted from the known sequence of the K54 capsule gene cluster containing IS$_{CP9}$*110* (1.4 kb). Insertion elements are well-known mediators of the evolution of genomic organization, and in other studies researchers have shown their role in the horizontal transfer of genes involved in polysaccharide synthesis (3, 55). The origin of the group III genes, however, remains speculative.

The 1.3 kb downstream from bp 6132 (bp 6133 to 7012 plus 0.5 kb 3′ to the submitted sequence) appeared, at least in part, to code for genes involved in capsule synthesis or transport. Although no homology to any known capsule genes was identified in this region, the transposon insertion in the K54 capsule-deficient strain CP9.C56 was located at bp 6179, which is 3′ to *kps*$_{K54}$*E*. Further evidence for the presence of these additional genes is suggested by the isolation of the capsule-negative strain CP9.C43 (42), which contains a Tn*phoA* insertion in an 18.5-kb *Xba*I fragment, *cl2* (Fig. 1). *cl2* and the *cl1* fragment sequenced here are cotransducible by phage T4, suggesting that the gene defined by CP9.C43 insertion and *cl2* are located 3′ to the end of *kps*$_{K54}$*E*.

Transcriptional organization of the K54 capsule gene locus. CP9.29, CP9.108, CP9.137, and CP9.171 possessed active *phoA* fusions with genes in the K54 capsule gene locus. The trans-

