Characterization of *hgpA*, a gene encoding a haemoglobin/haemoglobin-haptoglobin-binding protein of *Haemophilus influenzae*

Hongfan Jin, Zhen Ren, Paul W. Whitby, Daniel J. Morton and Terrence L. Stull

Author for correspondence: Terrence L. Stull. Tel: +1 405 271 4401. Fax: +1 405 271 8710. E-mail: Terrence-Stull@ouhsc.edu

*Haemophilus influenzae* binds haemoglobin and the haemoglobin-haptoglobin complex and utilizes either as a sole source of haem. Previously, a DNA fragment was cloned from *H. influenzae* that encodes an approximately 120 kDa protein (*HgpA*) expressing haemoglobin-binding activity in *Escherichia coli*. Partial sequence analysis revealed significant homology of *HgpA* with other bacterial haem- and iron-utilization proteins, and a length of CCAA repeating units immediately following the nucleotide sequence encoding the putative leader peptide. In the present study, the complete nucleotide sequence of the cloned DNA fragment was determined and the sequence was analysed. In addition to homology with other haem- and iron-utilization proteins, seven regions typical of TonB-dependent proteins were identified. The transcript of *hgpA* was determined to be monocistronic by RT-PCR. PCR performed with different colonies of a single *H. influenzae* strain at one CCAA-repeat-containing locus indicated varying lengths of CCAA repeats, suggesting that haemoglobin and haemoglobin–haptoglobin binding in *H. influenzae* is regulated by strand slippage across CCAA repeats, as well as by haem repression. *E. coli* containing cloned *hgpA* bound both haemoglobin and the haemoglobin-haptoglobin complex. A deletion/insertion mutation of *hgpA* was constructed in *H. influenzae* strain HI689. Mutation of *hgpA* did not affect the ability of *H. influenzae* either to bind or to utilize haemoglobin or haemoglobin–haptoglobin following growth in haem-deplete media. Affinity purification of haemoglobin-binding proteins from the mutant strain revealed loss of the 120 kDa protein and an increased amount of a 115 kDa protein, suggesting that at least one additional haemoglobin-binding protein exists.

**Keywords**: *Haemophilus influenzae*, haemoglobin, haemoglobin-binding protein, haem acquisition, outer-membrane protein

**INTRODUCTION**

The human-specific pathogen *Haemophilus influenzae* causes a range of infections including otitis media, meningitis, epiglottitis and pneumonia (Turk, 1984). The incidence of invasive disease caused by *H. influenzae* strains with the type b capsule has been radically reduced by vaccines based on the type b capsular polysaccharide (Madorne et al., 1990; Murphy et al., 1993). Currently available vaccines based on the type b capsule are not protective against disease caused by unencapsulated strains of *H. influenzae*, and such strains are a significant cause of otitis media in childhood, neonatal sepsis and pneumonia in adults (Falla et al., 1993; Wallace et al., 1978; Klein, 1997).

*H. influenzae* has an absolute growth requirement for protoporphyrin IX, the immediate precursor of haem (Evans et al., 1974). In *vitro*, the requirement may be satisfied by haem, haemoglobin, and complexes of haemoglobin–haptoglobin, haem–haemopexin and haem–albumin (Stull, 1987). In *vivo*, haem is intra-
cellular, in the form of haemoglobin or haem-containing enzymes and thus unavailable to invading micro-organisms (Bezkorovainy, 1987; Lee, 1995). Haemoglobin released by erythrocytes is avidly bound by the serum protein haptoglobin, and the haemoglobin–haptoglobin complex is rapidly cleared by hepatocytes (Bezkorovainy, 1987; Otto et al., 1992). Free haem, principally derived from the degradation of methaemoglobin, is bound by the serum proteins haemopexin and albumin and is also cleared from the circulatory system by hepatocytes (Bezkorovainy, 1987). We have shown that H. influenzae binds haemoglobin, possibly as an initial step in the utilization of haemoglobin-associated haem (Frangipane et al., 1994), and we have cloned a DNA fragment encoding an approximately 120 kDa haemoglobin-binding protein (HgpA) of H. influenzae (Jin et al., 1996).

The objective of the current study was to characterize the gene encoding the previously identified 120 kDa haemoglobin-binding protein (HgpA) of H. influenzae (Jin et al., 1996).

