Cha-Cha-Cha: Variable Adhesive Activity of the *Haemophilus* Cryptic Genospecies Trimeric Autotransporter Cha

by

Amanda Joan Sheets

Department of Molecular Genetics and Microbiology
Duke University

Date:_____
Approved:

_____
Dr. Joseph St. Geme III, Advisor

_____
Dr. Raphael Valdivia

_____
Dr. Soman Abraham

_____
Dr. Margarethe Kuehn

_____
Dr. Alejandro Aballay

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Molecular Genetics and Microbiology in the Graduate School
of Duke University

2009

ABSTRACT

Cha-Cha-Cha: Variable Adhesive Activity of the *Haemophilus* Cryptic Genospecies Trimeric Autotransporter Cha

by

Amanda Joan Sheets

Department of Molecular Genetics and Microbiology
Duke University

Date:_____
Approved:

_____
Dr. Joseph St. Geme III, Advisor

_____
Dr. Raphael Valdivia

_____
Dr. Soman Abraham

_____
Dr. Margarethe Kuehn

_____
Dr. Alejandro Aballay

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Department of
Molecular Genetics and Microbiology in the Graduate School
of Duke University

2009

Copyright by
Amanda Joan Sheets
2009

Abstract

Disease caused by the gram-negative *Haemophilus* cryptic genospecies begins with colonization of the maternal genital or neonatal respiratory tract. The primary goal of this work was to identify and characterize the molecular determinant(s) of *Haemophilus* cryptic genospecies adherence as a means to better understand the specific adaptation of this species to the urogenital tract and neonatal respiratory tract. Using transposon mutagenesis of prototype strain 1595, we identified a locus that is essential for *Haemophilus* cryptic genospecies adherence to a variety of epithelial cell lines of both genital and respiratory origin. This locus encodes a protein called Cha that shares homology with trimeric autotransporters. Trimeric autotransporters are composed of an N-terminal signal peptide, an internal passenger domain that harbors adhesive activity, and a short C-terminal membrane anchor domain and are classically characterized by head-stalk-anchor domain architecture. By generating chimeric proteins, we demonstrated that the C-terminus of Cha trimerizes in the bacterial outer membrane and is capable presenting a heterologous passenger domain (Hia) in a functional form, thus confirming that Cha is a trimeric autotransporter. Southern analysis revealed that *cha* is unique to the *Haemophilus* cryptic genospecies and is ubiquitous among these strains.

Similar to a number of trimeric autotransporters, the passenger domain of Cha contains scattered clusters of YadA-like head domains associated with head-to-stalk neck adaptor motifs, predicted coiled-coil stalks and a series of identical tandem coding repeats which are not required for adherence. By evaluating the adherence capacity of *H. influenzae* expressing Cha deletion derivatives, we established that the N-terminal 473 residues of Cha harbor the binding domains responsible for Cha-mediated adherence to epithelial cells. In additional studies, we demonstrated that this same N-terminal region mediates bacterial aggregation through inter-bacterial Cha-Cha binding.

Further analysis revealed that variable Cha-mediated adherence is linked to spontaneous changes in the number of identical tandem repeats predicted to comprise a coiled-coil stalk domain. Variation in repeat copy number has a direct effect on Cha adhesive and aggregative activity, independent of an impact on transcription of the *cha* locus or surface localization of Cha protein. Moreover, length of Cha surface fibers correlates with repeat copy number expansion. We propose two hypotheses to explain how repeat expansion inhibits bacterial aggregation and host cell binding: 1) Variation in the number of 28-amino acid repeats may influence the conformation of Cha, thus changing the surface accessibility of the Cha binding pocket. 2) Repeat expansion results in the formation of long, flexible Cha fibers on the bacterial cell surface that may have a greater propensity to interact with neighboring Cha trimers at the N-terminus, thereby precluding adherence to other bacteria or host epithelial cells.

In additional studies screening adherent cryptic genospecies isolates for expression of Cha protein, we identified an additional, antigenically-divergent Cha variant that we refer to as Cha2. Amino acid sequence and domain comparison of Cha2 with Cha (now Cha1) revealed that the structurally undefined N-terminal sequences (encompassing the Cha1 adhesive and aggregative domain) are strikingly divergent. Inspite of this, Cha2 mediates efficient adherence to human epithelial cells, similar to Cha1.

Identification of Cha offers insight into the apparent tissue tropism associated with the *Haemophilus* cryptic genospecies. We speculate that the unique regulation of Cha adhesive activity enhances the adaptive capability of this pathogenic organism in the human host.

Contents

Abstract iv

List of Tables xi

List of Figures xii

Acknowledgements xiv

Chapter 1. Introduction 1

1.1 Pathogenesis of *Haemophilus* genital infections 1

1.2 *Haemophilus* cryptic genospecies: classification and characteristics 2

1.3 Bacterial virulence factors 6

1.4 Autotransporters 13

1.5 Thesis overview 16

Chapter 2. Identification of a novel trimeric autotransporter adhesin in the cryptic genospecies of *Haemophilus* 20

2.1 Introduction 20

2.2 Materials and Methods 22

2.3 Results 31

2.3.1 Adherence of *Haemophilus* cryptic genospecies strain 1595 31

2.3.2 Identification of Mariner transposon mutants defective in adherence 33

2.3.3 Predicted protein domains of the cryptic *Haemophilus* adhesin (Cha)..... 34

2.3.4 Confirmation that Cha is essential for *H.* cryptic genospecies adherence ... 36

2.3.5 Demonstration that Cha is a trimeric autotransporter adhesin 38

2.3.6 Conservation of *cha* adhesin among cryptic genospecies isolates..... 40

2.4 Discussion 41

Chapter 3. The *Haemophilus* cryptic genospecies Cha autotransporter mediates adherence and microcolony formation via a novel adhesive domain in the N-terminal end of the passenger domain. 47

3.1 Introduction 47

3.2 Materials and Methods..... 50

3.3 Results..... 55

3.3.1 Binding domain localization of Cha..... 55

3.3.2 Cha-mediated bacterial aggregation..... 59

3.4 Discussion 64

Chapter 4. Variable Adhesive Activity of the *Haemophilus* Trimeric Autotransporter Cha 69

4.1 Introduction 69

4.2 Materials and Methods..... 72

4.3 Results..... 77

4.3.1 Size polymorphism localizes to 3' *cha* locus 77

4.3.2 Effect of tandem repeats on binding to host epithelial cells 79

4.3.3 Repeat expansion interferes with bacterial aggregation..... 81

4.3.4 Repeat variation does not impact *cha* transcript levels or surface localization of Cha 83

4.3.5 Visualization of Cha fibers..... 85

4.4 Discussion 90

Chapter 5. Identification of a smaller, antigenically-divergent Cha2 adhesin in the *Haemophilus* cryptic genospecies 95

5.1 Introduction 95

5.2 Materials and Methods..... 97

5.3 Results 102

5.3.1 Identification of Cha2..... 102

5.3.2 Cha1/Cha2 domain comparison..... 104

5.3.3 Cha2 host cell binding and aggregative properties 108

5.4 Discussion 112

Chapter 6. Conclusions, Additional Observations, and Future Directions..... 116

6.1 Biological function of Cha 116

6.2 Trimeric autotransporters are multifunctional adhesins 117

6.3 Relationship between repeats, aggregation, and adherence 123

6.4 Receptor specificity 128

References 130

Biography 146

List of Tables

Table 1: Strains and plasmids used in this study 25

Table 2: Cha-mediated adherence of *Haemophilus* cryptic genospecies and *H. influenzae* to genital and respiratory epithelial cells 38

Table 3: Strains and plasmid used in this study 54

Table 4: Primers used in this study 55

Table 5: Cha2 Sequencing primers 99

List of Figures

Figure 1: Adherence of *Haemophilus* cryptic genospecies strains 420, 422, 1595, and 1673 to Chang, Detroit 562, HEC-1-B, and HeLa epithelial cells.....32

Figure 2: Genomic organization of the *cha* locus and domain architecture of the Cha protein.....36

Figure 3: Western analysis of *Haemophilus* cryptic genospecies strains 1595-A9, 1595*cha*-, 420, and 420/pCha and *H. influenzae* strains DB117/pLS88P and DB117/pCha for expression of Cha.....37

Figure 4: Evidence that Cha is a trimeric autotransporter. 40

Figure 5: Southern hybridization analysis by probing for the *cha* gene among *Haemophilus* cryptic genospecies isolates.....41

Figure 6: Domain organization and expression of Cha deletion derivatives. 58

Figure 7: Adherence to epithelial cell monolayers by *H. influenzae* strain Rd expressing Cha in-frame deletion derivatives. 59

Figure 8: Cha-mediated bacterial flocculation. 63

Figure 9: A naturally occurring size variation in the 3′ region of cha locus. 78

Figure 10: Restriction analysis of recombinant pT7-7:cha3′.....79

Figure 11: Relationship between repeat copy number and Cha adhesive activity.....81

Figure 12: Impact of Cha tandem repeat expansion/contraction on bacterial aggregation. 83

Figure 13: Variation in repeats does not affect transcription or surface localization of Cha. 85

Figure 14: Visualization of *Haemophilus* cryptic genospecies surface fibers.....87

Figure 15: The unique Cha repeat unit. 87

Figure 16: Immunogold labeling of surface fibers.....88

Figure 17: Models for impact of repeat variation on host cell binding. 89

Figure 18: Model for inhibition of bacterial aggregation by Cha repeat expansion..... 89

Figure 19: Dot immunoblot analysis of cryptic genospecies clinical isolates probing for expression of Cha. 104

Figure 20: Comparison of the predicted amino acid sequences of Cha1 from strain 1595 and Cha2 from strain 1610. 107

Figure 21: Cha2-mediated adherence to Chang, Detroit 562, HeLa, and HEC-1-B epithelial cells. 110

Figure 22: Cha2-mediated bacterial aggregation. 111

Figure 23: Interaction of Cha fibers. Electron micrograph of *Haemophilus* cryptic genospecies strain 1673 negatively stained with uranyl acetate. (Scale bar = 200 nm).. 127

Acknowledgements

I would like to thank Dr. Joe St. Geme for giving me the opportunity to join his lab and for his encouragement and guidance as a mentor throughout my graduate school experience. His patience and professionalism as a teacher and role model were a significant part of why I chose to join him in moving from Washington University in St. Louis to Duke University in the summer of 2005. I would also like to thank him for introducing me to the *Haemophilus* cryptic genospecies story and giving me the support and freedom to determine the direction of my thesis project. Thank you especially to my committee members for challenging me and for your constructive advice over the years. I am also appreciative of three special investigators with whom we have shared lab meeting with over the years, Dr. Dave Haslam, Dr. Jeff McKinney, and Dr. Pat Seed, for their thoughtful questions and helpful suggestions. I would like to say thanks to all the St. Geme lab members along the way, and there have been many, for being great co-workers, both supportive and stimulating, and overall providing a really enjoyable work environment over the years. I want to specially acknowledge Sue Grass for technical advice and trouble-shooting throughout the course of my project and for help in making the first pCha construct which proved to be a frustrating challenge along the way. I also want to acknowledge Sue for facilitating our transition from WashU to Duke; without her we would have been lost. A special thanks to my Duke classmate, Jordan Cocchiaro,

whom I met when we both interviewed at WashU in 2003, for her friendship and help in adapting to life at Duke. Also, a huge thanks goes to her and Jessica McCann for a crash course in using Adobe Illustrator for editing figures used in this work. Thanks also to Stacey Prow Bateman, Nicole Spahich, and Jessica McCann for helpful input in writing parts of this thesis.

Finally, I would like to thank my family and especially my parents for their consistent support, particularly over the last few months. Without their faith and encouragement, I would not be where I am today. This is as much their accomplishment as it is mine.

Chapter 1. Introduction

1.1 Pathogenesis of *Haemophilus* genital infections

Haemophilus influenzae is a well-described human-specific gram-negative coccobacillus that causes a variety of infections ranging from localized respiratory infections to invasive diseases. More recently, *Haemophilus* species have been recognized as commensal organisms of the lower genital tracts (Quentin *et al.* 1989). Genital *Haemophilus influenzae* and *Haemophilus parainfluenzae* generally behave as opportunistic pathogens, capable of causing an array of infections, including cervicovaginitis, urethritis, endometritis, pelvic inflammatory disease (PID), and mother-infant infections (Quentin *et al.* 1989). Whereas *H. parainfluenzae* is more often isolated in male genital infections, *H. influenzae* prevails in female genital infections and during pregnancy (Quentin *et al.* 1989). It is presumed that *Haemophilus* species colonize the anogenital area and, under certain predisposing conditions, such as the immunosuppressive effects of pregnancy, are able to cause disease (Martel *et al.* 1989). Childbirth, leiomyomas, and intrauterine devices have also been implicated as potentiating factors in the development of upper genital tract *Haemophilus* infections (Quentin *et al.* 1989).

Although still under-recognized, the frequency of cases of neonatal, perinatal, and obstetric infections due to *Haemophilus* species has increased significantly since the 1970s (Quentin *et al.* 1989). In a study conducted between 1971-81 in Houston, TX Wallace and colleagues reported an increase in *H. influenzae*-positive blood cultures in

mothers and neonates from 0.08% (1971-75) to 2.5% (1976-81) and an incidence of neonatal *Haemophilus* bacteremia of 0.14 per 1000 live births (Wallace *et al.* 1983). Another study in Tours, France, reported that the rate of *Haemophilus* mother-infant infections increased from 0.64/1000 births in 1979-1984 to 2.83/1000 births in 1985-1987, reflecting a four-fold increase (Quentin *et al.* 1989). Neonatal infections due to *Haemophilus* species share clinical features similar to infections due to *Streptococcus agalactiae* (Group B Strep) aside from the infrequent occurrence of meningitis (Gill *et al.* 1995). Infections, often associated with premature onset of labor and preterm rupture of membranes, are believed to be acquired *in utero* through the ascending spread of bacteria from the lower genital tract (Martinez *et al.* 1999). *Haemophilus* species cause urinary tract infections, cervigovaginitis, and chorioamnionitis in expectant mothers, and babies may ingest or aspirate *Haemophilus*-containing amniotic or vaginal fluids during labor and delivery. Bacterial colonization of the neonatal airway may progress to respiratory distress syndrome and in some cases to systemic infection, with fever, shock, jaundice, and leukopenia (Quentin *et al.* 1989).

1.2 *Haemophilus* cryptic genospecies: classification and characteristics

There are eight recognized biotypes for *H. influenzae* clinical isolates. Biotyping classifies *Haemophilus* strains based on three enzymatic and biochemical properties: indole production, urea hydrolysis, and ornithine decarboxylase activity. Non-serotypable strains of *Haemophilus influenzae* associated with urogenital infections are

most commonly biotype II and rarely biotype IV; however, in a study by Quentin *et al.* 25% of urogenital *Haemophilus* isolates were biotype IV (i.e. indole-negative, urease-positive, ornithine decarboxylase-positive) (Quentin *et al.* 1989). More significantly, biotype IV *Haemophilus* strains seem to prevail in mother-infant infections (Martel *et al.* 1989; Quentin *et al.* 1989). Studies report that as many as 38% of nontypable *Haemophilus influenzae* maternal, genital, and neonatal infections are caused by biotype IV *Haemophilus* species (Wallace *et al.* 1983; Quentin *et al.* 1989; Martinez *et al.* 1999). The frequency of this biotype reported in mother-infant infections suggests that *Haemophilus* biotype IV represents a specific genital and neonatal pathogen (Quentin *et al.* 1989). It has been purported that this biotype distribution may be a consequence of a specific colonization tropism by this particular biotype as well as due to a selective advantage of biotype IV strains in the ecological environment of the anogenital area (Martel *et al.* 1989).

Quentin and colleagues first noted that the majority of biotype IV *Haemophilus influenzae* strains share unique genotypic and phenotypic characteristics and suggested these strains might represent a new species (Quentin *et al.* 1990). Multilocus enzyme electrophoresis estimates overall genetic relatedness among bacterial isolates by examination of allelic variation at a few structural gene loci on the chromosome. An allelic analysis of ten metabolic enzymes in *Haemophilus* using 117 isolates from genital and neonatal infections revealed a clear genetic divergence of a population of biotype IV

Haemophilus from all other biotypes examined (Quentin *et al.* 1990). In a comparison of electrophoretic mobility for all ten enzymes assayed, 86 distinctive multilocus combinations or electrophoretic types (ETs) were identified (Quentin *et al.* 1989; Quentin *et al.* 1990). The non-typable biotype IV isolates clustered within the most genetically divergent group of ETs (Quentin *et al.* 1990). Genomic DNA-DNA hybridization experiments revealed that the clustered strains of biotype IV *Haemophilus* were only ~60% related to *Haemophilus influenzae* strain NCTN 8143. This hybridization value is below the 70% overall genetic similarity threshold generally accepted as the criterion of species limits and therefore suggested that these biotype IV isolates represent a formerly unrecognized species of *Haemophilus*. This unique group of biotype IV isolates with divergent electrophoretic types was designated the *Haemophilus* cryptic genospecies (Quentin *et al.* 1990).

Quentin and colleagues reported that ribosomal DNA restriction fragment length polymorphism (rDNA RFLP) patterns obtained for 18 *Haemophilus* cryptic genospecies strains were relatively homogeneous and differ noticeably from reference strains of *Haemophilus influenzae sensu stricto* (Quentin *et al.* 1993). DNA-DNA hybridization and DNA reassociation experiments indicated that the *Haemophilus* cryptic genospecies is more closely related to *H. haemolyticus* than to *H. influenzae* (Quentin *et al.* 1993). Sequence analysis of the small subunit (16S) rRNA gene demonstrated that the cryptic genospecies isolates form a monophyletic unit with *H. haemolyticus* and *H. influenzae* and

reaffirmed that the cryptic genospecies is more closely related to *H. haemolyticus* (Quentin *et al.* 1996). In addition to the unique 16S rDNA sequence, expression of an active periplasmic Cu/Zn-superoxide dismutase by cryptic genospecies isolates differentiates them from *Haemophilus sensu stricto* (Langford *et al.* 2002). Although this species was typically identified as *H. influenzae* biotype IV in biochemical tests, it has been recently demonstrated that other biotypes can belong to this cryptic genospecies (Gill *et al.* 1995; Quentin *et al.* 1996; Martinez *et al.* 1999).

Outer membrane electrophoretic patterns are generally heterogeneous for *Haemophilus influenzae* biotypes I, II, III, V, and VI (Quentin *et al.* 1989). However, isolates of the cryptic genospecies display a unique and homogeneous outer membrane protein pattern, including expression of a unique 18 kDa protein (Quentin *et al.* 1989) and a variant P6 OMP, a peptidoglycan-associated lipoprotein that is highly conserved among *Haemophilus influenzae* strains (Quentin *et al.* 1989; Quentin *et al.* 1990). Cryptic genospecies strains are frequently characterized by abundant peritrichous piliation visualized by negative staining electron microscopy (Quentin *et al.* 1989). Rosenau and colleagues reported that most cryptic genospecies strains (12 of 17) were highly adherent to cultured HeLa (human cervical epithelioid carcinoma) cells, and adherence properties correlated with piliation (Rosenau *et al.* 1993). Cryptic genospecies isolates have an enhanced affinity for HeLa as compared to Hep-2 (human laryngeal epidermoid carcinoma) cells, an observation that may be interpreted as an apparent tropism of

cryptic genospecies strains for genital tissues (Rosenau *et al.* 1993). Despite high adhesive activity, low rates of invasion were reported, as less than 0.2% of associated bacteria invaded either HeLa or Hep-2 cells suggesting that pathogenic mechanisms other than cell invasion are involved in the development of disease (Rosenau *et al.* 1993).

1.3 Bacterial virulence factors

While the pathogenic mechanisms of *H. influenzae* have been studied extensively, very few virulence determinants have been identified in the cryptic genospecies. The following information reviews key bacterial virulence factors, in particular highlighting what has been reported for *H. influenzae* and where possible, what is known about the cryptic genospecies.

Evasion of Host Immune Response

Superoxide dismutases (SODs) play a key role in the enzymatic defense against oxygen toxicity. Broadly conserved Mn/Fe-SODs are found in both prokaryotes and eukaryotes and function in the dismutation of superoxide, produced endogenously during cellular respiration (Kroll *et al.* 1991). While Cu/Zn-SODs are widely conserved among eukaryotes, they have been rarely identified in bacteria (Kroll *et al.* 1991; Kroll *et al.* 1995). Periplasmic dismutases can defend gram-negative bacteria against exogenously derived superoxide radicals produced during the respiratory burst of phagocytic cells (Kroll *et al.* 1991). Consistent with this idea, *sodC* mutants of *Haemophilus ducreyi*, a sexually transmitted causative agent of genital ulcers (chancroid),

are more susceptible to extracellular superoxide than wild type (San Mateo *et al.* 1998). *Salmonella enterica* and *Neisseria meningitides sodC* mutants have also been shown to be less virulent in mouse models of infection (Farrant *et al.* 1997; Wilks *et al.* 1998). The expression of an active *sodC*-encoded, periplasmic Cu/Zn-superoxide dismutase in the cryptic genospecies raises an interesting question as to whether the enzyme enhances virulence by facilitating persistence in the genital environment. Despite conservation of a *sodC* locus, *H. influenzae* strains do not express a functional periplasmic SOD (Kroll *et al.* 1991). It is tempting to speculate that *sodC* expression may confer a survival advantage to the cryptic genospecies in the genital tract, although this hypothesis awaits the development of a suitable *in vivo* model.

The principal host immune defense to control bacterial colonization of mucosal surfaces in the respiratory and/or genital tract is secretory immunoglobulin A (S-IgA). S-IgA agglutinates bacteria, inhibiting adherence and facilitating clearance by peristaltic and mucociliary movements, a mechanism called immune exclusion. To combat this anti-colonization activity of S-IgA, numerous urogenital and oropharyngeal organisms have evolved IgA-specific proteases (Kilian 1996). An IgA1 protease secreted by *Haemophilus influenzae* inactivates human IgA1 via cleavage of the IgA1 hinge region (Plaut 1978).

As another means of evading the host immune response, several pathogens express proteins that bind immunoglobulins (IgA and IgG) via the Fc fragment in a

nonimmune manner (Anthony *et al.* 1990; Sandt and Hill 2001; Ramsland *et al.* 2007). Nonimmune binding inhibits Ig-dependent cellular effector functions mediated by the Fc receptor, such as phagocytosis, degranulation, and respiratory burst. This activity is likely to facilitate survival in mucosal environments and may contribute to systemic infections by attenuating serum sensitivity.

To survive in the bloodstream, bacteria need to inhibit complement-mediated bacteriolysis and avoid phagocytic uptake. Successful pathogens subvert the innate immune system by destroying host complement activation components, mimicking complement inhibitors, and binding of host complement inhibitors (Wurzner 1999; Hornef *et al.* 2002). To facilitate intravascular survival, the majority of systemic *H. influenzae* isolates express polysaccharide capsule (Sutton *et al.* 1982). Encapsulated *H. influenzae* belongs to one of six serotypes (a-f), of which serotype b (Hib) is the most virulent. Studies have shown that Hib bacteria express proteins that bind factor H and factor H-like protein 1, two inhibitors of the alternative complement activation pathway (Hallstrom *et al.* 2008) and that expression of type b capsule significantly decreases complement-dependent killing (Noel *et al.* 1996). Non-encapsulated organisms show greatly attenuated virulence in the infant rat model of systemic infection (Moxon and Vaughn 1981). While *Haemophilus* cryptic genospecies isolates are non-encapsulated (Musser *et al.* 1986), the association of these organisms with neonatal systemic infections

suggests that other non-capsule surface components function to inhibit complement and facilitate intravascular survival.

Nutrient Acquisition

Bacteria require iron to survive and thrive in the nutrient limiting environment of the human host. In addition to facilitating basic cellular processes, iron is involved in the regulation of a variety of virulence factors (Litwin and Calderwood 1993). In mammals, the level of free iron is too low to sustain bacterial growth as the majority is bound to ferric-binding proteins (transferrin, lactoferrin, and ferritin) and hemoproteins (hemoglobin, haptoglobin-hemoglobin, hemopexin, and albumin) (Ratledge and Dover 2000; Sebulsky *et al.* 2000). In response to iron limitation, successful pathogenic microorganisms have evolved diverse iron uptake systems with which to scavenge iron from the host environment. *Haemophilus influenzae* expresses two surface-associated transferrin-binding proteins (Tbp1 and Tbp2) that scavenge iron from the extracellular environment (Holland *et al.* 1992). *H. influenzae* also expresses a periplasmic iron-binding protein called FbpA that competes for transferrin-bound iron (Khan *et al.* 2007), which is subsequently transported into the bacterial cytoplasm by HitB and HitC (Adhikari *et al.* 1995).

Haemophilus species additionally require an exogenous heme source for aerobic growth. *H. influenzae* expresses a variety of proteins to aid in this process. Heme acquisition from hemoglobin-haptoglobin complexes is mediated by Hgp proteins (Jin *et*

al. 1996), while the HxuABC proteins facilitate binding and transport of hemopexin-associated heme-Fe into the bacterial cytoplasm (Cope *et al.* 1995). Mutations disrupting both the Hgp- and Hxu-mediated pathways of heme acquisition ablate virulence in the infant rat infection model of human invasive disease, indicating that these proteins represent key virulence determinants (Jin *et al.* 1996).

Adhesins

The pathogenesis of disease caused by *H. influenzae* and the cryptic genospecies begins with colonization of the genital and/or respiratory tract. *H. influenzae* expresses a variety of surface structures, including Hap, HMW1/HMW2, Hia/Hsf, Hif pili, OMP-P5, and lipooligosaccharide to mediate stable attachment to the respiratory epithelial cell surface (Rao *et al.* 1999). The haemagglutinating (Hif) pili of *H. influenzae* are polymeric helical structures displayed in a peritrichous distribution, which mediate agglutination of AnWj-positive human erythrocytes (Guerina *et al.* 1982; Pichichero *et al.* 1982). Hif pili promote adherence to human oropharyngeal epithelial cells, human nasopharyngeal and nasal tissue, respiratory mucus, and fibronectin (Loeb *et al.* 1988; Farley *et al.* 1990; Kubiet *et al.* 2000; Virkola *et al.* 2000). The *H. influenzae* pilus shaft consists of repeating HifA subunits joined to a short adhesive tip fibrillum composed of HifD and HifE (St Geme *et al.* 1996). Analysis of collections of non-typeable *H. influenzae* isolates indicate that approximately 15% of strains encode the intact *hif* pilus gene cluster and are able to express pili (Geluk *et al.* 1998; Krasan *et al.* 1999). Rosenau and colleagues have reported

that biotype IV cryptic genospecies strains are characterized by peritrichous piliation as well; however, these fibers are thinner, shorter, and denser than the *H. influenzae* Hif pili. Cryptic genospecies isolates are unable to agglutinate human erythrocytes (Clemans *et al.* 2001), and attempts to enrich for piliation through agglutination of human erythrocytes have failed (Rosenau *et al.* 1993). BLAST searches for the *hif*ABCDE genes in the genome of cryptic genospecies strain 1595 indicate that the entire *hif* locus is absent. Other groups have reported that among a subset of cryptic genospecies strains found to encode Hif pili, none express immunoreactive Hif pili (Gousset *et al.* 1999; Clemans *et al.* 2001; Bruant *et al.* 2002). Several investigators have reported a significant correlation between the expression of surface fibers and adherence; however, the identity and role of these fibers in the pathogenesis of the cryptic genospecies remains to be elucidated (Rosenau *et al.* 1993; Clemans *et al.* 2001).

The *H. influenzae* Hap protein mediates attachment to and invasion of a variety of cell types as well as to extracellular matrix proteins and is universally conserved among clinical isolates (St Geme *et al.* 1994; Fink *et al.* 2002). Hap is a serine protease capable of autoproteolysis and has also been shown to mediate bacterial aggregation and microcolony formation (Hendrixson and St Geme 1998). HMW1 and HMW2 are high molecular weight proteins that mediate attachment to a number of epithelial cell lines and are expressed by ~75% of NTHi strains (St Geme *et al.* 1993; St Geme *et al.* 1998). The Hia adhesin is expressed in most of the NTHi that do not express the HMW

adhesins (St Geme *et al.* 1998). The Hsf adhesin is a homolog of Hia that is larger and ubiquitous among encapsulated strains of *H. influenzae.* It forms thin, short surface fibers which confer the same binding specificity as Hia and mediate attachment to Chang conjunctival, HeLa cervical, and HEp-2 laryngeal epithelial cell lines (St Geme *et al.* 1996). The P2 protein in *H. influenzae* is an antigenically variable outer membrane porin that binds to nasopharyngeal mucin via sialic acid-containing oligosaccharides (Reddy *et al.* 1996) Frequent size and sequence variation associated with the P2 protein within clonal populations during the course of chronic infection enables the organism to evade clearance by potentially protective antibodies (Forbes *et al.* 1992; Smith-Vaughan *et al.* 1997). The P5 protein is another major outer membrane protein expressed by nontypable *H. influenzae* which is antigenically variable and has been shown to enhance adherence to oropharyngeal cells and mucin as well as mediate attachment to CHO and HeLa cells via CEACAM1 (Reddy *et al.* 1996; Hill *et al.* 2001; Bakaletz *et al.* 2005; Bookwalter *et al.* 2008). P5 shares homology with the *E. coli* OmpA and forms pilus-like surface appendages, leading to the designation 'P5 fimbrin' by some investigators (Sirakova *et al.* 1994). An *ompp5* homologue appears to be widely conserved among *H.* cryptic genospecies strains; however, this fimbrin-like gene is not responsible for the observed piliation, and it is doubtful that this protein constitutes the major adhesin expressed by these isolates (Gousset *et al.* 1999).

In recent work, Swords and colleagues demonstrated that lipooligosaccharide (LOS) expressed by *H. influenzae* also contributes to NTHi adherence and invasion of bronchial cells (Swords *et al.* 2000; Swords *et al.* 2001). This activity involves a specific interaction between a phosphorylcholine (ChoP) moiety within NTHi LOS and the PAF receptor and represents an example of molecular mimicry of host molecules. Although LOS is an effective adhesin, it is selected against in systemic isolates. However, *Haemophilus* (like other bacteria) is capable of modifying its LOS structure during infection (Mandrell and Apicella 1993; Weiser 1993). Findings by Weiser *et al.* show that expression of low molecular weight LOS by phase variants enhances serum resistance and intravascular survival (Weiser *et al.* 1990; Weiser 1993).

1.4 Autotransporters

The afimbrial Hap, Hia, and Hsf adhesins expressed by *H. influenzae* are all classified as autotransporter proteins (Type V secreted proteins). Autotransporters (ATs) are secreted by gram-negative bacteria and were historically considered to be unique in that their primary structure contains all the information necessary for transport from the periplasm to the cell surface (Wells *et al.* 2007). They contain an N-terminal signal sequence, an internal passenger domain, and a C-terminal membrane anchor domain and can be divided into two groups, namely, the classical autotransporters (Va) and the trimeric autotransporters (Vc). However, recently studies reporting that the Omp85/YaeT/Bam complex is required for integration of all outer

membrane proteins (including autotransporters) into the outer membrane have challenged the idea of spontaneous "autotransport" (Voulhoux *et al.* 2003; Bernstein 2007).

The N-terminal signal peptide directs secretion of the autotransporter across the inner membrane in a Sec-dependent manner. Signal sequences have an n-domain with positively charged amino acids followed by an h-domain with hydrophobic amino acids and a signal peptidase I cleavage site (Martoglio and Dobberstein 1998). Most signal peptides are ~25 amino acids in length; however, some autotransporters exhibit an atypical signal sequence with a conserved extension in the n-domain (Henderson *et al.* 1998; St Geme and Cutter 2000). Following translocation into the periplasm, the C-terminal domain inserts into the outer membrane and facilitates transport and presentation of the passenger domain. The membrane anchor domain of classical, monomeric autotransporters forms a β-barrel pore with 12 transmembrane β-strands and a transmembrane α-helical linker region (Oomen *et al.* 2004). Deletion of the linker region inhibits surface localization, suggesting the α-helix and β-barrel regions together function as the minimal translocation unit (Maurer *et al.* 1999; Oliver *et al.* 2003).

While the C-terminal domains associated with translocation of the mature protein are highly homologous among ATs, the extracellular passenger domains exhibit considerable sequence variation. Consequently, AT passenger domains confer a wide range of virulence-associated effector functions (sometimes in combination), including

adhesion, invasion, protease activity, cytotoxicity, and serum resistance. Some additionally mediate aggregation and are classified as self-associating autotransporters (SAATs). Antigen 43 (Ag43), AIDA, and the TibA adhesin/invasion expressed in *E. coli,* mediate homo- and heterologous interactions to facilitate formation of mixed bacterial aggregates and enhance biofilm formation (Klemm *et al.* 2006).

