THE UNIVERSITY OF OKLAHOMA HEALTH SCIENCES CENTER

GRADUATE COLLEGE

CHARACTERIZATION OF AN *HAEMOPHILUS INFLUENZAE* GENE, *hgpA*, ENCODING A HEMOGLOBIN AND HEMOGLOBIN-HAPTOGLOBIN BINDING PROTEIN

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

In partial fulfillment of requirements for the

degree of

DOCTOR OF PHILOSPHY

BY

Hongfan Jin

Oklahoma City, Oklahoma

1998

UNIVERSITY OF OKLAHOMA
LIBRARIES

)4
THESIS
JIN
cop. 2

CHARACTERIZATION OF AN *HAEMOPHILUS INFLUENZAE* GENE, *hgpA*, ENCODING A HEMOGLOBIN AND HEMOGLOBIN-HAPTOGLOBIN BINDING PROTEIN

APPROVED BY:

Terrence L. Stull, M.D., Chair

David W. Dyer, Ph.D.

Don C. Graves, Ph.D.

Michael S. Gilmore, Ph.D.

David R. McCarthy, Ph.D.

DISSERTATION COMMITTEE

ACKNOWLEDGEMENTS

I wish to express my deepest appreciation to my mentor and advisor, Dr. Terrence L. Stull. Without his guidance and unlimited encouragement, I would not have stay focused and finished my dissertation project. His patience, enthusiasm, and being positive made these years a pleasant experience. I am truly grateful to have worked with him.

I would like to extend my appreciation to my committee members, Drs. David Dyer, Don Graves, Michael Gilmore, and David McCarthy. Thank you all for your willingness to provide your time and input whenever it was needed. A special thank you goes to Dr. Dyer, who has constantly given me encouragement, support and advice

I would also like to thank Dr. Daniel Morton, who was not only a lab companion, but also a teacher and a friend. I can always count on him for help, suggestion, and friendship. I will always remember the many friends I came to know in Oklahoma. Dr. Paul Whitby, his help and sense of humor made it a pleasant experience to work with him; Dr. Robert McLaughlin, his genuine enthusiasm and insightful ideas about research are keenly appreciated.

I am thankful for my parents and my sister, without their love, spiritual support, I would not have been strong enough to walk this path.

Last but not least I dedicate this work to Zhen Ren, my husband, my classmate, and my colleague. Thank you for your understanding, love, and support.

COPYRIGHT

By

Hongfan Jin

December 11, 1998

ACKNOWLEDGEMENTS

I wish to express my deepest appreciation to my mentor and advisor, Dr. Terrence L. Stull. Without his guidance and unlimited encouragement, I would not have stay focused and finished my dissertation project. His patience, enthusiasm, and being positive made these years a pleasant experience. I am truly grateful to have worked with him.

I would like to extend my appreciation to my committee members, Drs. David Dyer, Don Graves, Michael Gilmore, and David McCarthy. Thank you all for your willingness to provide your time and input whenever it was needed. A special thank you goes to Dr. Dyer, who has constantly given me encouragement, support and advice.

I would also like to thank Dr. Daniel Morton, who was not only a lab companion, but also a teacher and a friend. I can always count on him for help, suggestion, and friendship. I will always remember the many friends I came to know in Oklahoma, Dr. Paul Whitby, his help and sense of humor made it a pleasant experience to work with him; Dr. Robert McLaughlin, his genuine enthusiasm and insightful ideas about research are keenly appreciated.

I am thankful for my parents and my sister, without their love, spiritual support, I would not have been strong enough to walk this path.

Last but not least I dedicate this work to Zhen Ren, my husband, my classmate, and my colleague. Thank you for your understanding, love, and support.

TABLE OF CONTENTS

List of Tables..vi

List of Figures...vii

Chapter I. Introduction..1

Chapter II. Cloning of a DNA fragment encoding a heme-repressible hemoglobin-binding outer membrane protein form *Haemophilus influenzae*..............30

Chapter III. Characterization of *hgpA*, a gene encoding a hemoglobin and hemoglobin-haptoglobin binding protein of *Haemophilus influe*..............................65

Chapter IV. Regulation by heme of the hemoglobin and hemoglobin-haptoglobin binding protein gene (*hgpA*) from *Haemophilus influenzae*...........................113

Chapter V. Summary...134

Appendix..137

LIST OF TABLES

Table Page

1. Strains and plasmids……………………………………………………………………34
2. Primers used in this work………………………………………………………………74
3. Comparison of HgpA with the proteins encoded by ORFs HI0635, HI0661, HI0712, and HI1566 in strain RdKW20 and with HhuA in strain TN106………………………91
4. The loci between HI0587 and HI0594 in RdKW20 and their assigned functions……92
5. Primers used in construction of the polycompetitor pJOK and in CQ-RT-PCR……126
6. LacZ activity under different growth conditions……………………………………127

LIST OF FIGURES

Figure Page

1. A hypothetical model for heme acquisition in *H. influenzae*........................9

2. Identification of a hemoglobin binding protein in *H. influenzae*........................42

3. Southern blot analysis of *H. influenzae* HI689 genomic digests........................44

4. Southern blot probed with ECL-labeled oligonucleotides specific for the N-terminus of the isolated hemoglobin binding protein........................45

5. Partial restriction map of the 4.2-kb insert of pHFJ2, the predicted ORF encoding Hgp........................46

6. Peptide alignments between the N-terminal region of the putative Hgp and the N-terminal region of other iron and heme related proteins........................47

7. Peptide alignments between the C-terminal region of the putative Hgp and the C-terminal region of other iron and heme related proteins........................48

8. Dot blot analysis of human hemoglobin binding to *E. coli* either harboring or without pXHGP........................50

9. Partial restriction enzyme maps of the 4.2-kb insert of pHFJ2, the deletion/insertion derivative pHFJC, and the clone encompassing the C-terminal region of *hgpA* and downstream sequence pHFJC'........................70

10. Southern blot analysis of *Rsa*I digested chromosomal DNA of *H. influenzae* strain HI689 and the mutant HI689hgpΔ*Bgl*II........................75

11. Comparison of the HgpA amino acid sequence with the sequences of other bacterial iron and heme related proteins over the 7 regions considered indicative of a TonB-dependent outer membrane protein........................81

12. Complete nucleotide sequence of *hgpA* and the deduced amino acid sequences of predicted ORF........................82

LIST OF FIGURES (Cont.)

Figure Page

13. Amino acid sequence of the N-terminal region of HgpA showing introduction of stop codons following removal of one or two CCAA repeats from the nucleotide sequence 85

14. 6% denaturing PAGE gel showing PCR products derived from amplification across the CCAA-containing locus of *H. influenzae* type b strain HI1715..........86

15. Whole cell hemoglobin-haptoglobin binding dot blot assay..........87

16. Comparison of the organization of the strain HI689 at the *hgpA* locus with the organization of RdKW20 genome..........89

17. RT-PCR analysis of transcripts of *hgpA*..........96

18. Plate bioassay to compare utilization of various heme sources by wild-type HI689 and the *hgpA* mutant strain..........97

19. SDS-PAGE gel shows affinity isolated proteins from wild-type and mutant strain of HI689..........98

20. Comparison of the amino acid sequence N-terminal region of HgpA from HI689 with the N-terminal region of HhuA and the N-terminal region of HhuA modified by removal of a CCAA repeat unit from *hhuA*..........105

21. Construction of DNA mimics and the polycompetitor plasmid pJOK..........120

22. CQ-RT-PCR analysis to quantitate *hgpA* transcripts of HI689..........124

23. SDS-PAGE gel showing whole proteins from *E. coli* either harboring or not harboring the cloned *hgpA* without CCAA repeats..........139

CHAPTER I

INTRODUCTION

Haemophilus influenzae: pathogenesis and disease

Haemophilus influenzae is a fastidious pleomorphic Gram-negative human specific pathogen which requires the X and V factors, present in blood, for growth (23). The heat-stable X factor is protoporphyrin IX, the immediate biosynthetic precursor of heme (100). The heat-labile V factor is nicotinamide mononucleotide, usually provided as nicotinamide adenine dinucleotide (NAD) or NAD phosphate (NADP), which functions as a coenzyme for pyridine-linked dehydrogenase.

H. influenzae was first isolated and described in 1892, however based on its frequent association with a clinical syndrome, it was erroneously named the "influenza bacillus" (74). The genus *Haemophilus* was originally proposed by the American Committee on Nomenclature in 1920 (101), to encompass a group of rod shaped Gram-negative cells which grew best in the presence of hemoglobin and in general required serum or ascitic fluid. Studies on the "influenzae bacillus" following the 1918 influenza pandemic revealed that true "influenza bacillus" is *Haemophilus influenzae* (80). Up to 80% of healthy individuals are colonized with nontypeable *H. influenzae*, which possesses no capsular polysaccharide; fewer than 5% are colonized with encapsulated *H. influenzae*, which are grouped into 6 serotypes, designated a through f, according to the

specific capsular polysaccharide (30,64,94). Type b strains, with capsule composed of repeating units of ribosyl-ribitol phosphate, have received most attention historically (89).

Colonization of the nasopharynx with *H. influenzae* occurs after the inhalation of infected droplets. The relatively common asymptomatic colonization occasionally develops into symptomatic disease that may spread contiguously to the sinuses, middle ear, or bronchi. Respiratory infections are important sources for seeding in the blood to produce invasive diseases in the meninges or joints and for the invasion of local tissues to cause epiglottis, pneumonia, or cellulitis (26,69). The type b capsule is the best studied virulence factor, and the role of other pathogenic mechanisms of *H. influenzae* is less well understood.

Before the development of a vaccine, *H. influenzae* type b (Hib) was responsible for virtually all invasive infections caused by *H. influenzae*, and it was the most common cause of bacterial meningitis among young children in the United States (28,65). It was also the cause of such invasive diseases as cellulitis, pneumonia, and acute epiglottitis (11). Since the introduction of protein-conjugated capsular vaccines in 1988, the incidence of invasive Hib infections has declined by at least 95% among infants and young children (8,22,70,83)

However, vaccine failures occur, and the current vaccines are ineffective against unencapsulated (nontypeable) *H. influenzae*, which is a leading cause of otitis media in children (29). Complications of otitis media include chronic otitis media, acute

mastoiditis, and facial nerve paralysis (45). Unencapsulated *H. influenzae* can also cause bacteremic pneumonia among adults (97).

Currently, the most important diseases caused by *H. influenzae* are otitis media and pneumonia. A more effective vaccine against both type b and unencapsulated *H. influenzae*, or an additional vaccine protecting against unencapsulated *H. influenzae* would contribute significantly to children's health. Until effective vaccines are developed, infections due to *H. influenzae* will remain an important public health issue for the near feature.

Virulence factors of *H. influenzae*

The ability of a microorganism to survive and produce pathogenic effects in the human host is dependent on virulence factors. Potential virulence determinants of *H. influenzae* include the polysaccharide capsule, lipooligosaccharide (LOS) of Hib strains, IgA protease, fimbriae and iron/heme acquisition mechanisms, which will be discussed in detail in the next section.

Capsule The polyribosylribitol phosphate (PRP) capsule of Hib is critical in the pathogenesis of invasive disease caused by this organism. Systemic infections are almost always caused by encapsulated strains and until recent vaccine development, the vast majority of these strains elaborate the type b capsular polysaccharide. The type b capsular polymer is unique in that it contains the pentose sugar, ribose and ribtol phosphate, instead of hexoses or hexosamines as found in other serotypes (17). Antibody

to this polyribosylribitol phosphate capsular antigen plays a key role in protection from Hib infection. Protein conjugate vaccines of type b capsule have radically reduced the incidence of disease caused by type b strains (70).

Lipooligosaccharide In addition to the capsule, LOS has been implicated as a major contributory factor to the pathogenicity of *H. influenzae*. The LOS of *Haemophilus* species is a major surface antigen and makes up approximately 4% of the organism's dry weight (26). *H. influenzae* LOS exerts a paralyzing action on the ciliated respiratory epithelium and promotes proliferation of the organism in the bronchial tree (91). A characteristic feature of *H. influenzae* LOS is the high frequency loss or gain of oligosaccharide epitopes, referred to as phase variation (47). Slipped-strand mispairing across the 5'-CAAT-3' repeating units at the 5' end of the genes involving LOS biosynthesis contribute to LOS phase variation (59,60,92). The ability to vary cell-surface composition may play a key role in the evasion of antigen-specific host immune defenses. In addition, it may promote the expression of the most appropriate cell-surface structures for a given environment or stage in pathogenesis, thereby optimizing the virulence potential of an organism (41).

IgA protease *H. influenzae* is one of the several bacterial species (*H. influenzae*, *Neisseria gonorrhoeae, N. meningitidis, Streptococcus pneumoniae, S. sanguis*) known to produce IgA protease, enzymes that have the unique ability to hydrolyze the human IgA of the IgA1 subclass heavy chain as their only known substrates (46). Human immunoglobulins of the IgA2 subclass are resistant to cleavage

because the primary structure of the heavy chain has a deletion of 13 amino acids in the hinge region that includes those peptide bonds attacked by the IgA protease (76). Because *H. influenzae* primarily infects human mucosal surfaces where host defense is mediated by secretory IgA, IgA protease may represent a significant virulence factor (66). *H. influenzae* produces three distinct types of IgA protease, each cleaving different sites within the hinge region (67). However a specific function or advantage of possession of IgA protease has not been determined in the pathogenesis of *H. influenzae*, and it has been demonstrated that IgA protease-deficient mutants behave in a similar manner to the wild-type strains in terms of mechanisms of attachment, invasion, and ability to remain within phagocytic vacuoles (25).

Pili (or Fimbriae) *H. influenzae* has evolved organelles, pili (or fimbriae), to facilitate its association with epithelial surfaces, thus preventing removal of the organism by natural mechanical and cleansing forces and sustaining successful colonization of the human respiratory tract (30). The expression of pili by *H. influenzae* is regulated by phase variation (24). Although the role of pili in mediating adherence of *H. influenzae* to respiratory epithelial cells is clear, their role in the establishment or maintenance of nasopharyngeal colonization is difficult to ascertain (1,98). Furthermore, it is suggested that piliated *H. influenzae* is cleared from the circulation more readily than are nonpiliated organisms (31,34,93). Thus, the most important roles of pili in the pathogenic pathway of *H. influenzae* infection appear to be their mediation of bacterial adherence to mucosal surfaces and their facilitation of respiratory tract colonization.

Heme and/or iron requirements by *H. influenzae*

Like many other transition elements, iron is an important biological catalyst because it possesses unfilled *d* atomic orbital and is able to undergo changes in oxidation states involving one electron. This easy access to two oxidation states, Fe^{2+} (ferrous iron) and Fe^{3+} (ferric iron), allows it to coordinate electron donors and to participate in redox process (6). Iron is the metal usually selected in electron transfer, not only because of its abundance, but also for its ability to interact with O_2, H_2, and H^-/H^+. However, the main advantage of iron as a catalytic element resides in the fact that the redox properties can be modified by its interaction with coordinating ligands so that the ease of electron acceptance can vary over a wide range (32). Iron is also a component of key molecules such as cytochromes, ribonucleotide reductase, and other metabolically linked compounds, and is, thus an essential nutrient for bacterial growth.

H. influenzae requires an exogenous source of porphyrin for growth because it lacks all the enzymes of heme synthesis pathway except ferrochelatase, which catalyzes the final insertion of iron into the porphyrin ring (100). Heme is comprised of a metal iron chelated in a porphyrin ring; thus heme can satisfy iron and porphyrin requirements. The biological reactivity of heme derives from the ability of iron to undergo reversible oxidative change. This property permits heme, as a prosthetic group of a variety of hemeproteins, to mediate biological functions such as oxygen transport, electron-transfer with the generation of cellular energy, and hydrogen peroxide use (95).

Under normal physiological conditions, the free ionic iron concentration is far too low to support the growth of most microorganisms due to poor water solubility of iron and the ability of host iron-binding proteins to scavenge iron (32). Intracellularly, iron is sequestered by ferritin, heme, and heme compounds such as hemoglobin. Extracellularly, iron is sequestered by the glycoproteins transferrin in serum and lactoferrin in secretions (62). These proteins have association constants for iron of about 10^{36} and they are normally only partly saturated. Thus, although there is an abundance of iron present in body fluids, the amount of free iron in equilibrium with iron-binding proteins can be calculated to be the order of 10^{-18} M (2). In addition, during infection the host reduces the total amount of iron bound to serum transferrin and increase intracellular iron storage (7,99), the so called hypoferremia of infection, possibly due to the release of lactoferrin from neutrophils and increased synthesis of ferritin. Therefore the availability of free iron to the microorganism is strictly limited, and this iron limitation is a major obstacle for invading organisms to survive in a human host.

In the human body, heme is sequestered intracellularly in heme proteins, such as hemoglobin and the cytochromes, or when present in the circulation, bound to the serum proteins hemopexin and albumin (in the case of heme) (84), and by haptoglobin (in the case of hemoglobin) (48,68). Thus, heme is unavailable to invading microorganisms without mechanisms of acquiring iron/heme from these sources.

To overcome iron/heme limitations, pathogenic bacteria have evolved iron and heme uptake mechanisms specific to its niche in humans. The best understood

bacterial iron-uptake system is the siderophore-dependent iron transport system. Siderophores are low molecular weight high affinity iron chelators produced by bacteria of genera such as *Escherichia*, *Salmonella*, *Shigella* and *Klebsiella* under iron restricted conditions (18,33). Siderophores can remove iron from iron-binding proteins lactoferrin and transferrin efficiently and transport it into the bacterial cell via specific outer membrane protein receptors (73). *H. influenzae* does not produce siderophores (40,75), but acquires iron through the direct interaction of outer membrane proteins and human iron binding proteins. *H. influenzae* can use ferric-transferrin, hemoglobin, hemoglobin-haptoglobin complex, heme-hemopexin and heme-albumin complex as iron sources (72,90). The mechanisms by which *H. influenzae* takes up iron from these proteins have not been fully elucidated.

In vitro experiments indicate that the heme requirement can be satisfied by heme, hemoglobin, hemoglobin complexed to haptoglobin and heme complexed to hemopexin or albumin (90). Heme acquisition in bacteria involves a receptor-mediated process in which the specific interaction with surface exposed components would constitute the initial step in heme uptake (5,52,71). It is proposed that heme uptake should encompass the following features: a way to access intracellular and extracellular hemeproteins; the expression of outer membrane proteins specific for the hemeproteins; the ability to transport heme intact across the cell membranes; the presence of energy-furnishing components; the mechanisms of regulating the expression of the network (52).

Accordingly, a hypothetical model for heme acquisition in *H. influenzae* is proposed (Fig.1).

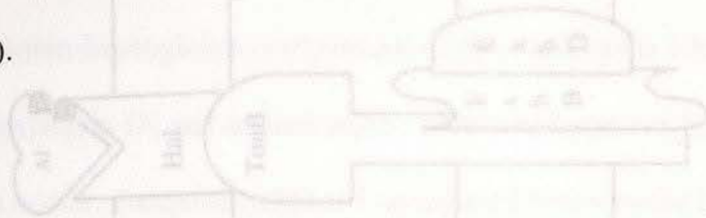
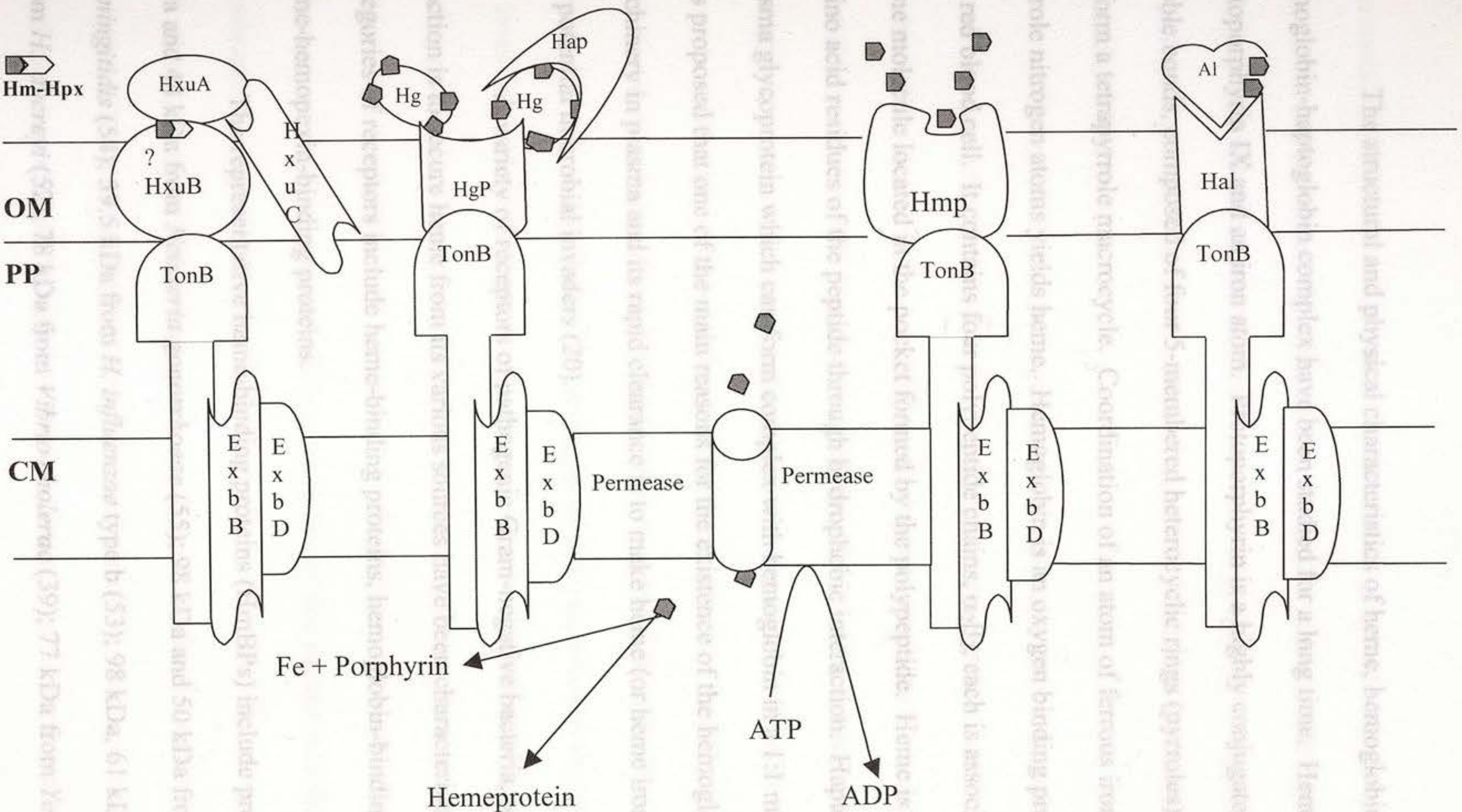


Figure. 1. A hypothetical model for heme acquisition of *H. influenzae*. Heme (Hm), heme-hemopexin (Hm-Hpx), hemoglobin (Hg), hemoglobin-haptoglobin (Hg-Hap) and Heme-Albumin (Hm-Alb) are bound by specific receptors at the cell surface (Although no specific Heme-Albumin receptor has been identified). The exact membrane topology and precise identity of each of the binding components are the subject of speculation. Following the specific recognition, heme is released from the receptor-ligand complex by an undefined mechanism and transported across the outer membrane into the periplasmic space by a TonB protein energized process. Subsequently, an ABC-type transported system conveys heme into the cell cytoplasm, where heme may be directly incorporated into hemeproteins or the iron removed from the porphyrin ring by the action of heme oxygenase homologue, with entry of iron into the intracellular iron pool (52). OM represents the bacterial outer membrane, CM denotes the inner membrane, and PP indicates the periplasmic space. HmP refers to heme binding protein, HgP refers to hemoglobin binding protein, HxuA,B, and C represent heme-hemopexin binding proteins, and Hal refers to heme-Albumin binding protein.

Figure 1.



The structural and physical characteristics of heme, hemoglobin, and hemoglobin-haptoglobin complex have been studied for a long time. Heme consists of protoporphyrin IX and an iron atom. Protoporphyrin is a highly conjugated system of double bonds, composed of four 5-membered heterocyclic rings (pyrroles) fused together to form a tetrapyrrole macrocycle. Coordination of an atom of ferrous iron by the four pyrrole nitrogen atoms yields heme. Hemoglobin is an oxygen binding protein found in the red blood cell. It contains four polypeptide chains, $\alpha_2\beta_2$, each is associated with a heme molecule located in the pocket formed by the polypeptide. Heme is anchored to the amino acid residues of the peptide through hydrophobic interaction. Haptoglobin is a plasma glycoprotein which can form complex with hemoglobin in a 1:1 molecule ratio. It is proposed that one of the main reasons for the existence of the hemoglobin sequesting machinery in plasma and its rapid clearance is to make heme (or heme iron) unavailable for potential microbial invaders (20).

A variety of receptors of pathogenic Gram-negative bacteria, whose proposed function is to secure heme from its various sources have been characterized. Three categories of receptors include heme-binding proteins, hemoglobin-binding proteins, and heme-hemopexin-binding proteins.

The representative heme-binding proteins (HmBPs) include proteins of 97 kDa and 44 kDa from *Neisseria gonorrhoeae* (55); 98 kDa and 50 kDa from *N. meningitidis* (54); 39.5 kDa from *H. influenzae* type b (53); 98 kDa, 61 kDa, and 43 kDa from *H. ducreyi* (52); 78 kDa from *Vibrio cholerae* (39); 77 kDa from *Yersinia*

enterocolitica (86); and 26 kDa from *Porphyromonas gingivalis* (3). In addition it has been shown that protein *e*(P4) is essential for utilization of heme in *H. influenzae* (78). Inherently hydrophobic, heme displays a pronounced propensity to aggregate at physiological pH (5), therefore the physiological role of a heme-binding protein in heme acquisition is unknown.

The characterized hemoglobin-binding proteins include proteins of 89.5 kDa from *N. meningitidis* (HmbR) (87); 100 kDa from *H. ducreyi* (HgbA) (21); 120 kDa from Hib strain 689 (HgpA) (43). A 117 kD hemoglobin-haptoglobin binding protein was isolated from nontypeable *H. influenzae* (NTHI) (HhuA) (58); and an 85 kDa from *N. meningitidis* to acquire hemoglobin-haptoglobin associated iron (HpuB) (56).

Heme-hemopexin-binding proteins have only been found in *H. influenzae*. A gene cluster encoding proteins of 100 kDa (HxuA), 60 kDa (HxuB), 78 kDa (HxuC) has been studied (12-14,36). Insertional mutagenesis analysis indicates that all three genes are involved in heme-hemopexin utilization. Evidence suggests that heme-hemopexin-binding protein HxuA is released into the environment; its secretion and activation may be controlled by HxuB protein. Alternatively, the presence of soluble HxuA protein in culture supernatants may reflect a loose association of this protein with the outer membrane, although HxuB probably functions in the release of HxuA from the bacterial cell. HxuC, localized in the outer membrane, showed similarities to the TonB-dependent receptors. The function of HxuC remains to be determined, although it may interact with HxuA to transport heme through the outer membrane (14). Using a heme-hemopexin

affinity resin, a set of three proteins of 29, 38 and 57 kDa independent of HxuA was isolated from Hib strain 706705 (102). Whether these genes constitute a distinct system for acquiring heme from heme–hemopexin has yet to be determined.

TonB dependent transport of heme and/or iron across the bacterial cell membrane

The mechanism by which iron and heme are transported across the Gram-negative bacterial membrane is not clear. However, it is clear that TonB is involved in the energy transduction from the cytoplasmic membrane to the outer membrane receptors. TonB was first related to the process of energy transduction during the study of the infection of phage T1 in *E. coli* (35). Phage resistant cells were mutated in two genes, designated *tonA* and *tonB* (ton from T one). TonA is an outer membrane protein and T1 does not bind to a *tonA* mutant. T1 adsorbs to *tonB* mutants reversibly and cellular energy is required for irreversible adsorption accompanied by infection, which suggests that TonB is somehow involved in energy transduction (35). The exact role TonB plays in the transport of molecules into the bacterial cell remains to be determined. Currently there are three hypotheses. First, the theory of periplasmic interaction, suggests TonB could energize outer membrane receptors directly by spanning the periplasm and interacting with the outer membrane receptors at the periplasmic face of the outer membrane (OM). Secondly, the theory of a mobile messenger suggests TonB could act by generating a high-energy intermediate molecule that diffuses across the periplasmic space to activate outer membrane receptors. Thirdly, the theory of adhesion

zones, suggests that the C-terminus of TonB may span the outer membrane bilayer at sites juxtaposing the inner and outer membrane (49). While there is evidence for and against each theory, the periplasmic interaction is most widely accepted (77).

TonB dependent outer membrane proteins share a series of highly conserved regions, one of which is called the TonB box, at their amino termini (38,82). Many studies emphasize the importance of this region as a mediator of the physical interaction between TonB and TonB-dependent receptors. Other evidence suggests that it is the conformation of the TonB box rather than the specific amino acid sequence that dictates productive interactions of energy coupling with TonB (44,50).

Other proteins are also involved in energy transduction. Transport of bound chelates across the outer membrane depends upon TonB-ExbB-ExbD, a cytoplasmic membrane-localized complex that transduces energy from the proton motive force to high affinity receptors in the outer membrane (44,51,77,88). TonB interacts preferentially and directly with ligand-loaded receptors. Such a mechanism ensures the productive use of cellular energy to drive active transport at the outer membrane (49,61,77). The TonB-ExbB-ExbD complex has been identified in many Gram-negative bacteria, and it probably represents a conserved mechanism to transduce energy to high-affinity transportors within the outer membrane (51,88). The cellular location of TonB is uncertain. The amino terminal end of TonB may remain embedded in the cytoplasmic membrane, while the rest of the protein extends into the periplasm.

Mutant analysis indicates *H. influenzae* TonB protein is required not only for heme utilization *in vitro*, but also for virulence of *H. influenzae* type b in an animal model (42).

Gene regulation of iron and/or heme related proteins

Iron is essential for most organisms. However, iron is also deleterious since it catalyzes hydroxyl free radicals generated through Haber-Weiss reactions, leading ultimately to cell death (16). Therefore, it is not surprising that production of the cellular components responsible for utilizing iron is tightly controlled under different environment conditions (19).

There are a multitude of mechanisms to regulate iron or heme related proteins. The best studied mechanism is the regulation of the aerobactin operon by the iron-Fur (Ferric uptake regulation) complex in *E. coli*. Fur is a 17 kDa DNA binding protein rich in histidine (37). At a sufficient intracellular iron content, Fe^{2+} associates with the Fur protein, and converts the Fur to a dimeric repressor that binds to a genomic control region, the so-called Fur box, upstream of iron-regulated genes. When the concentration of iron is decreased, the Fur-repressed genes become derepressed and initiation of transcription occurs. The Fur box in the aerobactin promoter of *E. coli* has a 19 bp symmetrical-dyad consensus sequence (GATAATGATAATCATTATC). Similar sequences are found in the promoter regions of a large number of genes whose expression is regulated by the Fur protein (4). A similar sequence was also found in the promoter

region of *fur*, suggesting the auto-regulation of the expression of the Fur protein (57). Genetic and biochemical evidence also indicates a moderate auto-regulation of *fur* expression by its gene product and a significant stimulation by the cAMP-CAP (catabolite-activator protein) system in *E. coli* (57). The auto-regulation may serve as a mechanism to maintain intracellular Fur levels below a certain limit, and catabolic activation suggests a relationship between iron regulation and the metabolic status of the cells (57).

Heme plays an important role in the regulation of the expression of heme and iron utilization proteins. In *H. influenzae*, heme satisfies not only the iron requirement but also an indispensable need for porphyrin (15). Therefore, in addition to elemental iron, there is need for control of the diverse array of receptors with specificity to the various hemeproteins that are expressed on the surface of *H. influenzae* (52). Dot blot assay using a human transferrin-horseradish peroxidase conjugate, indicates that expression of the *H. influenzae* transferrin receptor is repressible by heme but not elemental iron alone (63). In addition, hemoglobin binding activity is repressed by heme in *H. influenzae* (27) and *H. ducreyi* (21). The molecular mechanisms underlying heme-mediated regulation have not been elucidated. However, spectroscopic experimental data showed the tight binding of heme to Fur, suggesting a potential role for Fur-heme complexes in regulating the expression of heme acquisition proteins (85).

