JOURNAL OF BACTERIOLOGY, May 1996, p. 2613–2628 Vol. 178, No. 9
0021-9193/96/$04.00+0

Enteropathogenic *Escherichia coli*: Identification of a Gene Cluster Coding for Bundle-Forming Pilus Morphogenesis

INDIRA SOHEL,[1,2,3] JOSE LUIS PUENTE,[1,4] SANDRA W. RAMER,[1,2,3] DAVID BIEBER,[1,2,3] CHENG-YEN WU,[1,2,3] AND GARY K. SCHOOLNIK[1,2,3]*

Departments of Microbiology and Immunology[1] and Medicine,[2] Division of Infectious Diseases and Geographic Medicine, and the Howard Hughes Medical Institute,[3] Stanford University School of Medicine, Stanford, California 94305, and Department of Molecular Microbiology, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos 62271, Mexico[4]

Received 22 November 1995/Accepted 1 March 1996

Sequence flanking the *bfpA* locus on the enteroadherent factor plasmid of the enteropathogenic *Escherichia coli* (EPEC) strain B171-8 (O111:NM) was obtained to identify genes that might be required for bundle-forming pilus (BFP) biosynthesis. Deletion experiments led to the identification of a contiguous cluster of at least 12 open reading frames, including *bfpA*, that could direct the synthesis of a morphologically normal BFP filament. Within the *bfp* gene cluster, we identified open reading frames that share homology with other type IV pilus accessory genes and with genes required for transformation competence and protein secretion. Immediately upstream of the *bfp* gene cluster, we identified a potential replication origin including genes that are predicted to encode proteins homologous with replicase and resolvase. Restriction fragment length polymorphism analysis of DNA from six additional EPEC serotypes showed that the organization of the *bfp* gene cluster and its juxtaposition with a potential plasmid origin of replication are highly conserved features of the EPEC biotype.

Enteropathogenic *Escherichia coli* (EPEC) strains are a common cause of diarrhea, particularly among children living in developing countries (30, 31, 50, 52). Unlike the enterotoxigenic, enterohemorrhagic, and enteroinvasive *E. coli* biotypes, EPEC strains do not produce secretory enterotoxins or cytotoxins and do not invade and destroy intestinal epithelial cells (51). Instead, small-bowel biopsies of EPEC-infected children reveal discrete microcolonies of bacteria attached to mucous membranes (84, 85). Beneath adherent bacteria, close juxtaposition of the bacterial outer membrane and the underlying plasma membrane occurs, associated with localized elevation and invagination of the plasma membrane, loss of microvilli, and rearrangement of the cell's cytoskeleton (85). Together, these features compose the attaching and effacing lesion that epitomizes EPEC infection of epithelial surfaces (42, 43). Incubation of EPEC with tissue culture cell monolayers in vitro produces a similar pattern of adherence characterized by circumscribed clusters of bacteria attached to the epithelial cell surface (15, 61, 91). This pattern of attachment, termed localized adherence (LA), is now recognized to be a common property of the classic EPEC serotypes and requires the ~80-kb enteroadherent factor (EAF) plasmid (6, 53, 59, 60). Strains cured of the EAF plasmid no longer adhere to epithelial cells and when tested in human volunteers are found to be less virulent (6, 41, 53).

During growth in tissue culture media, EPEC strains express thin pilus filaments that emanate from localized sites on the bacterial surface, where they form laterally aligned aggregates or bundles of pilus filaments (29, 103). Designated bundle-forming pili (BFP), these structures are produced within adherent microcolonies of EPEC, where they appear to form a meshwork of interbacterial fibers that may physically stabilize the attached colony (29). BFP also mediate the autoaggregation phenotype, manifested by the formation of unattached bacterial aggregates that float above the epithelial cell surface and which may constitute the functional EPEC infectious unit (103).

Purified BFP, analyzed by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE), appear to be composed of a single, repeating 19.5-kDa polypeptide subunit (29). *bfpA*, the gene coding for this structural component of the BFP filament, has been cloned from the EAF plasmid of two EPEC serotypes (18, 93) and found to code for a 180-amino-acid polypeptide containing an N-terminal amino acid prepilin leader sequence and hydrophobic domain that demonstrates similarity to other members of the type IV pilin family (97).

Studies of the uropathogenic *E. coli* pyelonephritis-associated pilus (Pap) have demonstrated a contiguous cluster of genes that code for its pathogenic and biosynthetic properties, including a variety of accessory genes that are required for the production of a morphologically and functionally normal pilus filament (37). However, the amino acid sequence of PapA, the principal structural subunit of the Pap filament, is not homologous with the type IV pilins, and little similarity is evident with respect to genomic organization or biosynthetic process. Of the eight type IV pilin genes cloned thus far (97), only the toxin-coregulated pilus (Tcp) biosynthetic system of *Vibrio cholerae* was found to be arranged as a contiguous cluster of genes. The dispersal of type IV biosynthetic genes in many of the species studied thus far has impeded their morphogenic and functional characterization compared with the *E. coli* Pap system.

To determine if *bfpA* is closely linked to other genes required for the production of the BFP filament, we obtained ca. 17 kb of sequence from the EAF plasmid in the region flanking *bfpA*. Deletion experiments demonstrated that a region of DNA containing at least 12 contiguous open reading frames

* Corresponding author. Phone: (415) 723-8158. Fax: (415) 723-1399.

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

Strain or plasmid	Description	Reference(s)
Strains		
B171	EPEC O111:NM	69, 103
B171-8	EPEC O111:NM, containing the EAF plasmid	29
B171-4	EAF plasmid cured derivative of B171-8	82, 83
DH5α	Laboratory K-12 strain	87
BL21(DE3)	Expression host for cloned gene in pET plasmid	98
Plasmids		
pWKS130	Low-copy-number cloning vector (Km^r)	105
pBTA-BH1	pWKS130 (Km^r) derivative carrying 3.9-kb *Bam*HI EAF fragment containing *bfpT*	100
pSS9	pACYC177 (Ap^r) derivative carrying 4.0-kb *Hin*dIII EAF fragment containing *bfpA*	93
pSSD1	pJF119 (Ap^r) derivative carrying 0.6-kb fragment containing *bfpA*	93
pISP	pKK233 (Ap^r) derivative carrying PCR-amplified fragment of *bfpP* ORF at *Eco*RI-*Pst*I sites of the multicloning site	This study
pKK233-3	pBR322 (Ap^r)-derived vector for overexpression of proteins	4
pISD	pET23a+ (Ap^r) derivative carrying PCR-amplified fragment of *bfpD*	This study
pISE	pET23a+ (Ap^r) with insert of PCR-amplified fragment of *bfpE*	This study
pISF	pET23a+ (Ap^r) carrying PCR-amplified fragment of *bfpF*	This study
pET23a+	Expression vector with T7 promoter	99
pIS1	pACYC184 (Cm^r) derivative carrying 10-kb EAF *Sau*3AI partial library fragment containing *repI* to *bfpF*	This study
pIS2	pET23a+ (Ap^r) derivative carrying 11.5-kb *Hin*dIII fragment containing *bfpB* to the end of *bfpM*	This study
pERI-992	pUC18 (Ap^r) derivative carrying 10.9-kb *Eco*RI EAF fragment containing *bfpA* to the end of *bfpJ*	This study
pERI-493	pUC18 (Ap^r) derivative carrying 7.5-kb *Eco*RI EAF fragment containing *bfpA* to *bfpF*	This study
pISK	pBC KS+ (Cm^r) derivative carrying 11.3-kb *Eco*RI-*Ssp*I EAF fragment containing *bfpA* to the end of *bfpK*	This study
pISJ	pBC KS+ (Cm^r) derivative carrying 10.9-kb *Eco*RI EAF fragment containing *bfpA* to the end of *bfpJ*	This study
pISL	pBC KS+ (Cm^r) derivative carrying 12.1-kb *Eco*RI-*Eco*RV EAF fragment containing *bfpA* to the end of *bfpL*	This study
pISM	pUC18 (Ap^r) derivative carrying PCR-amplified 13.9-kb EAF fragment containing *rsv* to the end of *bfpM*	This study
pISB2	pWKS130 (Km^r) derivative carrying the ~9-kb *Bam*HI EAF fragment from nt 11764 to 20564	This study

(ORFs), which we have termed the *bfp* gene cluster, can direct the synthesis of morphologically normal BFP filaments. Analysis of these ORFs revealed similarities with accessory proteins involved with type IV pilin biogenesis, protein secretion, and DNA uptake.

MATERIALS AND METHODS

Bacterial strains, plasmids, and growth conditions. The bacterial strains and plasmids used in this study are listed in Table 1. Bacteria were grown in Luria-Bertani (LB) broth and agar or, to induce BFP expression, in Dulbecco's modified Eagle's medium containing 0.45% glucose (DMEM) or on Trypticase soy agar supplemented with 5% defibrinated sheep blood (TSA blood agar) as described by Vuopio-Varkila and Schoolnik (103) and Giron et al. (29).

DNA techniques. DNA manipulations were performed by standard genetic and molecular techniques (87). Restriction and DNA-modifying enzymes were obtained from New England Biolabs, Inc. (Beverly, Mass.), and used according to the manufacturer's instructions. Radiolabeled nucleotide [^{32}P]dCTP (3,000 Ci/mmol) was purchased from Amersham Corp. (Arlington Heights, Ill.). Oligonucleotides were prepared by Operon Technologies Inc. (Alameda, Calif.). PCRs were performed with AmpliTaq (Perkin-Elmer, Norwalk, Conn.) according to the manufacturer's instructions.

Construction of an EAF plasmid library. The EAF plasmid library used in this study was constructed as follows. Plasmid DNA was purified from strain B171-8 (Table 1) by using a Qiagen (Chatsworth, Calif.) plasmid DNA purification kit. A partial *Sau*3AI digestion of the plasmid DNA was conducted, and the resulting fragments were ligated to the *Bam*HI-digested vector pACYC184 (11). After transformation into *E. coli* DH5α (Table 1), recombinants were selected by growth on LB agar containing chloramphenicol (50 μg/ml). Clones containing overlapping fragments with *bfpA* were identified by using the labeled *bfpA* sequence (93) as a hybridization probe.

DNA sequence analysis. Double-stranded DNA sequencing of the 17-kb *bfp* gene cluster locus depicted in Fig. 1 was performed by the dideoxy-chain termination procedure (89). The resulting sequence was analyzed by using programs from Intelligenetics, Inc. (Mountain View, Calif.), the Center for Molecular and Genetic Medicine (Stanford University, Stanford, Calif.), and the Genetics Computer Group of the University of Wisconsin (19).

Construction of recombinant plasmids. The plasmids described in Table 1 were prepared as follows. pER1-992 was prepared by using an 11.5-kb gel-purified *Eco*RI fragment from the EAF plasmid containing ORFs *bfpA* to *bfpJ* ligated to the *Eco*RI site of the polylinker in pUC18; the resulting recombinant plasmid was transformed into *E. coli* DH5α. pER1-483 was derived similarly to pER1-992 in pUC18 and contains a 7.5-kb fragment with an *Eco*RI site only at the 5′ end of the insert; this fragment encompasses *bfpA* through the middle of *bfpF*. pISP was constructed by ligating a PCR-amplified fragment of *bfpP* with adapted *Eco*RI and *Pst*I ends to the *Eco*RI-*Pst*I-digested expression vector pKK233-3 (Pharmacia Biotech Inc., Uppsala, Sweden) (9). pBTA-BH1 contains a 3.9-kb *Bam*HI EAF plasmid fragment with *bfpT* (100), inserted into a *Bam*HI site of pWKS130 (105). pISD, pISE, and pISF were constructed by the PCR amplification of *bfpD*, *bfpE*, and *bfpF*, respectively, using *Eco*RI and *Hin*dIII restriction sites engineered at the 5′ and 3′ ends of each gene; the resulting fragments were then ligated to the unique *Eco*RI and *Hin*dIII sites of the polylinker of the T7 polymerase expression vector pET23a+ and transformed into *E. coli* BL21(DE3) (Novagen Inc., Madison, Wis.). *bfpD*, *bfpE*, and *bfpF* were also prepared with *Eco*RI sites engineered at both the 5′ and 3′ ends by PCR amplification; these were then ligated to the unique *Eco*RI site of pKK233-3, a high-level expression vector. pIS2 was constructed by gel purification of the 11.5-kb *Hin*dIII-digested EAF fragment carrying *bfpB* through *bfpM*, ligation of the fragment to *Hin*dIII-digested pET23a+, and transformation into *E. coli* BL21(DE3).

pISJ, pISK, and pISL (Table 1) were constructed in the following manner. pISJ was constructed with a 10.9-kb fragment encompassing nucleotides (nt) 2267 to 13179 (Fig. 1) from the EAF plasmid that had been digested with *Eco*RI; the resulting fragment was gel purified and ligated to *Eco*RI-digested vector pBC KS+ (Stratagene, La Jolla, Calif.), transformed into EPEC strain B171-4, and selected for chloramphenicol resistance. pISK was constructed by digesting pISJ with *Aat*II and *Sma*I; the resulting fragment was gel purified and ligated to a gel-purified 523-bp EAF plasmid fragment resulting from double digesting the EAF fragment in pISB2 with *Aat*II and *Ssp*I. pISB2 was prepared from an EAF plasmid *Bam*HI library and contains a ca. 9-kb fragment delimited by nt 11763 and 20565 (Table 1 and Fig. 2). The product of this ligation, pISK, therefore carries an insert of 11.3 kb from nt 2267 (an *Eco*RI site of the EAF plasmid) to 13571 (a *Ssp*I site of the EAF plasmid; Fig. 2). pISL was constructed as follows. pISB2 was digested with *Aat*II and *Eco*RV, and the gel-purified 1,333-bp fragment was ligated to *Aat*II- and *Sma*I-digested pISJ and transformed into EPEC strain B171-4. pISL therefore carries a 12.1-kb EAF plasmid insert extending from the *Eco*RI site at nt 2267 to the *Eco*RV site at nt 14381 (Table 1 and Fig. 2). pISJ, pISK, and pISL were examined for the presence of the correct insert by hybridization with appropriate probes. In addition, the junction points between the EAF plasmid fragments that constitute these inserts and between the inserts and the plasmid vector were verified by nucleotide sequencing.

pISM was constructed by PCR amplifying a 13.9-kb fragment of the EAF plasmid from EPEC strain B171-8. The amplification was carried out with a Takara LA PCR kit (PanVera Corp., Madison, Wis.) according to the manufacturer's instructions. The amplified fragment is delimited by nt 1150 and 15070

