Characterization of CEACAM binding by Human Specific Pathogens

by

Michael John Brooks

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy

Graduate Department of Molecular Genetics

University of Toronto

© Copyright by Michael John Brooks (2009)
Characterization of CEACAM binding by Human Specific Pathogens

Doctor of Philosophy (2009)

Michael John Brooks

Department of Molecular Genetics

University of Toronto

Abstract

Human restricted pathogens such as *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Moraxella catarrhalis* and *Haemophilus influenzae*, as well as bacteria that are responsible for human specific disease, such as adherent-invasive and diffusely adhering *E. coli*, have evolved unique mechanisms to target and exploit host cellular CEACAM receptors during the course of infection. To gain further insight into how pathogens interact with CEACAMs, my studies have focused on understanding the molecular basis of bacterial attachment to CEACAMs by *M. catarrhalis* and *N. gonorrhoeae*. It is now evident *M. catarrhalis* and *Neisseria* sp. use two structurally and phylogenetically unrelated adhesins to bind CEACAMs on host cells; UspA1 and Opa proteins, respectively.

In this thesis, I identify the sequence determinants for CEACAM binding in both *Neisseria* sp. Opa proteins and *M. catarrhalis* UspA1 proteins and show these sequences exist in distinct structural architectures within each adhesin, with the CEACAM binding
regions in UspA1 existing in the context of an α-helical coiled coil structure while Opa sequences are localized to surface exposed loops that are presumably flexible in nature. I also highlight the functional diversity that exists with respect to CEACAM binding, within both the Opa and UspA protein families, a finding important in furthering our understanding of the role of CEACAM in bacterial infections. Finally, these studies show that sequences within Opa are sufficient for adherence to CEACAM, providing important insight into the development of CEACAM-specific therapeutics that have potential applications for a variety of distinct conditions ranging from autoimmunity to cancer.
This work is dedicated to my parents, John and Marie Brooks. Without their support and encouragement I would not have been able to pursue this degree.
Acknowledgements

I have been fortunate to have had the support of many great people during the course of my Ph.D., who have contributed in a variety of ways to help me complete this degree. First, I feel incredibly lucky to have been mentored by Dr. Scott Gray-Owen. Scott is a brilliant scientist and fantastic person. His support for me has been unwavering, and he has let me explore science and find my niche within it. He created an amazing environment where I could learn and grow as a scientist with constant support and guidance. It is difficult to properly put into words how much Scott has contributed to my successful completion of this degree and I want to thank him for everything he has done for me over the years.

I would also like to thank my committee members Dr. Emil Pai and Dr. Martin McGavin for all their guidance and encouragement through this journey. During the course of my degree, I was able to establish a relationship with Dr. Eric Hansen through a collaborative project. Dr. Hansen ended up playing a huge role in my graduate career. His support, encouragement and generosity helped me grow as a scientist. He let me explore my ideas, supporting me and my work in every way possible. I want to thank him for all he has done to aid me during the course my scientific training.

The members of the Gray-Owen lab have been a constant source of advice and support over the years. I feel fortunate to have been able to get to know all of you and couldn’t have asked for a better group of people to have traveled along this path with.

I have been able to pursue and complete my Ph.D. because of the encouragement of many people in my personal life. I am lucky to have a close family that despite vast geographical separation, have both supported and encouraged me throughout this
journey. I’d like to thank all my friends especially Josh, Brandon and Eddy for all their friendship and advice over the years. I am thankful for having a great family and group of friends whose support helped me achieve this degree. I want to thank Stephanie for her love and support over the years, and her parents Linda and Danny for treating me like a son, as I struggled through my Ph.D.

Finally, I would like to thank my brother, Iain, and my sister, Christine, for their unwavering support during my pursuit of this degree. I’ve been able to rely on their support throughout the course of my Ph.D. and they have been instrumental in every step I’ve taken in my life. And to my parents, John and Marie, you have made sacrifices for me and my siblings that I will never fully be able to understand but that I truly appreciate. Without my parents support and continuous encouragement I would not have been able to pursue and complete my Ph.D. They have made many sacrifices in their own lives to allow me to follow my dreams. I cannot properly thank them for all they have given me.
Table of Contents

Abstract .............................................................................................................................. ii
Acknowledgements ........................................................................................................ iv
Table of Contents ........................................................................................................... vii
List of Figures ................................................................................................................ ix
List of Tables ................................................................................................................... ix
Abbreviations ................................................................................................................ x
Chapter 1: Introduction ................................................................................................... 1
  1.1 Carcinoembryonic antigen-related cellular adhesion molecules (CEACAMs) .......... 3
     1.1.1 Overview of CEACAM general architecture ................................................. 7
     1.1.2 CEACAM1 ..................................................................................................... 14
     1.1.3 CEACAM5 ..................................................................................................... 19
     1.1.4 CEACAM6 ..................................................................................................... 20
     1.1.5 CEACAM3 ..................................................................................................... 20
  1.2 Infection cycle .......................................................................................................... 21
     1.2.1 Epithelial cells ............................................................................................... 22
     1.2.2 Neutrophils ................................................................................................... 24
     1.2.3 Lymphocytes ................................................................................................. 25
     1.2.4 Dendritic cells ............................................................................................... 26
     1.2.5 Endothelial cells ............................................................................................ 27
  1.3 CEACAM binding pathogens. ................................................................................. 28
     1.3.1 Neisseria sp. ................................................................................................. 28
     1.3.2 Moraxella catarrhalis .................................................................................... 34
     1.3.3 Haemophilus influenzae ............................................................................... 37
     1.3.4 Pathogenic E. coli .......................................................................................... 40
  1.4 Thesis Rationale ....................................................................................................... 42
Chapter 2: Characterization of the Moraxella catarrhalis Opa-like protein, OlpA, reveals a phylogenetically conserved family of outer membrane proteins. ...................... 43
  2.1 Abstract .................................................................................................................. 44
  2.2 Introduction ............................................................................................................. 45
  2.3 Results .................................................................................................................... 46
     2.3.1 Identification of OlpA ................................................................................... 46
     2.3.2 OlpA is a putative outer membrane protein .................................................. 47
     2.3.3 Expression of OlpA ....................................................................................... 49
     2.3.4 Conservation of OlpA among M. catarrhalis strains ....................................... 51
     2.3.5 Surface accessibility and heat modifiability ................................................... 51
     2.3.6 M. catarrhalis OlpA does not bind to CEACAM receptors .......................... 53
     2.3.7 Homologues .................................................................................................. 54
  2.3 Discussion ............................................................................................................... 56
  2.4 Materials and Methods ......................................................................................... 59
Chapter 3: Moraxella catarrhalis binding to host cellular receptors is mediated by sequence-specific determinants not conserved between all UspA1 protein variants ...... 64
  3.1 Abstract .................................................................................................................. 65
  3.2 Introduction ............................................................................................................. 66
  3.2 Results .................................................................................................................... 67
     3.2.1 UspA1 proteins differ in cellular adherence through CEACAM ................. 67
3.2.2 CEACAM binding by UspA1 variants ................................................. 69
3.2.3 Sequence Analysis of CEACAM binding in UspA1 proteins.................. 73
3.2.4 Fibronectin and Chang cell binding.................................................... 74
3.3 Discussion .................................................................................................. 76
3.4 Materials and Methods ........................................................................... 79

Chapter 4: Modular arrangement of allelic variants explains divergence in *M. catarrhalis*
UspA protein function .................................................................................... 85
4.1 Abstract ..................................................................................................... 86
4.2 Introduction ................................................................................................ 87
4.2 Results ....................................................................................................... 89
  4.2.1 UspA proteins .................................................................................... 89
  4.2.2 UspA N-domains ............................................................................... 91
  4.2.3 UspA stalk region .............................................................................. 95
  4.2.4 UspA C-terminal regions ................................................................. 97
  4.2.5 Modular arrangement of UspA1 ......................................................... 99
  4.2.6 Modular arrangement of UspA2 ......................................................... 101
4.4 Discussion .................................................................................................. 104
4.5 Materials and Methods ........................................................................... 107

Chapter 5: Characterization of *Neisseria* Opa-CEACAM interactions .......... 109
5.1 Abstract ..................................................................................................... 110
5.2 Introduction ............................................................................................... 111
5.3 Results ....................................................................................................... 113
  5.3.1 Phage-display of Opa fragments to soluble CEACAM ...................... 113
  5.3.2 Exploiting the diversity of MS11 Opa proteins ................................... 115
  5.3.4 Exchange of CEACAM receptor specificity ..................................... 117
  5.3.6 Gain of CEACAM receptor binding .................................................. 119
5.4 Discussion .................................................................................................. 124
5.5 Materials and Methods ........................................................................... 128

Chapter 6: Summary and Future directions ................................................... 132
6.1 Summary ................................................................................................... 133
6.2 Future directions ....................................................................................... 137
List of Figures

Fig.1-1 Carcinoembryonic antigen-related cellular adhesion molecules ........................................5
Fig.1-2 CEACAM N-domains ........................................................................................................9
Fig.1-3 CEACAM IgC and membrane anchor ..............................................................................11
Fig.1-4 Human CEACAMs targeted by bacterial adhesins .........................................................16
Fig.1-5 Neisserial infection cycle ..................................................................................................23
Fig.1-6 Neisserial Opa proteins ..................................................................................................31
Fig.2-1 Sequence/Structural prediction of M. catarrhalis OlpA ..................................................48
Fig.2-2 Expression of OlpA ........................................................................................................50
Fig.2-3 Surface expression of OlpA .............................................................................................52
Fig.2-4 Phylogenetic and sequence alignments of OlpA homologue .........................................55
Fig.3-1 Cellular tropism of UspA1-O35E ...................................................................................68
Fig.3-2 CEACAM binding by recombinant E.coli expressing UspA1 variants ............................70
Fig.3-3 M. catarrhalis wild-type or UspA1-deficient binding to CEACAMs .................................72
Fig.3-4 UspA1 protein sequence alignments ..............................................................................75
Fig.4-1 Overview of UspA structure ............................................................................................90
Fig.4-2 UspA1 and UspA2H N-terminal domain arrangement .................................................92
Fig.4-3 UspA2 N-terminal domains ..........................................................................................94
Fig.4-4 UspA stalk repeats ..........................................................................................................96
Fig.4-5 C-terminal region motifs of UspA proteins .................................................................98
Fig.4-6 Modular arrangement of UspA1 proteins .................................................................100
Fig.4-7 Modular arrangement of UspA2 and UspA2H proteins ..............................................102
Fig.5-1 Phage display of neisserial loops .................................................................................114
Fig.5-2 Phylogenetic generation of Opa chimeras .................................................................116
Fig.5-3 CEACAM binding by Opa55 and Opa58 chimeras .......................................................118
Fig.5-4 CEACAM binding by Opa50 and Op58 chimeras ........................................................120
Fig.5-5 CEACAM binding by Op55 and Opa50 chimeras .......................................................122
Fig.6-1 Comparison of Opa and UspA1 CEACAM binding .....................................................135
Fig.6-2 Inhibition of T-cell proliferation and activation by Opa fragments .............................141

List of Tables

Table.1-1 Summary of CEACAM binding adhesins .................................................................38
Table.2-1 M. catarrhalis strains used in chapter 2 .....................................................................63
Table.2-2 Accession numbers of olpA genes/ PCR primers ....................................................63
Table.3-1 M. catarrhalis strains used in chapter 3 .....................................................................84
Table.4-1 Moraxella catarrhalis UspA1, Usp2A, and Usp2AH protein sequences .................108
Table.5-1 Primers for amplification of Opa surface loops .......................................................131
Abbreviations

AIEC – Adherent invasive Escherichia coli
BHI – Brain heart infusion
CEACAM – Carcinoembryonic antigen related cellular adhesion molecules
CHO – Chinese hamster ovary
COPD – Chronic obstructive pulmonary disease
ECM – Extracellular matrix
ELISA – Enzyme linked ImmunoSorbent Assay
ETEC – Enterotoxic Escherichia coli
EPEC – Enteropathogenic Escherichia coli
EHEC – Enterohemorrhagic Escherichia coli
EAEC – Enteroaggregative Escherichia coli
EIEC – Enteroinvasive Escherichia coli
Fbp – Ferric binding protein
GPI – glycolsylphosphatidylinositol
HV1 – Hypervariable region 1
HV2 – Hypervariable region 2
HSPG – Heparin sulfate proteoglycans
Ig – Immunoglobulin
IgV-like – Immunoglobulin variable like
IgC-like – Immunoglobulin constant like
IgSFRs – immunoglobulin super-family of receptors
ITAM – immunotyrosine-based activation motif
ITIM – immunotyrosine-based inhibitory motif
LB – Luria broth
MBP – Maltose binding protein
NFκB – nuclear factor-kappa B
NCAM – Neural cellular adhesion molecule
NspA – Neisserial surface protein A
Opa – Opacity associated proteins
OlpA – Opa-like protein A
PI3K – Phosphoinositide 3-kinases
SV – Semi-variable region
UPEC – Uropathogenic Escherichia coli
UspA1 – Ubiquitous surface protein A1
UspA2 – Ubiquitous surface protein A2
UspA2H – Ubiquitous surface protein A2H
VN – Vitronectin
Chapter 1: Introduction
Over the past 20 years, it has become evident that a variety of bacterial pathogens have evolved mechanisms to thrive in their restricted host environment by targeting a subset of surface glycoproteins involved in cell-cell adhesion called the Carcinoembryonic antigen-related cellular adhesion molecules (CEACAMs). CEACAMs are present on numerous cell types and tissues accessible at various stages of the infection cycle. Through interaction with CEACAMs, bacteria and viruses can colonize epithelia, invade host tissues and modulate the immune response. Interestingly, diverse bacterial species have convergently evolved unique adhesins that target CEACAMs. Human restricted pathogens such as *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Moraxella catarrhalis* and *Haemophilus influenzae*, as well as bacteria that are responsible for human specific disease, such as adherent-invasive and diffusely adhering *E. coli*, have evolved mechanisms to target and utilize CEACAMs during the course of infection. The focus of my thesis is to characterize the molecular basis by which diverse bacterial adhesins from *N. gonorrhoeae* and *M. catarrhalis* adhere to CEACAMs as host cellular receptors. The common link between these pathogens is their ability to target CEACAMs, so this introduction will describe the structure and function of the CEACAM family of receptors, highlighting the regions and residues targeted by the bacteria. I will then detail the tissue distribution and expression patterns of CEACAM on different cell types and tissues in the context of a bacterial infection. Finally, I will describe the different bacterial species that bind to CEACAMs and focus on the adhesins these bacterial species utilize to exploit CEACAM to their advantage.
1.1 Carcinoembryonic antigen-related cellular adhesion molecules (CEACAMs)

Carcinoembryonic antigen (CEA or CEACAM5) was first identified as a prominent tumor associated antigen in human colon cancer in the 1960s (Gold and Freedman, 1965) and has since been described to be upregulated in up to 70% of cancers (Chevinsky, 1991; Jantscheff et al., 2003). Since the identification of CEA, it has become apparent that it is but one member of the large family of proteins. The carcinoembryonic antigen-related cellular adhesion molecules (CEACAMs) are a family of surface glycoproteins involved in cell-cell interactions that influence cellular growth and differentiation and have been shown to effect vascular neogenesis (Ergun et al., 2000; Wagener and Ergun, 2000), diabetes (Poy et al., 2002), tumor development (Leung et al., 2006) and apoptosis (Nittka et al., 2004). CEACAMs are members of the immunoglobulin super-family of receptors (IgSFRs) and encoded within the carcinoembryonic antigen (CEA) family locus. IgSFRs are cell surface glycoproteins that contain at least one characteristic \( \beta \)-sheet fold of the immunoglobulin domain. These immunoglobulin-like domains are one of the most prominent protein folds in nature, conserved across many species despite impressive divergence in primary amino acid sequence. Ig-like domains have been found in species ranging from humans to bacteria providing a basic structural backbone for an almost limitless number of protein-protein interactions. In addition to the CEA family, the Ig-like fold exists in many other important cell surface glycoproteins, including immune receptors such as components of the T-cell receptor (TCR) (Garcia et al., 1996), CD4 (Ryu et al., 1990), CD8 (Leahy et al., 1992), Major histocompatibility complex (MHC) (Garrett et al., 1989) and
extracellular matrix (ECM) components such as fibronectin (Main *et al.*, 1992). Ig-like domains also form the structural unit of many bacterial proteins such as the tip protein of type I pili FimH (Choudhury *et al.*, 1999) as well as bacteriophage tail proteins (Fraser *et al.*, 2007).

Each CEACAM contains a single N-terminal immunoglobulin variable-like (IgV-like) domain linked to a variable number (from zero to 6) of immunoglobulin constant-like (IgC) domains and a membrane anchor (Fig. 1-1). The crystal structure of the murine CEACAM1 N-terminal domain confirmed that its Ig-like fold was intact (Tan *et al.*, 2002). Since then, the structures of the N-terminal IgV-like domain of human CEACAM1 (Fedarovich *et al.*, 2006) and CEACAM5 (Korotkova *et al.*, 2008) have also been solved. While individual CEACAMs may have from zero to six IgC-like domains it is important to appreciate that certain CEACAMs may also exist as distinct alternatively spliced isoforms that vary in the number of IgC-like domains. The effect that varying the number and type of IgC domains has on CEACAM function remains unclear. CEACAMs are anchored to the plasma membrane through either a transmembrane domain with a cytoplasmic tail or by being directly linked to glycosylphosphatidylinositol (GPI). CEACAMs with the proteinaceous membrane anchor may also undergo differential splicing to generate isoforms possessing different cell signaling motifs. The expression of these isoforms can thereby elicit differential cellular response upon CEACAM binding.

The 12 human CEACAMs (Fig. 1-1) have a wide variety of tissue distribution and attributed functions, but primarily function in cell-cell adhesion mediated through homophilic (e.g. CEACAM1-CEACAM1) and heterophilic (e.g. CEACAM1-
Fig. 1-1. Human carcinoembryonic antigen-related cellular adhesion molecules. (Adapted from http://www.carcinoembryonic-antigen.de/)
CEACAM5) interactions. CEACAMs are expressed primarily on the apical side of most epithelial cells; however different cells within a tissue typically express distinct subsets of the family rather than their expression being all or none. This is evident in the intestinal epithelia, where CEACAM5 and CEACAM6 are localized to the distal tips of microvilli while CEACAM1 and CEACAM7 were expressed primarily in the narrow crypts between microvilli (Hammarstrom and Baranov, 2001). CEACAM1, the evolutionary progenitor of the family, is also on endothelial cells and leukocytes, presumably playing a role in both extra-vascular interaction of leukocytes at sites of inflammation and antigen presenting cell binding to T cells during immunologic synapses (Gray-Owen and Blumberg, 2006). CEACAM3 and CEACAM4 expression is restricted to granulocytes, and are thought to be part of innate immune defenses against a variety of pathogens. More detailed discussion of those attributes of these receptors that pertain to my studies will follow below.

My primary interest in CEACAMs is derived from evidence that specific CEACAM family members can be hijacked by foreign microbes as receptors for colonization, invasion and immune cell modulation. CEACAMs were initially identified as host-cellular receptors targeted by colony opacity-associated (Opa) proteins expressed by human-restricted Neisseria sp. on neutrophils (Chen and Gotschlich, 1996; Virji et al., 1996a; Gray-Owen et al., 1997a). While the interaction with neutrophils is not beneficial to the bacteria, it has since been shown that Neisseria species are also capable of adhering to multiple members of the CEACAM family (CEACAM1, CEACAM5 and CEACAM6) present on the surface of epithelial cells the nasopharynx and in the urogenital tract (Gray-Owen et al., 1997b; Virji et al., 1996a). In recent years, it has
become clear that both *Moraxella catarrhalis* and non-typeable *Haemophilus influenzae* also interact with CEACAMs through distinct outer membrane-localized adhesins (Hill *et al.*, 2001; Hill and Virji, 2003). Both of these bacteria are considered asymptomatic commensal while in the nasopharynx of humans, but can spread from this sight to cause upper respiratory tract disease including middle ear infections in children. Evidence that two different pathotypes of *E. coli* capable of causing disease in humans also target CEACAMs, each utilizing a different adhesin to interact with CEACAMs in the urogenital or gastrointestinal tract (Berger *et al.*, 2004; Barnich *et al.*, 2007), has further expanded our understanding of the diversity of microbes that exploit CEACAMs during the infection process. It is remarkable that these evolutionary diverse pathogens are targeting CEACAMs as host receptors in the nasopharynx (*Neisseria* sp. *H influenzae*, *M. catarrhalis*), gut (AIEC, DAEC) and urogenital tract (DAEC, *N. gonorrhoeae*). This convergence is even more striking when considering these bacteria bind CEACAMs using adhesins that are both structurally and antigenically distinct, as detailed below.