Post transcriptional regulation of hemoglobin binding proteins via phase variation was found in *N. gonorrhoeae* and *H. influenzae*. In *N. gonorrhoeae*, the

expression of hemoglobin binding proteins HpuA and HpuB are controlled at the translational level by frame shift in a run of guanine residues in the *hpuA* sequence encoding the mature HpuA protein (9). Similarly in *H. influenzae*, the expression of hemoglobin binding protein (HgpA) may be regulated through strand slippage across multiple CCAA repeats within the structural gene (43,79). Such a mechanism may provide the advantage of protein expression in different host sites depending on the prevalent heme and/or iron source. Alternatively, strand slippage may provide a mechanism to evade host immunological response.

Another interesting regulatory mechanism has been demonstrated in the fish pathogen *V. anguillarum*, where an antisense RNA (RNAα) regulates the iron transport genes *fatA* and *fatB* (ferric anguibactin transport) (96). RNAα is encoded within the *fatB* gene and is preferentially expressed under iron-rich conditions. A high level of RNAα synthesis led to a reduced expression of the iron transport genes *fatA* and *fatB*(81). Experimental evidence indicated that it is not the promoter of RNAα which is regulated by iron, but that the stability of RNAα depends on the presence of iron (10), although the mechanism of stabilization is not clear.

Obviously, regulation mechanisms need to be studied in greater detail. It is not surprising that pathogenic bacteria have evolved a variety of schemes to regulate genes encoding virulence factors. The ability to produce virulence factors when they are needed and to stop producing factors that are no longer appropriate may be one of the most important characteristics of pathogenic bacteria.

Rationale and objectives of the current study

The current study was based on the finding in our laboratory of hemoglobin binding activity in *H. influenzae*. Using a dot blot assay with biotinylated hemoglobin as the primary ligand, it was shown that *H. influenzae* possessed a specific hemoglobin binding ability (27). The first objective of the current study was to identify a protein expressing hemoglobin binding activity and clone the gene encoding it. Using an affinity purification protocol, a 120 kDa putative hemoglobin binding protein was purified from outer membrane preparations *of H. influenzae* type b strain HI689. Based on the N-terminal amino acid sequence of the purified protein, two 20-mer oligonucleotides were designed for use as probes, and a 4.2 kb *EcoR*I fragment of HI689 genomic DNA encoding hemoglobin binding activity in *E. coli* was successfully cloned.

The second objective of the study was to characterize the gene encoding this hemoglobin binding activity. Nucleotide sequence of the 4.2 kb DNA fragment was determined. We found a putative ORF of 3.2 kb, designated *hgpA*, which would account for a protein of 120 kDa. Sequence analysis indicated that HgpA had significant homology with other iron or heme utilization proteins. The phenotype of a *hgpA* mutant indicated that there is more than one hemoglobin binding protein in HI689. Sequence analysis of *hgpA* led to the discovery of a family of putative hemoglobin binding proteins containing multiple CCAA repeats following the nucleotide sequence encoding the leader peptide. Accordingly, we proposed that these CCAA repeats are involved in gene regulation through a slip-strand mechanism.

The third objective of the research was to study the role of heme on the transcriptional and translational regulation of *hgpA*. Early experiments demonstrated that hemoglobin binding activity was repressed under heme replete conditions in *H. influenzae* (27). To accurately analyze the regulation of *hgpA* at the transcriptional level, we developed a Comparative Quantitative Reverse Transcriptase-Polymerase Chain Reaction (CQ-RT-PCR) methodology, and found that, *in vitro*, *hgpA* transcription was upregulated three-fold under heme deplete conditions compared to heme replete conditions. To further understand heme regulation on *hgpA* at the translational level, we assessed the β-galactosidase activity using a *hgpA-lacZ* protein fusion, and found that the HgpA fusion was produced three-fold more under heme deplete conditions as compared to heme replete conditions. Our results suggested that heme regulation of *hgpA* is at the transcriptional level, and there is a three fold up-regulation under heme restricted growth conditions.

In summary, the specific aims of this project were:

1. To affinity purify a hemoglobin binding protein. Results are presented in chapter II.
2. To characterize the hemoglobin binding protein gene. Results are presented in chapter III.
3. To study the transcriptional and translational regulation of the hemoglobin binding protein gene. Results are presented in chapter IV.

REFERENCES

1. **Anderson, P.W., M.E. Pichichero, and E.M. Connor**. 1985. Enhanced nasopharyngeal colonization of rats by piliated *Haemophilus influenzae* type b. Infect. Immun. **48**:565-568.

2. **Bezkorovainy, A.** 1981. Antimicrobial properties of iron-binding proteins. Adv.Exp.Med.Biol. **135**:139-154.

3. **Bramanti, T.E. and S.C. Holt**. 1993. Hemin uptake in Porphyromonas gingivalis:Omp26 is a hemin-binding surface protein. J.Bacteriol. **175**:7413-7420.

4. **Braun, V. and T. Focareta**. 1991. Pore-forming bacterial protein hemolysins (cytolysins). Crit. Rev. Microbiol. **18**:115-158.

5. **Brown, S.B. and R.F. King**. 1976. The structure of haem in pyridine/water mixtures and its implication in studies of haem catabolism. Biochem.J. **153**:479-483.

6. **Bullen, J.J.** 1987. Iron and the antibacterial function of polymorphonuclear lymphocytes, p.211-241. In J.J. Bullen and E. Griffith (ed.), Iron and infection, John Wiley & Sons, New York.

7. **Cartwright, G.E., A. Lauritsen, S. Humphreys, P.J. Jones, I.M. Merrill, and M.M. Wintrobe**. 1946. The anaemia of infection. II. The experimental production of hypoferremia and anemia in dogs. J.Clinic.Invest. **25**:81-86.

8. **CDC**. 1994. Reported vaccine-preventable diseases-United States, 1993, and the childhood immunization initiative. MMWR **43**:144-148.

9. **Chen, C.J., C. Elkins, and P.F. Sparling**. 1998. Phase Variation of Hemaglobin Utilization in *Neisseria gonorrhoeae*. Infect.Immun. **66**:987-993.

10. **Chen, Q. and J.H. Crosa**. 1996. Antisense RNA, Fur, iron and the regulation of iron transport gene in *Vibrio anguilllarium*. J.Biol.Chem. **271**:18885-18891.

11. **Cochi, S.L. and C.V. Broome**. 1986. Vaccine prevention of *Haemophilus influenzae* type b disease:past, present and future. Pediatr.Infect.Dis. **5**:12-19.

12. **Cope, L.D., S.E. Thomas, Z. Hrkal, and E.J. Hansen**. 1998. Binding of heme-hemopexin complexes by soluble HxuA protein allows utilization of this complexed heme by *Haemophilus influenzae*. Infect.Immun. **66**:4511-4516.

13. **Cope, L.D., S.E. Thomas, J.L. Latimer, C.A. Slaughter, U. Muller-Eberhard, and E.J. Hansen**. 1994. The 100 kDa haem:haemopexin-binding protein of *Haemophilus influenzae*: structure and localization. Mol.Microbiol. **13**:863-873.

14. **Cope, L.D., R. Yogev, U. Muller-Eberhard, and E.J. Hansen**. 1995. A gene cluster involved in the utilization of both free heme and heme:hemopexin by *Haemophilus influenzae* type b. J.Bacteriol. **177**:2644-2653.

15. **Coulton, J.W. and D.T. Wan**. 1983. The outer membrane of *Haemophilus influenzae* type b: cell envelope associations of major proteins. Can.J.Microbiol. **29**:280-287.

16. **Cox, C.D.** 1982. Iron uptake with ferric pyochetin and ferric citrate by *pseudomonas aeruginosa*. J.Bacteriol. **142**:581-587.

17. **Crisel, R.M., R.S. Baker, and D.E. Dorman**. 1975. Capsular polymer of *haemophilus influenzae* type b. I. Structural characterization of capsular polymer of strain Eagan. J.Biol.Chem. **250**:4926-4930.

18. **Crosa, J.H.** 1989. Genetics and molecular biology of siderophore-mediated iron transport in bacteria. Microbiol.Rev. **53**:517-530.

19. **Crosa, J.H.** 1997. Signal transduction and transcriptional and posttranscriptional control of iron-regulated genes in bacteria. Micro.Mol.Biol.Rev. **61**:319-336.

20. **Eaton, J.W., P. Brandt, J.R. Mahoney, and J.T. Lee Jr**. 1982. Haptoglobin: a natural bacteriostat. Science **215**:691-693.

21. **Elkins, C., C.J. Chen, and C.E. Thomas**. 1995. Characterization of the hgbA locus encoding a hemoglobin receptor from *Haemophilus ducreyi*. Infect.Immun. **63**:2194-2200.

22. **Eskola, J., H. Kayhty, H. Peltola, V. Karanko, P.H. Makela, J. Samuelson, and L.K. Gordon**. 1985. Antibody levels achieved in infants by course of *Haemophilus influenzae* type b polysaccharide/diphtheria toxoid conjugate vaccine. Lancet **k**:1184-1186.

23. **Evans, N.M., D.D. Smith, and A.J. Wicken**. 1974. Haemin and nicotinamide adenine dinucleotide requirements of *Haemophilus influenzae* and *Haemophilus parainfluenzae*. J.Med.Microbiol. 7:359-365.

24. **Farley, M.M., S.L. Stephenes, S.L. Kaplan, and E.O. Mason Jr.** 1990. Pilus- and non-pilus mediated interactions of *Haemophilus influenzae* type b with human erythrocytes and human nasopharyngeal mucos. J.Infect.Dis **161**:274-280.

25. **Farley, M.M., D.S. Stephens, M.H. Mulks, M.D. Cooper, J.V. Bricker, S.S. Mirra, and A. Wright**. 1986. Pathogenesis of IgA1 protease-producing and -nonproducing *Haemophilus influenzae* in human nasopharyngeal organ cultures. J.Infect.Dis. **154**:752-759.

26. **Foxwell, A.R., J.M. Kyd, and A.W. Cripps**. 1998. Nontypeable *Haemophilus influenzae* :pathogenesis and prevention. Micro.Mole.Biol.Rev. **62**:294-308.

27. **Frangipane, M.E., D.J. Morton, J.A. Wooten, J.M. Pozsgay, and T.L. Stull**. 1994. Binding of human hemoglobin by *Haemophilus influenzae*. FEMS.Microbiol.Lett **118**:243-248.

28. **Fraser, D.W.** 1981. *Haemophilus influenzae*: Epidemiology, Immunology, and Prevention of Disease, S.H. Sell and P.F. Wright (ed.), Elsevier Biomedical, New York.

29. **Giebink, G.S.** 1989. The microbiology of otitis media. Pediatr.Infect.Dis. **8**:S18-S20

30. **Gilsdorf, J.R., K.W. McCrea, and C.F. Marrs**. 1997. Role of pili in *Haemophilus influenzae* adherence and colonization. Infect.Immun. **65**:2997-3002.

31. **Gilsdorf, J.R., M. Tucci, L.J. Forney, W. Waston, C.F. Marrs, and E.J. Hansen**. 1993. Paradoxical effect of pilus expression on binding of antibodies by *Haemophilus influenzae*. Infect.Immun. **61**:3375-3381.

32. **Griffiths, E.** 1987. Iron in Biological Systems, p.1-26. In J.J. Bullen and E. Griffiths (ed.), Iron and Infection: Molecular, Physiological and Clinical Aspects, John Wiley & Sons, New York.

33. **Griffiths, E.** 1991. Iron and bacterial virulence-a brief overview. Biol.Metals **4**:7-13.

34. **Gutierrez, M.K., L.S. Joffe, L.J. Forney, and M.P. Glode**. 1990. Effect of pili on the frequency of bacteriumia following challenge with *Haemophilus influenzae* type b (HIB) in the infant rat, abstr.41. In Program and abstracts of the 30th Intersciences Conferrence on Antimicrobial Agents and Chemotherapy. American Society for Microbiology, Washington,D.C. ASM abstrct 93

35. **Hancock, R.E., K. Hantke, and V. Braun**. 1976. Iron trnasport of *Escherichia coli* K-12:involvement of the colicin B reeptor and of a citrate-inducible protein. J.Bacteriol. **127**:1370-1375.

36. **Hanson, M.S., C. Slaughter, and E.J. Hansen**. 1992. The *hbpA* gene of *Haemophilus influenzae* type b encodes a heme- binding lipoprotein conserved among heme-dependent *Haemophilus* species. Infect.Immun. **60**:2257-2266.

37. **Hantke, K.** 1984. Cloning of the repressor protein gene of iron-regulated systems in *Escherichia coli*. Mol.Gen.Genet. **197**:288-292.

38. **Heller, K., R.J. Kadner, and K. Gunther**. 1988. Suppression of the btuB451 mutation by mutatios in the *tonB* gene suggests a direct interaction between TonB and TonB-dependent receptor proteins in the outer membrane of *Escherichia coli*. Gene **64**:147-153.

39. **Henderson, D.P. and S.M. Payne**. 1993. Cloning and characterization of the *Vibrio cholerae* genes encoding the utilization of iron from haemin and haemoglobin. Mol.Microbiol. **7**:461-469.

40. **Herrington, D.A. and P.F. Sparling**. 1985. *Haemophilus influenzae* can use human transferrin as a sole source for required iron. Infect.Immun. **48**:248-251.

41. **High, N.J., M.P. Jennings, and E.R. Moxon**. 1996. Tandem repeats of the tetramer 5'-CAAT-3' present in *lic2A* are required for phase variation but not lipopolysaccharide biosynthesis in *Haemophilus influenzae*. Mol.Microbiol. **20**:165-174.

42. **Jarosik, G.P., J.D. Sanders, L.D. Cope, U. Muller-Eberhard, and E.J. Hansen**. 1994. A functional tonB gene is required for both utilization of heme and virulence expression by *Haemophilus influenzae* type b. Infect.Immun. **62**:2470-2477.

43. **Jin, H., Z. Ren, J.M. Pozsgay, C. Elkins, P.W. Whitby, D.J. Morton, and T.L. Stull**. 1996. Cloning of a DNA fragment encoding a heme-repressible

hemoglobin-binding outer membrane protein from *Haemophilus influenzae*. Infect.Immun **64**:3134-3141.

44. **Kadner, R.J.** 1990. Vitamine B12 transport in *Escherichia* coli: energy coupling between membranes. Mol.Microbiol. **4**:2027-2033.

45. **Kaplan, R.J.** 1976. Neurological complications of the head and neck. Otolaryngol.Clin.North.Am. **9**:729-749.

46. **Kilian, M., J. Mestecky, and R.E. Schrohenloher**. 1979. Pathogenic specieses of the genus *Hamophilus* and *Streptococcus pneumonia* produce immunoglobulin A1 protease. Infect.Immun. **26**:143-149.

47. **Kimura, A. and E.J. Hansen**. 1986. Antigenic and phenotypic variations of *Haemophilus influenzae* type b lipopolysaccharide and their relationship to virulence. Infect.Immun. **51**:69-79.

48. **Kino, K., H. Tsunoo, Y. Higa, M. Takami, H. Hamaguchi, and H. Nakajima**. 1980. Hemoglobin-haptoglobin receptor in rat liver plasma membrane. J.Biol.Chem. **255**:9616-9620.

49. **Klebba, P.E., J.M. Rutz, J. Liu, and C.K. Murphy**. 1993. Mechanisms of TonB-catalyzed iron transport through the enteric bacterial cell envelope. J.Bioenerg.Biomemb. **25**:603-611.

50. **Larsen, R.A., D. Foster-Hartnett, M.A. McIntosh, and K. Postle**. 1997. Regios of *Escherichia coli* TonB and FepA proteins essential for *in vivo* physical interactions. J.Bacteriol. **179**:3213-3221.

51. **Larsen, R.A., P.S. Myers, J.T. Skare, C.L. Seachord, R.P. Darveau, and K. Postle**. 1996. Identification of TonB homologs in the family Enterobacteriaceae and evidence for conservation of TonB-dependent energy transduction complexes. J.Bacteriol. **178** :1363-1373.

52. **Lee, B.C.** 1995. Quelling the red menace: haem capture by bacteria. Mol.Microbiol. **18**:383-390.

53. **Lee, B.C.** 1992. Isolation of haemin-binding proteins of *Neisseria gonorrhoeae*. J.Med.Microbiol. **36**:121-127.

54. **Lee, B.C.** 1994. Isolation and characterization of the haemin-binding proteins from *Neisseria meningitidis*. Microbiol. **140**:1473-1480.

55. **Lee, B.C. and P. Hill**. 1992. Identification of an outer-membrane haemoglobin-binding protein in *Neisseria meningitidis*. J.Gen.Microbiol. **138**:2647-2656.

56. **Lewis, L.A., E. Gray, Y.P. Wang, B.A. Roe, and D.W. Dyer**. 1997. Molecular characterization of *hpuAB*, the haemoglobin-haptoglobin-utilization operon of *Neisseria meningitidis*. Mol.Microbiol. **23**:737-749.

57. **Lorenzo, V.D., M. Herrero, F. Giovannini, and J.B. Neilands**. 1988. Fur (ferric uptake regulation) protein and CAP (catabolite-activator protein) modulate transcription of *fur* gene in *Escherichia coli*. Eur.J.Biochem. **173**:537-546.

58. **Maciver, I., J.L. Latimer, H.H. Liem, U. Muller-Eberhard, Z. Hrkal, and E.J. Hansen**. 1996. Identification of an outer membrane protein involved in utilization of hemoglobin-haptoglobin complexes by nontypeable *Haemophilus influenzae*. Infect.Immun. **64**:3703-3712.

59. **Maskell, D.J., M.J. Szabo, P.D. Butler, A.E. Williams, and E.R. Moxon**. 1991. Molecular analysis of a complex locus from *Haemophilus influenzae* involved in phase-variable lipopolysaccharide biosynthesis. Mol.Microbiol. **5**:1013-1022.

60. **Maskell, D.J., M.J. Szabo, P.D. Butler, A.E. Williams, and E.R. Moxon**. 1992. Molecular biology of phase-variable lipopolysaccharide biosynthesis by *Haemophilus influenzae*. J.Infect.Dis. **165 Suppl 1**:S90-2.

61. **Moeck, G.S., J.W. Coulton, and K. Postle**. 1997. Cell envelope signaling in *Escherichia coli*. Ligand binding to the ferrichrome-iron receptor fhua promotes interaction with the energy-transducing protein TonB. J.Biol.Chem. **272**:28391-28397.

62. **Morgan, E.H.** 1981. Inhibition of reticulocyte iron uptake by NH_4Cl and CH_3NH_2. Biochim.Biophys.Acta **642**:119-134.

63. **Morton, D.J., J.M. Musser, and T.L. Stull**. 1993. Expression of the *Haemophilus influenzae* transferrin receptor is repressible by hemin but not elemental iron alone. Infect.Immun. **61**:4033-4037.

64. **Moxon, E.R.** 1986. The carrier state: *Haemophilus influenzae*. J.Antimicrob.Chemother. **18**:1-16.

65. **Moxon, E.R. and K.A. Vaughan**. 1981. Type b capsular polysaccharide as a virulence determinant of *Haemophilus influenzae*: studies using clinical isolates and laboratory transformants. J.Infect.Dis. **143**:517-524.

66. **Mulkes, M.H., S.J. Kornfeld, B. Frangione, and A.G. Plaut.** 1982. Relationship between the specificity of IgA proteases and serotypes in *Haemophilus influenzae*. J.Infect.Dis. **146**:266-274.

67. **Mulks, M.H., S.J. Kornfeld, and A.G. Plavt.** 1980. Specific proteolysis of human IgA by *Streptococcus pneumonia* and *Haemophilus influenzae.* J.Infect.Dis. **141**:450-455.

68. **Muller-Eberhard, U. and H. Nikkia** . 1989. Transport of tetrapyrroles by proteins. Sem.Hematol. **26**:86-104.

69. **Murphy, T.F., J.M. Bernstein, D.M. Dryja, A.A. Campagnari, and M.A. Apicella**. 1987. Outer membrane protein and lipooligosaccharide analysis of paired nasopharyngeal and middle ear isolates in otitis media due to nontypable *Haemophilus influenzae*: pathogenetic and epidemiological observations. J.Infect.Dis.**156**:723-731.

70. **Murphy, T.V., K.E. White, and P. Pastor**. 1993. Declining incidence of *Haemophilus influenzae* type b disease since introduction of vaccination. JAMA **269**:246-248.

71. **Nikaido, H. and M. Vaara**. 1985. Molecular basis of bacterial outer membrane permeability. Microbiol.Rev. **49**:1-32.

72. **Otto, B.R., A.M. Verweij-van Vught, and D.M. MacLaren**. 1992. Transferrins and heme-compounds as iron sources for pathogenic bacteria. Crit.Rev.Microbiol. **18**:217-233.

73. **Payne, S.M.** 1988. Iron and virulence in the family *Enterobacteriaceae*. Crit.Rev.Microbiol. **16**:81-111.

74. **Pfeiffer, R.** 1892. Die aetiologie der Influenzae. Zeitscher F.Hyg. **13**:357-386.

75. **Pidcock, K.A. and T.L. Stull**. 1988. Tn5 transposon mutagenesis of *Pseudomonas cepacia* by an IncW suicide vector. J.Infect.Dis. **157**:1098-1099.

76. **Plaut, A.G., R. Wistar Jr, and J.D. Capra**. 1974. Differential susceptibility of human IgA immunoglobulins to streptococcal IgA protease. J.Clin.Invest. **54**:1295-1300.

77. **Postle, K.** 1993. TonB protein and energy transduction between membranes. J.Bioenerg.Biomemb. **25**:591-601.

78. **Reidl, J. and J.J. Mekalanos**. 1996. Lipoprotein *e*(P4) is essential for hemin uptake by *Haemophilus influenzae*. J.Exp.med. **183**:621-629.

79. **Ren, Z., H. Jin, D.J. Morton, and T.L. Stull**. 1997. Cloning of *hgpB*, a gene encoding a second hemoglobin binding protein of *Haemophilus influenzae*. Abstracts of the 97th General Meeting of the American Society for Microbiology

80. **Rivers, T.M.** 1922. Influenza-like Bacilli. Growth of influenza-like bacilli on media containing only an autoclave-labile substance as an accessory food factor. Bull.Johns.Hopkins.Hosp **33**:429-431.

81. **Saliase, P., L. Waldbeser, and J.H. Crosa**. 1993. Regulation of the expression of bacterial iron transport gene: possible role of an antisense RNA as a repressor. Gene **123**:33-38.

82. **Sauer, M., K. Hantke, and V. Braun**. 1987. Ferric-coprogen receptor FhuE of *Escherichia coli*: processing and sequence common to all TonB-dependent outer membrane receptor proteins. J.Bacteriol. **169**:2044-2049.

83. **Schoendorf, K.C., W.G. Adams, J.L. Kiely, and J.D. Wenger**. 1994. National Trends in *Haemophilus influenzae* Meningitis Mortality and Hospitalization Among Children, 1980 through 1991. Pediatr. **93**:663-668.

84. **Smith, A.** 1990. transport of tetrapyrroles: mechanisms and biological and regulatory consequences, p.435-489. In H.A. Dailey (ed.), Biosythesis and metabolism of heme and chlorophyll, McGraw Hill Publishing Co., New York.

85. **Smith, A., N.I. Hooper, N. Shipulina, and W.T. Morgan**. 1996. Heme binding by a bacterial repressor protein, the gene product of the ferric uptake regulation (*fur*) gene of *Escherichia coli*. J.Prot.Chem. **15**:575-583.

86. **Stojiljkovic, I. and K. Hantke**. 1992. Hemin uptake system of *Yersinia enterocolitica*: Similarties with other TonB-dependent system in gram-negative bacteria. EMBO J. **11**:4359-4367.

87. **Stojiljkovic, I., V. Hwa, L. de Saint Martin, P. O'Gaora, X. Nassif, Heffron, F, and M. So**. 1995. The *Neisseria meningitidis* haemoglobin receptor: its role in iron utilization and virulence. Mol.Microbiol. **15**:531-541.

88. **Stojiljkovic, I. and N. Srinivasan**. 1997. N*eisseria meningitidis tonB, exbB*, and *exbD* genes: Ton-dependent utilization of protein-bound iron in Neisseriae. J.Bacteriol. **179**:805-812.

89. **Strausbaugh, L.J.** 1997. *Haemophilus influenzae* infection in adults. Postgrad.Med. **101**:191-200.

90. **Stull, T.L.** 1987. Protein sources of heme for *Haemophilus influenzae*. Infect.Immun. **55**:148-153.

91. **Syrogiannopoulos, G.A., E.J. Hansen, A.L. Erwin, R.S. Munford, J. Rutledge, J.S. Reisch, and G.H. McCracken, Jr.** 1988. *Haemophilus influenzae* type b lipooligosaccharide induces meningeal inflammation. J.Infect.Dis. **157**:237-244.

92. **Szabo, M., D. Maskell, P. Butler, J. Love, and R. Moxon**. 1992. Use of chromosomal gene fusions to investigate the role of repetitive DNA in regulation of genes involved in lipopolysaccharide biosynthesis in *Haemophilus influenzae*. J.Bacteriol. **174**:7245-7252.

93. **Tosi, M.F., D.C. Anderson, J. Barrish, E.O. Mason Jr., and S.L. Kaplan**. 1985. Effect of pilation on interactions of *Haemophilus influenzae* type b with human polymophonuclear leukocytes. Infect.Immun. **47**:780-785.

94. **Turk, D.C.** 1984. The pathogenicity of *Haemophilus influenzae*. J.Med.Microbiol. **18**:1-16.

95. **Vincent, S.H.** 1989. Oxidative effects of heme and porphyrins on proteins and lipids. Sem.Hematol. **26**:105-113.

96. **Waldbeser, L., Q. Chen, and J.H. Crosa**. 1995. Antisense RNA regulation of the *fatB* iron transport protein gene in *Vibrio anguillarum*. Mol.Microbiol. **17**:747-756.

97. **Wallace, R.J., D.M. Musher, and R.J. Martin**. 1978. *Haemophilus influenzae* pneumonia in adults. Am.J.Med. **64**:87-93.

98. **Weber, A.K., S. Harris, S. Lohrke, L. Forney, and A.L. Smith**. 1991. Inability to express fimbriae results in impaired ability of *Haemophilus influenzae* b to colonizae the nasopharynex. Infect.Immun. **59**:4724-4728.

99. **Weinberg, E.D.** 1984. Iron withholding: a defense against infection and neoplasia. Physiol.Rev. **64**:65-102.

100. **White, D.C. and S. Granick**. 1963. Hemin biosynthesis in *Haemophilus*. J.Bacteriol. **85**:842-850.

101. **Winslow, C.E.A., J. Broadhurst, R.E. Buchanan, C. Krumwiede Jr, L.A. Rogers, and H.M. Smith**. 1920. The families and genera of the bacteria. J.Bacteriol. **5**:191-229.

102. **Wong, J.C., J. Holland, T. Parsons, A. Smith, and P. Williams**. 1994. Identification and characterization of an iron-regulated hemopexin receptor in *Haemophilus influenzae* type b. Infect.Immun. **62**:48-59.

CHAPTER II

CLONING OF A DNA FRAGMENT ENCODING A HEME REPRESSIBLE HEMOGLOBIN-BINDING OUTER MEMBRANE PROTEIN FROM *HAEMOPHILUS INFLUENZAE*

ABSTRACT

Haemophilus influenzae is able to use hemoglobin as a sole source of heme, and heme-repressible hemoglobin-binding to the cell surface has been demonstrated. Using an affinity purification methodology, a hemoglobin-binding protein (Hgp) of approximately 120 kDa was isolated from *H. influenzae* type b strain HI689 grown in heme-restricted but not in heme-replete conditions. The isolated protein was subjected to N-terminal amino acid sequencing, and the derived amino acid sequence was used to design corresponding oligonucleotides. The oligonucleotides were used to probe a Southern blot of *Eco*RI-digested HI689 genomic DNA. A hybridizing band of approximately 4.2 kb was successfully cloned into pUC19. Following partial nucleotide sequencing of the 4.2-kb insert a putative ORF was subcloned into an expression vector. The host *Escherichia coli* strain in which the cloned fragment was expressed bound biotinylated human hemoglobin, whereas binding of hemoglobin was not detected in *E.*

coli with the vector alone. In conclusion we hypothesize that the DNA fragment encoding the approximately 120-kDa heme-repressible Hgp mediates one step in the acquisition of hemoglobin by *H. influenzae in vivo*.

INTRODUCTION

Haemophilus influenzae is responsible for many human infections including otitis media, meningitis, epiglottitis and pneumonia (52). The incidence of invasive disease caused by strains with the type b capsule has been radically reduced following the introduction of vaccines based on the type b capsular polysaccharide (32,37,38,43). However vaccine failures occur and, although adults presumably have natural immunity, two thirds of *H. influenzae* blood isolates from adults are type b (10,24,46). Since currently available vaccines are based on the type b capsule, they provide no protection 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 (12,51,53,54). Viable vaccine candidates to provide protection against all *H. influenzae* disease would include surface exposed proteins which are widely distributed across the species and expressed during disease.

H. influenzae has an absolute growth requirement for an exogenous source of protoporphyrin IX (PPIX) which is the immediate precursor of heme (11). *In vivo* all heme is intracellular, in the form of hemoglobin or heme-containing enzymes, and thus

unavailable to invading microorganisms (3,18,26). Hemoglobin released by erythrocytes is avidly bound by the serum protein haptoglobin, and the hemoglobin-haptoglobin complex is rapidly cleared by hepatocytes (3,40). Free heme, principally derived from the degradation of methemoglobin, is bound by the serum proteins hemopexin and albumin and cleared from the circulatory system by hepatocytes (3). Hemoglobin, the hemoglobin-haptoglobin, heme-hemopexin, and heme-albumin complexes can all be utilized by *H. influenzae* as heme sources (48). We have recently shown that *H. influenzae* binds hemoglobin directly at the cell surface, possibly as an initial step in the utilization of hemoglobin-associated heme (14). We have also demonstrated that both hemoglobin binding and the binding of transferrin, an iron-binding glycoprotein which *H. influenzae* uses as an iron source, are repressible by heme but not by elemental iron alone (14,34). In addition, we have noted heme-repressible hemolytic activity expressed by *H. influenzae* (unreported observation), which may represent an important first step in acquiring hemoglobin *in vivo*.

The objectives of the current study were to clone and to characterize the gene(s) encoding the protein(s) mediating the binding of hemoglobin by *H. influenzae*.

MATERIALS AND METHODS

Bacterial strains and growth conditions. *H. influenzae* type b and Rd strains (Table 1) were kindly provided by J. Musser, Baylor University, Houston, TX and have

been described previously (34,39). Strains of *Haemophilus* were routinely maintained on brain heart infusion (BHI) agar (Difco, Detroit, MI) supplemented with 10 μg/ml of both hemin and β-NAD. For long term storage strains were stored at -70°C in skim milk. For experiments in heme-replete medium, *H. influenzae* was grown at 37°C in BHI broth (Difco) supplemented with 10 μg/ml β-NAD and 10 μg/ml heme (supplemented BHI; sBHI). Heme-restricted growth of *H. influenzae* was performed in BHI supplemented with 10 μg/ml β-NAD and 0.1 μg/ml hemin (hemin-restricted BHI; hrBHI). Table 1 also shows the *E. coli* strains which were maintained on Luria-Bertani (LB) medium supplemented with antibiotics, as appropriate, at the concentrations indicated in Table 1. Plasmids used in these studies and their relevant properties are listed in Table 1.