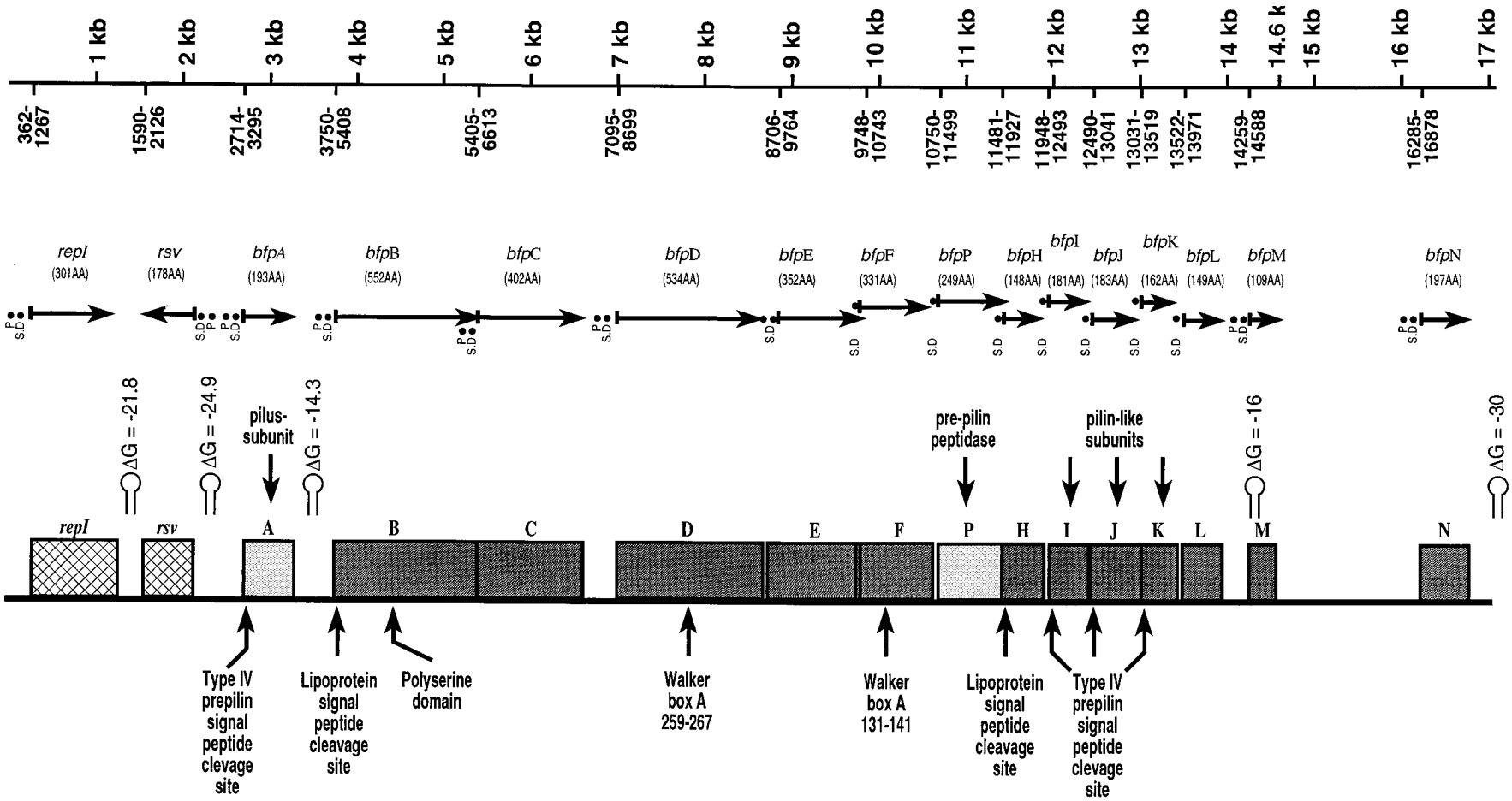

FIG. 1. Schematic representation of the organization of the *bfp* gene cluster locus on the EAF plasmid of EPEC strain B171-8. The 16 ORFs identified by computer analysis of a 17-kb segment of the EAF plasmid are indicated at the bottom as boxes. The direction of transcription of each ORF is depicted by a horizontal arrow, and the positions of putative promoter motifs (P) and Shine-Dalgarno (S.D) sequences are shown. Putative Rho-independent transcription terminators (stem-loop structures) are indicated together with the calculated free energy. Above the transcript arrows, the paired numbers correspond to the DNA base pairs at the beginning and end of each ORF. *rep*I and *rsv*, to the left of the *bfp* gene cluster, are predicted to encode RepI and resolvase, respectively, proteins required for the replication of plasmid members of the IncF incompatibility group.

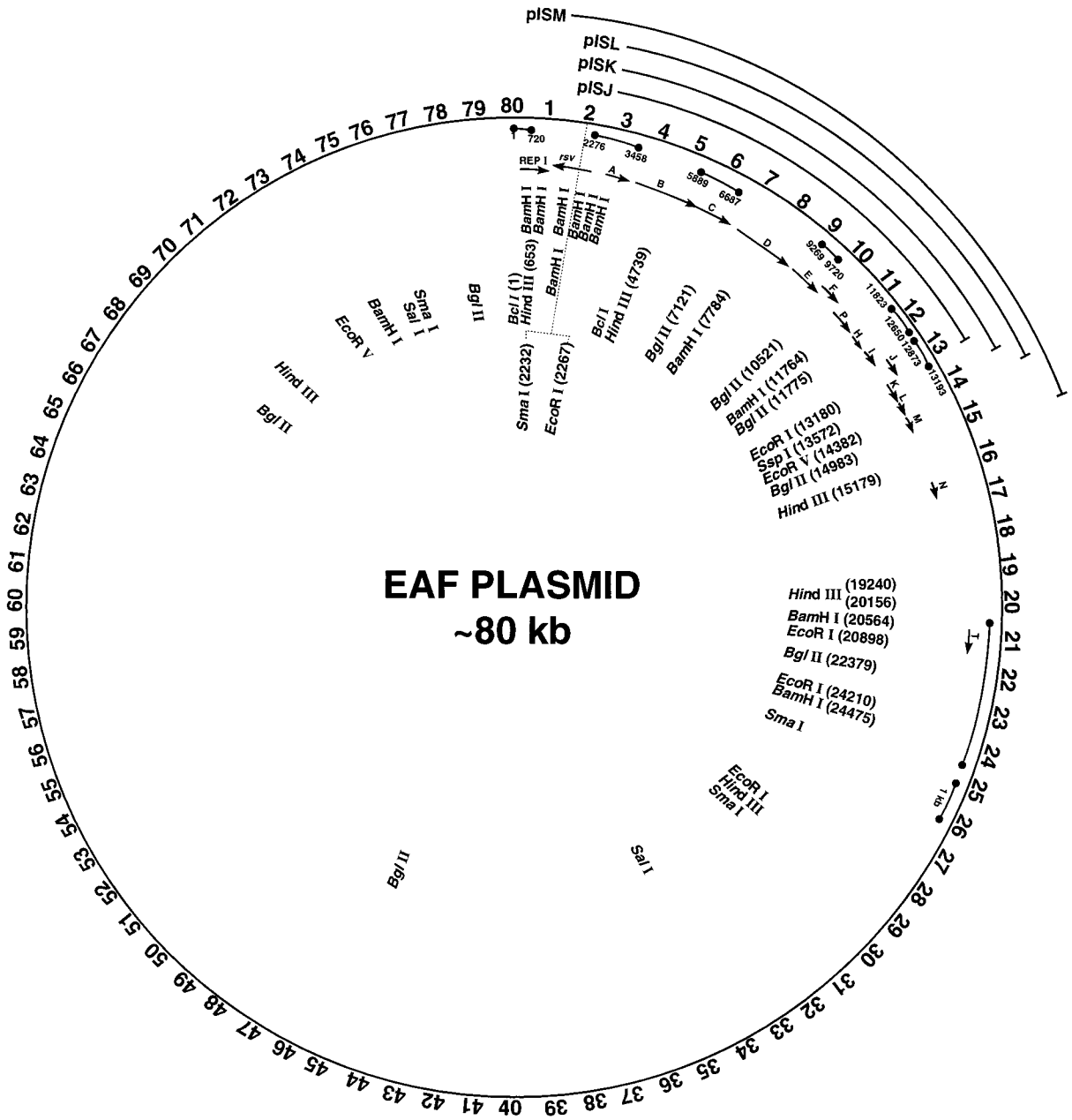

FIG. 2. EPEC strain B171-8 EAF plasmid map showing the relative positions of the *bfp* gene cluster, the *repI*, *rsv*, and *bfpT* loci, and the 1-kb EPEC diagnostic probe. Mapping was performed by Southern blot analysis using radiolabeled probes prepared from PCR-amplified fragments of structural genes (● - ●) and the purified 1-kb EAF fragment from plasmid pJPN16 (59). The arrows indicate the direction of transcription of the genes. Position 1 on the EAF map corresponds to the first *Bcl*I restriction site upstream of *repI*.

and thus encompasses *rsv* through *bfpM* (Fig. 2). Both the 5′ and 3′ primers were engineered with *Sal*I restriction sites, and accordingly, the amplified fragment was cloned into the *Sal*I-digested vector pUC18 (New England Biolabs). The resulting plasmid was transformed into EPEC strain B171-4 and selected for ampicillin resistance.

Southern blot analysis of the *bfp* gene cluster. EAF plasmid DNA was isolated as described above and digested to completion with *Bam*HI, *Bcl*I, *Bgl*II, *Eco*RI, *Hin*dIII, *Sal*I, or *Sma*I. Two micrograms of digested DNA was separated by electrophoresis, blotted onto a nylon membrane (Amersham), and hybridized to the labeled probes shown in Fig. 2. The blotted filters were incubated for 2 h at 42°C in a prehybridization solution containing 6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate), 5× Denhardt's solution (0.2% Ficoll, 0.2% polyvinylpyrrolidone, 0.2% bovine serum albumin [BSA]), 0.5% (wt/vol) SDS, and 0.1 mg of calf thymus DNA per ml. Hybridization was carried out in a fresh solution at 65°C for 12 to 16 h with 20% formamide, using 10^6 cpm of one of the radioactive probes described above. Filters were washed at 65°C, first in 1× SSC–0.1% SDS and then in 0.1× SSC–0.1% SDS, and subjected to autoradiography.

Western blot (immunoblot) analysis. Bacteria were grown in DMEM under *bfpA*-inducing conditions (75, 103) and centrifuged, the pellet was resuspended in water, the bacteria were lysed by sonication, and the total protein concentration was determined with a bicinchoninic acid protein assay reagent kit (Pierce, Rockford, Ill.), using BSA as the standard. Bacterial lysates (20 μg of protein) were denatured by boiling 5 min in 5% SDS–0.1% β-mercaptoethanol, the proteins were separated by SDS-PAGE (15% gel) (46), and the resolved proteins were then electrophoretically transferred to an Immobilon-P membrane (Millipore Co., Bedford, Mass.). The membranes were blocked and incubated with the indicated dilutions of a BFP-specific antiserum (29, 93), and the antibody-bound proteins were then detected with a horseradish peroxidase-labeled goat anti-

TABLE 2. Proteins homologous with predicted *bfp* gene cluster products[a]

Function	Source	Gene cluster	Predicted protein product(s)													Reference(s)
Pilin biogenesis	EPEC B171	*bfp*	A	B	C	D	E	F	P	H	I	J	K	L	M	This study
	Vibrio cholerae	*tcp*	A	C		T	E	E	J							25, 39, 40, 66, 67
	Pseudomonas aeruginosa	*pil*	A			B, T	C	U, T	D		V	V	V			3, 62, 63, 94, 96, 103
	Neisseria gonorrhoeae	*pil*	E			F			D			A				49, 56
	Dichelobacter nodosus	*fim*	A, B									A, B	B			34
Type II membrane traffic wardens	*Klebsiella oxytoca*	*pul*		D		E	F	E	O		H	I, J				17, 18, 77, 74, 81
	P. aeruginosa	*xcp*				R	S	R	A		T	T, V				7, 64, 65
	Erwinia spp.	*out*		D		E	F	E	O		H	I, G				14, 33, 54, 80
	Xanthomonas campestris	*xps*		D		E	F	E			G	I				21, 36
DNA uptake	*Bacillus subtilis*	*com*				G1	G2	G1	C							1, 57

[a] The homologous relationships shown are based on the presence of the following conserved structural and functional motifs, as judged from amino acid sequence similarity and in some cases from biochemical evidence for a conserved functional role. Each column is explained below by reference to the corresponding Bfp protein. BfpA proteins are type IV pilin subunit proteins, each with a conserved N-terminal prepilin domain recognized by a prepilin peptidase. Each of these proteins has been biochemically identified as the major component of a pilus polymeric filament. BfpB proteins are lipoproteins involved with the type II, *sec*-dependent secretion system, each containing the lipoprotein peptidase processing N-terminal motif. In addition, BfpB and OutD of *Erwinia* spp. both contain an internal polyserine domain of unknown function. No homologies with BfpC were identified. BfpD proteins are type II, *sec*-dependent proteins with conserved Walker box A motifs, conserved cysteine residues, and less well conserved Walker box B domains. All are hydrophilic and devoid of a signal peptide consensus sequence, suggesting that they are nucleotide-binding cytoplasmic proteins that provide energy by ATP hydrolysis. BfpE exhibits between 21 and 28% identity (47% similarity to PulF) with type II, *sec*-dependent proteins that lack signal peptides but appear to be associated with the cytoplasmic membrane. BfpF, like BfpD, contains a highly conserved Walker box A but lacks the conserved cysteine and Walker box B motifs found in the listed BfpD homologs. BfpP is highly homologous in structure and function with other type IV prepilin peptidases. BfpH, other than having a conserved lipoprotein processing site, is not homologous with other type II, *sec*-dependent proteins. BfpI, -J, and -K contain type IV pilin-like proteins with conserved N-terminal prepilin processing motifs and two cysteine residues in the middle to last one-third of the molecule, but without biochemical evidence for forming part of the pilus polymer. BfpL and BfpM are unrelated to other proteins in the GenBank database.

rabbit antibody as instructed for the ECL (enhanced chemiluminescence) Western blotting detection system (Amersham).

Electron microscopy. Transmission electron microscopy (TEM) was conducted with negatively stained (1% phosphotungstic acid [pH 7.5]) bacteria that had been cultivated overnight at 37°C on TSA blood agar. Immunogold electron microscopy was carried out with bacteria that had been grown in DMEM at 37°C for 4 h and then induced with 1 mM isopropylthiogalactopyranoside (IPTG) for 2 h. The bacteria were transferred to carbon-coated grids and incubated for 30 min in phosphate-buffered saline (PBS)–1% (wt/vol) BSA containing a 1/400 dilution of a BFP-specific rabbit antiserum. The grids were washed and then incubated with goat anti-rabbit immunoglobulin G conjugated with 30-nm gold particles for 30 min. After being washed with PBS, the bacteria were stained with 1% phosphotungstic acid and the grids were examined by TEM.

Nucleotide sequence accession number. The nucleotide sequence of the 17-kb EAF plasmid *bfp* locus of EPEC strain B171-8 has been deposited with GenBank with accession number U27184.

