JOURNAL OF BACTERIOLOGY, Nov. 1996, p. 6281–6287 Vol. 178, No. 21
0021-9193/96/$04.00+0

Characterization of the Genetic Locus Encoding *Haemophilus influenzae* Type b Surface Fibrils

JOSEPH W. ST. GEME III,[1]* DAVID CUTTER,[1] AND STEPHEN J. BARENKAMP[2]

Edward Mallinckrodt Department of Pediatrics and Department of Molecular Microbiology, Washington University School of Medicine, and Division of Infectious Diseases, St. Louis Children's Hospital, St. Louis, Missouri 63110,[1] and Department of Pediatrics, St. Louis University School of Medicine, and The Pediatric Research Institute, Cardinal Glennon Children's Hospital, St. Louis, Missouri 63104[2]

Received 1 March 1996/Accepted 27 August 1996

***Haemophilus influenzae* is a common gram-negative pathogen that initiates infection by colonizing the upper respiratory tract epithelium. In previous work, we reported the isolation of a locus involved in expression of short, thin surface fibrils by *H. influenzae* type b and presented evidence that surface fibrils promote attachment to human epithelial cells. In the present study, we determined that the fibril locus is composed of one long open reading frame, designated *hsf*, which encodes a protein (Hsf) with a molecular mass of ~240 kDa. The derived amino acid sequence of the *hsf* product demonstrated 81% similarity and 72% identity to a recently identified nontypeable *H. influenzae* adhesin referred to as Hia. In experiments with a panel of eight cultured cell lines, the Hsf and Hia proteins were found to confer the same binding specificities, suggesting that *hsf* and *hia* are alleles of the same locus. Southern analysis and mutagenesis studies reinforced this conclusion. Further investigation revealed that an *hsf* homolog is ubiquitous among encapsulated *H. influenzae* strains and is present in a subset of nontypeable *Haemophilus* strains as well. We speculate that the *hsf* gene product plays an important role in the process of respiratory tract colonization by *H. influenzae*.**

Haemophilus influenzae is a gram-negative bacterium that is frequently associated with human disease (43). In recent years, *H. influenzae* has been the leading cause of bacterial meningitis and a common cause of other bacteremic diseases (10). In addition, this organism accounts for a large proportion of otitis media, sinusitis, bronchitis, and pneumonia (31).

Isolates of *H. influenzae* are classified according to their polysaccharide capsule. There are six antigenically distinct capsular types, designated a through f (26). In addition, isolates can be nonencapsulated; these strains are defined by their failure to react with typing antisera directed against the known capsular types and are referred to as nontypeable (26). Most isolates recovered from patients with bacteremic illness express the serotype b capsule (42). In contrast, the majority of strains associated with localized respiratory tract infection are nontypeable (42).

The pathogenesis of disease due to *H. influenzae* is believed to begin with colonization of the upper respiratory tract mucosa (22). Both type b and nontypeable *H. influenzae* strains express pili, which promote agglutination of human erythrocytes and attachment to oropharyngeal epithelial cells (9, 16, 17, 25). In support of a role for pili in natural colonization, Weber and colleagues found that elimination of piliation resulted in a decreased density of nasopharyngeal colonization in 1-year-old monkeys (46). A number of reports suggest that nonpilus factors also facilitate *Haemophilus* colonization (13, 14, 21, 28, 29, 36–39, 47). Of particular note, in the study by Weber et al., nonpiliated organisms retained a capacity for colonization, although at reduced densities (46). In a recent study, we examined a series of *H. influenzae* type b isolates by transmission electron microscopy and visualized short, thin surface fibrils that were distinct from pili (33). In the same study, we reported the isolation of a large genetic locus involved in expression of these appendages and presented evidence that fibrils promote adherence to cultured human epithelial cells (33).

In the present study, we have characterized the *H. influenzae* fibril locus. Our results indicate that this locus consists of one long open reading frame, which encodes an ~240-kDa polypeptide with striking similarity to a recently described protein involved in in vitro adherence by a subset of nontypeable *H. influenzae* strains. Furthermore, a homolog of the fibril locus appears to be universally present among encapsulated strains of *H. influenzae*.

MATERIALS AND METHODS

Bacterial strains and plasmids. *H. influenzae* C54 is a type b strain originally described by Pichichero et al. (25). Strain C54b$^-$p$^-$ is a capsule-deficient, nonpiliated mutant that expresses fibrils and demonstrates efficient adherence to cultured epithelial cells (33). Strain C54-Tn*400*.23 is a derivative of C54b$^-$p$^-$ that contains a mini-Tn*10 kan* element in the *hsf* locus and demonstrates minimal in vitro adherence (33). Strains 1053, 1058, 1060, 1063, 1065, 1069, 1070, 1076, 1081, and 1084 are *H. influenzae* type b isolates generously provided by J. Musser (Baylor University, Houston, Tex.) (23). *H. influenzae* SM4 (type a), SM6 (type d), SM7 (type e), and SM72 (type c) are type strains obtained from R. Facklam at the Centers for Disease Control and Prevention (Atlanta, Ga.). Strains 142, 327, and 351 are *H. influenzae* type e isolates and strains 134, 219, 256, and 501 are *H. influenzae* type f isolates obtained from H. Kayhty (Finnish National Public Health Institute, Helsinki). Strain Rd (type d) and the 15 nontypeable isolates examined by Southern analysis have been described previously (1, 4). *Escherichia coli* DH5α is a nonadherent laboratory strain that was originally obtained from Gibco BRL. *E. coli* BL21(DE3) was a gift from F. W. Studier and contains a single copy of the T7 RNA polymerase gene under the control of the *lac* regulatory system (40). Plasmid pT7-7 was provided by S. Tabor and 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 (41). pUC19 is a high-copy-number plasmid that has been previously described (49). pDC400 is a pUC19 derivative that harbors the *H. influenzae* C54 surface fibril locus and is sufficient to promote in vitro adherence by laboratory strains of *E. coli* (33). pHMW8-5 is a pT7-7 derivative that contains the *H. influenzae* 11 *hia* locus and also promotes adherence by nonadherent laboratory strains of *E. coli* (6). pHMW8-6 contains the *H. influenzae hia* locus interrupted

* Corresponding author. Mailing address: Department of Molecular Microbiology, Washington University School of Medicine, 660 South Euclid Ave., Box 8230, St. Louis, MO 63110. Phone: (314) 362-5401. Fax: (314) 362-1232. Electronic mail address: stgeme@borcim.wustl.edu.

by a kanamycin cassette (6). pUC4K served as the source of the kanamycin resistance gene that was used as a probe in Southern analysis (44).

