Differences in genetic and transcriptional organization of the glpTQ operons between Haemophilus influenzae type b and nontypeable strains

Song, Xin-Ming; Janson, Håkan

Published in:
Journal of Bacteriology

DOI:
10.1128/JB.185.24.7285-7290.2003

2003

JOURNAL OF BACTERIOLOGY, Dec. 2003, p. 7285–7290 Vol. 185, No. 24
0021-9193/03/$08.00+0 DOI: 10.1128/JB.185.24.7285–7290.2003

Differences in Genetic and Transcriptional Organization of the *glpTQ* Operons between *Haemophilus influenzae* Type b and Nontypeable Strains

Xin-Ming Song[1] and Håkan Janson[2]*

Vaccine & Infectious Disease Organization, University of Saskatchewan, Saskatoon, Saskatchewan, S7N 5E3 Canada,[1] and Department of Medical Microbiology, Lund University, Malmö University Hospital, S-205 02 Malmö, Sweden[2]

Received 6 June 2003/Accepted 25 September 2003

The *glpTQ* operon of *Haemophilus influenzae* type b (Hib) and nontypeable *H. influenzae* (NTHi) strains is highly conserved, except for a 1.4-kb *glpTQ* intergenic region that was found in most Hib strains. The presence of this intergenic region results in divergent *glpTQ* transcriptional profiles for Hib and NTHi where Hib strains appear to have evolved an alternative promoter for *glpQ* expression. Based on the intergenic region's low G+C content, we speculate that this DNA fragment was acquired by lateral transfer.

Haemophilus influenzae is a common pathogen, especially among children, but the clinical manifestations are largely type specific. The encapsulated *H. influenzae* serotype b (Hib) usually causes invasive infections, such as meningitis and septicemia (2), whereas the much more common nonencapsulated, or nontypeable, *H. influenzae* (NTHi) is a major cause of otitis media, sinusitis, and pneumonia (8). General vaccination against Hib has reduced the incidence of Hib infection to a near minimum (10), while attempts to construct a vaccine against the costly NTHi infections have as yet been unsuccessful due to a high genetic heterogeneity among NTHi strains (20). An extensively studied virulence factor and potential vaccine candidate in *H. influenzae* is protein D, a 42-kDa conserved lipoprotein expressed on the bacterial surface (1, 13, 21, 24). An isogenic protein D-negative mutant has been shown to be less effective than its wild-type parental strain in its ability to (i) cause experimental otitis media in rats (14), (ii) cause damage to ciliated human respiratory epithelium (15), and (iii) promote internalization into human monocytic cells (17). The mechanism behind the virulence properties of protein D is unknown but may involve choline decoration of *H. influenzae* lipooligosaccharides (LOS), since protein D expression allows *H. influenzae* to obtain choline from cocultured host cells and subsequently incorporate this molecule into its LOS (6).

The ability of protein D to promote the incorporation of choline into LOS comes from its glycerophosphodiester phosphodiesterase activity, catalyzing the hydrolysis of glycerophosphodiesters into glycerol-3-phosphate (G3P) and an alcohol (14, 18). The gene encoding protein D (*hpd*) is homologous to *glpQ* of *Escherichia coli* (16) and other bacteria. The *glpQ* gene belongs to the *glp* regulon that is involved in the utilization of glycerol and of G3P and its precursors as energy sources and to supply precursors for phospholipid biosynthesis. In *E. coli* and *Bacillus subtilis* (19), *glpQ* is transcribed together with *glpT*, which is located upstream of *glpQ* and encodes a G3P permease that acts as a G3P-inorganic phosphate antiporter (5). The *E. coli glpTQ* operon is induced by G3P and repressed by the catabolic repressor glucose as well as a *glp*-specific repressor protein, GlpR (28). Available restriction fragment length polymorphism (RFLP) data and DNA sequences of the *glpTQ* region of *H. influenzae* suggest that its organization differs between strains (7, 13, 24, 25). RFLP analysis based on a combination of two restriction endonucleases and two probes complementary to the *glpQ* gene of *H. influenzae* (*hpd*) showed that the majority of Hib strains (34 of 39) carry the *glpQ* gene on a 3.3-kb fragment between two conserved *Pst*I sites (Fig. 1), whereas 24 of 35 NTHi strains display only a 1.9-kb fragment between the same *Pst*I sites (13). As in the case of *E. coli* and *B. subtilis*, *glpT* in most NTHi strains is located immediately upstream of *glpQ* (12, 13, 24). By contrast, *glpT* and *glpQ* are separated by a 1.4-kb fragment of unknown function in the majority of Hib strains and some other encapsulated *H. influenzae* strains, including strain Rd (Fig. 1) (7, 13, 24, 25). The present study was designed to characterize and compare the genetic structures and RNA transcriptional profiles of the *glpTQ* region in *H. influenzae* strains and to elucidate the potential function of the 1.4-kb *glpTQ* intergenic region that is present only in encapsulated *H. influenzae* strains.

