JOURNAL OF BACTERIOLOGY, Apr. 1985, p. 454–457
0021-9193/85/040454-04$02.00/0

Vol. 162, No. 1

Nucleotide Sequence of *pilA*, the Gene Encoding the Structural Component of Type 1 Pili in *Escherichia coli*

PAUL E. ORNDORFF†* AND STANLEY FALKOW
Department of Medical Microbiology, Stanford University, Stanford, California 94304

Received 17 December 1984/Accepted 2 January 1985

The *pilA* gene of *Escherichia coli* J96 encoding pilin, the structural component of type 1 pili, was sequenced and found to specify a polypeptide 159 amino acids long preceded by a 23-amino-acid signal peptide. As determined from the DNA sequence, the mature peptide lacked tryptophane and methionine, two amino acids previously shown to be lacking in type 1 pili from *E. coli*. Also, the amino-terminal sequence of amino acids inferred from the DNA sequence corresponded to earlier 20-amino-acid amino-terminal sequences determined by protein sequencing. In addition, piliation was abolished after a mutation was introduced into the *pilA* coding region in vitro. A possible site for initiation of transcription and a possible site encoding translation initiation were suggested 85 and 7 base pairs, respectively, from the *pilA* start codon. There appeared to be scant DNA sequence homology and scant amino acid sequence homology between type 1 pilin and other pilin species isolated from uropathogenic and enteropathogenic *E. coli*.

Type 1 pili are a class of filamentous, proteinaceous appendages of *Escherichia coli* and other gram-negative enteric bacteria (3) that mediate a mannose-sensitive attachment of the bacteria to a variety of eucaryotic cells (20). The novel genetic control (4, 7, 22), assembly (21), and receptor binding activity of type 1 pili provide a unique model system for studying the regulation of gene expression, membrane protein translocation, protein-protein interactions, and receptor-ligand interactions. In addition, examination of the adhesive properties of type 1 pili may contribute to understanding the role that adhesion plays in the pathogensis of extraintestinal infectious disease caused by *E. coli*.

We have previously located and characterized several genes involved in type 1 pilus expression (21, 22) (see Fig. 1). We found that at least two genes, *pilB* and *pilC*, encoded polypeptides that are involved in the assembly of a third gene product, pilin (the *pilA* gene product), into pili. A fourth gene, *hyp*, encodes a polypeptide involved in the regulation of piliation (22). A fifth gene, *pilD*, encodes a polypeptide that is related to pilin immunologically but is of uncertain function (21). In this report we present the nucleotide sequence of *pilA* and the primary sequence of amino acids composing the mature protein and the signal peptide.

The *pilA* gene was isolated by subcloning overlapping restriction endonuclease-generated DNA fragments from pSH2 containing Tn*5* insertions H16 and H21 on either side of the *pilA* gene (21), producing clones pORN119 and pORN120 (Fig. 1). The nucleotide sequence was determined by the procedure of Maxam and Gilbert (18). The regions sequenced are shown in Fig. 1. Figure 2 shows the nucleotide sequence as well as the predicted amino acid sequence for an open reading frame of 546 base pairs (bp).

The predicted start codon of *pilA* (bp 281; Fig. 2) was preceded by a five-nucleotide sequence (bp 269 to 273; Fig. 2) that has been associated with coding for ribosome binding (10). In addition, we suspect that transcription initiation may occur in the region from bp 41 to 267 (Fig. 2), since deletions of this region result in a drastic lowering of piliation (21, 22; and our unpublished observations). We have examined the region from bp 41 to 267 for possible transcription initiation sites and have suggested a site at bp 188 (−10 region) and bp 164 (−35 region) (Fig. 2), based upon the sequence similarity to other procaryotic initiation sites (25). However, we have no direct experimental evidence that the suggested promoter is used.

