

Comparative Analysis of the Structures of the Outer Membrane Protein P1 Genes from Major Clones of *Haemophilus influenzae* Type b

ROBERT MUNSON, JR.,^{1,2,3*} SUSAN GRASS,^{1,3} MENACHEM EINHORN,^{1,3} CAMELLA BAILEY,^{1,3}
AND CHRISTOPHER NEWELL^{1,3}

*Edward Mallinckrodt Department of Pediatrics*¹ and *Department of Molecular Microbiology,*² *Washington University School of Medicine, and Division of Infectious Diseases, Saint Louis Children's Hospital,*³ *St. Louis, Missouri 63110*

Received 31 May 1989/Accepted 18 July 1989

P1 outer membrane proteins from *Haemophilus influenzae* type b are heterogeneous antigenically and with respect to apparent molecular weight in sodium dodecyl sulfate-polyacrylamide gel electrophoresis. For determination of the molecular basis for the differences in the P1 proteins, the genes for the P1 proteins from strain 1613, representative of outer membrane protein subtype 3L, and strain 8358, representative of outer membrane protein subtype 6U, were cloned, sequenced, and compared with the previously reported gene for the P1 protein from strain MinnA, a strain with the outer membrane protein subtype 1H. These prototype strains are representatives of the three major clonal families of *H. influenzae* type b responsible for invasive disease in diverse areas of the world. The nucleotide sequences of the P1 genes from strains 1613 and 8358 were 94 and 90% identical to the MinnA sequence, respectively. The derived amino acid sequences were 91 and 86% identical, respectively. Heterogeneity between the MinnA and 1613 proteins was largely localized to two short variable regions; the protein from strain 8358 contained a third variable region not observed in the other P1 proteins. Thus, the outer membrane protein P1 genes are highly conserved; the variable regions may code for the previously demonstrated strain-specific antigenic determinants.

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 (8). Outer membrane protein (OMP) P1 (designated "a" by Loeb and Smith [12]) has been purified, and antibody directed against P1 has been shown to have protective activity in an infant rat bacteremic model (6, 11). Type b isolates contain P1 proteins of variable apparent molecular weight; differences in the mobilities of these proteins were used by Barenkamp et al. and Granoff et al. to subtype *H. influenzae* type b isolates (2, 3, 7). In the subtyping scheme, these proteins were designated H, L, or U. Strains with the OMP subtype designated 1H have the H form of the P1 protein and represent 42% of the isolates causing invasive disease in the United States (2, 18). Isolates with the common OMP subtype, 3L, are responsible for approximately 18% of the invasive disease in the United States and are the predominant isolates found in Europe (18, 26). These isolates have a P1 protein of slightly lower apparent molecular weight, designated L in the subtyping scheme. Originally, isolates having neither the H nor the L form of the P1 protein were designated U, for unclassified (7). However, in subsequent studies, most isolates with the P1 protein designated U were shown to be members of the OMP subtype 6U/12U group, a clonal family commonly observed in the developing world but rarely observed in the United States (18, 27). The P1 protein designated U is of lower apparent molecular weight than either the H or the L P1 protein.

The OMP P1 gene has been cloned (6, 16), and the gene for the P1 protein from an OMP subtype 1H isolate has been sequenced (16). We now report the molecular cloning, expression, DNA sequence, and derived amino acid se-

quence of the P1 proteins from prototype OMP subtype 3L and 6U isolates.

MATERIALS AND METHODS

Bacterial strains, plasmids, and phages. *H. influenzae* type b strain 1613 was isolated in St. Louis from the blood of a patient with periorbital cellulitis (17). It has the OMP subtype 3L. *H. influenzae* type b strain 8358 was a gift from Janet Montgomery. It was isolated in Papua New Guinea from the blood of a child and has the OMP subtype 6U. Selected isolates were examined by Southern hybridization (see below). Ten isolates with the OMP subtype 1H were examined; these isolates were from five U.S. states and Iceland. Eight OMP subtype 3L isolates were examined; they were isolated from four U.S. states, Washington, D.C., Papua New Guinea, and Iceland. The Washington, D.C., isolate was isolated in 1944 and is part of the Pittman collection (1). The five OMP subtype 6U/12U isolates examined were from two U.S. states as well as Papua New Guinea and Pakistan. The OMP subtype 13L isolate was from Missouri. Strain Rab, originally from the collection of Hattie Alexander, was isolated in 1938 and has the rare OMP subtype 17H (23). These *H. influenzae* type b isolates were obtained from the collection of Dan Granoff. *Escherichia coli* JM101 was obtained from New England BioLabs, Inc. (Beverly, Mass.), and *E. coli* LE392 and LE392 lysogenized with phage P2 were obtained from Stratagene (La Jolla, Calif.). Plasmid pGD103, a low-copy-number plasmid vector containing the multiple cloning site and α -complementing portion of the *lacZ* gene from pUC8 and the pSC101 origin of replication, was a kind gift from Bruce Green (5). M13mp18 and M13mp19 were obtained from New England BioLabs.

