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:

2003

Link to publication

Citation for published version (APA):

Total number of authors:
2

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
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\(^2\)*

\* 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.

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 monocytes (17). The mechanism behind the virulence properties of protein D is unknown but may involve choline decoration of 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-EcoRI 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 SacI-XbaI digest of the resulting plasmids, pE1.4pEI and pHK1.4pEI, and on *KpnI-BamHI*-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.
M13-u and M13-r (Table 1) by using an Applied Biosystems 377 automatic sequencer. The DNA sequences of \( \text{glp}T \) and the \( \text{glpTQ} \) operons of Hib HK695, Hib Eagan, Eagan-derived 1.4-kb \( \text{glpTQ} \) operon 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. The \( \text{glpT} \) sequences of Hib Eagan and NTHi 772 were almost as highly conserved as the previously sequenced \( \text{glpQ} \) gene (13, 24). The \( \text{glpT} \) 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 \( \text{glpT} \) gene was also identical to the corresponding region of Eagan, except that 772 contained one extra base pair located 91 bp upstream of the \( \text{glpT} \) start codon (Fig. 2). Potential catabolic and transcriptional regulation sites with homology to consensus sequences upstream of \( \text{glpT} \) in \( E. \ coli \) were also found upstream of \( \text{glpT} \) in Eagan and 772. They include two catabolic activator protein binding sites (4), two operator sites with the potential of binding to the \( E. \ coli \) GlpR 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 \( \text{ihfA} \) and \( \text{ihfB} \) (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 \( \text{glpTQ} \) 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 \( \text{glpTQ} \) intergenic region was found immediately upstream of \( \text{glpQ} \) (Fig. 1). The \( \text{glpTQ} \) 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 \( \text{glpTQ} \) intergenic region of the Hib strains was approximately 26%, which is much lower than the average of 38% for the \( \text{glpT} \) and \( \text{glpQ} \) 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 \( \text{glpTQ} \) intergenic region of NTHi 772 was only 103 bp long, out of which the first 52 bp was homologous to the \( \text{glpTQ} \) 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 \( \text{glpQ} \) expression. A similar conserved inverted repeat sequence was also identified immediately downstream of the \( \text{glpQ} \) stop codon in both Hib and NTHi strains (Fig. 1). This potential \( \text{glpQ} \) 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') cassette excised from pUC4K (Amersham Pharmacia Biotech) with HincII, and the ligation mixture was transformed into \( E. \ coli \) DH5α. Following the excision of the vector backbone by BamHI-SphI cleavage, the 3.5-kb linear DNA fragment was transformed into Eagan for homologous recombination (14). The chromosomal DNA of five Km' colonies was examined by Southern hybridization with digoxigenin (Roche Diagnostics)-labeled probes. The \( \text{glpT} \) and \( \text{glpQ} \) probes were labeled by PCR using a pXME10-derived nested deletion and

![FIG. 1. Schematic genetic map of the \( \text{glpABC} \), \( \text{glpTQ} \), and \( \text{glpFK} \) operons in \( H. \ influenzae \) Rd (7) and genetic map and RNA transcript analysis results of the \( \text{glpTQ} \) operons of Hib HK695, Hib Eagan, Eagan-derived 1.4-kb \( \text{glpTQ} \) 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.](image-url)
mycin probe was labeled by random priming from the gel-
hpd region (Fig. 1) in all tested Kmr clones. The growth rate of one
– pHIC348 (12) as templates with M13-u
VOL. 185, 2003 NOTES 7287

were analyzed on a 6% polyacrylamide gel containing 7 M urea
supplemented brain heart infusion broth. Labeled products
pared with an RNeasy kit (QIAGEN) from cultures grown in
and were used for the extension of 40
H9262 by T4 polynucleotide kinase (Amersham Pharmacia Biotech)

characterized 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-32P]ATP
by T4 polynucleotide kinase (Amersham Pharmacia Biotech)
and were used for the extension of 40 μg of total RNA, pre-
pared 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 soft-
ware (Molecular Dynamics). A total of four bands, designated
P1 to P4, corresponding to potential transcriptional start po-
sitions 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
positions of