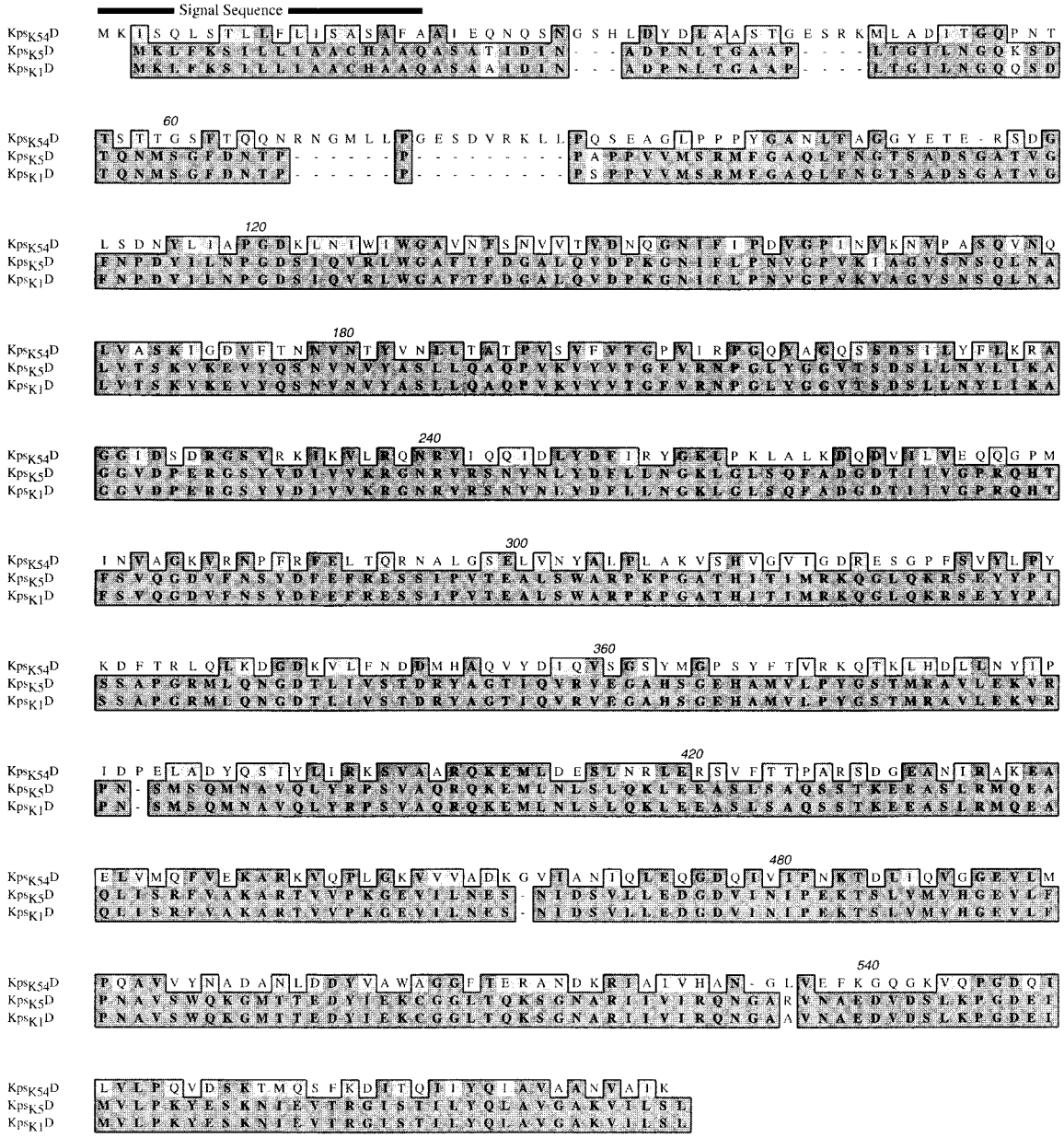

FIG. 3. CLUSTAL alignment of the predicted amino acid sequences of *E. coli kps*$_{K54}$*D* (this study), *kps*$_{K1}$*D*, and *kps*$_{K5}$*D*. The boxed sequence identifies amino acid residues that are functionally similar (lighter shading) or identical (darker shading). Numbers above the sequences are residue numbers. The predicted signal sequence of Kps$_{K54}$D is identified.

poson insertion in CP9.29 (bp 1799) and that in CP9.137 (bp 2801) were within *kps*$_{K54}$*D*, the transposon insertion in CP9.108 (bp 3878) was within *kps*$_{K54}$*M*, and the transposon insertion in CP9.171 (bp 5288) was within *kps*$_{K54}$*E* (Fig. 1 and 2). These active fusions confirm the direction of transcription. *kps*$_{K54}$*D* and *kps*$_{K54}$*M* are separated by 69 bp, *kps*$_{K54}$*M* and *kps*$_{K54}$*T* are separated by 14 bp, and *kps*$_{K54}$*T* and *kps*$_{K54}$*E* overlap by 29 bp. No predicted promoter regions were identified. Bp 6135 to 6157, located just 3 bp 3′ to *kps*$_{K54}$*E*, were consistent with a strong theoretical rho-independent RNA polymerase terminator. These base pairs formed a hairpin structure which was followed by a 6-bp poly(T) sequence (bp 6163 to 6169). The free energy of the stem-loop (ΔG [25°C]) was calculated to be −10.2 kcal (52). The K1 K5 capsule gene cluster regions 1 and 3 are organized as a single transcription unit. The organization of these regions for the K54 genes is not yet known.

Alkaline phosphatase assays on the subclones containing active Tn*phoA* insertions in either of two orientations were performed. In constructs in which the capsule gene insert was in the opposite orientation to the vector *lacZ* promoter (p29.1, p108.2, p137.2, and p171.2), PhoA activity likely reflects transcription from an insert promoter; however, the possibility of read-through transcription from a cryptic promoter cannot be excluded. In constructs with the insert in the same orientation as the *lacZ* promoter, PhoA activity reflects transcription from the insert and potentially the vector promoter. The similar degrees of PhoA activity seen with p29.1 (15.2 ± 1.6) (mean ± standard error) and p29.2 (17.7 ± 2.1) suggested that this

