INFECTION AND IMMUNITY, Apr. 1992, p. 1302–1313
0019-9567/92/041302-12$02.00/0

Vol. 60, No. 4

Cloning, Expression, and DNA Sequence Analysis of Genes Encoding Nontypeable *Haemophilus influenzae* High-Molecular-Weight Surface-Exposed Proteins Related to Filamentous Hemagglutinin of *Bordetella pertussis*

STEPHEN J. BARENKAMP[1]* AND ELIZABETH LEININGER[2]

Edward Mallinckrodt Department of Pediatrics, Washington University School of Medicine, and Division of Infectious Diseases, St. Louis Children's Hospital, St. Louis, Missouri 63110,[1] and Laboratory of Cellular Physiology, Division of Bacterial Products, Center for Biologics Evaluation and Research, Food and Drug Administration, Bethesda, Maryland[2]

Received 4 November 1991/Accepted 10 January 1992

A group of high-molecular-weight surface-exposed proteins of nontypeable *Haemophilus influenzae* are major targets of human serum antibody (S. J. Barenkamp and F. F. Bodor, Pediatr. Infect. Dis. J. 9:333–337, 1990). To further characterize these proteins, we cloned and sequenced genes encoding two related high-molecular-weight proteins from a prototype nontypeable *Haemophilus* strain. The gene encoding a 120-kDa *Haemophilus* protein consisted of a 4.4-kbp open reading frame, and the gene encoding a 125-kDa protein consisted of a 4.6-kbp open reading frame. The first 1,259 bp of the two genes were identical. Thereafter, the sequences began to diverge, but overall they were 80% identical, and the derived amino acid sequences showed 70% identity. A protein sequence homology search demonstrated similarity between the derived amino acid sequences of both cloned genes and the derived amino acid sequence of the gene encoding filamentous hemagglutinin, a surface protein produced by the gram-negative pathogen *Bordetella pertussis*. Antiserum raised against a recombinant protein encoded by the 4.6-kbp open reading frame recognized both the 120- and the 125-kDa proteins in the prototype strain as well as antigenically related high-molecular-weight proteins in 75% of a collection of 125 epidemiologically unrelated nontypeable *H. influenzae* strains. The antiserum directed against the recombinant protein also recognized purified filamentous hemagglutinin. A murine monoclonal antibody to filamentous hemagglutinin recognized both the 120-kDa and the 125-kDa protein in the prototype strain as well as proteins identical to those recognized by the recombinant-protein antiserum in 35% of the nontypeable *H. influenzae* strain collection. Thus, we have identified and partially characterized a group of highly immunogenic surface-exposed proteins of nontypeable *H. influenzae* which are related to the filamentous hemagglutinin of *B. pertussis*.

Nontypeable *Haemophilus influenzae* are unencapsulated organisms which commonly inhabit the upper respiratory tract of humans. These organisms are a frequent cause of mucosal surface infections, such as otitis media, sinusitis, and bronchitis, and occasionally cause invasive diseases such as bacteremia and meningitis (22). The interactions between the organism and the host which determine whether colonization will occur and whether disease will develop following colonization are not completely understood. However, host immunity is thought to play an important role in this sequence of events, and antibodies directed against surface antigens of these organisms are thought to be central to host protection (9).

The surface antigens of nontypeable *H. influenzae* against which protective antibodies are directed have not been completely defined. Bacterial outer membrane proteins appear to be targets of bactericidal antibody (9), and two of the major outer membrane proteins of nontypeable *H. influenzae*, designated P2 and P6, have been identified as targets of human serum bactericidal antibody (23, 24). Other outer membrane antigens are also likely to be targets of bactericidal antibodies. We previously identified a group of high-molecular-weight surface-exposed proteins as major targets of human serum antibody and observed a correlation between the development of serum bactericidal antibody and the appearance of antibody directed against these proteins (1). The structure and function of these high-molecular-weight proteins were unclear from these earlier studies. In an effort to further characterize these proteins, we cloned, expressed, and sequenced the genes encoding two immunodominant high-molecular-weight proteins from a prototype nontypeable *Haemophilus* strain. The nucleotide sequences of the two cloned genes were different but closely related, and the proteins encoded by these genes were antigenically related to filamentous hemagglutinin, a surface protein found on *Bordetella pertussis*. Furthermore, antigenically related high-molecular-weight proteins were present in the majority of heterologous nontypeable *H. influenzae* strains.

MATERIALS AND METHODS

Bacterial strains, plasmids, and phages. Nontypeable *H. influenzae* strain 12 was the clinical isolate chosen for study. The organism was isolated in pure culture from the middle ear fluid of a child with acute otitis media. The strain was identified as *H. influenzae* by standard methods (11) and was classified as nontypeable by its failure to agglutinate with a panel of typing antisera for *H. influenzae* types a to f (Burroughs Wellcome Co., Research Triangle Park, N.C.) and failure to show lines of precipitation with these antisera in counterimmunoelectrophoresis assays (42). Furthermore,

* Corresponding author.

genomic DNA purified from strain 12 failed to hybridize with a plasmid probe, pUO38, which contains DNA from the *cap* region of type b *H. influenzae* (7).

An additional 125 epidemiologically unrelated nontypeable *H. influenzae* strains were chosen for the studies of strain heterogeneity. These isolates were recovered during episodes of clinical disease from children and adults in Canada, Massachusetts, Maryland, Connecticut, Missouri, New York, Ohio, Texas, and Alabama. Of these isolates, 100 were cultured from middle ear effusions collected by tympanocentesis, 10 were from blood, 10 were from sputum, and 5 were from cerebrospinal fluid. Approximately half of these isolates were described previously (25). The bulk of the remainder were isolated from middle ear aspirates from children with acute otitis media, either by one of the authors (S.J.B.) or by his collaborators.

M13mp18 and M13mp19 were obtained from New England BioLabs, Inc. (Beverly, Mass.). λEMBL3 arms and *Escherichia coli* LE392 were obtained from Stratagene (La Jolla, Calif.). pT7-7 was the kind gift of Stanley Tabor (37). *E. coli* BL21(DE3)/pLysS was a gift from F. William Studier. Strain BL21(DE3) contains a single copy of the T7 RNA polymerase gene under the control of the *lac* regulatory system (36). Plasmid pLysS contains the T7 lysozyme gene. T7 lysozyme binds to the T7 RNA polymerase in vitro, and pLysS stabilizes many toxic T7 expression constructs, presumably by binding and inactivating the low quantities of T7 RNA polymerase produced by BL21(DE3) in the absence of isopropyl-β-D-thiogalactopyranoside (IPTG) (21).

Molecular cloning and plasmid subcloning. Chromosomal DNA from strain 12 was prepared by a modification of the method of Marmur (18). *Sau*3A partial restriction digests of chromosomal DNA were prepared and fractionated on sucrose gradients. Fractions containing DNA fragments in the 9- to 20-kbp range were pooled, and a library was prepared by ligation into λEMBL3 arms. Ligation mixtures were packaged in vitro with Gigapack (Stratagene) and plate-amplified in a P2 lysogen of *E. coli* LE392. Lambda plaque immunological screening was performed as described by Maniatis et al. (17). For plasmid subcloning studies, DNA from representative recombinant phage was subcloned into the T7 expression plasmid pT7-7 (37). This vector contains the T7 RNA polymerase promoter ϕ10, a ribosome-binding site, and the translational start site for the T7 gene 10 protein upstream from a multiple cloning site (see Fig. 2B). Standard DNA methods for manipulation of cloned DNA were as described by Maniatis et al. (17) or Silhavy et al. (32).

DNA sequence analysis. DNA sequence analysis was performed by the dideoxy method with the U.S. Biochemicals Sequenase kit as suggested by the manufacturer. [^{35}S]dATP was purchased from New England Nuclear (Boston, Mass.). Data were analyzed with Compugene software (2) and the Genetics Computer Group program from the University of Wisconsin (5) on a Digital VAX 8530 computer. Several 21-mer oligonucleotide primers were generated as necessary to complete the sequences. Both strands of the HMW1 gene and a single strand of the HMW2 gene were sequenced.

