INFECTION AND IMMUNITY, Dec. 1995, p. 4849–4856 Vol. 63, No. 12
0019-9567/95/$04.00+0

Genes for CS2 Pili of Enterotoxigenic *Escherichia coli* and Their Interchangeability with Those for CS1 Pili

BARBARA J. FROEHLICH, ALEXANDER KARAKASHIAN, HARRY SAKELLARIS, AND JUNE R. SCOTT*

Department of Microbiology and Immunology, Emory University, Atlanta, Georgia 30322

Received 3 July 1995/Returned for modification 4 August 1995/Accepted 11 September 1995

We have cloned and sequenced the DNA needed for production of CS2 pili in *Escherichia coli* K-12. The four open reading frames, *cotB*, *cotA*, *cotC*, and *cotD*, show homology with the genes needed for production of CS1 and CFA/I pili, which are also found on enterotoxigenic *E. coli* associated with human diarrheal disease. We also report that CotA plus CotB interact with the CS1 gene products CooC and CooD to form pili that can be visualized by electron microscopy and, conversely, that the CS1 gene products CooA and CooB interact with CotC and CotD to form pili.

Diarrheal diseases are an important cause of death in infants and young children in developing countries. One of the major causative agents of bacterial diarrhea is enterotoxigenic *Escherichia coli* (ETEC). For this reason, attention has been focused on efforts to develop a vaccine to protect against disease caused by these organisms. Pili are prime candidates for vaccine development, since they are presumably involved in the earliest step in the infectious process, attachment of the bacterium to the host. Since ETEC strains often produce serologically different pili, a vaccine would have to include epitopes from more than one type of pilus. For this reason, it would be helpful to analyze different members of the family of pili from ETEC strains associated with human disease to learn about their similarities and differences.

Among human ETEC strains, several types of pili have been distinguished on the basis of antigenic specificity and/or of subunit molecular weight. The best studied of these are CFA/I (colonization factor antigen I), CS1 (coli surface antigen 1), and CS2 (8, 21, 40). These three pilin types seem to form a family on the basis of several types of data. First, although present in different strains of ETEC, they are all positively regulated by *rns* or a closely related gene (1, 3, 35). Second, the N-terminal amino acid sequence of the major pilin protein is almost identical for these three pilins and for CS4 and PCF0166 as well (43, 49). Third, for CFA/I and CS1, the DNA sequence of the four genes needed for their synthesis in *E. coli* shows extensive homology between the proteins (10, 12, 17, 18, 27).

Of the genes encoding pili from human ETEC strains, so far only the genes for CS1 have been studied for function. Morphogenesis of CS1 pili appears to be completely different from that of other kinds of pili found on gram-negative bacteria, such as *pap*-related pili (which are found on *E. coli* strains associated with pyelonephritis) and the common type 1 pili (6, 14, 32). To express CS1 pili in *E. coli* K-12, only four genes are needed (10). The major pilin subunit, encoded by *cooA*, lacks the features common to the better-studied pilins, including disulfide bonds and a penultimate tyrosine (25, 27). No typical chaperonins can be identified by sequence comparisons (10, 38). It appears, therefore, that the pili found specifically on human ETEC strains constitute a new family, and their relationships to each other may help us understand their morphogenesis.

The phenotype of mutants with mutations in each of the *coo* genes needed for production of CS1 pili is absence of pili. Mutants with mutations in *cooB*, *cooC*, or *cooD* make pilin but do not produce structures identifiable as pili (10, 38). Thus, for CS1 and, by analogy, for the other pili of human ETEC strains, there is no clear candidate for a gene that might encode a separate adhesin which is present only at the tip of the pilus structure.

We have previously determined that CS2 pili are encoded by chromosomal genes (27), unlike both CFA/I and CS1 pili, which are encoded on plasmids. However, like CS1, CFA/I, and CS4, CS2 pili are positively regulated by *rns*, which is on a plasmid (1–3, 47, 48). We have cloned the CS2 genes, *cotB*, *cotA*, *cotC*, and *cotD*, and we present the translation of their DNA sequence here. Further, we have determined that cross-complementation occurs between some of the CS1 and CS2 proteins.

MATERIALS AND METHODS

Media. For most experiments, the bacteria were grown in Luria-Bertani broth (36). For hemagglutination and electron microscopy, cultures were grown on CFA agar (9). Antibiotics used were 50 μg of ampicillin per ml, 40 μg of chloramphenicol per ml, 40 μg of kanamycin per ml, and 10 μg of tetracycline per ml.

Bacterial strains and plasmids. *E. coli* K-12 strains MC4100 and DH5α (4, 33) were used for cloning. ETEC C91f-6 is a spontaneous CS pilus-minus derivative of the wild-type ETEC C91f, which expresses CS2 (40, 41). Strain LMC10 is a *lac* deletion, restriction-negative derivative of the ETEC-derived strain C921b-2 which expresses CS1 pili (1). Strain JEF100 is a derivative of LMC10 that contains the *cooB1* allele in place of the wild-type *cooB* (38). The *cooB1* allele is an insertion of a kanamycin resistance omega fragment, which inhibits transcription and translation. Strain FAK001 is a derivative of LMC10 which contains a tetracycline resistance omega fragment inserted into *cooC* (10). See Fig. 1 for the location of the omega fragment in JEF100 and in FAK001.

The cosmid vector pHC79, which carries ampicillin resistance (5), was used for cloning the CS2 gene cluster. The high-copy-number vector pUC19, which carries ampicillin resistance (23), and the low-copy-number pSC101-based vector pHSG576, which carries chloramphenicol resistance (44), were used for cloning. The *rns*-containing plasmids used were pEU2030, with *rns* cloned into pUC18 (1); pEU2040, with *rns* cloned into pHSG576 (27); and pEU2021, with *rns* cloned into pBR322 (1). Plasmid pEU478 contains *cooC* and *cooD* cloned under the *lac* promoter in pHSG576 (Fig. 1) (10).

The CS2 gene cluster was cloned as described previously for the CS1 gene cluster (27). Total DNA from *E. coli* C91f-6 was partially digested with *Sau*3A to give fragments 30 to 45 kb in size which were ligated to pHC79 cut with *Bam*HI. The ligated DNA was packaged into lambda by using the Gigapack Plus kit (Stratagene), and the lambda was used to transduce *E. coli* DH5α containing

* Corresponding author. Mailing address: Department of Microbiology and Immunology, Rollins Research Center, Emory University, 1510 Clifton Rd., Atlanta, GA 30322. Phone: (404) 727-0402. Fax: (404) 727-3659. Electronic mail address: Scott@microbio.emory.edu.

