JOURNAL OF BACTERIOLOGY, June 1985, p. 1285–1292
0021-9193/85/061285-08$02.00/0

Vol. 162, No. 3

Genetic Organization of the Afimbrial Adhesin Operon and Nucleotide Sequence from a Uropathogenic *Escherichia coli* Gene Encoding an Afimbrial Adhesin

AGNES LABIGNE-ROUSSEL,† M. ALEXANDER SCHMIDT, WALTRAUD WALZ, AND STANLEY FALKOW*

Department of Medical Microbiology, Stanford University School of Medicine, Stanford, California 94305

Received 16 November 1984/Accepted 4 March 1985

The uropathogenic *Escherichia coli* KS52 strain expresses a mannose-resistant hemagglutinin AFA-I, which recognizes a human erythrocyte site distinct from the α-digalactoside glycosphingolipid receptor common to uropathogenic *E. coli* strains specifying a P adhesin. A 6.7-kilobase chromosomal DNA fragment was cloned from KS52 into pBR322 and was shown to be necessary for host cell mannose-resistant hemagglutination expression and uroepithelial cell adherence (Labigne-Roussel et al., Infect. Immun. 46:251–259, 1984). The genetic organization of the 6.7-kilobase DNA fragment was investigated by generating derivative plasmids, and the polypeptides encoded by those plasmids in isolated minicells were analyzed on polyacrylamide gel. The 6.7-kilobase insert expresses five polypeptides of molecular mass 13,000, 16,000, 18,500, 30,000, and 100,000 daltons encoded, respectively, by the *afaA*, *afaE*, *afaD*, *afaB*, and *afaC* genes. The five genes were localized and were shown to belong to the same transcriptional unit. The *afaB*, *afaC*, and *afaE* gene products are required for mannose-resistant hemagglutination expression, whereas mutations in or deletions of either *afaA* or *afaD* do not modify host mannose-resistant hemagglutination expression. The *afaE* gene was identified as the structural gene encoding hemagglutinin. The gene has been sequenced; it encodes a 152-residue protein containing a typical 21-residue procaryote signal sequence and a 131-residue mature polypeptide, the AFA-I adhesin.

Uropathogenic *Escherichia coli* strains isolated from pyelonephritis or cystitis infection in humans usually express a variety of phenotypic properties that distinguish them from the random fecal flora (10). Among these properties, the capacity of *E. coli* strains to adhere to uroepithelial cells through bacterial surface adhesins that recognize specific cell surface receptors seems to be essential for the pathogenesis of acute urinary tract infection (30). This adhesin property correlates with the ability of the bacteria to agglutinate human erythrocytes in the presence of D-mannose (9) and is referred to as mannose-resistant hemagglutination (MRHA). Ninety percent of the MRHA *E. coli* strains isolated from patients with pyelonephritis (13–15) express an adhesin that specifically recognizes on the cell surface globoseries glycolipids that contain the disacharride α-D-galactosyl-(1-4)β-D-galactose; in those strains, this adhesin is chromosomally encoded by an operon that expresses P fimbriae (*pap* operon) (11, 17, 21, 22). However 10% of the MRHA *E. coli* pyelonephritic strains recognize receptors other than the digalactoside (32) and encode mannose-resistant binding specificities that have been referred in the literature as X adhesins (24, 31).

We recently described (16) the properties of an uropathogenic *E. coli* strain, KS52, serotype O2, that is representative of a group of strains that are responsible for urinary tract infections and that share the same characteristics. Those MRHA strains do not encode a P-specific adhesin and therefore can be called X adhesin-expressing strains (25, 31). Typically these strains mediate their MRHA in the absence of fimbrial structures, and they exhibit a specific hemagglutinin that agglutinates only human erythrocytes. Recently, we reported the cloning of a 6.7-kilobase (kb) chromosomal DNA fragment (Fig. 1) that was shown to be necessary for nonfimbriate host cell MRHA expression and adherence to human transitional and squamous uroepithelial cells (16). This DNA fragment encodes the production of a 16,000-dalton adhesin, AFA-I, which can be detected in the culture supernatant fluid of the parental strain and of the strain harboring the 6.7-kb DNA inserted in pBR322 pIL14.

In the present study, we describe the genetic organization of the 6.7-kb DNA fragment harboring the operon encoding AFA-I (AFA operon). Using Tn*5* insertions, in vitro generated deletion mutations, and subcloning experiments, we have delimited the genetic region required for the expression of hemagglutination and adhesion properties. The original hybrid plasmid, pIL14, and several derivative plasmids were introduced into a minicell-producing strain, and the polypeptides expressed by those plasmids in the minicells were identified on acrylamide gels. Both DNA and protein analyses allowed us to localize and associate genes and gene products required for the expression of MRHA; the adhesin gene (*afaE*) was identified and sequenced, allowing the determination of the amino acid sequence of the polypeptide adhesin.

MATERIALS AND METHODS

Bacterial strains and plasmids. *E. coli* KS52 (O2 serotype), isolated from human pyelonephritis infection, was obtained from G. Kallenius (13). *E. coli* HB101 (2) (*hsdR hsdM recA13 supE44 lacZ4 leuB6 proA2 thi-1* Sm^r) was used as a host in all transformation experiments and as a host for plasmid DNA and protein analyses. *E. coli* P678-54 (1) (F$^-$ *thr-1 leu-6 thi-1 lacY1 malA1 xyl-7 ara-13 mtl-2 tonA2 gal-6* λ$^-$ *rpsL minA minB*) was used for the preparation of minicells. Cells competent for transformation were prepared by the method of

* Corresponding author.
† Present address: Unite des Enterobacteries, Institut Pasteur, 75015 Paris, France.

