INFECTION AND IMMUNITY, Jan. 1989, p. 88–94
0019-9567/89/010088-07$02.00/0
Copyright © 1989, American Society for Microbiology
Vol. 57, No. 1

Molecular Cloning, Expression, and Primary Sequence of Outer Membrane Protein P2 of *Haemophilus influenzae* Type b

ROBERT MUNSON, JR.,* AND ROBERT W. TOLAN, JR.

Edward Mallinckrodt Department of Pediatrics and Department of Microbiology and Immunology, Washington University School of Medicine, and Division of Infectious Diseases, St. Louis Children's Hospital, St. Louis, Missouri 63110

Received 8 September 1988/Accepted 11 October 1988

The structural gene for the porin of *Haemophilus influenzae* type b, designated outer membrane protein P2, was cloned, and the DNA sequence was determined. An oligonucleotide probe generated by reverse translation of N-terminal amino acid sequence data from the purified protein was used to screen genomic DNA. The probe detected a single *Eco*RI fragment of ~1,700 base pairs which was cloned to λgt11 and then into M13 and partially sequenced. The derived amino acid sequence indicated that we had cloned the N-terminal portion of the P2 gene. An overlapping ~1,600-base-pair *Pvu*II genomic fragment was cloned into M13, and the sequence of the remainder of the P2 gene was determined. The gene for P2 was then reconstructed under the control of the T7 promoter and expressed in *Escherichia coli*. The N-terminal sequence of the purified protein corresponds to residues 21 through 34 of the derived amino acid sequence. Thus, the protein is synthesized with a 20-amino-acid leader peptide. The M_r of the processed protein is 37,782, in good agreement with the estimate of 37,000 from sodium dodecyl sulfate-polyacrylamide gel electrophoresis.

Haemophilus influenzae type b is a major cause of bacterial meningitis and other invasive infections in children under the age of 4 years in the United States. We previously purified and characterized the P2 protein (designated b/c by Loeb and Smith [19]) from several *H. influenzae* type b strains (25). Antibody directed against this protein has protective activity in the infant rat bacteremic model (25). Vachon and co-workers independently purified this protein and demonstrated that it had porin activity in liposomes and in planar lipid bilayer membranes (34, 35). The molecular size limit for oligosaccharides was estimated to be approximately 1,400. To further characterize this outer membrane protein, we cloned, expressed, and sequenced the structural gene for P2. Recently, Hansen and co-workers also reported the cloning of the P2 gene from another *Haemophilus* type b strain; they cloned a ~10-kilobase-pair (kbp) *Pst*I fragment on a shuttle vector and were able to express the P2 gene in an Rd *Haemophilus* strain (14a). Determination of the primary sequence and expression of the P2 gene is an important step in the characterization of this porin and in determining the role of this protein in the pathogenesis of and immunity to *Haemophilus* disease.

(This work was presented in part at the 28th Interscience Conference on Antimicrobial Agents and Chemotherapy, Los Angeles, Calif., 23 to 26 October 1988.)

MATERIALS AND METHODS

Bacterial strains, plasmids, and phages. *H. influenzae* type b MinnA was chosen for study (1). *Haemophilus* type b strains are clonally diverse, as analyzed by outer membrane protein profiles (3) and by alloenzyme analysis (27, 28). Strain MinnA has the outer membrane protein subtype 1H and the electrophoretic type 1; thus, this strain is representative of the most prevalent type b clone seen in the United States (2, 28).

M13mp18, M13mp19, *Eco*RI-digested and phosphatase-treated pBR322, and *Escherichia coli* JM101 were obtained from New England BioLabs, Inc. (Beverly, Mass.). λgt11 arms were obtained from Stratagene (La Jolla, Calif.) along with *E. coli* LE392 and Y1088. pT7-7 and mGP1-2 (an M13 derivative containing the phage T7 RNA polymerase gene) were the kind gifts of Stanley Tabor (32). pGD103 was the kind gift of Bruce Green (9).

DNA hybridization. Chromosomal DNA of *H. influenzae* type b MinnA was prepared by a modification of the method of Marmur (22). Restriction analyses were performed on 0.7% agarose gels. Enzymes were purchased from Bethesda Research Laboratories, Inc. (Gaithersburg, Md.), or New England BioLabs, the incubations were carried out as recommended by the vendors. Southern hybridization and lambda plaque hybridization were performed on nitrocellulose as described by Maniatis et al. (21). M13 plaque hybridization was performed by the method of Zoller and Smith (37). Oligonucleotides were synthesized on an Applied Biosystems model 380B DNA synthesizer. Probes were end labeled with [γ-^{32}P]ATP (New England Nuclear Corp., Boston, Mass.) by using T4 polynucleotide kinase (New England BioLabs). Hybridization was performed overnight at 42°C, and blots were washed three times at room temperature in 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% sodium dodecyl sulfate (SDS), dried, and autoradiographed.

Molecular cloning. *Eco*RI digests of chromosomal DNA were prepared and fractionated on sucrose gradients as described by Clark-Curtiss et al. (8). Fractions containing DNA fragments 1 to 2 kbp in size were pooled, and a partial library was generated by ligation into λgt11 arms. Ligation mixtures were packed in vitro with Packagene (Promega, Madison, Wis.) and plate amplified in *E. coli* Y1088. Plaques were screened by hybridization with the ^{32}P-labeled oligonucleotide. A single positive clone was further characterized and contained the N-terminal portion of the P2 gene (see below). The remainder of the gene was isolated on a *Pvu*II fragment of ~1,600 base pairs (bp). To clone this fragment, we isolated *Pvu*II genomic fragments of 1 to 2 kbp by preparative agarose gel electrophoresis by using the DEAE paper technique (21) and cloned them directly into *Sma*I-treated M13mp18. To prepare plasmid constructs containing

* Corresponding author.

the P2 gene, we prepared the replicative form of recombinant M13 phages by the alkaline-SDS method (30) and isolated restriction fragments. Oligonucleotide-directed site-specific mutagenesis was performed by the method of Kunkel (16) with strains and reagents from Bio-Rad Laboratories (Richmond, Calif.). Other DNA methodologies were as described by Maniatis et al. (21) or Silhavy et al. (30).