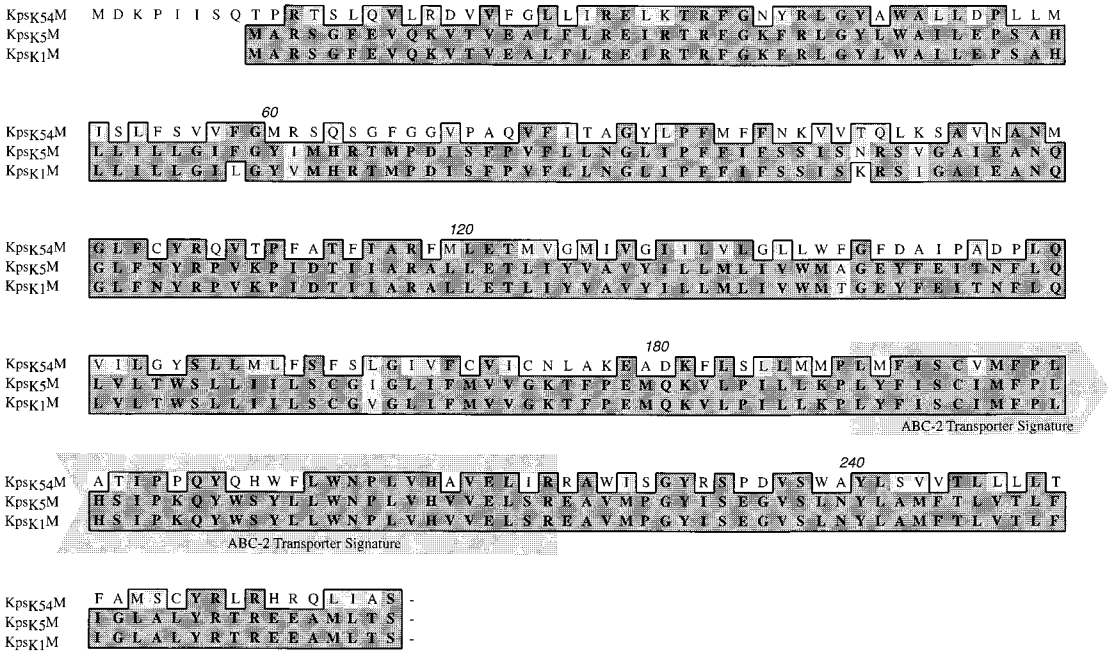

FIG. 4. CLUSTAL alignment of the predicted amino acid sequences of *E. coli* $kps_{K54}M$ (this study), $kps_{K5}M$, and $kps_{K1}M$. The boxed sequence identifies amino acid residues that are functionally similar (lighter shading) or identical (darker shading). The ABC-2 transporter system integral membrane protein signature is marked and corresponds to amino acid residues 190 to 224 (1, 18). Numbers above the sequences are residue numbers.

insert, not the vector, contained a promoter that is responsible for its *phoA* gene fusion activity. The inserts in p108 and p137 do not appear to contain promoters, based on the low level of PhoA activity produced by p108.2 (0.59 ± 0.2) and p137.2 (0.61 ± 0.1). p171 may possess a region with promoter activity, since p171.2 produced 1.93 ± 0.1 U of PhoA activity. The extent of these inserts (Fig. 1) and their activity or lack of activity suggest that an essential promoter element is upstream of bp 452, which is located within $kpsM_{truncated}$. Although the insert in p29 contains about 8 kb of DNA upstream of bp 452, it seems likely that the promoter for these genes lies within the 741 bp 5′ to the initiation site for $kpsM_{truncated}$ (bp 345), since the K5 region 3 promoter has been mapped to this location (50). No stem-loop terminators were identified between this region and the start of $kps_{K54}D$ (bp 1645). Also present in this region is the JUMPstart sequence (bp 279 to 317), which has been implicated in the regulation of a variety of polysaccharide genes and genes encoding secreted products (19, 29). This sequence is present 5′ to region 3 in both the K1 and K5 capsule gene sequences, in the promoter-operator region of

