JOURNAL OF BACTERIOLOGY, Nov. 1994, p. 6636–6643
0021-9193/94/$04.00+0

Vol. 176, No. 21

Molecular Cloning and Characterization of Outer Membrane Protein E of *Moraxella (Branhamella) catarrhalis*

REVA BHUSHAN,[1,2] ROBERT CRAIGIE,[1] AND TIMOTHY F. MURPHY[2,3,4]*

Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, Maryland 20892,[1] and Department of Microbiology[2] and Division of Infectious Diseases of the Department of Medicine,[3] State University of New York at Buffalo, and Department of Veterans Affairs Medical Center,[4] Buffalo, New York 14215

Received 2 May 1994/Accepted 29 August 1994

Outer membrane protein E (OMP E) is a 50-kDa protein of *Moraxella (Branhamella) catarrhalis*. It is a potential vaccine antigen because it is expressed on the surface of the bacterium and has antigenic determinants which are conserved among most strains of *M. catarrhalis*. To clone the gene encoding OMP E, an EMBL-3 genomic library of strain 25240 was screened with a family of degenerate oligonucleotides based on the amino-terminal protein sequence. The OMP E gene was identified in one of the six positive clones by Southern blot analysis. An open reading frame of 1,377 bp encoding a protein of 460 amino acids was identified. The calculated molecular mass of the mature protein of 436 amino acid residues was 47.03 kDa, which correlated well with the results of sodium dodecyl sulfate-polyacrylamide gel electrophoresis. The protein product of the OMP E gene had a leader peptide of 25 amino acids and a signal peptidase 1 cleavage site similar to those of known OMPs of *Escherichia coli*. The transcription initiation site of the OMP E gene was mapped by primer extension to be 78 nucleotides upstream of the ATG start codon. Borderline homology was found to the FadL protein of *E. coli* (49.1% similarity and 25.6% identity), which is involved in the binding and transport of fatty acids. Analysis of restriction fragment length polymorphisms of the OMP E genes of 19 different strains of *M. catarrhalis* showed that the OMP E gene is highly conserved. The high degree of conservation of sequences of the OMP E genes of *M. catarrhalis* from diverse sources, along with earlier observations that the protein contains antigenic determinants on the bacterial surface, indicates that OMP E should be studied further as a potential vaccine antigen.

Moraxella (Branhamella) catarrhalis is a gram-negative diplococcus. It is an important upper-respiratory-tract pathogen, causing otitis media and sinusitis in children (3). It is also a lower-respiratory-tract pathogen in adults with chronic bronchitis and chronic obstructive pulmonary disease (3). Outer membrane proteins (OMPs) of *M. catarrhalis* have been identified (OMP A to OMP H), and their molecular masses range from 98 to 21 kDa. Different strains have similar outer membrane profiles as determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (14). Little is known about the antigenic structure and function of the OMPs of *M. catarrhalis*. Studying OMPs will help elucidate antigenic structure, characterize the function of individual OMPs, and lead to an understanding of virulence factors and the immune response to infection.

OMP E is a heat-modifiable OMP with an apparent molecular mass (determined by SDS-PAGE) of 35 kDa at 25°C and 50 kDa when heated to 100°C (13). This heat-modifiable property is typical of several other OMPs (18). OMP E is expressed on the surface of the bacterium and has antigenic determinants which are conserved among most strains of *M. catarrhalis* from diverse clinical and geographic origins (15). Therefore, OMP E is a potential vaccine antigen.

In this paper we describe (i) cloning the gene encoding OMP E, (ii) determining its nucleotide sequence, (iii) mapping the transcription initiation site, and (iv) analyzing restriction fragment length polymorphisms of the OMP E genes of 19 different strains of *M. catarrhalis*.

* Corresponding author. Mailing address: Department of Veterans Affairs Medical Center, Medical Research 151, 3495 Bailey Ave., Buffalo, NY 14215.

MATERIALS AND METHODS

Bacterial strains, plasmids, media, and growth conditions. *M. catarrhalis* ATCC 25240 was obtained from the American Type Culture Collection (Rockville, Md.). Nineteen clinical isolates of *M. catarrhalis* used to study PCR restriction fragment length polymorphisms are listed in Table 1. *M. catarrhalis* was grown on chocolate agar plates at 37°C and 5% CO_2 or in Mueller-Hinton broth or brain heart infusion broth. Strains were stored in Mueller-Hinton broth with 10% glycerol at −80°C.

Escherichia coli LE392 (Stratagene, La Jolla, Calif.), used as the host strain for lambda (EMBL-3), was grown in tryptone broth supplemented with 0.2% maltose and 10 mM $MgSO_4$ at 37°C. The EMBL-3 library was plated on NZCYM plates for screening (22). *E. coli* JM101 and *E. coli* XL1-Blue MRF′ strains (Stratagene) were propagated in Luria-Bertani (22) broth at 37°C overnight with vigorous aeration. *E. coli* JM101 was used as the host strain for M13 mp18 phage (New England Biolabs, Beverly, Mass.). pET 22b$^+$ (Novagen, Madison, Wis.), pGEM 5ZF$^-$ (Promega Corp., Madison, Wis.), pACY6 (New England Biolabs), and pCR-Script SK(+) (Stratagene) were used for subcloning.

Determination of amino-terminal sequence of OMP E of *M. catarrhalis*. OMP E was electroeluted from a 10% SDS gel on which a preparation of purified outer membranes had been electrophoresed. Purified OMP E from strain ATCC 25240 was electrophoresed on a 10% acrylamide gel and transferred to a polyvinylidene difluoride membrane (10). OMP E was

TABLE 1. Sources of isolates of *M. catarrhalis*

Strain	Clinical source	Geographic source
5191	MEF[a]	Buffalo, N.Y.
555	MEF	Buffalo, N.Y.
585	Blood	Buffalo, N.Y.
7221	MEF	Buffalo, N.Y.
135	MEF	Buffalo, N.Y.
14	Sputum	Johnson City, Tenn.
701	Sputum	Buffalo, N.Y.
931	Sputum	Buffalo, N.Y.
42	Sputum	Johnson City, Tenn.
1	TTA[b]	Johnson City, Tenn.
56	Sputum	Johnson City, Tenn.
621	Sputum	Birmingham, England
690	Sputum	Birmingham, England
45	Sputum	Johnson City, Tenn.
9483	MEF	Buffalo, N.Y.
3584	MEF	Buffalo, N.Y.
58	Sputum	Johnson City, Tenn.
Tal2	Sinus	Philadelphia, Pa.
ATCC 25240		ATCC[c]

[a] MEF, middle ear fluid.
[b] TTA, transtracheal aspirate.
[c] ATCC, American Type Culture Collection.