DNA sequence analysis. DNA sequence analysis was performed by the dideoxy method with the US Biochemicals Sequenase kit as suggested by the manufacturer except that the reaction was performed at 42°C. [^{35}S]dATP was purchased from New England Nuclear. Data were analyzed with Compugene software (4) on a Digital VAX 8530 computer. 20-mer oligonucleotide primers were generated as necessary to complete the sequence.

Analytical techniques. P2 was purified as described previously (25). Tryptic peptides were generated and separated on a C18 Vydac column (4.6 mm by 25 cm, no. 218TP54; The Separations Group, Hesperia, Calif.) with an acetonitrile-H_2O gradient prepared in 0.1% trifluoroacetic acid (15, 24). Separations were performed on a Waters system (Millipore Corp., Itasca, Ill.) and monitored by measuring the A_{214}. A well-isolated peptide and the amino terminus of the purified protein were sequenced by automated Edman degradation at the Protein Chemistry Facility of Washington University (14).

Western blot (immunoblot) analysis was performed to determine whether the P2 gene was correctly reconstructed and expressed in *E. coli*. Strain JM101(pRSM478) was grown to an A_{600} of 0.6 in L broth containing 50 μg of ampicillin per ml and infected with mGP1-2, an M13 phage containing T7 RNA polymerase under the control of the *lac* promoter, at a multiplicity of infection of ~10. Isopropyl-β-D-thiogalactopyranoside was added to 1 mM. One hour later, cells were harvested, and crude extracts were prepared by sonication and compared with extracts prepared from uninfected strains JM101(pRSM478) and JM101(pT7-7), the strain containing the cloning vector. The protein concentrations of the extracts were determined by the bicinchoninic acid method (31) (BCA protein assay kit; Pierce Chemical Co., Rockford, Ill.) in accordance with the manufacturer's instructions. SDS-polyacrylamide gel electrophoresis was performed on 11% modified Laemmli gels as described by Lugtenberg et al. (20). Western blot conditions were as described previously (24). Nitrocellulose was probed sequentially with polyclonal rabbit anti-P2 antiserum (25) and alkaline phosphatase-conjugated goat antirabbit second antibody (Bio-Rad).

RESULTS

Outer membrane protein P2 was purified from strain MinnA as described previously (25), and the N-terminal sequence was determined by sequential Edman degradation. This sequence was reverse translated (Fig. 1). The nucleic acid sequence possibilities were numerous, and we had no codon usage data. Therefore, we synthesized a 23-mer oligonucleotide probe containing three inosines (29, 33) with a redundancy of 32. Genomic digests of MinnA DNA were prepared and analyzed by Southern hybridization. A single *Eco*RI fragment of ~1.7 kbp and a single *Pst*I fragment of ~10 kbp were observed (Fig. 2). Genomic DNA was digested to completion with *Eco*RI, and fragments of approximately 1 to 2 kbp were isolated by sucrose gradient centrifugation. A library of these fragments was generated in λgt11, and plaques were screened by hybridization. A single positive clone containing the 1,700-bp fragment was isolated. We were unable to subclone this fragment into pBR322 or pGD103, a low-copy-number plasmid vector (9). We were, however, able to clone this fragment directly into M13mp18. We used the mixed oligonucleotide as a sequencing primer and partially sequenced the *Eco*RI fragment. A 20-mer oligonucleotide complementary to the newly determined sequence was prepared. This oligonucleotide probe was used to screen our M13mp18 clones, and we isolated clones which hybridized to this oligonucleotide. These clones were partially sequenced with this oligonucleotide as a sequencing primer. We thus were able to sequence the DNA coding for the N-terminal portion of the protein and confirmed that we had cloned the 5′ portion of the P2 gene. The sequence of the remainder of the portion of the *Eco*RI fragment containing

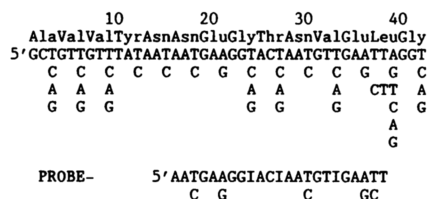

FIG. 1. N-terminal amino acid sequence of P2, possible nucleotide sequences derived by reverse translation, and the oligonucleotide probe used to isolate the P2 gene.