METHODS

Bacterial strains, culture conditions and plasmids. H. influenzae type b strain H1689 has been previously described (Jin et al., 1996). H. influenzae Rd KW20 (ATCC 51907), used in the complete sequencing of the H. influenzae genome (Fleischmann et al., 1995), was obtained from the American Type Culture Collection (ATCC). An additional Rd strain maintained in this laboratory was used in certain experiments. Non-typable H. influenzae TN106 has been described by Maciver et al. (1996) and was kindly provided by Eric Hansen, University of Texas Southwestern Medical Center, Dallas, TX, USA. H. influenzae type b strain HI1715 is a strain shown by Southern analysis to have a single CCAA-repeat-containing locus associated with a gene which is highly homologous to hgpA (data not shown). H. influenzae was routinely maintained on brain heart infusion (BHI) agar (Difco) supplemented with 10 µg haem ml⁻¹ and 10 µg β-NAD ml⁻¹. Haem-restricted growth of H. influenzae was performed in BHI broth supplemented with 10 µg β-NAD ml⁻¹ and 0.1 µg haem ml⁻¹. Escherichia coli DH5α was grown on LB medium. Ribostamycin was added where appropriate to a final concentration of 15 µg ml⁻¹ for H. influenzae and 50 µg ml⁻¹ for E. coli. Plasmids pHFJ2 and pUC19::TSTE, containing the aminoglycoside resistance cassette TSTE, have been previously described (Jin et al., 1991). Clones generated in the sequencing of the Rd KW20 genome (Fleischmann et al., 1995) encompassing the ORFs designated HI0586–HI0592 were obtained from the ATCC.

DNA methodology. Restriction endonucleases were obtained from New England Biolabs and were used as directed by the manufacturer. Genomic DNA was isolated using the DNA Now reagent (Biogentex) as directed by the manufacturer. Plasmid DNA was isolated using QiaGen plasmid kits according to the manufacturer’s recommendations. Sequencing of double-stranded template DNA was performed by automated sequencing on an ABI model 373A sequencer at the Recombinant DNA/Protein Resource Facility, Oklahoma State University, Stillwater, OK, USA. Oligonucleotides were synthesized at the Molecular Biology Resource Facility, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA.

Southern analysis was performed as described by Sambrook et al. (1989). The Enhanced Chemi-Luminescence (ECL) random prime labelling kit (Amersham) was used as directed by the manufacturer to label DNA fragments. Hybridization was performed overnight at 60 °C and stringency washes were 1 × SSC, 0.1% (w/v) SDS for 15 min and 0.5 × SSC, 0.1% SDS for 15 min, both at 60 °C (1 × SSC; 8.8 g NaCl l⁻¹, 4.5 g sodium citrate l⁻¹, pH 7.0). Hybridized probes were detected using ECL nuclear acid detection reagents (Amersham) as directed by the manufacturer.

Construction of an insertion/deletion mutation of hgpA. Mutation of hgpA was achieved as follows. Plasmid pHFJ2 (Jin et al., 1996) was completely digested with BglII and separated on a 1% (w/v) agarose gel. The 2.5 kbp internal BglII fragment was deleted and the remaining 4.5 kbp fragment was purified using the Prep-a-Gene kit (Bio-Rad). The 2.2 kbp TSTE fragment from pUC19::TSTE (Sharetzsky et al., 1991) was excised using BamHI and gel-purified. TSTE was ligated to the 4.5 kbp BglII fragment of pHFJ2, yielding plasmid pHFJC (Fig. 1). The plasmid pHFJC was transformed into H. influenzae strain H1689 made competent using the MII medium of Spenser & Herriott (1965). Recombinant colonies were initially selected by growth on BHI agar containing ribostamycin (15 µg ml⁻¹). The mutant strain was designated H1689gxpAΔbglII. Confirmation of the appropriate chromosomal rearrangements was obtained by Southern blot analysis.