TABLE 2. Family of degenerate oligonucleotides based on the amino-terminal sequence of OMP E of *M. catarrhalis*

Oligonucleotide	Sequence[a]
DG-1	CC(A/G)GA(T/C)GGAACATA(T/C)GCIGA(A/G)
DG-2	CA(A/G)GA(T/C)GGAACGTA(T/C)GCIGA(A/G)
DG-3	CA(A/G)GA(T/C)GGAACTTA(T/C)GCIGA(A/G)
DG-4	CA(A/G)GA(T/C)GGAACCTA(T/C)GCIGA(A/G)
DG-5	CA(A/G)GA(T/C)GGGACATA(T/C)GCIGA(A/G)
DG-6	CA(A/G)GA(T/C)GGGACGTA(T/C)GCIGA(A/G)
DG-7	CA(A/G)GA(T/C)GGGACTTA(T/C)GCIGA(A/G)
DG-8	CA(A/G)GA(T/C)GGGACCTA(T/C)GCIGA(A/G)
DG-9	CA(A/G)GA(T/C)GGTACATA(T/C)GCIGA(A/G)
DG-10	CA(A/G)GA(T/C)GGTACGTA(T/C)GCIGA(A/G)
DG-11	CA(A/G)GA(T/C)GGTACTTA(T/C)GCIGA(A/G)
DG-12	CA(A/G)GA(T/C)GGTACCTA(T/C)GCIGA(A/G)
DG-13	CA(A/G)GA(T/C)GGCACATA(T/C)GCIGA(A/G)
DG-14	CA(A/G)GA(T/C)GGCACGTA(T/C)GCIGA(A/G)
DG-15	CA(A/G)GA(T/C)GGCACTTA(T/C)GCIGA(A/G)
DG-16	CA(A/G)GA(T/C)GGCACCTA(T/C)GCIGA(A/G)

[a] Nucleotides in parentheses were mixed in equal amounts during synthesis of the oligonucleotide.

identified by staining with Coomassie brilliant blue R 250. The OMP E band was excised from the membrane, rinsed with distilled water, and air dried. The N-terminal sequence was determined by Edman degradation with an Applied Biosystems 471A automated protein sequencer.

Southern blot hybridization of *M. catarrhalis* genomic DNA and phage DNA with oligonucleotides. *M. catarrhalis* genomic DNA and lambda phage DNA were digested with restriction endonucleases and electrophoresed on a 0.6% agarose gel. DNA fragments were transferred to the GeneScreen hybridization transfer membrane (New England Nuclear Corp., Boston, Mass.), according to the manufacturer's instructions. The membranes were prehybridized, hybridized, and washed as described by Richet et al. (20), with some modifications. Briefly, the membranes were cross-linked in a UV Stratalinker 2400 at 120,000 μJ/cm^2 for approximately 30 s (Stratagene). The membranes were prehybridized for 4 h in buffer A (1 mM EDTA, 0.5 M $NaPO_4$ [pH 7.2], 7% SDS, 0.25 M NaCl, 2% bovine serum albumin) at 55 to 60°C. Forty picomoles of the oligonucleotide probe was end labeled with 100 μCi of [γ-^{32}P] ATP by using T4 polynucleotide kinase (Pharmacia, LKB Biotechnology, Piscataway, N.J.). Oligonucleotides were synthesized on an Applied Biosystems 380B DNA synthesizer. The ^{32}P-labeled oligonucleotide was added to fresh buffer A, and the membranes were incubated at 55°C overnight. The membranes were then washed three times with washing buffer (1 mM EDTA, 0.04 M sodium phosphate, 0.3 M NaCl, 1% SDS). The membranes were exposed to X-Omat AR film (Eastman Kodak Co., Rochester, N.Y.) with an intensifying screen for 24 to 36 h at −70°C and developed with a Kodak X-Omat automatic processor.

Screening the genomic library and analysis of the recombinant clones. An EMBL-3 genomic library of strain ATCC 25240 (16) was plated on 100 15-mm-diameter NZCYM plates with approximately 700 to 1,000 plaques per plate. The library was screened with a family of degenerate radiolabeled oligonucleotides corresponding to the amino-terminal sequence of OMP E (Table 2). The filters were hybridized and washed as described above. The initial positive plaques were picked and eluted in SM buffer (0.1 M NaCl, 8 mM $MgSO_4$, 0.05 M Tris-Cl, 0.01% gelatin [pH 7.5]). The plaques were purified by plating at low density and screening with the same oligonucleotide, until all the plaques were positive.

Lambda DNA of the positive clones was prepared from liquid lysates with the commercially available Qiagen system (Qiagen Inc., Chatsworth, Calif.). The lambda DNA was then digested with *Sal*I and subjected to electrophoresis on a 0.8% agarose gel to determine the presence of inserts.

Maxam-Gilbert base-specific chemical cleavage. Lambda DNA was prepared with the Qiagen Maxi kit. The DNA was cut with *Sal*I (New England Biolabs) and electrophoresed on a 0.8% agarose gel. The gel was stained with ethidium bromide, and the 12-kb insert fragment was excised from the gel and electroeluted with an IBI DNA electroeluter. The purified fragment was extracted with phenol and chloroform, precipitated with ethanol, and suspended in 10 mM Tris–1 mM EDTA (pH 8.0). The 12-kb purified fragment was again digested with *Nco*I, and a resulting 1.9-kb fragment was excised and eluted as described above. Fifty picomoles of the DNA fragment was dephosphorylated with 1 U of calf intestine alkaline phosphatase (Boehringer Mannheim, Indianapolis, Ind.) at 37°C for 60 min and end labeled with 100 μCi of [γ-^{32}P]ATP by using T4 polynucleotide kinase. This 1.9-kb fragment was digested with *Hin*dIII and electrophoresed on a 6% Tris-borate-EDTA acrylamide gel (Novex, San Diego, Calif.). The gel was stained with ethidium bromide and exposed to X-ray film. Two bands (1.1 and 0.8 kb) were excised, and the DNA was eluted in 0.5 M ammonium acetate–10 mM magnesium acetate–1 mM EDTA–0.1% SDS (crush soak buffer) overnight at 42°C (22). The eluted DNA was separated from the gel slices by passing through Poly-prep chromatography columns (Bio-Rad, Hercules, Calif.). The labeled DNA fragments were counted in a scintillation counter, and Maxam-Gilbert sequencing was performed (11).