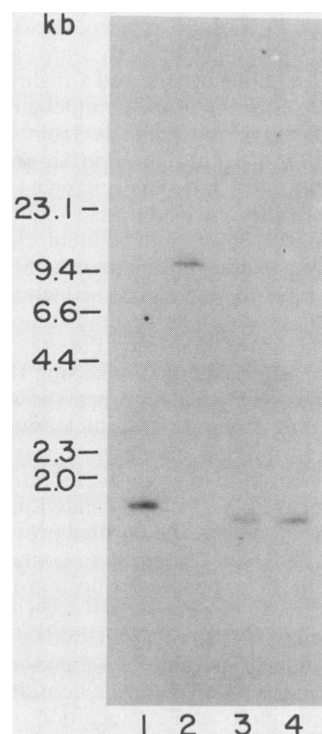

FIG. 2. Southern hybridization. Chromosomal DNA from *H. influenzae* type b MinnA (1 μg) was digested to completion with various restriction enzymes, electrophoresed on a 0.7% agarose gel, transferred to nitrocellulose, and probed with the ^{32}P-labeled mixed-oligonucleotide probe. Lanes: 1, *Eco*RI digest; 2, *Pst*I digest; 3, *Pst*I-*Pvu*II double digest; 4, *Pvu*II digest. The positions of λ *Hin*dIII size markers are shown in kilobase pairs (kb).

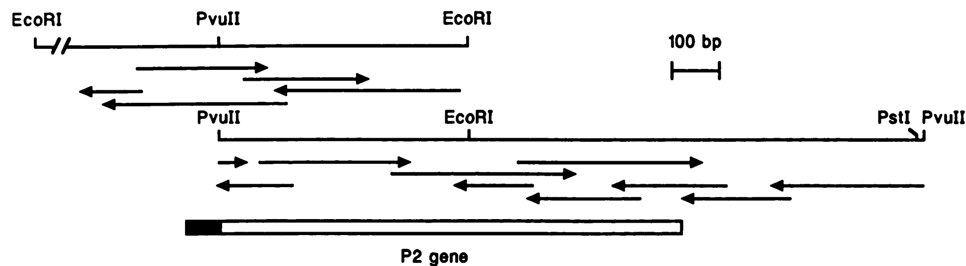

FIG. 3. Sequencing strategy for the P2 gene. The *Eco*RI and *Pvu*II fragments were cloned into M13 in both directions and sequenced by the dideoxy method with M13 primers and 20-mer oligonucleotide primers as indicated by the arrows. The coding region for the P2 gene is boxed; the open box represents the mature protein, and the solid box represents the signal peptide.

the N-terminal portion of the P2 gene was then completed (Fig. 3).

To clone the remainder of the P2 gene, we required a convenient overlapping genomic fragment containing sequences 3′ to those determined in the *Eco*RI fragment. Sequence analysis indicated that a *Pvu*II site was located near the N terminus of the P2 gene. We therefore analyzed genomic DNA cut with *Pvu*II and DNA cut with both *Pvu*II and *Pst*I by Southern hybridization (Fig. 2). A *Pvu*II fragment of ~1.6 kbp was observed. We estimated the P2 protein to have a molecular weight of 37,000 by SDS-polyacrylamide gel electrophoresis (25). A protein of this size would be coded for by a DNA fragment of approximately 1 kbp; we therefore concluded that the remainder of the P2 gene would be obtained by cloning this *Pvu*II fragment. Genomic DNA was digested with *Pvu*II, and 1- to 2-kbp fragments were isolated by preparative agarose gel electrophoresis with DEAE paper. The fragments were cloned into M13mp18, digested with *Sma*I, and screened by hybridization. One clone was saved for direct sequencing. To determine the sequence of the complementary strand, we isolated the replicative intermediate from an independent clone, digested it with *Kpn*I and *Hin*dIII, and cloned it into M13mp19 digested with the same enzymes. The *Kpn*I and *Hin*dIII sites were present in the M13 multiple cloning site but not in the insert. Thus, independent clones containing both strands were isolated. The sequence of the remainder of the P2 gene was then determined by sequencing these clones (Fig. 3). Throughout the open reading frame, the sequence was determined in both directions (Fig. 3).

The nucleotide sequence encompassing the P2 gene and the derived amino acid sequence are shown in Fig. 4. A single open reading frame spans nucleotides 234 through 1316. Multiple termination codons are present in all three reading frames downstream from nucleotide 1316. The calculated molecular weight of the translated protein is 39,683. The N-terminal sequence of the purified protein corresponds to residues 21 to 34. Thus, the protein contains a 20-amino-acid signal peptide. The processed protein has a molecular weight of 37,782, which agrees well with the estimate of 37,000 determined by SDS-polyacrylamide gel electrophoresis. A purified tryptic peptide was also sequenced; this peptide corresponds to translated nucleotides 1056 through 1121. The derived amino acid composition of the P2 protein was compared with the previously published composition determined for the protein purified from the same *Haemophilus* strain. These comparative data were in good agreement (Table 1) (25).