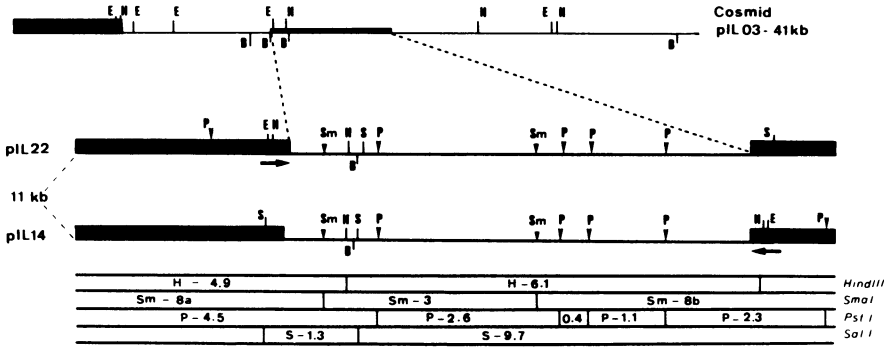

FIG. 1. Linear restriction map of pIL03, pIL14, and pIL22 hybrid plasmids. pIL14 and pIL22 (6.7 kb of DNA inserted in pBR322) were subcloned from the pIL03 cosmid (16) as the smallest DNA insert able to confer to *E. coli* HB101 the phenotypic properties of the parental clinical strain KS52 (16). Symbols and abbreviations: (■) pHC79 for pIL03 map and pBR322 for pIL14 and pIL22 maps, (——) DNA inserted in vitro; (B) *Bam*HI, (E) *Eco*RI, (H) *Hin*dIII, (P) *Pst*I, (Sm) *Sma*I, and (S) *Sal*I restriction sites.

Davis et al. (4). The plasmid hybrids used in these studies and their characteristics are listed in Table 1.

Bacteria were grown on L-broth without glucose (10 g of tryptone, 5 g of yeast extract, and 5 g of NaCl per liter [pH 7]) or on agar plates solidified with 1.5% agar (Difco Laboratories). Selection for antibiotic resistance was performed using the following concentrations: ampicillin, 100 μg/ml; tetracycline, 8 μg/ml; kanamycin, 20 μg/ml.

TABLE 1. Hybrid plasmids and their properties

Plasmid[a]	Size (kb)	MRHA[b]	Proteins expressed in P678-54[c]					Origin
			18K	16K[d]	100K[d]	30K	13K	
pIL03	41	+	+	+	+	+	+	(11)
pIL08	15.3	−	+	−	+	+	+	(11)
pIL14	11	+	+	+	+	+	+	(11)
pIL19	10.4	+	+	+	+	+	+	(11)
pIL22	11	+	+	+	+	+	+	(11)
pIL24	16.6	+	ND[e]	(+)	(+)	(+)	(+)	This paper
pIL42	16.6	−	ND	(+)	(+)	(+)	+	This paper
pIL46								
pIL27	16.6	−	ND	(+)	−	(+)	+	This paper
pIL28								
pIL30								
pIL34								
pIL36								
pIL37								
pIL38								
pIL40	16.6	+	ND	(+)	+	+	+	This paper
pIL41								
pIL45								
pIL47								
pIL31	16.6	−	ND	(+)	+	−	+	This paper
pIL39								
pIL48	6.7	−	−	−	−	−	+	This paper
pIL54	8	−	−	−	−	+	+	This paper
pIL56	8	−	−	−	−	+	+	This paper
pIL58	11	−	+	−	+	+	+	This paper
pIL80	9.4	+	(+)	(+)	(+)	(+)	−	This paper
pIL81	8	−	−	+	−	+	+	This paper

[a] The construction of the hybrid plasmids is illustrated in Fig. 1 and 3.

[b] Agglutination of human erythrocytes in presence of α-methyl mannoside was performed as described in the text.

[c] Minicells from the P678-54 strain harboring the appropriate hybrid plasmid were isolated and labeled with [^{35}S]methionine, and the expressed polypeptides were detected as described: +, clear presence of the polypeptide; (+), presence of the protein, but at a reduced level; −, total absence of the protein. K, Kilodaltons.

[d] Both the processed and unprocessed peptides were detected as labeled proteins in the P678-54 strain.

[e] ND, Not determined.

Hemagglutination assay and adherence assay. Hemagglutination in the presence of 2% (wt/vol) α-methyl mannoside and the adherence test to uroepithelial cells were performed as previously described (23).

Preparation of plasmid DNA. Plasmid DNA was isolated by alkaline lysis (18).

Plasmid DNA analysis. Restriction endonucleases *Bam*HI, *Eco*RI, *Hin*dIII, *Sal*I, *Sma*I, and *Pst*I and bacteriophage T4-induced polynucleotide ligase were purchased from New England BioLabs, Inc., Bethesda Research Laboratories, Inc., or Boehringer Mannheim Biochemicals. All of the enzymatic reaction conditions employed were according to the manufacters' recommendations. DNA fragments were separated by electrophoresis in vertical slab gels containing 5% polyacrylamide or in horizontal slab gels containing 0.7, 1, or 1.4% agarose and run in Tris-acetate buffer (17). Phage DNA and phage φX174 DNA cleaved with *Hin*dIII and *Hae*III, respectively, were used as molecular weight standards.

Extraction of restriction endonuclease-generated DNA fragments. Specific DNA fragments were isolated by elution from DEAE-nitrocellulose paper as previously described (16).

Mutagenesis. Transposon mutagenesis was performed with a plasmid ColE1::Tn*5* delivery system in conjunction with the incompatibility property of ColE1 with pBR322 derivative plasmids as a positive selection for transposition. Briefly, ColE1::Tn*5* plasmid DNA was transformed into an *E. coli* HB101 strain harboring the pIL14 plasmid. Transformants expressing ampicillin resistance and high-level resistance to kanamycin were selected on selective medium containing 100 μg of ampicillin per ml and 250 μg of kanamycin per ml for several passages. The transposition of Tn*5* from ColE1::Tn*5* into pIL14 and the disappearance of the donor of the transposon were confirmed by analyzing the plasmid DNA content of the transformants.

Small insertion mutations were generated by filling recessed 3′ ends of full-length pIL14 linear molecules generated by *Bam*HI digestion or partial *Pst*I digestion with the Klenow fragment of *E. coli* DNA polymerase I (18). The DNA was recircularized with T4 DNA ligase and was used to transform competent *E. coli* cells.

Antibody production. Antiserum against the hemagglutination AFA-I protein was raised as previously described (16).