H. influenzae type b was grown in supplemented brain heart infusion medium as previously described (3). *E. coli* strains were grown in L medium supplemented with 35 μ g of

* Corresponding author.

kanamycin or 50 μ g of ampicillin (Sigma Chemical Co., St. Louis, Mo.) per ml when appropriate.

Molecular cloning. Chromosomal DNA was prepared from strain 1613 by a modification of the method of Marmur (15). A λ EMBL3 genomic library was prepared from *Sau3A* partial digests of chromosomal DNA and screened immunologically as described previously (16). Three immunologically reactive phages were isolated and characterized. Southern hybridization was used to localize the P1 gene in the *H. influenzae* insert. Hybridization was performed on nitrocellulose (Bio-Rad Laboratories, Richmond, Calif.) as described by Maniatis et al. (14). Twenty-mer oligonucleotides corresponding to the 5' and 3' ends of the previously reported gene from strain Minna were synthesized on a model 380B Applied Biosystems DNA synthesizer. The oligonucleotides 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 at 42°C overnight. Blots were washed with 2 \times SSC (1 \times SSC is 0.15 M NaCl plus 0.015 M sodium citrate) containing 0.1% sodium dodecyl sulfate at room temperature, dried, and autoradiographed. The P1 gene from the strain 1613 library was localized to a 5.3-kilobase-pair (kbp) *Bgl*II-*Eco*RI fragment, which was subcloned into pGD103, which was digested with *Bam*HI and *Eco*RI.

The P1 gene from strain 1613 was over 90% identical to the previously reported gene from strain Minna (see below); therefore, we used the polymerase chain reaction (PCR) to isolate the P1 gene from chromosomal DNA of strain 8358. Chromosomal DNA (5 μ g) was incubated with 500 ng of oligonucleotides, and the P1 gene was amplified by 30 rounds of denaturation and polymerization. Because of uncertainty in the homology between the primers and the genomic DNA, annealing was performed at 37°C and extension was performed at 50°C for 15 min for cycles 1 through 5. Annealing was performed at 40°C and extension was performed at 72°C for 10 min for cycles 6 through 30. *Taq* polymerase and other PCR reagents were purchased from Cetus (Norwalk, Conn.). Other enzymes were purchased from New England BioLabs or Bethesda Research Laboratories, Inc. (Gaithersburg, Md.), and used in accordance with manufacturer instructions. Agarose gel electrophoresis was performed on 0.7% gels with λ *Hind*III size standards (Bethesda Research Laboratories).

Restriction length polymorphisms. Chromosomal DNA from *H. influenzae* type b strains was isolated essentially as described by Silhavy et al. (24). DNA (2 μ g) was digested to completion with *Eco*RI or *Pst*I. Fragments were electrophoresed on 0.7% agarose gels and transferred to Hybond-N (Amersham Corp., Arlington Heights, Ill.) in accordance with manufacturer instructions. For generation of a probe containing the entire P1 gene but devoid of 3' sequences, pRSM188 (16) was digested with *Dra*III, blunt ended with the Klenow fragment of DNA polymerase, ligated to *Eco*RI linkers, digested with *Eco*RI, religated, and transformed into *E. coli* LE392. A single *Dra*III site is present in pRSM188 at nucleotide 1472 of the P1 sequence (see Fig. 3). Thus, the resulting plasmid contained the complete P1 gene, but the ~1,200 base pairs (bp) of *H. influenzae* DNA 3' to the P1 gene in pRSM188 were deleted. The *Pst*I-*Eco*RI fragment containing the P1 gene was isolated, purified from vector sequences by preparative agarose electrophoresis, and nick translated with [α - 32 P]CTP (New England Nuclear Corp.) by using a kit from Boehringer Mannheim Biochemicals, Indianapolis, Ind.) in accordance with manufacturer instructions. Hybridization was performed in accordance with