DNA sequence analysis. In addition to using previously reported sequences (12, 24, 25), we used sequences determined in the course of this study, namely, the entire DNA sequences of the *glpTQ* operons of Hib Eagan and NTHi 772, as well as the 1.4-kb *glpTQ* intergenic region of Hib HK695 (26). The 1.4-kb *glpTQ* intergenic regions of Hib Eagan and HK695 were cloned into pBluescript II KS by excising a 2.2-kb *Pst*I-*Eco*RI fragment (Fig. 1) from plasmids pXME10 and pXMHK10, which harbor the 3.3-kb *Pst*I fragments that were used to clone the *hpd* genes (24) and subsequently transformed into *E. coli* XL1-Blue. Nested deletions with exonuclease III and mung bean nuclease (22) were performed after a *Sac*I-*Xba*I digestion of the resulting plasmids, pE1.4pe1 and pHK1.4pel, and on *Kpn*I-*Bam*HI-cleaved pXME10 and pXMHK10, respectively. The DNA sequences of the 1.4-kb intergenic regions of Eagan and HK695 were determined in both orientations from the constructed nested deletion plasmids with vector primers

* Corresponding author. Mailing address: Department of Medical Microbiology, Lund University, Malmö University Hospital, S-205 02 Malmö, Sweden. Phone: 46 (40) 331321. Fax: 46 (40) 336234. E-mail: Hakan.Janson@mikrobiol.mas.lu.se.

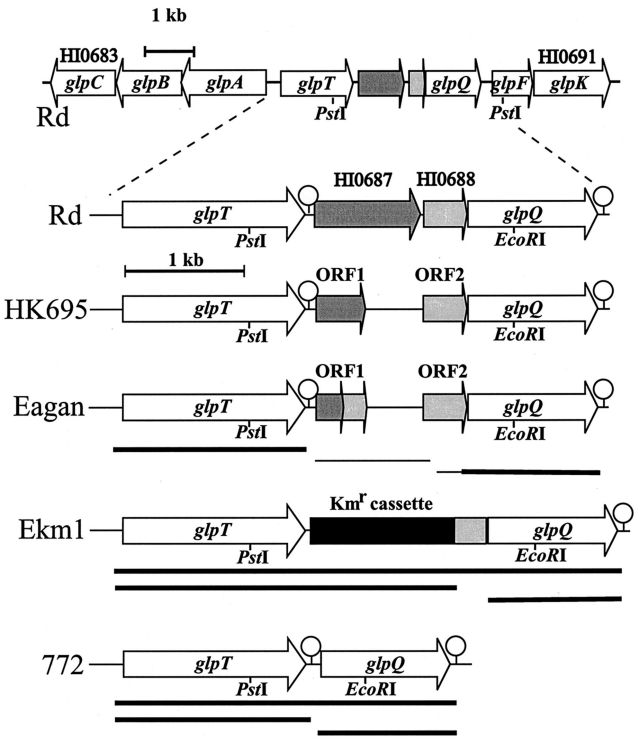

FIG. 1. Schematic genetic map of the *glpABC*, *glpTQ*, and *glpFK* operons in *H. influenzae* Rd (7) and genetic map and RNA transcript analysis results of the *glpTQ* operons of Hib HK695, Hib Eagan, Eagan-derived 1.4-kb *glpTQ* intergenic region mutant Ekm1, and NTHi 772. Arrows labeled with gene names and ORFs denote transcriptional orientations. Potential transcriptional termination loop structures are marked with circles. Conserved restriction endonuclease sites are indicated. RNA transcripts detected by Northern blotting or RT-PCR are marked under the respective genes or ORFs with thick and fine lines, respectively.

M13-u and M13-r (Table 1) by using an Applied Biosystems 377 automatic sequencer. The DNA sequences of *glpT* and the *glpTQ* intergenic region of NTHi 772 were determined from a 2.0-kb PCR product amplified with primers *glpA*-1 and *hpd*-10 (Table 1) and sequenced with primers specific for the *glpT* gene of Hib Eagan (25). The DNA sequence upstream of *glpT* of Hib Eagan was obtained from a PCR amplicon by using primers *glpA*-1 and *glpT*-9 (Table 1).

The *glpT* sequences of Hib Eagan and NTHi 772 were almost as highly conserved as the previously sequenced *glpQ* gene (13, 24). The *glpT* genes from the two strains differed in 15 bp and displayed identical deduced amino acid sequences of the same length. The regulatory region upstream of the NTHi 772 *glpT* gene was also identical to the corresponding region of Eagan, except that 772 contained one extra base pair located 91 bp upstream of the *glpT* start codon (Fig. 2). Potential catabolic and transcriptional regulation sites with homology to consensus sequences upstream of *glpT* in *E. coli* were also found upstream of *glpT* in Eagan and 772. They include two catabolic activator protein binding sites (4), two operator sites with the potential of binding to the *E. coli glp* repressor protein (29), and a sequence homologous to the integration host factor (IHF) (9) (Fig. 2). The binding of IHF to this site may bend the DNA to facilitate the formation of a complex structure that is involved in the inhibition of GlpT expression, mediated by the GlpR repressor in *E. coli* (28). However, despite the existence of two GlpR homologues, no homologues of *ihfA* and *ihfB* (R. A. Weisberg, M. Freundlich, D. Friedman, J. Gardner, N. Goosen, H. Nash, A. Oppenheim, and J. Rouvière-Yaniv, Letter, Mol. Microbiol. **19:**642, 1996) have been identified in the *H. influenzae* Rd genome (7).