Analysis of proteins expressed in minicells. Minicells from the P678-54-producing strains harboring the appropriate plasmid were isolated and labeled with [^{35}S]methionine or [^{35}S]cysteine (7). Approximately 100,000 cpm of acetone-pre-

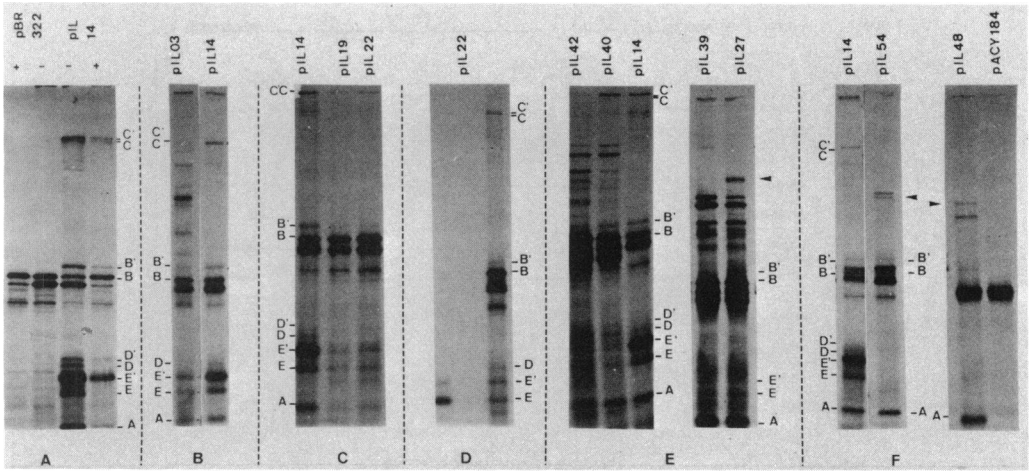

FIG. 2. Fluorographs of plasmid-mediated polypeptides expressed in minicells labeled with [^{35}S]methionine in the absence (no notation or −) or in the presence (+) of 9.5% ethanol and analyzed on 12.5% sodium dodecyl sulfate–polyacrylamide gels. The six polypeptides and their respective precursor (in daltons) encoded by pIL14 were designated as follows: C′ and C, 102,000 and 100,000; B′ and B, 33,000 and 30,000; D′ and D, 20,000 and 18,500; E′ and E, 17,500 and 16,000; A, 13,000. (A) Polypeptides encoded by the pIL14 hybrid compared with that expressed by the pBR322 vector. (B) Polypeptides expressed from the pIL03 cosmid in comparison with those of pIL14. (C) Effect of the orientation of the DNA inserted into the pBR322 vector on the level of polypeptide synthesis. (D) Immunoprecipitation of pIL22 labeled polypeptides precipitated with AFA-I antibodies. (E) Illustrations of the polar effect of the Tn*5* insertion into the 6.7-kb insert. (F) Polypeptides expressed from two different subcloned pIL14 restriction fragments. The arrowheads indicate the truncated form of the 100,000-dalton polypeptide.

cipitate material was subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis with a 12.5% sodium dodecyl sulfate gel (27). Standard proteins with molecular weights ranging from 94,000 to 14,000 (low-molecular-weight kit from Bio-Rad Laboratories) were run in parallel. The gel was stained and examined by fluorography with En^{3}Hance (New England Nuclear Corp.).

Immunoprecipitation. Immunoprecipitation of [^{35}S]methionine-labeled plasmid hybrid-encoded proteins was performed essentially as described by Dallas and Falkow (3), except that protein A bound to Sepharose was used instead of *Straphylococcus aureus* cells.

Nucleotide sequence. Appropriate DNA fragments were cloned into M13mp8 and M13mp9 (19). Plaques containing inserts were identified by using X-gal and isopropyl-β-D-thiogalactopyranoside. Single-stranded DNA templates were prepared by the polyethylene glycol method (27), and the sequence was determined by the dideoxynucleotide chain termination method of Sanger et al. (28).

RESULTS

Polypeptides expressed from pIL14. The pIL14 plasmid corresponds to the insertion of 6.7 kb of *E. coli* KS52 chromosomal DNA fragment into pBR322. It represented the smallest in vitro-constructed hybrid that expressed all of the hemagglutination and uroepithelial cell adhesion properties of the parental KS52 strain (Fig. 1) (16). Consequently we used this prototype plasmid for the analysis of genes and gene products required for the expression of those adherence properties. pBR322 and pIL14 were introduced into a minicell-producing strain, P678-54, and plasmid-dependent protein synthesis was examined by using purified minicells labeled with [^{35}S]methionine or [^{35}S]cysteine. Labeling with either amino acid demonstrated eight pIL14-specific polypeptides having apparent molecular masses of 102,000, 100,000, 33,000, 20,000, 18,500, 17,500, 16,000, and 13,000 daltons (Fig. 2A). These polypeptides were also expressed by the pIL03 cosmid (Fig. 2B), and we therefore concluded that none of them corresponded to a fusion product generated from the cloning procedure. As we previously described (16) (Fig. 1), pIL14 and pIL22 possess the same chromosomal DNA insert, but differ from each other by their orientation of insertion into the pBR322 vector. pIL22 expressed the same polypeptides as pIL14, but the intensity of those polypeptides relative to that of the pBR322-specific β-lactamase product was much lower in pIL22 than in pIL14 (Fig. 2C). This observation suggested that the level of expression of all the proteins encoded by the 6.7-kb insert was orientation dependent; the higher level of synthesis observed with pIL14 by comparison to that of pIL22 might be due, therefore, to the involvement of a promotor within the cloning vector.

