INFECTION AND IMMUNITY, June 1993, p. 2505–2512
0019-9567/93/062505-08$02.00/0

Vol. 61, No. 6

Genetic Analysis of the Gene Cluster Encoding Nonfimbrial Adhesin I from an *Escherichia coli* Uropathogen

RALPH AHRENS,[1] MANFRED OTT,[2] ANGELIKA RITTER,[2] HEINZ HOSCHÜTZKY,[1] THOMAS BÜHLER,[1] FRIEDRICH LOTTSPEICH,[3] GRAHAM J. BOULNOIS,[4] KLAUS JANN,[1] AND JÖRG HACKER[2]*

Max-Planck Institut für Immunbiologie, D-W-7800 Freiburg,[1] Lehrstuhl für Mikrobiologie der Universität Würzburg, Röntgenring 11, D-W-8700 Würzburg,[2] and Max-Planck Institut für Biochemie, Munich,[3] Germany, and Department of Microbiology, University of Leicester, Leicester LE1 9HN, United Kingdom[4]

Received 29 December 1992/Accepted 16 March 1993

The chromosomally encoded nonfimbrial adhesin I (NFA-I) from *Escherichia coli* urinary tract isolate 827 (O83:K1:H4) mediates agglutination of human erythrocytes. Subclones were constructed from an NFA-I-expressing recombinant *E. coli* K-12 clone, derived from a genomic library of *E. coli* 827. Minicell analysis and nucleotide sequencing revealed that proteins of 30.5, 9, 80, 15, and 19 kDa encoded on a stretch of approximately 6 kb are involved in the expression of NFA-I. NFA-I exhibits a polymeric structure, which disintegrates with elevated temperature into a 19-kDa monomer but with some relatively stable dimers. By using gold-conjugated monoclonal antibodies directed against NFA-I in electron microscopy, the adhesin could be localized on the outer surface of the recombinant *E. coli* K-12 bacteria. The nucleotide sequence of the *nfaA* gene encoding the monomeric structural subunit of the adhesin was determined. An open reading frame of 184 amino acids encoding the NfaA precursor, which is processed to the mature protein, was found; it consisted of 156 amino acids with a calculated molecular weight of 16,000. Peptide sequencing of the NFA-I subunit protein confirmed that this open reading frame corresponds to the NfaA coding locus. Furthermore, the nucleotide sequence of the open reading frame termed NfaE, located at the proximal part of the DNA stretch responsible for NFA-I expression, was elaborated. NfaE consists of 247 amino acids, including a presumptive 29-amino-acid signal peptide, leading to a molecular weight of 24,000 for the mature protein. The *nfaE* sequence shares homology with the 27-kDa CS3 protein, which is involved in the assembly of CS3 fibrillae, and might encode the 30.5-kDa protein, detected in minicells.

Escherichia coli isolates may be the causative agents of urinary tract infections and cases of sepsis and newborn meningitis (23). Adherence of bacteria to eukaryotic cells contributes to the pathogenicity of these bacteria (29). The binding to the eukaryotic cell surface is often mediated by fimbrial adhesins, rodlike structures, which can be easily detected by electron microscopy (11). Adhesion properties can be tested by hemagglutination of strains, using various types of erythrocytes. In addition to fimbriated strains, *E. coli* isolates which expressed non- or A-fimbrial adhesins were described (5, 15). These adhesins were also able to confer hemagglutination but were devoid of any fimbrial structures (9, 10).

In the last years, various non- or A-fimbrial adhesins have been characterized (9, 12, 16, 17, 24). It has become evident that the nonfimbrial specific adhesion is mediated by single proteins located on the bacterial surface. To gain insight into the genetic basis of non- or A-fimbrial adhesins, the determinants coding for some of these adhesins have been cloned (12, 17, 24). Cosmid clones derived from a genomic library of *E. coli* 827 (O83:K1:H4) which were able to express nonfimbrial adhesin I (NFA-I) were isolated by Hales et al. (12). In this study, we have further characterized the gene cluster coding for NFA-I by subcloning and analyzing the gene products. In addition, the nucleotide sequences of the structural gene *nfaA* and another gene, termed *nfaE*, were determined.

* Corresponding author.

MATERIALS AND METHODS

Bacterial strains and plasmids. *E. coli* 827 (O83:K1:H4) was isolated from urine and blood of an elderly patient (9). Cosmids pGB3002 and pGB3004 were described recently (12). A 15.5-kb *Bam*HI-*Eco*RI fragment derived from pGB3002 was cloned into pBR328, resulting in plasmid pGB3003 (12). After Tn*1000* (γδ) mutagenesis of this plasmid, a 10-kb *Eco*RI fragment of a γδ insertion mutant was further subcloned into pUC19, leading to plasmid pPS3, which contained parts of the γδ element. Plasmid pPS3 then was used to generate various subclones, depicted in Fig. 1. As recipients, *E. coli* K-12 CC118 and JM109 were used. Fragments were inserted into the cloning vector pUC18 or pUC19 (35).

Media and reagents. Bacterial strains were grown on Luria-Bertani (LB) agar plates or in liquid LB medium, as described before (12). For maintenance of plasmids, 50 μg of ampicillin per ml of medium was added. For detection of the Lac phenotype, LB medium was supplemented with IPTG (isopropylthiogalactoside; 0.05 mM) and X-Gal (5-bromo-4-chloro-3-indolyl-galactoside; 0.01%). Restriction enzymes, T4 ligase, and Klenow enzyme were purchased from Boehringer, Mannheim, Germany. Antibiotics were from Bayer, Leverkusen, Germany. Reagents for growth media were obtained from Oxoid, Wesel, Germany. All other chemicals were obtained from Sigma, Munich, Germany. Radiochemicals were purchased from NEN, Dreieich, Germany.