Culture conditions. *H. influenzae* strains were grown on chocolate agar supplemented with 1% IsovitaleX, on brain heart infusion agar supplemented with hemin and NAD, or in brain heart infusion broth supplemented with hemin and NAD (3). These strains were stored at −80°C in brain heart infusion broth with 25% glycerol. *E. coli* strains were grown on Luria-Bertani agar or in Luria-Bertani broth and were stored at −80°C in Luria-Bertani broth with 50% glycerol. For *H. influenzae*, kanamycin was used at a concentration of 25 μg/ml. Antibiotic concentrations for *E. coli* were as follows: ampicillin or carbenicillin at 100 μg/ml and kanamycin at 50 μg/ml.

Induction of plasmid-encoded proteins. To identify plasmid-encoded proteins, the bacteriophage T7 expression vector pT7-7 was employed and the relevant pT7-7 derivatives were transformed into *E. coli* BL21(DE3). Activation of the T7 promoter was achieved by inducing expression of T7 RNA polymerase with isopropyl-β-D-thiogalactopyranoside (final concentration, 1 mM). After induction for 30 min at 37°C, rifampin was added to a final concentration of 200 μg/ml. Thirty minutes later, 1 ml of culture was pulsed with 50 μCi of Trans^{35}S-label (ICN, Irvine, Calif.) for 5 min. Bacteria were harvested, and whole-cell lysates were resuspended in Laemmli buffer for analysis by sodium dodecyl sulfate-polyacrylamide gel electrophoresis on 7.5% acrylamide gels (20). Autoradiography was performed with Kodak XAR-5 film.

Recombinant DNA methods. DNA ligations, restriction endonuclease digestions, and gel electrophoresis were performed by standard techniques (30). Plasmids were introduced into *E. coli* strains by either chemical transformation or electroporation, as previously described (12, 30). Transformation in *H. influenzae* was performed by the M-IV method of Herriott et al. (18).

Nucleotide sequencing. Nucleotide sequence was determined with a Sequenase kit and double-stranded plasmid template. DNA fragments were subcloned into pUC19 and sequenced along both strands by primer walking. DNA sequence analysis was performed using the Genetics Computer Group software package from the University of Wisconsin (11). Sequence similarity searches were carried out with the BLAST program of the National Center for Biotechnology Information (2).

Adherence assays. Adherence assays were performed with tissue culture cells which were seeded into wells of 24-well tissue culture plates as previously described (37). Adherence was measured after incubating bacteria with epithelial cell monolayers for 30 min as previously described (39). Tissue culture cells included Chang epithelial cells (Wong-Kilbourne derivative, clone 1-5c-4 [human conjunctiva]) (ATCC CCL 20.2), KB cells (human oral epidermoid carcinoma) (ATCC CCL 17), HEp-2 cells (human laryngeal epidermoid carcinoma) (ATCC CCL 23), A549 cells (human lung carcinoma) (ATCC CCL 185), Intestine 407 cells (human embryonic intestine) (ATCC CCL 6), HeLa cells (human cervical epitheloid carcinoma) (ATCC CCL 2), ME-180 cells (human cervical epidermoid carcinoma) (ATCC HTB 33), HEC-IB cells (human endometrium) (ATCC HTB 113), and CHO-K1 cells (Chinese hamster ovary) (ATCC CCL 61). Chang, KB, Intestine 407, HeLa, and HEC-IB cells were maintained in modified Eagle medium with Earle's salts and nonessential amino acids. HEp-2 cells were maintained in Dulbecco's modified Eagle medium, A549 cells and CHO-K1 cells were maintained in F-12 medium (Ham), and ME-180 cells were maintained in McCoy 5A medium. All media were supplemented with 10% heat-inactivated fetal bovine serum.

Southern analysis. Southern blotting was performed under high-stringency conditions as previously described (38).

Microscopy. Samples of epithelial cells with associated bacteria were stained with Giemsa stain and examined by light microscopy as previously described (37).

For negative-staining electron microscopy, bacteria were stained with 0.5% aqueous uranyl acetate (38) and examined with a Zeiss 10A microscope.

Nucleotide sequence accession number. The DNA sequence described here has been deposited in the GenBank library and has been assigned accession number U41852.

RESULTS

Initial characterization of the cloned fibril locus. In a previous study, we demonstrated that laboratory *E. coli* strains harboring the plasmid pDC400 were capable of efficient attachment to cultured human epithelial cells (33). Subcloning studies and transposon mutagenesis indicated that the relevant coding region of pDC400 was present within an 8.3-kb *Xba*I fragment (33) (Fig. 1). To confirm this conclusion, in the present study this *Xba*I fragment was subcloned into pT7-7, generating plasmids designated pDC601 and pDC602, which contained the insert in opposite orientations (Fig. 1). As predicted, expression of these plasmids in *E. coli* DH5α was associated with a capacity for high-level in vitro attachment (Table 1).