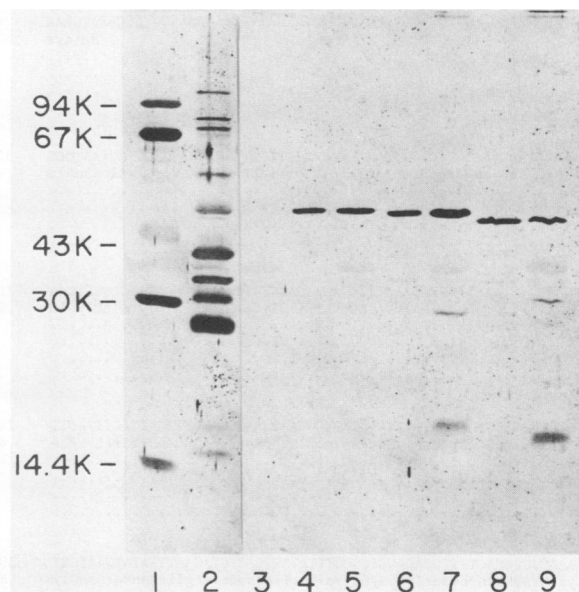


FIG. 1. Western blot of sarcosyl-insoluble fractions demonstrating the synthesis of the recombinant P1 proteins. Lane 1 contains molecular weight standards, and lane 2 contains 10 μ g of a sarcosyl-insoluble fraction from strain Minna. These lanes were stained with India ink (10) to visualize all proteins. Lanes 3 through 9 contain sarcosyl-insoluble protein and were developed immunologically with rabbit anti-P1 antibody. Lane 3, 10 μ g of *E. coli* LE392(pBR322); lane 4, 5 μ g of *H. influenzae* Minna (OMP subtype 1H); lane 5, 10 μ g of LE392(pRSM188); lane 6, 5 μ g of *H. influenzae* 1613 (OMP subtype 3L); lane 7, 20 μ g of JM101(pRSM206); lane 8, 5 μ g of *H. influenzae* 8358 (OMP subtype 6U); lane 9, 20 μ g of JM101(pRSM652). All samples were solubilized at 100°C for 5 min.

Amersham Corp. instructions. After hybridization, the blots were washed twice with 2 \times SSC and once with 2 \times SSC containing 0.1% sodium dodecyl sulfate at 65°C and autoradiographed.

DNA sequence analysis. M13 subcloning and dideoxy sequencing were performed with New England BioLabs re-

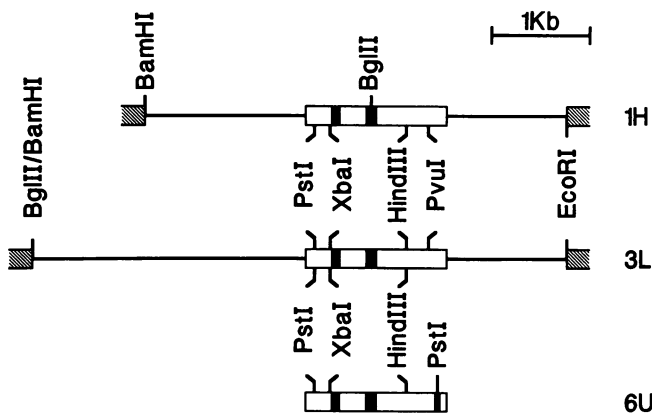


FIG. 2. Comparative restriction analysis of the 1H, 3L, and 6U P1 genes and flanking regions. Symbols: □, constant coding regions of the P1 genes; ■, variable coding regions of the P1 genes; ▨, vector sequences. The OMP subtype 1H P1 gene was cloned into *Bam*HI-*Eco*RI-digested pBR322. The OMP subtype 3L P1 gene was cloned into *Bgl*II-*Eco*RI-digested pGD103. The OMP subtype 6U P1 gene was cloned after amplification of genomic DNA with the PCR. Relevant restriction sites are indicated. Kb, Kilobase pairs.