RESULTS

Identification and organization of a *bfp* gene cluster. To identify genes required for the biogenesis of the BFP fiber, approximately 17 kb of EAF plasmid DNA sequence was obtained in the region flanking *bfpA*. Computer-assisted analysis of this sequence revealed 16 ORFs which fulfill the following criteria: an ORF that would encode a polypeptide of 100 or more translated amino acids; ATG as the translational initiation codon; and an *E. coli* consensus Shine-Dalgarno sequence, located at an optimal distance upstream of the ATG (92). These ORFs are depicted in Fig. 1. Potential ORFs with start codons other than ATG and ORFs predicted to encode proteins of less than 100 amino acids were also identified but are not included in this analysis.

Upstream of *bfpA*, we found two ORFs that are homologous with genes required for the maintenance and replication of several other *E. coli* and *Salmonella* plasmids: *repI* is predicted to encode replicase I (RepI), a component of the replication region of the IncF plasmid group (71, 72, 90); *rsv* is predicted to encode resolvase, a protein that affects the resolution of cointegrates formed by plasmid replication (5, 13, 44, 47). Because *repI* and *rsv* may be components of the EAF plasmid replicon, as discussed below, sequence obtained during the course of this study was numbered beginning with a *Bcl*I restriction site located 361 bp upstream from the *repI* translation initiation codon; this numbering system is used in the physical map shown in Fig. 1, in the EAF plasmid map shown in Fig. 2, and in the sequence data deposited in GenBank.

Downstream of *bfpA*, 13 additional ORFs were identified and are designated alphabetically *bfpB* through *bfpN*, with one exception: the ORF between *bfpF* and *bfpH* is designated *bfpP* instead of *bfpG* because its homolog was previously cloned from EPEC strain 2348/69, serotype O127:H6, by Zhang et al. (110) and named *bfpP* because it was found to encode the prepilin peptidase. We have retained their nomenclature for this ORF. As demonstrated below, *bfpA*, together with the next 11 ORFs (i.e., *bfpA* through *bfpL*; Fig. 1) can encode a morphologically normal BFP filament when cotransformed with *bfpT*, a positive transcriptional regulator of *bfpA* that is located on the EAF plasmid outside the contiguous cluster of genes discussed in this report (100) (Fig. 2). Thus, for the purpose of this study, the *bfp* gene cluster is defined as the 11,257-bp segment of the EAF plasmid that encompasses at least 12 contiguous ORFs delimited by the translation initiation codon of *bfpA* on the left (corresponding to nt 2714; Fig. 1) and by the stop codon of *bfpL* on the right (corresponding to nt 13971; Fig. 1).

Three of the 12 ORFs analyzed here that comprise the *bfp* gene cluster are known to be required for the biogenesis of the BFP fiber: *bfpA* encodes the major structural pilus subunit (20, 93), *bfpB* encodes a lipoprotein required for the export and/or assembly of the signal peptidase-processed form of BfpA (79), and *bfpP* encodes the prepilin signal peptidase that processes BfpA (reference 110 and this report). The remaining nine ORFs were assigned to the *bfp* gene cluster because they are contained within a fragment that can direct the production of the BFP filament, and their orientation and spacial juxtaposition suggest that they could be part of the same transcriptional unit. In addition, as summarized in Table 2, the predicted products of some of these ORFs are homologous with proteins believed to be required for the biogenesis of type IV pili by other bacterial species (97), with proteins that comprise the

type II *sec*-dependent general secretory pathway involved in the secretion of enzymes and/or toxins by diverse gram-negative bacteria (35, 76, 86), or with proteins that mediate DNA uptake (1, 2, 48, 57, 101, 106). *bfpM* and *bfpN* are not included in the *bfp* gene cluster because they are not required for BFP filament production (as shown below) and their predicted protein sequences do not share homology with the accessory proteins of other members of the type II *sec*-dependent general secretory pathway family.

Each of the 12 analyzed ORFs of the *bfp* gene cluster was predicted by sequence analysis to be transcribed in the same direction and to be preceded by a ribosome binding site. From *bfpD* through *bfpL*, the stop codon of each ORF and the start codon of the next ORF either overlap by between 3 and 18 nt (*bfpE-F*, *bfpP-H*, *bfpI-J*, and *bfpJ-K*; Fig. 1) or are separated by only 3 to 21 nt (*bfpD-E*, *bfpF-P*, *bfpH-I*, and *bfpK-L*; Fig. 1). The *bfpA* stop codon is followed by an inverted repeat that might form a stem-loop structure ($\Delta G = -14.3$ kcal [1 cal = 4.184 J]) that could function as a potential Rho-independent transcriptional terminator or attenuator (Fig. 1). At the end of the *bfp* gene cluster, 287 bp separate *bfpM* from the tightly grouped ORFs *bfpD* through *bfpL*, and downstream from the translational stop of *bfpL* is a potential Rho-independent terminator ($\Delta G = -16$ kcal; Fig. 1).

Expression of BFP filaments by the *bfpA-bfpL* cluster of genes. To determine the downstream boundary of the *bfp* gene cluster, a series of deletions was made to identify the last ORF in the cluster required for BFP filament production. For this purpose, four constructs were prepared (Fig. 2 and Table 1). pISJ carries a 10,912-bp EAF plasmid fragment delimited by nt 2267 and 13179 (Fig. 1) which includes the entire ORFs *bfpA* through *bfpJ* and part of *bfpK*. pISK carries an 11,305-bp EAF plasmid fragment spanning nt 2267 to 13572 which includes full-length ORFs from *bfpA* through *bfpK* and part of *bfpL*. pISL carries a 12,115-bp EAF plasmid fragment corresponding to nt 2267 through 14382 which encompasses all of *bfpA* through *bfpL* and part of *bfpM*. pISM carries a 13.9-kb PCR-amplified fragment of the EAF plasmid corresponding to nt 1150 to 15070 comprising *rsv* through *bfpM* (Fig. 1 and 2). Each of these constructs was transformed into the EAF plasmid-cured strain B171-4 (Table 1) together with the compatible plasmid pBTA-BH1, which contains *bfpT*, an EAF plasmid gene that codes for a *trans*-acting transcriptional activator of *bfpA* expression (100). The four resulting recombinant strains were analyzed for BfpA pilin expression by Western blotting and for the production of BFP filaments by TEM.

The wild-type EPEC parent strain B171-8 served as the BFP-positive control for these experiments. It produced BFP filaments and signal peptidase-processed pilin (Fig. 3A and H, lane 1). Serving as an EAF plasmid-negative control, B171-4 was found to produce neither BFP filaments nor pilin (Fig. 3D and H, lane 2). In contrast, B171-4 carrying pISJ, pISK, pISL, and pISM each produced processed pilin (data not shown and Fig. 3H, lanes 3 to 5). However, while pISL and pISM also produced BFP filaments (Fig. 3E and G), pISJ and pISK did not (Fig. 3F and data not shown). Thus, *bfpL*, but not the complete *bfpM* ORF, is required for the biogenesis of the BFP filament, suggesting that the 3′ boundary of the *bfp* gene cluster is the termination codon of *bfpL* and that the *bfp* gene cluster encompasses approximately 14 kb. This deletion study also suggests that *bfpJ*, *bfpK*, and *bfpL* are necessary for the biogenesis of the BFP filament and that they act as steps that follow signal peptidase processing of BfpA.

The experiments described above were conducted with an EPEC host strain, B171-4, that had been cured of the EAF plasmid. To determine if the *bfp* gene cluster could also direct the synthesis of BFP filaments in a non-EPEC *E. coli* host, two compatible plasmids (pIS1 and pIS2; Table 1), which together contain *repI* through *bfpM*, were cotransformed into the non-EPEC strain BL21(DE3). pIS1 is a derivative of pACYC184 that carries a 10-kb *Sau*3AI fragment encompassing *repI* to the middle of *bfpF*. pIS2 is a derivative of pET23a+ that carries a 11.5-kb *Hin*dIII fragment encompassing *bfpB* to *bfpM*. A third compatible plasmid, pBTA-BH1, carrying *bfpT* on a 3.9-kb *Bam*HI fragment of the EAF plasmid (Table 1), was also transformed to provide the positive transcriptional regulator for *bfpA* expression (100).

As in the deletion experiments described above, expression of the processed form of the BfpA protein by this recombinant strain was demonstrated by Western blotting (data not shown). Immunogold TEM of the negatively stained recombinant using a BFP-specific antiserum showed colloidal gold-labeled pilus fibers (Fig. 3C), indicating that a contiguous series of genes (*repI* through *bfpM*; Fig. 1) together with *bfpT* can direct BFP filament biogenesis in a non-EPEC host. These results and the findings from the deletion study led us to concentrate on the *bfpA-bfpL* cluster of genes in the remaining sections of this report.

***bfpB* and *bfpH* code for lipoproteins.** *bfpB*, the second ORF of the *bfp* gene cluster, is located to the right of *repI*, *rsv*, and *bfpA* (Fig. 1). Transcribed in the same direction as *bfpA*, it contains 1,656 bp and is predicted to encode a 552-amino-acid, 58.2-kDa protein. Analysis of this sequence with the Prosite database of structural motifs revealed a prokaryotic membrane lipoprotein lipid attachment site, including a precursor signal peptide, at the N-terminal region of BfpB. This leader sequence conforms closely to the lipoprotein signal peptidase recognition site including the presence of the essential cysteine to which a glyceride fatty acid lipid would be attached. Intrinsic labeling experiments in our laboratory have proven BfpB to be a lipoprotein, given its capacity to incorporate [^{3}H]palmitate during growth (79). Between residues 216 and 243 of the deduced BfpB amino acid sequence is a centrally located polyserine domain (Fig. 1) predicted by the Chou-Fasman algorithm (12) to have a high β-turn probability; the function of this region is unknown, but it might serve to separate N-terminal and C-terminal functional domains.

bfpH, the seventh ORF downstream of *bfpA*, is also predicted to encode a lipoprotein, given the presence of a characteristic N-terminal cysteine lipid attachment site. However, unlike most other prokaryotic membrane lipoproteins, it does not contain a conserved lysine or arginine within the first seven amino acids of the signal peptide. The possibility that mature BfpH incorporates a covalently linked fatty acid is under investigation.

***bfpC* is predicted to encode a 45.4-kDa protein.** *bfpC* comprises 1,209 bp and would encode a 402-amino-acid, 45.4-kDa protein; no signal peptide motif is evident. A GenBank database search for homologous sequences showed that *bfpC* is 51% identical at the nucleotide sequence level with *tcpT* of *V. cholerae*, but no significant amino acid sequence homology was detected with this or any other protein in the database. Because of this disparity and evidence that TcpT may be required for biogenesis of the type IV Tcp pilus filament (10, 39, 40, 66, 67), the sequencing gels were reexamined for the possibility that an erroneous reading may have mistakenly deleted or added a base, leading to an apparent frameshift. However, no such error was detected, and translation of the sequence in each of the remaining five reading frames failed to disclose an amino acid sequence with homology to TcpT.

***bfpD* through *bfpL* comprise a cluster of nine closely spaced ORFs.** Following an intergenic segment of 482 bp containing

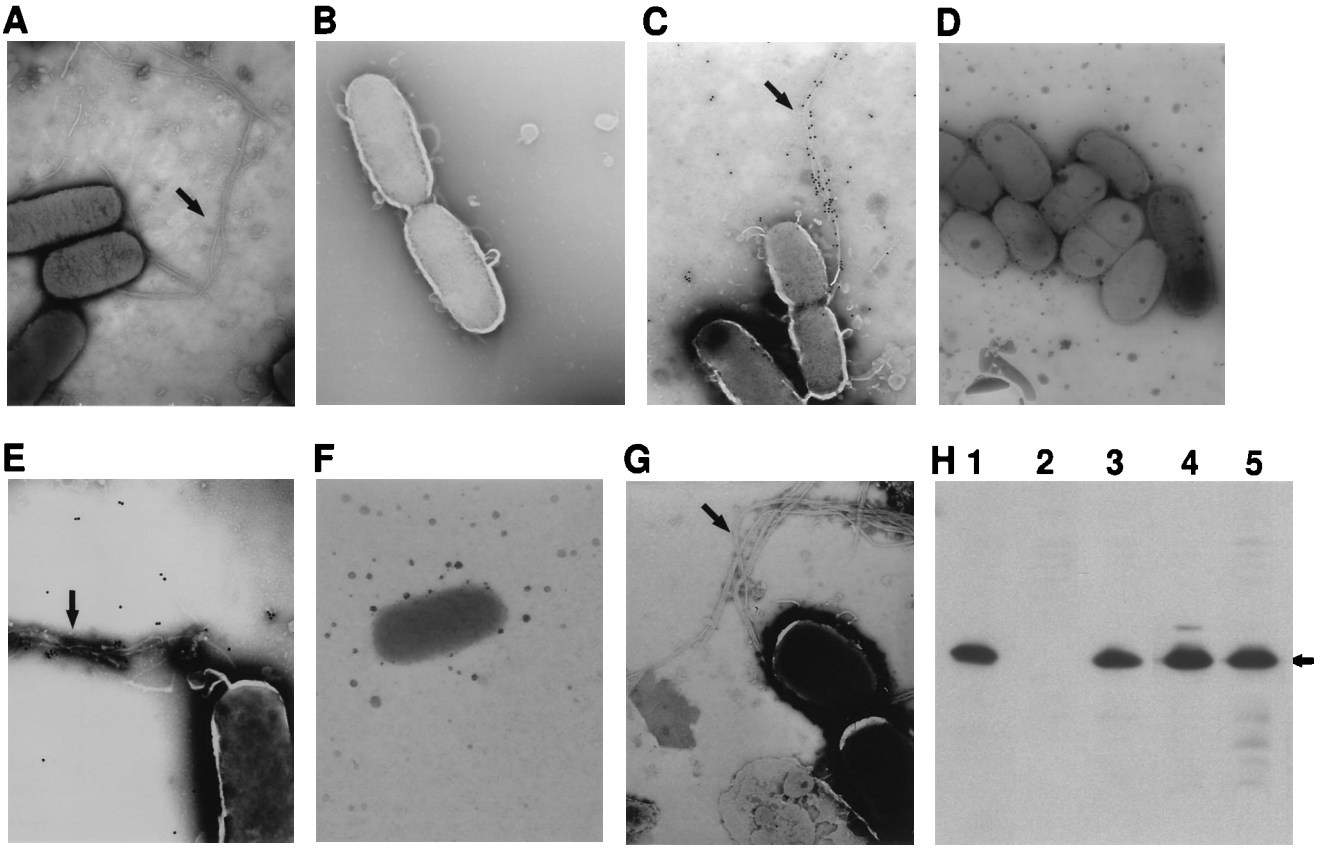

FIG. 3. Expression of the BfpA protein and BFP fibers by the *bfp* gene cluster. After overnight growth at 37°C on TSA blood agar, the bacteria were examined for production of the BfpA subunit by Western blotting and for the production of BFP filaments by TEM. (A) TEM of a negatively stained sample of EPEC strain B171-8 (serotype O111:NM) showing BFP filaments. (B) Negatively stained specimen of the *bfp* gene cluster-negative *E. coli* strain BL21(DE3) carrying plasmid pET23a+ (Table 1) showing no pilus filaments. (C) *bfp* gene cluster-positive recombinant strain BL21(DE3) carrying pIS1, pIS2, and pBTA-BH1. Immunogold TEM of this negatively stained specimen shows gold-labeled BFP. (D) TEM of a negatively stained sample of EPEC strain B171-4 (EAF plasmid cured) showing no BFP. (E) TEM of B171-4 carrying pISM and pBTA-BH1. Production of BFP is demonstrated by immunogold TEM. (F) TEM of B171-4 carrying pISK and pBTA-BH1. No BFP are seen. (G) TEM of B171-4 carrying pISL and pBTA-BH1 showing production of BFP. (H) Production of the BfpA subunit detected by Western blotting of bacterial lysates, using a BFP-specific antiserum. Each lane was loaded with 20 μg of total protein. Lanes: 1, wild-type EPEC strain B171-8 showing the production of the processed 19-kDa BfpA subunit; 2, EPEC strain B171-4 (EAF plasmid cured) showing no production of BfpA; 3, EPEC strain B171-4 carrying pISK and pBTA-BH1; 4, EPEC strain B171-4 carrying pISL and pBTA-BH1; and 5, EPEC strain B171-4 carrying pISM and pBTA-BH1. The arrow indicates the position of the band specifically recognized by the anti-BfpA antiserum.

an upstream ribosome binding site at an optimal distance from the *bfpD* translational initiation codon, there is a succession of nine ORFs, each predicted to be transcribed in the same orientation and which either overlap or are separated by 21 or fewer nt (Fig. 1).

bfpD comprises 1,605 bp and is predicted to encode a 534-amino-acid, 60.5-kDa protein. Proof that this ORF can express a protein of the predicted molecular mass was sought by cloning *bfpD* in pKK233-3, a high-level expression vector, and in pET23a+, the inducible T7 promoter-RNA polymerase expression system described in Materials and Methods. Following induction by IPTG, SDS-PAGE analysis of *E. coli* DH5α carrying *bfpD* in pKK233-3 (data not shown) and *E. coli* BL21(DE3) carrying *bfpD* cloned into pET23a+ (pISD; Table 1) revealed production of a new protein with an estimated molecular mass of 61 kDa, in good agreement with the predicted molecular mass of BfpD (Fig. 4A, lane 5).