In the *glpTQ* intergenic region, both Hib strains contained practically identical 1,428-bp sequences, except that HK695 contained one more tetrameric repeat (TTTA) in addition to the five repeats present in Eagan. The repeats were localized in a putative open reading frame (ORF) of 414 bp in HK695, designated ORF1 for these Hib strains (Fig. 1). The lack of one tetramer in Eagan caused a frameshift that led to a stop codon after 228 bp and resulted in two smaller ORFs (Fig. 1). Most tetrameric DNA repeats in *H. influenzae* have been associated with the regulation of potential virulence factors such as iron acquisition factors, adhesion molecules, and LOS biosynthesis (11, 27), but the function of ORF1 is unknown. Another putative 390-bp ORF (ORF2) of the Hib *glpTQ* intergenic region was found immediately upstream of *glpQ* (Fig. 1). The *glpTQ* intergenic region of the two Hib strains was highly homologous to the corresponding region of *H. influenzae* Rd (7). Two ORFs with unknown functions, HI0687 and HI0688, correspond to ORF1 and ORF2 of the Hib strains with the exception that HI0687 of *H. influenzae* Rd exhibits a larger size than ORF1 (Fig. 1). The G+C contents of the 1.4-kb *glpTQ* intergenic region of the Hib strains was approximately 26%, which is much lower than the average of 38% for the *glpT* and *glpQ* genes and 39% for the *H. influenzae* Rd genome in general. Although no typical insertion-like sequences were found in this region, the 1.4-kb fragment might have been acquired by lateral transfer from an organism with a lower G+C content. The *glpTQ* intergenic region of NTHi 772 was only 103 bp long, out of which the first 52 bp was homologous to the *glpTQ* intergenic regions of the Hib strains Eagan and HK695. This homologous part contains an inverted repeat sequence with the ability to form a typical hairpin loop structure exhibiting a lowest free energy value of −9.8 kcal/mol (23) that may function as a potential transcription terminator for *glpT* expression. A similar conserved inverted repeat sequence was also identified immediately downstream of the *glpQ* stop codon in both Hib and NTHi strains (Fig. 1). This potential *glpQ* transcriptional terminator exhibited a lowest free energy value of −16.8 kcal/mol (23).

Construction of a 1.4-kb mutant in *H. influenzae*. Reverse PCR with a *Pwo* PCR kit (Roche Diagnostics) was used to amplify the fragment flanking the 1.4-kb intergenic region of pXME10 with primers Hibup-3 and Hibup-4 (Table 1), resulting in a 1.2-kb deletion of the 1.4-kb intergenic region. The PCR product was ligated to the 1.3-kb kanamycin resistance (Km^r) cassette excised from pUC4K (Amersham Pharmacia Biotech) with *Hin*cII, and the ligation mixture was transformed into *E. coli* DH5α. Following the excision of the vector backbone by *Bam*HI-*Sph*I cleavage, the 3.5-kb linear DNA fragment was transformed into Eagan for homologous recombination (14). The chromosomal DNA of five Km^r colonies was examined by Southern hybridization with digoxigenin (Roche Diagnostics)-labeled probes. The *glpT* and *glpQ* probes were labeled by PCR using a pXME10-derived nested deletion and