Analytical techniques. Western immunoblot analysis was performed to identify the recombinant proteins being produced by reactive phage clones. Phage lysates grown in LE392 cells or plaques picked directly from a lawn of LE392 cells on YT plates were solubilized in gel electrophoresis sample buffer prior to electrophoresis. Sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis was performed on 7.5% or 11% polyacrylamide modified Laemmli gels as described by Lugtenberg et al. (16). After transfer of the proteins to nitrocellulose sheets, the sheets were probed sequentially with an *E. coli*-absorbed human serum sample containing high-titer antibody to the high-molecular-weight proteins and then with alkaline phosphatase-conjugated goat anti-human immunoglobulin G (IgG) second antibody (Tago, Burlingame, Calif.). In previous studies, we noted that sera from healthy adults contain high-titer antibody directed against surface-exposed high-molecular-weight proteins of nontypeable *H. influenzae* (1). Serum from one of these adults was used as the screening antiserum after having been extensively absorbed with LE392 cells.

To identify recombinant proteins being produced by *E. coli* transformed with recombinant plasmids, the plasmids of interest were used to transform *E. coli* BL21(DE3)/pLysS. The transformed strains were grown to an A_{600} of 0.5 in L broth containing 50 μg of ampicillin per ml. IPTG was then added to 1 mM. One hour later, cells were harvested, and a sonicate of the cells was prepared. The protein concentrations of the samples were determined by the bicinchoninic acid method (33) (BCA protein assay kit; Pierce Chemical Co., Rockford, Ill.) in accordance with the manufacturer's instructions. Cell sonicates containing 100 μg of total protein were solubilized in electrophoresis sample buffer, subjected to SDS-polyacrylamide gel electrophoresis, and transferred to nitrocellulose. The nitrocellulose was then probed sequentially with the *E. coli*-absorbed adult serum sample and then with alkaline phosphatase-conjugated goat anti-human IgG second antibody.

Western immunoblot analysis was also performed to determine whether homologous and heterologous nontypeable *H. influenzae* strains expressed high-molecular-weight proteins antigenically related to the protein encoded by the cloned HMW1 gene (rHMW1). Cell sonicates of bacterial cells were solubilized in electrophoresis sample buffer, subjected to SDS-polyacrylamide gel electrophoresis, and transferred to nitrocellulose. Nitrocellulose was probed sequentially with polyclonal rabbit rHMW1 antiserum (see below) and then with alkaline phosphatase-conjugated goat anti-rabbit IgG second antibody (Bio-Rad, Gaithersburg, Md.).

Finally, Western immunoblot analysis was performed to determine whether nontypeable *Haemophilus* strains expressed proteins antigenically related to the filamentous hemagglutinin protein of *Bordetella pertussis*. Monoclonal antibody X3C, a murine immunoglobulin G (IgG) antibody which recognizes filamentous hemagglutinin (13), was used to probe cell sonicates by Western blot. An alkaline phosphatase-conjugated goat anti-mouse IgG second antibody (Tago) was used for detection.

Preparation of HMW1 recombinant-protein antiserum. *E. coli* BL21(DE3)/pLysS was transformed with pHMW1-4, and expression of recombinant protein was induced with IPTG, as described above. A cell sonicate of the bacterial cells was prepared and separated into a supernatant and pellet fraction by centrifugation at 10,000 × *g* for 30 min. The recombinant protein fractionated with the pellet fraction. A rabbit was subcutaneously immunized on a biweekly schedule with 1 mg of protein from the pellet fraction, the first dose given with Freund's complete adjuvant and subsequent doses with Freund's incomplete adjuvant. Following the fourth injection, the rabbit was bled. Prior to use in the Western blot assay, the antiserum was absorbed extensively with sonicates of the host *E. coli* strain transformed with the cloning vector alone.

ELISA with rHMW1 antiserum and purified filamentous hemagglutinin. Purified filamentous hemagglutinin was a gift from Pasteur Merieux Sérums et Vaccines, Marcy l'Étoile,

France. Enzyme-linked immunosorbent assay (ELISA) plates (Costar, Cambridge, Mass.) were coated with 60 μl of a 4-μg/ml solution of filamentous hemagglutinin in Dulbecco's phosphate-buffered saline per well for 2 h at room temperature. Wells were blocked for 1 h with 1% bovine serum albumin in Dulbecco's phosphate-buffered saline prior to addition of serum dilutions. rHMW1 antiserum was serially diluted in 0.1% Brij (Sigma, St. Louis, Mo.) in Dulbecco's phosphate-buffered saline and incubated for 3 h at room temperature. After being washed, the plates were incubated with peroxidase-conjugated goat anti-rabbit IgG antibody (Bio-Rad) for 2 h at room temperature and subsequently developed with 2,2'-azino-bis(3-ethylbenzthiazoline-6-sulfonic acid) (Sigma) at a concentration of 0.54 mg/ml in 0.1 M sodium citrate buffer, pH 4.2, containing 0.03% H_2O_2. Absorbances were read on an automated ELISA reader (Dynatech Laboratories, Chantilly, Va.).

Nucleotide sequence accession numbers. The sequences of the HMW1 and HMW2 genes have been submitted to GenBank. The accession numbers are M84616 for the HMW1 gene and M84615 for the HMW2 gene.

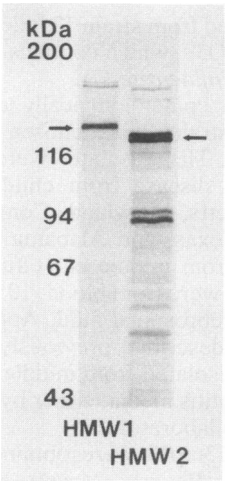


FIG. 1. Western immunoblot assay of phage lysates containing either the HMW1 or HMW2 recombinant proteins. Lysates were probed with an *E. coli*-absorbed adult serum sample with high-titer antibody against high-molecular-weight proteins. The arrows indicate the major immunoreactive protein bands of 125 and 120 kDa in the HMW1 and HMW2 lysates, respectively.

RESULTS

Isolation and restriction mapping of recombinant phage expressing HMW1 or HMW2 high-molecular-weight recombinant proteins. The nontypeable *H. influenzae* strain 12 genomic library was screened for clones expressing high-molecular-weight proteins with an *E. coli*-absorbed human serum sample containing a high titer of antibodies directed against the high-molecular-weight proteins (1). Although this human serum sample contained antibodies against a number of other *Haemophilus* proteins, the undefined nature of the high-molecular-weight proteins and our lack of specific antisera precluded the use of a more specific screening antiserum.

Numerous strongly reactive clones were identified along with more weakly reactive ones. Twenty strongly reactive clones were plaque-purified and examined by Western blot for expression of recombinant proteins. Each of the strongly reactive clones expressed one of two types of high-molecular-weight proteins, designated HMW1 and HMW2 (Fig. 1). The major immunoreactive protein bands in the HMW1 and HMW2 lysates migrated with apparent molecular masses of 125 and 120 kDa, respectively. In addition to the major bands, each lysate contained minor protein bands of higher apparent molecular weight. Protein bands seen in the HMW2 lysates at molecular masses of less than 120 kDa were not regularly observed and presumably represent proteolytic degradation products. Lysates of LE392 infected with the λEMBL3 cloning vector alone were nonreactive when immunologically screened with the same serum sample (data not shown). Thus, the observed activity was not due to cross-reactive *E. coli* proteins or λEMBL3-encoded proteins. Furthermore, the recombinant proteins were not simply binding immunoglobulin nonspecifically, since the proteins were not reactive with the goat anti-human IgG conjugate alone, with normal rabbit sera, or with sera from a number of healthy young infants.

Representative clones expressing either the HMW1 or HMW2 recombinant proteins were characterized further. The restriction maps of the two phage types were different from each other, including the regions encoding the HMW1 and HMW2 structural genes. Figure 2A shows restriction maps of representative recombinant phage which contained the HMW1 or HMW2 structural genes. The locations of the structural genes are indicated by the shaded bars.

Subcloning and expression of the HMW1 structural gene in recombinant plasmids. HMW1 plasmid subclones were constructed by using the T7 expression plasmid pT7-7 (Fig. 2A and B). HMW2 plasmid subclones were also constructed, but the results with these latter subclones were similar to those observed with the HMW1 constructs; so for the purposes of discussion, only the HMW1 subcloning data will be presented.

The approximate location and direction of transcription of the HMW1 structural gene were initially determined by using plasmid pHMW1 (Fig. 2A). This plasmid was constructed by inserting the 8.5-kb *Bam*HI-*Sal*I fragment from λHMW1 into *Bam*HI- and *Sal*I-cut pT7-7. *E. coli* transformed with pHMW1 expressed an immunoreactive recombinant protein with an apparent molecular mass of 115 kDa, which was strongly inducible with IPTG (data not shown). This protein was significantly smaller than the 125-kDa major protein expressed by the parent phage, suggesting that it either was being expressed as a fusion protein or was truncated at the carboxy terminus.