<table>
<thead>
<tr>
<th>Primer</th>
<th>Nucleotide sequence</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>glpA-1</td>
<td>5'-TTGCCGatCCACCGATAATGATC-3'</td>
<td>Reverse primer annealing to nt 85–60 of Rd glpA, used for amplification and sequencing</td>
</tr>
<tr>
<td>glpT-9</td>
<td>5'-ATGCGetGCaGGTTATGAAGTCCA-3'</td>
<td>Reverse primer annealing to nt 29–62 of Eagan glpT, used for amplification, sequencing, and primer extension</td>
</tr>
<tr>
<td>Hibup-1</td>
<td>5'-AAAAAGGCrGcaGGCAATGCTAGTC-3'</td>
<td>Forward primer annealing to nt 8–35 of downstream glpT stop codon of Eagan, used for RT-PCR</td>
</tr>
<tr>
<td>Hibup-2</td>
<td>5'-CCTACTCAaggTCCAATGCGCACATTTG-3'</td>
<td>Reverse primer annealing to nt 264–234 of downstream ORF1 stop codon of HK695, used for RT-PCR</td>
</tr>
<tr>
<td>Hibup-3</td>
<td>5'-AAAGGCACTAGCAATTTGCCACCGCTTTTAAAG-3'</td>
<td>Reverse primer annealing to nt 41–5 of downstream glpT stop codon of Eagan, used for construction of 1.4-kb mutant</td>
</tr>
<tr>
<td>Hibup-4</td>
<td>5'-TGGCTGAAATGATTATAAAT-3'</td>
<td>Forward primer annealing to nt 98–117 of Eagan ORF2, used for construction of 1.4-kb mutant and RT-PCR</td>
</tr>
<tr>
<td>Hibup-5</td>
<td>5'-AATGGGGTTAGATTCTATTA-3'</td>
<td>Forward primer annealing to nt 41–63 of Eagan ORF1, used for RT-PCR</td>
</tr>
<tr>
<td>Hibup-6</td>
<td>5'-AATCTCCCGATAAAAAATCTAACA-3'</td>
<td>Reverse primer annealing to nt 130–107 of Eagan ORF1, used for primer extension</td>
</tr>
<tr>
<td>Hibup-7</td>
<td>5'-GGCATAGCAAGTATTCCAAAGATAC-3'</td>
<td>Reverse primer annealing to nt 328–303 of HK695 ORF1, used for RT-PCR</td>
</tr>
<tr>
<td>Hibup-8</td>
<td>5'-TTCCGTGAATAATTTTAATAGT-3'</td>
<td>Reverse primer annealing to nt 267–246 of Eagan ORF2, used for RT-PCR</td>
</tr>
<tr>
<td>Hibup-10</td>
<td>5'-CTCTcaAgtCTTCCAATAATCTAATT-3'</td>
<td>Reverse primer annealing to nt 35–8 of Eagan ORF2, used for primer extension</td>
</tr>
<tr>
<td>Hibup-11</td>
<td>5'-ATTTtgATcCtTATTACCTTCTTG-3'</td>
<td>Reverse primer annealing to nt 185–156 of Eagan ORF2, used for primer extension</td>
</tr>
<tr>
<td>hpd-10</td>
<td>5'-CCTGCgAaTtCGCCAGCTGCTAATA-3'</td>
<td>Reverse primer annealing to nt 52–28 of glpQ, used for amplification and primer extension</td>
</tr>
<tr>
<td>M13-u</td>
<td>5'-GTTTCCCAGTCACGAC-3'</td>
<td>M13 universal primer for pBluescript II vectors, used for sequencing</td>
</tr>
<tr>
<td>M13-r</td>
<td>5'-AACAGCTATGACCAGT-3'</td>
<td>M13 reverse primer for pBluescript II vectors, used for sequencing</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Primer</th>
<th>Nucleotide sequence</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>M13-u</td>
<td>5'-AACAGCTATGACCAGT-3'</td>
<td>M13 reverse primer for pBluescript II vectors, used for sequencing</td>
</tr>
</tbody>
</table>

*p Lowercase letters indicate nucleotides substituted for the creation of restriction sites.