A signal peptide of 22 to 25 amino acids was predicted for type 1 pili from our earlier biochemical studies (21), and indeed, a 23-amino-acid signal sequence was indicated from the DNA sequence (bp 281 to 350; Fig. 2). The signal sequence of type 1 pilin is similar in amino acid composition, hydrophobicity, and the position of charged residues to signal sequences of several other procaryotic envelope proteins (12). The start of the mature protein (bp 350, amino acid position +1; Fig. 2) was inferred from published 20-amino-acid, amino-terminal sequences of *E. coli* type 1 pilin (9, 11) and from the unpublished 20-amino-acid, amino-terminal sequence of pilin produced from pSH2 (G. Schoolnik, Department of Medical Microbiology, Stanford University, Stanford, Calif., personal communication). The 20-amino-acid sequences from the literature sources and from our cloned gene were the same. The amino acids specified by the DNA sequence were identical to those found for the first 20 amino acids determined by protein sequencing. The mature pilin protein, as predicted from the DNA sequence, lacks methionine and tryptophane, correlating exactly with what has been found in amino acid composition determinations (3; G. Schoolnik, personal communication). Also, as predicted from the DNA sequence, the mature protein has two cysteine residues, at amino acid positions 21 and 61 (Fig. 2), which may form a disulfide bridge similar to that inferred for gonococcal pili (24), K99 pili (23), and Pap pili (1).

To confirm that the sequenced region did indeed encode the structural subunit of type 1 pili, we took advantage of the nucleotide sequence to introduce a mutation into the *pilA* gene. We note that in our prior description of the genes for type 1 piliation, the structural gene was inferred by subcloning of the region containing *pilA* and identification of the gene product (21). No mutations in *pilA* were obtained at that time. A genetic lesion was produced in *pilA* by the

* Corresponding author.
† Present address: Department of Microbiology, Pathology, and Parasitology, School of Veterinary Medicine, North Carolina State University, Raleigh, NC 27606.