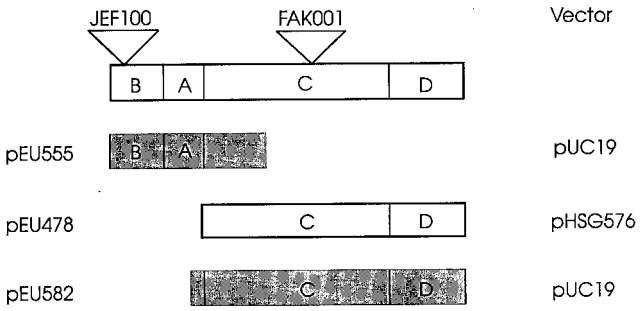

FIG. 1. Map of CS1 and CS2 genes in mutant strains and clones. The locations of omega insertions into the *coo* genes on the wild-type plasmid are shown above the map and labeled with the name of the mutant strains. The boxes labeled pEU555, pEU478, and pEU582 show the extent of the *coo* genes (white boxes) or *cot* genes (gray boxes) in each clone (further details are given in Materials and Methods). The genes are cloned under the *lac* promoter in each plasmid.

pEU2040, which encodes Rns. Of the 452 ampicillin- and chloramphenicol-resistant colonies tested by colony immunoblot analysis with antiserum specific for CS2 (see below), 3 were positive. One clone which contained CS2 antigen in heat extracts as determined by Western blot (immunoblot) analysis was chosen for further study. The plasmid contained in this CS2-positive strain was designated pEU5006. Plasmid pEU5006 was digested with *Pst*I and *Eco*RV, and the 5.7-kb fragment was isolated and ligated to pUC19 digested with *Pst*I and *Sma*I to give pEU588. This 5.7-kb fragment corresponds to bases 6 to 5698 of the sequenced DNA (see below).

The plasmids described below were cloned so that the pilin genes were expressed from the *lac* promoter in the vector (Fig. 1). To make plasmid pEU555, which contains *cotB* and *cotA*, a 6.0-kb *Cla*I-*Eco*RI fragment was blunted and ligated to pUC19 digested with *Sma*I. The *Eco*RI site of the 6.0-kb fragment is at bp 2753 in the DNA sequence. To make plasmid pEU582 (which contains *cotC* and *cotD*), a 4.0-kb *Pvu*II fragment, extending from bp 1642 to 5591, was purified and ligated to pUC19 digested with *Sma*I. The presence and orientation of all inserts were confirmed by restriction analysis.

Antiserum preparation. The anti-CS2 antiserum was prepared from C91f-6/pEU2021 cells as described by Scott et al. (38). Nonspecific antibodies were removed by adsorption with sonicated *E. coli* MC4100 and LMC10. The anti-CS1 antiserum was prepared as described previously (38).

Western blots and hemagglutination. Western blots were performed on whole-cell extracts as described by Perez-Casal et al. (27), and hemagglutination was performed in the presence of mannose as described by Caron et al. (3). Bovine erythrocytes (obtained from D. Jacobsen, University of Georgia) were used for the hemagglutinations.

DNA sequencing. The sequence of both strands was determined for bases 360 to 5798, and the sequence of the first 360 bases, which were 99% homologous to IS*3*, was determined on only one strand.

Electron microscopy. Strains were grown at 37°C on CFA agar containing the proper antibiotics. A small sample of each strain was suspended in 25 μl of a solution containing 1 mM Tris-HCl (pH 7.5) and 10 mM $MgCl_2$. The bacteria were allowed to float on the surface for 60 min at room temperature. A 10-μl volume of the sample was diluted into 15 μl of the same solution, and Formvar-coated carbon grids (200 mesh) were floated on the diluted sample for 30 min. Excess medium was removed with a paper wick, and the grids were stained for 10 to 15 s with ammonium phosphotungstate (pH 7.0). The grids were examined with a Philips CM-10 transmission electron microscope. For immunogold electron microscopy of pili, bacteria were adsorbed to Formvar-coated nickel grids as above and the grids were blocked in TBS (20 mM Tris-HCl, 150 mM NaCl [pH 7.8]) containing 3% bovine serum albumin (BSA). Grids were then floated on anti-CS1 antibody or anti-CS2 antibody for 2 h, washed six times in TBS–1% BSA, floated for 1 h on pools of 1:10-diluted gold-conjugated goat anti-rabbit immunoglobulin G (20-nm-diameter gold particles; EY Laboratories), washed six times in TBS, and counterstained as above for up to 40 s.

Nucleotide sequence accession number. The CS2 gene sequence has been submitted to the EMBL database under accession number Z47800.

RESULTS

CS2 gene cluster. To begin the analysis of the CS2 gene cluster, we constructed a cosmid library of total DNA from an ETEC strain that expressed functional CS2 pili. The library was screened by colony immunoblot with anti-CS2 antiserum, and a cosmid clone that expressed functional CS2 pili, as measured by hemagglutination, was isolated and subcloned. The subclone, plasmid pEU588, which contains an insert of about 5.7 kb, produces pili in *E. coli* K-12 strain MC4100 (Fig. 2), which were shown by hemagglutination to be functional (data not shown). Sequence analysis of this DNA region revealed that it contains four open reading frames, which we named *cot* (for coli surface antigen two).