Because in minicells the proteolytic processing (which corresponds to the cleavage of a signal peptide required for the transport of membrane protein across the cytoplasmic membrane of *E. coli*) is not fully effective (6), we presumed that some of the eight polypeptides might represent precursor forms of mature peptides. In subsequent experiments we inhibited the proteolytic processing in minicells by labeling them in the presence of 9.5% ethanol (20, 24). The pBR322-encoded β-lactamase, which was present as both an unprocessed polypeptide (32,000 daltons) and a processed peptide (29,000 daltons) in the absence of ethanol, was recovered as almost solely unprocessed peptide in the presence of ethanol (Fig. 2A). In the same way, the disappearance of the 16,000-, 18,500-, and 100,000-dalton pIL14-specific polypeptides correlated with the increase in intensity of the 20,000-, 17,500-, and 102,000-dalton bands, respectively (Fig. 2A). We concluded that the 17,500-, 20,000-, and 102,000-dalton peptides represented the precursors of the 16,000-, 18,500-, and 100,000-dalton polypeptides, respectively. The 33,000-dalton polypeptide, at this point, can be considered either as a cytoplasmic unprocessed protein or as the precursor of a 30,000-dalton polypeptide that would comigrate with the unprocessed precursor of the pBR322 β-lactamase. To answer this question, the 6.1-kb *Hin*dIII fragment of pIL14 was cloned into pACY184. This hybrid plasmid, introduced into a minicell-producing strain, expressed a 33,000-dalton poly-

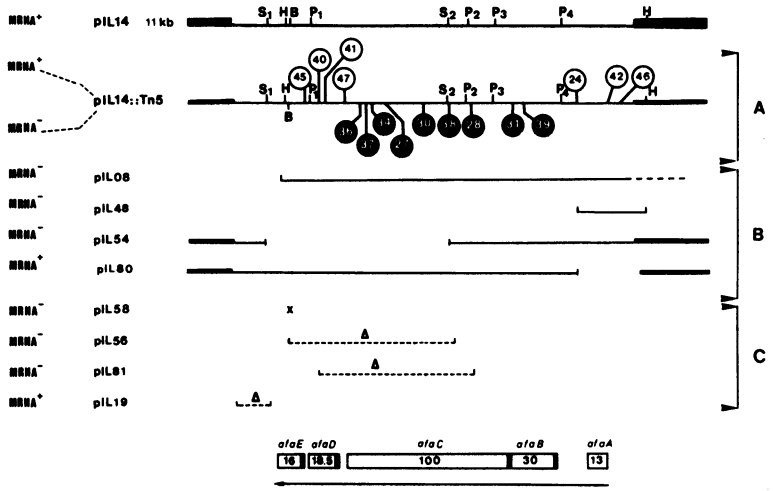

FIG. 3. Genetic organization of the *afa* operon was determined by the following approaches: (A) random insertion of Tn*5*, (B) subcloning of various part of the operon, or (C) in vitro generation of small insertion (×) or deletion (△) as described in the text. The letters used in the physical map correspond to those used in Fig. 1. The arrow indicates the direction of transcription. The dark circles indicate Tn*5* insertions that abolish MRHA expression (MRHA$^-$); the open circles indicate Tn*5* insertions in pIL14 that, when harbored in *E. coli* K-12 HB101 or P678-54, do not alter the MRHA expression (MRHA$^+$). The locations of the genes are indicated by bars (□). The lengths of the bars correspond to the lengths of DNA required to code for the polypeptides. The numbers in the bars indicate the apparent molecular masses of the mature polypeptides in kilodaltons. The solid parts of the bars (■) indicate the parts of the genes encoding the signal peptides. Shown are the pIL08 11-kb *Hin*dIII fragment of pIL03 cloned in pBR322 (11) and the pIL48 2.7-kb fragment of pIL24 cloned in pACY184.

peptide processed as a 30,000-dalton mature polypeptide. In contrast, the 13,000-dalton pIL14 polypeptide did not disappear when labeled in the presence of ethanol, and we believe that it is a cytoplasmic protein. The immunoprecipitation of labeled polypeptides from minicells with an antiserum raised against the purified hemagglutinin, AFA-I, isolated from culture supernatant fluid, revealed that only the 16,000-dalton polypeptide and its 17,500-dalton precursor polypeptide were specifically precipitated with the AFA-I antibodies (Fig. 2D). This result confirms the identification of the 16,000-dalton polypeptide and its precursor as the hemagglutinin protein (16). Our data are consistent with the view that five polypeptides were expressed from the 6.7-kb cloned DNA fragment. We therefore wanted to identify among the polypeptide-encoding genes those required for MRHA expression.

Identification and location of genes required for the MRHA expression. To determine the DNA region required for pIL14-mediated MRHA expression, pIL14 derivatives carrying the Tn*5* insertion were isolated and analyzed for the expression of MRHA phenotype in an HB101 host strain. The precise site of Tn*5* insertion was identified by digesting the pIL14 derivatives with endonucleases *Hin*dIII and *Pst*I. Out of 16 insertions selected for study, 9 derivatives had entirely lost the capacity for MRHA, whereas 7 still expressed the MRHA phenotype (Fig. 3.) These results defined a large region absolutely required for the expression of the hemagglutination phenotype. In addition other available data permitted us to define the regions of pIL14 essential for the MRHA phenotype. These data included the following. (i) The pIL19 hybrid plasmid (16) is formed of a DNA insert identical to that of pIL22 (Fig. 1), except that the DNA sequence from the left end of the pIL22 6.7-kb insert to the *Sma*I-1 site is missing in pIL19. pIL19 still expresses MRHA (Fig. 3C) and expresses the same polypeptides as pIL22 and pIL14. (ii) In contrast, the pIL08 hybrid plasmid missing the DNA sequence between the *Sma*I-1 site and the *Hin*dIII site of the pIL14 insert is MRHA negative. Therefore some DNA sequences between the *Sma*I-1 site and the *Hin*dIII site are required for MRHA expression. Two regions of the 6.7-kb insert of pIL14 seemed not to be involved in MRHA expression; Tn*5* insertions into the right end of the pIL14 insert or around the *Pst*I-1 site did not affect the MRHA expression of the host strain.

The five genes (*afaA*, *afaB*, *afaC*, *afaD*, *afaE*) encoding the pIL14 products, the 100,000-, 30,000-, 18,000-, 16,000-, and 13,000-dalton polypeptides, were then localized by transforming pIL14 derivative hybrids into minicell-producing strains and analyzing each of their products after labeled amino acid in vitro synthesis on sodium dodecyl sulfate-polyacrylamide gels (Table 1, Fig. 3).