PCR. One microgram of lambda DNA of clone B was used as the template, and the reaction was performed in the presence of 2 U of Vent R DNA polymerase (with exonuclease activity) (New England Biolabs). The reactions were carried out in a 50-μl volume with 0.25 μg of primers and 2.5 mM deoxyribonucleoside triphosphate. Predenaturing was done at 95°C for 3 min. Denaturing at 96°C for 15 s, annealing at 62°C for 1 min, and polymerization at 74°C for 1 min were done for

TABLE 3. Oligonucleotide primers corresponding to sequence of OMP E of *M. catarrhalis*

Primer	Sequence[a]
MG-1	5′-GGCTTGGGCAACTTTGTCATCACCCTCC-3′
MG-2	5′-GTTGAATTCACACCAGTTTGAAAATCCAAG-3′
A-PCR	5′-GGCCGGGAATTCCA(A/G)GA(T/C)GGCACTTA(T/C)GCIGA(A/G)-3′
B-PCR	5′-GCGCGCGAATTCGTTGGATTCACACAAGTTGAAAATCCAAG-3′
A-SC5′	5′-CGCATAAAAATTGTAAGAAAATATATATATTTTAC-3′
A-SC3′	5′-GCTATTTTTTATATTTAGACATTTTTGTATGATTTAGC-3′
B-SC5′	5′-GTGATGAGCTTAAAATTTGGATACAAAGCGCTGAG-3′
B-SC3′	5′-GCATGAGATTATTTGGCGTGATAAGCAAGC-3′
C-SC5′	5′-GCAGGCCTGGATCGCTCAGGGCAAGATGTGACTG-3′
PE-1	5′-CGCCAAACTCAGCGCTTTGTATCC-3′
PE-2	5′-GTCAGTCCTTACAATATGTAAAAC-3′

[a] Nucleotides in parentheses were mixed in equal amounts during synthesis of the oligonucleotide primer.

15 cycles in the presence of 3 mM $MgSO_4$. PCR was performed in the DNA Thermal Cycler GeneAmp PCR system 9600 (Perkin-Elmer Cetus, Norwalk, Conn.).

Subcloning and sequencing. The 0.8-kb PCR product was subjected to electrophoresis on a 6% Tris-borate-EDTA acrylamide gel. The gel was stained with ethidium bromide, and the 0.8-kb band was excised and treated as described above. The 0.8-kb fragment was subcloned at the *Eco*RI site into M13 mp18. Replicative-form M13 DNA was purified (22), digested with *Eco*RI, and subjected to electrophoresis on a 1% agarose gel to confirm the presence of the insert. Single-stranded M13 DNA was used to determine the nucleotide sequence by using the dideoxy-chain termination method of Sanger et al. (23). Sequencing reactions were performed with α-^{35}S-dATP and Sequenase (United States Biochemicals, Cleveland, Ohio). OMP E-specific oligonucleotides were synthesized and used for sequencing reactions. The rest of the OMP E gene was sequenced from the lambda clone directly with OMP E-specific primers.

On the basis of the nucleotide sequence of the gene encoding OMP E, three sets of primers were synthesized and used to amplify the OMP E gene. Primers A-SC5′ and A-SC3′ were used to amplify 1.573 kb of the OMP E gene which contained the complete OMP E gene with the promoter region. The second set of primers, B-SC5′ and B-SC3′, were used to amplify 1.391 kb of the OMP E gene containing the leader peptide and the rest of the OMP E gene. The third set of primers, C-SC5′ and B-SC3′, were used to amplify 1.313 kb of the OMP E gene from the first amino acid of the mature protein to the end of the carboxy terminus. The 3′ primers were complementary to the 3′ end of the OMP E gene.

All of the above-described PCR products were subcloned into the pCR-Script SK(+) phagemid (Stratagene) by using standard procedures (22). Plasmid DNA of the transformed colonies was isolated by the standard alkaline lysis method with the commercially available Qiagen system. The plasmid DNA was digested with *Sac*I and *Pst*I to determine the size of the insert.

Plasmid constructs pRB2 and pRB3 are pBluescript SK+ phagemid containing 1.391- and 1.313-kb fragments of the OMP E gene, respectively. The other strand of the OMP E gene was sequenced from pRB2.

Analysis of the sequence. Sequencing information was analyzed on an Apple Macintosh computer with DNA Strider, version 1.1 (9). Comparison of the sequence with sequences of other known genes was done with the BLAST algorithm developed by the National Center for Biotechnology Information at the National Library of Medicine (1), and sequence alignments were done with programs of the Genetics Computer Group sequence analysis software package, version 7.1-UNIX (5).

RNA isolation and mapping of the initiation site of the OMP E gene by primer extension. *M. catarrhalis* ATCC 25240 was grown in Mueller-Hinton broth overnight at 37°C. The cells were harvested and total RNA was extracted by the guanidine thiocyanate method (22). The RNA was subjected to agarose gel electrophoresis to verify its purity. Primer PE-1 (Table 3), specific for bp 114 to 137 of the OMP E gene, and primer PE-2 (Table 3), specific for bp 168 to 191, were 5′ end labeled with [γ-^{32}P]ATP by using T4 polynucleotide kinase. For primer extension, 50 μg of the total RNA was annealed with 100 fmol of the labeled primers and incubated at 55°C for 45 min. This step was followed by extension with 1 U of avian myeloblastosis virus reverse transcriptase (Promega Corp.) in the presence of deoxyribonucleoside triphosphates for 1 h at 42°C. The primer extension product was analyzed on an 8% urea acrylamide sequencing gel with labeled ϕX174 standards. Products of dideoxy nucleotide sequencing reactions generating a sequencing ladder and primed with the same primers were also electrophoresed in adjacent lanes to assess the exact base for the initiation of the OMP E transcript.