TABLE 1. Oligonucleotides used in this study

Primer	Nucleotide sequence[a]	Comments[b]
glpA-1	5′-TTGCGgatCCACCGATAATGATCAC-3′	Reverse primer annealing to nt 85–60 of Rd *glpA*, used for amplification and sequencing
glpT-9	5′-ATGCGctGCaGGTTTGAATGGTCCA-3′	Reverse primer annealing to nt 29–6 of Eagan *glpT*, used for amplification, sequencing, and primer extension
Hibup-1	5′-AAAAAAGGCtGcaGGCAAATTGCTTAGTC-3′	Forward primer annealing to nt 8–35 of downstream *glpT* stop codon of Eagan, used for RT-PCR
Hibup-2	5′-CCTATCAAggaTCCAATGCCAACATCTTG-3′	Reverse primer annealing to nt 264–234 of downstream ORF1 stop codon of HK695, used for RT-PCR
Hibup-3	5′-AAAGGCGACTAAGCAATTTGCCCACCGCCTTTTTTAAG-3′	Reverse primer annealing to nt 41–5 of downstream *glpT* stop codon of Eagan, used for construction of 1.4-kb mutant
Hibup-4	5′-TGCTGGAAATGATTATAAAT-3′	Forward primer annealing to nt 98–117 of Eagan ORF2, used for construction of 1.4-kb mutant and RT-PCR
Hibup-5	5′-AATGGGGTTAGGATTTCCTATTA-3′	Forward primer annealing to nt 41–63 of Eagan ORF1, used for RT-PCR
Hibup-6	5′-AATCCTCCAGATAAAAATCTAACA-3′	Reverse primer annealing to nt 130–107 of Eagan ORF1, used for primer extension
Hibup-7	5′-GGCATAGCAAGTATTCCAAAGATAC-3′	Reverse primer annealing to nt 328–303 of HK695 ORF1, used for RT-PCR
Hibup-8	5′-TTCCGTGAATAATTTTAATAGT-3′	Reverse primer annealing to nt 267–246 of Eagan ORF2, used for RT-PCR
Hibup-10	5′-CTTCaAgCTTTTCCCAAAATATCTAATT-3′	Reverse primer annealing to nt 35–8 of Eagan ORF2, used for primer extension
Hibup-11	5′-ATTTTggATccCTTTAATCATTCTTTCTGA-3′	Reverse primer annealing to nt 185–156 of Eagan ORF2, used for primer extension
hpd-10	5′-CCTGCgAaTtCGCCAGCTGCTAATA-3′	Reverse primer annealing to nt 52–28 of *glpQ*, used for amplification and primer extension
M13-u	5′-GTTTTCCCAGTCACGAC-3′	M13 universal primer for pBluescript II vectors, used for sequencing
M13-r	5′-AACAGCTATGACCATG-3′	M13 reverse primer for pBluescript II vectors, used for sequencing

[a] Lowercase letters indicate nucleotides substituted for the creation of restriction sites.
[b] nt, nucleotide.

pHIC348 (12) as templates with M13-u–*glpTx*1 and *hpd*-6B–*hpd*-4 (Table 1) (24) as primer pairs, respectively. The kanamycin probe was labeled by random priming from the gel-purified Km^r cassette that had been excised from pUC4K with *Hin*cII. Southern blot analysis (data not shown) of *Eco*RI- or *Hae*III-digested chromosomal Hib DNA revealed that the Km^r cassette had been inserted just downstream of the *glpT* coding region (Fig. 1) in all tested Km^r clones. The growth rate of one such mutant, Ekm1, was compared to that of the parent strain Eagan, and metabolic fingerprints were analyzed with GN MicroPlate and ES MicroPlate kits (Biolog, Hayward, Calif.) either in brain heart infusion broth supplemented with NAD and hemin or in the chemically defined BCM1 medium (25). No significant difference in cell densities, generation times, or metabolic properties was observed for triplicate experiments (data not shown).

Characterization of promoter positions of the *glpTQ* operons. Primer extension analysis (22) was applied to identify potential transcriptional start positions by using antisense DNA primers (Table 1) that bound just downstream of the respective start codons. Primers were labeled with [γ-^{32}P]ATP by T4 polynucleotide kinase (Amersham Pharmacia Biotech) and were used for the extension of 40 μg of total RNA, prepared with an RNeasy kit (QIAGEN) from cultures grown in supplemented brain heart infusion broth. Labeled products were analyzed on a 6% polyacrylamide gel containing 7 M urea in a Sequi-Gen sequencing cell (Bio-Rad Laboratories). The signal in the gel was measured either by exposure to X-ray film or by a PhosphorImager and analyzed with ImageQuant software (Molecular Dynamics). A total of four bands, designated P1 to P4, corresponding to potential transcriptional start positions of *glpT* were detected in NTHi 772, whereas only three bands were detected in Hib Eagan (P1, P2, and P4) (Fig. 2). The lack of P3 in Hib Eagan was unexpected since potential corresponding −10 and −35 promoter sequences were found in both strains. The existence of one more promoter (P3) in 772 suggests that the regulation conditions and expression levels of *glpT* might differ for NTHi and Hib strains. This hypothesis was supported by the observation that *glpT* of NTHi 772 that was cloned into the pACYC184-based vector pSU18 (3) failed to complement the G3P uptake levels in the Hib Eagan *glpT rec-1* double mutant ETM22r2 (unpublished data), although both strains contain almost identical *glpT* genes. We further analyzed the total RNA of ETM22r2 that had been transformed with pXMS1 (25), carrying Eagan *glpT* plus a partial regulatory region ending 156 bp upstream of the *glpT* start codon cloned into pSU18. Only one band corresponding to position P1 was found (data not shown), although the P2 site and its putative −10 promoter sequence plus the region where a −35 consensus sequence normally is found (Fig. 2) was included in pXMS1. This result suggests that the P1 site may be essential for *glpT* transcription in Eagan since pXMS1 is able to complement the *glpT* mutant ETM22r2 to a higher level of GlpT expression than that of the wild-type strain, Eagan (25).

```
                                                  glpA ←|                729799
Rd,Eagan,772    GATCTCCAACATCTTTATAGATACTAGGCGATAATCCCATTGTTTACCCCCGATTCCCAA
                L  D  G  V  D  K  Y  I  S  P  S  L  G  M

Rd              AATGTAAGAAAAAATAATATAAAAAGTGTAATCATTGTAGGGAAGTGCGGCCAAAATTAA
                                          ↓ P4
Eagan,772       ----------------------------------------A---------¤---------

                                                                  |←-------O2-
                                          |←------------CAP2---------→|
Rd,Eagan,772    CAACATTTTGAGTGTAAAAGAAAGCGATTTTGTGATATTGATCACAATAATTTTCGCATT

                -------→|        |←---------CAP1--------→|↓ P2
Rd,Eagan        CGCTCAAACGGTGTAAAGATTTGTGAAACACTTCACATTTTTGATTTTTTTTACTTTCTT
                    ↓ P3                                  ↓ P2
772             ------------------------------------------------------------

                                          ↓ P1
                                  |←--------O1--------→|
Rd,Eagan        A¤TTTTTTATTTTCTTTTATGATTCGCAAATCCCTTACTGTTTCAAAAGGGTAAATTTAGA
                                          ↓ P1
772             -T----------------------------------------------------------

                        |←----IHF--->|          |→ glpT                730099
Rd,Eagan,772    TGACACATAATTAAAAAGTTGGAGAAAAAATTATGTTTGGACCATTCAAACCCGCTCCGC
                                   rbs          M  F  G  P  F  K  P  A  P  H
```