Biotinylation of hemoglobin. Human hemoglobin (Sigma, St. Louis, MO) was biotinylated as described previously (14). Hemoglobin (1 mg/ml) was dissolved in phosphate-buffered saline (PBS), pH 7.4, and NHS-LC biotin (Pierce, Rockford, IL) was dissolved to a concentration of 1 mg/ml in water at 50°C. To 5 ml of the hemoglobin solution was added 430 μl of the biotin solution and, following incubation for 2 hr at room temperature, unbound biotin was removed by passage through a Sephadex G-15 column (Pharmacia, Piscataway, NJ).

Preparation of outer membrane proteins. Outer membrane proteins were isolated by selective solubilization with Triton X-100 essentially as previously described (49). *H. influenzae* HI689 were grown to late logarithmic phase in sBHI or hrBHI and harvested by centrifugation at 6,000 x g for 10 min. The cell pellets were resuspended in

Table 1. Strains and plasmids

Strain or plasmids	Relevant characteristic(s)[a]	Reference or resource
Strains		
H. influenzae		
689	Type b, ET22	39
Rd	Capsule-deficient type d	ATCC
E. coli		
DH5α	Δ(*lacZYA-argF*)U169	BRL[b]
BL21(DE3)pLysS	F^-, *ompT*, (DE3)pLysS, Cm^r	Novagen
Plasmids		
pRESTA	Amp^r, carrying T7 promoter, metal binding domiain polylinker, F1 origion	Invitrogen
pHFJ1, pHFJ2	pUC19, carrying a 4.2 kbp *Eco*RI frgment from *H. influenzae*	This work
pXHGP	pRESETA, carrying 3.2 kbp PCR product derived from pHFJ2	This work

[a]ET, multilocus enzyme electrophoretic type; F^-, F episome negative; Amp^r, ampicillian resistance (50 μg/ml); Cm^r, chloraphenicol resistance (50 μg/ml).
[b]BRL, Bethesda Research Laboratories.

distilled water to an A_{605} of 0.6 and 10 ml of the suspension were sonicated (Heat Systems, Farmingdale, NY; Model CL4 with microtip, set to microtip limit). Sonication was carried out in an ice/NaCl bath for a total time of 2 min in 10 sec bursts with 50 sec. between bursts. Triton X-100 was added to the sonicate at a final concentration of 2 % v/v, and following incubation for 5 min on ice the mixture was sonicated as above or until the solution was clear. The sonicate was centrifuged at 35,000 x g for 1hr at 4°C and the pellet was resuspended in 2 ml of 10 mM Tris-HCl (pH 7.4).

Affinity chromatography purification of *H. influenzae* hemoglobin-binding protein. Resuspended outer membranes, as described above, were divided equally into two microcentrifuge tubes and 100 µl of biotinylated hemoglobin (1 mg/ml) were added. Following incubation for 1 h with gentle agitation, mixtures were centrifuged at 13,000 x g for 10 min. The pellet was resuspended in 1 ml of buffer I (0.75 % w/v sarkosyl, 100 mM NaCl, 100 mM EDTA, 50 mM Tris-HCl, [pH 8.0]). The resuspended pellet was incubated for 1 h with 100 µl of streptavidin-agarose (Sigma, St. Louis, MO; diluted 1:1 in sterile distilled water). Samples were centrifuged at 1,000 x g for 30 sec and the supernatant was carefully removed. One ml of buffer II (0.5% w/v sarkosyl, 100 mM NaCl, 100 mM EDTA, 50 mM Tris–HCl, [pH 8.0]) was added to the beads and following incubation for 5 min with gentle agitation the mixture was centrifuged at 1,000 x g and the supernatant removed. The beads were washed twice more in buffer II and once in buffer III (100 mM NaCl, 50 mM Tris-HCl, [pH 8.0]) as described above. All incubations and washes were performed at room temperature. Following the final wash the beads were resuspended in 20 µl of SDS-PAGE sample buffer (50 mM Tris-HCl [pH 6.8], 2 % v/v 2-mercaptoethanol, 10 % v/v glycerol, 1 % w/v SDS, 0.01 % w/v Bromophenol Blue) and heated in a 100 °C water bath for 5 min to elute bound proteins. Eluted proteins were separated by SDS-PAGE on 6.5 % acrylamide gels using the discontinuous buffer system of Laemmli (25). Approximately 30 µl of protein preparation were loaded per lane.

N-terminal amino acid sequencing. Affinity chromatography purified proteins from heme-restricted *H. influenzae* HI689 were separated by SDS-PAGE on 7.5%

acrylamide gels and transferred to membranes for N-terminal amino acid sequencing as described by Moos (33). Thioglycolic acid (sodium salt) 0.066 % w/v was added to the upper buffer reservoir during SDS-PAGE to remove the by-products of acrylamide polymerization (33). Proteins were transferred in 10 mM 3-cyclohexylamino-1-propanesulfonic acid (CAPS) (pH 10.5), 1% v/v methanol, 0.05 % w/v dithiothreitol, to a polyvinyldene difluoride (PVDF) membrane (Millipore, Bedford, MA) and visualized by staining with 0.1% Coomassie blue in 50% v/v methanol. The entire membrane was submitted to the UCLA Medical School Protein Microsequencing Facility where the N-terminal amino acid sequence of the 120-kDa protein was determined.

DNA isolation. Bacterial genomic DNA was isolated by standard techniques as previously described (44), or using the DNA Now reagent (Biogentex, Seabrook, TX) as directed by the manufacturer. Plasmid DNA was isolated by the use of Qiagen plasmid kits (Qiagen, Chatsworth, CA) as directed by the manufacturer. DNA concentrations were assessed spectrophotometrically using a Shimadzu UV-1201S Spectrophotometer with DNA/Protein proGram pack (Shimadzu, Kyoto, Japan).

Southern blot and DNA hybridization. DNA was digested with restriction enzymes as directed by the manufacturers, separated on agarose gels (0.8 % w/v agarose) in TBE buffer (0.045 M Tris-borate, 0.001 M EDTA) and transferred to Magnagraph nylon membranes (MSI, Westbrook, MA) by the method of Southern as described by Sambrook *et al.* (44). For screening of partial *H. influenzae* libraries in *E. coli*, bacterial colonies were

transferred to a Magnagraph nylon membrane and prepared for hybridization as described by Sambrook *et al.* (44).

The ECL 3'-oligolabeling system (Amersham Life Science, Arlington Heights, IL) was used as directed by the manufacturer to label the 3' end of the oligonucleotides. The ECL random prime labeling kit (Amersham) was used as directed by the manufacturer to label DNA probes.

Labeled oligonucleotides or DNA were used to probe Southern blots or colony blots. For oligonucleotide probes hybridization was performed at 44 °C followed by stringency washes as follows: 5 X SSC, 0.1% w/v SDS twice for 5 min at room temperature and 1 X SSC, 0.1% w/v SDS twice for 15 min at 44 °C. For DNA probes the hybridization temperature was 60 °C and stringency washes were 1 X SSC, 0.1% w/v SDS for 15 min and 0.5 X SSC, 0.1% w/v SDS for 15 min both at 60 °C (1 X SSC: NaCl 8.8 g/l, sodium citrate 4.4 g/l [pH 7.0]). Hybridization was detected using ECL nucleic acid detection reagents (Amersham) as directed by the manufacturer. Blots were subsequently exposed to X-ray film (Fuji Photo Film Co., Japan).

Cloning of the Gene Encoding Hgp. *H. influenzae* HI689 chromosomal DNA was digested with *Eco*RI and separated on a 0.8 % w/v agarose gel. Bands in the region 4-6 kb were excised and purified from the gel using the Prep-A-Gene DNA purification kit (Bio-Rad, Hercules, CA) as directed by the manufacturer. The isolated DNA fragments were ligated using standard techniques to dephosphorylated *Eco*RI digested pUC19 (44). The ligation mixture was transformed into competent *E. coli* DH5α, and transformants were

selected on LB plates containing 50 μg/ml ampicillin. Ampicillin resistant colonies were subcultured on selective media and colonies were probed with the N-terminal oligonucleotides to identify positive clones. Putative positive plasmids were isolated and confirmed by Southern blot and hybridization with the oligonucleotide probes.

Partial automated sequencing (ABI model 373A, Recombinant DNA/Protein Resource Facility, Oklahoma State University, Stillwater, OK) of a positive clone allowed us to identify a putative leader sequence and start codon and also a presumed C-terminal sequence showing significant homology with other bacterial iron and heme related proteins at the amino acid level. The length of DNA flanked by the proposed start codon and C-terminal sequence would account for a protein of approximately 120 kDa. Primers were designed for use in PCR to allow for cloning of the putative coding region into a controlled expression vector. Primers were synthesized (Molecular Biology Resource Facility, University of Oklahoma Health Sciences Center, Oklahoma City, OK) with the sequences of primer pRSETA1 5'-GACCAGGGATCCATGACCAATTTTAG-3', and primer pRSETA2 5'-GGAAGGGGTACCCTAGAATTCAAACTG-3'. A *Bam*HI site upstream of the start codon, and a *Kpn*I site downstream of the stop codon were included in the primers to allow for directional cloning of the PCR product in the vector pRSETA in the correct reading frame. PCR was performed in 50 μl reactions using 30 ng of pHFJ2, linearized by digestion with *Hin*dIII, as template. PCR reactions contained 2 mM $MgCl_2$, 200 μM of each dNTP, 100 ng of each primer and 2 units of Taq DNA polymerase. PCR was carried out for 30 cycles with each cycle consisting of denaturation at 95 °C for 1 min, annealing at 55 °C

for 1 min, and primer extension at 72 °C for 3.5 min, with a final extension time of 10 min. Amplicons of the expected size (3.2 kb) were gel purified and mapped by digestion with *Bgl*II. The PCR products were digested with *Bam*HI and *Kpn*I, and ligated to gel purified *Bam*HI and *Kpn*I digested pRSETA. The ligation mixture was transformed into *E. coli* BL21(DE3)pLysS and recombinants were selected on LB agar containing 50 µg/ml ampicillin. Plasmids were isolated from ampicillin resistant colonies and mapped by restriction enzyme digestion to identify clones containing the expected product. A positive clone was identified and designated pXHGP.

Expression of Hgp in *E. coli*. A hemoglobin-binding dot-blot assay was used to determine whether *E. coli* containing pXHGP bound hemoglobin. In a dot-blot manifold (Bio-Rad) 100 µl of cell suspension (10^7 CFU) were filtered onto nitrocellulose membranes. Membranes were air-dried for 15 min and then incubated in 10 % w/v skim milk in PBS for 1 h. Membranes were incubated in 500 ng/ml biotinylated human hemoglobin in PBS for 1 h, and then washed three times for 10 min each in PBS. A second blocking step in 10 % w/v skim milk in PBS was performed for 1 h, followed by incubation in 25 ng/ml streptavidin-horseradish peroxidase conjugate (Jackson Immunoresearch, West Grove, PA) in PBS for 1 h. Membranes were washed three times for 10 min each in PBS and developed with ECL western blot detection reagents (Amersham) as directed by the manufacturer. All steps were performed at room temperature.

In some experiments IPTG-induced bacteria were harvested, resuspended to 10^8 cfu/ml and 5 ml samples were sonicated (Heat Systems, Farmingdale, NY; Model CL4 with

microtip, set to microtip limit). Sonication was carried out in an ice/NaCl bath for a total time of 2 min in 10 sec bursts with 50 sec. between bursts. Sonicates (100 µl) were filtered onto nitrocellulose membranes in a dot blot manifold and the membranes were probed and developed as above.

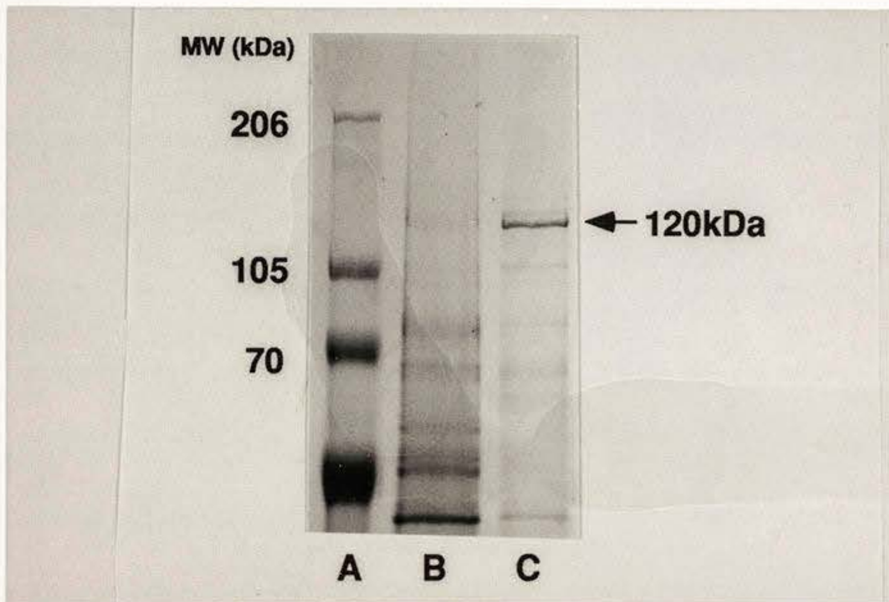
Western Immunoblot Analysis. The 100-kDa hemoglobin-binding protein of *H. ducreyi* 35000 was affinity purified as previously described (9). Recombinant protein from IPTG induced BL21(DE3)pLysS harboring pXHGP was purified using the Xpress protein purification system (Invitrogen) as directed by the manufacturer. These proteins and the affinity purified 120-kDa protein from *H. influenzae* HI689 were separated by SDS-PAGE on 7.5 % acrylamide gels and electrophoretically transferred to nitrocellulose membranes. Membranes were blocked in 1 % w/v skim milk in PBS for one hour and then probed with a 1:1,000 dilution in PBS containing 1 % w/v skim milk of an antibody raised against the 100-kDa HgbA of *H. ducreyi* in rabbits (9). Following four washes in PBS for 15 min each membranes were probed with 1:10,000 donkey anti-rabbit Ig-horseradish peroxidase conjugate (Amersham) in PBS with 1 % w/v skim milk. Membranes were washed 4 times for 15 min each in PBS and developed using ECL detection reagents as directed by the manufacturer.

RESULTS

Affinity purification of a hemoglobin-binding protein (Hgp). *H. influenzae* is heme-dependent and utilizes hemoglobin as a heme source *in vitro* (48). Using a whole cell dot blot assay we have previously shown that hemoglobin binds to the *H. influenzae* cell surface in a manner indicative of a specific receptor, and demonstrated that binding is induced by limiting heme levels in the growth media (14). To investigate further the possibility that hemoglobin-binding to *H. influenzae* is mediated through an outer membrane protein regulated by heme levels we subjected resuspended outer membrane proteins of strain HI689 to affinity purification using biotinylated hemoglobin. A 120-kDa protein was isolated from outer membranes derived from HI689 grown under heme-restricted (hrBHI) (Fig. 2, lane C) but not from those grown under heme replete conditions (sBHI) (Fig. 2, lane B).

N-terminal amino acid sequencing of Hgp. An N-terminal amino acid sequence of Hgp was obtained from affinity purified outer membranes of *H. influenzae* HI689. The determined N-terminal amino acid sequence was AQPTNQPTNQ. The synthetic oligonucleotides 5'-CAACCAACTAATCAACCAAC-3' and 5'-CAGCCTACAAATCAACCAAC-3', synthesized by Ransom Hill Bioscience, Inc., Ramona, CA, were designed on the basis of the N-terminal amino acid sequence of the 120-kDa Hgp after consulting the *H. influenzae* codon preference data of Gilsdorf *et al.* (15).

Figure 2. Identification of a hemoglobin-binding protein in *H. influenzae*. SDS-PAGE (6.5 % acrylamide) gel stained with Coomassie blue. Lane A, molecular weight marker; lane B, affinity purified outer membrane protein from *H. influenzae* type b grown in sBHI; lane C, affinity purified outer membrane protein from *H. influenzae* type b grown in hrBHI. Numbers represent molecular weights.



These two oligonucleotides were used in mixture to probe Southern blots of *H. influenzae* strain HI689 genomic DNA digests (Fig. 3). The oligonucleotides hybridized to an approximately 4.2-kb *Eco*RI fragment (Fig. 3, lane B) which was deemed of appropriate size for cloning. No hybridization of the oligonucleotides to *E. coli* chromosomal DNA (data not shown) was detected, suggesting that the oligonucleotides could be used to identify recombinant clones containing the structural gene encoding the120-kDa Hgp.

Cloning of the DNA fragment encoding Hgp. A limited genomic library of *H. influenzae* strain HI689 was constructed in the vector pUC19 as described in Materials and Methods. Two recombinant clones were isolated from this library by using the N-terminal amino acid sequence-derived oligonucleotides as probes. Both clones contained inserts of approximately 4.2 kb which hybridized in Southern analyses with the oligonucleotide probes (Fig. 4, lanes D & F). No hybridization of the oligonucleotides to pUC19 was detected (Fig. 4, lane C). These plasmids were designated pHFJ1 and pHFJ2. Partial mapping of the two clones revealed them to contain the same fragment in reverse orientation (Fig. 5). From preliminary sequencing analysis the N-terminal nucleotide sequence and a putative C-terminal sequence flanking a region of the correct size to encode a 120-kDa protein were identified. The N-terminal and C-terminal regions showed significant homology at the amino acid level with other bacterial iron and heme related proteins (Figs. 6 & 7), and a predicted leader peptide of 23 amino acids was identified (Fig. 6). Based on the sequence data, primers were designed to amplify the entire putative coding region by PCR.

Figure 3. Southern blot analysis of *H. influenzae* strain HI689 genomic digests, probed with ECL labeled oligonucleotide specific for the 120-kDa Hgp N-terminus, and developed with ECL detection reagents as directed by the manufacturer (Amersham). Lane A, labeled λ *Hin*dIII fragments; Lane B-G are *H. influenzae* HI689 chromosomal DNA digested by *Eco*RI (Lane B), *Pst*I (Lane C), *Eco*RI/*Pst*I (lane D), *Hinc*II (lane E), *Sac*II (lane F), *Pst*I/*Hinc*II (Lane G).

Figure 4. Southern blot probed with ECL labeled oligonucleotide specific for the 120-kDa Hgp N-terminus, and developed with ECL detection reagents as directed by the manufacturer (Amersham). Lane A, labeled λ *Hin*dIII fragments; Lane B, *H. influenzae* HI689 chromosomal DNA *Eco*RI digest; Lane C, pUC19 *Eco*RI digested; Lane D, pHFJ1 *Eco*RI digested; Lane E, pHFJ1 undigested; Lane F, pHFJ2 *Eco*RI digested; Lane G, pHFJ2 undigested.

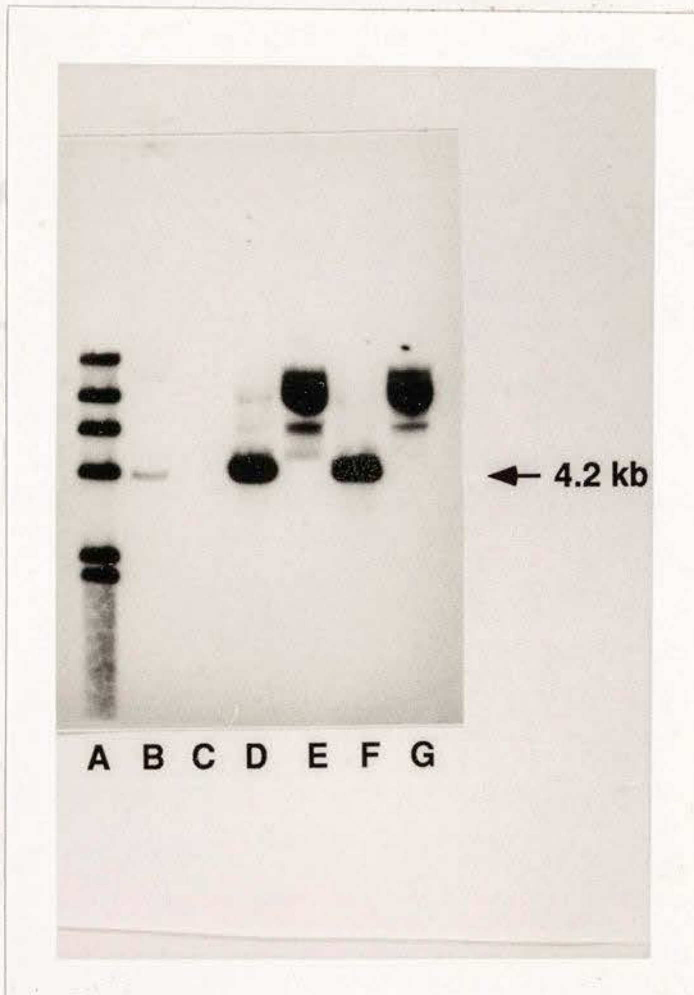


Figure 5. Partial restriction enzyme map of the 4.2-kb insert of pHFJ2, the predicted ORF encoding Hgp. The boxed area represents the putative ORF encoding Hgp and the insert of pXHGP. Numbers represent size in kilobase pairs.

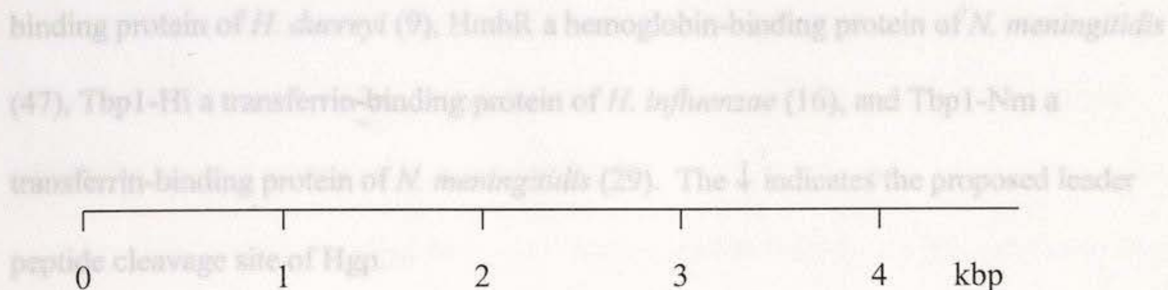


Figure 6. Peptide alignments between the N-terminal region of the putative Hgp and the N-terminal regions of other iron and heme related proteins. Residues which are 100% conserved between the five proteins are bold. The other proteins are: HgbA a hemoglobin-binding protein of *H. ducreyi* (9), HmbR a hemoglobin-binding protein of *N. meningitidis* (47), Tbp1-Hi a transferrin-binding protein of *H. influenzae* (16), and Tbp1-Nm a transferrin-binding protein of *N. meningitidis* (29). The ↓ indicates the proposed leader peptide cleavage site of Hgp.

```
                                        ↓
    HgpA  ...MTNFRLN VLAYSVMLGL TASVAYAEPT NQPTNQPTNQ PTNQPTNQPT
    HgbA  ...MKANKLS AITLCIL.G. YAHTVYAESN MQT....... ..........
    HmbR  ...MKPLQML PIAALVGSIF GNPVFAADEA AT........ ..........
    Tbp1  MTKKPYFRLS IISCLLISCY VKAETQSIKD TKEAISSEVD T.........
 Tbp1-Nm  MQQQHLFRLN ILC...LSLM TALPVYAENV QA........ ..........

    HgpA  NQPTNQPTNQ PTNQPTNQPT NQNSNASEQL EQINVSGSTE NTDTKAPPKI
    HgbA  .......... .......... .......EKL ETIVVSSEDD SVHNK...NV
    HmbR  .......... .......... .......... ETTPVKAEVK AVRVKGQRNA
    Tbp1  .......... .......... ..QSTEDSEL ETISVTAEKV RDR.KDNEVT
 Tbp1-Nm  .......... .......... ..EQAQEKQL DTIQVKAKKQ KTR.RDNEVT

    HgpA  AETVK..TAK KLEKEQAQDV KDLVRYETGI TVVEAGRFGN SGFAVRGVEE
    HgbA  GEIKK..NSK ALSKQQVQDS RDLVRYETGV TVVEKGRFGS SGYAIRGVDE
    HmbR  PAAVERVNLN RIKQEMIRDN KDLVRYSTDV GLSDSGRHQK .GFAVRGVEG
    Tbp1  GLGKIIKTSE SISREQVLNI RDLTRYDPGI SVVEQGRGAS SGYSIRGMDR
 Tbp1-Nm  GLGKLVKSSD TLSKEQVLNI RDLTRYDPGI AVVEQGRGAS SGYSIRGMDK
```

Figure 7. Peptide alignments between the C-terminal region of the putative Hgp and the C-terminal regions of other iron and heme related proteins. Residues which are 100% conserved between the five proteins are bold. The other proteins are: HgpA a hemoglobin-binding protein of *H. ducreyi* (9), HmbR a hemoglobin-binding protein of *N. meningitidis* (47), Tbp1-Hi a transferrin-binding protein of *H. influenzae* (16), and Tbp1-Nm a transferrin-binding protein of *N. meningitidis* (29).

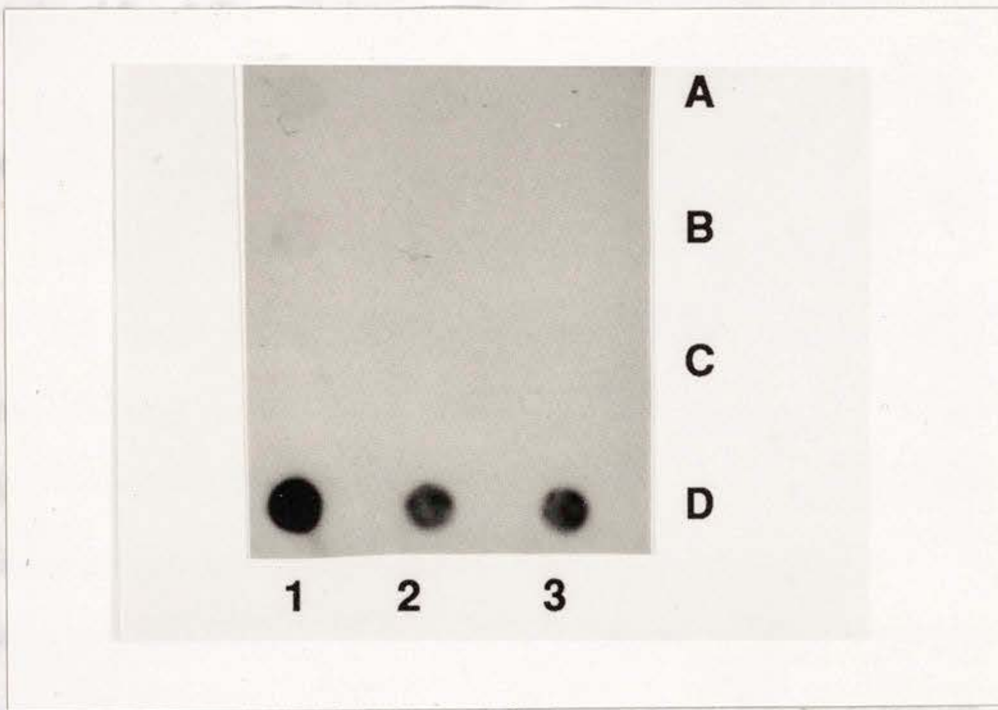
```
    HgpA  KEEG...... KTDSTIKWRS KSYTTIDLLG YIKPIKNLTL RAGVYNLTNR
    HgbA  DGKD...... QKDQHIKWRS DRYTLVDLIA YVKPVKNVTL RAGVYNLTNR
    HmbR  ENKGWGTPLQ KKVKDYPWLN KSAYVFDMYG FYKPVKNLTL RAGVYNVFNR
    Tbp1  GKRALGNNSR NV.KSTRKLT RAWHILDVSG YYMVNRSILF RLGVYNLLNY
 Tbp1-Nm  GSQALLNGNA NAKKAASRRT RPWYVTDVSG YYNIKKHLTL RAGVYNLLNY

    HgpA  KYITWDSARS IRPFGTSNMI NQDTGLGI.N RFYAPERNYR MSVQFEF.
    HgbA  EYGTWDSIRS IRPFGTTNLI NQETGKGI.K RFNAPGRNFR VNAEITF*
    HmbR  KYTTWDSLRG LYSYSTTNSV DRD.GKGL.D RYRAPSRNYA VSLEWKF*
    Tbp1  RYVTWEAVRQ ....TAQGAV NQHQNVGNYT RYAASGRNYT LTLEMKF*
 Tbp1-Nm  RYVTWENVRQ ....TAGGAV NQHKNVGVYN RYAAPGRNYT FSLEMKF*
```

Restriction enzyme sites were included in the primers to permit directional cloning in the expression vector pRSETA. Using pHFJ2 as the template in the PCR an approximately 3.2-kb product was amplified; restriction analysis of the amplicon revealed it to be the expected product. The gel-purified PCR product was successfully cloned into pRSETA as demonstrated by restriction mapping (Fig. 5), and the resulting plasmid was designated pXHGP

Expression of Hgp in *E. coli*. A hemoglobin-binding dot-blot assay was used to investigate whether the cloned hemoglobin-binding protein gene in *E. coli* resulted in expression of a hemoglobin-binding phenotype. Sonicated *E. coli* BL21(DE3)pLysS harboring pXHGP bound biotinylated human hemoglobin following induction with IPTG (Fig. 8, row D). Whole cells of IPTG-induced *E. coli* BL21(DE3)pLysS harboring pXHGP did not bind hemoglobin (Fig. 8, row B), probably because the recombinant protein is not expressed at the *E. coli* cell surface. *E. coli* BL21(DE3)pLysS harboring pRSETA alone and induced with IPTG did not bind hemoglobin whether or not the cells were lysed by sonication (Fig. 8, rows A & C), and all tested strains were unable to bind hemoglobin when uninduced (data not shown). These data demonstrate that pXHGP contained a fragment of DNA from *H. influenzae* HI689 which encodes a hemoglobin-binding protein, Hgp, which is expressed in *E. coli*.

Figure 8. Dot-blot analysis of human hemoglobin binding to *E. coli* either harboring or without pXHGP. Columns 1, 2 and 3 represent serial dilutions of whole cells or sonicated cells. Rows A, whole cells of *E. coli* BL21(DE3) pLysS pRSETA; B, whole cells of *E. coli* BL21(DE3) pLysS pXHGP; C, sonicates of *E. coli* BL21(DE3) pLysS pRSETA; D, sonicates of *E. coli* BL21(DE3) pLysS pXHGP.



Lack of cross-reactivity between HgbA of *H. ducreyi* and Hgp of *H. influenzae*. A conserved 100-kDa hemoglobin-binding protein from *H. ducreyi*, HgbA, has been recently identified by Elkins (8,9). To determine if antibodies generated against HgbA would recognize the 120-kDa Hgp described in this paper, Western blots were probed with antibodies raised against the 100-kDa hemoglobin-binding protein of *H. ducreyi*. The antibodies reacted with the 100-kDa protein isolated from *H. ducreyi*, but not with either the 120-kDa protein isolated from *H. influenzae*, or the recombinant protein isolated from IPTG-induced *E. coli* (Data not shown).

DISCUSSION

H. influenzae has an absolute growth requirement for PPIX, the immediate precursor of heme (11). The availability of heme within the human host to invading pathogens is strictly limited. Heme is contained largely within intracellular hemoglobin; free hemoglobin and heme are rapidly bound by the serum proteins haptoglobin (in the case of hemoglobin) and hemopexin or albumin (in the case of heme) and cleared from the circulation (3). The heme requirement of *H. influenzae* can be satisfied *in vitro* by hemoglobin, hemoglobin complexed to haptoglobin, heme complexed to either hemopexin or albumin, and protoporphyrin IX in the presence of an iron source such as ferritransferrin (3,35,36). The mechanism(s) by which *H. influenzae* takes up iron and/or heme from these protein sources has not been fully elucidated. It is clear, however, that *H. influenzae* does

not produce siderophores (7,35,36,41). Certain uptake mechanisms involve a direct interaction between the protein and the bacterial cell surface (14,36,55). The utilization of heme and the acquisition of heme from hemoglobin, the hemoglobin-haptoglobin complex, and the heme-hemopexin complex is dependent on a functional *tonB* gene, indicating that uptake is mediated by an outer membrane TonB-dependent protein(s) (23,42), and a *tonB* homologue has been reported in the recently sequenced genome of *H. influenzae* strain Rd (13). Based on the sequence currently available Hgp exhibits significant homology with other TonB proteins over regions that are highly conserved amongst this class of proteins (6). In the case of acquisition of iron from transferrin two outer membrane proteins, Tbp1 and Tbp2, are involved in binding of transferrin to *H. influenzae* (16). Subsequent steps in the process have not been fully elucidated although there is evidence for a periplasmic iron transport system encoded by an operon of three genes *hitABC* (1). Proteins binding the heme-hemopexin complex have been described, one of which is apparently secreted into the growth media (4,5,20,56). A heme-binding outer membrane protein has been isolated (27), and a heme-binding lipoprotein with significant homology to a periplasmic transport protein of *E. coli* has also been characterized (19,21).