Analysis of the BfpD amino acid sequence showed a highly conserved Walker box A motif (104) between residues 259 and 267 (Fig. 1 and 5) and the associated Asp boxes and four conserved cysteine residues characteristic of nucleotide-binding proteins (Fig. 5). A poorly conserved Walker box B is also evident just beyond the cluster of conserved cysteines (Fig. 5). Further analysis of BfpD showed a relatively hydrophilic Kyte-Doolittle hydropathy profile (data not shown) and the absence of a signal peptide consensus sequence, features that suggest that it may be localized to the cytoplasm of the bacteria. A search of the GenBank database revealed that BfpD is homologous with proteins involved in type IV pili biogenesis, protein secretion, and DNA uptake (Table 2).

bfpE comprises 1,059 bp and is predicted to encode a 352-amino-acid, 39.8-kDa protein. Proof that *bfpE* can direct the synthesis of a protein of this mass was sought by using the T7 promoter expression system described above. Following IPTG induction and SDS-PAGE analysis, the production of a novel, 34.7-kDa polypeptide could be detected by Coomassie blue staining (Fig. 4A, lane 7).

Examination of the BfpE deduced amino acid sequence showed no signal peptide consensus sequence but four prominent hydropathic segments, located between residues 116 and 150, 162 and 187, 205 and 230, and 324 and 342 (data not shown), that could be membrane-spanning domains. These features suggest that BfpE may be an integral cytoplasmic membrane protein. A search of the databases identified pro-

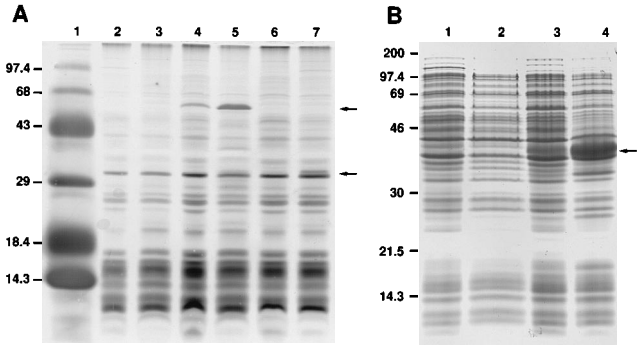

FIG. 4. Protein expression by ORFs *bfpD*, *bfpE*, and *bfpF*. *E. coli* BL21(DE3) harboring the T7 expression vector pET23a+ containing the PCR-amplified fragment corresponding to *bfpD* (pISD), *bfpE* (pISE), or *bfpF* (pISF) was cultivated overnight in LB broth at 37°C, and the bacteria were then transferred to fresh LB broth and grown at 37°C for 4 h before being induced by the addition of 1 mM IPTG for 2 h. Bacterial lysates (25 μg of total protein per lane) were analyzed by SDS-PAGE (15% gel), and the protein bands were visualized by Coomassie blue staining. The estimated molecular masses (in kilodaltons) are indicated at the left of each gel, and the positions of the ORF-specific expressed polypeptides are indicated by arrows on the right. (A) Lanes: 1, protein size standards; 2, *E. coli* BL21(DE3); 3, BL21(DE3)/pET23a+; 4, BL21(DE3)/pISD, uninduced; 5, BL21(DE3)/pISD, induced with IPTG; 6, BL21(DE3)/pISE, uninduced; 7, BL21(DE3)/pISE, induced with IPTG. (B) Lanes: 1, BL21(DE3); 2, BL21(DE3)/pET23a+; 3, BL21(DE3)/pISF, uninduced; 4, BL21(DE3)/pISF, induced with IPTG. (Note that uninduced cultures express low levels of BfpD and BfpF.)

teins homologous with BfpE that are involved in the biogenesis of type IV pili, protein secretion, and DNA uptake (Table 2).

bfpF comprises 996 bp and is predicted to encode a 331-amino-acid, 36.7-kDa protein. To determine if *bfpF* can encode a polypeptide of this expected mass, a subclone of *bfpF* was prepared in the T7 expression system. Upon IPTG induction, expression of a new polypeptide with an estimated molecular mass of 36.5 kDa was detected by SDS-PAGE and Coomassie blue staining (Fig. 4B, lane 4). Like BfpD, BfpF contains a Walker box A nucleotide-binding motif located between residues 131 and 141 (Fig. 1 and 5). BfpF also contains two Asp boxes but lacks the conserved four cysteine residues and the Walker box B motif found in BfpD.

BfpF was found to be homologous with the TcpE protein of *V. cholerae* and the PilT protein of *Pseudomonas aeruginosa* (Table 2). PilT is required for pilus-mediated twitching motility, possibly providing energy through ATP hydrolysis for the depolymerization-repolymerization of pilin subunits (16, 109).

The deduced amino acid sequence of the next downstream ORF was found to be identical with the prepilin signal peptidase cloned from EPEC strain E2348/69 (O127:H6) by Zhang et al. (110), who designated it *bfpP*. The *bfpP* homolog identified in the *bfpA* gene cluster of EPEC strain B171-8 commences 7 nt downstream from the *bfpF* termination codon, is preceded by a probable ribosome binding site, and consists of 750 bp that would encode a 249-amino-acid, 28-kDa protein. Analysis of this sequence by the Kyte-Doolittle (45) and Chou-Fasman (12) algorithms predicted eight hydrophobic β sheets; this finding and the absence of a signal peptide consensus sequence indicate that BfpP may be an integral protein of the inner membrane. Comparison of the deduced amino acid sequence of BfpP with entries in the GenBank database showed that it is 40% identical to TcpJ (the *V. cholerae* toxin-coregulated prepilin peptidase [39]), 44% identical to the prepilin peptidase (PilD) of *Neisseria gonorrhoeae* (49), and 41% identical to PilD of *P. aeruginosa* (63) (Table 2).

Proof that the BfpP protein of EPEC strain B171-8 functions as a prepilin peptidase for BfpA was sought by determining if subclones containing *bfpP* would process BfpA from the 22-kDa precursor protein to the 19-kDa mature pilus subunit (Fig. 6). A 11.5-kb *Eco*RI fragment of the EAF plasmid was cloned in pUC18, yielding pERI-992 (Fig. 6B). This fragment, extending from 447 bp upstream of the *bfpA* translation initiation codon to the middle of *bfpJ* (Fig. 1 and 6B), resulted in normal processing of the BfpA subunit (Fig. 6A, lane 3) after induction with IPTG compared with wild-type EPEC strain B171-8, which contains an intact EAF plasmid and processes BfpA normally (Fig. 6A, lane 4). In contrast, the same plasmid vector containing a smaller, 7.5-kb fragment of the EAF plasmid (pERI-493), which extends from the same nucleotide upstream of *bfpA* but ends in the middle of *bfpF* (Fig. 6B), expresses unprocessed pilin (Fig. 6A, lane 1). These findings pointed to *bfpP*, *bfpH*, or *bfpI* as the prepilin peptidase gene. To identify which of these ORFs is responsible for the processing activity of pERI-992, three additional constructs were prepared. The first of these, pIS1, which contains a 10-kbp EAF plasmid fragment that extends from sequences upstream of the natural promoter of *bfpA* through the first one-half of *bfpF* (Fig. 6B), produced neither prepilin nor pilin (Fig. 6A, lane 5). Upon addition of a second compatible plasmid (pBTA-BH1) containing *bfpT*, which codes for a transcriptional activator for *bfpA* (100), unprocessed pilin was produced (Fig. 6A, lane 6). Finally, addition of a third compatible plasmid (pISP; Fig. 6B) containing the PCR-amplified *bfpP* coding sequence in an inducible expression vector led to the production of processed pilin (Fig. 6A; lane 7) after IPTG induction. These results and the homology of BfpA to other type IV prepilin peptidases indicate that BfpP is the BfpA prepilin peptidase of EPEC strain B171-8 and confirm the findings of Zhang et al. (110).

BfpI, BfpJ, and BfpK are type IV pilin-like proteins. *bfpI*, *bfpJ*, and *bfpK* are predicted to code for proteins ranging from 162 to 183 amino acids that share four common structural features with BfpA and other type IV pilins: similar sizes of the polypeptides; a conserved prepilin peptidase processing site, including the presence of a glutamate residue at position +5 which is required for posttranslational modification of the mature protein's amino terminus (95) (Fig. 7); a conserved hydrophobic N-terminal region of 30 amino acids; and the presence of two cysteine residues (Fig. 7). These similarities in primary structure were supported by the presence of nearly identical hydropathy profiles through position 30 (Fig. 7). However, these sequences and the corresponding hydropathy profiles diverge markedly toward the carboxy termini of the four proteins.

BfpL is not homologous with other type IV pilus biogenic proteins. *bfpL* is predicted to encode a 149-amino-acid protein that is remarkable for the presence of five cysteine residues and for a hydrophobic N-terminal segment of 18 amino acids that is predicted by the Chou-Fasman algorithm (12) to comprise a β sheet. *bfpL* was identified as a component of the *bfp* gene cluster on the basis of the deletion experiments depicted in Fig. 3. In addition, its proximity to other *bfp* genes and its predicted transcriptional orientation suggested that it might be part of the same functional genetic unit. However, computer analysis of its deduced amino acid sequence failed to identify similarities with other proteins associated with type IV pilus assembly or regulation.

Location of the *bfp* gene cluster near a putative EAF plasmid replicon. Analysis of sequence upstream of *bfpA* revealed two ORFs that are components of plasmid replicons: *rsv* and *repI* (Fig. 1). *rsv* is located 587 bp upstream of *bfpA* and is transcribed in the opposite direction. The predicted start codon is preceded by potential −35 and −10 promoter sequences and

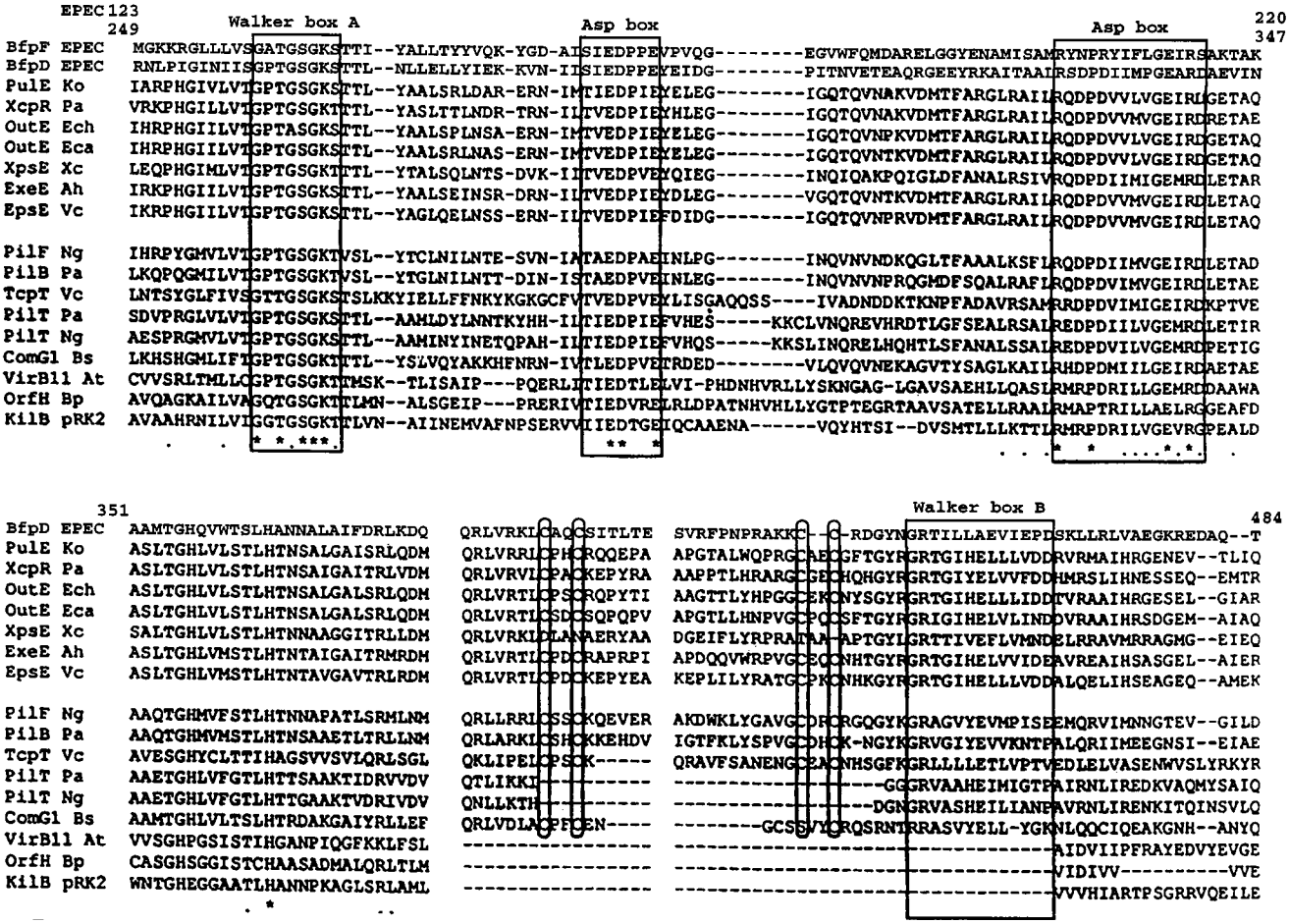

FIG. 5. Partial alignment of highly conserved regions around the putative nucleotide-binding domains of BfpD, BfpF, and related proteins. The Walker box A, Asp boxes, and Walker box B consensus sequences and the positions of the four conserved cysteine residues of BfpD are shown above the alignment (104). The numbers 123 and 220 on the top indicate the positions of the translated amino acids of BfpF, and the numbers 249 and 347 on the top and 351 and 484 on the bottom indicate the positions of the translated amino acids of BfpD. Identical residues in all proteins are indicated by asterisks, and conserved residues are indicated by dots. Dashes show the positions of gaps that were introduced to improve alignments. The following proteins were aligned: BfpF and BfpD (this study), PulE of *K. oxytoca* (Ko) (73), XcpR of *P. aeruginosa* (Pa) (7), OutE of *Erwinia chrysanthemi* (Ech) (54), OutE of *Erwinia carotovora* (Eca) (80), XpsE of *Xanthomonas campestris* (Xc) (21), ExeE of *Aeromonas hydrophila* (Ah) (38), EpsE of *V. cholerae* (Vc) (88), PilF of *N. gonorrhoeae* (Ng) (49), PilB of *P. aeruginosa* (Pa) (62), TcpT of *V. cholerae* (Vc) (40, 67), PilT of *P. aeruginosa* (Pa) (109), PilT of *N. gonorrhoeae* (Ng) (49), ComG1 of *B. subtilis* (Bs) (2), VirB11 of *A. tumefaciens* (At) (106), OrfH of *Bordetella pertussis* (Bp) (108), and KilB from plasmid pRK2 (58). This alignment is modified by the addition of BfpD and BfpF to an alignment of the other proteins prepared by Possot and Pugsley (74).