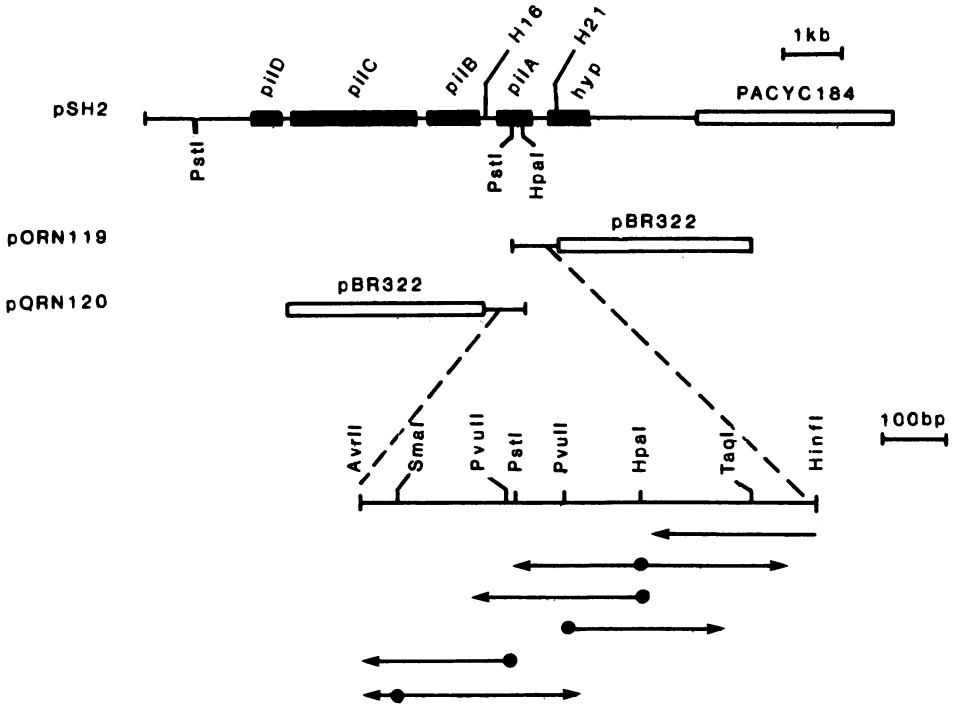

FIG. 1. The pSH2 plasmid and the sequenced region. Restriction endonuclease sites in IS50 (13) of Tn*5* insertion mutants H16 and H21 were used to subclone the region containing *pilA*. The pORN120 plasmid contained the carboxyterminal region of *pilA* from the *Hpa*I site to the *Hin*dIII site in IS50 of Tn*5* insertion H16. The pORN119 plasmid contained the aminoterminus of *pilA* and encompassed the region from *Pst*I in *pilA* to the *Xho*I site in IS50 of insertion H21. Both plasmids were constructed by using pBR322 (2) as the plasmid replicon. The two regions were ligated to the *Eco*RI site in pBR322 after treatment of the restriction fragments with DNA polymerase I and addition of synthetic *Eco*RI "linkers" (17). The lowest bar in the figure represents a portion of the region sequenced covering the *pilA* structural gene and the presumptive promoter region (see Fig. 2). The arrows designate the sequencing strategy.

```
       10                                                  50
ACGTTTCTGT GGCTCGACGC ATCTTCCTCA TTCTTCTCTC CAAAAACCAC CTCATGCAAT ATAAACATCT

                              100
ATAAATAAAG ATAACAAATA GAATATTAAG CCAACAAATA AACTGAAAAA GTTTGTCCGC GATGCTTTAC

      150 HinfI                                                   200
CTCTATGAGT CAAAATGGCC CCAATGTTTC ATCTTTTGGG GGAAACTGTG CAGTGTTGGC AGTCAAACTC
                              -35                         -10
                                      TaqI250
GTTGACAAAC AAAGTGTACA GAACGACTGC CCATGTCGAT TTAGAAATAG TTTTTTGAAA GGAAAGCAGC
                                                          S.D.
                     300
ATG AAA ATT AAA ACT CTG GCA ATC GTT GTT CTG TCG GCT CTG TCC CTC AGT TCT ACG
MET Lys Ile Lys Thr Leu Ala Ile Val Val Leu Ser Ala Leu Ser Leu Ser Ser Thr
-23         -20                                     -10
               350
ACG GCT CTG GCC GCT GCC ACG ACG GTT AAT GGT GGG ACC GTT CAC TTT AAA GGG GAA
Thr Ala Leu Ala Ala Ala Thr Thr Val Asn Gly Gly Thr Val His Phe Lys Gly Glu
                -1  +1                                  10
    400 HpaI          MstI                                BclI              450
GTT GTT AAC GCC GCT TGC GCA GTT GAT GCA GGC TCT GTT GAT CAA ACC GTT CAG TTA
Val Val Asn Ala Ala Cys Ala Val Asp Ala Gly Ser Val Asp Gln Thr Val Gln Leu
                    20                                      30
                                                                    500
GGA CAG GTT CGT ACC GCA TCG CTG GCA CAG GAA GGA GCA ACC AGT TCT GCT GTC GGT
Gly Gln Val Arg Thr Ala Ser Leu Ala Gln Glu Gly Ala Thr Ser Ser Ala Val Gly
                    40                                      50
              PvuII                                 550
TTT AAC ATT CAG CTG AAT GAT TGC GAT ACC AAT GTT GCA TCT AAA GCC GCT GTT GCC
Phe Asn Ile Gln Leu Asn Asp Cys Asp Thr Asn Val Ala Ser Lys Ala Ala Val Ala
                        60                                      70
                                        600                 PstI
TTT TTA GGT ACG GCG ATT GAT GCG GGT CAT ACC AAC GTT CTG GCT CTG CAG AGT TCA
Phe Leu Gly Thr Ala Ile Asp Ala Gly His Thr Asn Val Leu Ala Leu Gln Ser Ser
                        80                                      90
PvuII                                 650
GCT GCG GGT AGC GCA ACA AAC GTT GGT GTG CAG ATC CTG GAC AGA ACG GGT GCT GCG
Ala Ala Gly Ser Ala Thr Asn Val Gly Val Gln Ile Leu Asp Arg Thr Gly Ala Ala
                                100                                     110
                     700
CTG ACG CTG GAT GGT GCG ACA TTT AGT TCA GAA ACA ACC CTG AAT AAC GGA ACC AAT
Leu Thr Leu Asp Gly Ala Thr Phe Ser Ser Glu Thr Thr Leu Asn Asn Gly Thr Asn
                                120
               750                                      SmaI
ACC ATT CCG TTC CAG GCG CGT TAT TTT GCA ACC GGG GCC GCA ACC CCG GGT GCT GCT
Thr Ile Pro Phe Gln Ala Arg Tyr Phe Ala Thr Gly Ala Ala Thr Pro Gly Ala Ala
130                             140
      800                                           AvrII         850
AAT GCG GAT GCG ACC TTC AAG GTT CAG TAT CAA TAA CCT ACC TAG GTTCAGGGACGTTCA
Asn Ala Asp Ala Thr Phe Lys Val Gln Tyr Gln  .  Pro Thr
    150                                 160
```