To determine the direction of transcription and identify

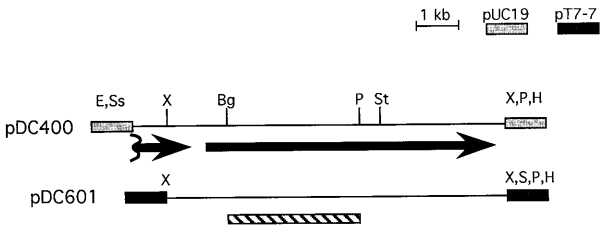

FIG. 1. Restriction map of pDC400 and its derivatives. pDC400 contains a 9.1-kb insert from strain C54 cloned into pUC19. Vector sequences are represented by shaded boxes. Letters above the top horizontal line indicate restriction enzyme sites: Bg, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; S, *Sal*I; Ss, *Sst*I; St, *Sty*I; X, *Xba*I. The longer dark arrow represents the location of the *hsf* locus within pDC400 and the direction of transcription; the shorter arrow represents the 3′ end of the putative exoribonuclease II gene. The striated horizontal line represents the 3.3-kb intragenic fragment used as a probe for Southern analysis. The plasmid pDC602, which is not shown, contains the same insert as pDC601 but in the opposite orientation.

plasmid-encoded proteins, pDC601 and pDC602 were introduced into *E. coli* BL21(DE3), producing BL21(DE3)/pDC601 and BL21(DE3)/pDC602, respectively. As a negative control, pT7-7 was also transformed into BL21(DE3). As shown in Fig. 2, induction of BL21(DE3)/pDC601 resulted in increased expression of a large protein, over 200 kDa in size, along with several slightly smaller proteins, which presumably represent degradation products. In contrast, when BL21(DE3)/pDC602 and BL21(DE3)/pT7-7 were induced, there was no expression of these proteins. This experiment indicated that the genetic material contained in the 8.3-kb *Xba*I fragment is transcribed from left to right, as shown in Fig. 1, and suggested that a single long open reading frame may be present.

Nucleotide sequence. Sequencing of the 8.3-kb *Xba*I fragment revealed a 7,059-bp gene, which we have designated *hsf* (for *Haemophilus* surface fibrils). This gene encodes a 2,353-amino-acid polypeptide, referred to as Hsf, with a calculated molecular mass of 243.8 kDa, which is similar in size to the observed protein species detected after induction of BL21(DE3)/pDC601. The *hsf* gene has a G+C content of 42.8%, somewhat greater than the published estimate of 38 to 39% for the whole genome (15, 19). A putative ribosomal binding site with the sequence AAGGTA begins 7 bp upstream of the presumed initiation codon. A sequence similar to a *rho*-independent transcription terminator is present beginning 20 nucleotides beyond the stop codon and contains interrupted inverted repeats with the potential for forming a hairpin struc-

TABLE 1. Adherence to Chang conjunctival cells

Strain	Adherence (% of inoculum)[a]
DH5α/pT7-7	0.4 ± 0.1
DH5α/pDC400	25.3 ± 1.2
DH5α/pDC601	54.3 ± 7.5
DH5α/pDC602	55.5 ± 4.3
C54b$^-$p$^-$	98.7 ± 9.5
C54-*hia*::*kan*[b]	1.5 ± 0.2
C54-Tn400.23[c]	3.3 ± 0.4

[a] Adherence was measured in a 30-min assay and was calculated by dividing the number of adherent bacteria by the number of inoculated bacteria. Values are the means ± the standard errors of the means of measurements made in triplicate from representative experiments.

[b] Strain C54-*hia*::*kan* was constructed by transforming C54b$^-$p$^-$ with linearized pHMW8-6, which contains the *hia* gene with an intragenic kanamycin cassette.

[c] Strain C54-Tn400.23 contains a mini-Tn*10 kan* element in the *hsf* locus (33).

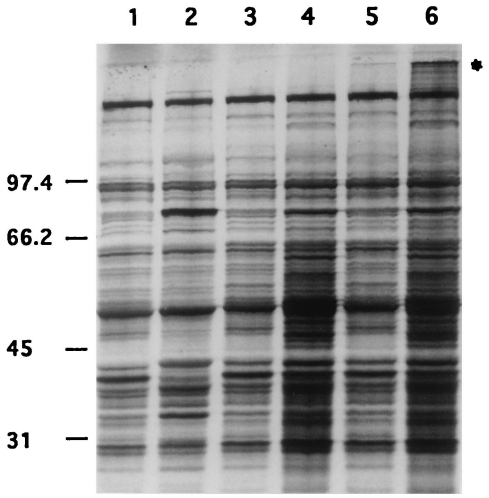

FIG. 2. Identification of plasmid-encoded proteins by using the bacteriophage T7 expression system. Bacteria were radiolabelled with Tran^{35}S-label, and whole-cell lysates were resolved on a sodium dodecyl sulfate–7.5% polyacrylamide gel. Proteins were visualized by autoradiography. Lane 1, *E. coli* BL21(DE3)/pT7-7, uninduced; lane 2, BL21(DE3)/pT7-7, induced; lane 3, BL21(DE3)/pDC602, uninduced; lane 4, BL21(DE3)/pDC602, induced; lane 5, BL21(DE3)/pDC601, uninduced; lane 6, BL21(DE3)/pDC601, induced. Plasmids pDC602 and pDC601 are derivatives of pT7-7 that contain the 8.3-kb *Xba*I fragment from pDC400 in opposite orientations. The asterisk indicates the overexpressed protein in BL21(DE3)/pDC601. The positions of molecular mass standards (in kilodaltons) are shown on the left.

ture containing a loop of 2 bases and a stem of 11 bases. It is of note that a string of 29 thymines spans the region from 149 to 121 nucleotides upstream of *hsf*.

Analysis of the predicted amino acid sequence of Hsf revealed several notable features, including an unusual N-terminal extremity (described below) and a putative Walker A box nucleotide-binding motif (residues 1943 to 1950). In addition, the protein is relatively acidic, with an isoelectric point of 5.4, and has a single cysteine residue, indicating that the folded form exists in the absence of any internal disulfide bond.

Homology to *hia* and Hia. We recently identified a nontypeable *H. influenzae* nonpilus protein called Hia which promotes attachment to cultured human epithelial cells (6). Comparison of the predicted amino acid sequence of the *hsf* product and the sequence of Hia revealed 81% similarity and 72% identity overall. As depicted in Fig. 3, the two sequences are highly conserved at their N- and C-terminal ends, and both contain a nucleotide-binding motif. Interestingly, Hia is encoded by a

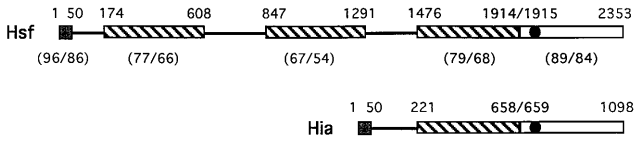

FIG. 3. Schematic alignment of Hsf and Hia. Regions of sequence similarity are indicated by shaded, hatched, and open bars, corresponding to N-terminal domains, internal domains, and C-terminal domains, respectively. The solid circles represent conserved Walker box ATP-binding motifs (GINVSGKT). The numbers above the bars refer to amino acid residue positions in the full-length proteins. The pairs of numbers in parentheses below the Hsf bars represent the percentages of sequence similarity and identity (in that order) between these domains and the corresponding Hia domains. The regions of Hsf defined by amino acid residues 51 to 173, 609 to 846, and 1292 to 1475 show minimal similarity to amino acids 51 to 220 of Hia.