Trimeric autotransporters (sometimes referred to as the oligomeric coiled-coil adhesin (Oca) family) are a rapidly growing subfamily of autotransporters which have been identified in both animal and plant pathogenic bacteria. The prototypic members of this subfamily are the *Yersinia enterocolitica* YadA adhesin and the *Haemophilus influenzae* Hia adhesin. The multifaceted YadA protein mediates serum resistance and adherence to epithelial cells and extracellular matrix proteins (El Tahir and Skurnik 2001). Like classical autotransporters, all trimeric autotransporters contain an N-terminal signal sequence and an internal passenger domain that is presented on the bacterial cell surface and harbors the effector functions of the protein. Unlike the classic ATs, trimeric autotransporters have an unusually short C-terminal region consisting of an α-helical linker and 4 β-strands. This conserved sequence motif (defined as pFam03895) differs from the β-domain of monomeric TAs (pFam03797) and trimerizes in the outer membrane, forming a β-barrel pore consisting of 12 β-strands capable of translocating the passenger domain (Roggenkamp *et al.* 2003; Surana *et al.* 2004; Ackermann *et al.* 2008). All TAAs form heat-resistant oligomers and contain a passenger

domain that remains uncleaved (Elkins *et al.* 2000; Sandt and Hill 2000; Ray *et al.* 2002; Riess *et al.* 2004). Further studies with Hia have demonstrated that trimerization of the passenger domain is required for protein stability and adhesive activity (Cotter *et al.* 2006). While TAAs are defined by conserved membrane anchor homology, the passengers domains contain a variety of structural motifs categorized into head, neck, and stalk domains (Szczesny and Lupas 2008). Similar to a number of TAAs, YadA molecules are visible by negative staining electron microscopy and have been described as having a lollipop appearance (Hoiczyk *et al.* 2000; Szczesny *et al.* 2008). A cluster Hep_Hag degenerate repeats, a β-roll-forming sequence element named for the *Moraxella* haemagglutinin Hag adhesin (Pearson *et al.* 2002), comprises the adhesive head domain in YadA. Conservation of common structural motifs in different arrangements with varied activities among TAA passenger domains may reflect bacterial adaptations to different physiological niches and binding targets. TAA domains have been given different names throughout the literature. In this thesis, Hep_Hag and YadA-like head repeats refer to the same sequence motif, and membrane anchor and C-terminal domain are used interchangeably.

1.5 Thesis overview

Disease caused by the *Haemophilus* cryptic genospecies begins with colonization of the maternal genital or neonatal respiratory tract. The primary goal of this work was to identify the molecular determinant(s) of *Haemophilus* cryptic genospecies adherence.

In pursuit of this aim, this thesis focuses on the characterization of the novel adhesin expressed by the *Haemophilus* cryptic genospecies as a means to better understand the specific adaptation of this species to the urogenital tract and neonatal respiratory tract. Using transposon mutagenesis of prototype strain 1595, we identified a locus that encodes a new trimeric autotransporter, designated Cha (for cryptic *Haemophilus* adhesin). Studies demonstrated that expression of Cha (later referred to as Cha1) is essential for cryptic genospecies adherence to a variety of epithelial cell lines of both genital and respiratory origin. Further analysis confirmed that the C-terminus of Cha trimerizes in the outer membrane and is capable presenting a heterologous passenger domain (Hia) in a functional form, confirming that Cha is a trimeric autotransporter. Southern analysis revealed that *cha* genes are conserved among cryptic genospecies strains and absent from the closely related *Haemophilus influenzae*. This finding leads us to speculate that expression of Cha contributes to the apparent tissue tropism associated with the *Haemophilus* cryptic genospecies.

Similar to a number of other bacterial adhesins, the passenger domain of Cha contains scattered clusters of YadA-like head domains, each associated with head-to-stalk neck adaptor motifs. In addition, the passenger domain contains predicted coiled-coil stalks and a series of identical tandem coding repeats which are not required for adherence. By evaluating the adherence capacity of *H. influenzae* expressing Cha with internal deletions, we established that the N-terminal 473 residues of Cha harbor the

binding domain responsible for Cha-mediated adherence to epithelial cells. This same N-terminal region facilitates bacterial aggregation through inter-bacterial Cha-Cha binding. In the process of screening our adherent cryptic genospecies strains for expression of Cha, we identified an antigenically divergent *cha* allele, designated *cha2,* present at the same locus in a subset of our clinical isolates. Interestingly, based on predicted amino acid sequences, Cha1 and Cha2 are strikingly divergent in the N-terminal 315 residues of the passenger domain.

In this work, we report that variable Cha-mediated adherence is linked to spontaneous changes in the number of 28-amino acid repeats. Variation in repeat copy number has a direct effect on Cha adhesive and aggregative activity, independent of an impact on transcription of the *cha* locus or surface localization of Cha protein. Moreover, visualization and length of Cha surface fibers correlates with repeat copy number expansion. We propose two hypotheses to explain how repeat expansion inhibits bacterial aggregation and host cell binding: 1) Variation in the number of 28-amino acid repeats may influence the conformation of Cha, thus changing the surface accessibility of the Cha binding pocket, and/or 2) Repeat expansion results in the formation of long, flexible Cha fibers on the bacterial cell surface which may have a greater propensity to interact with neighboring Cha trimers at the N-terminus, thereby precluding adherence to other bacteria or host epithelial cells. We speculate that this unique system of

regulation enhances the pathogenicity of the organism by increasing its adaptive capability in the human host.

Chapter 2. Identification of a novel trimeric autotransporter adhesin in the cryptic genospecies of *Haemophilus*

2.1 Introduction

Over the past few decades the recognition of *Haemophilus* urogenital and neonatal infections has risen significantly (Wallace *et al.* 1983; Campognone and Singer 1986; Friesen and Cho 1986; Quentin *et al.* 1989). *Haemophilus* strains belonging to biotype IV account for up to 38% of maternal genital tract and neonatal *Haemophilus* infections but are rarely isolated from invasive or respiratory infections in older children or adults, suggesting a genital and neonatal tropism (Albritton *et al.* 1982; Wallace *et al.* 1983; Friesen and Cho 1986; Martel *et al.* 1989; Rosenau *et al.* 1993). Multilocus enzyme electrophoretic analysis, DNA-DNA hybridization, and 16S RNA sequence analysis have shown that these biotype IV *Haemophilus* strains are genetically homogeneous and are sufficiently distinct from *H. influenzae* to represent a new species, designated the *Haemophilus* cryptic genospecies (Musser *et al.* 1986; Quentin *et al.* 1993; Rosenau *et al.* 1993; Quentin *et al.* 1996). These biotype IV isolates are further characterized by a distinctive outer membrane protein profile and a specific P6 outer membrane protein amino acid sequence (Quentin *et al.* 1989; Murphy *et al.* 1992; Quentin *et al.* 1996). Although isolates belonging to the cryptic genospecies are typically identified as *H. influenzae* biotype IV in the clinical microbiology laboratory (Quentin *et al.* 1989), recent

data suggest that the cryptic genospecies also includes other biotypes (Gill *et al.* 1995; Martinez *et al.* 1999).

Colonization of the genital tract is a necessary prerequisite for development of *Haemophilus* cryptic genospecies urogenital infections, including urethritis, vaginitis, cervicitis, salpingitis, endometritis, and bartholin abscess, among others (Turk 1984; Quentin *et al.* 1989; St Geme 1993). Infected neonates likely acquire the *Haemophilus* cryptic genospecies during passage through the birth canal, leading to colonization of the respiratory tract and then respiratory distress and sepsis, typically within hours of birth (Wallace *et al.* 1983; Campognone and Singer 1986; Quentin *et al.* 1989; Rusin *et al.* 1991). Neonatal disease occurs most commonly in premature infants (Wallace *et al.* 1983; Friesen and Cho 1986).

Previous studies have reported a correlation between the appearance of peritrichous pilus-like structures in *Haemophilus* cryptic genospecies strains and an ability to adhere to HeLa cells and to a lesser extent to Hep-2 cells (Quentin *et al.* 1989; Rosenau *et al.* 1993; Gousset *et al.* 1999). Initial work suggested that these fibers might be related to *H. influenzae* hemagglutinating pili, which are encoded by the *hif* gene cluster consisting of 5 genes, designated *hifA-E*. In particular, PCR analysis has demonstrated that a subset of *Haemophilus* cryptic genospecies isolates contain *hif*-like genes (Gousset *et al.* 1999; Clemans *et al.* 2001; Bruant *et al.* 2002). However, antisera against *H. influenzae* HifA and HifE proteins are non-reactive with these isolates (Clemans *et al.*

2001). Furthermore, the short, thin surface fibers observed in the *Haemophilus* cryptic genospecies differ in appearance from *H. influenzae* hemagglutinating pili, which are longer and thicker (Bruant *et al.* 2002). Taken together, these observations suggest that *Haemophilus* cryptic genospecies surface fibers are distinct from *H. influenzae* hemagglutinating Hif pili and raise the possibility that a novel molecule accounts for *Haemophilus* cryptic genospecies adherence.

In this study we sought to elucidate the genetic determinants of *Haemophilus* cryptic genospecies adherence. Using transposon mutagenesis, we identified a new locus that encodes a trimeric autotransporter adhesin which we designated Cha. Disruption of this locus resulted in loss of adherence. Further analysis confirmed that the Cha C-terminus trimerizes in the outer membrane and is both necessary and sufficient for surface presentation of a heterologous passenger domain. Of note, the Cha-encoding locus appears to be uniformly conserved among isolates belonging to the *Haemophilus* cryptic genospecies, perhaps accounting for the tissue tropism and distinctive disease associations.

2.2 Materials and Methods

Culture and storage conditions

E. coli strains were grown on Luria-Bertani (LB) agar or in LB broth at 37°C and were stored at −80°C in LB broth with 50% glycerol. *Haemophilus* strains were grown on chocolate agar plates (Becton-Dickson, Franklin, NJ) at 37°C with 5% CO_2 and were

stored at −80°C in brain heart infusion broth with 25% glycerol. For *E. coli*, 100 μg/ml ampicillin or 50 μg/ml kanamycin was used as appropriate for plasmid selection. For *H. influenzae* and the *Haemophilus* cryptic genospecies, kanamycin was used at 25 μg/ml and 50 μg/ml, respectively.

Bacterial strains and plasmid construction

Bacterial strains and plasmids are described in Table 1. *Haemophilus* cryptic genospecies strain 1595 exhibits phase variable adherence, allowing enrichment of adherent variants for loss of adherence and enrichment of non-adherent variants for re-acquisition of adherence using epithelial cell monolayers. In the process of screening revertant adherent variants for loss of adherence, we identified a clone designated 1595-A9 that is stably adherent to Chang cells, refractory to further phase variation.

To clone the *cha* gene, coding sequence for Cha and upstream sequence (including a native NcoI site 383 nucleotides upstream of the translational start site) was amplified from wildtype strain 1595 using primers CAATGAAACGAATTTACAAAGCTACTCTATTCTC and AAGAGCCATATGTTTACCAAATAAATGACAAAAATACCGCTC and was ligated into pCR®-XL-TOPO (Invitrogen), generating pTOPO::Cha. The *cha* gene was excised from pTOPO::Cha by digesting with NcoI and BamHI and was ligated into NcoI/BamHI digested pLS88P, generating pCha.

The plasmid pHAT::HiaSS contains the Hia promoter and signal peptide coding

sequence (amino acids 1 to 49) fused to the HAT epitope tag and was generated previously (Surana *et al.* 2004). The plasmid pHAT::Hia$_{(P\text{-}49/50\text{-}779)}$ (also called pHiaPD) encodes the HAT-tagged Hia signal peptide (amino acids 1 to 49) fused to residues 50 to 779 of the passenger domain of Hia (containing both Hia binding domains). To generate this plasmid, pHAT::HiaSS was digested with SalI and BamHI and ligated to a SalI/BamHI-digested fragment amplified from *H. influenzae* strain 11 and encoding Hia residues 50-779. The plasmid pHAT::Hia$_{(P\text{-}49/50\text{-}779/977\text{-}1098)}$ (also called pHiaPD-HiaC') encodes the Hia signal peptide and passenger domain (amino acids 50-779) fused to the Hia translocator domain (amino acids 977-1098). To generate this plasmid, pHAT::Hia$_{(P\text{-}49/50\text{-}779)}$ was digested with SacI and EcoRI and ligated to a SacI/EcoRI-digested fragment encoding the C-terminal 122 residues of Hia (amino acids 977 to 1098). The chimeric plasmid pHAT::Hia$_{(P\text{-}49/50\text{-}779)}$-Cha$_{(1491\text{-}1610)}$ (also called pHiaPD-ChaC') encodes a chimeric protein containing the Hia signal peptide and passenger domain fused to the Cha C-terminus. To generate this plasmid, pHAT::Hia$_{(P\text{-}49/50\text{-}779)}$ was digested with SacI and EcoRI and ligated to a SacI/EcoRI-digested fragment encoding the C-terminal 120 residues of Cha (amino acids 1491 to 1610) amplified by PCR from pTOPO::Cha using primers GGGAGCTCATCGGTTCACAAGGTAATGAACGTCGA and GGAATTCTTACCATTGGTAACCAATACCAGCACC containing SacI and EcoRI restriction sites, respectively. The plasmid pHAT::ChaC' encodes the C-terminal 120 amino acids and was generated by ligating the SacI/EcoRI-digested Cha fragment

encoding Cha amino acids 1491-1610 to SacI/EcoRI-digested pHAT::HiaSS.

Table 1: Strains and plasmids used in this study

Strains	Description	Source/Reference
420	*Haemophilus* cryptic genospecies neonatal blood isolate	J. Musser
421	*Haemophilus* cryptic genospecies neonatal blood isolate	(Quentin *et al.* 1993)
422	*Haemophilus* cryptic genospecies neonatal blood isolate	(Quentin *et al.* 1993)
427	*Haemophilus* cryptic genospecies neonatal gastric isolate	(Quentin *et al.* 1990)
1595	*Haemophilus* cryptic genospecies neonatal blood isolate	(Musser *et al.* 1986)
1610	*Haemophilus* cryptic genospecies neonatal blood isolate	(Musser *et al.* 1986)
1673	*Haemophilus* cryptic genospecies neonatal gastric isolate	(Musser *et al.* 1986)
2446	*Haemophilus* cryptic genospecies isolate	J. Musser
2452	*Haemophilus* cryptic genospecies isolate	J. Musser
1595-A9	Stably adherent variant of strain 1595	This study
1595*cha-*	Non-adherent transposon mutant of 1595-A9	This study
DH5α	*E. coli* F- Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*) *U169 deoR recA1 endA1 hsdR17*(rK-mK+) *phoA supE441 thi-1 gyrA96 relA1*	(Sambrook *et al.* 1989)
11	*H. influenzae* strain containing *hia* locus	(St Geme *et al.* 1996)
DB117	Derivative of *H. influenzae* strain Rd that is *rec-*	(Setlow *et al.* 1968)
Plasmids		
pTOPO::Cha	pCR®-XL-TOPO (Invitrogen) containing *cha* promoter and coding sequence	This study
pLS88P	derivative of pLS88 (ATCC 86980) with BamHI site engineered at EcoRI site	This study
pCha	pLS88P containing *cha* promoter and coding sequence	This study
pHAT::HiaSS	pHAT10 containing *hia* promoter and coding sequence for Hia signal peptide: Hia(P-49)	(Surana *et al.* 2004)
pHiaPD	pHAT10 containing coding sequence for Hia signal peptide and passenger domain: Hia(P-49/50-779)	This study
pHiaPD-HiaC′	pHAT10 containing coding sequence for Hia signal peptide, passenger domain, and C-terminal 122 residues: Hia(P-49/50-779/977-1098)	This study
pHiaPD-ChaC′	pHAT10 containing coding sequence for Hia signal peptide, passenger domain, and coding sequence for Cha C-terminal 120 residues: Hia(P-49/50-779)-Cha(1491-1610)	This study
pChaC′	pHAT10 containing coding sequence for Hia signal peptide and coding sequence for Cha C-terminal 120 residues: Hia(P-49)-Cha(1491-1610)	This study

Molecular Biology Techniques.

DNA ligations, restriction endonuclease digestions, and gel electrophoresis were performed according to standard techniques (Sambrook *et al.* 1989). Plasmids were introduced into *E. coli* by electroporation (Dower *et al.* 1988). In *H. influenzae*, transformation was performed using the MII/MIV method of Herriott *et al.* (Herriott *et al.* 1970). In the *Haemophilus* cryptic genospecies, transformation was accomplished by

incubating a 500 μl suspension of bacteria in Schaedler broth with approximately 1 μg of transforming DNA at room temperature for 30 min, then supplementing with clarified horse blood and 2 μg/ml NAD and incubating for 1 hour at 37°C with aeration. Ultimately, transformation reaction mixtures were plated onto chocolate agar containing appropriate antibiotics to select for transformants.

Generation of Cha antiserum

A 1212-bp fragment of the *cha* gene encoding residues 70-473 from the N-terminal region of Cha was amplified by PCR using primers CAGAATTCGCCTCTTTTACAGATAACTACACTGAGGGT and GAGAATCCCTAACCTGTCGCTGTCTTGCCTTTATTACC. This fragment was digested with EcoRI and BamHI and then ligated to EcoRI/BamHI-digested pGEX-6P-1 (GE Healthcare, Piscataway, NJ), generating pGEX::Cha70-473, which was transformed into *E. coli* DH5α. Cultures of DH5α/pGEX::Cha70-473were incubated at 37ºC to an optical density of 0.4-0.5, and then expression of the GST-Cha fusion protein was induced with 0.1 mM isopropyl ß-D thiogalactopyranoside at 30ºC for 4 hours. Purification of the GST-Cha protein was performed as described previously with minor modifications (Yeo *et al.* 2004). Briefly, cells were harvested and lysed by sonication in lysis buffer (5 mM EDTA + 1 mM Pefabloc SC (Roche) in PBS pH 7.4), and the protein was isolated from clarified supernatant by affinity chromatography using glutathione-Sepharose beads (Pierce) according to manufacturer's instructions. Following extensive

washing with PBS, the Cha fragment was cleaved from the GST moiety using 80 U/ml PreScission Protease (Amersham Pharmacia Biotech) in cleavage buffer (50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, pH 7.5) and was injected into a guinea pig to raise a polyclonal antiserum that is reactive with formic acid denatured Cha (Cocalico Biologicals, Inc.).

Transposon library construction

To create a transposon library in *Haemophilus* cryptic genospecies strain 1595, chromosomal DNA was prepared from a stably adherent variant of strain 1595 (designated 1595-A9) using a Wizard genomic DNA kit (Promega, Madison, WI) according to the manufacturer's instructions and was mutagenized as described by Hendrixson *et al*. (Hendrixson *et al.* 2001), with modifications as described by Kehl-Fie *et al.* (Kehl-Fie and St Geme 2007). Briefly, DNA was mutagenized using Himar1 transposase and pFalcon2, a plasmid that contains a minitransposon derivative called *solo*, which carries the *aphA3* kanamycin resistance gene. In preparation for transformation, bacteria were grown for 16 h on chocolate agar and were resuspended to an optical density at 600 nm of 0.8 in Schaedler broth (BD Biosciences). Aliquots of bacteria and mutagenized DNA were mixed in eppendorf tubes, incubated standing for 30 min at room temperature, supplemented with 2% horse plasma and NAD, then incubated for 1 h at 37°C with aeration. Transformants were recovered by plating on chocolate agar containing 50 μg/ml kanamycin. Mutants obtained after transformation

with DNA from each of 10 transposition reactions were combined to create a library of ~20,000 individual random transposon mutants. To assess the randomness of transposon insertion, chromosomal DNA was extracted from individual transformants, digested with BglII, and examined by Southern hybridization, using the *aphA3* cassette from pFalcon2 as a probe.

Adherence assays

Quantitative adherence assays were performed with epithelial cells as previously described (St Geme *et al.* 1993). Percent adherence was calculated by dividing the number of adherent CFU per epithelial cell monolayer by the number of inoculated CFU. Each strain was examined in triplicate in a given assay, and each assay was performed a minimum of three times. For qualitative adherence analysis, epithelial cell monolayers with associated bacteria were stained with Geimsa stain and examined by light microscopy (St Geme and Falkow 1990). Tissue culture cells included Chang cells (human conjunctiva, Wong-Kilbourne derivative, clone 1-5c-4; ATCC CCL-20.2), Detroit 562 cells (human pharyngeal carcinoma; ATCC CCL-138), HeLa cells (human cervical epidermoid carcinoma; ATCC CCL-2), and HEC-1-B cells (human endometrium; ATCC HTB-113). Chang cells were maintained in modified Eagle medium with Earle's salts and nonessential amino acids. Detroit 562, HeLa, and HEC-1-B cells were maintained in the same medium with addition of 1 mM sodium pyruvate. All media were supplemented with 10% fetal bovine serum.

Identification of non-adherent transposon mutants

The transposon mutant library of *Haemophilus* cryptic genospecies strain 1595 was resuspended in Schaedler broth to a density of approximately 3 x 10^{10} CFU/ml. A 10 μl volume of this bacterial suspension was inoculated onto confluent Chang epithelial cell monolayers in 24-well plates. To facilitate bacterial contact with the epithelial cell monolayers, plates were centrifuged at 165g for 5 min. After a 25 minute incubation at 37ºC in 5% CO_2, the supernatant (containing non-adherent bacteria) was transferred to a new confluent Chang cell monolayer. Passage of supernatant bacteria over confluent monolayers was repeated six times to enrich for non-adherent mutants and then dilutions of the supernatant were plated to yield individual colonies (Geme and Cutter 1995). To confirm that the loss of adherence in individual mutants was due to the transposon insertion, chromosomal DNA was extracted from these mutants and retransformed into the parent strain 1595-A9. The resulting transformants were then screened for the ability to adhere to Chang cells.

Localization of mariner insertion by arbitrary PCR

Localization of the *solo* insertion among the stably non-adherent mutants was achieved by arbitrary PCR as described previously (Kehl-Fie and St Geme 2007). Sequence flanking the *aphA3* gene (*solo*) insertion was examined for homology with the genomic sequence of strain 1595 (E. Mardis, R. Fulton, A. J. Sheets, J. W. St. Geme III, unpublished data) using BLAST analysis.

Protein analysis

Western blot analysis of outer membrane proteins was performed as described previously (Cotter *et al.* 2006). Outer membrane fractions of whole-cell bacterial sonicates were prepared on the basis of sarkosyl insolubility as described by Carlone *et al.* (Carlone *et al.* 1986). Where noted, outer membrane fractions were treated with 95% formic acid overnight (St Geme *et al.* 1998). Proteins were resolved by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis, and Western blots were probed using a guinea pig polyclonal antiserum raised against an N-terminal fragment of Cha or a rabbit polyclonal antiserum raised against the HAT epitope (Clontech).

Southern Hybridization

Approximately 1 μg of chromosomal DNA from each strain was digested with BglII, separated by agarose gel electrophoresis, and transferred to nitrocellulose. The DNA probe was generated by PCR amplification of the *cha* locus from *Haemophilus* cryptic genospecies strain 1595 using primers TATGGCAAAATACCATACGCCACTCC and AAGAGCCATATGTTTACCAAATAAATGACAAAAATACCGCTC. The probe was labeled using the ECL nucleic acid labeling system (GE Healthcare, Piscataway, NJ) and incubated with the UV-crosslinked membrane at 42ºC in blocking solution. The membrane was washed twice with 0.5X SSC containing 0.4% sodium dodecyl sulfate at 55ºC for 10 min and rinsed twice with 2X SSC at room temperature for 5 min (1X SSC is

0.15 M NaCl + 0.015 M sodium citrate, pH 7.4). Hybridization was detected with Supersignal West Pico (Pierce, Rockford, IL).

Nucleotide sequence accession number

The sequence corresponding to the *Haemophilus* cryptic genospecies strain 1595 *cha* gene has been deposited in the GenBank and has been assigned accession number EU309721.

2.3 Results

2.3.1 Adherence of *Haemophilus* cryptic genospecies strain 1595

As a first step toward understanding the tissue tropism of the *Haemophilus* cryptic genospecies, we extended previous studies by Clemans *et al.* and Rosenau *et al.* and examined the adherence of four strains to both genital (HeLa and HEC-1-B) and respiratory (Chang and Detroit 562) epithelial cell lines (Rosenau *et al.* 1993; Clemans *et al.* 2001). As shown in Figure 1, strains 422, 1595, and 1673 were highly adherent to all of these cell lines, consistent with published results (Rosenau *et al.* 1993; Clemans *et al.* 2001). The fourth clinical isolate, strain 420, was non-adherent with all cell lines.

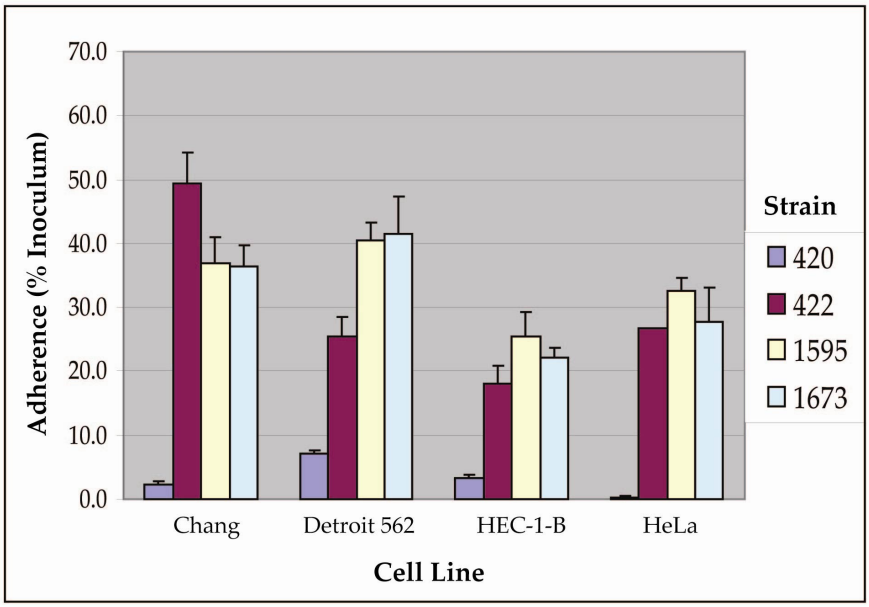

Figure 1: Adherence of *Haemophilus* cryptic genospecies strains 420, 422, 1595, and 1673 to Chang, Detroit 562, HEC-1-B, and HeLa epithelial cells.

Adherence was measured after incubating bacteria with epithelial cells monolayers for 30 min and was calculated by dividing the number of adherent colony-forming units per epithelial monolayer by the number of inoculated colony-forming units. Adherence values represent the mean of measurements from representative experiments performed in triplicate. Error bars represent standard error.

To define the mechanism of adherence, we examined our collection of isolates by Western blot analysis for expression of the major *H. influenzae* adhesins, including HMW1/HMW2 (St Geme *et al.* 1993), Hia (St Geme and Cutter 2000), and Hap (St Geme 2002). Using appropriate positive and negative controls, we found that none of the *Haemophilus* cryptic genospecies strains expressed homologs of these proteins (data not shown). PCR and Southern hybridization analysis revealed that the *hmw, hia,* and *hap* genes were absent among these strains (data not shown). Similarly, the *hifA* and *hifC*

genes involved in pilus expression were not detected (data not shown). Consistent with these results, examination of the total genome of strain 1595 (E. Mardis, R. Fulton, A. J. Sheets, J. W. St. Geme III, unpublished data) revealed none of these adhesin genes.

2.3.2 Identification of Mariner transposon mutants defective in adherence

To define the molecular determinants of adherence by the *Haemophilus* cryptic genospecies, we focused on strain 1595 based on the availability of genomic sequence and the adherence properties of this isolate. Using a stably adherent variant called strain 1595-A9 (E. Mardis, R. Fulton, A. J. Sheets, J. W. St. Geme III, unpublished data), we constructed a Mariner transposon library. Southern hybridization analysis of 11 mutants revealed that the *solo* transposon inserted randomly into the chromosome (data not shown). To identify non-adherent mutants, we enriched the library for nonadherence by passing the library over Chang epithelial cell monolayers a total of 6 times. Following enrichment, we picked 32 individual colonies and screened these clones individually for adherence to Chang cells as assessed by Giemsa staining and light microscopy. Among these 32 colonies, 16 were non-adherent. To confirm the lack of adherence by these 16 mutants was a consequence of the *solo* transposon insertion, chromosomal DNA was extracted from each of these mutants and was transformed into the parent strain 1595-A9 to select for kanamycin-resistant transformants, resulting from a double crossover event. In all cases the retransformants were non-adherent.

To localize the *solo* transposon insertions in the non-adherent mutants, we began by performing arbitrary PCR on 2 of the mutants and then sequencing the transposon junctions. The sequences were used to scan the genome sequence of strain 1595 for associated ORFs, and the insertions in both mutants were localized to the same novel open reading frame. Using this sequence information, we performed directed PCR on the remaining 14 mutants and observed that all the insertions localized to 8 distinct positions either within or immediately upstream of the same open reading frame, suggesting that some mutants represented siblings. Based on the strain 1595 genome sequence, the single disrupted open reading frame is flanked upstream by a putative thioredoxin gene and a cytochrome oxidase gene and is flanked downstream by two genes with unknown homology or function. We designated the disrupted gene *cha* for cryptic *Haemophilus* adhesin (Fig. 2A).

2.3.3 Predicted protein domains of the cryptic *Haemophilus* adhesin (Cha)

The *cha* gene encodes a protein with 1610 amino acids and a predicted molecular mass of 157 kDa. This protein has a putative long signal peptide and a predicted signal peptidase cleavage site between residues 69 and 70 (Bendtsen *et al.* 2004), consistent with secretion from the cytoplasm and surface localization. BLAST analysis revealed that the C-terminal 80 residues (amino acids 1531-1610) share significant homology with the C-termini of the *Haemophilus influenzae* Hia and *Yersinia enterocolitica* YadA proteins, defined by the conserved domain Pfam03895. Hia and YadA are prototype trimeric

autotransporter proteins associated with *H. influenzae* and *Y. enterocolitica* adherence, respectively (St Geme and Cutter 2000; El Tahir and Skurnik 2001; Roggenkamp *et al.* 2003). Also notable is the presence of an incomplete series of contiguous 28-residue repeats (amino acids 1138-1399) for which the repeat unit is entirely conserved at the nucleotide level. Other features of the predicted Cha protein include two unlinked regions of repetitive sequence, several clusters of Hep_Hag domains, and 4 scattered HIM domains, identified using the Pfam protein family database (Bateman *et al.* 2002). Hep_Hag domains (Pfam05658) are comprised of 14-residue degenerate repeats that are present in a number of bacterial adhesins, including YadA (El Tahir and Skurnik 2001), the *Moraxella* Hag protein (Bullard *et al.* 2005; Bullard *et al.* 2007), and the *Burkholderia* BuHA proteins (Tiyawisutsri *et al.* 2007). HIM domains (Pfam05662) are degenerate repeats often located adjacent to Hep_Hag domains (Nummelin *et al.* 2004; Linke *et al.* 2006). The predicted domains of the Cha protein are depicted schematically in Figure 2B.

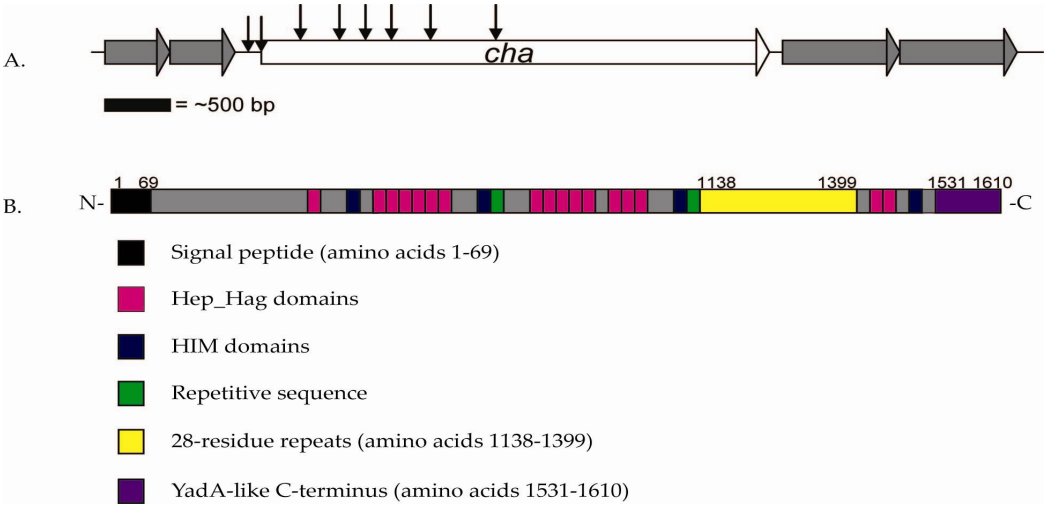

Figure 2: Genomic organization of the *cha* locus and domain architecture of the Cha protein.

Arrows indicate the locations of transposon insertions among the non-adherent mutants (1595*cha*-). The disrupted open reading frame, which was designated *cha* for cryptic *Haemophilus* adhesin, is flanked by putative upstream theoredoxin and cytochrome oxidase genes and is flanked downstream by two genes of unknown function. (B) The domain organization of Cha from strain 1595 was evaluated with SignalP and Pfam analyses.