Recombinant DNA techniques. Plasmid DNA was isolated by the method of Birnboim and Doly (2). DNA was cleaved with restriction enzymes in accordance with the manufacturer's instructions and separated on 1% agarose gels, as described by Sambrook et al. (26). Isolation of DNA frag-

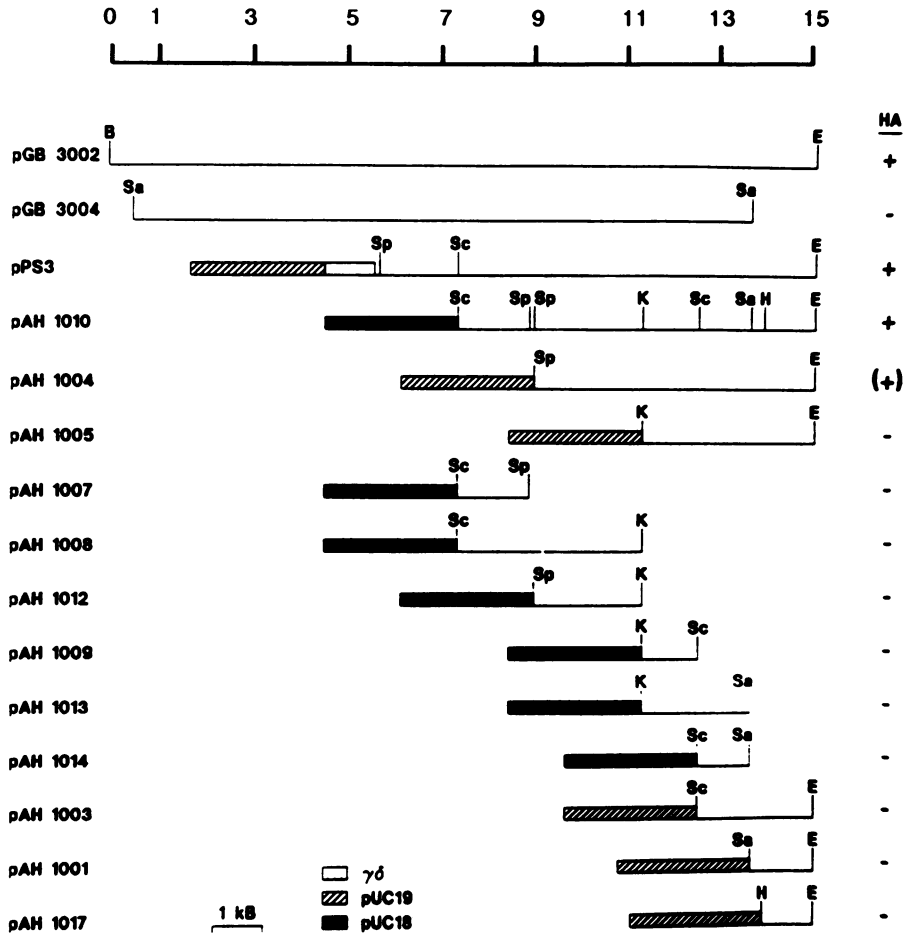

FIG. 1. Physical maps of plasmids pGB3002, pGB3004, and pPS3 and derived subclones. Coordinates are given at the top. Plasmids are given on the left. On the right side hemagglutination activity (HA) with human erythrocytes is indicated. Restriction sites: Sp, *Sph*I; Sc, *Sac*I; Sa, *Sal*I; K, *Kpn*I; H, *Hin*cII; E, *Eco*RI; B, *Bam*HI.

ments from agarose gels was performed by the freeze-squeeze method of Thuring et al. (31). Ligation of DNA fragments was carried out as described previously (26). Competent cells for DNA transformation were prepared by the $CaCl_2$ method (19).

Southern hybridization. DNA fragments separated by agarose gel electrophoresis were transferred to nitrocellulose paper as described previously (26). Synthetic oligonucleotides were prepared as described by Becauge and Caruthers (1a) and labelled by T4 polynucleotide kinase reaction (26), using [γ-^{32}P]ATP. The sequence of the oligonucleotide DNA probe was deduced from the N-terminal amino acid sequence of the NFA-I structural subunit and reads as follows: 5'-AACGTAAACGCTGGCGATGG-3'.

DNA sequencing. DNA sequencing was performed by the dideoxy chain termination method of Sanger et al. (27), using [α-^{32}P]dATP. Double-stranded DNA of recombinant pUC plasmids was used for sequencing. The primers were synthetic oligodeoxyribonucleotides, prepared by an automated phosphoramidite coupling method (1a). Sequencing was carried out with a T7 sequencing kit (Pharmacia, Freiburg, Germany), including universal primer (35).

Computer analysis. General compilation and analysis of DNA sequences were performed with the UWGCG programs obtained from Devereux (3). The software package of PC Gene was also used (Intelli-Genetics, Geneva, Switzerland [13]).

Minicell analysis and SDS-PAGE. For analysis of plasmid-encoded proteins, the recombinant *nfaI*-specific DNAs were transformed in *E. coli* DS410 (4). Proteins were labelled with ^{35}S-methionine (21). Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed by the method of Laemmli (18). For prevention of signal peptide processing, 9% ethanol was added to the assay medium.

Isolation of NFA-I and Western blot analysis. Bacteria grown on LB agar at 37°C for 24 h were harvested into phosphate-buffered saline (PBS; pH 7.2). Preparation of NFA-I from this suspension was carried out as described recently (12). Western blot (immunoblot) analysis was performed as described by Towbin et al. (32).

Preparation of MAbs. Anti-NFA-I monoclonal antibodies (MAbs) were prepared by immunization with purified NFA-I as described before (20). MAb 1C3 was isolated as an NFA-I-specific MAb.

Peptide sequencing. Isolated NFA-I was subjected to *Staphylococcus aureus* (V8) endoprotease digestion (8). The peptides obtained were separated by SDS-PAGE and electroblotted onto a siliconized glass-fiber sheet (glassy bond; Biometra, Göttingen, Germany) as described by Eckerskorn

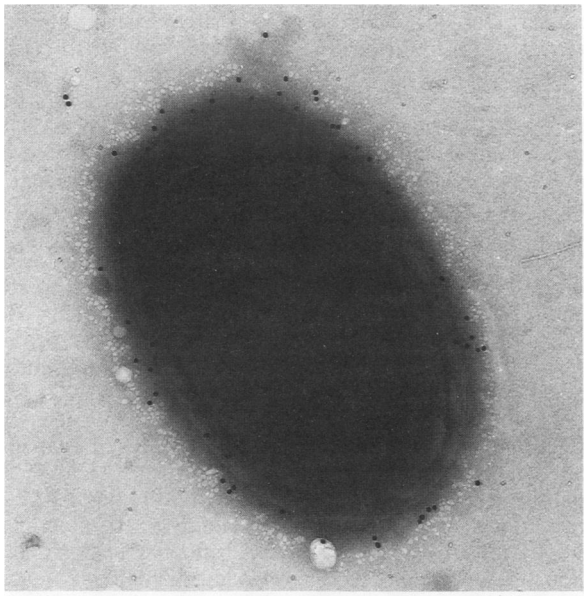

FIG. 2. Electron micrograph of *E. coli* JM109 (pAH1010). Immunogold labelling was performed with MAb 1C3, detecting NFA-I. Magnification, ×40,000.

et al. (6). The bands were excised and sequenced in an Applied Biosystems 477 A gas phase sequencer.

