Sequence Stability of the Gene Encoding Outer Membrane Protein P2 of Nontypeable Haemophilus influenzae in the Human Respiratory Tract

Thomas J. Hiltke, Sanjay Sethi, and Timothy F. Murphy

Nontypeable Haemophilus influenzae (NTHI) is a major cause of otitis media in children and of acute exacerbation in adult patients with chronic obstructive pulmonary disease (COPD) [1–3]. The cost of managing infections caused by NTHI is substantial [4, 5]. Therefore, a vaccine to prevent NTHI infection in humans is desired. Realization of this goal, however, has been impeded by the extensive genetic diversity of NTHI strains [6, 7]. Most notably, the outer membrane protein (OMP) expression pattern is highly diverse among strains of NTHI [8].

P2 is the major OMP of NTHI, constituting as much as 50% of the protein of the outer membrane, and it is a target of bactericidal antibodies [9, 10]. The gene encoding P2 (ompP2) displays marked heterogeneity among strains [11, 12]. Almost all diversity occurs in the regions encoding 8 surface-exposed loops [11]. Some of these loops, however, are relatively conserved and may be useful in generating antibodies that bind to P2 of many strains. We recently reported the induction and purification of antibodies to a peptide mimic of the loop-6 region of a strain of NTHI and showed functional activity of these antibodies against other strains [13]. The success of a P2 loop–based vaccine, however, would rely on the stability of ompP2 loop sequences during colonization. The goal of the present study is to examine the intrastrain variability of ompP2 during colonization of the human respiratory tract.

The Buffalo Veterans Affairs Medical Center COPD study clinic is investigating the role of bacterial colonization and infection during the progression of COPD in the adult. The study has generated a collection of sets of NTHI strains that have been isolated during sequential visits from the same patient and subjected to molecular typing. In this report, we examine the temporal intrastrain variability of ompP2 obtained from 13 sets of strains that have continuously colonized patients with COPD for varying lengths of time.

Materials and Methods

Bacteria and culture conditions. All isolates of NTHI were maintained on chocolate agar plates at 35°C in a 5% CO₂ atmosphere. NTHI were also grown in brain-heart infusion broth supplemented with hemin and nicotinamide adenine dinucleotide (both at 10 µg/mL) at 35°C–37°C.

Collection of isolates. Detailed description of the Buffalo COPD clinic and the collection of isolates can be found elsewhere [14]. In brief, spontaneous morning sputum samples were collected and homogenized by incubation at 37°C for 15 min with an equal volume of 0.1% dithiothreitol (Sputolysin; Calbiochem). Serial dilutions of homogenized sputum were prepared in PBS and plated on blood, chocolate, and MacConkey agar plates. Bacteria were identified by standard techniques. If H. influenzae was present, we attempted to isolate and characterize as many as 10 individual colonies. All isolates were stored at −70°C. Isolate names refer to patient, visit, and colony number: for example, isolate 23P2H1 refers to patient 23, clinic visit 2, and H. influenzae colony 1. Identical iso-
lates were identified by OMP profiles in SDS-PAGE, as described elsewhere [15]. Isolate identity was confirmed by use of pulsed-field gel electrophoresis analysis of Smal-digested DNA [16].

Polymerase chain reaction (PCR) and sequencing. Genomic DNA was isolated from 400 μL of an overnight NTHI broth culture by use of the Wizard genomic DNA purification kit (Promega Biotech). Approximately 75 ng of DNA was amplified with synthetic primers (Genosys Biotechnologies; table 1) and the reagents provided in the GeneAmp kit (Applied Biosystems). Alternatively, 1 colony of NTHI was suspended in 100 μL of sterile water, and 1 μL of this suspension was used as the PCR template. Reactions were done in either a GeneAmp PCR System 9600 thermocycler (PerkinElmer) or a Mastercycler personal cycler (Eppendorf). The reactions consisted of a hold for 3 min at 94°C, followed by 35 cycles of 94°C for 30 s, 58°C for 30 s, and 72°C for 90 s. Amplicons were purified with a PCR purification kit (Qiagen) and subjected to automated DNA sequencing (Biopolymer Facility, Roswell Park Cancer Institute, Buffalo, NY) with use of synthetic primers (Genosys Biotechnologies; table 1).