PCR restriction fragment length polymorphism mapping. Genomic DNA from 19 strains of *M. catarrhalis* was used as a template for the PCR. The two OMP E-specific primers used were B-SC5′ and B-SC3′ (excluding the promoter region). The OMP E gene was amplified by using the protocol described above. The PCR product was analyzed on a 0.8% agarose gel. Ten microliters of the amplified product was then digested with different restriction enzymes (*Hin*dIII, *Sau*96I, *Bsl*I, and *Bsg*I) and precipitated with ethanol. The digested product was again subjected to 6% acrylamide gel electrophoresis, and bands were visualized by staining with ethidium bromide.

Nucleotide sequence accession number. The GenBank accession number for the OMP E gene of *M. catarrhalis* is L31788.

RESULTS

Cloning and localization of the OMP E gene in genomic DNA and recombinant clones. The amino-terminal sequence determined for OMP E, eluted from a polyacrylamide gel, was AGLDRSGQDVTAFLQDGTYAE-VI (amino acid 22 was equivocal). On the basis of this sequence a series of 16 different degenerate oligonucleotides was synthesized, as shown in Table 2. To identify which of these oligonucleotides were best for screening the genomic library, Southern blot assays with *M. catarrhalis* and *E. coli* genomic DNA digested with different restriction enzymes were performed (Fig. 1A). The oligonu-

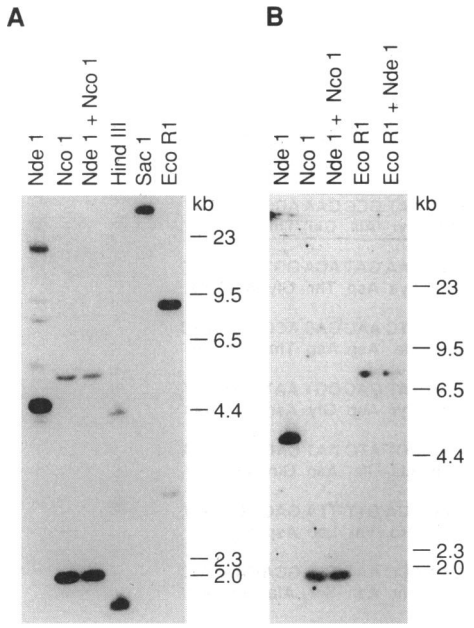

FIG. 1. Southern blot assays. (A) Genomic DNA of *M. catarrhalis* ATCC 25240 digested with the indicated restriction enzymes; (B) Purified phage DNA of clone B digested with the indicated restriction enzymes. The blots in panels A and B were hybridized with end-labeled oligonucleotides (9 to 16) (Table 2) specific for the sequence encoding the amino terminus of the OMP E gene. (The less prominent bands are the result of partial digestion of genomic DNA in panel B).

cleotides were labeled and assigned to two pools (oligonucleotides 1 to 8 and 9 to 16). The family of oligonucleotides 1 to 8 faintly hybridized to the *M. catarrhalis* and the *E. coli* genomic DNA, while oligonucleotides 9 to 16 hybridized strongly to *M. catarrhalis* DNA but not to *E. coli* DNA (data not shown). Each restriction digest produced a single strong band; this result was consistent with a single copy of the OMP E gene in the *M. catarrhalis* genome.

An EMBL-3 genomic library of *M. catarrhalis* ATCC 25240 was screened with the family of oligonucleotides 9 to 16 as the probe. A total of 8,100 plaques was screened, and six clones that hybridized strongly were isolated. Phage DNA of the clones digested with *Sal*I revealed insert sizes of 12 to 17 kb. To localize the OMP E gene, the insert of clone B was subjected to restriction enzyme digestion and Southern blot analysis with oligonucleotides 9 to 16 as the probe. The region of the OMP E gene that is complementary to the oligonucleotide 9-to-16 probe was located in fragments, as shown in Fig. 1B.

Several attempts to subclone the 4.4-kb *Nde*I (or *Nde*I-*Sal*I) and the 1.9-kb *Nco*I (or *Nco*I-*Sal*I) fragments into pET22b$^+$, pGEM 5zf$^-$, pACYC, and M13 mp18 vectors were unsuccessful. Since parallel attempts to clone a restriction fragment with the same ends into pET22b$^+$ and pGEM5zf$^-$ were successful, we suspect that the OMP E gene or the other unknown genes on the 4.4- or 1.9-kb fragment are toxic to *E. coli*.

Since we could not subclone the OMP E gene from the lambda clone, an alternative approach was taken. The nucleotide sequence of the ends of the 1.9-kb fragment was obtained by the method of Maxam and Gilbert. Two primers based on the Maxam-Gilbert sequence, MG-1 and MG-2 (Table 3), along with the degenerate oligonucleotide corresponding to the amino-terminal sequence of OMP E, were used to amplify the 5′ region of the OMP E gene by PCR. Oligonucleotide 15 (of the 16 degenerate oligonucleotides) hybridized most strongly to the genomic DNA in a Southern dot blot hybridization, and this and MG-1 and MG-2 were used as primers.

Primers MG-1, MG-2, and DG-15 (Table 2) were selected for the following reasons. Only one of the primers corresponding to the sequence in the 1.9-kb fragment should yield a product with the degenerate oligonucleotide corresponding to the amino-terminal protein sequence. Since the OMP E gene is approximately 1.4 kb, this strategy would allow us to locate the OMP E gene in the 1.9-kb fragment and to amplify all or part of the gene. In addition, since the PCR product will not contain the promoter region, it may be possible to subclone the OMP E gene, because it may be less toxic without the promoter. A DNA fragment of 0.8 kb was obtained by PCR with primers A-PCR (based on primer DG-15) and B-PCR (based on primer MG-2) (Table 3). This product was subcloned into M13 mp18 phage for sequencing. The promoter region and the carboxy-terminal region were sequenced directly from the lambda DNA.

Since the complete sequence of the OMP E gene was obtained, it was now possible to amplify the entire OMP E gene by PCR. Attempts to subclone the PCR product A (the entire OMP E gene including the promoter region) into pCR-Script SK(+) were unsuccessful, suggesting that the OMP E gene is toxic to *E. coli*. PCR products B (the entire open reading frame without the promoter) and C (the mature OMP E) were subcloned into pCR-Script SK(+). This was confirmed by determining the nucleotide sequences of the ends of the inserts in pRB2 and pRB3. All clones contained the OMP E coding sequence in the wrong orientation for expression by a plasmid promoter, again consistent with expression of OMP E being toxic to *E. coli*.