Immunoelectron microscopy. Bacteria grown overnight at 37°C on agar were suspended in PBS and transferred to Formvar-coated copper grids. Labelling with anti-NFA-I MAbs and protein A-conjugated gold spheres was carried out as described previously (20). Electron microscopy was performed with a Phillips EM450 TIST electron microscope.

Hemagglutination assay. For rapid detection of NFA-I expression, recombinant *E. coli* clones grown on LB agar at 37°C were removed with a sterile toothpick and mixed with 20 μl of a 10% suspension of fresh human erythrocytes diluted in PBS. Agglutination was performed on glass slides, which were incubated on ice to enhance the agglutination reaction.

RESULTS

Subcloning of the *nfaI* determinant. Cosmid pBG3002 derived from a gene library of *E. coli* 827 (12) carries the DNA region responsible for expression of NFA-I. Subcloning and Tn*1000* mutagenesis revealed that the genes necessary for NFA-I production must be located between the *Sph*I site (coordinate 9.0) and beyond the *Sal*I site (coordinate 13.5; Fig. 1; cf. reference 12). We have constructed subclones starting with plasmid pPS3 consisting of vector pUC19 and an *Eco*RI fragment of 10.0 kb (see Materials and Methods). The recombinant *E. coli* K-12 clones were tested for hemagglutination of human erythrocytes, indicating the expression of NFA-I. Of the constructed plasmids (Fig. 1), only pPS3 and pAH1010 are able to confer strong hemagglutination to the *E. coli* K-12 host strain. Plasmid pAH1004 confers only weak hemagglutination activity. Thus, expression of NFA-I seems to be mediated by a stretch of approximately 6 to 8 kb, located between the *Sac*I site (coordinate 7.0) and the *Eco*RI site at coordinate 15.0.

Characterization of clone JM109 (pAH1010). With the anti-NFA-I MAb 1C3 in immunoelectron microscopic analysis (Fig. 2), NFA-I could be detected on the surface of the recombinant *E. coli* K-12 strain JM109 (pAH1010). Furthermore, NFA-I was isolated from the recombinant clone. The preparation was applied to SDS-PAGE (Fig. 3) both untreated (lane 1) and after treatment for 10 min at either 85°C (lane 2) or 100°C (lane 3). It can be seen that the high-molecular-weight NFA-I polymer disintegrates with elevated temperature to the 19-kDa monomer, which forms the structural subunit of NFA-I. Interestingly, a relatively stable NFA-I dimer could be observed; this was confirmed by peptide sequencing of the respective protein band (Fig. 3, lane 3; see below).

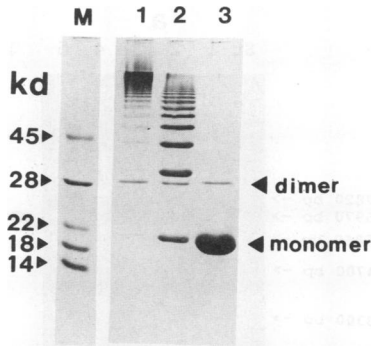

FIG. 3. SDS-PAGE of isolated NFA-I from recombinant *E. coli* JM109 (pAH1010). The preparation was suspended in Laemmli buffer and applied untreated (lane 1) and after treatment for 10 min at either 85°C (lane 2) or 100°C (lane 3). Molecular size standards are separated in lane M. Protein sizes are indicated.

Genetic organization of the *nfaI* determinant. To analyze the gene products of the *nfaI* determinant, plasmids pAH1010 and pAH1013 (Fig. 1) were transformed into *E. coli* DS410. In Fig. 4 the minicell products of the recombi-

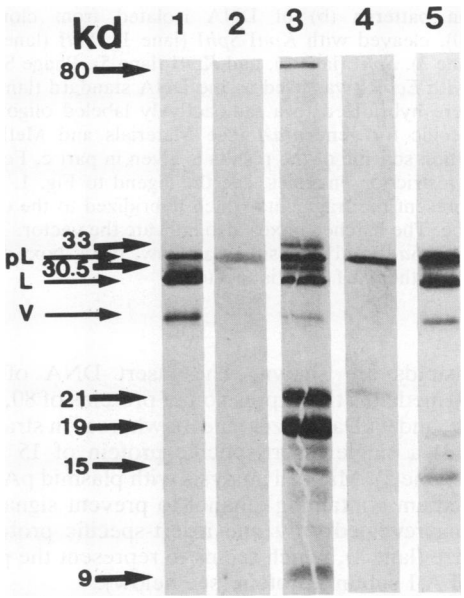

FIG. 4. Minicell analysis of DS410 carrying plasmid pUC18 without (lane 1) and with (lane 2) ethanol, plasmid pAH1010 without (lane 3) and with (lane 4) ethanol, and plasmid pAH1013 without ethanol (lane 5). Protein sizes are indicated. Prominent vector-specific proteins are indicated as pL (pre-lactamase), L (lactamase), and V (vector).