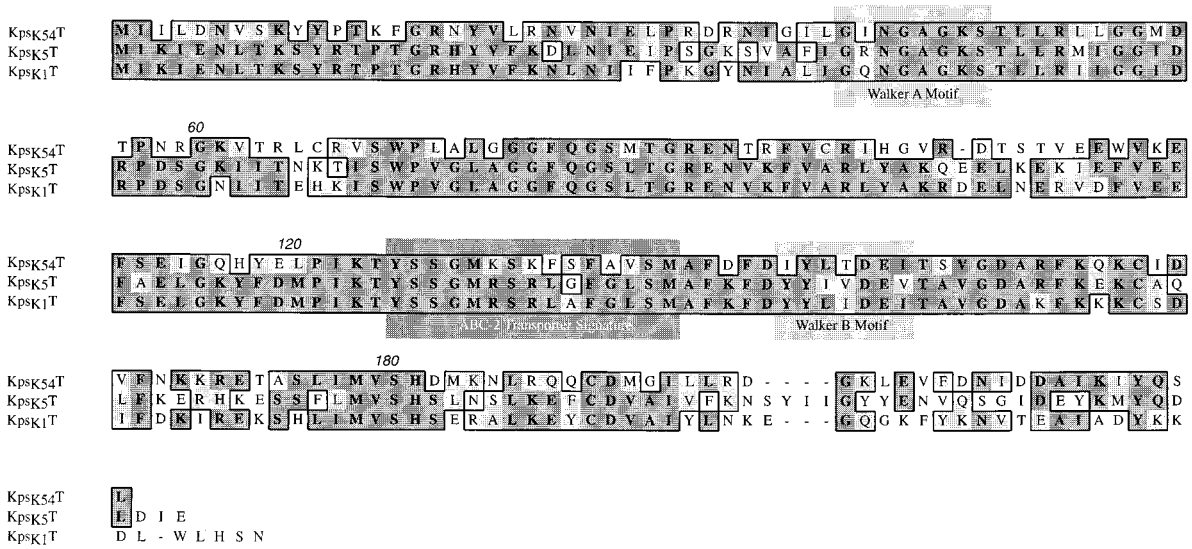

FIG. 5. CLUSTAL alignment of the predicted amino acid sequences of *E. coli* $kps_{K54}T$ (this study), $kps_{K5}T$, and $kps_{K1}T$. The boxed sequence identifies amino acid residues that are functionally similar (lighter shading) or identical (darker shading). The ATP-binding domain Walker A (residues 38 to 45), and Walker B (residues 145 to 151) motifs (34) and the ABC-2 transporter signature sequence (residues 125 to 139) are marked (1, 18). Numbers above the sequences are residue numbers.

