Identification of New hmwA Alleles from Nontypeable Haemophilus influenzae

I. Zafer Ecevit,1 Kirk W. McCrea,1 Carl F. Marrs,2 and Janet R. Gilsdorf1,2*

Department of Pediatrics and Communicable Diseases1 and Department of Epidemiology,2 University of Michigan, Ann Arbor, Michigan

Received 20 April 2004/Returned for modification 3 June 2004/Accepted 13 October 2004

High-molecular-weight proteins of Haemophilus influenzae mediate attachment to epithelial cells. Previous reports describe several allelic versions of hmwA genes that have different adherence properties. Here we report three new alleles of hmwA (hmwA from strain AAr96, hmwA from strain AAr105, and hmwA from strain G822), demonstrating the high degree of DNA variation of these genes among different strains.

Nontypeable Haemophilus influenzae (NTHi), a gram-negative, nonencapsulated, human-specific microorganism, commonly inhabits the upper respiratory tract and causes otitis media, conjunctivitis, sinusitis, pneumonia, and acute exacerbation of chronic bronchitis. Occasionally, NTHi causes severe invasive diseases, such as meningitis, endocarditis, and bacteremia (14).

NTHi high-molecular-weight (HMW) proteins mediate bacterial attachment to epithelial cells (7, 11, 17, 18) and have been implicated as possible virulence factors for otitis media (12, 20) or chronic obstructive pulmonary disease (21). This adhesin is produced by the action of three genes (hmwA, hmwB, and hmwC) located in the hmw locus (1, 2). Many H. influenzae strains contain two distinct hmw loci, hmw-1 and hmw-2 (3, 4). The hmwA genes of these loci encode the HMWA adhesive proteins, which are 52 to 62% identical at the amino acid level (4) among several NTHi strains. hmwB genes encode outer membrane translocator proteins, which are 99% identical in NTHi strain 12. HMWB proteins are located in the outer membrane and serve to translocate HMWA across the outer membrane and prevent degradation by periplasmic proteases. hmwC genes encode cytoplasmic proteins, which are 97% identical in NTHi strain 12, and appear to stabilize HMWA (2, 17) and to influence glycosylation of HMWA1 (10). HMWA1 of strain 12 mediates binding to α-2,3-linked sialylated glycoproteins, and the epithelial cell receptor structure for HMWA2 of this strain is unknown (17).

While initial studies of hmwA described the alleles hmwA1 and hmwA2, one in each hmw locus of strain 12 (3), Van Schilfgaarde et al. (21) described a third hmwA allele from the chronic obstructive pulmonary disease H. influenzae strain A95006, whose 4,671-bp gene encodes a predicted protein with 70% homology to HMWA1 and 68% homology to HMWA2 of strain 12. More recently, Buscher et al. (4) have identified four additional hmwA alleles with either HMWA1- or HMWA2-like binding characteristics from two NTHi strains and showed that the genes encoding these differential binding characteristics were variably located downstream of either HI01679 or HI01598 in the Rd genome.

Although type b strains lack hmw loci, 55 to 80% of NTHi strains have these genes (1, 8, 20). Among the other encapsulated H. influenzae types, hmw loci were detected in 26% of type a, 8% of type e, and 5% of type f strains (16). In a previous study by our group using dot blot hybridization, 51% of NTHi isolates hybridized with strain 12 hmwA1-specific probes, 23% hybridized with hmwA2-specific probes, and 48% hybridized with hmwC-specific probes (8). While 18% hybridized with all three probes, 23% hybridized with only hmwA1 and hmwC probes, 1% hybridized with only hmwA2 and hmwC probes, and 6% (10 isolates) hybridized with only the hmwC probe, suggesting that many strains may contain hmwA genes that do not hybridize the strain 12-specific probes. Specifically, NTHi strains AAr96, AAr105, and G822 failed to hybridize with gene probes targeting the unique regions hmwA1 and hmwA2 of strain 12 but hybridized with a probe targeting the conserved hmwC genes. In the present study, we investigate the possibility that these three strains contain allelic versions of hmwA that failed to hybridize with the strain 12-specific probes.

Table 1 lists the primers synthesized by the University of Michigan Biomedical Research Core Facility that were used to amplify regions of the hmwA genes from NTHi strains AAr96, AAr105, and G822. For all PCRs, NTHi strain 12 served as the positive control and the hmwA-deficient NTHi strain 11 served as the negative control (3). All PCRs were carried out as previously described (8). The resulting PCR products were cloned into the plasmid vector pCR4-TOPO (Invitrogen, Carlsbad, Calif.), and the resulting recombinant plasmids were transformed into TOP10 Escherichia coli host cells (Invitrogen). Insert regions were sequenced at the University of Michigan Sequencing Core Facility, using Applied Biosystems model 3700 and 3730 automated sequencers. DNA and protein sequences were aligned and compared by using Lasergene Bioinformatics software (DNASTAR, Inc., Madison, Wis.). Multiple alignments were performed with the CLUSTAL W program. The Lipman-Pearson algorithm (Ktuple = 2; gap penalty = 4; gap length penalty = 12) was used for pairwise alignment.

hmwA genes from strains AAr96, AAr105, and G822 were 4,428, 4,476, and 4,449 bp in length, respectively. Figure 1

* Corresponding author. Mailing address: Department of Pediatrics and Communicable Diseases, University of Michigan, Ann Arbor, MI 48109-0244. Phone: (734) 763-2440. Fax: (734) 936-7635. E-mail: gilsdorf@umich.edu.