Cloning of DNA downstream of hgpA. The sequence of the insert of pHFJ2 was analysed, and a putative gene (hgpA) was identified based on the existence of an ORF and homology to other sequences in the databases. Since a cloned EcoRI fragment was used to generate pHFJ2 (Jin et al., 1996), the ORF ended with GAATTC, encoding glutamic acid (E) and phenylalanine (F). Sequence comparison with other homologous haem- and iron-utilization proteins suggested a stop codon following the nucleotides encoding EF. To confirm the presence of a stop codon, the region downstream of the proposed ORF was cloned. The primers HFJ25 and HFJTERMS (Table 1) were used in the PCR to amplify a 315 bp DNA fragment encoding the C terminus of hgpA. This gene fragment was used as a probe. H1689 chromosomal DNA was digested with EcoRV, and Southern blot analysis was performed using the 315 bp DNA probe. The probe hybridized with a DNA fragment of approximately 1.4 kbp. Restriction mapping of hgpA indicated an EcoRV site internal to hgpA (Fig. 1), approximately 0.6 kbp upstream of the C terminus; thus the 1.4 kbp hybridizing fragment would contain about 0.8 kbp downstream of the C terminus of hgpA. An H1689 mini-library was constructed by gel-purification of the DNA fragment of approximately 1.4 kbp in an EcoRV digest of H1689 chromosomal DNA and ligation of the purified fragments to the pCR-Blunt vector (Invitrogen). The library was screened with the 315 bp probe and a positive clone was identified and designated pHFJC'.

PCR amplification across a region of CCAA repeats. A pair of oligonucleotide primers designed PBUS and PBDS were designed 5 bp upstream and downstream of the CCAA repeats of HI1715, respectively (Table 1). Following repeated subculture, single colonies of H. influenzae HI1715 were picked from chocolate agar plates and chromosomal DNA extracted using Instagene kit (BIO-RAD) as directed by the manufacturer. Using HI1715 chromosomal DNA as template, the PCR amplification was performed in a total volume of 50 µl containing 5 µl of 10 × 30 mM Mg PCR buffer (Idaho Technology), [α-32P]dATP (10 µCi; 3.7 × 10⁸ Bq), 1 µl 10 mM dNTPs, 1 U Taq DNA polymerase and 10 µM of each primer. PCR was performed in a RapidCycler thermocycler (Idaho
Haemophilus influenzae haemoglobin binding

**(Fig. 1.** Partial restriction map of the insert of pHFJ2, the deletion/insertion derivative pHFJC and the clone encompassing the C-terminal region of hgpA and downstream sequence, pHFJC. The coding sequence of hgpA is shown by the boxed area. The TSTE insertion element encoding ribostamycin resistance is shown by the black area. The scale indicates size in kbp.)

**Table 1.** Primers used in this work

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5'→3')</th>
<th>Primer location</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBUS</td>
<td>GGGCTAACGGCAAGTGTTGCTT</td>
<td>5 bp upstream of ccaa repeats</td>
</tr>
<tr>
<td>PBDS</td>
<td>GTTCTAGTTGTTGTTCAAGAAGCATCC</td>
<td>5 bp downstream of ccaa repeats</td>
</tr>
<tr>
<td>HFJF4</td>
<td>GAAGTACCCTAGAATGATGC</td>
<td>Intergenic region between HI0594 and hgpA.</td>
</tr>
<tr>
<td>HFJ16</td>
<td>CGG GCAATTTCGGGGGAGGCC</td>
<td>hgpA</td>
</tr>
<tr>
<td>HFJ25</td>
<td>TATATCTTCATGCGATCAG</td>
<td>hgpA</td>
</tr>
<tr>
<td>HFJCTERMS</td>
<td>GAAATCAACTGAACGATGAC</td>
<td>hgpA</td>
</tr>
<tr>
<td>HFJ6</td>
<td>GCAGTTCTGTTTGAATTTC</td>
<td>hgpA</td>
</tr>
<tr>
<td>HFJ9</td>
<td>CATATTCATGAAAGGTTGGC</td>
<td>H10587</td>
</tr>
<tr>
<td>HFJXN</td>
<td>GACCGGAGGATCCATAGTATGCTTCT</td>
<td>hgpA</td>
</tr>
<tr>
<td>HFJXC</td>
<td>GGAAGGGGATCCTTGGAAATGAATGAC</td>
<td>hgpA</td>
</tr>
</tbody>
</table>

Technology). The reactions were cycled 30 times with an annealing step of 58 °C for 10 s, extension at 72 °C for 30 s and a denaturing step at 95 °C for 10 s. PCR products were visualized by running on a 6% denaturing polyacrylamide urea gel, dried and exposed to X-ray film as required. Sizing of the PCR products was achieved by running a sequencing reaction in lanes adjacent to the PCR products.