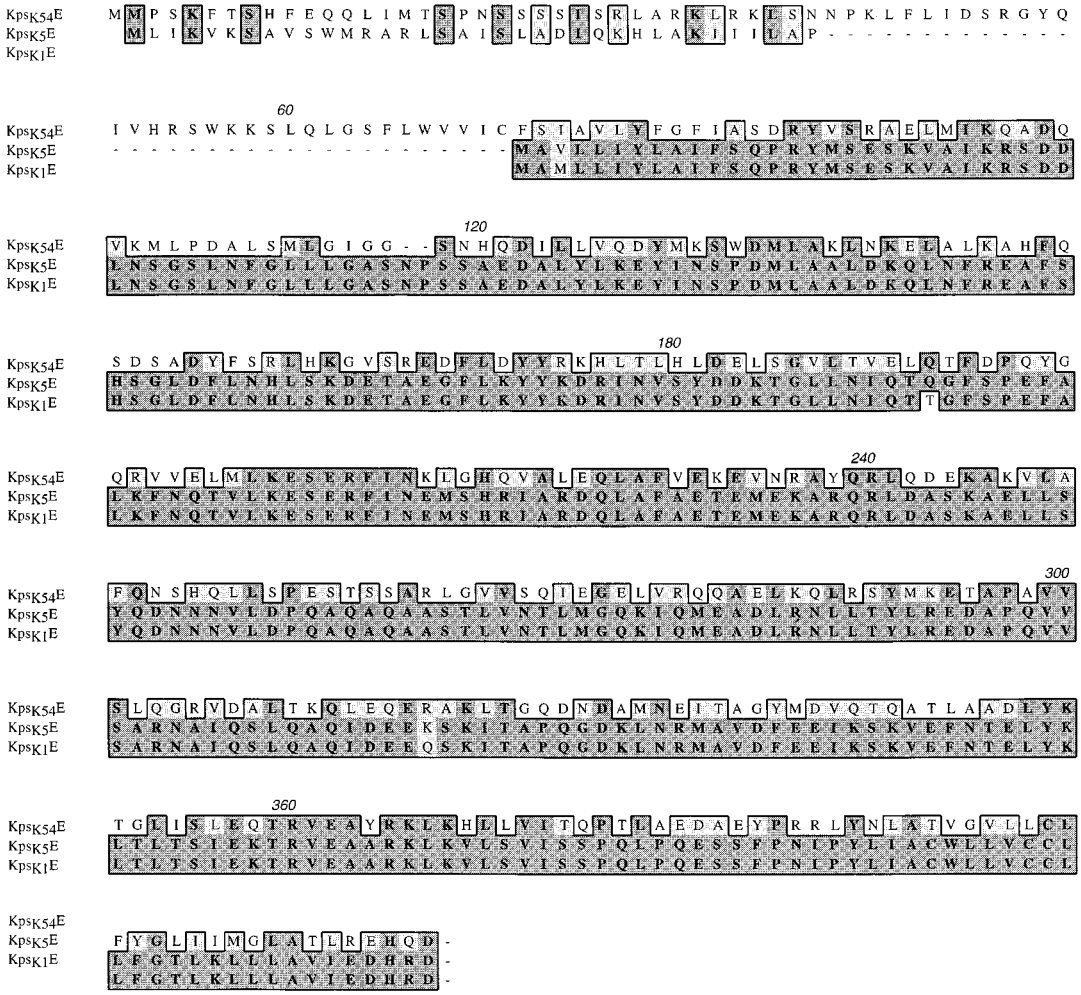

FIG. 6. CLUSTAL alignment of the predicted amino acid sequences of *E. coli kps*$_{K54}$*E* (this study), *kps*$_{K1}$*E*, and *kps*$_{K5}$*E*. The boxed sequence identifies amino acid residues that are functionally similar (lighter shading) or identical (darker shading). Numbers above the sequences are residue numbers.

the *cps* genes coding for the group I colanic acid capsule (51), and in the 5′ noncoding region of several lipopolysaccharide gene clusters (19) and the *hly* and *tra* genes (29). It has been shown that this site is required for the up-regulation of *hly* genes and *kps*$_{K5}$ region 2 genes by RfaH via antitermination (29, 50). The effect of RfaH on K54 capsule gene activity has not yet been evaluated. However, in previous studies we have demonstrated that RcsA is a negative regulator of group III capsule genes, which we have now identified as *kps*$_{K54}$*DME* (46). Although the precise promoter-operator region for these group III capsule genes has not been definitively established, it is intriguing that the JUMPstart site is present within both the insert in p29 and the promoter-operator region of the *cps* genes encoding for the group I colanic acid capsule. Further, RcsA is a positive regulator of group I capsule production and interacts with the promoter-operator region of the *cps* genes (51). Since RcsA divergently regulates the group 1 colanic acid and the K54 capsular polysaccharides, it is tempting to speculate that this regulation is mediated by RcsA either via direct interaction with the JUMPstart site or by interaction with protein complexes assembled there.

In summary, the genomic location, the novel organization of a portion of the group III, K54 capsule gene cluster, and the DNA sequences of *kps*$_{K54}$*DMTE* have been determined. De-

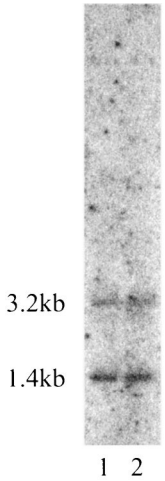

FIG. 7. Southern analysis of DNA from CP9 and J96 to detect the copy number of IS$_{CP9}$*110* element. *Acc*I-restricted whole-cell DNA from CP9 and J96 was separated by conventional electrophoresis, blotted onto nylon, and subjected to Southern analysis under high-stringency conditions as described in Materials and Methods, with the IS$_{CP9}$*110* element as the probe. Lane 1, J96; lane 2, CP9.

spite a divergence of these genes at the nucleotide and protein levels, analysis suggests that their function is similar to that of described homologs. Further, this study lends insight into the phylogenetic evolution of a group III capsule gene cluster. Findings support the hypothesis that CP9 previously possessed group II capsule genes and acquired group III capsule genes via IS*110*-mediated horizontal transfer. Future studies will be focused on the remaining genes in the K54 capsule gene cluster, in particular those that do not possess homology to reported capsule genes. The products of these genes may function differently from those responsible for group II capsule expression.