In this report we identify the gene encoding a 120-kDa hemoglobin-binding protein (Hgp) of *H. influenzae* type b. Hgp was isolated from *H. influenzae* using affinity chromatography and an N-terminal amino acid sequence obtained (AQPTNQPTNQ). Using oligonucleotides derived from the N-terminal sequence as probes a DNA fragment from *H. influenzae* was cloned, to yield pHFJ2. Partial sequencing of pHFJ2 revealed that

the N-terminal nucleotide sequence would encode a peptide of sequence AEPTNQPTNQ. On the basis of sequence homology and Southern analysis there is no gene corresponding to *hgp* present in the Rd chromosomal sequence recently reported. Although pHFJ2 contains a region of greater than 95% identity to the HI0594 locus of *H. influenzae* upstream of *hgp* there is no sequence homologous to *hgp* downstream of the HI0594 locus in the Rd chromosomal sequence (13). However the nucleotide sequence encoding the N-terminal region of Hgp is highly homologous with the N-terminal regions of the putative products of 3 separate open reading frames (ORFs) in the Rd chromosomal sequence (13). Each of these areas in the Rd sequence (designated HI0661, HI0712, and HI1566 by Fleischmann *et al.* [13]) and the corresponding region of *hgp* contains multiple CCAA repeats, with the number of repeats varying between 18 and 36, giving rise to proteins having 6 to 12 PTNQ repeats. In two of the three putative ORFs in Rd the CCAA repeat is followed by an in frame stop codon, whilst the third constitutes an ORF of approximately 3 kb. The function of the repeat region is not known although since conservation is 100% at the nucleotide level it seems likely that their importance lies in the CCAA unit rather than the encoded peptide sequence. It is possible that the CCAA repeats may have a regulatory function, perhaps in a manner analogous to the slip-strand regulation involved in phase variation of lipopolysaccharide expression in *H. influenzae* (51). In the two ORFs identified in Rd which have a stop codon downstream of the CCAA region, slippage across one CCAA unit would eliminate the stop codon. The N-terminal and C-terminal amino acid sequences of Hgp have been compared with the corresponding amino acid regions of the proteins

encoded by HI0661, HI0712 and HI1566. To facilitate this the reading frame of HI0661 and HI1566 were altered by addition or removal of a CCAA unit from the published nucleotide sequence as appropriate. Over the 76 amino acids directly subsequent to the final QPTN unit of Hgp identity was calculated as 80% for HI0661, 70% for HI0712, and 72% for HI1566. Over the final 90 amino acids of Hgp the respective identities were 73% (HI0661), 63% (HI0712), and 62% (HI1566). These analyses indicate that the protein encoded by ORF HI0661 is the most closely related to Hgp. The ORF HI0661 does not correspond to *hgp* since we have partially cloned a homolog of HI0661 from strain *H. influenzae* HI689 (data not shown). In addition we have partially cloned a HI0712 homologue from HI689 (data not shown).

The discrepancy between the microsequencing derived N-terminal sequence (AQPTNQPTNQ) and the nucleotide sequence derived N-terminal (AEPTNQPTNQ) is unlikely to be due to microsequencing errors since the misread of a glutamine (Q) for a glutamic acid (E) is highly unlikely (Audree Fowler, John Keyte; personal communications). The possibility of nucleotide sequencing errors was minimized by repeated sequencing across the area of interest. An alternative explanation for the difference is that the protein originally isolated by affinity chromatography was not Hgp, but rather one of the three ORF products identified in *H. influenzae* Rd. In particular the product of HI0712 has a predicted molecular mass of 124 kDa compared to approximately 120 kDa for Hgp, and the nucleotide sequence derived N-terminal amino acid sequence of HI0712 is AQPTNQPTN. The product of HI0661 has a predicted molecular mass of 115 kDa

and an N-terminal amino acid sequence of AQPTNQPTN whilst the data for the HI1566 gene product would be 114 kDa and AEPTNQPTN respectively. Thus it is possible that an alternate gene product was originally isolated (i.e., ORF HI0712 or ORF HI0661) whilst the oligonucleotides designed from the amino acid sequence led to cloning of *hgp* based on the 100% conservation of the CCAA region. This possibility indicates that there is more than one hemoglobin binding protein expressed by *H. influenzae*, and that one (or more) of HI0661, HI0712 and HI1566 may represent this additional protein(s). Work is underway to clarify how many hemoglobin binding proteins *H. influenzae* expresses, and to define the functions of the gene products of HI0661, HI0712, and HI1566.

A putative coding sequence from pHFJ2 was subcloned into *E. coli*, and expression of this gene resulted in the recombinant strain binding hemoglobin. Hemoglobin-binding proteins have been reported in *H. ducreyi*, and *Neisseria meningitidis*, (8,28,30,47). Using affinity chromatography Elkins isolated a protein of 100 kDa from *H. ducreyi*, and also reported the isolation of an approximately 115-kDa protein from *H. influenzae* strain DL42 (8,9). The 115-kDa protein from *H. influenzae* was not recognized by antibodies raised to either the entire *H. ducreyi* 100-kDa protein or to the N-terminal peptide of the 100-kDa protein. In addition, the N-terminal sequence reported for the *H. ducreyi* protein (ESNMQTEKLETIVV) is highly dissimilar to the N-terminal of the 120-kDa protein reported here (AEPTNQPTNQ). Neither the recombinant Hgp nor the wild-type Hgp react with an antibody raised against the 100-kDa protein of *H. ducreyi*, however it remains to be

clarified whether Hgp is related to the 115-kDa protein isolated from *H. influenzae* DL42 (9). Two hemoglobin-binding proteins have been described in *N. meningitidis*, one of 85-kDa (30) and a second of 89.5-kDa (47); the former binds the hemoglobin-haptoglobin complex. These two proteins were isolated from different strains and it is not known whether both proteins are expressed in the same strain or whether they interact with each other in the acquisition of heme from hemoglobin, although Lee and Hill have suggested that meningococcus possesses two hemoglobin receptors with different affinities for the ligand (28). In the case of transferrin binding by *H. influenzae* and *N. gonorrhoeae* it is clear that expression of both Tbp1 and Tbp2 is necessary for maximal binding of the ligand (2,16), although partial binding occurs when either protein is present alone. There is no direct evidence to date for the existence of more than one hemoglobin-binding protein in either *H. influenzae* or *H. ducreyi*. In previous studies (14) we showed that binding of hemoglobin by *H. influenzae* was blocked by the hemoglobin-haptoglobin complex. It is possible that, similarly to the 85-kDa *N. meningitidis* protein (30), Hgp also binds the hemoglobin-haptoglobin complex, and studies will be performed to answer this question.

Hgp was isolated from heme-restricted but not from heme-replete *H. influenzae*, indicating that expression of the protein may be regulated by levels of heme. Previously we have shown that the hemoglobin-binding phenotype in *H. influenzae* is repressible by heme (14), and that transferrin binding expressed by *H. influenzae* is repressible by heme as opposed to elemental iron levels (34). The 100-kDa hemoglobin-binding protein of *H. ducreyi* and the 39.5-kDa heme-binding outer membrane protein of *H. influenzae* are

similarly repressible by heme, rather than elemental iron (8,9,27). In many bacterial species the Fur (Ferric uptake regulator) protein is a regulatory element which in the presence of iron binds to a conserved binding sequence (the Fur-box) upstream of iron regulated genes, preventing transcription (31). Sequences homologous to the consensus Fur binding sequence (31) have been reported upstream of the *H. ducreyi* hemoglobin-binding protein (9), the *H. influenzae* transferrin-binding proteins (16), and a heme-hemopexin utilization gene cluster of *H. influenzae* (5). In none of these cases has the putative Fur binding site been shown to be functional; indeed the 100-kDa heme-hemopexin-binding protein described by Cope *et al.* appears to be expressed constitutively (5). That this protein is expressed in a heme and iron sufficient media despite a putative Fur box (5), raises further questions regarding the functionality of this putative Fur-box. Fleischmann *et al.* have recently identified an ORF with strong homology (61.4 % identity; 75 % similarity) to the *E. coli fur* gene in the genome of *H. influenzae* strain Rd (13,45). Thus a Fur analog may be a regulatory factor for expression of the *H. influenzae* iron and heme related proteins, although it is clear that clarification is necessary. In addition the 57-kDa heme-hemopexin-binding protein of Wong *et al.* is apparently repressible by iron (55,56), although the heme-hemopexin phenotype is variously reported as inducible by iron-restricted growth or heme-restricted anaerobic growth (21,55). The regulation of heme and iron related proteins in *H. influenzae* remains poorly characterized, and further studies should be undertaken to clarify the roles of iron and heme in regulating their expression.

In conclusion we have identified a 120-kDa hemoglobin-binding protein (Hgp) of *H. influenzae*, the expression of which is repressible by heme, and cloned the gene encoding Hgp. We speculate that Hgp represents an important step in the acquisition of heme from hemoglobin. Further studies will clarify the role of Hgp in hemoglobin utilization, investigate the possibility of other hemoglobin binding proteins, and define the mechanism of regulation of this protein(s).

REFERENCES

1. **Adhikari, P., S. D. Kirby, A. J. Nowalk, K. L. Veraldi, A. B. Schryvers, and T. A. Mietzner.** 1995. Biochemical characterization of a *Haemophilus influenzae* periplasmic iron transport operon. J. Biol. Chem. **270:**25142-25149.

2. **Anderson, J. E., P. F. Sparling, and C. N. Cornelissen.** 1994. Gonococcal transferrin-binding protein 2 facilitates but is not essential for transferrin utilization. J. Bacteriol. **176:**3162-3170.

3. **Bezkorovainy, A.** 1987. Iron Proteins, p. 27-68. *In* J. J. Bullen and E. Griffiths (ed), Iron and Infection: Molecular, Physiological and Clinical Aspects. John Wiley & Sons, New York, NY.

4. **Cope, L. D., S. E. Thomas, J. L. Latimer, C. A. Slaughter, U. Muller-Eberhard, and E. J. Hansen.** 1994. The 100 kDa haem:haemopexin-binding protein of *Haemophilus influenzae*: structure and localization. Mol. Microbiol. **13:**863-873.

5. **Cope, L. D., R. Yogev, U. Muller-Eberhard, and E. J. Hansen.** 1995. A gene cluster involved in the utilization of both free heme and heme:hemopexin by *Haemophilus influenzae* type b. J. Bacteriol. **177:**2644-2653.

6. **Cornelissen, C. N., G. D. Biswas, J. Tsai, D. K. Paruchuri, S. A. Thompson, and P. F. Sparling.** 1992. Gonococcal transferrin-binding protein 1 is required for transferrin utilization and is homologous to TonB-dependent outer membrane receptors. J. Bacteriol. **174:**5788-5797.

7. **Cornelissen, C. N. and P. F. Sparling.** 1994. Iron piracy: acquisition of transferrin-bound iron by bacterial pathogens. Mol. Microbiol. **14:**843-850.

8. **Elkins, C.** 1995. Identification and purification of a conserved heme-regulated hemoglobin-binding outer membrane protein from *Haemophilus ducreyi*. Infect. Immun. **63:**1241-1245.

9. **Elkins, C., C. J. Chen, and C. E. Thomas.** 1995. Characterization of the *hgbA* locus encoding a hemoglobin receptor from *Haemophilus ducreyi*. Infect. Immun. **63:**2194-2200.

10. **Eskola, J., H. Kayhty, H. Peltola, V. Karanko, P. H. Makela, J. Samuelson, and L.K. Gordon.** 1985. Antibody levels achieved in infants by course of *Haemophilus influenzae* type b polysaccharide/diphtheria toxoid conjugate vaccine. Lancet. **i:**1184-1186.

11. **Evans, N. M., D. D. Smith, and A. J. Wicken.** 1974. Haemin and nicotinamide adenine dinucleotide requirements of *Haemophilus influenzae* and *Haemophilus parainfluenzae*. J. Med. Microbiol. **7:**359-365.

12. **Falla, T. J., S. R. Dobson, D. W. Crook, W. A. Kraak, W. W. Nichols, E. C. Anderson, J. Z. Jordens, M. P. Slack, E. Mayon-White, and E. R. Moxon.** 1993. Population-based study of non-typable *Haemophilus influenzae* invasive disease in children and neonates. Lancet. **341:**851-854.

13. **Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. Tomb, B. A. Dougherty, J. M. Merrick, K. McKenney, G. Sutton, W. FitzHugh, C. Fields, J. D. Gocayne, J. Scott, R. Shirley, L. Liu, A. Glodek, J. M. Kelley, J. F. Weidman, C. A. Phillips, T. Spriggs, E. Hedblom, M. D. Cotton, R. C. Utterback, M. C. Hanna, D. T. Nguyen, D. M. Saudek, R. C. Brandon, L. D. Fine, J. L. Fritchman, J. L. Fuhrmann, N. S. M. Geoghagen, C. L. Gnehm, L. A. McDonald, K. V. Small, C. M. Fraser, H. O. Smith, and J. C. Venter.** 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.Science. **269:**496-512.

14. **Frangipane, M. E., D. J. Morton, J. A. Wooten, J. M. Pozsgay, and T .L. Stull.** 1994. Binding of human hemoglobin by *Haemophilus influenzae*. FEMS. Microbiol. Lett. **118:**243-248.

15. **Gilsdorf, J. R., K. McCrea, and L. Forney.** 1990. Conserved and nonconserved epitopes among *Haemophilus influenzae* type b pili. Infect. Immun. **58:**2252-2257.

16. **Gray-Owen, S. D., S. Loosmore, and A. B. Schryvers.** 1995. Identification and characterization of genes encoding the human transferrin-binding proteins from *Haemophilus influenzae*. Infect. Immun. **63:**1201-1210.

17. **Gray-Owen, S. D. and A. B. Schryvers.** 1995. Characterization of transferrin binding proteins 1 and 2 in invasive type b and nontypeable strains of *Haemophilus influenzae*. Infect. Immun. **63:**3809-3815.

18. **Griffiths, E.** 1987. Iron in Biological Systems, p. 1-26. *In* J. J. Bullen and E. Griffiths (ed), Iron and Infection: Molecular, Physiological and Clinical Aspects. John Wiley & Sons, New York, NY.

19. **Hanson, M. S. and E. J. Hansen.** 1991. Molecular cloning, partial purification, and characterization of a haemin-binding lipoprotein from *Haemophilus influenzae* type b. Mol. Microbiol. **5:**267-278.

20. **Hanson, M. S., S. E. Pelzel, J. Latimer, U. Muller-Eberhard, and E. J. Hansen.** 1992. Identification of a genetic locus of *Haemophilus influenzae* type b necessary for the binding and utilization of heme bound to human hemopexin. Proc. Natl. Acad. Sci. USA. **89:**1973-1977.

21. **Hanson, M. S., C. Slaughter, and E. J. Hansen.** 1992. The *hbpA* gene of *Haemophilus influenzae* type b encodes a heme- binding lipoprotein conserved among heme-dependent *Haemophilus* species. Infect. Immun. **60:**2257-2266.

22. **Hardie, K. R., R. A. Adams, and K. J. Towner.** 1993. Transferrin-binding ability of invasive and commensal isolates of *Haemophilus* spp. J. Med. Microbiol. **39:**218-224.

23. **Jarosik, G. P., J. D. Sanders, L. D. Cope, U. Muller-Eberhard, and E. J. Hansen.** 1994. A functional *tonB* gene is required for both utilization of heme and virulence expression by *Haemophilus influenzae* type b. Infect. Immun. **62:**2470-2477.

24. **Kostman, J. R., B. L. Sherry, C. L. Flinger, S. Egaas, P. Sheeran, L. Baken, J. E. Bauwens, C. Clausen, D. M. Sherer, J. J. Plorde, T. L. Stull, and P. M. Mendelmen.** 1993. Invasive *Haemophilus influenzae* infections in older children and adults in Seattle. Clin. Inf. Dis. **17:**389-396.

25. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. **227:**680-685.

26. **Lee, B. C.** 1995. Quelling the red menace: haem capture by bacteria. Mol. Microbiol. **18:**383-390.

27. **Lee, B. C.** 1992. Isolation of an outer membrane hemin-binding protein of *Haemophilus influenzae* type b. Infect. Immun. **60:**810-816.

28. **Lee, B. C. and P. Hill.** 1992. Identification of an outer-membrane haemoglobin-binding protein in *Neisseria meningitidis*. J. Gen. Microbiol. **138:**2647-2656.

29. **Legrain, M., V. Mazarin, S. W. Irwin, B. Bouchon, M. J. Quentin-Millet, E. Jacobs, and A. B. Schryvers.** 1993. Cloning and characterization of *Neisseria meningitidis* genes encoding the transferrin-binding proteins Tbp1 and Tbp2. Gene. **130:**73-80.

30. **Lewis, L. A. and D. W. Dyer.** 1995. Identification of an iron-regulated outer membrane protein of *Neisseria meningitidis* involved in the utilization of hemoglobin complexed to haptoglobin. J. Bacteriol. **177:**1299-1306.

31. **Litwin, C. M. and S. B. Calderwood.** 1993. Role of iron in regulation of virulence genes. Clin. Microbiol. Rev. **6:**137-149.

32. **Madorne, D. V., C. L. Johnson, D. C. Phipps, Pennridge Pediatric Associates, L. A. Popejoy, R. Eby, and D. H. Smith.** 1990. Safety and immunologic response to *Haemophilus influenzae* type b oligosaccharide-CRM197 conjugate vaccine in 1 to 6-month-old infants. Pediatrics. **85:**331-337.

33. **Moos, M.** 1993. Isolation of proteins for microsequence analysis, p. 8.71-8.712. *In* J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, and W. Strober (ed), Current Protocols in Immunology. John Wiley & Sons, New York. NY.

34. **Morton, D. J., J. M. Musser, and T. L. Stull.** 1993. Expression of the *Haemophilus influenzae* transferrin receptor is repressible by hemin but not elemental iron alone. Infect. Immun. **61:**4033-4037.

35. **Morton, D. J. and P. Williams.** 1989. Utilization of transferrin-bound iron by *Haemophilus* species of human and porcine origins. FEMS. Microbiol. Lett. **53:**123-127.

36. **Morton, D. J. and P. Williams.** 1990. Siderophore-independent acquisition of transferrin-bound iron by *Haemophilus influenzae* type b. J. Gen. Microbiol. **136:**927-933.

37. **Moxon, E. R. and K. A. Vaughan.** 1981. Type b capsular polysaccharide as a virulence determinant of *Haemophilus influenzae:* studies using clinical isolates and laboratory transformants. J. Infect. Dis. **143:**517-524.

38. **Murphy, T. V., K. E. White, and P. Pastor.** 1993. Declining incidence of *Haemophilus influenzae* type b disease since introduction of vaccination. JAMA. **269:**246-248.

39. **Musser, J. M., S. J. Barenkamp, D. M. Granoff, and R. K. Selander.** 1986. Genetic relationships of serologically nontypable and serotype b strains of *Haemophilus influenzae*. Infect. Immun. **52:**183-191.

40. **Otto, B. R., A. M. Verweij-van Vught, and D. M. MacLaren.** 1992. Transferrins and heme-compounds as iron sources for pathogenic bacteria. Crit. Rev. Microbiol. **18:**217-233.

41. **Pidcock, K. A., J. A. Wooten, B. A. Daley, and T. L. Stull.** 1988. Iron acquisition by *Haemophilus influenzae*. Infect. Immun. **56:**721-725.

42. **Postle, K.** 1990. TonB and the Gram-negative dilemma. Mol. Microbiol. **4:**2019-2025.

43. **Roberts, M., T. L. Stull, and A. L. Smith.** 1981. Comparative virulence of *Haemophilus influenzae* with a type b or type d capsule. Infect. Immun. **32:**518-524.

44. **Sambrook, T., E.F. Fritsch, and T. Maniatis.** 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

45. **Schaffer, S., K. Hantke, and V. Braun.** 1985. Nucleotide sequence of the iron regulatory gene *fur*. Mol. Gen. Genet. **200:**110-113.

46. **Shapiro, E. D., L. S. Capobianco, A. T. Berg, and M. Q. Zitt.** 1989. The immunogenicity of *Haemophilus influenzae* type b polysaccharide-*Neisseria meningitidis* group B outer membrane protein complex vaccine in infants and young children. J. Infect. Dis. **160:**1064-1066.

47. **Stojiljkovic, I., V. Hwa, L. de Saint Martin, P. O'Gaora, X. Nassif, Heffron, F, and M. So.** 1995. The *Neisseria meningitidis* haemoglobin receptor: its role in iron utilization and virulence. Mol. Microbiol. **15:**531-541.

48. **Stull, T. L.** 1987. Protein sources of heme for *Haemophilus influenzae*. Infect. Immun. **55:**148-153.

49. **Stull, T. L., K. Mack, J. E. Haas, J. Smit, and A. L. Smith.** 1985. A comparison of techniques for isolation of the outer membrane proteins of *Haemophilus influenzae* type b. Anal. Biochem. **150:**471-480.

50. **Szabo, M., D. Maskell, P. Butler, J. Love, and R. Moxon.** 1992. Use of chromosomal gene fusions to investigate the role of repetitive DNA in regulation of genes involved in lipopolysaccharide biosynthesis in *Haemophilus influenzae*. J. Bacteriol. **174:**7245-7252.

51. **Teele, D. W., J. O. Klein, V. Rosner, and The Greater Boston Otitis Media Study Group.** 1989. Epidemiology of otitis media during the first seven years of life in children in Greater Boston: a prospective, cohort study. J. Infect. Dis. **160:**83-94.

52. **Turk, D. C.** 1984. The pathogenicity of *Haemophilus influenzae*. J. Med. Microbiol. **18:**1-16.

53. **Wallace, R. J., D. M. Musher, and R. J. Martin.** 1978. *Haemophilus influenzae* pneumonia in adults. Am. J. Med. **64:**87-93.

54. **Wenger, J. D., R. Pierce, K. Deaver, R. Fraklin, G. Bosley, N. Pigott, and C. V. Broome.** 1992. Invasive *Haemophilus influenzae* disease: a population-based evaluation of the role of capsular polysaccharide serotype. J Infect. Dis. **165:**S34-S35.

55. **Wong, J. C. Y., J. Holland, T. Parsons, A. Smith, and P. Williams.** 1994. Identification and characterization of an iron-regulated hemopexin receptor in *Haemophilus influenzae* type b. Infect. Immun. **62:**48-59.

56. **Wong, J. C. Y., R. Patel, D. Kendall, P. W. Whitby, A. Smith, J. Holland, and P. Williams.** 1995. Affinity, conservation, and surface exposure of hemopexin-binding proteins in *Haemophilus influenzae*. Infect. Immun. **63:**2327-2333.

CHAPTER III

CHARACTERIZATION OF *hgpA*, A GENE ENCODING A HEMOGLOBIN/ HEMOGLOBIN-HAPTOGLOBIN-BINDING PROTEIN OF *HAEMOPHILUS INFLUENZAE*

ABSTRACT

Haemophilus influenzae binds hemoglobin and the hemoglobin-haptoglobin complex and utilizes either as a sole source of heme. Previously, we cloned a *H. influenzae* DNA fragment, encoding an approximately 120 kDa protein, expressing hemoglobin-binding protein (HgpA) activity in *Escherichia coli.* Partial sequence analysis revealed significant homology of HgpA with other bacterial heme 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 analyzed. In addition to homology with other heme 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 repeats-containing locus indicated varying lengths of CCAA repeats, suggesting that hemoglobin binding in

H. influenzae is also regulated by strand slippage across CCAA repeats, other than by heme repression. *E. coli* containing cloned *hgpA* bound both hemoglobin and the hemoglobin-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 hemoglobin or hemoglobin-haptoglobin complex following growth in heme deplete media. Affinity purification of hemoglobin 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 hemoglobin binding protein exists.

INTRODUCTION

The human specific pathogen *Haemophilus influenzae* causes a range of infections including otitis media, meningitis, epiglottitis and pneumonia (39). 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 (24,27). 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 (8,17,41).

H. influenzae has an absolute growth requirement for protoporphyrin IX (PPIX), the immediate precursor of heme (7). *In vitro*, the requirement may be satisfied by heme, hemoglobin, complexes of hemoglobin-haptoglobin, heme-hemopexin and heme-albumin (37). *In vivo* heme is intracellular, in the form of hemoglobin or heme-containing enzymes, and thus unavailable to invading microorganisms (1,20). Hemoglobin released by erythrocytes is avidly bound by the serum protein haptoglobin, and the hemoglobin-haptoglobin complex is rapidly cleared by hepatocytes (1,28). Free heme, principally derived from the degradation of methemoglobin, is bound by the serum proteins hemopexin and albumin and cleared from the circulatory system by hepatocytes (1). We have shown that *H. influenzae* binds hemoglobin, possibly as an initial step in the utilization of hemoglobin-associated heme (10), and we have cloned a DNA fragment encoding an approximately 120-kDa hemoglobin-binding protein (HgpA) of *H. influenzae* (16).

The objective of the current study was to characterize the gene encoding the previously identified 120-kDa hemoglobin-binding protein (HgpA) of *H. influenzae* (16).

MATERIALS AND METHODS

Bacterial strains, culture conditions and plasmids. *H. influenzae* type b strain HI689 has been previously described (16). *H. influenzae* Rd KW20 (ATCC 51907) was the strain used in the complete sequencing of the *H. influenzae* genome (9) and was obtained from the American Type Culture Collection (ATCC), Rockville, Maryland. An additional Rd strain, maintained in this laboratory was used in some experiments. Non-typable *H. influenzae* TN106 has been previously described (23) and was kindly provided by Eric Hansen, University of Texas Southwestern Medical Center, Dallas, TX. *H. influenzae* type b strain HI1715 is a strain shown by Southern analysis to have a single CCAA repeats-containing locus associated with a gene which is highly homologous to *hgpA*. *H. influenzae* was routinely maintained on brain heart infusion (BHI) agar (Difco, Detroit, Mich.) supplemented with 10 $\mu g\ ml^{-1}$ of both heme and β-NAD. Heme-restricted growth of *H. influenzae* was performed in BHI broth supplemented with 10 $\mu g\ ml^{-1}$ of β-NAD and 0.1 $\mu g\ ml^{-1}$ of heme. *E. coli* DH5α were grown on Luria-Bertani (LB) medium. Ribostamycin was added where appropriate to a final concentration of 15 $\mu g\ ml^{-1}$ for *H. influenzae* and 50 $\mu g\ ml^{-1}$ for *E. coli*. Plasmids pHFJ2 and pUC19::TSTE have been previously described (16,35). Clones generated in the sequencing of the Rd KW20 genome (9), encompassing the ORFs designated HI0588 to HI0592 were obtained from the ATCC.

DNA Methodology. Restriction endonucleases were obtained from New England Biolabs (Beverly, Mass.) and were used as directed by the manufacturer. Genomic DNA was isolated using the DNA Now reagent (Biogentex, Seabrook, TX.) as directed by the manufacturer. Plasmid DNA was isolated using Qiagen plasmid kits (Qiagen, Chatsworth, Calif.) 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. Oligonucleotides were synthesized at the Molecular Biology Resource Facility, University of Oklahoma Health Sciences Center, Oklahoma City, OK. Southern Analysis was performed using the method of Southern as described by Sambrook *et al.* (34). The Enhanced Chemi-luminescene(ECL) random prime labeling kit (Amersham, Arlington Heights, Ill.) was used as directed by the manufacturer to label DNA fragments. Hybridization was performed overnight at 60 °C and stringency washes were 1 X SSC, 0.1% w/v SDS for 15 min and 0.5 X SSC, 0.1% w/v SDS for 15 min both at 60 °C (1 X SSC: NaCl 8.8 g/l, sodium citrate 4.4 g/l [pH 7.0]). Hybridized probes were detected using ECL nucleic acid detection reagents (Amersham) as directed by the manufacturer. Blots were subsequently exposed to X-ray film (Fuji Photo Film Co., Tokyo, Japan).

Construction of an insertion/deletion mutant in *hgpA*. Mutation of *hgpA* was achieved as follows. Plasmid pHFJ2 (16) was completely digested with *Bgl*II,

separated on a 1% w/v agarose gel. The 2.5 kbp internal *Bgl*II fragment was deleted and the remaining 4.5 kbp fragment was purified using the Prep-a-Gene kit (Bio-Rad, Hercules, Calif.). The 2.2 kbp TSTE fragment from pUC19::TSTE (35) was excised using *Bam*HI and gel purified. TSTE was ligated to the 4.5 kbp *Bgl*II fragment of pHFJ2, yielding plasmid pHFJC (Fig. 9). The plasmid pHFJC was transformed into *H. influenzae* strain HI689 made competent using the MII medium of Spenser and Herriott (36). Recombinant clones were selected initially by growth on BHI agar containing ribostamycin (15 μg l^{-1}). The mutant strain was designated HI689*hgpA*Δ*Bgl*II.

Confirmation that appropriate chromosomal rearrangements had occurred was obtained by Southern blot analysis (Fig. 9b).

Figure 9. a. 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 shaded area. The scale indicates size in kilobase pairs. b. Southern blot of *Eco*RI digested chromosomal DNA of *H. influenzae* strain HI689 (C and E) and *H. influenzae* strain HI689*hgp*Δ*Bgl*II (B and D) probed with either the labeled insert of pHFJ2 (B and C) or with the labeled TSTE insertion element (D and E). Lane A contains labeled λ *Hind*III digest; size of bands top to bottom are 23 kbp, 9.5 kbp, 6.5 kbp, 4.4 kbp, 2.3 kbp, and 2 kbp.

Figure 9a.