2.3.4 Confirmation that Cha is essential for *H.* cryptic genospecies adherence

To verify that Cha is essential for the broad range adherence of the *Haemophilus* cryptic genospecies, we examined the level of adherence of a representative transposon mutant (1595*cha*-) to each of the genital and respiratory cell lines used in this study. As shown in Table 2, the mutant demonstrated negligible adherence levels compared to the parent strain (1595-A9). To confirm that Cha is an adhesin, we cloned the *cha* gene into the plasmid pLS88P and then introduced the resulting plasmid into both a non-adherent laboratory strain of *H. influenzae* called DB117 and *Haemophilus* cryptic genospecies

strain 420, which is stably non-adherent. (The *cha* gene in this recombinant plasmid contains a spontaneous internal deletion of coding sequence for residues 1138-1389.) As shown in Figure 3, Western analysis using an antiserum against an N-terminal fragment of Cha demonstrated reactivity with formic acid denatured outer membrane fractions from 1595-A9, 420/pCha, and DB117/pCha and no reactivity with 1595*cha-*, 420, and DB117/pLS88P. The size difference of the monomeric Cha protein between 1595-A9 (lane 1) and 420/pCha (lane 4) and DB117/pCha (lane 6) is likely due to the absence of repeats in the recombinant strains. As shown in Table 2, both DB117/pCha and 420/pCha were capable of efficient adherence to Chang, Detroit 562, HeLa, and HEC-1-B cells, similar to adherence by strain 1595 and demonstrating that Cha is adhesive in spite of the deletion of the nine contiguous 28-residue repeats.

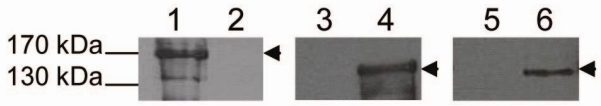

Figure 3: Western analysis of *Haemophilus* cryptic genospecies strains 1595-A9, 1595*cha-*, 420, and 420/pCha and *H. influenzae* strains DB117/pLS88P and DB117/pCha for expression of Cha.

Outer membrane samples were denatured with formic acid and examined by Western analysis with an antiserum that was raised against Cha residues 70-473 and diluted 1:1000. Samples were loaded as follows: lane 1, 1595-A9; lane 2, 1595*cha-*; lane 3, 420; lane 4, 420/pCha; lane 5, DB117/pLS88P; lane 6, DB117/pCha. Arrowheads indicate monomeric Cha protein.

Table 2: Cha-mediated adherence of *Haemophilus* cryptic genospecies and *H. influenzae* to genital and respiratory epithelial cells

Strain	Adherence* (% Inoculum)			
	Chang Cells	Detroit 562 Cells	HeLa Cells	HEC-1-B Cells
1595-A9	33.0 ± 1.3	32.9 ± 4.9	39.4 ± 2.2	26.7 ± 1.4
1595*cha-*	0.3 ± 0.1	1.2 ± 0.04	0.5 ± 0.04	0.7 ± 0.1
420†	2.3 ± 0.2	0.6 ± 0.1	3.5 ± 0.9	3.3 ± 0.3
420/pCha	34.9 ± 3.2	53.0 ± 4.5	47.5 ± 4.1	30.2 ± 2.3
DB117/pLS88P†	0.2 ± 0.04	0.3 ± 0.01	0.6 ± 0.1	0.3 ± 0.1
DB117/pCha	37.4 ± 2.6	45.7 ± 3.8	44.4 ± 1.1	29.5 ± 1.2

*Adherence was measured after incubating bacteria with epithelial cell monolayers for 30 min as described previously . Values represent mean (± SE) of measurements in triplicate from representative experiments.
† *H.* cryptic genospecies strain 420 and *H. influenzae* strain DB117/pLS88P were used as negative controls.

2.3.5 Demonstration that Cha is a trimeric autotransporter adhesin

Given the homology between the C-terminus of Cha and the C-terminus of YadA and Hia, we hypothesized that Cha is a trimeric autotransporter (Cotter *et al.* 2005). To test this hypothesis, we began by generating a plasmid encoding a signal peptide, the HAT epitope tag, and the C-terminal 120 residues of Cha (pChaC′), then introduced this plasmid into *E. coli* DH5α. Outer membrane fractions of DH5α/pChaC′ were examined by Western blot analysis using antiserum against the HAT epitope tag. As shown in Figure 4A, under standard denaturing conditions, the Cha C-terminus migrated at ~48 kDa, approximately three times the predicted molecular mass. Following more stringent denaturation with formic acid, the protein migrated at ~16 kDa, the predicted molecular mass. This result demonstrated that the final 120 amino acids of Cha are

capable of forming a trimeric structure in the bacterial outer membrane, characteristic of trimeric autotransporters.

To extend this result, we generated a chimeric protein containing a signal peptide, the HAT epitope, and the passenger domain of the *H. influenzae* Hia trimeric autotransporter fused to the C-terminus of Cha and then expressed this protein in *E. coli* DH5α (DH5α/pHiaPD-ChaC′). As shown in Figure 4B, DH5α/pHiaPD-ChaC′ was capable of high level adherence to Chang cells, similar to adherence by a control strain expressing the Hia passenger domain fused to the Hia C-terminus (DH5α/pHiaPD-HiaC′). *E. coli* expressing either the Hia passenger domain (DH5α/pHiaPD) or the Cha C-terminus (DH5α/pCha′) alone did not adhere to Chang cells. Evidence of adherence by DH5α/pHiaPD-ChaC′ demonstrates that the Cha C-terminus was able to present the Hia passenger domain stably on the bacterial cell surface in a functional form. Considered together, these data demonstrate that Cha is a member of the trimeric autotransporter family.

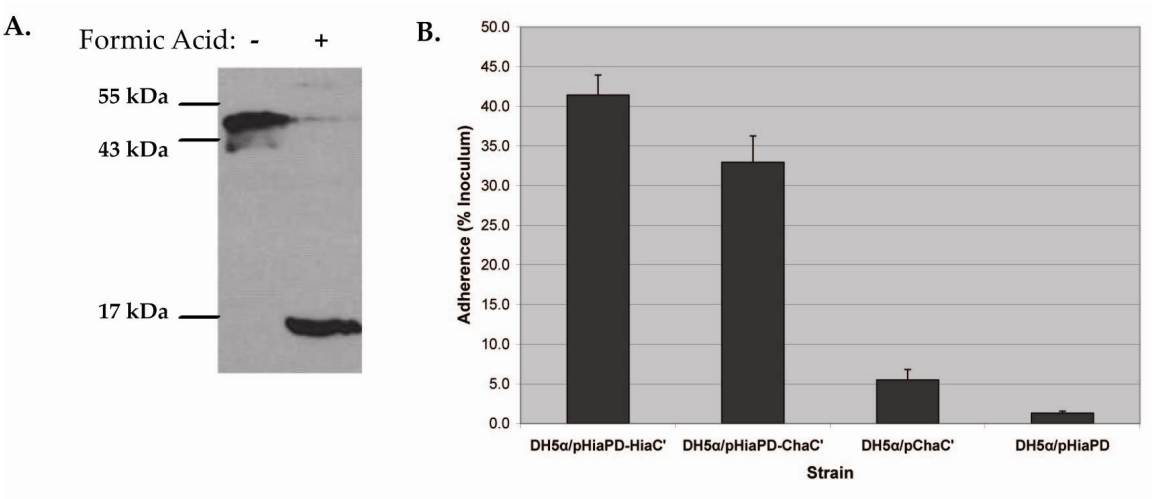

Figure 4: Evidence that Cha is a trimeric autotransporter.

(A) Western blot assay of outer membrane fractions of *E. coli* DH5α/pChaC′ without (-) or with (+) formic acid denaturation by probing with anti-HAT antibody. (B) Quantitative adherence assay with Chang cells examining the ability of the Cha C terminus to present the Hia passenger domain on the bacterial cell surface in a functional form. Plasmid pHiaPD-HiaC′ encodes the Hia passenger domain fused to the Hia C terminus of Cha, and plasmid pHiaPD encodes only the passenger domain of Hia. Error bars represent standard errors.

2.3.6 Conservation of *cha* adhesin among cryptic genospecies isolates

To assess the prevalence of *cha* among *Haemophilus* cryptic genospecies isolates, we examined a collection of 9 clinical isolates by Southern hybridization, probing with the PCR-amplified *cha* gene from strain 1595. As shown in Figure 5, all 9 strains had hybridizing fragments. In contrast, *H. influenzae* strains 11 and 12 failed to hybridize.

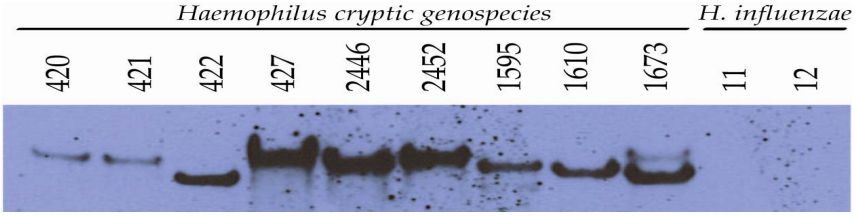

Figure 5: Southern hybridization analysis by probing for the *cha* gene among *Haemophilus* cryptic genospecies isolates.

Chromosomal DNA was isolated from nine cryptic genospecies strains and two *H. influenzae* strains and digested to completion with BglII. An ECL-labeled probe consisting of the full-length *cha* PCR amplicon from strain 1595 was used for hybridization.

2.4 Discussion

Maternal genital tract and neonatal disease caused by the *Haemophilus* cryptic genospecies requires colonization of the maternal genital epithelium and the neonatal respiratory tract, respectively. Presumably, heavy colonization of the vaginal tract predisposes to neonatal inhalation of contaminated amniotic fluid or mucosal secretions during birth, facilitating bacterial colonization of the respiratory tract, in some cases leading to respiratory distress and invasive disease. The mechanism by which the *Haemophilus* cryptic genospecies attaches to mucosal surfaces has not yet been reported. In this study, we found that epidemiologically diverse isolates of the *Haemophilus* cryptic genospecies were capable of adherence to epithelial cells of both respiratory and genital origin. Further analysis led to the identification of a novel trimeric autotransporter

protein called Cha that is responsible for the broad range adherence of the *Haemophilus* cryptic genospecies.

Adhesins classified as trimeric autotransporters have a characteristic head-stalk-anchor domain architecture and are composed of an N-terminal signal peptide, an internal passenger domain that harbors adhesive activity, and a short C-terminal autotransport domain that anchors the protein in the membrane (Koretke *et al.* 2006; Linke *et al.* 2006). Based on studies of YadA and Hia it appears that stable trimer formation is required for normal folding, stability, and adhesive activity of trimeric autotransporters (Cotter *et al.* 2006). The highly-conserved C-terminal anchor domain is important for oligomerization and contains 4 transmembrane ß-strands which assemble into a 12-stranded ß-barrel that is filled by the C-terminal end of the passenger domain (Hoiczyk *et al.* 2000; Linke *et al.* 2006). In this study, we demonstrated that the Cha C-terminal region encompassing residues 1491-1610 is inserted into the bacterial outer membrane in a trimeric form. This trimer is stable to heat and detergent and requires formic acid denaturation for dissociation into individual monomers. Consistent with studies of other trimeric autotransporters, we found that the Cha C-terminus was capable of presenting a heterologous passenger domain on the bacterial surface in a functional form. In particular, we demonstrated that *E. coli* expressing a chimeric protein containing the Hia passenger domain fused to the Cha C-terminus was capable of adherence to cultured epithelial cells.

While the anchor domains of trimeric autotransporters are highly homologous, the passenger domains are diverse and frequently have repetitive domains in varying combinations. Repetitive sequences within the passenger domains are believed to facilitate recombination of domains, thereby altering the specificity of bacterial adherence to host epithelium, extracellular matrix proteins, and in some cases circulating complement factors or immunoglobulins. In YadA, four Hep_Hag repeats (comprised of degenerate 14-residue repeats sometimes referred to as NSVAIGXXS motifs) in the amino-terminal half of the passenger domain form a ß-roll head domain (Nummelin *et al.* 2004). The head domain is linked to the coiled-coil stalk by a neck adaptor domain composed of a HIM sequence motif (Nummelin *et al.* 2004). The association of Hep_Hag repeats and HIM motifs is common to a number of bacterial adhesins, including the *Moraxella* Hag (Bullard *et al.* 2007) and UspA1 proteins (Hoiczyk *et al.* 2000), *Xanthomonas* XadA (Ray *et al.* 2002), *Burkholderia* BuHA (Tiyawisutsri *et al.* 2007), and *Bartonella* BadA (Riess *et al.* 2004). Interestingly, the Hep_Hag repeats present in YadA are required for binding to collagen (Nummelin *et al.* 2004), while the Hep_Hag repeats in Hag are not important for Hag-mediated adherence to either collagen or epithelial cells (Bullard *et al.* 2007). Additionally, the role of a single HIM motif as a head-to-stalk adaptor domain seems be unique to the YadA and UspA1 proteins. The XadA, Hag, and BuHA proteins contain scattered clusters of Hep_Hag repeats and multiple non-contiguous HIM motifs throughout the passenger domain. Similar to

these proteins, the *Haemophilus* cryptic genospecies Cha protein contains 17 clustered Hep_Hag repeats and 4 scattered HIM motifs, including two HIM motifs that are linked to unique 14-residue repeats.

Cha is also characterized by contiguous unique 28-amino acid repeats. Based on the size and location of this alanine-rich repeat region in the passenger domain and the characteristic architecture of trimeric autotransporters, it is likely that this region forms the stalk domain and is responsible for extending binding domains away from the bacterial cell surface. The genomic sequence of strain 1595 suggests that 9 full repeats plus a 10-residue partial repeat represent the wildtype number of repeats. However, it is likely that the number of repeat units varies, perhaps explaining the small size differences of the *cha*-hybridizing genomic fragments among cryptic genospecies strains (Fig. 5). Changes in repeat number may alter the adhesive activity of Cha, although among our isolates we did not observe a clear correlation between a capacity for adherence to Chang cells and fragment size in our Southern analysis. Changes in the repeat domain may play an additional role in antigenic variation as a means for evading the immune system during invasive systemic disease, a phenomenon observed with the alpha C protein in group B streptococci (Gravekamp *et al.* 1997; Gravekamp *et al.* 1998).

We found that the *cha* locus was uniformly conserved among our collection of 9 strains belonging to the cryptic genospecies. In contrast, a *cha*-hybridizating genomic fragment was not detected in either *H. influenzae* strain 11 or *H. influenzae* 12. The

absence of *cha* among 5 additional *H. influenzae* strains (86-028NP, Rd KW20, R2866, PittEE, and PittGG) for which genomic sequencing has been completed suggests that *cha* is uniformly lacking among *H. influenzae* strains and may be unique to the cryptic genospecies. With this information in mind, we speculate that the *cha* locus was acquired after *H. influenzae* and the *Haemophilus* cryptic genospecies diverged from each other evolutionarily and contributes to the apparent adaptation of the *Haemophilus* cryptic genospecies to the urogenital tract.

At this point, it is unclear if Cha corresponds to the peritrichous fibers that have been detected on the surface of some strains of the *Haemophilus* cryptic genospecies. On the one hand, a variety of investigators have observed a correlation between the presence of fibers and capacity for adherence among diverse isolates of the *Haemophilus* cryptic genospecies. On the other hand, we have recovered one variant of *Haemophilus* cryptic genospecies strain 1595 that possesses fibers and lacks Cha and another variant that lacks fibers and expresses Cha (A. Sheets and J.W. St. Geme, unpublished data). In addition, examination of strain DB117/pCha and strain 420/pCha by transmission electron microscopy revealed no fibers (data not shown). In future work, we will attempt to address this question definitively.

In this work we report the identification of Cha, a trimeric autotransporter protein that is the major adhesin of the *Haemophilus* cryptic genospecies and presumably plays a critical role in colonization of the maternal genital tract and neonatal respiratory

tract. In ongoing work, we are examining whether Cha has adhesive activity that extends beyond adherence to epithelial cells, perhaps mediating binding to extracellular matrix proteins and facilitating binding of circulating host proteins, as observed with other members of the trimeric autotransporter family.

Acknowledgements

This work presented in this chapter was supported by NIH grant RO1-AI44167 to J. W. S. and has been published in *The Journal of Bacteriology*:

Sheets, A. J., S. A. Grass, S. E. Miller and J. W. St Geme, 3rd (2008). "Identification of a novel trimeric autotransporter adhesin in the cryptic genospecies of *Haemophilus*." *J Bacteriol* **190**(12): 4313-20. Copyright © American Society for Microbiology

Chapter 3. The *Haemophilus* cryptic genospecies Cha autotransporter mediates adherence and microcolony formation via a novel adhesive domain in the N-terminal end of the passenger domain.

3.1 Introduction

Most bacterial infections begin with colonization of a relevant mucosal surface. Attachment of bacteria to host epithelium is mediated by specific interactions between bacterial surface proteins and cognate receptors on epithelial cells and oftentimes extracellular matrix components. In addition, bacterial aggregation aids in colonization, protects against adverse environmental conditions such as oxidative stress and host defenses, and possibly represents an intermediate step in biofilm formation.

The pathogenesis of disease due to the *Haemophilus* cryptic genospecies begins with bacterial attachment to genital epithelial mucosa. Outgrowth of these bacteria in the maternal genital tract can lead to cervicovaginitis, chorioamionitis, and premature rupture of membranes. Aspiration of contaminated amniotic fluid or mucosal secretions during the process of delivery results in bacterial colonization of the neonatal respiratory tract and can lead to invasive disease. We recently reported the identification of a novel adhesin called Cha, which is conserved among strains belonging to the cryptic genospecies and mediates efficient bacterial attachment to epithelial cells of both genital and respiratory origin (Sheets *et al.* 2008). Presumably

expression of Cha facilitates bacterial colonization of the genital tract and contributes to the apparent genital/neonatal-tissue tropism of this pathogen.

The Cha adhesin belongs to the trimeric autotransporter family of proteins. All trimeric autotransporters (TAAs) share highly repetitive, modular architecture including an N-terminal signal peptide and a C-terminal membrane anchor. The defining C-terminal domain is the only universally conserved domain among all TAAs; and we have previously shown that the C-terminal domain of Cha is capable of presenting a heterologous passenger domain (Hia) on the bacterial cell surface (Sheets *et al.* 2008). This domain targets the protein to the outer membrane of gram-negative bacteria where it homo-trimerizes into a β-barrel and provides a pore for autotransport of the passenger domain. Thus far 3 different functional classes within the passenger domain have been annotated: head, connector, and stalks (Szczesny 2008). Connector motifs allow for the transition from one type of structure to another. Most common among these is the neck or HIM motif (Pfam05662). Head domains include Ylhead degenerate repeats (forming a β-solenoid), HIN(1,2) domains (two Ylhead insert motifs, defined by GYDP and FxG motifs), Trp-rings (β-meander), GIN motif (β-prism), and an FxG motif (unknown structure). Unlike the classic head-stalk-anchor organization exemplified by the *Yersinia* YadA (Koretke *et al.* 2006), the *Neisseria* NadA (Comanducci *et al.* 2002), and the *Bartonella* BadA (Kaiser *et al.* 2008) TAAs, head and stalk domains alternate throughout the Cha passenger domain. Cha contains 4 scattered clusters of Ylhead degenerate

repeats (previously referred to as Hep_Hag), each associated with HIM/neck domains, and 2 predicted alanine-rich, left-handed coiled-coil stalk domains (Sheets *et al.* 2008).

In spite of the highly conserved structural domains, binding domains of trimeric autotransporters vary greatly in their composition. While Ylhead degenerate repeats [NSVAIGXXS] are critical to the adhesive activity of both the YadA and BadA head domains, these repeats are located outside the region important for adherence by the *Moraxella catarrhalis* Hag protein (Bullard *et al.* 2007; Kaiser *et al.* 2008). It is therefore difficult to predict the requirement for the broadly-conserved Ylhead clusters in the adhesive function of Cha.

In this study we sought to identify the regions of Cha responsible for attachment to epithelial cells. We observed that the adhesive domain of Cha localizes to the N-terminal head of the protein. We further observed that the same N-terminal region of Cha mediates microcolony formation through inter-bacterial, homo-oligomeric Cha-Cha interections. This adhesive region largely contains unique, structurally undefined sequence, linked to a short cluster of Ylhead repeats followed by a neck motif. These results extend our understanding of the functional organization of the Cha autotransporter.

3.2 Materials and Methods

Culture and storage conditions

Escherichia coli strains were grown on Luria-Bertani (LB) agar or in LB broth at 37°C and were stored at −80°C in LB broth with 50% glycerol. *Haemophilus* strains were grown on chocolate agar plates (BD, Franklin Lakes, NJ) at 37°C with 5% CO_2 and were stored at −80°C in brain heart infusion broth with 25% glycerol.

Molecular Biology Techniques

DNA ligations, restriction endonuclease digestions, and gel electrophoresis were performed according to standard techniques (Sambrook *et al.* 1989). Plasmids were introduced into *E. coli* by electroporation (Dower *et al.* 1988). In *H. influenzae*, transformation was performed by the MII/MIV method of Herriott *et al.* (Herriott *et al.* 1970). Transformation reaction mixtures were plated onto chocolate agar containing 2 μg/ml chloramphenicol to select for transformants. *E. coli* and *H. influenzae* strains were grown with 50 μg/ml kanamycin to select for plasmids.

Bacterial strain and plasmid construction

Cha derivatives were inserted into the chromosome of non-adherent *Haemophilus influenzae* strain Rd (which does not encode a Cha homologue). The Rd *hap* locus contains a nonsense mutation at codon 710 that results in a truncated open reading frame, presenting a useful location for insertion of foreign DNA. DNA encoding deletion derivatives of Cha was cloned next to a chloramphenical-resistance marker in

the middle of the *hap* pseudogene. Chloramphenical-resistant transformants resulting from double crossover events at the *hap* locus were screened for Cha expression by whole cell dot immunoblot.

The bacterial strains and plasmids used in this study are described in Table 3. Amino acid number designations reflect the presence of one arbitrary 28-residue repeat to highlight their existence and physical location, however genomic DNA from variant strain 1595-A lacking these repeats was used as template in the generation of Cha derivatives. To clone the C-terminal domain of Cha (ChaCTD), the coding sequence for residues 1298-1386 was amplified from wild-type *Haemophilus* cryptic genospecies strain 1595 with primers ChaCTDF and ChaCTDR containing engineered 5′-MluI and 3′-BamHI restriction sites, respectively. This resulting fragment was digested with MluI and BamHI and was ligated to MluI/BamHI-digested pLS88P (Sheets *et al.* 2008), generating pChaCTD. The Cha passenger domain including upstream native promoter sequence was amplified from a variant strain 1595-A using primers ChaP1F and ChaFL1R, digested with NcoI/MluI and ligated to NcoI/MluI-digested pChaCTD to create pChaΔ1138-1165. To insert an additional selectable marker downstream of the *cha* gene, the chloramphenicol resistance cassette from pACYC184 was amplified using primers CmF and CmR containing engineered BglII sites, was digested with BglII, and was then ligated to BamHI-digested pChaΔ1138-1165. Primers ChaP2F and CmF-ClaI were used to amplify the *cha-cam*^{R} linked genes, and the ClaI-digested product was

ligated to AclI-digested pXL:hap. The plasmid pXL:hap was created by amplifying the inactive *hap* locus from *Haemophilus influenzae* strain Rd using primers RdhapF and RdhapR and then ligating this fragment into pCR-XL-TOPO (Invitrogen). A single, native AclI site is located near the middle of the *hap* locus. Prior to insertion of the *cha-cam*^{R} genes, the BamHI and MluI restriction sites were removed from pXL:hap by digestion with BamHI and MluI, blunting with T4 DNA polymerase, and ligating the blunt ends. This sequence of cloning steps generates a base plasmid (designated pXL:hap-ChaΔ1138-1165-Cm^R) in which unique BamHI and MluI restriction sites flank the upstream and coding sequence for the passenger domain of Cha (P-1298), thus facilitating the exchange of passenger domains containing various internal in-frame deletions. Relevant primer sets for each deletion mutant are listed in Table 4. An internal 207-amino acid deletion (Δ83-290) was created by digesting Cha$_{(290-474)}$ coding sequence (amplified using primers 155F/109R) with SalI and MluI and ligating this fragment to SalI/MluI-digested pXL:hap-ChaΔ474-1297-Cm^R. The *cha-cam*^{R}-containing *hap* locus was amplified using primers RdhapF and RdhapR for each deletion construct, and the resulting DNA amplicon was transformed into *H. influenzae* strain Rd.

Plasmid pGFP is a pLS88-derivative containing the *hap* non-coding region fused to a promoterless green fluorescent protein cassette as described previously (Fink and St Geme 2003).

Quantitative Adherence Assays

Quantitative adherence assays using epithelial cells were performed as described previously (St Geme 1993). Percent adherence was calculated by dividing the number of adherent CFU per epithelial cell monolayer by the number of inoculated CFU. Each strain was examined in triplicate in a given assay, and each assay was performed a minimum of three times. Chang conjunctival cells (Wong-Kilbourne derivative, clone 1-5c-4; ATCC CCL-20.2) were maintained in modified Eagle medium with Earle's salts and supplemented with nonessential amino acids and 10% fetal bovine serum.

Protein analysis

Western blot analysis of outer membrane proteins was performed as described previously (Cotter *et al.* 2006). Outer membrane fractions of whole-cell bacterial sonicates were prepared on the basis of Sarkosyl insolubility (Carlone *et al.* 1986) and treated with 95% formic acid overnight (St Geme *et al.* 1998). Proteins were resolved by sodium dodecyl sulfate-polyacrylamide gel electrophoresis, and Western blots were probed with a guinea pig polyclonal antiserum raised against an N-terminal fragment of Cha (aa70-473). Surface localization of recombinant Cha protein was qualitatively assessed by whole cell dot immunoblot using anti-Cha antiserum diluted 1:5000.

Quantitative and co-culture aggregation assays

Haemophilus strains were grown overnight on plates and were then resuspended in 3 ml BHI to an OD_{600} of 0.8. Bacterial suspensions were incubated at room

temperature for 4 hr during which time the optical density at 600 nm was measured at 30 min intervals. In some cases, 1.5 ml aliquots of strain Rd or Rd/ChaΔ474-1297 were mixed equally with Rd/ChaΔ474-1297 + pGFP. Co-culture bacterial aggregates were removed from the bottoms of the culture tubes and spotted onto glass slides with coverslips. Visualization of aggregates was accomplished using DIC and fluorescence microscopy.

Table 3: Strains and plasmid used in this study

Strains	Description	Source /Reference
Haemophilus influenzae		
Rd	Non-adherent, capsule-deficient serotype d laboratory strain	Setlow *et al.* (1968)
Rd/ChaΔ1138-1165	Strain Rd expressing chromosomally-encoded Cha protein from cryptic genospecies strain 1595 with deletion of repeat residues 1138-1165, passenger domain cloned using primer set (ChaP1F/ChaFL1R)	This study
Rd/ChaΔ1094-1297	Strain Rd expressing Cha aa 1-1093/1298-1386, primer set (ChaP2F/Cha1bp3279R)	This study
Rd/ChaΔ804-1297	Strain Rd expressing Cha aa 1-803/1298-1386, primer set (ChaP2F/Cha1bp2409R)	This study
Rd/ChaΔ474-1297	Strain Rd expressing Cha aa 1-473/1298-1386, primer set (ChaP2F/Cha1bp1419R)	This study
Rd/ChaΔ382-1297	Strain Rd expressing Cha aa 1-381/1298-1386, primer set (ChaP2F/Cha1bp1143R)	This study
Rd/ChaΔ83-290/474-1297	Strain Rd expressing Cha aa 1-82/291-473/1298-1386, primer set (ChaP2F/109R)	This study
***Haemophilus* cryptic genospecies**		
1595-A	neonatal blood isolate, colony variant with 84-nucleotide repeats within *cha* gene spontaneously deleted, template for Rd/Cha derivatives	This study
E. coli		
DH5α	*E. coli* F- Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*) *U169 deoR recA1 endA1 hsdR17*(rK-mK+) *phoA supE441 thi-1 gyrA96 relA1*	Sambrook *et al.* (2001)
Plasmids		
pGFP	pLS88 (ATCC 86980) containing green fluorescent protein cassette, KanR	Fink *et al.* (2003)

Table 4: Primers used in this study

Primers/Amplified Target	Nucleotide Sequence (5'-3')	Restriction sites
Cha C-terminal domain		
ChaCTDF	ATACGCGTGAAAACCGTATCAATAATAAACTTGGTAAATTTG	MluI
ChaCTDR	AGGATCCTTTACCAAATAAATGACAAAAATACCGCTC	BamHI
Cha passenger domain derivatives		
ChaP1F	CAATGAAACGAATTTACAAAGCTACTCTATTCTC	-
ChaP2F	CCATGG-ATCGAT-GGATCC-CCGCTGAAAAATCAGCATTGGGAC	NcoI-ClaI-BamHI
Cha1bp871F	GCGGTCGACAGGCGTTCTTTTGGGTACTAACTTG	SalI
ChaFL1R	CGACGCGTTACCGCATTGACTTGACCAACATTAGCCGC	MluI
Cha1bp3279R	CTACGCGTATCATTTCTACCAACAGAAACAACCCC	MluI
Cha1bp2409R	GGACGCGTATTCGCTGTCTCATTTAAGGCAGCAC	MluI
Cha1bp1419R	GGACGCGTACCTGTCGCTGTCTTGCCTTTATTACC	MluI
Cha1bp1143R	GAACGCGTACTGTTGGTGTTTGCGGTTGTCGC	MluI
Cm resistance cassette		
CmF	TAAAGATCTCCTGGTGTCCTGTTGATACCGGG	BglII
CmR	CGGAGATCTAACGACCCTGCCCTGAACCGACG	BglII
CmF-ClaI	CATATGATCGATCCTGGTGTCCCTGTTGATACCGGG	ClaI
Rd *hap* locus		
RdhapF	GCGCGTGTACGTATAGGCTCAGGACG	-
RdhapR	GTTGCAGACAACTCCGATAACGCAC	-

3.3 Results

3.3.1 Binding domain localization of Cha

Like most trimeric autotransporter adhesins, Cha is characterized by highly repetitive modular architecture. Predicted head and stalk domains alternate throughout the Cha passenger domain (which harbors the adhesive activity), suggesting a more complex architecture than is exhibited by other well-described trimeric autotransporter adhesins (Koretke *et al.* 2006; Szczesny *et al.* 2008). Cha contains 4 scattered clusters of Ylhead repeats, neck/HIM adaptor domains, and two predicted alanine-rich, left-handed coiled-coil stalk domains. The second coiled-coil domain contains unique 84-nucleotide

contiguous repeats which are not required for adherence (Sheets *et al.* 2008). To explore the domains in Cha involved in binding to host cells, a panel of internal-deletion mutant strains was tested for their ability to bind to Chang conjunctival cells. We created sequential in-frame deletions throughout the Cha passenger domain, in each case retaining the 89-residue C-terminal membrane anchor domain. Each chromosomally-encoded derivative was expressed in the outer membrane of *Haemophilus influenzae* strain Rd, which does not express any of the major *H. influenzae* adhesins (Hap, HMW, Hia) and does not adhere to Chang cells. Formic acid-denatured outer membrane fractions from each Rd/Cha strain were examined by Western blot analysis to confirm surface expression of each Cha protein at the appropriate, predicted monomeric molecular weight (Fig. 6B). In the absence of formic acid denaturation, each deletion-derivative protein migrated at a molecular weight consistent with trimerization (not shown).

In quantitative assays, adherence to Chang cells was similar for Rd expressing ChaΔ1138-1165 (full-length minus repeats), ChaΔ1094-1297, and ChaΔ804-1297, as shown in Figure 7. Rd expressing ChaΔ474-1297 adhered most efficiently, which we speculate is because of higher expression levels or stability of the protein. (Outer membrane fractions repeatedly contained significantly higher levels of ChaΔ474-1297 protein, and therefore one fifth of the sample volume relative to other Rd/Cha strains was used for Western analysis.) Adherence by Rd/ChaΔ474-1297 suggests that the N-

terminal 473 residues of Cha contain the domain responsible for adherence to epithelial cells. Cha containing further deletions within the N-terminus was unable to mediate adherence. The N-terminal region of Cha containing the epithelial cell binding domain includes Ylhead/β-roll motifs, a putative head insert motif (HIN2), and a single neck/HIM domain. Preceding these domains, Cha is characterized by a significant amount of unique and undefined sequence (Fig. 6A). The inability of Rd/ChaΔ83-290/1138-1297 to adhere indicates that the undefined N-terminal sequence is critical to adhesive activity.