The introduction of the encoding Tn*5* sequences into pIL14 had two major consequences on the peptide synthesis. (i) It led to the addition of several Tn*5*-specific polypeptides, among which some comigrated with the investigated pIL14 polypeptides (i.e., comigration of Tn*5*-encoded polypeptides with the 20,000- and 18,000-dalton pIL14 polypeptides; Fig. 2E). (ii) All of the Tn*5* insertions into the 6.7-kb DNA fragment had a strong polar effect on the level of synthesis of some polypeptides adjacent to the site of Tn*5* insertion (Fig. 2E). These observations suggested the existence of a large transcriptional unit read from the right to the left end of the pIL14 6.7-kb insert, relative to the genetic map shown in Fig. 3.

No disappearance or modification of the investigated pIL14 polypeptides was detected in the analysis of the seven MRHA-positive pIL14::Tn*5* derivatives. However, with the exception of pIL31 and pIL39, the sodium dodecyl sulfate-polyacrylamide peptide analyses of the MRHA-negative pIL14::Tn*5* derivatives showed the disappearance of the 100,000- to 102,000-dalton polypeptides. The other polypeptides remained unchanged in size. Nonetheless, whereas the 13,000- and 30,000-dalton polypeptides were present with the same relative concentration as in pIL14, the amount of the 16,000-dalton polypeptide synthesized was strongly reduced (Fig. 2E). The analysis of the polypeptides expressed

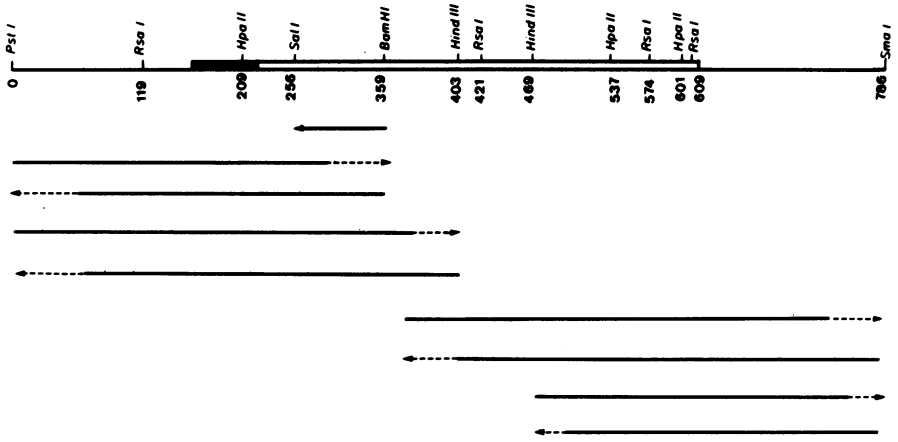

FIG. 4. Strategy for sequencing the AFA-I structural gene (*afaE*) located in the pIL14 hybrid plasmid. DNA restriction fragments were separated on agarose slab gels and electroeluted as described in the text. The indicated fragments were cloned into M13 phage vectors mp8 and mp9. The DNA sequence was determined on both strands by the dideoxynucleotide chain termination method (28). Symbols: (■) AFA-I signal peptide-encoding sequence, (□) AFA-I mature polypeptide-encoding sequence.

from pIL36, pIL37, pIL34, and pIL27 revealed the presence of new truncated polypeptides with decreasing molecular weights (Fig. 2E), attesting that the *afaC* gene, encoding the 100,000-dalton polypeptide, was transcribed from right to left. We believe that the Tn*5* insertion in pIL36 localized a site corresponding to the C terminal of the 100,000-dalton polypeptide, whereas the Tn*5* insertion within pIL28 corresponded to the N terminal of this polypeptide.

The pIL31 and pIL39 polypeptide patterns showed the disappearance of the 30,000-dalton polypeptide encoded by a gene we have designated as *afaB*. These insertions also caused reduced synthesis of the 100,000- and 16,000-dalton polypeptides, whereas the 13,000-dalton protein intensity was unchanged (data not shown).

None of the analyzed pIL14::Tn*5* derivatives showed any obvious effect on the 16,000- and 13,000-dalton polypeptides. Because the pIL08 hybrid did not synthesize the 16,000-dalton polypeptides in minicells, we suspected that the gene specifying the 16,000-dalton hemagglutinin was located on the left side of the 6.7-kb insert (Fig. 3). A plasmid, pIL58, was constructed by filling the recessed 3′ ends of *Bam*HI-cut molecules and had a MRHA-negative phenotype. Only the 16,000-dalton polypeptide (as well as its 17,500-dalton precursor form) disappeared in the polypeptide profile analysis of pIL58 (Fig. 2F). This result permitted us to localize *afaE* as the hemagglutinin structural gene.

The 13,000-dalton polypeptide was synthesized by all of the deleted pIL14 derivatives. This indicated that the gene *afaA* was not located between the *Sma*I-1 and the *Pst*I-4 sites of the insert. Using the additional *Hin*dIII site provided by the insertion of Tn*5* in the pIL24 derivative, we subcloned, into pACY184, a 2.7-kb *Hin*dIII fragment corresponding to the right end of the 6.7-kb insert (pIL48) (Fig. 3). This hybrid plasmid synthesized only the 13,000-dalton polypeptide (Fig. 2F). The pIL24, pIL42, and pIL46 plasmid derivatives also synthesized this polypeptide (data not shown). Our data led us to believe that the *afaA* gene should be located between the sites of insertion of Tn*5* in pIL24 and pIL42. This was confirmed by constructing the pIL80 hybrid plasmid, which resulted from the ligation of a partial *Hin*dIII digest of pIL24. It synthesized, at a very low level, all but the 13,000-dalton polypeptides encoded by pIL14 and still retained the MRHA-positive phenotype. Hence, the requirement of the synthesis of the 13,000-dalton polypeptide did not seem essential for MRHA expression, but a potential role for this protein is discussed below.