Sequence analysis of the gene encoding OMP E. The complete nucleotide sequence of the OMP E gene and the deduced amino acid sequence are shown in Fig. 2. An open reading frame of 1,377 bp (460 amino acids), starting with the codon ATG at position 154 and ending with TAA at position 1531, has been identified. A potential ribosome binding site, GGAGA, was located five bases upstream of the ATG translation initiation codon. The sequence GGAGA and the location that is five bases upstream of the ATG codon agree with the preferred spacing of the ribosome binding site in *E. coli* (24). Thirty bases downstream of the TAA stop codon was the sequence ATAAAAAATAGCTTGAATTTCAAGCTATTTT TTAT, a palindrome that could form a stem-loop structure which might serve as a potential transcriptional terminator (21).

The open reading frame defined a protein with a molecular mass of 49,334 Da. The predicted amino acids of OMP E suggested the presence of a signal peptide with a probable cleavage site between amino acids 25 and 26 (Ala-Gln-Ala/Ala; slash indicates cleavage site), which follows the "(−3−1)" rule for signal peptide cleavage (25). The deduced amino acid sequence of the first 24 amino acids from the putative cleavage site corresponds precisely to the N-terminal protein sequence determined from the purified OMP E. These results indicate that OMP E is synthesized as a precursor possessing a signal peptide composed of 25 amino acid residues. The leader peptide has a hydrophobic segment similar to known *E. coli* OMPs (19, 20). The predicted amino acid sequence of OMP E was subjected to a hydrophobicity profile with the algorithms of Kyte and Doolittle (8) and showed a strong hydrophobic portion corresponding to the signal peptide. The predicted antigenic determinants correspond to the hydrophilic region. The amino-terminal alanine residue of the mature OMP E protein has been designated the first amino acid. The predicted

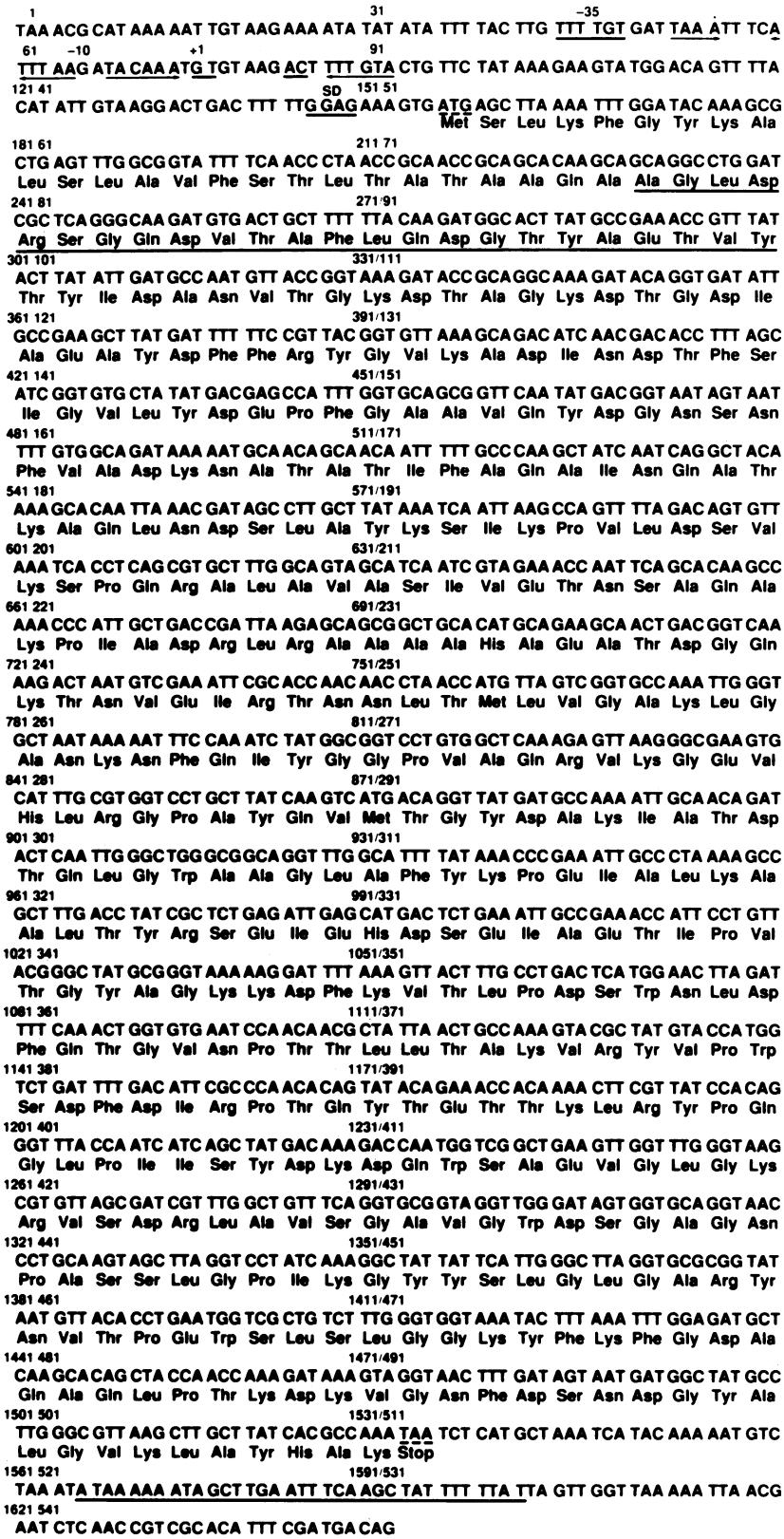

FIG. 2. Complete nucleotide sequence and the deduced amino acid sequence of the gene encoding OMP E of *M. catarrhalis* ATCC 25240. The numbers above the sequence indicate the nucleotide and amino acid positions. The putative Shine-Dalgarno ribosome binding sequence (SD) and the potential promoter sites −10 and −35 are noted. The translation initiation codon, ATG, and the stop codon, TAA, are indicated by dashed lines. The transcription site defined by primer extension is labeled +1. The inverted repeats of the potential transcription terminator stem-loop and the amino acids confirmed by N-terminal automated protein sequencing are underlined. The two regions with hyphenated dyad symmetry upstream of the gene are indicated by convergent arrows.