RT-PCR. This was performed essentially as described previously (Whitby et al., 1997). RNA (1 μg) prepared from haem-starved H1689 grown to mid-exponential phase (OD660 0.64) using the RNaseasy total RNA kit (Qiagen) as directed by the manufacturer was treated with 1 U DNase I (Gibco-BRL) in a total volume of 12.5 μl for 15 min at room temperature. The DNase I was inactivated by the addition of 1 μl 25 mM EDTA solution to the reaction mixture at 94 °C for 5 min. The cDNA or the negative-control reaction (3 μl) was used as template in the PCR. In addition, reactions using either H1689 chromosomal DNA or ddH2O as template were used as positive and negative controls, respectively. Three sets of primers were designed for PCR (Table 1). Primer pair 1 comprised HFJF4 and HFJ16, primer pair 2 comprised HFJ25 and HFJCTERMS, and primer pair 3 comprised HFJ6 and HFJ9.

PCR was performed for 30 cycles with the following parameters: for primer pair 1, each cycle consisted of denaturation at 95 °C for 1 min, annealing at 58 °C for 1 min and extension at 72 °C for 45 s; for primer pair 2, each cycle consisted of denaturation at 95 °C for 1 min, annealing at 50 °C for 1 min and extension at 72 °C for 20 s; for primer pair 3, each cycle consisted of denaturation at 95 °C for 1 min, annealing at 50 °C for 1 min and extension at 72 °C for 45 s.

The reactions were analysed on a 2% agarose gel.

**Growth studies.** Plate bioassays of haem source utilization
were performed on haem-deplete *H. influenzae* strains as previously described (Stull, 1987).

**Dot-blot assay for haemoglobin–haptoglobin binding.** A haemoglobin–haptoglobin dot-blot assay was performed as previously described (Jin et al., 1996). Haptoglobin was complexed with biotinylated haemoglobin as previously described (Stull, 1987). The complex equivalent to 500 ng haemoglobin ml⁻¹ was used in the dot-blot assay.

**Affinity isolation of haemoglobin-binding proteins.** Outer-membrane proteins were subjected to haemoglobin affinity purification as previously described (Jin et al., 1996). Eluted proteins were separated by SDS-PAGE on 7.5% acrylamide gels using the discontinuous buffer system of Laemmli (1970). Approximately 30 μl of protein preparation, representing proteins isolated from approximately 10⁶ organisms, was loaded per lane.

**RESULTS**

**Sequencing of hgpA and analysis of the deduced HgpA amino acid sequence**

The insert of the recombinant plasmid, pHFJ2, encoding the previously identified haemoglobin-binding activity of *H. influenzae* (Jin et al., 1996), was 4358 bp as determined by automated nucleotide sequencing. Analysis of pHFJ2 and the downstream region cloned in pHFJC' indicated an ORF of 3234 bp encoding a putative protein of 1077 amino acids with a molecular mass, after cleavage of the proposed leader sequence, of 120296 Da, which is consistent with the size of the affinity-purified protein. The predicted haemoglobin-binding protein (HgpA) encoded by the gene *hgpA* showed significant homology with other bacterial iron- and haem-related outer-membrane proteins, particularly across the seven regions considered typical of TonB-dependent outer-membrane proteins (Lundrigan & Kadner, 1986).

The genomic sequence of *H. influenzae* strain Rd KW20 reported by Fleischmann et al. (1995) contains four putative genes designated HI0635, HI0661, HI0712 and HI1566, which have significant homology with *hgpA*. The accession numbers of the GenBank entries corresponding to these four genes are U32746, U32749, U32754 and U32831, respectively. One particularly interesting feature of *hgpA* and its Rd KW20 homologues is a length of CCAA repeating units directly following the nucleotide sequence encoding the putative leader sequence (Table 2). We proposed that strand slippage across the CCAA region may be a mechanism of regulation (Jin et al., 1996), similar to that of the *H. influenzae* lipo-oligosaccharide (Weiser et al., 1989; Szabo et al., 1992). In the *hgpA* clone a full-length protein is encoded by the ORF; however, addition or loss of CCAA repeat units could change the reading frame and result in introduction of stop codons downstream of the CCAA repeat region (Fig. 2). To determine if variation in length over the CCAA region occurs in *H. influenzae*, PCR reactions were performed on chromosomal DNA across this region. Strain HI689 from which *hgpA* was cloned contains three genes which have lengths of CCAA repeating nucleotides immediately following the leader-peptide-encoding sequence. All three of these genes are highly homologous to each other and it has proven impossible to design primers which will amplify a *hgpA*-specific product of an appropriate size to demonstrate a change in CCAA length which may be as small as four nucleotides. Because of these complications, we selected strain HI1715, which con-

---

**Table 2. Comparison of HgpA with the proteins encoded by ORFs HI0635, HI0661, HI0712 and HI1566 in strain Rd and with HhuA**

