Proc. Natl. Acad. Sci. USA
Vol. 96, pp. 1077–1082, February 1999
Microbiology

Variation in expression of the *Haemophilus influenzae* HMW adhesins: A prokaryotic system reminiscent of eukaryotes

SUZANNE DAWID*, STEPHEN J. BARENKAMP†, AND JOSEPH W. ST. GEME III*‡

*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, MO 63110; and †Department of Pediatrics, St. Louis University School of Medicine, and The Pediatric Research Institute, Cardinal Glennon Children's Hospital, St. Louis, MO 63104

Edited by John J. Mekalanos, Harvard Medical School, Boston, MA, and approved November 30, 1998 (received for review September 1, 1998)

ABSTRACT **Expression of a number of eukaryotic genes is regulated by long stretches of tandem repeats located within the 5′ untranslated region of the particular gene. In this study, we describe a regulatory system in *Haemophilus influenzae* with striking similarities to those found in eukaryotes. We show that expression of the HMW1 and HMW2 adhesins varies based on the number of 7-bp tandem repeats in the *hmw1A* and *hmw2A* promoters. The repeats lie between two separate transcription initiation sites and exert a repressive effect, such that increases in repeat number result in step-wise decreases in levels of specific mRNA and protein production and vice versa. The range of expression of HMW1 and HMW2 varies between very weak and very strong, with a series of gradations in between. Variation in the number of repeats in the *hmw1A* and *hmw2A* promoters occurs in individual colonies passaged *in vitro*, in an animal model of infection, and during natural infection in humans. This system of regulation is unique in prokaryotes and likely enhances the pathogenicity of the organism by increasing adaptive potential.**

In recent years, long stretches of tandem oligonucleotide repeats have been found increasingly in the upstream untranslated region of eukaryotic genes. In many cases, the number of repeats influences gene expression, reflecting the fact that transcriptional regulators often bind to the repeat region and regulate transcription (1). For example, the gene encoding human epidermal growth factor receptor is preceded by a series of TCC repeats that are recognized by a transcriptional activator. As these repeats are sequentially deleted, transcriptional activity and protein expression steadily decrease (2). Other examples in which upstream repeats function as transcriptional activating elements include the yeast *Adr2p* gene, the *Drosophila hsp26* and *actin 5C* genes, the chicken α-2(I)collagen gene, and the rat *neu* gene (3–7). Less commonly, tandem repeats have a negative effect on downstream gene activity (8). Typically, the number of repeats undergoes high-frequency spontaneous variation via slipped-strand mispairing, serving as a source of quantitative genetic variation. Such quantitative variation is believed to confer an evolutionary advantage to a population, facilitating adaptation to novel ecological challenges.

Nontypable *Haemophilus influenzae* is a Gram-negative bacterium that represents a common commensal organism in the human upper respiratory tract and an important cause of localized respiratory tract disease (8). The pathogenesis of disease begins with colonization of the upper respiratory mucosa, followed by contiguous spread to the middle ear, the sinuses, or the lungs, usually precipitated by respiratory viral infection or exacerbation of underlying allergic disease (9, 10). Similar to the situation with other bacterial pathogens, adherence to respiratory epithelium is a fundamental step in the process of colonization and is mediated by bacterial surface factors called adhesins.

Among diverse strains of nontypable *H. influenzae*, both pilus and nonpilus adhesins exist (11–14). Approximately 75% of clinical isolates express proteins that are immunologically and functionally related to the HMW1 and HMW2 adhesins produced by *H. influenzae* strain 12 (11). HMW1 is encoded by the *hmw1A* gene, which is flanked downstream by accessory genes called *hmw1B* and *hmw1C*, whereas HMW2 is encoded by the *hmw2A* gene, which is flanked downstream by genes called *hmw2B* and *hmw2C* (15). The accessory genes encode proteins involved in the processing and surface localization of the HMW adhesins (16).

Comparison of the predicted amino acid sequences of HMW1 and HMW2 reveals 71% identity and 80% similarity overall (11). Interestingly, despite their sequence similarity, the HMW1 and HMW2 adhesins exhibit different cellular binding specificities, suggesting recognition of distinct receptor structures (17). In early studies, Barenkamp and Bodor (18) examined sera from patients recovering from disease caused by nontypable *H. influenzae* and found that the HMW adhesins were the predominant target of the serum antibody response. Thus, although the HMW1 and HMW2 adhesins facilitate bacterial colonization, they also potentially contribute to immune clearance. Given these divergent functions, there is reason to believe that the organism might have evolved a mechanism to vary expression of the HMW1 and HMW2 proteins.

In recent work, Barenkamp (19) used the chinchilla otitis media model and found that immunization with purified protein preparations containing primarily HMW1 or primarily HMW2 provided partial protection against challenge of the middle ear with wild-type strain 12. Interestingly, among animals that were immunized with HMW1 and developed otitis media, isolates from the middle ear uniformly expressed relatively decreased amounts of HMW1. Similarly, isolates from animals immunized with HMW2 expressed diminished quantities of HMW2. These observations suggest that expression of HMW1 and HMW2 can be down-modulated.

Inspection of the nucleotide sequence upstream of the *hmw1A* and *hmw2A* genes reveals absolute identity up to 311 bp from the start codon, including 17 7-bp direct repeats arranged in a tandem array. In this study we characterized the *hmw1A* and *hmw2A* promoters and examined the relationship between the number of 7-bp repeats and the level of expression of HMW1 and HMW2. Our results demonstrated that the number of 7-bp repeats undergoes spontaneous variation in a Rec-independent manner, presumably via slipped-strand mispairing. Increases in the number of repeats resulted in a graded

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

This paper was submitted directly (Track II) to the *Proceedings* office.