### 1.1.1 Overview of CEACAM general architecture.

**A. N-terminal Domain:**

All CEACAM-specific bacterial adhesins characterized to date bind specifically to the receptor’s N-terminal domain. To understand the molecular basis of bacterial-mediated interactions with CEACAMs, it is important to understand the structure and residues in CEACAM targeted by the various adhesins. Considering that the IgV fold is the basis for the CEACAM N-domain, I will provide a brief overview of this structure.
Despite deriving its namesake from immunoglobulin molecules, the Ig-fold is one of the most prominent folds in nature. Its prevalence seems to stem from that fact that it provides an ideal scaffold that can be varied to accommodate a large array of binding interactions, as evident by the seemingly endless array of molecules that can specifically be targeted by antibodies. The Ig-fold is made up of two anti-parallel β-sheets that form a tight and compact structure based on the classical Greek key β-sheet (Barclay, 2003) (Fig. 1-2A). A typical Ig-like domain contains between seven and nine anti-parallel β-strands arranged into two sheets linked by a disulfide bond between strands B and F. IgV-like domains do not contain a disulfide bond, are more variable in sequence, and contain two extra strands (C’ and C’’ ) not present in typical IgC-like domains (Fig. 1-2A and 1-3A).

CEACAM IgV-like domains are differentially glycosylated, with the ABEG face (face comprised of strands A, B, E and G) being extensively glycosylated while the opposite face comprised of strands C, C’, C’’, F and G (CFG) face lacks any glycosylation. The unglycosylated CFG face has been shown to mediate CEACAM-CEACAM interactions (Markel et al., 2004) and also is the region of CEACAM targeted by most bacteria adhesins (Fig.1-2A). Consistent with the fact that the neisserial Opa proteins bind to CEACAM3 (see Fig. 1-1), the N-domains alone have been shown sufficient for binding to *Neisseria meningitis* Opa proteins (Virji et al., 1996b). Site-directed mutagenesis-based studies have yielded a variety of residues in the CFG face involved in both CEACAM-CEACAM binding and pathogen-mediated adherence. Antibodies that specifically target the CFG face of the N-domain block bacterial adherence, and highly conserved residues Tyr34 on the C’ strand and Ile91 on strand F
Fig. 1-2. CEACAM N-domains. A. Structure and organization of CEACAM N-domain β-strands. Strands that make up the CFG face are highlighted in blue (PDB:2GK2). B. Space filling model of CEACAM N-domains targeted by pathogens. Positive charged residues are coloured red, negative charged residues are coloured green, hydrophobic residues are black and polar residues are coloured blue. Conserved residues important to pathogen mediated adhesins are circled in white. Residues that contribute to the differential specificities are labeled in white. C. Multiple sequence alignment of CEACAM N-domains targeted by bacterial adhesins with residues important to differential CEACAM specificities indicated.
are crucial to Opa-mediated CEACAM binding (Virji et al., 1999). However, these residues also exist in CEACAMs (such as CEACAM8) that do not bind to bacterial adhesins, suggesting that are required but not sufficient to mediate binding (Fig. 1-2B).

A variety of residues in the CFG face have been shown to be important in binding by Opa proteins (Fig. 1-2B and C). Ser32 is conserved amongst CEACAMs and mutations result in disruption of binding by a number of Opa variants (Popp et al., 1999). Additionally, residues 27-29, which are at the edge of the C-strand, vary amongst CEACAMs and contribute to the specificity of each Opa protein for different CEACAMs (Popp et al., 1999) (Fig.1-2B and C). Other residues that play a role in CEACAM trans-dimerization, including L44 in CEACAM3 and CEACAM6 or Q44 in CEACAM1 and CEACAM5 (Markel et al., 2004), also contribute to the specificity of CEACAMs and affect the ability of Opa binding. Additionally residue 89 seems to play a role in Opa mediated adhesion, as some Opa variants show either a dramatic increase or decrease in binding dependent upon the residue substituted at this position (Villullas et al., 2007). By comparing the amino acid sequence and translating the differences to the structure of each N-domain it is clear that a select few amino acids with quite different physical properties are targeted by bacteria, some of which are at the core of binding while others mediate specificity (Fig. 1-2B and C).

**IgC domains:**

While the focus of my thesis is the interaction of bacterial adhesins with N-terminal IgV-like domain, I think that it is important to consider that differences in the number of IgC domains may also influence receptor accessibility and/or function. More than 50% of the mass of most CEACAMs is attributed to N-linked glycosylation. Other than the
Fig.1-3. CEACAM IgC and membrane anchors. A. IgC β-strand organization and differential glycosylation patterns. B. CEACAM anchors are either type I membrane spanning domains with cytoplasmic signaling motifs or GPI linked. (Adapted from http://www.carcinoembryonic-antigen.de/)
CEACAM3 and CEACAM4, which lack IgC domains, the IgC domains are the primary site of N-glycosylation in the individual CEACAMs. CEACAMs contain IgC type C2 domains, which have sequence similarity to IgV domains but are similar in size to IgC domains (i.e. lacking the C’ and C’’ strands) (Barclay, 2003). CEACAMs can have different versions of the Ig-C-like domains (A or B types) that differ in length (A 92-96 a.a. or B 86 a.a.) (Thompson et al., 1991), primary sequence, and glycosylation pattern (Fig. 1-3A). Glycosylation is amongst the most complicated type of post-translational modification, and can be controlled by numerous enzymatic reactions and pathways that differ between cell types. In general, glycosylation has been shown to be important for proper protein folding and can increase the solubility of proteins exposed to aqueous environments. N-glycosylation of proteins during biogenesis in the ER can affect localization to the apical surface of polarized cells (Potter et al., 2006). While most bacteria that bind CEACAMs recognize protein residues, it has recently emerged that the adherent and invasive E. coli instead target mannose residues linked to CEACAM, and this interaction is sufficient to mediate invasion into epithelia of the gastrointestinal tract (Barnich et al., 2007). While the role of the IgC in CEACAM function remains unclear, variation in the number of IgC domains and their subsequent glycosylation patterns in different cell types provide CEACAMs with high level of structural complexity.

C. Membrane anchors:

The extracellular Ig-like domains of CEACAMs are anchored to the cell membrane by either a typical type I transmembrane domain or by a glycosylphosphatidylinositol (GPI) anchor (Fig. 1-3B). While the Ig-like domains determine specificity of binding to extracellular ligands, the membrane linkage determines what type of intracellular signal is
initiated and, thereby, how the cell will respond. CEACAM1, CEACAM3 and CEACAM4 possess typical alpha-helical transmembrane domains that contain cytoplasmic tail with sequence motifs important in initiating intracellular signal cascades. The cytoplasmic domains of CEACAM1 and CEACAM3 can be differentially spliced into short isoforms, displaying truncated cytoplasmic tails. This highlights that functionally distinct isoforms of the same molecule may exist in different cell types depending on the cell-specific gene regulation of CEACAMs and the ratio of these isoforms may be important in response to CEACAM ligation. Most obvious of these are the immunotyrosine-based activation motifs (ITAM) present in the cytoplasmic domain of CEACAM3 and CEACAM4 and the immunotyrosine-based activation inhibitory motifs (ITIM), present in CEACAM1. As their names suggest, ITAM phosphorylation initiates a potent, Src family kinase-dependent signaling cascade that stimulates a response whereas ITIM motifs typically promote phosphatase-dependent inhibition of activation signals.

The basic structure of the GPI anchor is a phosphoethanolamine linker, a highly conserved glycan core and a phospholipid anchor. CEACAM5, CEACAM6, CEACAM7 and CEACAM8 are post-translationally linked to the ethanolamine linker based on a ‘GPI anchor sequence’ within the C-terminal region, which is cleaved prior to the addition of the anchor. While GPI anchors are often considered simple anchors, they are structurally complex molecules with the glycan core differentially modified with side chains, such as a phosphoethanolamine group, mannose, galactose, sialic acid, or other sugars (Paulick and Bertozzi, 2008). Depending on the protein and species of origin, the lipid anchor of the phosphoinositol ring is a diacylglycerol, an alkylacylglycerol, or a
ceramide (Paulick and Bertozzi, 2008). The lipid species can also vary in length, ranging from 14 to 28 carbons, and can be either saturated or unsaturated.

While the traditional function of GPI linkage was thought to be a stable membrane anchoring device, it is now thought that these differences in GPI anchors may directly contribute to a variety of processes. GPI-linked protein association with lipid rafts can also influence intracellular signaling by affecting the organization of membrane micro-domains. GPI-linked proteins are also thought to be capable of inducing signaling cascades, by physical association of the membrane anchor with neighboring transmembrane domains of signaling proteins. Interestingly, antibody cross-linking of certain GPI proteins can result in cellular activation or inhibition signals that are deficient when the respective GPI anchors are replaced with a transmembrane domain (Jones and Varela-Nieto, 1998) (Robinson, 1997). In fact, there is direct evidence that the CEACAM5 GPI anchor attached to a protein is critical to its function, since replacement of the GPI anchor sequence of NCAM with that of CEACAM5 resulted in an NCAM molecule with tumorigenic effects mimicking CEACAM5 while a chimera consisting of the CEACAM5 extracellular domain with the NCAM GPI anchor had no similar effect (Nicholson and Stanners, 2007).

With the variable nature of CEACAMs in mind, I will now discuss the physiological role of CEACAM family members that are targeted by the various bacterial adhesins (Fig.1-4).

1.1.2 CEACAM1

CEACAM1 (previously known as biliary glycoprotein, C-CAM and CD66a) is expressed on the apical side of most epithelial and endothelial cells and is also expressed
by leukocytes. CEACAM1 mediates cell-cell interactions by both trans-homotypic and -heterotypic binding with CEACAMs on the adjacent cells or other cell types (Hunter et al., 1996) through interactions between the N-domains (Markel et al., 2004). Interestingly, CEACAM1 can also engage in cis- (parallel) homo-dimerization (Hunter et al., 1996), however it remains unclear as to the domain(s) involved. CEACAM1 is the only CEACAM with orthologs in species as diverse as mice and fish. In humans, CEACAM1 is made up of 9 exons that can be spliced into 11 different splice variants, each of which contain the N-terminal IgV-like domain but may differ in the number (3 or 4) and type of IgC like domains as well as the length of the cytoplasmic tail. The cytoplasmic tail of full length CEACAM1 contains two ITIM motifs, whereas these are absent in the short isoforms.

CEACAM1 is found primarily as a cis-homodimer at the cell surface. This dimeric state is proposed to regulate the ability of Src-family kinases and the tyrosine phosphatase Src homology 2 (SH2)-domain containing phosphatases 1 and 2 (SHP1 and SHP2). Trans-ligation and resultant phosphorylation of CEACAM1’s ITIM results in the recruitment of the SHP1 and SHP2 leading to the dephosphorylation of Zap70 (Boulton and Gray-Owen, 2002;Lee et al., 2008). It has been suggested that CEACAM1 isoforms that lack the cytoplasmic ITIMs may induce activation signals (Chen et al., 2004), implying that the relative ratio of the long to short isoforms may play an important role in the type of signal induced upon CEACAM1 binding (Watt et al., 2001).

CEACAM1 in Immunity: The full length CEACAM1 molecule has been shown to play a co-inhibitory role in the activation and proliferation in a variety of cell types. This inhibition is contact dependant, mediated by homo/heterotypic trans-binding interactions
Fig. 1-4. CEACAM receptors targeted by bacterial adhesins with described tissue expression profiles. Adapted with permission from Macmillan Publishers Ltd: (Nature Reviews Immunology 6, 433-446), (June 2006).
between cells. In T-cells, CEACAM1 surface expression is low, however upon TCR stimulation or exposure to IL-2, CEACAM1 expression is strongly upregulated. Ligation of CEACAM1 leads to a SHP1-dependent inhibition of activating signals, blocking activation and proliferation of CD4+ T-cells (Boulton and Gray-Owen, 2002; Chen et al., 2004; Nagaishi et al., 2006; Lee et al., 2008). Additionally, this role of co-inhibition of activation can be seen on Natural killer (NK) cell function, as trans-ligation of CEACAM1 on the surface of activated NK-cells results in an inhibition of their cytotoxic effects (Markel et al., 2002). Interestingly, CEACAM1 knockout mice do not exhibit dramatic phenotypes. However, T-cells isolated from CEACAM1 -/- mice show defective proliferation and cytokine secretion in comparison to wild type mice (N. Beauchemin, personal communication).

**CEACAM1 and Growth Factor Signaling:** CEACAM1 has also been shown to regulate insulin metabolism as CEACAM1 phosphorylation is required for efficient insulin clearance and degradations (Poy et al., 2002). Dominant negative forms of CEACAM1 expressed in mice induced hyperinsulinemia from impaired insulin clearance (Poy et al., 2002). Insulin clearance does not appear to be mediated by cell adhesion as the N-domain of CEACAM1 is not required for insulin receptor mediated endocytosis (Soni et al., 1999). It is thought that CEACAM1 phosphorylation by the insulin receptor kinase promotes an indirect interaction between the insulin receptor and CEACAM1, which become internalized via clathrin-coated vesicles (Poy et al., 2002).

CEACAM1 has also been implicated to influence angiogenesis. Vascular endothelial growth factor (VEGF) upregulates CEACAM1 expression *in vitro* and ligation of CEACAM1 stimulates proliferation, chemotaxis and capillary-like formation
(Ergun et al., 2000; Wagener and Ergun, 2000). Expression of intact CEACAM1 in murine endothelial cells was required for the regulation of endothelial proliferation and invasion and additional endothelial expression of CEACAM1 enhanced vascular remodeling in vivo (Horst et al., 2006). CEACAM1 -/- mice have defective vascular remodeling in vivo, whereas endothelial overexpression of CEACAM1 in mice induced extensive vascular growth (Horst et al., 2006), suggesting that CEACAM1 plays a central role in neovascularization.

**CEACAM1 in Cancer:** Unlike the other CEACAMs, CEACAM1 is implicated as a tumour suppressor, involved in control of growth and cell cycle progression. CEACAM1 has been reported to be lost or down regulated in various cancers (Neumaier et al., 1993) while expression of CEACAM1 on tumour cells can inhibit tumour progression (Hsieh et al., 1995). CEACAM1 -/- mice do not seem to be susceptible to increased tumour formation, however tumours appear and progress more quickly in CEACAM -/- mice than in mice that express CEACAM1 (Leung et al., 2006). CEACAM1 has been implicated to induce apoptosis in epithelial cells and relies on the cytoplasmic ITIMs, presumably via SHP-1 and SHP-2 dependant control of cell growth and survival and reductions in CEACAM1 expression are associated with reduced apoptosis in early colon tumour lesions in vivo, and CEACAM1-mediated apoptosis (Nittka et al., 2004; Nittka et al., 2008). These studies demonstrate a variety of process where CEACAM1-dependent control can regulate cellular activation, division and differentiation, highlight its global importance in regulation of cellular activation.
1.1.3 CEACAM5

CEACAM5 (or CEA) was initially identified over 40 years ago as a prominent tumor marker up-regulated on numerous cancers in a variety of different tissues (Gold and Freedman, 1965). Despite the fact that its normal function remains unclear, CEACAM5 can mediate homophilic and heterophilic interactions between adjacent cells (Benchimol et al., 1989; Oikawa et al., 1989). CEACAM5 expression is typically limited to apical epithelial membranes in the gastrointestinal mucosa, nasopharynx and urogenital tract (Hammarstrom, 1999). Mice lack a CEACAM5 ortholog and do not contain any GPI-linked CEACAMs. Transgenic mice expressing levels and distribution patterns reflecting normal human CEACAM5 expression do not display any abnormal phenotypes (Chan and Stanners, 2004). However, over-expression of CEACAM5 in mice increases their susceptibility to oncogenic agents, consistent with its heightened expression in a variety of human tumours (Chan et al., 2006). These results suggest CEACAM5 contributes to the oncogenic process rather than simply being a marker for oncogenesis. This effect is undoubtedly related to that fact that over-expression of CEACAM5 inhibits cellular differentiation and prevents normal programmed cell death upon detachment of the cell from the extra-cellular matrix, a process termed anoikis (Ordóñez et al., 2000).

Curiously, CEACAM5 has recently been shown to also be over-expressed in the ileal mucosa of Crohn’s diseased patients, although its role in the progression of the disease remains unclear (Barnich et al., 2007).

CEACAM5 is post-translationally linked to a GPI anchor that directs it to specific lipid rafts or microdomains. As mentioned above, it has been shown that the GPI anchor signal of CEACAM5 plays an important role in its function as NCAM chimeric
proteins that express the CEACAM5-anchor mimic CEACAM5 tumorigenic activity (Nicholson and Stanners, 2007). CEACAM5 mediated inhibition of anoikis involves inactivation of caspase-9 and activation of the survival PI3K/Akt pathway and required the CEA-specific GPI anchor (Camacho-Leal and Stanners, 2008).

1.1.4 CEACAM6

CEACAM6 (or NCA) expression closely mimics that of CEACAM5 in terms of normal tissue distribution except that it is also expressed on neutrophils. CEACAM6 is also co-expressed with CEACAM5 on the apical surface of most epithelial cells and over-expressed in many cancers, including colorectal cancers (Jantscheff et al., 2003). Despite this broad distribution, the normal function of CEACAM6 and the reason for its vast tissue expression pattern remains a mystery. The CEACAM6 N-terminal IgV-like domain is linked to two IgC-like domains, and is also post-translationally linked to a GPI anchor, which appears functionally similar to that of CEACAM5 (Fig. 1-1). CEACAM6 is also involved in inhibition of anoikis (Duxbury et al., 2004) and is over expressed in the ileal mucosa of crohn’s diseased patients, facilitating bacterial mediated disease progression (Barnich et al., 2007).

1.1.5 CEACAM3

In contrast to the vast tissue distribution exhibited by some of the other CEACAM family members, CEACAM3 (previously known as Cgm1) is uniquely expressed on human neutrophils and appears to be restricted to humans, with no obvious orthologs even in non-human primates. CEACAM3 was first identified as the receptor on neutrophils responsible for opsonin-independent uptake of Opa+ Neisseria gonorrhoeae
(Chen and Gotschlich, 1996). CEACAM3 does not appear to function in cellular-adhesion, but is rather thought to act as an innate immune receptor that functions as a decoy to capture microbes that target other CEACAMs. Rather than benefiting the bacteria, CEACAM3 binding causes a Src family kinase-dependant activation of the phagocytes and engulfment and killing of the bound bacteria. As detailed below, this process closely resembles that which is normally elicited upon Ig-specific receptor binding to antibody coated bacteria except that in this case CEACAM3 directly interacts with bacterial surface.

### 1.2 Infection cycle

As I previously mentioned, a variety of distinct pathogens have now been recognized to utilize CEACAM at different stages of the infection cycle. While the routes of infection and the pathogenic process of each of these bacteria may be distinct, there are obvious parallels concerning the effect of CEACAM binding on colonization and disease. As such, it is useful to consider a generalized infection cycle whereby attachment to CEACAMs on the apical surface of mucosal epithelia leads to bacterial entry and cellular transcytosis through the epithelial cells to allow their emergence at the basolateral surface without disruption of the epithelial barrier (Fig. 1-5). From here bacteria may either persist at this site and proliferate in the sub-epithelial space or, less frequently, may disseminate through the bloodstream to distal sites. Regardless, the bacteria will encounter immune cells responding to the infection. In the following sections, I will discuss each of these cell types encountered during infection and highlight CEACAM’s role and the consequence of bacterial adhesion to these cell surface molecules in the context of each. Considering that most of work in this regard
has focused on the interaction between *Neisseria gonorrhoeae* and CEACAM receptors, this will be used as the paradigm with obvious difference indicated where they have been described.

### 1.2.1 Epithelial cells

The epithelia is an important host defense, acting as a mechanical barrier from external environmental factors. The epithelia are made up of polarized cells joined by a variety of intercellular adhesion molecules to create a tight barrier from the extracellular space. Disruption of this barrier leads to inflammation as foreign bodies infiltrate into sensitive host tissues. The apical side of the cell interacts with the extracellular environment while the basolateral side interacts with extracellular matrix and host immune defenses.

CEACAM1 has been shown to be expressed in most epithelial cells including cervical, nasopharyngeal and intestinal cells that are colonized by bacteria known to target CEACAMs. CEACAM5 and CEACAM6 are also expressed in gastrointestinal mucosal membranes and in epithelia in the nasopharynx, providing multiple receptor targets at these sites. Localization of CEACAMs to the apical side of polarized epithelial is important in the context of bacterial infections, since these receptors are exposed in intact tissues. Bacterial adhesins that specifically bind to either CEACAM1 or the GPI-linked CEACAM5 and CEACAM6 can mediate an indistinguishable invasion and transcellular transcytosis through polarized epithelial cells (Fig. 1-5) despite their different membrane anchors (McCaw *et al.*, 2004). Bacterial engulfment by these receptors occurs independent of actin rearrangements, does not require tyrosine phosphorylation and occurs in the absence of the CEACAM1 cytoplasmic tail,
Fig. 1-5. Neisseria infection cycle. Adapted with permission from Macmillan Publishers Ltd: (Nature Reviews Immunology 6, 433-446), (June 2006).
suggesting a zipper-like process whereby the bacteria ‘pull’ the host membrane around themselves until they are internalized. It has been recently demonstrated that CEACAM1 on epithelial cells may regulate TLR2 expression, highlighting a novel mechanism of bacterial immune evasion (Slevogt et al., 2008).