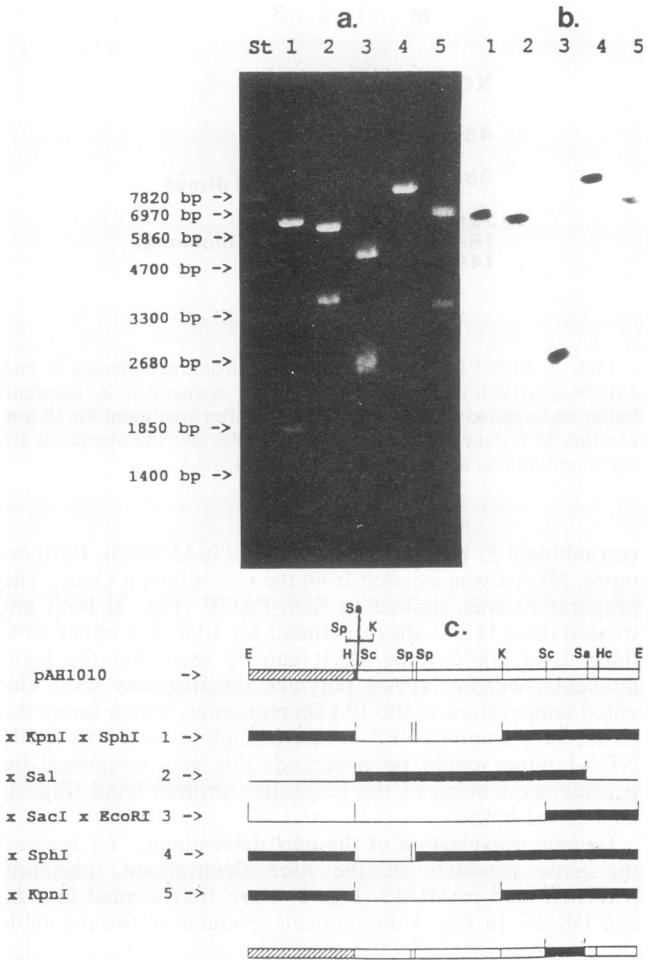

FIG. 5. Agarose gel electrophoresis (a) and Southern blot hybridization patterns (b) of DNA isolated from clone JM109 (pAH1010), cleaved with *Kpn*I-*Sph*I (lane 1), *Sal*I (lane 2), *Sac*I-*Eco*RI (lane 3), *Sph*I (lane 4), and *Kpn*I (lane 5). Phage SPP1 DNA cleaved with *Eco*RI was used as the DNA standard (lane St). The DNAs were hybridized to a radioactively labeled oligonucleotide probe specific for gene *nfaA* (see Materials and Methods). An interpretation scheme of the results is given in part c. For abbreviations of restriction enzymes, see the legend to Fig. 1. The black boxes represent the fragments which hybridized to the oligonucleotide probe. The hatched boxes also indicate the vector. The region between the *Sac*I and *Sal*I sites (last row, black box) defines the region where the *nfaA* gene is located.

nant plasmids are shown. The insert DNA of plasmid pAH1010 mediates the expression of proteins of 80, 33, 30.5, 21, 19, 15, and 9 kDa in size (lane 3), whereas in strain DS410 (pAH1013) a single insert-specific protein of 15 kDa was detected (lane 5). Minicell analysis with plasmid pAH1010 in assay medium containing ethanol to prevent signal peptide processing revealed only one insert-specific protein of 21 kDa in size (lane 4), which seems to represent the precursor of the NFA-I subunit protein (see below).

To determine the location of the gene encoding the 19-kDa NFA-I subunit, oligonucleotides generated according to the N-terminal amino acid sequence of the NFA-I structural protein (see Materials and Methods) were used in Southern hybridizations with cleaved DNA of plasmid pAH1010. As shown in Fig. 5, DNA fragments making up the distal part of the insert DNA of pAH1010 gave positive hybridization. By comparing the hybridization patterns of pAH1010 cleaved with different restriction enzymes or combinations of restriction enzymes, we concluded that the NFA-I structural gene is located at the distal end of the insert DNA of pAH1010 between map positions 13 and 15 in Fig. 1. From minicell analysis with plasmid pAH1013, it can be concluded that the gene encoding the 15-kDa protein should be located near the subunit-specific gene.

Sequence analysis of the region at the *Kpn*I site (coordinates 11.0 to 11.5; Fig. 1) revealed part of an open reading frame (ORF) sharing extended homology to proteins SfaF, PapC, and FanD, which represent the large anchor proteins (70 to 90 kDa) of the determinants encoding S-fimbrial adhesins, Pap fimbriae, and K99 fimbriae (22, 25, 28), respectively (data not shown). From these data it is concluded that the 80-kDa protein found in minicell analysis is located next to the 15-kDa protein, described above. The nucleotide sequence coding for the two proteins of 30.5 and 9 kDa in size would fit into the remaining space extending to the *Sph*I sites (coordinate 9). DNA sequence analysis of the region at coordinates 9 to 10.5 revealed an ORF of 247 amino acids, with a calculated molecular weight of approximately 27,000 (see below). After cleavage of a putative signal peptide, a protein of 24 kDa would be generated. It is suggested that this ORF corresponds to the 30.5-kDa protein detected in minicells, even though there are discrepancies in the exact molecular weight. Thus, we conclude that the genes encoding the proteins involved in NFA-I expression are located in the following order: 30.5, 9, 80, 15, and 19 kDa, which we termed NfaE, -D, -C, -B, and -A, respectively (Fig. 6).

Nucleotide sequence of the structural NFA-I subunit gene *nfaA*. The region between map positions 13.0 and 15.0 kb (Fig. 1 and 6) was sequenced and analyzed for the presence of an ORF encoding the NFA-I subunit protein. One ORF was found in this region. The nucleotide sequence and the deduced amino acids are given in Fig. 7. The precursor protein starts at an -ATG- codon at positions 139 to 141, and the gene ends at a stop codon at positions 691 to 693. Thus, *nfaA* is presumed to encode a protein of 184 amino acids.

Protein sequencing of NfaA. The 21 N-terminal amino acids of the mature NFA-I subunit protein (15) are identical to amino acids 29 to 50 deduced from the nucleotide sequence of *nfaA*, suggesting cleavage of a 28-amino-acid signal peptide. To confirm that the ORF of *nfaA* and the NFA-I structural protein are identical, peptides were generated following cleavage of NfaA with V8 endopeptidase. The amino acids at positions 52 and 53 as well as those at positions 74 and 75 were used as cleavage sites. N-terminal sequencing of the peptides showed amino acid sequences which are in full agreement with the amino acid sequence calculated on the basis of the nucleotide sequence (Fig. 7). It is therefore concluded that *nfaA* encodes the NFA-I subunit protein NfaA. As mentioned above, the polymeric NFA-I disintegrates into monomers upon heat treatment. Aside from the 19-kDa monomers, a protein band of 30 kDa is detected in preparations even after treatment at 100°C (Fig. 3). The N-terminal amino acids of this protein were sequenced. The sequence corresponds exactly to the N-terminal amino acid sequence elaborated for the 19-kDa NFA-I monomer (data not shown; cf. Fig. 7). Thus, this 30-kDa protein is thought to be a dimer of the 19-kDa NFA-I monomer.