introduction of a synthetic *Xho*I recognition site into the site normally recognized by the *Pst*I endonuclease at bp 611 (Fig. 2). To accomplish this, pSH2 was partially digested with *Pst*I and linear, full-length DNA isolated (22). The staggard ends of the *Pst*I site were made flush, and synthetic *Xho*I linkers (Collaborative Research, Inc.) were introduced and ligated to the flush ends as described by Maniatis et al. (17). This manipulation is predicted to result in the substitution of two amino acids and the formation of a frame shift mutation at the *Xho*I site. The *Xho*I-linked linear plasmid was ligated and introduced, by transformation (21), into strain ORN103 (ORN103 is a *recA lacU*169 derivative of P678-54) (21; and our unpublished data). Since the pSH2 plasmid contains two *Pst*I sites, one in *pilA* and the other outside the Pil region (Fig. 1), transformants could contain plasmid DNA possessing an *Xho*I site at either one of the *Pst*I sites. All transformants that were scored as Pil$^-$ (21) contained the *Xho*I site in *pilA* (10 Pil$^-$ isolates were examined). One plasmid containing the unique *Xho*I site in *pilA* was chosen for further study and was called pORN117. As evidence of the Pil$^-$ phenotype, an electron micrograph (22) of strain ORN103 containing pORN117 is shown together with ORN103 containing the parental (pSH2) plasmid for comparison (Fig. 3).

FIG. 2. The nucleotide sequence of *pilA*. The nucleotide positions are designated on the top of the sequence. The amino acid positions are given below. Only the coding region of *pilA* is translated. The amino acids of the signal sequence are given negative numbers, and the mature protein starts at +1 and ends with the first stop codon. A possible −10 and −35 region is denoted with double lines, as is a possible region encoding ribosome binding site (S.D.). Selected restriction endonuclease sites are underlined and designated by their names above the coding sequence.

Computer-assisted comparison of the nucleotide sequence and the amino acid sequence of type 1 pilin with K88 pilin (14), K99 pilin (23), and Pap pilin (1) revealed no striking regions of similarity in either the nucleotide sequence or the primary amino acid sequence. Also, comparison of the amino acid sequence of CFAI pilin (15) with the inferred amino acid sequence of type 1 pilin revealed no striking regions of commonality between the two pilin molecules (data not shown).

The hydrophobic character of type 1 pili was reflected in the computer-generated hydropathicity profile of pilin (Fig. 4). The hydrophobic nature of type 1 pili has been recognized for some time (3), and it is thought that this hydrophobicity contributes to the resistance of type 1 pili to dissociation into its constituent pilin monomers under conditions that dissociate other pili (8, 19). Also, the hydrophobicity of type 1 pili probably contributes to the tendency of type 1 piliated cells to accumulate at the air-water interface during growth in static broth, forming a pellicle at the surface of the growth medium (3, 6). One might suspect that the unique properties of type 1 pili would be reflected when the hydropathicity and secondary structure (5) profiles of type 1 pili were