Invasive bacterial species that engage CEACAM and transcytose across the epithelial cell layer exit at the basolateral membrane and persist in the sub-epithelial space. While the focus of my thesis is the interaction of microbes with CEACAMs, it is important to note that most of the bacterial species that I will discuss in the following sections have many adhesins, some of which interact with various extracellular matrix (ECM) proteins. It is generally assumed that interaction with the ECM allows attachment and penetration into damaged epithelial cell layers, however it is interesting to speculate that it allows basolateral to apical trafficking permitting re-entry into the lumen and spread to the next host. Both the *Neisseria* sp. Opa proteins and *Moraxella catarrhalis* UspA1 proteins can also interact with extracellular matrix proteins such as fibronectin, laminin, vitronectin and are able to induce integrin-based signaling cascades (van Putten and Paul, 1995; Duensing and van Putten, 1997; Tan et al., 2005; Tan et al., 2006).

### 1.2.2 Neutrophils

Neutrophils are the first responders of the innate immune system and one of the most prominent cells types at the site of bacterial infection. Neutrophils recognize bacterial-derived product and efficiently trigger a bactericidal response. Massive infiltration of neutrophils at the site of infection is characteristic of gonococcal infections. Neutrophil activation in response to pathogenic *Neisseria* has been largely attributed to CEACAM3, as there is little response to *N. gonorrhoeae* that do not express CEACAM3-specific
adhesins. CEACAM3 binding by Opa-expressing gonococci leads to a Src family kinase-dependent phosphorylation of both tyrosines within the CEACAM3 ITAM (McCaw et al., 2003; Schmitter et al., 2004), actin re-arrangements (Billker et al., 2002; McCaw et al., 2003; Schmitter et al., 2004) and PI3K-mediated uptake into phagosomes and subsequent killing of the engulfed bacteria (Booth et al., 2003) via a process that involves neutrophil degranulation and oxidative burst (Sarantis and Gray-Owen, 2007) (H. Sarantis, personal communication). The devastating fate encountered by bacteria that engage CEACAM3 on neutrophils suggest that bacteria are not targeting this molecule in a beneficial manner, but rather that humans have evolved a mechanism to destroy pathogens that specifically adhere to CEACAM during the course of infection. CEACAM1 and CEACAM6 also are present on neutrophils. However, while bacterial engagement of these CEACAMs results in bacterial uptake, there is no degranulation or oxidative burst response. Instead, these receptors appear to facilitate neutrophil binding to immune-activated endothelial cells to promote neutrophil movement from the blood towards infected tissues.

1.2.3 Lymphocytes

Lymphocytes play an important role in the establishment of the acquired immune response against foreign organisms. These cells can be found at sites of infection and can also encounter bacteria or bacteria-derived products within the lymph nodes. T and B lymphocytes exclusively express CEACAM1 in the absence of other CEACAMs. While lymphocytes have been shown to express varying ratios of the different isoforms depending on the state of cellular activation, the contribution of short variants remains unclear. The level of surface CEACAM1 is low in resting cells, but is strongly
upregulated in response to cellular activation. CEACAM1 has been shown to play a role in modulation of immune activation with CEACAM1 ligation initiating an inhibitory signaling cascade in CD4+ T-cells. Bacterial engagement of CEACAM1 results in phosphorylation of its cytoplasmic ITIMs, which recruits the tyrosine phosphatases SHP1 and SHP2. These effectors prevent sustained phosphorylation of the CD3zeta-chain and ZAP-70 tyrosine kinase upon T-cell stimulation (Lee et al., 2008). This effect is induced by *N. gonorrhoeae* in a CEACAM-specific manner, and results in inhibition of CD4+ T-cell activation and proliferation in response to a variety of stimuli (Boulton and Gray-Owen, 2002). This interaction represents the first example of a bacterium arresting an immune response through the direct engagement of a co-inhibitor receptor. Moreover, this activity presumably contributes to the weak immune response to *N. gonorrhoeae*, which allows re-infections due to the apparent lack of immunological memory.

### 1.2.4 Dendritic cells

During the course of infection bacteria can also come into contact with dendritic cells (DC), the primary professional antigen presenting cells of the immune system. Immature DCs are found in tissues exposed to the external environment and can also be found in the blood. Immature DCs display highly phagocytic activity and are constantly sampling their surrounding environment for pathogens such as bacteria and virus. Exposure to microbial-derived products results in activation and maturation of DCs, resulting in their expression of various surface glycoproteins that act as T-cell co-receptors, and simultaneous migration to the lymphoid tissues where they can present antigen to naïve or memory T-cells. Immature DCs express CEACAM1 and Opa-dependent binding to
this receptor inhibits the normal maturation process, causing a significant reduction in the DCs ability to stimulate a T cell response (Q. Yu, E. Chow, M. Ostrowski and S. Gray-Owen, personal communication). Direct inhibition of maturation of professional antigen presenting cells is an exciting idea, but research into bacteria engagement of CEACAM1 on the DCs is still in its infancy.

1.2.5 Endothelial cells

The accumulation of bacteria systemically can induce in endotoxic shock and can lead to bacterial meningitis through trafficking and infiltration of the inflammation of the meninges. Both invasive encapsulated *N. meningitis* and typeable *H. influenzae* can cause systemic disease including bacteremia and meningitis, however the means by which they gain access into the vasculature remains poorly defined. Interesting in this regard is the observation that neisserial infection induces an NF-kappa B-dependent expression of CEACAM1 by primary endothelial cells (Muenzner et al., 2001), apparently inducing receptor on cells that otherwise lack any CEACAM receptor. However, considering that certain endothelial cells express CEACAM1 *in vivo*, it remains unclear whether the induced expression of CEACAM1 facilitates heightened entry into the bloodstream, association with the endothelia and/or extravascularization in a manner that effects disease.
1.3 CEACAM binding pathogens.

1.3.1 Neisseria sp.

A. General characteristics:

*Neisseria* species include causative agents of gonorrhea and meningitis, and also contain various species that make up the commensal flora. Despite presenting as vastly different clinical manifestations, these organisms share a close evolutionary relationship.

*Neisseria* sp. are Gram negative diplococci that are human restricted microbes. These organisms are uniquely adapted to thrive within humans and, as such, have small genomes encoding limited metabolic capabilities.

B. *N. gonorrhoeae* pathogenesis:

*N. gonorrhoeae* infections are the second most common sexually infection, behind chlamydia, with the World Health Organization reporting over 62 million cases worldwide. Despite being treatable with antibiotics, gonorrhea infections are on the rise in the US and Canada, with an increase in antibiotic resistant strains. In men, gonorrhea symptoms typically occur 2-5 days after contact, and present as acute urethritis. Women typically are asymptomatic, or only present minor symptoms, however bacterial spread to the upper genital tract causes pelvic inflammatory disease (PID) caused by immune-mediated scarring of the reproductive tract and may lead to sterility. Disseminated gonococcal infection is a relatively rare occurrence, estimated to develop in only 0.5-3% of untreated infections.

Pathogenic *N. gonorrhoeae* primarily infect the urogenital mucosa following sexual contact, but can also infect anorectal mucosa and cases of gonococcal pharyngitis have been described. The mechanisms involved in pathogenesis and
virulence of *Neisseria* involve an initial attachment to host mucosal surfaces through a type IV pilus. Urethral infection of male volunteers suggests expression of pilus is important the initial stages of bacterial colonization (Jerse *et al.*, 1994). Once the pilus retracts to allow close contact with the mucosal cell, neisserial Opa proteins can mediate tight interaction with host cellular CEACAM receptors.

**C. *N. meningitidis* pathogenesis:**

*Neisseria meningitidis* is one of the leading etiological agents of bacterial meningitis, however it resides in the nasopharynx of the majority of people without incident. The progression to disease state correlates with an invasive phenotype. A major factor in the virulence and progression of *N. meningitidis* is the presence of a polysaccharide capsule, which allows resistance to phagocytosis and complement mediated killing. The capsule also contributes to the organism’s ability to spread between hosts via an aerosol route.

The development of effective capsular-based vaccines A, C, Y and W135 in the early 1990s has lead to the virtual eradication of these strains from western societies as they were implemented, although they still contribute to significant morbidity and mortality in third world countries. A capsule-based approach will not work for serogroup B strains since this carbohydrate possesses cross-reactivity with human neural cell adhesion molecules (NCAM). Considering that serogroup B strains are the primary cause of infectious meningitis outbreaks in industrialized nations, a vaccine that protects against these strains remains a major goal of industry and public health efforts.
D. Neisserial Opa proteins:

Opa proteins were first identified in the late 1970s by virtue of the fact that their phase-variable expression resulted in opaque *N. gonorrhoeae* colonies (Swanson, 1978). During the 1980s, it became clear that Opa proteins played important functions in attachment to epithelial cells and appeared to be a primary mediator if interactions with neutrophils. In the mid-1990s, it was revealed that these interactions were mediated by Opa binding to CEACAMs (Chen and Gotschlich, 1996; Virji *et al.*, 1996a; Gray-Owen *et al.*, 1997a).

A single *N. gonorrhoeae* bacterium possesses up to 11 different *opa* alleles that are independently regulated and antigenically distinct, while *N. meningitidis* typically possess 3 or 4 *opa* alleles. Commensal *Neisseria* *sp.* encode 1 or 2 *opa* alleles that can adhere to CEACAM receptors, but can be out competed in vitro by pathogenic strains (Toleman *et al.*, 2001). Opa proteins are thought to play a primary role in the pathogenicity of infection, since gonococci recovered after natural urogenital, cervical or rectal infections typically express at least one Opa protein (Swanson *et al.*, 1988). The importance of the Opa-CEACAM interactions is evident by the fact that over 95% of gonococcal clinical isolates express Opa proteins that bind to CEACAM, despite the fact that Opa protein expression may phase variably switch on or off *in vitro* (H. Wong, personal communication).
Fig. 1-6. Neisserial Opa proteins. Variable regions are labeled as semi-variable (SV), hypervariable region 1 or 2 (HV-1 or HV-2), while the fourth loop is highly conserved (CL). A. Predicted secondary and tertiary structures of Opa proteins. B. Multiple sequence alignment of *Neisseria gonorrhoeae* MS11 Opa proteins. C. CEACAM receptor specificities of *Neisseria gonorrhoeae* MS11 Opa proteins.
Opa’s are integral outer membrane proteins with a predicted structure containing four surface-exposed loops (Malorny et al., 1998). The secondary structure was predicted through rules derived from porin crystal structures and the demonstration that highly variable sequences are present at the cell surface while regions of high sequence identity lie within transmembrane β-strands (Fig.1-6A) (Bhat et al., 1991). Further evidence for this predicted structure is attributed to the solved crystal structure of another Neisseria outer membrane protein, NspA, which shares an extremely high degree of sequence conservation with the transmembrane regions of Opa, allowing me to model Opa protein structure (Fig. 1-6A) (Vandeputte-Rutten et al., 2003). The surface-exposed loops primarily contain regions of high amino acid sequence variability with loop 1 displaying a semi-variable region (SV) while loops 2 (HV-1) and 3 (HV-2) contain hyper-variable regions (Fig. 1-6A and B). The fourth surface exposed loop is the shortest and is highly conserved. Each loop is likely flexible in nature and one or more may directly interact with the CEACAMs.

Opa proteins can specifically target the N-terminal IgV-like domain of CEACAM1, CEACAM3, CEACAM5 and CEACAM6. Individual Opa protein variants can bind to different combinations of these four CEACAMs (Fig. 1-6C). Throughout my thesis I will use the term ‘receptor specificity’ when describing the unique combination of CEACAMs targeted by a specific Opa variant. While the residues on individual CEACAM receptors that bind Opa proteins have largely been resolved, the Opa protein structures that mediate Opa-CEACAM binding remain elusive, in part due to the extreme sequence diversity in the surface loops of Opa proteins. The difficulty in understanding these interactions is evident from the fact that Opa variants that bind to
the same combination of CEACAM receptors often have little similarity by linear sequence comparison. Two different groups have tried to address this problem. Bos et al. (2002) suggested that distinct combinations of HV loops are required for Opa to adhere to the CEACAMs, a conclusion based upon the fact that chimeric Opa variants expressing certain combinations of HV1 and HV2 loops lost their ability to bind CEACAMs, although some combinations did to retain Opa’s binding ability (Bos et al., 2002). Separate site-directed mutagenesis-based studies targeted conserved residues in the HV1 and HV2 loops of functionally distinct Opa proteins. While most of these mutations resulted in a loss of binding function, certain mutations seemed to differentially affect Opa binding to CEACAM1 or CEACAM5 (de Jonge et al., 2003). This result is consistent with at least some difference in the contact points between Opa proteins and individual CEACAM receptors.

While the majority of Opa variants (i.e. 10 of 11 encoded by N. gonorrhoeae strain MS11) bind to CEACAM receptors, the remaining have been shown to bind heparan sulfate proteoglycan (HSPG) receptors (van Putten and Paul, 1995) and can also bind to vitronectin (VN) inducing bacterial uptake by bridging αvβ3 or αvβ5 integrins (van Putten and Paul, 1995; Gomez-Duarte et al., 1997). In contrast to the apical expression of CEACAM receptors, HSPG-containing cellular receptors predominate on the basolateral surface of polarized epithelia. This suggests that HSPG and/or VN binding may be important in basolateral to apical trafficking of the bacteria, allowing them to re-enter the lumen to infect the next host, although functions in other tissues during the course of infection cannot be ruled out.
1.3.2 *Moraxella catarrhalis*

**A. General characteristics:**

*Moraxella catarrhalis* is a Gram-negative, unencapsulated bacterium and is an obligate parasite of humans that can cause disease in both the upper and lower respiratory tract. Previously considered to be a commensal organism, it is now recognized as an important mucosal surface pathogen. In the upper respiratory tract, *M. catarrhalis* is responsible for cases of sinusitis and is the third leading cause of otitis media in infants after only non-typeable *Haemophilus influenzae* and *Streptococcus pneumoniae* (Christensen, 1999; Faden *et al.*, 1992; Karalus and Campagnari, 2000; Murphy, 1996). In the lower respiratory tract, *M. catarrhalis* can cause respiratory infections in adults with chronic obstructive pulmonary disease (COPD) resulting in increased morbidity and mortality of these patients (Murphy and Sethi, 1997; Murphy and Sethi, 1992). *M. catarrhalis* is thought to be responsible for 2-4 million infectious exacerbations of COPD in the U.S. each year (Murphy, 2005), and there is significant cost in treating disease related to *M. catarrhalis*, complicated in part by a rise in antibiotic-resistant strains (Bandak *et al.*, 2001; Turnak *et al.*, 2001). This makes development of a vaccine an exciting and important goal.

**B. *M. catarrhalis* pathogenesis:**

The mechanisms involved in pathogenesis and virulence of *M. catarrhalis* remain poorly defined. Attachment to host mucosal surfaces is an important step in colonization. Recently, *M. catarrhalis* genes encoding the type IV pilus were identified and characterized (Luke *et al.*, 2004), leading to speculation that this structure mediates
initial attachment to mucosal epithelia in a manner similar to that of the closely related pathogenic *Neisseria* sp. (Craig *et al.*, 2004). *M. catarrhalis* can also interact with the human cell lines through a variety of bacterial adhesins that include the outer membrane protein CD (Holm *et al.*, 2004) and Hag (Holm *et al.*, 2003) adhesins. Another adhesin, UspA1 promotes adherence to Chang human conjunctiva-derived epithelial cells (Lafontaine *et al.*, 2000), and was recently shown to mediate adherence via the host cellular receptor CEACAM1 (Hill and Virji, 2003). As such, this protein will be discussed in detail below.

**C. *M. catarrhalis* UspA proteins:**

Perhaps the most intensively studied potential virulence factors of *M. catarrhalis* are the non-fimbrial ubiquitous surface protein A (UspA) molecules, initially described about fourteen years ago. At first, these two proteins were thought to be a single gene product (Helminen *et al.*, 1994; Klingman and Murphy, 1994) until it was discovered that two genes encode distinct proteins sharing some epitopes (Aebi *et al.*, 1998b). UspA1 and UspA2 are now known to be encoded by ORFs located far apart in the *M. catarrhalis* genome (Wang *et al.*, 2007).

The UspA proteins are subdivided into three basic groups based on conserved amino acids motifs within the N- and C-termini of these macromolecules: UspA1 (~88 kDa) (Aebi *et al.*, 1997; Cope *et al.*, 1999), UspA2 (~62 kDa) (Aebi *et al.*, 1997; Cope *et al.*, 1999), and the hybrid UspA2H (~92 kDa) (Lafontaine *et al.*, 2000). The UspA proteins are homologues of the YadA protein of *Yersina* sp. (Koretke *et al.*, 2006; Hoiczyk *et al.*, 2000) and, as such, are predicted to be autotransporter proteins (Henderson *et al.*, 2004) that are present on the bacterial cell surface. Each protein
consists of three distinct structural domains (Hoiczyk et al., 2000; Koretke et al., 2006). The N-terminal region is considered to form a β-sheet-based globular head, while the C-terminal region is predicted to form a membrane spanning β-barrel structure (Koretke et al., 2006) (Table.1-1). The stalk region that connects the head to the membrane-anchoring domains consists of a coiled-coil structure. Their large size and extended structure can be visualized by transmission electron microscopy as a ‘fuzzy’ layer surrounding the bacterial cell (Pearson et al., 2002; Hoiczyk et al., 2000).

Early studies of *M. catarrhalis uspA1* and *uspA2* mutants indicated that the former protein had adhesive properties and functioned to bind *M. catarrhalis* to Chang conjunctival epithelial cells *in vitro* whereas expression of UspA2 was essential for *M. catarrhalis* to resist the bactericidal activity of normal human serum (Aebi et al., 1998b). Subsequent studies have shown that both of these proteins have the potential to express multiple different functional activities, ranging from attachment to host gene products to affecting biofilm development by some strains of *M. catarrhalis* growing *in vitro* (Pearson et al., 2006).

UspA1 and UspA2 have been shown to interact with a variety of human-derived targets normally found on certain eukaryotic cell surfaces, or extracellular matrix components such as fibronectin (Tan et al., 2005) and laminin (Tan et al., 2006). In addition to its role as an adhesin, UspA1 has been reported to confer serum resistance by interacting with both C3 (Nordstrom et al., 2005) and C4b (Nordstrom et al., 2004) components of the complement cascade. UspA2 also binds the serum protein vitronectin and apparently uses this mechanism to help some *M. catarrhalis* strains resist killing by normal human serum (Attia et al., 2006; Attia et al., 2005).
Interestingly, it has been demonstrated that UspA1 allows cellular adherence via CEACAM1 (Hill and Virji, 2003), and was recently shown to be associated with a CEACAM1-dependent development of apoptosis in epithelial cells (N'Guessan et al., 2007b). Remarkably, the interaction with each of these host proteins has been localized to different regions of UspA1, suggesting a potential for this bacterial macromolecule to simultaneously co-ordinate multiple ligands, although the molecular requirements for each of these functions has largely remained unclear.

Throughout my thesis I will describe UspA protein variants from a variety of strains. The nomenclature I will use to describe each variant will be the name of the protein (UspA1, UspA2 or UspA2H) followed by the strain name in subscript (e.g. UspA1O35E).

1.3.3 Haemophilus influenzae

A. H. influenzae pathogenesis:

*Haemophilus influenzae* is also a human-restricted bacterium that resides in the nasopharynx and characterized as a member of the commensal flora. Encapsulated *H. influenzae* serotype b had been the primary cause of bacterial meningitis in children, causing significant death and morbidity in children prior to the success of the Hib vaccine introduced in the early 1990s. Unencapsulated, or non-typeable, *H. influenzae* remains an opportunistic pathogen that is a major etiologic agent of childhood otitis media and is often associated with mucosal infections of patients with respiratory disease. The disease course is characterized by initial colonization of the mucosal
Table 1-1. Summary of CEACAM binding adhesins. Predictive tertiary structures of Opa and P5 were generated using 3D-prediciton (ESyPred3D) (Lambert et al., 2002). UspA1 structure was modified from predicted structure of YadA non-fimbrial adhesion (Hoiczyk et al., 2000;Koretke et al., 2006). Diffusely adhering E. coli Dr adhesin is modified based on the structural model for the complex between the Dr adhesins and CEACAM5(Korotkova et al., 2008). CEACAM receptor specificity of each adhesin is indicated where known.

<table>
<thead>
<tr>
<th>Pathogen</th>
<th>Disease</th>
<th>Adhesin</th>
<th>CEACAM1</th>
<th>CEACAM3</th>
<th>CEACAM5</th>
<th>CEACAM6</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>N. gonorrhoeae</em>&lt;br&gt;<em>N. meningitidis</em></td>
<td>Gonorrhea&lt;br&gt;Meningitis</td>
<td>Opa - Integral membrane protein</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td><em>M. catarrhalis</em></td>
<td>Otitis Media&lt;br&gt;COPD</td>
<td>UspA1 Trimeric autotransporter</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>ND</td>
</tr>
<tr>
<td><em>H. influenzae</em>&lt;br&gt;Typeable Hif</td>
<td>Otitis Media&lt;br&gt;Meningitis</td>
<td>P5 - Integral membrane protein</td>
<td>+</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td><em>E. coli</em>&lt;br&gt;DAEC</td>
<td>Diarrhea Urinary tract infections</td>
<td>Dr/afa fimbrial or afimbrial surface adhesin</td>
<td>+</td>
<td>ND</td>
<td>ND</td>
<td>+</td>
</tr>
<tr>
<td><em>EAEC</em></td>
<td>Diarrheal to dysentery</td>
<td>Type I pilus</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>+</td>
</tr>
</tbody>
</table>
membranes in the nasopharynx, with the infection moving up the eustachian tube towards the inner ear or into the lungs of patients with respiratory illnesses.