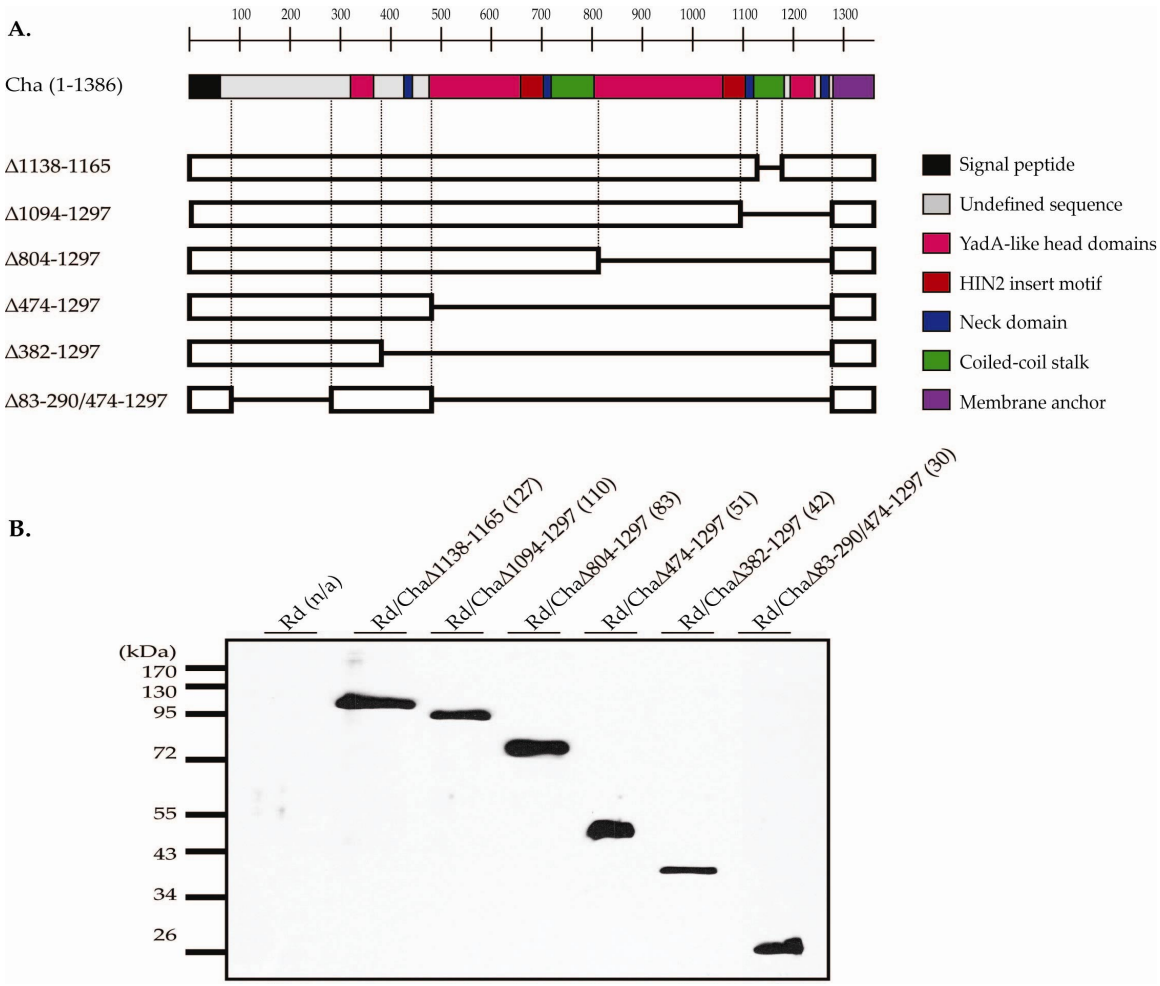

Figure 6: Domain organization and expression of Cha deletion derivatives.

(A) Schematic diagram of recombinantly expressed Cha. Color-coded features highlight domain organization of each Cha protein consisting of YadA-like head (Hep_Hag) repeats, HIN2 domains, neck (HIM) domains, helical stalk regions, and membrane anchor. Numbers provided represent amino acid positions. One 28-residue repeat is arbitrarily depicted starting at amino acid 1138 to show physical location of these variable repeats. Genomic DNA from repeat variant strain 1595-A lacking these repeats was used as template in the construction of Cha derivatives. (B) Western blot analysis of formic acid-denatured outer membrane proteins from *H. influenzae* strain Rd expressing Cha in-frame deletion derivatives. Samples were loaded as indicated. The numbers in parentheses represent the predicted molecular weight of each monomeric Cha protein. Immunoblotting was performed with anti-Cha guinea pig polyclonal antiserum.

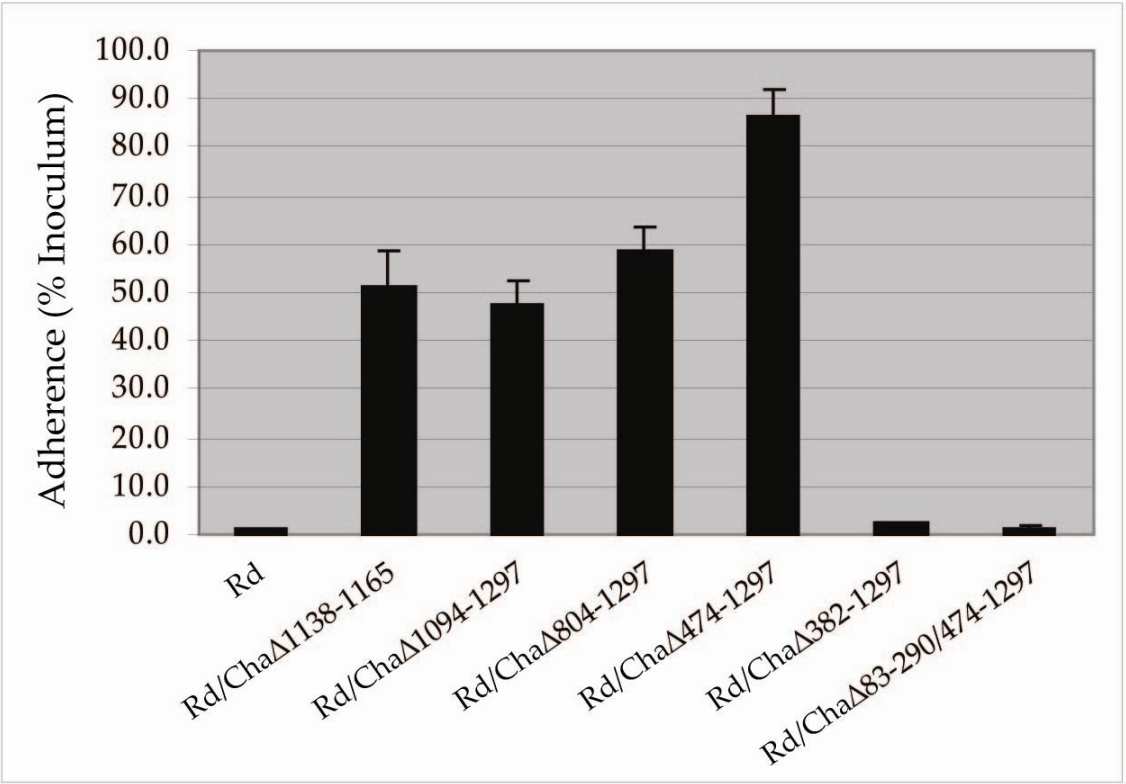

Figure 7: Adherence to epithelial cell monolayers by *H. influenzae* strain Rd expressing Cha in-frame deletion derivatives.

Adherence was measured after incubating bacteria with epithelial cell monolayers for 30 min and was calculated by dividing the number of adherent CFU per epithelial cell monolayer by the number of inoculated CFU. Adherence values represent the mean of measurements from representative experiments performed in triplicate. Error bars represent standard errors.

3.3.2 Cha-mediated bacterial aggregation

Beyond mediating interaction with host cells, a number of adhesins have been shown to play a role in bacterial aggregation and microcolony formation (Sherlock *et al.* 2005; Valle *et al.* 2008), a phenomenon that likely facilitates persistence of organisms on mucosal epithelial surfaces. Visual examination of strain 1595 interacting with epithelial

cell monolayers revealed adherence by large bacterial clusters among single organisms. Similar bacterial microcolonies have also been observed among some clinical isolates visualized by negative staining electron microscopy. To further investigate the functional role of Cha at the bacterial cell surface, we tested the contribution of the Cha adhesin to bacterial aggregation by several methods. First, we visually assessed bacterial clumping and settling of a suspension of stably adherent strain 1595-A versus 1595*cha*- mutant. Rates of bacterial aggregation were quantified by measuring absorbance as aggregates settled to the bottom of culture tubes. As shown in part A of Figure 8, over a four hour time period, Cha-expressing strain 1595-A rapidly aggregated and settled whereas mutant strain 1595*cha*- remained in suspension. To identify which regions of Cha might be responsible for bacterial aggregation, we performed tube settling assays with suspensions of strain Rd expressing the Cha in-frame deletion derivatives. Over time, suspensions of Rd expressing ChaΔ1138-1165 (full-length minus the repeats) steadily lost turbidity, while strain Rd alone remained in suspension. Strains expressing ChaΔ1094-1297, ChaΔ804-1297, and ChaΔ474-1297 aggregated at increasingly rapid rates. Rd/ChaΔ474-1297 expressing the N-terminal 473 residues of Cha fused to the C-terminal membrane anchor domain settled most rapidly, suggesting that the presence of the entire Ylhead/HIM-containing passenger domain may sterically interfere with the aggregative domain. Alternatively, this enhanced aggregation may again reflect higher Cha protein density on the bacterial cell surface. Strain

Rd/ChaΔ382-1297 expressing residues 1-381 of the passenger domain failed to settle in a four hour time period. The region between residues 382 and 437 contains a neck adapter domain, which connects β structures with coiled-coils. Deletion of this motif may disrupt proper folding of the N-terminus of Cha, thereby interfering with bacterial aggregation. Additionally, Rd/ChaΔ83-290/474-1297, a derivative of the highly aggregative Rd/ChaΔ474-1297 that lacks residues 83-290 did not undergo aggregation and settling (Fig. 8A). Taken together, these results suggest that the autoaggregative domain of Cha lies within the N-terminal 473 amino acids and is thus far not separable from the Cha host cell binding domain.

Cha-mediated bacterial aggregation may occur by homo-oligomerization of Cha trimers on neighboring cells or by Cha binding to a heterologous structure on neighboring bacteria. To distinguish between the two possibilities, we evaluated the rate of settling of a mixed suspension of strain Rd with Rd/ChaΔ474-1297 (50/50). The observed rate of settling was markedly reduced relative to settling of Rd/ChaΔ474-1297 alone (Fig. 8A). This interference of strain Rd with Rd/ChaΔ474-1297-settling suggests that inter-bacterial Cha-Cha interactions are responsible for bacterial aggregation. To obtain further evidence for this conclusion, we prepared suspensions of bacteria expressing ChaΔ474-1297 plus green fluorescent protein (GFP) mixed with unlabeled bacteria expressing ChaΔ474-1297 (or unlabeled strain Rd). Bacterial suspensions were incubated standing for 4 h, and bacterial aggregates collected from the bottoms of

culture tubes were examined by differential interference contrast and fluorescence microscopy (Fig. 8B). Mosaic aggregates from Rd/ChaΔ474-1297 + pGFP resuspended with Rd/ChaΔ474-1297 suggests that both GFP-labeled and unlabelled bacteria aggregated. Aggregates from Rd mixed with Rd/ChaΔ474-1297 + pGFP were less abundant and uniformly green, suggesting that only bacteria expressing the N-terminus of Cha participated in aggregate formation. As expected, large aggregates of ChaΔ474-1297 + pGFP (Fig. 8B) were uniformly green, and unlabeled strain Rd did not aggregate (not shown). These results indicate that each Cha(aa70-473) molecule must contain both receptor and receptor-recognition faculties.

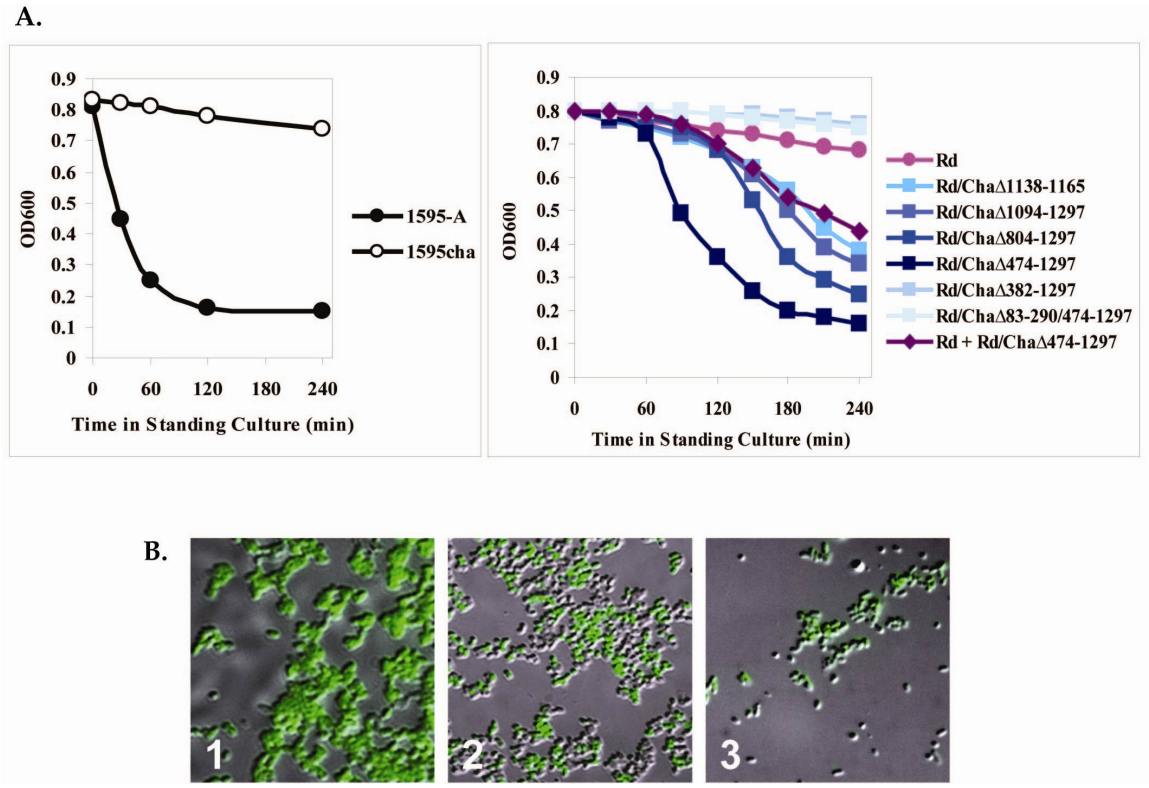

Figure 8: Cha-mediated bacterial flocculation.

(A) Settling assays with cryptic genospecies strain 1595, 1595*cha* mutant, and *H. influenzae strain* Rd expressing Cha in-frame deletion derivatives. Bacteria were resuspended in BHI to an optical density of 0.8 at 600 nm. Bacterial aggregation and settling was evaluated by measuring the absorbance every 30 min for 4 hours. A co-suspension of strain Rd with Rd/ChaΔ474-1297 reflects a 50/50 mixture. (B) Bacterial aggregation mediated by inter-bacterial Cha-Cha binding. Fluorescence and differential interference contrast (DIC) micrograph overlays of aggregates from *H. influenzae* strain Rd expressing ChaΔ474-1297. Bacterial suspensions were left standing at room temperature for 4 hours prior to collection and inspection of settled aggregates. Panels depict as follows: 1) Rd/ChaΔ474-1297 plus pGFP 2) Rd/ChaΔ474-1297 plus pGFP mixed equally with unlabeled Rd/ChaΔ474-1297, 3) Rd/ChaΔ474-1297 plus pGFP mixed equally with unlabeled strain Rd.

3.4 Discussion

Isolates of the *Haemophilus* cryptic genospecies are recovered almost exclusively from genital sites or from infected neonates who acquired the bacteria *in utero* or during passage through the vagina during birth. These organisms appear to occupy a distinct ecological niche different from that of nontypable *Haemophilus influenzae* respiratory strains which colonize the pharynx and nasopharynx. Cryptic genospecies strains adhere to a variety of human epithelial cells of both respiratory and genital origin (Rosenau *et al.* 1993; Sheets *et al.* 2008) but do not express classic NTHI adhesins (hemagglutinating pili, Hia, HMW1/2, and Hap). Instead, these strains express a unique adhesin called Cha, which mediates attachment to human epithelial cells (Sheets *et al.* 2008). Cha is classified as a trimeric autotransporter based on C-terminal homology of the membrane anchor domain. However, the passenger domain, which harbors the adhesive domain of the protein, bares little resemblance to Hia, consistent with different binding targets and tissue specificity associated with this species.

In this study, we sought to localize the adhesive region of Cha. Several in-frame deletions were generated within the *cha* gene from cryptic genospecies strain 1595. Derivatives of *H. influenzae* strain Rd expressing the resulting proteins were tested for their ability to bind Chang cell monolayers. Using this approach, we observed that residues 70-473 of the 1386-amino acid Cha protein are critical for adherence. Binding domains located at the tip of a protein are easily accessible to host cell receptor

structures and recognition of an N-terminal binding domain is a common theme among trimeric autotransporters (Nummelin *et al.* 2004; Kaiser *et al.* 2008; Yu *et al.* 2008). This region in Cha contains mostly unique, undefined sequence followed by a small cluster of Ylhead β-roll repeats and a neck motif. The Ylhead repeats present in this minimal adhesive fragment (and throughout the passenger domain of Cha), are widely conserved among trimeric autotransporter adhesins (TAAs), including YadA of *Yersinia entercolitica* (Roggenkamp *et al.* 2003), BadA of *Bartonella henselae (Riess et al. 2004),* Vomps of *Bartonella quintana* (Zhang *et al.* 2004), XadA of *Xanthomonas oryzae* (Ray *et al.* 2002), EmaA of *Aggregatibacter actinomycetemcomitans* (Yu *et al.* 2008), Hag of *Moraxella catarrhalis* (Bullard *et al.* 2005), and the BuHA proteins of *Burkholderia sp.* (Tiyawisutsri *et al.* 2007). Conservation of these structural building blocks in different arrangements with varied activities among bacterial pathogens may reflect adaptations to different physiological niches. The Ylhead repeats contribute to the binding activity in several TAAs but not exclusively, reflecting disparate receptor targets, and these conserved sequences are not present in Hia (Laarmann *et al.* 2002; Nummelin *et al.* 2004; Bullard *et al.* 2007; Kaiser *et al.* 2008; Yu *et al.* 2008). Our data suggests that the N-terminal Ylhead repeats are required for epithelial cell binding; however, it is unclear whether this β-roll domain represents an adhesive motif or a key structural domain necessary for stability of a neighboring adhesive pocket. Deletion within the N-terminal structurally undefined sequence disrupts adherence suggesting the adhesive pocket may reside in a

novel structural motif conferred by this sequence. It is likely that the first neck/HIM domain included in the minimal binding fragment (aa 70-473) does not represent part of the host cell binding domain but is required for proper folding and transition of the N-terminal β-roll structure into the coiled-coil linker of the membrane anchor.

Crystal structure information from other trimeric autotransporter adhesins has shown in each case that the adhesive domains have threefold symmetry, conferred by such sequences as Ylhead/HIM motifs (Cotter *et al.* 2005). An adhesive structure with three identical faces provides the potential for three binding pockets to adhere with strong affinity to receptor targets and may explain the requirement for the N-terminal trimeric Ylhead cluster and HIM domain for Cha adhesive activity. The function of the internal Ylhead clusters at this time remains unclear. These repeats may represent additional binding sites which increase the avidity of the Cha protein for its cognate receptor, in a manner similar to the two Hia binding pockets (Laarmann *et al.* 2002). Ylhead repeats in YadA and EmaA function as collagen binding motifs (Nummelin *et al.* 2004; Yu *et al.* 2008), which raises the possibility that Cha may mediate adherence to ECMs such as collagen though interactions with other internal domains of the passenger domain. These repeats may also function to extend the distal binding domain further beyond the cell surface. The functional significance of the predicted coiled-coil stalk domain located towards the C-terminus of the passenger domain remains unclear. In this study and previous work (Sheets *et al.* 2008), we have shown it is not required for

adhesive activity. In ongoing studies, we aim to directly address the functional impact the predicted stalk domain has on Cha adhesive activity.

In addition to being a host cell adhesin, Cha mediates intebacterial interactions, demonstrated in this study. The ability to form aggregates appears to be a common theme among many bacterial pathogens. A diverse array of bacteria, including *E. coli* (Schembri *et al.* 2001; Sherlock *et al.* 2005; Valle *et al.* 2008), *Bordetella pertussis* (Menozzi *et al.* 1994), *Staphylococcus aureus (McDevitt et al. 1994)*, *Streptococcus pyogenes (*Caparon *et al.* 1991; Frick *et al.* 2000), and *Haemophilus influenzae* (Hendrixson and St Geme 1998) can form aggregates. Bacterial aggregation and microcolony formation enables pathogens to resist harsh environmental conditions (e.g. hydrogen peroxidase treatment) and host defenses (e.g. complement attack and phagocytosis) better than solitary bacteria (Ochiai *et al.* 1993; Schembri *et al.* 2003). Some aggregation factors such as the polymeric type 1 pili (Schembri and Klemm 2001) and curli (Vidal *et al.* 1998) further mediate biofilm formation. Considering this information, the ability of bacteria to autoaggregate represents an important virulence mechanism. We observed that expression of Cha either by cryptic genospecies wildtype strain 1595 or by *H. influenzae* strain Rd promotes rapid autoagglutination of bacterial cells similar to the *Haemophilus* Hap adhesin (Hendrixson and St Geme 1998). Our data indicate that Cha is self-associating at the distal head of the passenger domain (between amino acids 70-473). Cha therefore joins a growing group of self-associating autotransporters that includes the *E. coli* UpaG, TibA,

Ag43, and AIDA adhesins (Klemm *et al.* 2006; Valle *et al.* 2008). The same N-terminal 473-residues of Cha capable of promoting adherence to epithelial cells are sufficient for mediating bacterial aggregation, and thus far these two adhesive domains are not separable.

In summary, Cha, the major adhesin of the *Haemophilus* cryptic genospecies, appears to be a multipurpose protein like other members of the trimeric autotransporter family. Expression of Cha presumably aids in bacterial colonization and microcolony formation in the maternal genital tract and the neonatal respiratory tract. Both adhesive activities reside in the N-terminus of the protein which includes extensive unique sequence, a well-conserved Ylhead repeat cluster, and a neck motif. In ongoing studies, we are examining the functional significance of other structural domains within the passenger domain including the predicted stalk domain.

Chapter 4. Variable Adhesive Activity of the *Haemophilus* Trimeric Autotransporter Cha

4.1 Introduction

The *Haemophilus* cryptic genospecies is a gram-negative, unencapsulated bacterium that is naturally maintained in the female genital tract of humans. It has recently been recognized as an important mucosal surface pathogen responsible for causing a variety of urogenital infections, including urethritis, vaginitis, cervicitis, salpingitis, endometritis, and Bartholin's gland abscess (Turk 1984; Quentin *et al.* 1989; St Geme 1993). Colonization of the maternal genital tract can lead to chorioamnionitis and premature rupture of membranes. Aspiration of contaminated amniotic fluid and mucosal secretions during passage through the birth canal leads to bacterial colonization of the neonatal respiratory tract and in some cases respiratory distress and systemic infection (Albritton *et al.* 1982; Wallace *et al.* 1983; Friesen and Cho 1986; Martel *et al.* 1989; Rosenau *et al.* 1993). Strains belonging to this unique species account for 25-38% of reported maternal, genital, and neonatal *Haemophilus* infections (Albritton *et al.* 1982; Wallace *et al.* 1983; Friesen and Cho 1986; Rosenau *et al.* 1993). We recently reported identification of the novel Cha adhesin, which is conserved among strains belonging to the cryptic genospecies and mediates efficient attachment of bacteria to epithelial cells of both genital and respiratory origin (Sheets *et al.* 2008). Given that colonization represents a necessary step in the development of maternal and infant infections, Cha

presumably contributes to the apparent genital and respiratory tissue tropism of this pathogen.

The Cha adhesin belongs to the trimeric autotransporter (TAA) family of proteins. The defining feature of this group is the C-terminal membrane anchor domain. An N-terminal signal peptide directs secretion of the protein through the Sec apparatus into the periplasm. Subsequently the C-terminal domain targets the protein to the outer membrane where it trimerizes and facilitates surface presentation of the passenger domain. The β-barrel is occluded by a short right-handed coiled-coil which connects with the passenger domain of the protein. In all cases, the passenger domain harbors the adhesive activity of the protein. Structural mapping studies indicate that trimeric autotransporter proteins are highly modular and are characterized by well-conserved structural and sequence motifs (Linke *et al.* 2006). Generally, TAAs adopt a head-stalk-anchor organization with neck adaptor motifs facilitating transition between large head motifs and coiled-coil stalk domains (Szczesny and Lupas 2008). The most well-described head groups include the β-roll degenerate repeats of the *Yersinia* YadA adhesive head (YadA-like head) (Nummelin *et al.* 2004) and the Trp-ring and GIN motifs of the *Haemophilus influenzae* Hia binding domains (Yeo *et al.* 2004).

The Cha protein contains several clusters of YadA-like head (Ylhead) degenerate repeats, head insert motifs (HIN2), neck adaptor domains, and coiled-coil stalk domains throughout the passenger domain. We have recently localized the host cell binding

domain of Cha to the N-terminal 473 residues of the protein, a region that contains unique, structurally undefined sequence followed by a small cluster of Ylhead repeats and a neck motif (Sheets and St. Geme, unpublished). We also observed that Cha mediates bacterial aggregation via Cha-Cha interactions. Interestingly, the autoaggregative domain and the host cell binding domain of Cha appear to co-localize (Sheets and St. Geme, unpublished).

The sequence composition of left-handed coiled-coiled stalk domains varies from protein to protein but is usually characterized by repetitive sequences with heptad periodicity, whose seven positions are denoted a-g, with small hydrophobic residues (commonly alanine) preferentially occurring in the first (a) and fourth (d) positions. The Cha protein is predicted to contain two trimeric coiled-coil domains within the passenger domain. Notably, the second stalk region is associated with the presence of large, unique compound repeats (Sheets *et al.* 2008). These tandem identical repeats encode alanine/threonine-rich 28-residue repeats and genomic sequencing of strain 1595 (Mardis *et al.*, unpublished) suggests this unique sequence unit within the *cha* gene is repeated nine times and is followed by a 30-bp partial repeat (Sheets *et al.* 2008).

Consistent with previous reports, we have observed a general correlation between the appearance of peritrichous fibers in cryptic genospecies strains and an ability to adhere to epithelial cells, suggesting an adhesive function (Quentin *et al.* 1989; Rosenau *et al.* 1993; Gousset *et al.* 1999). Previous studies concluded that these fibers are

morphologically distinct from *H. influenzae* heamagglutinating Hif pili (Bruant *et al.* 2002). We observed a disappearance of surface fibers associated with the disruption of the *cha* locus; however, expression of Cha in *H. influenzae* strain DB117 did not yield piliation (Sheets *et al.* 2008). While it is possible the fibers are encoded by an unidentified locus dependent on the expression of Cha, we speculated that visualization of Cha fibers requires the presence of a threshold number of the 28-amino acid Cha repeats which were absent in our recombinant DB117/pCha strain.

In this study, we localize a size polymorphism associated with the *cha* gene from strain 1595 to the repeat-containing 3′ region. To study the role of the unique repeat domain of Cha, we isolated strain variants with different numbers of repeats and evaluated the adhesive function of Cha by quantitative adherence assays. The relationship between repeats and self-association of Cha was further evaluated using settling assays. We report that expansion in Cha coding repeats results in fiber elongation and negatively impacts bacterial aggregation and adherence to Chang epithelial cells.

4.2 Materials and Methods

Culture and storage conditions

Haemophilus cryptic genospecies strains were grown on chocolate agar plates (BD, Franklin Lakes, NJ) at 37°C with 5% CO_2 and were stored at -80°C in brain heart infusion broth with 25% glycerol.

Southern hybridization

Southern analysis of cryptic genospecies repeat variant strains was performed as described previously (Sheets *et al.* 2008). Approximately 1 μg of chromosomal DNA from each repeat variant was double-digested with DraI/HindIII, separated by agarose gel electrophoresis, and transferred to nitrocellulose. An ECL-labeled DNA probe was generated by PCR amplification of the partial *cha* locus from strain 1595 with primers 5′-ACTGCTGAAGTTAATCCTATTTCAGAACTTAC-3′ and 5′-CCATGTTAGCTCCAGGGATGG-3′ [bp3112-4678]

Isolation of Repeat Variant Strains

Haemophilus cryptic genospecies clinical strain 1595, a neonatal blood isolate, was resuspended in PBS to an optical density of 0.8 at 600nm. A 10-μl volume of this bacterial suspension was inoculated onto a confluent Chang epithelial cell monolayer in a 24-well plate. To facilitate bacterial contact with the epithelial cells, plates were centrifuged at 165 x g for 5 min. After a 25-min incubation at 37°C in 5% CO_2, the supernatant was transferred to a new confluent Chang cell monolayer. To enrich for non-adherent variants, passage of supernatant bacteria over the cell monolayers was repeated six times and then dilutions of the supernatant were plated to yield individual colonies. To enrich for adherent variants, monolayers were inoculated and incubated with strain 1595 bacteria, rinsed four times with PBS, and trypsinized to release the cells. Dilutions containing adherent bacteria were plated, passaged overnight, and

reinoculated six successive times onto confluent Chang cell monolayers for further enrichment. Isolated colonies were evaluated by Southern analysis to identify variants and estimate the approximate number of 84-nucleotide repeats. Non-adherent variants were reinoculated onto Chang cell monolayers to re-enrich for adherence and identify additional repeat variants.

Quantitative Adherence Assays

Quantitative adherence assays were performed with Chang conjunctiva cells (Wong-Kilbourne derivative, clone 1-5c-4; ATCC CCL-20.2) as previously described (St Geme *et al.* 1993). Percent adherence was calculated by dividing the number or adherent CFU per epithelial cell monolayer by the number of inoculated CFU. Each repeat variant of strain 1595 was examined in triplicate in a given assay, and generally, several independently-derived repeats variants representing the same number of repeats were evaluated. Chang cells were maintained in modified Eagle medium with Earle's salts and nonessential amino acids supplemented with 10% fetal bovine serum.

Electron Microscopy

Bacteria were allowed to absorb onto carbon/formvar-coated grids for 1 min and were then washed twice in distilled H_2O and stained with 1% aqueous uranyl acetate. Bacteria were prepared for immunoelectron microscopy as described by Roberts *et al.* (Roberts *et al.* 1994) with minor modifications. Briefly, 200-nm mesh, carbon/formvar-coated nickel grids (LADD) were each overturned on a drop of bacterial suspension in

0.2 M ammonium acetate, pH 7.4 and allowed to absorb for 1 min. Subsequently, samples were blocked for 1 h with PBS containing 2% goat serum and 0.1% gelatin and were then incubated for 2 h with a 1:250 dilution of rabbit polyclonal antiserum (anti-Cha aa70-473). Samples were then washed three times with PBS and incubated for 1 h with a 1:30 dilution of goat anti-rabbit IgG conjugated to 12-nm colloidal gold beads (Jackson ImmunoResearch). Samples were washed again, fixed with 1% glutaraldehyde in PBS, washed with PBS followed by distilled H_2O, and stained with 1% aqueous uranyl acetate for 30 seconds. Excess liquid was gently wicked away, and the grids were air-dried. Samples were viewed with a JEOL 1200EX transmission electron microscope (JEOL USA, Peabody, MA) at an accelerating voltage of 80 kV.

Quantitative Real time PCR

Bacteria representing 6 repeat variants (A, C, I, J, K, L) were grown overnight on chocolate agar, resuspended in PBS to equal density, pelleted and lysed using TRI® reagent (Sigma). Total RNA was isolated using the RNeasy mini kit (Qiagen) according to the manufacturer's instructions for lipid-rich tissue, and residual DNA was degraded using DNase I (New England Biolabs). RNA integrity was verified by gel electrophoresis and RNA concentration was quantified spectrophotometrically. Using 1-μg of template RNA, cDNA was generated with random hexamers and iSCRIPT reverse transcriptase (Bio-Rad) following the manufacturer's protocol.

The primers used were as follows: ftsZ-sense: 5′-ATTCCGTGATGGGAATTTCA-3′, ftsZ-antisense: 5′-CACAGCTGAACCAAAACCAA-3′, cha-sense: 5′-GCCGTAGCTGTCGGTTACTC-3′, cha-antisense: 5′-CAATACCAGCACCGAATGTG-3′. These primers were designed using the Primer3 program tool (Rozen and Skaletsky 2000). The *cha* primers were designed to amplify a 106-bp fragment of DNA 3′ to the 84-nucleotide repeats. In each variant strain, the expression level of *cha* was normalized to the internal control, *ftsZ*, a house-keeping gene.

The relative quantification of target gene expression was done using the comparative cycle threshold (C_T) method. C_T is defined as the cycle number at which SYBR green I (Invitrogen) fluorescence intensity exceeded 10X the standard deviation of the baseline fluorescence. The relative target expression was given by the formula: $2^{-\Delta\Delta CT}$, where $\Delta\Delta C_T = \Delta C_{T\ \text{target sample}} - \Delta C_{T\ \text{calibrator sample}}$ with $\Delta C_T = \Delta C_{T\ \text{cha}} - \Delta C_{T\ \text{ftsZ}}$. Samples were run in triplicate and results are presented as the mean fold-change in *cha* transcript relative to the zero-repeat variant (1595-A) ± standard deviation.

Dot immunoblot analysis

Bacteria grown overnight on chocolate agar were resuspended in BHI to an OD_{600} of 0.3, and 10-ul suspensions were diluted 1:10 and inoculated into wells of a 96-well dot blot manifold apparatus holding a nitrocellulose membrane. Samples were pulled through the filter by vacuum suction for 5 min. After blocking for 1 hr with Tris-

buffered saline plus 5% milk, surface-exposed Cha was detected with guinea pig anti-Cha(aa70-473) antiserum diluted 1:10,000.