<table>
<thead>
<tr>
<th>Gene/locus</th>
<th>Strain cloned from:</th>
<th>Binding activity*</th>
<th>% Homology to HgpA from strain HI689 (identity/similarity)†</th>
<th>Predicted prototypic gene product size (kDa)‡</th>
<th>No. of ccaa repeats</th>
<th>In/out of frame as cloned</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>hgpA</em></td>
<td>HI689</td>
<td>Hb/Hb–Hp</td>
<td>100/100</td>
<td>120</td>
<td>33</td>
<td>In</td>
</tr>
<tr>
<td>HI0635</td>
<td>Rd KW20</td>
<td>ND</td>
<td>49/65</td>
<td>119</td>
<td>21</td>
<td>Out</td>
</tr>
<tr>
<td>HI0661</td>
<td>Rd KW20</td>
<td>ND</td>
<td>59/74</td>
<td>112</td>
<td>20</td>
<td>Out</td>
</tr>
<tr>
<td>HI0712</td>
<td>Rd KW20</td>
<td>ND</td>
<td>49/67</td>
<td>121</td>
<td>37</td>
<td>In</td>
</tr>
<tr>
<td>HI1566</td>
<td>Rd KW20</td>
<td>ND</td>
<td>54/72</td>
<td>112</td>
<td>19</td>
<td>Out</td>
</tr>
<tr>
<td>HhuA</td>
<td>TN 106</td>
<td>Hb–Hp</td>
<td>84/90</td>
<td>117</td>
<td>13</td>
<td>Out§</td>
</tr>
</tbody>
</table>

ND, Not determined.

* Hb, haemoglobin; Hb–Hp, haemoglobin–haptoglobin complex.
† Percentage homology to HgpA is given as identical amino acids and functionally similar amino acids, based on translation of the sequence subsequent to the final ccaa repeat unit. Percentage homologies were generated using the GAP program in the GCG Program Package (Devereux et al., 1984).
‡ Gene product size was calculated following alteration of the number of ccaa units as appropriate and cleavage of the putative leader sequence.
§ Out of frame with the alternative start codon proposed in this paper.
A gene (hhuA) involved in haemoglobin–haptoglobin utilization by nontypable *H. influenzae* TN106 was recently reported by Maciver et al. (1996). HhuA is homologous to HgpA, showing 84% identity and 90% similarity; however, HhuA bound only haemoglobin–haptoglobin complex and not haemoglobin. To determine whether HgpA binds the haemoglobin–haptoglobin complex, a dot-blot assay was performed. The results demonstrated that HgpA bound the haemoglobin–haptoglobin complex (Fig. 4), in addition to the previously demonstrated haemoglobin binding (Jin et al., 1996).

Although HgpA is highly homologous to the products of the four ORFs in Rd KW20, it is unlikely that any of these ORFs represent *hgpA*. The nucleotide sequence upstream of *hgpA* in strain HI689 is highly homologous to the ORF designated HI0594 (95.2% identity) from the Rd KW20 genome. However, there is no homologue of *hgpA* at this site in the Rd KW20 genome (Fig. 5). The nucleotide sequence immediately downstream of *hgpA* is highly homologous to ORF HI0587 (84.4% identity), which is approximately 6.6 kbp downstream of ORF HI0594 in the Rd KW20 genome (Fig. 5). In Rd KW20 the 6.6 kbp sequence between HI0594 and HI0587 contains five ORFs designated HI0588–HI0592. Using a clone (GHIFY44) containing portions of the ORFs HI0591 and HI0592 as a probe, Southern blot analysis indicated that the locus between ORFs HI0591 and HI0592 is absent in *H. influenzae* strains HI689 and TN106 (Fig. 5). Similar results were obtained using probes encompassing the ORFs HI0588, HI0589 and HI0590 (data not shown).

H0661 is the ORF in Rd KW20 with greatest homology to *hgpA*; its gene product exhibits 59% identity and 74% similarity (Table 2). We have cloned a homologue of HI0661 from strain HI689 (Ren et al., 1998), designated *hgpB*, and shown that it also binds haemoglobin and the haemoglobin–haptoglobin complex. In addition, a direct repeat of 49 bp is present in the upstream region of *hgpA* and between the ORFs HI0594

---

**Fig. 2.** Nucleotide and amino acid sequence of the N-terminal region of HgpA showing the introduction of stop codons following removal of one or two CCAA repeats from the nucleotide sequence. Bold ccaa indicates removal of one CCAA unit; underlined ccaa indicates removal of two CCAA units. Arrow indicates the leader peptide cleavage site.