Copyright © 2005, American Society for Microbiology. All Rights Reserved.
FIG. 1. Comparison of the predicted amino acid sequences of HMWA from the strains AAr96, AAr105, and G822 with the deduced amino acid sequences of HMW-1 and HMW-2 from nontypeable Hi strains 12, 5, and A95006 previously reported. The HMWA protein sequences were aligned by using the CLUSTAL algorithm in the DNAStar software. Identical amino acids in all HMWA sequences are indicated with dots, and gaps introduced to maximize alignments are indicated by dashed lines. At the top in bold is a consensus amino acid sequence in which all letters except X represent an amino acid present in at least three of the six predicted proteins, and X means there is no consensus amino acid at that position. The RGD sequences are underlined.
compares the predicted amino acid sequences of these three hmwA genes to the five previously described and available in GenBank. In all eight proteins, the N-terminal 440 amino acids of HMWA are highly conserved and the C-terminal 100 amino acids are moderately conserved, but the region between, which corresponds to the epithelial cell binding domain (7), is extremely variable. In pairwise comparisons, using the Lipman-Pearson algorithm, amino acid identities between the eight alleles varied from 66 to 77% (Table 2). DNA sequence analysis of each hmwA gene in this study revealed that the genes from strains AAr105 and G822 were adjacent to the homologous Rd gene HI01679, while hmwA gene in strain AAr96 was adjacent to the HI01598 homolog (Table 2).

HMW proteins are structural and functional analogs of the filamentous hemagglutinin of Bordetella pertussis (1), which mediates binding by an arginine-glycine-aspartic acid (RGD) motif. Because this motif is also present in the derived amino acid sequence of HMWA2 of strain 12, RGD-mediated adherence of HMWA2 to the integrin CR3 has been suggested (15, 21). The RGD tripeptide motif of HMWA2 from strain 12 compares the predicted amino acid sequences of these three hmwA genes to the five previously described and available in GenBank. In all eight proteins, the N-terminal 440 amino acids of HMWA are highly conserved and the C-terminal 100 amino acids are moderately conserved, but the region between, which corresponds to the epithelial cell binding domain (7), is extremely variable. In pairwise comparisons, using the Lipman-Pearson algorithm, amino acid identities between the eight alleles varied from 66 to 77% (Table 2). DNA sequence analysis of each hmwA gene in this study revealed that the genes from strains AAr105 and G822 were adjacent to the homologous Rd gene HI01679, while hmwA gene in strain AAr96 was adjacent to the HI01598 homolog (Table 2).

HMW proteins are structural and functional analogs of the filamentous hemagglutinin of Bordetella pertussis (1), which mediates binding by an arginine-glycine-aspartic acid (RGD) motif. Because this motif is also present in the derived amino acid sequence of HMWA2 of strain 12, RGD-mediated adherence of HMWA2 to the integrin CR3 has been suggested (15, 21). The RGD tripeptide motif of HMWA2 from strain 12

FIG. 1—Continued.
The 5′ regions of the amino acid sequences of strain AAr96, AAr105, and G822 are identical except for one base difference. In conclusion, we identified three additional hmwA alleles of H. influenzae, which is reminiscent of the allelic diversity of genes encoding both the structural protein (hmwA) and the adhesin (hifE) of H. influenzae hemagglutinating pili (5, 13). The most conserved domain of the HMW-like proteins is the N terminus regions of the immature protein, which traffics the proteins to the cell surface and is then cleaved from the mature proteins, while the most variable domain is the receptor binding region (6). While the antigenic domains corresponding to protective antibodies of HMW remain unknown, the extreme sequence diversity of the binding domain suggests that these proteins may vary antigenically as well as functionally.

Nucleotide sequence accession numbers. Sequences determined in this work have been submitted to GenBank with the following accession numbers: for hmwA from strain AAr96, AY601284; for hmwA from strain AAr105, AY601283; and for hmwA from strain G822, AY601282.

We thank May Patel for her expert technical assistance. This study was supported by awards AI 25630 and DC05840 to J.R.G. from the National Institutes of Health.

REFERENCES


---

**Editor:** J. N. Weiser