Nucleotide sequence of the *nfaE* gene. As mentioned above, analysis of the nucleotide sequences between map positions

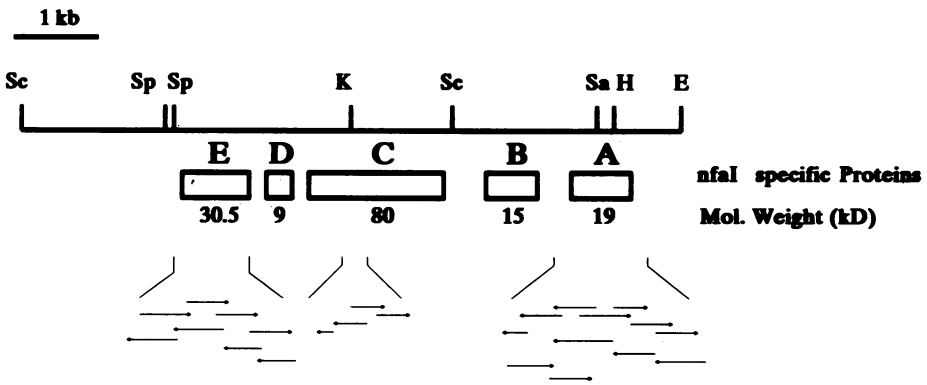

FIG. 6. Genetic organization of the *nfaI* gene cluster. The positions of the *nfaI*-specific genes (*nfaA* to *nafE*) are shown, and the respective molecular weights of the proteins are given. The strategy of sequencing is depicted below. For map positions and abbreviations of restriction enzymes, see the legend to Fig. 1.

```
TCAGAACTTACTTATATGCAATGAACAGTCTCTGCTGCGGGTACAGACATCTGTGAACGG   60

TGGTTAATGTGGGGTAATAGACAGCTTACTGATTCTGGGATGGATTAACAGAACACAACT  120

                    -28
GGCTTGTCCATAAGCAAAATGAAGGCAAAAAAATATGAAAATCAAATATACAATGAAAAT  180
          SD        MetLysAlaLysLysTyrGluAsnGlnIleTyrAsnGluAsn
                                                 -1 ↓+1
GGCCGCCGTTGCCAGCGTCATGGTCGCCGGCTAGCTATCGCGGATGCAAATGGGCTCAAC  240
GlyArgArgCysGlnArgHisGlyArgArgLeuAlaIleAlaAspAlaAsnGlyLeuAsn

ACTGTGAACGCCGGGGATGGCAAGAATCTGGGCACCGCAACCGCGACGATCACCACTCTG  300
ThrValAsnAlaGlyAspGlyLysAsnLeuGlyThrAlaThrAlaIleIleThrThrLeu

          SalI
CAGAGCTGCTCTGTCGACCTGAATCTCGTTACCCCGAACGCGACAGTGAACAGAGCAGGA  360
GlnSerCysSerValAspLeuAsnLeuValThrProAsnAlaThrValAsnArgAlaGly

ATGCTAGCAAACCGCGAAATCACTAAATTTTCGGTGGGGAGTAAGGATTGCCCTAGCGAC  420
MetLeuAlaAsnArgGluIleThrLysPheSerValGlySerLysAspCysProSerAsp

ACCTATGCTGTATGGTTTAAAGAGATCGATGGCGAAGGACAGGGGGTCGCGCAGGGCACT  480
ThrTyrAlaValTrpPheLysGluIleAspGlyGluGlyGlnGlyValAlaGlnGlyThr

ACGGTGACCAACAAGTTTTACCTTAAAATGACATCGGCCGACGGGACCGCGAGCGTAGGG  540
ThrValThrAsnLysPheTyrLeuLysMetThrSerAlaAspGlyThrAlaSerValGly

                                        .HincII
GACATCAACATAGGAACCAAATCAGGCAAAGGCCTGAGTGGTCAACTGGTAGGGGGAAAA  600
AspIleAsnIleGlyThrLysSerGlyLysGlyLeuSerGlyGlnLeuValGlyGlyLys

TTCGACGGAAAAATAACGGTGGCATATGACTCGGCCACCGCCCCCGCCGACGTTTATACA  660
PheAspGlyLysIleThrValAlaTyrAspSerAlaThrAlaProAlaAspValTyrThr

TACGATCTCATGGCCGCCGTGTACGTCCAATAATGCTTACCCACATAACTAACACTGACC  720
TyrAspLeuMetAlaAlaValTyrValAsn***

GCCGCCCGCCGGTTTGCTTTCCTCGCGGGAATCGACCTCGTG                    762
```

FIG. 7. Nucleotide sequence of the *nfaA* gene encoding the NFA-I structural subunit protein. The amino acid sequence is given underneath. The cleavage site for the signal peptidase is marked by an arrow. The start codon -ATG- of the precursor (amino acid position, −28) is boxed. The 21 sequenced amino acids of the N terminus of the mature NfaA protein and the N-terminal sequences of the V8-generated peptides are underlined. Restriction enzyme recognition sites for *Sal*I and *Hin*cII are shown in the nucleotide sequence. The ribosome-binding site (Shine-Dalgarno [SD]) is indicated.

```
AGAGGCAGTTCCGGGTGGGGATGGTGAATGAATATATATAAAAGACAACGTGTATCTGTC  60
          SD    -29
TGTTCCGGAGGGAGTATGAAAATGCGGGCTGTGGCTGTGTTCACCGGCATGCTGACGGGA 120
                MetLysMetArgAlaValAlaValPheThrGlyMetLeuThrGly
                                      -1 ↓ +1
GTGTTATCAGTGACAGGTTTGCTGTCAGCGGGGGCATATGCCGCCGGGGGAGAAGGGAAT 180
ValLeuSerValThrGlyLeuLeuSerAlaGlyAlaTyrAlaAlaGlyGlyGluGlyAsn

ATGTCTGCATCCGCGACGGAGACAAACGCCAGAGTATTCTCGCTGCATCTGGGGGCCACC 240
MetSerAlaSerAlaThrGluThrAsnAlaArgValPheSerLeuHisLeuGlyAlaThr

CGGGTGGTGTACAACCCGGCCTCGTCGGGGGAGACGCTGACTGTGATTAATGACCAGGAC 300
ArgValValTyrAsnProAlaSerSerGlyGluThrLeuThrValIleAnsAspGlnAsp

TATCCGATGCTGGTGCAGTCGGAGGTGCTGAGTGAGGACCAGAAGAGTCCGGCGCCTTTT 360
TyrProMetLeuValGlnSerGluValLeuSerGluAspGlnLysSerProAlaProPhe

TGGTGGACACCACCGTTGTTCCGTCGTGATGGTCAGCAGTCGAGTCGTCGTCGTAGTGTC 420
ValValThrProProLeuPheArgLeuAspGlyGlnGlnSerSerArgLeuArgIleVal

AGTACGGGCGGGGAGTTTCCGTCAGACCGTGAGAGTCGGCAGTGGATTTGCGTGAAAGGC 480
ArgThrGlyGlyGluPheProSerAspArgGluSerLeuGlnTrpIleCysValLysGly

ATTCCGCCGAAGGAGGATGACAGGTGGGCGGAAGGGAAGGACGGGGAGAAGAAGGCTGAC 540
IleProProLysGluAspAspArgTrpAlaGluGlyLysAspGlyGluLysLysAlaAsp

AAAGTCTCCCTGAATGTACAGCTTTCAGTGAGCAGCTGCATCAAGCTGTTTGTTCGTCCG 600
LysValSerLeuAsnValGlnLeuSerValSerSerCysIleLysLeuPheValArgPro

CCGGCGGTGAAGGGGCGACCGGATGATGTGGCCGGCAAGGTGGAGTGGCAGAGGGCCGGC 660
ProAlaValLysGlyArgProAspAspValAlaGlyLysValGluTrpGlnArgAlaGly

AACAGGCTGAAGGGGGTTAACCCGACGCCGTTTTACATCAACCTGTCCACGCTGACGGTG 720
AsnArgLeuLysGlyValAsnProThrProPheTyrIleAsnLeuSerThrLeuThrVal

GGGGGTAAGGAAGTGAAGGAGCGTGAATATATTGCGCCGTTTTCCTCCCGTGAATATCCG 780
GlyGlyLysGluValLysGluArgGluTyrIleAlaProPheSerSerArgGluTyrPro

CTGCCTGCGGGGCATCGGGTAAGGTTCAGTGGAAGGTGATAACGGATTACGGCGGGACCA 840
LeuProAlaGlyHisArgValArgPheSerGlyArg***
```