3.3-kb gene and is only ~115 kDa. In this context, it is noteworthy that three separate stretches of Hsf (corresponding to amino acids 174 to 608, 847 to 1291, and 1476 to 1914) show significant homology to the region of Hia defined by amino acids 221 to 658 (Fig. 3). The percentages of similarity and identity, respectively, among these three stretches of Hsf are as follows: amino acids 174 to 608 versus amino acids 847 to 1291, 65 and 53%; amino acids 847 to 1291 versus amino acids 1476 to 1914, 70 and 56%; and amino acids 1476 to 1914 versus amino acids 174 to 608, 76 and 60%. The suggestion is that the larger size of Hsf may relate in part to the presence of a repeated domain which is present in a single copy in Hia.

To determine if *hsf* and *hia* are alleles of the same locus, we began by performing a series of Southern blot analyses. Samples of chromosomal DNA from strains C54 and 11 were subjected to digestion with *Bgl*II, *Cla*I, and either *Pst*I or *Xba*I. The resulting DNA fragments were separated by agarose electrophoresis and transferred bidirectionally to nitrocellulose membranes. One membrane was probed with a 3.3-kb internal fragment of the *hsf* gene (Fig. 1), and the other membrane was probed with a 1.6-kb intragenic fragment of the *hia* gene. As shown in Fig. 4, the two probes recognized exactly the same chromosomal fragments.

To obtain additional evidence that the *hsf* and *hia* genes are homologs, we sought to inactivate *hsf* by transformation of *H. influenzae* C54b$^-$p$^-$ with insertionally inactivated *hia*. The plasmid pHMW8-6 (6), which contains the *hia* gene with an intragenic kanamycin cassette, was linearized with *Nde*I and introduced into competent C54 cells. Southern hybridization confirmed insertion of the kanamycin cassette into *hsf* (data not shown). Furthermore, examination of the C54 mutant by negative-staining transmission electron microscopy revealed the loss of surface fibrils (data not shown). Consistent with these findings, the mutant strain demonstrated minimal attachment to Chang conjunctival cells (Table 1).

In additional experiments, we compared the cellular binding specificities conferred by the Hsf and Hia proteins. As shown in Fig. 5, DH5α/pDC601 (expressing *hsf*) demonstrated high-level attachment to Chang cells, KB cells, HeLa cells, and Intestine 407 cells, moderate-level attachment to HEp-2 cells, and minimal attachment to HEC-IB cells, ME-180 cells, and CHO-K1 cells. DH5α harboring pHMW8-5 (expressing *hia*) showed virtually the same pattern of attachment. Giemsa staining and subsequent examination by light microscopy confirmed these viable-count adherence assay results.

Homology to other bacterial extracellular proteins. A protein sequence similarity search was performed with the predicted Hsf amino acid sequence by using the BLAST network service of the National Center for Biotechnology Information (2). This search revealed low-level sequence similarity to a series of other bacterial adherence factors, including HMW1 and HMW2 (the proteins we previously identified as being important adhesins in Hia-deficient nontypeable *H. influenzae* strains [39]), AIDA-I (an adhesion protein expressed by some diarrheagenic *E. coli* strains [8]), and Tsh (a hemagglutinin produced by an avian-pathogenic *E. coli* strain [27]). In addition, Hsf showed homology to SepA, a *Shigella flexneri* secreted protein that appears to play a role in tissue invasion (7). Alignment of Hsf with HMW1, HMW2, AIDA-I, Tsh, and SepA revealed a highly conserved N-terminal domain (Fig. 6). In AIDA-I, Tsh, and SepA, this N-terminal extremity precedes a typical procaryotic signal sequence (7). Similarly, in Hsf, this conserved domain precedes a 26-amino-acid segment that is characterized by a positively charged region, followed by a string of hydrophobic residues and then alanine-asparagine-alanine.

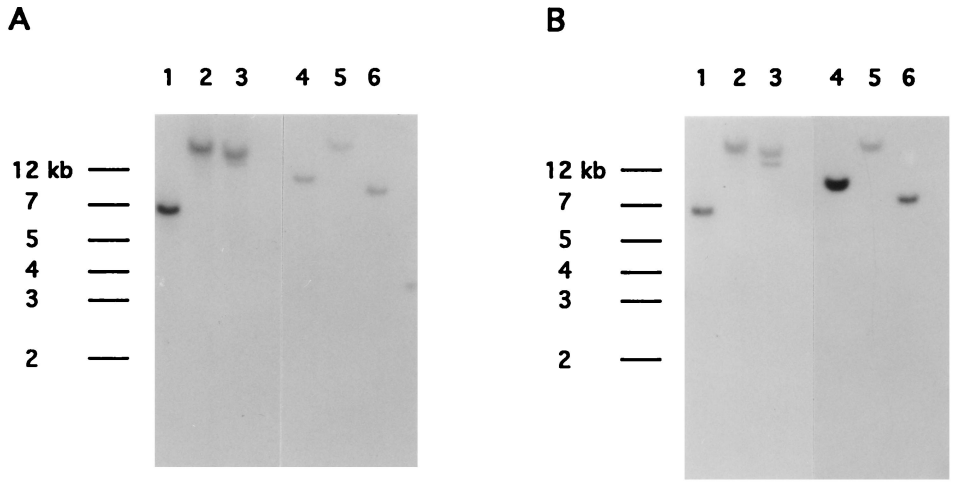

FIG. 4. Southern analysis of chromosomal DNA from *H. influenzae* C54 and 11, probing with *hsf* or *hia*. (A) Hybridization with the 3.3-kb *Pst*I-*Bgl*II intragenic fragment of *hsf* from strain C54. (B) Hybridization with the 1.6-kb *Sty*I-*Ssp*I intragenic fragment of *hia* from strain 11. DNA fragments were separated on a 0.7% agarose gel and transferred bidirectionally to nitrocellulose membranes prior to probing with either *hsf* or *hia*. Lanes 1, strain C54 chromosomal DNA digested with *Bgl*II; lanes 2, strain C54 chromosomal DNA digested with *Cla*I; lanes 3, strain C54 chromosomal DNA digested with *Pst*I; lanes 4, strain 11 chromosomal DNA digested with *Bgl*II; lanes 5, strain 11 chromosomal DNA digested with *Cla*I; lanes 6, strain 11 chromosomal DNA digested with *Xba*I.