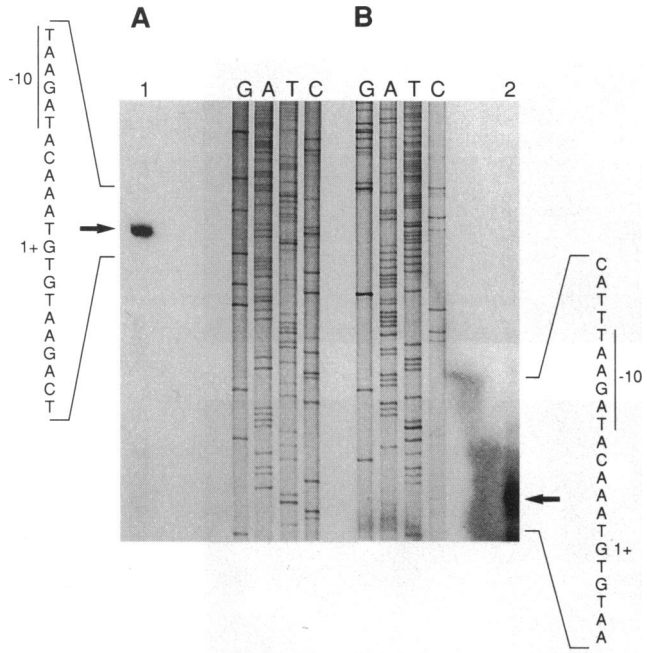

FIG. 3. Mapping of the transcription initiation site of the OMP E gene of *M. catarrhalis* by primer extension with primers PE1 (A) and PE2 (B) (Table 3). The same primers were also used to generate a dideoxy sequencing ladder with lambda DNA containing the OMP E gene as the template (lanes G, A, T, and C). The primer extension products in lane 1 (A) and lane 2 (B) (arrows) and the product of the sequencing reaction were resolved by electrophoresis. The complementary sequences are shown on the sides; +1 denotes the start of transcription of the OMP E gene.

molecular mass of the mature protein is 47,030 Da, which correlates well with the migration of the OMP E in SDS-PAGE.

The overall guanine and cytosine (G+C) content of the coding region of OMP E is 43.4%, which is similar to the reported G+C content of 41% for the *M. catarrhalis* genome (3). The 5′ end of the noncoding region including the promoter region is very AT rich. The A+T content of the noncoding region (1 to 153 bases) is 75.2%. AT rich promoters have been reported for several *E. coli* genes (17). Analysis of the amino acid composition of OMP E indicated that alanine, glycine, leucine, and valine are most abundant (range, 13 to 8%) and that no cysteine residues are present. Analysis of the codon usage of the OMP E coding region indicated a strong preference for T in the third position.

Mapping of the transcriptional initiation site of the OMP E gene by primer extension. To determine the transcriptional start site of the OMP E gene, primer extension analysis was performed with two different primers (PE1 and PE2) hybridizing to the 5′ region of the OMP E mRNA. This analysis indicated that the transcript starts with a guanine residue at position 75, which is 78 bases upstream of the ATG codon (Fig. 3). The potential −10 TAAGAT sequence, or the Pribnow box (nucleotide positions 63 to 68), was located six bases upstream of the +1 start site of transcription. The −35 (positions 40 to 45) TTGTTT sequence was located 17 bases upstream of the −10 sequence (Fig. 2). These hexameric sequences have similarity to the consensus sequences of −10 TATAAT and −35 TTGACA of promoters recognized by *E. coli* RNA polymerase. The spacing between the two promoters is also consistent with that for the *E. coli* promoters.

```
  1 MSLKFGYKALSLAVFSTLTATAAQAAGLD....RSGQDVTAFLQDGTYAE 46
    || |  :.  .|||  .|..|.| .||::    .|:. . |: .:|..|:
  3 MSQKTLFTKSALAVAVALISTQAWSAGFQLNEFSSSGLGRAYSGEGAIAD 52

 47 TVYTYIDANVTGKDTAGKDTGDIAEAYDFFRYGVKADINDTFSIGVLYDE 96
       ||. .:::.|      : . :|          ..||| |.:| :
 53 ......DAGNVSRNPA......LITMFD.........RPTFSAGAVYID 80

 97 PFGAAVQYDGNSNFVADKNATATIFAQAINQATKAQLNDSLAY.KSIKPV 145
    | ..  . ...:. : ..| .:| :..  |    |.:|| :::  ||..
 81 PDVNISGTSPSGRSLKADNIAPTAWVP..NMHFVAPINDQFGWGASITSN 128

146 LDSVKSPQRALAVASIVETNSAQAKPI....ADRLRAAAAHAEATDGQ.. 189
    .: ... . ..|.:|:.:|.. :.. :    | ||..| . : : ::
129 YGLATEFNDTYAGGSVGGTTDLETMNLNLSGAYRLNNAWSFGLGFNAVYV 178

190 KTNVEIRTNNLTMLVGAKLGANKNFQIYGGPVAQRVKGEVHLRGPAYQVM 239
    :..:|  ..:|. ||::         || .:|.:|  .|:. | :.| .:
179 RAKIERFAGDLGQLVAG........QIMQSPAGQTQQGQA.LAATANGI. 218

240 TGYDAKIA....TDTQLGWAAGLAFYKPEIALKAALTYRSEIEHD..... 280
     : :.|||    .: .:||.||: :|. : . : |||||||:. |
219 .DSNTKIAHLNGNQWGFGWNAGI.LYELDKNNRYALTYRSEVKIDFKGNY 266

281 .SEIAE.....TIPVTGYAGKKD....FKVTLPDSWNLDFQTGVNPTTLL 320
     |::.     .:|:.. .| :.    :.:.||: |::.  . |:|   :
267 SSDLNRAFNNYGLPIPTATGGRTQSGYLTLNLPEMWEVSGYNRVDPQWAI 316

321 TAKVRYVPWSDF.DIRPTQYTETTKLRYPQGLPIISYDKDQWSAEVGLGK 369
      .: |..||:| ::::|  .:.| :. .:|:      ||.:. .:| .
317 HYSLAYTSWSQFQQLKATSTSGDTLFQKHEGF......KDAYRIALGTTY 360

370 RVSDRLAVSGAVGWDSG...AGNPASSLGPIKGYYSLGLGARYNVTPEWS 416
      .|.:....::::|.:   |.|.. |: | .: :.|: |. |....: |
361 YYDDNWTFRTGIAFDDSPVPAQNRSISI.PDQDRFWLSAGTTYAFNKDAS 409

417 LSLGGKYF.....KFGDAQAQLPTKDKVGNFDSNDGYAL 450
    :.:|..|:      |:.::. |:...:|.  |:.| .||:
410 VDVGVSYMHGQSVKINEGPYQFESEGKAWLFGTNFNYAF 448
```

FIG. 4. Alignment of the deduced amino acid sequences of OMP E of *M. catarrhalis* and FadL of *E. coli*. Vertical lines indicate amino acid identity. Colons and periods indicate amino acids with comparative values greater than or equal to 0.5 and 0.1, respectively. The top sequence represents OMP E, and the bottom sequence represents FadL. The numbers to the left and right of the sequences are amino acid residue numbers. The GenBank nucleotide accession number for FadL is M37714.