1H	ATAGATTCGGGGTTTATAATTGCCAGATTTTATTATAACAAAGGGTTCCAAATGAAA	60	1H	CGCCAGTGAATGATAAATTCGGCTGGGTGCTGGAATGAATTCGGTCTAAAA	480
3L	MetLys	-21	3L	AlaProValAsnAspLysPheAlaLeuGlyAlaGlyMetAsnValAsnPheGlyLeuLys	120
6U		6U	..T....A.....C.A.....A.....	
1H	AAATTTAATCAATCTCTATTAGCAACTGCAATGTTGGCTGAGGTGTCAAATGCG	120	1H	AGTGAATATGACGATAGTTATGATGCTGGTGTATTGGTGGAAAACTGACTTGACTGCT	540
3L	LysPheAsnGlnSerLeuLeuAlaThrAlaMetLeuLeuAlaAlaGlyGlyAlaAsnAla	-1	3L	SerGluTyrAspAspSerTyrAspAlaGlyValPheGlyGlyLysThrAspLeuSerAla	140
6UA.....G.....		6UA.....G.....C.....	
1H	GCAGCGTTTCAATTGGCGGAAGTTTCTACTTCAGGCTTGGTCTGCCTATGCGGGTAA	180	1H	ATCAACTTAAATTTAAGTGGTCTTATCGAGTAACAGAAAGTTGAGCCTAGGTTTAGGG	600
3L	AlaAlaPheGlnLeuAlaGluValSerThrSerGlyLeuGlyAlaTyrAlaGlyGlu	20	3L	IleAsnLeuAsnLeuSerGlyAlaTyrArgValThrGluLeuSerLeuGlyLeuGly	160
6UA.....		6UC.....	
1H	GCGCGGATTGCAGATAATGCTCTCTGCTGGCACTAACCCAGCTTTGATGAGTTTATT	240	1H	GTAATCGGGTTTATGCTAAAGCCCAAGTTGAACGGAATCTGCTCTTATTGGGGATAGT	660
3L	AlaAlaIleAlaAspAsnAlaSerValValAlaThrAsnProAlaLeuMetSerLeuPhe	40	3L	ValAsnAlaValTyrAlaLysAlaGlnValGluArgAsnAlaGlyLeuIleAlaAspSer	180
6UA.G.....		6UA.....A.....	
1H	AAAACGGCAGACGTTTCCACAGTGGCGTTTATATTGATTCTAGAAATTAATGAATGGT	300	1H	GTTAAGGATAACCAAATAACAACGGCACTCTCAACAGCAGAAACCAATTCAGAGATCTT	720
3L	LysThrAlaGlnPheSerThrGlyGlyValTyrIleAspSerArgIleAsnMetAsnGly	60	3L	ValLysAspAsnGlnIleThrSerAlaLeuSerThrGlnGlnGluProPheArgAspLeu	200
6UG.....Val.....		6UT....CA..G..A..CT.....GT..T.GC.CC...C..A..GG... ...Asn...Thr...ValLysThr.....ValLeuAlaPro...LeuLysGly...AT.GCA... ..A.....AA... ..CT..TT...GGA.C.CCAA...IleAla... ..Asn.....Lys... ..ValVal...GlyThrProIle	
1H	GATGTA ^{ACTTCTTATGCTCAGATAATAACAATCAGATTGGAATGAAAGCAATAAAGGAC}	360	1H	AAGAAGTATTGCCCTCTAAGGAC ^{AAATCTGTCTCTCATTACAAGATAGAGCCGCTGG}	780
3L	AspValThrSerTyrAlaGlnIleIleThrAsnGlnIleGlyMetLysAlaIleLysAsp	80	3L	LysLysTyrLeuProSerLysAspLysSerValValSerLeuGlnAspArgAlaAlaTrp	220
6UG...C.....T.GCAACT.C.AA.....CT..GC...T... ...AlaSer.....IleAlaThrThrLys...AsnSerAla...TyrG...G...C.....T...AGGTGC.AC.....CA...C.....AlaAlaSer.....IleLysGlyAlaThr...AsnThrThr.....		6U	G.TC..A.....A.....A.....A..... AspGlnAsn..... CCTG.C...A.....AA...G.....A..... ProAsp.....Thr.....Asn.....	
1H	GGCTCAGCTTCACAGCGTAATGTTGTTCCCGGTGCTTTTGCCAAATGTTATTTCGTT	420	1H	GGCTTGGCTGCAATGACAGGTGTAATGATCAATTTAATGAAGCTAACAGAAATGGTTTA	840
3L	GlySerAlaSerGlnArgAsnValValProGlyAlaPheValProAsnLeuTyrPheVal	100	3L	GlyPheGlyTrpAsnAlaGlyValMetTyrGlnPheAsnGluAlaAsnArgIleGlyLeu	240
6UG.....T.....		6UG.....C.....	

FIG. 3. Complete nucleotide and amino acid sequences of the 3L and 6U P1 genes compared with those of the previously reported 1H P1 gene. The variable regions are boxed.

agents or Sequenase (U.S. Biochemical Corp., Cleveland, Ohio) in accordance with manufacturer instructions. [³⁵S]dATP was purchased from New England Nuclear Corp. Data were analyzed with Compugene software (4) on a Digital VAX 8530 computer. We used an M13 universal primer as well as oligonucleotide primers, which were synthesized on a model 380B Applied Biosystems DNA synthesizer. The nucleotide sequence data reported in this paper will appear in the EMBL, GenBank, and DDBJ nucleotide sequence data bases. The accession numbers are J03359, M27682, and M27683 for the P1 genes from strains MinnA, 1613, and 8358, respectively.

Immunologic methods. Sarcosyl-insoluble preparations enriched in OMPs were prepared as described previously (3). Protein concentrations were determined by the bicinchoninic acid method (25) (BCA protein assay kit; Pierce Chemical Co., Rockford, Ill.) in accordance with manufacturer instructions. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis was performed on modified Laemmli 11% gels as described by Lugtenberg et al. (13). Western blot (immunoblot) analysis was performed with rabbit anti-P1 antisera prepared against the protein isolated from strain MinnA as described previously (16).

RESULTS

Rabbit antisera prepared by immunization with the MinnA P1 protein were used to immunologically screen the λ

EMBL3 genomic library from strain 1613, the prototype OMP subtype 3L isolate. Three reactive phages were isolated and characterized. Phage lysates were prepared in LE392 and examined by Western blot analysis. Each produced the full-size *H. influenzae* P1 protein. The restriction map of each phage was unique; however, each contained an *EcoRI* fragment of approximately 8.8 kbp. Hybridization analysis confirmed that the P1 gene was localized to this *EcoRI* fragment. For further localization of the gene, the 8.8-kbp *EcoRI* fragment was cut with a number of restriction enzymes and analyzed by hybridization with 20-mer oligonucleotides generated from the sequence of the 5' and 3' portions of the 1H P1 gene. A 5.3-kbp *BglII-EcoRI* fragment hybridized to both probes. This fragment was subcloned into pGD103; the recombinant was designated pRSM206.

