INFECTION AND IMMUNITY, Apr. 1989, p. 1100–1107
0019-9567/89/041100-08$02.00/0

Vol. 57, No. 4

Primary Structure of the Porin Protein of *Haemophilus influenzae* Type b Determined by Nucleotide Sequence Analysis

ERIC J. HANSEN,[1,2]* CHARLES HASEMANN,[1] ADRIAN CLAUSELL,[1] J. DONALD CAPRA,[1] KIM ORTH,[3] CAROLYN R. MOOMAW,[3] CLIVE A. SLAUGHTER,[3] JO L. LATIMER,[1] AND ELIZABETH E. MILLER[1]

Department of Microbiology,[1] The Howard Hughes Medical Institute,[3] and The Program in Cell and Molecular Biology,[2] The University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75235

Received 13 October 1988/Accepted 15 December 1988

Sequencing techniques for single- and double-stranded DNA were used to determine the nucleotide sequence of the gene encoding P2, the major outer membrane (porin) protein of *Haemophilus influenzae* type b (Hib). The open reading frame encoding the P2 protein comprised 361 amino acid codons. Comparison of the inferred amino acid sequence with data obtained by amino acid sequencing of the N terminus of the mature or fully processed P2 protein revealed that this protein has a signal peptide composed of 20 amino acids. N-terminal amino acid sequencing of tryptic peptides derived from purified P2 allowed direct identification of 158 of the 341 amino acids in the fully processed P2 protein; there was 100% correlation between these amino acid sequences and that inferred from the nucleotide sequence. The amino acid sequence of Hib P2 protein had 23 to 25% homology with the sequence of the OmpF porin of *Escherichia coli* and with that of the *Neisseria gonorrhoeae* porin P.IA. Codon usage in the Hib P2 gene was significantly different from that observed for a gene encoding a porin of *E. coli*. DNA hybridization studies indicated that there is a single copy of the P2 gene in the Hib chromosome. The availability of the nucleotide and amino acid sequences for the Hib P2 protein will facilitate investigation of the antigenic characteristics and structure-function relationship of this porin.

The encapsulated bacterium *Haemophilus influenzae* type b (Hib) is the most important cause of meningitis in the United States (7). Development of a vaccine against systemic Hib disease has involved primarily investigation of the type b capsular polysaccharide (12, 38). However, several surface-exposed outer membrane proteins of Hib have been shown to be targets for antibodies protective against experimental Hib disease (10, 13, 23, 24, 31, 33, 39), and evaluation of the vaccinogenic potential of these proteins has been facilitated by the recent cloning of the genes encoding some of these polypeptides (8, 10, 18, 32).

The P2 major outer membrane protein of Hib has been shown to be a porin and to induce the synthesis of serum antibodies protective against Hib disease in an animal model (33, 42, 43). We cloned the gene encoding this porin by using a plasmid shuttle vector capable of replication in both *H. influenzae* and *Escherichia coli* (18). To facilitate the eventual elucidation of the structural and functional characteristics of this porin protein, we determined the nucleotide sequence of the Hib gene encoding P2 and used these data to deduce the amino acid sequence of P2.

MATERIALS AND METHODS

Bacterial strains and culture media. Hib DL42 and *H. influenzae* Rd DB117 (rec-1) have been described previously (15, 18). Hib DL26, CH100, OA104, SL103, Madigan, DV102, and H234 are all systemic disease isolates of this pathogen. *H. influenzae* strains were grown in brain heart infusion broth (Difco Laboratories, Detroit, Mich.) supplemented with Levinthal base as described elsewhere (17). The *E. coli* strain used in this series of experiments was BSJ72, a streptomycin-resistant and λ-resistant derivative of TG1, which itself is an *hsdV5* derivative of JM101 (Δ[*lac-proAB*] *supE thi*/F′ *traD36 proAB lacI*^{q}*Z* ΔM15) (28).

Bacterial plasmids and subcloning. The recombinant plasmid pEJH39-1 containing the Hib DNA insert encoding the Hib P2 protein has been described previously (18). The phagemid vectors pTZ18U and pTZ19U were obtained as part of a DNA sequencing kit (Genescribe-Z Kit; U.S. Biochemical Corp., Cleveland, Ohio). The phagemid vector pVZ1 is a derivative of the pBluescript vector (Stratagene, La Jolla, Calif.) that contains an extra polylinker with restriction sites for additional enzymes which cut infrequently. Ligation mixtures were transformed into BSJ72 cells made competent for transformation by the method of Hanahan (16); recombinant strains were identified by blue-versus-white colony screening (47).

Southern blot analysis. Southern blot analysis of restriction enzyme digests of pEJH39-1 was accomplished as described by Maniatis et al. (28) with a ^{32}P-labeled oligonucleotide probe for the DNA region encoding the N terminus of the Hib P2 protein as described previously (18). Southern blot analysis of complete restriction enzyme digests of Hib chromosomal DNA was accomplished as described previously (18), except that the DNA probe was radiolabeled with [α-^{32}P]dCTP by the random-primers DNA-labeling system (Bethesda Research Laboratories, Inc., Gaithersburg, Md.).

Production of single-stranded DNA and dideoxy sequencing. Single-stranded DNA was produced from phagemid vectors essentially as described by Dente et al. (9). BSJ72 cells containing the desired phagemid were maintained on M9 minimal agar plates containing ampicillin (50 μg/ml) and thiamine (0.001%). Five milliliters of LB broth containing ampicillin (150 μg/ml) was inoculated with a single colony and grown to log phase (optical density at 600 nm, ~0.500). A 2-ml portion of this culture was infected with 1 μl (~10^{11} PFU/ml) of helper phage M13KO7 and grown for an additional hour. A 400-μl portion of this infected culture was then used to inoculate 10 ml of 2× YT (8 g of tryptone, 5 g of yeast extract, 5 g of NaCl per liter [pH 7.4]) containing ampicillin (150 μg/ml) and kanamycin (70 μg/ml), and the culture was grown overnight (12 to 14 h). Single-stranded

* Corresponding author.