The PCR primers were derived from the sequence of ompP2 from strain 1479 (GenBank accession no. M93268) to amplify specifically the region between and including the codon that encodes the first amino acid of the mature protein (GCT) and the termination codon (TAA) [11]. In ompP2 of strain 1479, this corresponds to the region between bp 61 and 79.

Long-run SDS-PAGE. Proteins in whole-cell lysates were separated overnight on 15% polyacrylamide gels (16 × 16 cm plates) in a Protean II xiCell (BioRad) at 4°C and 125 V. Electrophoresis was continued until the prestained 28-kDa marker (BioRad) was near the bottom of the gel.

Nucleotide sequence accession numbers. All ompP2 sequences were submitted to GenBank, and the following accession numbers were issued: AY051383 (isolate 13P24H1), AY051384 (13P36H1), AY051385 (14P14H1), AY051386 (14P1H1), AY051387 (14P-23H1), AY051388 (14P5H1), AY051389 (18P1H1), AY051390 (18P4H1), AY051391 (22P1H1), AY051392 (22P5H1), AY051393 (23P2H1), AY051394 (25P4H1), AY051395 (31P12H1), AY051396 (31P7H1), AY051397 (47P12H1), AY051398 (47P17H1), AY051399 (48P29H1), AY051400 (48P45H1), AY051401 (54P24H1), AY051402 (54P33H1), AY051403 (5P-30H1), AY051404 (5P41H1), AY051405 (73P2H1), AY051406 (73P4H1), AY051407 (74P15H1), AY051408 (74P1H1), and AY051409 (74P4H1).

### Results

**Molecular typing of isolates.** Twelve pairs of isolates and 1 set of triplet isolates were selected for this study, allowing us to compare 14 pairs of isolates. The duration of colonization (i.e., the length of time between collection of the first and the second isolate from the patient) ranged from 2 to 16 months (mean, 7 months; table 2). That both isolates in a pair were of the same strain was confirmed by comparison of OMP profiles, using SDS-PAGE (data not shown). To further confirm isolate identity, genomic DNA from each isolate was digested with SmaI and subjected to pulsed-field gel electrophoresis. Each pair had banding patterns consistent with isolate identity (figure 1). Therefore, each pair represented a single isolate that had persistently colonized the human host.

**ompP2 sequence analysis.** Examination of the sequence of ompP2 derived from the isolates revealed little sequence heterogeneity within each set of isolates. In 9 pairs of isolates (duration of colonization, 2–12 months), the ompP2 sequences were identical in every base, indicating that ompP2 was stable for as long as 12 months in vivo (table 2).

Every change observed in 4 sets of isolates was nonsynonymous (each change resulted in an alteration of the predicted amino acid sequence of the mature protein; figure 2). Furthermore, all changes were localized exclusively to regions encoding putative surface-exposed loops (figure 2). In each case, changes occurred in areas of repetitive DNA. For example, most of the changes in ompP2 that was derived from isolates 18P1H1 and 18P4H1 (duration of colonization, 3 months) occurred in an

### Table 2. Comparison of ompP2 sequences among nontypeable Haemophilus influenzae isolates.

<table>
<thead>
<tr>
<th>Isolate pairs</th>
<th>Duration of colonization, months</th>
<th>Changes in ompP2 sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>23P2H1/23P4H1</td>
<td></td>
<td>Unchanged</td>
</tr>
<tr>
<td>73P2H1/73P4H1</td>
<td></td>
<td>Unchanged</td>
</tr>
<tr>
<td>18P1H1/18P4H1</td>
<td>3</td>
<td>4 Point mutations</td>
</tr>
<tr>
<td>14P1H1/14P5H1</td>
<td>4</td>
<td>Unchanged</td>
</tr>
<tr>
<td>22P1H1/22P5H1</td>
<td>4</td>
<td>Unchanged</td>
</tr>
<tr>
<td>47P12H1/47P17H1</td>
<td>5</td>
<td>Unchanged</td>
</tr>
<tr>
<td>31P7H7/31P12H1</td>
<td>6</td>
<td>Unchanged</td>
</tr>
<tr>
<td>14P14H1/14P23H1</td>
<td>9</td>
<td>Unchanged</td>
</tr>
<tr>
<td>54P24H1/54P33H1</td>
<td>10</td>
<td>Unchanged</td>
</tr>
<tr>
<td>5P30H1/5P41H1</td>
<td>10</td>
<td>Addition of 2 repeats in loop 8</td>
</tr>
<tr>
<td>13P24H1/13P36H1</td>
<td>12</td>
<td>Unchanged</td>
</tr>
<tr>
<td>74P1H1/74P4H1</td>
<td>4</td>
<td>Lost 3 repeats in loop 4</td>
</tr>
<tr>
<td>74P4H1/74P15H1</td>
<td>11</td>
<td>Regained repeats in loop 4</td>
</tr>
<tr>
<td>48P29H1/48P45H1</td>
<td>16</td>
<td>Multiple changes</td>
</tr>
</tbody>
</table>
ACC repeat region of DNA encoding loop 6. Similarly, the ompP2 sequences of isolates 5P30H1 and 5P41H1 (duration of colonization, 10 months) differed only in the insertion of 2 additional copies of the hexanucleotide TAGAAC in the region encoding loop 8 (figure 2). A remarkable modification was observed in the ompP2 sequence from patient 74: During a 4-month colonization period, isolate 74P4H1 had lost 3 CTG repeats in the region encoding loop 4. Apparently, these repeats were regained at some time during the next 10 months of colonization, so that ompP2 sequences from isolates 74P1H1 and 74P15H1 were identical (figure 2).