The location of the 18,500-dalton encoding gene (*afaD*) was investigated and deduced from the protein synthesis analyses of the pIL54, pIL56, pIL58, and pIL81 derivatives. Unlike pIL14 and pIL58, these latter plasmid derivatives did not synthesize the 18,500-dalton polypeptide or its 20,000-dalton precursor. Thus, the *afaD* gene should be located between those encoding the 100,000- and 16,000-dalton polypeptides, a domain where none of the Tn*5* insertions inactivated the MRHA phenotype. Therefore, the *afaD* gene did not seem to be required for MRHA expression.

Nucleotide sequence of the *afaE* gene. From our analysis of the available plasmid derivatives, we concluded that the *afaE* gene encoding the 16,000-dalton hemagglutinin, AFA-I, was located between the *Sma*I-1 site and the *Pst*I-1 site of pIL14. Various asymmetrical cloning of *Sma*I-*Hin*dIII, *Sma*I-*Bam*HI, and *Pst*I-*Bam*HI fragments in M13mp8 and M13mp9 allowed insertions in both orientation and the determination, on both strands, of the nucleotide sequence of the entire *Sma*I-*Pst*I fragment (Fig. 4.).

The search for initiation and termination codons on the two strands and in the three reading frames revealed a unique open reading frame in the sequenced DNA fragment. It started (ATG) at position 161 and ended (TGA) at position 616 (Fig. 5). This 456-base-pair open reading frame codes for a 152-residue protein that contains a 21-residue peptide with all the features of a procaryotic signal sequence (12); namely, a short hydrophobic region with two positively charged lysines in positions 2 and 3, a long hydrophobic region (9 of 12 amino acids, from residue 3 to 15, are hydrophobic), and a likely cleavage site adjacent to an alanine residue. The precise position of the first residue of the 132-amino-acid mature protein has been determined by peptide analysis of the purified hemagglutinin (M. A. Schmidt, W. Walz, A. Labigne-Roussel, S. Falkow, and G. Schoolnik, submitted for publication). The first 20 amino acids of this purified material corresponded exactly to the translation of the 20 codons in the open reading frame from base 224 to 283, and the amino acid compositions deduced from the DNA sequence and from the protein hydrolysis were similar.

The codon ATG at bases 161 to 163 is preceded by a potential ribosomal binding site at positions 146 to 152 (-TGAATAGG-) (29) and by a sequence (TAxxxT at posi-

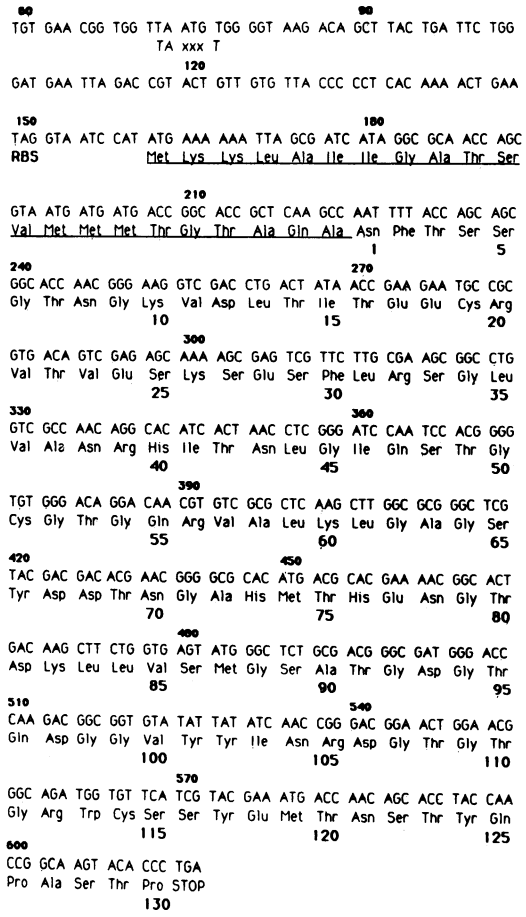

FIG. 5. Nucleotide sequence of the AFA-I structural gene (*afaE*). The amino acid sequence of the deduced polypeptide product is shown beneath the nucleotide sequence. Numbering of the nucleotides begins at the *Pst*I site (above each line); numbers below each line refer to the amino acid position. RBS, Ribosome-binding site.

tions 72 to 77) showing homologies with a consensus promoter sequence (8).

DISCUSSION

The uropathogenic *E. coli* strain KS52 expresses hemagglutination of human erythrocytes and adhesion to human uroepithelial cells. We previously showed that a 6.7-kb fragment of chromosomal DNA cloned into pBR322 pIL14 was required for the expression of this phenotype. The cloned fragment encodes the production of an extracellular hemagglutinin, AFA-I, having the same properties as the KS52 strain. Unlike other *E. coli* operons expressing MRHA (*pap* operon and CFA-I, K88, K99 antigen-encoding operons [5, 13, 20]), no obvious fimbriae or pili were associated with the expression of the KS52 hemagglutinin and adherence properties. In this study we describe the genetic and molecular organization of the KS52 genetic elements of the plasmid pIL14 involved in MRHA and adherence expression. The pIL14 plasmid was found to encode at least five polypeptides. The location and the relative order of the five genes coding for the polypeptides identified was determined, and each gene was associated with a gene product (Fig. 3). The DNA sequence included between the *Sma*I-1 site and the Tn*5* insertion site in pIL42 has a coding capacity that corresponds to the length of the DNA required to encode the identified five polypeptides.

The level of synthesis of the five polypeptides appeared to be coordinated; in pIL14 the five polypeptides, presumably under the control of the pBR322 promoter, are overproduced compared with their level of expression in the original pIL03 cosmid. In contrast the same insert cloned in pBR322 in the opposite direction (pIL22) showed a drastic reduction of the level of synthesis of all of the five polypeptides in minicells. This suggests that the five genes belong to the same transcriptional unit and that we did not clone in the 6.7-kb DNA fragment, the regulatory region of the *afa* operon. However, the fact that pIL22, in the absence of the pBR322 promoter, still produces a low amount of the five polypeptides indicated that each gene or group of genes can be transcribed from weak promoters. The presence of such a weak promoter was revealed in the nucleotide sequence data: the *afaE* gene is preceded by a region exhibiting only weak complementarity to the 3′ end of the 16S RNA (8). The presence of a large transcriptional unit and the direction of transcription of the different genes (from *afaA* to *afaE*, Fig. 3) was confirmed by the strong polar effect observed downstream from any Tn*5* insertion and the sequence of the *afaE* gene.