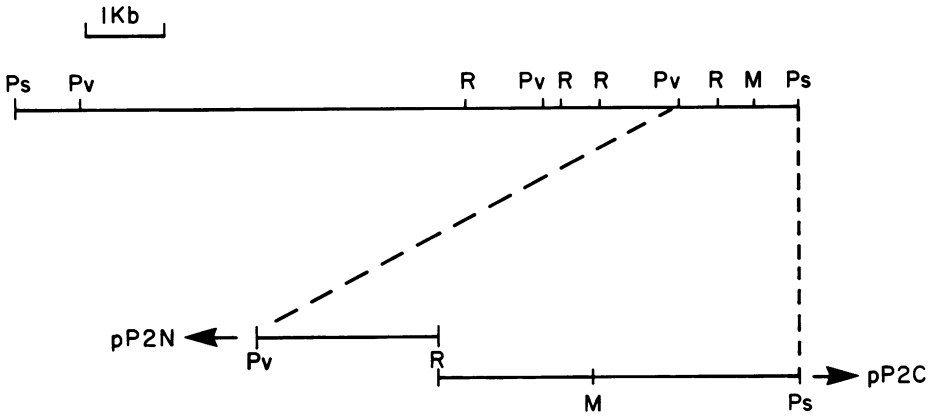

FIG. 1. Partial restriction enzyme map of the Hib DNA insert in pEJH39-1 and the plasmid subclones containing fragments of the P2 gene. The *Pvu*II-*Eco*RI fragment was subcloned into the vectors pTZ18U and pTZ19U (in both orientations) to yield the pP2N-18 and pP2N-19 subclones, respectively. The *Eco*RI-*Pst*I fragment was subcloned into the vector pVZ1 to yield pP2C. Abbreviations: Ps, *Pst*I; R, *Eco*RI; Pv, *Pvu*II; M, *Mlu*I.

DNA was isolated from culture supernatants by polyethylene glycol precipitation, treated with RNase A, and extensively extracted to remove contaminants which might interfere with subsequent sequencing reactions.

Single-stranded templates were sequenced by using reagents and protocols supplied with the Sequenase kit (U.S. Biochemical). The M13 universal primer was purchased from U.S. Biochemical. Other oligonucleotide primers were synthesized on a model 380 DNA synthesizer (Applied Biosystems, Foster City, Calif.) in the Oligonucleotide Synthesis Laboratory at this institution. Double-stranded templates were prepared from pEJH39-1 by alkaline denaturation as described by Zhang et al. (48); Sequenase was used for the dideoxy sequencing reactions.

Amino acid sequencing of the Hib P2 protein. The amino acid sequence of the N terminus of native Hib P2 protein was determined as described previously (18). In addition, proteins in outer membrane vesicles of Hib DL42 were resolved by sodium dodecyl sulfate-polyacrylamide gel electrophoresis in a 12.5% (wt/vol) polyacrylamide separating gel (18). Approximately 500 pmol of P2 protein was transferred from this gel to nitrocellulose paper for solid-phase tryptic digestion by the method of Aebersold et al. (1). Fifty percent of the digest was then separated by reverse-phase high-performance liquid chromatography on a model 130A high-performance liquid chromatograph (Applied Biosystems) with a Brownlee RP300 C_8 column (2.1 by 100 mm). Separation was performed in 0.1% trifluoroacetic acid with a 0 to 50% acetonitrile gradient over a period of 120 min at a flow rate of 50 μl/min. Peaks were collected manually onto 1-cm disks of Whatman GF/C filter paper (Whatman International, Maidstone, England). Cysteine residues were reduced and alkylated by the method of Andrews and Dixon (4). Peptides were sequenced directly on a model 470A sequencer with an on-line model 120A high-performance liquid chromatograph (both from Applied Biosystems).

Computer analyses. Analyses of nucleotide and amino acid sequences were performed with the DNASTAR sequence analysis program (DNASTAR, Inc., Madison, Wis.) and with the Microgenie sequence analysis program (Beckman Instruments, Palo Alto, Calif.).

RESULTS

Subcloning of the P2 gene. Southern blot analysis was used to localize the structural gene encoding the Hib P2 protein in the 10-kilobase (kb) Hib DNA insert in recombinant plasmid pEJH39-1 (Fig. 1). The 60-base oligonucleotide probe, which was designed originally from the N-terminal amino acid sequence of Hib DL42 P2 protein (18), was used to probe complete restriction enzyme digests of pEJH39-1 (Fig. 2). This oligonucleotide probe bound to the 10-kb Hib DNA insert obtained from a *Pst*I digest (Fig. 2, lane A), to a 1.6- to 1.7-kb band from a *Pst*I-*Eco*RI double digest (lane B), to a 1.5- to 1.6-kb fragment from a *Pst*I-*Pvu*II double digest (lane C), to a 1.6- to 1.7-kb band from an *Eco*RI digest (lane D), and to a 0.5- to 0.6-kb band from an *Eco*RI-*Pvu*II digest (lane E). Analysis of these results indicated that the N terminus of the P2 gene was located in the small (0.5- to 0.6-kb) *Pvu*II-*Eco*RI fragment located near the *Mlu*I restriction site on the Hib DNA insert.

Recombinant plasmid pEJH39-1 was digested to completion, first with *Pvu*II and then with *Eco*RI, and the DNA fragments in this doubly digested preparation were resolved by agarose gel electrophoresis. DNA fragments ranging in size from 0.4 to 0.6 kb were electroeluted from the gel and

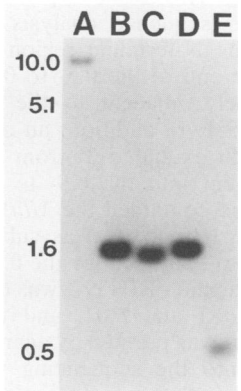

FIG. 2. Localization of the N terminus of the P2 gene in the Hib DNA insert in pEJH39-1 by Southern blot analysis. pEJH39-1 was digested to completion with the restriction enzymes listed below. The resultant DNA fragments were resolved by agarose gel electrophoresis and probed in Southern blot analysis with the ^{32}P-labeled oligonucleotide which hybridizes to the region encoding the N terminus of the P2 protein (18). Lane A, *Pst*I; lane B, *Pst*I and *Eco*RI; lane C, *Pst*I and *Pvu*II; lane D, *Eco*RI; lane E, *Eco*RI and *Pvu*II. Numbers on the left indicate molecular sizes in kilobases.