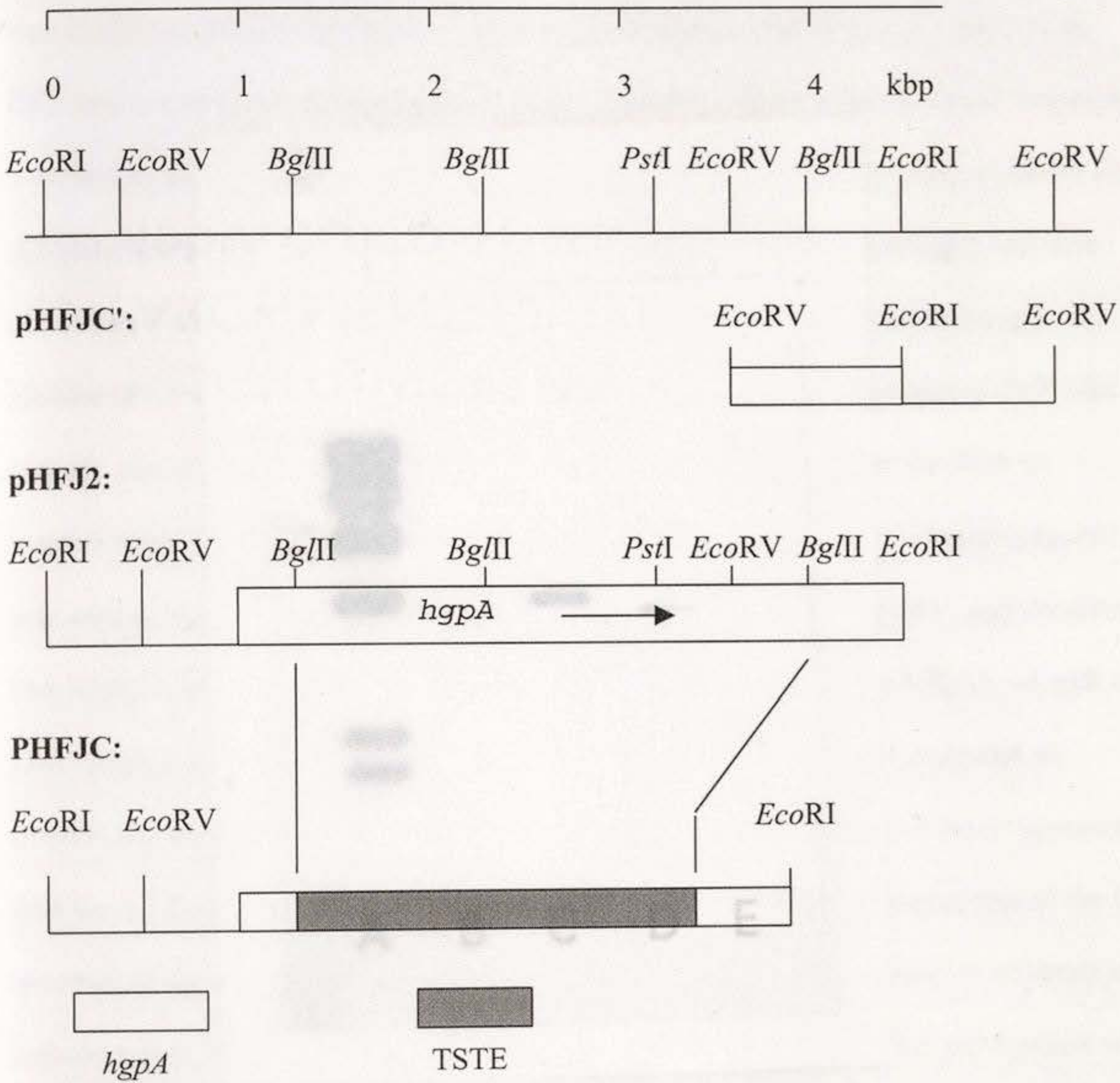
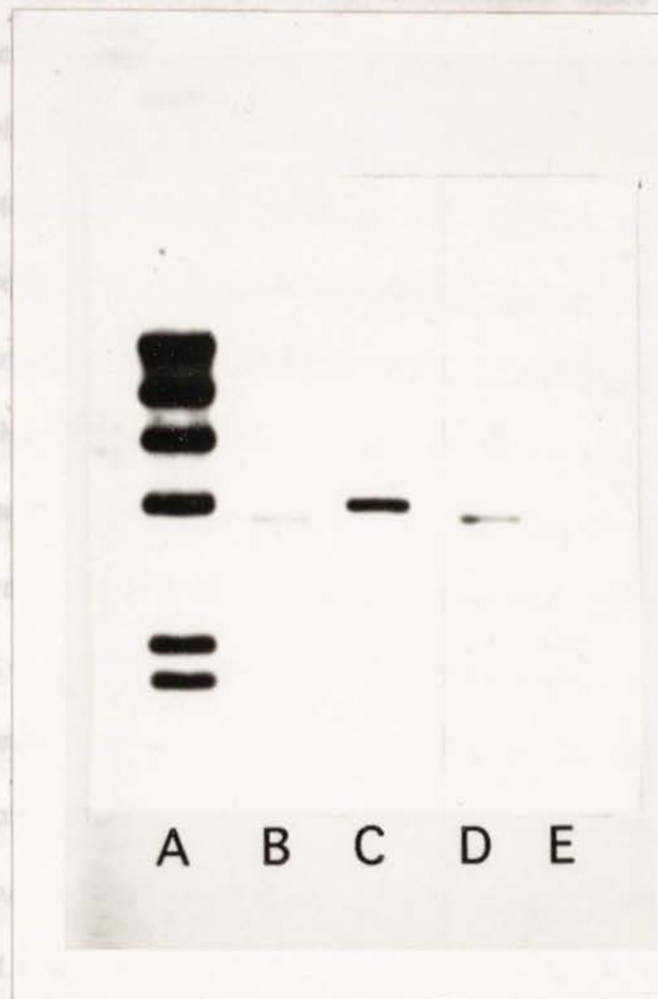


Figure 9b.



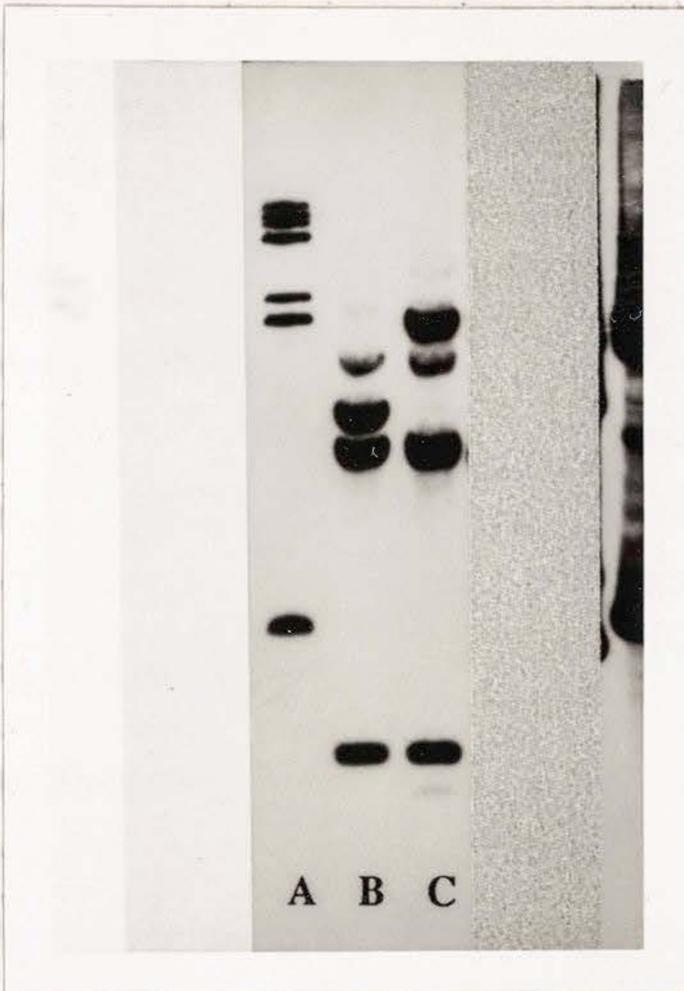
Cloning of DNA downstream of *hgpA*. The sequence of the insert of pHFJ2 was analyzed, and a putative gene (*hgpA*) was identified based on the existence of an ORF and homology to other sequences in the databases. Since a cloned *Eco*RI fragment was used to generate pHFJ2 (16), the ORF ended with GAATTC, encoding glutamic acid (E) and phenylalanine (F). Sequence comparison to other homologous heme 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 HFJCTERMS (Table 2) 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. HI689 chromosomal DNA was digested with *Eco*RV, and Southern blot analysis was performed using the 315 bp DNA probe. The probe hybridized with a DNA fragment of approximately1.4 kbp. Restriction mapping of *hgpA* indicated an *Eco*RV site internal to *hgpA* (Fig. 9), 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*. A HI689 mini-library was constructed by gel-purification of bands of approximately 1.4 kbp in an *Eco*RV digest of HI689 chromosomal DNA and ligation of the purified fragments to the pCR-Blunt vector (Invitrogen, Carlsbad, CA). The library was screened with the 315 bp probe. A positive clone was identified and designated pHFJC'.

Table 2. Primers used in this work.

Name	Sequence (5'→3')	Primer Location
HFJ2F4	GAAGTACGCATAGAATGATGC	intergenic region between HI0594 and *hgpA*
HFJ16	CGG GCAATTTTTGGCGGAGC	*hgpA*
HFJ25	TATATCTCTCATGCATCAG	*hgpA*
HFJCTERMS	GAATTCAAACTGAACTGAC	*hgpA*
HFJ26	GTCAGTTCAGTTTGAATTC	*hgpA*
HFJ29	CATATTCATCAAAAGTGCGG	HI0587
HFJXN	GACCAGGGATCCAATAGTAATGCTTCT	*hgpA*
HFJXC	GGAAGGGGTACCTTAGAATTCAAACTG	*hgpA*
PBUS	GGGCTAACGGCAAGTGTTGCTT	5 bp upstream of CCAA repeats
PBDS	GTTCTAGTTGTTGTTCAGAAGCATCC	5 bp downstream of CCAA repeats

Cloning of an independent copy of the CCAA region associated with *hgpA* from HI689. To identify the CCAA locus associated with *hgpA*, Southern analyses of the wildtype and the insertion/deletion mutant were performed (Fig.10). Chromosomal DNA from *H. influenzae* strains HI689 and HI689*hgpA*Δ*Bgl*II was digested with *Rsa*I. The digested fragments were separated on a 0.8% agarose gel, transferred onto a nylon membrane and probed with a $(CCAA)_6$ probe. Comparison of the hybridization pattern revealed that a 1.3 kbp band of HI689 had apparently shifted to 1.9 kbp in the mutant strain HI689*hgpA*Δ*Bgl*II, indicating that the 1.3 kbp hybridizing band in HI689 was associated with *hgpA* (Fig. 10). A limited library of HI689 *Rsa*I chromosomal fragments of approximately 1.3 kbp was constructed in the pCR-Blunt vector. The library was screened using the $(CCAA)_6$ oligonucleotide probe and positive clones were identified and sequenced by automated sequencing.

Figure 10. Southern blot of *Rsa*I digested chromosomal DNA of *H. influenzae* strain HI689 (Lane B) and *H. influenzae* strain HI689*hgp*Δ*Bgl*II (Lane C) probed with $(CCAA)_6$ repeats. Lane A contains labeled λ *Hind*III digest; size of DNA fragments top to bottom are 23 kbp, 9.5 kbp, 6.5 kbp, 4.4 kbp, 2.3 kbp, and 2 kbp.



PCR amplification across a region of CCAA repeats. A pair of oligonucleotide primers designated PBUS and PBDS were designed 5 bp upstream and downstream of the CCAA repeats respectively (Table 2). Following repeated subculture, single colonies of *H. influenzae* HI1715 were picked from chocolate agar plates and chromosomal DNA extracted using the Instagene kit (Bio-Rad) as directed by the manufacturer. Using HI1715 chromosomal DNA as templates, the PCR amplification was performed in a total volume of 50 µl containing 5 µl of 10x 30mM Mg^{2+} PCR buffer (Idaho Technology Inc., Idaho Falls, Idaho), 10 µCi α^{32}P dATP, 1 µl of 10 mM dNTPs, 1 unit of Taq DNA polymerase, and 10 pM of each primer. PCR was performed in a RapidCycler Thermocycler (Idaho Technology Inc.). The samples were cycled 30 times with an annealing step of 58 °C for 10 sec, extension at 72 °C for 30 sec, and a denaturing step at 95 °C for 10 sec. PCR products were visualized by running on a 6% denaturing PAGE 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.

Reverse Transcriptase-Polymerase chain reaction (RT-PCR). RT-PCR was performed essentially as described (43). One microgram of RNA prepared from heme-starved HI689 using the RNeasy total RNA kit (Qiagen) as directed by the manufacturer was treated with 1 U of 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 of 25 mM EDTA solution to the reaction mixture and heating for 10 min at 65 °C. A 12.5 µl

aliquot of the DNase-treated RNA sample was used for annealing with 1 μl of random nonamer primer (Amersham Random Prime Labeling kit) by heating to 70 °C for 2 min, followed by rapid chilling on ice. Complementary DNA was synthesized at 42 °C using Superscript II reverse transcriptase (Gibco BRL) in a 20 μl reaction containing 100 U of enzyme, 1x first strand buffer, 2.5 mM of each dNTP (dATP, dCTP, dGTP, dTTP), and 5U of RNase inhibitor. Reverse transcriptase was omitted in negative-control tubes. After 1 h the reaction mixture was heated at 94 °C for 5 min and subsequently stored at -20 °C. The cDNA or the negative-control reaction (3 μl) were used as template in the PCR. In addition, reactions using either HI689 chromosomal DNA or ddH_20 as template were used as positive and negative controls respectively. Three sets of primers were designed for PCR (Table 2). Primer pair 1 comprised HFJ2F4 and HFJ16, primer pair 2 comprised HFJ25 and HFJCTERMS, and primer pair 3 comprised HFJ26 and HFJ29. 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 sec; 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 sec; 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 sec. The reactions were analyzed on a 2% agarose gel.

Growth studies. Plate bioassays of heme source utilization were performed on heme-deplete *H. influenzae* strains as previously described (37).

Dot blot assay for hemoglobin-haptoglobin binding. The hemoglobin-haptoglobin binding dot blot assay was performed as previously described(16). Haptoglobin was complexed with biotinylated hemoglobin as previously described (37). Complex at a concentration equivalent to 500 ng ml^{-1} of hemoglobin was used in the dot blot assay.

Affinity isolation of hemoglobin binding proteins. Human hemoglobin (Sigma, St. Louis, Missouri) was biotinylated as described previously (10). Outer membrane proteins were subjected to hemoglobin affinity purification as previously described (16). Eluted proteins were separated by SDS-PAGE on 7.5 % acrylamide gels using the discontinuous buffer system of Laemmli (18). Approximately 30 µl of protein preparation, representing proteins isolated from equal numbers of organisms, were 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 hemoglobin binding activity of *H. influenzae* (16), 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 weight, after cleavage of the proposed leader sequence, of 120,296 Da, which is consistent with the size of affinity purified protein. The predicted

hemoglobin binding protein (HgpA) encoded by the gene *hgpA* showed significant homology with other bacterial iron and heme related outer membrane proteins, particularly across the seven regions considered typical of Ton-B dependent outer membrane proteins (22) (Fig. 11). The nucleotide sequence has been deposited in Genbank with the Accession No. U51922 (Fig. 12).

One particularly interesting feature of *hgpA* is the length of CCAA repeating units directly following the nucleotide sequence encoding the putative leader peptide. We propose that strand slippage across the CCAA region may be a mechanism of regulation, analogous to the strand slippage reported across the CAAT repeat region of the *H. influenzae lic* locus (38). In the *hgpA* clone a full-length protein is encoded from the gene; however, addition or loss of one or two CCAA repeat units would result in stop codons downstream of the CCAA repeat region (Fig. 13). By screening of a second library of HI689 chromosomal DNA, the 1.3 kbp 5' portion of *hgpA* from strain HI689 was cloned (Fig. 10). One such clone contained one less CCAA repeat than the original clone, pHFJ2. In this clone, a stop codon immediately followed the CCAA region. The existence of a clone with a different number of CCAA repeats supports the theory of strand slippage. However it is possible that the difference in CCAA length is an artifact occurring in the *E. coli* host strain. To determine if variation in length over the CCAA region occurs in *H. influenzae*, PCR reactions were performed on chromosome DNA derived from 3 single colonies of a *H. influenzae* strain HI1715, which contains a single

region of CCAA repeats. In two of the colonies the size of the CCAA region was the same; whilst the size of the CCAA region in the third colony was different (Fig. 14). To investigate the possibility that the change in CCAA lengths occurred as a result of the PCR reactions, repeated amplifications were performed across a cloned CCAA region. The plasmid containing the cloned CCAA region was used as the template in these reactions and no variation in the length of the PCR product was seen in repeated experiments (data not shown). In addition Van Belkum *et al.* have performed PCR reactions across areas of tetranucleotide repeats in *H. influenzae* and have seen no variation in length of the repeating regions as a result of PCR (40). These data indicate that alteration in the length of the CCAA repeat region occurs in *H. influenzae*.

A gene (*hhuA*) involved in hemoglobin-haptoglobin utilization by nontypeable *H. influenzae* TN106 was recently reported by Maciver *et al.* (23). HhuA is homologous to HgpA, showing 84% identity and 90% similarity. However, HhuA bound only the hemoglobin-haptoglobin complex, and not hemoglobin. To determine whether HgpA binds the hemoglobin-haptoglobin complex, a dot blot assay was performed. The results demonstrated that HgpA bound the hemoglobin-haptoglobin complex (Fig.15), in addition to the previously demonstrated hemoglobin binding (16).

Figure 11. Comparison of the HgpA amino acid sequence from *H. influenzae* strain HI689 with the sequences of other bacterial iron and heme related proteins over the 7 regions considered indicative of a TonB-dependent outer membrane protein (Lundrigan and Kadner, 1986). Residues within these regions which are 100% conserved between the five proteins are marked with an asterisk. The proteins identified in the genbank database as highly homologous to *hgpA* are: HgbA, a hemoglobin binding protein of *H. ducreyi* (Elkins *et al.*,1995); HmbR, a hemoglobin binding protein of *Neisseria meningitidis* (Stojiljkovic *et al.*,1995); Tbp1, a transferrin binding protein of *H. influenzae* (Gray-Owen and Schryvers, 1995), and Tbp1-Nm, a transferrin binding protein of *N. meningitidis* (Legrain *et al.*,1993).

```
          *                                     **    *   **   **    **
HgpA 78 EQINVSGSTENTD 90     HgpA  122 ETGITVVEAGRFGNSGFAVRGVEENRVAVQIDGLHQA 158
HgbA    ETIVVSSEDDSVH        HgbA      ETGVTVVEKGRFGSSGYAIRGVDENRVAVVVDGLHQA
HmbR    ETTPVKAEVKAVR        HmbR      STDVGLSDSGRHQK.GFAVRGVEGNRVGVSIDGVNLP
Tbp1    ETISVTAEKVRDR        Tbp1      DPGISVVEQGRGASSGYSIRGMDRNRVALLVDGLPQT
Tbp1-Nm DTIQVKAKKQKTR        Tbp1-Nm   DPGIAVVEQGRGASSGYSIRGMDKNRVSLTVDGVSQI

            * **       * *** * *                  *
HgpA 189 QVTIRKGADSLKSGSGALGGSVSLDTKD 217   HgpA 838 LQPETAKTKELSLT 851
HgbA     QAVIQKGADAIRTGSGSLGGTVSFESKD       HgbA     LKAERSKNKEVAIT
HmbR     NIDIVKGADSFNTGSGALGGGVNYQTLQ       HmbR     LKAERTTTHTLSLQ
Tbp1     AVEISKGGSSSEYGNGALAGSVTFQSKS       Tbp1     FKPETSRNQEFGLA
Tbp1-Nm  AVEISKGSNSSEYGNGALAGSVAFQTKT       Tbp1-Nm  LKPEKSFNREAGIV

           *   *                          * ****  *  *
HgpA 900 PNAKVTGFE 908         HgpA 1020 LRAGVYNLTNRKY 1033
HgbA     DKARVKGLE             HgbA      LRAGVYNLTNREY
HmbR     DKARIRGIE             HmbR      LRAGVYNVFNRKY
Tbp1     QNAKLVGVN             Tbp1      FRLGVYNLLNYRY
Tbp1-Nm  QNARIAGIN             Tbp1-Nm   LRAGVYNLLNYRY

           **        *
HgpA 1066 ERNYRMSVQFEF* 1077
HgbA      GRNFRVNAEITF*
HmbR      SRNYAVSLEWKF*
Tbp1      GRNYTLTLEMKF*
Tbp1-Nm   GRNYTFSLEMKF*
```

Figure 12. Complete nucleotide sequence of *hgpA* and the deduced amino acid sequences of predicted ORF. While the sequences of both strands were determined, only that of the 5' to 3' top strand is shown. The stop codon is marked with an asterisk. The putative -35 region, -10 region, and ribosomal binding site (S.D.) are underlined. The pair of direct repeats are underlined, and the repeat direction is indicated by arrows. The dash lines indicate proposed rho-independent terminator.

```
   1  GATCCGAACACAACTATCCCATTCTCAGCTCGTCGCAAATTAATCTTAACGCTTTTCTGTATCTCATTCCCTATTATGATTTGGGTGTAA

  91  TGATTGGCGGATGGTGGTTCCCTAAATGGCTGATCTTTCCTTGCCATTACCATCATCATTATGTTTATTAGTGGATTGTCTGAAAAAGAT

 181  GTTGTGGAATCTTTCACTGAAGGTGCATCAGAATTAGTAGGCGTATCTTAATCATCGGTCTTGCTCGTGGCGTAAACTTAGTACTCGAAC

 271  AAGGTATGATTTCTGACACTATCCTGATTATATGTCTAATGTGGTTAGGGTATGCCAGGTAGCGTATTCATCTTAGGCAATTAGTCGGAT

 361  TTATTTTCCTAGGTTTATCCGTACCATCTTCTTCTGGTTACAGCTACTTCCACTGACAATTATGGCGCCACTTGCTGACTCAGTGGGTAT

 451  TCCACGCGATATCGTGGTTTCCGCTTACAACTGGGGACAATATGCAATGCTATTCTTGGCTCCGACAGGATTAGTTTGGGTGACACTCCA

 541  AATGTTGCATATTCCATTTGATAGATGGGTTAAATTCGTCATGCCAATGATTGGATGCTTATTGCTAATTGGTTCCATTTTATTGGTAGT

 631  ACAAGTATCTTTATATAGTGTTTAATTAGtcaaaaatacctatcttaattctaggcttggcattaaaaccttaaaatcagaagttactct
            end of HI0594 ——▶ *
 721  aaaaacggtcaaatcaacaactgagttgcctttatagaatcatcataatgattgctcctaaaataaaaagaagtacgcatagaatgatgc

 811  atcaacgtcgagtggacttcagcccaccaaaatatctcacgattgtttgttggtgggctaaagcccagcctacaactactaaagtccacc
                                                       DR1 ——▶
 901  aaatatctcacgattgtttgttggtgggctgaagcccaccctacaacgactaatatcaacaaccaacggctcgcattataggaaacaagc
                   DR2 ——▶                                                  -35
 991  aggtttatctacagaaaataataagaatagtttctatttcttgattttatcaaaatcaatctttatttttgtgagaaaatacttcctc
       -10                         S.D.
1081  gtataatatatttgagaattattattatttgttataggactaaatatatgaccaatttagattaaacgtgcttgcctattccgttatg
                                                      M  T  N  F  R  L  N  V  L  A  Y  S  V  M

1171  cttgggctaacggcaagtgttgcttatgctgaaccaaccaaccaaccaaccaaccaaccaaccaaccaaccaaccaaccaaccaaccaac
      L  G  L  T  A  S  V  A  Y  A  E  P  T  N  Q  P  T  N  Q  P  T  N  Q  P  T  N  Q  P  T  N

1261  caaccaaccaaccaaccaaccaaccaaccaaccaaccaaccaaccaaccaaccaaccaaccaaccaaccaaccaaaatagtaatgcttct
      Q  P  T  N  Q  P  T  N  Q  P  T  N  Q  P  T  N  Q  P  T  N  Q  P  T  N  Q  N  S  N  A  S

1351  gaacaactagaacaaataaatgtttctggctctaccgaaaatactgatacgaaagctccgccaaaaattgccgaaactgtaaaaacagct
      E  Q  L  E  Q  I  N  V  S  G  S  T  E  N  T  D  T  K  A  P  P  K  I  A  E  T  V  K  T  A

1441  aaaaaattagaaaaagaacaagcacaagatgttaaagatctcgtgcgttatgaaaccggaattactgtcgtagaagcaggacgctttgga
      K  K  L  E  K  E  Q  A  Q  D  V  K  D  L  V  R  Y  E  T  G  I  T  V  V  E  A  G  R  F  G

1531  aatagtggctttgcagttcgaggagtggaagaaaaccgtgtagccgttcaaatagatggacttcatcaagcggaaactatctcttcacaa
      N  S  G  F  A  V  R  G  V  E  E  N  R  V  A  V  Q  I  D  G  L  H  Q  A  E  T  I  S  S  Q

1621  gggtttaaagaattatttgaaggatatggaaatttaataatacgcgtaatagcgcagaaatagaaacgctaaaacaagttacaattcga
      G  F  K  E  L  F  E  G  Y  G  N  F  N  N  T  R  N  S  A  E  I  E  T  L  K  Q  V  T  I  R
```

```
1711  aaaggagctgattctttaaaatctggtagtggtgcattaggtggttctgttagtttggatacaaaagatgctagagattatttacttaac
       K  G  A  D  S  L  K  S  G  S  G  A  L  G  G  S  V  S  L  D  T  K  D  A  R  D  Y  L  L  N

1801  aaaaattactacgcttcctataaaagaggctataacacagcagataaccaaaatctcaatacactaacgcttggcggtcgctataaatat
       K  N  Y  Y  A  S  Y  K  R  G  Y  N  T  A  D  N  Q  N  L  N  T  L  T  L  G  G  R  Y  K  Y

1891  tttgacgcaattgctgttcttacatcacgcaaagggcatgaattagagaactttggttataaaaattataacgataaaattcaaggaaaa
       F  D  A  I  A  V  L  T  S  R  K  G  H  E  L  E  N  F  G  Y  K  N  Y  N  D  K  I  Q  G  K

1981  acgagagaaaaagcagatccttatagaagaacccaagatagtgcactttaaaaattggtttccaaccaacagaaaatcaccgcttctca
       T  R  E  K  A  D  P  Y  R  R  T  Q  D  S  A  L  L  K  I  G  F  Q  P  T  E  N  H  R  F  S

2071  gttgttgcagatttatataaacaaacttctaaaggtcatgactttcttatactctaaaaccgaatacacagtacatgacatatgatgaa
       V  V  A  D  L  Y  K  Q  T  S  K  G  H  D  F  S  Y  T  L  K  P  N  T  Q  Y  M  T  Y  D  E

2161  aaagaattacgtcatactaacgataaagtagaacgtaaaaatatcgctttgtttatgagaattttactgaaacaccatttgggatacg
       K  E  L  R  H  T  N  D  K  V  E  R  K  N  I  A  F  V  Y  E  N  F  T  E  T  P  F  W  D  T

2251  ttaaaaatcacctattcccaccaaaaaattactacaagcgcaagaacagacgattattgcgatgggaatgacaaatgtgccttggcggga
       L  K  I  T  Y  S  H  Q  K  I  T  T  S  A  R  T  D  D  Y  C  D  G  N  D  K  C  A  L  A  G

2341  aatccgcttggaatgaaatacaatcaggataatcaacttgttgggaaagatggtaagtcagctaagtatcaagatataaataaaacacaa
       N  P  L  G  M  K  Y  N  Q  D  N  Q  L  V  G  K  D  G  K  S  A  K  Y  Q  D  I  N  K  T  Q

2431  gttattaaagaaagactacctttttactaaaccaaatggaagatggagattccacaaagttgattgggatgcactcaagaaaaaataccca
       V  I  K  E  R  L  P  F  T  K  P  N  G  R  W  R  F  H  K  V  D  W  D  A  L  K  K  K  Y  P

2521  ggcgttccgatctatgcttcttgtttagaagaagataatgatcctagtgagtttgtacttatgaagtaaaaaccacaaaaaaagaaaac
       G  V  P  I  Y  A  S  C  L  E  E  D  N  D  P  S  E  F  C  T  Y  E  V  K  T  T  K  K  E  N

2611  acctttgaaattaatggtaaaagatacgatctttatctgaagcagataaaaacgttatctctgatgaacaaagattgccaactaacgtc
       T  F  E  I  N  G  K  R  Y  D  L  L  S  E  A  D  K  N  V  I  S  D  E  Q  R  L  P  T  N  V

2701  agctatctttttagttgtgatggcttaaattgcgataagaaaacaattctgggttttaaaaaaaggcggaacctactgaaaatattcctt
       S  Y  L  F  S  C  D  G  L  N  C  D  K  K  T  I  L  G  F  K  K  R  R  N  L  L  K  I  F  L

2791  tttgaggtaattgaaaaaagatgccaaaaatatggcaaaacaaaagtaaaagctaatgatcagctatctggtccatacctatttatgcca
       F  E  V  I  E  K  R  C  Q  K  Y  G  K  T  K  V  K  A  N  D  Q  L  S  G  P  Y  L  F  M  P

2881  aataaaaagggttatcaggcgaacctttggtcgcaacgtgacttaacaagtgaaaccaaacaaattaaccttgacttaacaaaacatcta
       N  K  K  G  Y  Q  A  N  L  W  S  Q  R  D  L  T  S  E  T  K  Q  I  N  L  D  L  T  K  H  L

2971  gaattaggtaaaacacagcacgatttgtcttatggtggtttatggtctgaaatggaaaaatcaatgaccaatctagctggagatactcct
       E  L  G  K  T  Q  H  D  L  S  Y  G  G  L  W  S  E  M  E  K  S  M  T  N  L  A  G  D  T  P

3061  ctaaatgtaaaatggtgggctcaatatccacataattgtgcgacctttttgccaccttcaacaatgactcctaatgcgaaaccaacgtta
       L  N  V  K  W  W  A  Q  Y  P  H  N  C  A  T  F  L  P  P  S  T  M  T  P  N  A  K  P  T  L

3151  aatccagaacggacgagtacattatgtaataatgtcaatgtgttctctttcctcattccagtgaaaacgaaaacaggcgcgttatatttt
       N  P  E  R  T  S  T  L  C  N  N  V  N  V  F  S  F  L  I  P  V  K  T  K  T  G  A  L  Y  F

3241  attaatgatttccgtgtgaataactatgtagcttttaatttaggttatcgttatgatcgagtgaaatatgaaccagaatatattcctgga
       I  N  D  F  R  V  N  N  Y  V  A  F  N  L  G  Y  R  Y  D  R  V  K  Y  E  P  E  Y  I  P  G

3331  aaaacaccaaaaattcctgatgatatggtgacaaatctttatataaaaacgccagaatttgatgcaagtaaagcagattcagatcctgat
       K  T  P  K  I  P  D  D  M  V  T  N  L  Y  I  K  T  P  E  F  D  A  S  K  A  D  S  D  P  D

3421  gaattatcaaaaaaagaagctaatgctgcagcaaatattaaagaaattgcacaaccgaaaaaatttctgcaagttcctactcttttggt
       E  L  S  K  K  E  A  N  A  A  A  N  I  K  E  I  A  Q  P  K  K  F  S  A  S  S  Y  S  F  G

3511  acaacacttgatccgctaaattggttacgtttacaggctaaatatagtaagggattccgagcaccaacaagtgatgaaatctactttaca
       T  T  L  D  P  L  N  W  L  R  L  Q  A  K  Y  S  K  G  F  R  A  P  T  S  D  E  I  Y  F  T

3601  ttcaaacatccagattttctattcaaccaaatagagatcttcaaccagaaacagcaaaaaccaaagagttatctttaactgtgcataat
       F  K  H  P  D  F  S  I  Q  P  N  R  D  L  Q  P  E  T  A  K  T  K  E  L  S  L  T  V  H  N

3691  gacatgggatatattacaacttctgttttgatacccgatatcaaaactttattgatttatcctatcaagggcgtcgtgatgttcatgga
       D  M  G  Y  I  T  T  S  V  F  D  T  R  Y  Q  N  F  I  D  L  S  Y  Q  G  R  R  D  V  H  G

3781  cactcaaaactaataccatttcatttttatcaaaatgtaaatagaccaaatgctaaagtaactggttttgaaattgcttcacaaatatcc
       H  S  K  L  I  P  F  H  F  Y  Q  N  V  N  R  P  N  A  K  V  T  G  F  E  I  A  S  Q  I  S
```

```
3871  ttaggaaatattactaaattattcaatggtttcagtttaagctataaatacacctatcaaaaaggcagaataaatggcaatataccaatg
      L  G  N  I  T  K  L  F  N  G  F  S  L  S  Y  K  Y  T  Y  Q  K  G  R  I  N  G  N  I  P  M

3961  aatgcaattcagcctagaacagctgtttatggagtaagttatgttcaccctgatgataaatatggtctagatctttatatctctcatgca
      N  A  I  Q  P  R  T  A  V  Y  G  V  S  Y  V  H  P  D  D  K  Y  G  L  D  L  Y  I  S  H  A

4051  tcagctaaaaatgcagaagatacttataatatgttctataaagaagaaggtaaaacagacagcacaattaaatggagaagcaaatcttat
      S  A  K  N  A  E  D  T  Y  N  M  F  Y  K  E  E  G  K  T  D  S  T  I  K  W  R  S  K  S  Y

4141  accaccattgatttgttgggatatatcaaaccaattaaaaatcttactttaagagctggtgtgtataatttaactaatagaaaatatatt
      T  T  I  D  L  L  G  Y  I  K  P  I  K  N  L  T  L  R  A  G  V  Y  N  L  T  N  R  K  Y  I

4231  acttgggattcagcgagatcaattcgaccatttggtacaagtaatatgattaatcaagatactggattaggtatcaaccgtttctatgca
      T  W  D  S  A  R  S  I  R  P  F  G  T  S  N  M  I  N  Q  D  T  G  L  G  I  N  R  F  Y  A

4321  cctgaacgaaactatagaatgtcagttcagtttgaattctaatctgttgtaaaaataaaatacgaggatttcctgtttagattgagtttg
      P  E  R  N  Y  R  M  S  V  Q  F  E  F  *

4411  acactcatatactaaacaggaaatgctcatttttattgtgaagttttatttttcacatagttttgtatgtagataatctttccttact

4501  aagtgcttttctaaaacaagaaattgcgttgctgataacatataatcctacttatcacttgacatttgaattatttaccatagactcc

4591  gccccgacgatagggaacgttggtcactgatagcattatcaaccccctacaatattgagacaaaagccacctaaaggtggcttttgtc
                                                                      Terminator
4681  atttttaattcctgaaaagcatcgttttcaactaaatcaataatgctacgcaatccataaaaatccagcttttatgctggatttgtttt

4771  tagcttaggcataatagcaataacttttcatccctaaaaaggattataaatATGAAAAATATGTTACTGCTCAGTAGCTCAAAATACAAG
                                                 start of HI0587
4861  AACACAGGTTATCTTGAACACACTCTACCTTGGTTACAAAATTTCTTGGCGGATTATCGTGGTAAAACAATTGCTTTTGTGCCTTATGCT

4951  GGTGTACGCCGCACTTTTGATGAATATGAAAAAACTGTGCAAAATGCACTTTCCGATTTAGGGATGGATATTGTTTCCGTTCATCATGGT

5041  AAACAGCATCGAGAAATCATTGAACAAGCAGACGTCATCGCTATAGGCGGTGGGAATACATTTTGTTTGTTAAAACAGCTTTATGAACAT

5131  AATTTGATTGAT
```

Figure 13. Amino acid sequence of the N-terminal region of HgpA showing 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.

```
  1  atgaccaatttagattaaacgtgcttgcctattccgttatgcttgggctaacggcaagt   60
     M  T  N  F  R  L  N  V  L  A  Y  S  V  M  L  G  L  T  A  S

     M  T  N  F  R  L  N  V  L  A  Y  S  V  M  L  G  L  T  A  S

     M  T  N  F  R  L  N  V  L  A  Y  S  V  M  L  G  L  T  A  S

                ↓
 61  gttgcttatgctgaaccaaccaaccaaccaaccaaccaaccaaccaaccaaccaaccaac  120
     V  A  Y  A  E  P  T  N  Q  P  T  N  Q  P  T  N  Q  P  T  N

     V  A  Y  A  E  P  T  N  Q  P  T  N  Q  P  T  N  Q  P  T  N

     V  A  Y  A  E  P  T  N  Q  P  T  N  Q  P  T  N  Q  P  T  N

121  caaccaaccaaccaaccaaccaaccaaccaaccaaccaaccaaccaaccaaccaaccaac  180
     Q  P  T  N  Q  P  T  N  Q  P  T  N  Q  P  T  N  Q  P  T  N

     Q  P  T  N  Q  P  T  N  Q  P  T  N  Q  P  T  N  Q  P  T  N

     Q  P  T  N  Q  P  T  N  Q  P  T  N  Q  P  T  N  Q  P  T  N

181  caaccaaccaaccaaccaaccaaccaaaatagtaatgcttctgaacaactagaacaaata  240
     Q  P  T  N  Q  P  T  N  Q  N  S  N  A  S  E  Q  L  E  Q  I

     Q  P  T  N  Q  P  T  k  _     I  V  M  L  L  N  N  * -1 CCAA

     Q  P  T  N  Q  P  K  _  _  _ * -2 CCAA
```

Figure 14. 6% denaturing PAGE gel separating PCR products derived from amplification across the CCAA-containing locus of *H. influenzae* type b strain HI1715. Each lane contains the PCR product derived from a single independent colony of the same strain. Numbers to the left are sizes in bases derived from a sequencing reaction run in lanes adjacent to the PCR products.