b nt, nucleotide.
It also indicates that the signal at the P2 site in primer extension analysis may be a degradation product of a \( \text{glpT} \) mRNA initiated from the upstream promoter site P4. Since ETM22r2 is a \( \text{glpT} \) 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 \( \text{glpQ} \) were analyzed, only one band was identified 28 bp upstream of the \( \text{glpQ} \) start codon in Hib Eagan as well as in \( \text{E. coli} \) (pXME10), which contains \( \text{glpQ} \) from Eagan. A primer extension product was found 64 bp upstream of the \( \text{glpQ} \) start codon in NTHi 772 and from \( \text{E. coli} \) (pHIC348), which contains \( \text{glpQ} \) from 772 (data not shown). The transcriptional start of Hib Eagan ORF1 in the \( \text{glpTQ} \) intergenic region was mapped to 59 bp upstream of the \( \text{glpQ} \) start codon in NTHi 772 and from \( \text{E. coli} \) (pHIC348), which contains \( \text{glpQ} \) from 772 (data not shown). The transcriptional start of Hib Eagan ORF1 in the \( \text{glpTQ} \) intergenic region was mapped to 59 bp upstream of the \( \text{glpQ} \) start codon in NTHi 772 and from \( \text{E. coli} \) (pHIC348), which contains \( \text{glpQ} \) from 772 (data not shown). The transcriptional start of Hib Eagan ORF1 in the \( \text{glpTQ} \) intergenic region was mapped to 59 bp upstream of the \( \text{glpQ} \) start codon in NTHi 772 and from \( \text{E. coli} \) (pHIC348), which contains \( \text{glpQ} \) from 772 (data not shown). It is important to note that the hairpin loop structure immediately downstream of the \( \text{glpT} \) stop codon terminates the \( \text{glpT} \) transcription in Hib Eagan, we examined the \( \text{glpTQ} \) transcription patterns of the Eagan 1.4-kb mutant Ekml, in which the loop structure was destroyed during mutant construction (Fig. 1 and 4). Two Ekml transcripts (2.7 and 3.9 kb) were found hybridizing to the \( \text{glpT} \) probe. It is likely that the 2.7-kb band is a combination of the 1.4-kb \( \text{glpT} \) and the 1.3-kb Kmr cassette and that the 3.9-kb fragment is the 2.7-kb band plus the downstream 1.1-kb \( \text{glpQ} \) transcript.
to the 1.1-kb \textit{glpQ} transcript, the 3.9-kb fragment was also found when the same blot was reprobed with a \textit{glpQ} probe, suggesting that \textit{glpT} transcription could continue downstream once the loop structure was removed. However, the practically identical loop structure downstream of 772 \textit{glpT} only partially terminated \textit{glpT} transcription because a cotranscribed \textit{glpTQ} transcript was found in this strain. We therefore conclude that a complete interruption of the \textit{glpTQ} cotranscript requires the existence of the 1.4-kb \textit{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 \textit{glpQ}. This result is in contrast to the 1.1-kb \textit{glpQ} transcript previously identified by Northern hybridization and to the potential promoter positions of \textit{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 \textit{glpQ} promoter situated within ORF2 is low. ORF2 may act as a transcription regulatory region for \textit{glpQ} in Hib, since it is located immediately upstream of the \textit{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 \textit{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 \textit{glpT} and \textit{glpQ} have evolved an alternative promoter region for \textit{glpQ} transcription and that the old \textit{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 \textit{glpQ} mRNA was detectable by Northern blot analysis. The promoters are located 1 bp apart within the potential transcriptional terminator loop for \textit{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 \textit{glpQ} gene is unknown because it transcribed independently of \textit{glpT}, whereas the regulation of the \textit{glpQ} gene in NTHi 772 might be dependent upon the cotranscribed upstream \textit{glpT} gene.

Overall, this study shows the effect the presence of the 1.4-kb \textit{glpTQ} intergenic region that exists only in Hib and other encapsulated \textit{H. influenzae} strains has on \textit{glpT} and \textit{glpQ} transcription in these strains compared with their transcription in NTHi 772. For NTHi strains, the \textit{glpQ} gene (encoding the virulence factor protein D) was in part cotranscribed and possibly coregulated with the upstream \textit{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 determined in this study and submitted to GenBank under the 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 Malmo Medical Services Foundation, and the Forssmans Foundation.

**REFERENCES**