4.3 Results

4.3.1 Size polymorphism localizes to 3' *cha* locus

Previously we reported the universal presence of the *cha* locus in our collection of clinical isolates belonging to the *Haemophilus* cryptic genospecies. However, Southern analysis revealed a slight size variation among *cha*-hybridizing fragments (Sheets *et al.* 2008). Available genomic sequence for strain 1595 indicates that compound repeats are located towards the C-terminal end of the passenger domain of Cha. With this information in mind, we hypothesized that the size polymorphism from strain to strain reflects spontaneous repeat variation. To further localize the polymorphic region, we performed additional Southern analysis using DraI/HindIII-digested genomic DNA from isolated colonies of strain 1595 and probed with a DNA fragment that hybridizes to *cha* bp3112-4678 (assuming 9 repeats). As shown in figure 9, the size polymorphism localizes to the repeat-containing 3′ region of the *cha* gene.

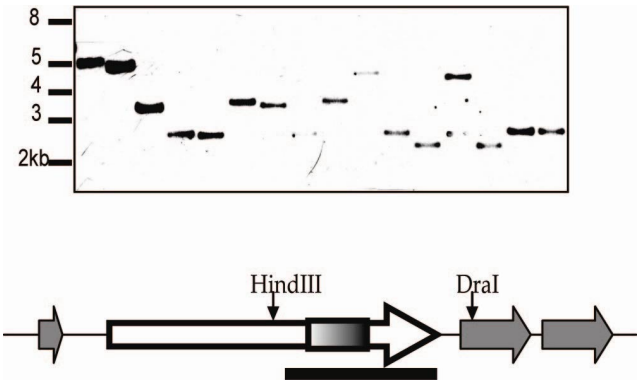

Figure 9: A naturally occurring size variation in the 3′ region of cha locus.

Southern hybridization analysis of *Haemophilus* cryptic genospecies strain 1595 colony variants. Chromosomal DNA was isolated from 16 isolated variants and digested to completion with DraI/HindIII. The schematic depicts the location of the relevant restriction sites flanking the repeat-containing 3′-region of the *cha* gene. The black bar represents the ECL-labeled probe used for hybridization and was generated by PCR amplification of the partial *cha* locus (nucleotides 3112-4678, including 9 repeats) from strain 1595.

The observed size variation associated with *cha* could reflect repeat addition/deletion or could be the result of a sequence element insertion for which the repeats represent a recognition site. To distinguish between these two possibilities, we cloned ~11kb HindIII-digested, *cha3′*-containing fragments from both strain 1595 and strain 1673 into pT7-7, verified insertion by dot blot using a *cha3′* probe, and performed restriction analysis with HincII on the recombinant plasmid. A single HincII site occurs once within each 84-nucleotide repeat, and the remaining fragments can easily be assigned to flanking regions of *cha* and the vector backbone. The observed restriction patterns for both clones indicated that the size polymorphism was indeed the result of repeat expansion/contraction (Figure 10). Reflecting on the Southern analysis of

apparent repeat variants in Figure 9, our observation of single hybridizing repeat-containing fragments suggests that only one copy of the *cha* gene is present in the genome of strain 1595. This finding further indicates that there is no reservoir of *cha* repeats (other than intragenic copies) elsewhere in the genome, a conclusion that was confirmed by examination of the total genome of strain 1595 (Mardis *et al.*, unpublished).

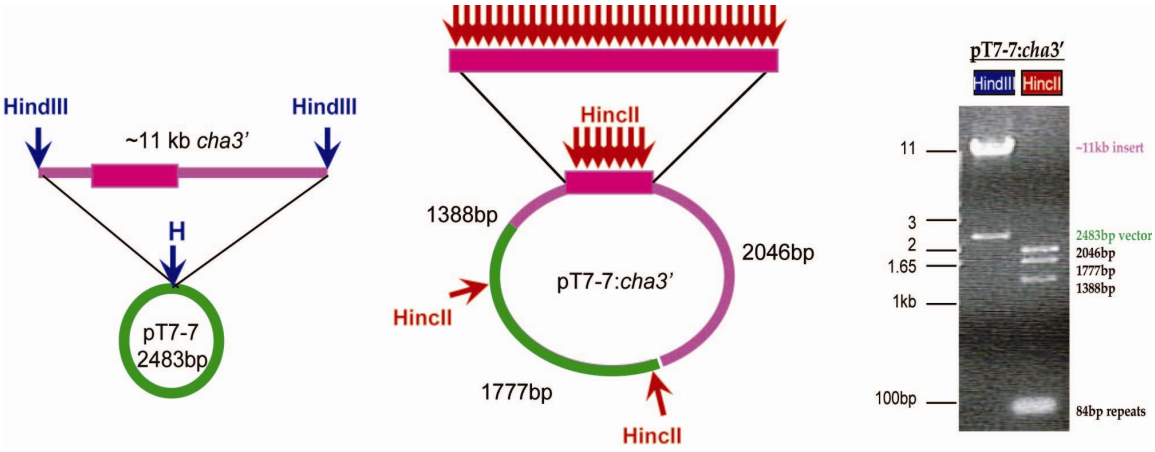

Figure 10: Restriction analysis of recombinant pT7-7:cha3′.

A HindIII-digested, *cha3′*-containing ~11kb fragment from strain 1595 colony variant was cloned into HindIII-digested pT7-7. Positive clones were identified by dot blot using a *cha3′* probe. Arrows in the schematic diagram denote the locations of HindIII (blue) and HincII (red) restriction sites. A single HincII site occurs once within each 84-bp repeat (pink). The HincII restriction pattern suggests extensive expansion of repeats (~100).

4.3.2 Effect of tandem repeats on binding to host epithelial cells

Variation in 84-bp coding tandem repeats does not alter the reading frame; however, changes in repeat copy number may alter the expression or function of Cha. Spontaneous deletion of the 84-bp repeats upon amplification and cloning of the *cha*

gene did not interfere with the adhesive function of Cha when expressed in *H. influenzae* lab strain DB117 (Sheets *et al.* 2008), indicating that this tandem repeat domain is not required for proper processing or folding of the Cha protein. To study the impact of these repeats in more detail, an enrichment scheme was developed to isolate repeat copy number variants (see Materials and Methods). Numerous variants representing zero to ~100 repeats were identified by Southern analysis (Figure 11A). We tested the adherence of 12 representative variants to Chang conjunctival cells. As shown in figure 11B, we observed an inverse correlation between repeat copy number and percent adherence. Adherence by the zero-repeat variant 1595-A was nearly 83% of the inoculum. Variants 1595-K and 1595-L contain ~76 and ~95 repeats, respectively, and were non-adherent. Similar colony variants were obtained for strain 1673 and we observed the same inverse correlation between repeat copy number and binding affinity (not shown), demonstrating that this phenomenon is not unique to strain 1595 but presumably represents a conserved mechanism of modulating the adhesive activity of the major adhesin of the *Haemophilus* cryptic genospecies.

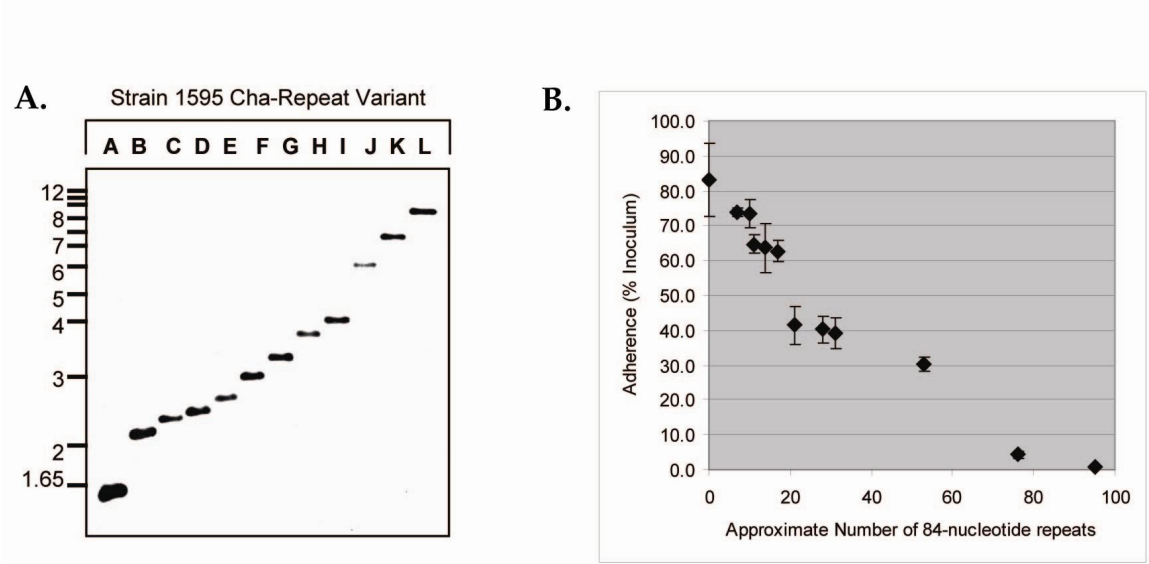

Figure 11: Relationship between repeat copy number and Cha adhesive activity.

(A) Southern hybridization analysis of 12 representative strain 1595 repeat variants. Isolated colony variants (A-L) were identified after passing pools of bacteria over Chang cell monolayers, enriching for more-adherent and less-adherent variants. Chromosomal DNA was isolated and digested to completion with DraI/HindIII. An ECL-labeled probe used for hybridization was generated by PCR amplification of the partial *cha* locus (nucleotides 3112-4678, including 9 repeats). A 1.58 kb-hybridizing fragment is consistent with complete deletion of the 84-nucleotide repeats (variant 1595-A). (B) Quantitative adherence assay with Chang cells examining adhesive activity of each representative repeat variant. Repeat copy numbers ranging from zero to ~95 were estimated by Southern analysis according to *cha*-hybridizing fragment size. Values represent mean percent adherence by strain variants examined in triplicate in at least three different assays. Error bars represent standard error.

4.3.3 Repeat expansion interferes with bacterial aggregation

To further study the functional significance of the 84-bp repeats in this study, we evaluated settling among our collection of strain 1595 repeat variants (Fig. 12). Repeat variants A(0) through I(31) demonstrated rapid settling at progressively decreasing rates with increasing repeat copy number. Both mutant strain 1595*cha-* and strain 420 which

does not express Cha protein (Sheets *et al.* 2008), gradually lost turbidity, suggesting that a cell surface factor with lower affinity than Cha also mediates cell-cell interactions. Notably, 1595 variant strains J(53), K(76), and L(95) did not aggregate and settle in the 6.5 hr assay. This observation suggests there may be a threshold for repeat expansion beyond which, Cha protein becomes anti-aggregative by sterically interfering with cell-cell interactions mediated by a non-Cha molecule. Given that the host cell binding and autoaggregative domains co-localize to the N-terminus of Cha, it is noteworthy that strain J(53) is adherent to Chang cells yet poorly aggregative suggesting that the repeat copy number threshold for loss of self-association is lower than that for loss of adhesion to host cells.

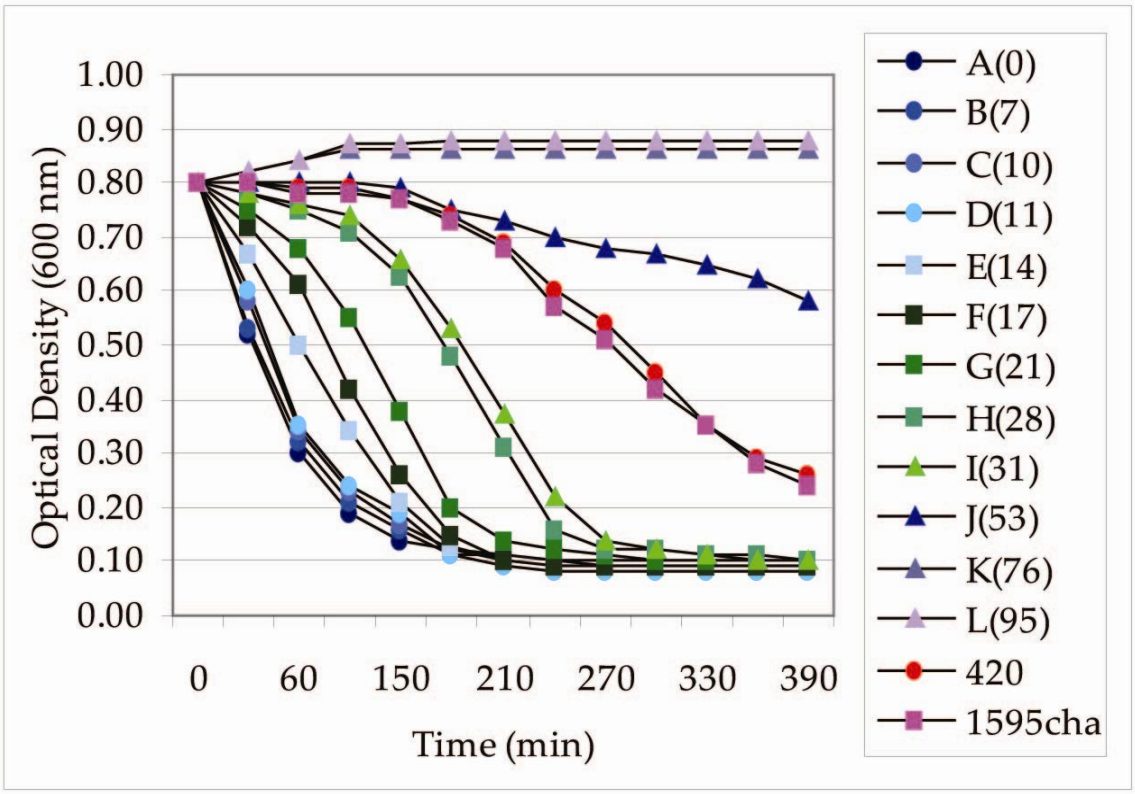

Figure 12: Impact of Cha tandem repeat expansion/contraction on bacterial aggregation.

Quantitative analysis of autoagglutination of suspensions of *Haemophilus* cryptic genospecies strain 1595 repeat variants A-L, a 1595*cha*- mutant, and non-adherent strain 420(Cha-). Bacteria were resuspended in BHI to an optical density of 0.8 at 600 nm. Bacterial aggregation and settling was evaluated by measuring the absorbance every 30 min for 6.5 hours. Numbers in parenthesis represent approximate repeat copy number.

4.3.4 Repeat variation does not impact *cha* transcript levels or surface localization of Cha

To investigate the mechanism by which repeat expansion interferes with Cha adhesive function, we compared *cha* transcript levels among our repeat variant strains. Quantitative real time PCR demonstrated that no significant differences in *cha* mRNA

can be appreciated among repeat variants C(10), I(31), J(53), K(76), and L(95) relative to zero-repeat variant A(0) (Fig. 13A). Noticing an overrepresentation of 2 particular codons within the repeat tract, we considered that codon usage might impact translation efficiency and protein expression levels of Cha. The repeat unit is highly enriched for the less common CGA(Ala) and ACT(Thr) codons compared to *cha* coding sequence and *H. influenzae* strain Rd codon usage frequencies (evaluated in the absence of an annotated cryptic genospecies genome). In semi-quantitative experiments evaluating abundance of surface-localized Cha protein by dot immunoblot with whole bacteria, we did not observe diminished quantities of Cha protein associated with repeat expansion (Fig 13B). High repeat variants J(53), K(76), and L(95) exhibited robust expression of Cha protein similar to that of zero-repeat A(0). Lack of signal for mutant strain 1595*cha*- demonstrated that binding of Cha antiserum was highly specific. Taken together, these data suggest that repeat variation does not significantly impact expression or surface localization of Cha protein and that a transcription-independent mechanism accounts for the interference of repeat expansion with adhesive activity.

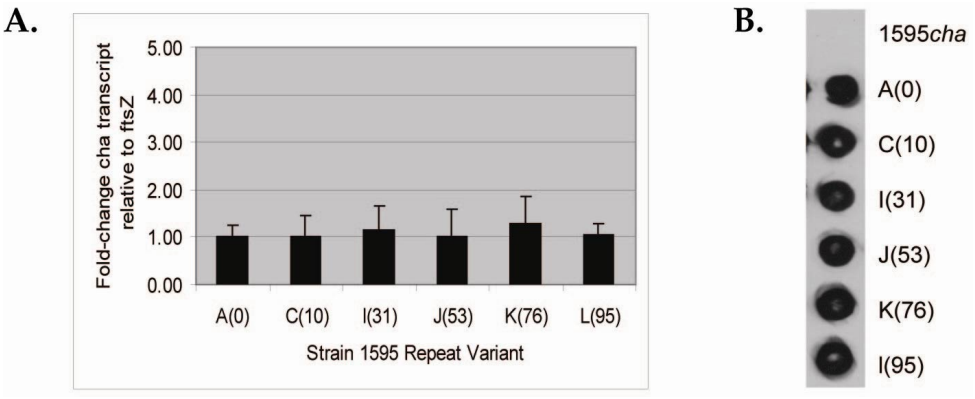

Figure 13: Variation in repeats does not affect transcription or surface localization of Cha.

(A) The expression level of *cha* transcript in strain 1595 repeat variants C(10), I(31), J(53), K(76), and I(95) relative to zero-repeat strain A(0) was measured by quantitative real-time PCR. Vertical bars represent the mean fold-change ± SD (N = 3, P>0.05 using the unpaired t test.) (B) Dot immunoblot assay for detection of Cha on bacterial cell surface. Reactivity of whole-cell variant strains A(0), C(10), I(31), J(53), K(76), and I(95) with Cha antiserum demonstrates no discernible difference in protein expression.

4.3.5 Visualization of Cha fibers

Clinical isolates belonging to the cryptic genospecies have routinely been reported to display abundant peritrichous pili. These reports prompted investigators to probe isolates for conservation of *H. influenzae* haemagglutinating Hif pili genes. Although rare isolates appear to contain components of the *hif* gene cluster (Gousset *et al.* 1999), these isolates fail to agglutinate human red blood cells and do not express immunoreactive HifA (major subunit) or HifE (tip adhesin) protein (Clemans *et al.* 2001). Furthermore, in our own studies, pilus-like structures visualized by negative staining electron microscopy appear to be shorter and thinner than *Haemophilus* Hif pili (Fig. 14),

suggesting that a novel gene is responsible for the observed fibers. Consistent with previous reports (Clemans *et al.* 2001), we observed surface fibers in strain 1595, the parent strain from which *cha* was identified. Disruption of this locus resulted in loss of surface fibers, suggesting *cha* encodes the fibers (Sheets *et al.* 2008). However, expression of Cha by *H. influenzae* strain DB117 and *H.* cryptic genospecies strain 420 to demonstrate adhesive activity did not result in the appearance of surface fibers, presumably the result of spontaneous deletion of the tandem repeats during cloning (Sheets *et al.* 2008). Using the daTAA web-based domain annotation tool for trimeric autotransporters, the sequence encompassing the Cha repeats is predicted to form a left-handed trimeric coiled-coiled stalk domain (http://toolkit.tuebingen.mpg.de/dataa) (Szczesny and Lupas 2008). This structural motif is defined by heptad periodicity of small hydrophobic residues, a periodicity that is present upon expansion of the unique 28-residue Cha repeat (Fig. 15). Given this information and the previously reported correlation between adherence and surface fibers observed among clinical isolates, we investigated the identity of the surface fibers present on repeat variants by immunogold labeling using rabbit polyclonal anti-Cha antiserum. As shown in Figure 16, the Cha antiserum specifically labeled the surface fibers of strain 1595. Repeat variant L (~95 repeats) had long, thin tangled fibers that appear to interact at the distal tip representing the N-terminus of the protein. Thin, flexible fibers were also observed on repeat variant J (~65 repeats) albeit seemingly shorter in length. Variant A (zero repeat) did not exhibit

fibers, but gold particles appeared intimately associated with the bacterial outer membrane. The negative staining outlines the presence of electron dense globular structures that likely represent the β-roll head motifs within the Cha passenger domain.

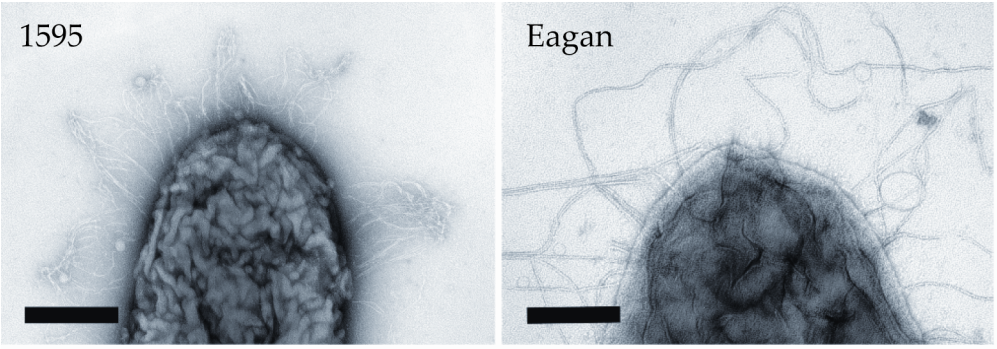

Figure 14: Visualization of *Haemophilus* cryptic genospecies surface fibers

Electron micrographs of *Haemophilus* cryptic genospecies strain 1595 and *H. influenzae* strain Eagan negatively stained with uranyl acetate. Abundant cryptic genospecies surface fibers are morphologically distinct from the *Haemophilus* Hif pili expressed by strain Eagan. Scale bar=200nm.

```
abcdefg abcdefg abcdefg abcdefg abcdefg abcdefg abcdefg
*  *    *  *    *  *    *  *    *  *    *  *    *  *
ATALNTL ATDAKTA AAVADSK ATVASTA ASTADSK ATVASTA AAVADSK
```

Figure 15: The unique Cha repeat unit.

Sequence of a single repeat unit (underlined) is presented with flanking passenger domain sequence. This predicted coiled-coiled stalk domain exhibits heptad periodicity for which the seven positions are denoted a-g with small hydrophobic residues (commonly alanine) occurring every first (a) and fourth (d) position.

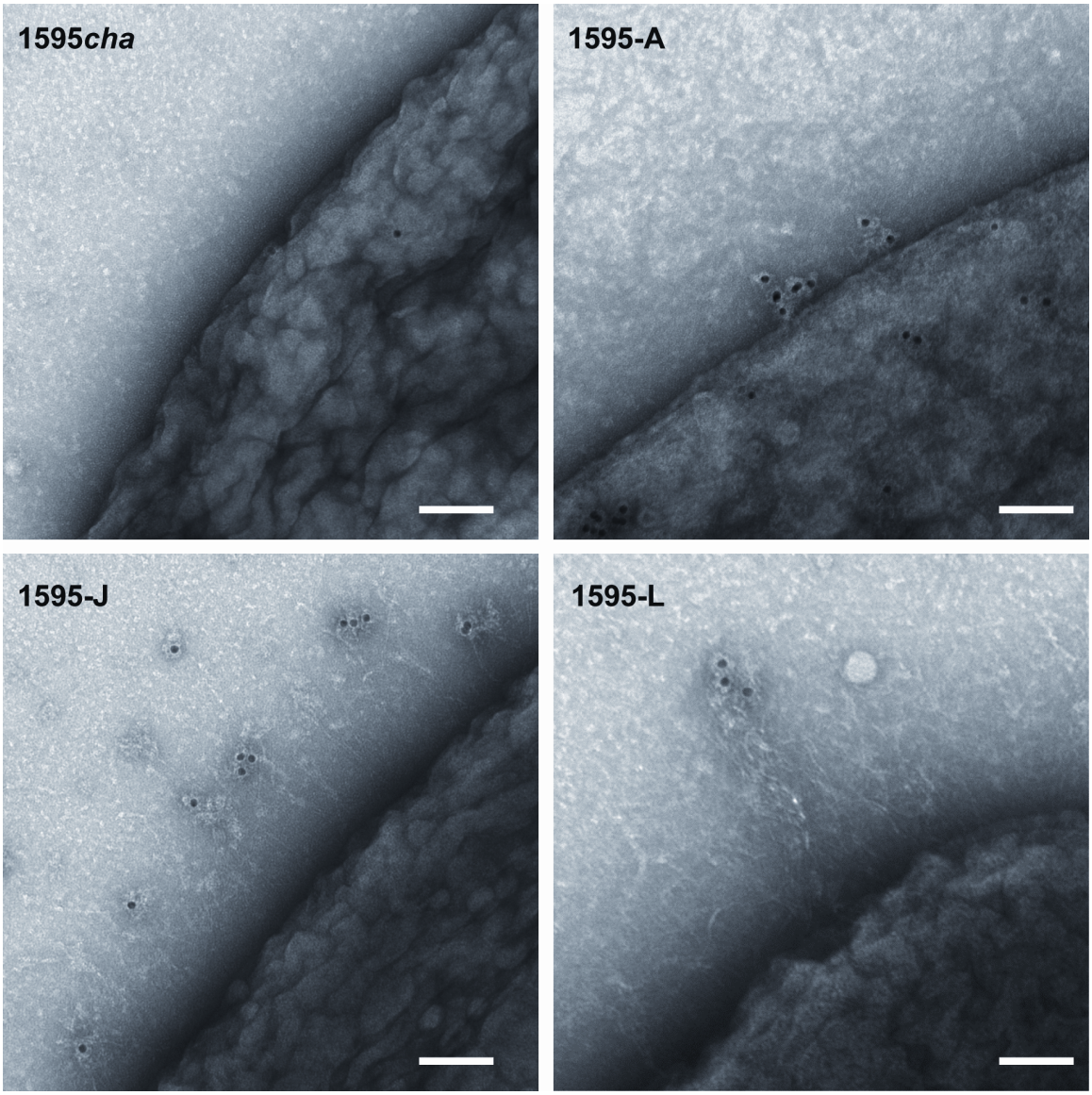

Figure 16: Immunogold labeling of surface fibers.

Electron micrographs of strain 1595 repeat variants negatively stained with uranyl acetate. Antibody raised against the N-terminal head fragment of Cha exclusively labels the distal tips of thin, tangled fibers visualized on both non-adherent strain 1595-L as well as adherent strain 1595-J. Abundant labeling of Cha protein is not associated with surface fibers on the highly-adherent, zero-repeat strain 1595-A. Non-adherent 1595*cha* mutant exhibits no surface fibers and no immunoreactive protein. Scale bar=100nm.

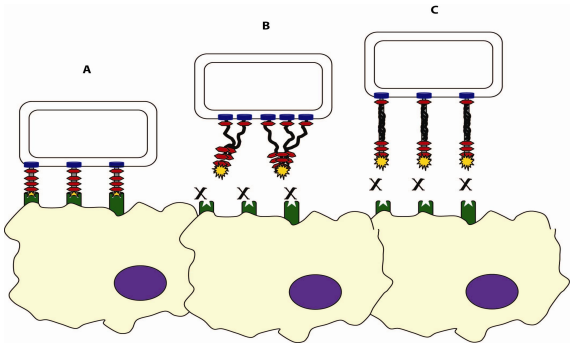

Figure 17: Models for impact of repeat variation on host cell binding.

(A) Cha mediates bacterial attachment to host epithelial cells via an N-terminal binding domain. (B) Repeat expansion results in the formation of long, flexible Cha fibers that have a propensity to interact at the distal tip. This aggregative interaction of intrabacterial Cha fibers occludes the host cell binding domain and inhibits adherence. (C) Expansion of the repeat-encoded coiled-coil stalk domain has structural consequences on neighboring motifs and alters the N-terminal binding-pocket conformation.

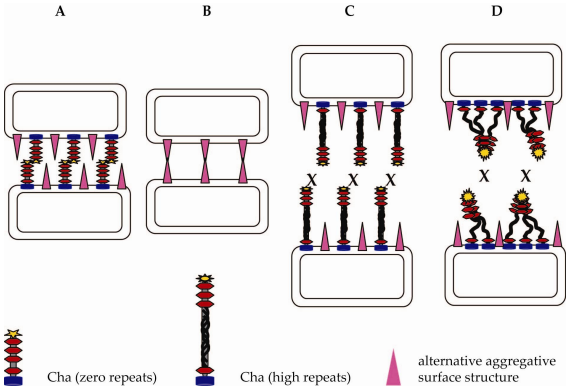

Figure 18: Model for inhibition of bacterial aggregation by Cha repeat expansion

(A) Cha molecules (lacking repeats) mediate intimate bacterial interactions. (B) In the absence of Cha on the bacterial cell surface, an alternative surface structure mediates bacterial aggregation. (C) Repeat expansion alters the conformation of the N-terminal aggregative domain, making it nonfunctional. In addition, the presence of long Cha fibers masks the function of an alternative aggregative surface structure. (D) Upon repeat expansion, N-terminal Cha aggregative domains interact with neighboring fibers, thus inhibiting interbacterial Cha-Cha interactions. Again, the presence of long Cha fibers sterically interferes with the function of an alternative aggregative surface structure.

4.4 Discussion

Genes containing several tandem repeats represent hypervariable components of bacterial genomes. Slipped-strand mispairing of non-coding tandem repeats can impact expression of key virulence factors, such as the *H. influenzae* Hif pili (van Ham *et al.* 1993) and HMW adhesins (Dawid *et al.* 1999) by altering promoter affinity for RNA polymerase. Changes in tandem repeats can also mediate phase variation if the repeated units change the translational reading frame, resulting in altered amino acid sequence and oftentimes truncation of the encoded protein (De Bolle *et al.* 2000), as exemplified by the *opa* locus in *Neisseria gonorrhoeae* (Murphy *et al.* 1989) and the *lic* loci in *H. influenzae* (Weiser *et al.* 1990). Expansion and contraction of coding repeats does not alter the reading frame with copy number and is associated with functional diversity among numerous outer membrane and cell wall associated proteins. Gravekamp and colleagues showed that variation in repeat number in the Alpha C protein among clinical Group B streptococci isolates alters the antigenicity of immunogenic epitopes (Gravekamp *et al.* 1996). Such repeat variation facilitates efficient adaptation to the host environment. Consistent with this, numbers of coding repeat copies frequently vary among clinical isolates (Jordan *et al.* 2003). Size-variant Vaa antigens expressed by *Mycoplasma hominis* have been linked to changes in 121-residue repeats within clonal populations of this organism, suggesting frequent recombination events (Zhang and Wise 1996). Repeat variation has also been reported in fungi, highlighting a broadly

conserved mechanism of generating diversity (Jordan *et al.* 2003; Levdansky *et al.* 2007). In the yeast, *Candida albicans,* an increase in the number of 108-nucleotide tandem repeats in the *ALS5* gene modulates the structure of the substrate-binding domain and augments adhesion to fibronectin as well as promoting cell-cell interactions (Rauceo *et al.* 2006).

In this study, we found that spontaneous changes in identical tandem repeats located in the passenger domain of the mature protein alters the adhesive activity of Cha. These Cha repeats are unique for several reasons: 1) These repeats are identical at the nucleic acid level. 2) These 84-nucleotide repeats are coding, and therefore copy number variation does not alter the reading frame. 3) The repeat unit is significantly larger than those found in most other gram-negative bacterial surface proteins. 4) There is a striking GCA and ACT codon bias resulting in an alanine/threonine-rich protein domain. 5) An invariant alanine occurs at every first and fourth residue within repeating heptads. Examination of our collection of 12 representative repeat variants revealed a inverse linear correlation between repeat copy number and host cell adherence; however, we cannot rule out the possibility of an underlying periodic impact of repeat expansion. Quantitative RT-PCR analysis demonstrated that the mechanism by which repeat variation influences adhesive function of Cha does not function at the level of transcription. Additionally, no significant decrease in Cha surface expression could be detected among high repeat variants. Using the daTAA web-based domain

annotation tool for trimeric autotransporters, we predict that this repeat tract forms a left-handed trimeric coiled-coiled stalk domain. Consistent with this possibility, immunogold labeling and negative staining of high repeat variants highlighted the presence of abundant Cha surface fibers.

Repeat expansion seemingly extends the N-terminal clusters of Ylhead head domains of Cha away from the bacterial cell surface. Expansion to one hundred 84-bp repeats would yield a massive 386 kDa protein, 2/3 of which is dominated by the repeat-encoded domain, raising the possibility that significant repeat expansion may modulate neighboring structural motifs as reported for other adhesins such as the Candida Als5p protein (Rauceo *et al.* 2006). Elongation of Cha fibers may alternatively shield adhesive epitopes proximal to the cell surface. However, in recent work using in-frame deletion derivatives, we localized the host epithelial cell binding domain to the N-terminal 473-residues of Cha. Considering this information, expansion of the repeat-encoded coiled-coil stalk domain would effectively extend the distal binding domains away from the bacterial cell surface. We therefore would not expect that addition of repeats sterically masks access to the functional binding domain. However, considering that addition of 75-100 repeats creates a protein that is dominated by the repeat tract (corresponding to 58-65% of the total protein), it is possible this has structural consequences for the protein. This overwhelming repeat motif may alter folding at the N-terminus from a binding-competent conformation (Fig. 17).

Recent studies indicated that Cha contributes to bacterial microcolony formation via interactions between the N-termini of interbacterial Cha molecules. The interference of repeat expansion with host cell adhesion prompted us to investigate the relationship between repeat copy number and bacterial aggregation. Our settling assay observations indicate that expansion of the tandem 84-bp repeats decreases the rate of bacterial aggregation. Taking into account that the host cell binding and aggregative domains co-localize to the N-terminus, it is tempting to consider that extensive repeat expansion and Cha fiber elongation enables the N-termini of intrabacterial Cha fibers to interact. This interaction could interfere with both adherence and bacterial aggregation by occluding the adhesive domains (Fig. 17 & 18). This idea is supported by negative-stained images showing a tangling of fibers with evident interactions at the distal tips of neighboring fibers. Conversely, we suspect that the low-repeat Cha aggregative domains are not accessible to intrabacterial Cha molecules and therefore are more accessible to mediate intimate interactions to host epithelium and neighboring bacteria. We do not consider our hypotheses explaining the fuctional impact of repeat variation to be mutually exclusive, as conformational changes and accessibility of binding domains could both factor into the binding target preference.