Full-size 3L protein was detectable in sarcosyl-insoluble preparations of *E. coli* JM101(pRSM206) (Fig. 1). The protein from strain 1613 and the native and recombinant OMP subtype 1H proteins are shown for comparison. The recombinant protein was localized in the sarcosyl-insoluble fraction. Antibody labeling experiments performed as described previously (16) demonstrated that epitopes of the 3L protein were present at the surface of intact *E. coli* cells, indicating that the protein was processed normally and was localized to the outer membrane. Small amounts of two lower-molecular-weight degradation products, which have not been further characterized, were also observed (Fig. 1).

1H	GCCTATCATTCTAAAGTGGACATTGATTTTGGCTGACCGCACTGCTACTAGTTTAGAAGCA	900
3L	AlaTyrHisSerLysValAspIleAspPheAlaAspArgThrAlaThrSerLeuGluAla	260
6U	..T.....A.....C.....G.....	
1H	AATGTCATCAAAGAAGTAAAAAGGTAATTTAACCTTTACATTGCCAGATTACTTAGAA	960
3L	AsnValIleLysGluGlyLysLysGlyAsnLeuThrPheThrLeuProAspTyrLeuGlu	280
6UC.....	
1H	CTTTCTGGTTTCCATCAATTAACGTGACAACTGGCAGTGCATTATAGTTATAAATATACC	1020
3L	LeuSerGlyPheHisGlnLeuThrAspLysLeuAlaValHisTyrSerTyrLysTyrThr	300
6UG.....T.....Phe.....	
1H	CATTGGAGTGGTTTAAACAAAATTACATGCCAGCTTCGAAGATGGTAAAAAGCTTTTGAT	1080
3L	HisTrpSerArgLeuThrLysLeuHisAlaSerPheGluAspGlyLysLysAlaPheAsp	320
6UC..T.....G.....	
1H	AAAGAATTACAATACAGTAATAACTCTCGTGGCATTAGGGGCAAGTTATAATCTTTAT	1140
3L	LysGluLeuGlnTyrSerAsnAsnSerArgValAlaLeuGlyAlaSerTyrAsnLeuTyr	340
6UG.....A.....G.....	
1H	GAAAAATTGACCTTACGTGGCGGTATTGCTTACGATCAAGCGGCATCTCGTCATCACCGT	1200
3L	GluLysLeuThrLeuArgAlaGlyIleAlaTyrAspGlnAlaAlaSerArgHisHisArg	360
6UA.....T.....	
1H	AGTGTGCAATTCAGATACCGATCCGACTGGTATAGTTTAGGTGCAACCTATAAATTC	1260
3L	SerAlaAlaIleProAspThrAspArgThrTrpTyrSerLeuGlyAlaThrTyrLysPhe	380
6UA.....G.....	
1H	ACGCCGAATTTATCTGTGATCTTGGCTATGCTTACTTAAAAGGCAAAAAAGTCTCACTT	1320
3L	ThrProAsnLeuSerValAspLeuGlyTyrAlaTyrLeuLysGlyLysLysValHisPhe	400
6U	
1H	AAAGAACTAAAAACAATAGGTGACAAACGTACATTGACATTGAATACAACTGCAAATTAT	1380
3L	LysGluValLysThrIleGlyAspLysArgThrLeuThrLeuAsnThrThrAlaAsnTyr	420
6UC.C..CA.GCT.CA.GTGGCTTC.....A.A.CA.....C.....C	
1H	ACTTCTCAAGCAGCAAGCAATCTTACGGTTGAATTTAAATTATAGTTTCTAATCCGTT	1440
3L	ThrSerGlnAlaHisAlaAsnLeuTyrGlyLeuAsnLeuAsnTyrSerPhe	437
6UC..A..C.....	
1H	AAAAAATTAGCATAATAAAGCACAAATCCACACTAAGTGTGCTTTTCTTTATAAAACA	1500
1H	AGGCGAAAAATGACCGCACTTTATTACACTTATTACCCCTCGCCAGTCGGACGGCTTTTG	1560

FIG. 3—Continued

sequences of the P1 gene from the 6U strain were also highly conserved we could clone the 6U P1 gene directly after amplification of the gene from chromosomal DNA with the PCR.