Our inability to subclone the *Eco*RI fragment into plasmid vectors suggested that expression of the porin gene was toxic in *E. coli*. We therefore separated the P2 gene from its promoter and expressed the gene under the control of the T7 promoter. No convenient restriction site was available between the ATG start and the putative upstream promoter. Therefore, the M13 phage containing the 1,700-bp *Eco*RI fragment (designated mP2A, Fig. 5) was subjected to oligonucleotide-directed mutagenesis. An *Nde*I site was created at nucleotides 231 to 236. Clones containing this *Nde*I site were identified and confirmed by DNA sequence analysis. A replicative form was isolated from one clone (designated mP2A1). The double-stranded intermediate was treated sequentially with *Eco*RI and *Nde*I. The fragment of ~600 bp containing the N-terminal portion of the P2 gene was cloned into pT7-7 cut with the same restriction enzymes; ligation mixtures were transformed into *E. coli* LE392, and ampicillin-resistant clones were isolated. Plasmids were isolated and characterized by restriction analysis; the plasmid designated pRSM432 had the predicted restriction pattern and was retained. The portion of the P2 gene 3′ to the *Eco*RI site was isolated as an *Eco*RI-*Pst*I fragment from the replicative form of an M13 clone containing the *Pvu*II fragment (designated mP2B). This fragment was ligated into *Eco*RI-*Pst*I-treated pRSM432 and transformed into *E. coli* JM101. Plasmids were prepared from several ampicillin-resistant strains, and those containing the complete P2 gene were identified by restriction analysis. Expression of the P2 gene was analyzed in several strains which contained plasmids with the correct restriction pattern. As the genetic construction was designed to place the P2 gene under the control of the T7 promoter, we infected these strains with an M13 phage containing the T7 RNA polymerase gene. During infection, the T7 RNA polymerase was synthesized and transcripts were produced downstream from the T7 promoter. Extracts were prepared from cells after 1 h of infection and analyzed for P2 protein by Western blotting. The full-sized P2 protein was produced in these constructs. Western blot analysis of JM101 (pRSM478) is shown in Fig. 6. Surprisingly, P2 protein was detectable in the absence of T7 RNA polymerase, indicating some *E. coli* promoter activity upstream of the P2 gene. Western blots of extracts of JM101(pT7-7) were negative (data not shown), indicating that the observed activity was not due to a cross-reactive *E. coli* protein. DNA sequence analysis of the 5′ portion of pRSM478 confirmed that the genetic construction was correct. The sequence has been submitted to GenBank/EMBL; the accession number is J03359.

DISCUSSION

Outer membrane antigens of *H. influenzae* type b are possible components of future vaccines (13). Both protein and lipooligosaccharide epitopes have been demonstrated on

```
       10        20        30        40        50        60        70
TTTTTCGATACCTAAAAATATTTATAGGTAATACTTAAACCTAAAACAGAATAAAAAATAATCAAATTCA

       80        90       100       110       120       130       140
TTAAAAAATGTGATCTCGATCAGATTTCAAGAAAATTAAAATTTTGGAGTATTGACATCAAAATTTTTTT

      150       160       170       180       190       200       210
GTAAAGATGCAGACCGTCCATTCTGGCGATTGGACAATTCTATTGGAGAAAAGTTCAATCATAGA"AGTA

      220       230       240       250       260       270       280
AACAACCATAAGGAATACAAATTATGAAAAAAACACTTGCAGCATTAATCGTTGGTGCATTCGCAGCTTC
                       MetLysLysThrLeuAlaAlaLeuIleValGlyAlaPheAlaAlaSe
     ____

      290       300       310       320       330       340       350
AGCAGCAAACGCAGCTGTTGTTTATAACAACGAAGGGACTAACGTAGAATTAGGTGGTCGTTTAAGCATT
rAlaAlaAsnAlaAlaValValTyrAsnAsnGluGlyThrAsnValGluLeuGlyGlyArgLeuSerIle
      ____________________________________________
    PvuII

      360       370       380       390       400       410       420
ATCGCAGAACAAAGTAATAGCACTGTAGATAATCAAAAACAGCAACACGGTGCATTACGCAATCAAGGTT
IleAlaGluGlnSerAsnSerThrValAspAsnGlnLysGlnGlnHisGlyAlaLeuArgAsnGlnGlyS

      430       440       450       460       470       480       490
CACGTTTCCACATTAAAGCAACTCATAACTTCGGTGATGGTTTCTATGCACAAGGTTATTTAGAAACTCG
erArgPheHisIleLysAlaThrHisAsnPheGlyAspGlyPheTyrAlaGlnGlyTyrLeuGluThrAr

      500       510       520       530       540       550       560
TTTTGTTACAAAAGCCTCTGAAAACGGTTCAGATAACTTCGGTGATATTACAAGCAAATATGCTTATGTT
gPheValThrLysAlaSerGluAsnGlySerAspAsnPheGlyAspIleThrSerLysTyrAlaTyrVal

      570       580       590       600       610       620       630
ACTTTAGGAAATAAAGCATTCGGTGAAGTAAAACTTGGTCGTGCGAAAACTATTGCTGATGGCATAACAA
ThrLeuGlyAsnLysAlaPheGlyGluValLysLeuGlyArgAlaLysThrIleAlaAspGlyIleThrS

      640       650       660       670       680       690       700
GTGCAGAAGATAAAGAATATGGCGTTCTCAACAATAGTGACTATATTCCTACTAGTGGTAATACCGTTGG
erAlaGluAspLysGluTyrGlyValLeuAsnAsnSerAspTyrIleProThrSerGlyAsnThrValGl

      710       720       730       740       750       760       770
CTATACTTTTAAAGGTATTGATGGTTTAGTATTAGGCGCTAATTATTTATTAGCACAAAAGCGTGAGGGT
yTyrThrPheLysGlyIleAspGlyLeuValLeuGlyAlaAsnTyrLeuLeuAlaGlnLysArgGluGly

      780       790       800       810       820       830       840
GCAAAAGGTGAAAATAAGCGGCCTAATGATAAGGCTGGTGAAGTACGTATAGGTGAAATCAATAATGGAA
AlaLysGlyGluAsnLysArgProAsnAspLysAlaGlyGluValArgIleGlyGluIleAsnAsnGlyI
                                                                   Eco

      850       860       870       880       890       900       910
TTCAAGTTGGTGCAAAATATGATGCAAACGACATCGTTGCAAAAATTGCTTATGGTAGAACTAACTACAA
leGlnValGlyAlaLysTyrAspAlaAsnAspIleValAlaLysIleAlaTyrGlyArgThrAsnTyrLy
RI

      920       930       940       950       960       970       980
ATATAACGAATCTGACGAGCATAAACAGCAATTAAATGGTGTATTAGCAACTTTAGGCTATCGTTTTAGT
sTyrAsnGluSerAspGluHisLysGlnGlnLeuAsnGlyValLeuAlaThrLeuGlyTyrArgPheSer

      990      1000      1010      1020      1030      1040      1050
GATTTAGGCTTATTAGTGTCTCTAGATAGTGGCTATGCAAAAACTAAAAACTATAAAATTAAACACGAAA
AspLeuGlyLeuLeuValSerLeuAspSerGlyTyrAlaLysThrLysAsnTyrLysIleLysHisGluL
                     XbaI

     1060      1070      1080      1090      1100      1110      1120
AACGCTATTTCGTATCTCCAGGTTTCCAATATGAATTAATGGAAGATACTAATGTCTATGGCAACTTCAA
ysArgTyrPheValSerProGlyPheGlnTyrGluLeuMetGluAspThrAsnValTyrGlyAsnPheLy
    __________________________________________________________________

     1130      1140      1150      1160      1170      1180      1190
ATATGAACGCACTTCTGTAGATCAAGGTGAAAAAACACGTGAACAAGCAGTATTATTCGGTGTAGATCAT
sTyrGluArgThrSerValAspGlnGlyGluLysThrArgGluGlnAlaValLeuPheGlyValAspHis
_

     1200      1210      1220      1230      1240      1250      1260
AAACTTCACAAACAACTATTAACCTATATTGAAGGTGCTTACGCTAGAACTAGAACAACTGAGACAGGTA
LysLeuHisLysGlnLeuLeuThrTyrIleGluGlyAlaTyrAlaArgThrArgThrThrGluThrGlyL

     1270      1280      1290      1300      1310      1320      1330
AAGGCGTAAAAACTGAAAAAGAAAAATCAGTGGGTGTAGGTTTACGCGTTTACTTCTAATCATTTGTTAG
ysGlyValLysThrGluLysGluLysSerValGlyValGlyLeuArgValTyrPhe

     1340      1350      1360      1370      1380      1390      1400
AAATACATTATTAAAAGCAAGGCGAATCGAAAGATTCGCTTTTTTTGCTCAAAATCAAGTTAAAAAATGA
                 _________    _________
                      _  ___    ___  _

     1410      1420      1430      1440      1450      1460      1470
TTAAGTTAAAAGTGTATAAATATTTAGGCTATTTTATAAGTAACAAAATATTAATAAAAAATCTGTGACA

     1480      1490      1500      1510      1520      1530      1540
TATATCACAGATTTTTAAATCAATTAACTATTTAAGTGTTTACTATTAATTCTCTTTCCACTTTCCGTTT

     1550      1560      1570      1580      1590      1600      1610
ACTACTGTGCCGATTACTTGGTAATTTGGCGTAAACACGGCTAAGTTTGCTATCTTACCTTTTTCTACCG

     1620      1630      1640      1650      1660      1670      1680
AACCTAAACGATCATCTATACCAATTGCTCGTGCAGGATAAAGGTTACTCATTCTAATGGCTTCCGCAAA

     1690      1700      1710      1720      1730      1740      1750
GAGGAATTTCAACATATTCTACCGCATTTTTAATAGATTCCATCATTGTGATTGATGCCCCTGCAATTGT

     1760      1770      1780      1790      1800      1810      1820
GCCATTTGCATCATAGCAACGCCCTTCTTTGATATAAATTGTTTTTCCTACAAAGGTAAAACTTTCTAGT

     1830      1840      1850      1860      1870      1880      1890
TCTGGCGGTGCGCCTGCAGGGG
           PstI
              PvuII/SmaI
```