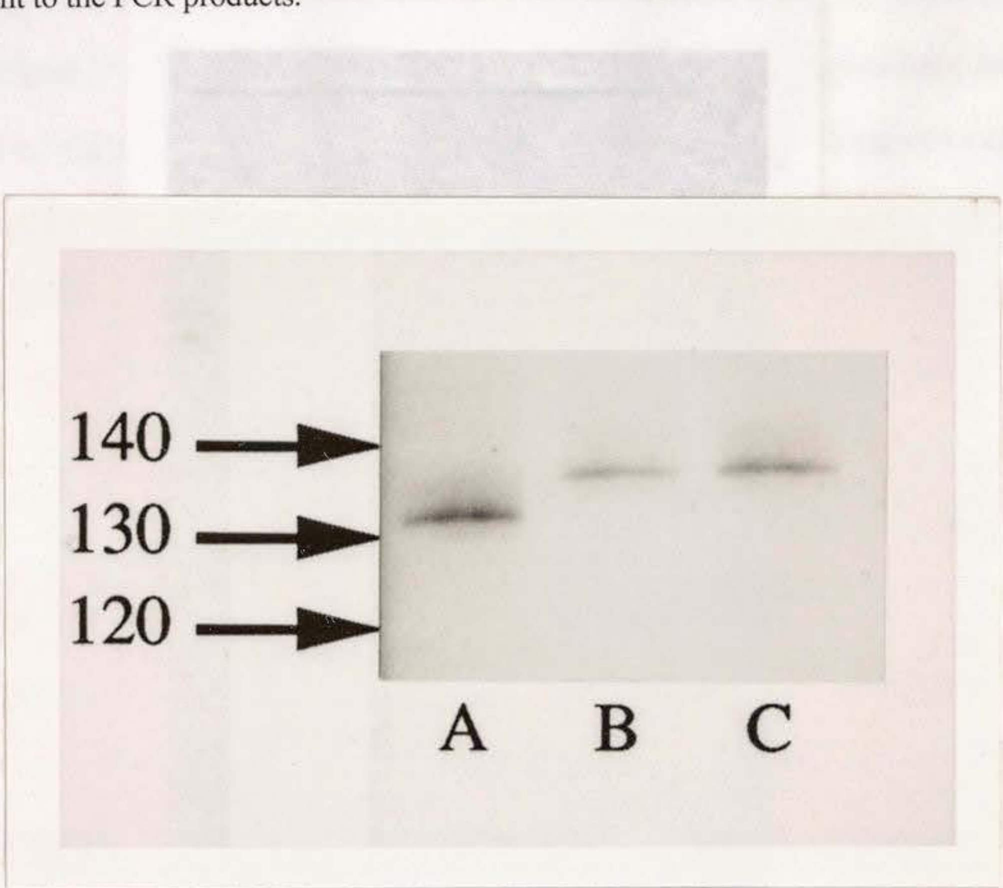
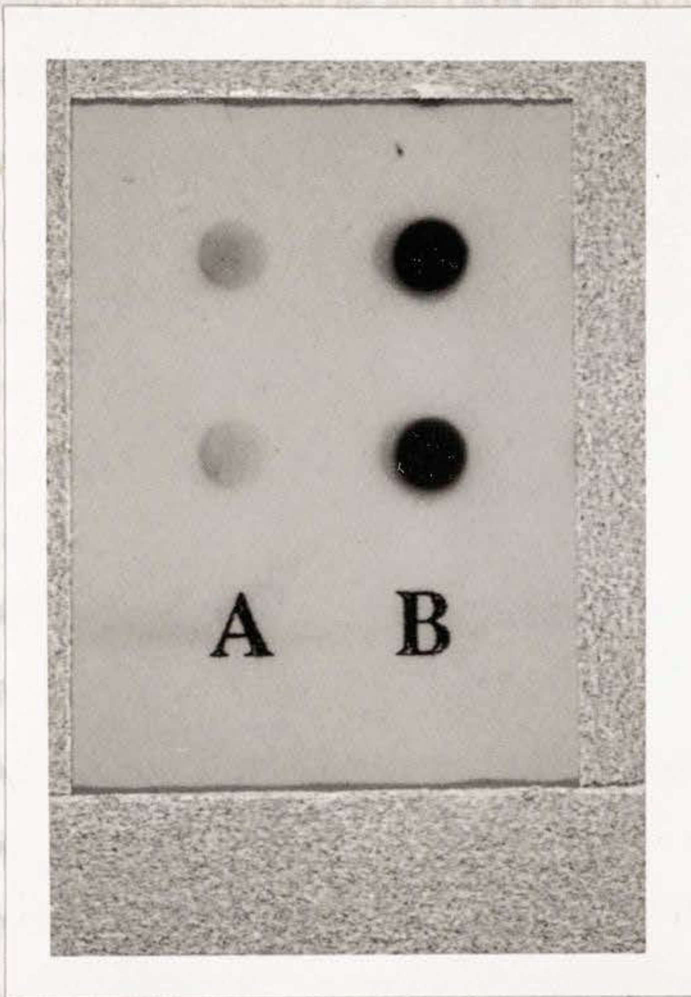


Figure 15. Whole cell hemoglobin-haptoglobin binding dot blot assay. Column A, duplicates of *E. coli* BL21(DE3)pLysS harboring vector pRESETA; Column B, duplicates of sonicates of *E. coli* BL21(DE3)PlysS harboring HgpA on the expression vector pRESTA.



The genomic sequence of *H. influenzae* strain Rd KW20 reported by Fleischmann *et al.* contains four putative ORFs (HI0635, HI0661, HI0712, and HI1566) which have significant homology with *hgpA* (9). Their common feature is a length of CCAA repeats following the leader peptide coding sequence (Table 3). 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 ORF HI0594 from the Rd KW20 genome and a direct repeat of 49 bp present in Rd KW20 between the ORFs HI0594 and HI0592 (Fig.12) ,(9). However, there is no homologue of *hgpA* at this site in the Rd KW20 genome (Fig. 16). The nucleotide sequence immediately downstream of *hgpA* is highly homologous to ORF HI0587, which is approximately 6.6 kbp downstream of ORF HI0594 in the Rd KW20 genome (Fig. 16). In Rd KW20 this 6.6 kbp sequence between HI0594 and HI0587 contains five ORFs designated HI0588 through HI0592. Using a clone (GHIFY44) containing portions of the ORFs HI0591 and HI0592 of RdKW20 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. 16). Similar results were obtained using probes encompassing the ORFs HI0588, HI0589, and HI0590 (data not shown). These data suggest that in the strains HI689 and TN106 a deletion may have occurred with the insertion of *hgpA* into the genome, and that the function of the deleted ORFs are apparently not essential for the survival of HI689 and TN106 (Table 4).

Figure 16. a. Comparison of the organization of the strain HI689 at the *hgpA* locus (top) with the organization of Rd KW20 genome (bottom) based on nucleotide sequence analysis, showing that *hgpA* is not present at this locus in the Rd KW20 genome. HI0588, HI0589, HI0592 and HI0594 are ORFs from the Rd KW20 genome with no assigned function (Fleischmann *et al.*,1995). b. Southern blot probed with the labeled clone GHIFY44. Lane A, labeled λ *Hind*III digest; size of DNA fragments top to bottom are 23 kbp, 9.5 kbp, 6.5 kbp, 4.4 kbp, 2.3 kbp, and 2 kbp. Lane B, *H. influenzae* Rd KW20 genomic DNA digested with *Eco*RI. Lane C, *H. influenzae* HI689 genomic DNA digested with *Eco*RI. Lane D, *H. influenzae* TN106 genomic DNA digested with *Eco*RI. Lane E, *H. influenzae* Rd genomic DNA digested with *Eco*RI.

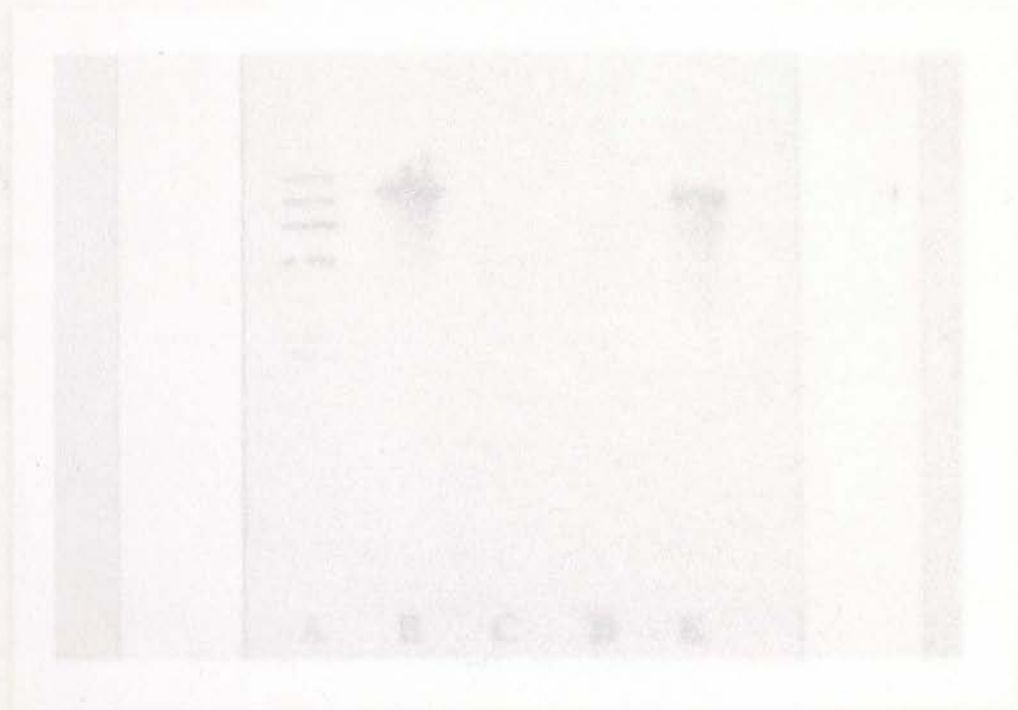


Figure 16a.

***Haemophilus influenzae* HI689**

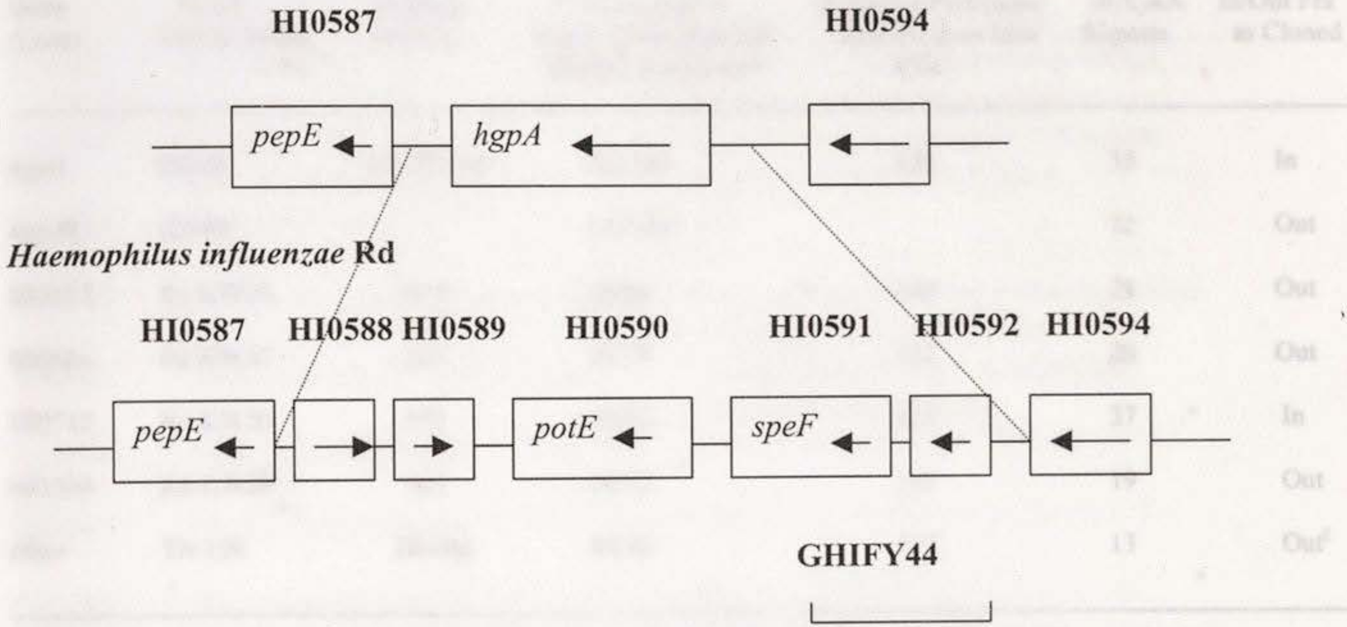


Figure 16b.

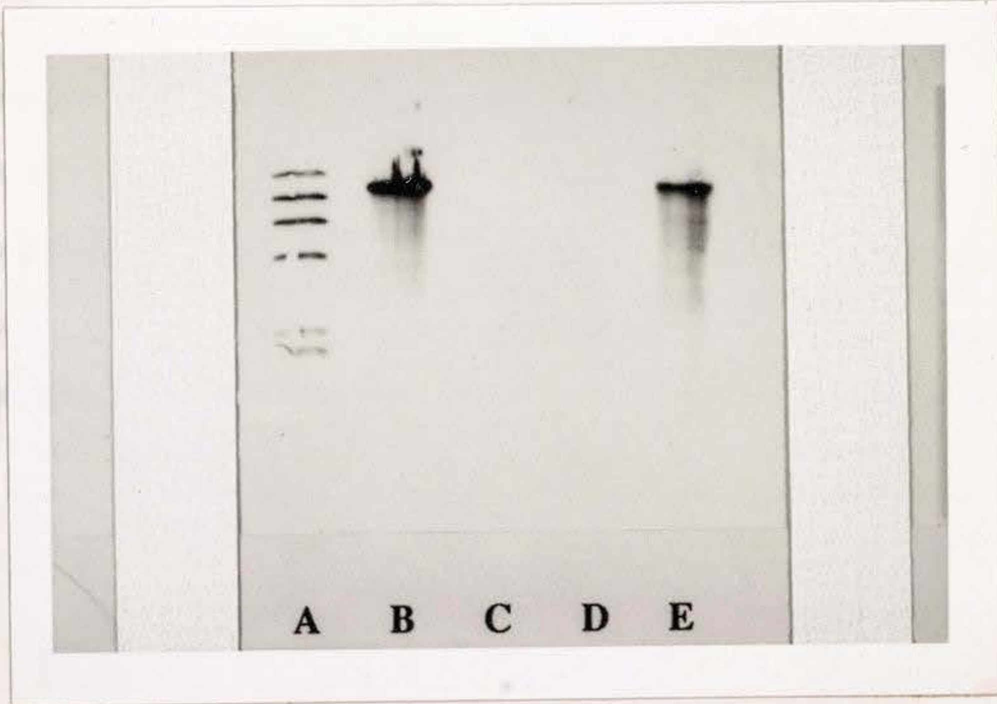


Table 3. Comparison of HgpA with the proteins encoded by ORFs HI0635, HI0661, HI0712, and HI1566 in strain Rd and with HhuA.

Gene /Locus	Strain Cloned From	Binding Activity	% Homology to HgpA from strain 689 Identity/Similarity*	Predicted Prototypic Gene Product Size kDa†	#CCAA Repeats	In/Out Fra as Cloned
hgpA	HI689	Hb/Hb-Hp‡	100/100	120	33	In
hgpA¶	HI689		100/100		32	Out
HI0635	Rd KW20	ND§	49/65	119	21	Out
HI0661	Rd KW20	ND	59/74	112	20	Out
HI0712	Rd KW20	ND	49/67	121	37	In
HI1566	Rd KW20	ND	54/72	112	19	Out
hhuA	TN 106	Hb-Hp	84/90	117	13	Out‖

* % 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. % 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.

‡ Hb, hemoglobin; Hb-Hp, the hemoglobin haptoglobin complex.

§ Not determined

‖ Out of frame with the alternative start codon proposed in this paper.

¶ The independent copy of *hgpA*, which has one less CCAA unit.

Table 4. Loci between HI0587 and HI0594, their putative identifications and roles

HI#	Putative Identification	Role
HI0587	Peptidase E	Translation, degradation of proteins, peptides, and glycopeptides.
HI0588	N-carbamyl-L-amino acid amidohydrolase	Translation, degradation of proteins, peptides, and glycopeptides.
HI0589	Hypothetical protein	NA*
HI0590	Putrescine transport protein (potE)	Transport and binding proteins, amino acids, peptides and amines.
HI0591	Ornithine decarboxylase (speF)	Central intermediary metabolism, polyamine biosynthesis
HI0592	Hypothetical protein	NA
HI0593	Hypothetical protein	NA
HI0594	Hypothetical protein	NA

* NA, Not applicable

HI0661 is the ORF in Rd KW20 with greatest homology to *hgpA*; it exhibits 59% identity and 74% similarity (Table 3). We have cloned a homologue of HI0661 from strain HI689 (33) and shown that it binds hemoglobin and the hemoglobin-haptoglobin complex. In addition, regions homologous to the 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.

Analysis of the *hgpA* transcript. Analysis of the sequences flanking *hgpA* indicated the presence of a putative promoter region; the hairpin structure and a polyT tract following the stop codon indicated a rho-independent terminator, suggesting that *hgpA* is transcribed as a monocistronic message. Reverse transcriptase-PCR (RT-PCR) was used to investigate this possibility. Random nonamer primers were annealed to RNA prepared from heme-restricted *H. influenzae* HI689, and reverse transcriptase was used to generate complementary DNA (cDNA). The cDNA was used as the template for PCR reactions with 3 sets of primers (Table 2 and Fig. 17) 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 HFJCTERMS) 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. 17). All primer pairs

amplified a DNA fragment of the correct size when the genomic DNA was used as template.

Construction of a mutation in *hgpA*. Analysis of nucleotide sequence of the clone pHFJ2 revealed the presence of three *Bgl*II restriction sites within the ORF encoding *hgpA*. Following complete digestion of pHFJ2 with *Bgl*II, the 2.5 kbp *Bgl*II fragment was deleted, the remaining 4.5 kbp DNA fragment was ligated to TSTE, an aminoglycoside resistance cassette designed for mutant construction in *H. influenzae* (35). The resulting plasmid pHFJC (Fig. 9), was used to transform *H. influenzae* strain HI689, and recombinants were selected on 15 μg ml^{-1} Rb. One Rb resistant colony was selected for further investigation. Southern hybridization, using labeled TSTE as the probe confirmed insertion of a single copy of the TSTE element into the genome (Fig. 9b, lane D). The labeled insert of pHFJ2 hybridized to an approximately 4.3 kbp *Eco*RI fragment in the wild type strain (Fig. 9b, lane C) and to a band of slightly smaller size in the mutant (Fig. 9b, lane B). The reduction in size and intensity of the hybridizing band was consistent with deletion of 2553 bp and insertion of the 2222 bp TSTE element, resulting in an overall size reduction of 330 bp (Fig. 9b). The mutant strain was designated HI689*hgpA*Δ*Bgl*II.

Heme utilization and binding. The mutant strain HI689*hgp*Δ*Bgl*II was unaltered in the ability to grow with either hemoglobin or the hemoglobin-haptoglobin complex as the sole source of heme on plate bioassays (Fig. 18). Following growth in

heme-starved conditions the mutant was not altered in the ability to bind either biotinylated hemoglobin or labeled hemoglobin-haptoglobin complex in a dot blot assay compared to the wild type strain HI689 (data not shown).

Affinity isolation of hemoglobin binding proteins from an *hgpA* mutant. Since there was no apparent alteration in either utilization or binding of hemoglobin or hemoglobin-haptoglobin, bacteria grown in heme restricted media to mid log phase were subjected to the affinity purification procedure using hemoglobin as the primary ligand. The wild type strain HI689 resulted in isolation of an approximately 120 kDa band, while the mutant strain HI689*hgp*Δ*Bgl*II did not yield the 120 kDa band; however, a band at approximately 115 kDa was present in increased amounts (Fig. 19). 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 (16).

Figure 17. RT-PCR analysis of transcription of *hgpA*. The positions of the 3 primer pairs within the studied locus are shown in the upper panel. Lane 1 is the negative control lacking template DNA. Lanes 2, 3, and 4 use the primer set 1; lanes 5, 6, and 7 use the primer set 2; lanes 8, 9, and 10 use primer set 3. Lanes 2, 5, and 8 use *H. influenzae* HI689 chromosomal DNA as template; lanes 3, 6, and 9 use RNA sample without RT as template; lanes 4, 7, and 10 use RNA sample with RT as template. Lane 11 is 100 bp ladder marker.

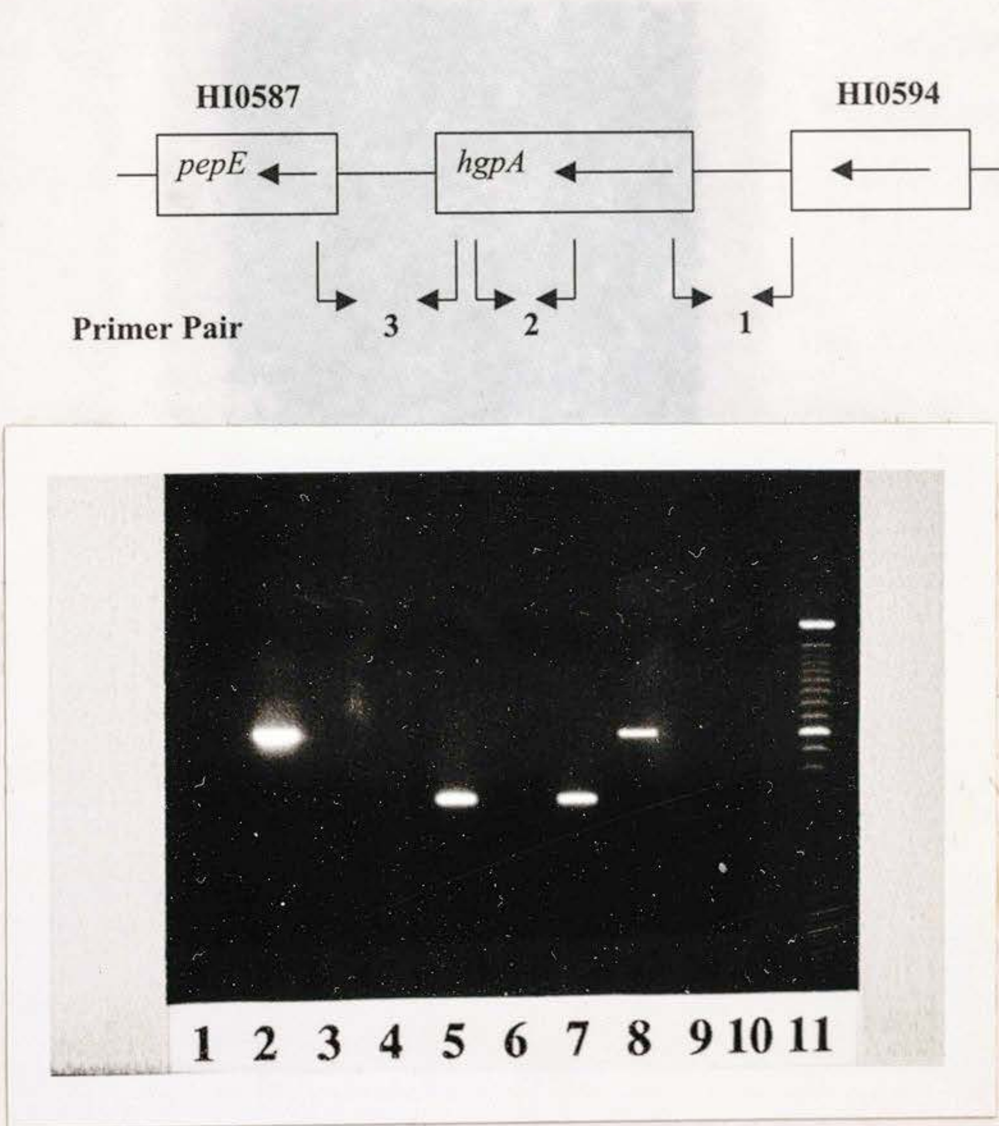


Figure 18. Plate bioassay showing utilization of the hemoglobin-haptoglobin complex (4 nanomoles of hemoglobin equivalent) (well 1), heme (0.04 μmole) (well 2) or hemoglobin (4 nanomole) (well 3) by *H. influenzae* HI689 (panel A) or the insertion/deletion mutant HI689*hgp*Δ*Bgl*II (panel B).

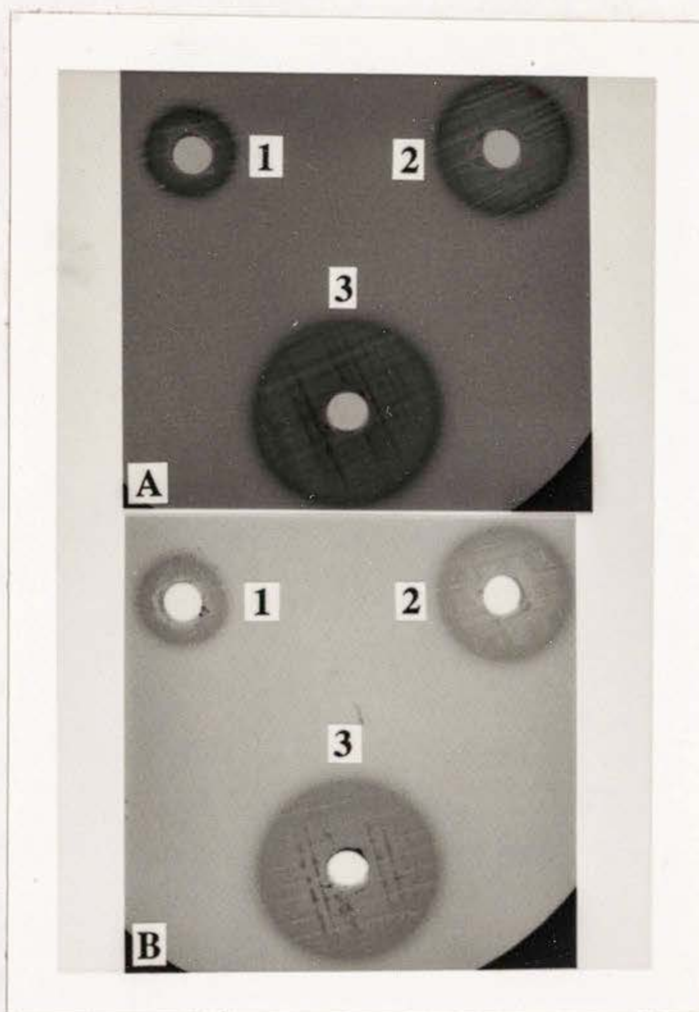
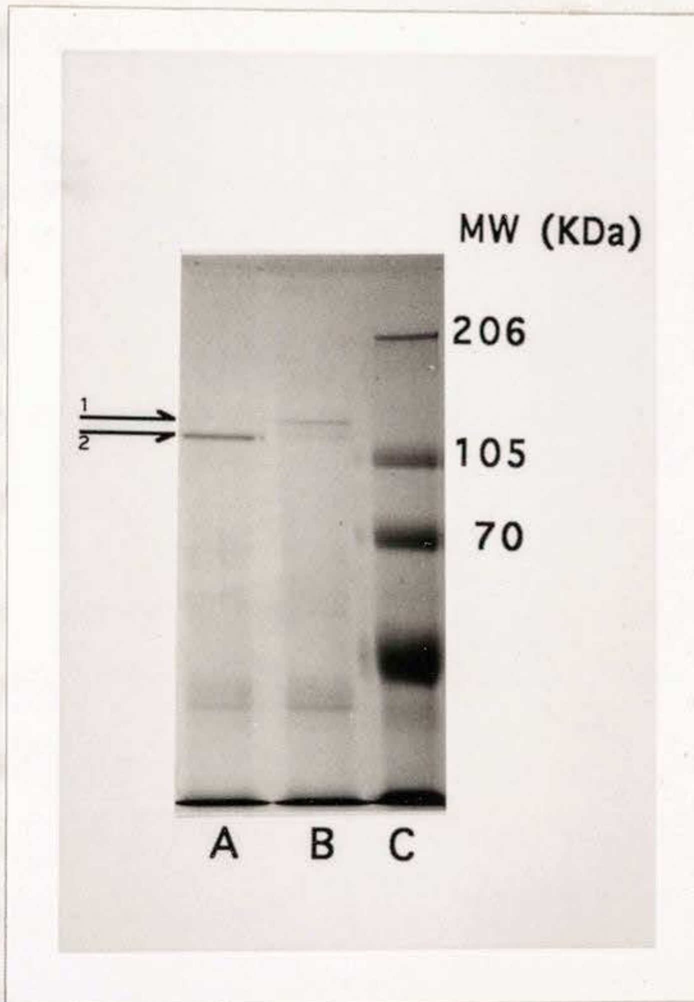


Figure 19. SDS-PAGE (7.5 % acrylamide) gel stained with Coomassie blue. Lane A, affinity-purified outer membrane protein from HI689*hgp*Δ*Bgl*II. Lane B, affinity-purified outer membrane protein from HI689. Lane C, molecular weight markers. Arrow 1 points to a band at approximately 120-kDa in lane B. Arrow 2 points to a band at approximately 115-kDa present in lane A and in lower amounts in lane B. Numbers at right indicate molecular masses.