cotB. The first open reading frame in the CS2 cluster, *cotB*

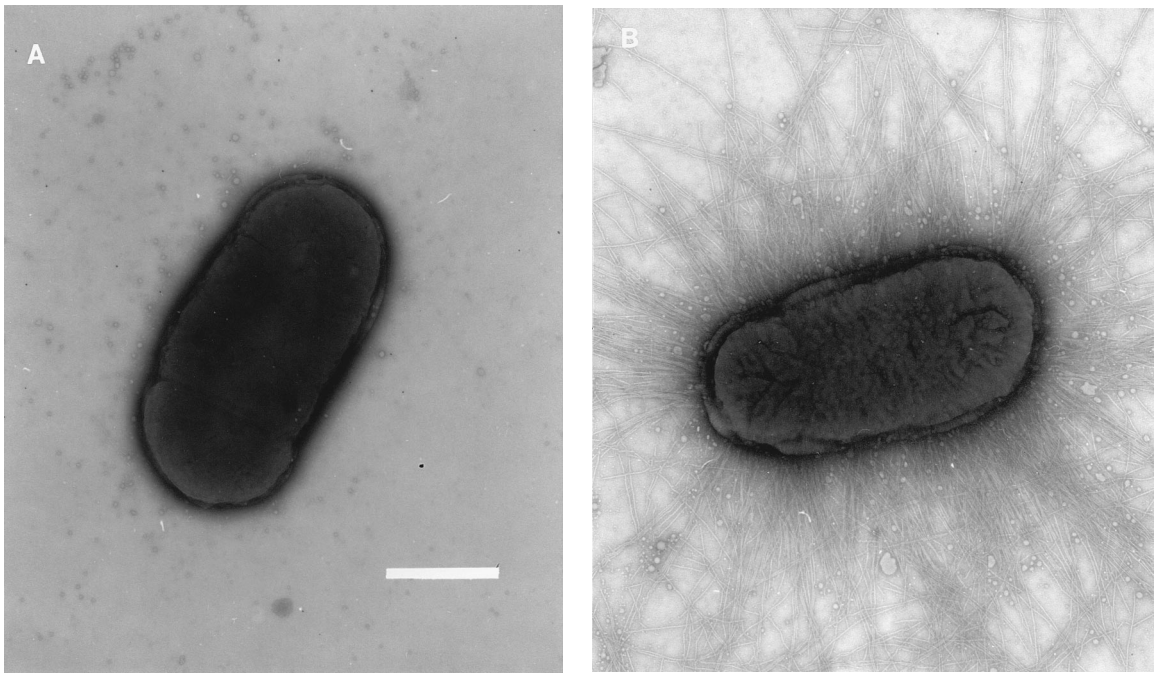

FIG. 2. Electron micrographs of MC4100 (A) and MC4100/pEU588 (B). Bar, 0.5 μm.

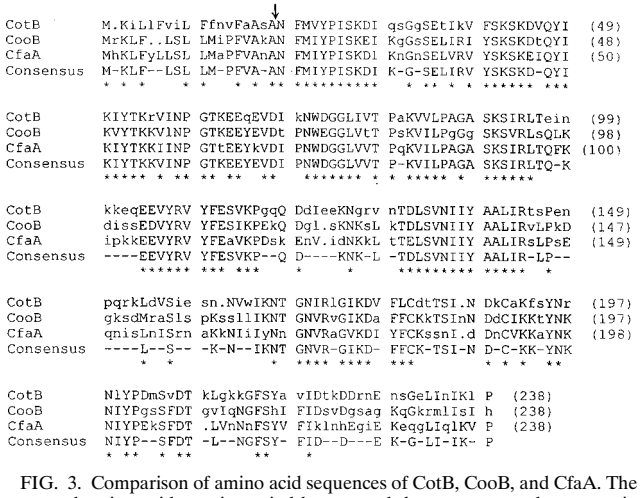

FIG. 3. Comparison of amino acid sequences of CotB, CooB, and CfaA. The conserved amino acids are in capital letters, and the nonconserved ones are in lowercase letters. Amino acids which are conserved in all three proteins are indicated by a * under the consensus sequence. The arrow above the line marks the predicted site of cleavage of the signal peptide. The analysis was done with the PILEUP algorithm in the Genetics Computer Group software.

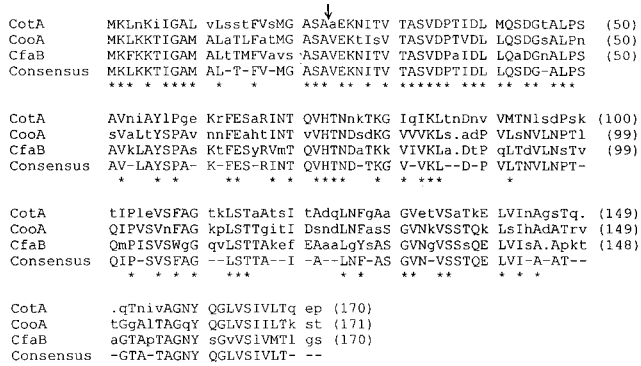

FIG. 4. Comparison of amino acid sequences of CotA, CooA, and CfaB. The symbols and analyses are as in Fig. 3.

(bp 499 to 1212), which is 714 nucleotides in length, is preceded by a potential ribosome-binding site (AAGG [11, 39]) beginning 16 bases upstream of the AUG start codon. The predicted CotB protein is 238 amino acids in length and is expected to have an 18-amino-acid signal sequence (46). Processing at the end of this sequence would generate a 24.8-kDa protein in the periplasm.

cotA. The second gene in the CS2 cluster, called *cotA*, is 510 nucleotides in length (bases 1255 to 1764). The *cotA* reading frame is preceded by a potential ribosome-binding site (TA AGG [11, 39]), which starts 14 bases upstream of the AUG start codon. The predicted CotA protein is 170 amino acids in length and is expected to have a 23-amino-acid signal sequence (46). Processing at the end of the signal sequence would generate a 15.4-kDa protein in the periplasm.

cotC. The third gene in the CS2 cluster, called *cotC*, is 2598 nucleotides in length (bases 1836 to 4433). There is a potential-ribosomal binding site (AAG [11, 39]) starting 14 bases upstream of the AUG start codon. The mature protein is expected to be 94.6 kDa in the periplasm after cleavage of a predicted 26-amino-acid signal sequence (46).

cotD. The fourth open reading frame, *cotD*, is 1,092 nucleotides in length (bases 4451 to 5542). The only start codon with a region predicted to be a ribosome-binding site at the proper distance upstream is the less frequently used start codon UUG. The potential ribosomal-binding site (GAGGT [11, 39]) starts 10 bases upstream from this. The predicted 364-amino-acid protein is expected to have an 18-amino-acid signal sequence (46), whose cleavage would generate a 38.9-kDa protein in the periplasm.

Homologies. A search of the combined GenBank databases with the Blast software showed that CotB, CotA, CotC, and CotD have significant homology with the gene products of the CS1 and CFA/I pilin gene clusters (10, 12, 17, 18, 27).

The product of the first gene in the cluster, CotB, is homologous to the products of the first genes in the CS1 and CFA/I gene clusters, CooB and CfaA, respectively (12, 17). Mature CotB is 54% identical and 72% similar to mature CooB and 52% identical and 71% similar to mature CfaA (Fig. 3).

CotA shows significant homology to the major pilin proteins of CS1 (CooA) and CFA/I (CfaB) (Fig. 4) (12, 27). All three are coded for by the second gene in their respective gene clusters. Mature CotA is 50% identical and 64% similar to mature CooA and 51% identical and 68% similar to mature CfaB (Fig. 4). The first 30 amino acids of the predicted mature CotA protein correspond to the N-terminal 30 amino acids determined for the CS2 antigen (42). Unlike many other pilins, CotA, CooA, and CfaB have no cysteines.

The predicted CotC protein shows homology with proteins encoded by the third genes in the CS1 and CFA/I gene clusters, CooC and CfaC, respectively (Fig. 5) (10, 18). As was found with CooC and CfaC (10), there is not much homology between CotC and its relatives in the regions containing the signal sequence (Fig. 5). However, mature CotC is 58% identical and 73% similar to mature CooC and 56% identical and 72% similar to mature CfaC (Fig. 5).

The product of the fourth gene in the CS2 cluster, CotD, is homologous to the proteins encoded by the fourth genes of the CS1 and CFA/I gene clusters, CooD and CfaE, respectively (10, 18). Mature CotD is 52% identical and 67% similar to mature CooD and 50% identical and 68% similar to mature CfaE (Fig. 6).