FIG. 4. Complete nucleotide and derived amino acid sequences of the P2 gene and DNA sequences 5′ and 3′ to the coding region. Sequences of interest are underlined. The putative ribosome-binding site corresponds to nucleotides 220 to 224. A single open reading frame is present (nucleotides 234 to 1316). The N-terminal sequence of the purified protein and the sequence of a tryptic peptide are underlined. A putative transcriptional terminator is also underlined (nucleotides 1349 to 1372). The *Pvu*II and *Eco*RI sites and an internal *Xba*I site are indicated.

the surfaces of intact cells, and antibodies against certain outer membrane proteins and lipooligosaccharide epitopes have been shown to be protective in an infant rat bacteremic model. We have demonstrated that an antibody directed against the P2 protein is protective in an infant rat bacteremic model (25). Recently, Murphy and Bartos (26) isolated a monoclonal antibody directed against the P2 protein of a nontypeable *Haemophilus* isolate. This antibody recognizes a surface-exposed epitope and has in vitro bactericidal activity.

Vachon and co-workers (34, 35) demonstrated that the P2 protein has porin activity both in liposome preparations and in planar lipid bilayer membranes. Liposomes containing P2 were permeable to oligosaccharides, and an estimate of the M_r exclusion limit of approximately 1,400 was determined. Burns and co-workers (6) demonstrated that certain chloramphenicol-resistant strains lack a major protein with an M_r of approximately 40,000, possibly P2. Presumably, the resistance of these strains to chloramphenicol was due to changes in outer membrane permeability.