by a putative ribosome binding site located 21 bp upstream, not an optimal distance for effective expression (92). Consisting of 537 bp, *rsv* is predicted to encode a 178-amino-acid polypeptide that is 43% identical with the *Salmonella dublin* resolvase protein encoded by *rsd* on the 80-kDa pSDL2 virulence plasmid (44). Like resolvase of *S. dublin*, the predicted protein product of the *rsv* locus on the EPEC EAF plasmid contains a highly conserved C-terminal domain having residues in common with other site-specific recombinases of the integrase family (Fig. 8A).

Further upstream of *bfpA* and transcribed in the opposite direction as *rsv* is a 906-bp ORF that is predicted to encode a 301-amino-acid protein (Fig. 1). Analysis of the deduced amino acid sequence revealed significant homology with two related *E. coli* plasmid-encoded replicase proteins: 98% identity with RepI of pCO1V-K30 (71, 72) and 83% identity with RepF1B of the P307 replicon (90). These related sequences each contain the same helix-turn-helix motif that is characteristic of DNA-binding proteins (68). Because of its near identity with RepI, this EAF plasmid ORF was designated *repI*. Upstream of *repI* are putative ribosome binding site and promoter sequences, and on either side of *repI* are sequence motifs common to *E. coli* plasmid replicons (Fig. 8B). These include two putative DnaA-binding sites 219 and 311 bp upstream of the *repI* translation start codon; the second has perfect homology with the DnaA-binding consensus sequence (5′ TTAT[A/C]CA[A/C]A 3′), and the first is identical except in the first and sixth positions (27). In addition, the adenine methylation (*dam*) tetranucleotide 5′ GATC 3′ is found in high frequency in the *repI* upstream and downstream regions (55). This sequence motif is also commonly found within plasmid origins of replication. Eight 18-bp direct repeats are also noted in the *repI* upstream region and overlap the putative −35 and −10 RNA polymerase binding sites.

Downstream of the *repI* stop codon are two indirect repeats which could form a stem-loop structure for transcription termination. Also found in the downstream *repI* region are three copies of the 18-bp repeat 5′ ATAA(A/G)CTGT(A/G)GTAA 3′. Finally, four copies of a 19-bp repeat (5′ TACTNACTAC TACAGCTTATAT 3′) are located at the 3′ end of *repI*; they

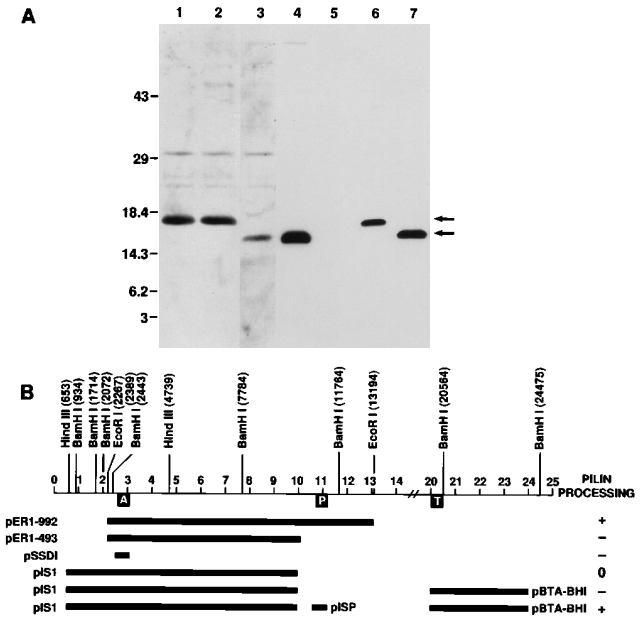

FIG. 6. Identification of *bfpP*, the gene coding for the BfpA prepilin peptidase of EPEC strain B171-8. (A) Western blot analysis of BfpA processing by cloned prepilin peptidase, BfpP. Bacteria were grown under inducing conditions in DMEM or in Luria broth with 1 mM IPTG at 37°C, and their lysates were analyzed by SDS-PAGE (15% gel) and Western blotting with a BFP-specific antiserum (29). Each of the recombinant plasmids was expressed in the non-EPEC strain DH5α. Lanes: 1, pERI-493; 2, pSSD1; 3, pERI-992; 4, EPEC strain B171-8; 5, pIS1 (derivative of the nonexpression vector pACYC184; Table 1); 6, pIS1 and pBTA-BH1; 7, pIS1, pBTA-BH1, and pISP. (Plasmid pBTA-BH1 carries *bfpT*, which codes for the transcription factor necessary for the expression of *bfpA*.) The numbers on the left denote the molecular masses (kilodaltons) of the BfpA polypeptides estimated from the migration of molecular weight standards (not shown). The arrows on the right indicate the locations of the preprocessed (top arrow) and processed (lower arrow) pilin polypeptides. (B) Schematic representation of the EAF plasmid fragments used to deduce the location of *bfpP*. A 24-kb DNA region of the EAF plasmid is shown; the *bfp* gene cluster, positions of restriction sites used to construct the recombinant clusters used in this analysis, and the locations of *bfpA*, *bfpP*, and *bfpT* are indicated. The dark horizontal bars beneath the map show the DNA fragments included in each plasmid. The symbols + and − denote the presence and absence of pilin processing to the correct size by BfpP; 0 indicates no detectable pilin subunit synthesis.

complement one of the direct repeats and thus form part of the above-hypothesized stem-loop structure. These sequences are part of the *incE* and *incF* incompatibility motifs found in *repI*-containing plasmids (28, 32, 102).

Physical map of the EAF plasmid with respect to the *bfp* gene cluster. To locate the *bfp* gene cluster on the EAF plasmid, a physical map of the ca. 80-kb EAF plasmid from EPEC strain B171-8 (O111:NM) was constructed (Fig. 2) by using 12 restriction endonucleases and eight probes that hybridize with *repI*, *rsv-bfpA*, *bfpC*, *bfpE*, *bfpH-bfpI*, or *bfpJ-bfpK* or with two loci beyond the sequence reported here: *bfpT*, which corresponds to the positive transcriptional activator of *bfpA* expression discussed above (100); and the 1-kb segment that corresponds to the EAF diagnostic probe described by Nataro et al. (59). *bfpT* was determined to be within a 3.9-kb *Bam*HI fragment and was located approximately 18 kb downstream of *bfpA* and oriented in the same transcriptional direction (100); the 1-kb noncoding EPEC diagnostic probe locus was located approximately 4 kb beyond *bfpT* or about 23 kb from *bfpA* (Fig. 2).

Comparison of the organizations of *bfp* gene clusters from seven EPEC serotypes. Eleven *E. coli* O-antigen EPEC serogroups have been identified as etiologic agents of diarrhea and found to contain an EAF plasmid and to exhibit the localized adherence and actin condensation phenotypes (50, 52). In this study, nine EPEC O:H serotypes were subjected to restriction fragment length analysis to determine if they were identical with respect to organization of their *bfp* gene clusters and their proximity to *repI* and *rsv*. Genomic DNA from these strains was digested with *Hin*dIII, and the fragments were analyzed by Southern blot hybridization (Fig. 9) using six probes corresponding to *repI*, *rsv*, *bfpA*, *bfpB*, *bfpC*, *bfpF*, and *bfpI-J*. According to the physical map of the *repI-rsv-bfp* gene cluster on the EAF plasmid of EPEC strain B171-8 (Fig. 1 and 2), the *Hin*dIII digest was expected to yield three fragments: a ca. 11,154-bp fragment (from nt ~69500 to 654), a 4,085-bp fragment (from nt 655 to 4739), and a ca. 11,061-bp fragment (from nt 4740 to ~15800). On the basis of the locations of the probe sequences with respect to these fragments (Fig. 1 and 2), hybridization was predicted to be as follows: the *repI* probe with the 11,154- and 4,085-bp fragments, the *rsv* and *bfpA* probes with the 4,085-bp fragment only, the *bfpB* probe with the 4,085- and 11,061-bp fragments, and the *bfpC* and *bfpI-J* probes with the 11,061-bp fragment only.

These predicted patterns are evident in lane 11 of the six Southern blots shown in Fig. 9, which contains DNA from the O111:NM prototype strain B171-8. Sharing this pattern are strains prototypic of EPEC serotypes O55:NM, O127:H6, and O111:H2. In the region flanked by *rsv* on the left and *bfpI-J* on the right, similar but nonidentical patterns are seen for EPEC serotypes O119:H6 and O142:H6 (which appear to have the same pattern) and for serotype O126:NM, indicating minor degrees of polymorphism in this region with respect to the *Hin*dIII restriction sites. However, more dramatic differences are evident around *repI* for these three strains.

As expected, no hybridized fragments were seen in the lane containing DNA of B171-4, a derivative of strain B171-8 that was cured of the EAF plasmid. The absence of hybridized fragments in the lanes containing DNA from EPEC serotypes O114:NM and O26:NM was unexpected. However, further analysis of these strains showed that each had spontaneously lost the EAF plasmid during storage.

To determine if the restriction fragment length polymorphism patterns shown in Fig. 9 are indicative of the juxtaposition of the *bfp* cluster with *repI* and *rsv* as was demonstrated by sequence analysis of the B171-8 EAF plasmid (Fig. 1), PCR experiments were conducted with forward and reverse primers that in B171-8 led to the amplification of a fragment spanning *repI*, *rsv*, and *bfpA*. Each of the strains shown to contain these genes by restriction fragment length polymorphism analysis was also shown to yield a common, amplified fragment containing *repI*, *rsv*, and *bfpA*, demonstrating that the *bfp* gene cluster is located next to a potential replicon in each of the tested EPEC serotypes (data not shown).

DISCUSSION

The work reported here has identified a ca. 11-kb region of the EAF plasmid containing a cluster of at least 12 tandemly arrayed ORFs that is sufficient to direct the production of a morphologically normal BFP filament. Our analysis has not proven that every ORF within this cluster is required for BFP biogenesis. However, knockout mutations and/or functional studies have demonstrated that 3 of the 12 cluster genes are required for pilus fiber production: *bfpA* encodes pilin, the repeating subunit that comprises the pilus filament; *bfpB* encodes an outer membrane lipoprotein that is required for steps in the biogenesis of the pilus filament that follow signal pep-

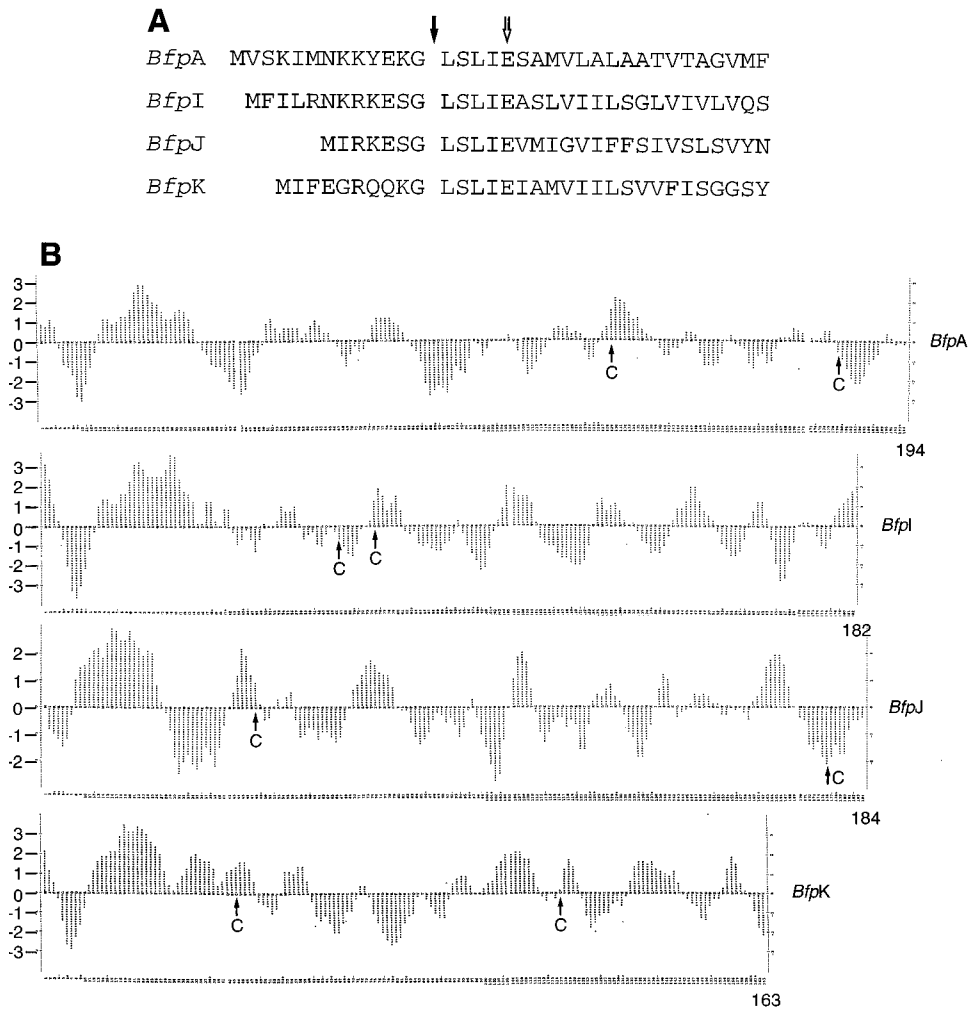

FIG. 7. Comparison of BfpA with BfpI, BfpJ, and BfpK, pilin-like proteins of the *bfp* gene cluster. (A) Alignment of the amino-terminal sequence of the EPEC B171-8 BfpA prepilin (unprocessed BfpA) with the translated amino acid sequences of putative BfpA-like proteins. The solid arrow indicates the predicted prepilin peptidase cleavage site in each of these sequences; cleavage has been demonstrated only in BfpA. The open arrow indicates the invariant Glu residue at position +5, demonstrated for type IV pilins to be required for signal peptidase processing and N-terminal modification (49). (B) Hydropathy plots of BfpA and the three BfpA-like proteins (BfpI, BfpJ, and BfpK) according to the algorithm of Kyte and Doolittle (45), calculated by using a window of six amino acids. The relative hydrophobicity is shown on the vertical axis (>0, increasing hydrophobicity; <0, increasing hydrophilicity). The amino acid number is shown on the horizontal axis of each plot. C's denote the positions of cysteines.