**Fig. 3.** Denaturing PAGE gel (6%) separating PCR products derived from amplification across a CCAA-containing locus of *H. influenzae* HI1715. Each lane contains the PCR product derived from a single independent colony of the same strain. Numbers on the left are sizes in bases derived from a sequencing reaction run in lanes adjacent to the PCR products.
and HI0592 in RdKW20 (Fleischmann et al., 1995). Regions homologous to the 49 bp direct repeat upstream of hgpA can be identified upstream of ORFs HI0635, HI0661 and HI1566 (data not shown). The function of this direct repeat is unknown, although it may be an element regulating expression of these genes. Additional regulatory elements may include a putative fur box identified 16 bp upstream of the ATG start codon of hgpA. The fur box is a conserved DNA sequence which binds the ferric uptake regulator (Fur) protein and is implicated in regulation of expression of iron- and/or haem-uptake-related proteins (Bagg & Neilands, 1987), although a specific role for Fur in the regulation of hgpA has not been demonstrated.

**Analysis of the hgpA transcript**

The sequences flanking hgpA indicated the presence of a putative promoter region. The hairpin structure and a poly(T) tract following the stop codon indicated a rho-independent terminator. This sequence analysis suggested that hgpA is transcribed as a monocistronic message. RT-PCR was used to investigate this possibility. Random nonamer primers were annealed to RNA prepared from haem-restricted H. influenzae HI689, and reverse transcriptase was used to generate cDNA. The cDNA was used as the template for PCR with three pairs of primers (Table 1, Fig. 6) designed such that primer pairs 1 (HFJ2F4 and HFJ16) and 3 (HFJ26 and HFJ29) would amplify PCR products of 650 bp and 660 bp, respectively, if the hgpA transcript was polycistronic. The control primer pair 2 (HFJ25 and HFJCTERS) was designed to amplify an internal 300 bp fragment of hgpA. With cDNA as template, only amplification with primer pair 2 resulted in a product, indicating that the hgpA transcript is monocistronic (Fig. 6). All primer pairs amplified a DNA fragment of the correct size when genomic DNA was used as template.

**Construction of a mutation in hgpA**

A mutation in hgpA was constructed by deletion of the internal 2.5 kbp BglII fragment and insertion of the 2.2 kbp TSTE antibiotic resistance marker (Sharetzsky et al., 1991). That the mutant HI689 strain was correctly constructed was confirmed by Southern analysis (Ren et
Haemophilus influenzae haemoglobin binding

Fig. 7. Plate bioassay showing utilization of the haemoglobin–haptoglobin complex (4 nmol haemoglobin equivalent) (well 1), haem (0.04 μmol) (well 2) or haemoglobin (4 nmol) (well 3) by H. influenzae HI689 (a) or the insertion/deletion mutant HI689hgpΔBgII (b).

al., 1998). The mutant strain was designated HI689hgpΔBgII.

Haem utilization and binding

The mutant strain HI689hgpΔBgII was unaltered in its ability to grow with either haemoglobin or the haemoglobin–haptoglobin complex as the sole source of haem in plate bioassays (Fig. 7). Following growth in haem-starved conditions the mutant was unaltered in its ability to bind either biotinylated haemoglobin or haemoglobin–haptoglobin complex in a dot-blot assay compared to the wild-type strain HI689 (data not shown).

Affinity isolation of haemoglobin-binding proteins from an hgpA mutant

Since there was no apparent alteration in either utilization or binding of haemoglobin or haemoglobin–haptoglobin, bacteria grown in haem-restricted media to mid-exponential phase were subjected to affinity purification using biotinylated haemoglobin as the primary ligand. The wild-type strain HI689 resulted in isolation of an approximately 120 kDa band, whilst the mutant strain HI689hgpΔBgII did not yield the 120 kDa band; however, a band at approximately 115 kDa was present in increased amounts (Fig. 8). Although the 115 kDa band has been detected at low levels in the wild-type strain, the band is clearly below the limit of Coomassie detection in some affinity isolations performed with the wild-type strain (Jin et al., 1996).