To more precisely localize the 3′ end of the structural gene, we constructed additional plasmids with progressive deletions from the 3′ end of the pHMW1 construct. Plasmid pHMW1-1 was constructed by digestion of pHMW1 with *Pst*I, isolation of the resulting 8.8-kb fragment, and religation. Plasmid pHMW1-2 was constructed by digestion of pHMW1 with *Hin*dIII, isolation of the resulting 7.5-kb fragment, and religation. *E. coli* transformed with either plasmid pHMW1-1 or pHMW1-2 also expressed an immunoreactive recombinant protein with an apparent molecular mass of 115 kDa. These results suggested that the 3′ end of the structural gene was 5′ of the *Hin*dIII site. Figure 3 demonstrates the Western blot results with pHMW1-2-transformed cells before and after IPTG induction (lanes 3 and 4, respectively). The 115-kDa recombinant protein is indicated by the arrow. Transformants also demonstrated cross-reactive bands of lower apparent molecular weight, which presumably represent partial degradation products. Shown for

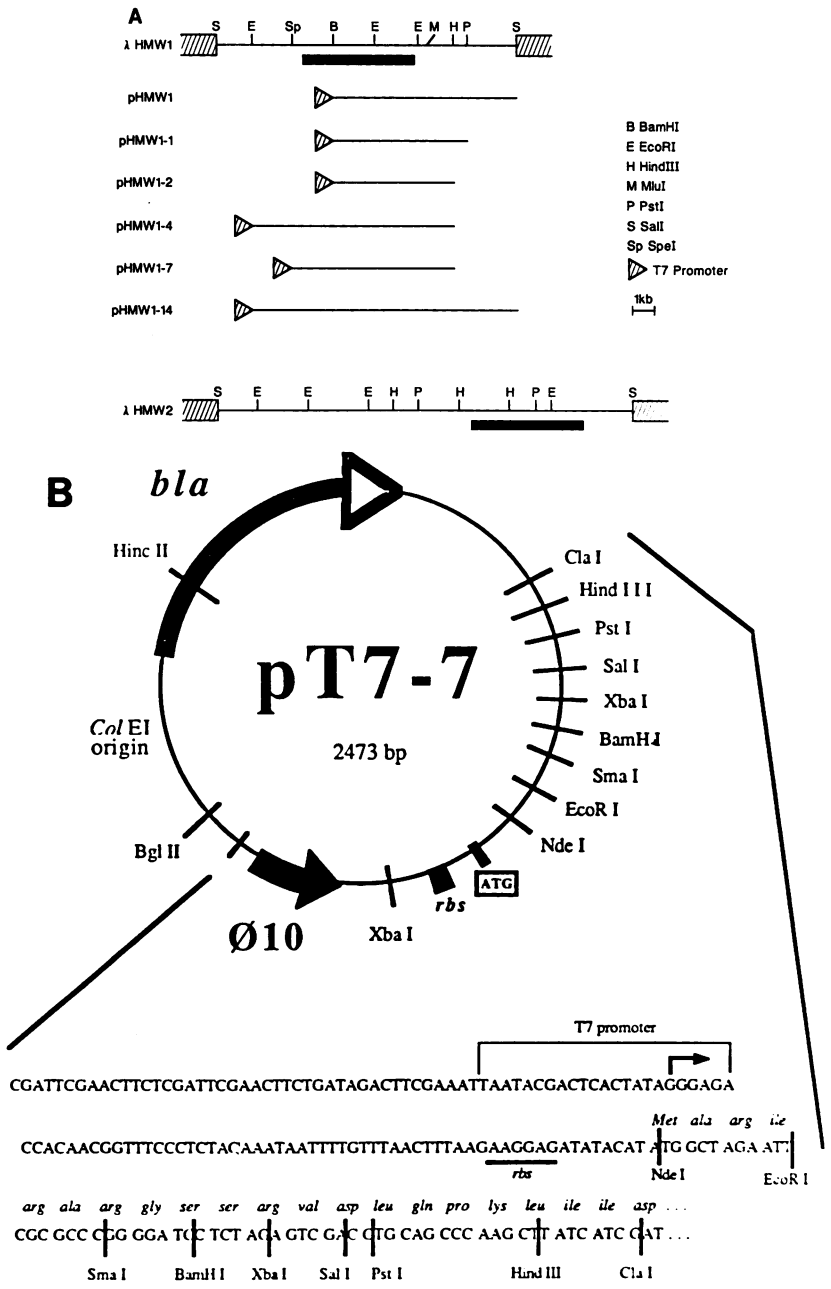


FIG. 2. (A) Partial restriction maps of representative HMW1 and HMW2 recombinant phage and of HMW1 plasmid subclones. The shaded boxes indicate the locations of the structural genes. In the recombinant phage, transcription proceeds from left to right for the HMW1 gene and from right to left for the HMW2 gene. The methods used for construction of the plasmids shown are described in the text. (B) Restriction map of the T7 expression vector pT7-7. This vector contains the T7 RNA polymerase promoter ϕ10, a ribosome-binding site (*rbs*), and the translational start site for the T7 gene 10 protein upstream from a multiple cloning site (37).

comparison are the results for *E. coli* transformed with the pT7-7 cloning vector alone (Fig. 3, lanes 1 and 2).

To more precisely localize the 5′ end of the gene, plasmids pHMW1-4 and pHMW1-7 were constructed. Plasmid pHMW1-4 was constructed by cloning the 5.1-kb *Bam*HI-*Hin*dIII fragment from λHMW1 into a pT7-7-derived plasmid containing the upstream 3.8-kb *Eco*RI-*Bam*HI fragment. *E. coli* transformed with pHMW1-4 expressed an immunoreactive protein with an apparent molecular mass of approximately 160 kDa (Fig. 3, lane 6). Although protein production was inducible with IPTG, the levels of protein production in these transformants were substantially lower than those with the pHMW1-2 transformants described above. Plasmid pHMW1-7 was constructed by digesting pHMW1-4 with *Nde*I and *Spe*I. The 9.0-kbp fragment generated by this double digestion was isolated, blunt ended, and religated. *E. coli* transformed with pHMW1-7 also expressed an immunoreactive protein with an apparent molecular mass of 160 kDa (data not shown), a protein identical in size to that expressed by the pHMW1-4 transformants. This result sug-

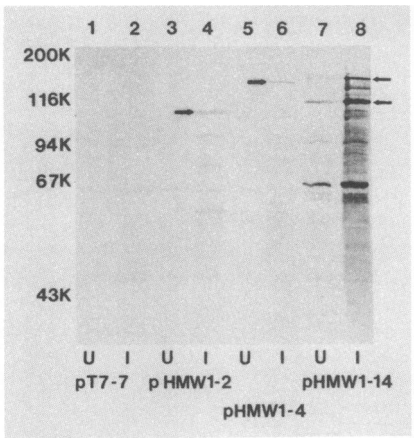


FIG. 3. Western immunoblot assay of cell sonicates prepared from *E. coli* transformed with plasmid pT7-7 (lanes 1 and 2), pHMW1-2 (lanes 3 and 4), pHMW1-4 (lanes 5 and 6), or pHMW1-14 (lanes 7 and 8). The sonicates were probed with an *E. coli*-absorbed adult serum sample with high-titer antibody against high-molecular-weight proteins. Lanes labeled U and I represent sonicates prepared before and after induction of the growing samples with IPTG, respectively. The arrows indicate protein bands of interest as described in the text.

gested that the initiation codon for the HMW1 structural gene was 3′ of the *Spe*I site. DNA sequence analysis (described below) confirmed this conclusion.

As noted above, the λHMW1 phage clones expressed a major immunoreactive band of 125 kDa, whereas the HMW1 plasmid clones pHMW1-4 and pHMW1-7, which contained what was believed to be the full-length gene, expressed an immunoreactive protein of approximately 160 kDa. This size discrepancy was disconcerting. One possible explanation was that an additional gene or genes necessary for correct processing of the HMW1 gene product were deleted in the process of subcloning. To address this possibility, plasmid pHMW1-14 was constructed. This construct was generated by digesting pHMW1 with *Nde*I and *Mlu*I and inserting the 7.6-kbp *Nde*I-*Mlu*I fragment isolated from pHMW1-4. Such a construct would contain the full-length HMW1 gene as well as the DNA 3′ of the HMW1 gene which was present in the original HMW1 phage. *E. coli* transformed with this plasmid expressed major immunoreactive proteins with apparent molecular masses of 125 and 160 kDa as well as additional degradation products (Fig. 3, lanes 7 and 8). The 125- and 160-kDa bands were identical to the major and minor immunoreactive bands detected in the HMW1 phage lysates (compare Fig. 1 and 3). Interestingly, the pHMW1-14 construct also expressed significant amounts of protein in the uninduced condition, a situation not observed with the earlier constructs.