A 20-mer sequencing primer immediately 5' to the reported sequence (16) and another 20-mer sequencing primer complementary to nucleotides 1525 to 1562 (3' to the Minna gene) were used. After 30 rounds of amplification of genomic DNA from strain Minna, the PCR product was analyzed by agarose gel electrophoresis. A predominant band of the predicted size (approximately 1,600 bp) was observed. When genomic DNA from strain 8358, our prototype OMP subtype 6U isolate, was amplified in the same manner, the predominant product was approximately 3,000 bp. Restriction analysis indicated that this PCR product contained a *Pst*I fragment of approximately 1,200 bp. The *Pst*I fragment was cloned into M13 and partially sequenced. Sequence data indicated that it was the P1 gene and that the downstream *Pst*I site was near the 3' end of the gene (Fig. 2). The *Hind*III site observed in the 1H and 3L P1 genes had been conserved. Oligonucleotides corresponding to the 5'- and 3'-flanking sequences were generated with a *Bam*HI site (5' oligonucleotide) and an *Eco*RI site (complementary 3' oligonucleotide). A 3,000-bp PCR product was again generated, and the *Bam*HI-*Pst*I, *Pst*I-*Hind*III, and *Hind*III-*Eco*RI fragments were cloned into M13. The sequences of both strands of the 6U P1 gene were determined, beginning with the 5' *Pst*I site. The sequence of the leader peptide was determined in one direction only from the *Bam*HI-*Pst*I fragment. The comparative restriction maps of the P1 genes are shown in Fig. 2.

For expression of the 6U P1 gene, an independent PCR was performed, and the *Xba*I-*Eco*RI fragment was prepared and purified. The *Xba*I site was present at base 280 (Fig. 2 and 3). The *Eco*RI site 3' to the gene was part of the PCR primer (see above). As the 5' portions of the 3L and 6U P1 genes code for the identical peptide, we cloned the *Xba*I-*Eco*RI-digested PCR product into *Xba*I-*Eco*RI-treated pRSM206. The resulting plasmid was designated pRSM652; the full-size OMP subtype 6U P1 gene product was produced by this construct (Fig. 1). The sequence of the *Xba*I-*Eco*RI fragment was independently determined and agreed with the sequence determined from the M13 derivatives.

The nucleotide sequences of the three genes were highly conserved. Overall, 94% of the base pairs coding for the 1H and 3L proteins were identical. The 6U P1 gene was 90% identical to the 1H P1 gene and 92% identical to the 3L P1 gene. Sequence divergence of the 1H and 3L P1 genes was largely localized to two regions (nucleotides 307 to 360 and nucleotides 646 to 744; Fig. 2 and 3). The 6U P1 gene was also divergent in these two regions and had a third region of divergence (nucleotides 1327 to 1365). The 3L and 6U proteins were 91 and 86% identical to the 1H protein, respectively. The 3L and 6U protein sequences were 88% identical to each other.

We previously demonstrated that the N-terminal sequence of the purified 1H protein was AAFQLAEVSTSGLG (16). The 3L and 6U proteins contained this sequence. The 1H protein contained a 22-amino-acid signal peptide. The 3L and 6U proteins had the same signal peptide as the 1H protein, with the exception of a substitution of Ile for Leu at position -15. Assuming that these proteins are processed identically to the 1H protein, the M_r of the 3L protein is 46,937 and the M_r of the 6U protein is 46,111, as compared with an M_r of 47,752 for the 1H protein.

The P1 gene from strain Minna was present on a genomic *Eco*RI fragment of approximately 6 kbp (16; Fig. 4). We used

DNA sequence analysis was performed by the dideoxy method after cloning of the *Pst*I-*Hind*III and *Hind*III-*Eco*RI restriction fragments into M13mp18 and M13mp19 (Fig. 2). M13 primers as well as 20-mer oligonucleotide primers were used to complete the sequences of both strands. The sequence 5' to the *Pst*I site was sequenced directly from pRSM206, in one direction only.

The DNA sequences of the 1H and 3L P1 genes and their flanking sequences were greater than 90% identical (see below). We reasoned that if the upstream and downstream

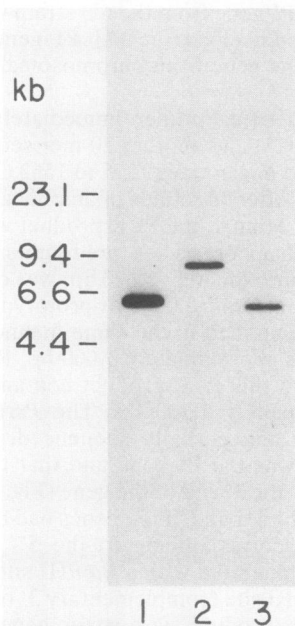


FIG. 4. Southern hybridization analysis of *H. influenzae* type b chromosomal DNA. Chromosomal DNA was digested with *EcoRI*, separated on a 0.7% agarose gel, transferred to Hybond-N, and probed with the nick-translated P1 gene. Lanes: 1, strain MinnA (OMP subtype 1H); 2, strain 1613 (OMP subtype 3L); lane 3, strain 8358 (OMP subtype 6U). The positions of λ *HindIII*-digested DNA (Bethesda Research Laboratories) size markers are shown in kilobase pairs (kb).