ACKNOWLEDGMENTS

This work was supported by Research for Health in Erie County (T.A.R.) and the Office of Research and Development, Department of Veterans Affairs (A.J.L.).

We appreciate the continued support of Tim Murphy.

REFERENCES

1. **Ames, G. F.-L., C. S. Mimura, S. R. Holbrook, and V. Shyamala.** 1992. Traffic ATPases: a superfamily of transport proteins operating from *Escherichia coli* to humans. Adv. Enzymol. **65:**1–47.
2. **Bairoch, A., P. Bucher, and K. Hofmann.** 1995. The PROSITE database, its status in 1995. Nucleic Acids Res. **24:**189–196.
3. **Bik, E. M., A. E. Bunschoten, R. D. Gouw, and F. Mooi.** 1995. Genesis of the novel epidemic *Vibrio cholerae* O139 strain: evidence for horizontal transfer of genes involved in polysaccharide synthesis. EMBO J. **14:**209–216.
4. **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.
5. **Brendel, V., and E. N. Trifonov.** 1984. A computer algorithm for testing potential prokaryotic terminators. Nucleic Acids Res. **12:**4411–4427.
6. **Bronner, D., V. Sieberth, C. Pazzani, I. S. Roberts, G. J. Boulnois, B. Jann, and F. Jann.** 1993. Expression of the capsular K5 polysaccharide of *Escherichia coli*: biochemical and electron microscopic analyses of mutants with defects in region 1 of the K5 gene cluster. J. Bacteriol. **175:**5984–5992.
7. **Bronner, D., V. Sieberth, C. Pazzani, A. N. Smith, G. J. Boulnois, I. S. Roberts, B. Jann, and F. Jann.** 1993. Synthesis of the K5 (group II) capsular polysaccharide in transport-deficient recombinant *Escherichia coli*. FEMS Microbiol. Lett. **113:**279–284.
8. **Bruton, C. J., and K. F. Chater.** 1987. Nucleotide sequence of *IS110*, an insertion sequence of *Streptomyces coelicolor* A3(2). Nucleic Acids Res. **15:** 7053–7065.
9. **Cieslewicz, M. J., S. M. Steenbergen, and E. R. Vimr.** 1993. Cloning, sequencing, expression, and complementation analysis of the *Escherichia coli* K1 *kps* region 1 gene, *kpsE*, and identification of an upstream open reading frame encoding a protein with homology to GutQ. J. Bacteriol. **175:**8018–8023.
10. **Cross, A. S.** 1990. The biologic significance of bacterial encapsulation. Curr. Top. Microbiol. Immunol. **150:**87–95.
11. **Dinh, T., I. T. Paulsen, and M. H. Saier, Jr.** 1994. A family of extracytoplasmic proteins that allow transport of large molecules across the outer membranes of gram-negative bacteria. J. Bacteriol. **176:**3825–3831.
12. **Drake, C. R., G. J. Boulnois, and I. S. Roberts.** 1993. The *Escherichia coli ser-A*-linked capsule locus and its flanking sequences are polymorphic, genetic evidence of more than two groups of capsule gene clusters. J. Gen. Microbiol. **139:**1707–1714.
13. **Drake, C. R., I. S. Roberts, B. Jann, K. Jann, and G. J. Boulnois.** 1990. Molecular cloning and expression of the genes encoding the *Escherichia coli* K4 capsular polysaccharide, a fructose-substituted chondroitin. FEMS Microbiol. Lett. **66:**227–230.
14. **Finke, A., B. Jann, and K. Jann.** 1990. CMP-KDO synthetase activity in *Escherichia coli* expressing capsular polysaccharide. FEMS Microbiol. Lett. **69:**129–134.
15. **Finlay, B. B., and S. Falkow.** 1997. Common themes in microbial pathogenicity revisited. Microbiol. Mol. Biol. Rev. **61:**136–169.
16. **Godoy, V. G., M. M. Dallas, T. A. Russo, and M. H. Malamy.** 1993. A role for *Bacteroides fragilis* neuraminidase in bacterial growth in two model systems. Infect. Immun. **61:**4415–4426.
17. **Hadad, J. J., and C. L. Gyles.** 1982. The role of K antigens of enteropathogenic *Escherichia coli* in colonization of the small intestine of calves. Can. J. Comp. Med. **46:**21–26.
18. **Higgins, C. F.** 1992. ABC transporters: from microorganisms to man. Annu. Rev. Cell Biol. **8:**67–113.
19. **Hobbs, M., and P. Reeves.** 1994. The JUMPstart sequence: a 39 bp element common to several polysaccharide gene clusters. Mol. Microbiol. **12:**855–856.
20. **Jann, B., and K. Jann.** 1990. Structure and biosynthesis of the capsular antigens of *Escherichia coli*. Curr. Top. Microbiol. Immunol. **150:**19–42.
21. **Jann, K., and B. Jann.** 1987. Polysaccharide antigens of *Escherichia coli*. Rev. Infect. Dis. **9:**S517–S526.
22. **Jann, K., and B. Jann.** 1992. Capsules of *Escherichia coli*, expression and biologic significance. Can. J. Microbiol. **38:**705–710.