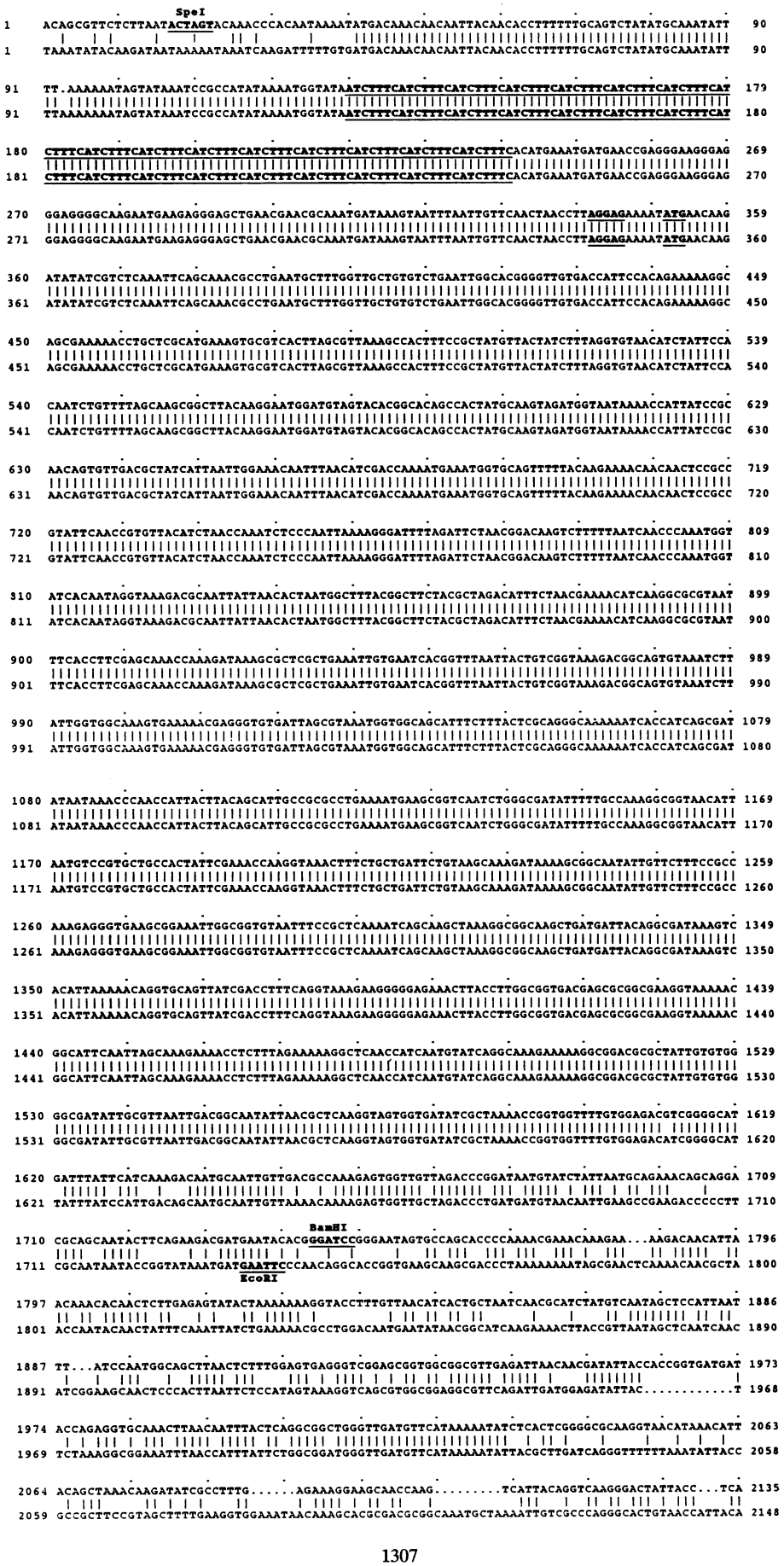
The relationship between the 125- and 160-kDa proteins remains somewhat unclear. Sequence analysis, described below, reveals that the HMW1 gene would be predicted to encode a protein of 159 kDa. Our speculation is that the 160-kDa protein is a precursor form of the mature 125-kDa protein, with the conversion from one protein to the other being dependent on the products of a downstream gene or genes.

Sequence analysis of the HMW1 and HMW2 genes. To further characterize the nature of the HMW1 and HMW2 gene products, the sequences of both genes were determined.

HMW1 gene. The nucleotide sequences of both strands of the HMW1 gene and flanking DNA were determined by subcloning restriction fragments into phages M13mp18 and M13mp19. Figure 4 shows the entire region sequenced and compares the HMW1 sequence with the HMW2 sequence by using the GAP program (5). Also indicated are several of the restriction sites used for restriction mapping and for generation of plasmid subclones. Sequence analysis of the HMW1 gene revealed a 4,608-bp open reading frame (ORF), beginning with an ATG codon at nucleotide 351 and ending with a TAG stop codon at nucleotide 4959. A putative ribosome-binding site with the sequence AGGAG begins 10 bp upstream of the putative initiation codon. Five other in-frame ATG codons are located within 250 bp of the beginning of the ORF, but none of these is preceded by a typical ribosome-binding site. The 5′-flanking region of the ORF contains a series of direct tandem repeats, with the 7-bp sequence ATCTTTC repeated 16 times. These tandem repeats stop 100 bp 5′ of the putative initiation codon. An 8-bp inverted repeat characteristic of a rho-independent transcriptional terminator (3, 5) is present, beginning at nucleotide 4983, 25 bp 3′ of the presumed translational stop. Multiple termination codons are present in all three reading frames both upstream and downstream of the ORF. The derived amino acid sequence of the protein encoded by the HMW1 gene has a molecular weight of 159,000, in good agreement with the apparent molecular weights of the proteins expressed by the HMW1-4 and HMW1-7 transformants (Fig. 3). The derived amino acid sequence of the amino terminus does not demonstrate the characteristics of a typical signal sequence (43). The *Bam*HI site used in generation of pHMW1 comprises bp 1743 through 1748 of the nucleotide sequence. The ORF downstream of the *Bam*HI site would be predicted to encode a protein of 111 kDa, in good agreement with the 115 kDa we estimated for the apparent molecular mass of the pHMW1-encoded fusion protein.

HMW2 gene. The nucleotide sequence of the HMW2 gene was similarly determined by subcloning restriction fragments from a representative HMW2 phage into M13mp18 and M13mp19 phages. The sequence of the HMW2 gene consists of a 4,431-bp ORF, beginning with an ATG codon at nucleotide 352 and ending with a TAG stop codon at nucleotide 4783 (Fig. 4). The first 1,259 bp of the ORF of the HMW2 gene are identical to those of the HMW1 gene. Thereafter, the sequences begin to diverge but are 80% identical overall. With the exception of a single base addition at nucleotide 93 of the HMW2 sequence, the 5′-flanking

FIG. 4. Complete nucleotide sequences of the HMW1 and HMW2 genes and DNA sequences 5′ and 3′ to the coding regions. The sequences were aligned by using the GAP program to identify regions of identity and nonidentity. The putative ribosome-binding sites correspond to nucleotides 341 to 345 in HMW1 and 342 to 346 in HMW2. A single ORF is present in the HMW1 sequence from the initiation codon at nucleotide 351 to the termination codon at nucleotide 4959. A single ORF is present in HMW2 from the initiation codon at nucleotide 352 to the termination codon at nucleotide 4783. The 5′-flanking regions of the two genes contain an identical set of direct tandem repeats from positions 129 through 240, with the sequence ATCTTTC repeated 16 times. Restriction sites of relevance used in restriction mapping, subcloning, and sequencing are shown in boldface.