Presence of an *hsf* homolog in other encapsulated and nonencapsulated strains. In previous work, we demonstrated that an *hsf* homolog is present in *H. influenzae* type b strains M42 and Eagan (33). To define the extent to which the *hsf* locus is shared by other type b strains, we examined a panel of evolutionarily diverse type b isolates by Southern analysis. Among these strains were six belonging to phylogenic division I (representing four different electromorph clusters) and four belonging to phylogenic division II (representing two different electromorph clusters) (24). Chromosomal DNA was digested with *Bgl*II and then probed with the intragenic 3.3-kb fragment of the *hsf* gene. As shown in Fig. 7A, all 10 strains showed hybridization, with some variation in the size of the hybridizing band and a correlation between the pattern of hybridization and evolutionary lineage. The universal presence of this locus among *H. influenzae* type b strains raised the question of its prevalence in other, non-type b encapsulated *H. influenzae* strains. Southern analysis of a series of type a, c, d, e, and f isolates again demonstrated a homolog in all cases, again with some variability in the size of the hybridizing band (Fig. 7B). It is of note that all four type f strains contained a hybridizing band of the same size.

Recently Fleischmann et al. (15) reported the genome sequence of *H. influenzae* Rd, which was one of the two serotype d strains we examined by Southern analysis. In accord with our Southern blotting results, a search of the Rd genome revealed an open reading frame with striking sequence similarity to *hsf*. From the published sequence it is evident that the Rd locus is only 894 nucleotides in length and encodes a protein of 298 amino acids that is 63% identical and 75% similar to the N-terminal region of Hsf. Interestingly, sequence immediately downstream of the Rd gene also shows similarity to C54 *hsf*. Further analysis of the Rd gene suggests that a base pair has been deleted at position 887, creating a frameshift and a premature termination codon. To confirm the accuracy of the Rd sequence in the region with the possible deletion, we used PCR and amplified a 514-bp fragment that overlaps this area; nucleotide sequencing of the amplified fragment corroborated the published sequence. As predicted from these findings, Rd lacks fibrils and demonstrates minimal in vitro adherence (data

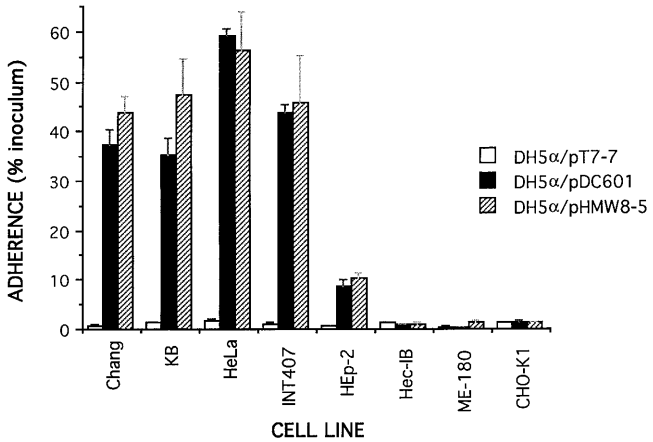

FIG. 5. Comparison of cellular binding specificities of *E. coli* DH5α harboring *hsf* or *hia*. Adherence was measured after incubating bacteria with eucaryotic cell monolayers for 30 min as previously described (39) and was calculated by dividing the number of adherent CFUs by the number of inoculated CFUs (39). Values are the means ± the standard errors of the means of measurements made in triplicate from representative experiments. Plasmid pDC601 contains the *hsf* gene from *H. influenzae* C54, while pHMW8-5 contains the *hia* gene from nontypeable *H. influenzae* 11. Both pDC601 and pHMW8-5 were prepared with pT7-7 as the cloning vector. INT407, Intestine 407.

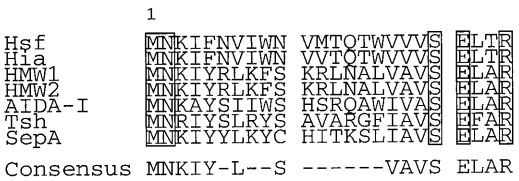

FIG. 6. Comparison of the N-terminal extremities of Hsf, HMW1, HMW2, AIDA-I, Tsh, and SepA. The N-terminal sequence of Hsf is aligned with those of Hia (6), HMW1 (4), HMW2 (4), AIDA-I (8), Tsh (27), and SepA (7). Amino acids that are invariant among the seven proteins are enclosed in boxes. A consensus sequence, defined as the presence of the same amino acid in four or more of the seven sequences, is shown on the lower line.