While the polysaccharide capsule is a major virulence factor in the course of invasive disease, the factors associated with disease in opportunistic infections has been more elusive. Many adhesins have been identified and characterized as playing roles in colonization and immune evasion. Of these the outer membrane protein P5 has been shown essential for colonization and middle ear infections in chinchillas, and is an effective vaccine in the chinchilla model of otitis media. More recently, P5 was demonstrated to specifically interact with CEACAM1 (Hill et al., 2001).

**B. H. influenzae P5 proteins:**

P5 proteins are predicted to form a β-barrel in the outer membrane in a manner similar to neisserial Opa proteins. Interestingly, despite adopting a similar structure, P5 proteins share no similarity to Opa proteins, even within Opa protein’s highly conserved membrane spanning regions (Table. 1-1), and are highly conserved among strains. Although P5 has been shown to mediate bacterial attachment by interaction with host cell CEACAMs (Hill et al., 2001), some P5 deficient strains were reported to be still able to adhere to CEACAM, suggesting a potential secondary CEACAM adhesin (Hill et al., 2001).

The interaction with CEACAM has been shown to play an important role in the *H. influenzae* infections *in vivo*. Chinchillas infected with wild type (P5-expressing) *H. influenzae* progress to develop otitis media, with bacteria forming a biofilm moving up the eustachian tube and into the middle ear. However, pre-treatment of the nasopharyngeal area with anti-CEACAM antibodies result in an inhibition of bacterial
colonization and subsequent infection (Bookwalter et al., 2008). Considering that no models for *Neisseria* or *M. catarrhalis* exist, this represents the first *in vivo* demonstration of a role for CEACAM binding during infection.

### 1.3.4 Pathogenic *E. coli*

#### A. General characteristics:

Pathogenic *E. coli* are represented by a diverse group of strains that have been divided into six major categories based on both virulence factors and the clinical disease symptoms. The six major groups responsible for diarrheal disease are enterotoxigenic (ETEC), enteropathogenic (EPEC), enterohemorrhagic (EHEC), enteroaggregative (EAEC), enteroinvasive (EIEC) and diffusely adhering (DAEC) *E. coli*. Urinary tract infections are mediated by uropathogenic *E. coli* (UPEC). A more recently recognized *E. coli* pathotype has been identified and termed adherent invasive *E. coli* (AIEC).

The differences in the virulence factors expressed by these different groups undoubtedly contribute to their different types of disease. As with most infections, the adhesins expressed define the tissue tropism. Interestingly most pathogenic *E. coli* strains possess a type I pili that is involved in host cell colonization by binding primarily to mannose-containing carbohydrates expressed on the cell surface. Recent studies have demonstrated that strains from two diverse *E. coli* groups adhere to host cells by binding to CEACAMs. Remarkably, the diffusely adhering (DAEC) and adherent-invasive (AIEC) *E. coli* express unrelated adhesins to bind CEACAMs.
**B. Diffusely adhering *E.coli* Dr/Afa adhesins:**

Dr/Afa adhesins are adhesins expressed by a subgroup of DAEC, but also by UPEC. While they share a similar basic structure, Dr adhesins are fimbrial, while Afa are afimbrial adhesins that do not polymerize the basic structural unit. DAEC express adhesins from the Dr/Afa family to target the proteinaceous CFG face on CEACAM N-domains (Berger *et al.*, 2004), indicating that binding may mimic CEACAM homo/heterotypic interactions. Numerous Dr/Afa structures have been solved and they interestingly display a typical Ig-like fold (Table. 1-1). While the family of Dr/Afa adhesin is extensive, only a specific subgroup of Dr/Afa family members including Dr, AfaE-III and f1845 have been demonstrated to bind CEACAM receptors.

**C. Adherent invasive *E.coli* Type I pilus:**

While all *E. coli* pathotypes express a type I pili that targets mannose, a specific pilus variant AIEC has been shown to bind CEACAM6 (Table. 1-1) (Barnich *et al.*, 2007). The prototypical AIEC strain LF82 appears to express a variant type I pilus that preferentially interacts with CEACAM6 in a mannose-dependant fashion. A provocative finding has been the recent observation that CEACAM5 and CEACAM6 are expressed in the ileum of patients with Crohn’s disease but not in normal individuals, and that this effect correlates with increased colonization of the Crohn’s tissues with AIEC (Barnich *et al.*, 2007). The causal role for CEACAM upregulation versus AIEC colonization remains to be worked out, and the contribution of each to Crohn’s disease must still be established.
1.4 Thesis Rationale

The focus of my thesis is to characterize the adhesins that mediate interaction with CEACAMs from *M. catarrhalis* and *Neisseria* sp. I have organized my thesis to initially focus on the identification of the molecular determinants that mediate adherence of *M. catarrhalis* to CEACAM (Chapters 2-4) followed by an investigation into the molecular basis of CEACAM binding by the hypervariable Opa proteins of pathogenic *Neisseria* (Chapter 5).


Author Contributions:

Michael Brooks, Dr. Scott Gray-Owen and Dr. Eric Hansen designed the experiments, analyzed research and wrote the paper.

Michael Brooks contributed Fig.2-1, Fig.2-2B, Fig.2-3 and Fig.2-4. Cassie Laurence contributed Fig.2-A and C.
2.1 Abstract

*Moraxella catarrhalis* is a human restricted pathogen that can cause respiratory tract infections. In this study, I identify a previously uncharacterized 24 kDa outer membrane protein with a high degree of similarity to *Neisseria* sp. Opa protein adhesins, with a predicted β-barrel structure consisting of 8 anti-parallel β-sheets with 4 surface-exposed loops. In striking contrast to the antigenically variable Opa proteins, the *M. catarrhalis* Opa-like protein (OlpA) is highly conserved and constitutively expressed, with 25 of 27 strains corresponding to a single variant. Protease treatment of intact bacteria and isolation of outer membrane vesicles confirm that the protein is surface exposed, yet does not bind host cellular receptors recognized by neisserial Opa proteins. Genome-based analyses indicate OlpA and Opa derive from a conserved family of proteins shared by a broad array of Gram negative bacteria.
2.2 Introduction

*Moraxella catarrhalis* is an obligate parasite of humans that can cause disease in both the upper and lower respiratory tract. In the upper respiratory tract, *M. catarrhalis* is responsible for cases of sinusitis and is the third leading cause of otitis media in infants after only *Haemophilus influenzae* and *Streptococcus pneumoniae* (Christensen, 1999; Faden *et al*., 1992; Karalus and Campagnari, 2000; Murphy, 1996). In the lower respiratory tract, *M. catarrhalis* can cause respiratory infections in adults with chronic obstructive pulmonary disease (COPD) resulting in increased morbidity and mortality of these patients (Murphy and Sethi, 1997; Murphy and Sethi, 1992). *M. catarrhalis* is thought to be responsible for 2-4 million infectious exacerbations of COPD in the U.S. each year (Murphy, 2005), and there is significant cost in treating disease related to *M. catarrhalis*, complicated in part by a rise in antibiotic resistant strains (Bandak *et al*., 2001; Turnak *et al*., 2001). This makes development of a vaccine an exciting and important goal.

The mechanisms involved in pathogenesis and virulence of *M. catarrhalis* remain poorly defined and at the onset of my work it had just become evident that *M. catarrhalis* was able to interact with CEACAM on the surface of transfected CHO cells (Hill and Virji, 2003). The data indicated that UspA1, a large multifunctional adhesin with a structure clearly different from both previously characterized CEACAM binding bacterial adhesins Opa and P5 mediated this interaction. This study peaked our interest and we began a collaboration with Dr. Eric Hansen’s group, a leading expert in both *M. catarrhalis* and UspA proteins.
A preliminary experiment conducted at the University of Texas Southwestern Medical Center showed that UspA1 cloned from the prototypical strain 035E, which was originally derived from a case otitis media, did not use CEACAM receptors to adhere to either chang or A549 cells. Considering the importance of the Opa family of outer membrane proteins in neisserial pathogenesis and the close phylogenetic relationship between Neisseria and Moraxella coupled with the fact that H. influenzae adhered to CEACAM using an integral membrane protein reminiscent of Opa proteins and that in addition to P5, had a second CEACAM binding adhesion (Hill et al., 2001), this lead me to examine whether homologues of neisserial Opa proteins existed in M. catarrhalis. A genome-wide screen of M. catarrhalis revealed a single gene with high sequence similarity to the neisserial Opa proteins. I characterized this previously unidentified 24 kDa protein I termed Opa-like protein A of M. catarrhalis (OlpA), and established it as an outer membrane protein that is highly conserved among clinical strains. I then investigated whether OlpA was capable of binding, and although it was clear that it was a homologue of Opa, OlpA lacked CEACAM binding activity. Broader analyses revealed that OlpA is a member of a highly conserved family of proteins conserved across a wide variety of Gram negative bacteria.

2.3 Results

2.3.1 Identification of OlpA.

Given the close relationship between M. catarrhalis and the Neisseria sp., and that UspA1<br>035E mediated its attachment to chang cells in a CEACAM-independent manner, I sought to determine whether proteins related to the neisserial Opa proteins are present in
*M. catarrhalis* and whether they were capable of binding to CEACAMs. A search of the *M. catarrhalis* ATCC43617 genome was conducted using Gene Tool software, by searching with a combination of neisserial Opa protein antigenic variants. The screen identified an open reading frame encoding a predicted protein with a high degree of sequence identity to neisserial Opa proteins. The predicted protein is 234 amino acids in length with a possible leader peptide of 30 residues. Oligonucleotide primers (Table 2-2) were designed based on this predicted ORF, and used to screen a variety of *M. catarrhalis* laboratory strains by PCR. A single fragment of ~1.6 kb in size was amplified from each of the strains and then sequenced. Genbank accession numbers for each sequence are listed in Table 2-2. The various *M. catarrhalis* alleles encode proteins displaying a remarkably high level of sequence identity, with the most obvious difference being that the strains 035E and 046E have a two amino acid (NT) insertion in the second predicted surface exposed region (Fig. 2-1A-B). Besides this insertion, the sequences were almost 100% identical to each other, yet diverge from the original sequence identified from the genome screen (compare ATCC43617, Fig. 2-1A). To distinguish the relatively high conservation of the *M. catarrhalis* proteins from the hyper-variable nature of the neisserial Opa protein, I termed the former as Opa-like protein A (OlpA)

### 2.3.2 OlpA is a putative outer membrane protein.

Neisserial Opa proteins and *H. influenzae* P5 are outer membrane proteins containing 8 anti-parallel β-sheets connected by short (1-2 residue) periplasmic loops and larger (10-25 residue) variable surface-exposed loops. In addition to a relationship with the Opa proteins, BLASTp analysis revealed that OlpA also shares homology with various other
Fig. 2-1. Moraxella catarrhalis OlpA. A. Alignment of OlpA proteins from M. catarrhalis strains ATTC4 3617, 035E, 046E, 4223, ATTC 25238,7169, and TTA37. Red indicates high (>90%) conservation; blue indicates >50% conservation. !, I or V; $, indicates L or M; %, F or Y; #, any one of N, D, Q, or E. Gene sequence accession numbers are listed in Table 2. B. Predicted 2D structure of OlpA. Residues depicted in filled circles with white lettering are identical in OlpA and Neisseria NspA; those in black are unique to OlpA. The two-residue insertion in the second extracellular loops of M. catarrhalis strains O35E and O46E are encircled in black and green. C. Predicted OlpA 3D structure. The structure was derived by threading over the crystal structure of the neisserial NspA.
outer membrane proteins, including the neisserial NspA. Comparison of linear sequences and their respective predicted secondary and tertiary structures suggest that OlpA also forms a β-barrel structure within the outer membrane protein (Fig. 2-1B). This allowed a model of OlpA structure to be created by threading its sequence over that of NspA (Fig. 2-1C), which has recently been crystallized (Vandeputte-Rutten et al., 2003). This analysis revealed that OlpA possesses significantly larger surface-exposed loops than NspA, making them more reminiscent of the neisserial Opa proteins. The few differences between the sequences obtained in this study all lay within regions predicted to be surface-exposed. Moreover, the strain ATCC43617 genomic OlpA sequence divergence also lies primarily within the predicted surface-exposed regions.

### 2.3.3 Expression of OlpA.

To monitor expression of OlpA, mouse polyclonal antisera to an OlpA-derived peptide was generated. Immunoblot analysis indicated that this antiserum detected a 24 kDa protein expressed by all *Moraxella* strains (Fig. 2-2A), except the genome-sequenced strain ATCC43617, which diverges within the sequence used in generation of the polyclonal antibody. A 1.6 kb PCR product containing the *olpA* gene from each *M. catarrhalis* strain was then cloned into the pCC1 copy control plasmid and expressed in *E. coli* strain EPI300, as outlined in the Materials and Methods. *E. coli* cell lysates were probed with the polyclonal OlpA-specific antiserum to confirm that the recombinant protein was expressed in *E. coli* (Fig. 2-2B). The OlpA protein was expressed in *E. coli* without any induction, indicating expression is likely constitutively driven by its native *M. catarrhalis* promoter (Fig. 2-2B and data not shown).
Fig. 2-2. Detection of OlpA protein by immunoblotting. Western blots of bacterial lysates were probed with the OlpA-specific antiserum. A. *M. catarrhalis* cell lysates. B. Cell lysates from recombinant *E. coli* strains expressing either OlpA cloned from the indicated *M. catarrhalis* strains or a kanamycin cassette (Kan) as a negative control. C. Immunoblot analysis to detect reactivity with *M. catarrhalis* strain O35E-specific OlpA antiserum with lysates prepared from the indicated strains. Molecular mass position markers (in kilodaltons) are shown on the left.
2.3.4 Conservation of OlpA among *M. catarrhalis* strains.

Considering that the cloned OlpA gene sequences were highly conserved amongst *M. catarrhalis* strains tested in this study, it is remarkable that the strain ATCC43617 genome sequence-derived allele is so distinct. Since the OlpA antiserum distinguishes between these two variants, *M. catarrhalis* strains from a variety of geographic locations were collected (Table. 2-1) and bacterial lysates probed by immunoblot. In total, 25 of the 27 strains tested during this study react with the *M. catarrhalis* antibody (Figs. 2-2A and 2-2C), with only strains ATCC43617 and V1156 displaying no cross-reactive OlpA. To understand the relationship between these two proteins, the V1156 allele was sequenced. As expected, the V1156 OlpA (Genbank Accession # DQ996464) is closely related to that of ATCC 43617, suggesting that these two represent an uncommon second variant of OlpA.

2.3.5 Surface accessibility and heat modifiability

The structural analyses suggested that OlpA is a surface-exposed outer membrane protein (Fig. 2-1B-C). Consistent with its specificity for a peptide sequence predicted to lie within the transmembrane sequence (Fig. 2-1B), the OlpA-specific antisera did not react with intact bacteria (data not shown). To localize the OlpA protein, *M. catarrhalis* strain 035E was treated with heat and EDTA to liberate outer membrane vesicles (OMVs) (Murphy and Loeb, 1989). Immunoblot analysis confirmed that the OMV preparations contained OlpA and other well-characterized outer membrane proteins, including CopB and UspA1 (Fig. 2-3A), while the abundant periplasmic ferric binding
Fig. 2-3. A. Detection of OlpA in EDTA-induced outer membrane vesicles of *M. catarrhalis*. OMVs prepared from either strain O35E or the USPA1-deficient O35E mutant (O35EΔU) were probed with specific antibodies to detect OlpA, USPA1, and CopB. B. Surface exposure of OlpA. Intact *E. coli* expressing either OlpA or the periplasmic MBP were subjected to increasing doses of trypsin and proteinase K. Lysates were resolved and probed for either OlpA or MBP, as indicated. C. Heat-modifiability of OlpA was detected by using SDS-PAGE and immunoblot analysis following incubation of bacterial cell lysates for 20 min at either 100°C or 23°C. Molecular mass position markers (in kilodaltons) are shown on the left.
protein (FbpA) was barely detectable (data not shown), consistent with an outer membrane localization of OlpA. To confirm exposure at the bacterial cell surface recombinant OlpA-expressing *E. coli* and *M. catarrhalis* were treated with proteinase K and then were subjected to immunoblot analysis. Protease treatment caused a dose-dependant degradation of OlpA without affecting the integrity of the periplasmic maltose binding protein (MbpA) or FbpA (Fig. 2-3B), confirming surface exposure of OlpA.

The β-barrel structure of many Gram negative outer membrane proteins, including the neisserial Opa proteins, tends to resist unfolding in SDS-containing sample buffer unless they are boiled. The electrophoretic mobility of OlpA was affected by heating (Fig. 2-3C), consistent with a β-barrel conformation. The heat-modifiable nature of recombinant OlpA expressed in *E. coli* is indistinguishable from that observed in the respective *M. catarrhalis* strains, suggesting that the expression and folding of OlpA in *E. coli* is comparable to its native form.

**2.3.6 *M. catarrhalis* OlpA does not bind to CEACAM receptors.**

The *M. catarrhalis* ATCC25238 UspA1 protein attaches to human CEACAM receptors (Hill and Virji, 2003), whereas the O35E UspA1 protein does not (Chapter 3). Given the similarity between OlpA and the neisserial Opa proteins, I sought to determine whether these Opa homologues might represent another mechanism by which *M. catarrhalis* binds to CEACAM and/or other host receptors. *E. coli* expressing the OlpA from strain 046E or ATCC25238 did not interact with either soluble CEACAM1 or CEACAM5 in either bacterial pull-down or solid phase binding assays, while *E. coli* expressing the *N. gonorrhoeae* Opa57 protein bound both proteins (data not shown). Moreover, a UspA1-deficient derivative of *M. catarrhalis* 035E does not bind to CEACAM-expressing cell
lines (Chapter 3) despite expressing OlpA (Figs. 2-2 and 2-3). Other Opa-like proteins, such as the *N. gonorrhoeae* Opa<sub>50</sub> (van Putten and Paul, 1995), the *N. meningitidis* OpcA (Virji *et al*., 1995) and the enterotoxigenic *E. coli* (ETEC) Tia proteins (Fleckenstein *et al*., 2002) bind HSPGs so I looked for heparin binding but I was also unable to detect any difference in heparin binding between the recombinant OlpA-expressing *E. coli* versus the parental strain. To ascertain whether any adhesin function could be attributed to OlpA in *M. catarrhalis*, an *olpA*-deficient strain of *M. catarrhalis* O35E was generated by insertional mutagenesis. *In vitro* infection assays to compare wild type, UspA1-deficient and OlpA-deficient *M. catarrhalis* 035E, as well as the OlpA-expressing versus parental (empty expression vector-containing) *E. coli* strains. These studies provided no evidence that OlpA facilitates bacterial association with Hela (human endocervical), A549 (human lung-derived), Caco-2 (human colon-derived) or Lec11 (Chinese hamster ovary-derived) cell lines, either in the presence or absence of serum.

**2.3.7 Homologues**

Considering the presence of the OlpA in *M. catarrhalis* and the presence of the closely related Opa and NspA proteins in *Neisseria* sp., I next sought to determine the phylogenetic distribution of Opa-like proteins via mining available genome sequences. Interestingly, *H. influenzae*, which also exclusively colonize the human upper respiratory tract and causes a localized and disseminated infection, encode a protein that has significant sequence identity to OlpA but otherwise remains uncharacterized (Accession #ZP_00154421; Fig. 4). Other respiratory pathogens, such as *Actinobacillus*
Fig. 2-4. Relationship between diverse Opa-like proteins. A. The neighbor-joining tree diagram depicted is based on ClustalW protein alignments using Phylodraw, as outlined in Materials and Methods. Listed are bacterial species, previously used names and functions of OlpA-related proteins, and sequence accession numbers. ETEC, enterotoxigenic E. coli. B. Sequence alignment of diverse Opa-like proteins representing the indicated bacterial species illustrated in the tree diagram (A). Predicted surface loops are indicated as L1 to L4, while transmembrane regions are indicated by TM1 to TM8. Amino acid sequence identities are indicated by red (>90%) or blue (>50%).
succinogenes and Pasteurella multocida, also appear to have highly related homologues (Fig. 2-4). Each of these Opa-like proteins is more closely related to the M. catarrhalis OlpA than they are to the neisserial Opa proteins (Fig. 2-4A). Related proteins are also evident in the Enterobacteriaceae, including a homologue present in enterotoxigenic E. coli (ETEC) that functions as an adhesin involved in pathogenesis (Fleckenstein et al., 2002).