‡To whom reprint requests should be addressed at: Department of Molecular Microbiology, Washington University School of Medicine, 660 South Euclid Avenue, Box 8230, St. Louis, MO 63110. e-mail: stgeme@borcim.wustl.edu.

decrease in protein expression, and decreases in the number of repeats were associated with a step-wise increase in protein expression. Changes in protein expression reflected changes in the level of specific mRNA, suggesting that variation in repeat number influences either transcriptional activity or transcript stability. We speculate that spontaneous variation in the number of tandem repeats enables the organism to vary between states associated with efficient adherence versus effective immune evasion.

MATERIALS AND METHODS

Bacterial Strains and Growth Conditions. The primary bacterial strains used in this study included: nontypable *H. influenzae* strain 12, a nontypable clinical isolate (11); strain 12 *rec-1*, a *rec*− derivative of strain 12 constructed as described (20); *H. influenzae* strain DB117, a laboratory strain that is *rec*− (21); matched nasopharyngeal and middle ear isolates of nontypable *H. influenzae* from patients FL and RA (22); and *Escherichia coli* strain DH5α (GIBCO/BRL). Matched clinical isolates were obtained as described by Krasan *et al.* (22) and were determined to represent the same strain based on genomic fingerprints and total protein profiles. *H. influenzae* and *E. coli* strains were grown as described (23, 24).

Analysis of Repeat Number in Chinchilla and Strain 12 Variants. Relevant strains were inoculated onto chocolate agar plates, and individual colonies were picked and analyzed for repeat number by using two consecutive PCRs. The first reaction separated the *hmw1* locus from the *hmw2* locus by using unique upstream sequences, and the second reaction used primers immediately flanking the repeat region. The forward primers for the first reaction were 5′-TCTTAATAC-TAGTACAAACCC-3′ for *hmw1A* and 5′-CCACTCAAAT-CAACTGGTTAA-3′ for *hmw2A*, and the common reverse primer was 5′-AGTAACATAGCGGAAAGTGGC-3′. The primers for the second reaction were 5′-CCGCCATATA-AAATGGTATAATC-3′ and 5′-CCCTCGGTTCATCATT-TCAT-3′. The final PCR products of the wild-type strain 12 founder colonies and the original chinchilla isolates initially were analyzed for repeat number by nucleotide sequencing by using the dideoxy method with the U.S. Biochemicals Sequenase kit. Once the sizes were determined, subsequent analysis of *in vitro*-derived variants was performed by running samples next to a 50-bp ladder (GIBCO/BRL) on 3% metaphore gels (FMC).

Western Blotting. To determine the amount of HMW1 and HMW2 produced by *H. influenzae* strain 12 variants and the isolates from patients FL and RA, bacteria were grown to mid-log phase in brain heart infusion broth supplemented with hemin and NAD (23) and then pelleted by centrifugation at $10{,}000 \times g$ for 5 min. Pellets were resuspended in 50 mM Hepes buffer, pH 8.0 and sonicated. Equal quantities of protein were subjected to SDS/PAGE on 7.5% polyacrylamide gels (25), and Western analysis was performed as described (16).

Purification and Analysis of RNA. RNA was purified for Northern blot by using a modified acid phenol protocol (26). RNA was quantitated spectrophotometrically, and 30 μg was loaded per lane on a formaldehyde-containing gel. Samples were separated by electrophoresis, transferred to a nylon membrane by capillary transfer, and probed with a divergent region of either *hmw1A* or *hmw2A*. As a control for loading, the membrane also was probed with an internal fragment of the *H. influenzae porA* gene, which is constitutively expressed (27). *hmw1A* and *hmw2A* transcripts were quantitated by using a Storm PhosphorImager (Molecular Dynamics) and were corrected according to corresponding *porA* transcript levels.

RNA was purified for primer extension by using the Rneasy kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. Primer extension was performed by using 10 μg of RNA, the appropriate oligonucleotide primer, and the Primer Extension Kit (Promega), as described by the manufacturer. Products were run next to a DNA ladder (Promega) or a sequencing ladder on a denaturing 8% acrylamide gel. The oligonucleotide primer used to characterize P1 was extension 1: 5′-CAGGCGTTTGCTGAATTTGA-3′, and the primer used to optimally characterize P2 was extension 2: 5′-CCCTCGGTTCATCATTTCATG-3′.

Construction of the Promoter Knockout Vector and Transformation into *H. influenzae*. To construct a vector for elimination of the *hmv1A* promoter, we began with pHMW1–14, a derivative of pT7–7 containing the entire *hmw1* gene cluster, including a portion of the upstream ORF (11). pHMW1–14 was digested with *Spe*I and *Bam*HI to delete a 1.7-kb fragment corresponding to the *hmw1A* promoter and 5′ coding sequence. After blunt-ending with T4 DNA polymerase (GIBCO/BRL), this piece of DNA was ligated to a *Hinc*II fragment containing the kanamycin cassette from pUC4K (28), generating pHMW1–14Δ*hmw1p*. For transformation into *H. influenzae*, pHMW1–14Δ*hmw1p* was digested with *Nde*I and *Sal*I to generate an 11-kb fragment, which was gel purified and incubated with *H. influenzae* made competent by the MIV method (29). Transformants were selected on brain heart infusion-DB agar plates containing kanamycin and were confirmed by Western analysis to lack expression of HMW1.