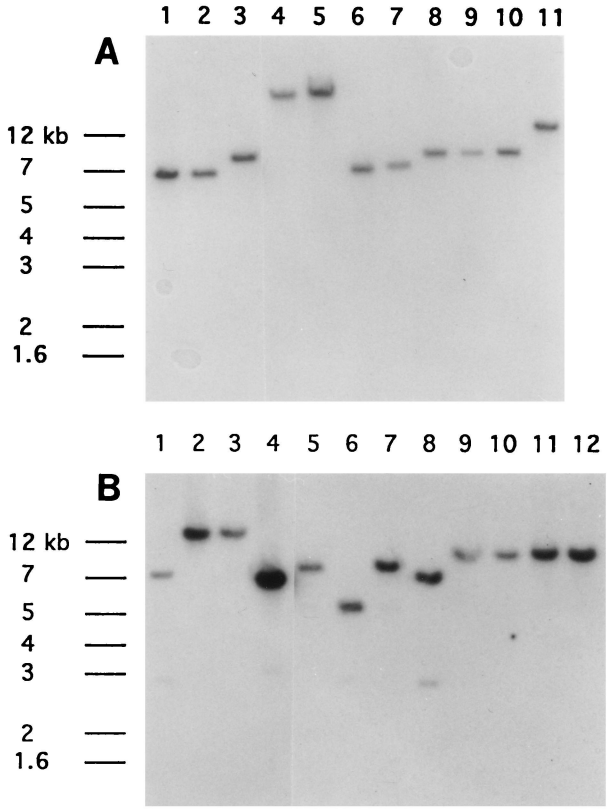

FIG. 7. Southern analysis of chromosomal DNA from epidemiologically distinct strains of type b *H. influenzae* and from non-type b encapsulated strains of *H. influenzae*. Chromosomal DNA was digested with *Bgl*II, separated on a 0.7% agarose gel, transferred to nitrocellulose, and probed with the 3.3-kb *Pst*I-*Bgl*II intragenic fragment of *hsf* from strain C54. (A) Lane 1, strain C54; lane 2, strain 1081; lane 3, strain 1065; lane 4, strain 1058; lane 5, strain 1060; lane 6, strain 1053; lane 7, strain 1063; lane 8, strain 1069; lane 9, strain 1070; lane 10, strain 1076; lane 11, strain 1084. (B) Lane 1, SM4 (type a); lane 2, SM72 (type c); lane 3, SM6 (type d); lane 4, Rd (type d); lane 5, SM7 (type e); lane 6, 142 (type e); lane 7, 327 (type e); lane 8, 351 (type e); lane 9, 134 (type f); lane 10, 219 (type f); lane 11, 346 (type f); lane 12, 503 (type f).

not shown). Theoretical correction of the frameshift results in a predicted protein at least 1,324 amino acids in length, depending on whether other, larger deletions also exist. The smaller size of the 1,324-amino-acid Rd protein compared with Hsf reflects the presence of one and one-half internal repeats rather than three.

Previous experiments revealed that 13 of 15 nontypeable strains lacking an HMW1- or HMW2-related protein had evidence of an *hia* homolog (6). Consistent with the demonstration that *hsf* and *hia* are homologous, Southern analysis of these 15 strains, probing with the 3.3-kb fragment of *hsf*, demonstrated hybridization in 12 of the same 13 (data not shown).

Chromosomal location of the *hsf* locus. In earlier work, the *hia* locus in nontypeable strain 11 was found to be flanked upstream by an open reading frame with significant homology to the gene coding for *E. coli* exoribonuclease II (6). In the present investigation, we determined that the *hsf* locus in strain C54 likewise is flanked on the 5′ side by an open reading frame with similarity to the gene coding for *E. coli* exoribonuclease II. This gene terminates 357 bp before the *hsf* start codon and encodes a protein with a predicted amino acid sequence that is 61% similar and 33% identical at its C-terminal end to exoribonuclease II. It is of note that the Rd *hsf* homolog is also flanked upstream by the exoribonuclease II locus.

DISCUSSION

In the present study we found that the *H. influenzae* type b surface fibril (*hsf*) locus contains a single long open reading frame that codes for a protein over 200 kDa in size. Analysis of the predicted amino acid sequence of the *hsf* product (Hsf) revealed significant similarity to a recently identified nontypeable *H. influenzae* adhesin called Hia. Consistent with this finding, *E. coli* transformants expressing Hsf and derivatives expressing Hia demonstrated the same cellular binding specificity. Furthermore, Southern blotting demonstrated that an *hsf* homolog is ubiquitous among encapsulated *H. influenzae* strains.

Overall, Hsf and Hia are 81% similar to one another. On the basis of the results of this study, these two proteins appear to recognize the same host cell receptor structure. Interestingly, the Hsf protein is roughly 240 kDa and is associated with expression of short surface fibrils; in contrast, Hia is only 114 to 115 kDa and has not been associated with fibril expression (6, 32). With this information in mind, it is notable that Hsf was identified in a type b strain while Hia was discovered in a nontypeable strain. One possibility is that a larger protein that forms a filamentous structure and extends beyond the surface of the bacterium is necessary in an encapsulated strain but is unnecessary in a nontypeable (nonencapsulated) strain. In support of this hypothesis, fibril-mediated attachment by encapsulated strain C54 is greatest under growth conditions that result in diminished encapsulation (34); in contrast, Hia-mediated attachment by nontypeable strain 11 is slightly reduced under the same growth conditions (35).

Examination of a series of evolutionarily diverse *H. influenzae* type b strains demonstrated the universal presence of a homolog to the *hsf* locus. Similarly, analysis of 11 non-type b encapsulated *H. influenzae* strains again revealed the existence of a homolog in each case. These findings indicate that the *hsf* locus is highly conserved among encapsulated *H. influenzae* strains and suggest that the *hsf* gene product may play a critical role in the life cycle of these organisms. It is interesting that an *hsf* homolog is present in some but not all nontypeable strains. Analysis of a large number of nontypeable strains showed that roughly one-quarter appear to contain a homolog of the *hsf* locus (32). Furthermore, in *H. influenzae* type b strain C54 and nontypeable strain 11, the *hsf* or *hia* gene is positioned in the same location in the genome, flanked upstream by the gene coding for the exoribonuclease II protein. It is intriguing to speculate that a subgroup of nontypeable strains, defined by the presence of an *hsf* homolog, are more highly related to encapsulated strains.

Analysis of the predicted amino acid sequence of the *hsf* gene product revealed low-level sequence homology to several bacterial adhesins, including the HMW1 and HMW2 proteins of nontypeable *H. influenzae* strains (4, 39), the AIDA-I protein of some diarrheagenic *E. coli* strains (8), and the hemagglutinin called Tsh of an avian *E. coli* strain (27). In addition, we noted homology to an *S. flexneri* secreted protein called SepA (7). The HMW1 and HMW2 proteins and the AIDA-I adhesin are encoded by gene clusters that include a structural gene along with accessory genes required for correct processing and surface localization of the respective adhesive protein (5, 8). On the other hand, Tsh and SepA are capable of autosecretion via a mechanism similar to that described for the *Neisseria* and *H. influenzae* IgA1 serine proteases (7). On the basis of adherence data from *E. coli* transformants expressing the Hsf protein, it appears that Hsf is capable of reaching the bacterial surface and mediating in vitro adherence independent of other *H. influenzae* gene products. Along these lines, it

is notable that Hsf contains a potential Walker A box ATP-binding domain (45), which may provide the energy for autosecretion. It is also possible, however, that *E. coli* proteins influence surface localization of Hsf.