DISCUSSION

H. influenzae requires a porphyrin source for growth (7). 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 heme and/or iron has not been fully elucidated (25,26,29). Proteins binding the heme-hemopexin complex, one of which is apparently secreted into the growth media, have been described (3,4,13,45). Heme-binding outer membrane protein has been isolated (21), and a heme-binding lipoprotein with significant homology to a periplasmic transport protein of *E. coli* has also been characterized (12). In addition Reidl and Mekalanos have recently shown that protein *e*(P4) of *H. influenzae* is essential for utilization of heme, PPIX, and heme from hemoglobin (32). A short heme binding motif homologous to regions of other heme binding proteins was identified in protein *e*(P4) (32)

The acquisition and utilization of heme from hemoglobin, the hemoglobin-haptoglobin complex, and the heme-hemopexin complex is dependent on a functional *tonB* gene, indicating that uptake is mediated by an outer membrane TonB-dependent protein(s) (15,30). In this study the complete nucleotide sequence of the gene encoding a heme repressible 120-kDa hemoglobin-binding protein (HgpA) of *H. influenzae* type b has been determined. HgpA exhibits significant homology with other iron and heme related bacterial TonB-dependent proteins over regions that are highly conserved among this class of

proteins (22) (Fig. 11). TonB-dependent outer membrane protein receptors have been reported in a number of bacterial species, and are commonly involved in transport of iron and heme across the bacterial membrane (2,5,11,22,23,31). TonB is believed to act in energy transduction from the cytoplasmic membrane to the outer membrane (19,30). HgpA did not contain a region homologous to the heme-binding motif identified in protein *e* (P4) by Reidl and Mekalanos (32).

While RdKW20 chromosome has homology to the DNA sequences flanking *hgpA*, there is no homologous *hgpA* at this locus, which instead contains the ORFs of HI0588, HI059, HI0591, and HI0592 (9). 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. 16).

There are four ORFs in the *H. influenzae* strain Rd KW20 chromosomal sequence, HI0635, HI0661, HI0712, and HI1566, with significant homology to *hgpA*. All four ORFs in Rd KW20 and the corresponding region of *hgpA* contain multiple CCAA repeats, with the number of repeats varying between 19 and 37 (Table 3). In three of the four ORFs in strain Rd KW20 the CCAA repeat is followed by an in-frame stop codon, while the fourth (HI0712) constitutes an ORF of approximately 3 kbp. Our data indicate that the number of CCAA repeats in a given locus varies in *H. influenzae* (Fig. 14). In accordance with the slip-strand hypothesis, we have modified several of the published

sequences to alter the number of CCAA units in our sequence analysis, thus allowing in-frame comparisons of full-length proteins from the reported open reading frames (ORFs).

One of the CCAA containing regions in Rd KW20, locus HI0712, is in frame as sequenced, and alteration in the number of CCAA repeats would lead to introduction of in-frame stop codons.

The second set of CCAA repeats, locus HI0661, was initially reported as a small ORF encoding a protein of 65 amino acids (Genbank Accession No. U32749) (9). The nucleotide sequence was subsequently modified by the authors by inclusion of one additional nucleotide to bring the CCAA region into frame with a downstream gene encoding an apparent iron related protein (Genbank Accession No. U32695). The additional nucleotide was hypothetical, and the ORF generated by the modification did not contain a typical leader peptide sequence. In our analysis of the locus, adding a single CCAA repeat to the original sequence would result in a protein with a leader sequence highly homologous to the leader sequence identified in HgpA.

The third CCAA repeat locus of strain Rd KW20, HI1566, was also modified subsequent to initial publication by the addition of two hypothetical nucleotides to produce a protein lacking a leader peptide (Genbank Accession No. U32777). Without need for the added bases, deletion of one CCAA repeat would produce a protein with a leader peptide homologous to that of HgpA.

The fourth CCAA locus in strain Rd KW20 was not included as an ORF by Fleischmann *et al*. However an ORF (designated HI0635) homologous to bacterial iron related proteins was identified downstream of the CCAA region (9). Removal of a CCAA repeat would result in an ORF with a different start codon, encoding a protein with a leader sequence and significant homology to HgpA (9).

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 heme source. Alternatively, strand slippage may provide a mechanism to avoid the immunological response of the host. An analogous mechanism has been demonstrated to mediate phase variation of lipooligosaccharide (LOS) in *H. influenzae*, with strand slippage across a CAAT repeat motif placing potential initiation codons in or out of frame with the remainder of the ORF (38,42). High *et al.* have recently shown that the CAAT repeat motif is required for phase variation but not for biosynthesis of LOS (14). Similarly, the *hgpA* structural gene without CCAA was expressed in *E. coli* and the recombinant *E. coli* bound to hemoglobin and hemoglobin-haptoglobin complex (data not shown). The hypothesis that slip-strand occurs in *hgpA* is supported by the isolation of an independent copy of this gene from strain HI689 containing a different number of CCAA repeats. In addition, PCR amplification across a CCAA region from different colonies of the same strain yielded varying sized products (Fig.

14). These data provide direct evidence for the proposed slip-strand mechanism in *H. influenzae*.

Insertional inactivation of *hgpA* did not abrogate the ability of the mutant strain to either bind hemoglobin or the hemoglobin-haptoglobin complex or to utilize either as a heme source. Affinity purification of hemoglobin binding proteins from the mutant strain demonstrated the loss of the 120-kDa protein and increase of a 115-kDa protein (Fig.19). These data suggest that additional hemoglobin binding proteins exist. We have shown *that H. influenzae* HI689 contains a HI0661 homologue, *hgpB*, which encodes a 115 kDa hemoglobin and hemoglobin-haptoglobin binding protein. Elkins has also identified a 115-kDa hemoglobin binding protein from *H. influenzae* strain DL42 (6), 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 (16), consistent with the hypothesis of phase variation of hemoglobin and hemoglobin-haptoglobin-binding protein gene 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.

Hemoglobin-haptoglobin binding was expressed by a fusion of *hhuA* lacking the CCAA repeats, indicating that the repeats are not essential for function of the protein (23). The clone of *hgpA* bound both hemoglobin (16) and the hemoglobin-haptoglobin complex

(Fig.15), whereas the clone of *hhuA* bound only the hemoglobin-haptoglobin complex (23). Maciver *et al.* reported a start site immediately preceding the CCAA repeat units which would lead to expression of a protein lacking a leader peptide (23). 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 (Fig. 20).

Mutation of *hhuA* resulted in a decreased ability of the mutant to bind hemoglobin-haptoglobin when grown in a heme-replete medium (23). However, no effect was seen on hemoglobin-haptoglobin binding when bacteria were grown in a heme-restricted medium. The mutant showed a reduced ability to utilize hemoglobin-haptoglobin as a heme source, but was unaltered in the ability to utilize hemoglobin (23). Thus *hhuA* appears to encode constitutive hemoglobin-haptoglobin binding ability, while there is also an independent heme-repressible hemoglobin-haptoglobin binding activity expressed by this strain. Similarly there may be constitutive and heme or iron repressible hemopexin binding proteins expressed by *H. influenzae* (4,44,45). In contrast hemoglobin binding appears to be heme-repressible (10), although no extensive survey has been reported among organisms grown under heme-replete conditions nor have other conditions been investigated.

The similarity between *hgpA* and *hhuA* is 90%, and *H. influenzae* TN106, the strain *hhuA* was cloned from, apparently lacks the same 6.6 kbp sequence missing in HI689 (23) (Fig. 16). Thus it is likely that *hgpA* and *hhuA* represent alleles of the same gene in

Figure 20. Comparison of the amino acid sequence N-terminal region of HgpA from *H. influenzae* strain HI689 with the N-terminal region of HhuA (Maciver et al., 1996) from *H. influenzae* strain TN106 and the N-terminal region of HhuA modified by removal of a CCAA repeat unit from hhuA (mHhuA).

```
        1                                                      50
 Hhua   .......... .......... .......... .......... ..........
mHhua   MTNFRLNLLA YSVMLGLTAG VAYAA..... .......... ..........
 HgpA   MTNFRLNVLA YSVMLGLTAS VAYAE..... PTNQPTNQPT NQPTNQPTNQ

        51                                                    100
 Hhua   ...MQLNQPT NQPTNQPTNQ PTNQDGNVSE QLEQINVLGS DNHNDNTPPK
mHhua   .......QPT NQPTNQPTNQ PTNQDGNVSE QLEQINVLGS DNHNDNTPPK
 HgpA   PTNQPTNQPT NQPTNQPTNQ PTNQNSNASE QLEQINVSGS TENTDTKAPP
```

different strains. However expression of clones of *hhuA* and *hgpA* in *E. coli* apparently resulted in different phenotypes; *hhuA* bound only the hemoglobin-haptoglobin complex and *hgpA* bound both the complex and hemoglobin alone. This discrepancy 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 (23). It is possible that the missing 150 amino acids are essential for hemoglobin binding, or alternatively that the fusion inhibits hemoglobin binding but not binding of the hemoglobin-haptoglobin complex. Additionally, insertion mutants of *hhuA* and *hgpA* in TN106 and HI689 respectively result in different phenotypes with the *hhuA* showing reduced utilization of the hemoglobin-haptoglobin complex, while the *hgpA* mutant is unaltered in this respect. Neither mutant is apparently altered in the ability to bind the hemoglobin-haptoglobin complex following heme-restricted growth. These data may reflect differences in the growth techniques or alternatively may reflect varying complements of hemoglobin and/or hemoglobin-haptoglobin binding proteins between these strains. In order to elucidate the hemoglobin and hemoglobin-haptoglobin utilization pathways *of H. influenzae*, it will be necessary to define the functions of the additional CCAA containing genes and to define the gene complements of the strains under investigation.

In conclusion, a 120-kDa hemoglobin and hemoglobin-haptoglobin binding protein HgpA of *H. influenzae* and several genes possessing CCAA-repeating units at the 5' ends and encoding proteins highly homologous 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.

REFERENCES

1. **Bezkorovainy, A.** 1987. Iron Proteins, p. 27-68. *In* J.J. Bullen and E. Griffiths (ed), Iron and Infection: Molecular, Physiological and Clinical Aspects. John Wiley & Sons, New York, NY.

2. **Bitter, W., J.D. Marugg, L.A. de Weger, J. Tommassen, and P.J. Weisbeek.** 1991. The ferric-pseudobactin receptor PupA of *Pseudomonas putida* WCS358: homology to TonB-dependent *Escherichia coli* receptors and specificity of the protein. Mol. Microbiol. **5:**647-655.

3. **Cope, L.D., S.E. Thomas, J.L. Latimer, C.A. Slaughter, U. Muller-Eberhard, and E.J. Hansen.** 1994. The 100 kDa haem:haemopexin-binding protein of *Haemophilus influenzae*: structure and localization. Mol. Microbiol. **13:**863-873.

4. **Cope, L.D., R. Yogev, U. Muller-Eberhard, and E.J. Hansen.** 1995. A gene cluster involved in the utilization of both free heme and heme:hemopexin by *Haemophilus influenzae* type b. J. Bacteriol. **177:**2644-2653.

5. **Cornelissen, C.N., G.D. Biswas, J. Tsai, D.K. Paruchuri, S.A. Thompson, and P.F. Sparling.** 1992. Gonococcal transferrin-binding protein 1 is required for transferrin utilization and is homologous to TonB-dependent outer membrane receptors. J. Bacteriol. **174:**5788-5797.

6. **Elkins, C., C.J. Chen, and C.E. Thomas.** 1995. Characterization of the hgbA locus encoding a hemoglobin receptor from *Haemophilus ducreyi*. Infect. Immun. **63:**2194-2200.

7. **Evans, N.M., D.D. Smith, and A.J. Wicken.** 1974. Haemin and nicotinamide adenine dinucleotide requirements of *Haemophilus influenzae* and *Haemophilus parainfluenzae*. J. Med. Microbiol. **7:**359-365.

8. **Falla, T.J., S.R. Dobson, D.W. Crook, W.A. Kraak, W.W. Nichols, E.C. Anderson, J.Z. Jordens, M.P. Slack, E. Mayon-White, and E.R. Moxon.** 1993. Population-based study of non-typable *Haemophilus influenzae* invasive disease in children and neonates. Lancet. **341:**851-854.

9. **Fleischmann, R.D., M.D. Adams, O. White, R.A. Clayton, E.F. Kirkness, A.R. Kerlavage, C.J. Bult, J. Tomb, B.A. Dougherty, J.M. Merrick, K.**

McKenney, G. Sutton, W. FitzHugh, C. Fields, J.D. Gocayne, J. Scott, R. Shirley, L. Liu, A. Glodek, J.M. Kelley, J.F. Weidman, C.A. Phillips, T. Spriggs, E. Hedblom, M.D. Cotton, R.C. Utterback, M.C. Hanna, D.T. Nguyen, D.M. Saudek, R.C. Brandon, L.D. Fine, J.L. Fritchman, J.L. Fuhrmann, N.S.M. Geoghagen, C.L. Gnehm, L.A. McDonald, K.V. Small, C.M. Fraser, H.O. Smith, and J.C. Venter. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science. **269:**496-512.

10. **Frangipane, M.E., D.J. Morton, J.A. Wooten, J.M. Pozsgay, and T.L. Stull.** 1994. Binding of human hemoglobin by *Haemophilus influenzae*. FEMS. Microbiol. Lett. **118:**243-248.

11. **Gray-Owen, S.D. and A.B. Schryvers.** 1995. Characterization of transferrin binding proteins 1 and 2 in invasive type b and nontypeable strains of *Haemophilus influenzae*. Infect. Immun. **63:**3809-3815.

12. **Hanson, M.S. and E.J. Hansen.** 1991. Molecular cloning, partial purification, and characterization of a haemin-binding lipoprotein from *Haemophilus influenzae* type b. Mol. Microbiol. **5:**267-278.

13. **Hanson, M.S., S.E. Pelzel, J. Latimer, U. Muller-Eberhard, and E.J. Hansen.** 1992. Identification of a genetic locus of *Haemophilus influenzae* type b necessary for the binding and utilization of heme bound to human hemopexin. Proc. Natl. Acad. Sci. USA. **89:**1973-1977.

14. **High, N.J., M.P. Jennings, and E.R. Moxon.** 1996. Tandem repeats of the tetramer 5'-CAAT-3' present in *lic2A* are required for phase variation but not lipopolysaccharide biosynthesis in *Haemophilus influenzae*. Mol. Microbiol. **20:**165-174.

15. **Jarosik, G.P., J.D. Sanders, L.D. Cope, U. Muller-Eberhard, and E.J. Hansen.** 1994. A functional tonB gene is required for both utilization of heme and virulence expression by *Haemophilus influenzae* type b. Infect. Immun. **62:**2470-2477.

16. **Jin, H., Z. Ren, J.M. Pozsgay, C. Elkins, P.W. Whitby, D.J. Morton, and T.L. Stull.** 1996. Cloning of a DNA fragment encoding a heme-repressible hemoglobin-binding outer membrane protein from *Haemophilus influenzae*. Infect. Immun. **64:**3134-3141.

17. **Klein, J.O.** 1997. Role of nontypeable *Haemophilus influenzae* in pediatric respiratory tract infections. Pediatr. Infect. Dis. J. **16:**S5-S8.

18. **Laemmli, U.K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. **227:**680-685.

19. **Larsen, R.A., G.E. Wood, and K. Postle.** 1994. The conserved proline-rich motif is not essential for energy transduction by *Escherichia coli* TonB protein. Mol. Microbiol. **12:**857

20. **Lee, B.C.** 1995. Quelling the red menace: haem capture by bacteria. Mol. Microbiol. **18:**383-390.

21. **Lee, B.C.** 1992. Isolation of an outer membrane hemin-binding protein of *Haemophilus influenzae* type b. Infect. Immun. **60:**810-816.

22. **Lundrigan, M.D. and R.J. Kadner.** 1986. Nucleotide sequence of the gene for the ferrienterochelin receptor FepA in *Escherichia coli*. Homology among outer membrane receptors that interact with TonB. J. Biol. Chem. **261:**10797-10801.

23. **Maciver, I., J.L. Latimer, H.H. Liem, U. Muller-Eberhard, Z. Hrkal, and E.J. Hansen.** 1996. Identification of an outer membrane protein involved in utilization of hemoglobin-haptoglobin complexes by nontypeable *Haemophilus influenzae*. Infect. Immun. **64:**3703-3712.

24. **Madorne, D.V., C.L. Johnson, D.C. Phipps, Pennridge Pediatric Associates, L.A. Popejoy, R. Eby, and D.H. Smith.** 1990. Safety and immunologic response to *Haemophilus influenzae* type b oligosaccharide-CRM197 conjugate vaccine in 1 to 6-month-old infants. Pediatrics. **85:**331-337.

25. **Morton, D.J. and P. Williams.** 1989. Utilization of transferrin-bound iron by *Haemophilus* species of human and porcine origins. FEMS. Microbiol. Lett. **53:**123-127.

26. **Morton, D.J. and P. Williams.** 1990. Siderophore-independent acquisition of transferrin-bound iron by *Haemophilus influenzae* type b. J. Gen. Microbiol. **136:**927-933.

27. **Murphy, T.V., K.E. White, and P. Pastor.** 1993. Declining incidence of *Haemophilus influenzae* type b disease since introduction of vaccination. JAMA. **269:**246-248.

28. **Otto, B.R., A.M. Verweij-van Vught, and D.M. MacLaren.** 1992. Transferrins and heme-compounds as iron sources for pathogenic bacteria. Crit. Rev. Microbiol. **18:**217-233.

29. **Pidcock, K.A. and T.L. Stull.** 1988. Tn5 transposon mutagenesis of *Pseudomonas cepacia* by an IncW suicide vector. J. Infect. Dis. **157:**1098-1099.

30. **Postle, K.** 1990. TonB and the Gram-negative dilemma. Mol. Microbiol. **4:**2019-2025.

31. **Pressler, U., H. Staudenmaier, L. Zimmermann, and V. Braun.** 1988. Genetics of the iron dicitrate transport system of *Escherichia coli*. J. Bacteriol. **170:**2716-2724.

32. **Reidl, J. and J.J. Mekalanos.** 1996. Lipoprotein *e*(P4) is essential for hemin uptake by *Haemophilus influenzae*. J. Exp. Med. **183:**621-629.

33. **Ren, Z., H. Jin, D.J. Morton and T.L Stull.** 1997. Cloning of *hgpB*, a gene encoding a second hemoglobin binding protein of *Haemophilus influenzae*, abstr. B-227. P. 67. *In* Abstracts of the 97th General Meeting of the American Society for Microbiology 1997, Washington D.C.

34. **Sambrook, T., E.F. Fritsch, and T. Maniatis.** 1989.*In* C. Nolan (ed), Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

35. **Sharetzsky, C., T.D. Edlind, J.J. LiPuma, and T.L. Stull.** 1991. A novel approach to insertional mutagenesis of *Haemophilus influenzae*. J. Bacteriol. **173:**1561-1564.

36. **Spenser, H.T. and R.M. Herriott.** 1965. Development of competence in *Haemophilus influenzae*. J. Bacteriol. **90:**911-920.

37. **Stull, T.L.** 1987. Protein sources of heme for *Haemophilus influenzae*. Infect. Immun. **55:**148-153.

38. **Szabo, M., D. Maskell, P. Butler, J. Love, and R. Moxon.** 1992. Use of chromosomal gene fusions to investigate the role of repetitive DNA in regulation of genes involved in lipopolysaccharide biosynthesis in *Haemophilus influenzae*. J. Bacteriol. **174:**7245-7252.

39. **Turk, D.C.** 1984. The pathogenicity of *Haemophilus influenzae*. J. Med. Microbiol. **18:**1-16.

40. **van Belkum, A., S. Scherer, W. van Leeuwen, D. Willemse, L. van Alphen, and H. Verbrugh.** 1997. Variable Number of Tandem Repeats in Clinical Strains of *Haemophilus influenzae*. Infect. Immun. **65:**5017-5027.

41. **Wallace, R.J., D.M. Musher, and R.J. Martin.** 1978. *Haemophilus influenzae* pneumonia in adults. Am. J. Med. **64:**87-93.

42. **Weiser, J.N., J.M. Love, and E.R. Moxon.** 1989. The molecular mechanism of phase variation of *H. influenzae* lipopolysaccharide. Cell. **59:**657-665.

43. **Whitby, P.W., K.E. Sim, D.J. Morton, J.A. Patel, and T.L. Stull.** 1997. Transcription of genes encoding iron and heme acquisition proteins of *Haemophilus influenzae* during acute otitis media. Infect. Immun. **65:**4696-4700.

44. **Wong, J.C., J. Holland, T. Parsons, A. Smith, and P. Williams.** 1994. Identification and characterization of an iron-regulated hemopexin receptor in *Haemophilus influenzae* type b. Infect. Immun. **62:**48-59.

45. **Wong, J.C.Y., R. Patel, D. Kendall, P.W. Whitby, A. Smith, J. Holland, and P. Williams.** 1995. Affinity, conservation, and surface exposure of hemopexin-binding proteins in *Haemophilus influenzae*. Infect. Immun. **63:**2327-2333.

CHAPTER IV

HEME REGULATION OF AN *HAEMOPHILUS INFLUENZAE* HEMOGLOBIN AND HEMOGLOBIN-HAPTOGLOBIN BINDING PROTEIN GENE (*hgpA*)

ABSTRACT

Several heme acquisition mechanisms by *Haemophilus influenzae* have been characterized and shown to be heme repressible *in vitro*. However, little is known about the regulatory mechanism of the expression of heme acquisition proteins. We have cloned a gene, *hgpA*, encoding a hemoglobin and hemoglobin-haptoglobin binding outer membrane protein from *H. influenzae* strain HI689. The expression of *hgpA* is heme repressible. A comparative quantitative reverse transcriptase polymerase chain reaction (CQ-RT-PCR) method was developed to quantitate and compare *hgpA* transcripts from bacterium after growth in heme replete or heme deplete growth conditions. DNA mimics of *hgpA* and the *hindIII* methyltransferase gene (*hindIIIM*) were constructed. Since *hindIIIM* is expressed constitutively in heme replete and heme deplete media, comparing the *hgpA*/*hindIIIM* ratio after growth in heme replete and heme deplete conditions provides a measurement of *hgpA* transcriptional regulation by heme. HgpA expression is increased three fold in heme restricted medium compared to expression in heme replete

medium. To verify the method, a *hgpA-lacZ* protein fusion was used to assay expression of HgpA under the same conditions. The results were comparable to those of CQ-RT-PCR. Therefore, CQ-RT-PCR may prove useful to investigate *in vitro* and *in vivo* transcriptional regulation of genes.

INTRODUCTION

Haemophilus influenzae is a small pleomorphic Gram-negative bacterium which causes both invasive and mucosal disease in humans including meningitis, otitis media and pneumonia (7,13,25,26). Since *H. influenzae* is unable to synthesize the porphyrin ring, it requires an exogenous porphyrin source for aerobic growth. Heme can satisfy not only elemental iron requirements but also the demand of porphyrin. In humans, heme is sequestered intracellularly in heme proteins, hemoglobin and cytochromes, or when present in the circulation, it is bound to the serum proteins hemopexin and albumin (in the case of heme), and by haptoglobin (in the case of hemoglobin). Thus, heme is generally unavailable to invading microorganisms (17). The ability of bacteria to capture iron and heme is critical for pathogenicity (3,28). A number of iron and heme acquisition mechanisms have been characterized and been shown to be heme repressible *in vitro*, including transferrin binding by *H. influenzae* (19), and hemoglobin binding by both *H. influenzae* (9), and *H. ducreyi* (6). However, little is known about the regulatory mechanism. Previously, we cloned a gene, *hgpA*, encoding a

hemoglobin and hemoglobin-haptoglobin binding outer membrane protein from *H. influenzae* strain HI689 (12). Based on a whole cell dot blot assay, binding of hemoglobin by *H. influenzae* was repressed by heme in *vitro*, although the extent of repression was not characterized (9,12).

Further understanding of gene regulation of *hgpA* at the transcriptional level would be aided by an accurate, sensitive measurement of *hgpA* transcripts under different physiological and pathological conditions. The method of RT-PCR enables rapid amplification of low copy number of specific mRNA. However, PCR amplification has an exponential nature, the extent of amplification (N) is given by the equation $N=N_0(1+\text{eff})^n$, where N_0 is the initial amount of material, eff is the efficiency, and n is the cycle number. Any variation affecting amplification efficiency could result in dramatic changes in product yield (11). In addition, the plateau effect after many cycles at a high concentration of PCR products can lead to inaccurate estimates of mRNA (18). These drawbacks can be overcome by co-amplification of a target and a mimic which share the same primer binding sequence. Since amplification efficiency is primarily determined by the primer sequence (11,27,31), and since the ratio of target DNA to mimic DNA remains constant during the amplification, it is not necessary to obtain data before the reaction reaches the plateau phase.

Currently, it is difficult to determine the expression of specific proteins *in vivo* because culturing a clinical sample prior to PCR changes the conditions for expression. This is an important barrier to characterizing the pathogenic importance of proteins which

are not constitutively expressed. The objective of this work was to develop a method which is useful for studying *in vivo* heme regulation of *hgpA*, the hemoglobin and hemoglobin-haptoglobin binding protein gene of *H. influenzae*. In this study, an accurate and sensitive comparative quantitative-reverse transcriptase-polymerase chain reaction (CQ-RT-PCR) protocol was developed by constructing DNA mimics of *hgpA* and the *hind*III methyltransferase gene (*hindIIIM*). Since *hindIIIM* is expressed constitutively in heme replete and heme deplete media (29), the *hgpA*/*hindIIIM* ratio is a measurement of *hgpA* transcriptional regulation by heme. In addition, a protein fusion of *hgpA-lacZ* was used to study the heme regulation of *hgpA*'s expression. The technique of CQ-RT-PCR may provide important information on expression of specific genes from small samples obtained during disease episodes.

MATERIAL AND METHODS

Bacterial strains and growth conditions. *H. influenzae* type b strain HI689 has been previously described (12). *H. influenzae* Rd KW20 (ATCC 51907) was the strain used in the complete sequencing of the *H. influenzae* genome (8) and was obtained from the American Type Culture Collection (ATCC), Rockville, Maryland. *H. influenzae* was routinely maintained on brain heart infusion (BHI) agar (Difco, Detroit, Mich.) supplemented with 10 μg ml^{-1} of both heme and β-NAD (sBHI). Heme-restricted growth of *H. influenzae* was performed in BHI broth supplemented with 10 μg ml^{-1} of β-

NAD and 0.1 μg ml^{-1} of heme (dBHI). The *hgpA-LacZ* protein fusion inserted into the *H. influenzae* RdKW20 chromosome in a single copy (23) was used to monitor expression of HgpA under different growth conditions as specified in Table 6.

DNA Methodology. Restriction endonucleases were obtained from New England Biolabs (Beverly, Mass.) and were used as directed by the manufacturer. Genomic DNA was isolated using the DNA Now reagent (Biogentex, Seabrook, TX) as directed by the manufacturer. Plasmid DNA was isolated using Qiagen plasmid kits (Qiagen, Chatsworth, CA) according to the manufacturer's recommendations. Oligonucleotides were synthesized at the Molecular Biology Resource Facility, University of Oklahoma Health Sciences Center, Oklahoma City, OK.

Construction of a polycompetitor plasmid pJOK. Primer pairs designed to perform CQ-RT-PCR are shown in Table 5. Using *H. influenzae* strain HI689 genomic DNA as template, the primer pair for *hgpA* would amplify an internal *hgpA* specific gene fragment. The primer pair for *hgpAm* would amplify an *hgpA* DNA mimic. Similarly, the primer pair *hindIIIM* would amplify a *hindIIIM* specific DNA fragment, while the primer pair for *hindIIIMm* would amplify an *hindIIIM* DNA mimic. Since the primers were designed such that they have similar melting temperature, a single program was used to perform PCR. PCR was performed in a total volume of 50 μl in a RapidCycler Thermocycler (Idaho Technology Inc.). The PCR reaction mixture contained 5 μl of 10X 20 mM Mg PCR buffer (Idaho Technology Inc., Idaho Falls, Idaho), 1μl of 10 mM dNTPs, 1 unit of Taq DNA polymerase, 10 pM of each primer, and 1 μl of template with

UNIVERSITY OF OKLAHOMA
LIBRARIES

different concentration. The samples were cycled 30 times with a denaturing step at 95 °C for 10 sec., annealing step at 51 °C for 10 sec, and an extension at 72 °C for 30 sec. The PCR products were analyzed on a 1.2% agarose gel.

Figure 21a shows the synthesis of DNA mimics using composite primers. Figure 21b shows the polycompetitor plasmid pJOK. Plasmid pJOK was constructed using the PCR products derived from amplification with the *hgpA* mimic primers and the *hindIIIM* mimic primers, which were cloned into the pCR2.1-TOPO TA cloning vector (Invitrogen) separately. The insert of *hgpA*m was excised from the vector using *Eco*RI, the fragment was blunt ended using the Klenow fragment, and inserted into the *Eco*RV site of pZEro-1 (Invitrogen). The insert of *hindIIIM*m was similarly excised using *Eco*RI and ligated into the *Eco*RI site of pZEro-1 containing the *hgpA*m to create the polycompetitor plasmid pJOK (Figure 21).

Reverse Transcriptase-Polymerase chain reaction (RT-PCR). RT-PCR was performed as follows. One microgram of RNA isolated using the RNase total RNA kit (Qiagen) was treated with 1 U of 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 of 25 mM EDTA and heating for 10 min at 65 °C. Twelve point five microlitre of the RNA sample was annealed to 1 µl of random nanomer primer by heating to 70 °C for 2 min, followed by rapid chilling on ice. Complementary DNA (cDNA) was synthesized at 42 °C using Superscript II reverse transcriptase (RT) (Gibco BRL) in a 20 µl reaction mixture containing 100 U of enzyme, 1x first strand buffer, 2.5 mM of each dNTP

(dATP, dCTP, dGTP, dTTP), and 5U of RNase inhibitor. Reverse transcriptase was omitted in negative-control tubes to assure that chromosomal DNA template was not present. After 1 h, the reaction mixture was heated at 94 °C for 5 min and subsequently stored at -20 °C. The cDNA or the negative-control (reaction without RT) was used as template in the PCR. In addition, reactions using either the plasmid pJOK or ddH_20 as template were used as positive and negative controls respectively. PCR conditions were the same as those described above.