Construction of *lacZ* Fusions. To facilitate studies of the *hmwA* promoters, we constructed an *E. coli–H. influenzae* shuttle vector that contains a promoterless *lacZ* gene. Initially, the plasmid pGJB103 (30) was digested with *Pst*I, blunt-ended, and then digested with *Bgl*II to remove a 565-bp fragment. The resulting DNA was ligated to a *Bgl*II—*Sna*BI fragment containing the *lacZ* gene from a derivative of pRS552 (31), generating the plasmid pSD100.

By using pSD100, we constructed derivatives with the *hmwA* promoter containing either 17 or 22 7-bp repeats fused to *lacZ*. To construct the derivative with 17 repeats, the *hmwA* promoter region was amplified from *H. influenzae* strain 12 chromosomal DNA and inserted into *Bgl*II—*Sal*I-digested pSD100, generating the plasmid pSD117. To construct the derivative with 22 repeats, the *hmwA* promoter region was amplified from the plasmid pHMW1–14, which is known to contain 22 repeats, and again inserted into *Bgl*II—*Sal*I digested pSD100, creating a plasmid designated pSD122. pSD117 and pSD122 were confirmed to contain 17 and 22 repeats, respectively, as described above. To construct the negative control plasmid, pSD201, the ribosomal binding site from the *hmwA* gene was fused to a region just upstream of the 5′-most repeat, removing both potential promoter sequences.

RESULTS

Tandem Repeats Upstream of *hmw1A* and *hmw2A* Lie Between Two Transcriptional Start Sites. By using RNA extracted from *E. coli* DH5α/pHMW1–14 along with an oligonucleotide primer corresponding to the 5′ end of the coding region of *hmw1A*, primer extension analysis demonstrated two transcriptional start sites, designated P1 and P2 (Fig. 1). Identical results were obtained with RNA from *H. influenzae* strain 12 and the same oligonucleotide primer and in studies of *E. coli* harboring the *hmw2* gene cluster (data not shown). In both *E. coli* and *H. influenzae*, P1 mapped downstream of the tandem repeats. Examination of the sequence immediately preceding P1 revealed a highly conserved σ^{70} −35 sequence just overlapping the first (3′-most) repeat unit, but no classic σ^{70} −10 sequence. Localization of P2 was more difficult because of alterations in the number of repeats within a given population of organisms, giving rise to a ladder of bands at 7-bp intervals; nevertheless, based on multiple experiments and the position of the most prominent product, the P2 transcriptional start site appeared to map at the upstream end of the repeats, within the 5′-most repeat. Examination of

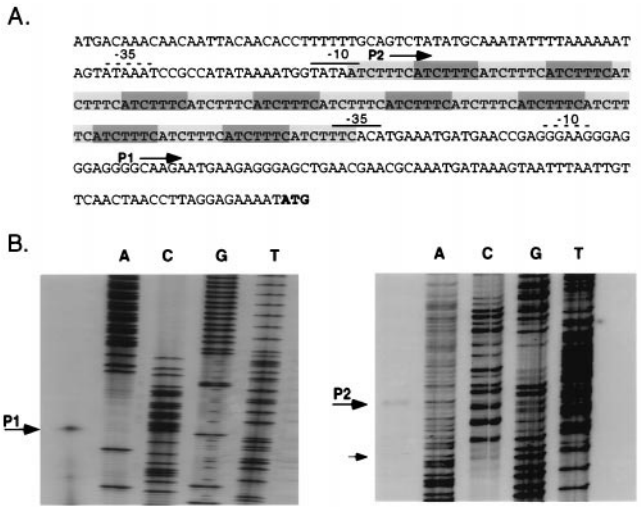

FIG. 1. Diagrammatic representation of the *hmwA* promoters and primer extension results. (*A*) The 311 nucleotides upstream of the *hmwA* start codon are shown. The 7-bp tandem repeats are highlighted by gray shadowing, the start codon is indicated in bold, and the two transcriptional start sites, P1 and P2, are denoted by arrows. The putative −10 and −35 regions for both promoters are delineated by lines (similar to σ^{70} consensus sequences) or dotted lines (lacking homology to σ^{70} consensus sequences) above the relevant sequence. (*B*) Primer extension results for RNA derived from DH5α expressing the *hmw1A* gene from a multicopy plasmid. The small arrow below P2 is a minor product that is smaller by two repeat units.

the sequence just upstream of P2 revealed a classic σ^{70} −10 sequence but no identifiable −35 region. Thus, the tandem repeats lie between two separate transcription initiation sites: P2, which should give rise to a transcript that contains the repeats, and P1, which should generate a transcript that lacks the repeats.

Increase in Repeat Number Correlates with Decrease in Protein Expression in Animal Isolates. In experiments with chinchillas, Barenkamp (19) found that among animals previously immunized with HMW1 and then challenged with *H. influenzae* strain 12, colonies recovered from the middle ear fluid produced relatively diminished quantities of HMW1. To begin to define the mechanism of down-modulation of HMW1 expression, we selected a representative variant, designated 12.267D, for more detailed studies. Initially, variant 12.267D was inoculated onto chocolate agar, and individual colonies were selected for analysis by nested PCR. By using this approach, we identified three different colonies containing 19, 21, and 22 repeats, respectively (data not shown). Each of these colonies was inoculated into broth and incubated to mid-log phase, then sonicated and subjected to Western analysis with mAb 10C5, which reacts specifically with HMW1. As shown in Fig. 2*A*, examination of comparable amounts of total protein revealed an inverse correlation between the number of repeats within the *hmw1A* promoter and the quantity of HMW1 produced.