SpeI
BamHI
EcoRI

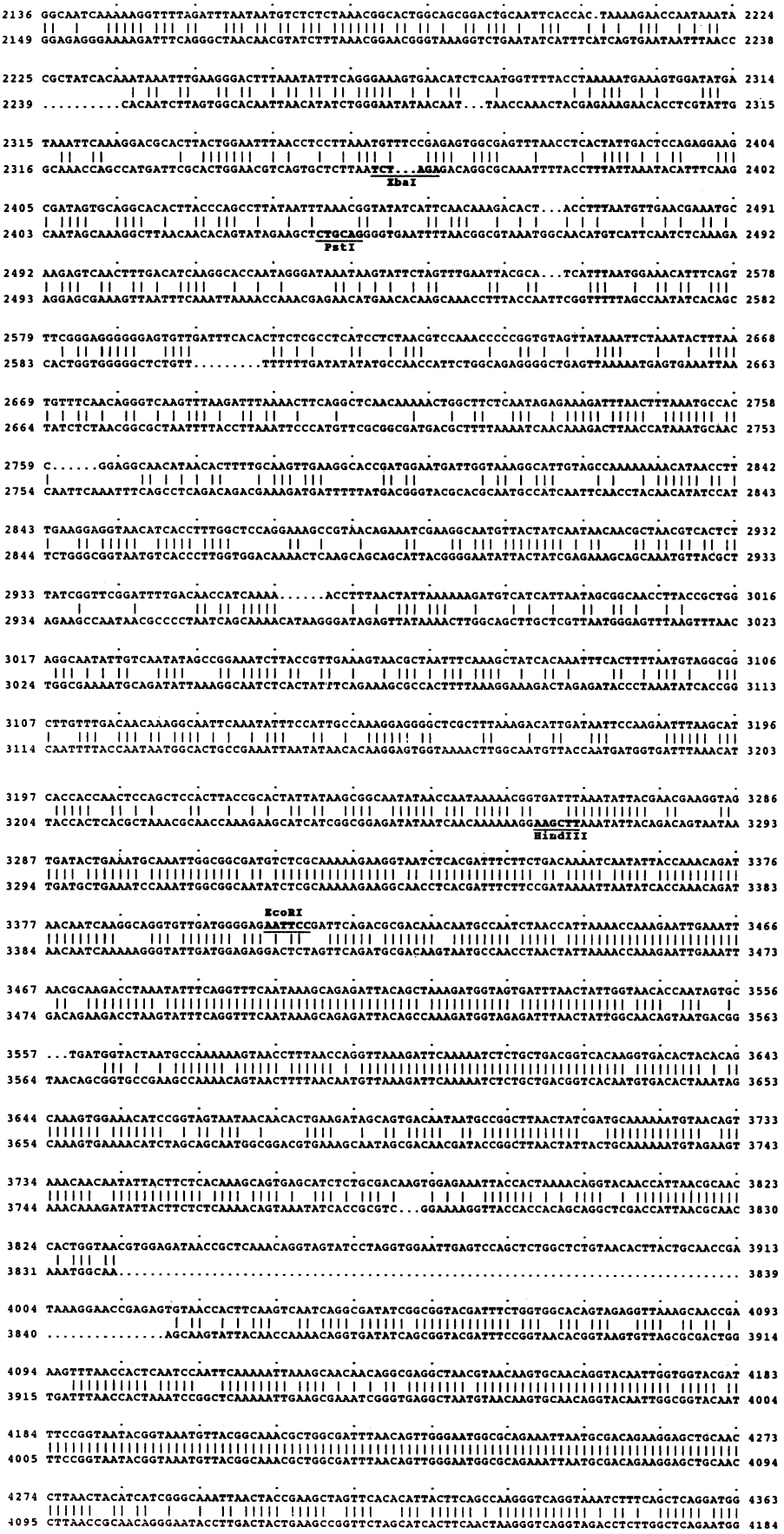


FIG. 4—*Continued.*

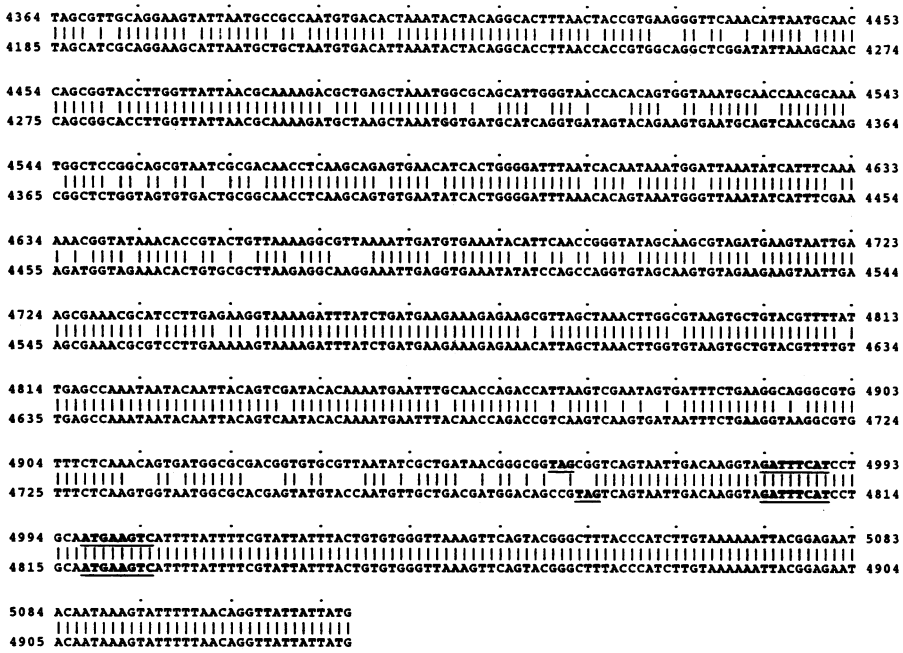


FIG. 4—*Continued.*

regions of the HMW1 and HMW2 genes are identical for 310 bp upstream from the respective initiation codons. Thus, the HMW2 gene is preceded by the same set of tandem repeats and the same putative ribosome-binding site which lie 5′ of the HMW1 gene. A putative transcriptional terminator identical to that identified 3′ of the HMW1 ORF is noted, beginning at nucleotide 4804. The discrepancy in the lengths of the two genes is principally accounted for by a 186-bp gap in the HMW2 sequence, beginning at nucleotide position 3839. The derived amino acid sequence of the protein encoded by the HMW2 gene has a molecular weight of 155,000 and is 71% identical with the derived amino acid sequence of the HMW1 gene. Figure 5 compares the derived amino acid sequences of the two genes by using the GAP program (5).

Protein sequence similarity search. The structure and function of the proteins encoded by the HMW1 and HMW2 genes are unknown. A protein sequence similarity search was performed with the derived amino acid sequences of the two genes by using the Lipman-Pearson TFASTA algorithm and the GenBank/EMBL nucleotide sequence library (5, 14). The derived amino acid sequences of both the HMW1 and HMW2 genes demonstrated sequence similarity with the derived amino acid sequence of filamentous hemagglutinin of *Bordetella pertussis*, a surface-associated protein of this organism (4, 6, 29). The initial and optimized TFASTA scores for the HMW1-filamentous hemagglutinin sequence comparison were 87 and 186, respectively, with a word size of 2. The *z* score for the comparison was 45.8. The initial and optimized TFASTA scores for the HMW2-filamentous hemagglutinin sequence comparison were 68 and 196, respectively. The *z* score for the latter comparison was 48.7. The magnitudes of the initial and optomized TFASTA scores and the *z* scores suggested that a biologically significant relationship existed between the HMW1 and HMW2 gene products and filamentous hemagglutinin (14). When the derived amino acid sequences of the HMW1, HMW2, and filamentous hemagglutinin genes were aligned and compared, the similarities were most notable at the amino-terminal ends of the three sequences. Twelve of the first 22 amino acids in the predicted peptide sequences were identical. In addition, the sequences demonstrated a common five-amino-acid stretch, Asn-Pro-Asn-Gly-Ile, and several shorter stretches of sequence identity within the first 200 amino acids.

Reactivity of HMW1 recombinant-protein antiserum with purified filamentous hemagglutinin. To further explore the HMW1-filamentous hemagglutinin relationship, we assessed the ability of antiserum prepared against the HMW1-4 recombinant protein (rHMW1) to recognize purified filamentous hemagglutinin (Fig. 6). As can be seen, the rHMW1 antiserum demonstrated ELISA reactivity with filamentous hemagglutinin in a dose-dependent manner. Preimmune rabbit serum had minimal reactivity in this assay. The rHMW1 antiserum was also examined in a Western blot assay and demonstrated weak but positive reactivity with purified filamentous hemagglutinin in this system also (data not shown).