Interestingly, we also observed that while the high repeat variants inhibit settling, the 1595*cha*- mutant strain exhibits a slow rate of settling. This observation suggests that the presence of an expansive repeat tract yields Cha fibers that mask the

function of an additional aggregative factor (Fig. 18). A similar phenomenon has previously been associated with TibA-mediated aggregation of Enterotoxigenic *E. coli* (Sherlock *et al.* 2005). Expression of either capsule or polymeric pili effectively masks the monomeric TibA adhesin, which is much more intimately associated with the bacterial outer membrane. We speculate that bacterial aggregation mediated by zero-repeat Cha as well as another unknown surface structure results in a more intimate cell-cell contact than is possible by the long, filamentous projections of medium to high-repeat Cha.

This study demonstrates that Cha repeat variation has broad functional consequences. While loss of repeats may facilitate colonization and survival in the genital or respiratory tract, repeat expansion may facilitate dispersion, invasion, and possible immune escape. Repeat expansion may also confer adhesive activity toward different, as yet unidentified, cell types or extracellular matrix components. While trimeric autotransporters are commonly characterized by highly repetitive architecture, to our knowledge this is the first instance of a trimeric autotransporter being modulated by long, identical, tandem repeats and may be a unique mechanism by which the cryptic genospecies enhances its adaptive capability in the human host.

Chapter 5. Identification of a smaller, antigenically-divergent Cha2 adhesin in the *Haemophilus* cryptic genospecies

5.1 Introduction

Female genital tract and neonatal disease caused by the *Haemophilus* cryptic genospecies begins with colonization of the vaginal tract and neonatal respiratory tract, respectively. Heavy colonization of the vaginal tract facilitates ascending spread of bacteria and can lead to upper genital tract infections, chorioamnionitis, and premature rupture of membranes in the maternal host (Wallace *et al.* 1983; Quentin *et al.* 1989). Aspiration of contaminated amniotic fluid or mucosal secretions during birth predisposes neonates to respiratory distress and in some cases bacteremia (Campognone and Singer 1986; Quentin *et al.* 1989; Rusin *et al.* 1991). Bacterial attachment to mucosal surfaces is a critical step in the progression to both maternal and infant disease. In recent studies we have identified and characterized a novel trimeric autotransporter adhesin called Cha, which is responsible for the broad-range adherence of these organisms (Sheets *et al.* 2008). The passenger domain of Cha harbors the adhesive activity of the protein and is distinct from all other characterized trimeric autotransporters. It is distinguished by a unique arrangement of several clusters of well-conserved Ylhead β-roll repeats, 4 neck motifs, and 2 predicted coiled-coiled stalk domains as well as considerable unique, undefined sequence. The defining C-terminal translocator domain of Cha shows high homology with that of the *Yersinia* YadA and the

H. influenzae Hia adhesins and anchors the protein in the outer membrane in a trimeric conformation (Sheets *et al.* 2008).

The Cha adhesin mediates attachment to host cells through intimate interactions at the N-terminal, distal tip of the molecule (Sheets and St. Geme, unpublished). This region (amino acids 70-473) is comprised of mostly structurally undefined sequence followed by a short cluster of Ylhead β-roll repeats and a neck motif. Additional studies have identified the Cha adhesin as an aggregative protein like a number of other autotransporters (Hendrixson and St Geme 1998; Pearson *et al.* 2002; Klemm *et al.* 2006). Such aggregation and microcolony formation is believed to enhance colonization and bacterial survival on mucosal surfaces. Our studies have shown that the aggregative domain of Cha overlaps the binding domain at the N-terminus, and bacterial flocculation occurs through homo-oligomeric Cha-Cha interactions (Sheets and St. Geme, unpublished). Expansion of 28-residue tandem, identical repeats located in the passenger domain has been shown to progressively interfere with Cha adhesive activity. Repeat variation does not influence transcription of *cha* or surface localization of Cha protein but alters the appearance of thin, flexible Cha surface fibers (Sheets and St. Geme, unpublished). Consequently, it has been hypothesized that proximal repeat expansion has structural consequences on the distal host cell binding domain, although the precise mechanism of this effect remains to be elucidated.

We previously reported conservation of the *cha* locus among our collection of cryptic genospecies isolates and speculated that acquisition of this unique gene contributed to the apparent adaptation of the *Haemophilus* cryptic genospecies to the urogentital tract (Sheets *et al.* 2008). Recent data also indicates that only one copy of the *cha* gene occurs in cryptic genospecies strains (Sheets and St. Geme, unpublished). In the current study, in the process of probing our collection of cryptic genospecies isolates for expression of Cha protein, we identified an adherent strain that did not react with anti-Cha antiserum, leading to the identification of a variant *cha2* allele encoded by a subset of our cryptic genospecies isolates. Sequence analysis revealed that this gene encodes a smaller, antigenically-divergent Cha protein. In spite of striking N-terminal sequence divergence, we demonstrated that Cha2 mediates bacterial aggregation and adherence to epithelial cells of both genital and respiratory origin, similar to the Cha protein from strain 1595 (hereafter Cha1).

5.2 Materials and Methods

Culture and storage conditions

Escherichia coli strains were grown on Luria-Bertani (LB) agar or in LB broth at 37°C and were stored at −80°C in LB broth with 50% glycerol. *Haemophilus* strains were grown on chocolate agar plates (BD, Franklin Lakes, NJ) at 37°C with 5% CO_2 and were stored at −80°C in brain heart infusion broth with 25% glycerol.

Molecular Biology Techniques

DNA ligations, restriction endonuclease digestions, and gel electrophoresis were performed according to standard techniques (Sambrook *et al.* 1989). Plasmids were introduced into *E. coli* by electroporation (Dower *et al.* 1988). In *H. influenzae*, transformation was performed using the MII/MIV method of Herriott *et al.* (Herriott *et al.* 1970). In the *Haemophilus* cryptic genospecies, transformation was accomplished by incubating a 500 μl suspension of bacteria in Schaedler broth with approximately 1 μg of transforming DNA at room temperature for 30 min, then supplementing with clarified horse blood and 2 μg/ml NAD and incubating for 1 hour at 37°C with aeration. Ultimately, transformation reaction mixtures were plated onto chocolate agar containing appropriate antibiotics to select for transformants.

Sequencing of Cha2 gene from *H.* cryptic genospecies strain 1610

To facilitate sequencing of the *cha* locus from strain 1610, coding sequence was amplified using three primer sets (1F/1R, cha2seqG/H, cha2seqI/J), and each fragment was ligated into pCR-XL-TOPO (Invitrogen). The first two primer sets amplified coding sequence upstream of the 84-nucleotide repeats, and the primer set cha2seqI/J amplified sequence 3′ to the variable repeats. Sequence walking was accomplished using primers M13F, M13R, and cha2seqA-J (see Table 5). Each insert was sequenced with 2X coverage and assembled using the ContigExpress program (VectorNTI).

Table 5: Cha2 Sequencing primers

Primer Name	Nucleotide Sequence (5'-3')
1F	CAATGAAACGAATTTACAAAGCTACTCTATTCTC
1R	ATCATTTCTACCAACAGAAACAACCCC
M13F (Invitrogen)	GTAAAACGACGGCCAG
M13R (Invitrogen)	CAGGAAACAGCTATGAC
cha2seqA	GATTGTATATTTATCATCATTTTTATGGC
cha2seqB	GGAATAAAACGACCCAACGATTGGAAG
cha2seqC	GAATGGGCTGACAATATTAGATGGAATAAC
cha2seqD	GAACAACTCTCCATTTGGTATTGGTAATG
cha2seqE	CATTACCAATACCAAATGGAGAGTTGTTC
cha2seqF	GGAGTCGTTTCAGTAGGTAGTTCTAGTGTTAG
cha2seqG	CGTGCTGCATTAAATGAGACAGCG
cha2seqH	TGCTGCAGTTTTGGCATCTGT
cha2seqI	GCAGTAGCTGACTCTAAAGCAGAAGCAGTG
cha2seqJ	TTTACCAAATAAATGACAAAAATACCGCTC

Bacterial strain construction

The *cha* gene in mutant strain 1595*cha-* is disrupted by a *solo* transposon cassette containing the *aphA3* kanamycin resistance gene and was isolated from a Mariner transposon mutant library generated previously (Sheets *et al.* 2008). Genomic DNA was extracted from strain 1595*cha-* and used to transform wild type strain 1610 to generate 1610*cha-*. Kanamycin-resistant transformants resulting from double-crossover events at the *cha* locus screened for the ability to adhere to Chang cells.

The *cha2* gene was inserted into the chromosome of non-adherent *H. influenzae* strain Rd. The Rd *hap* locus contains a nonsense mutation at codon 710 that results in a truncated open reading frame, presenting a useful location for insertion of foreign DNA. DNA encoding Cha2 was cloned next to a chloramphenical-resistance marker in the middle of the *hap* pseudogene. Chloramphenical-resistant transformants resulting from double crossover events at the *hap* locus were screened for Cha expression by whole cell

dot immunoblot. Cloning of Cha2 was performed using genomic DNA from wildtype strain 1610. In the absence of a reactive antibody to Cha2, a FLAG-tag [M-DYKDDDDK] was inserted at the 3′ end of the passenger domain coding sequence. The plasmid pXL:hap-Cha1Δ474-1297-Cm^{R} (generated previously, see Chapter 3) encodes Cha from strain 1595 linked to a chloramphenicol resistance marker with 5′-BamHI and 3′-MluI sites engineered on either side of the promoter-signal peptide-passenger domain (P-473). Wildtype strain 1610 contains a very long tract of tandem 84-bp repeats that inhibited direct amplification of the entire *cha2* passenger domain. Unable to identify a zero-repeat variant of strain 1610, primers Cha2F-up (ATCGATGGATCCCCGCTGAAAAATCAGCATTGGGAC) and Cha2R-FLAG(TTGCGCGC*CTTGTCGTCATCGTCTTTGTAGTCCAT*TTGACGTTTAATTCTGGT TTTATCATTTCTAC) were used to amplify Cha2 coding sequence (aa 1-806) plus 383-bp of upstream sequence yielding a product containing an engineered 5′-BamHI site and a terminal FLAG-BssHII sequence motif. The resulting product was digested with BamHI and BssHII and ligated to BamHI/MluI-digested pXL:hap-Cha1Δ474-1297-Cm^{R}, creating pXL:hap-Cha2$_{(P\text{-}806/1004\text{-}1092)}$-Cm^{R}. The Cha2 membrane anchor domain labeled as amino acids 1004-1092 reflects the same sequence previously designated 1298-1386 for Cha1. The *cha2-cam^{R}*-containing *hap* locus was amplified using primers RdhapF (GCGCGTGTACGTATAGGCTCAGGACG) and RdhapR

(GTTGCAGACAACTCCGATAACGCAC), and the resulting DNA amplicon was transformed into *H. influenzae* strain Rd to create Rd/Cha2(1-806/1004-1092).

Quantitative Adherence Assays

Quantitative adherence assays using epithelial cells were performed as described previously (St Geme 1993). Percent adherence was calculated by dividing the number of adherent CFU per epithelial cell monolayer by the number of inoculated CFU. Each strain was examined in triplicate in a given assay, and each assay was performed a minimum of three times. Chang conjunctival cells (Wong-Kilbourne derivative, clone 1-5c-4; ATCC CCL-20.2) were maintained in modified Eagle medium with Earle's salts and supplemented with nonessential amino acids and 10% fetal bovine serum. Detroit 562 (human pharyngeal carcinoma; ATCC CCL-138), HeLa (human cervical epidermoid carcinoma; ATCC CCL-2), and HEC-1-B cells (human endometrium; ATCC HTB-113) were maintained in the same medium supplemented additionally with 1 mM sodium pyruvate.

Protein analysis

Surface localization of Cha protein was assessed qualitatively by whole cell dot immunoblot using guinea pig polyclonal anti-Cha (aa70-473) antiserum diluted 1:5000. Western blot analysis of outer membrane proteins was performed as described previously (Cotter *et al.* 2006). Outer membrane fractions of whole-cell bacterial sonicates were prepared on the basis of Sarkosyl insolubility (Carlone *et al.* 1986) and

were treated with 95% formic acid overnight (St Geme *et al.* 1998). Proteins were resolved by sodium dodecyl sulfate-polyacrylamide gel electrophoresis, and Western blots were probed with mouse monoclonal anti-FLAG M2 antibody diluted 1:500 (Sigma) to detect FLAG-tagged Cha2 protein.

Quantitative aggregation and settling assays

Haemophilus strains were grown overnight on plates and were then resuspended in 3 ml BHI to an OD_{600} of 0.8. Bacterial suspensions were incubated at room temperature for 4 hr during which time the optical density at 600 nm was measured at 30 min intervals.

Nucleotide sequence accession number

The sequence corresponding to the *Haemophilus* cryptic genospecies strain 1610 *cha2* gene has been deposited in the GenBank database and assigned accession number GU171279.

5.3 Results

5.3.1 Identification of Cha2

Previously we reported that the *cha* locus was uniformly present among our collection of clinical isolates of the *Haemophilus* cryptic genospecies (Sheets *et al.* 2008). In an effort to examine the correlation between Cha expression and adherence among these isolates, we performed dot immunoblots of whole bacteria with anti-Cha antiserum (raised against the N-terminal amino acids 70-473 of Cha from strain 1595).

As shown in Figure 19, Cha protein was detected in adherent strains 421, 422, 1595, and 1673 but was not detected in non-adherent strains 420, 427, and 2446. Cha was also not detected in strain 1610, which adheres to Chang cells. To address whether strain 1610 contains a functional *cha* gene, *cha* was disrupted. Interestingly, this mutation abrogated adherence to Chang cells, suggesting that this locus encodes a Cha-like adhesin (Fig. 21A). The gene was successfully cloned in 3 fragments (necessary due to its repetitive nature) to facilitate sequencing of the entire locus. We confirmed that this gene, designated *cha2*, is physically located at the same locus as *cha* from strain 1595 (hereafter *cha1*) by performing PCR amplification using external primers which anneal within the upstream cytochrome oxidase gene and downstream predicted ORF of unknown function (not shown). PCR amplification and sequencing of *cha* coding sequence from each clinical isolate revealed that strain 2446 also contains *cha2,* whereas strains 420, 421, 422, 427, 2452, and 1673 all contain *cha1*, suggesting that the *cha1* allele may be more prevalent than *cha2* (Fig. 19).

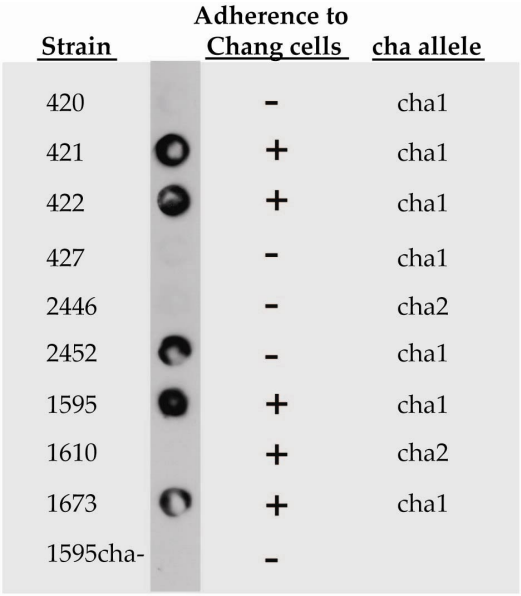

Figure 19: Dot immunoblot analysis of cryptic genospecies clinical isolates probing for expression of Cha.

Whole bacteria from strains 420, 421, 422, 427, 2446, 2452, 1595, 1610, 1673, and a 1595*cha-* mutant were absorbed onto nitrocellulose and probed with anti-Cha antiserum (aa 70-473) diluted 1:5000. Adherence by each strain to Chang epithelial cells is indicated by +/-. Adherence levels below 5% of the inoculum were scored as negative. The *cha* coding sequence from each clinical isolate was sequenced and categorized as cha1/cha2 allele.

5.3.2 Cha1/Cha2 domain comparison

The *cha2* gene from strain 1610 encodes a 1092-residue protein, counting one of the variable 28-amino acids repeats that start at residue 844 (GenBank accession no. GU171279), with a predicted signal peptide of 64 amino acids (Bendtsen *et al.* 2004). Southern analysis of wildtype strain 1610 suggests the presence of ~60 tandem, identical repeats (not shown), consistent with a moderate adherence phenotype and the appearance of surface fibers, previously reported by Rosenau and colleagues (Rosenau *et*

al. 1993). The C-terminal 89 amino acids of Cha2 are 100% identical to the membrane anchor domain of Cha1. This region trimerizes to form a β-barrel pore that is predicted to be occluded by a short coiled-coil motif that connects with the surface exposed passenger domain (Sheets *et al.* 2008). As shown in Figure 20, alignment of Cha1 with Cha2 highlights striking N-terminal sequence divergence and C-terminal conservation between these proteins. The C-terminal 666-residues (including one 28-residue repeat) are 98% identical and are characterized by 2 Ylhead β-roll clusters, 2 neck domains, a coiled-coil stalk region, and the membrane anchor domain. Interestingly, the sequence between the first and second HIM/neck domains of Cha1 appears to be absent in Cha2 (including 1 HIM motif), suggesting a deletion event via homologous recombination at the repetitive HIM sequences. This region in Cha1 is characterized by a large Ylhead repeat domain. Following the predicted signal peptides, residues 70-418 of Cha1 and residues 65-402 of Cha2 are strikingly divergent, sharing only 56% similarity and 35% identity. Highlighted in Figure 20, this region of maximal divergence contains mostly undefined sequence, followed by a short Ylhead β-roll motif and a putative head insert motif (HIN2) predicted using the daTAA domain annotation program (http://toolkit.tuebingen.mpg.de/dataa) (Szczesny and Lupas 2008). Despite sequence conservation with HIN2 domains from other TAAs, it should be noted that the defining FxG motif is absent and therefore the HIN2 domain is predicted with low confidence

(and it is not predicted for Cha1). The structural significance of this motif has not been reported.

```
Cha1       MNKIFKIIWNKTTQRLEVVSELAKSQGKATVSTDKRAEVTSAVDFKKSTLFRAISLALGLSAISLPAG-AASFTDNYTEG
Cha2       MNKIFKIIWNKTTQRLEVVSELAKSQGKATVSTDKRSPISFTPKFKLSLIS---SLIIGITSIPVIAATPAEWADNIRW-
Consensus  ************************************: :: : .** * :    ** :*:::*.: *. .*.::**

Cha1       WSTAWGAGNKYGERISAKDFPIVVGRNNNLAPKNSRAEGNPASATVMGRDNNITSGLAFTVFGDRNQLNTGNVSTVVGIY
Cha2       --------NNYPDNATIKG-AFELGTSNKYGPELSRDKEKAFIVGADNRLNTPASGLSLNVVGRGNLVKYGNRVNIFGDS
Consensus          *:* :. : *. .: :* .*: .*: ** : :.  . . .* *. :***::.*.*  * :: **  .:.*

Cha1       NRITG-GKTNKYLTVLGNNIYAQNPTESGLYIGNNLEANASPIAIGNNVYSEIGGSVIGHNASVRARHANGGGFLTQDSG
Cha2       NNIYGDNGVHGDLLVLGTNNKLENPTNYGLVIGHKNNLNNSPFGIGNELTLSGGGRAIGRNITLDKLKFNGGFRPDDSAG
Consensus  *.* * . .:  * ***.*   :***: ** **:: : * **:.***::  . ** .**:* ::   : ***    :.:*

Cha1       FAVGDNIKGYGNHVFFGRGINLEGGKIANVATADNKKGVYVGQDITSKGAKEGVLLGTNLTSNDNVAHYSVGVGSNINMT
Cha2       LAFGESINAFGNHLVVGRDINVD--KKNNSANTNNPNDKNLARSTVLR---RSVFFGDSITSDG--AINSVFIGRSITSS
Consensus  :*.*:.*:.:***:..**.**::  *  * *.::* :.  :.:. . :   ..*::* .:**:.  *  ** :* .*. :

Cha1       G--NAENATAVGQHITLNG--KKSVAVGHNITVSAEDSVSLGSGSRAAKQGDVVLGAGSATTANTNSNLSGWTVSGTNLT
Cha2       GSESTKNQVGVGSNITFSSGLSNATAIGSHLTVKATDATAIGTRATAENIGDIALGADSTTTGATNT-LSAWNVGGTNLA
Consensus  *  .::* ..**.:**:..  .::.*:* ::**.* *:.::*: : * : **:.***.*:**. **: **.*.*.****:

Cha1       ASYATSVKQTAGKGVVSVGSDSVKRQIKNVATGDVSATSTDAVNGAQLYALSTKIGESSSEYFHVNSNKTGNKGKTATGA
Cha2       NSYNANIKSTN-NGVVSVGSSSVRRQIKNVAAGDVSATSTDAVNGAQL-------------------------------
Consensus   ** :.:*.*  :*******.**:*******:****************

Cha1       TGSNSVAVGPNAKASSESGIAIGDESKAAGSQGALAVGQNANAAHNNATAIGRNSVANGTDSVAVGLNAIASGSSAIAMG
Cha2       ------------------------------------------------------------------------------
Consensus

Cha1       KSANSKNNSTVALGQEANATGIAGTALGYRSQSTGDWTVAAGPDARALKNFAVAVGNNANATEVNAVAVGKSSTAGKDSS
Cha2       ------------------------------------------------------------------------------
Consensus

Cha1       LALGNNATANAASGDIALGASSVTATANNITDFTIGTTKVSDSFVKAKGNQGVVSVGSTTVKRQIQNVGAGDVSAISTDA
Cha2       ------------------------------------------------------------------------------
Consensus

Cha1       INGSQLYHVATALNTLATDAKTAAVAADEKAGAASTAASTAASKADAAQTTANKAKTAADGAAAAVNNSYLHVNSSKTGN
Cha2       ------YHAATALNKLATDAKTTAVAADKKAGAASTAASTAASKADAAQTTANKAKTAADGAAAAVNNSYLHVNSSKTGN
Consensus        **.*****.*******:*****:**************************************************

Cha1       RAALNETANATGENSIALGPNATASGTSSIAIGDEASSTASGTVSIGQNAINRANNSTVIGRNAQVTKDSTDSVAIGLGS
Cha2       RAALNETASATGENSIALGPNATASGASSIAIGDEASSTAAGTVSIGQNAINRANNSTVIGRNAQVTKDSTDSVAIGLGS
Consensus  ********.*****************:*************:***************************************

Cha1       IVTSTSATAIGKNANASGASSISIGQNTLTKNSATIAVGYNITAVANNATAVGSYTNATGEASVALGYKIEAGSWATAVG
Cha2       IVTSTSATAIGKNANASGASSISIGQNTLTKNSATIAVGYNITAVANNATAVGSYTNATGEASVALGYKIEAGSWATAVG
Consensus  ******************************************************************************

Cha1       SNTKATGSTSAAFGNNVIASGESSLALGYQSKATANWAVSVGGGSQANGVNSTSVGRGALAAGEQATAIGAVANAGKDKS
Cha2       SNTKATGSTSAAFGNNVIASGESSLALGYQSKATANWAVSVGGGSQANGVNSTSVGRGALAAGEQATAIGAVANAGKDKS
Consensus  ******************************************************************************

Cha1       LALGRGATANTAEGDVAIGDSSLTAEVNPISELTIGNTKVSDSFQKATKNGVVSVGRNDKTRIKRQIQNVGAGDVSATST
Cha2       LALGRGATANTAEGDVAIGDSSLTAEVNPISELTIGNTKVSDSFQKATKNGVVSVGRNDKTRIKRQIQNVGAGDVSATST
Consensus  ******************************************************************************

Cha1       DAINGSQLYHVATALNTLATDAKTAAAVADSKATVASTAASTADSKATVASTAAAVADSKAEAVKETINRSGLEIDGQNT
Cha2       DAINGSQLYHVATALNTLATDAKTAAAVADSKATVASTAASTADSKATVASTAAAVADSKAEAVKETINRSGLEIDGQNT
Consensus  ******************************************************************************

Cha1       KIGSGSTYAGGSNATAIGGSAKATAENTTAVGQGSEATAKGATAIGQGAKATAENAVAIGTGSVAAEANTVSIGSQGNER
Cha2       KIGSGSTYAGGSNATAIGGSAKATAENTTAVGQGSEATAKGATAIGQGAKATAENAVAIGTGSVAAEANTVSIGSQGNER
Consensus  ******************************************************************************

Cha1       RITNVAAGVNPTDAANVGQVNAVENRINNKLGKFDKKLRSGVASAVAVANIPQVTIPGANMVALGVGNYKGQSAVAVGYS
Cha2       RITNVAAGVNPTDAANVGQVNAVENRINNKLGKFDKKLRSGVASAVAVANIPQVTIPGANMVALGVGNYKGQSAVAVGYS
Consensus  ******************************************************************************

Cha1       RASDSNRVIIKMTAGASTQRDYTFGAGIGYQW
Cha2       RASDSNRVIIKMTAGASTQRDYTFGAGIGYQW
Consensus  ********************************
```

Figure 20: Comparison of the predicted amino acid sequences of Cha1 from strain 1595 and Cha2 from strain 1610.

The protein sequences were aligned using the ClustalW algorithm with slight manual modifications. For clarity, key trimeric autotransporter structural domains, predicted using SignalP analysis and the daTAA annotation server, are highlighted (black: signal peptide, purple: Ylhead repeats, red: HIN2 head insert motif, light blue: neck/HIM domain, green: coiled-coil, navy: membrane anchor). A single (variable) repeat unit is depicted with yellow text. Identical residues are marked with asterisks, very similar amino acids marked by two dots, and slightly similar amino acids denoted by single dots. The underlined sequence highlights the region of greatest divergence (35% identity and 56% similarity).

5.3.3 Cha2 host cell binding and aggregative properties

Interestingly, the N-terminal region of greatest sequence divergence between Cha1 and Cha2 encompasses the adhesive region of Cha1, raising the possibility of differing tissue specificities. To address whether Cha2 demonstrates similar adhesive properties as Cha1 from strain 1595, we examined the levels of adherence of wildtype strain 1610 and mutant strain 1610*cha-* to each of the genital and respiratory cell lines previously used to demonstrate Cha1-adherence (Sheets *et al.* 2008). As shown in figure 21A, the mutant demonstrated negligible adherence levels compared to wildtype strain 1610. We speculate that the moderate (as opposed to high) adherence levels exhibited by strain 1610 reflect the number of tandem repeats (~60). To confirm that Cha2 functions as an adhesin, we sought to express the protein in non-adherent *H. influenzae* strain Rd. Unable to identify a zero-repeat strain 1610 variant, residues 1-806 (preceding the tandem repeats) were linked to the C-terminal membrane anchor domain (aa 1004-1092). In the absence of a reactive Cha2 antibody, we elected to place a FLAG-epitope tag on the Cha2 passenger domain. We verified expression of Cha2 by Western analysis

of formic acid-denatured outer membrane fractions. Shown in Fig. 21C, monoclonal anti-FLAG M2 antibody (Sigma) reacted specifically with an ~84 kDa protein from strain Rd/Cha2(1-806/1004-1092). In quantitative adherence assays, strain Rd/Cha2(1-806/1004-1092) was capable of significant adherence to Chang, Detroit 562, HeLa, and HEC-1-B cells unlike strain Rd alone (Fig. 21B). The significance of the reduced levels of adherence to Detroit 562 and HEC-1-B cells compared with Chang and HeLa cells is uncertain, but may reflect a lower concentration of Cha2 receptor. However, overall both Cha1 and Cha2 appear to mediate attachment to the same cell types. While we have not observed differences in cell line specificity, comparison of previously reported strain 1595 Cha1-mediated adherence with strain 1610 Cha2-mediated adherence suggests the variant Cha2 protein may represent a lower affinity adhesin.

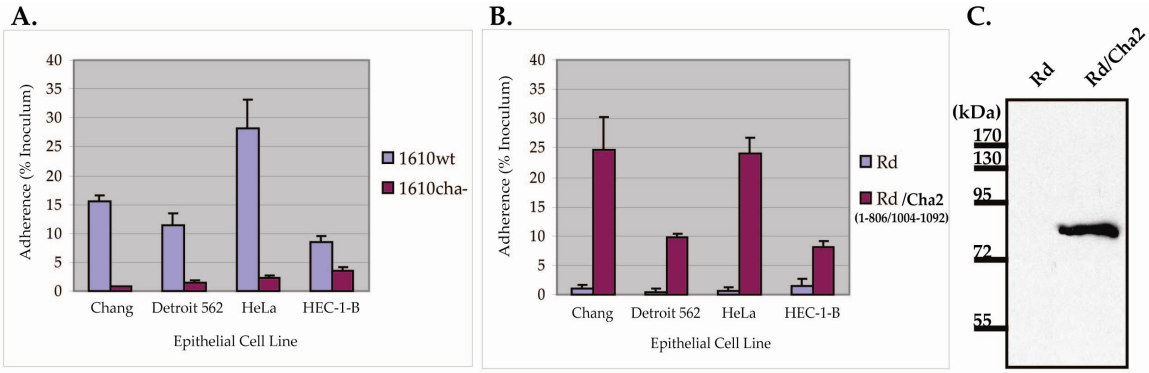

Figure 21: Cha2-mediated adherence to Chang, Detroit 562, HeLa, and HEC-1-B epithelial cells.

(A,B) Adherence by *Haemophilus* cryptic genospecies wild type strain 1610, 1610*cha*- mutant, *H. influenzae* strain Rd, and Rd/Cha2(1-806/1004-1092) was measured after incubating bacteria with epithelial cell monolayers for 30 min and was calculated by dividing the number of adherent colony-forming units per epithelial monolayer by the number of inoculated colony-forming units. Values represent mean (± SE) of measurements in triplicate from representative experiments. (C) Western analysis of *H. influenzae* strain Rd and Rd/Cha2(1-806/1004-1092). Outer membrane fractions were formic acid denatured and probed with monoclonal anti-FLAG M2 antibody diluted 1:500 (Sigma).

In previous studies, we demonstrated that the Cha1 protein from strain 1595 promotes rapid bacterial aggregation through homo-oligomeric Cha-Cha interactions at the N-terminus and that this activity is inhibited by tandem, repeat expansion (Sheets and St. Geme, unpublished). We tested the aggregative activity of Cha2 by performing settling assays in which bacteria were resuspended in broth culture and rates of bacterial aggregation were quantified by measuring culture absorbance at 600 nm as aggregates settled to the bottoms of the culture tubes. Wildtype strain 1610 remained in suspension beyond 24 h (not shown), consistent with the presence of ~60 tandem repeats and similar

to high-repeat 1595 strain variants (Figure 12). However, a suspension of *H. influenzae* strain Rd expressing Cha2 protein lost turbidity as aggregates settled within 4 h, whereas a suspension of Rd alone remained turbid (Figure 22). This result demonstrates that Cha2 mediates bacterial flocculation at a rate comparable to that of Cha1 (Figure 8) and also indicates that the N-terminal 806 residues of Cha2 are sufficient for aggregation. In ongoing work, we aim to further localize the aggregative domain of Cha2.

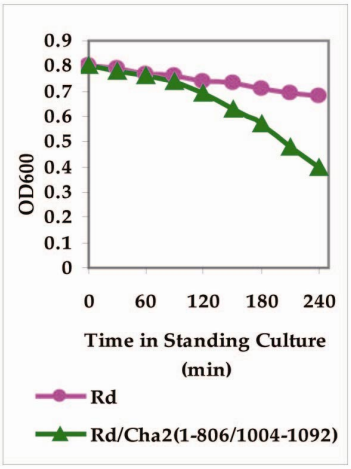

Figure 22: Cha2-mediated bacterial aggregation.

Settling assay with *H. influenzae strain* Rd expressing Cha2. This in-frame deletion derivative contains the N-terminal signal peptide and residues 65-806 of the passenger domain linked to the C-terminal membrane anchor (aa 1004-1092). Bacteria were resuspended in BHI to an optical density of 0.8 at 600 nm. Bacterial aggregation and settling was evaluated by measuring the absorbance every 30 min for 4 hours.

5.4 Discussion

The *Haemophilus* cryptic genospecies Cha protein is a large trimeric autotransporter that mediates its own insertion, trimerization, and subsequent surface localization of an adhesive extracellular domain (Sheets *et al.* 2008). Like other trimeric autotransporters, the surface-exposed Cha passenger domain exhibits modular architecture consisting of repetitive YadA-like head, neck, and stalk structural domains. In earlier work, we demonstrated that the Cha protein promotes intimate attachment to human epithelial cells and facilitates bacterial aggregation. Presentation of N-terminal residues 70-473 is sufficient to confer adhesive and aggregative activities to bacteria; however Cha-mediated adherence is progressively inhibited by expansion of identical tandem repeats located at residue 1138. Since all characterization studies have focused on Cha protein expressed by prototypic strain 1595, we have not previously determined whether the described sequence modules are subject to mutational variation.