DISCUSSION

H. influenzae requires a porphyrin source for growth (Evans et al., 1974). However the availability of this essential nutrient is strictly limited within the human host. H. influenzae does not produce siderophores, and the mechanism(s) by which H. influenzae acquires haem and/or iron has not been fully elucidated. Proteins binding the haem–haemopexin complex, of which one is apparently secreted into the growth media, have been described (Cope et al., 1994, 1995, 1998; Hanson et al., 1992; Wong et al., 1995). Haem-binding outer-membrane proteins have been isolated (Lee, 1992; Thomas et al., 1998), and a haem-binding lipoprotein with significant homology to a periplasmic transport protein of E. coli has also been characterized (Hanson & Hansen, 1991). In addition, Reidl & Mekalanos (1996) have recently shown that protein e(P4) of H. influenzae is essential for utilization of haem, protoporphyrin IX and haem from haemoglobin. A short haem-binding motif homologous to regions of other haem-binding proteins was identified in protein e(P4) (Reidl & Mekalanos, 1996).

The acquisition and utilization of haem from haemoglobin, the haemoglobin–haptoglobin complex and the haem–haemopexin complex is dependent on a
functionalTonB gene, indicating that uptake is mediated by an outer-membrane TonB-dependent protein(s) (Postle, 1990; Jarosik et al., 1994). In this study the complete nucleotide sequence of the gene encoding a haem-repressible 120 kDa haemoglobin and haemoglobin–haptoglobin-binding protein (HgpA) of H. influenzae type b was determined. HgpA exhibits significant homology with other iron- and haem-related bacterial TonB-dependent proteins over regions that are highly conserved among this class of proteins (Lundrigan & Kadner, 1986). TonB-dependent outer-membrane protein receptors have been reported in a number of bacterial species, and are commonly involved in transport of iron and haem across the bacterial membrane (Bitter et al., 1991; Cornelissen et al., 1992; Gray-Owen & Schryvers, 1995; Lundrigan & Kadner, 1986; Pressler et al., 1988; Maciver et al., 1996). TonB is believed to act in energy transduction from the cytoplasmic membrane to the outer membrane (Postle, 1990; Larsen et al., 1994). HgpA did not contain a region homologous to the haem-binding motif identified in protein e(P4) by Reidl & Mekalanos (1996).

Whilst RdKW20 chromosome has homology to the DNA sequences flanking bgpA, there is no homologous bgpA at this locus, which instead contains the ORFs of H10588, H10589, H10590, H10591 and H10592 (Fleischmann et al., 1995). Southern hybridization analyses indicate that a 6.6 kbp DNA sequence present at this position in Rd KW20 is missing in both Hi689 and TN106 (Fig. 5).

There are four ORFs in the H. influenzae strain Rd KW20 chromosomal sequence, H10635, H10661, H10712 and H11566, with significant homology to bgpA (Fleischmann et al., 1995; Hood et al., 1996). All four ORFs in Rd KW20 and bgpA contain multiple CCAA repeats, where the number of repeats in each published reported gene varies between 19 and 37 (Table 2). Our data indicate that the number of CCAA repeats in a given locus may vary in H. influenzae (Fig. 3). We propose that the CCAA repeat region may regulate expression of these genes through a slip-strand mechanism. Such a mechanism may modulate expression of proteins in different host sites depending on the prevalent haem source. Alternatively, strand slippage may provide a mechanism to avoid the immunological response of the host. A similar mechanism has been demonstrated to mediate phase variation of the H. influenzae lic locus (Szabo et al., 1992). High et al. (1996) have recently shown that the CAAT repeat motif in the lic locus is required for phase variation but not for biosynthesis of lipo-oligosaccharide. The PCR amplification across a CCAA region from different colonies of the same strain yielded varying sized products (Fig. 3). These data provide evidence for the proposed slip-strand mechanism in H. influenzae.

Insertional inactivation of bgpA did not abrogate the ability of the mutant strain to either bind haemoglobin or the haemoglobin–haptoglobin complex, or to utilize either as a sole haem source. Affinity purification of haemoglobin-binding proteins from the mutant strain demonstrated the loss of the 120 kDa protein and increased isolation of a 115 kDa protein. These data suggest that at least one more haemoglobin-binding protein exists. We have shown that H. influenzae Hi689 contains a H10661 homologue, bgpB, which encodes a 115 kDa haemoglobin and haemoglobin–haptoglobin-binding protein (Ren et al., 1998). Elkins et al. (1995) also identified a 115 kDa haemoglobin-binding protein from H. influenzae strain DL42, although the relationship of this protein to either HgpA, HgpB or the ORFs in strain Rd KW20 is unclear. In some affinity isolations from the wild-type strain Hi689, there was no apparent expression of a 115 kDa protein (Jin et al., 1996), consistent with the hypothesis of phase variation of haemoglobin/haemoglobin–haptoglobin-binding protein expression. The mechanism of this apparent shift is as yet unknown, although loss of one protein may result in selection of populations in which strand slippage across the CCAA region has led to expression of an alternative protein.