FIG. 8. Nucleotide sequence of the *nfaE* gene. The deduced amino acid sequence is given underneath. The presumptive cleavage site for the signal peptidase is marked by an arrow. The ribosome-binding site (Shine-Dalgarno [SD]) is indicated.

9 and 10.5 (Fig. 1) revealed an ORF of 247 amino acids with a putative signal sequence of 29 amino acids (Fig. 8). The ORF starts at positions 76 to 78 with -ATG- and ends at positions 817 to 819 with -TGA-. The calculated molecular weight of the precursor protein is 27,000 and that of the mature protein is 24,000. Cleavage by the signal peptidase (7) might occur between Ala and Ala at positions 162 to 163 (arrow). Upstream of the ORF a ribosome-binding site (Shine-Dalgarno) sequence is found. The 30.5-kDa protein found in minicells might be referred to this ORF (see above). Computer analysis revealed a similarity of 60% between this ORF and the 27-kDa protein involved in the assembly of CS3 fibrillae (Fig. 9) (14).

DISCUSSION

NFA-I which mediates agglutination of human erythrocytes is produced by uropathogenic *E. coli* strains (9). It is suggested that NFA-I expression contributes to the virulence of strains because it may facilitate their colonization on host cells during urinary tract infections. Recent data (10) furthermore indicated binding of *E. coli* to polymorphonuclear leukocytes via NFA-I, a process which also may be relevant to pathogenicity.

We have analyzed the genetic structure of the *nfaI* determinant and have identified and sequenced the subunit-specific gene *nfaA*. The 19-kDa NFA-I subunit NfaA seems to be a unique protein because its deduced amino acid sequence is unrelated to sequences of other nonfimbrial adhesins such as the A-fimbrial adhesin I adhesin (AfaE) and the blood group M agglutinin (BmaE) or to the sequences of the Dr fimbrial agglutinin (DraA) and the F1845 fimbrial adhesin (DaaE [17, 24, 30]).

It has been shown recently (15) that the mature NFA-I subunit protein starts with amino acids Asp-Ala-Asn. The

```
NfaE: LTGVLSVTGLLSAGAYAAGGEGNMSASATETNARVFSLHLGATRVVYNPA
      *| *  * * ||  **     *|* |    |    |   ||||||*|*
CS3 : MTPIKLIFAALSLFPC.....SNIYANNITTQK..FEAILGATRVIYHLD

 63   SSGETLTVINDQDYPMLVQSEVLSEDQKSPAPFVVTPPLFRLDGQQSSRL
      * || | | | |  |*|*|| |* |* |  | |*|||||||||*     *
 44   GNGESLRVKNPQISPILIQSKVMDEGSKDNADFIVTPPLFRLDAKRETDI

113   RIVRTGGEFPSDRESLQWICVKGIPPKEDDRWAEGKDGEKKADKVSLNVQ
|||    | *| |||||  *||*|||||**| ||*         * |** *
 94   RIVMVNGLYPKDRESLKTLCVRGIPPKQGDLWANNEK.....EFVGMKLN

163   LSVSSCIKLFVRPPAVKGRPDDVAGKVEWQRAGNRLKGVNPTPFYINLST
*|*  ||||**||  *     *  | *||   *  | * | ||*|* *
139   VSINTCIKLILRPHNLPKLDINSEGQIEWGIRDGNLVAKNKTPYYFTIVN

213   LTVGGKEVKEREYIAPFSSREYPLPAGHRVRFSG
          || *|  * **|*  * | ||   **  ||
189   ASFNGKALKTPGTLGPYEQKLYTLPS..KISVSG
```

FIG. 9. Comparison of the NfaE ORF and the 27-kDa CS3 protein involved in the assembly of CS3 fibrillae (14). Lines indicate identical amino acids; asterisks indicate functionally similar amino acids. The overall similarity was calculated to be 60%.