Complementation. Since the gene products needed for CS2 and CS1 pili production are homologous, we wished to determine whether they could complement each other for production of pili. Compatible plasmids containing the A and B genes from one cluster and the C and D genes from the other were constructed (see Materials and Methods). In these plasmids, the *coo* and *cot* genes were expressed from the *lac* promoter (see Materials and Methods) (Table 1; Fig. 1).

All complementation experiments were done with mutants of the ETEC-derived strain LMC10, which has the genes for CS1 pili. Two mutants of LMC10 were used. The first, JEF100, expresses no *coo* gene products because it contains an omega insertion in the first gene in the cluster, *cooB*, which is polar on the downstream genes (Fig. 1) (38). JEF100 does not provide CooA detectable on Western blots of whole-cell extracts (Fig. 7A, lane 1). We showed previously that JEF100 does not express pili and that the *coo* genes can act in *trans* to complement this mutant for production of CS1 pili (38). In the complementation test, the *cotA* and *cotB* products were produced from pEU555 and the *cooC* and *cooD* products were produced from pEU478 in the JEF100 background (Fig. 1). Western blot analysis indicates that JEF100/pEU555/pEU478 produced no detectable CooA (Fig. 7A, lane 4) but produced CotA as expected (Fig. 7B, lane 3). The JEF100 strain with either plasmid alone produced no CooA protein (Fig. 7A, lanes 2 and 3) and produced CotA protein only when pEU555 was present

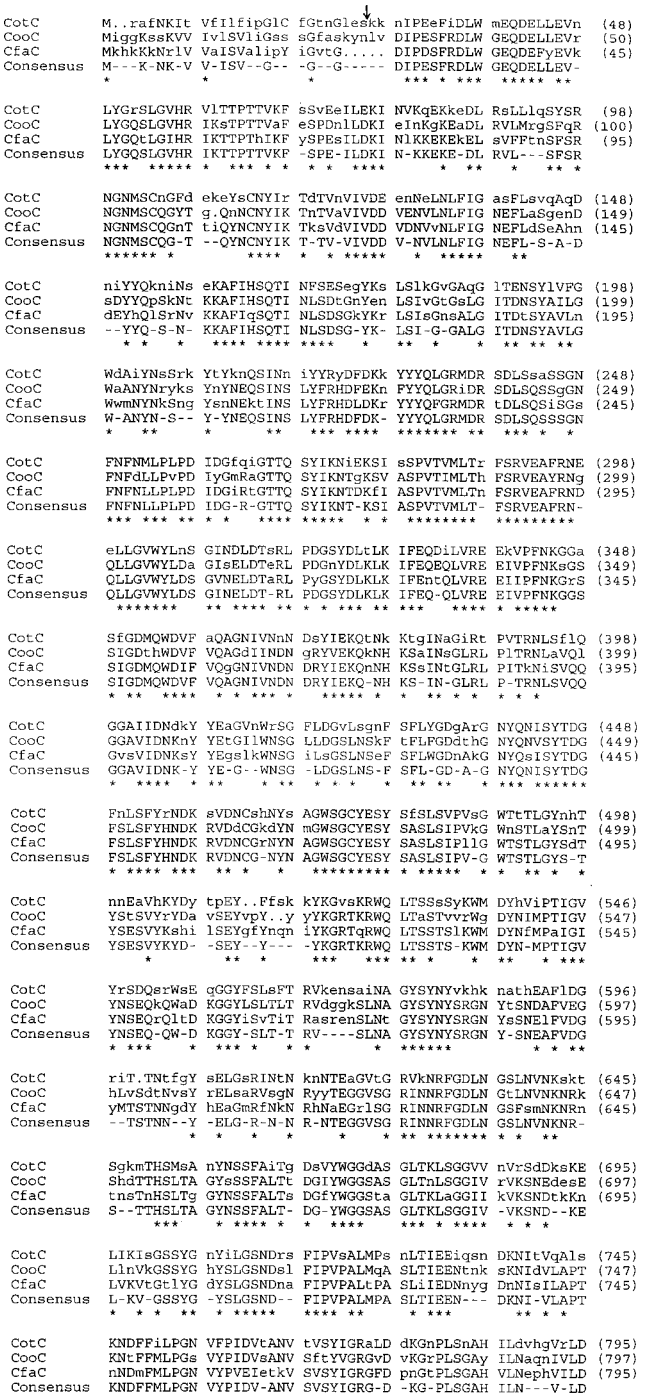

FIG. 5. Comparison of amino acid sequences of CotC, CooC, and CfaC. The symbols and analyses are as in Fig. 3, except that the arrow above the line marks the predicted site of cleavage of the signal peptide only for CotC. The predicted cleavage sites for CooC and CfaC have been described previously (10, 18).

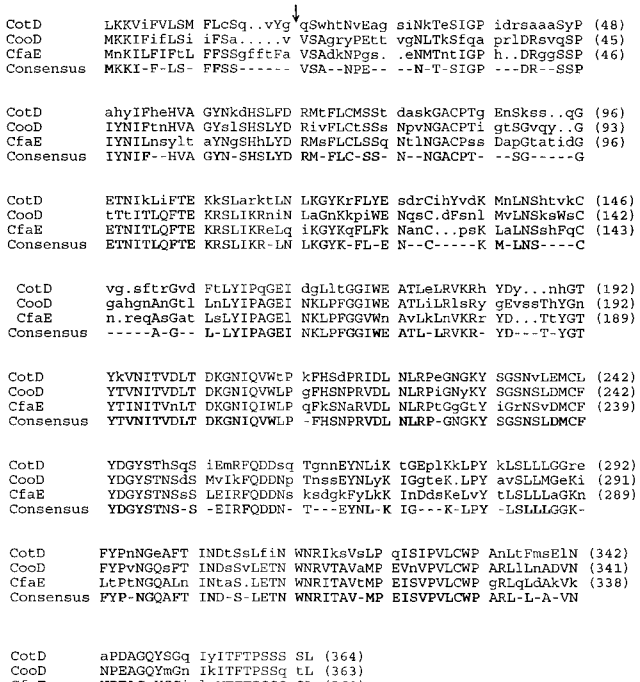

FIG. 6. Comparison of amino acid sequences of CotD, CooD, and CfaE. The symbols and analyses are as in Fig. 3, except that the arrow above the line marks the predicted site of cleavage of the signal peptide only for CotD; those for CooD and CfaE have been identified previously (10, 18).

(Fig. 7B, lanes 1 and 2). Although JEF100 carrying either plasmid alone produced no visible pili (data not shown), JEF100/pEU555/pEU478 was highly piliated (Fig. 8A). This indicates that *cotA* and *cotB* can interact with *cooC* and *cooD* to form visible pili. That JEF100/pEU555/pEU478 has CS2 and not CS1 pili was confirmed by immunogold electron microscopy (Fig. 8A; data not shown for anti-CS1).