In additional experiments we examined strain 12 variants recovered from chinchillas previously immunized with HMW2. Compared with wild-type strain 12, these variants expressed normal amounts of HMW1 but diminished quantities of HMW2 (data not shown). We chose one representative variant, designated 12.234, for further study. After inoculation onto chocolate agar, isolated colonies again were picked for analysis by nested PCR, this time with primers that resulted in amplification of only the *hmw2A* promoter. In all cases, colonies contained more than 17 repeats in the *hmw2A* promoter, with the number ranging between 23 and 28. Similar to the situation with HMW1, there was an inverse correlation

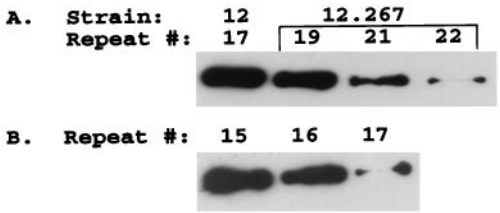

FIG. 2. Relationship between number of repeats in the *hmw1A* and *hmw2A* promoters and quantity of HMW1 and HMW2 produced. (*A*) Sonicates prepared from variants of wild-type strain 12 and chinchilla isolate 12.267 containing different numbers of repeats in the *hmw1A* promoter were examined by Western blot with mAb 10C5, which recognizes HMW1. (*B*) Sonicates of strain 12 variants containing different numbers of repeats in the *hmw2A* promoter were examined by Western blot with mAb AD6, which recognizes HMW2 and shows minimal reactivity with HMW1. Equal quantities of total protein were loaded in each lane.

between the number of repeats in the *hmw2A* promoter and the quantity of HMW2 produced (data not shown).

Variation in Repeat Number Occurs During Growth *in Vitro*. To determine whether the number of repeats in the *hmw1A* and *hmw2A* promoters varies at a detectable frequency when organisms are grown in the absence of immune selection, we inoculated a single colony of wild-type strain 12 onto chocolate agar, then picked 48 of the resulting colonies for analysis by PCR as described above. The original colony contained 16 repeats in the *hmw2A* promoter. Among the 48 derivatives, we isolated colonies containing 15, 16, and 17 repeats in the *hmw2A* promoter. Analysis of the *hmw1A* promoter also revealed variation in repeat number, though at a somewhat lower frequency. By using mAb AD6, which reacts preferentially with HMW2, Western analysis of HMW2 variants with 15, 16, or 17 repeats demonstrated an inverse correlation between the number of repeats in the *hmw2A* promoter and the quantity of HMW2 produced (Fig. 2*B*). Similarly, based on Western analysis with mAb 10C5, HMW1 variants with 16 repeats produced more HMW1 than did colonies with 17 repeats (not shown).

In other bacterial systems characterized by tandem nucleotide repeats that vary in number, addition or subtraction of a repeat is believed to occur via slipped-strand mispairing, a process that is independent of RecA and homologous recombination (32). To examine whether RecA is required for variation in repeat number in the *hmw1A* and *hmw2A* promoters, we insertionally inactivated the *recA* locus (also called *rec-1* in *H. influenzae*) in *H. influenzae* strain 12. Subsequently, we picked individual colonies and performed colony PCR to determine the number of repeats preceding *hmw2A*. The frequency of variation in repeat number was similar for wild-type strain 12 and the strain 12 RecA mutant (data not shown), arguing against a role for RecA and homologous recombination.

Increase in Repeat Number in the *hmw1A* and *hmw2A* Promoters Correlates with a Decrease in Specific mRNA Levels. To examine the relationship between quantity of protein and level of specific mRNA among strain 12 variants with different numbers of repeats in the *hmw2A* promoter, we performed Northern analysis on the *in vitro* variants with 15 or 17 repeats and on variants of chinchilla isolate 12.234 with 23, 26, and 28 repeats. To address transcription of *hmw2A* independent of *hmw1A*, the *hmw1A* promoter and 5′ coding region first were deleted in each variant by transforming with pHMW1–14Δ*hmw1p* and selecting for kanamycin resistance. PCR and Western analysis of the resulting transformants confirmed maintenance of 15, 17, 23, 26, and 28 repeats, and an inverse correlation between the number of repeats and the level of HMW2 (Fig. 3*A*). As shown in Fig. 3*B*, Northern analysis with a probe corresponding to an internal fragment of