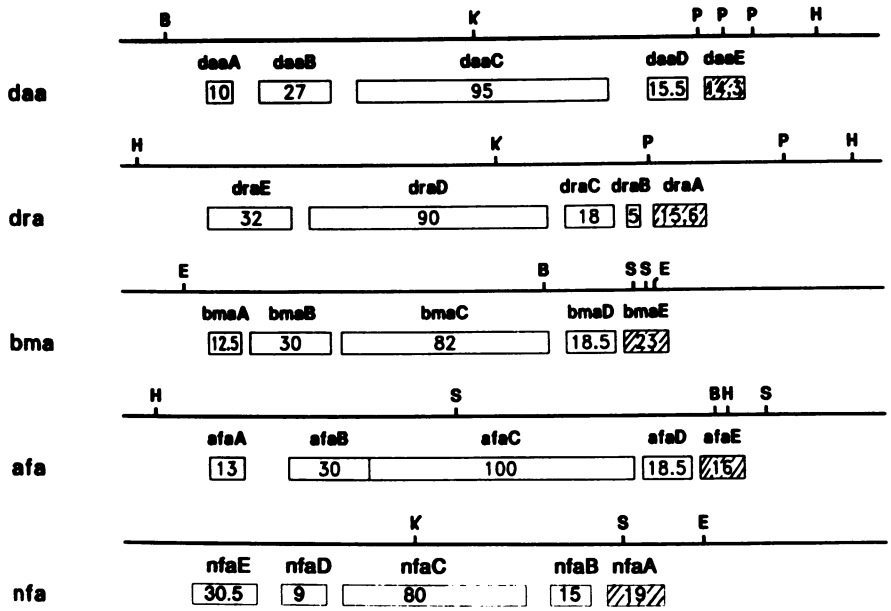

FIG. 10. Comparison of physical maps and the gene products of the F1845 gene cluster (*daa*), Dr agglutinin gene cluster (*dra*), blood group M gene cluster (*bma*), A fimbrial adhesin I gene cluster (*afa*), and NFA-I gene cluster (*nfaI*). The boxes represent the gene products, the numbers (10^3) indicate the molecular weights of the proteins, and the hatched boxes represent adhesin proteins. The main direction of transcription is from left to right. Data are from references 17, 24, and 30 and from this study. Restriction enzymes: B, *Bam*HI; K, *Kpn*I; P, *Pst*I; H, *Hin*dIII; E, *Eco*RI; S, *Sal*I.

NfaA amino acid sequence based on the nucleic acid sequence of the corresponding gene is in full agreement with the analyzed N terminus of the protein. From these data it can be concluded that a relatively hydrophilic 28-amino-acid signal peptide is cleaved from the NfaA precursor protein, leading to the mature protein consisting of 156 amino acids. Cleavage occurs between Ala and Asp residues that is not typical for the *E. coli* signal peptidase (Fig. 7). The molecular weight of the mature protein calculated by the amino acid sequence is 16,000, differing from the size of 19 kDa estimated by SDS-PAGE.

The NfaA protein contains two cysteine residues, which may form a disulfide bridge. The positions of the two cysteine residues in NfaA (Cys-29 and Cys-63) are nearly the same as the position of cysteine residues, which are found in the fimbrillins FsoA and FstA of P fimbriae (33, 34) as well as in the structural subunit protein AfaE or DraA (17, 30). The NfaA protein seems to be highly hydrophilic as predicted by Kyte-Doolittle and Goldman et al. (3, 13). The amino acid compositions of NfaA, AfaE, BmaE, and DraA and of the fimbrillins FsoA and FstA surprisingly have a high content of glycine (around 12 mol%), pointing to a relatively flexible protein, and the very high content of serine and threonine (16 to 23 mol%) indicates that hydrogen bonds are probably important in building up high-molecular-weight polymeric structures. In this respect, it is noteworthy that upon heat disintegration a certain portion of NFA-I remains as relatively stable dimers. From the primary amino acid sequence of NfaA monomers the prediction of the secondary structures revealed 17% α-helices and a high content (35 to 45%) of β-sheets. Analyses of AmidI and AmidI′ infrared absorption bands of NFA-I confirm the preponderance (~50%) of β-sheets (1).

The ORF, termed *nfaE*, detected at the proximal end of the *nfaI* gene cluster exhibits 60% similarity to the 27-kDa CS3 protein, which might be involved in transport and assembly of CS3 fibrillae (14). A similar function might be attributed to the NfaE protein that could correspond to the 30.5-kDa protein found in minicell analysis.

Besides the *nfaI* determinant, gene clusters coding for other nonfimbrial adhesins have been cloned and analyzed recently (16, 17, 24). Also, the Dr fimbrial hemagglutinin, originally described as a nonfimbrial adhesin (11), and the closely related F1845 fimbrial adhesin have been cloned and genetically characterized (30). Despite the fact that the gene coding for the structural protein of NFA-I does not exhibit any homology to loci responsible for subunits of the other adhesins, the gene structures of the corresponding determinants share common features (Fig. 10). In all cases, five genes are responsible for full expression of the adhesion phenotype. These genes are located on a stretch of DNA of approximately 6 to 8 kb. In addition, the genes encoding the respective adhesin structural proteins are located at the distal (3′) end of the gene clusters.

Interestingly, large proteins like the NfaC gene product are found in all nonfimbrial adhesins as well as in fimbrial adhesin complexes. It is shown that such large anchor proteins are involved in the biogenesis of fimbriae and fibrillae (22, 25, 28). The occurrence of such proteins in nonfimbrial adhesins which exhibit homologies to the anchor proteins of fimbriae (see above) may argue for common functions of these gene products in both types of attachment factors.

ACKNOWLEDGMENTS

We thank Laurence R. Phillips (Würzburg) for critical reading of the manuscript.

This work was supported by the Deutsche Forschungsgemeinschaft (DFG grant Ha 1434/1-7) and the Fonds der Chemischen Industrie.