Reactivity of HMW1 recombinant-protein antiserum with homologous and heterologous nontypeable *H. influenzae* strains. To identify the native *Haemophilus* protein corresponding to the HMW1 gene product and to determine the extent to which proteins antigenically related to the HMW1 cloned gene product were common among other nontypeable *H. influenzae* strains, we screened a panel of *Haemophilus* strains by Western blot with the rHMW1 antiserum. The antiserum recognized both a 125- and a 120-kDa protein band in the homologous strain 12 (Fig. 7), the putative mature protein products of the HMW1 and HMW2 genes, respectively. The 120-kDa protein appears as a single band in Fig. 7, whereas it appeared as a doublet in the HMW2 phage lysates (Fig. 1). The reasons for this discrepancy are unclear at present.

When used to screen heterologous nontypeable *H. influenzae* strains, rHMW1 antiserum recognized high-molecular-weight proteins in 75% of 125 epidemiologically unrelated strains. In general, the antiserum reacted with one or two

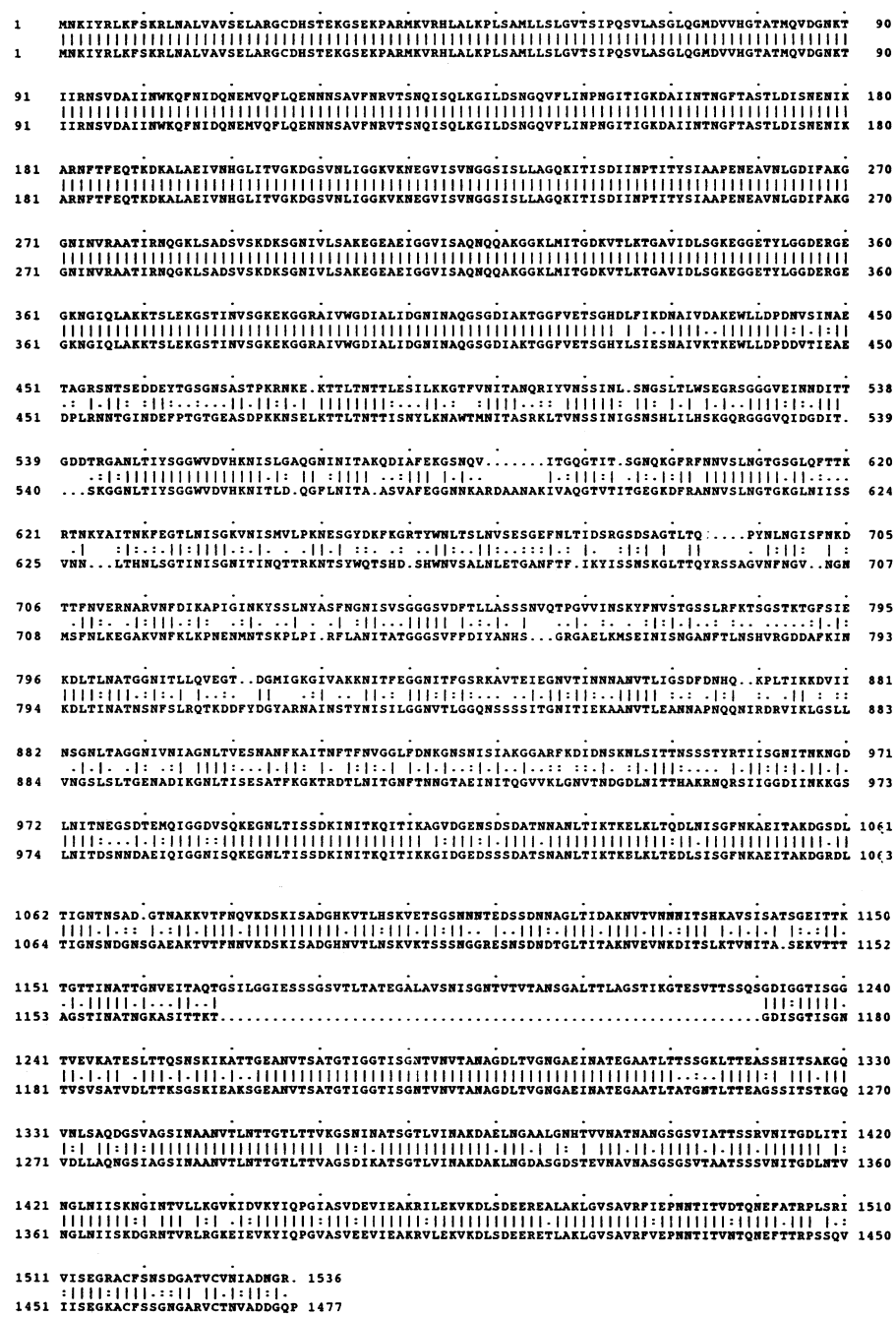


FIG. 5. Derived amino acid sequences of the HMW1 and HMW2 genes, aligned and compared by using the GAP program. | , identical amino acids; : and ., similar amino acids (5).

protein bands in the 100- to 150-kDa range in each of the heterologous strains in a pattern similar but not identical to that seen in the homologous strain (Fig. 7).

Reactivity of a monoclonal antibody to filamentous hemagglutinin with nontypeable *H. influenzae*. Monoclonal antibody X3C is a murine IgG antibody directed against the filamentous hemagglutinin protein of *B. pertussis* (13). This monoclonal antibody will prevent adhesion of *B. pertussis* cells to Chinese hamster ovary cells and HeLa cells in culture and will inhibit hemagglutination of erythrocytes by purified filamentous hemagglutinin. We were interested in knowing whether the *Haemophilus* high-molecular-weight proteins expressed epitopes recognized by this monoclonal antibody. Figure 8 shows the results of a Western blot assay in which this monoclonal antibody was screened against the same panel of nontypeable *H. influenzae* strains examined in Fig. 7. Monoclonal antibody X3C recognized both the high-molecular-weight proteins in nontypeable *H. influenzae* strain 12 which were recognized by the recombinant-protein antiserum. In addition, the monoclonal antibody recognized protein bands in a subset of heterologous nontypeable *H. influenzae* strains which were identical to those recognized

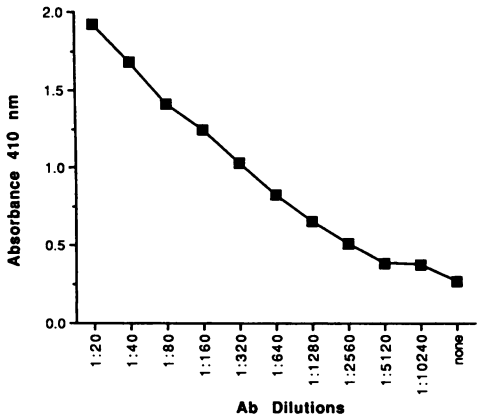


FIG. 6. ELISA with rHMW1 antiserum assayed against purified filamentous hemagglutinin of *B. pertussis*. Ab, antibody.

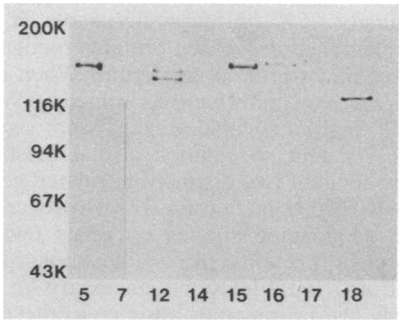


FIG. 8. Western immunoblot assay of cell sonicates from a panel of epidemiologically unrelated nontypeable *H. influenzae* strains. The sonicates were probed with monoclonal antibody X3C, a murine IgG antibody which recognizes the filamentous hemagglutinin of *B. pertussis* (13). The strain designations are indicated by the numbers below each lane.

by the recombinant-protein antiserum (compare Fig. 7 and 8). On occasion, the filamentous hemagglutinin monoclonal antibody appeared to recognize only one of the two bands which had been recognized by the recombinant-protein antiserum (compare strain 18 lanes in Fig. 7 and 8, for example). Overall, monoclonal antibody X3C recognized high-molecular-weight protein bands identical to those recognized by the rHMW1 antiserum in approximately 35% of our collection of nontypeable *H. influenzae* strains.