FIG. 2. Nucleotide sequence alignment of the *glpT* promoter region of *H. influenzae* Rd (7), Hib Eagan, and NTHi 772. The numbered positions of nucleotides refer to the Rd genome. A sequence gap is indicated with the symbol ¤. The *glpT* and *glpA* start codons are indicated in boldface and italics, and arrows mark the transcriptional directions. Putative −10 and −35 sequences are underlined with double and single solid lines, respectively. Proposed catabolic activator protein (CAP), IHF, and operator binding (O) sites are indicated with dashed arrows above the sequences. Potential promoter sites (P) for the *glpT* genes of Eagan and 772, based on the primer extension assay, are marked with vertical arrows.

It also indicates that the signal at the P2 site in primer extension analysis may be a degradation product of a *glpT* mRNA initiated from the upstream promoter site P4. Since ETM22r2 is a *glpT* mutant that was created by the selection of spontaneous fosfomycin-resistant Hib Eagan clones (25), it is likely that the P4 promoter site probably exists in ETM22r2. No signal from P4 could be detected in ETM22r2(pXMS1), which may be due to the fact that this potential signal is too weak in comparison to the P1 signal from the medium-copy-number plasmid pXMS1. When potential promoter sites for *glpQ* were analyzed, only one band was identified 28 bp upstream of the *glpQ* start codon in Hib Eagan as well as in *E. coli*(pXME10), which contains *glpQ* from Eagan. A primer extension product was found 64 bp upstream of the *glpQ* start codon in NTHi 772 and from *E. coli*(pHIC348), which contains *glpQ* from 772 (data not shown). The transcriptional start of Hib Eagan ORF1 in the *glpTQ* intergenic region was mapped to 59 bp upstream of the ORF1 start codon for both Eagan and *E. coli* harboring pXME10 (data not shown). Since the *glpTQ* intergenic region of NTHi 772 is only 103 bp, the potential transcriptional starts of NTHi 772 *glpQ* and Hib Eagan ORF1 must be located within 1 bp from each other in relation to the ends of their respective *glpT* genes. Both promoters share the same putative −10 and −35 sequences, which are situated within the loop structure downstream of *glpT*. Furthermore, in spite of the divergent sequences immediately upstream of the *glpQ* start codons of Eagan and 772 (24), the two strains appear to share the same putative ribosomal binding site. Attempts to locate a transcriptional start for ORF2 of the 1.4-kb region were unsuccessful, but two bands were found within the ORF2 coding region for both Eagan and *E. coli* harboring pXME10 (Fig. 3). These were located 85 and 89 bp, respectively, downstream of the ORF2 start codon. Shared putative −10 and −35 sequences were also found upstream of these sites.

RNA transcript analysis. *glpT* (1.4 kb) and *glpQ* (1.1 kb) transcripts from the total RNAs of NTHi 772, Hib Eagan, and ETM22r2(pXMS1) were visualized by using Northern hybridization (25) (Fig. 4). Hib Eagan and ETM22r2(pXMS1) displayed a single band for *glpT* and *glpQ*, respectively. NTHi 772 exhibited a band of approximately 2.6 kb that hybridized to both *glpT* and *glpQ* probes in addition to the respective smaller bands that were detected in Eagan. The possibility that monomeric *glpT*- and *glpQ*-specific bands are a result of the mRNA processing of the larger 2.6-kb transcript cannot be ruled out, but since *E. coli* harboring pHIC348 expresses GlpQ from its endogenous NTHi 772 promoter (12), and since we successfully mapped the *glpQ* promoter with primer extension analysis, it is concluded that the *glpQ* gene of NTHi 772 is partially cotranscribed with *glpT*. No variation in mRNA lengths due to alternative transcriptional start points was detected by Northern blot analysis. Based on the assumption that the hairpin loop structure immediately downstream of the *glpT* stop codon terminates the *glpT* transcription in Hib Eagan, we examined the *glpTQ* transcription patterns of the Eagan 1.4-kb mutant Ekm1, in which the loop structure was destroyed during mutant construction (Fig. 1 and 4). Two Ekm1 transcripts (2.7 and 3.9 kb) were found hybridizing to the *glpT* probe. It is likely that the 2.7-kb band is a combination of the 1.4-kb *glpT* and the 1.3-kb Km^r cassette and that the 3.9-kb fragment is the 2.7-kb band plus the downstream 1.1-kb *glpQ* transcript. In addition