REFERENCES

1. **Ahrens, R., and P. Hübner.** Unpublished data.
1a. **Becauge, S. L., and M. H. Caruthers.** 1981. Deoxynucleoside phosphoramidites: a new class of key intermediate for deoxypolynucleotide synthesis. Tetrahedron Lett. **22:**1859–1862.
2. **Birnboim, H. C., and J. Doly.** 1979. A rapid extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7:**1513–1523.
3. **Devereux, J.** 1984. UWGCG computer programs. University of Wisconsin, Madison.
4. **Dougan, G., and D. Sherratt.** 1977. The transposon Tn1 as a probe for studying ColE1 structure and function. Mol. Gen. Genet. **151:**151–160.
5. **Duguid, J. P., S. Clegg, and M. I. Wilson.** 1979. The fimbrial and non-fimbrial haemagglutinins of *Escherichia coli*. J. Med. Microbiol. **12:**213–228.
6. **Eckerskorn, C., W. Mewes, H. Goretzki, and F. Lottspeich.** 1988. A new siliconized glass fibre as support to protein chemical analysis of electro-blotted proteins. Eur. J. Biochem. **175:** 509–519.
7. **Ferenci, T., and T. J. Silhavy.** 1987. Sequence information required for protein translocation from the cytoplasm. J. Bacteriol. **169:**5339–5342.
8. **Findlay, J. B. C., and M. J. Geisow (ed.).** 1989. Protein sequencing: a practical approach. IRL Press, Oxford.
9. **Goldhar, J., R. Perry, J. R. Golecki, H. Hoschützky, B. Jann, and K. Jann.** 1987. Nonfimbrial, mannose-resistant adhesins from uropathogenic *Escherichia coli* O83:K1:H4 and O14:K?: H11. Infect. Immun. **55:**1837–1842.
10. **Goldhar, J., M. Yarzori, Y. Keisari, and I. Ofek.** 1991. Phagocytosis of *Escherichia coli* mediated by mannose-resistant nonfimbrial haemagglutinin (NFA-1). Microb. Pathog. **11:**171–178.
11. **Hacker, J.** 1990. Genetic determinants coding for fimbriae and adhesins of extraintestinal *Escherichia coli*. Curr. Top. Microbiol. Immunol. **151:**1–27.
12. **Hales, B. A., H. Beverly-Clarke, N. J. High, K. Jann, R. Perry, J. Goldhar, and G. J. Boulnois.** 1988. Molecular cloning and characterization of the genes for a non-fimbrial adhesin from *Escherichia coli*. Microb. Pathog. **5:**9–17.
13. **IntelliGenetics Inc.** 1986. Genofit soft ware package. IntelliGenetics, Inc., Geneva.
14. **Jalajakumari, M. B., C. J. Thomas, R. Halter, and P. A. Manning.** 1989. Genes for biosynthesis and assembly of CS3 pili of CFA/II enterotoxigenic *Escherichia coli*: novel regulation of pilus production by bypassing an amber codon. Mol. Microbiol. **3:**1685–1695.
15. **Jann, K., and H. Hoschützky.** 1990. Nature and organization of adhesins. Curr. Top. Microbiol. Immunol. **151:**55–70.
16. **Labigne-Roussell, A., and S. Falkow.** 1988. Distribution and degree of heterogeneity of the afimbrial-adhesin-encoding operon (*afa*) among uropathogenic *Escherichia coli* isolates. Infect. Immun. **56:**640–648.
17. **Labigne-Roussel, A., M. A. Schmidt, W. Walz, and S. Falkow.** 1985. Genetic organization of the afimbrial adhesin operon and nucleotide sequence from a uropathogenic *Escherichia coli* gene, encoding an afimbrial adhesin. J. Bacteriol. **162:**1285–1292.
18. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.
19. **Lederberg, E., and S. N. Cohen.** 1974. Transformation of *Salmonella typhimurium* by plasmid deoxyribonucleic acid. J. Bacteriol. **119:**1072–1074.
20. **Moch, T., H. Hoschützky, J. Hacker, K. D. Kröncke, and K. Jann.** 1987. Isolation and characterization of the α-sialyl-β-2-3 galactosyl specific adhesin from fimbriated *Escherichia coli*. Proc. Natl. Acad. Sci. USA **84:**3462–3466.
21. **Noegel, A., U. Rdest, W. Springer, and W. Goebel.** 1979. Plasmid cistrons controlling synthesis and excretion of the exotoxic hemolysin of *Escherichia coli*. Mol. Gen. Genet. **175:**343–350.
22. **Norgren, M., M. Baga, J. M. Tennent, and S. Normark.** 1987. Nucleotide sequence, regulation and functional analysis of the *papC* gene required for cell surface localization of Pap pili of uropathogenic *Escherichia coli*. Mol. Microbiol. **1:**169–178.
23. **Orskov, I., and F. Orskov.** 1988. *Escherichia coli* in extraintestinal infections. J. Hyg. **95:**551–575.
24. **Rhen, M., V. Väisänen-Rhen, M. Sarasta, and T. K. Korhonen.** 1986. Organization of genes expressing the blood group M-specific hemagglutination of *Escherichia coli*: identification and nucleotide sequence of the M-agglutinin subunit gene. Gene **49:**351–360.
25. **Roosendahl, B., and F. K. de Graaf.** 1989. The nucleotide sequence of the *fan*D gene encoding the large outer membrane protein involved in the biosynthesis of K99 fimbriae. Nucleic Acids Res. **17:**1263.
26. **Sambrook, S., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
27. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5464.
28. **Schmoll, T., J. Morschhäuser, M. Ott, B. Ludwig, I. van Die, and J. Hacker.** 1990. Complete genetic organization and functional aspects of the *Escherichia coli* S fimbrial adhesin determinant: nucleotide sequence of the genes *sfa* B, C, D, E, F. Microb. Pathog. **9:**331–343.
29. **Svanborg-Eden, C., L. Hagberg, L. A. Hanson, S. Hull, R. Hull, U. Jodal, H. Leffler, H. Lomberg, and E. Straube.** 1983. Bacterial adherence—a pathogenic mechanism in urinary tract infections caused by *Escherichia coli*. Prog. Allergy **33:**175–188.
30. **Swanson, T. N., S. S. Bilge, B. Nowicki, and S. L. Moseley.** 1991. Molecular structure of the Dr adhesin: nucleotide sequence and mapping of receptor-binding domain by use of fusion constructs. Infect. Immun. **59:**261–268.
31. **Thuring, R. W. F., J. P. Sander, and B. Borst.** 1975. A freeze squeeze method for recovering long DNA from agarose gels. Anal. Biochem. **66:**2113–2120.
32. **Towbin, H., T. Staehelin, and J. Gordon.** 1979. Electrophoretic transfer from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA **76:**4350–4354.
33. **Van Die, I., and H. Bergmans.** 1984. Nucleotide sequence of the gene encoding the $F7_2$ fimbrial subunit of a uropathogenic *Escherichia coli* strain. Gene **32:**83–90.
34. **Van Die, I., G. Spierings, I. van Megen, E. Zuidweg, W. Hoekstra, and H. Bergmans.** 1985. Cloning and genetic organization of the gene cluster encoding $F7_1$ fimbriae of a uropathogenic *Escherichia coli* and comparison with the $F7_2$ gene cluster. FEMS Microbiol. Lett. **28:**329–334.
35. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13 mp7 derived system for insertion and mutagenesis and sequencing with synthetic universal primers. Gene **19:**259–268.