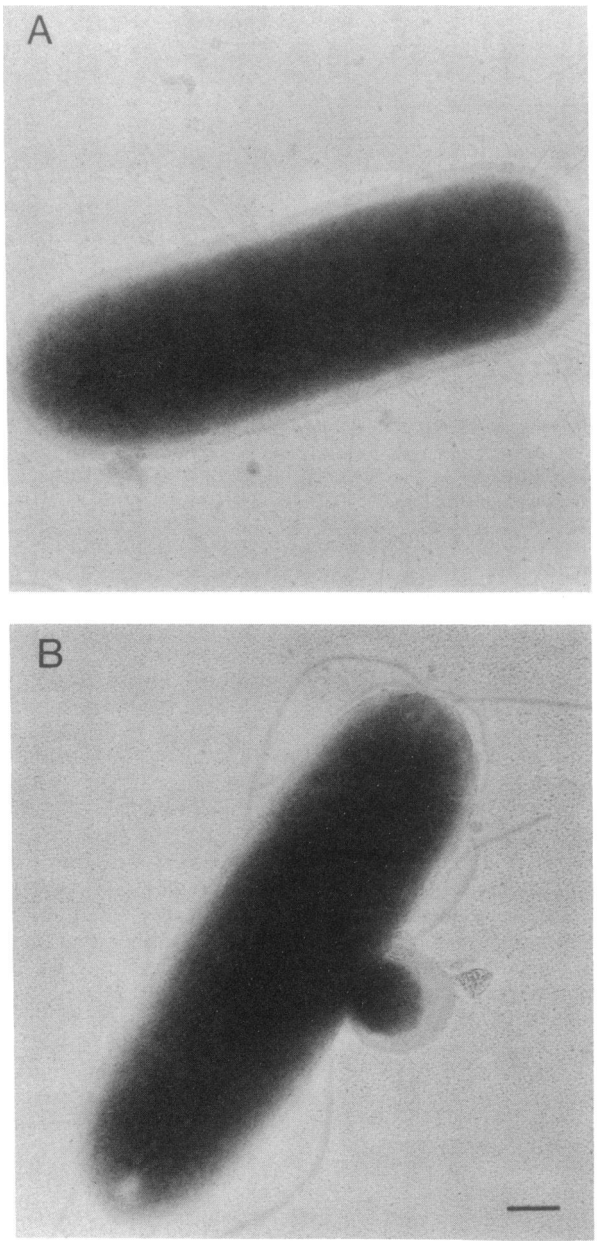

FIG. 3. Electron micrographs of a *pilA* mutant and a strain containing the parental plasmid. Electron micrographs were prepared from strain ORN103 containing pSH2 (A) and strain ORN103 containing pORN117 (B) as described previously (22).

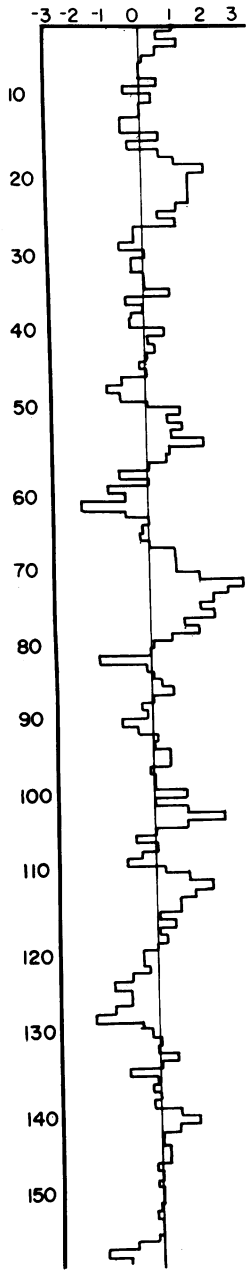

FIG. 4. The hydropathicity profile of the mature *pilA* protein. Hydropathicity was measured by the method of Kyte and Doolittle (16), and a computer-assisted tracing was made. Areas above the midpoint are more hydrophobic than those below the midpoint. Values of hydropathicity were calculated across overlapping six amino acid residues. The values on the *y* axis were calculated by Kyte and Doolittle (16) and represent a relative measure of hydropathicity.

compared with the profiles of other pilin species. This may indeed be true; however, the profiles suggested to us no noteworthy dissimilarities in type 1 pilin when compared with the other pilin species. We hope that our current studies involving site-directed mutagenesis will better define the areas of pilin involved in pilin-to-pilin interaction and receptor binding.

We thank G. Schoolnik for providing the amino acid composition and the amino terminal sequence of type 1 pili.

This work was supported by grant PCM-8306654 from the National Science Foundation.

ADDENDUM IN PROOF

While this manuscript was in preparation, the nucleotide sequence of a different type 1 pilin gene was published (P. Klemm, Eur. J. Biochem. **143**:395–399, 1984). It differed from the sequence herein described in that it lacked one codon (ACC) specifying tyrosine at amino acid position 140 (Fig. 2) and had alanine rather than threonine at amino acid position −4 (Fig. 2) due to a single nucleotide difference.