Changes in the region encoding loop 4 were also observed in ompP2 sequences obtained from isolates 48P29H1 and 48P45H1 (duration of colonization, 16 months) in the form of an insertion and 5 point mutations in the area of imperfectly repeated CTG residues (figure 2). If additional changes in regions encoding loops 1 and 5 are included, ompP2 obtained from these 2 isolates contained the most changes of any pair studied.

To exclude the possibility that PCR or automated sequencing artifacts caused the observed differences, PCR and sequence analysis were repeated for all pairs of isolates that showed ompP2 differences, using bacteria from the original freezer stocks as starting material. This analysis showed sequence changes identical to those depicted in figure 2 in all 4 sets of isolates.

**SDS-PAGE analysis.** To determine whether P2 phenotype reflected the changes observed in ompP2 sequences, whole-cell lysates were subjected to long-run SDS-PAGE analysis. This technique allowed us to visualize small changes in the mobility of P2 that were not apparent on the standard gels used to determine strain identity. As expected, P2 from isolate 5P41H1 had a slower mobility in SDS-PAGE than did P2 from 5P30H1, because 5P41H1 contained 4 additional amino acids, as determined by ompP2 sequence analysis (figure 3). Shifts in mobility were also detected in P2 from isolates 18P1H1 and 18P4H1 and isolates 48P29H1 and 48P45H1, as predicted by the DNA sequence data. Most notably, the mobility of P2 of isolate 74P4H1 was slightly faster than that of P2 of isolates 74P1H1 and 74P15H1, reflecting the loss of 3 Ala residues in loop 4. The data showed that the changes in P2 that would be expected on the basis of ompP2 sequence analysis were observable in the whole organism, indicating that the changes detected in ompP2 that occurred during colonization of the human host were not due to artifacts of PCR or sequencing.

**Discussion**

In this study, the stability of ompP2 in NTHI during colonization of the human host was examined. No changes in ompP2 were detected in 9 of 13 sets of isolates that colonized the human respiratory tract for 2–12 months. In 4 isolates, nonsynonymous point mutations, deletions, and insertions occurred in regions encoding surface-exposed loops. Most of the changes in ompP2 were confined to regions of repetitive DNA, suggesting that this type of DNA has a role in the antigenic variation of P2. We conclude that ompP2 is relatively stable in NTHI during human colonization.

Comparisons of ompP2 from several epidemiologically unrelated strains of NTHI have clearly established that ompP2 is
highly diverse in the regions that encode several of the 8 potentially surface-exposed loops of the protein [11, 12, 17]. P2 is also a major target of strain-specific bactericidal antibodies present in serum samples from patients with COPD [10]. These data suggest that polymorphism in ompP2 may be a mechanism of immune evasion. The use of P2 in vaccines designed to prevent NTHI colonization or infection, therefore, has not been actively pursued because of heterogeneity of P2 among strains. Closer examination, however, of ompP2 sequences has revealed that regions that encode selected loops are relatively conserved among strains. We recently reported that antibodies made in rabbits to a conserved loop-6 peptide mimic showed binding to P2 in 14 of 15 strains tested [13]. These antibodies also directed the complement-mediated killing of 8 of the 15 strains. Therefore, it is possible to breach strain specificity of the antibody response to P2, and further work in this area may lead to effective cross-strain vaccines.