Haemoglobin–haptoglobin binding was expressed by a fusion of hhuA lacking the CCAA repeats, indicating that the repeats are not essential for haemoglobin–haptoglobin binding (Maciver et al., 1996). The clone of bgpA bound both haemoglobin (Jin et al., 1996) and the haemoglobin–haptoglobin complex, whereas the clone of hhuA bound only the haemoglobin–haptoglobin complex (Maciver et al., 1996). Maciver et al. (1996) reported a start site immediately preceding the CCAA repeat units, which would lead to expression of a protein lacking a leader peptide. Removal of a CCAA repeat from their reported nucleotide sequence would bring the protein into frame with an alternative start codon, giving rise to a protein with a leader sequence homologous to that of HgpA.

Mutation of hhuA resulted in a decreased ability of the mutant to bind haemoglobin–haptoglobin when grown in a haem-replete medium (Maciver et al., 1996). However, no effect was seen on haemoglobin–haptoglobin binding when bacteria were grown in a haem-restricted medium. The mutant also showed a reduced ability to utilize haemoglobin–haptoglobin as a haem source, but was unaltered in the ability to utilize haemoglobin (Maciver et al., 1996). Thus hhuA appears to encode constitutive haemoglobin–haptoglobin binding ability, whilst there is also an independent haem-repressible haemoglobin–haptoglobin-binding activity expressed by this strain. Similarly, there may be constitutive and haem- or iron-repressible haemopexin-binding proteins expressed by H. influenzae (Cope et al., 1995; Wong et al., 1994, 1995). In contrast, haemoglobin binding appears to be haem-repressible (Frangipane et al., 1994), although no extensive survey has been reported among organisms grown under haem-replete conditions nor have other conditions been investigated.

The similarity between bgpA and hhuA is 90%, and H. influenzae TN106, the strain from which hhuA was cloned, apparently lacks the same 6.6 kbp sequence missing in Hi689 (Maciver et al., 1996) (Fig. 5). Thus it
is likely that hgpA and hhuA represent alleles of the same gene in different strains. However, expression of clones of hhuA and hgpA in E. coli apparently resulted in different phenotypes; hhuA bound only the haemoglobin–haptoglobin complex and hgpA bound both the complex and haemoglobin alone. Additionally, insertion mutants of hhuA and hgpA in TN106 and H1689, respectively, result in different phenotypes, with the hhuA mutant showing reduced utilization of the haemoglobin–haptoglobin complex, whereas the hgpA mutant is unaltered in this respect. Neither mutant is apparently altered in the ability to bind the haemoglobin–haptoglobin complex following haem-restricted growth.

These discrepancies may arise from the reported construction of the hhuA clone as a fusion protein lacking approximately 150 amino acids at the N-terminal end of the protein (Maciver et al., 1996). It is possible that the missing 150 amino acids are essential for haemoglobin binding, or alternatively that the fusion inhibits haemoglobin binding but not binding of the haemoglobin–haptoglobin complex. These data may reflect differences in the growth techniques, or alternatively may reflect varying complements of haemoglobin and/or haemoglobin–haptoglobin-binding proteins between these strains. In order to elucidate the haemoglobin and haemoglobin–haptoglobin utilization pathways of H. influenzae, it will be necessary to define the functions of the additional CCAA-containing ORFs and to define the gene complements of the strains under investigation.

In conclusion, a haemoglobin-binding protein gene, hgpA, of H. influenzae and several ORFs possessing CCAA repeating units at the 5' ends and encoding putative proteins of high homology to HgpA have been identified. Expression of these genes may be regulated via a slip-strand mechanism. Based on homology at the amino acid level, these proteins are likely to have similar functions. Work is in progress to define the function of each of the gene products, to elucidate the mechanisms of regulation and to define the potential role of the gene products in the pathogenesis of H. influenzae disease.

ACKNOWLEDGEMENTS

This work was supported by the Children’s Medical Research Institute, Public Health service grant AI29611 from the National Institute of Allergy and Infectious Disease, USA, to T.L.S., and by Health Research contract HN5-055 from the Oklahoma Center for the Advancement of Science and Technology, USA, to D.J.M.

REFERENCES