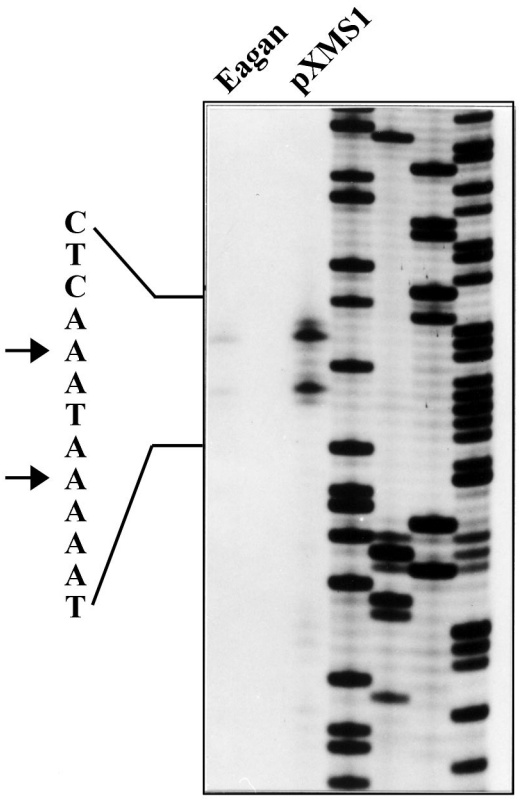

FIG. 3. Primer extension analysis of potential ORF2 transcriptional start positions (indicated by arrows) of total RNAs of Hib Eagan and *E. coli* harboring pXME10 by using primer Hibup-11. The DNA ladder (T, G, C, and A from left to right) was obtained from pXME10 that had been sequenced with Hibup-11 and separated in the same polyacrylamide gel.

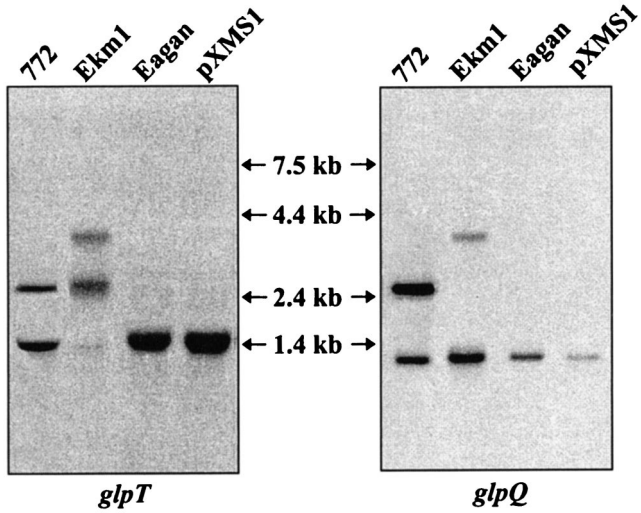

FIG. 4. Northern blot analysis of *glpT*- and *glpQ*-specific transcripts of total RNAs of NTHi 772, Hib Eagan, Eagan-derived 1.4-kb *glpTQ* intergenic region mutant Ekm1, and the cloned *glpT* gene in ETM22r2(pXMS1). The same membrane was probed with the *glpT* probe and reprobed with the *glpQ* probe (indicated at the bottom of the blots) after the membrane was stripped. The molecular sizes (indicated between the blots) were estimated according to the RNA ladder separated in the same gel.

to the 1.1-kb *glpQ* transcript, the 3.9-kb fragment was also found when the same blot was reprobed with a *glpQ* probe, suggesting that *glpT* transcription could continue downstream once the loop structure was removed. However, the practically identical loop structure downstream of 772 *glpT* only partially terminated *glpT* transcription because a cotranscribed *glpTQ* transcript was found in this strain. We therefore conclude that a complete interruption of the *glpTQ* cotranscript requires the existence of the 1.4-kb *glpTQ* intergenic region.