Two regions of hyphenated dyad symmetry were identified downstream of the −35 region: 5′-TTAATTTCATTTAA-3′ and 5′-TACAAATGTGTAAGACTTTTGTA-3′. These regions may play a role in the regulation of expression of the OMP E gene.

Homology of OMP E with other known proteins. The deduced amino acid sequence of OMP E was compared with the sequences of other known proteins in the GenBank database by using the BLAST algorithms (blastp, blastx, and tblastn). This comparison indicated borderline homology between OMP E and OMP FadL of *E. coli*. BESTFIT and GAP analyses of OMP E and FadL revealed 49.1% similarity and 25.6% identity over 489 amino acid residues (Fig. 4).

Analysis of PCR restriction fragment length polymorphisms. To determine the degree of conservation of the OMP E gene among strains of *M. catarrhalis* from diverse geographic and clinical sources, restriction fragment length polymorphisms of the PCR products of the OMP E gene from 19 different strains were analyzed. The OMP E gene was present in all strains tested and was the same size (1.4 kb) in all the strains. The PCR product, when digested with *Hin*dIII, *Sau*96I, *Bsl*I, and *Bsg*I, showed no variation among the different strains tested with regard to the presence of restriction sites (Fig. 5). For each enzyme, the banding patterns were similar in all the strains (Fig. 5). These experiments indicate that the OMP E gene is highly conserved among strains of *M. catarrhalis*.

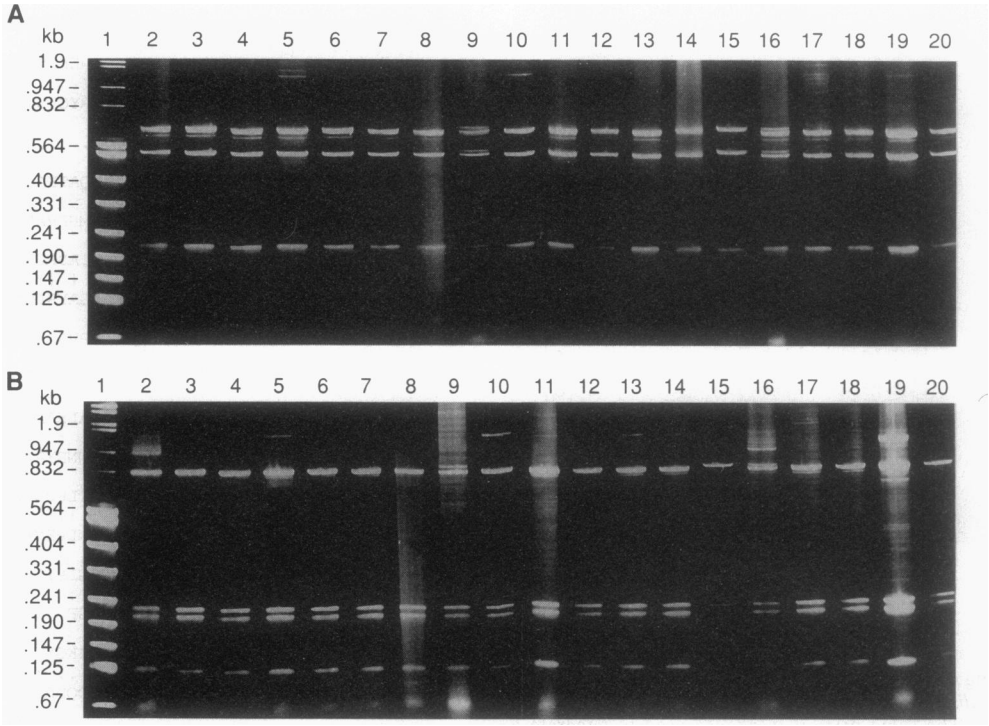

FIG. 5. Polyacrylamide gels stained with ethidium bromide. OMP E genes were amplified by PCR and digested with *Sau*96I (A) and *Bsl*I (B). Lane 1, molecular size standards. Lanes 2 through 20 contain the OMP E gene from different strains: 2, Tal2; 3, 58; 4, 3584; 5, 9483; 6, 45; 7, 690; 8, 621; 9, 56; 10, 1; 11, 42; 12, 931; 13, 701; 14, 14; 15, 135; 16, 7221; 17, 585; 18, 555; 19, ATCC 25240; 20, 5191.

DISCUSSION

In this study we have cloned the gene encoding OMP E, determined its nucleotide sequence, and defined the start site of transcription of the OMP E gene. Sequence analysis demonstrates that OMP E comprises 460 amino acid residues and encodes a protein with a molecular mass of 49.334 kDa. The predicted amino acid sequence of OMP E suggested the presence of a signal peptide of 25 amino acids, and this was confirmed by the N-terminal amino acid sequence of the protein. On the basis of the last three amino acids of the signal peptide, we predict that the precursor OMP E is cleaved by signal peptidase 1 (19). The −10 and −35 regions were identified by comparison with the preferred spacing and consensus sequences for RNA polymerase and sigma 70 (21). The high A+T content of the OMP E promoter may be important for the efficient transcription of the gene, because the presence of the AT base pair is thought to facilitate unwinding of the duplex DNA helical structure, which results in efficient initiation of mRNA transcription (4). The region of hyphenated dyad symmetry found in the OMP E gene located downstream of the −35 region may be involved in regulation of the OMP E gene. Such sequences are involved in the regulation of OMP genes in *E. coli*. For example, regions of hyphenated dyad symmetry are found in *fadL* downstream of the −10 region. These have similarity to the dyads defined in the *fadBA* gene (encoding the fatty acid-oxidation multienzyme complex) which represent potential binding sites for the FadR repressor (2). Similarly, at 5′ ends of the noncoding regions of *E. coli* porin genes, *ompC* and *ompF* are inverted repeats which may be involved as the recognition sites for the OmpR gene product (12).