The stretch of 29 thymine residues upstream of the *hsf* start codon is reminiscent of the A tract upstream of the *Mycoplasma hyorhinis vlp* genes and the C tract upstream of the *Bordetella pertussis fim* genes. In the case of the *vlp* genes, the poly(A) region is positioned between the −10 box and the −35 sequence and the length of the A tract controls Vlp phase variation, presumably by influencing transcription initiation (50). Similarly, the poly(C) region upstream of the *fim* genes is located between the −10 box and the binding site for an activator, and shifts in the length of the C tract correlate with altered levels of *fim* expression (48). By analogy, it is possible that the *hsf* T tract can vary in length and that such variation can influence *hsf* transcription. Future studies will examine this issue.

To summarize, we have identified a locus designated *hsf* which is involved in *H. influenzae* adherence to human epithelial cells. This locus appears to be common to all encapsulated *H. influenzae* strains and to a subset of nontypeable *Haemophilus* strains. We are presently studying whether the *hsf* product plays a role in in vivo colonization.

ACKNOWLEDGMENTS

We thank Linda Winter for expert technical assistance.

This work was supported by Public Health Service grants 1RO1 DC-02873-01 from the National Institute on Deafness and Other Communication Disorders (to J.W.S.) and AI-21707 from the National Institute on Allergy and Infectious Diseases (to S.J.B.) and by an American Heart Association Grant-in-Aid and an Infectious Diseases Society of America Young Investigator Award to J.W.S.

REFERENCES

1. **Alexander, H. E., and G. Leidy.** 1951. Determination of inherited traits of *H. influenzae* by deoxyribonucleic acid fractions isolated from type-specific cells. J. Exp. Med. **83:**345–359.
2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basis local alignment search tool. J. Mol. Biol. **215:**403–410.
3. **Anderson, P., R. B. Johnston, Jr., and D. H. Smith.** 1972. Human serum activity against *Haemophilus influenzae* type b. J. Clin. Invest. **51:**31–38.
4. **Barenkamp, S. J., and E. Leininger.** 1992. 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*. Infect. Immun. **60:**1302–1313.
5. **Barenkamp, S. J., and J. W. St. Geme III.** 1994. Genes encoding high-molecular-weight adhesion proteins of nontypeable *Haemophilus influenzae* are part of gene clusters. Infect. Immun. **62:**3320–3328.
6. **Barenkamp, S. J., and J. W. St. Geme III.** 1996. Identification of a second family of high molecular weight adhesion proteins expressed by nontypeable *Haemophilus influenzae*. Mol. Microbiol. **19:**1215–1223.
7. **Benjelloun-Touimi, Z., P. J. Sansonetti, and C. Parsot.** 1995. SepA, the major extracellular protein of *Shigella flexneri*: autonomous secretion and involvement in tissue invasion. Mol. Microbiol. **17:**123–135.
8. **Benz, I., and M. A. Schmidt.** 1992. AIDA-I, the adhesin involved in diffuse adherence of the diarrhoeagenic *Escherichia coli* strain 2787 (O126:H27), is synthesized via a precursor molecule. Mol. Microbiol. **6:**1539–1546.
9. **Brinton, C. C., M. J. Carter, D. B. Derber, S. Kar, J. A. Kramarik, A. C.-C. To, S. C.-M. To, and S. W. Wood.** 1989. Design and development of pilus vaccines for *Haemophilus influenzae* diseases. Pediatr. Infect. Dis. J. **8:**S54–S61.
10. **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.
11. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
12. **Dower, W. J., J. F. Miller, and C. W. Ragsdale.** 1988. High efficiency transformation of *E. coli* by high voltage electroporation. Nucleic Acids Res. **16:**6127–6145.
13. **Farley, M. M., D. S. Stephens, S. L. Kaplan, and E. O. Mason.** 1990. Pilus- and non-pilus-mediated interactions of *Haemophilus influenzae* type b with human erythrocytes and human nasopharyngeal mucosa. J. Infect. Dis. **154:**752–759.
14. **Farley, M. M., D. S. Stephens, M. H. Mulks, M. D. Cooper, J. V. Bricker, S. S. Mirra, and A. Wright.** 1986. Pathogenesis of IgA1 protease-producing and -nonproducing *Haemophilus influenzae* in human nasopharyngeal organ cultures. J. Infect. Dis. **161:**274–280.
15. **Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J.-F. Tomb, B. A. Dougherty, J. M. Merrick, K. McKenney, G. Sutton, W. FitzHugh, C. Fields, J. D. Gocayne, J. Scott, R. Shirley, L.-I. Liu, A. Glodek, J. M. Kelley, J. F. Weidman, C. A. Phillips, T. Spriggs, E. Hedblom, M. D. Cotton, T. R. Utterback, M. C. Hanna, D. T. Nguyen, D. M. Saudek, R. C. Brandon, L. D. Fine, J. L. Fritchman, J. L. Fuhrmann, N. S. M. Geoghagen, C. L. Gnehm, L. A. McDonald, K. V. Small, C. M. Fraser, H. O. Smith, and J. C. Venter.** 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science **269:**496–512.
16. **Gilsdorf, J. R., H. Y. Chang, K. W. McCrea, and L. O. Bakaletz.** 1992. Comparison of hemagglutinating pili of *Haemophilus influenzae* type b with similar structures of nontypeable *H. influenzae*. Infect. Immun. **60:**374–379.
17. **Guerina, N. G., S. Langermann, H. W. Clegg, T. W. Kessler, D. A. Goldmann, and J. R. Gilsdorf.** 1982. Adherence of piliated *Haemophilus influenzae* type b to human oropharyngeal cells. J. Infect. Dis. **146:**564.
18. **Herriott, R. M., E. M. Meyer, and M. Vogt.** 1970. Defined nongrowth media for stage II development of competence in *Haemophilus influenzae*. J. Bacteriol. **101:**517–524.
19. **Kilian, M.** 1976. A taxonomic study of the genus *Haemophilus*, with proposal of a new species. J. Gen. Microbiol. **93:**9–62.
20. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.
21. **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.
22. **Murphy, T. F., J. M. Bernstein, D. M. Dryja, A. A. Campagnari, and M. A. Apicella.** 1987. Outer membrane protein and lipooligosaccharide analysis of paired nasopharyngeal and middle ear isolates in otitis media due to nontypeable *Haemophilus influenzae*: pathogenic and epidemiologic observations. J. Infect. Dis. **5:**723–731.
23. **Musser, J. M., J. S. Kroll, D. M. Granoff, E. R. Moxon, B. R. Brodeur, J. Campos, H. Dabernat, W. Fredriksen, J. Hamel, G. Hammond, E. A. Hoiby, K. E. Jonsdottir, M. Kabeer, I. Kallings, W. N. Khan, M. Kilian, K. Knowles, H. J. Koornhof, B. Law, K. I. Li, J. Montgomery, P. E. Pattison, J.-C. Piffaretti, A. K. Takala, M. L. Thong, R. A. Wall, J. I. Ward, and R. K. Selander.** 1990. Global genetic structure and molecular epidemiology of encapsulated *Haemophilus influenzae*. Rev. Infect. Dis. **12:**75–111.
24. **Musser, J. M., J. S. Kroll, E. R. Moxon, and R. K. Selander.** 1988. Evolutionary genetics of the encapsulated strains of *Haemophilus influenzae*. Proc. Natl. Acad. Sci. USA **85:**7758–7762.
25. **Pichichero, M. E., P. Anderson, M. Loeb, and D. H. Smith.** 1982. Do pili play a role in pathogenicity of *Haemophilus influenzae* type b? Lancet **ii:**960–962.
26. **Pittman, M.** 1931. Variation and type specificity in the bacterial species *Haemophilus influenzae*. J. Exp. Med. **53:**471–493.
27. **Provence, D., and R. Curtiss III.** 1994. Isolation and characterization of a gene involved in hemagglutination by an avian pathogenic *Escherichia coli* strain. Infect. Immun. **62:**1369–1380.
28. **Read, R. C., R. Wilson, A. Rutman, V. Lund, H. C. Todd, A. P. R. Brain, P. K. Jeffery, and P. J. Cole.** 1991. Interaction of nontypeable *Haemophilus influenzae* with human respiratory mucosa in vitro. J. Infect. Dis. **163:**549–558.
29. **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.
30. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
31. **St. Geme, J. W., III.** 1993. Nontypeable *Haemophilus influenzae* disease: epidemiology, pathogenesis, and prospects for prevention. Infect. Agents Dis. **2:**1–16.
32. **St. Geme, J. W., III, and S. J. Barenkamp.** Unpublished data.
33. **St. Geme, J. W., III, and D. Cutter.** 1995. Evidence that surface fibrils expressed by *Haemophilus influenzae* type b promote attachment to human epithelial cells. Mol. Microbiol. **15:**77–85.
34. **St. Geme, J. W., III, and D. Cutter.** 1996. Influence of fibrils, pili, and capsule on *Haemophilus influenzae* type b *in vitro* adherence. Mol. Microbiol. **21:**21–31.
35. **St. Geme, J. W., III, and D. Cutter.** Unpublished data.
36. **St. Geme, J. W., III, M. L. de la Morena, and S. Falkow.** 1994. A *Haemophilus influenzae* IgA protease-like protein promotes intimate interaction with human epithelial cells. Mol. Microbiol. **14:**217–233.
37. **St. Geme, J. W., III, and S. Falkow.** 1990. *Haemophilus influenzae* adheres to and enters cultured human epithelial cells. Infect. Immun. **58:**4036–4044.
38. **St. Geme, J. W., III, and S. Falkow.** 1991. Loss of capsule expression by *Haemophilus influenzae* type b results in enhanced adherence to and invasion of human cells. Infect. Immun. **59:**1325–1333.
39. **St. Geme, J. W., III, S. Falkow, and S. J. Barenkamp.** 1993. High-molecular-weight proteins of nontypeable *Haemophilus influenzae* mediate attachment