For the reciprocal experiment, a different mutant derived from LMC10 was used. Strain FAK001 contains an omega insertion in *cooC* (Fig. 1), so it expresses no CooC and little or no CooD (10). In the presence of a plasmid producing Rns, the positive regulator of the CS1 and CS2 genes, FAK001 expresses CooB and CooA (10). FAK001/pEU2040 (a plasmid producing Rns) expresses *cooA* (Fig. 7A, lane 5) but makes no visible pili (Fig. 8B). In the homologous complementation, pili are visible. Strain FAK001 carrying pEU2030 (which expresses Rns) and pEU478 (which expresses CooC and CooD) ex-

TABLE 1. Complementation

Strain	Gene[a]				Pili present
	B	A	C	D	
JEF100	−	−	−	−	−
JEF100/pEU555/pEU478	*cot**	*cot**	*coo**	*coo**	+
FAK001/pEU2040	*coo*	*coo*			−
FAK001/pEU2030/pEU478	*coo*	*coo*	*coo**	*coo**	+
FAK001/pEU2040/pEU582	*coo*	*coo*	*cot**	*cot**	+

[a] The column headings list the genes in the order in which they occur in the gene clusters. For each complementation, the allele present (*coo* or *cot*) is indicated for each gene. The alleles labeled with * are cloned under the *lac* promoter. The other genes are in their wild-type location and require *rns*, provided on pEU2040 or pEU2030, for expression.

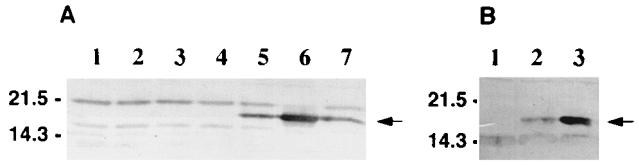

FIG. 7. Western immunoblots with anti-CS1 and anti-CS2 antisera. Molecular weights (in thousands) determined by a lane containing markers are indicated on the left of the figure. The locations of CooA and CotA are indicated by arrows on the right of the figure. (A) Whole-cell extracts reacted with anti-CS1. Lanes contain extracts of 1. JEF100 (lane 1), JEF100/pEU478 (lane 2), JEF100/pEU555 (lane 3), JEF100/pEU555/pEU478 (lane 4), FAK001/pEU2040 (lane 5), FAK001/pEU2030/pEU478 (lane 6), and FAK001/pEU2040/pEU582 (lane 7). (B) Whole-cell extracts reacted with anti-CS2. Lanes contain extracts of JEF100/pEU478 (lane 1), JEF100/pEU555 (lane 2), and JEF100/pEU555/pEU478 (lane 3).

presses CooA (Fig. 7A, lane 6) and is highly piliated (Fig. 8C) (10). In the heterologous complementation, when pEU582, which encodes *cotC* and *cotD*, is present in FAK001/pEU2040, CooA is expressed (Fig. 7A, lane 7) and many pili are also visible (Fig. 8D). Thus, *cooA* and *cooB* can also interact with *cotC* and *cotD* to form pili. Immunogold electron microscopy confirmed the presence of CS1 and absence of CS2 pili on FAK001/pEU2030/pEU478 (Fig. 8C) and FAK001/pEU2040/pEU582 (Fig. 8D; data not shown for anti-CS2).

DNA sequence upstream and downstream of the CS2 gene cluster. The region upstream of the CS2 gene cluster is of interest because it contains the promoter for at least the first gene (*cotB*) of the CS2 gene cluster and possibly includes regulatory elements. Part of this region (bases 1 to 369) is 99% homologous with the last 369 bases of the IS*3* insertion sequence (45). Directly following the IS*3* homology is a region (bases 369 to 401) which is 100% homologous with bases 73 to 41 of IS*1* (26). There is a potential promoter (bases 422 to 451) upstream of the CotB open reading frame (which begins at base 499). This promoter is very likely to be utilized, since both the −35 (gTGACA) and the −10 (TATcAT) regions match the consensus in 5 of 6 bases and the spacing of 18 bases between the two regions is acceptable (13, 28).

Downstream of the *cot* gene cluster, 253 bp was sequenced. There are no open reading frames in this region, and a search of the GenBank database showed no significant homologies with it.

DISCUSSION

The sequence of the four *cot* gene products needed for production of CS2 pili by *E. coli* K-12 shows them to be related to the *coo* gene products for CS1 pili and the *cfa* gene products for CFA/I pili. The predicted hydrophobicity plot of each of the *cot* gene products is very similar to the hydrophobicity plot predicted for its homologs. These three types of pili found on different ETEC strains isolated from human infections are clearly members of the same family. For all four genes in this family, there are regions conserved in all three members of the pilus family interspersed with nonhomologous regions (Fig. 3 to 6). These highly conserved regions are likely to be important for protein structure and function.

In addition to the homology found between CotA and the CS1 and CFA/I major pilin proteins CooA and CfaB, the N-terminal amino acid sequences of the CS4 and PCF0166 pilin proteins are also homologous (43, 49). All of these pili are found on ETEC strains associated with human disease. Amino acids 2 to 20 of the mature CotA protein are 73% identical and 84% similar to amino acids 2 to 20 of the CS4 pilin. Comparing PCF0166 with CotA, amino acids 2 to 25 of the mature proteins are 79% identical and 100% similar. At this time, only the N-terminal sequences of the CS4 and PCF0166 pilin proteins are known, so it is possible that the homology extends over the entire proteins. These homologies suggest that CS4 and PCF0166 pili may be additional members of the same family as CS1, CS2, and CFA/I.

The major pilin proteins, CotA, CooA, and CfaB, are serologically different, although weak cross-reactivity has been reported among some of them (22). In addition, the predicted sequences of the proteins encoded by each of the other genes needed to produce pili show some differences (Fig. 3, 5, and 6). For the major pilin antigen (the second gene in each cluster) and the protein encoded by the third gene of each cluster, the predicted pI values of the homologs are similar (second gene: CotA, 4.86; CooA, 5.5; CfaB, 5.09; and third gene: CotC, 6.53; CooC, 5.36; CfaC, 6.88). However, for the first gene in each cluster, there is a large difference in the predicted pI values between CotB (pI 8.6) and the other two, CooB (pI 10.3) and CfaA (pI 10.2). Similarly, for the last gene in the clusters, there is a large difference in pI values predicted for one of the family members. The predicted pI of CfaE is 9.2, while the predicted values for the CooD and CotD proteins are 7.1 and 7.0, respectively. Because neither the three-dimensional structures nor the functions of these proteins are known, it is too early to suggest the possible significance of the differences in predicted pI values.