DISCUSSION

We previously identified a group of high-molecular-weight surface-exposed proteins of nontypeable *H. influenzae* which are major targets of antibodies present in the sera of children convalescing from acute *Haemophilus* otitis and in the sera of healthy adults (1). In an effort to further characterize these proteins, we have cloned and sequenced the structural genes for two immunogenic high-molecular-weight proteins from a prototype nontypeable *H. influenzae* strain, proteins we have designated HMW1 and HMW2. A protein sequence similarity search performed with the derived amino acid sequences of the HMW1 and HMW2 genes revealed sequence similarity with the derived amino acid sequence of the filamentous hemagglutinin protein of *B. pertussis*. Antiserum raised against HMW1 recombinant protein reacted with purified filamentous hemagglutinin in ELISA and Western blot assays. In addition, a monoclonal antibody directed against filamentous hemagglutinin reacted with high-molecular-weight proteins in a panel of nontypeable *H. influenzae* strains which were identical to those recognized by the recombinant-protein antiserum. These data raise the possibility that the HMW1 and HMW2 gene products are structural and functional analogs of the filamentous hemagglutinin protein of *B. pertussis*. Filamentous hemagglutinin is a surface-associated protein which appears to have a central role in adherence (29, 38, 39, 41). It also is highly immunogenic in humans during natural disease (10) and is a potent protective immunogen in some animal models of pertussis infection (12, 26).

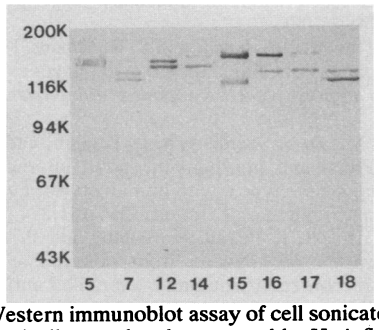


FIG. 7. Western immunoblot assay of cell sonicates from a panel of epidemiologically unrelated nontypeable *H. influenzae* strains. The sonicates were probed with rabbit antiserum prepared against HMW1-4 recombinant protein. The strain designations are indicated by the numbers below each lane.

The adherence factors of *H. influenzae* have not yet been fully characterized. Although surface pili mediate adherence under certain circumstances, non-pilus-mediated adherence of *H. influenzae* has been described in several different in vitro systems (8, 15, 28, 31, 34). Whether the HMW1 and HMW2 gene products have a role in promoting the adherence of *H. influenzae* is unknown at present. Comparative analysis of the derived amino acid sequences of the filamentous hemagglutinin, HMW1, and HMW2 genes could provide some insight into this question. At least two distinct eucaryotic cell-binding sites have been described for filamentous hemagglutinin (30, 38). One appears to be a carbohydrate-binding site, but its localization on the filamentous hemagglutinin molecule has not yet been precisely defined. The other appears to reside in or around an Arg-Gly-Asp (RGD) sequence (30). No RGD sequences were found in the HMW1-derived amino acid sequence, but the HMW2 sequence contains a predicted RGD sequence from amino acids 785 to 787.

The amino-terminal portion of filamentous hemagglutinin demonstrates substantial homology with the amino-terminal regions of the hemolysins A of *Serratia marcescens* and *Proteus mirabilis* (4, 27, 40). It has been suggested that the region of homology may be involved in binding of these three proteins to eucaryotic cell receptors (4). The derived amino acid sequences of the two hemolysin genes also demonstrate sequence similarity with the derived amino acid sequences of the HMW1 and HMW2 genes when compared by the TFASTA program, but the scores are lower than those observed with the filamentous hemagglutinin comparison. One finding of potential interest is that the two hemolysins have the same five-amino-acid sequence near their amino termini, Asn-Pro-Asn-Gly-Ile, which was noted to be common to the HMW1, HMW2, and filamentous hemagglutinin sequences.

Approximately 75% of the nontypeable *H. influenzae* isolates we examined expressed proteins recognized by the HMW1 recombinant-protein antiserum. When sonicates of these cross-reactive strains were examined by Western blot, two discrete high-molecular-weight bands were usually present (Fig. 7). Our speculation is that these two bands represent products of two distinct but related genes, as was the case with prototype strain 12. However, additional cloning and sequencing studies of genes encoding high-molecular-weight proteins from other nontypeable *Haemophilus* strains will likely be necessary to clarify the relationships. The presence or absence of these proteins in type b *H. influenzae* will also be important to define. The recombinant-protein antiserum did not appear to recognize antigenically related high-molecular-weight proteins in a small number of type b strains we have examined to date. However, when the HMW1 gene and associated 3′-flanking region were used to probe genomic DNA from a panel of type b strains by Southern hybridization, weak but definite hybridization signals were observed (unpublished observations).

The nucleotide sequences of the HMW1 and HMW2 genes and flanking DNA were notable for several reasons. Most provocative was the presence of the 7-bp direct tandem repeats in the regions immediately upstream of the putative initiation codons. This finding is reminiscent of the situation in *Vibrio cholerae*, in which a series of three to eight direct repeats of the 7-bp sequence TTTTGAT are present in the promoter region of the cholera toxin gene (19, 20). In *V. cholerae*, this repeat region is believed to serve as a binding site for the ToxR protein in a step necessary for transcriptional activation of the cholera toxin promoter (20). Whether the set of tandem repeats in *H. influenzae* serve a similar function in terms of transcriptional control of the HMW1 and HMW2 genes is unknown at present. Preliminary data suggest that the tandem repeat sequence is highly conserved among other nontypeable *H. influenzae* organisms. As noted above, approximately 75% of unrelated nontypeable *H. influenzae* strains have proteins antigenically related to the HMW1 gene product. When genomic DNA from a panel of such strains was probed by Southern hybridization with a 21-mer oligonucleotide consisting of three serial repeats of the 7-bp *Haemophilus* sequence, each demonstrated two to four strong hybridization signals (unpublished observation). Our speculation is that these signals represent hybridization with tandem repeats upstream of copies of HMW1-related genes in the heterologous strains. Interestingly, DNA from strains nonreactive on Western blot were similarly nonreactive on Southern hybridization with the above-mentioned oligonucleotide.

In the subcloning studies of the HMW1 gene, we presented data suggesting that the HMW1 gene product was processed in some manner by the proteins encoded by downstream genes. Precedent for such processing exists in studies of filamentous hemagglutinin (6). Translation of the entire ORF of the filamentous hemagglutinin gene predicts the expression of a polypeptide with a molecular mass of 367 kDa, yet the mature polypeptide is estimated to be approximately 220 kDa in size. The precise events which occur during processing of filamentous hemagglutinin are unknown (6, 35), but the fact that filamentous hemagglutinin is processed from a large precursor supports our contention that the HMW1 and HMW1 gene products may also be processed prior to generation of the mature protein. Of note in this regard, protocols were recently developed in our laboratory for partial purification of the native 125- and 120-kDa proteins of strain 12. N-terminal amino acid sequence analysis revealed the sequences --NV-INAETAG and P--VTIE-AEDPL for the 125- and 120-kDa proteins, respectively. These sequences correspond to the derived amino acid sequences of the HMW1 and HMW2 genes, each beginning at peptide 442 (Fig. 5). Thus, it appears that lengthy segments are being removed from the amino termini of the HMW1 and HMW2 gene products in generation of the mature proteins.

The vaccine potential of the nontypeable *H. influenzae* high-molecular-weight proteins has yet to be determined. The proteins are surface exposed and highly immunogenic, and antigenically related proteins are present in the majority of nontypeable *H. influenzae* strains. However, the epitopes on the proteins that are potentially capable of interacting with protective antibodies have yet to be defined, and the degree to which such epitopes are common among strains is unknown. Such issues will be critical to address in future studies.

ACKNOWLEDGMENTS

We thank Denise Jadali and Linda Winter for expert technical assistance. We thank Michael J. Brennan, Theresa M. Finn, Robert S. Munson, Jr., and Dan M. Granoff for critical review of the manuscript. We also thank Robert S. Munson, Jr., for thoughtful discussions and guidance throughout the project.

This work was supported in part by Public Health Service grant AI21707 and Thrasher Research Fund grant 2804-2.