### 2.3 Discussion

In searching for homologues of the well characterized Opa protein adhesins from the pathogenic Neisseria, I identified a previously unrecognized but highly conserved outer membrane protein in M. catarrhalis. This OlpA protein shares a high degree of similarity to neisserial Opa proteins, but I found no evidence that they share a similar function. While neisserial Opa proteins bind to CEACAM and/or HSPG receptors (Gray-Owen et al., 1997a; Virji et al., 1996a; van Putten and Paul, 1995), neither OlpA-expressing but UspA1-deficient M. catarrhalis nor the OlpA-expressing E. coli displayed any binding to either of these receptor types. In fact, I did not detect OlpA-mediated adherence to any of a variety of cell lines tested.

Sequence alignments and our experimental results expressed herein indicate that OlpA forms an eight stranded β-barrel in the M. catarrhalis outer membrane. This allowed us to predict 3D structure by threading over the closely related N. meningitidis NspA, which has been previously crystallized (Vandeputte-Rutten et al., 2003). Curiously, the OlpA proteins appear to have surface-exposed loops that are much larger than NspA, yet shorter than those found in the Opa adhesins. The similarity of OlpA to the well-characterized NspA and Opa proteins make it very likely that these structural
predictions provide an accurate depiction of OlpA in the membrane-spanning regions. However, these software cannot predict the tertiary structure within the surface-exposed loops.

While the search for a vaccine protecting against *M. catarrhalis* infection has lead to the characterization of other outer membrane proteins, this is the first description of OlpA. To ensure that the OlpA gene is expressed, an antibody was generated against a peptide in the predicted OlpA gene and detected expression in *M. catarrhalis* strains and recombinant *E. coli* in which the gene is being expressed from its native promoter. I also confirmed that OlpA was present within outer membrane fractions, has a heat-modifiable native characteristic of β barrel structures, and was surface exposed.

The level of conservation within OlpA genes sequenced was remarkably high, with only two different variants apparent among the 27 strains tested. Of note, these variants are highly conserved among membrane-spanning sequences, with significant deviation only apparent within the surface-exposed loops. However, 25 of 27 strains contained the variant typified by strain 035E, with >99% identity among the 6 sequenced alleles.

Considering the lack of CEACAM and HSPG binding and the relatively high conservation of OlpA as compared to neisserial Opa proteins, I wondered whether similar proteins exist in more distantly related bacterial species. Our survey of available genome sequences indicates that there are uncharacterized outer membrane proteins with a high degree of similarity to OlpA in a wide variety of species (Fig. 2-4). Alignments indicated that each of these is an outer membrane protein with 8 anti-parallel β-strands and four surfaced exposed loops, and some species contain homologues in 2 different
genomic loci. As with the *M. catarrhalis* OlpA variants, these homologues possess high sequence conservation within the transmembrane sequences, but vary in size and linear sequence within the surface-exposed loops. It will be interesting to establish whether the strict conservation of certain residues within the membrane-spanning sequences reveals residues that are required for proper folding or membrane insertion of the basic eight-stranded β-barrel structure versus those that are evolutionarily related.

While the function of OlpA remains unknown, their conservation across phylogenetically disparate bacterial species suggests a basic essential function. This, along with the fact that the *M. catarrhalis* OlpA protein is surface exposed and highly conserved, make it an obvious vaccine candidate. Finally, when considering the evolutionary relationship among the Opa and Opa-like proteins, it is enticing to speculate that evolution has taken advantage of the minimal β-barrel structure represented by proteins such as NspA to provide a basic framework on which host cell binding adhesins or proteins possessing other necessary functions can be derived.
2.4 Materials and Methods

**Bacterial strains and culture conditions.** The *M. catarrhalis* strains employed in this study are listed in Table 2-1. Several of these strains, including O35E (Aebi *et al.*, 1997), O46E (Lafontaine *et al.*, 2000), TTA37 (Lafontaine *et al.*, 2000), 7169 (Luke and Campagnari, 1999), and 4223 (Kyd *et al.*, 1998) have been described previously. For preparation of whole cell lysates, *M. catarrhalis* strains were grown at 37°C on brain heart infusion (Difco/Becton Dickinson, Sparks, MD) agar plates in an atmosphere of 5% CO₂ or in liquid broth. *E. coli* strains employed in this study were strain EPI300 with plasmids indicated in Table 2-2. *E. coli* strains were grown in Luria-Bertani (LB) agar plates or broth with standard antibiotic supplementation as required.

**General DNA methods.** *M. catarrhalis* genomic DNA was isolated from agar plate-grown cells using the Easy-DNA kit (Invitrogen, Carlsbad, CA) with minor modifications. The preparation of plasmid and purification of PCR products was performed using kits manufactured by QIAGEN (Santa Clarita, CA).

**DNA sequencing and analysis.** The *M. catarrhalis* strain ATCC 43617 genome sequence was previously published (NCBI patent number WO0078968). This sequenced genome was scanned using Gene Tool (http://www.biotools.com/products/genetool.html) in all ORFs to search for proteins with homology to neisserial Opa protein sequences. The identified sequence was submitted to BLASTP (Entrez-pubmed) to examine for homologous proteins. Where indicated, DNA fragments were sequenced by the York University DNA Sequencing Facility (http://www.biol.yorku.ca/cm/autoseq.htm) using the same primers. Sequences were submitted via the Entrez-Pubmed protein/nucleotide database, with relevant
accession numbers being listed in Table 2-2. Mult-Alin (Corpet, 1988) was used to generate sequence alignments. 2D and 3D modeling was performed using 3D-JIGSAW (Contreras-Moreira and Bates, 2002; Bates et al., 2001; Bates and Sternberg, 1999). N-J tree diagrams were designed using ClustalW (http://clustalw.genome.jp/) based sequence alignments and Phlyodraw v0.82.

**Cloning the gene.** PCR Primers Sgo114 and Sgo115 (Table 2-2) were designed using the putative *M. catarrhalis* OlpA gene identified on the basis of our analyses of the *M. catarrhalis* genome (NCBI patent number WO0078968). A PCR product was amplified using primers Sgo114 and Sgo115 from chromosomal DNA of strains 035E, 046E, ATCC 25238, 7169, 4223, and TTA37, and then cloned into the pCC1™ (Blunt Cloning-Ready) Vector (Epicentre Biotechnologies) to create pCC.035E, pCC.046E, pCC.25238, pCC.7169, pCC.4223, and pCC.TTA37, respectively. A kanamycin resistance cassette was obtained and cloned into pCC1 to generate the pCCKan control. Plasmids were electroporated into Electrocompetent TransforMax™ *E. coli* EPI300 as described by Epicentre CopyControl™ cDNA, Gene & PCR Cloning Kit to create OlpA-expressing *E. coli* strains. The periplasmic MBP-encoding plasmid pMAL was obtained from New England Biolabs and introduced into *E. coli* EPI300 by electroporation.

**Preparation of polyclonal OlpA antiserum.** The translated OlpA protein sequence from *M. catarrhalis* strain O35E was analyzed for hydrophilic sequence stretches using the Kyte-Doolittle method (Kyte and Doolittle, 1982). The peptide SNLEAKYNDNDKLEDK was selected and synthesized with a cysteine residue added to its N-terminus, by the Protein Chemistry Technology Center at UT
Southwestern Medical Center. This peptide was covalently coupled to Imject Maleimide-Activated Mariculture Keyhole Limpet Hemocyanin (Pierce, Rockford, IL). A 100 μg portion of the peptide-KLH conjugate was mixed 50:50 with Freund’s complete adjuvant (Difco, Detroit, MI) and used to immunize five BALB/c mice. Four weeks later, these mice received a booster immunization with 30 μg of this conjugate mixed 1:1 with Freund’s incomplete adjuvant. These animals were euthanized and exsanguinated approximately two weeks later, and serum was then prepared from this blood by standard methods.

**Western blot analysis.** For detection of the OlpA protein, whole cell lysates (Aebi et al., 1997) were resolved by SDS-PAGE in 12.5 % (wt/vol) polyacrylamide separating gels, transferred to Immobilon-P membranes (Millipore, Bedford, MA) and probed with the indicated primary antibody. OlpA was detected using the mouse polyclonal OlpA antisera described herein. UspA1 and CopB were detected using the monoclonal antibodies 24B5 (Cope et al., 1999) and 10F3 (Helminen et al., 1993), respectively. Maltose binding protein was detected using the MBP-specific monoclonal antibody (New England Biolabs). The FbpA-specific rabbit polyclonal antisera was generously provided by Dr. Anthony Schryvers (University of Calgary, Alberta, Canada). The secondary antibody used was goat anti-mouse IgG or anti-rabbit IgG conjugated to horseradish peroxidase (Jackson ImmunoResearch, West Grove, PA). Antigen-antibody complexes were visualized by Chemiluminescence, achieved via the use of the Western Lightning Chemiluminescence Reagent Plus (New England Nuclear, Boston, MA).

**Protease treatment of intact cells.** *M. catarrhalis* or *E. coli* cells were grown in broth overnight at 37°C. Cells were pelleted and washed three times in phosphate-buffered
saline (PBS) containing 10 mM MgCl$_2$ and 5 mM CaCl$_2$ (PBS-Mg/Ca) and re-suspended to a concentration of $1 \times 10^9$ cells/ml. Then, 250 µl of cells was incubated with increasing proteinase K (Pharmacia) before incubation at room temperature (RT) for 10 minutes. The reaction was stopped by adding 10 µl of a 10 mg/ml stock of phenylmethylsulfonyl fluoride (in isopropanol) to stop the reaction. The cells were then centrifuged and then resuspended in PBS-Mg/Ca twice, and the final washed pellet then boiled in SDS-PAGE sample buffer containing β-mercaptoethanol.

**Heat-modifiability.** Bacterial cultures were grown overnight at 37°C. Cells were then pelleted and re-suspended in SDS-PAGE sample buffer. Samples were then incubated at either RT or in a boiling water bath for 20 minutes before western blot analysis.

**Outer membrane vesicles.** Outer membrane vesicles of *M. catarrhalis* were isolated through heating in the presence of EDTA, as described by Murphy et al. (Murphy and Loeb, 1989). Briefly, 50 ml BHI broth was inoculated with *M. catarrhalis* and cultured overnight with shaking at 37°C. The cells were centrifuged at 10,000 × g for 15 minutes at 4°C and the resulting pellet then re-suspended in 0.05 M Na$_2$HPO$_4$, 0.15 M NaCl, 0.01 M EDTA, pH 7.4 before incubation at 56°C with shaking. The bacteria were removed by two consecutive spins of 10,000 × g for 15 minutes at 4°C, and the resulting supernatant then centrifuged at 50,000 × g for 90 minutes at 4°C to recover the outer membrane vesicles, which were then re-suspended in PBS.
<table>
<thead>
<tr>
<th>Strain</th>
<th>Source</th>
<th>Strain</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATCC43617</td>
<td>ATCC</td>
<td>FR3227</td>
<td>Richard J. Wallace</td>
</tr>
<tr>
<td>O35E</td>
<td>John D. Nelson</td>
<td>FR2336</td>
<td>Richard J. Wallace</td>
</tr>
<tr>
<td>O46E</td>
<td>John D. Nelson</td>
<td>FR2213</td>
<td>Richard J. Wallace</td>
</tr>
<tr>
<td>TTA37</td>
<td>Steven L. Berk</td>
<td>FIN2406</td>
<td>Merja Helminen</td>
</tr>
<tr>
<td>7169</td>
<td>Anthony A. Campagnari</td>
<td>FIN2344</td>
<td>Merja Helminen</td>
</tr>
<tr>
<td>ATCC25238</td>
<td>ATCC</td>
<td>FIN2341</td>
<td>Merja Helminen</td>
</tr>
<tr>
<td>4223</td>
<td>Timothy F. Murphy</td>
<td>FIN2265</td>
<td>Merja Helminen</td>
</tr>
<tr>
<td>V1171</td>
<td>Frederick Henderson</td>
<td>ETSU-26</td>
<td>Steven L. Berk</td>
</tr>
<tr>
<td>V1156</td>
<td>Frederick Henderson</td>
<td>ETSU-22</td>
<td>Steven L. Berk</td>
</tr>
<tr>
<td>V1145</td>
<td>Frederick Henderson</td>
<td>ETSU-17</td>
<td>Steven L. Berk</td>
</tr>
<tr>
<td>V1120</td>
<td>Frederick Henderson</td>
<td>ETSU-5</td>
<td>Steven L. Berk</td>
</tr>
<tr>
<td>V1118</td>
<td>Frederick Henderson</td>
<td>B59911</td>
<td>David Goldblatt</td>
</tr>
<tr>
<td>V1069</td>
<td>Frederick Henderson</td>
<td>B59504</td>
<td>David Goldblatt</td>
</tr>
<tr>
<td>P44</td>
<td>Steven L. Berk</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table. 2-1. *M. catarrhalis* strains used in this chapter.

<table>
<thead>
<tr>
<th><em>M. catarrhalis</em></th>
<th><em>olpA</em> Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>035E</td>
<td>DQ996458</td>
</tr>
<tr>
<td>046E</td>
<td>DQ996459</td>
</tr>
<tr>
<td>25238</td>
<td>DQ996460</td>
</tr>
<tr>
<td>7169</td>
<td>DQ996461</td>
</tr>
<tr>
<td>4223</td>
<td>DQ996462</td>
</tr>
<tr>
<td>TTA37</td>
<td>DQ996463</td>
</tr>
<tr>
<td>V1156</td>
<td>DQ996464</td>
</tr>
</tbody>
</table>

Primers

<table>
<thead>
<tr>
<th>Primers</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sgo114</td>
<td>AACCCAATGCCACAAGGAC</td>
</tr>
<tr>
<td>Sgo115</td>
<td>AGCAATTGGACTGTGTTGG</td>
</tr>
</tbody>
</table>

Table. 2-2. Accession numbers of *olpA* genes from strains used for cloning.
Chapter 3: *Moraxella catarrhalis* binding to host cellular receptors is mediated by sequence-specific determinants not conserved between all UspA1 protein variants

A version of this chapter has been previously published:
*Moraxella catarrhalis* binding to host cellular receptors is mediated by sequence-specific determinants not conserved among all UspA1 protein variants. Infection and Immunity, 2008 Nov;76(11):5322-9, doi:10.1128/IAI.00572-08 (reproduced/amended with permission from American Society for Microbiology)

Author Contributions:
Michael Brooks, Dr. Scott Gray-Owen and Dr. Eric Hansen designed the experiments, analyzed research and wrote the paper.

Michael Brooks contributed Fig.3-2, Fig.3-3 and Fig.3-4A and B. Jennifer Sedillo contributed Fig.3-1 and Fig.3-4C.

Henry Wong, Cassie Laurence and Nikki Wagner performed preliminary experiments involved in the initiation of the project.

3.1 Abstract

The *Moraxella catarrhalis* ubiquitous surface proteins (UspAs) are autotransporter molecules reported to interact with a variety of different host proteins and to affect processes ranging from serum resistance to cellular adhesion. The role of UspA1 as an adhesin has been confirmed with a number of different human cell types, and is mediated by binding to eukaryotic proteins including Carcinoembryonic Antigen-related Cellular Adhesion Molecules (CEACAMs), fibronectin, and laminin. A distinct difference in the ability of prototypical *M. catarrhalis* strains to adhere to CEACAM-expressing cell lines prompted us to perform strain-specific, structure-function analysis of UspA1 proteins. In this study, I characterized CEACAM binding by a diverse set of UspA1 proteins and showed that 3 out of 10 UspA1 proteins were incapable of binding CEACAM. This difference resulted from the absence of a distinct CEACAM binding motif in the non-adhering strains. Our sequence analysis also revealed a single *M. catarrhalis* isolate that lacked the fibronectin-binding motif and was defective in adherence to Chang conjunctival epithelial cells. These results clearly demonstrate that UspA1-associated adhesive functions are not universally conserved. Instead, UspA1 proteins must be considered as variants with the potential to confer both different cell tropism and host cell responses.
3.2 Introduction

As indicated in the previous chapter, it was clear that OlpA did not mediate attachment to CEACAMs, but it still remained unclear why UspA1_035E did not bind to CEACAM. As I previously mentioned (Chapter 1), the UspA proteins are subdivided into three basic groups based on conserved amino acids motifs within the N- and C-termini of these macromolecules: UspA1 (~88 kDa) (Aebi et al., 1997; Cope et al., 1999), UspA2 (~62 kDa) (Aebi et al., 1997; Cope et al., 1999), and the hybrid UspA2H (~92 kDa) (Lafontaine et al., 2000). The UspA proteins are homologues of the YadA protein of Yersina sp. (Koretke et al., 2006; Hoiczyk et al., 2000) and, as such, are predicted to be autotransporter proteins (Henderson et al., 2004) that are present on the bacterial cell surface. Each protein consists of three distinct structural domains (Hoiczyk et al., 2000; Koretke et al., 2006): the N-terminal region is considered to form a β-sheet-based globular head, the C-terminal region is predicted to form a membrane spanning β-barrel structure (Koretke et al., 2006), and the stalk region that connects the head to the membrane-anchoring domains consists of a coiled-coil structure.

UspA1 and UspA2 have been shown to interact with a variety of human-derived targets normally found on certain eukaryotic cell surfaces or in the extracellular matrix including fibronectin (Tan et al., 2005), laminin (Tan et al., 2006), and CEACAM1 (Hill and Virji, 2003). Each binding function contributes to a varying degree to M. catarrhalis interactions with different cell types (Lafontaine et al., 2000; McMichael et al., 1998; Hill and Virji, 2003). In addition to its role as an adhesin, UspA1 has been reported to confer serum resistance by interacting with both C3 (Nordstrom et al., 2005) and C4b (Nordstrom et al., 2004) components of the complement cascade. Most recently, UspA1
was shown to be associated with the onset of apoptosis in epithelial cells exposed to *M. catarrhalis* (N'Guessan et al., 2007b) and to be involved in invasion of human epithelial cells *in vitro* (Spaniol et al., 2007; Slevogt et al., 2007). Remarkably, the interaction with each of these host proteins has been localized to a different region of UspA1, suggesting a potential for this bacterial macromolecule to simultaneously co-ordinate multiple ligands. However, the studies that have attributed each individual phenotype to UspA1 have typically been performed using different strains and each binding activity is frequently considered in isolation, making it difficult to reconcile sometimes contradictory observations regarding UspA1 function. This prompted us to perform a functional analysis with a diverse set of *M. catarrhalis* isolates, thereby allowing us to demonstrate a previously unappreciated link between the variability in UspA1 primary amino acid sequence and cellular tropism.

### 3.2 Results

#### 3.2.1 UspA1 proteins differ in cellular adherence through CEACAM

Recent work clearly demonstrates that the UspA1 protein from *M. catarrhalis* strain ATCC25238 mediates bacterial binding to CEACAM1 (Hill and Virji, 2003). However, the cellular tropism of the prototypical *M. catarrhalis* strain O35E did not correlate with CEACAM expression. Specifically, the UspA1-expressing *M. catarrhalis* strain O35E demonstrated effective binding to Chang conjunctival epithelial cells (Fig. 3-1A) which express relatively little CEACAM (Fig. 3-1B), while demonstrating little association with A549 respiratory epithelial cells (Fig. 3-1A) (p<0.05) that express relatively large quantities of several different CEACAMs (Fig. 3-1B).
Fig. 3-1. Cellular tropism of the *M. catarrhalis* O35E UspA1 protein. (A) Attachment of the wild-type O35E.118 strain and the mutant strain O35EΔUspA1 to A549 and Chang cells in vitro using a 30 min incubation period for attachment. (B) Expression of CEACAMs by Chang and A549 cells as determined by Western blot analysis. Whole cell lysates of these two human cell lines were probed with antibody to CEACAM (panel 1) or glyceraldehyde-3-phosphate dehydrogenase (panel 2) followed by HRP-conjugated secondary antibody. (C) Attachment of wild-type *M. catarrhalis* O35E.118 and recombinant *E. coli* DH5α strains containing either pELU1-10G, which expresses the O35E UspA1 or the empty pACYC184 vector to A549 and Chang cells using a 30 minute incubation period for attachment. (D) Effect of rabbit antibody to CEACAM (αCEACAM - DAKO) and normal rabbit antibody (NRS) on binding of the two recombinant *E. coli* strains to Chang cells. Representative experiments are shown in Fig. 1A, 1C and 1D. Statistical analysis of attachment data involved use of one-way or two-way ANOVA with a Bonferroni post-test.
An isogenic UspA1-deficient strain of *M. catarrhalis* O35E (O35EΔUspA1) did not adhere to Chang cells, allowing us to attribute O35E adherence to UspA1 expression, yet the UspAI mutant binding to the A549 cells was greater than that of the wild-type strain (p<0.05) (Fig. 3-1A), suggesting that other adhesin(s) contributed to this interaction. This prompted us to express the UspA1 protein from *M. catarrhalis* O35E (UspA1O35E) in *E. coli* so that host cellular binding by UspA1 could be characterized in the absence of other *M. catarrhalis* adhesins.