The only one of the proteins encoded by the CS1, CS2, and CFA/I gene clusters whose function can be predicted from sequence analysis is CotC and its homologs CooC and CfaC. These proteins have several properties in common with those of outer membrane proteins (15, 24). All three proteins are predicted to have high beta-sheet and low alpha-helix content (30, 31, 34). They also have a high content of charged amino acids and are predicted to have no extensive hydrophobic regions. Therefore, we expect that CotC, CooC, and CfaC are outer membrane proteins.

Some limited homology was found between CotC and some characterized outer membrane proteins. Among these, the gene product with the most homology to CotC (and also to CooC and CfaC) is Cfa1A of *Yersinia pestis*, which encodes a protein involved in capsule biogenesis. This process is thought to be similar to formation of pili (19), and the Cfa1A protein has some homology to PapC, an outer membrane protein required for dissociation of the *pap* pilin from the PapD chaperonin and for addition of the pilin subunits in the correct order to form a pilus structure (7). Two other gene products with some limited homology to CotC are FanD and FimD. These are large outer membrane proteins required, respectively, for assembly of K99 pili of ETEC strains from animals and type 1 pili of *Salmonella typhimurium* (29; Swenson and Clegg accession number P37924). Both FanD and FimD also show homology to PapC. These homologies suggest that CotC is an outer membrane protein needed for export and assembly of the CS2 pili. In agreement with this hypothesis, we have shown previously that CooC is required for assembly of the CS1 pili (10). Because of its probable localization in the outer membrane and because of the predicted large size of CotC, comparison with other pilus systems would suggest that CotC is probably involved in transport of the growing pili through the outer membrane of *E. coli.*

The success of cross-complementation in the ETEC strain between pilus proteins with analogous functions presumably depends on local similarities at the regions important for their interaction with the other proteins needed for pilus formation. Even in systems in which the function of the proteins is better

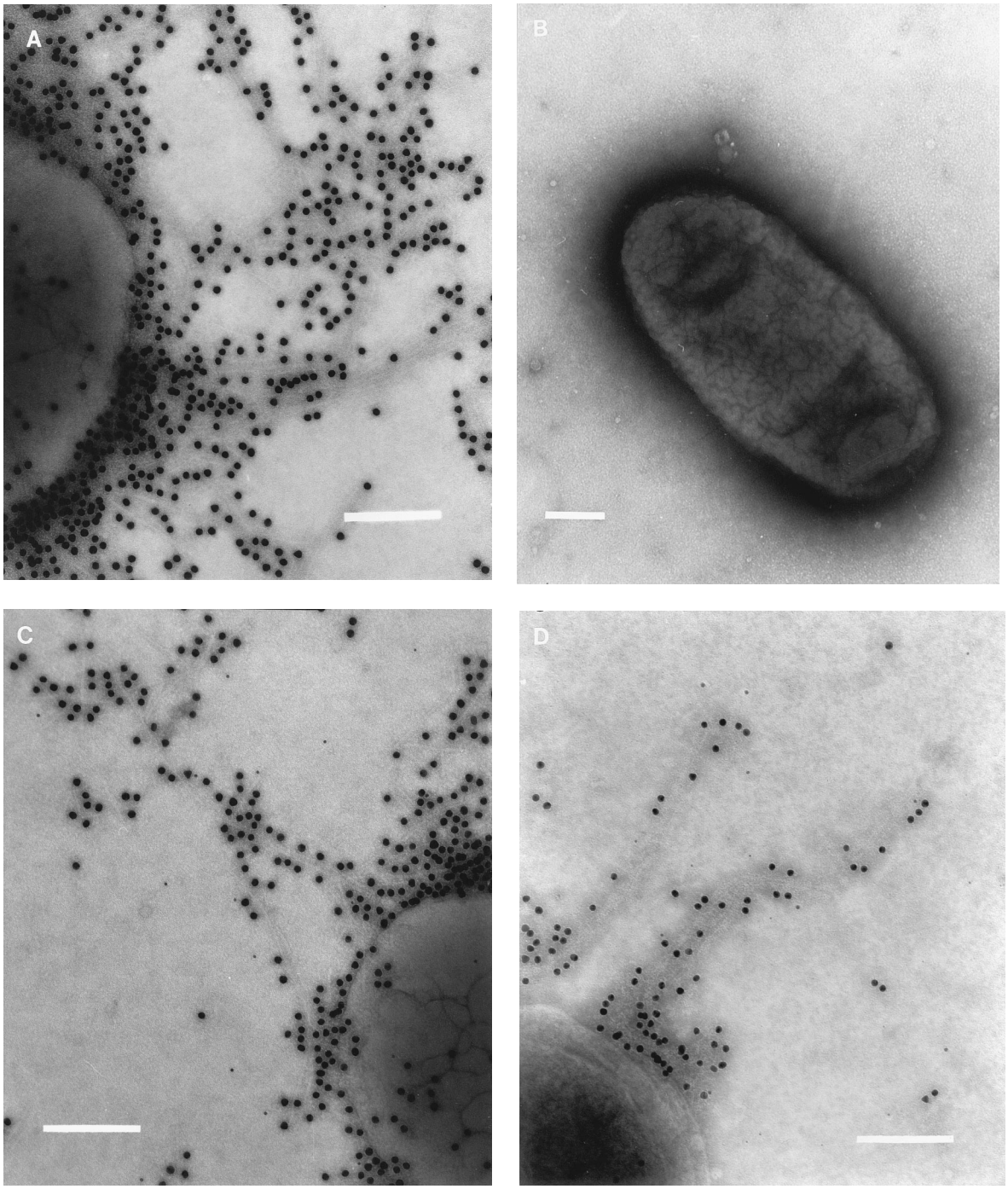

FIG. 8. Electron micrographs. JEF100/pEU555/pEU478 was incubated with anti-CS2 as discussed in Materials and Methods. FAK001/pEU2030/pEU478 and FAK001/pEU2040/pEU582 were incubated with anti-CS1. All three strains were then incubated with gold-conjugated goat anti-rabbit immunoglobulin G. (A) JEF100/pEU555/pEU478; (B) FAK001/pEU2040; (C) FAK001/pEU2030/pEU478; (D) FAK001/pEU2040/pEU582. For a description of which CS1 and CS2 genes are present in each strain, see Table 1 and Fig. 1. Bar, 0.2 μm.

understood than that of human ETEC pili, it has not been possible to predict the results of cross-complementation experiments. For example, the genes for the pap chaperone (*papD*) and the type 1 chaperone (*fimC*), which are only 32% identical, complement in only one direction. The wild-type *papD* complements a mutation in *fimC*, resulting in production of type 1 pili, but wild-type *fimC* is unable to complement a *papD* mutation (16). On the other hand, although their products are 68% identical, *fimC* and *focC* (which encodes the chaperone of F1C pili) cannot replace each other unless their cognate usher protein is present (20). In this work, we have shown that CotC and CotD can substitute functionally for CooC and CooD, and vice versa, although they share only about 50% amino acid identity. This suggests that the major pilins of CS1 and CS2 are

similar enough for either predicted transport protein to interact with both.