23. **Johnson, J. R., et al.** 1997. Discovery of a disseminated J96-like clone of uropathogenic *Escherichia coli* O4:H5 containing both $papG_{J96}$ ("Class I") and $prsG_{J96}$ ("Class III") Gal(alpha1-4)Gal-binding adhesion sequences. J. Infect. Dis. **175:**983–988.
24. **Johnson, J. R., A. E. Stapleton, T. A. Russo, F. Scheutz, J. J. Brown, and J. N. Maslow.** 1997. Characteristics and prevalence within serogroup O4 of a J96-like clonal group of uropathogenic *Escherichia coli* O4:H5 containing the class I and class III alleles of *papG*. Infect. Immun. **65:**2153–2159.
25. **Keenleyside, W. J., P. Jayaratne, P. R. MacLachlan, and C. Whitfield.** 1992. The *rcsA* gene of *Escherichia coli* O9:K30:H12 is involved in the expression of the serotype-specific group I K (capsular) antigen. J. Bacteriol. **174:**8–16.
26. **Moxon, E. R., and J. S. Kroll.** 1990. The role of bacterial polysaccharide capsules as virulence factors. Curr. Top. Microbiol. Immunol. **150:**65–85.
27. **Nagy, B., H. W. Moon, and R. E. Isaacson.** 1976. Colonization of porcine small intestine by *Escherichia coli*: ileal colonization and adhesion by pig enteropathogens that lack K88 antigen and by some acapsular mutants. Infect. Immun. **13:**1214–1220.
28. **Nielsen, H., J. Engelbrecht, S. Brunak, and G. von Heijne.** 1997. Identification of procaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. **10:**1–6.
29. **Nieto, J. M., M. J. A. Bailey, C. Hughes, 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.
30. **Opir, T., and D. L. Gutnick.** 1994. A role for exopolysaccharides in the protection of microorganisms from desiccation. Appl. Environ. Microbiol. **60:**740–745.
31. **Orskov, F., I. Orskov, B. Jann, and K. Jann.** 1977. Serology, chemistry, and genetics of O and K antigens of *Escherichia coli*. Bacteriol. Rev. **41:**667–710.
32. **Orskov, 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.
33. **Orskov, I., V. Sharina, and F. Orskov.** 1976. Genetic mapping of the K1 and K4 antigens (L) of *Escherichia coli*. Acta Pathol. Microbiol. Scand. Sect. B **84:**125–131.
34. **Pavelka, M. S., 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.
35. **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.
36. **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.
37. **Pigeon, R. P., and R. P. Silver.** 1994. Topological and mutational analysis of KpsM, the hydrophobic component of the ABC-transporter involved in the export of polysialic acid in *Escherichia coli* K1. Mol. Microbiol. **14:**871–881.
38. **Roberts, I., R. Mountford, N. High, D. Bitter-Suermann, K. Jann, K. Timmis, and G. Boulnois.** 1986. Molecular cloning and analysis of genes for production of K5, K7, K12, and K92 capsular polysaccharides in *Escherichia coli*. J. Bacteriol. **168:**1228–1233.
39. **Roberts, I. S.** 1996. The biochemistry and genetics of capsular polysaccharide production in bacteria. Annu. Rev. Microbiol. **50:**285–315.
40. **Roberts, I. S., R. Mountford, R. Hodge, K. B. Jann, and G. J. Boulnois.** 1988. Common organization of gene clusters for production of different capsular polysaccharides (K antigens) in *Escherichia coli*. J. Bacteriol. **170:**1305–1310.
41. **Russo, T. A., J. J. Brown, S. T. Jodush, and J. R. Johnson.** 1996. The O4 specific antigen moiety of lipopolysaccharide but not the K54 group 2 capsule is important for urovirulence of an extraintestinal isolate of *Escherichia coli*. Infect. Immun. **64:**2343–2348.
42. **Russo, T. A., J. E. Guenther, S. Wenderoth, and M. M. Frank.** 1993. Generation of isogenic K54 capsule-deficient *Escherichia coli* strains through Tn*phoA*-mediated gene disruption. Mol. Microbiol. **9:**357–364.
43. **Russo, T. A., Y. Liang, and A. S. Cross.** 1994. The presence of K54 capsular polysaccharide increases the pathogenicity of *Escherichia coli in vivo*. J. Infect. Dis. **169:**112–118.
44. **Russo, T. A., M. C. Moffitt, C. H. Hammer, and M. M. Frank.** 1993. Tn*phoA*-mediated disruption of K54 capsular polysaccharide genes in *Escherichia coli* confers serum sensitivity. Infect. Immun. **61:**3578–3582.
45. **Russo, T. A., G. Sharma, J. Weiss, and C. Brown.** 1995. The construction and characterization of colanic acid deficient mutants in an extraintestinal isolate of *Escherichia coli* (O4/K54/H5). Microb. Pathog. **18:**269–278.
46. **Russo, T. A., and G. Singh.** 1993. An extraintestinal, pathogenic isolate of *Escherichia coli* (O4/K54/H5) can produce a group 1 capsule which is diver-