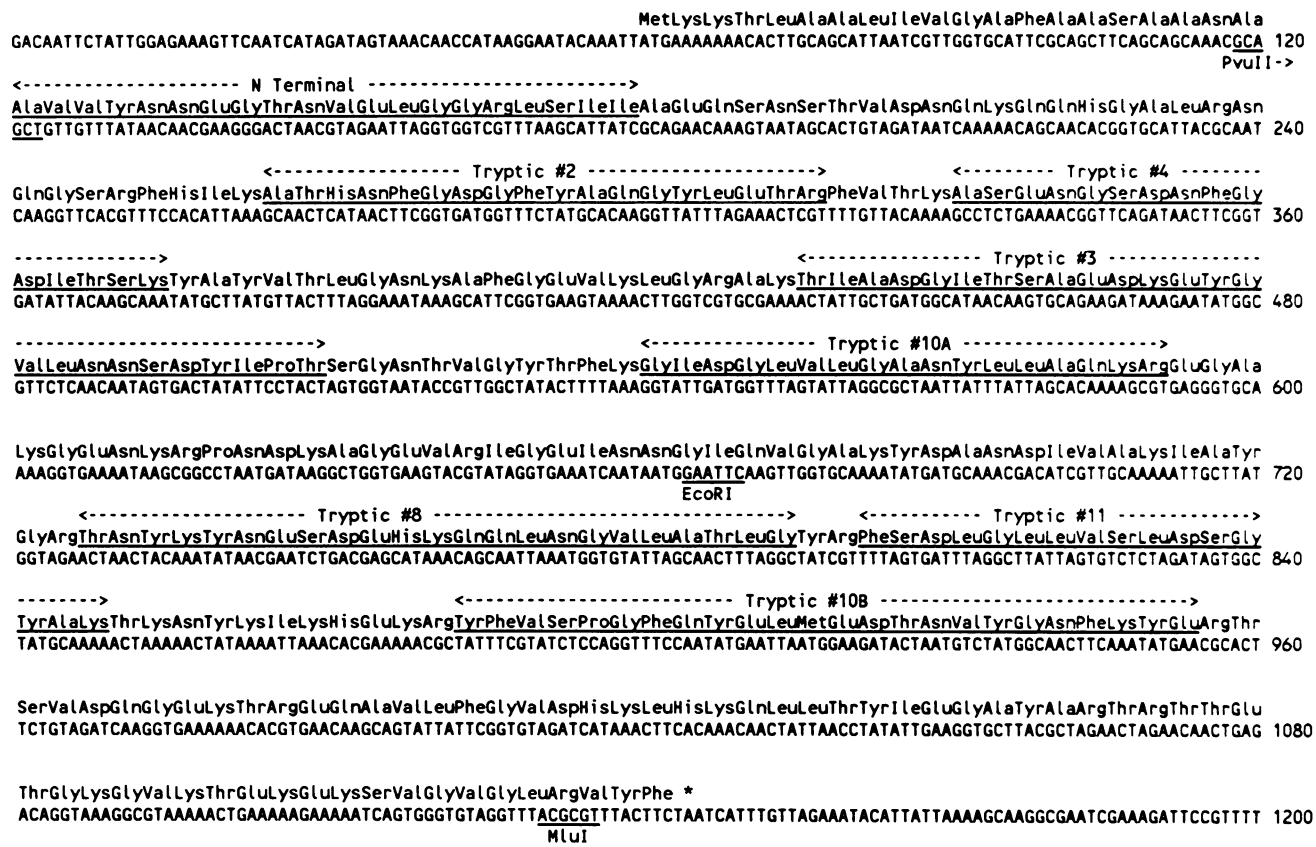

FIG. 3. DNA sequence of the P2 gene of Hib DL42. The predicted amino acid sequence is shown above the DNA sequence. The first 20 amino acids, starting with Met, represent the signal peptide; the N-terminal region of the mature or fully processed P2 protein is indicated. The underlined areas represent tryptic peptides that were analyzed by N-terminal amino acid sequencing. The restriction sites for *Pvu*II, *Eco*RI, and *Mlu*I are indicated and underlined. The asterisk indicates the position of a stop codon.

subcloned directly into the multiple cloning site in the *Eco*RI- and *Hinc*II-cleaved sequencing vectors pTZ18U and pTZ19U (representing the two different insert orientations), with *E. coli* BSJ72 as the host for transformation. These two recombinant plasmids were designated pP2N-18 and pP2N-19, respectively (Fig. 1).

Preliminary DNA sequence analysis revealed that the codons encoding the N-terminal region of the P2 protein were located at the end of the 0.5- to 0.6-kb *Pvu*II-*Eco*RI fragment immediately adjacent to the 1-kb *Eco*RI-*Pvu*II fragment in pEJH39-1. In addition, an open reading frame was identified which extended 3′ from this 0.5- to 0.6-kb *Pvu*II-*Eco*RI fragment into the 0.9- to 1.0-kb *Eco*RI-*Pst*I fragment, which also contained the *Mlu*I site (Fig. 1). This particular *Eco*RI-*Pst*I fragment was subcloned by using a strategy analogous to that used for the 0.5- to 0.6-kb *Pvu*II-*Eco*RI fragment. Plasmid pEJH39-1 was digested to completion with both *Eco*RI and *Pst*I, and the 0.9- to 1.0-kb *Eco*RI-*Pst*I fragment was purified by agarose gel electrophoresis and ligated into the sequencing vector pVZ1. The identity of this insert in the recombinant plasmid was verified by restriction digest mapping of the unique *Mlu*I site in the 0.9- to 1.0-kb *Eco*RI-*Pst*I fragment. This pVZ1-derived recombinant plasmid was designated pP2C (Fig. 1). Preliminary nucleotide sequence analysis of this *Eco*I-*Pst*I fragment suggested that it encoded all of the remaining amino acid residues and the carboxy-terminal end of the Hib P2 protein.