To further characterize the protein, we attempted to isolate the structural gene for P2. We were unable to detect clones expressing P2 in our λEMBL3 libraries, libraries which contained the genes for outer membrane proteins P1 (24) and P6 (M. B. Nelson, M. A. Apicella, J. P. Molleston, R. S. Munson, Jr., and T. F. Murphy, Program Abstr. Annu. Meet. Am. Soc. Microbiol. 1988, D69, p. 82). We therefore concluded that the expression of the P2 gene was toxic and that the clones of interest were deleted from the library. We reasoned that it should be possible to isolate portions of the gene by screening a library of small DNA fragments with an oligonucleotide probe. A mixed-oligonucleotide probe which corresponded to the reverse-translated N-terminal sequence was prepared. We isolated an *Eco*RI fragment containing the N-terminal portion of the P2 gene in λgt11 and cloned the remainder of the P2 gene as a *Pvu*II fragment. The gene was then reconstructed and expressed in the T7 expression vector pT7-7. Similar strategies were used by Carbonetti and Sparling (7) and by Gotschlich et al. (12) for the cloning of the porin genes from *Neisseria gonorrhoeae*. Carbonetti and Sparling also expressed the gonococcal porin gene in a T7 expression system.

The sequences 5′ and 3′ to the coding region were analyzed. Surprisingly, the portion of the sequence 5′ to the coding region contains no strong consensus promoter sequence (23) and no open reading frame. A putative ribo-

TABLE 1. Comparison of the amino acid composition of the P2 protein with the derived composition of the P2 protein from DNA sequence data

Residue	Data from: Derived amino acid sequence No.	Mol%	Acid hydrolysate (mol%)[a]
Ala	24	7.0	7.7
Arg	16	4.7	4.7
Asn	25	7.3	
Asp	17	5.0	
Asx			12.6[b]
Cys	0	0	ND[c]
Gln	14	4.1	
Glu	24	7.0	
Glx			11.7[d]
Gly	40	11.7	12.2
His	7	2.0	2.1
Ile	15	4.4	5.9
Leu	24	7.0	8.4
Lys	30	8.8	8.8
Met	1	0.3	0.1
Phe	13	3.8	3.1
Pro	3	0.9	0.6
Ser	17	5.0	5.2
Thr	24	7.0	6.9
Trp	0	0	ND
Tyr	23	6.7	4.5
Val	24	7.0	5.3

[a] Data are for purified P2 from strain MinnA (25).
[b] Sum of values for Asn and Asp.
[c] ND, Not determined.
[d] Sum of values for Gln and Glu.

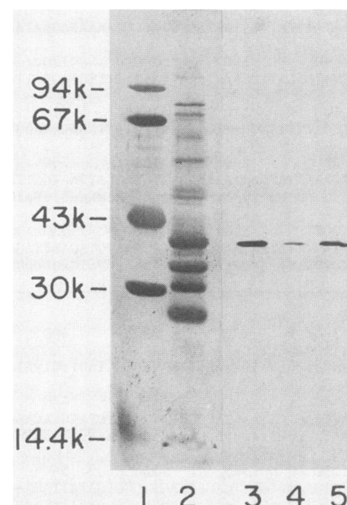

FIG. 6. Western blot analysis of JM101(pRSM478). Lanes 1 and 2 contain india ink-stained molecular weight markers (lane 1) and a detergent-insoluble fraction enriched in outer membrane proteins from *H. influenzae* type b MinnA (5 μg) (lane 2). Lanes 3 to 5 contain 20 μg of sonicate from the following strains and were developed immunologically with anti-P2 antiserum: *H. influenzae* type b MinnA (lane 3), JM101(pRSM478), uninduced (lane 4), and JM101(pRSM478), induced for 1 h by infection with mGP1-2 (lane 5). k, Kilodaltons.

some-binding site, AAGG, is located 9 bases upstream from the translational start site (11). Several potential factor-independent transcriptional terminators were identified downstream from the translational termination site. The strongest transcriptional terminator, as determined by the

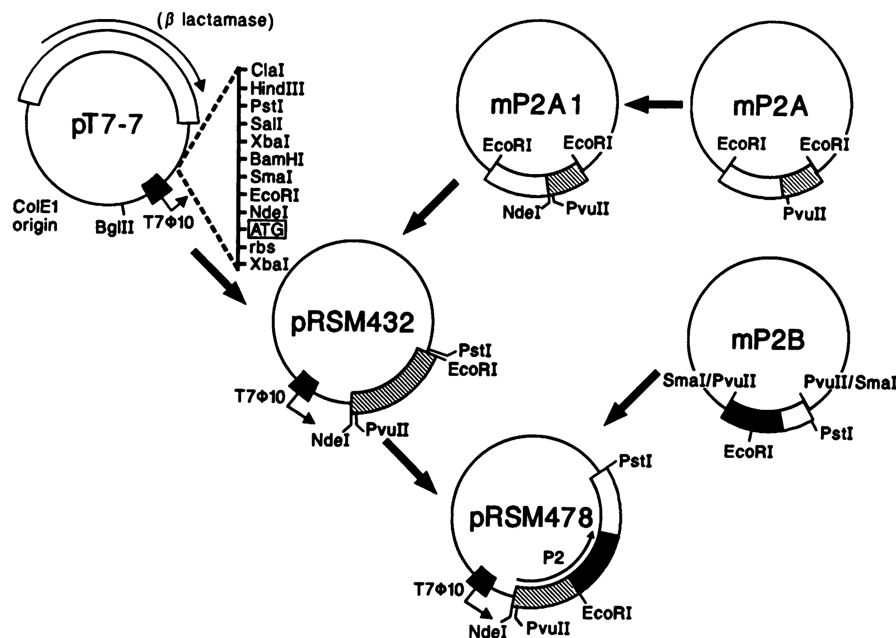

FIG. 5. Reconstruction of the P2 gene. The M13 phage containing the ~1,700-bp *Eco*RI fragment (mP2A) was subjected to oligonucleotide-directed mutagenesis to create an *Nde*I site at codons 231 to 236. This phage was designated mP2A1. A replicative form of mP2A1 was isolated, and the fragment of ~600 bp containing the N-terminal portion of the P2 gene (hatched bars) was cloned into pT7-7 to create pRSM432. The ~1-kbp *Eco*RI-*Pst*I fragment containing the 3′ portion of the P2 gene (solid bars) and approximately 500 bp 3′ to the gene (open bars) was obtained from the replicative form of mP2B, the M13 phage containing the *Pvu*II fragment, and cloned into pRSM432 to create pRSM478. Vector sequences are shown as lines. The position of the P2 gene and the direction of transcription from the T7ϕ10 promoter are shown by arrows.