tidase cleavage of prepilin (79); and *bfpP* encodes the prepilin signal peptidase (reference 110 and Fig. 6). Beyond the identification of these essential cistrons, we have not yet sought to determine if the other internal genes are required for pilus biogenesis. However, according to the 3′ deletion analysis depicted in Fig. 3, the boundaries of the essential *bfp* gene cluster are delimited by *bfpA* on the left of Fig. 1 and by *bfpL* on the right, because *bfpL*, but not *bfpM*, was found to be required for BFP filament biogenesis. Results from the same experiment also suggest that the requirement of *bfpJ*, *bfpK*, and *bfpL* for filament production occurs after signal peptidase cleavage of BfpA (Fig. 3). Studies are now in progress to determine if the remaining genes in the cluster are required for BFP biogenesis and function by using allelic exchange to generate in-frame mutations that do not interrupt the transcription of downstream genes.

Genes outside the *bfp* gene cluster are also required for BFP biogenesis. *bfpT*, located on the EAF plasmid 6,777 bp beyond the last downstream gene of the *bfp* gene cluster, encodes an AraC-like *trans*-acting protein that activates *bfpA* transcription (100). In addition, several chromosomal genes may also participate in BFP biogenesis. Among these are the product of *dsbA*, a periplasmic chaperone protein that catalyzes intramolecular disulfide bond formation in exported proteins (70); DsbA may be required for the correct folding of BfpA and the pilin-like proteins of the cluster (BfpI, BfpJ, and BfpK; Fig. 7), each of which contains two cysteine residues which probably form intramolecular disulfide bonds. *lspA*, a chromosomal gene that codes for lipoprotein signal peptidase (signal peptidase II), together with the glyceryl and fatty acyl transferase genes, probably is necessary to process BfpB and BfpH, which have lipoprotein protease processing consensus sequences (76). While type IV pilin export across the cytoplasmic membrane appears to be SecA independent (39), it is likely that the six chromosomally encoded proteins that constitute the Sec system (SecA, -B, -D, -E, -F, and -Y) are necessary for BFP biogenesis because they are required for lipoprotein export (76) and the lipoprotein BfpB is required for pilus filament production (79). If so, the biogenesis of BFP probably requires

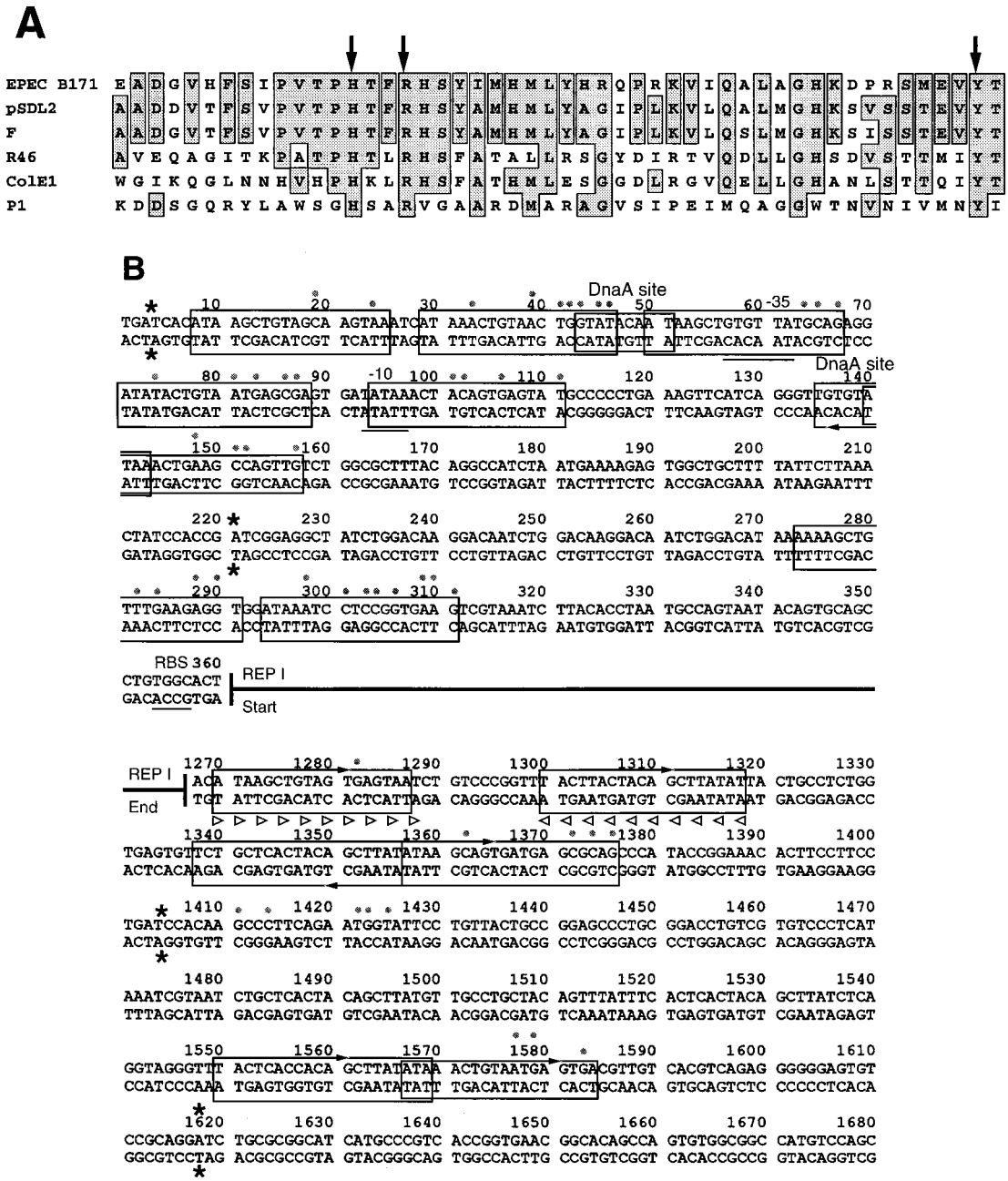

FIG. 8. Analysis of the deduced amino acid and nucleotide sequences of a possible origin of replication region of the EAF plasmid upstream of *bfpA* (including the putative resolvase and replicase loci) from EPEC strain B171-8. (A) Alignment of amino acids of the C-terminal domain of site-specific recombinases of the integrase family. The following sequences are depicted: Rsv protein of the EPEC B171-8 EAF plasmid (this work); Rsd protein of pSDL2 from *S. dublin* (44); D protein of F plasmid from *E. coli* K-12 (47); integrase of plasmid R46 (32); *xerC* product involved in the resolution of ColE1 plasmids (13); and Cre of bacteriophage P1 (5). The shadowed residues are conserved in at least three of the six proteins. Vertical arrows denote invariant amino acids in all recombinases of the integrase family. This figure was adapted from an alignment of site-specific recombinases in reference 44 by the addition of the EAF plasmid *rsv* locus of EPEC strain B171-8. (B) Nucleotide sequences flanking *repI* showing the *incE* and *incF* incompatibility regions. Nucleotide numbers are indicated at the top of the sequence and correspond with the numbering system used in Fig. 1 and 2. The ORF of *repI* is represented by the dark horizontal bar; the nucleotide sequences 360 bp upstream and 480 bp downstream of *repI* are shown. The three potential DNA methylation sites around the origin of replication (5′ GATC 3′ [*]), the two DnaA boxes and the direct repeats (27, 102) discussed in the text, are highlighted in the sequence. The 18-bp direct repeat, 5′ ATAA(A/G)CTGT(A/G)GTAAGTTA 3′, and the 19-bp motif, 5′ TACTNACTA CAGCTTATAT 3′, are boxed, and the arrow depicts the direct and inverse orientation of the repeats. The dots above these boxed sequences denote the base pairs that are different from the consensus sequences. The direct repeats and the DnaA boxes overlap each other. Possible RNA polymerase-binding sites are marked above the lines as −35 and −10 and are underlined. RBS refers to the putative ribosome binding site (underlined). The *incE* and *incF* incompatibility regions are shown between nt 1414 and 1429, denoted by dots over the relevant base pairs. The ORF *repI* is followed by an inverted repeat shown as split arrows under the sequence.

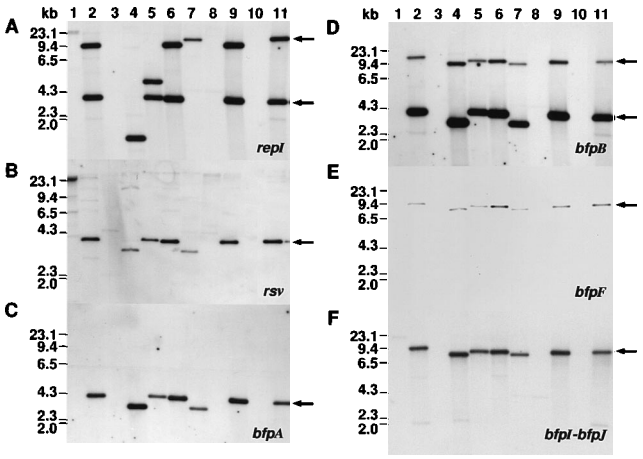

FIG. 9. Southern blot analysis of the organization of *bfp* gene clusters in *E. coli* strains representing different EPEC serotypes. *Hin*dIII-digested plasmid DNA from class I and II EPEC strains (52) was analyzed by Southern blotting using radiolabeled probes of PCR-amplified EAF plasmid loci *repI*, *rsv*, *bfpA*, *bfpB*, *bfpF*, and *bfpI-bfpJ*, as indicated. Lanes: 1, size standards; 2, EPEC strain 6170-50 (O111:H2); 3, strain 4393-57 (O114:NM) (class II); 4, strain 1157-54 (O119:H6); 5, strain 1929-55 (O126:NM); 6, strain E2348/69 (O127:H6); 7, strain 1181-83 (O142:H6) (class II); 8, strain 1914-55 (O26:NM); 9, strain 5513-51 (O55:NM); 10, strain B171-4 (O111:NM); 11, strain B171-8 (O111:NM). The DNA size standards are shown on the left. The arrows on the right indicate the *Hin*dIII bands hybridizing to the respective probes.

in excess of 22 genes, provided that all of the *bfp* gene cluster cistrons are needed.

The organization of the *bfp* gene cluster depicted in Fig. 2 most resembles the Tcp gene cluster of *V. cholerae* (10, 40, 66, 67) and the pullulanase gene cluster of *Klebsiella oxytoca* (77). Both of these in turn have been assigned by Salmond and Reeves (86) to the type II, *sec*-dependent general secretion pathway family of secreted proteins. Several other members of this family are shown in Table 2, which also depicts homologous relationships between the accessory proteins of these secretion systems. Besides the identification of homologous proteins, support for the assignment of the type IV pili to this family comes from heterogeneric complementation experiments in which the prepilin-like peptidase of the *K. oxytoca* pullulanase system (PulO) was shown to correctly process *N. gonorrhoeae* and *P. aeruginosa* type IV pilins (22, 97).

The most fundamental attribute of the type II secretion system is its capacity to move macromolecules across the cell envelope, into or out of the bacterial cytoplasmic space. Beyond the secreted enzymes and type IV pilins described above, this system also is associated with the translocation of DNA across the cell envelope and thus is required for transformation competence in *Haemophilus influenzae* (101, 48) and *Bacillus subtilis* (1, 2, 57) and with the transport of *Agrobacterium tumefaciens* Ti plasmid DNA to host plant tissues (106, 109). However, until the discovery of the BFP system, *E. coli* was not known to secrete proteins via the type II secretion pathway, even though several cryptic genes that could potentially code for proteins with type IV prepilin signal sequences (*ppdA* to *ppdC*) had been identified (26), and some processing of *N. gonorrhoeae* and *Dichelobacter nodosus* type IV prepilins reportedly occurs in *E. coli* (23, 24). Nonetheless, it seems likely that *E. coli* normally lacks several essential genes required to secrete proteins by the type II secretion pathway (78) and that these genes are provided through acquisition of the EAF plasmid and its *bfpA* gene cluster.

In addition to the translocated molecule (e.g., pilins, degradative enzymes, and DNA), the type II secretion system also includes the secretion apparatus itself, a hypothetical assembly that is believed to be composed of both structural and catalytic proteins that together have been termed type II membrane traffic wardens (86). The exact protein composition, size, and location of the pilus secretion apparatus are unknown, but there is no evidence to suggest that it is similar to the well-characterized *E. coli* Pap pilus assembly system (37) since none of the *bfp* gene cluster proteins are homologous to Pap pilus accessory proteins. Except for the catalytic activity of BfpP and the other prepilin peptidases shown in Table 2, which are all probably bifunctional enzymes that both cleave and N methylate prepilin and prepilin-like proteins (63, 96), little is known about either the subcellular locations or the functions of the other traffic wardens, although functional and topological models have been proposed (35, 76, 86). These models and our own concepts based on preliminary localization studies of BfpA and BfpB in EPEC strain B171-8 suggest that the membrane traffic wardens encoded by the *bfp* gene cluster compose topologically discrete pilus assembly sites that are associated with the bacterial cell envelope. The localized nature of these sites would account for the observation that type IV pili often emanate from the bacterial poles, in contrast to the peritrichous distribution of *E. coli* somatic pili. In addition to processing prepilin, the traffic wardens within the assembly complex would need to perform two additional functions necessary for pilus biogenesis and function: provide energy for the secretion and assembly phase of biogenesis, and form a periplasmic and/or outer membrane channel through which the pilus polymer would pass. Energy from ATP hydrolysis could be supplied by BfpD or BfpF, both of which have well-conserved nucleotide-binding motifs (104) (Fig. 5). ATP-binding traffic wardens that are homologous to BfpD and BfpF by virtue of the same nucleotide-binding motifs have been proposed to provide energy not only for pilin export and assembly but also for the twitching motility phenotype (16, 109). Twitching motility, which provides lateral mobility along a surface, is readily observed in type IV piliated *N. gonorrhoeae*, *P. aeruginosa*, and *Moraxella nonliquefaciens* but has not yet been reported for BFP-producing EPEC grown on solid media. However, preliminary observations now suggest that the autoaggregation phenotype, evident when EPEC strains are grown in tissue culture media (103), may be the consequence of non-flagellum-mediated movement in liquid that is equivalent to the twitching motility phenotype reported for most other type IV piliated species (8). Studies are in progress to determine if BfpD and BfpF are required for the autoaggregation phenotype.