the 1H P1 gene as a hybridization probe to assess the restriction length polymorphisms of the *EcoRI* fragment containing the P1 gene from 25 *H. influenzae* type b isolates from various global locations. Genomic DNA from 10 of 10 isolates with the OMP subtype 1H contained a 6-kbp *EcoRI* genomic fragment which hybridized to the probe. The P1 gene was localized to an 8.8-kbp *EcoRI* fragment in eight of eight OMP subtype 3L isolates examined. Isolates with the OMP subtype 13L appear to be closely related to OMP subtype 3L isolates but were more frequently associated with nasopharyngeal colonization rather than invasive disease (9). Genomic DNA from one subtype 13L isolate was examined; the *EcoRI* genomic fragment hybridizing to the P1 probe was 8.8 kbp. Strain Rab has the rare OMP subtype 17H. A 5-kbp *EcoRI* genomic fragment from this strain hybridized to the probe. Lastly, five isolates in the OMP subtype 6U/12U group were examined; the P1 gene was localized to a 5.7-kbp *EcoRI* genomic fragment. The DNAs from the OMP subtype 6U/12U isolates were also digested with *PstI* and probed. All five isolates hybridized to a *PstI* genomic fragment of 1,200 bp, indicating that these isolates all contained a P1 gene with a *PstI* site near the 3' end of the gene.

As indicated above, genomic DNA from the OMP subtype 6U isolate contained 1,200 bp of DNA immediately 5' to the P1 gene which was not observed 5' to the MinnA P1 gene. To determine whether this DNA was present in other isolates, we probed *EcoRI*-digested chromosomal DNA from the 25 isolates with the *BamHI-PstI* fragment cloned from the PCR product. The probe bound to a single restriction fragment in four of the five OMP subtype 6U/12U isolates as well as to a single fragment in strain Rab. DNA from the other 20 isolates did not hybridize to this fragment. The significance

of this 1,200-bp fragment remains to be determined; our data indicate, however, that it is not widely distributed among the major type b clone families and is not a multicopy transposable genetic element.

DISCUSSION

Alloenzyme electrophoretic typing studies (19–21) indicate that the natural population of disease-producing *H. influenzae* type b isolates is clonal in nature. Five lineages of type b isolates have been identified. Studies done with restriction length polymorphisms or OMP profiles are consistent with this proposal (20–22). The major electrophoretic typing clonal family of type b isolates seen in the United States is group A1. Isolates with the OMP subtype 1H belong to this clonal family and account for 42% of isolates causing invasive disease in the United States. We have previously reported the molecular cloning and primary sequence of the P1 protein from a strain with the OMP subtype 1H (16).

H. influenzae type b isolates with the OMP subtype 3L are the most common organisms in the electrophoretic typing clonal family A2. OMP subtype 3L isolates are the predominant organisms observed in Europe and cause 18% of the invasive disease in the United States. In this study, the P1 gene from a prototype 3L isolate was cloned and expressed in *E. coli*. The *PstI*, *XbaI*, *HindIII*, and *PvuI* restriction sites in the 1H and 3L P1 genes were conserved, as was the *EcoRI* site 3' to the gene (Fig. 2). However, the *BglII* site in the 1H P1 gene was not seen in the 3L P1 gene. The *BamHI* site 5' to the 1H P1 gene was also not present 5' to the 3L P1 gene.

H. influenzae type b isolates with OMP subtype 6U (and the closely related subtype, 12U), belong to electrophoretic typing clonal family B1. These isolates are a common cause of invasive disease in several areas of the developing world but have rarely been observed in the United States. The P1 gene from a prototype OMP subtype 6U isolate was cloned after amplification of the structural gene with the PCR. The restriction map of the 6U P1 gene was notable for the *PstI* site near the 3' end of the gene and the loss of the *PvuI* site near the 3' end of the 1H and 3L P1 genes.

The three P1 genes are highly conserved. Heterogeneity of the gene and derived protein sequences is largely localized to three variable regions. Outside of these variable regions, the majority of nucleotide changes are silent. When the conserved regions of the 1H and 3L amino acid sequences are compared, there is a single amino acid change in the signal sequence and nine other amino acid changes scattered throughout the protein. Similarly, when the conserved regions of the 1H and 6U proteins are compared, a single amino acid change is observed in the signal peptide and 16 other changes are scattered throughout the protein. Seven of these 16 amino acid differences are common to the 3L and 6U proteins.

Southern hybridization was performed to determine if more than one copy of the P1 gene was present in an isolate and, further, to determine if restriction polymorphisms could be identified. Each isolate contained a single *EcoRI* genomic fragment which hybridized to the P1 probe, consistent with the production of a single P1 protein by each isolate. Each group of isolates, as defined by OMP subtype, had a unique *EcoRI* genomic fragment which contained the P1 gene.