DNA sequence analysis. DNA sequence analysis of the pP2N and pP2C subclones, combined with N-terminal amino acid sequence analysis of tryptic peptides from P2 (see below), suggested that the 1.5-kb *Pvu*II-*Pst*I region (Fig. 1) contained the coding sequence for all of the mature P2 protein except the N-terminal amino acid (Ala). Therefore, to obtain the DNA encoding this N-terminal amino acid, the putative signal peptide, and any adjacent regulatory regions, we attempted to subclone the 1-kb *Eco*RI-*Pvu*II fragment located 5′ in pEJH39-1 to the 0.5- to 0.6-kb *Pvu*II-*Eco*RI fragment in plasmid pP2N (Fig. 1). Numerous attempts at subcloning this fragment into *E. coli* vectors were unsuccessful; therefore, double-strand sequencing of this region in plasmid pEJH39-1 was performed. Overall, DNA sequencing was carried out for both strands of the majority (two-thirds) of the structural gene. In addition, overlapping sequence determinations were performed from different starting points over the entire 1,200-nucleotide sequence shown in Fig. 3.

Computer analysis revealed a large open reading frame starting at nucleotide 60 and terminating in a stop codon (TAA) at nucleotide 1143. The direction of transcription in this open reading frame was inferred to be from within the 1-kb *Eco*RI-*Pvu*II fragment, through the 0.5- to 0.6-kb *Pvu*II-*Eco*RI fragment, and into the *Eco*RI-*Pst*I fragment (Fig. 1). Transcription was terminated 12 nucleotides downstream from the *Mlu*I restriction site in pEJH39-1.

Primary structure of P2. The calculated molecular weight of the P2 protein encoded by nucleotides 60 to 1143 was 39,705. N-terminal amino acid analysis of purified P2 protein

TABLE 1. Comparison of the amino acid compositions of Hib DL42 P2, *E. coli* OmpF, and *N. gonorrhoeae* P.IA

Amino acid group	No. of residues/mol[a] in:		
	P2[b]	OmpF[c]	P.IA[d]
Basic			
Lys	30 (2)	18 (1)	20 (2)
Arg	16	11 (1)	16
His	7	1	12
Acidic			
Asp	17	27	17
Glu	24	14	13
Polar			
Asn	25 (1)	30 (2)	16
Cys	0	0	0
Gln	14	13	15
Gly	40 (1)	48 (1)	37
Ser	17 (1)	16	23 (1)
Thr	24 (1)	21 (1)	18 (1)
Tyr	23	29	18
Hydrophobic			
Ala	24 (8)	29 (5)	27 (6)
Ile	15 (1)	12 (2)	6 (1)
Leu	24 (2)	20 (3)	16 (4)
Met	1 (1)	3 (2)	0 (2)
Phe	13 (1)	19	15
Pro	3	4 (1)	5 (1)
Trp	0	2	5
Val	24 (1)	23 (3)	28 (1)

[a] In the mature or fully processed protein. Numbers in parentheses indicate the amino acid composition of the signal peptide.
[b] Data derived from the nucleotide sequence of the Hib P2 gene.
[c] Data obtained from Inokuchi et al. (22).
[d] Data obtained from Carbonetti and Sparling (6).

(18) indicated that the N terminus of the mature protein was encoded by nucleotides 121 through 1143. This in turn indicated that the preceding 20 amino acids encoded by nucleotides 60 through 120 represent the expected signal peptide (35). Mature P2 protein was composed of 341 amino acids and had a calculated molecular weight of 37,805. This molecular weight compares favorably with the apparent molecular weight of approximately 38,000 to 39,000 determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (15, 18). The amino acid compositions of both the mature protein and the signal peptide are listed in Table 1. N-terminal amino acid sequencing was carried out on seven individual trypsin cleavage fragments of the P2 protein. When the N terminus was included, 158 (46%) of the 341 amino acids in P2 were identified by amino acid sequencing (Fig. 3). All of these amino acid sequences obtained by peptide analysis matched exactly the amino acid sequences derived from the nucleotide sequence (Fig. 3).

The primary amino acid sequence of the entire Hib P2 protein was compared with those of the *E. coli* porins OmpC (30), OmpF (22), and PhoE (36) and with the sequence of the gonococcal porin (P.IA) (6). The greatest degree of homology was found between P2 and OmpF (24.2% identity) and between P2 and the gonococcal porin (23.5% identity) (Fig. 4). The amino acid composition of the P2 protein was unusual in that there were 48 strongly basic (Arg and Lys) residues (Table 1). Computer analysis of the amino acid sequence of P2 for regions of hydrophilicity with the Hopp and Woods algorithm (20) revealed that the overall profile of this protein was similar to those of OmpF and the gonococcal porin protein (Fig. 5).

Analysis of P2 gene copy number. DNA hybridization studies were performed to determine the number of P2 genes present in Hib. Chromosomal DNA from strain DL42 and seven other Hib strains was digested to completion with *Pst*I and *Bam*HI, separated by agarose gel electrophoresis, and transferred to nitrocellulose. A P2 gene probe was constructed by purifying and radiolabeling the 1,010-base-pair *Pvu*II-*Mlu*I fragment from pEJH39-1, which comprises all but 13 nucleotides of the DNA encoding the mature P2 protein. Southern blot analysis revealed that all eight Hib strains had a single *Pst*I band of approximately 10 kb which hybridized to this P2 gene probe (Fig. 6A). Similarly, all eight strains had a single 45- to 50-kb *Bam*HI fragment which hybridized to the P2 gene probe (Fig. 6B).