Consistent with the phenotype of the parental *M. catarrhalis* strains, Chang cell binding of *E. coli* expressing UspA1O35E was much greater than that by A549 cells (Fig. 3-1C)(p<0.05), suggesting that CEACAM receptors are not the primary determinants of A549 binding by this UspA1 variant. The contribution of CEACAM to adherence in attachment assays was characterized in the presence of CEACAM-specific polyclonal antiserum, which effectively blocks binding of *M. catarrhalis* ATCC25238 (Hill and Virji, 2003). There was no effect of the CEACAM antiserum on Chang cell adherence conferred by UspA1O35E (p>0.05)(Fig. 3-1D), suggesting that CEACAM binding was not essential for adherence to this cell line.

### 3.2.2 CEACAM binding by UspA1 variants

Previous studies have confirmed that UspA1 proteins are expressed in recombinant *E. coli* strains, and that their function reflects that in the parental *M. catarrhalis* strain (Lafontaine *et al.*, 2000). In an attempt to compare CEACAM binding by a diverse set of UspA1 proteins, a total of 10 UspA1 proteins were expressed in *E. coli* (Table. 3-1). To confirm UspA1 expression, bacterial lysates were probed with UspA1-specific antibodies. Consistent with previous reports (Meier *et al.*, 2005), individual UspA1
Fig. 3-2. CEACAM binding by recombinant *E. coli* expressing UspA1 variants. **A.** *E. coli* whole cell lysates were Western blotted and then probed to detect UspA1 protein (upper panel) with MAb 24B5 and soluble CEACAM5 binding (lower panel) was detected with anti-CEACAM (Dako). CEACAM5 binding associated with UspA1 in 7 of 10 strains, while no binding was evident in the other 3 strains. **B.** *E. coli* expressing indicated UspA1 variants were spotted onto nitrocellulose and then incubated with soluble CEACAM5, followed by antibodies to detect bound CEACAM5. **C.** *E. coli* expressing indicated UspA1 variants were adsorbed to ELISA plates and then incubated with soluble CEACAM5, which was then detected using specific antibodies.
proteins are expressed at varying levels (Fig. 3-2A). In addition, all of these cloned UspA1 proteins could be detected on the surface of their respective recombinant E. coli strains when they were probed with UspA1-specific MAb 24B5 in flow cytometry (data not shown).

It was previously demonstrated that denatured UspA1 bound to purified soluble CEACAM (Hill and Virji, 2003). The blots containing the lysates were, therefore, re-probed using soluble CEACAM5. Seven of 10 UspA1 proteins tested revealed CEACAM5 binding, while the other three (UspA1_{ATCC43617}, UspA1_{TTA37}, UspA1_{O35E}) did not (Fig. 3-2A). To confirm that binding in the western blot overlay assays reflected the phenotype of UspA1 in the recombinant bacterial membrane, the E. coli strains were spotted onto nitrocellulose (Fig. 3-2B) and adsorbed on to wells of an ELISA plate (Fig. 3-2C). In each case, binding reflected that of the overlay assays, with the same seven UspA1 variants conferring CEACAM5 binding in each assay (compare Fig. 3-2A, B and C). These results suggest that CEACAM binding is not an inherent property of all UspA1 variants, as certain strains express UspA1 variants that do not bind CEACAMs.

In order to confirm that the CEACAM-binding phenotype of UspA1 expressed by E. coli reflected that of the parental M. catarrhalis strains, I used a panel of UspA1-deficient M. catarrhalis strains (Aebi et al., 1998b; Lafontaine et al., 2000; Attia et al., 2006). Lysates derived from the wild-type and UspA1-deleted strains were probed with MAb to UspA1 protein to confirm expected UspA1 expression (Fig. 3-3A and B, top panel). These blots were subsequently re-probed with soluble CEACAM1 or CEACAM5 to assess the CEACAM-binding activity of blotted proteins. CEACAM binding co-localized with the UspA1 variants of M. catarrhalis strains O12E, FIN2344, V1145, and
**Fig. 3-3.** *M. catarrhalis* wild-type (wt) or UspA1-deficient (Δ) strain binding to CEACAMs. *M. catarrhalis* whole cell lysates were separated by electrophoresis and were immunoblotted with Mab24B5 to detect UspA1 (top panel – A. and B.) or binding to soluble CEACAM5 (lower panel A.), or CEACAM1 (lower panel B.). UspA proteins were detected using Cy5-conjugated secondary antibodies, while bound CEACAM was using Bodipy-conjugated antisera. *M. catarrhalis* ferric binding (Fbp) was probed using mouse anti-Fbp antisera as a loading control (lower panels). C. *M. catarrhalis* strains were adsorbed into ELISA plates and then incubated with soluble CEACAM1 or CEACAM5. The bound CEACAMs were then detected using the CEACAM cross-specific antisera (DAKO CEA) followed by HRP-conjugated goat anti-rabbit immunoglobulin.
V1156, but not with that from strain O35E (Fig. 3-3A and B, middle panel). To confirm that overlay-based experiments reflect the phenotype of the intact *M. catarrhalis* strains, wild-type and mutant *M. catarrhalis* strain binding to soluble CEACAM1 and CEACAM5 was assessed using the ELISA-based assay (Fig. 3-3C). While different variants of the adhesins from other pathogenic bacteria may be specific for either CEACAM1 and/or CEACAM5 (Gray-Owen et al., 1997b), the *M. catarrhalis* UspA1 proteins appear to bind either both CEACAMs or neither one. Altogether, these results demonstrate that CEACAM binding is conferred by certain UspA1 protein variants, and that no other CEACAM-binding adhesins are expressed by *M. catarrhalis* that are detectable in this assay system.

### 3.2.3 Sequence Analysis of CEACAM binding in UspA1 proteins.

To reveal the basis of the differences in CEACAM binding for UspA1, the UspA1 amino acid sequence for each of the 10 strains tested were aligned. Previous work by Hill *et al.* (2005) localized CEACAM1 binding to a 141 amino acid polypeptide fragment derived from the C-terminal stalk region of UspA1$_{ATCC25238}$. Sequence alignments showed that each UspA1 variant that lacks CEACAM binding activity is missing a portion of the CEACAM binding region (Fig. 3-4A). Consistent with the loss of these sequences not being an isolated event, the size of the deletion varied amongst the three proteins. Of these the UspA1$_{TTA37}$ had the smallest deletion, lacking only 33 amino acids of the sequence corresponding to the CEACAM binding region described by Hill *et al.* (2005).
### 3.2.4 Fibronectin and Chang cell binding.

In addition to CEACAM binding, various UspA1 proteins have been shown to bind the extracellular matrix protein fibronectin (McMichael *et al.*, 1998). In particular, a 153 amino acid peptide derived from the UspA1 protein from *M. catarrhalis* strain Bc5 was shown to be sufficient to mediate fibronectin binding and could inhibit UspA1-dependant binding to Chang cells (Tan *et al.*, 2005). The minimal sequence capable of binding fibronectin spans a region containing a so-called ‘NINNY’ repeat flanked on either side by a ‘VEEG’ repeat (Fig. 3-4B), which was conserved between the UspA1 and UspA2 proteins expressed by the Bc5 strain (Tan *et al.*, 2005). Whether or not this entire region was necessary for fibronectin binding has remained undetermined. In order to understand whether the presence of these fibronectin binding motifs correlated with CEACAM binding, I compared the UspA1 sequences (Fig. 3-4B). Based on this analysis, I observed that UspA1$_{\text{ATCC43617}}$ was unique in that it lacked the complete VEEG-NINNY-VEEG repeat. While the NINNY repeat is considered to represent the core of the fibronectin binding sequence, the flanking regions have been suggested to play a role in binding. The ordered VEEG-NINNY-VEEG sequence is highly conserved amongst UspA1 proteins but is less well conserved in related UspA2 proteins. Subsequently, UspA2 proteins lacking portions of this ordered repeat were recognized to be deficient in adherence to fibronectin and in Chang cell binding (McMichael *et al.*, 1998; Lafontaine *et al.*, 2000). Based upon these analyses, I predicted that UspA1$_{\text{ATCC43617}}$ would be unable to bind to Chang cells.

In order to determine whether the absence of one VEEG repeat would affect fibronectin-based Chang cell binding, Chang cell binding by recombinant *E. coli*
Fig. 3-4. UspA1 protein sequence alignments. A. Alignment of variant sequences spanning the CEACAM-binding region of UspA1, defined by Hill et al. (2005), which is delineated with a horizontal black line above the sequences. CEACAM binding by *M. catarrhalis* and/or recombinant *E. coli* strains, as determined in the aforementioned studies, is indicated. +, CEACAM binding; -, does not bind CEACAM. B. Alignment of UspA1 variant sequences spanning the region reported to contribute to fibronectin and Chang cell binding (Tan et al., 2005). Relevant peptides from the UspA1 and UspA2 proteins of *M. catarrhalis* strain Bc5, which was originally used to define the fibronectin-binding sequences, are included for comparison with the sequences considered in this study. C. Attachment of recombinant *E. coli* expressing UspA1O35E or UspA1ATCC43617 to Chang conjunctival epithelial cells in the presence or absence of fetal calf serum (FCS). The negative control is *E. coli* containing the pCC1-Kan construct (Kan).
bacteria expressing \( \text{UspA1}_{\text{ATCC43617}} \) was compared with the prototypical \( \text{UspA1}_{\text{O35E}} \) variant. While \( \text{UspA1}_{\text{O35E}} \) expression conferred binding to Chang cells, bacteria expressing \( \text{UspA1}_{\text{ATCC43617}} \) mirrored that of the background control \( E. \text{coli} \) strain (Fig. 3-4C). Adherence of the \( \text{UspA1}_{\text{O35E}} \) bacteria was further increased in the presence of serum, consistent with the ability of serum components such as fibronectin to facilitate binding (Tan et al., 2005). These results indicate that the first VEEG repeat is essential for \( \text{UspA1} \)-dependent binding to Chang epithelial cells, and naturally-occurring variants of \( \text{UspA1} \) can lack this binding function.

### 3.3 Discussion

The surface protein \( \text{UspA1} \) has been shown to play a primary role in \( M. \text{catarrhalis} \) attachment to a variety of different cell types (Aebi et al., 1998b; McMichael et al., 1998; Hill and Virji, 2003). This association appears to result from its ability to bind host-derived molecules including fibronectin (Tan et al., 2005), laminin (Tan et al., 2006), and members of the CEACAM family of intercellular adhesion molecules (Hill and Virji, 2003). While not considered in this study, \( \text{UspA1} \) also interacts with components of both the classical and alternative complement pathways by interacting with both C3 (Nordstrom et al., 2005) and C4b (Nordstrom et al., 2004).

The diversity of functions attributed to \( \text{UspA1} \) is remarkable. However, the various binding activities have largely been considered independent of each other, and often with different strains of \( M. \text{catarrhalis} \). While it was generally assumed that each function would be conserved among \( \text{UspA1} \) variants, the current study was prompted by our finding that the prototypical \( M. \text{catarrhalis} \) strain O35E does not bind to certain CEACAM-expressing cell lines. Since the sequence of variant \( \text{UspA1} \) proteins is
frequently unavailable, several additional *uspA1* genes were cloned and sequenced to obtain a total of ten genes encoding different UspA1 proteins. Three of the ten UspA1 variants tested did not bind to CEACAMs (Fig. 4-2), and each of these lacked a portion of the sequence that has been shown to confer CEACAM1 binding (Hill et al., 2005). While *M. catarrhalis* expresses a number of different adhesins (Bullard et al., 2005; Balder et al., 2007; Lipski et al., 2007; Holm et al., 2004; Reddy et al., 1997; Luke et al., 2007b; Plamondon et al., 2007; Lafontaine et al., 2000) our UspA1-deficient mutants clearly demonstrated that UspA1 was the only adhesin with the potential to confer CEACAM binding.

The lack of fibronectin binding exhibited by the UspA1 protein from *M. catarrhalis* ATCC 43617 (Fig. 3-4C) is a clear demonstration of the variant nature of individual UspA1 proteins. Fibronectin binding has been shown to mediate adherence of *M. catarrhalis* to Chang cells and, unlike CEACAM binding, has been associated with both UspA1 and UspA2 (Tan et al., 2005). The amino acid sequence considered to contribute to fibronectin binding is present in UspA1, UspA2 and the naturally occurring chimera, UspA2H; however its association with each UspA protein is strain-specific. The fact that both UspA1 and UspA2 can confer fibronectin-mediated binding to Chang cells makes it difficult to test the relative contribution of each in the context of *M. catarrhalis*. However, our nucleotide sequence analyses indicated that the UspA1 from strain ATCC43617 contained a deletion within the region that was previously shown to contribute to fibronectin binding (Tan et al., 2005), and our binding assays demonstrated that this variant could not adhere to Chang epithelial cells.
The existence of three different *M. catarrhalis* isolates (O35E, TTA37, and ATCC 43617) that are unable to bind CEACAM1 makes it intriguing to speculate how this phenotype may contribute to *M. catarrhalis* colonization of the nasopharynx or the production of disease in the respiratory tract. *M. catarrhalis* can be found as a member of the commensal flora in healthy individuals or associated with distinct diseases in either children (i.e., otitis media) or adults (i.e., exacerbations of COPD). The specific location (nasopharynx, trachea or middle ear) and manner (swab, tympanocentesis or aspirate) in which *M. catarrhalis* isolates are obtained must be considered when analyzing strain-specific differences in phenotype (Table. 3-1). It is interesting in this regard that two of the three strains that lack CEACAM binding ability were isolated from adult tracheal aspirates.

While much remains to be learned about how each UspA1-associated function contributes to colonization or disease production (or both), it is clear that the ability to bind CEACAM and fibronectin are not universally conserved functions among *M. catarrhalis* strains. Sequence analyses and functional assays must be performed before the spectrum of potential functions can be deduced. We must also consider that other functions, including complement binding, may also vary between *M. catarrhalis* isolates (Attia et al., 2006). It is, therefore, enticing to consider that the ability of UspA1 to engage different combinations of host receptors will elicit different cellular responses (Slevogt et al., 2006; Rosseau et al., 2005; N'Guessan et al., 2007a), and that the different combination of activities may define the colonization ability and virulence potential of each strain.
3.4 Materials and Methods

**Bacterial strains, plasmids, and culture conditions.** The *M. catarrhalis* wild-type strains and mutants employed in this study are listed in Table.3-1. The UspA1 mutants used in this study were previously described (Aebi et al., 1997; Lafontaine et al., 2000; Attia et al., 2006) and were designated as ΔUspA1 in the present study because they contain an internal deletion in the UspA1 ORF. For preparation of whole cell lysates, *M. catarrhalis* strains were grown at 37°C on Brain Heart Infusion agar (Difco/Becton Dickinson, Sparks, MD) in an atmosphere containing 5% CO₂ or in broth. Recombinant *E. coli* DH5α strains containing pACYC184-based recombinant plasmids were grown on Luria-Bertani (LB) medium containing chloramphenicol (10 μg/ml) at 37°C. *E. coli* strain EPI300 cells (Epicentre Biotechnologies, Madison, WI) expressing recombinant UspA1 proteins were grown at 37°C on LB medium with standard antibiotic supplementation and in LB medium with chloramphenicol and CopyControl induction solution (Epicentre) as required.

**General DNA methods.** *M. catarrhalis* genomic DNA was isolated from agar plate-grown cells using the Easy-DNA kit (Invitrogen, Carlsbad, CA). The preparation of plasmid DNA and the purification of PCR products were performed using kits manufactured by QIAGEN (Santa Clarita, CA). Nucleotide sequence data obtained from automated sequencing systems were analyzed using the MacVector analysis package (version 6.5, Oxford Molecular Group, Campbell, CA). Mult-Alin (Corpet, 1988) was used to generate sequence alignments.

**DNA cloning.** Plasmid pELU1-10G which encodes the uspA1 gene from *M. catarrhalis* O35E has been described (Lafontaine et al., 2001). To clone uspA1 genes into the
CopyControl vector (Epicentre), PCR was used to amplify DNA encoding the entire 
uspA1 ORF and approximately 40-200 nt of DNA 5' from the uspA1 translational start 
codon, using bacterial chromosomal DNA as the template for PCR. This PCR product 
was then cloned into the Blunt Cloning-Ready pCC1 Vector (Epicentre) to generate 
recombinant plasmids expressing each different UspA1 protein. Each cloned uspA1 
gene was fully sequenced and confirmed to be identical to that in the chromosome of the 
relevant M. catarrhalis strain. A kanamycin resistance cassette was cloned into pCC1 to 
generate the pCC-Kan negative control plasmid. These CopyControl-based plasmids 
were electroporated into Electrocompetent TransforMax E. coli EPI300 as described in 
the Epicentre CopyControl cDNA, Gene & PCR Cloning Kit.

**Measurement of bacterial attachment to human cells.** Attachment of wild-type and 
mutant M. catarrhalis strains and recombinant E. coli strains expressing M. catarrhalis 
UspA1 proteins to Chang human conjunctival epithelial cells (ATCC CCL20.2) and to 
A549 human lung epithelial cells (ATCC CCL-185) was measured by using a 
modification of a previously described attachment assay (Aebi et al., 1998b). It must be 
noted that these Chang cells purchased from the American Type Culture Collection were 
accompanied by an ATCC notice that this line has been contaminated with HeLa cells. 
Briefly, tissue culture medium was removed from human cells grown overnight in 24-
Well plates and replaced with 0.5 ml of fresh tissue culture medium to which was then 
added 25 μl of M. catarrhalis cells (~10^6 cfu) or recombinant E. coli cells (~10^6 cfu). 
These bacterial suspensions were obtained by harvesting growth from agar plates that 
had been grown overnight and suspending these cells in either BHI (for M. catarrhalis) 
or LB (for E. coli) to a density of approximately 300 Klett units as measured in a Klett-
Summerson colorimeter (VWR Scientific, West Chester, PA). The tissue culture plates were then subjected to centrifugation at 200 x g for 5 min and incubated at 37°C in an atmosphere of 95% air-5% CO₂ for 30 min or 1 hr after which the wells were washed 5 times with BHI or LB medium. The human cells and attached bacteria were removed from the wells by treatment with trypsin (0.05%-0.25%), subjected to vigorous mixing, and plated onto either BHI or LB agar to determine the number of attached bacteria.

Additional bacterial attachment experiments with Chang cells examined the effect of specific antibodies or serum components on attachment. To determine the effect of CEACAM antibodies on bacterial attachment, tissue culture medium was removed from Chang cells grown overnight in 24-well plates and replaced with 1 ml of fresh tissue culture medium containing 50 μl of rabbit antiserum to CEACAM (Dako North America, Carpinteria, CA; catalog A0115) or normal rabbit serum (Dako catalog X0903) or PBS. These two antisera had been previously dialyzed overnight against PBS to remove azide. These plates were subjected to centrifugation at 200 x g for 5 min and then incubated at 37°C for 1 hr. Next, the medium was removed from each Well and replaced with fresh tissue culture medium to which was then added 25 μl of recombinant E. coli cells containing plasmid pACYC184 or pELU1-10G (suspended to 250 Klett units) and the attachment assay was carried out as described above except that attachment was allowed to proceed for 2 hr.

To determine the effect of serum components on attachment, Chang cell monolayers grown in the presence of 10% (vol/vol) fetal calf serum (FCS) were washed once with serum-free tissue culture medium and then tissue culture medium with or without 10% FCS was added to each well. Recombinant E. coli cells containing M.
*catarrhalis uspA1* genes cloned into the CopyControl vector were induced for 3 hrs with CopyControl Induction solution, adjusted to a density of 300 Klett units, and then 25 μl portions of this suspension were added to the wells and the attachment assay performed as described above with a 1 hr incubation period.

**Western blot and overlay analyses.** Bacterial whole cell lysates (Aebi *et al.*, 1997) were boiled and resolved by SDS-PAGE in 12.5 % (wt/vol) polyacrylamide separating gels, transferred to Immobilon-P membranes (Millipore, Bedford, MA) and probed with the appropriate antibody. UspA1 was detected using monoclonal antibody (MAb) 24B5 (Cope *et al.*, 1999). Whole cell lysates of Chang and A549 cells were prepared from confluent monolayers grown in 24-well tissue culture plates. These monolayers were trypsinized, suspended in PBS, and then solublized in SDS-PAGE digestion buffer. Rabbit antiserum to CEACAM (described above) and mouse antiserum to glyceraldehyde-3-phosphate dehydrogenase was used as the primary antibodies. CEACAMs were detected using CEA-specific rabbit polyclonal antisera (DAKO, Glostrup, Denmark). In overlay experiments, the blots were probed with soluble CEACAM5 (Fitzgerald, Concord, MA) or soluble CEACAM1 (Schumann *et al.*, 2004), followed by the CEA-specific antisera, which was then detected using secondary antisera. The secondary antibodies were goat anti-mouse IgG or anti-rabbit IgG conjugated to either Bodipy or Cy5 fluorophores (Jackson ImmunoResearch, West Grove, PA). Antigen-antibody complexes were visualized by immunofluorescence using a Typhoon 9400™ (Amersham Biosciences).