In summary, the OMP P1 gene has been analyzed from prototype strains representative of three major clonal families of organisms responsible for invasive disease. Each strain produces a single P1 protein. The P1 genes are highly

conserved; three variable regions account for the majority of the sequence heterogeneity. Studies are in progress to epitope map these proteins and determine if the previously described strain-specific surface-exposed antigenic determinants of OMP P1 are localized to the variable regions.

ACKNOWLEDGMENTS

We thank Dan Granoff for stimulating discussions and continued enthusiastic support. We also thank Kathleen Lottenbach for preparation of the genomic DNAs used in the RFLP analysis.

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

LITERATURE CITED

- Barenkamp, S. J., D. M. Granoff, and M. Pittman. 1983. Outer membrane protein subtypes and biotypes of *Haemophilus influenzae* type b: relation between strains isolated in 1934-1954 and 1977-1980. *J. Infect. Dis.* **148**:1127.
- 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.
- 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.
- Barnes, W. M. 1987. Sequencing DNA with dideoxyribonucleotides as chain terminators: hints and strategies for big projects. *Methods Enzymol.* **152**:538-556.
- 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.
- Gonzales, F. R., S. Leachman, M. V. Norgard, J. D. Radolf, G. H. McCracken, Jr., C. Evans, and E. J. Hansen. 1987. Cloning and expression in *Escherichia coli* of the gene encoding the heat-modifiable major outer membrane protein of *Haemophilus influenzae* type b. *Infect. Immun.* **55**:2993-3000.
- Granoff, D. M., S. J. Barenkamp, and R. S. Munson, Jr. 1982. Outer membrane protein subtypes for epidemiologic investigation of *Haemophilus influenzae* type b disease, p. 43-55. In S. H. Sell and P. F. Wright (ed.), *Haemophilus influenzae: epidemiology, immunology, and prevention of disease*. Elsevier Science Publishing, Inc., New York.
- 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.
- Hampton, C. M., S. J. Barenkamp, and D. M. Granoff. 1983. Comparison of outer membrane protein subtypes of *Haemophilus influenzae* type b isolates from healthy children in the general population and from diseased patients. *J. Clin. Microbiol.* **18**:596-600.
- Hancock, K., and V. C. W. Tsang. 1983. India ink staining of proteins on nitrocellulose paper. *Anal. Biochem.* **133**:157-162.
- Loeb, M. R. 1987. Protection of infant rats from *Haemophilus influenzae* type b infection by antiserum to purified outer membrane protein a. *Infect. Immun.* **55**:2612-2618.
- 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.
- 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.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Marmur, J. 1961. A procedure for the isolation of deoxyribonucleic acid from micro-organisms. *J. Mol. Biol.* **3**:208-218.
- 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.
- Munson, R., Jr., and A. Hunt. 1989. Isolation and characterization of a mutant of *Haemophilus influenzae* type b deficient in outer membrane protein P1. *Infect. Immun.* **57**:1002-1004.
- Munson, R. S., Jr., M. H. Kabeer, A. A. Lenoir, and D. M. Granoff. 1989. Epidemiology and prospects for prevention of disease due to *Haemophilus influenzae* in developing countries. *Rev. Infect. Dis.* **11**(Suppl. 3):S588-S597.
- 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.
- Musser, J. M., J. S. Kroll, E. R. Moxon, and R. K. Selander. 1988. Clonal population structure of encapsulated *Haemophilus influenzae*. *Infect. Immun.* **56**:1837-1845.
- Musser, J. M., J. S. Kroll, E. R. Moxon, and R. K. Selander. 1988. Evolutionary genetics of the encapsulated strains of *Haemophilus influenzae*. *Proc. Natl. Acad. Sci. USA* **85**:7758-7762.
- Poulsen, K., J. P. Hjorth, and M. Kilian. 1988. Limited diversity of the immunoglobulin A1 protease gene (*iga*) among *Haemophilus influenzae* serotype b strains. *Infect. Immun.* **56**:987-992.
- Rodrigues, L. P., R. Schneerson, and J. B. Robbins. 1971. Immunity to *Haemophilus influenzae* type b. *J. Immunol.* **107**:1071-1080.
- Silhavy, T. J., M. L. Berman, and L. W. Enquist. 1984. Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- 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.
- van Alphen, L., L. Geelen, K. Jonsdottir, A. K. Takala, H. Kayhty, and H. C. Zanen. 1987. Distinct geographic distribution of subtypes of *Haemophilus influenzae* type b in western Europe. *J. Infect. Dis.* **156**:216-218.
- Weinberg, G. A., A. Ghaffoor, Z. Ishaq, N. K. Nomani, M. Kabeer, F. Anwar, M. I. Burney, A. W. Qureshi, J. M. Musser, R. K. Selander, and D. M. Granoff. 1989. Clonal analysis of *Haemophilus influenzae* isolated from children from Pakistan with lower respiratory infections. *J. Infect. Dis.* **160**:634-643.