Genes required for MRHA expression. Among the five genes belonging to the *afa* operon, only three are required for MRHA expression: *afaB*, *afaC*, and *afaE*.

The *afaE* encodes the adhesin or hemagglutinin AFA-I polypeptide. The adhesin polypeptide was identified by precipitation with immune serum raised against the hemagglutinin purified from the supernatant of *E. coli* K-12 strains harboring pIL14 (16). The processed form of this structural gene product comigrates with the purified hemagglutinin and was found to have on polyacrylamide gel an apparent molecular weight of 16,000. It is synthesized as a 17,500-dalton precursor detectable in minicells. A small insertion in this structural gene (pIL58) or a deletion of the *afaE* gene (pIL08) results in the abolition of the MRHA phenotype and of the adherence properties of *E. coli* K-12 harboring pIL58 and in the disappearance of the 16,000-dalton polypeptide in minicells. These results clearly identify the *afaE* gene product as the AFA-I hemagglutinin, responsible for the specific adhesion to uroepithelial cells and to human erythrocytes.

The nucleotide sequence of the *afaE* region shows a unique open reading frame of 152 amino acids; the first 21 amino acids constitute the signal sequence. The presence of the four methionine residues in this leader peptide accounts for the very strong radioactive signal observed at the level of the unprocessed 17,500-dalton polypeptide, compared with that of the mature protein, when [^{35}S]methionine was used to label the minicells. The amino acid sequence deduced from the DNA sequence led to a mature protein with a calculated molecular mass of 13,100 daltons (131 residues). Schmidt et al. (in press), by analyzing the amino acid composition of the purified AFA-I protein, obtained the same estimate of the molecular weight of the polypeptide. The discrepancy observed between the AFA-I molecular weight deduced from the DNA sequence or the amino acid composition and that visualized from the polyacrylamide gels suggests that AFA-I might be modified after its translation. Comparison of the DNA sequence and the deduced amino acid sequence of AFA-I with sequences of fimbriae subunits, encoded by various MRHA operons (data not shown), did not reveal homology or relevant common features between the sequences.

The *afaB* gene is also required for the MRHA expression.

A 30,000-dalton polypeptide synthesized as a 33,000-dalton precursor is clearly associated with the expression of *afaB*. *E. coli* strains harboring derivative plasmids missing the 30,000-dalton polypeptide synthesized a normal, processed 16,000-dalton hemagglutinin in minicells. The *afaB* gene product, therefore, does not play an obvious role in the biosynthesis or the maturation of the AFA-I hemagglutinin.

The *afaC* gene codes for a 100,000-dalton polypeptide synthesized as a 102,000-dalton precursor. This indicates that the *afaC* gene product (required for MRHA expression) is transported through the cytoplasmic membrane by means of a signal sequence. Preliminary results showed that in *E. coli* strains harboring a derivative plasmid missing the 100,000-dalton polypeptide, the AFA-I subunit is synthesized normally. However, in these latter strains the AFA-I subunit is not transported to or associated within the outer membrane. Hence, the 100,000-dalton polypeptide might be involved in the anchorage of the adhesin in the outer membrane, a function which has been suggested for the large outer membrane protein identified in other MRHA-encoding operons (21). Two other genes belong to the *afa* operon: the *afaA* gene encoding the cytoplasmic 13,000-dalton polypeptide and the *afaD* gene expressing an 18,500-dalton polypeptide synthesized as a 20,000-dalton precursor. The inactivation of *afaA* or *afaD* was not correlated with the modification of MRHA or the adherence phenotype. The functions of these two proteins are unknown. However, the properties associated with the *afaA* gene product (location of the encoding gene at the beginning of the operon, size of the product) are reminiscent of the properties of the regulatory proteins in the operons expressing MRHA: in the *pap* operon, the B gene product (13,000 daltons) is described as a positive regulator activating the *papA* promotor (21). A regulatory protein involved in the expression of the K88ab genes was described as an antiterminator of transcription (F. R. Mool, Ph.D. thesis, Vrije Universiteit, Amsterdam, The Netherlands, 1982). The experiments that we carried out with pIL14 and its derivative plasmids do not allow us to define a regulatory function of the *afaA* gene. Experiments are in progress to clone the regulatory region of the *afa* operon and to determine the nature and role of the 13,000-dalton polypeptide.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grant AI18719 from the National Institutes of Health. A.L.-R. was supported by a fellowship from the Fondation de la Recherche Medicale and from the Philippe Foundation and is a fellow of the Pasteur-Institute (France).

We thank David Low for continuing interest and suggestions throughout the course of this work.