to human epithelial cells. Proc. Natl. Acad. Sci. USA **90:**2875–2879.
40. **Studier, F. W., and B. A. Moffatt.** 1986. Use of bacteriophage T7 RNA polymerase to direct high-level expression of cloned genes. J. Mol. Biol. **189:**113–130.
41. **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.
42. **Turk, D. C.** 1982. Clinical importance of *Haemophilus influenzae*, p. 3–9. *In* S. H. Sell and P. F. Wright (ed.), *Haemophilus influenzae* epidemiology, immunology, and prevention of disease. Elsevier/North-Holland Publishing Co., New York.
43. **Turk, D. C.** 1984. The pathogenicity of *Haemophilus influenzae*. J. Med. Microbiol. **18:**1–16.
44. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19:**259–268.
45. **Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay.** 1982. Distantly related sequences in the alpha and beta subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1:**945–951.
46. **Weber, A., K. Harris, S. Lohrke, L. Forney, and A. L. Smith.** 1991. Inability to express fimbriae results in impaired ability of *Haemophilus influenzae* b to colonize the nasopharynx. Infect. Immun. **59:**4724–4728.
47. **Weiser, J. N., S. T. H. Chong, D. Greenberg, and W. Fong.** 1995. Identification and characterization of a cell envelope protein of *Haemophilus influenzae* contributing to phase variation in colony opacity and nasopharyngeal colonization. Mol. Microbiol. **17:**555–564.
48. **Willems, R., A. Paul, H. G. J. van der Heide, A. R. ter Avest, and F. R. Mooi.** 1990. Fimbrial phase variation in *Bordetella pertussis*: a novel mechanism for transcriptional regulation. EMBO J. **9:**2803–2809.
49. **Yanish-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.
50. **Yogev, D., R. Rosengarten, R. Watson-McKown, and K. S. Wise.** 1991. Molecular basis of *Mycoplasma* surface antigenic variation: a novel set of divergent genes undergo spontaneous mutation of periodic coding regions and 5′ regulatory sequences. EMBO J. **10:**4069–4079.