REFERENCES

1. **Barenkamp, S. J., and F. F. Bodor.** 1990. Development of serum bactericidal activity following nontypable *Haemophilus influenzae* acute otitis media. Pediatr. Infect. Dis. J. **9:**333–339.
2. **Barnes, W. M.** 1987. Sequencing DNA with deoxyribonucleotides as chain terminators: hints and strategies for big projects. Methods Enzymol. **152:**538–556.
3. **Brendel, V., and E. N. Trifonov.** 1984. A computer algorithm for testing potential prokaryotic terminators. Nucleic Acids Res. **12:**4411–4427.
4. **Delisse-Gathoye, A.-M., C. Locht, F. Jacob, M. Raaschou-Nielsen, I. Heron, J.-L. Ruelle, M. DeWilde, and T. Cabezon.** 1990. Cloning, partial sequence, expression, and antigenic analysis of the filamentous hemagglutinin gene of *Bordetella pertussis*. Infect. Immun. **58:**2895–2905.
5. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
6. **Domenighini, M., D. Relman, C. Capiau, S. Falkow, A. Prugnola, V. Scarlato, and R. Rappuoli.** 1990. Genetic characterization of *Bordetella pertussis* filamentous haemagglutinin: a protein processed from an unusually large precursor. Mol. Microbiol. **4:**787–800.
7. **Ely, S., J. Tippett, J. S. Kroll, and E. R. Moxon.** 1986. Mutations affecting expression and maintenance of genes encoding the serotype b capsule of *Haemophilus influenzae*. J. Bacteriol. **167:**44–48.
8. **Farley, M. M., D. S. Stephens, S. L. Kaplan, and E. O. Mason, Jr.** 1990. Pilus- and non-pilus-mediated interactions of *Haemophilus influenzae* type b with human erythrocytes and human nasopharyngeal mucosa. J. Infect. Dis. **161:**274–280.
9. **Gnehm, H. E., S. I. Pelton, S. Gulati, and P. A. Rice.** 1985. Characterization of antigens from nontypable *Haemophilus influenzae* recognized by human bactericidal antibodies: role of *Haemophilus* outer membrane proteins. J. Clin. Invest. **75:**1645–1658.
10. **Hallander, H. O., J. Storsaeter, and R. Mollby.** 1991. Evaluation of serology and nasopharyngeal cultures for diagnosis of pertussis in a vaccine efficacy trial. J. Infect. Dis. **163:**1046–1054.

11. **Kilian, M.** 1985. *Haemophilus*, p. 387–393. *In* E. H. Lennette, A. Balows, W. J. Hausler, Jr., and H. J. Shadomy (ed.), Manual of clinical microbiology, 4th ed. American Society for Microbiology, Washington, D.C.
12. **Kimura, A., K. T. Mountzouros, D. A. Relman, S. Falkow, and J. L. Cowell.** 1990. *Bordetella pertussis* filamentous hemagglutinin: evaluation as a protective antigen and colonization factor in a mouse respiratory infection model. Infect. Immun. **58:**7–16.
13. **Leininger, E., J. G. Kenimer, and M. J. Brennan.** 1990. Surface proteins of *Bordetella pertussis*: role in adherence, p. 100–104. *In* C. R. Manclark (ed.), Proceedings of the Sixth International Symposium on Pertussis. Department of Health and Human Services, United States Public Health Service, Bethesda, Md.
14. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227:**1435–1441.
15. **Loeb, M. R., E. Connor, and D. Penney.** 1988. A comparison of the adherence of fimbriated and nonfimbriated *Haemophilus influenzae* type b to human adenoids in organ culture. Infect. Immun. **56:**484–489.
16. **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.
17. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
18. **Marmur, J.** 1961. A procedure for the isolation of deoxyribonucleic acid from micro-organisms. J. Mol. Biol. **3:**208–218.
19. **Miller, V. L., and J. J. Mekalanos.** 1984. Synthesis of cholera toxin is positively regulated at the transcriptional level by *toxR*. Proc. Natl. Acad. Sci. USA **81:**3471–3475.
20. **Miller, V. L., R. K. Taylor, and J. J. Mekalanos.** 1987. Cholera toxin transcriptional activator *toxR* is a transmembrane DNA binding protein. Cell **48:**271–279.
21. **Moffatt, B. A., and F. W. Studier.** 1987. T7 lysozyme inhibits transcription by T7 RNA polymerase. Cell **49:**221–227.
22. **Murphy, T. F., and M. A. Apicella.** 1987. Nontypable *Haemophilus influenzae*: a review of clinical aspects, surface antigens, and the human immune response to infection. Rev. Infect. Dis. **9:**1–15.
23. **Murphy, T. F., and L. C. Bartos.** 1988. Human bactericidal antibody response to outer membrane protein P2 of nontypeable *Haemophilus influenzae*. Infect. Immun. **56:**2673–2679.
24. **Murphy, T. F., L. C. Bartos, P. A. Rice, M. B. Nelson, K. C. Dudas, and M. A. Apicella.** 1986. Identification of a 16,600-dalton outer membrane protein on nontypable *Haemophilus influenzae* as a target for human serum bactericidal antibody. J. Clin. Invest. **78:**1020–1027.
25. **Musser, J. M., S. J. Barenkamp, D. M. Granoff, and R. K. Selander.** 1986. Genetic relationships of serologically nontypeable and serotype b strains of *Haemophilus influenzae*. Infect. Immun. **52:**183–191.
26. **Oda, M., J. L. Cowell, D. G. Burstyn, and C. R. Manclark.** 1984. Protective activities of the filamentous hemagglutinin and the lymphocytosis-promoting factor of *Bordetella pertussis* in mice. J. Infect. Dis. **150:**823–833.
27. **Poole, K., E. Schiebel, and V. Braun.** 1988. Molecular characterization of the hemolysin determinant of *Serratia marcescens*. J. Bacteriol. **170:**3177–3188.
28. **Read, R. C., R. Wilson, A. Rutman, V. Lund, H. C. Todd, A. P. R. Brain, P. K. Jeffrey, and P. J. Cole.** 1991. Interaction of nontypable *Haemophilus influenzae* with human respiratory mucosa *in vitro*. J. Infect. Dis. **163:**549–558.
29. **Relman, D. A., M. Domenighini, E. Tuomanen, R. Rappuoli, and S. Falkow.** 1989. Filamentous hemagglutinin of *Bordetella pertussis*: nucleotide sequence and crucial role in adherence. Proc. Natl. Acad. Sci. USA **86:**2637–2641.
30. **Relman, D., E. Tuomanen, S. Falkow, D. T. Golenbock, K. Saukkonen, and S. D. Wright.** 1990. Recognition of a bacterial adhesin by an integrin: macrophage CR3 ($\alpha_M\beta_2$,CD11b/CD18) binds filamentous hemagglutinin of *Bordetella pertussis*. Cell **61:**1375–1382.
31. **Sable, N. S., E. M. Connor, C. B. Hall, and M. R. Loeb.** 1985. Variable adherence of fimbriated *Haemophilus influenzae* type b to human cells. Infect. Immun. **48:**119–123.
32. **Silhavy, T. J., M. L. Berman, and L. W. Enquist.** 1984. Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
33. **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.
34. **St. Geme, J. W., III, and S. Falkow.** 1990. *Haemophilus influenzae* adheres to and enters cultured human epithelial cells. Infect. Immun. **58:**4036–4044.
35. **Stibitz, S., A. A. Weiss, and S. Falkow.** 1988. Genetic analysis of a region of the *Bordetella pertussis* chromosome encoding filamentous hemagglutinin and the pleiotropic regulatory locus *vir*. J. Bacteriol. **170:**2904–2913.
36. **Studier, F. W., and B. A. Moffatt.** 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. **189:**113–130.
37. **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.
38. **Tuomanen, E., H. Towbin, G. Rosenfelder, D. Braun, G. Larson, G. C. Hansson, and R. Hill.** 1988. Receptor analogs and monoclonal antibodies that inhibit adherence of *Bordetella pertussis* to human ciliated respiratory epithelial cells. J. Exp. Med. **168:**267–277.
39. **Tuomanen, E., and A. Weiss.** 1985. Characterization of two adhesins of *Bordetella pertussis* for human ciliated respiratory-epithelial cells. J. Infect. Dis. **152:**118–125.
40. **Uphoff, T. S., and R. A. Welch.** 1990. Nucleotide sequencing of the *Proteus mirabilis* calcium-independent hemolysin genes (*hpmA* and *hpmB*) reveals sequence similarity with the *Serratia marcescens* hemolysin genes (*shlA* and *shlB*). J. Bacteriol. **172:**1206–1216.
41. **Urisu, A., J. L. Cowell, and C. R. Manclark.** 1986. Filamentous hemagglutinin has a major role in mediating adherence of *Bordetella pertussis* to human WiDr cells. Infect. Immun. **52:**695–701.
42. **Wallace, R. J., Jr., D. M. Musher, E. J. Septimus, J. E. McGowan, Jr., F. J. Quinones, K. Wiss, P. H. Vance, and P. A. Trier.** 1981. *Haemophilus influenzae* infection in adults: characterization of strains by serotypes, biotypes, and beta-lactamase production. J. Infect. Dis. **144:**101–106.
43. **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.