**CEACAM binding assays.** Overnight cultures of *E. coli* expressing, *M. catarrhalis* UspA1 proteins were grown in LB chloramphenicol (12.5μg/ml) and Copycontrol
induction solution (Epicentre™) for 5 hours. Cells were washed with PBS/Mg/Ca (10 mM MgCl₂ and 5 mM CaCl₂). For bacterial dot blots, bacterial density was calculated by measuring optical density at 550 nm and then diluting to 10⁶ bacteria/ml before spotting onto nitrocellulose. Cells were dried overnight at Room Temperature (RT) and the membranes were blocked in PBS containing 4% skim milk. These membranes were probed with soluble CEACAM5, which was subsequently detected using CEA-specific antiserum and HRP-coupled secondary antibodies. For ELISAs, Nunc plates™ were coated with heat-killed bacteria and dried at 37°C overnight. Wells were blocked using PBS containing 4% skim milk, followed by incubation with soluble CEACAM5 or CEACAM1. CEACAM was detected as described for the dot blots experiments except that HRP-based colorimetric detection system (Sure Blue™) was used to measure CEACAM binding by detection at 650 nm.
<table>
<thead>
<tr>
<th>Strain</th>
<th>Adult/Child</th>
<th>Disease</th>
<th>Location isolated</th>
<th>Geographical location</th>
<th>UspA1 Accession</th>
<th>Reference/Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATCC 25238</td>
<td>Adult</td>
<td>Unknown</td>
<td>Unclear – nasopharynx (Neisseria isolate)</td>
<td>University of Rochester</td>
<td>AAD43465</td>
<td>(Cope et al., 1999) / ATCC</td>
</tr>
<tr>
<td>O35E</td>
<td>Child</td>
<td>Otitis media</td>
<td>tympanocentesis-derived isolate</td>
<td>Dallas, TX, USA</td>
<td>AAB96359</td>
<td>(Aebi et al., 1997) / John D. Nelson</td>
</tr>
<tr>
<td>O35E.118</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Lafontaine et al., 2001)</td>
</tr>
<tr>
<td>O35EΔUspA1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O12E</td>
<td>Child</td>
<td>Otitis media</td>
<td>tympanocentesis-derived isolate</td>
<td>Dallas, TX, USA</td>
<td>AAF40118</td>
<td>(Lafontaine et al., 2000) / John D. Nelson</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Aebi et al., 1997)</td>
</tr>
<tr>
<td>O12EΔUspA1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V1145</td>
<td>Child</td>
<td>Healthy</td>
<td>Nasopharynx</td>
<td>North Carolina, USA</td>
<td>ABY64882</td>
<td>(Attia et al., 2006) / Frederick Henderson</td>
</tr>
<tr>
<td>V1145 ΔUspA1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Attia et al., 2006)</td>
</tr>
<tr>
<td>V1156</td>
<td>Child</td>
<td>Healthy</td>
<td>Nasopharynx</td>
<td>North Carolina, USA</td>
<td>ACC44785</td>
<td>(Attia et al., 2006) / Frederick Henderson</td>
</tr>
<tr>
<td>V1156 ΔUspA1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Attia et al., 2006)</td>
</tr>
<tr>
<td>V1118</td>
<td>Child</td>
<td>Healthy</td>
<td>Nasopharynx</td>
<td>North Carolina, USA</td>
<td>ACR58308</td>
<td>(Attia et al., 2006) / Frederick Henderson</td>
</tr>
<tr>
<td>7169</td>
<td>Child</td>
<td>Otitis media</td>
<td>tympanocentesis-derived isolate</td>
<td>Buffalo, NY, USA</td>
<td>ACR59385</td>
<td>(Luke and Campagnari, 1999) / Anthony Campagnani</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Attia et al., 2006)</td>
</tr>
<tr>
<td>7169 ΔUspA1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TTA37</td>
<td>Adult</td>
<td>COPD</td>
<td>Transtracheal aspirate</td>
<td>Tennessee, USA</td>
<td>AAF40122</td>
<td>(Lafontaine et al., 2000) / Steven Berk</td>
</tr>
<tr>
<td>FIN2344</td>
<td>Child</td>
<td>unclear</td>
<td>Unknown</td>
<td>Finland</td>
<td>ACC44784</td>
<td>(Attia et al., 2006) / Merja Helminen</td>
</tr>
<tr>
<td>FIN2344 ΔUspA1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Attia et al., 2006)</td>
</tr>
<tr>
<td>ATCC 43617</td>
<td>Adult</td>
<td>COPD</td>
<td>Transtracheal aspirate from coal miner with chronic bronchitis</td>
<td>Belgium</td>
<td>ACR58310</td>
<td>(Wallace, Jt. et al., 1989) / ATCC</td>
</tr>
</tbody>
</table>

Table 3-1. *M. catarrhalis* strains used in this chapter.
Chapter 4: Modular arrangement of allelic variants explains divergence in *M. catarrhalis* UspA protein function.

The contents of this chapter have been previously published:
Modular arrangement of allelic variants explains divergence in Moraxella catarrhalis UspA protein function.
(reproduced/amended with permission from American Society for Microbiology)

Author Contributions:
Michael Brooks, Scott Gray-Owen and Eric Hansen designed the experiments, analyzed research and wrote the paper.

Michael Brooks contributed all figures.

Jennifer Sedillo, Nikki Wagner, Cassie Laurence, Wei Wang and Ahmed Attila all contributed various UspA sequences.
4.1 Abstract

The ubiquitous surface protein Δ molecules (UspAs) of *Moraxella catarrhalis* are large, non-fimbrial autotransporter proteins that can be visualized as a “fuzzy” layer on the bacterial surface by transmission electron microscopy. Previous studies have attributed a wide array of functions and binding activities to the closely related UspA1, UspA2 and/or UspA2H proteins, yet the molecular and phylogenetic relationship between these activities remains largely unexplored. To address this issue, the nucleotide sequence of the *uspA1* genes from a variety of independent *M. catarrhalis* isolates were sequenced and compared to the deduced amino acid sequences of previously characterized UspA1, UspA2, and UspA2H proteins. Rather than being conserved proteins, I observed a striking divergence of individual UspA1, UspA2 and UspA2H proteins resulting from the modular assortment of unrelated ‘cassettes’ of peptide sequence. The exchange of certain variant cassettes correlates with strain-specific differences in UspA protein function, and confers differing phenotypes upon these mucosal surface pathogens.
4.2 Introduction

Over the past decade, much of the research effort concerning *M. catarrhalis* has focused on identification of proteinaceous surface antigens and determination of their potential role in virulence expression (Karalu's and Campagnari, 2000; Verduin et al., 2002). Of these, the Hag (Holm et al., 2003; Forsgren et al., 2003), OmpCD (Reddy et al., 1997; Holm et al., 2004), filamentous hemagglutinin-like proteins (Balder et al., 2007; Plamondon et al., 2007), McaP (Timpe et al., 2003) and type IV pilus proteins (Luke et al., 2007a) have been shown to be involved in bacterial adherence to human cells or components of the respiratory tract in vitro, while CopB (Campagnari et al., 1994; Aebi et al., 1996), transferrin binding proteins (Schryvers and Lee, 1989; Luke and Campagnari, 1999), and lactoferrin-binding proteins (Du et al., 1998; Bonnah et al., 1999) are likely essential for iron acquisition.

Perhaps the most intensively studied potential virulence factors of *M. catarrhalis* are the non-fimbrial ubiquitous surface protein A (UspA) molecules, initially described about fourteen years ago. At first, these two proteins were thought to be a single gene product (Helminen et al., 1994; Klingman and Murphy, 1994) until it was discovered that two genes encode distinct proteins sharing certain epitopes (Aebi et al., 1998b). UspA1 and UspA2 are encoded by ORFs located far apart in the *M. catarrhalis* genome (Wang et al., 2007).

Early studies of *M. catarrhalis* uspA1 and uspA2 mutants indicated that the former protein had adhesive properties and functioned to bind *M. catarrhalis* to Chang conjunctival epithelial cells in vitro, whereas expression of UspA2 was essential for *M. catarrhalis* to resist the bactericidal activity of normal human serum (Aebi et al., 1998b).
Subsequent studies have shown that both of these proteins have the potential to express multiple different functional activities, ranging from attachment to host gene products to affecting biofilm development by some strains of *M. catarrhalis* growing in vitro (Pearson *et al.*, 2006). UspA1 has been shown to bind CEACAM1 (Hill and Virji, 2003; Hill *et al.*, 2005), fibronectin (Tan *et al.*, 2005), laminin (Tan *et al.*, 2006), and the serum complement factors C3 (Nordstrom *et al.*, 2005) and C4b-binding protein (Nordstrom *et al.*, 2004). UspA2 has also been reported to bind most of these same proteins except CEACAM1 binding has not been described. UspA2 also binds the serum protein vitronectin and apparently uses this mechanism to help some strains of *M. catarrhalis* resist killing by normal human serum (Attia *et al.*, 2006; Attia *et al.*, 2005).

Previously, a limited molecular characterization of the UspA proteins revealed that UspA1, UspA2, and UspA2H share a number of homologous repeats and ‘motifs’ (Cope *et al.*, 1999; Aebi *et al.*, 1998b; Lafontaine *et al.*, 2000), a finding which suggested that the make-up of UspA proteins might be interchangeable. Despite the large number of functions attributed to the UspA proteins, the molecular characterization of these proteins has largely been limited to alleles encoded by few prototypical strains. I (Brooks *et al*. Chapter 3) and others (Lafontaine *et al.*, 2000) have noted strain-specific differences in UspA protein function, yet no systematic structure-function analyses have been performed to date. In the present report, I characterized the ‘motifs’ and repeats of a large number of UspA proteins and further defined these shared peptide sequences. In addition, I showed that individual UspA proteins consist of a modular arrangement of sequence cassettes, some of which encode recently defined functional motifs.
4.2 Results

4.2.1 UspA proteins.

While the crystal structure of a UspA protein has not been solved to date, the basic structure of these proteins can be inferred from structural studies of other autotransporter proteins, due to their sequence similarity to the *Neisseria meningitidis* NalP (Oomen *et al.*, 2004; Attia *et al.*, 2005), *Yersina enterocolitica* YadA (Oomen *et al.*, 2004) and *Haemophilus influenzae* Hia proteins (Meng *et al.*, 2006), the crystal structures of which have been solved in whole or in part. NalP is a conventional autotransporter whereas YadA is the prototype of a trimeric autotransporter (Koretke *et al.*, 2006). A model of the entire YadA structure has been generated by combining a crystal structure of the amino (N)-terminal domain (Cotter *et al.*, 2005) with predicted structures of the stalk and the C-terminal membranes-spanning domains (Nummelin *et al.*, 2004). As such, the UspA proteins are also predicted to possess three distinct domains (Fig. 4-1A): the N-terminal region is considered to form a β-sheet-based globular head, the stalk region that connects the head to the membrane-anchoring domains consists of a coiled-coil structure and a C-terminal translocation domain responsible for formation of a pore in the outer membrane to allow passage of the passenger domain to the cell surface (Koretke *et al.*, 2006). By analogy to YadA, the N-terminal portion of the passenger domain likely forms a “head” that is connected to an extended “stalk” region which itself is bound to the membrane-spanning region of the translocation domain.

The UspA proteins have been subdivided into three basic types: UspA1, UspA2 and UspA2H (Lafontaine *et al.*, 2000; Aebi *et al.*, 1997). UspA1 and UspA2 can be
Fig. 4-1. A. Predicted structure of YadA non-fimbrial adhesin (Hoiczyk et al., 2000; Koretke et al., 2006). Non-fimbrial adhesins such as YadA and UspA are predicted to form three distinct structural domains: the head, stalk and membrane-spanning regions. B. UspA proteins have been categorized into three groups (UspA1, UspA2 and UspA2H) based on similarity within the head and Membrane-spanning domains.
distinguished by differences in amino acid sequence within the head and membrane-spanning regions, yet they share homology within the stalk region (Fig. 4-1B). UspA2H is a ‘hybrid’ protein containing a head region similar to that of UspA1 while having the UspA2-like C-terminal region (Fig. 4-1B). Based on analysis of the C-terminal region of these three proteins, all three appear to be members of the trimeric autotransporter family (Cotter et al., 2006; Kim et al., 2006).

Our recent demonstration that UspA1 proteins from different *M. catarrhalis* strains differ with respect to host cell receptor specificities (Chapter 3) prompted us to analyze the diversity within each UspA protein group, considering the clear modular arrangement of variant sequences apparent from these analyses. Herein, I will discuss structural considerations of the variant sequences on each domain separately before discussing their functional implications.

### 4.2.2 UspA N-domains

UspA1 and UspA2H have previously been shown to possess homology within their N-terminal domains. This region is characterized by a number of sequence motifs, including a series of ‘GGG repeats’ followed by a ‘FAAG domain’ (Lafontaine et al., 2000). Five additional UspA1 genes were sequenced, and combined these data with the previously sequenced UspA1 and UspA2H genes (Table. 4-1) to further define these ‘motifs’ (Lafontaine et al., 2000). Using “WebLogo” (Crooks et al., 2004) software, I refined the N-terminal GGG repeat consensus sequence using 150 repeats taken from all available UspA1 and UspA2H sequences (19 variants); (Fig. 4-2A), leading to the generation of a logo showing a highly conserved consensus sequence. The number of GGG repeats varies between 5-16 repeats within individual UspA1/UspA2H proteins,
Fig. 4-2. UspA1 and UspA2H N-terminal domain arrangement and composition. A. Multiple sequence alignment of N-terminal GGG repeats and consensus sequence revealed by comparison of the 149 available GGG repeats. B. Multiple sequence alignment of FAAG domains from UspA1 and UspA2H. Highlighted are FAAG core and transition repeats. Predicted structural repeat (arrow) and linking loop (green) are indicated. C. Consensus sequences for the FAAG core repeats and the YadA SVAIG-S repeats. D. Structure of the YadA head domain (PDB:1P9H). Overlay of a consensus of UspA1 and YadA repeat sequence on YadA left-handed parallel β-roll. Amino acid sequence identities are indicated by red (>90%) or blue (>50%).
with each GGG repeat being predicted to form an anti-parallel β-strand based on alignments with the N-terminal YadA structure (Hoiczyk et al., 2000; Nummelin et al., 2004; Crooks et al., 2004). This observation highlights a substantial variation in the size of the distal portion of the head region resulting from a difference in the number of conserved core repeats among individual alleles.

Downstream of the GGG repeat lies the previously defined FAAG motif (Lafontaine et al., 2000) (Koretke et al., 2006) (Fig. 4-2B), which contains a high degree of sequence similarity to the head region of YadA (Lafontaine et al., 2000). A recently solved crystal structure of the YadA head reveals a structural repeat of 11 amino acids containing an NSVAIG-S sequence which forms two short anti-parallel strands (Hoiczyk et al., 2000). While the link between the NSVAIG-S repeats varies slightly in size and composition, a strikingly conserved G residue occurs at the fifth position of each repeat, allowing the formation of a left handed parallel β-roll (Fig. 4-2C). I analyzed the M. catarrhalis FAAG motif for NSVAIG-S like sequences. Each UspA FAAG contained five core NSVAIGS-like sequences, each being linked to the next by short sequences that varied in size and composition to a greater degree than is apparent among YadA variants. Using the WebLogo software, I created consensus sequences (Fig. 4-2D) for the available YadA SVAIG-S (42 repeats) and UspA FAAG (91 repeats) sequences. The homology between these sequences is primarily linked to the structural regions, suggesting a similar fold, with a varying length and composition of the linker sequences. While the head of YadA is responsible for the collagen binding properties of YadA (Koretke et al., 2006), it remains unclear whether such linker variability contributes to antigenic or functional differences among UspA1 and/or UspA2H variants.
**N-terminal UspA2**

### NTER2A motif

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Consensus</th>
</tr>
</thead>
</table>
| 412EugR2 | NTER2A |}
| 3P2Z-10aR2 | NTER2A |
| SP1-10aR2 | NTER2A |
| HuUspA2 | NTER2A |
| NestA2 | NTER2A |
| FtsZLpR2 | NTER2A |
| L46-10aR2 | NTER2A |
| 3P1Z-10aR2 | NTER2A |
| Consensus | NTER2A |

### NTER2B motif

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Consensus</th>
</tr>
</thead>
</table>
| 1112TspR2 | NTER2B |}
| T12204aR2 | NTER2B |
| T12206aR2 | NTER2B |
| T12209aR2 | NTER2B |
| RICE 5228010aR2 | NTER2B |
| Consensus | NTER2B |

Fig. 4-3. N-terminal UspA2 domain arrangements. Sequence alignment of distinct UspA2 N-terminal motifs, defined as NTER1 or NTER2.
While the UspA1 N-terminal region has a high degree of overall similarity to that of YadA, the N-terminal region of UspA2 does not share similarity in either sequence or predicted structure with either YadA or UspA1. The N-terminal region of UspA2 itself has also remained largely uncharacterized due to an apparent lack of homology between the few fully sequenced UspA2 variants. To understand this diversity, additional uspA2 alleles were sequenced and compared to previously sequenced variants. This analysis revealed that the N-terminal domains can be clearly divided into two different groups, which I have termed NTER2A and NTER2B (Fig. 4-3). The UspA2 protein from strain P44 is the only variant that does not belong to either group; it appears to be NTER2-related, but the N-terminal domain has been replaced by a duplicated ‘HDD’ stalk motif (described below). It is clear that UspA2 proteins can have significantly different N-terminal regions based on amino acid sequence alone. Any function ascribed to the N-terminal domain of UspA2 must be considered in the context of these two protein families.

4.2.3 UspA stalk region

While the N-terminal regions of individual UspA proteins are clearly distinct based upon primary amino acid sequence, the stalk region consists of both a combination of repeats that occur in all UspA proteins and distinct sequence ‘motifs’ that are primarily associated with either UspA1 or UspA2. I further refined each of the previously described ‘motifs’ by using all available sequences to create consensus logos. Our inclusion of new alleles has also allowed us to define previously uncharacterized sequence motifs shared amongst the UspA proteins. Each ‘motif’ located within the stalk region is depicted as sequence logos within Fig. 4-4 including the number of repeats.
Fig. 4. UspA stalk repeats/motifs. Consensus sequences were generated for each region, with the number of repeats used to generate each consensus indicated.
used to define each consensus. Adjacent stalk repeats are often connected by short (typically 4-8 residues - K/QADIAKN) linker sequences, which tend to be conserved at junctions between specific combinations of stalk repeats, but vary when an adjacent repeat is different. Whether linker sequences are determined by protein structural requirements or required for nucleotide recombination remains unclear. While the specific arrangement of individual stalk repeats will be detailed below, it is pertinent to note that certain ‘motifs’ appear to be restricted to either UspA1, UspA2 and/or UspA2H whereas others are shared among these different groups.

4.2.4 UspA C-terminal regions

The C-terminal regions of UspAs form a structurally conserved membrane-spanning domain (Wollmann et al., 2006). Each contains a coiled-coil region that links the stalk region to the membrane-spanning translocation domain. In contrast to the primary stalk domain, there is almost complete sequence identity shared among the C-terminal domains (CTER1) of UspA1 proteins and among the C-terminal domains (CTER2) of the UspA2 and UspA2H proteins, although CTER1 and CTER2 are clearly distinct from each other (Fig. 4-5). In each case, the most carboxyl-terminal sequence encodes 4 β-strands that would be predicted to weave back and forth through the outer membrane, reminiscent of both Hia and YadA. Structural modeling of YadA illustrates that three membrane-spanning monomers coalesce to from a single β-barrel that spans the membrane 12 times. The α-helical portion of each monomer passes up through the center of this trimer, forming a coiled-coil structure that forms the base of the stalk domain. This trimeric structure is remarkable considering that the proteins are auto-transporters that direct their own insertion and assembly within the outer membrane.
Fig. 4-5. Carboxyl-terminal region motifs of UspA proteins. Multiple sequence alignment of UspA1 CTER1 (A) and UspA2 CTER2 (B) motifs, with predicted transmembrane β-strands indicated by horizontal arrows over the relevant sequence corresponding to either Hia or YadA. Structure of predicted 12-strand model of C-terminal auto-transporter membrane anchor domain created by Koretke et al. 2006 was modified to hi-light monomeric CTER motif (Blue).
4.2.5 Modular arrangement of UspA1

By defining the conserved motif and repeat sequences, I was able to consider the composition of distinct UspA variants. The C-terminal region of UspA1 variants is highly conserved among diverse clinical isolates, with all 14 available UspA1 proteins possessing a NINNY-KASS-FET sequence immediately adjacent to the CTER1 motif (Fig. 4-6). The first variability in modular arrangement involves the CEACAM motif that has been shown to mediate adherence to CEACAM (Hill et al., 2005). This motif is either truncated or completely absent from some UspA1 variants and both of these changes disrupt CEACAM-mediated adherence (Brooks et al., Chapter 3). The CEACAM binding phenotype of all of the M. catarrhalis strains included in the present study strictly correlates with the presence of an intact CCM motif (Fig. 4-6), highlighting the importance of allelic sequence-specific analyses in the assignment of function to UspA proteins.