The lack of specificity of the assembly functions of the human ETEC pili may be useful for design of a vaccine consisting of an enteric bacterium that can produce several pilus types. The complementation tests (Table 1) demonstrated that pili containing CotA can be made by using the CooC and CooD proteins and, conversely, that pili containing CooA can be made by using the CotC and CotD proteins. Thus, it appears that the presence of either version of the C and D proteins will allow the polymerization of the nonhomologous major pilin protein. We therefore anticipate that it will be possible for a vaccine strain expressing either version of the C and D proteins to produce either both kinds of pili and/or chimeric pili containing mixed subunits. If this is correct and if this complementation can be extended to the other members of the human ETEC pilus family, this might greatly facilitate development of a pilus-based ETEC vaccine.

Many virulence gene clusters appear to have been imported as a unit into bacteria that may not have previously been pathogenic. This is deduced from their unusual G+C content and/or from the presence of insertion sequences flanking them. The CS2 gene cluster is 39% G+C, which is significantly lower than the average for *E. coli* (50%). Therefore, as we previously suggested for the CS1 gene cluster (10), the CS2 region may have been introduced recently into *E. coli* as a cassette by transposition.

We had noted previously that the CS1 gene cluster is bounded by IS-related sequences (10, 37). We have now found that upstream of the CS2 gene cluster is the 3′ end of an IS*3* element (45), and it is possible that a complete IS*3* is present. However, within the 253 bases that have been sequenced downstream of CS2, no IS-related sequences have been found. This might indicate that the CS2 gene cluster was not introduced into *E. coli* on a transposon or that it has been present in *E. coli* long enough for the downstream IS to have diverged more than the sequences flanking the CS1 gene cluster. It is also possible that an IS remnant is present downstream of the *cot* genes but that it does not resemble any currently identified IS elements. Finally, it is possible that there is an IS element present further downstream from the CS2 gene cluster.

ACKNOWLEDGMENTS

We thank Judy Caron for assistance in constructing the CS2 cosmid clone and Lawrence Melson for help with electron microscopy. We are grateful to the Biomolecular Computing Resource at Emory for facilitating sequence analyses.

This work was supported by Public Health Service grant AI24870 from the National Institutes of Health.

REFERENCES

1. **Caron, J., L. Coffield, and J. Scott.** 1989. A plasmid-encoded regulatory gene, *rns*, required for expression of the CS1 and CS2 adhesins of enterotoxigenic *Escherichia coli*. Proc. Natl. Acad. Sci. USA **86:**963–967.
2. **Caron, J., D. R. Maneval, J. B. Kaper, and J. R. Scott.** 1990. Association of *rns* homologs with colonization factor antigens in clinical *Escherichia coli* isolates. Infect. Immun. **58:**3442–3444.
3. **Caron, J., and J. Scott.** 1990. A *rns*-like regulatory gene in CFA/I that controls expression of CFA/I pilin. Infect. Immun. **58:**874–878.
4. **Casadaban, M.** 1976. Regulation of the regulatory gene for the arabinose pathway, *araC*. J. Mol. Biol. **104:**557–566.
5. **Collins, J., and B. Hohn.** 1980. A small cosmid for efficient cloning of large DNA fragments. Gene **11:**291–298.
6. **de Graaf, F.** 1990. Genetics of adhesive fimbriae of intestinal *Escherichia coli*. Curr. Top. Microbiol. Immunol. **151:**29–53.
7. **Dodson, K. W., F. Jacob-Dubuisson, R. T. Striker, and S. Hultgren.** 1993. Outer-membrane PapC molecular usher discriminately recognizes periplasmic chaperone-pilus subunit complexes. Proc. Natl. Acad. Sci. USA **90:**3670–3674.
8. **Evans, D. G., and D. J. Evans.** 1978. New surface-associated heat-labile colonization factor antigen (CFA/II) produced by enterotoxigenic *Escherichia coli* of serogroups O6 and O8. Infect. Immun. **21:**638–647.
9. **Evans, D. G., D. J. Evans, and W. Tjoa.** 1977. Hemagglutination of human group A erythrocytes by enterotoxigenic *Escherichia coli* isolated from adults with diarrhea: correlation with colonization factor. Infect. Immun. **18:**330–337.
10. **Froehlich, B. J., A. Karakashian, L. R. Melsen, J. C. Wakefield, and J. R. Scott.** 1995. CooC and CooD are required for assembly of CS1 pili. Mol. Microbiol. **12:**387–401.
11. **Gold, L.** 1988. Post-transcriptional regulatory mechanisms in *Escherichia coli*. Annu. Rev. Biochem. **57:**199–233.
12. **Hamers, A., H. Pel, G. Willshaw, J. Kusters, B. Van Der Zeijst, and W. Gaastra.** 1989. The nucleotide sequence of the first two genes of the CFA/I fimbrial operon of human enterotoxigenic *Escherichia coli*. Microb. Pathog. **6:**297–309.
13. **Hawley, D. K., and W. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11:**2237–2255.
14. **Hultgren, S. J., S. Abraham, M. Caparon, P. Falk, J. W. St. Geme III, and S. Normark.** 1993. Pilus and nonpilus bacterial adhesins: assembly and function in cell recognition. Cell **73:**887–901.
15. **Jap, B., and P. Walian.** 1990. Biophysics of the structure and function of porins. Q. Rev. Biophys. **23:**367–403.
16. **Jones, C. H., J. S. Pinkner, A. V. Nicholes, L. N. Slonin, S. N. Abraham, and S. J. Hultgren.** 1993. FimC is a periplasmic PapD-like chaperone that directs assembly of type 1 pili in bacteria. Proc. Natl. Acad. Sci. USA **90:**8397–8401.
17. **Jordi, B. J. A. M., A. H. M. Van Vliet, G. A. Willshaw, B. A. M. Van Der Zeijst, and W. Gaastra.** 1991. Analysis of the first two genes of the CS1 fimbrial operon in human enterotoxigenic *Escherichia coli* of serotype O139:H28. FEMS Microbiol. Lett. **80:**265–270.
18. **Jordi, B. J., G. A. Willshaw, B. A. Van Der Zeijst, and W. Gaastra.** 1992. The complete nucleotide sequence of region 1 of the CFA/I fimbrial operon of human enterotoxigenic *Escherichia coli*. DNA Sequence **2:**257–263.
19. **Karlyshev, A. V., E. E. Galyov, O. Smirnov, O. Yu, A. P. Guzayez, V. M. Abramov, and V. P. Zav'yalov.** 1994. A new gene of the f1 operon of *Y. pestis* involved in capsule biogenesis. FEBS Lett. **297:**77–80.
20. **Klemm, P., B. J. Jorgensen, B. Kreft, and G. Christiansen.** 1995. The export systems of type 1 and F1C fimbriae are interchangeable but work in parental pairs. J. Bacteriol. **177:**621–627.
21. **Levine, M. M., P. Ristaino, G. Marley, C. Smyth, S. Knutton, E. Boedeker, R. Black, C. Young, M. L. Clements, C. Cheney, and R. Patnaik.** 1984. Coli surface antigens 1 and 3 of colonization factor antigen II-positive enterotoxigenic *Escherichia coli*: morphology, purification, and immune responses in humans. Infect. Immun. **44:**409–420.
22. **McConnell, M. M., H. Chart, and B. Rowe.** 1989. Antigenic homology within human enterotoxigenic *Escherichia coli* fimbrial colonization factor antigens: CFA/I, coli-surface-associated antigens (CS)1, CS2, CS4 and CS17. FEMS Microbiol. Lett. **61:**105–108.
23. **Messing, J., and J. Vieira.** 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene **19:**269–276.
24. **Nikaido, H., and M. Vaara.** 1985. Molecular basis of bacterial outer membrane permeability. Microbiol. Rev. **49:**1–32.
25. **Normark, S., M. Baga, M. Goransson, F. Lindberg, B. Lund, M. Norgren, and B.-E. Uhlin.** 1986. Genetics and biogenesis of *Escherichia coli* adhesins, p. 113–143. *In* D. Mirelman (ed.), Microbial lectins and agglutinins: properties and biological activity. John Wiley & Sons, Inc., New York.
26. **Ohtsubo, E., and H. Ohtsubo.** 1978. Nucleotide sequence of an insertion element. Proc. Natl. Acad. Sci. USA **75:**615–619.
27. **Perez-Casal, J., J. Swartley, and J. R. Scott.** 1990. Gene encoding the major subunit of CS1 pili of human enterotoxigenic *Escherichia coli*. Infect. Immun. **58:**3594–3600.
28. **Pribnow, D.** 1979. Genetic control signals in DNA, p. 219–277. *In* R. F. Goldberger (ed.), Biological regulation and development, vol. 1. Plenum Press, New York.
29. **Roosendall, B., D. Bakker, and F. D. De Graaf.** 1989. The nucleotide sequence of the *fanD* gene encoding the large outer membrane protein involved in the biosynthesis of K99 fimbriae. Nucleic Acids Res. **17:**1263.
30. **Rost, B., and C. Sander.** 1992. Jury returns on structure prediction. Nature (London) **360:**540.
31. **Rost, B., and C. Sander.** 1993. Prediction of protein secondary structure at better than 70% accuracy. J. Mol. Biol. **232:**584–599.
32. **Russell, P. W., and P. E. Orndorff.** 1992. Lesions in two *Escherichia coli* type 1 pilus genes alter pilus number and length without affecting receptor binding. J. Bacteriol. **174:**5923–5935.
33. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
34. **Sander, C., and R. Schneider.** 1991. Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins Struct. Funct. Genet. **9:**56–68.
35. **Savelkoul, P., G. Willshaw, M. McConnell, H. Smith, A. Hamers, B. Van Der Zeijst, and W. Gaastra.** 1990. Expression of CFA/I fimbriae is positively regulated. Microb. Pathog. **8:**91–99.