As we failed to detect any signals by using ORF1 or ORF2 probes in Northern hybridization (data not shown), RNA transcripts of the 1.4-kb intergenic region of Eagan were further characterized with reverse transcription (RT)-PCR by using Ready-To-Go beads (Amersham Pharmacia Biotech) and different combinations of RT-PCR primers (Table 1). An RNA transcript that covered the entire ORF1 and the ORF1-ORF2 intergenic region was identified, whereas another was found covering most parts of ORF2 (Fig. 1), corresponding to the potential promoter positions found 85 to 89 bp within ORF2 in our primer extension assay. This latter ORF2 transcript appeared to be cotranscribed with the downstream *glpQ*. This result is in contrast to the 1.1-kb *glpQ* transcript previously identified by Northern hybridization and to the potential promoter positions of *glpQ* obtained by the primer extension assay, but as we could detect the transcriptional signal only by RT-PCR and not by Northern blotting, it is assumed that the transcription level of this alternative *glpQ* promoter situated within ORF2 is low. ORF2 may act as a transcription regulatory region for *glpQ* in Hib, since it is located immediately upstream of the *glpQ* coding region, separated by only 11 bp. No apparent catabolic regulation sequence homologues were found in this region. We could not find any putative catabolic regulation sites upstream of NTHi 772 *glpQ* either, but its putative promoter, found in the primer extension analysis, is located at practically the same position as the potential promoter of ORF1 in Hib Eagan. We therefore speculate that Hib strains that contain the 1.4-kb fragment between *glpT* and *glpQ* have evolved an alternative promoter region for *glpQ* transcription and that the old *glpQ* promoter now serves as a promoter for ORF1. The transcriptional level of Hib ORF1 appears to be low, since an ORF1 mRNA was detected only by RT-PCR analysis, whereas an NTHi 772 *glpQ* mRNA was detectable by Northern blot analysis. The promoters are located 1 bp apart within the potential transcriptional terminator loop for *glpT* (Fig. 1), but it is unlikely that this difference would be the explanation for the observed differences in transcriptional levels between these promoters. The regulation of the Eagan *glpQ* gene is unknown because it transcribed independently of *glpT*, whereas the regulation of the *glpQ* gene in NTHi 772 might be dependent upon the cotranscribed upstream *glpT* gene.

Overall, this study shows the effect the presence of the 1.4-kb *glpTQ* intergenic region that exists only in Hib and other encapsulated *H. influenzae* strains has on *glpT* and *glpQ* transcription in these strains compared with their transcription in NTHi 772. For NTHi strains, the *glpQ* gene (encoding the virulence factor protein D) was in part cotranscribed and possibly coregulated with the upstream *glpT* gene, encoding G3P per-

mease, and in part transcribed by a promoter that is situated within a hairpin loop structure. In Hib strains, the same promoter appears to transcribe at least one unique ORF (ORF1) that may have the ability to phase vary its expression due to the variation of the number of tetramer repeats situated within the ORF. It would be very interesting to further study and determine how the regulation of genes in the *glpTQ* region of NTHi strains differs from that in Hib strains and to investigate what role the ORFs in the 1.4-kb intergenic region of Hib strains have in pathogenesis.

Nucleotide sequence accession numbers. The entire DNA sequences of the *glpT* gene of NTHi 772, as well as the 1.4-kb *glpTQ* intergenic region of Hib Eagan and HK695 (26), were determined in this study and submitted to GenBank under the following accession numbers: for the 1.4-kb intergenic region of Eagan, accession no. AF132899; for the 1.4-kb intergenic region of HK695, accession no. AF132900; and for *glpT* of NTHi 772, accession no. AF132901.

We thank Arne Forsgren for generous support and critical review, Christel Ekman for DNA sequencing, and Elisabeth Glatz of the Department of Microbiology and Andreas Hillarp of the Department of Clinical Chemistry, Lund University, for technical assistance in the RNA analyses.

This work was supported by grants from the Alfred Österlund Foundation, the Malmö Medical Services Foundation, and the Forssmans Foundation.