How type IV pilins transit the periplasmic space and outer membrane is unknown. Localization studies of BfpA suggest that a BfpA pilin pool is associated with the cytoplasmic membrane and that BfpB is an outer membrane protein (79). It is possible that BfpB, either as a homooligomer or in association with other accessory proteins, forms a transmembrane channel for the pilus fiber. Furthermore, BfpA, BfpI, BfpJ, and BfpK all share the conserved hydrophobic N-terminal segment that was shown for *P. aeruginosa* pili to be involved with pilin subunit-subunit association within the pilus polymer (107). As a result, it is possible that BfpI, BfpJ, and BfpK interact through this domain with BfpA and through their nonhomologous C-terminal domains with other proteins comprising the pilus assembly complex, including the hypothesized pilus channel.

The stoichiometric relationships between the components of the pilus assembly complex are probably regulated within a narrow range in order to form a functional entity (97). How

optimal molar ratios between these proteins are achieved is not known, but the mechanism may relate to the transcriptional organization of the *bfp* gene cluster. Potential stem-loop structures that might function as mRNA transcriptional terminators or attenuators are located 3′ to *bfpA* and *bfpL*. Northern (RNA) blotting experiments (75) and preliminary RNase protection studies (79) show that *bfpA* mRNA is much more abundant than *bfpB* mRNA, a finding that agrees with Western blot estimates of the relative amounts of the corresponding proteins that are produced. However, at present it is not known if the transcription of units within the *bfp* gene cluster is differentially regulated or if protein production is also governed at the translational level.

The *bfp* gene cluster is located next to a possible origin of replication of the EAF plasmid (Fig. 2), and this arrangement is conserved in the EAF plasmids of six additional EPEC serotypes (Fig. 9). This juxtaposition may constitute what Perez-Casal et al. have termed a virulence factor replication unit (71, 72) because it brings together two potentially crucial plasmid loci, one which may be required for plasmid replication (*repI* and *rsv*) and the other (the *bfp* gene cluster) which may confer a survival advantage to the plasmids and strains that harbor it. This arrangement would reduce the probability that these loci would become dissociated as a result of deletions sustained by the EAF plasmid during recombination events. As a result, the *bfp* gene cluster would retain a replication region and thus survive as a functional and evolutionary entity.

ACKNOWLEDGMENTS

This work was supported by the Howard Hughes Medical Institute and in part by Digestive Disease Center grant 38707 to G. K. Schoolnik. Jose Luis Puente was the recipient of International Research Fellowship 1F05 TWO4697-01 from the Fogarty International Center, NIH.

We thank Toru Tobe for advice, critical reading of the manuscript, and stimulating discussions, Jorge Giron for providing BFP antiserum, and Rosemary Fernandes for preparing the electron micrographs.

ADDENDUM IN PROOF

The nomenclature of the *bfp* locus described in this manuscript was developed by agreement between the laboratories of G.K. Schoolnik and M. S. Donnenberg to facilitate consistency in the literature. In the course of developing this nomenclature, we became aware of two open reading frames in the *bfp* region with putative start sites other than ATG. By convention, these genes are designated *bfpG* and *bfpU*. Antipeptide antibodies generated to BfpG (S. Ramer, unpublished observation) have been used to demonstrate that *bfpG* encodes an EAF plasmid-dependent protein in B171-8.

REFERENCES

1. **Albano, M., R. Breitling, and D. A. Dubnau.** 1989. Nucleotide sequence and genetic organization of the *Bacillus subtilis comG* operon. J. Bacteriol. **171:**5386–5404.
2. **Albano, M., and D. A. Dubnau.** 1989. Cloning and characterization of a cluster of linked *Bacillus subtilis* late competence mutations. J. Bacteriol. **171:**5376–5385.
3. **Alm, R. A., and J. S. Mattick.** 1995. Identification of a gene, *pilV*, required for type 4 fimbrial biogenesis in *Pseudomonas aeruginosa*, whose product possesses a pre-pilin-like leader sequence. Mol. Microbiol. **16:**485–496.
4. **Amann, E., J. Brosius, and M. Ptashne.** 1983. Vectors bearing a hybrid *trp-lac* promoter useful for regulated expression of cloned genes in *Escherichia coli*. Gene **25:**167–178.
5. **Argos, P., A. Landy, K. Abremski, J. B. Egan, E. Haggard-Ljungquist, R. H. Hoess, M. L. Kahn, B. Kalionis, S. V. L. Narayama, L. S. Pierson, N. Sternberg, and J. M. Leong.** 1986. The integrase family of site-specific recombinases: regional similarities and global diversity. EMBO J. **5:**433–440.
6. **Baldini, M. M., J. B. Kaper, M. M. Levine, D. C. Candy, and H. W. Moon.** 1983. Plasmid-mediated adhesion in enteropathogenic *Escherichia coli*. J. Pediatr. Gastroenterol. Nutr. **2:**534–538.
7. **Bally, M., A. Filloux, M. Akrim, G. Ball, A. Lazdunski, and J. Tommassen.** 1992. Protein secretion in *Pseudomonas aeruginosa*: characterization of seven *xcp* genes and processing of secretory apparatus components by prepilin peptidase. Mol. Microbiol. **6:**1121–1131.
8. **Bieber, D.** Unpublished data.
9. **Brosius, J., and A. Holy.** 1984. Regulation of ribosomal RNA promoters with a synthetic *lac* operator. Proc. Natl. Acad. Sci. USA **81:**6929–6933.
10. **Brown, R. C., and R. K. Taylor.** 1995. Organization of *tcp*, *acf*, and *toxT* genes within a ToxT-dependent operon. Mol. Microbiol. **16:**425–439.
11. **Chang, A. C. Y., and S. N. Cohen.** 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. J. Bacteriol. **134:**1141–1156.
12. **Chou, P. Y., and G. D. Fasman.** 1978. Empiric predictions of protein conformation. Annu. Rev. Biochem. **47:**251–276.
13. **Colloms, S. D., P. Sykora, G. Szatmari, and D. Sherratt.** 1990. Recombination at ColE1 *cer* requires the *Escherichia coli xerC* gene product, a member of the lamba integrase family of site-specific recombinases. J. Bacteriol. **171:**6973–6980.
14. **Condemine, G., C. Dorel, N. Hugouvieux-Cotte-Pattat, and J. Robert-Baudouy.** 1992. Some of the *out* genes involved in the secretion of pectate lyases in *Erwinia chrysanthemi* are regulated by *kdgR*. Mol. Microbiol. **6:**3199–3211.
15. **Cravioto, A., R. J. Gross, S. M. Scotland, and B. Rowe.** 1979. An adhesive factor found in strains of *Escherichia coli* belonging to the traditional infantile enteropathogenic serotypes. Curr. Microbiol. **3:**95–99.
16. **Darzins, A.** 1994. Characterization of a *Pseudomonas aeruginosa* gene cluster involved in pilus biosynthesis and twitching motility: sequence similarity to the chemotaxis proteins of enterics and the gliding bacterium *Myxococcus xanthus*. Mol. Microbiol. **11:**137–153.
17. **d'Enfert, C., and A. P. Pugsley.** 1989. *Klebsiella pneumoniae pulS* gene encodes an outer membrane lipoprotein required for pullulanase secretion. J. Bacteriol. **171:**3673–3679.
18. **d'Enfert, C., I. Reyss, C. Wandersman, and A. P. Pugsley.** 1989. Protein secretion by Gram-negative bacteria: characterization of two membrane proteins required for pullulanase secretion by *Escherichia coli* K-12. J. Biol. Chem. **264:**17462–17468.
19. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
20. **Donnenberg, M. S., J. A. Giron, J. P. Nataro, and J. B. Kaper.** 1992. A plasmid-encoded type IV fimbrial gene of enteropathogenic *Escherichia coli* associated with localized adherence. Mol. Microbiol. **6:**3427–3437.
21. **Dums, F., J. M. Dow, and M. J. Daniels.** 1991. Structural characterization of protein secretion genes of the bacterial phytopathogen *Xanthomonas campestris* pathovar *campestris*: relatedness to secretion systems of other gram-negative bacteria. Mol. Gen. Genet. **229:**357–364.
22. **Dupuy, B., M. K. Taha, O. Possot, C. Marchal, and A. P. Pugsley.** 1992. PulO, a component of the pullulanase secretion pathway of *Klebsiella oxytoca*, correctly and efficiently processes gonococcal type IV prepilin in *Escherichia coli*. Mol. Microbiol. **6:**1887–1894.
23. **Dupuy, B., M. K. Taha, A. P. Pugsley, and C. Marchal.** 1991. *Neisseria gonorrhoeae* prepilin export studied in *Escherichia coli*. J. Bacteriol. **173:**7589–7598.
24. **Elleman, T. C., P. Hoyne, D. L. Emery, D. J. Stewart, and B. L. Clark.** 1984. Isolation of the gene encoding pilin of *Bacteroides nodosus* (strain 198), the causal organism of ovine footrot. FEBS Lett. **173:**103–107.
25. **Faast, R., M. A. Ogierman, U. H. Stroeher, and P. A. Manning.** 1989. Nucleotide sequence of the structural gene, *tcpA*, for a major pilin subunit of *Vibrio cholerae*. Gene **85:**227–231.
26. **Finch, P. W., R. E. Wilson, K. Brown, I. D. Hickson, A. E. Tomkinson, and P. Emmerson.** 1986. Complete nucleotide sequence of the *Escherichia coli recC* gene and the *thyA-recC* intergenic region. Nucleic Acids Res. **14:**4437–4450.
27. **Fuller, R. S., and A. Kornberg.** 1989. Purified dnaA protein in initiation of replication at the *Escherichia coli* chromosomal origin of replication. Proc. Natl. Acad. Sci. USA **80:**5817–5821.
28. **Gardner, R. C., L. Malcolm, P. L. Bergquist, and H. E. D. Lane.** 1982. *IncD*, a genetic locus in F responsible for incompatibility with several plasmids of the *IncFI* group. Mol. Gen. Genet. **188:**345–352.
29. **Giron, J. A., A. S. Y. Ho, and G. K. Schoolnik.** 1991. An inducible bundle-forming pilus of enteropathogenic *Escherichia coli*. Science **254:**710–713.
30. **Gomes, T. A., M. A. Vieira, I. K. Wachsmuth, P. A. Blake, and I. R. Trabulsi.** 1989. Serotype-specific prevalence of *Escherichia coli* strains with EPEC adherence factor genes in infants with and without diarrhea in Sao Paulo, Brazil. J. Infect. Dis. **160:**131–135.
31. **Gomes, T. A. T., P. A. Blake, and L. R. Trabulsi.** 1989. Prevalence of *Escherichia coli* strains with localized, diffuse, and aggregative adherence to HeLa cells in infants with diarrhea and matched controls. J. Clin. Microbiol. **27:**266–269.