36. **Scott, J. R.** 1974. A turbid plaque-forming mutant of phage P1 that cannot lysogenize *Escherichia coli*. Virology **62:**344–349.
37. **Scott, J. R., and B. J. Froehlich.** 1994. CS1 pili of enterotoxigenic *E. coli*, p. 17–30. *In* C. I. Kado and J. H. Crosa (ed.), Molecular mechanisms of bacterial virulence. Kluwer Academic Publishers, Dordrecht, The Netherlands.
38. **Scott, J. R., J. C. Wakefield, P. W. Russell, P. E. Orndorff, and B. J. Froehlich.** 1992. CooB is required for assembly but not transport of CS1 pilin. Mol. Microbiol. **6:**293–300.
39. **Shine, J., and L. Dalgarno.** 1974. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71:**1342–1346.
40. **Smyth, C. J.** 1982. Two mannose-resistant haemagglutinins on enterotoxigenic *E. coli* of serotype O6:K15:H16 or H^- isolated from travellers' and infantile diarrhoea. J. Gen. Microbiol. **128:**2081–2096.
41. **Smyth, C. J.** 1984. Serologically distinct fimbriae on enterotoxigenic *Escherichia coli* of serotype O6:K15:H16 or H−. FEMS Microbiol. Lett. **21:**51–57.
42. **Smyth, C. J., M. Boylan, H. M. Matthews, and D. C. Coleman.** 1991. Fimbriae of human enterotoxigenic *Escherichia coli* and control of their expression. FEMS Symp. **51:**37–54.
43. **Sommerfelt, H., H. M. Grewal, A. M. Svennerholm, W. Gaastra, P. R. Frod, G. Vidous, and M. K. An Bhan.** 1992. Genetic relationship of putative colonization factor 0166 to colonization factor I and coli surface antigen 4 of enterotoxigenic *Escherichia coli*. Infect. Immun. **60:**3799–3806.
44. **Takeshita, S., M. Sato, M. Toba, W. Masahashi, and T. Hashimoto-Gotoh.** 1987. High-copy-number and low-copy-number plasmid vectors for *lacZ* alpha-complementation and chloramphenicol- or kanamycin-resistance selection. Gene **61:**63–74.
45. **Timmerman, K. P., and C. Tu.** 1985. Complete sequence of IS3. Nucleic Acids Res. **13:**2127–2139.
46. **von Heijne, G.** 1987. Sequence analysis in molecular biology: treasure trove or trivial pursuit, p. 92–93. Academic Press, Inc., New York.
47. **Willshaw, G., M. McConnell, W. Gaastra, A. Thomas, M. Hibberd, and B. Row.** 1990. Structural and regulatory genes for coli surface antigens 4 (CS4) are encoded by separate plasmids in enterotoxigenic *Escherichia coli* strains of serotype O25.H42. FEMS Microbiol. Lett. **68:**255–260.
48. **Willshaw, G. A., H. R. Smith, M. M. McConnell, and B. Rowe.** 1991. Cloning of regulator genes controlling fimbrial production by enterotoxigenic *Escherichia coli*. FEMS Microbiol. Lett. **82:**125–130.
49. **Wolf, M. K., G. P. Andrews, B. D. Tall, M. M. McConnell, M. M. Levine, and E. C. Boedeker.** 1989. Characterization of CS4 and CS6 antigenic components of PCF8775, a putative colonization factor complex from enterotoxigenic *Escherichia coli* E8775. Infect. Immun. **57:**164–173.

Editor: P. E. Orndorff