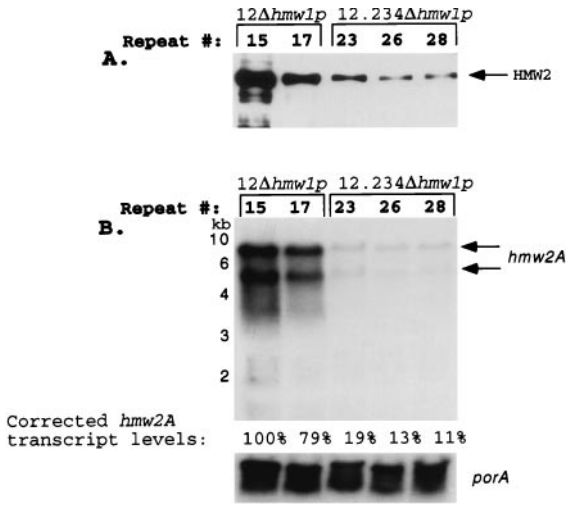

FIG. 3. Relationship between number of repeats in the *hmw2A* promoter, quantity of HMW2 produced, and the level of *hmw2A* transcript. In the blots in *A* and *B*, lanes were loaded as labeled with either 12Δ*hmw1p* variants containing 15 and 17 repeats or 12.234Δ*hmw1p* variants containing 23, 26, and 28 repeats. (*A*) Variants were analyzed for levels of HMW2 by Western blot by using the polyclonal antiserum 28G. (*B*) Variants were analyzed for the level of *hmw2A* mRNA by Northern blot, probing with a specific internal fragment of *hmw2A*. The resulting signals were quantitated by a PhosphorImager, and the *hmw2A* transcript levels were corrected for slight differences in loading based on the *porA* transcript, which is expressed constitutively. The percentages were determined by setting the isolate with 15 repeats at 100% and comparing the relative intensity of each subsequent isolate.

hmw2A revealed two separate bands for all five variants, one corresponding to the size of *hmw2A* plus the downstream *hmw2B* gene (≈8 kb) and the other matching the size of *hmw2A* alone (≈5 kb). Probing with a fragment of the *H. influenzae porA* gene, which is constitutively expressed (27), allowed quantitation of the *hmw2A* transcripts, with adjustment for small differences in RNA loading. Transcript levels were maximal in the variant with 15 repeats and decreased progressively with each increment in repeat number (note corrected *hmw2A* transcript levels, Fig. 3*B*). Interestingly, there was little difference in transcript levels between the variant with 26 repeats and the variant with 28 repeats, suggesting that the effect of variation in repeat number tapers off as the repeat number approaches 26.

In additional experiments, we performed Northern analysis on the strain 12 variants with either 17 or 22 repeats in the *hmw1A* promoter. Consistent with our results related to *hmw2A*, *hmw1A* transcript levels were markedly reduced in the variant with 22 repeats (data not shown).

Both P1 and P2 Are Affected by Variation in Repeat Number. In considering the observation that the number of repeats in the *hmw1A* and *hmw2A* promoters is inversely correlated with the level of specific mRNA, we wondered whether both the P1 and P2 transcriptional start sites are involved. To determine the effect of changes in repeat number on P1 and P2, we used the strain 12 variant that contains 15 repeats in the *hmw2A* promoter but lacks the *hmw1A* promoter and 5′ coding sequence (12Δ*hmw1p*). For comparison, we used the 12.234 variants that contain 23 or 26 repeats in the *hmw2A* promoter and again lack the *hmw1A* promoter region (12.234Δ*hmw1p*). RNA was isolated from these variants, and the *hmw2A* promoter was examined by primer extension. As shown in Fig. 4, both P1 and P2 were affected by a change in the number of repeats in the *hmw2A* promoter. In both cases, increasing repeat number was associated with a decrease in the level of transcription.

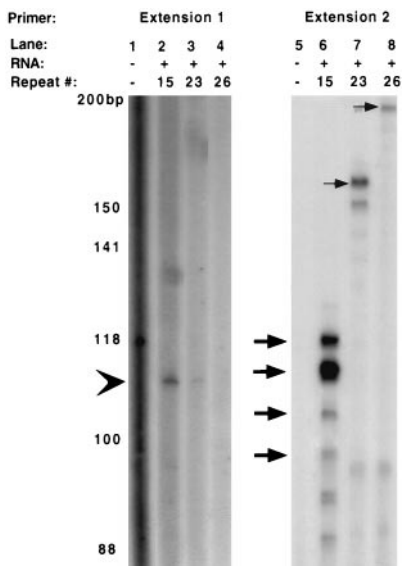

FIG. 4. Primer extension on strain 12 derivatives containing different numbers of repeats, comparing the effect of increasing repeat number on P1 and P2. The gel was loaded as follows: lanes 1 and 5, reactions without RNA; lanes 2 and 6, reactions using RNA from 12Δ*hmw1p* with 15 repeats; lanes 3 and 7, reactions using RNA from 12.234Δ*hmw1p* with 23 repeats; lanes 4 and 8, reactions using RNA from 12.234Δ*hmw1p* with 26 repeats. The samples in lanes 1–4 were analyzed for P1 levels by using the primer designated extension 1, and samples in lanes 5–8 were analyzed for P2 levels by using the primer designated extension 2. The arrowhead shows the location of P1. The large arrows show the ladder of P2 products separated by single repeat units. The smaller arrows show the location of P2 in the isolates with 23 and 26 repeats, respectively. The minor bands at 95 bp in lanes 7 and 8 are most likely artifactual, because they would imply transcription initiation well within the repeat region, and they are absent when primer extension is performed with other primers (not shown).