DISCUSSION

Although extensive research has been done on the porin proteins of *E. coli* (34), relatively little is known about the porins of other bacterial pathogens. Genes encoding the porin proteins of other bacterial species, including *Chlamydia trachomatis* (3, 41), *Neisseria gonorrhoeae* (6, 11), *Pseudomonas aeruginosa* (46), and *Salmonella typhi* (2), have been cloned and expressed in *E. coli*, thus providing the opportunity for detailed molecular biological investigation of the structure and function of the products of these

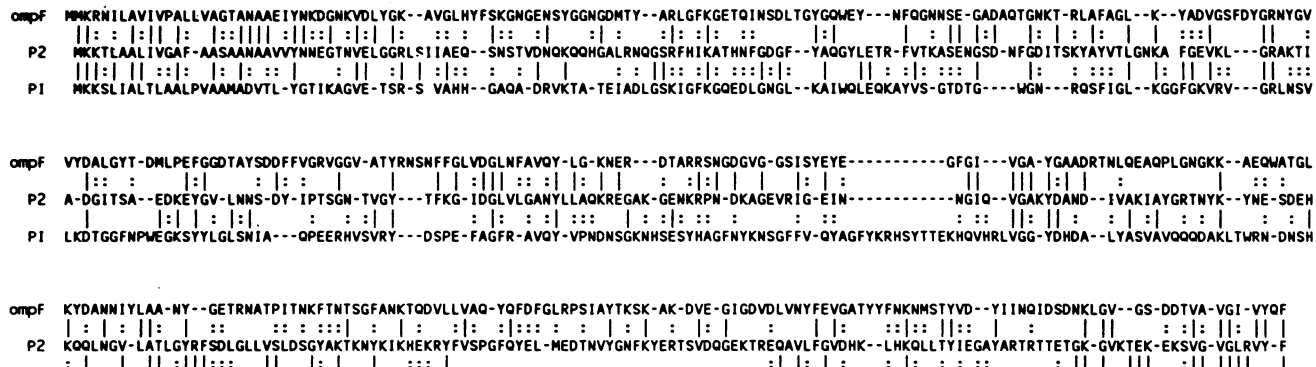

FIG. 4. Homology of the predicted primary structure of the Hib P2 protein to both the *E. coli* OmpF protein (22) and the P.IA protein of *N. gonorrhoeae* (6). The homology was analyzed by the AALIGN modification (DNASTAR) of the FASTP algorithm (38), with a gap penalty of 2 and a deletion penalty of 4. Vertical bars indicate homologies; colons indicate conservative substitutions.

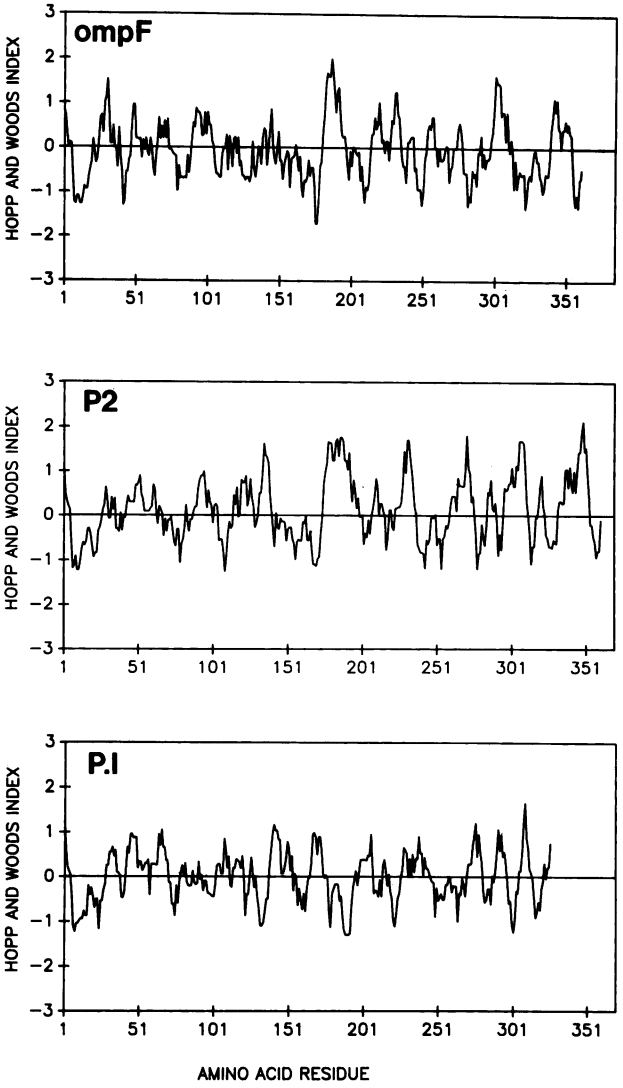

FIG. 5. Hydrophilicity analysis of *E. coli* OmpF, Hib P2, and the gonococcal P.IA protein with the algorithm of Hopp and Woods (20) at a window setting of 6. The positive index values indicate hydrophilicity.

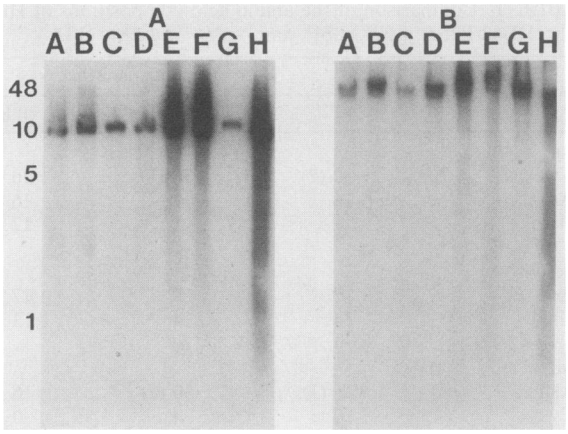

FIG. 6. Southern blot analysis of restriction enzyme digests of chromosomal DNA from eight Hib strains with a P2 gene probe. Chromosomal DNA from eight Hib strains digested to completion with *Pst*I (A) or with *Bam*HI (B) was resolved by agarose gel electrophoresis, transferred to nitrocellulose, and probed with the 1,010-base-pair *Pvu*II-*Mlu*I fragment from pEJH39-1. The Hib strains used in this experiment were DL42 (lanes A), DL26 (lanes B), CH100 (lanes C), OA104 (lanes D), SL103 (lanes E), Madigan (lanes F), DV102 (lanes G), and H234 (lanes H). Numbers on the left indicate molecular sizes in kilobases.

particular genes. Similarly, information about the P2 porin protein of Hib has only recently begun to accumulate (14, 15, 18, 26, 27, 33, 42–45), and the fact that this protein can be a target for antibodies protective against experimental Hib disease (33) underscores the importance of elucidating its immunogenic, antigenic, and functional properties. Cloning of the gene encoding the Hib P2 protein (18) made possible the exact determination of the amino acid sequence of this protein.