gently regulated from its constitutively produced group 2, K54 capsular polysaccharide. J. Bacteriol. **175:**7617–7623.
47. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
48. **Silver, R. P., C. W. Finn, W. F. Vann, W. Aaronson, R. Schneerson, P. J. Kretschmer, and C. F. Garon.** 1981. Molecular cloning of the K1 capsular polysaccharide genes of *E. coli*. Nature **289:**696–698.
49. **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.
50. **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.
51. **Stout, V.** 1996. Identification of the promoter region for the colanic acid polysaccharide biosynthetic genes in *Escherichia coli* K-12. J. Bacteriol. **178:** 4273–4280.
52. **Tinoco, I., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla.** 1973. Improved estimation of secondary structure in ribonucleic acids. Nature **246:**40–41.
53. **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.
54. **Wunder, D. E., W. Aaronson, S. F. Hayes, J. M. Bliss, and R. P. Silver.** 1994. Nucleotide sequence and mutational analysis of the gene encoding KpsD, a periplasmic protein involved in transport of polysialic acid in *Escherichia coli* K1. J. Bacteriol. **176:**4025–4033.
55. **Zhao, S., C. H. Sandt, G. Feulner, D. A. Vlazny, J. A. Gray, and C. W. Hill.** 1993. *Rhs* elements of *Escherichia coli* K-12: complex composites of shared and unique components that have different evolutionary histories. J. Bacteriol. **175:**2799–2808.