Increasing Repeat Number Correlates with Decreasing β-Galactosidase Levels in Reporter Fusion Constructs in *H. influenzae* Strain DB117. To study the *hmwA* promoters more quantitatively, we constructed an *E. coli–H. influenzae* shuttle vector designated pSD100, which harbors a promoterless *lacZ* gene. By using PCR, *hmw1A* promoters containing either 17 or 22 repeats were amplified together with the *hmw1A* ribosomal binding site and start codon and were inserted upstream of *lacZ* in pSD100, generating pSD117 (17 repeats) and pSD122 (22 repeats). Each of these plasmids was introduced into DB117, and the resulting transformants were examined for β-galactosidase activity by using the Miller assay (33). DB117 containing the plasmid with 17 repeats had nearly 5-fold more activity than did DB117 carrying the plasmid with 22 repeats (803.2 ± 15.6 Miller units vs. 136.3 ± 2.9 Miller units). The control plasmid, pSD201, had minimal activity (6.74 ± 1.7 Miller units), thus eliminating the possibility of upstream effects from a cryptic promoter.

Variation in Repeat Number Occurs During Natural Human Infection. To assess whether variation in repeat number in the *hmw1A* and *hmw2A* promoters occurs during natural infection, we characterized matched middle ear and nasopharyngeal isolates of nontypable *H. influenzae* recovered from two children with acute middle ear infection. In previous work we established that both isolates from patient FL represent a single strain and that both isolates from patient RA represent a separate strain (22). As shown in Fig. 5*A*, by using a polyclonal antiserum that recognizes both HMW1 and HMW2, we detected only a single reactive band in all four isolates, suggesting that these strains express either two HMW adhesins of similar size or only one HMW adhesin. Interestingly, with both FL and RA, the middle ear isolate produced

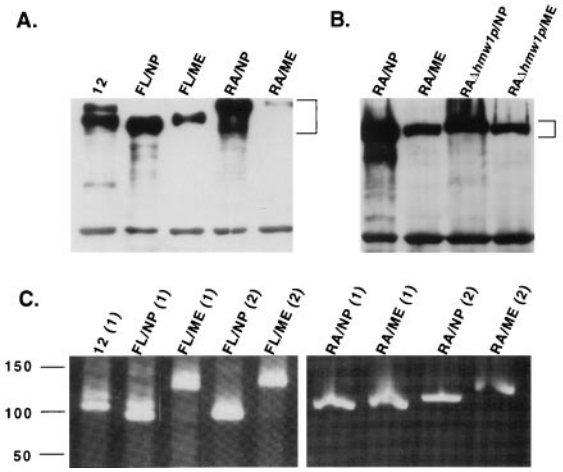

FIG. 5. Analysis of matched clinical isolates of nontypable *H. influenzae* from two different patients. Isolates FL/NP and FL/ME were recovered from the nasopharynx and middle ear, respectively, of patient FL. Isolates RA/NP and RA/ME were recovered from the nasopharynx and middle ear, respectively, of patient RA. (*A*) Western blot of sonicates of *H. influenzae* strain 12 and the four clinical isolates, using polyclonal antiserum 28G, which recognizes both HMW1 and HMW2. The bracket indicates the location of the HMW proteins. (*B*) Western blot showing the pattern of HMW expression after deletion of the *hmw1A* promoter in the isolates from patient RA by using the polyclonal antiserum 28G. (*C*) Agarose gel showing the sizes of the PCR products generated from the four clinical isolates by using primers that amplify the repeat region in the *hmw1A* promoter (1) or the *hmw2A* promoter (2). Samples were resolved next to a 50-bp ladder, using the strain 12 *hmw1A* locus as a control, to determine the number of repeats present.

substantially less HMW protein than the matched nasopharyngeal isolate.

To determine the number of repeats in both the *hmw1A* and the *hmw2A* promoters in the paired isolates, we performed colony PCR with two new sets of primers. Because there is some degree of sequence heterogeneity in the *hmwA* structural genes among diverse strains, in one case we used a 5′ primer within the ORF immediately upstream of *hmw1A* and a 3′ primer just downstream of the repeats, and in the second case we used a 5′ primer within the ORF upstream of *hmw2A* and the same 3′ primer downstream of the repeats. This approach allowed us to separate the *hmw1A* promoter from the *hmw2A* promoter by virtue of the corresponding upstream ORF. As shown in Fig. 5*C*, the middle ear isolate from patient FL contained 21 repeats in both the *hmw1A* and the *hmw2A* promoters, whereas the paired nasopharyngeal isolate contained only 15 repeats in these promoters. In the case of patient RA, the middle ear isolate contained 17 repeats in the *hmw1A* promoter and 19 repeats in the *hmw2A* promoter, whereas the matched nasopharyngeal isolate contained 16 repeats in both the *hmw1A* and *hmw2A* promoters. Because the polyclonal antiserum used to analyze these isolates does not distinguish between HMW1 and HMW2, we deleted the *hmw1A* promoter and 5′ coding region in both isolates from patient RA, intending to confirm that the decrease in detectable HMW protein was a result of increased repeats at the *hmw2* locus. Examination of the resulting strains by PCR confirmed maintenance of the same repeat pattern seen in the parent strains. As shown in Fig. 5*B*, the pattern of reactivity by Western analysis was identical for the deletion mutants and the wild-type isolates, suggesting that the increase in repeat number in the *hmw2A* promoter is at least partly responsible for the decrease in protein expression. Thus, variation in the number of repeats in the *hmwA* promoters occurs during human infection, again with an inverse correlation between the number of repeats and the level of detectable protein.