32. **Hall, R. M., and C. Vockler.** 1987. The region of the IncN plasmid R46 coding for resistance to βlactam antibiotics, streptomycin/spectinomycin and sulphonamides is closely related to antibiotic resistance segments found in IncW plasmids and in Tn21-like transposons. Nucleic Acids Res. **15:**7491–7501.
33. **He, S. Y., M. Lindeberg, A. K. Chatterjee, and A. Collmer.** 1991. Cloned *Erwinia chrysanthemi out* genes enable *Escherichia coli* to selectively secrete a diverse family of heterologous proteins to its milieu. Proc. Natl. Acad. Sci. USA **88:**1079–1083.
34. **Hobbs, M., B. P. Dalrymple, P. T. Cox, S. P. Livingstone, S. F. Delaney, and J. S. Mattick.** 1991. Organization of the fimbrial gene region of *Bacteroides nodosus*: class I and class II strains. Mol. Microbiol. **5:**543–560.
35. **Hobbs, M., and J. S. Mattick.** 1993. Common components in the assembly of type 4 fimbriae, DNA transfer systems, filamentous phage and protein-secretion apparatus: a general system for the formation of surface-associated protein complexes. Mol. Microbiol. **10:**233–243.
36. **Hu, N.-T., M.-N. Hung, S.-J. Chiou, F. Tang, D.-C. Chiang, H.-Y. Huang, and C.-Y. Wu.** 1992. Cloning and characterization of a gene required for the secretion of extracellular enzymes across the outer membrane by *Xanthomonas campestris* pv. campestris. J. Bacteriol. **174:**2679–2687.
37. **Hultgren, S., and S. Normark.** 1991. Biogenesis of the bacterial pilus. Curr. Opin. Genet. Dev. **1:**313–318.
38. **Jiang, B., and S. P. Howard.** 1992. The *Aeromonas hydrophila exeE* gene, required both for protein secretion and normal outer membrane biogenesis, is a member of a general secretion pathway. Mol. Microbiol. **6:**1351–1361.
39. **Kaufman, M. R., J. M. Seyer, and R. T. Taylor.** 1991. Processing of TCP pilin by TcpJ typifies a common step intrinsic to a newly recognized pathway of extracellular protein secretion by gram-negative bacteria. Genes Dev. **5:**1834–1846.
40. **Kaufman, M. R., C. E. Shaw, I. D. Jones, and R. K. Taylor.** 1993. Biogenesis and regulation of the *Vibrio cholerae* toxin-coregulated pilus: analogies to other virulence factor secretory systems. Gene **126:**43–49.
41. **Knutton, S., M. Baldini, J. Kaper, and A. McNeish.** 1987. Role of plasmid-encoded adherence factors in adhesion of enteropathogenic *Escherichia coli* to HEp-2 cells. Infect. Immun. **55:**78–85.
42. **Knutton, S., T. Baldwin, P. H. Williams, and A. S. McNeish.** 1989. Actin accumulation at sites of bacterial adhesion to tissue culture cells: basis of a new diagnostic test for enteropathogenic and enterohemorrhagic *Escherichia coli*. Infect. Immun. **57:**1290–1298.
43. **Knutton, S., D. Lloyd, and A. McNeish.** 1987. Adhesion of enteropathogenic *Escherichia coli* to human intestinal enterocytes and cultured human intestinal mucosa. Infect. Immun. **55:**69–77.
44. **Krause, M., and D. G. Guiney.** 1991. Identification of a multimer resolution system involved in stabilization of the *Salmonella dublin* virulence plasmid pSDL2. J. Bacteriol. **173:**5754–5792.
45. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:**105–132.
46. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.
47. **Lane, D., R. de Feyteer, M. Kennedy, S.-H. Phua, and D. Semon.** 1986. D protein of miniF plasmid acts as a repressor of transcription and as a site-specific resolvase. Nucleic Acids Res. **14:**9713–9728.
48. **Larson, T. G., and S. H. Goodgal.** 1991. Sequence and transcriptional regulation of *com101A*, a locus required for genetic transformation in *Haemophilus influenzae*. J. Bacteriol. **171:**4683–4691.
49. **Lauer, P., N. H. Albertson, and M. Koomey.** 1993. Conservation of genes encoding components of a type IV pilus assembly/two-step protein export pathway in *Neisseria gonorrhoeae*. Mol. Microbiol. **8:**357–368.
50. **Levine, M. M.** 1987. *Escherichia coli* that cause diarrhea: enterotoxigenic, enteropathogenic, enteroinvasive, enterohemorrhagic, and enteroadherent. J. Infect. Dis. **155:**377–385.
51. **Levine, M. M., E. J. Bergquist, D. R. Nalin, D. H. Waterman, R. B. Hornick, C. R. Young, S. Sotman, and B. Rowe.** 1978. *Escherichia coli* strains that cause diarrhoea but do not produce heat-labile or heat-stable enterotoxins and are non-invasive. Lancet **i:**1119–1122.
52. **Levine, M. M., and R. Edelman.** 1984. Enteropathogenic *Escherichia coli* of classic serotypes associated with infant diarrhea: epidemiology and pathogenesis. Epidemiol. Rev. **6:**31–84.
53. **Levine, M. M., J. P. Nataro, H. Karch, M. M. Baldini, J. B. Kaper, R. E. Black, M. L. Clements, and A. D. O'Brien.** 1985. The diarrheal response of humans to some classic serotypes of enteropathogenic *Escherichia coli* is dependent on a plasmid encoding an enteroadhesiveness factor. J. Infect. Dis. **152:**550–559.
54. **Lindeberg, M., and A. Collmer.** 1992. Analysis of eight *out* genes in a cluster required for pectic enzyme secretion by *Erwinia chrysanthemi*: sequence comparison with secretion genes from other gram-negative bacteria. J. Bacteriol. **174:**7385–7397.
55. **Lu, A.-L.** 1987. Influence of GATC sequences on *Escherichia coli* DNA mismatch repair in vitro. J. Bacteriol. **169:**1254–1259.
56. **Meyer, T. F., E. Billyard, R. Haas, S. Storzbach, and M. So.** 1984. Pilus genes of *Neisseria gonorrhoeae*: chromosomal organization and DNA sequence. Proc. Natl. Acad. Sci. USA **81:**6110–6114.
57. **Mohan, S., J. Aghion, N. Guillen, and D. Dubnau.** 1989. Molecular cloning and characterization of *comC*, a late competence gene of *Bacillus subtilis*. J. Bacteriol. **171:**6043–6051.
58. **Motallebe-Veshareh, M., D. Balzer, E. Lanka, G. Jagura-Burdzy, and C. M. Thomas.** 1992. Conjugative transfer function of broad-host-range plasmid RK2 are coregulated with vegetative replication. Mol. Microbiol. **6:**907–920.
59. **Nataro, J. P., M. M. Baldini, J. B. Kaper, R. E. Black, N. Bravo, and M. M. Levine.** 1985. Detection of an adherence factor of enteropathogenic *Escherichia coli* with a DNA probe. J. Infect. Dis. **162:**550–552.
60. **Nataro, J. P., K. O. Maher, P. Mackie, and J. B. Kaper.** 1987. Characterization of plasmids encoding the adherence factor of enteropathogenic *Escherichia coli*. Infect. Immun. **55:**2370–2377.
61. **Nataro, J. P., I. C. A. Scaletsky, J. B. Kaper, M. M. Levine, and L. R. Trabulsi.** 1985. Plasmid-mediated factors conferring diffuse and localized adherence of enteropathogenic *Escherichia coli*. Infect. Immun. **48:**378–383.
62. **Nunn, D., S. Bergman, and S. Lory.** 1990. Products of three accessory genes, *pilB*, *pilC*, and *pilD*, are required for biogenesis of *Pseudomonas aeruginosa* pili. J. Bacteriol. **172:**2911–2919.
63. **Nunn, D. N., and S. Lory.** 1991. Product of the *Pseudomonas aeruginosa* gene *pilD* is a prepilin leader peptidase. Proc. Natl. Acad. Sci. USA **88:**3281–3285.
64. **Nunn, D. N., and S. Lory.** 1992. Components of the protein-excretion apparatus of *Pseudomonas aeruginosa* are processed by the type IV prepilin peptidase. Proc. Natl. Acad. Sci. USA **89:**47–51.
65. **Nunn, D. N., and S. Lory.** 1993. Cleavage, methylation, and localization of the *Pseudomonas aeruginosa* export proteins XcpT, -U, -V, and -W. J. Bacteriol. **175:**4375–4382.
66. **Ogierman, M. A., and P. A. Manning.** 1992. TCP pilus biosynthesis in *Vibrio cholerae* O1: gene sequence of *tcpC* encoding an outer membrane lipoprotein. FEMS Microbiol. Lett. **97:**179–184.
67. **Ogierman, M. A., S. Zabihi, L. Mourtzios, and P. A. Manning.** 1993. Genetic organization and sequence of the promoter-distal region of the *tcp* gene cluster of *Vibrio cholerae*. Gene **126:**51–60.
68. **Pabo, C. O., and R. T. Sauer.** 1984. Protein-DNA recognition. Annu. Rev. Biochem. **53:**293–321.
69. **Paulozzi, L. J., K. E. Johnson, L. M. Kamahele, C. R. Clausen, L. W. Riley, and S. D. Helgerson.** 1986. Diarrhea associated with adherent enteropathogenic *Escherichia coli* in an infant and toddler center, Seattle, Washington. Pediatrics **77:**296–299.
70. **Peek, J. A., and R. K. Taylor.** 1992. Characterization of periplasmic thiol: disulfide interchange protein required for the functional maturation of secreted virulence factors of *Vibrio cholerae*. Proc. Natl. Acad. Sci. USA **89:**6210–6214.
71. **Perez-Casal, J. F., and J. H. Crosa.** 1987. Novel incompatibility and partition loci for the REPI replication region of plasmid ColV-K30. J. Bacteriol. **169:**5078–5086.
72. **Perez-Casal, J. F., A. E. Gammie, and J. H. Crosa.** 1989. Nucleotide sequence analysis and expression of the minimum REPI replication region and incompatibility determinants of pColV-K30. J. Bacteriol. **171:**2195–2201.
73. **Possot, O., C. d'Enfert, I. Reyss, and A. P. Pugsley.** 1992. Pullulanase secretion in *Escherichia coli* K-12 requires a cytoplasmic protein and a putative polytopic cytoplasmic membrane protein. Mol. Microbiol. **6:**95–105.
74. **Possot, O., and A. P. Pugsley.** 1994. Molecular characterization of PulE, a protein required for pullulanase secretion. Mol. Microbiol. **12:**287–299.
75. **Puente, J. L., D. Bieber, S. W. Ramer, W. Murray, and G. K. Schoolnik.** The bundle-forming pili of enteropathogenic *Escherichia coli*: transcriptional regulation by environmental signals. Mol. Microbiol., in press.
76. **Pugsley, A. P.** 1993. The complete general secretory pathway in gram-negative bacteria. Microbiol. Rev. **57:**50–108.
77. **Pugsley, A. P., and B. Dupuy.** 1992. An enzyme with type IV prepilin peptidase activity is required to process components of the general extracellular protein secretion pathway of *Klebsiella oxytoca*. Mol. Microbiol. **6:**751–760.
78. **Pugsley, A. P., M. G. Kornacker, and I. Poquet.** 1991. The general protein-export pathway is directly required for extracellular pullulanase secretion in *Escherichia coli* K12. Mol. Microbiol. **5:**343–352.
79. **Ramer, S. W., D. Bieber, I. Sohel, and G. K. Schoolnik.** 1995. BfpB is a lipoprotein required for bundle-forming pilus production in enteropathogenic *Escherichia coli*, abstr. B-160, p. 193. *In* Abstracts of the 95th General Meeting of the American Society for Microbiology 1995. American Society for Microbiology, Washington, D.C.
80. **Reeves, P. J., D. Whitcombe, S. Wharam, M. Gibson, G. Allison, N. Bunce, R. Barallon, P. Douglas, V. Mulholland, S. Stevens, D. Walker, and G. P. C. Salmond.** 1993. Molecular cloning and characterization of 13 *out* genes from *Erwinia carotovora* subspecies *carotovora*: genes encoding members of a general secretion pathway (GSP) widespread in Gram-negative bacteria. Mol. Microbiol. **8:**443–456.
81. **Reyss, I., and A. P. Pugsley.** 1990. Five additional genes in the *pulC-O*

operon of the gram-negative bacterium *Klebsiella oxytoca* UNF5023 which are required for pullulanase secretion. Mol. Gen. Genet. **222:**176–184.
82. **Riley, L. W., L. N. Junio, L. B. Libaek, and G. K. Schoolnik.** 1987. Plasmid-encoded expression of lipopolysaccharide O-antigenic polysaccharide in enteropathogenic *Escherichia coli*. Infect. Immun. **55:**2052–2056.
83. **Riley, L. W., L. N. Junio, and G. K. Schoolnik.** 1990. HeLa cell invasion by a strain of enteropathogenic *Escherichia coli* that lacks the O-antigenic polysaccharide. Mol. Microbiol. **4:**1661–1666.
84. **Rothbaum, R., A. McAdams, R. Giannella, and J. Partin.** 1982. A clinicopathologic study of enterocyte-adherent *Escherichia coli*: a cause of protracted diarrhea in infants. Gastroenterology **83:**441–454.
85. **Rothbaum, R., J. Partin, K. Saalfield, and A. McAdams.** 1983. An ultrastructural study of enteropathogenic *Escherichia coli* infection in human infants. Ultrastruct. Pathol. **4:**291–304.
86. **Salmond, G. P. C., and P. J. Reeves.** 1993. Membrane traffic wardens and protein secretion in Gram-negative bacteria. Trends Biochem. Sci. **18:**7–12.
87. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
88. **Sandkvist, M., V. Morales, and M. Bagdasarian.** 1993. A protein required for secretion of cholera toxin through the outer membrane of *Vibrio cholerae*. Gene **123:**81–86.
89. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
90. **Saul, D., A. J. Spiers, J. McAnulty, M. G. Gibbs, P. L. Bergquist, and D. F. Hill.** 1989. Nucleotide sequence and replication characteristics of RepFIB, a basic replicon of IncF plasmids. J. Bacteriol. **171:**2697–2707.
91. **Scaletsky, I. C. A., M. L. M. Silva, and L. R. Trabulsi.** 1984. Distinctive patterns of adherence of enteropathogenic *Escherichia coli* to HeLa cells. Infect. Immun. **45:**534–536.
92. **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.
93. **Sohel, I., J. L. Puente, W. J. Murray, J. Vuopio-Varkila, and G. K. Schoolnik.** 1993. Cloning and characterization of the bundle-forming pilin gene of enteropathogenic *Escherichia coli* and its distribution in *Salmonella* serotypes. Mol. Microbiol. **7:**563–575.
94. **Spanfenberg, C., R. Fislage, W. Sierralta, B. Tummler, and U. Romling.** 1995. Comparison of type IV-pilin genes of *Pseudomonas aeruginosa* of various habitats has uncovered a novel unusual sequence. FEMS Microbiol. Lett. **125:**265–274.
95. **Strom, M. S., D. Nunn, and S. Lory.** 1991. Multiple roles of the pilus biogenesis protein PilD: involvement of PilD in excretion of enzymes from *Pseudomonas aeruginosa*. J. Bacteriol. **173:**1175–1180.
96. **Strom, M. S., D. Nunn, and S. Lory.** 1993. A single bifunctional enzyme, PilD, catalyzes cleavage and N-methylation of proteins belonging to the type IV pilin family. Proc. Natl. Acad. Sci. USA **90:**2404–2408.
97. **Strom, M. S., and S. Lory.** 1993. Structure and biogenesis of the type IV pili. Annu. Rev. Microbiol. **47:**565–596.
98. **Studier, F. W., and B. A. Moffatt.** 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. **189:**113–130.
99. **Studier, F. W., A. H. Rosenberg, J. J. Dunn, and J. W. Dubendorff.** 1990. Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol. **185:**60–89.
100. **Tobe, T., V. H. Bustamante, I. Sohel, G. K. Schoolnik, and J. L. Puente.** Cloning and characterization of *bfpT*, an EAF plasmid locus that activates *bfpA* transcription in enteropathogenic *Escherichia coli*. Submitted.
101. **Tomb, J.-F., H. El-Hajj, and H. O. Smith.** 1991. Nucleotide sequence of a cluster of genes involved in the transformation of *Haemophilus influenzae* Rd. Gene **104:**1–10.
102. **Tsutsui, H., A. Fujiyama, T. Murotsu, and K. Matsubara.** 1983. Role of nine repeating sequences of the mini-F genome for expression of F-specific incompatibility phenotype and copy number control. J. Bacteriol. **155:**337–344.
103. **Vuopio-Varkila, J., and G. K. Schoolnik.** 1991. Localized adherence by enteropathogenic Escherichia coli is an inducible phenotype associated with the expression of new outer membrane proteins. J. Exp. Med. **174:** 1167–1177.
104. **Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay.** 1982. Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1:**945–951.
105. **Wang, F. F., and S. R. Kushner.** 1991. Construction of versatile low-copy-number vectors for cloning, sequencing and gene expression in *Escherichia coli*. Gene **100:**195–199.
106. **Ward, J. E., D. E. Akiyoshi, D. Regier, A. Datta, M. P. Gordon, and E. W. Nester.** 1988. Characterization of the *virB* operon from an *Agrobacterium tumefaciens* Ti plasmid. J. Biol. Chem. **262:**5804–5814.
107. **Watts, T. H., C. M. Kay, and W. Paranchych.** 1983. Spectral properties of three quaternary arrangements of *Pseudomonas* pilin. Biochemistry **22:** 3640–3646.
108. **Weiss, A. A., F. D. Johnson, and D. L. Burns.** 1993. Molecular characterization of an operon required for pertussis toxin secretion. Proc. Natl. Acad. Sci. USA **90:**2970–2974.
109. **Whitchurch, C. B., M. Hobbs, S. P. Livingston, V. Krishnapillai, and S. Mattick.** 1990. Characterization of a *Pseudomonas aeruginosa* twitching motility gene and evidence for a specialized protein export system widespread in eubacteria. Gene **101:**33–44.
110. **Zhang, H.-Z., S. Lory, and M. S. Donnenberg.** 1994. A plasmid-encoded prepilin peptidase gene from enteropathogenic *Escherichia coli*. J. Bacteriol. **176:**6885–6891.