Figure 21. Construction of DNA mimics (a) and the polycompetitor plasmid pJOK (b). Gray lines indicate gene specific primers, black lines indicate heterologous DNA specific primers. Solid gray boxes indicate gene specific DNA fragment, solid black boxes indicate heterologous DNA fragment.

Figure 21 a

Composite primers

heterologous DNA fragment

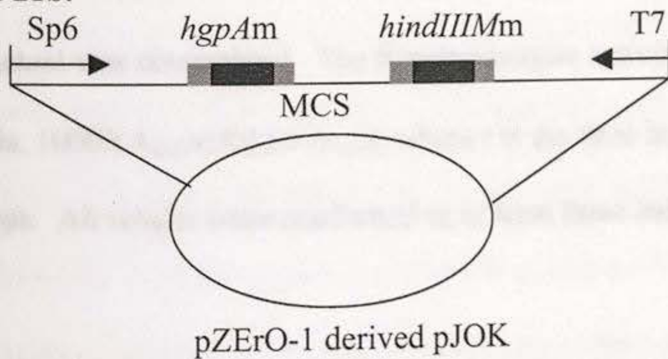
PCR with composite primers

DNA mimics

PCR with gene specific primer

Gene specific PCR products

Figure 21b.



pZErO-1 derived pJOK

LacZ activity assay for the *hgpA- lacZ* protein fusion. RdKW20 cells containing the *hgpA-lacZ* fusion in the genome were grown in 10 ml of sBHI broth overnight. The overnight culture was used to inoculate 25 ml of dBHI broth (4% inoculum), which was incubated to mid log phase and used to inoculate (10% inoculum) 25 ml of sBHI and 25 ml of dBHI, or other medium as specified in Table 6. One millilitre of cell culture was taken to assay β-galactosidase activity by monitoring the hydrolysis of *0*-nitrophenyl-β-D-galactopyranoside (ONPG) as described by Miller (Miller, 1972). Briefly, 1ml cell culture was resuspended in 1 ml A medium (10.5 g/l of K_2HPO_4, 4.5 g/l of KH_2PO_4, 1.0g/l of $(NH_4)_2SO_4$, 0.5 g/l of Na citrate·$2H_2O$, pH7.0) to assay OD_{600}. Then 0.5 ml of the cell culture in A medium was mixed with 0.5 ml of Z buffer (16.1 g/l of $Na_2HPO_4 \cdot 7H_2O$, 5.5 g/l of $NaH_2PO_4 \cdot H_2O$, 0.75 g/l of KCl, 0.246 g/l of $MgSO_4 \cdot 7H_2O$, and 3.57 ml of 14.5M β-mercaptoethanol, pH 7.0), 40 µl of $CHCl_3$, 20µl of 0.1 % SDS. The above mixture was vortexed for 10 sec, and equilibrated to 28 °C for 15 min. Then 0.2 ml of 4 mg/ml of ONPG in A medium was added to the equilibrated mixture. The reaction was stopped by adding 0.5 ml of 1M Na_2CO_3 after the yellow color had developed. The reaction mixture was briefly centrifuged and the OD_{420} of the supernatant was determined. The β-galactosidase activity was calculated according to the formula, $1000XA_{420}/(tX0.5XA_{600})$, where t is the time in min at which the yellow color develops. All results were confirmed in at least three independent experiments.

RESULTS

Construction of the polycompetitor pJOK. DNA mimics of *hgpA* were generated in a PCR amplification using a pair of composite primers (Table 5). Each primer contained a 20 mer which would anneal to the locus of HI0147 homologue in HI689, which has no assigned function in the *H. influenzae* Rd KW20 genome. Each primer also contained a 20 mer tail identical to the *hgpA* specific primers. During the PCR reaction, a 500 bp HI0147 DNA fragment was amplified, and *hgpA*-specific primer sequences were incorporated into the ends of the PCR product (Fig. 21a). This PCR product is a DNA mimic of *hgpA*, designated *hgpA*m, which is 94 bp longer than the *hgpA* specific PCR products. The *hgpA*m was cloned into the vector pZEr0-1 to yield a competitor. Similarly, a DNA mimic of *hindIIIM* was amplified, containing a 500 bp DNA fragment of the HI0730 homologue of HI689 with *hindIIIM* specific primer binding sequences incorporated into each end of the PCR product and 97 bp longer than *hindIIIM* specific PCR products. The *hindIIIM* mimic was cloned into the *hgpA* competitor plasmid to yield the polycompetitor pJOK (Fig. 21b).

CQ-RT-PCR to study *in vitro* transcriptional regulation of *hgpA* by heme. HI689 were grown in heme replete (10μg/ml heme) and heme deplete (0.1μg/ml heme) medium to early log phase, the total RNA was extracted and cDNA synthesis was performed. Ten fold serial dilutions of pJOK were made, and 1 μl of each dilution was mixed with 1μl of cDNA either from HI689 grown in heme replete medium or from heme deplete medium. PCR amplification was performed using either *hgpA* specific

primers or *hindIIIM* specific primers (Table 5) to determine the region of equivalence of PCR products derived from cDNA and pJOK based on intensity on the agarose gel. A second PCR was performed as above using appropriate dilutions of pJOK. The quantity of *hgpA* and *hindIIIM* in each sample was determined by observing the dilutions which yield equivalent quantity of PCR products derived from cDNA and pJOK based on the intensity on the agarose gel analysis (Fig. 22b). Since *hindIIIM* is constitutively expressed under heme replete and heme deplete conditions (29), the ratio of *hgpA/hindIIIM* PCR products is a measurement of relative expression of *hgpA* under these conditions.

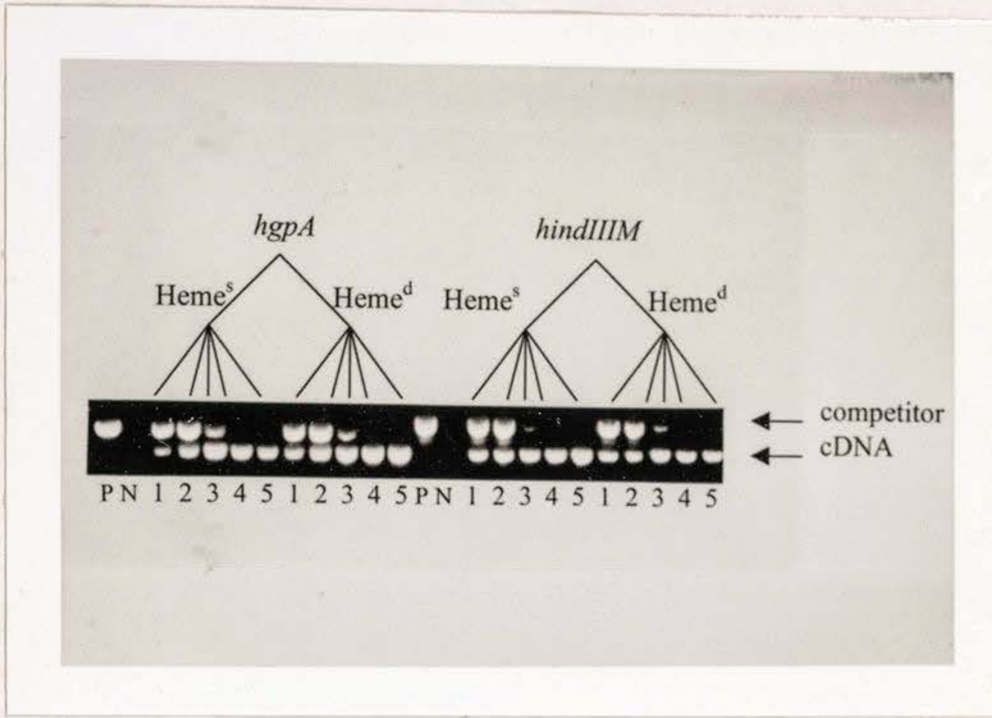
The quantity of *hgpA* is equivalent to $7.5X10^{-5}$ dilution of pJOK, *hindIIIM* is equivalent to $7.5X10^{-5}$ dilution of pJOK under heme deplete condition. Similarly, the quantity of *hgpA* is equivalent to $2.5X10^{-5}$ dilution of pJOK, and *hindIIIM* is equivalent to $7.5X10^{-5}$ dilution of pJOK under heme replete condition. A comparison of the ratios of *hgpA/hindIIIM* PCR products in heme deplete (*hgpA/hindIIIM*=1) and heme replete (*hgpA/hindIIIM*=1/3) conditions indicated that there was an approximately 3 fold up-regulation of *hgpA* transcription in heme deplete growth conditions.

***LacZ* activity assay for the *hgpA* protein fusion.** To compare transcriptional regulation with gene expression, LacZ activity was assayed using a *hgpA-lacZ* protein fusion constructed in *H. influenzae* Rd KW20 (23). The organisms were grown to early log phase in heme replete or heme deplete media. The β-galactosidase activity was assayed and 3.9 fold increase in gene expression was observed in those organisms grown

in heme deplete media compared to those cultured in heme replete conditions (Table 6). Similar experiments were performed where the elemental iron level was restricted in the presence of heme or the β-NAD was restricted. There was no apparent regulation of *hgpA* expression under either iron or β-NAD restricted condition (Table 6).

Figure 22. CQ-RT-PCR analysis to quantitate *hgpA* transcription of HI689. A constant amount of cDNA from total RNA of HI689 grown in the heme replete or heme deplete medium was mixed with serial dilution of the competitor plasmid pJOK. PCR was performed with either primer pair for *hgpA* or *hindIIIM*. Upper panel shows the primary PCR reaction in which the competitor plasmid pJOK was diluted in a series of 10 fold dilutions to determine the correct dilution range to use for a more precise determination of *hgpA* transcriptional regulation. Lane 1, 2, 3, 4, 5 represent 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}, 10^{-8} dilutions of pJOK. Lower panel shows the results of the PCR reaction in which the competitor plasmid pJOK was diluted in the appropriate ranges for analysis of *hgpA* and *hindIIIM* cDNA levels. Lane 1, 2, 3 represent $7.5X10^{-5}$, $5.0X10^{-5}$, $2.5X10^{-5}$ dilutions of pJOK. $Heme^{s}$ are heme replete samples and $Heme^{d}$ are heme deplete samples. N represents negative controls of PCR without any templates. P represents positive controls with pJOK as templates (upper panel) or HI689 chromosomal DNA as templates (lower panel).

Figure 22.



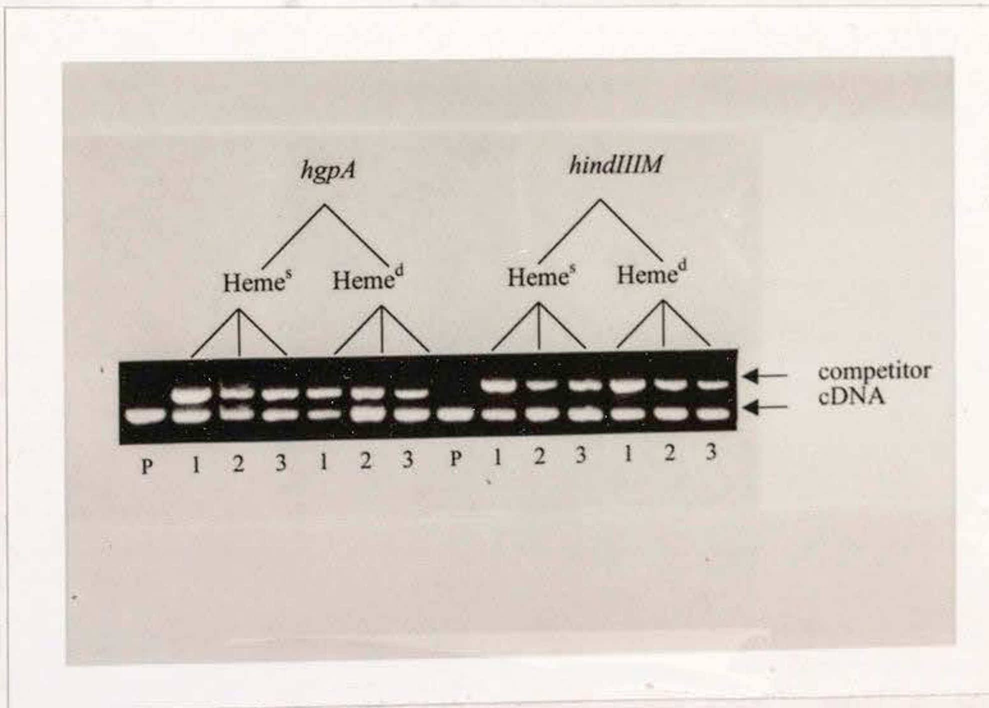


Table 5. Primers used in construction of pJOK and in the CQ-RT-PCR

Product	Sense sequence (5' to 3')	Antisense sequence (5' to 3')	Size(bp)
hgpA	**agatggagattccacaaagt**	**cataaataggtatggaccag**	406
*hgpA*m	**agatggagattccacaaagt**attcatttcggtattcgtact	**cataaataggtatggaccag**aaaatggcacggcaagtaat	500
hindIIIM	**ggatagtaattctattcatgc**	**gacgttcaaacacacaagaa**	403
*hindIIIM*m	**ggatagtaattctattcatgc**tagtggtttagatcgaatttc	**gacgttcaaacacacaagaa**acgcctagatattgttctac	500

Table 6. Expression of *hgpA-lacZ* in various growth conditions

Medium	Growth condition	β-Gal. act. (Miller units)	Fold Induction
sBHI	BHI+NAD (10μg /ml) + Heme (10μg /ml)	24.50	1.0
Heme deplete	BHI+NAD (10μg /ml) + Heme (0.1μg/ml)	97.77	3.9
Iron deplete	BHI+NAD (10μg /ml) + Heme (10μg /ml) + EDDA (250μM)	31.03	1.3
NAD deplete	BHI+NAD (0.02μg/ml) + Heme (10μg /ml)	26.23	1.1

DISCUSSION

In this study, the effect of heme availability on the transcriptional regulation of an *H. influenzae* hemoglobin and hemoglobin-haptoglobin binding protein gene (*hgpA*) was examined by use of a polycompetitor plasmid pJOK in CQ-RT-PCR. The polycompetitor plasmid was designed such that the target and the mimic compete for the same primers but yield different sized PCR products during PCR amplification (Fig.21, Table 5). There are different methods to construct DNA mimics (1), including 1) restriction enzyme digestion of a PCR product at two enzyme recognition sites, and subsequent re-ligation to produce a mimic of modified size (14); 2) inserting a fragment into a previously cloned sequence (22); and 3) multiple step amplification of genomic DNA to remove an internal fragment (4). All of the above methods have the advantage that the PCR products can be distinguished easily by size in agarose gels. However, the first two methods require the presence of unique restriction sites within the PCR products. In addition, all three methods encourage formation of a heteroduplex containing one target strand and one mimic strand (2). Heteroduplex formation complicates the electrophoretic pattern as well as quantitation(31). In the present study, DNA mimics were constructed using composite primers which share the same primer binding sequence but amplify an heterologous DNA fragment of different size (Fig.21, Table 5). Thus the DNA mimics are amplified at the same efficiency as its target but avoid the formation of heteroduplex.

To quantitate *hgpA* transcripts, a constant amount of the *hgpA* cDNA was mixed with 10 fold serial dilutions of the mimic DNA (polycompetitor plasmid) and PCR performed. Since the initial ratio of target DNA to mimic DNA is equal to the ratio of their PCR products, the amount of *hgpA* cDNA present in the sample can be determined by observing the point of equivalence of final PCR products. Because the ratio of target DNA to mimic DNA remains constant during the amplification, it is not necessary to obtain data before the reaction reaches the plateau phase. The polycompetitor plasmid also incorporates a mimic to compete for primers specific to the *H. influenzae hindIIIM* (Fig.21). Since *hindIIIM* expression is not regulated by heme, the ratio of *hgpA/hindIIIM* is a measurement of *hgpA* transcriptional regulation by heme.

The CQ-RT-PCR protocol may be applied to bacterial mRNA isolated from clinical samples to determine the level of gene regulation *in vivo*. Since culturing bacteria from clinical samples alters gene expression, direct quantitation of transcription may be more useful in revealing *in vivo* gene expression.

CQ-RT-PCR experiments were performed on *in vitro* grown organisms, and ratios of *hgpA* to *hindIIIM* demonstrated expression of *hgpA* to be 3-fold in heme deplete conditions compared to heme replete conditions (Fig.22). To rule out the possibility that significant regulation occurred at post-transcriptional level, we measured the β-galactosidase activity of a *hgpA-lacZ* protein fusion. The activity increased 3.9 fold in response to heme limitation (Table 6). These comparable results suggest that *hgpA* is regulated by heme at the transcriptional level and CQ-RT-PCR may be useful to

investigate *in vivo* transcriptional regulation of genes from pathogenic organisms with similar regulation. These results also indicate that *hgpA* gene expression is not regulated by elemental iron but by heme (Table 6), which are consistent with the earlier observations of Morton with respect to transferrin binding(19)and are in contrast to the conclusions of Hasan *et al* (10).

Whether the heme regulatory effect is direct or indirect is unclear. In yeast, heme induces the DNA binding of the transcriptional activator HAP1 by promoting protein dimerization (20,30). The heme-binding domain of HAP1 contains a Cys-Pro-X-Asp-His sequence. The presence of a similar sequence, termed a heme regulatory motif, in other heme-responsive proteins suggests that the regulatory action of heme is mediated by binding to this motif (15,16). However, there is no such motif present in HgpA. Smith *et al*. demonstrated that heme binds *E. coli* Fur tightly (24), although Fur does not contain the proposed heme binding motif and the biological consequences of heme-Fur binding have not been elucidated(5,21). Therefore, it is possible that heme binds directly to the transcription factors in both eukaryotes and prokaryotes to regulate heme-responsible genes.

REFERENCES

1. **Anderson, K.M., P.H. Cheung, and M.D. Kelly**. 1997. Rapid generation of homologous internal standards and evaluation of data for quantitation of messenger RNA by competitive polymerase chain reaction. J.Pharmacol.Toxicol.Meth. **38**:133-140.

2. **Araki, N., F.D. Robinson, and S.K. Nishimoto**. 1993. Rapid and sensitive method of quantitation of bone gla protein mRNA using competitive polymerase chain reaction. J.Bone & Mineral Res. **8**:313-322.

3. **Bullen, J.J.** 1981. The significance of iron in infection. Rev.Infect.Dis. **3**:1127-1138.

4. **Celi, F.S., M.E. Zenilman, and A.R. Shuldiner**. 1993. A rapid and versatile methods to synthesize internal standards for competitive PCR. Nucleic.Acids.Res. **21**:1047

5. **Creusot, F., J. Verdiere, M. Gaisne, and P.P. Slonimski**. 1988. CYP1(HAP1) regulator of oxygen-dependent gene expression in yeast. I. Overall organiztion of the protein sequence displays several novel structural domains. J.Mol.Biol. **204**:263-276.

6. **Elkins, C., C.J. Chen, and C.E. Thomas**. 1995. Characterization of the hgbA locus encoding a hemoglobin receptor from *Haemophilus ducreyi*. Infect.Immun. **63**:2194-2200.

7. **Falla, T.J., S.R. Dobson, D.W. Crook, W.A. Kraak, W.W. Nichols, E.C. Anderson, J.Z. Jordens, M.P. Slack, E. Mayon-White, and E.R. Moxon**. 1993. Population-based study of non-typable *Haemophilus influenzae* invasive disease in children and neonates. Lancet **341**:851-854.

8. **Fleischmann, R.D., M.D. Adams, O. White, R.A. Clayton, E.F. Kirkness, A.R. Kerlavage, C.J. Bult, J. Tomb, B.A. Dougherty, J.M. Merrick, K. McKenney, G. Sutton, W. FitzHugh, C. Fields, J.D. Gocayne, J. Scott, R. Shirley, L. Liu, A. Glodek, J.M. Kelley, J.F. Weidman, C.A. Phillips, T. Spriggs, E. Hedblom, M.D. Cotton, R.C. Utterback, M.C. Hanna, D.T. Nguyen, D.M. Saudek, R.C. Brandon, L.D. Fine, J.L. Fritchman, J.L. Fuhrmann, N.S.M. Geoghagen, C.L. Gnehm, L.A. McDonald, K.V. Small, C.M. Fraser, H.O. Smith, and J.C. Venter**. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science **269**:496-512.

9. **Frangipane, M.E., D.J. Morton, J.A. Wooten, J.M. Pozsgay, and T.L. Stull.** 1994. Binding of human hemoglobin by *Haemophilus influenzae*. FEMS Microbiol.Lett **118**:243-248.

10. **Hasan, A.A., J. Holland, A. Smith, and P. Williams**. 1997. Elemental iron does repress transferrin, haemopexin and haemoglobin receptor expression in *Haemophilus influenzae*. FEMS Microbiol.Lett. **150**:19-26.

11. **Innis, M.A. and D.H. Gelfand**. 1990. Optimization of PCRs, p.3M.A. Innis, D.H. Gelfand, J.J. Sninsky, and T.J. White (ed.), In PCR Protocols: A guide to methods and applications, Academic Press Inc., San Diego.

12. **Jin, H., Z. Ren, J.M. Pozsgay, C. Elkins, P.W. Whitby, D.J. Morton, and T.L. Stull**. 1996. Cloning of a DNA fragment encoding a heme-repressible hemoglobin-binding outer membrane protein from *Haemophilus influenzae*. Infect.Immun **64**:3134-3141.

13. **Klein, J.O.** 1997. Role of nontypeable *Haemophilus influenzae* in pediatric respiratory tract infections. Pediatr.Infect.Dis.J. **16**:S5-S8

14. **Kozbor, D., E. Hyjek, R. Wiaderkiewicz, Z. Wang, M. Wang, and E. Loh**. 1993. Competitor mRNA fragments for quantitation of cytokine specific transcripts in cell lysates. Mol.Immunol. **30**:1-7.

15. **Lathrop, J.T. and M.P. Timko**. 1993. Regulation by heme of mitochondrial protein transport through a conserved amino acid motif. Science **259**:522-525.

16. **Lee, B.C.** 1995. Quelling the red menace: haem capture by bacteria. Mol.Microbiol. **18**:383-390.

17. **Lee, B.C.** 1992. Isolation of an outer membrane hemin-binding protein of *Haemophilus influenzae* type b. Infect.Immun. **60**:810-816.

18. **Mocharla, H., R. Mocharla, and M.E. Hodes**. 1990. Coupled reverse transcription-polymerase chain (RT-PCR) as a sensitive and rapid method for isozyme genotyping. Gene **93** :271-275.

19. **Morton, D.J., J.M. Musser, and T.L. Stull**. 1993. Expression of the *Haemophilus influenzae* transferrin receptor is repressible by hemin but not elemental iron alone. Infect.Immun. **61**:4033-4037.

20. **O'Halloran, T.V.** 1993. Transition metals in control of gene expression. Science **261**:715-725.

21. **Pfeifer, K., K.S. Kim, S. Kogan, and L. Guarente**. 1989. Functional dissection and sequence of yeast HAP1 activator. Cell **56**:291-301.

22. **Reiner, S.L., S. Zheng, D.B. Corry, and R.M. Locksley**. 1993. Constructing polycompetitor cDNAs for quantitative PCR. J.Immunol.Methods **165**:37-46.

23. **Ren, Z., H. Jin, D.J. Morton, and T.L. Stull**. 1998. Evidence for Slip-strand Mispairing as a Mechanism of Regulation of Hemoglobin-binding Proteins in *Haemophilus influenzae*, abstr. In Abstracts of the 98th General Meeting of the American Socitey for Microbiology 1998. American Society for Microbiology, Washington, D.C.

24. **Smith, A., N.I. Hooper, N. Shipulina, and W.T. Morgan**. 1996. Heme binding by a bacterial repressor protein, the gene product of the ferric uptake regulation (*fur*) gene of *Escherichia coli*. J.Prot.Chem. **15**:575-583.

25. **Turk, D.C.** 1984. The pathogenicity of *Haemophilus influenzae*. J.Med.Microbiol. **18**:1-16.

26. **Wallace, R.J., D.M. Musher, and R.J. Martin**. 1978. *Haemophilus influenzae* pneumonia in adults. Am.J.Med. **64**:87-93.

27. **Wang, A.M., M.V. Doyle, and D.F. Mark**. 1989. Quantitation of mRNA by the polymerase chain reaction. Proc.Natl.Acad.Sci.USA. **86**:9717-9721.

28. **Weinberg, E.D.** 1978. Iron and infction. Microbiol.Rev. **42**:45-66.

29. **Whitby, P.W., K.E. Sim, D.J. Morton, J.A. Patel, and T.L. Stull**. 1997. Transcription of genes encoding iron and heme acquisition proteins of *Haemophilus influenzae* during acute otitis media. Infect.Immun. **65**:4696-4700.

30. **Zhang, L., O. Bermingham-McDonogh, B. Turcotte, and L. Guarente**. 1993. Antibody-promoted dimerization bypasses the regulation of DNA binding by the heme domain of the yeast transcriptional activator HAP1. Proc.Natl.Acad.Sci.USA **90**:2851-2855.

31. **Zhao, G., C. Pijoan, K. Choi, S.K. Maheswaran, and E. Trigo**. 1995. Expression of iron-regulated outer membrane proteins by porcine strains of *Pasteurella multocida*. Can.J.Vet.Res **59**:46-50.

CHAPTER V

SUMMARY

The dependence of *Haemophilus influenzae* on exogenous heme sources for aerobic growth distinguishes this pathogen from nearly all other bacterial species. Since hemoglobin can satisfy this requirement *in vitro*, it is reasonable to predict that outer membrane receptor(s) of *H. influenzae* may mediate the binding of hemoglobin. Using biotin-labeled human hemoglobin, a 120 kDa putative hemoglobin binding protein (HgpA) was purified from *H. influenzae* strain HI689 through an affinity chromatography. Following N-terminal microsequencing, a gene (*hgpA*) was cloned from the *H. influenzae* genomic library using probes designed according to the N-terminal amino acid sequence of HgpA and *Haemophilus* codon preference. Sequence analysis indicated *hgpA* had significant homology with other bacterial iron and heme utilization proteins. Like other TonB dependent proteins, *hgpA* has a TonB box at its N-terminus and six other characteristic TonB regions in the sequence. The function of HgpA was further established by cloning of *hgpA* in an expression vector, with demonstration of hemoglobin binding activity in *E. coli* harboring the recombinant protein. Similarly, hemoglobin-haptoglobin activity was also detected in this assay. Mutation of the gene in HI689 revealed no phenotypic change in the binding and utilization of hemoglobin or hemoglobin-haptoglobin, suggesting that there maybe more than one hemoglobin/hemoglobin-haptoglobin-binding protein in HI689.

A search of the complete genome sequence of *H. influenzae* strain RdKW20 indicated four ORFs with a significant homology with *hgpA*. An interesting feature of *hgpA* and its four homologues is that they contain a multiple CCAA repeats in the structural gene following the leader peptide coding sequence. Preliminary data and ongoing studies have shown there are two more CCAA containing genes with homology to *hgpA* in HI689. We propose that CCAA mediated strand slippage may regulate *hgpA* expression by addition or deletion of CCAA units and therefore cause the introduction of stop codons immediately downstream of the CCAA sequence.

Using a novel Comparative Quantitative-Reverse Transcriptase-Polymerase Chain Reaction (CQ-RT-PCR) method and a *hgpA-LacZ* protein fusion construct, repression of *hgpA* by heme *in vitro* at the transcriptional level, but not at the translational level was characterized. The method of CQ-RT-PCR involved construction of DNA mimics, which compete with genes of interest for primers, thus allowing quantification of target sequences. More importantly, this method used a constitutively expressed gene as a standard to avoid counting bacterial number, which may have significant application to the *in vivo* transcriptional regulation studies for a variety of pathogens. This method provides an opportunity to investigate gene regulation *in vivo*.

The findings obtained from work presented in this dissertation have improved our understanding of heme uptake by *H. influenzae*. The information gained from these studies paved the way for finding and understanding other hemoglobin/hemoglobin-

haptoglobin-binding proteins. Furthermore, the proposal of CCAA slippage mechanism opened a new avenue for post-transcriptional regulation of heme utilization proteins.

APPENDIX

CCAA REPEATS ARE NOT REQUIRED FOR HgpA TO BIND HEMOGLOBIN AND HEMOGLOBIN-HAPTOGLOBIN

We proposed that CCAA repeats are involved in gene regulation via slipped strand mechanism, similar to the CAAT repeats in the *lic* locus in *H. influenzae*. CAAT repeats are not required for biosynthesis of lipooligosaccharide {High1996}. To investigate whether CCAA repeats are required in HgpA's function, a construct was made to insert *hgpA* gene without CCAA repeats in an expression vector to test its ability to bind hemoglobin and hemoglobin-haptoglobin in *E. coli*.

PCR amplification of *hgpA* without CCAA. A pair of primers HFJXN and HFJXC was designed to clone the gene *hgpA* without CCAA repeats. The primer HFJXN (5'-GACCAGGGATCCAATAGTAATGCTTCT-3') was immediately following the CCAA repeats, primer HFJXC (5'-GGAAGGGGTACCTTAGAATTCAAACTG-3') was designed at the C-terminal of *hgpA*. Each primer has a restriction site (*Bam*HI and *Kpn*I respectively) incorporated into the 5' end to directionally clone the PCR products into the expression vector pRESETA in the correct reading frame. Using 30 ng pHFJ2 as template, PCR was performed in 50 μl reactions containing 2 mM $MgCl_2$, 200 μM of each dNTP, 100 ng of each primer and 2 units of Taq DNA polymerase. PCR was

carried out for 30 cycles with each cycle consisting of denaturation at 95 °C for 1 min, annealing at 48 °C for 1 min, and primer extension time 72 °C for 3.5 min, with a final extension at 72 °C for 10 min.

Amplicons of the expected size (3.0 kbp) were gel purified. The PCR products were digested with *Bam*HI and *Kpn*I, and ligated to gel purified *Bam*HI and *Kpn*I digested pRSETA. The ligation mixture was transformed into *E. coli* BL21(DE3)pLysS and recombinants were selected on LB agar containing 50 μg/ml ampicillin. Plasmids were isolated from ampicillin resistant colonies and mapped by restriction enzyme digestion to identify clones containing the expected product. A positive clone was identified and designated pXHGP'.

Expression of Hgp in *E. coli*. Recombinant *E. coli* BL21(DE3)pLysS were grown to mid-logarithmic phase in LB medium with appropriate antibiotic supplementation. After induction with 1 mM isopropyl β-D-thiogalactopyranoside (IPTG) for 4 h cells were harvested and resuspended in PBS to a concentration of 10^8 cfu/ml. Expression of the recombinant protein was analyzed by running the whole cell lysates on a SDS-PAGE and stained with Coomassie blue (Fig 23).

Dot blot assays for hemoglobin and hemoglobin-haptoglobin binding are performed essentially the same as described in chapter II and III. Results indicate that *E. coli* harboring pXHGP' binds hemoglobin and hemoglobin-haptoglobin complex (data not shown). Therefore, CCAA repeats are not required for HgpA to bind hemoglobin and hemoglobin-haptoglobin complex in *E. coli*.

Figure 23. Analysis of the recombinant protein expression in *E. coli* BL21(DE3)pLysS. 6.5% SDS-PAGE gel stained with Coomassie blue. Lane 1, affinity purified hemoglobin binding protein from HI689. Lane 2, molecular weight marker; lane 3, whole cell lysates of *E. coli* BL21(DE3)pLysS harboring pXHGP'. Lane 4, whole cell lysates of *E. coli* BL21(DE3)pLysS harboring pRESTA.