DISCUSSION

During the course of natural infection, pathogenic bacteria encounter varied and changing environments. To survive these diverse conditions, many pathogens have developed mechanisms that facilitate efficient adaptation. Phase variation represents one such mechanism and is characterized by the reversible loss or gain of a defined structure. In most cases, the involved structure is expressed on the surface of the organism and varies between two states, namely off and on (34, 35). Sometimes, three states exist, including off (−), weakly on (+), and strongly on (+++) (35, 36). In our study, we found that the *H. influenzae* HMW1 and HMW2 adhesins are subject to phase variation. In contrast to other known examples, the phase-variable expression of HMW1 and HMW2 involves multiple states that range from very weakly on to very strongly with a series of gradations in between.

In considering bacterial systems in which tandem repeats affect gene expression, three mechanisms have been described previously. In the most common situation, tandem repeats are located within the coding region of a gene, and addition or subtraction of a repeat unit alters the reading frame. Examples include the *opa* locus in *Neisseria gonorrhoeae* (37) and the *lic* loci in *H. influenzae* (35). The second mechanism is unique to the *H. influenzae* pilus gene cluster, which contains a series of dinucleotide repeats in the overlapping −10 and −35 regions of the divergent promoters of *hifA* and *hifB*, two genes essential for pilus biogenesis (36). In this case addition or subtraction of a TA repeat unit changes the spacing between the −10 and −35 regions and alters promoter affinity for RNA polymerase. The third mechanism has been identified only in the *Vibrio cholerae ctxA* gene, a gene that is characterized by 7-bp tandem repeats upstream of the start site for transcription (38). These repeats are located far upstream of the RNA polymerase binding site and influence the binding of ToxR, a positive regulator of toxin expression, to a region downstream of the repeats (39). Most isolates of *V. cholerae* contain eight repeats and express normal quantities of cholera toxin, but in isolates with only three repeats, expression of cholera toxin is minimal. Our studies of the HMW1 and HMW2 adhesins demonstrate a unique system of phase variation. In particular, 7-bp repeats are present in the promoters of the HMW1 and HMW2 structural genes (*hmw1A* and *hmw2A*) and exert a step-wise repressive effect on gene expression and protein production. Increases in repeat number result in a decrease in specific mRNA levels and protein expression, and decreases in repeat number are associated with an increase in specific mRNA expression and protein production.

Analysis of the *hmw1A* and *hmw2A* promoters demonstrated two transcriptional start sites, designated P1 and P2. P1 lies downstream of the repeats, whereas P2 is located within the 5′-most repeat. By using primer extension on individual isolates lacking the *hmw1A* promoter and containing a variable number of repeats in the *hmw2A* promoter, we established that addition of repeat units diminishes the activity of both P1 and P2. This observation suggests several potential explanations for the mechanism by which the repeats down-modulate expression of *hmw1A* and *hmw2A*. One possibility is that the repeats are recognized by a negative regulator, which interferes with RNA polymerase binding to both the P1 and P2 promoters. Consistent with this consideration, at least for P1, the 3′-most repeat overlaps with the −35 region of the P1 promoter. As a second possibility, the sequence of the repeats may form a tertiary structure that interferes with RNA polymerase affinity for the two promoters, with an increase in repeat number resulting in more efficient interference. Alternatively, the repeats may act differently at P1 and P2. For

example, P2 may be affected by repeat number via binding of a negative regulator, whereas P1 activity is determined by the distance between the P1 RNA polymerase binding site and an activator. Still another possibility is that the repeats serve to destabilize the message originating at P2.

To examine whether variation in repeat number in the *hmw1A* and *hmw2A* promoters occurs during natural *H. influenzae* disease, we characterized paired nasopharyngeal and middle ear isolates of *H. influenzae* from two different children with acute middle ear infection. Interestingly, in both of these patients we found that the middle ear isolate contained an increased number of repeats and produced less protein than the matched nasopharyngeal isolate, thus mimicking observations in immunized chinchillas with experimental otitis media. Previous work has established that human infection with nontypable *H. influenzae* results in a vigorous antibody response against the HMW proteins (18). Furthermore, middle ear fluid from children with acute otitis media is known to contain high titers of serum-derived antibody (40, 41). Thus, as the organism spreads from the nasopharynx to the middle ear, it is likely to encounter an environment rich in antibody against HMW1 and HMW2. With this information in mind, our results suggest that in the middle ear organisms expressing relatively little HMW1 and HMW2 have a survival advantage and are able to persist, whereas organisms expressing high levels of HMW1 and HMW2 are eliminated. Thus, variation of HMW expression may enable the organism to balance the need to colonize the respiratory tract against the ability to evade the immune response.

In summary, we have found that the *H. influenzae* HMW1 and HMW2 adhesins undergo phase variation. The system of phase variation is unique in bacteria and involves changes in the number of 7-bp repeats in the *hmw1A* and *hmw2A* promoters, with an inverse, step-wise relationship between the number of repeats and the level of protein expression. We speculate that phase variation of HMW1 and HMW2 plays a fundamental role in enabling the organism to survive in diverse environments, including the nasopharynx early during infection and the middle ear at later points.

We thank M. Caparon for critical reading of the manuscript. This work was supported by Public Health Service Grant 1RO1 DC-02873 from the National Institute on Deafness and Other Communication Disorders (J.W.S.), by a training grant from the National Institute of Allergy and Infectious Diseases (S.D.), and by funds from Pasteur-Merieux Connaught Canada (J.W.S. and S.J.B.).