In this study, examination of our cryptic genospecies isolates for Cha expression led us to the recognition of a smaller, variant Cha2 protein. The anti-Cha antiserum raised against an N-terminal fragment of Cha protein strain 1595 was unable to recognize Cha2 epitopes in strain 1610, consistent with N-terminal sequence divergence between these strains. Genetic analysis indicated that strains 1610 and 2446 contain the variant *cha2* gene and that the *cha1* and *cha2* genes encode distinct proteins sharing certain epitopes. The divergent *cha2* gene appears to be present at the same locus as *cha1*

in strain 1595, consistent with previous data indicating that only one copy of the *cha* gene occurs in cryptic genospecies strains (Chapter 4). Cryptic genospecies strains appear to contain either of these alleles, but not both (unlike the *Moraxella catarrhalis* UspA proteins (Brooks *et al.* 2008)). While among our collection of isolates, we only found either the *cha1* or *cha2* genes, genetic analysis of additional cryptic genospecies isolates might reveal the existence of additional *cha* alleles.

Domain comparison revealed that, like Cha1, the Cha2 protein contains identical clusters of Ylhead degenerate repeats, HIM motifs and coiled-coil domains, which are situated proximal to the cell surface and linked to the membrane-spanning translocator domain. Furthermore, although the number of tandem 28-aa repeats varies, the amino acid sequence identity throughout the C-terminal region is maintained at the nucleotide level. Interestingly, Cha1 residues 443-696 encompassing the second Ylhead cluster are absent in Cha2, likely the result of recombination and deletion of the sequence between the first and second HIM motifs. However, the functional significance of this deletion event is unclear considering that the adhesive domain of Cha1 lies distal to this Ylhead group. Examination of the N-terminal Cha2 sequence revealed striking sequence divergence from the N-terminal adherence-mediating module of Cha1 (Chapter 3). While a small Ylhead domain is maintained, the majority of this region is distinguished by unique sequence, which is only 56% similar and 35% identical to the corresponding Cha1 sequence.

Due to the presence of extensive *cha* repeats in wildtype strain 1610 template, we were restricted in cloning and expressing full-length Cha2 protein in *H. influenzae* strain Rd for functional assays. The Rd/Cha2 strain included only amino acids 1-806 of Cha2 linked in-frame to the membrane anchor domain. Nevertheless, we successfully demonstrated that this fragment of the passenger domain (aa 65-806) is sufficient for adherence to epithelial cells and bacterial aggregation (similar to what we observed with Cha1). In ongoing work, we are working to further localize the aggregative and host cell binding domains of Cha2. Taking into account the absence of the second Ylhead cluster of Cha1, and the adhesive capability of Rd/Cha2(1-806/1004-1092), we predict that the adhesive domains will localize to the distal tip of the Cha2 molecule. Although we were not able to recover a zero-repeat variant in this study, we predict that the tandem repeats vary in copy number and play a similar role in altering the adhesive activities of the Cha2 protein as they do with Cha1.

At this point, it does not appear that the absence of the Ylhead cluster and N-terminal sequence divergence confers different tissue specificities or binding phenotypes on the Cha1 and Cha2 proteins, as both conferred adherence to Chang, Detroit 562, HeLa and HEC-1-B cells. Conservation of the *cha* genes among clinical isolates and Cha1 and Cha2 adhesive and aggregative functions underscores the importance of the role of the Cha proteins in *Haemophilus* cryptic genospecies pathogenesis. Additional or divergent functions attributed to the novel head structure of the Cha2 protein remain to

be determined in future studies. Little information is available as to the identity of the receptor bound by Cha. (Preliminary evidence from unpublished studies indicates the binding partner is not proteinaceous, and Clemans *et al.* reported that the pilus receptor analog GM-1 does not inhibit binding of strain 1595 to 16HBE14o- (bronchial), ME180 (cervical), HEp-2 (laryngeal), or HeLa (cervical) cells (Clemans *et al.* 2000).) Nevertheless, we hypothesize that variation in the N-terminal domain could impact the specificity or affinity of adherence and enable the organism to adapt to different host environments such as the urogenital tract or respiratory tract, where distinctive receptors may be required for optimal colonization.

Chapter 6. Conclusions, Additional Observations, and Future Directions

6.1 Biological function of Cha

In the past few decades, recognition of *Haemophilus* urogenital and neonatal infections has risen significantly (Wallace *et al.* 1983; Campognone and Singer 1986; Friesen and Cho 1986; Quentin *et al.* 1989) and led to the identification of a genetically divergent and novel cryptic genospecies. A significant increase in recovery of these strains associated with maternal/infant infections highlighted an apparent adaptation of these strains for the genital tract and led to their distinction as a maternal, genital, and neonatal pathogen (Albritton *et al.* 1982; Wallace *et al.* 1983; Friesen and Cho 1986; Martel *et al.* 1989; Rosenau *et al.* 1993). The work presented here identifies Cha as the major adhesin expressed by the *Haemophilus* cryptic genospecies and reveals that Cha expression is essential for cryptic genospecies adherence to human epithelial cells. We have shown that the *cha* gene is uniformly present among cryptic genospecies isolates and is universally lacking among *H. influenzae* strains, suggesting that it may be unique to the cryptic genospecies. The Cha adhesin is composed of an N-terminal signal peptide, an internal passenger domain that harbors adhesive activity, and a C-terminal translocator domain and was classified as a trimeric autotransporter adhesin on the basis of domain homology and functional assays. We demonstrated that the C-terminal domain of Cha targets the protein to the bacterial outer membrane in a trimeric form and that both adherence and aggregation are mediated by the N-terminal 473 amino

acids of the passenger domain. Perhaps most interestingly, we found that expansion of internal coding repeats within the passenger domain gradually modulates the adhesive and aggregative activities of the Cha protein. At the start of this project, little was known about the unique pathogenic determinants of the cryptic genospecies. Considering the importance of adherence and colonization of epithelial cell surfaces in the establishment of disease, the Cha protein represents the first characterized potential virulence factor of this organism. Our findings provide insight into the apparent adaptation of this novel species to the genital tract and leave us with new, interesting, and challenging questions to answer. The following discussions will reflect on what we have discovered in our characterization of the Cha1 and Cha2 adhesins and considerations of future directions of this project.

6.2 Trimeric autotransporters are multifunctional adhesins

***Structural domain organization* (Cha1)**

Ranging from 1386 to ~4200 amino acids, Cha is an incredibly large trimeric autotransporter. Full expansion of the tandem repeat domain (the largest we have witnessed being ~100 units) yields an impressive 389 kDa protein, which is comparable to BadA from *Bartonella henselae,* perhaps the best known large TAA to date (Szczesny *et al.* 2008). Like most trimeric autotransporters, the Cha proteins are distinguished by a unique modular arrangement of highly repetitive architecture. The basic structure of the Cha proteins can be inferred from structural studies of other autotransporter adhesins,

namely YadA (Koretke *et al.* 2006) and Hia (Meng *et al.* 2006), due to their domain and sequence similarities. Koretke *et al.* generated a model structure of the entire YadA protein by combining the crystal structure of the N-terminal β-roll head domain with predicted structures of the coiled-coil stalk and C-terminal translocation domain (Nummelin *et al.* 2004; Cotter *et al.* 2005). Szczesny *et al.* constructed a model structure of the entire BadA protein by similar means after crystallizing the unique BadA head domain (which contains a Ylhead β-roll, neck, Trp-ring, and GIN domain) (Szczesny *et al.* 2008). Analogous to YadA and BadA, the Cha proteins contain a predicted stalk domain that is connected to an N-terminal head motif, extending it away from the bacterial cell surface. However, unlike the simple arrangement exemplified by YadA, both Cha proteins are characterized by a more complicated arrangement of alternating YadA-like head and stalk domains yielding a predicted head-stalk-head-stalk-head-anchor domain arrangement. Additionally, there are significant size differences between each head domain, resulting from differences in the number of degenerate [NSVAIGXXS] repeats, ranging from clusters of 4 to a cluster of 18.

Studies with the YadA and Hia trimeric autotransporters report that trimerization of not only the membrane anchor but of the passenger domain is necessary for adhesive function of these proteins (Cotter *et al.* 2006). The same appears to be true of Cha. In unpublished experiments, we found that co-expression of Cha proteins with different numbers of tandem repeats had a dominant negative effect on adherence,

presumably a consequence of the unequally-sized coiled-coil stalk domains disrupting trimerization of the passenger domain. However, interestingly the requirement for trimerization of the structurally undefined sequence at the N-terminus remains an open question worth addressing considering that this region of the protein represents part of the adhesive-module of Cha. In future studies, this question can be tested by evaluating the adherence phenotype of bacteria co-expressing full-length and an in-frame N-terminal deletion mutant of Cha. Cotter *et al.* have speculated that stable trimerization is essential for native folding, stability, and full-level adhesive activity among TAAs and that therapeutics designed to disrupt the trimeric architecture of trimeric autotransporters may be an effective strategy to eliminate adhesive activity (Cotter *et al.* 2006).

Bacterial Aggregation

Beyond mediating adherence to host cells, we have shown that Cha can function as an avid bacterial aggregator much like the *H. influenzae* Hap adhesin (Hendrixson and St Geme 1998; Fink *et al.* 2003). We demonstrated that expression of Cha in *H. influenzae* strain Rd (which does not express Hap) results in rapid aggregation and settling of bacterial suspensions. We further determined that Cha is self-associating and the autoaggregative domain localizes to the same N-terminal fragment of the passenger domain as the adhesive domain (amino acids 70-473). Interestingly, however, expansion of the 28-residue identical tandem repeats modulates this activity and actually inhibits

intimate bacterial interactions mediated by another unidentified surface structure. We speculate that aggregation and microcolony formation of these organisms enhances bacterial colonization and survival in the vaginal tract and that repeat variation represents a unique mechanism by which bacteria disseminate and potentially become more invasive. Another interesting point to consider is whether Cha can mediate heterologous interactions with other proteins, similar to the *E. coli* self-associating autotransporters, AIDA, TibA, and Ag43 (Klemm *et al.* 2006), whether it be to the other cryptic genospecies aggregative surface structure or to proteins expressed by other bacterial species. It is intriguing to consider that Cha may mediate aggregation to other commensal bacterial species within the vaginal tract which may facilitate persistence or aid mixed biofilm formation.

Binding to extracellular matrix components

At this time, we have only determined functional significance of the N-terminal domains (adhesive) and the variable repeat/stalk domain (anti-adhesive); however, the significance or functional role of the additional head domains throughout the passenger domain is not known. Perhaps they merely function to extend the adhesive N-terminus away from the bacterial cell surface, making it more accessible to the target receptor. Alternatively, the other domains within the passenger domain confer additional activities on the Cha adhesin. A common feature among trimeric autotransporter adhesins is an affinity for extracellular matrix components. In particular, numerous

TAAs with Ylhead repeats mediate adherence to collagen IV (Nummelin *et al.* 2004; Zhang *et al.* 2004; Kaiser *et al.* 2008; Yu *et al.* 2008); however, in unpublished experiments we have not observed binding of strain 1595 to collagen IV, laminin, or fibronectin. In ongoing work, we are examining whether Cha has adhesive activity beyond adherence to epithelial cells and autoaggregation, perhaps mediating binding to fibrinogen, entactin, heparin sulfate, vitronectin, and other types of collagen, as observed with other TAAs (Leduc *et al.* 2009).

Immune Evasion

Four out of five of our neonatal bloodstream isolates exhibit Cha surface fibers, which is surprising considering that Cha likely would be a target of antibody-mediated immunity. It is tempting to speculate that expansion of the repeats and continued expression of Cha fibers in the bloodstream confers some sort of survival advantage to the organism. In pilot experiments, we have found that prototype strain 1595 is considerably resistant to the bactericidal activity of normal human serum. However, mutant strain 1595*cha-* is similarly resistant, indicating that Cha expression is not essential for this activity. Expression of Cha fibers in several bloodstream isolates suggests that the protein plays some as yet unrecognized role or has a means by which to evade immune recognition.

Cha variants

Ubiquitous conservation of the *cha* locus among cryptic genospecies strains demonstrated by Southern analysis indicates that each isolate has the capacity to express the Cha protein and underscores the importance of this unique adhesin. Interestingly, in probing our adherent strains for expression of Cha, we encountered an antigenic variant of the adhesin (designated Cha2) expressed by strain 1610. Specifically, the N-terminal residues 70-418 of Cha1 and residues 65-402 of Cha2 are only 56% similar and 35% identical, while the C-terminal region is virtually identical. This observation led us to examine the remaining 7 isolates among our strain collection by sequence analysis, and compare them to Cha1 from strain 1595 and Cha2 from strain 1610. This analysis revealed that strain 2446 also encodes *cha2*, whereas strains 420, 421, 422, 427, 2452, and 1673 all encode *cha1*, suggesting that the *cha1* allele may be more prevalent than *cha2.* Beyond N-terminal divergence, we also found that Cha2 is distinguished by the absence of a large Ylhead domain between the first and second HIM motifs, suggesting that the variant *cha2* gene was derived from *cha1* after extensive sequence mutations in combination with an apparent recombination/deletion event. Recognition of the divergent Cha2 allele raises the possibility of additional Cha variants expressed by cryptic genospecies isolates. Future studies should include a thorough evaluation of the potential diversity among Cha proteins expressed by clinical isolates, aiming to identify variant sequences, domain arrangements, and the functional implications of such.

Studies with *Moraxella caterrhalis* UspA proteins have demonstrated that allelic variants with different modular arrangements are associated with divergent protein functions, including receptor specificity, fibronectin-binding, and serum resistance (Brooks *et al.* 2008) Considering that both Cha1 and Cha2 are expressed by neonatal blood isolates, at this time it does not appear that one allele confers more invasiveness than the other. However, in future studies evaluating Cha sequence and expression profiles among additional clinical isolates, an interesting question to consider is whether there is conservation of domain architecture and sequences among more virulent/invasive strains and divergence among commensal organisms.

6.3 Relationship between repeats, aggregation, and adherence

The aim of this work was to identify the major determinant of adherence by the *Haemophilus* cryptic genospecies. Upon identification of the Cha trimeric autotransporter, we demonstrated its ability to function as both an adhesin and an aggregator and began to elucidate the unique domain organization on a functional level. Expression of Cha1 protein containing in-frame deletions throughout the passenger domain in non-adherent *H. influenzae* strain Rd enabled us to localize both the host cell binding domain and autoaggregative domain to amino acids 70-473. This region is distinguished by extensive unique, structurally undefined sequence followed by a small Ylhead domain and HIM neck motif which likely functions to connect the N-terminal head domain to the membrane anchor in a structurally stable way. We observed that 84-

nucleotide, identical, tandem repeats located towards the C-terminus of the passenger domain spontaneously vary in number and that this variation does not have an impact on transcription of *cha* or surface localization of Cha protein. It does, however, have a clear impact on the adhesive and aggregative activities of Cha. In quantitative analyses, strain 1595 variants with ~ 76 and ~95 Cha repeats were unable to adhere to Chang cells. Variants lacking the tandem repeats adhered most dramatically. The impact of repeat expansion interfering with adherence was gradual and mirrored the inhibitory effect on aggregation. With one exception, variant strains that did not aggregate did not adhere to epithelial cells and vice versa. Variant strain 1595-J (~53 repeats) did not aggregate efficiently but was adherent to Chang cells, suggesting that the repeat copy number threshold for interfering with self-association is lower than that for adherence. Considering the predicted coiled-coil stalk structure of this repeat domain and previous reports of a general correlation between bacterial fimbriation and adherent phenotypes (Rosenau *et al.* 1993), we sought to immunogold label Cha protein on the bacterial cell surface. In doing so, we demonstrated that repeat expansion results in the appearance and elongation of thin, peritrichous Cha surface fibers. Considering the co-localization of the adhesive and aggregative domains to the N-terminus of Cha, we proposed two models (which are not mutually exclusive) to explain how repeat expansion modulates the adhesive function of Cha.

1) Variation in the number of 28-amino acid repeats may alter the conformation of the N-terminal binding module.

Hypothetically, changes in tertiary structure induced by increased length (and repeat number) of the Cha protein could expose or occlude conformational epitopes near the host cell binding pocket. These changes could be due to alterations in the angles of peptide bonds or to an interaction of the repeat region with other N-terminal domains. While the precise mechanism by which a repetitive coiled-coil stalk domain could modulate distal conformational epitopes is somewhat perplexing, further localization of the host cell binding domain in addition to structural modeling or crystallization of the minimal binding fragment will be necessary to begin to evaluate conformational changes associated with repeat expansion. Generation of monoclonal antibodies that bind epitopes near or within the host cell binding pocket can then be evaluated for differing affinities to Cha repeat variant proteins. The Alpha C protein expressed by Group B streptococci (GBS) is similarly characterized by identical tandem repeats, each 246 bp in length, with a single partial repeat of 33 bp that spontaneously varies during infection (Michel *et al.* 1992; Madoff *et al.* 1996). Interestingly, Gravekamp *et al.* reported that antibodies to 9-repeat antigen predominantly recognized conformational epitopes contained in proteins with higher repeat copy number (9-16) but demonstrated considerably lower binding affinities for epitopes within 1- and 2-repeat alpha C protein (Gravekamp *et al.* 1996). Additional studies reported the occurrence of alpha repeat

deletions during transmission of GBS from mother to neonate and demonstrated that mutants with fewer repeats in their alpha C proteins can escape antibody-mediated immunity (Madoff *et al.* 1996). It is tantalizing to speculate a similar (albeit reciprocal) phenomenon occurring with the Cha adhesin.

2) Repeat expansion, which results in the formation of long, flexible Cha fibers on the bacterial cell surface, may facilitate an interaction between neighboring Cha fibers at the N-terminus, thereby masking the binding domains and inhibiting adherence to other bacteria or host epithelial cells.

Examination of strain 1673 by negative staining electron microscopy highlights an apparent interaction of the distal tips of the Cha fibers, consistent with N-terminal localization of the self-associating, aggregative domain. Similar examination of strain 1595 repeat variants in future work will provide additional evidence as to the validity of this hypothesis, in particular if high-repeat Cha fibers exhibit more tip interactions than shorter, low-repeat fibers. Another means by which to test this second hypothesis involves identifying conditions that interfere with self-association of high-repeat Cha fibers and testing the impact of this on adherence levels. Ionic forces play an important role in many protein-protein interactions and self-association of *E. coli* Antigen 43 (Ag43) protein has been shown to be pH-dependent. Ag43-mediated aggregation is optimal at pH 4 and Ag43-Ag43 interactions are abolished at pH 11 (Klemm *et al.* 2004).

Considering the acidic environment of the vaginal tract, it would be intriguing if Cha-mediated bacterial aggregation was similarly enhanced by a low pH. Regardless, identification of environmental conditions (pH, osmolarity, etc.) that disrupt self-association of Cha-fibers would enable testing the significance of these interactions in precluding adherence to host epithelial cells.

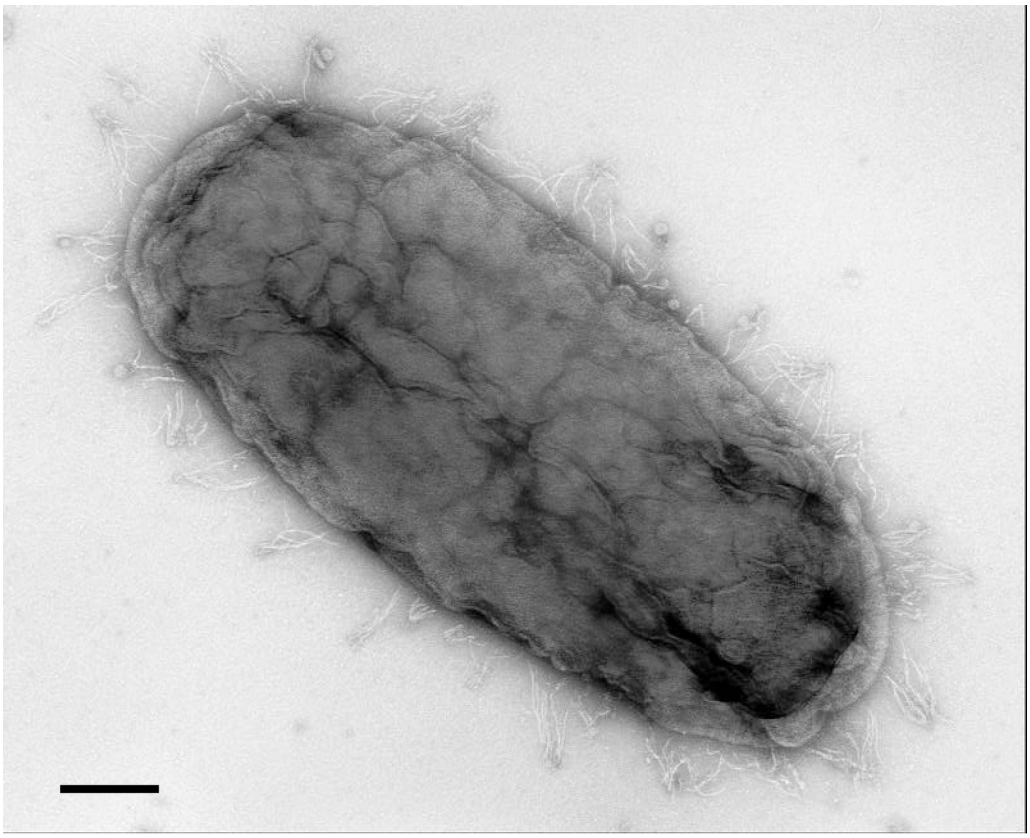

Figure 23: Interaction of Cha fibers. Electron micrograph of *Haemophilus* cryptic genospecies strain 1673 negatively stained with uranyl acetate. (Scale bar = 200 nm)

6.4 Receptor specificity

Recent identification of a smaller, antigenically-divergent Cha2 allele expressed by *Haemophilus* cryptic genospecies strain 1610 was an interesting and unexpected finding in this study. In spite of extensive sequence identity throughout the C-terminal region, a large medial Ylhead cluster present in Cha1 is absent from Cha2, and the N-terminal region of each passenger domain exhibits significant sequence divergence. The divergent region encompasses the adhesive module of Cha1 (and we expect Cha2) and we speculate that this variation may alter affinity or specificity of adherence, although this has yet to be demonstrated. From what we observed, both Cha1 and Cha2 mediate broad-range adherence to human epithelial cells of genital and respiratory origin. Screening adherence properties with additional genital cell lines may uncover differences in tissue specificity. A pilot experiment designed to test whether Cha1 and Cha2 recognize the same host cell receptor proved inconclusive. We aimed to test if binding by Cha1 would saturate receptors and inhibit subsequent binding by Cha2. Unfortunately, preincubation of Chang epithelial cell monolayers with Cha1-expressing bacteria enhanced adherence by Cha2-expressing bacteria, a result that we presume was a consequence of the cross-aggregative activity of these proteins. Preincubation of monolayers with purified Cha1 may or may not have the same effect considering the trimeric nature of the protein. Perhaps the best approach for comparing receptor specificities is direct identification of the Cha receptor, a goal that we intend to actively

pursue. In preliminary experiments, we have observed that proteinase K-treatment of fixed epithelial cell monolayers does not abrogate adherence, indicating that the receptor is not proteinaceous and perhaps may be a glycolipid. Although challenging, identification of the Cha-receptor will ultimately enhance our understanding of the apparent tissue tropism of this bacterial species.

References

Ackermann, N., M. Tiller, G. Anding, A. Roggenkamp and J. Heesemann (2008). "Contribution of trimeric autotransporter C-terminal domains of oligomeric coiled-coil adhesin (Oca) family members YadA, UspA1, EibA, and Hia to translocation of the YadA passenger domain and virulence of *Yersinia enterocolitica*." *J Bacteriol* **190**(14): 5031-43.

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**(42): 25142-9.

Albritton, W. L., J. L. Brunton, M. Meier, M. N. Bowman and L. A. Slaney (1982). "*Haemophilus influenzae*: comparison of respiratory tract isolates with genitourinary tract isolates." *J Clin Microbiol* **16**(5): 826-31.

Anthony, B. F., N. F. Concepcion, S. M. Puentes and N. R. Payne (1990). "Nonimmune binding of human immunoglobulin A to type II group B streptococci." *Infect Immun* **58**(6): 1789-95.

Bakaletz, L. O., B. D. Baker, J. A. Jurcisek, A. Harrison, L. A. Novotny, J. E. Bookwalter, R. Mungur and R. S. Munson, Jr. (2005). "Demonstration of Type IV pilus expression and a twitching phenotype by *Haemophilus influenzae*." *Infect Immun* **73**(3): 1635-43.

Barenkamp, S. J. and J. W. St Geme, 3rd (1996). "Identification of a second family of high-molecular-weight adhesion proteins expressed by non-typable *Haemophilus influenzae*." *Mol Microbiol* **19**(6): 1215-23.

Bateman, A., E. Birney, L. Cerruti, R. Durbin, L. Etwiller, S. R. Eddy, S. Griffiths-Jones, K. L. Howe, M. Marshall and E. L. Sonnhammer (2002). "The Pfam protein families database." *Nucleic Acids Res* **30**(1): 276-80.

Bendtsen, J. D., H. Nielsen, G. von Heijne and S. Brunak (2004). "Improved prediction of signal peptides: SignalP 3.0." *J Mol Biol* **340**(4): 783-95.

Bernstein, H. D. (2007). "Are bacterial 'autotransporters' really transporters?" *Trends Microbiol* **15**(10): 441-7.

Bookwalter, J. E., J. A. Jurcisek, S. D. Gray-Owen, S. Fernandez, G. McGillivary and L. O. Bakaletz (2008). "A carcinoembryonic antigen-related cell adhesion molecule 1

homologue plays a pivotal role in nontypeable *Haemophilus influenzae* colonization of the chinchilla nasopharynx via the outer membrane protein P5-homologous adhesin." *Infect Immun* **76**(1): 48-55.

Brooks, M. J., J. L. Sedillo, N. Wagner, C. A. Laurence, W. Wang, A. S. Attia, E. J. Hansen and S. D. Gray-Owen (2008). "Modular arrangement of allelic variants explains the divergence in *Moraxella* catarrhalis UspA protein function." *Infect Immun* **76**(11): 5330-40.

Bruant, G., N. Gousset, R. Quentin and A. Rosenau (2002). "Fimbrial ghf gene cluster of genital strains of *Haemophilus* spp." *Infect Immun* **70**(10): 5438-45.

Bullard, B., S. Lipski and E. R. Lafontaine (2007). "Regions important for the adhesin activity of *Moraxella catarrhalis* Hag." *BMC Microbiol* **7**: 65.

Bullard, B., S. L. Lipski and E. R. Lafontaine (2005). "Hag directly mediates the adherence of *Moraxella catarrhalis* to human middle ear cells." *Infect Immun* **73**(8): 5127-36.

Campognone, P. and D. B. Singer (1986). "Neonatal sepsis due to nontypable *Haemophilus influenzae*." *Am J Dis Child* **140**(2): 117-21.

Caparon, M. G., D. S. Stephens, A. Olsen and J. R. Scott (1991). "Role of M protein in adherence of group A streptococci." *Infect Immun* **59**(5): 1811-7.

Carlone, G. M., M. L. Thomas, H. S. Rumschlag and F. O. Sottnek (1986). "Rapid microprocedure for isolating detergent-insoluble outer membrane proteins from *Haemophilus* species." *J Clin Microbiol* **24**(3): 330-2.

Clemans, D. L., R. J. Bauer, J. A. Hanson, M. V. Hobbs, J. W. St Geme, 3rd, C. F. Marrs and J. R. Gilsdorf (2000). "Induction of proinflammatory cytokines from human respiratory epithelial cells after stimulation by nontypeable *Haemophilus influenzae*." *Infect Immun* **68**(8): 4430-40.

Clemans, D. L., C. F. Marrs, R. J. Bauer, M. Patel and J. R. Gilsdorf (2001). "Analysis of pilus adhesins from *Haemophilus influenzae* biotype IV strains." *Infect Immun* **69**(11): 7010-9.

Comanducci, M., S. Bambini, B. Brunelli, J. Adu-Bobie, B. Arico, B. Capecchi, M. M. Giuliani, V. Masignani, L. Santini, S. Savino, D. M. Granoff, D. A. Caugant, M. Pizza, R. Rappuoli and M. Mora (2002). "NadA, a novel vaccine candidate of *Neisseria meningitidis*." *J Exp Med* **195**(11): 1445-54.

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**(10): 2644-53.

Cotter, S. E., N. K. Surana, S. Grass and J. W. St Geme, 3rd (2006). "Trimeric autotransporters require trimerization of the passenger domain for stability and adhesive activity." *J Bacteriol* **188**(15): 5400-7.

Cotter, S. E., N. K. Surana and J. W. St Geme, 3rd (2005). "Trimeric autotransporters: a distinct subfamily of autotransporter proteins." *Trends Microbiol* **13**(5): 199-205.

Cotter, S. E., H. J. Yeo, T. Juehne and J. W. St Geme, 3rd (2005). "Architecture and adhesive activity of the *Haemophilus influenzae* Hsf adhesin." *J Bacteriol* **187**(13): 4656-64.

Dawid, S., S. J. Barenkamp and J. W. St Geme, 3rd (1999). "Variation in expression of the *Haemophilus influenzae* HMW adhesins: a prokaryotic system reminiscent of eukaryotes." *Proc Natl Acad Sci U S A* **96**(3): 1077-82.

De Bolle, X., C. D. Bayliss, D. Field, T. van de Ven, N. J. Saunders, D. W. Hood and E. R. Moxon (2000). "The length of a tetranucleotide repeat tract in *Haemophilus influenzae* determines the phase variation rate of a gene with homology to type III DNA methyltransferases." *Mol Microbiol* **35**(1): 211-22.

Dower, W. J., J. F. Miller and C. W. Ragsdale (1988). "High efficiency transformation of *E. coli* by high voltage electroporation." *Nucleic Acids Res* **16**(13): 6127-45.

El Tahir, Y. and M. Skurnik (2001). "YadA, the multifaceted *Yersinia* adhesin." *Int J Med Microbiol* **291**(3): 209-18.

Elkins, C., K. J. Morrow, Jr. and B. Olsen (2000). "Serum resistance in *Haemophilus ducreyi* requires outer membrane protein DsrA." *Infect Immun* **68**(3): 1608-19.

Farley, M. M., D. S. Stephens, 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 mucosa." *J Infect Dis* **161**(2): 274-80.

Farrant, J. L., A. Sansone, J. R. Canvin, M. J. Pallen, P. R. Langford, T. S. Wallis, G. Dougan and J. S. Kroll (1997). "Bacterial copper- and zinc-cofactored superoxide

dismutase contributes to the pathogenesis of systemic salmonellosis." *Mol Microbiol* **25**(4): 785-96.

Fink, D. L., A. Z. Buscher, B. Green, P. Fernsten and J. W. St Geme, 3rd (2003). "The *Haemophilus influenzae* Hap autotransporter mediates microcolony formation and adherence to epithelial cells and extracellular matrix via binding regions in the C-terminal end of the passenger domain." *Cell Microbiol* **5**(3): 175-86.

Fink, D. L., B. A. Green and J. W. St Geme, 3rd (2002). "The *Haemophilus influenzae* Hap autotransporter binds to fibronectin, laminin, and collagen IV." *Infect Immun* **70**(9): 4902-7.

Fink, D. L. and J. W. St Geme, 3rd (2003). "Chromosomal expression of the *Haemophilus influenzae* Hap autotransporter allows fine-tuned regulation of adhesive potential via inhibition of intermolecular autoproteolysis." *J Bacteriol* **185**(5): 1608-15.

Forbes, K. J., K. D. Bruce, A. Ball and T. H. Pennington (1992). "Variation in length and sequence of porin (ompP2) alleles of non-capsulate *Haemophilus influenzae*." *Mol Microbiol* **6**(15): 2107-12.

Frick, I. M., M. Morgelin and L. Bjorck (2000). "Virulent aggregates of *Streptococcus pyogenes* are generated by homophilic protein-protein interactions." *Mol Microbiol* **37**(5): 1232-47.

Friesen, C. A. and C. T. Cho (1986). "Characteristic features of neonatal sepsis due to *Haemophilus influenzae*." *Rev Infect Dis* **8**(5): 777-80.

Geluk, F., P. P. Eijk, S. M. van Ham, H. M. Jansen and L. van Alphen (1998). "The fimbria gene cluster of nonencapsulated *Haemophilus influenzae*." *Infect Immun* **66**(2): 406-17.

Geme, J. W., 3rd and D. Cutter (1995). "Evidence that surface fibrils expressed by *Haemophilus influenzae* type b promote attachment to human epithelial cells." *Mol Microbiol* **15**(1): 77-85.

Gill, M. V., P. E. Schoch, J. M. Musser and B. A. Cunha (1995). "Bacteremia and chorioamnionitis due to cryptic genospecies of *Haemophilus Influenzae* biotype I." *Eur J Clin Microbiol Infect Dis* **14**(12): 1088-90.

Gousset, N., A. Rosenau, P. Y. Sizaret and R. Quentin (1999). "Nucleotide sequences of genes coding for fimbrial proteins in a cryptic genospecies of *Haemophilus* spp. isolated from neonatal and genital tract infections." *Infect Immun* **67**(1): 8-15.

Gravekamp, C., D. S. Horensky, J. L. Michel and L. C. Madoff (1996). "Variation in repeat number within the alpha C protein of group B streptococci alters antigenicity and protective epitopes." *Infect Immun* **64**(9): 3576-83.