algorithm of Brendel et al. (5), was localized to two regions of overlapping diad symmetry spanning bases 1349 through 1372. Three other potential terminators were localized between bases 1400 and 1600.

The derived amino acid sequence of the protein was determined. The N-terminal sequence was determined chemically and corresponds to amino acid residues 21 to 34. The 20-amino-acid signal peptide shares many structural features with other procaryotic signal peptides (36). Lysine is present as the second and third amino acids. The positively charged, hydrophilic N terminus is followed by a stretch of hydrophobic residues. Alanine is present at residues 18 and 20 (positions −1 and −3 of the mature protein). Alanine is also present at positions −1 and −3 of the *H. influenzae* type b outer membrane protein P1 signal peptide (24) and is a common feature of many procaryotic signal peptides (36). The derived amino acid composition of the protein is in good agreement with data obtained from acid hydrolysates of the purified protein. The P2 protein contains one methionine, consistent with the previously observed two cyanogen bromide peptides (25). The protein is also notable for the absence of cysteine and tryptophan.

The sequence was compared with the available sequences in the National Biomedical Research Foundation library Protein Sequence Database (Release 14, Protein Identification Resource, 1987) by using FASTP software (18). The sequence was also compared with the two published gonococcal porin sequences by using the algorithm of Lipman and Pearson (18). The overall homology between P2 and other porin proteins was low, 20% or less. No transmembrane domains were apparent when the sequence was analyzed by the hydropathy method of Kyte and Doolittle (17) or by the method of Engelman et al. (10). However, structural features similar to those seen in the *E. coli* OmpF porin were discernible by comparison of the Kyte-Doolittle plots.

The cloned P2 gene will be used to map the surface-exposed epitopes of the protein. Current studies are also directed toward overexpression and purification of the recombinant antigen for immunogenicity and functional studies.

ACKNOWLEDGMENTS

We thank Camella Bailey and Susan Grass for excellent technical assistance and Roy Curtiss III and Charlotte Parker for helpful discussions. We also thank Geoffrey Weinberg for the isolation of the tryptic peptides, Gregory Grant and Mark Frazier of the Washington University Protein Chemistry Laboratory for the protein sequencing and analysis, and Jane Gold for the preparation of oligonucleotides.

This work was supported by Public Health Service grant R01 AI 17572 from the National Institutes of Health.

LITERATURE CITED

1. **Barenkamp, S. J., D. M. Granoff, and R. S. Munson, Jr.** 1981. Outer-membrane protein subtypes of *Haemophilus influenzae* type b and spread of disease in day-care centers. J. Infect. Dis. **144**:210–217.
2. **Barenkamp, S. J., R. S. Munson, Jr., and D. M. Granoff.** 1981. Comparison of outer-membrane protein subtypes and biotypes of isolates of *Haemophilus influenzae* type b. J. Infect. Dis. **144**: 480.
3. **Barenkamp, S. J., R. S. Munson, Jr., and D. M. Granoff.** 1981. Subtyping isolates of *Haemophilus influenzae* type b by outer-membrane protein profiles. J. Infect. Dis. **143**:668–676.
4. **Barnes, W. M.** 1987. Sequencing DNA with dideoxyribonucleotides as chain terminators: hints and strategies for big projects. Methods Enzymol. **152**:538–556.
5. **Brendel, V., G. H. Hamm, and E. N. Trifonov.** 1986. Terminators of transcription with RNA polymerase from *Escherichia coli*: what they look like and how to find them. J. Biomol. Struct. Dyn. **3**:705–723.
6. **Burns, J. L., P. M. Mendelman, J. Levy, T. L. Stull, and A. L. Smith.** 1985. A permeability barrier as a mechanism of chloramphenicol resistance in *Haemophilus influenzae*. Antimicrob. Agents Chemother. **27**:46–54.
7. **Carbonetti, N. H., and P. F. Sparling.** 1987. Molecular cloning and characterization of the structural gene for protein I, the major outer membrane protein of *Neisseria gonorrhoeae*. Proc. Natl. Acad. Sci. USA **84**:9084–9088.
8. **Clark-Curtiss, J. E., W. R. Jacobs, M. A. Docherty, L. R. Ritchie, and R. Curtiss III.** 1985. Molecular analysis of DNA and construction of genomic libraries of *Mycobacterium leprae*. J. Bacteriol. **161**:1093–1102.
9. **Deich, R. A., B. J. Metcalf, C. W. Finn, J. E. Farley, and B. A. Green.** 1988. Cloning of genes encoding a 15,000-dalton peptidoglycan-associated outer membrane lipoprotein and an antigenically related 15,000-dalton protein from *Haemophilus influenzae*. J. Bacteriol. **170**:489–498.
10. **Engelman, D. M., T. A. Steitz, and A. Goldman.** 1986. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. Annu. Rev. Biophys. Biophys. Chem. **15**: 321–353.
11. **Gold, L., and G. Stormo.** 1987. Translation initiation, p. 1302–1307. *In* J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
12. **Gotschlich, E. C., M. E. Seiff, M. S. Blake, and M. Koomey.** 1987. Porin protein of *Neisseria gonorrhoeae*: cloning and gene structure. Proc. Natl. Acad. Sci. USA **84**:8135–8139.
13. **Granoff, D. M., and R. S. Munson, Jr.** 1986. Prospects for prevention of *Haemophilus influenzae* type b disease by immunization. J. Infect. Dis. **153**:448–461.
14. **Grant, G. A., and V. A. Chiappinelli.** 1985. κ-Bungarotoxin: complete amino acid sequence of a neuronal nicotinic receptor probe. Biochemistry **24**:1532–1537.