LITERATURE CITED

1. **Alder, H. I., D. Fisher, A. Cohen, and A. A. Hardlget.** 1967. Miniature *E. coli* cells deficient in DNA. Proc. Natl. Acad. Sci. U.S.A. **57**:321–326.
2. **Boyer, H. W., and D. Roulland-Dussoix.** 1969. A complementation analysis of the restriction and modification of DNA in *E. coli*. J. Mol. Biol. **41**:459–472.
3. **Dallas, W., and S. Falkow.** 1979. The molecular nature of heat-labile enterotoxin (LT) of *Escherichia coli*. Nature (London) **277**:406–407.
4. **Davis, R. W., D. Bostein, and J. R. Roth.** 1980. Advanced bacterial genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
5. **de Graaf, F. K., B. E. Krenn, and P. Klaasen.** 1983. Organization and expression of genes involved in the biosynthesis of K99 fimbriae. Infect. Immun. **43**:508–514.
6. **Dougan, G., M. Saul, A. Twigg, R. Gill, and D. Sherratt.** 1979. Polypeptides expressed in *Escherichia coli* K-12 minicells by transposition elements Tn*1* and Tn*3*. J. Bacteriol. **138**:48–54.
7. **Gill, R. E., F. Heffron, and S. Falkow.** 1979. Identification of the protein encoded by the transposable element Tn3 which is required for its transposition. Nature (London) **282**:797–801.
8. **Gold, L., D. Pribnow, T. Schneider, S. Shineding, B. S. Singer, and G. Stormo.** 1981. Translational initiation in procaryotes. Annu. Rev. Microbiol. **35**:365–403.
9. **Hagberg, L., U. Jodal, T. K. Korhonen, G. Lidin-Johnson, U. Linberg, and C. Svanborg-Eden.** 1981. Adhesion, hemagglutination, and virulence of *Escherichia coli* causing urinary tract infections. Infect. Immun. **31**:564–570.
10. **Hugues, C., R. Philips, and A. P. Roberts.** 1982. Serum resistance among *Escherichia coli* strains causing urinary tract infection in relation to O type and the carriage of hemolysin, colicin, and antibiotic resistance determinants. Infect. Immun. **35**: 270–275.
11. **Hull, R. A., R. E. Gill, P. Hsu, B. H. Minshew, and S. Falkow.** 1981. Construction and expression of recombinant plasmids encoding type 1 or D-mannose-resistant pili from a urinary tract infection *Escherichia coli* isolate. Infect. Immun. **33**:933–938.
12. **Inouye, M., and S. Halegoua.** 1980. Secretion and membrane localization of proteins in *Escherichia coli*. Crit. Rev. Biochem. **7**:339–371.
13. **Kallenius, G., R. Mollby, S. B. Svenson, J. Windberg, A. Lundblad, S. Svenson, and B. Cedergren.** 1980. The P^k antigen as receptor for the haemagglutination of pyelonephritic *Escherichia coli*. FEMS Microbiol. Lett. **7**:297–302.
14. **Kallenius, G., S. B. Svenson, R. Mollby, B. Cedergren, and H. Hultberg.** 1981. Structure of carbohydrate part of receptor on human uroepithelial for pyelonephritogenic *Escherichia coli*. Lancet **ii**:604–606.
15. **Kallenius, G., S. B. Svenson, R. Mollby, T. Korhonen, J. Winberg, B. Cedergren, I. Helin, and H. Hultberg.** 1982. Carbohydrate receptor structures recognized by uropathogenic *E. coli*. Scand. J. Infect. Dis. Suppl. **33**:52–60.
16. **Labigne-Roussel, A., D. Lark, G. Schoolnik, and S. Falkow.** 1984. Cloning and expression of an afimbrial adhesin (AFA-I) responsible for P-blood-group independent mannose resistant hemagglutination from a pyelonephritic *Escherichia coli* strain. Infect. Immun. **46**:251–259.
17. **Lindberg, F. P., B. Lund, and S. Normark.** 1984. Genes of pyelonephritogenic *E. coli* required for digalactoside specific agglutination of human cells. EMBO J. **3**:1167–1173.
18. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1983. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
19. **Messing, J., and J. Vieira.** 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene **19**:269–276.
20. **Mooi, F., R. G. Koopman, B. Roosendaal, and F. K. de Graaf.** 1984. K88ab gene of *Escherichia coli* encodes a fimbria-like protein distinct from K88ab fimbrial adhesin. J. Bacteriol. **159**:482–487.
21. **Norgren, M., S. Normark, D. Lark, P. O'Hanley, G. Schoolnik, S. Falkow, C. Svanborg-Eden, M. Baga, and B. E. Uhlin.** 1984. Mutations in *E. coli* cistrons affecting adhesion to human cells do not abolish Pap pili fiber formation. EMBO J. **3**:1159–1169.
22. **Normark, S., D. Lark, R. Hull, M. Norgren, M. Baga, P. O'Hanley, G. Schoolnik, and S. Falkow.** 1983. Genetics of digalactoside-binding adhesion from a uropathogenic *Escherichia coli* strain. Infect. Immun. **41**:942–949.
23. **O'Hanley, P., D. Lark, S. Normark, S. Falkow, and G. Schoolnik.** 1983. Mannose-sensitive and Gal-Gal binding *Escherichia coli* pili from recombinant strains. J. Exp. Med. **158**:1713–1719.
24. **Palva, E. T., J. R. Hirst, S. J. S. Harfy, J. Holmgren, and L. Randall.** 1981. Synthesis of a precursor to the B subunit of heat labile enterotoxin in *Escherichia coli*. J. Bacteriol. **146**:325–330.
25. **Parkkinen, J., J. Finne, M. Achtman, V. Vaisanen, and T. K. Korhonen.** 1983. *Escherichia coli* strains binding neuraminyl 2-3

galactosides. Biochem. Biophys. Res. Commun. **111**:456–461.
26. **Portnoy, D. A., S. L. Moseley, and S. Falkow.** 1981. Characterization of plasmids and plasmid-associated determinants of *Yersinia enterocolitica* pathogenesis. Infect. Immun. **31**:775–782.
27. **Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe.** 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143**:161–178.
28. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463–5467.
29. **Stormo, D. G., T. D. Schneider, and L. M. Gold.** 1982. Characterization of translational initiation sites in *E. coli*. Nucleic Acids Res. **10**:2971–2996.
30. **Svanborg-Eden, C., L. A. Hanson, U. Jodal, U. Lindberg, and A. S. Akerlund.** 1976. Variable adhesion to normal urinary tract epithelial cells of *Escherichia coli* strains associated with various forms of urinary tract infection. Lancet **ii**:490–492.
31. **Vaisanen, V., T. K. Korhonen, M. Jokinen, C. G. G. Ahmberg, and C. Ehnholm.** 1982. Blood group M specific haemagglutinin in pyelonephritogenic *Escherichia coli*. Lancet **i**:1192.
32. **Vaisanen, V., L. G. Tallgren, P. H. Makela, G. Kallenius, H. Hultberg, J. Elo, A. Siitonenm, C. Svanborg-Eden, S. B. Svenson, and T. Korhonen.** 1981. Mannose-resistant haemagglutination and P antigen recognition are characteristic of *Escherichia coli* causing primary pyelonephritis. Lancet **ii**:1366–1369.