We calculated (from the published sequence) the G+C contents of the known *M. catarrhalis* OMP CD gene (45.5%) and OMP B gene (43.8%), which are similar to the G+C content of the OMP E gene (43.4%) and the reported G+C content of the *M. catarrhalis* genome. Codon usage of OMP E shows that T is preferred over C. According to Grosjean and Fiers, the preferential use of T in the third position, rather than C, is a characteristic of highly expressed genes in *E. coli* (6).

A homology search revealed some similarity between OMP E and FadL of *E. coli*. FadL, like OMP E, is a heat-modifiable OMP (2). FadL acts as a receptor for the bacteriophage T2. The carboxy terminus of FadL is required for specific binding and transport of exogenous long-chain fatty acids with high affinity and facilitates their transfer across the outer membrane (7). Sequence similarity between OMP E and FadL suggests a similarity in function. The regions of closest homology between OMP E and FadL are from amino acids 270 to 280 of OMP E and 251 to 261 of FadL; there is 72% identity in this region. It is known that the binding and transport activity is located in the 3′ end of the *fadL* gene, but it is not yet known exactly where the functional activity for the phage receptor is located. It is not yet clear whether the homology of these two proteins indicates functional or structural similarity. Until studies of the function of OMP E are performed we cannot draw definite conclusions regarding the significance of the homology between OMP E and FadL.

OMP E is present in all strains of *M. catarrhalis* tested. Previous studies have shown that OMP E has determinants that are exposed on the surface of the intact bacterium and that these determinants are antigenically conserved among most strains of *M. catarrhalis*. These findings were demonstrated by adsorption experiments. Two polyclonal antisera made against whole-cell lysates of two different strains of *M. catarrhalis* were adsorbed individually with cells of 20 different

strains. The immunoblot assay with adsorbed and unadsorbed antisera showed that OMP E expressed surface-exposed determinants on 17 of 20 strains used in the study (15).

PCR restriction fragment length polymorphism analysis of different *M. catarrhalis* strains also demonstrated that OMP E is highly conserved in all strains tested. Four different enzymes were tested; three of them cut at three different sites in the gene and one cuts at two different sites. Therefore, the similar results for all strains indicate that the sequences recognized at the 11 sites are identical among the strains tested. Differences in restriction patterns among different strains may exist, but differences were not seen with the enzymes we tested.

When a gram-negative OMP is being considered as a potential vaccine antigen, three characteristics are important: (i) antigenic determinants should be present on the bacterial surface, (ii) an immune response to the OMP should be protective and, (iii) the protein should be conserved among the strains of the species. OMP E contains antigenic determinants on the bacterial surface (15). The present study establishes that OMP E is conserved among strains. Future work should focus on determining whether an immune response to OMP E is protective against infection.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grant AI28304 and by the Department of Veterans Affairs.

We thank Prem Seth for his scientific discussions and suggestions. We also thank Judah L. Rosner for his contributions and Moshe Soudasky for assistance with computer analysis. We also thank Adeline Thurston for help in preparing the manuscript and Prem Seth and Judah L. Rosner for critically reviewing the manuscript.

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403–410.
2. **Black, P. N.** 1991. Primary sequence of the *Escherichia coli fadL* gene encoding an outer membrane protein required for long-chain fatty acid transport. J. Bacteriol. **173**:435–442.
3. **Catlin, B. W.** 1990. *Branhamella catarrhalis*: an organism gaining respect as a pathogen. Clin. Microbiol. Rev. **3**:293–320.
4. **Chamberlin, M. J.** 1976. Interaction of RNA polymerase with DNA template. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
5. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
6. **Grosjean, H., and W. Fiers.** 1982. Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene **18**:199–209.
7. **Kumar, G. B., and P. N. Black.** 1993. Bacterial long-chain fatty acid transport. Identification of amino acid residues within the outer membrane protein FadL required for activity. J. Biol. Chem. **268**:15469–15476.
8. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
9. **Marck, C.** 1988. 'DNA Strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. Nucleic Acids Res. **16**:1829–1836.
10. **Matsudaira, P.** 1987. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. J. Biol. Chem. **262**:10035–10038.
11. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA **74**:560–564.
12. **Mizuno, T., M. Y. Chou, and M. Inouye.** 1983. A comparative study of the genes for three porins of the *Escherichia coli* outer membrane. J. Biol. Chem. **258**:6932–6940.
13. **Murphy, T. F.** 1989. The surface of *Branhamella catarrhalis*: a systematic approach to the surface antigens of an emerging pathogen. Pediatr. Infect. Dis. J. **8**:S75–S77.
14. **Murphy, T. F.** 1990. Studies of the outer membrane proteins of *Branhamella catarrhalis*. Am. J. Med. **88**(5A):41S–45S.
15. **Murphy, T. F., and L. C. Bartos.** 1989. Surface-exposed and antigenically conserved determinants of outer membrane proteins of *Branhamella catarrhalis*. Infect. Immun. **57**:2938–2941.
16. **Murphy, T. F., C. Kirkham, and A. J. Lesse.** 1993. The major heat-modifiable outer membrane protein CD is highly conserved among strains of *Branhamella catarrhalis*. Mol. Microbiol. **10**:87–98.
17. **Nakamura, K., and M. Inouye.** 1979. DNA sequence of the gene for the outer membrane lipoprotein of *E. coli*: an extremely AT-rich promoter. Cell **18**:1109–1117.
18. **Nakamura, K., and S. Mizushima.** 1976. Effects of heating in dodecyl sulfate solution on the conformation and electrophoretic mobility of isolated outer membrane proteins from *Escherichia coli* K-12. J. Biochem. **80**:1411–1422.
19. **Oliver, D.** 1985. Protein secretion in *Escherichia coli*. Annu. Rev. Microbiol. **39**:615–648.
20. **Richet, E., P. Abcarian, and H. A. Nash.** 1986. The interaction of recombination proteins with supercoiled DNA: defining the role of supercoiling in lambda integrative recombination. Cell **46**:1011–1021.
21. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
22. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
23. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
24. **Shine, J., and L. Dalgarno.** 1975. Determinant of cistron specificity in bacterial ribosomes. Nature (London) **254**:34–38.
25. **von Heijne, G.** 1984. How signal sequences maintain cleavage specificity. J. Mol. Biol. **173**:243–251.