14a. **Hansen, E. J., F. R. Gonzales, N. R. Chamberlain, M. V. Norgard, E. E. Miller, L. D. Cope, S. E. Pelzel, B. Gaddy, and A. Clausell.** 1988. Cloning of the gene encoding the major outer membrane protein of *Haemophilus influenzae* type b. Infect. Immun. **56**:2709–2716.

15. **Hortin, G., H. Sims, and A. W. Strauss.** 1986. Identification of the site of sulfation of the fourth component of human complement. J. Biol. Chem. **261**:1786–1793.
16. **Kunkel, T. A.** 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc. Natl. Acad. Sci. USA **82**: 488–492.
17. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
18. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.
19. **Loeb, M. R., and D. H. Smith.** 1980. Outer membrane protein composition in disease isolates of *Haemophilus influenzae*: pathogenic and epidemiological implications. Infect. Immun. **30**:709–717.
20. **Lugtenberg, B., J. Meijers, R. Peters, P. van der Hoek, and L. van Alphen.** 1975. Electrophoretic resolution of the major outer membrane protein of *Escherichia coli* K12 into four bands. FEBS Lett. **58**:254–258.
21. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. **Marmur, J.** 1961. A procedure for the isolation of deoxyribonucleic acid from micro-organisms. J. Mol. Biol. **3**:208–218.
23. **McClure, W. R.** 1985. Mechanism and control of transcription initiation in prokaryotes. Annu. Rev. Biochem. **54**:171–204.
24. **Munson, R., Jr., and S. Grass.** 1988. Purification, cloning, and sequence of outer membrane protein P1 of *Haemophilus influenzae* type b. Infect. Immun. **56**:2235–2242.
25. **Munson, R. S., Jr., J. L. Shenep, S. J. Barenkamp, and D. M.**

Granoff. 1983. Purification and comparison of outer membrane protein P2 from *Haemophilus influenzae* type b isolates. J. Clin. Invest. **72**:677–684.

26. **Murphy, T. F., and L. C. Bartos.** 1988. Purification and analysis with monoclonal antibodies of P2, the major outer membrane protein of nontypable *Haemophilus influenzae*. Infect. Immun. **56**:1084–1089.
27. **Musser, J. M., S. J. Barenkamp, D. M. Granoff, and R. K. Selander.** 1986. Genetic relationships of serologically nontypable and serotype b strains of *Haemophilus influenzae*. Infect. Immun. **52**:183–191.
28. **Musser, J. M., D. M. Granoff, P. E. Pattison, and R. K. Selander.** 1985. A population genetic framework for the study of invasive diseases caused by serotype b strains of *Haemophilus influenzae*. Proc. Natl. Acad. Sci. USA **82**:5078–5082.
29. **Ohtsuka, E., S. Matsuki, M. Ikehara, Y. Takahashi, and K. Matsubara.** 1985. An alternative approach to deoxyoligonucleotides as hybridization probes by insertion of deoxyinosine at ambiguous codon positions. J. Biol. Chem. **260**:2605–2608.
30. **Silhavy, T. J., M. L. Berman, and L. W. Enquist.** 1984. Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
31. **Smith, P. K., R. I. Krohn, G. T. Hermanson, A. K. Mallia, F. H. Gartner, M. D. Provenzano, E. K. Fujimoto, N. M. Goeke, B. J. Olson, and D. C. Klenk.** 1985. Measurement of protein using bicinchoninic acid. Anal. Biochem. **150**:76–85.
32. **Tabor, S., and C. C. Richardson.** 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. Proc. Natl. Acad. Sci. USA **82**: 1074–1078.
33. **Takahashi, Y., K. Kato, Y. Hayashizaki, T. Wakabayashi, E. Ohtsuka, S. Matsuki, M. Ikehara, and K. Matsubara.** 1985. Molecular cloning of the human cholecystokinin gene by use of a synthetic probe containing deoxyinosine. Proc. Natl. Acad. Sci. USA **82**:1931–1935.
34. **Vachon, V., R. Laprade, and J. W. Coulton.** 1986. Properties of the porin of *Haemophilus influenzae* type b in planar lipid bilayer membranes. Biochim. Biophys. Acta **861**:74–82.
35. **Vachon, V., D. J. Lyew, and J. W. Coulton.** 1985. Transmembrane permeability channels across the outer membrane of *Haemophilus influenzae* type b. J. Bacteriol. **162**:918–924.
36. **Wu, H. C.** 1986. Proteolytic processing of signal peptides, p. 33–59. *In* A. W. Strauss, I. Boime, and G. Kreil (ed.), Protein compartmentalization. Springer-Verlag, New York.
37. **Zoller, M. J., and M. Smith.** 1984. Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template. DNA **3**:479–488.