Examination of the 5′ untranslated region of the P2 sequence revealed the presence of the putative promoter sequence TTGAGA, located 44 nucleotides upstream from the transcriptional start site. This proposed −35 sequence differs by only one nucleotide from the *E. coli* consensus sequence TTGACA (29). The proposed ribosome-binding site AGGAA is located just 8 nucleotides upstream from the ATG (Met) codon and contains four of the bases found in the consensus sequence AGGAGG (40). Several putative Pribnow boxes were identified in the region between the proposed −35 sequence and the proposed ribosome-binding site; these were sufficiently different from the *E. coli* consensus sequence TATAAT (29) that no assignment was made for the −10 sequence.

The deduced amino acid sequence of P2 included a 20-amino-acid signal peptide, beginning with Met and terminating with Ala (Fig. 3). This signal peptide appears to have the common characteristics of these sequences, including a long stretch of hydrophobic amino acids, a disproportionately large number of alanine residues (8 Ala in 20 amino acids), and an Ala-X-Ala cleavage site (34). The signal peptide for P2 has an amino acid composition different from that of the signal peptide for the P1 heat-modifiable major outer membrane protein of Hib (32).

The deduced amino acid composition of P2 from Hib DL42 (Table 1) was similar to those reported for other Hib strains whose purified P2 proteins were subjected to amino acid analysis (33, 43). Again, the Hib P2 protein of DL42 had a markedly higher content of strongly basic amino acids (Lys and Arg) than did the OmpF porin of *E. coli* (30) and the porin of *N. gonorrhoeae* (6, 11) (Table 1). It has been suggested that the relatively high number of positively charged residues in P2 may contribute to a greater degree of hydrophilicity in this Hib porin (42). The Hib P2 protein also differed from both *E. coli* OmpF and the gonococcal P.IA protein in that it lacked tryptophan (Table 1).

There is a limited degree of homology (23 to 25%) between the Hib P2 protein and the porins of *E. coli* (OmpF) and the gonococcus (Fig. 4). Computer analysis of these amino acid sequences indicated that the Hib P2 protein was slightly more similar to OmpF than to the gonococcal porin (Fig. 4). The hydrophilicity profile of the Hib P2 protein, as determined with the Hopp and Woods algorithm (20), was very similar to those of OmpF and P.IA in that it was characterized by several relatively long hydrophilic regions and lacked substantial hydrophobic stretches (Fig. 5).

The A+T content of the Hib P2 gene was 64%, similar to the 62% A+T content of *H. influenzae* Rd chromosomal DNA (37). Codon usage in the Hib P2 protein was compared with that observed for the heat-modifiable major outer

TABLE 2. Codon usage in Hib P2 and other major outer membrane proteins

Amino acid	Codon	Codon frequency[a] in:		
		Hib P2[b]	Hib P1[c]	*E. coli* OmpF[d]
Ala	GCA	58	46	30
	GCC	3	11	6
	GCG	3	26	14
	GCU	25	44	44
Arg	AGA	8	9	0
	AGG	0	0	0
	CGA	0	2	0
	CGC	11	4	6
	CGG	3	2	0
	CGU	22	17	28
Asn	AAC	36	11	77
	AAU	36	57	11
Asp	GAC	8	17	39
	GAU	39	39	36
Cys	UGC	0	0	0
	UGU	0	0	0
Gln	CAA	33	24	17
	CAG	6	11	19
Glu	GAA	58	31	36
	GAG	8	0	3
Gly	GGA	6	6	0
	GGC	25	13	41
	GGG	3	4	3
	GGU	80	57	91
His	CAC	11	7	0
	CAU	8	13	3
Ile	AUA	6	11	0
	AUC	11	4	36
	AUU	28	20	3
Leu	CUA	6	7	0
	CUC	3	2	0
	CUG	0	2	50
	CUU	8	20	8
	UUA	55	44	6
	UUG	0	26	0
Lys	AAA	80	59	44
	AAG	8	11	8
Met	AUG	6	15	14
Phe	UUC	30	17	33
	UUU	8	26	19
Pro	CCA	3	13	3
	CCC	0	4	0
	CCG	0	2	6
	CCU	6	0	6
Ser	AGC	8	7	8
	AGU	17	31	3
	UCA	11	11	3
	UCC	0	2	14
	UCG	0	0	0
	UCU	14	28	17
Thr	ACA	19	24	3
	ACC	6	11	30
	ACG	0	4	8
	ACU	44	24	19
Trp	UGG	0	9	6
Tyr	UAC	8	11	55
	UAU	55	46	25
Val	GUA	33	13	11
	GUC	3	7	3
	GUG	6	13	14
	GUU	28	26	44

[a] Based on the number of codons used per 1,000 codons.
[b] Data derived from the nucleotide sequence of P2.
[c] Data obtained from Munson and Grass (32).
[d] Data obtained from Inokuchi et al. (22).

membrane protein (P1) of Hib and for the *E. coli* OmpF protein (Table 2). In general, codons utilized in the synthesis of the Hib P2 protein had an A or U in position 3; this finding is again in keeping with the A+T-rich nature of *H. influenzae* chromosomal DNA (37). Codon usage differed between P2 and P1 in several ways. For example, the leucine codon UUG was used frequently in the P1 gene but not at all in the P2 gene. Similarly, the phenylalanine codon UUC was used in the P2 gene much more frequently than was UUU; in the P1 gene, the latter phenylalanine codon was predominant. While P2 had relatively more threonine residues than did P1, the threonine codon ACG was not used at all in the P2 gene, but it was used in the P1 gene. Codon usage was much more similar between the Hib P1 and P2 genes, however, than it was between the P2 gene and the *E. coli* OmpF gene. As might be expected from a previous study involving the Hib P1 gene (32), certain codons used relatively frequently in the OmpF gene, such as the proline codon CCG, the leucine codon CUG, and the serine codon UCC, were not used at all in the Hib P2 gene. All three of these codons lack an A or U in position 3. These findings, together with the fact that *E. coli* genes encoding abundant proteins tend to use major isoaccepting transfer RNA molecules selectively (21, 30), suggest that the distribution of major transfer RNA species in Hib may be different from that in *E. coli*.