Gravekamp, C., D. L. Kasper, J. L. Michel, D. E. Kling, V. Carey and L. C. Madoff (1997). "Immunogenicity and protective efficacy of the alpha C protein of group B streptococci are inversely related to the number of repeats." *Infect Immun* **65**(12): 5216-21.

Gravekamp, C., B. Rosner and L. C. Madoff (1998). "Deletion of repeats in the alpha C protein enhances the pathogenicity of group B streptococci in immune mice." *Infect Immun* **66**(9): 4347-54.

Guerina, N. G., S. Langermann, H. W. Clegg, T. W. Kessler, D. A. Goldman and J. R. Gilsdorf (1982). "Adherence of piliated *Haemophilus influenzae* type b to human oropharyngeal cells." *J Infect Dis* **146**(4): 564.

Hallstrom, T., P. F. Zipfel, A. M. Blom, N. Lauer, A. Forsgren and K. Riesbeck (2008). "*Haemophilus influenzae* interacts with the human complement inhibitor factor H." *J Immunol* **181**(1): 537-45.

Henderson, I. R., F. Navarro-Garcia and J. P. Nataro (1998). "The great escape: structure and function of the autotransporter proteins." *Trends Microbiol* **6**(9): 370-8.

Hendrixson, D. R., B. J. Akerley and V. J. DiRita (2001). "Transposon mutagenesis of *Campylobacter jejuni* identifies a bipartite energy taxis system required for motility." *Mol Microbiol* **40**(1): 214-24.

Hendrixson, D. R. and J. W. St Geme, 3rd (1998). "The *Haemophilus influenzae* Hap serine protease promotes adherence and microcolony formation, potentiated by a soluble host protein." *Mol Cell* **2**(6): 841-50.

Herriott, R. M., E. M. Meyer and M. Vogt (1970). "Defined nongrowth media for stage II development of competence in *Haemophilus influenzae*." *J Bacteriol* **101**(2): 517-24.

Herriott, R. M., E. Y. Meyer, M. Vogt and M. Modan (1970). "Defined medium for growth of *Haemophilus influenzae*." *J Bacteriol* **101**(2): 513-6.

Hill, D. J., M. A. Toleman, D. J. Evans, S. Villullas, L. Van Alphen and M. Virji (2001). "The variable P5 proteins of typeable and non-typeable *Haemophilus influenzae* target human CEACAM1." *Mol Microbiol* **39**(4): 850-62.

Hoiczyk, E., A. Roggenkamp, M. Reichenbecher, A. Lupas and J. Heesemann (2000). "Structure and sequence analysis of *Yersinia* YadA and *Moraxella* UspAs reveal a novel class of adhesins." *EMBO J* **19**(22): 5989-99.

Holland, J., P. R. Langford, K. J. Towner and P. Williams (1992). "Evidence for *in vivo* expression of transferrin-binding proteins in *Haemophilus influenzae* type b." *Infect Immun* **60**(7): 2986-91.

Hornef, M. W., M. J. Wick, M. Rhen and S. Normark (2002). "Bacterial strategies for overcoming host innate and adaptive immune responses." *Nat Immunol* **3**(11): 1033-40.

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**(8): 3134-41.

Jordan, P., L. A. Snyder and N. J. Saunders (2003). "Diversity in coding tandem repeats in related *Neisseria* spp." *BMC Microbiol* **3**: 23.

Kaiser, P. O., T. Riess, C. L. Wagner, D. Linke, A. N. Lupas, H. Schwarz, G. Raddatz, A. Schafer and V. A. Kempf (2008). "The head of *Bartonella* adhesin A is crucial for host cell interaction of *Bartonella henselae*." *Cell Microbiol* **10**(11): 2223-34.

Kehl-Fie, T. E. and J. W. St Geme, 3rd (2007). "Identification and characterization of an RTX toxin in the emerging pathogen *Kingella kingae*." *J Bacteriol* **189**(2): 430-6.

Khan, A. G., S. R. Shouldice, S. D. Kirby, R. H. Yu, L. W. Tari and A. B. Schryvers (2007). "High-affinity binding by the periplasmic iron-binding protein from *Haemophilus influenzae* is required for acquiring iron from transferrin." *Biochem J* **404**(2): 217-25.

Klemm, P., L. Hjerrild, M. Gjermansen and M. A. Schembri (2004). "Structure-function analysis of the self-recognizing Antigen 43 autotransporter protein from *Escherichia coli*." *Mol Microbiol* **51**(1): 283-96.

Klemm, P., R. M. Vejborg and O. Sherlock (2006). "Self-associating autotransporters, SAATs: functional and structural similarities." *Int J Med Microbiol* **296**(4-5): 187-95.

Koretke, K. K., P. Szczesny, M. Gruber and A. N. Lupas (2006). "Model structure of the prototypical non-fimbrial adhesin YadA of *Yersinia enterocolitica*." *J Struct Biol* **155**(2): 154-61.

Krasan, G. P., D. Cutter, S. L. Block and J. W. St Geme, 3rd (1999). "Adhesin expression in matched nasopharyngeal and middle ear isolates of nontypeable *Haemophilus influenzae* from children with acute otitis media." *Infect Immun* **67**(1): 449-54.

Kroll, J. S., P. R. Langford and B. M. Loynds (1991). "Copper-zinc superoxide dismutase of *Haemophilus influenzae* and *H. parainfluenzae*." *J Bacteriol* **173**(23): 7449-57.

Kroll, J. S., P. R. Langford, K. E. Wilks and A. D. Keil (1995). "Bacterial [Cu,Zn]-superoxide dismutase: phylogenetically distinct from the eukaryotic enzyme, and not so rare after all!" *Microbiology* **141 (Pt 9)**: 2271-9.

Kubiet, M., R. Ramphal, A. Weber and A. Smith (2000). "Pilus-mediated adherence of *Haemophilus influenzae* to human respiratory mucins." *Infect Immun* **68**(6): 3362-7.

Laarmann, S., D. Cutter, T. Juehne, S. J. Barenkamp and J. W. St Geme (2002). "The *Haemophilus influenzae* Hia autotransporter harbours two adhesive pockets that reside in the passenger domain and recognize the same host cell receptor." *Mol Microbiol* **46**(3): 731-43.

Langford, P. R., B. J. Sheehan, T. Shaikh and J. S. Kroll (2002). "Active copper- and zinc-containing superoxide dismutase in the cryptic genospecies of *Haemophilus* causing urogenital and neonatal infections discriminates them from *Haemophilus influenzae* sensu stricto." *J Clin Microbiol* **40**(1): 268-70.

Leduc, I., B. Olsen and C. Elkins (2009). "Localization of the domains of the *Haemophilus ducreyi* trimeric autotransporter DsrA involved in serum resistance and binding to the extracellular matrix proteins fibronectin and vitronectin." *Infect Immun* **77**(2): 657-66.

Levdansky, E., J. Romano, Y. Shadkchan, H. Sharon, K. J. Verstrepen, G. R. Fink and N. Osherov (2007). "Coding tandem repeats generate diversity in *Aspergillus fumigatus* genes." *Eukaryot Cell* **6**(8): 1380-91.

Linke, D., T. Riess, I. B. Autenrieth, A. Lupas and V. A. Kempf (2006). "Trimeric autotransporter adhesins: variable structure, common function." *Trends Microbiol* **14**(6): 264-70.

Litwin, C. M. and S. B. Calderwood (1993). "Role of iron in regulation of virulence genes." *Clin Microbiol Rev* **6**(2): 137-49.

Loeb, M. R., E. Connor and D. Penney (1988). "A comparison of the adherence of fimbriated and nonfimbriated *Haemophilus influenzae* type b to human adenoids in organ culture." *Infect Immun* **56**(2): 484-9.

Madoff, L. C., J. L. Michel, E. W. Gong, D. E. Kling and D. L. Kasper (1996). "Group B streptococci escape host immunity by deletion of tandem repeat elements of the alpha C protein." *Proc Natl Acad Sci U S A* **93**(9): 4131-6.

Mandrell, R. E. and M. A. Apicella (1993). "Lipo-oligosaccharides (LOS) of mucosal pathogens: molecular mimicry and host-modification of LOS." *Immunobiology* **187**(3-5): 382-402.

Martel, A. Y., G. St-Laurent, L. A. Dansereau and M. G. Bergeron (1989). "Isolation and biochemical characterization of *Haemophilus* species isolated simultaneously from the oropharyngeal and anogenital areas." *J Clin Microbiol* **27**(7): 1486-9.

Martinez, M. A., A. Ovalle, M. T. Ulloa and R. M. Vidal (1999). "Role of *Haemophilus influenzae* in intra-amniotic infection in patients with preterm rupture of membranes." *Eur J Clin Microbiol Infect Dis* **18**(12): 890-2.

Martoglio, B. and B. Dobberstein (1998). "Signal sequences: more than just greasy peptides." *Trends Cell Biol* **8**(10): 410-5.

Maurer, J., J. Jose and T. F. Meyer (1999). "Characterization of the essential transport function of the AIDA-I autotransporter and evidence supporting structural predictions." *J Bacteriol* **181**(22): 7014-20.

McDevitt, D., P. Francois, P. Vaudaux and T. J. Foster (1994). "Molecular characterization of the clumping factor (fibrinogen receptor) of *Staphylococcus aureus*." *Mol Microbiol* **11**(2): 237-48.

Meng, G., N. K. Surana, J. W. St Geme, 3rd and G. Waksman (2006). "Structure of the outer membrane translocator domain of the *Haemophilus influenzae* Hia trimeric autotransporter." *EMBO J* **25**(11): 2297-304.

Menozzi, F. D., P. E. Boucher, G. Riveau, C. Gantiez and C. Locht (1994). "Surface-associated filamentous hemagglutinin induces autoagglutination of *Bordetella pertussis*." *Infect Immun* **62**(10): 4261-9.

Michel, J. L., L. C. Madoff, K. Olson, D. E. Kling, D. L. Kasper and F. M. Ausubel (1992). "Large, identical, tandem repeating units in the C protein alpha antigen gene, bca, of group B streptococci." *Proc Natl Acad Sci U S A* **89**(21): 10060-4.

Moxon, E. R. and K. A. Vaughn (1981). "The type b capsular polysaccharide as a virulence determinant of *Haemophilus influenzae*: studies using clinical isolates and laboratory transformants." *J Infect Dis* **143**(4): 517-24.

Murphy, G. L., T. D. Connell, D. S. Barritt, M. Koomey and J. G. Cannon (1989). "Phase variation of gonococcal protein II: regulation of gene expression by slipped-strand mispairing of a repetitive DNA sequence." *Cell* **56**(4): 539-47.

Murphy, T. F., C. Kirkham and D. J. Sikkema (1992). "Neonatal, urogenital isolates of biotype 4 nontypeable *Haemophilus influenzae* express a variant P6 outer membrane protein molecule." *Infect Immun* **60**(5): 2016-22.

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**(1): 183-91.

Noel, G. J., A. Brittingham, A. A. Granato and D. M. Mosser (1996). "Effect of amplification of the Cap b locus on complement-mediated bacteriolysis and opsonization of type b *Haemophilus influenzae*." *Infect Immun* **64**(11): 4769-75.

Nummelin, H., M. C. Merckel, J. C. Leo, H. Lankinen, M. Skurnik and A. Goldman (2004). "The *Yersinia* adhesin YadA collagen-binding domain structure is a novel left-handed parallel beta-roll." *Embo J* **23**(4): 701-11.

Ochiai, K., T. Kurita-Ochiai, Y. Kamino and T. Ikeda (1993). "Effect of co-aggregation on the pathogenicity of oral bacteria." *J Med Microbiol* **39**(3): 183-90.

Oliver, D. C., G. Huang and R. C. Fernandez (2003). "Identification of secretion determinants of the *Bordetella pertussis* BrkA autotransporter." *J Bacteriol* **185**(2): 489-95.

Oomen, C. J., P. van Ulsen, P. van Gelder, M. Feijen, J. Tommassen and P. Gros (2004). "Structure of the translocator domain of a bacterial autotransporter." *EMBO J* **23**(6): 1257-66.

Pearson, M. M., E. R. Lafontaine, N. J. Wagner, J. W. St Geme, 3rd and E. J. Hansen (2002). "A hag mutant of *Moraxella catarrhalis* strain O35E is deficient in hemagglutination, autoagglutination, and immunoglobulin D-binding activities." *Infect Immun* **70**(8): 4523-33.

Pichichero, M. E., M. Loeb, Anderson and D. H. Smith (1982). "Do pili play a role in pathogenicity of *Haemophilus influenzae* type B?" *Lancet* **2**(8305): 960-2.

Quentin, R., A. Goudeau, R. J. Wallace, Jr., A. L. Smith, R. K. Selander and J. M. Musser (1990). "Urogenital, maternal and neonatal isolates of *Haemophilus influenzae*: identification of unusually virulent serologically non-typable clone families and evidence for a new *Haemophilus* species." *J Gen Microbiol* **136**(7): 1203-9.

Quentin, R., C. Martin, J. M. Musser, N. Pasquier-Picard and A. Goudeau (1993). "Genetic characterization of a cryptic genospecies of *Haemophilus* causing urogenital and neonatal infections." *J Clin Microbiol* **31**(5): 1111-6.

Quentin, R., J. M. Musser, M. Mellouet, P. Y. Sizaret, R. K. Selander and A. Goudeau (1989). "Typing of urogenital, maternal, and neonatal isolates of *Haemophilus influenzae* and Haemophilus parainfluenzae in correlation with clinical source of isolation and evidence for a genital specificity of H. influenzae biotype IV." *J Clin Microbiol*. **27**(10): 2286-94.

Quentin, R., R. Ruimy, A. Rosenau, J. M. Musser and R. Christen (1996). "Genetic identification of cryptic genospecies of *Haemophilus* causing urogenital and neonatal infections by PCR using specific primers targeting genes coding for 16S rRNA." *J Clin Microbiol* **34**(6): 1380-5.

Ramsland, P. A., N. Willoughby, H. M. Trist, W. Farrugia, P. M. Hogarth, J. D. Fraser and B. D. Wines (2007). "Structural basis for evasion of IgA immunity by *Staphylococcus aureus* revealed in the complex of SSL7 with Fc of human IgA1." *Proc Natl Acad Sci U S A* **104**(38): 15051-6.

Rao, V. K., G. P. Krasan, D. R. Hendrixson, S. Dawid and J. W. St Geme, 3rd (1999). "Molecular determinants of the pathogenesis of disease due to non-typable *Haemophilus influenzae*." *FEMS Microbiol Rev* **23**(2): 99-129.

Ratledge, C. and L. G. Dover (2000). "Iron metabolism in pathogenic bacteria." *Annu Rev Microbiol* **54**: 881-941.

Rauceo, J. M., R. De Armond, H. Otoo, P. C. Kahn, S. A. Klotz, N. K. Gaur and P. N. Lipke (2006). "Threonine-rich repeats increase fibronectin binding in the *Candida albicans* adhesin Als5p." *Eukaryot Cell* **5**(10): 1664-73.

Ray, S. K., R. Rajeshwari, Y. Sharma and R. V. Sonti (2002). "A high-molecular-weight outer membrane protein of *Xanthomonas oryzae* pv. oryzae exhibits similarity to non-fimbrial adhesins of animal pathogenic bacteria and is required for optimum virulence." *Mol Microbiol* **46**(3): 637-47.

Reddy, M. S., J. M. Bernstein, T. F. Murphy and H. S. Faden (1996). "Binding between outer membrane proteins of nontypeable *Haemophilus influenzae* and human nasopharyngeal mucin." *Infect Immun* **64**(4): 1477-9.

Riess, T., S. G. Andersson, A. Lupas, M. Schaller, A. Schafer, P. Kyme, J. Martin, J. H. Walzlein, U. Ehehalt, H. Lindroos, M. Schirle, A. Nordheim, I. B. Autenrieth and V. A. Kempf (2004). "*Bartonella* adhesin a mediates a proangiogenic host cell response." *J Exp Med* **200**(10): 1267-78.

Roberts, J. A., B. I. Marklund, D. Ilver, D. Haslam, M. B. Kaack, G. Baskin, M. Louis, R. Mollby, J. Winberg and S. Normark (1994). "The Gal(alpha 1-4)Gal-specific tip adhesin of *Escherichia coli* P-fimbriae is needed for pyelonephritis to occur in the normal urinary tract." *Proc Natl Acad Sci U S A* **91**(25): 11889-93.

Roggenkamp, A., N. Ackermann, C. A. Jacobi, K. Truelzsch, H. Hoffmann and J. Heesemann (2003). "Molecular analysis of transport and oligomerization of the *Yersinia enterocolitica* adhesin YadA." *J Bacteriol* **185**(13): 3735-44.

Rosenau, A., P. Y. Sizaret, J. M. Musser, A. Goudeau and R. Quentin (1993). "Adherence to human cells of a cryptic *Haemophilus* genospecies responsible for genital and neonatal infections." *Infect Immun* **61**(10): 4112-8.

Rozen, S. and H. Skaletsky (2000). "Primer3 on the WWW for general users and for biologist programmers." *Methods Mol Biol* **132**: 365-86.

Rusin, P., R. D. Adam, E. A. Peterson, K. J. Ryan, N. A. Sinclair and L. Weinstein (1991). "*Haemophilus influenzae*: an important cause of maternal and neonatal infections." *Obstet Gynecol* **77**(1): 92-6.

Sambrook, J., E. F. Fritsch and T. Maniatis (1989). Molecular cloning : a laboratory manual. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press.

San Mateo, L. R., M. M. Hobbs and T. H. Kawula (1998). "Periplasmic copper-zinc superoxide dismutase protects *Haemophilus ducreyi* from exogenous superoxide." *Mol Microbiol* **27**(2): 391-404.

Sandt, C. H. and C. W. Hill (2000). "Four different genes responsible for nonimmune immunoglobulin-binding activities within a single strain of *Escherichia coli*." *Infect Immun* **68**(4): 2205-14.

Sandt, C. H. and C. W. Hill (2001). "Nonimmune binding of human immunoglobulin A (IgA) and IgG Fc by distinct sequence segments of the EibF cell surface protein of *Escherichia coli*." *Infect Immun* **69**(12): 7293-303.

Schembri, M. A., G. Christiansen and P. Klemm (2001). "FimH-mediated autoaggregation of *Escherichia coli*." *Mol Microbiol* **41**(6): 1419-30.

Schembri, M. A., L. Hjerrild, M. Gjermansen and P. Klemm (2003). "Differential expression of the *Escherichia coli* autoaggregation factor antigen 43." *J Bacteriol* **185**(7): 2236-42.

Schembri, M. A. and P. Klemm (2001). "Biofilm formation in a hydrodynamic environment by novel fimh variants and ramifications for virulence." *Infect Immun* **69**(3): 1322-8.

Sebulsky, M. T., D. Hohnstein, M. D. Hunter and D. E. Heinrichs (2000). "Identification and characterization of a membrane permease involved in iron-hydroxamate transport in *Staphylococcus aureus*." *J Bacteriol* **182**(16): 4394-400.

Serruto, D., T. Spadafina, M. Scarselli, S. Bambini, M. Comanducci, S. Hohle, M. Kilian, E. Veiga, P. Cossart, M. R. Oggioni, S. Savino, I. Ferlenghi, A. R. Taddei, R. Rappuoli, M. Pizza, V. Masignani and B. Arico (2009). "HadA is an atypical new multifunctional trimeric coiled-coil adhesin of *Haemophilus influenzae* biogroup aegyptius, which promotes entry into host cells." *Cell Microbiol* **11**(7): 1044-63.

Setlow, J. K., D. C. Brown, M. E. Boling, A. Mattingly and M. P. Gordon (1968). "Repair of deoxyribonucleic acid in *Haemophilus influenzae*. I. X-ray sensitivity of ultraviolet-sensitive mutants and their behavior as hosts to ultraviolet-irradiated bacteriophage and transforming deoxyribonucleic acid." *J Bacteriol* **95**(2): 546-58.

Sheets, A. J., S. A. Grass, S. E. Miller and J. W. St Geme, 3rd (2008). "Identification of a novel trimeric autotransporter adhesin in the cryptic genospecies of *Haemophilus*." *J Bacteriol* **190**(12): 4313-20.

Sherlock, O., R. M. Vejborg and P. Klemm (2005). "The TibA adhesin/invasin from enterotoxigenic *Escherichia coli* is self recognizing and induces bacterial aggregation and biofilm formation." *Infect Immun* **73**(4): 1954-63.

Sirakova, T., P. E. Kolattukudy, D. Murwin, J. Billy, E. Leake, D. Lim, T. DeMaria and L. Bakaletz (1994). "Role of fimbriae expressed by nontypeable *Haemophilus influenzae* in pathogenesis of and protection against otitis media and relatedness of the fimbrin subunit to outer membrane protein A." *Infect Immun* **62**(5): 2002-20.

Smith-Vaughan, H. C., K. S. Sriprakash, J. D. Mathews and D. J. Kemp (1997). "Nonencapsulated *Haemophilus influenzae* in Aboriginal infants with otitis media: prolonged carriage of P2 porin variants and evidence for horizontal P2 gene transfer." *Infect Immun* **65**(4): 1468-74.

St Geme, J. W., 3rd (1993). "Nontypeable *Haemophilus influenzae* disease: epidemiology, pathogenesis, and prospects for prevention." *Infect Agents Dis* **2**(1): 1-16.

St Geme, J. W., 3rd (2002). "Molecular and cellular determinants of non-typeable *Haemophilus influenzae* adherence and invasion." *Cell Microbiol* **4**(4): 191-200.

St Geme, J. W., 3rd and D. Cutter (2000). "The *Haemophilus influenzae* Hia adhesin is an autotransporter protein that remains uncleaved at the C terminus and fully cell associated." *J Bacteriol* **182**(21): 6005-13.

St Geme, J. W., 3rd, D. Cutter and S. J. Barenkamp (1996). "Characterization of the genetic locus encoding *Haemophilus influenzae* type b surface fibrils." *J Bacteriol* **178**(21): 6281-7.

St Geme, J. W., 3rd, M. L. de la Morena and S. Falkow (1994). "A *Haemophilus influenzae* IgA protease-like protein promotes intimate interaction with human epithelial cells." *Mol Microbiol* **14**(2): 217-33.

St Geme, J. W., 3rd and S. Falkow (1990). "*Haemophilus influenzae* adheres to and enters cultured human epithelial cells." *Infect Immun* **58**(12): 4036-44.

St Geme, J. W., 3rd, S. Falkow and S. J. Barenkamp (1993). "High-molecular-weight proteins of nontypable *Haemophilus influenzae* mediate attachment to human epithelial cells." *Proc Natl Acad Sci U S A* **90**(7): 2875-9.

St Geme, J. W., 3rd, V. V. Kumar, D. Cutter and S. J. Barenkamp (1998). "Prevalence and distribution of the hmw and hia genes and the HMW and Hia adhesins among genetically diverse strains of nontypeable *Haemophilus influenzae*." *Infect Immun* **66**(1): 364-8.

St Geme, J. W., J. S. Pinkner, 3rd, G. P. Krasan, J. Heuser, E. Bullitt, A. L. Smith and S. J. Hultgren (1996). "*Haemophilus influenzae* pili are composite structures assembled via the HifB chaperone." *Proc Natl Acad Sci U S A* **93**(21): 11913-8.

Surana, N. K., D. Cutter, S. J. Barenkamp and J. W. St Geme, 3rd (2004). "The *Haemophilus influenzae* Hia autotransporter contains an unusually short trimeric translocator domain." *J Biol Chem* **279**(15): 14679-85.

Sutton, A., R. Schneerson, S. Kendall-Morris and J. B. Robbins (1982). "Differential complement resistance mediates virulence of *Haemophilus influenzae* type b." *Infect Immun* **35**(1): 95-104.

Swords, W. E., B. A. Buscher, K. Ver Steeg Ii, A. Preston, W. A. Nichols, J. N. Weiser, B. W. Gibson and M. A. Apicella (2000). "Non-typeable *Haemophilus influenzae* adhere to and invade human bronchial epithelial cells via an interaction of lipooligosaccharide with the PAF receptor." *Mol Microbiol* **37**(1): 13-27.

Swords, W. E., M. R. Ketterer, J. Shao, C. A. Campbell, J. N. Weiser and M. A. Apicella (2001). "Binding of the non-typeable *Haemophilus influenzae* lipooligosaccharide to the PAF receptor initiates host cell signalling." *Cell Microbiol* **3**(8): 525-36.

Szczesny, P., D. Linke, A. Ursinus, K. Bar, H. Schwarz, T. M. Riess, V. A. Kempf, A. N. Lupas, J. Martin and K. Zeth (2008). "Structure of the head of the *Bartonella* adhesin BadA." *PLoS Pathog* **4**(8): e1000119.

Szczesny, P. and A. Lupas (2008). "Domain annotation of trimeric autotransporter adhesins--daTAA." *Bioinformatics* **24**(10): 1251-6.

Tiyawisutsri, R., M. T. Holden, S. Tumapa, S. Rengpipat, S. R. Clarke, S. J. Foster, W. C. Nierman, N. P. Day and S. J. Peacock (2007). "Burkholderia Hep_Hag autotransporter (BuHA) proteins elicit a strong antibody response during experimental glanders but not human melioidosis." *BMC Microbiol* **7**: 19.

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

Valle, J., A. N. Mabbett, G. C. Ulett, A. Toledo-Arana, K. Wecker, M. Totsika, M. A. Schembri, J. M. Ghigo and C. Beloin (2008). "UpaG, a new member of the trimeric autotransporter family of adhesins in uropathogenic *Escherichia coli*." *J Bacteriol* **190**(12): 4147-61.

van Ham, S. M., L. van Alphen, F. R. Mooi and J. P. van Putten (1993). "Phase variation of *H. influenzae* fimbriae: transcriptional control of two divergent genes through a variable combined promoter region." *Cell* **73**(6): 1187-96.

Vidal, O., R. Longin, C. Prigent-Combaret, C. Dorel, M. Hooreman and P. Lejeune (1998). "Isolation of an *Escherichia coli* K-12 mutant strain able to form biofilms on inert surfaces: involvement of a new ompR allele that increases curli expression." *J Bacteriol* **180**(9): 2442-9.

Virkola, R., M. Brummer, H. Rauvala, L. van Alphen and T. K. Korhonen (2000). "Interaction of fimbriae of *Haemophilus influenzae* type B with heparin-binding extracellular matrix proteins." *Infect Immun* **68**(10): 5696-701.

Voulhoux, R., M. P. Bos, J. Geurtsen, M. Mols and J. Tommassen (2003). "Role of a highly conserved bacterial protein in outer membrane protein assembly." *Science* **299**(5604): 262-5.

Wallace, R. J., Jr., C. J. Baker, F. J. Quinones, D. G. Hollis, R. E. Weaver and K. Wiss (1983). "Nontypable *Haemophilus influenzae* (biotype 4) as a neonatal, maternal, and genital pathogen." *Rev Infect Dis* **5**(1): 123-36.

Weiser, J. N. (1993). "Relationship between colony morphology and the life cycle of *Haemophilus influenzae*: the contribution of lipopolysaccharide phase variation to pathogenesis." *J Infect Dis* **168**(3): 672-80.

Weiser, J. N., D. J. Maskell, P. D. Butler, A. A. Lindberg and E. R. Moxon (1990). "Characterization of repetitive sequences controlling phase variation of *Haemophilus influenzae* lipopolysaccharide." *J Bacteriol* **172**(6): 3304-9.

Wells, T. J., J. J. Tree, G. C. Ulett and M. A. Schembri (2007). "Autotransporter proteins: novel targets at the bacterial cell surface." *FEMS Microbiol Lett* **274**(2): 163-72.

Wilks, K. E., K. L. Dunn, J. L. Farrant, K. M. Reddin, A. R. Gorringe, P. R. Langford and J. S. Kroll (1998). "Periplasmic superoxide dismutase in meningococcal pathogenicity." *Infect Immun* **66**(1): 213-7.

Wurzner, R. (1999). "Evasion of pathogens by avoiding recognition or eradication by complement, in part via molecular mimicry." *Mol Immunol* **36**(4-5): 249-60.

Yeo, H. J., S. E. Cotter, S. Laarmann, T. Juehne, J. W. St Geme, 3rd and G. Waksman (2004). "Structural basis for host recognition by the *Haemophilus influenzae* Hia autotransporter." *EMBO J* **23**(6): 1245-56.

Yu, C., T. Ruiz, C. Lenox and K. P. Mintz (2008). "Functional mapping of an oligomeric autotransporter adhesin of *Aggregatibacter actinomycetemcomitans*." *J Bacteriol* **190**(9): 3098-109.

Zhang, P., B. B. Chomel, M. K. Schau, J. S. Goo, S. Droz, K. L. Kelminson, S. S. George, N. W. Lerche and J. E. Koehler (2004). "A family of variably expressed outer-membrane proteins (Vomp) mediates adhesion and autoaggregation in *Bartonella quintana*." *Proc Natl Acad Sci U S A* **101**(37): 13630-5.

Zhang, Q. and K. S. Wise (1996). "Molecular basis of size and antigenic variation of a *Mycoplasma hominis* adhesin encoded by divergent vaa genes." *Infect Immun* **64**(7): 2737-44.

Biography

Amanda Joan Sheets was born on April 19, 1981 in Princeton, NJ. She grew up in Bedminster, PA and graduated from Pennridge high school in 1999. She attended Gettysburg College in Gettysburg, Pennsylvania graduating *suma cum laude* and first in her major with a B.S. in Biochemistry and Molecular Biology in May 2003. During the spring semester of 2002, she studied abroad at the University of Queensland in Australia. Upon returning to the United States, she completed a summer research internship at the National Institute of Standards and Technology (NIST), studying the mechanism producing mitochondrial DNA heteroplasmies under the guidance of Dr. Koren Holland and as part of her senior thesis project. The summer of 2003, following graduation from Gettysburg, she entered the Molecular Microbiology and Microbial Pathogenesis Program within the Division of Biology and Biomedical Sciences at Washington University in St. Louis. She began her graduate work in the lab of Dr. Joe St. Geme studying autotransporter adhesins of *H. influenzae* and the cryptic genospecies. After two years of study at WashU, she moved with the St. Geme lab to Duke University as a result of her thesis mentor accepting a new position as Chairman of the Department of Pediatrics. Upon transferring to Duke, she joined the Molecular Genetics and Microbiology Program where she completed the remainder of her thesis project, and in her spare time, began competing with the Duke club lacrosse team. In June 2008, she received the fellow/graduate student award for best poster presentation at the

Department of Pediatrics Retreat at Duke University Medical Center. In 2009, she received a Bacteriology Meritorious Research Travel Award from the Center for Microbial Pathogenesis at Duke to present her work at the 2009 Microbial Pathogenesis and Host Response Research Meeting at CSH in Cold Spring Harbor, NY.

Peer-Reviewed Publications

Deckman KH, Kishore R, Albanetti T, Peery S, Knipe A, **Sheets A**, Boire N, Reiner JE, DeLongchamp DM, Vogt BD, Brooks CM, Kano K, Obrzut J, Richter CA, Kirillov OA, Lin EK. (2005) Influence of a water rinse on the structure and properties of poly(3,4-ethylene dioxythiophene):poly(styrene sulfonate) films. Langmuir 21 (24): 11480-11483.

Sheets AJ, Grass SA, Miller SE, St Geme JW III. (2008) Identification of a novel trimeric autotransporter adhesin in the cryptic genospecies of *Haemophilus*. *J Bacteriol.* 190(12):4313-20.

Sheets AJ, and St Geme JW III. The *Haemophilus* cryptic genospecies Cha autotransporter mediates adherence to epithelial cells and microcolony formation via binding to regions in the N-terminal end of the passenger domain. (manuscript in preparation)

Sheets AJ, and St Geme JW III. Mechanism of phase variable Cha-mediated adherence in the *Haemophilus* cryptic genospecies. (manuscript in preparation)

Book Chapters

Sheets AJ, and St Geme JW III. Bacterial Adhesion to the Cell Surface and Extracellular Matrix of Host Tissues: Non-pilus adhesins. In: Bacterial Pathogenesis. Locht C and Simonet M (eds.) Horizon Press; Norwich, UK. In Preparation.

Abstracts, Oral and/or Poster Presentations:

1. Reiner J, Kishore R, Albanetti T, Peery S, Knipe A, **Sheets A**, Boire N, Deckman K, Levin B, and Helmerson K. (2006) A methodology to separate single cells and single mitochondria to determine the location of heteroplasmy in mitochondrial DNA. American Physical Society Meeting, Baltimore, MD

2. **Sheets AJ**, and St. Geme JW III. (2007) Identification of a novel trimeric autotransporter adhesin in the cryptic genospecies of *Haemophilus*. Department of Pediatrics Retreat, Duke University Medical Center.

3. **Sheets AJ**, and St. Geme JW III. (2008) Mechanism of phase variable adherence in the *Haemophilus* cryptic genospecies. Department of Pediatrics Retreat, Duke University Medical Center.

4. **Sheets AJ**, and St. Geme JW III. (2009) Mechanism of phase variable adherence in the *Haemophilus* cryptic genospecies. Mid-Atlantic Microbial Pathogenesis Meeting, Wintergreen, VA.

5. **Sheets AJ**, and St. Geme JW III. (2009) Phase-variable adhesive activity of the *Haemophilus* cryptic genospecies trimeric autotransporter Cha. Microbial Pathogenesis and Host Response Meeting, Cold Spring Harbor, NY.