LITERATURE CITED

1. **Baga, M., S. Normark, J. Hardy, P. O'Hanley, D. Lark, O. Olsson, G. Schoolnik, and S. Falkow.** 1984. Nucleotide sequence of the *papA* gene encoding the Pap pilus subunit of human uropathogenic *Escherichia coli*. J. Bacteriol. **157**:330–333.
2. **Bolivar, F.** 1978. Construction and characterization of new cloning vehicles. III. Derivatives of plasmid pBR322 carrying unique *Eco*RI sites for selecting *Eco*RI-generated recombinant molecules. Gene **4**:121–136.
3. **Brinton, C. C., Jr.** 1965. The structure, function, synthesis, and genetic control of bacterial pili and a molecular model for DNA and RNA transport in gram negative bacteria. Trans. N.Y. Acad. Sci. **27**:1003–1054.
4. **Brinton, C. C., A. Buzzell, and M. A. Lauffer.** 1954. Electrophoresis and phage susceptibility studies on a filament-producing variant of the *E. coli* bacterium. Biochem. Biophys. Acta **15**: 533–542.
5. **Chou, P. Y., and G. D. Fasman.** 1978. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. **47**:45–148.
6. **Duguid, J. P., and D. C. Old.** 1980. Adhesive properties of enterobacteriaceae, p. 185–217. *In* E. H. Beachey (ed.), Bacterial adherence. Receptors and recognition, series B., vol. 6. Chapman and Hall, London.
7. **Eisenstein, B. I.** 1981. Phase variation of type 1 fimbriae in *Escherichia coli* is under transcriptional control. Science **214**: 337–339.
8. **Eshdat, Y., F. J. Silverblatt, and N. Sharon.** 1981. Dissociation and reassembly of *Escherichia coli* type 1 pili. J. Bacteriol. **148**: 308–314.
9. **Gaastra, W., and F. K. de Graaf.** 1982. Host-specific fimbrial adhesins of noninvasive enterotoxogenic *Escherichia coli* strains. Microbiol. Rev. **46**: 129–161.
10. **Gold, L., D. Pribnow, T. Schneider, S. Shenedling, B. Swebilius Singer, and G. Stormo.** 1981. Translation initiation in procaryotes. Annu. Rev. Microbiol. **35**:365–403.
11. **Hermodson, M. A., K. C. S. Chen, and T. M. Buchanan.** 1978. Neisseria pili proteins: amino-terminal amino acid sequences and identification of an unusual amino acid. Biochemistry **17**: 442–445.
12. **Inouye, M., and S. Halegoua.** 1980. Secretion and membrane localization of proteins in *Escherichia coli*. Crit. Rev. Biochem. **7**:339–371.
13. **Jorgensen, R. A., S. J. Rothstein, and W. S. Reznikoff.** 1979. A restriction enzyme cleavage map of Tn5 and localization of a region encoding neomycin resistance. Mol. Gen. Genet. **177**: 6047–6051.
14. **Klemm, P.** 1981. The complete amino-acid sequence of the K88 antigen, a fimbrial protein from *Escherichia coli*. Eur. J. Biochem. **124**:339–348.
15. **Klemm, P.** 1982. Primary structure of the CFA1 fimbrial protein from human enterotoxigenic *Escherichia coli* strains. Eur. J. Biochem. **124**:339–348.
16. **Kyte, J., and R. F. Doolittle.** 1982. A method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
17. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning, p. 219. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
18. **Maxam, A., and W. Gilbert.** 1980. Sequencing the DNA by labeling the end and breaking at bases. Methods Enzymol. **65**: 499–560.
19. **McMichael, J. C., and J. T. Ou.** 1979. Structure of common pili from *Escherichia coli*. J. Bacteriol. **138**:969–975.
20. **Old, D. C.** 1972. Inhibition of the interaction between fimbrial hemagglutinins and erythrocytes by D-mannose and other carbohydrates. J. Gen. Microbiol. **71**:149–157.
21. **Orndorff, P. E., and S. Falkow.** 1984. Organization and expression of genes specifying type 1 piliation in *Escherichia coli*. J. Bacteriol. **159**:736–744.
22. **Orndorff, P. E., and S. Falkow.** 1984. Identification and characterization of a gene product that regulates type 1 piliation. J. Bacteriol. **160**:61–66.
23. **Roosendaal, B., W. Graastra, and F. K. deGraaf.** 1984. The nucelotide sequence of the gene encoding the K99 subunit of enterotoxigenic *Escherichia coli*. FEMS Lett. **22**:253–258.
24. **Schoolnik, G. K., J. Y. Tai, and E. C. Gotschlich.** 1983. A pilus peptide vaccine for the prevention of gonorrhea. Prog. Allergy **33**:314–331.
25. **Siebenlist, U., R. B. Simpson, and W. Gilbert.** 1980. *E. coli* RNA polymerase interacts homologously with two different promoters. Cell **20**:269–281.