1. Kashi, Y., King, D. & Soller, M. (1997) *Trends Genet.* **13,** 74–78.
2. Johnson, A. C., Jinno, Y. & Merlino, G. T. (1988) *Mol. Cell. Biol.* **8,** 4174–4184.
3. Chung, Y. T. & Keller, E. B. (1990) *Mol. Cell. Biol.* **10,** 206–216.
4. McKeon, C., Schmidt, A. & de Crombrugghe, B. (1984) *J. Biol. Chem.* **259,** 6636–6640.
5. Suen, T. C. & Hung, M. C. (1990) *Mol. Cell. Biol.* **10,** 6306–6315.
6. Sandaltzopoulos, R., Mitchelmore, C., Bonte, E., Wall, G. & Becker, P. B. (1995) *Nucleic Acids Res.* **23,** 2479–2487.
7. Russell, D. W., Smith, M., Cox, D., Williamson, V. M. & Young, E. T. (1983) *Nature (London)* **304,** 652–654.
8. Turk, D. C. (1984) *J. Med. Microbiol.* **18,** 1–16.
9. St. Geme, J. W., III (1993) *Infect. Agents Dis.* **2,** 1–16.
10. Murphy, T. F. & Apicella, M. A. (1987) *Rev. Infect. Dis.* **9,** 1–15.
11. Barenkamp, S. J. & Leininger, E. (1992) *Infect. Immun.* **60,** 1302–1313.
12. Barenkamp, S. J. & St. Geme, J. W., III (1996) *Mol. Microbiol.* **19,** 1215–1223.
13. Brinton, C. C., Jr., Carter, M. J., Derber, D. B., Kar, S., Kramarik, J. A., To, A. C.-C., To, S.C.-M. & Wood, S. W. (1989) *Pediatr. Infect. Dis. J.* **8,** S54–S61.
14. Geluk, F., Eijk, P. P., van Ham, S. M., Jansen, H. M. & van Alphen, L. (1998) *Infect. Immun.* **66,** 406–417.
15. Barenkamp, S. J. & St. Geme, J. W., III (1994) *Infect. Immun.* **62,** 3320–3328.
16. St. Geme, J. W., III & Grass, S. (1998) *Mol. Microbiol.* **27,** 617–630.
17. Hultgren, S. J., Abraham, S., Caparon, M., Falk, P., St. Geme, J. W. & Normark, S. (1993) *Cell* **73,** 887–901.
18. Barenkamp, S. J. & Bodor, F. F. (1990) *Pediatr. Infect. Dis. J.* **9,** 333–339.
19. Barenkamp, S. J. (1996) *Infect. Immun.* **64,** 1246–1251.
20. Hoiseth, S. K., Moxon, E. R. & Silver, R. P. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 1106–1110.
21. Setlow, J. K., Brown, D. C., Boling, M. E., Mattingly, A. & Gordon, M. P. (1968) *J. Bacteriol.* **95,** 546–558.
22. Krasan, G. P., Cutter, D., Block, S. L. & St. Geme, J. W., III (1999) *Infect. Immun.,* in press.
23. St. Geme, J. W., III & Falkow, S. (1990) *Infect. Immun.* **58,** 4036–4044.
24. Anderson, P., Johnston, R. B., Jr. & Smith, D. H. (1972) *J. Clin. Invest.* **51,** 31–38.
25. Laemmli, U. K. (1970) *Nature (London)* **227,** 680–685.
26. Ausubel, F. M., Kingston, R., Moore, D., Seidman, J., Smith, J. & Struhl, K., eds. (1998) *Current Protocols in Molecular Biology* (Wiley, New York).
27. Tomb, J.-F. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 10252–10256.
28. Oka, A., Sugisaki, H. & Takanami, M. (1981) *J. Mol. Biol.* **147,** 217–226.
29. Herriott, R. M., Meyer, E. M. & Vogt, M. (1970) *J. Bacteriol.* **101,** 517–524.
30. Tomb, J. F., Barcak, J. J., Chandler, M. S., Redfield, R. J. & Smith, H. O. (1989) *J. Bacteriol.* **171,** 3796–3802.
31. Simons, R. W., Houman, F. & Kleckner, N. (1987) *Gene* **53,** 85–96.
32. Levinson, G. & Gutman, G. A. (1987) *Mol. Biol. Evol.* **4,** 203–221.
33. Miller, J. H. (1977) *Experiments in Molecular Genetics* (Cold Spring Harbor Lab. Press, Plainview, NY).
34. Stern, A., Brown, M., Nickel, P. & Meyer, T. F. (1986) *Cell* **47,** 61–71.
35. Weiser, J. N., Maskell, D. J., Butler, P. D., Lindberg, A. A. & Moxon, E. R. (1990) *J. Bacteriol.* **172,** 3304–3309.
36. van Ham, S. M., van Alphen, L., Mooi, F. R. & van Putten, J. P. M. (1993) *Cell* **73,** 1187–1196.
37. Murphy, G. L., Connell, T., Barritt, D. S., Koomey, M. & Cannon, J. G. (1989) *Cell* **56,** 539–547.
38. Miller, V. L., Taylor, R. K. & Mekalanos, J. J. (1987) *Cell* **48,** 271–279.
39. Pfau, J. D. & Taylor, R. K. (1996) *Mol. Microbiol.* **20,** 213–222.
40. Faden, H., Brodsky, L. & Bernstein, J. (1989) *Infect. Immun.* **1983,** 297–304.
41. Sloyer, J. L., Cate, C. C., Howie, V. M., Ploussard, J. H. & Johnston, R. B., Jr. (1975) *J. Infect. Dis.* **151,** 273–279.