REFERENCES

1. **Akkoyunlu, M., M. Ruan, and A. Forsgren.** 1991. Distribution of protein D, an immunoglobulin D-binding protein, in *Haemophilus* strains. Infect. Immun. **59:**1231–1238.
2. **Aubrey, R., and C. Tang.** 2003. The pathogenesis of disease due to type b *Haemophilus influenzae.* Methods Mol. Med. **71:**29–50.
3. **Bartolomé, B., Y. Jubete, E. Martinez, and F. de la Cruz.** 1991. Construction and properties of a family of pACYC184-derived cloning vectors compatible with pBR322 and its derivatives. Gene **102:**75–78.
4. **de Crombrugghe, B., S. Busby, and H. Buc.** 1984. Cyclic AMP receptor protein: role in transcription activation. Science **224:**831–838.
5. **Elvin, C. M., C. M. Hardy, and H. Rosenberg.** 1985. P_i exchange mediated by the GlpT-dependent *sn*-glycerol-3-phosphate transport system in *Escherichia coli.* J. Bacteriol. **161:**1054–1058.
6. **Fan, X., H. Goldfine, E. Lysenko, and J. N. Weiser.** 2001. The transfer of choline from the host to the bacterial cell surface requires *glpQ* in *Haemophilus influenzae.* Mol. Microbiol. **41:**1029–1036.
7. **Fleischmann, R. D., et al.** 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science **269:**496–512.
8. **Foxwell, A. R., J. M. Kyd, and A. W. Cripps.** 1998. Nontypeable *Haemophilus influenzae*: pathogenesis and prevention. Microbiol. Mol. Biol. Rev. **62:**294–308.
9. **Goodrich, J. A., M. L. Schwartz, and W. R. McClure.** 1990. Searching for and predicting the activity of sites for DNA binding proteins: compilation and analysis of the binding sites for *Escherichia coli* integration host factor (IHF). Nucleic Acids Res. **18:**4993–5000.
10. **Heath, P. T.** 1998. *Haemophilus influenzae* type b conjugate vaccines: a review of efficacy data. Pediatr. Infect. Dis. J. **17:**S117–S122.
11. **Hood, D. W., M. E. Deadman, M. P. Jennings, M. Bisercic, R. D. Fleischmann, J. C. Venter, and E. R. Moxon.** 1996. DNA repeats identify novel virulence genes in *Haemophilus influenzae.* Proc. Natl. Acad. Sci. USA **93:** 11121–11125.
12. **Janson, H., L.-O. Hedén, A. Grubb, M. Ruan, and A. Forsgren.** 1991. Protein D, an immunoglobulin D-binding protein of *Haemophilus influenzae*: cloning, nucleotide sequence, and expression in *Escherichia coli.* Infect. Immun. **59:**119–125.
13. **Janson, H., M. Ruan, and A. Forsgren.** 1993. Limited diversity of the protein D gene (*hpd*) among encapsulated and nonencapsulated *Haemophilus influenzae* strains. Infect. Immun. **61:**4546–4552.
14. **Janson, H., Å. Melhus, A. Hermansson, and A. Forsgren.** 1994. Protein D, the glycerophosphodiester phosphodiesterase from *Haemophilus influenzae* with affinity for human immunoglobulin D, influences virulence in a rat otitis model. Infect. Immun. **62:**4848–4854.
15. **Janson, H., B. Carlén, A. Cervin, A. Forsgren, A. B. Magnusdottir, S. Lindberg, and T. Runer.** 1999. Effects on the ciliated epithelium of protein D-producing and -nonproducing nontypeable *Haemophilus influenzae* in nasopharyngeal tissue cultures. J. Infect. Dis. **180:**737–746.
16. **Larson, T. J., M. Ehrmann, and W. Boos.** 1983. Periplasmic glycerophosphodiester phosphodiesterase of *Escherichia coli*, a new enzyme of the *glp* regulon. J. Biol. Chem. **258:**5428–5432.
17. **Lazou-Ahrén, I., H. Janson, A. Forsgren, and K. Riesbeck.** 2001. Protein D expression promotes the adherence and internalization of non-typeable *Haemophilus influenzae* into human monocytic cells. Microb. Pathog. **31:**151–158.
18. **Munson, R. S., Jr., and K. Sasaki.** 1993. Protein D, a putative immunoglobulin D-binding protein produced by *Haemophilus influenzae*, is glycerophosphodiester phosphodiesterase. J. Bacteriol. **175:**4569–4571.
19. **Nilsson, R.-P., L. Beijer, and B. Rutberg.** 1994. The *glpT* and *glpQ* genes of the glycerol regulon in *Bacillus subtilis.* Microbiology **140:**723–730.
20. **Poolman, J. T., L. Bakaletz, A. Cripps, P. A. Denoel, A. Forsgren, J. Kyd, and Y. Lobet.** 2001. Developing a nontypeable *Haemophilus influenzae* (NTHi) vaccine. Vaccine **19:**S108–S115.
21. **Ruan, M., M. Akkoyunlu, A. Grubb, and A. Forsgren.** 1990. Protein D of *Haemophilus influenzae.* A novel bacterial surface protein with affinity for human IgD. J. Immunol. **145:**3379–3384.
22. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
23. **SantaLucia, J., Jr.** 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc. Natl. Acad. Sci. USA **95:**1460–1465.
24. **Song, X.-M., A. Forsgren, and H. Janson.** 1995. The gene encoding protein D (*hpd*) is highly conserved among *Haemophilus influenzae* type b and nontypeable strains. Infect. Immun. **63:**696–699.
25. **Song, X.-M., A. Forsgren, and H. Janson.** 1998. Glycerol-3-phosphate transport in *Haemophilus influenzae*: cloning, sequencing, and transcription analysis of the *glpT* gene. Gene **215:**381–388.
26. **Song, X.-M., A. Forsgren, and H. Janson.** 1999. Fragmentation heterogeneity of 23S ribosomal RNA in *Haemophilus* species. Gene **230:**287–293.
27. **van Belkum, A.** 1999. Short sequence repeats in microbial pathogenesis and evolution. Cell. Mol. Life Sci. **56:**729–734.
28. **Yang, B., S. G. Gerhardt, and T. J. Larson.** 1997. Action at a distance for *glp* repressor control of *glpTQ* transcription in *Escherichia coli* K-12. Mol. Microbiol. **24:**511–521.
29. **Zhao, N., W. Oh, D. Trybul, K. S. Thrasher, T. J. Kingsbury, and T. J. Larson.** 1994. Characterization of the interaction of the *glp* repressor of *Escherichia coli* K-12 with single and tandem *glp* operator variants. J. Bacteriol. **176:**2393–2397.