P2 has been the subject of much research; however, most studies stress the diversity or antigenic variation of P2 among strains. For example, Groeneveld et al. [18] showed variation over time in P2 and other OMPs in several NTHI strains that colonized patients with COPD; however, careful examination of the data indicated that SDS-PAGE demonstrated no change in the mobility of P2 in 7 of 12 persistent strains. In a study of persistent colonization of infants with NTHI, Smith-Vaughan et al. [19] found little variation in ompP2 in 5 of 6 strains (mean duration of colonization, 5 months). This result was overshadowed by the evidence of horizontal transfer of ompP2 among the strains studied, which undoubtedly contributes to the heterogeneity of P2. Most of the data concerning antigenic drift in P2 are derived from in-depth study of NTHI strain d1 and its variants. During a 6-month colonization period, strain d1 was isolated 4 times from a patient with COPD (variants d1–d4), and a change was detected in the apparent mobility of P2 each time [18, 20]. The change in mobility was a result of the stepwise accumulation of a total of 18 point mutations and 1 codon insertion in the loop regions, most notably in the region encoding loop 6 [21]. In our study, we found less change in ompP2 than was found between strain d1 and its variants. Interestingly, we found only 3 point mutations that did not lie in regions of repetitive DNA (loop 5 of isolates 18P1H1 and 18P4H1 and loops 1 and 5 of isolates 48P29H1 and 48P45H1; figure 2). The data presented in this report are in agreement with previous findings that ompP2 remains relatively stable during colonization of the human respiratory tract [18, 19] and that NTHI strain d1 may represent a type of hypermutable strain that we did not encounter in our study.

It is becoming increasingly apparent that bacterial genomes contain substantial tracts of repetitive DNA [22]. Repetitive DNA is involved in the regulation of several bacterial virulence genes. For example, the expression of NTHI genes encoding fimbriae or lipooligosaccharide biosynthesis enzymes are turned on and off by the insertion or deletion of tandem repeats of DNA located within the genes [6]. These deletions and insertions occur mainly through strand slippage events during DNA replication, in which the 2 single strands are displaced by an integral number of repeats [6, 23]. We propose here that NTHI may also use strand slippage replication events to elicit antigenic variation of P2. In 2 patients, the hallmark of strand slippage during DNA replication—the deletion or insertion of an integral number of repeats—was observed (isolates 5P30H1 and 5P41H1 and isolates 74P1H1, 74P4H1, and 74P15H1). In 2 other patients, strand slippage may have induced point mutations, because the region of repetitive DNA was imperfect (loop 6 of isolates 18P1H1 and 18P4H1 and loop 4 of isolates 48P29H1 and 48P45H1; figure 2). Although these changes in ompP2 occurred in regions encoding surface-exposed portions of P2, whether this antigenic variation is a mechanism of immune evasion used by NTHI remains to be determined.

Several other studies have reported finding bacterial genes that have variable tandem repeats in the coding regions as a mechanism of antigenic variation; however, we know of no instance in which a trinucleotide, such as those found in the present study, is the source of the repeated DNA [24–26]. Interestingly, the relative instability of trinucleotide repeats during DNA replication, in particular CTG and CAG, has been investigated because of the involvement of these repeats in several heritable human diseases [27]. In Escherichia coli, all lengths of CTG or CAG repeats are subject to length changes mediated by strand slippage and other mechanisms [23, 27, 28]. Many of the changes in ompP2 reported here occurred in tracts of perfect or nearly perfect CTG repeats, suggesting that NTHI may use the inherent instability of this DNA as a source of antigenic variation. It must be noted, however, that we did not address experimentally whether slip-strand mispairing was the mechanism responsible for the changes in ompP2 among pairs of isolates.

In conclusion, ompP2 is relatively stable during colonization by NTHI of the human respiratory tract. Nine of 13 sets of isolates that colonized patients with COPD for 2–12 months showed no changes in ompP2 during colonization. These data are im-
portant for future studies of use of P2 as a vaccine to prevent infections due to NTHI.

Acknowledgments

We thank Norine Kuhn for performing pulsed-field gel electrophoresis analysis, Nancy Evans for sample collection, Cookie Thurston for secretarial assistance, and the volunteer patients of the Buffalo Veterans Affairs Medical Center Chronic Obstructive Pulmonary Disease study clinic. We also thank Alan Lesse for helpful criticism of the manuscript.

References