It is known that some bacteria, including *E. coli*, can synthesize more than a single type of porin (34). In *E. coli* there is considerable identity (69%) between the DNA regions encoding the mature forms of the porins OmpF and OmpC (30). When a Hib P2 probe representing essentially all of the DNA encoding the P2 protein of strain DL42 was used in Southern blot analysis with complete restriction enzyme digests of chromosomal DNA from DL42 and seven other Hib strains, only a single band in each digest hybridized to this probe (Fig. 6). The restriction enzymes used in this experiment (*Pst*I and *Bam*HI) do not cut within the P2 gene but instead yield relatively large DNA fragments (10 and 45 to 50 kb, respectively) that hybridized to the P2 gene probe. These results suggest that there is a single copy of the Hib P2 gene in the Hib chromosome and that if other porins exist in Hib, their nucleotide sequences are significantly different from that encoding the P2 protein characterized in this report. Similarly, it should be noted that minor nucleotide

sequence variations may exist among the P2 genes detected in the seven other Hib strains and that these differences may not be sufficient to prevent hybridization to the DL42-derived, 1,010-base-pair P2 probe used in these Southern blot analyses.

A previous study suggested that unregulated expression of intact Hib P2 protein might be lethal to *E. coli* (18). It is interesting to note that we were unable to subclone into *E. coli* the 1-kb *Eco*RI-*Pvu*II fragment of pEJH39-1 that contains the DNA encoding the signal peptide and presumed promoter and regulatory regions of P2. The fact that Hib promoters function effectively in *E. coli* (18, 19, 32) indicates that this fragment could be used to synthesize the P2 signal peptide, which might be lethal if overproduced in *E. coli*. Alternatively, our inability to subclone this particular Hib DNA fragment into *E. coli* may have resulted from some property of either the phagemid sequencing vectors or the *E. coli* host used in the subcloning experiments.

Availability of the nucleotide sequence for P2 will facilitate detailed analyses of the structure-function relationships of this protein. It is known that Hib can synthesize P2 proteins with apparent molecular weights (in sodium dodecyl sulfate-polyacrylamide gel electrophoresis) of approximately 38,000 to 39,000 or 39,000 to 40,000 and that each strain apparently expresses one or the other, but not both, of these two basic types of P2 (5, 14, 15, 25, 44). This characteristic is also involved in the subtyping of Hib strains on the basis of their outer membrane protein profiles in sodium dodecyl sulfate-polyacrylamide gel electrophoresis (5). N-terminal amino acid analysis of P2 proteins from the two types of Hib strains has shown that the sequences of the first 24 amino acids in both types of P2 are identical (E. J. Hansen et al., manuscript in preparation). The results of immunoprotection studies involving antisera raised against purified P2 protein, however, indicate that some degree of antigenic heterogeneity may exist among Hib strains with respect to the surface epitopes of this protein which are targets for protective antibodies (33). Cloning and sequencing of genes encoding the other type of P2 protein will be required to identify those differences in amino acid sequences between the two types of P2 which account not only for the molecular weight variation but also for the apparent difference in immunogenic surface epitopes. These findings can be used with site-directed mutagenesis techniques to identify surface-exposed P2 epitopes which are targets for antibodies protective against experimental Hib disease. Similarly, it may be possible to use Hib P2 DNA sequence information and site-directed mutagenesis to identify the region(s) of the P2 protein involved in the avid binding of this protein to Hib lipo-oligosaccharide (15).

ACKNOWLEDGMENTS

This work was supported by Public Health Service grant AI-17621 from the National Institute of Allergy and Infectious Diseases to E.J.H.

We thank Porter Anderson for Hib Madigan and H234 and Martin Goldberg for assistance with computer analyses and graphics. We also thank Mark Hanson, Leslie Cope, Michael Norgard, Justin Radolf, and Jacob Kupersztoch for their comments on the manuscript, which was expertly typed by Cindy Baselski.

LITERATURE CITED

1. **Aebersold, R. H., J. Leavitt, R. A. Saavedra, L. E. Hood, and S. B. H. Kent.** 1987. Internal amino acid sequence analysis of proteins separated by one- or two-dimensional gel electrophoresis by *in situ* protease digestion on nitrocellulose. Proc. Natl. Acad. Sci. USA **84**:6970–6974.
2. **Aguero, J., G. Mora, M. J. Mroczenski-Wildey, M. E. Fernandez-Beros, L. Aron, and F. C. Cabello.** 1987. Cloning, expressing, and characterization of the 36 KDal *Salmonella typhi* porin gene in *Escherichia coli*. Microb. Pathogen. **3**:399–407.
3. **Allan, I., T. M. Cunningham, and M. A. Lovett.** 1984. Molecular cloning of the major outer membrane protein of *Chlamydia trachomatis*. Infect. Immun. **45**:637–641.
4. **Andrews, P. C., and J. E. Dixon.** 1987. A procedure for *in situ* alkylation of cysteine residues on glass fiber prior to protein microsequence analysis. Anal. Biochem. **161**:524–528.
5. **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.
6. **Carbonetti, N. H., and P. F. Sparling.** 1987. Molecular cloning and characterization of the structural gene for protein I, the major outer membrane protein in *Neisseria gonorrhoeae*. Proc. Natl. Acad. Sci. USA **84**:9084–9088.
7. **Cochi, S. L., and C. V. Broome.** 1986. Vaccine prevention of *Haemophilus influenzae* type b disease: past, present, and future. Pediatr. Infect. Dis. **5**:12–19.
8. **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.
9. **Dente, L., G. Cesareni, and R. Cortese.** 1983. PEMBL: a new family of single-stranded plasmids. Nucleic Acids Res. **11**: 1645–1655.
10. **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.
11. **Gotschlich, E. L., 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.
12. **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.
13. **Green, B. A., T. Quinn-Dey, and G. W. Zlotnick.** 1987. Biologic activities of antibody to a peptidoglycan-associated lipoprotein of *Haemophilus influenzae* against multiple clinical isolates of *H. influenzae* type b. Infect. Immun. **55**:2878–2883.
14. **Gulig, P. A., C. F. Frisch, and E. J. Hansen.** 1983. A set of two monoclonal antibodies specific for the cell surface-exposed 39K major outer membrane protein of *Haemophilus influenzae* type b defines all strains of this pathogen. Infect. Immun. **42**:516–524.
15. **Gulig, P. A., and E. J. Hansen.** 1985. Coprecipitation of lipopolysaccharide and the 39,000-molecular-weight major outer membrane protein of *Haemophilus influenzae* type b by lipopolysaccharide-directed monoclonal antibody. Infect. Immun. **49**:819–827.
16. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**:557–580.
17. **Hansen, E. J., C. F. Frisch, R. L. McDade, Jr., and K. H. Johnston.** 1981. Identification of immunogenic outer membrane proteins of *Haemophilus influenzae* type b in the infant rat model system. Infect. Immun. **32**:1084–1092.
18. **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.
19. **Holmans, P. L., T. A. Loftus, and E. J. Hansen.** 1985. Cloning and surface expression in *Escherichia coli* of a structural gene encoding a surface protein of *Haemophilus influenzae* type b. Infect. Immun. **50**:236–242.
20. **Hopp, T. P., and K. R. Woods.** 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA **78**:3824–3828.
21. **Ikemura, T.** 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon

choice that is optimal for the *E. coli* translational system. J. Mol. Biol. **151**:389–409.
22. **Inokuchi, K., N. Mutoh, S. Matsuyama, and S. Mizushima.** 1982. Primary structure of the ompF gene that codes for a major outer membrane protein of *Escherichia coli* K-12. Nucleic Acids Res. **10**:6957–6968.
23. **Kimura, A., P. A. Gulig, G. H. McCracken, Jr., T. A. Loftus, and E. J. Hansen.** 1985. A minor high-molecular-weight outer membrane protein of *Haemophilus influenzae* type b is a protective antigen. Infect. Immun. **47**:253–259.
24. **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.
25. **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.
26. **Loeb, M. R., and D. H. Smith.** 1982. Human antibody response to individual outer membrane proteins of *Haemophilus influenzae* type b. Infect. Immun. **37**:1032–1036.
27. **Loeb, M. R., and K. A. Woodlin.** 1987. Cross-reactivity of surface-exposed epitopes of outer membrane antigens of *Haemophilus influenzae* type b. Infect. Immun. **55**:2977–2983.
28. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
29. **McClure, W. R.** 1985. Mechanism and control of transcriptional initiation in prokaryotes. Annu. Rev. Biochem. **54**:171–204.
30. **Mizuno, T., M.-Y. Chou, and M. Inouye.** 1983. A comparative study on the genes for three porins of the *Escherichia coli* outer membrane. DNA sequence of the osmoregulated ompC gene. J. Biol. Chem. **258**:6932–6940.
31. **Munson, R. S., Jr., and D. M. Granoff.** 1985. Purification and partial characterization of outer membrane proteins P5 and P6 from *Haemophilus influenzae* type b. Infect. Immun. **49**:544–599.
32. **Munson, R. S., Jr., and S. Grass.** 1988. Purification, cloning, and sequence of outer membrane protein P1 of *Haemophilus influenzae* type b. Infect. Immun. **56**:2235–2242.
33. **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.
34. **Nikaido, H., and M. Vaara.** 1985. Molecular basis of bacterial outer membrane permeability. Microbiol. Rev. **49**:1–32.
35. **Oliver, D.** 1985. Protein secretion in *Escherichia coli*. Annu. Rev. Microbiol. **39**:615–648.
36. **Overbeeke, N., H. Bergmans, F. Van Mansfeld, and B. Lugtenberg.** 1983. Complete nucleotide sequence of *phoE*, the structural gene for the phosphate limitation inducible outer membrane pore protein of *Escherichia coli* K12. J. Mol. Biol. **163**:513–532.
37. **Roy, P. H., and H. O. Smith.** 1973. DNA methylases of *H. influenzae* Rd. J. Mol. Biol. **81**:427–444.
38. **Sel, S. H. W., and D. T. Karzon.** 1973. *Haemophilus influenzae*. Vanderbilt University Press, Nashville, Tenn.
39. **Shenep, J. L., R. S. Munson, Jr., S. J. Barenkamp, and D. M. Granoff.** 1983. Further studies of the role of noncapsular antibody in protection against experimental *Haemophilus influenzae* type b bacteremia. Infect. Immun. **42**:257–263.
40. **Shine, J., and L. Dalgarno.** 1975. Determinant of cystron specificity in bacterial ribosomes. Nature (London) **254**:34–38.
41. **Stephens, R. S., G. Mullenbach, R. Sanchez-Pescador, and N. Agabian.** 1986. Sequence analysis of the major outer membrane protein gene of *Chlamydia trachomatis* serovar L_2. J. Bacteriol. **168**:1277–1282.
42. **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.
43. **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.
44. **Van Alphen, L., T. Riemens, J. Poolman, and H. C. Zanen.** 1983. Characteristics of major outer membrane proteins of *Haemophilus influenzae*. J. Bacteriol. **155**:878–885.
45. **Van Alphen, L., T. Riemens, and H. C. Zanen.** 1983. Antibody response against outer membrane components of *Haemophilus influenzae* type b strains in patients with meningitis. FEMS Microbiol. Lett. **18**:189–195.
46. **Woodruff, W. A., T. R. Parr, Jr., R. E. W. Hancock, L. F. Hanne, T. I. Nicas, and B. H. Iglewski.** 1983. Expression in *Escherichia coli* and function of *Pseudomonas aeruginosa* outer membrane porin protein F. J. Bacteriol. **167**:473–479.
47. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.
48. **Zhang, H., R. Scholl, J. Browse, and C. Somerville.** 1988. Double-stranded DNA sequencing as a choice for DNA sequencing. Nucleic Acids Res. **16**:1220.