The CEACAM motif is preceded by a highly conserved LAAY-KASS sequence (Fig. 4-4 and Fig. 4-6). The stalk regions of UspA1 proteins then contain a variable number of VEEG repeats, a NINNY repeat, and then an additional VEEG sequence (Fig. 4-6). This region has previously been shown to mediate binding to the extracellular matrix protein fibronectin and/or attachment to Chang epithelial cells (Tan et al., 2005). This binding appears to correlate with a complete ordered VEEG-NINNY-VEEG sequence, as UspA1\textsubscript{ATCC43617} lacks the N-terminal VEEG and its binding to Chang cells \textit{in vitro} is diminished (Brooks et al., Chapter 3). As discussed below, the presence or absence of an intact VEEG-NINNY-VEEG sequence strictly correlates with the ability
Fig. 4-6. Modular arrangement of UspA1 proteins. A. Molecular basis of UspA1 proteins, with regions of variability indicated. Motif designations indicated above the filled boxes reflect sequences defined in Figures 2 through 5. B. Modular arrangement of UspA1 variants with function attributed to a particular variant indicated where it has been determined.
of UspA2 binding to fibronectin and UspA2H variants to adhere to Chang cells, indicating that shared sequences explain the overlapping function of UspA proteins from these various groups.

The N-terminus of UspA1 (Fig. 4-6) and UspA2H (Fig. 4-7B) both consist of a variable number (5-16) of GGG repeats followed by a FAAG motif. However, the region between the FAAG motif and the fibronectin-binding sequence is remarkably divergent, and can be sub-divided based upon apparent phylogenetic relationships within this region. For this reason, I have termed this region as the UspA1-variable region (U1VR). Phenotypic analysis of strains expressing diverse U1VR domains may reveal sequence motifs that mediate other functions attributed to UspA1, including its ability to bind laminin (Tan et al., 2006) and at least two different proteins involved in the complement cascade (Nordstrom et al., 2004; Nordstrom et al., 2005).

4.3.6 Modular arrangement of UspA2

UspA2 proteins follow a very similar trend to UspA1 proteins in terms of sequence conservation. The CTER2 motif (Fig. 4-5B) is very highly conserved among UspA2 and UspA2H variants, and consists of a membrane spanning region linked to a coiled-coil stalk. CTER2 is linked to a FET motif (Fig. 4-7B) in all cases but one: UspA2TTA24 contains a portion of the CEACAM domain in place of the FET sequence. Based upon previous work (Hill and Virji, 2003; Hill et al., 2005) (Brooks-Chapter 3), it is not likely that this small portion of the CEACAM motif is sufficient for CEACAM binding. However, this arrangement highlights the potential for exchange of cassette-like sequences between UspA1 and UspA2 coding sequences, and clearly indicates that function must be assigned to individual alleles rather than to UspA protein groups.
Fig. 4-7. Modular arrangement of UspA2 and UspA2H proteins. UspA2 (A) or UspA2H (B) proteins, with regions of variability indicated. Motif designations indicated above the filled boxes reflect sequences defined in Figures 2 through 5. Color coding is also consistent with that used for UspA1 in Figure 6. Modular arrangement of UspA2 and UspA2H variants with functions indicated where they have been identified. (+) indicates UspA2 proteins shown to mediate serum resistance. (–) indicates UspA2 proteins unable to mediate serum resistance. (*) indicates strains previously shown to be serum resistant, but have not attributed the function to any adhesin.
The C-terminal stalk region of the UspA2 protein mimics what is seen in the UspA1 protein with a NINNY-KASS repeat except that UspA2 proteins completely lack the CEACAM binding sequence, generating a LAAY-KASS-NINNY-KASS sequence.

Rather than the single LAAY-KASS repeat present in UspA1 proteins (Fig. 4-6), UspA2 proteins have a variable number (2-4 repeats) of LAAY-KASS closely linked to one or more VEEG motifs (Fig. 4-7A). While VEEG and NINNY repeats exist in all complete UspA2 sequences available, they only rarely exist in a VEEG-NINNY-VEEG format that correlates with fibronectin binding. This is consistent with the fact that only a small subset of UspA2 variants bind to fibronectin, and those that do display an obvious fibronectin/Chang cell binding motif (Fig. 4-7A).

As mentioned previously, the UspA2 variants tend to display one of two distinct N-terminal ‘motifs’ (NTER1 or NTER2; Fig. 4-3). These are linked to a variable region similar to that observed in UspA1, highlighting the diversity in UspA2 proteins among different strains. To highlight the fact that most stalk motifs found in the UspA2 variable region differ from those found in UspA1, I have termed this region the UspA2 variable region (U2VR; Fig. 4-7).

It was previously demonstrated that serum resistance could be conferred on the serum-sensitive *M. catarrhalis* strain 317 by exchange of the UspA2\textsubscript{MC317} variable region (HDD-SIE) with that of UspA2\textsubscript{O35E} (LAAY-KASS-TAEER) from the serum-resistant *M. catarrhalis* strain O35E (Attia \textit{et al.}, 2005). However, both of these UspA2 proteins contain a NTER1 motif and the specific requirements for serum resistance remain unclear because both UspA2\textsubscript{7169} and UspA2\textsubscript{O12E} also confer serum resistance on
their respective *M. catarrhalis* strains despite having what are clearly different UspA2-variable regions.

**4.4 Discussion.**

UspA proteins have been previously assigned to one of the three main groups: UspA1, UspA2 or UspA2H. Our detailed analysis of 34 UspA amino acid sequences has revealed a remarkable diversity among the three UspA groups, as well as the potential to exchange variable motifs between them. Structurally, the inter-strain variability appears most evident within the N-terminal region, where the UspA1 and UspA2H proteins possess a wide variability in the number of GGG repeats, each of which form an anti-parallel β-strand, followed by a FAAG motif. This variability will undoubtedly cause large changes in the size of the head domain (Fig. 4-1 and Fig. 4-2), which is otherwise analogous to that of the *Yersinia* YadA protein (Tahir *et al.*, 2000; Nummelin *et al.*, 2004). In stark contrast, the UspA2 N-terminus lacks any similarity to the UspA1/UspA2H/YadA structure. This is the greatest difference among these proteins, yet its impact awaits definition of a function attributable to the novel head structure. At the amino acid sequence level, the highly variable U1VR and U2VR regions (Fig. 4-7) extend the overall diversity by changing the length of the coiled-coil stalk and by conferring different binding phenotypes on UspA1 and UspA2H proteins.

While the modularity of UspA proteins may facilitate immune escape due to the variation of peptide sequences exposed at the bacterial surface, it also clearly affects bacterial phenotype. Perhaps the most striking examples of this involve host cellular attachment via CEACAM1 receptors and adherence to Chang cells, with this latter trait appearing to be mediated by binding to the extracellular matrix protein fibronectin.
While CEACAM binding is common to a variety of UspA1 variants (Hill and Virji, 2003), it is clearly not a property observed in all clinical isolates (Brooks et al., Chapter 3). Sequence analysis reveals a direct link between the presence of a complete CEACAM binding motif and CEACAM binding. While CEACAM binding is not apparent in any of the UspA2 variants tested (Brooks et al., Chapter 3), the presence of a portion of the CEACAM binding domain on UspA2 visits clearly illustrates the potential for exchange between UspA classes.

Rather than CEACAM receptor binding, Chang cell binding by M. catarrhalis is associated with the ability of UspA proteins to bind the extracellular matrix component fibronectin (Tan et al., 2005). This activity has been attributed to UspA1 and/or UspA2, yet the majority of UspA1 and UspA2H variants characterized contain the fibronectin binding motif whereas only a minority of the UspA2 proteins characterized in this study possessed it. Laminin binding has previously linked to the N-terminal regions of UspA1 and UspA2 (Tan et al., 2006), however ascribing this function to a particular sequence is less clear because the amino acid sequences of laminin-binding UspA protein variants are generally unavailable.

Serum resistance is primarily associated with UspA2 proteins (Attia et al., 2006;Aebi et al., 1998b), but it is not inherent to all UspA2 variants (Attia et al., 2006;Attia et al., 2005;Aebi et al., 1998b) and different UspA2 variants appear to have different effects on the complement cascade (Attia et al., 2005;Nordstrom et al., 2004;Nordstrom et al., 2005). For example, while the complement components C4b (Attia et al., 2006;Nordstrom et al., 2004;Nordstrom et al., 2005) and C3 (Nordstrom et al., 2004) both bind to UspAs, it is clear that C4b binding is restricted to certain M.
catarrhalis strains (Nordstrom et al., 2005), and that either UspA1 or UspA2 is of primary importance in serum resistance of different strains (Attia et al., 2006; Nordstrom et al., 2004). The issue of serum resistance in M. catarrhalis is further complicated by the fact that other mutations have been reported to have an adverse effect on serum resistance of some strains. Inactivation of the genes encoding the CopB (Aebi et al., 1998a), OMP CD (Holm et al., 2004), and OMP E (Murphy et al., 2000) outer membrane proteins have been shown to reduce serum resistance. In addition, expression of at least three genes encoding LOS biosynthesis enzymes (Zaleski et al., 2000; Luke et al., 2003; Peng et al., 2005) is required for wild-type levels of serum resistance in M. catarrhalis. To date, however, only the UspA2 protein has been shown to be directly involved in the expression of serum resistance (Attia et al., 2005; Attia et al., 2006).

In summary, while various studies have revealed novel UspA protein functions in a prototypical strain, it is clear that the function conferred by different UspA variants may differ widely. Moreover, it is enticing to speculate that the various sequence ‘cassettes’ evident in the UspA variants characterized to date may confer heretofore unrecognized functions. Future analyses must obviously consider the sequence of a particular UspA variant being studied, and place the results in the context of known allelic differences. Considering the natural genetic competence of M. catarrhalis (Catlin and Cunningham, 1964), it seems likely that the structural and functional determinants of each UspA are phylogenetically fluid, allowing the acquisition of various combinations of binding functions. This combinatorial nature of UspA proteins makes it essential to understand how these different functions interact, with an aim to ascertaining whether
certain bacterial phenotypes differentiate between asymptomatic commensalism and pathogenesis upon *M. catarrhalis* infection.

### 4.5 Materials and Methods.

**PCR and nucleotide sequence analysis.** PCR was used to amplify DNA encoding the *uspA1* and *uspA2* genes from *M. catarrhalis* chromosomal DNA as described (Lafontaine *et al.*, 2000; Attia *et al.*, 2006; Pearson *et al.*, 2006). Nucleotide sequence data derived from automated sequencing systems were initially subjected to analysis by using MacVector software (version 6.5, Oxford Molecular Group, Campbell, CA).

**Protein sequence analysis.** Protein sequence alignments were performed using MultAlin (Corpet, 1988). 3D structures were viewed and modified using Deep View Swiss-PDB viewer. 2D and 3D comparative predictions were performed by 3D-Jigsaw (Bates *et al.*, 2001).
<table>
<thead>
<tr>
<th>Strain</th>
<th>Adult/Child</th>
<th>Disease Associated</th>
<th>Location isolated</th>
<th>Geographical location</th>
<th>UspA1 Accession</th>
<th>UspA2 Accession</th>
<th>UspA2H Accession</th>
<th>Reference/Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>O35E</td>
<td>Child</td>
<td>Otitis media</td>
<td>tympanocentesis-derived isolate</td>
<td>Dallas, TX, USA</td>
<td>AAB96359</td>
<td>AAB96391</td>
<td></td>
<td>(Adjiri et al., 1997)</td>
</tr>
<tr>
<td>O12E</td>
<td>Child</td>
<td>Otitis media</td>
<td>tympanocentesis-derived isolate</td>
<td>Dallas, TX, USA</td>
<td>AAF40118</td>
<td>AAF40119</td>
<td></td>
<td>(Lafontaine et al., 2000)</td>
</tr>
<tr>
<td>O46E</td>
<td>Child</td>
<td>Otitis media</td>
<td>tympanocentesis-derived isolate</td>
<td>Dallas, TX, USA</td>
<td>AAF36416</td>
<td>AAF40120</td>
<td></td>
<td>(Lafontaine et al., 2000)</td>
</tr>
<tr>
<td>V1145</td>
<td>Child</td>
<td>Healthy nasopharynx</td>
<td>tympanocentesis-derived isolate</td>
<td>Dallas, TX, USA</td>
<td>ABY64882</td>
<td>ACH55489</td>
<td></td>
<td>(Attia et al., 2006)</td>
</tr>
<tr>
<td>V1156</td>
<td>Child</td>
<td>Healthy nasopharynx</td>
<td>North Carolina, USA</td>
<td>ACC44785</td>
<td>ACC55488</td>
<td></td>
<td>(Attia et al., 2006)</td>
<td></td>
</tr>
<tr>
<td>V1118</td>
<td>Child</td>
<td>Healthy nasopharynx</td>
<td>North Carolina, USA</td>
<td>ACB58308</td>
<td>ACB55487</td>
<td></td>
<td>(Attia et al., 2006)</td>
<td></td>
</tr>
<tr>
<td>V1122</td>
<td>Child</td>
<td>Healthy nasopharynx</td>
<td>North Carolina, USA</td>
<td>AAO59378</td>
<td></td>
<td></td>
<td>(Attia et al., 2006)</td>
<td></td>
</tr>
<tr>
<td>V1171</td>
<td>Child</td>
<td>Healthy nasopharynx</td>
<td>North Carolina, USA</td>
<td>AAD43469</td>
<td>AAD43470</td>
<td></td>
<td>(Frederick Henderson)</td>
<td></td>
</tr>
<tr>
<td>P44</td>
<td>Adult</td>
<td>Pneumonia Sputum</td>
<td>Tennessee, USA</td>
<td>AAN84895</td>
<td>AAN84896.1</td>
<td></td>
<td>(Jordan et al., 1999)</td>
<td></td>
</tr>
<tr>
<td>7169</td>
<td>Child</td>
<td>Otitis media</td>
<td>tympanocentesis-derived isolate</td>
<td>Buffalo, NY, USA</td>
<td>ACB59385</td>
<td>ACB58309</td>
<td></td>
<td>(Lake and Campagnari, 1998)</td>
</tr>
<tr>
<td>FIN2344</td>
<td>Child</td>
<td>Unclear nasopharynx</td>
<td>Finland</td>
<td>ACC44784</td>
<td>ACB59384</td>
<td></td>
<td>(Attia et al., 2006)</td>
<td></td>
</tr>
<tr>
<td>TTA24</td>
<td>Adult</td>
<td>COPD transitral aspirate</td>
<td>Tennessee, USA</td>
<td>AAD43467</td>
<td>AAD43468</td>
<td></td>
<td>(Cope et al., 1999)</td>
<td></td>
</tr>
<tr>
<td>TTA37</td>
<td>Adult</td>
<td>COPD transtracheal aspirate</td>
<td>Tennessee, USA</td>
<td>AAF40122</td>
<td>AAF40121</td>
<td></td>
<td>(Lafontaine et al., 2000)</td>
<td></td>
</tr>
<tr>
<td>ATCC 25238</td>
<td>Adult</td>
<td>Unknown nasopharynx</td>
<td>University of Rochester</td>
<td>AAD43465</td>
<td>AAD43466</td>
<td></td>
<td>(Cope et al., 1999)</td>
<td></td>
</tr>
<tr>
<td>ATCC 43617</td>
<td>Adult</td>
<td>COPD transtracheal aspirate from coal mine</td>
<td>Belgium</td>
<td>ACB58310</td>
<td></td>
<td></td>
<td>(Wallace, Jr. et al., 1989)</td>
<td></td>
</tr>
<tr>
<td>MC317</td>
<td>Child</td>
<td>Unknown nasopharynx</td>
<td>Greece</td>
<td>AAW62383</td>
<td></td>
<td></td>
<td>(Attia et al., 2005; Meier et al., 2003)</td>
<td></td>
</tr>
<tr>
<td>SP-12-5</td>
<td>Child</td>
<td>Otitis prone nasopharynx</td>
<td>USA: New York</td>
<td>AAL76059</td>
<td></td>
<td></td>
<td>(Hays et al., 2003)</td>
<td></td>
</tr>
<tr>
<td>SP-12-6</td>
<td>Child</td>
<td>Otitis prone nasopharynx</td>
<td>USA: New York</td>
<td>AAL76060</td>
<td></td>
<td></td>
<td>(Hays et al., 2003)</td>
<td></td>
</tr>
<tr>
<td>SP12-2</td>
<td>Child</td>
<td>Otitis prone nasopharynx</td>
<td>USA: New York</td>
<td>AAL76062</td>
<td></td>
<td></td>
<td>(Hays et al., 2003)</td>
<td></td>
</tr>
<tr>
<td>SP12-1</td>
<td>Child</td>
<td>Otitis prone nasopharynx</td>
<td>USA: New York</td>
<td>AAL76061</td>
<td></td>
<td></td>
<td>(Hays et al., 2003)</td>
<td></td>
</tr>
<tr>
<td>AL5-1</td>
<td>Child</td>
<td>Otitis non-prone nasopharynx</td>
<td>USA: New York</td>
<td>AAL76064</td>
<td></td>
<td></td>
<td>(Hays et al., 2003)</td>
<td></td>
</tr>
<tr>
<td>LK4-2</td>
<td>Child</td>
<td>Otitis non-prone nasopharynx</td>
<td>USA: New York</td>
<td>AAL76063</td>
<td></td>
<td></td>
<td>(Hays et al., 2003)</td>
<td></td>
</tr>
<tr>
<td>ETSU-9</td>
<td>Adult</td>
<td>COPD transtracheal aspirate</td>
<td>Tennessee, USA</td>
<td>ACB59386</td>
<td>ABB07416</td>
<td></td>
<td>(Pearson and Hansen, 2007)</td>
<td></td>
</tr>
<tr>
<td>E22</td>
<td>Child</td>
<td>Otitis Media</td>
<td>tympanocentesis-derived isolate</td>
<td>Canada</td>
<td>AAO59379</td>
<td></td>
<td></td>
<td>(Beaulieu et al., 1988)</td>
</tr>
</tbody>
</table>

**Table 4.1. Moraxella catarrhalis** UspA1, Usp2A, and Usp2AH protein sequences used in this chapter.
Chapter 5: Characterization of *Neisseria* Opa-CEACAM interactions

Brooks MJ and Gray-Owen SD

Michael Brooks contributed all figures.
5.1 Abstract

*Neisseria* sp. possess a large repertoire of alleles encoding Opa proteins that have vast linear protein sequence variability within the surface regions of the protein. Despite this, Opa proteins maintain remarkable functional conservation. In order to investigate the molecular basis of adherence to CEACAM receptors, I examined the ability of Opa fragments to mediate adherence to CEACAM5. I demonstrated that fragments from the HV2 region of Opa could bind to CEACAM5 despite having no primary amino acid sequence conservation. I was able to translate these findings into the native architecture of Opa proteins, with the HV2 mediating both exchange and gain of CEACAM receptor binding by Opa proteins that previously were not able to adhere to CEACAMs. Interestingly, the HV1 region of the HSPG-specific Opa protein appeared to interfere with the ability of Opa to interact with CEACAMs, suggesting that CEACAM and HSPG binding are mutually exclusive. HV2 regions that mediated interactions with CEACAM did not share a consensus binding sequence, suggesting that the combination CEACAM receptors targeted and the affinity at which they interact may influence the molecular determinants of the individual binding sites. This interchange of antigenically distinct surface loops within the neisserial Opa proteins allow the bacteria to maintain CEACAM binding specificities while still evading host immune responses by continuously shuffling in antigenically disparate sequences.
5.2 Introduction

*Neisseria gonorrhoeae* and *Neisseria meningitidis* are the causative agents of gonorrhea and meningitis, respectively. Both of these species encode a family of antigenic and phase variable outer membrane proteins, known as the colony opacity-associated (Opa) proteins, which have been shown to play a primary role in neisserial infection by adhering and invading through carcinoembryonic cellular adhesion molecules (CEACAM) receptors (Chen and Gotschlich, 1996; Gray-Owen et al., 1997a; Virji et al., 1996a). Each *N. gonorrhoeae* strain possesses up to 11 unique *opa* alleles, and a single bacterium can phase-vary expression of the encoded proteins, displaying different combinations of these proteins on the bacterial surface at any given point in the infection process. The surface exposed loops of each Opa protein variant display vast primary amino acid sequence variability, yet retain the ability to bind to one or more CEACAMs via specific protein-protein interactions. Together, these features provide antigenic variation and allow avoidance of adaptive immunity while still maintaining function.

While the CEACAM residues recognized by the Opa proteins have been well described (Popp et al., 1999; Virji et al., 1999), the molecular determinants that mediate the CEACAM binding still remain elusive. The semi-variable (SV) loop has been demonstrated to be dispensable in CEACAM binding, and the particular combination of hypervariable loops 1 and 2 (HV1 and HV2) that exist within each Opa variant can influence their ability to adhere to CEACAMs (Bos et al., 2002). Interestingly, mutagenesis studies involving *N. meningitidis* Opa proteins showed that mutations in the HV1 and HV2 differentially affected the bacteria’s ability to bind to CEACAM1 or CEACAM5 (de Jonge et al., 2003). For example, Opa-CEACAM5 interactions were
disrupted by various mutations in the HV1 region, while these same mutations did not negatively affect Opa-CEACAM1 binding. The meningococcal Opa protein’s ability to adhere to CEACAM1 was instead disrupted by mutations within the HV2 region, however the sequences identified are not present within *N. gonorrhoeae* Opa proteins (de Jonge *et al.*, 2003).

The vast sequence diversity of Opa has made understanding the molecular basis by which divergent alleles can conserve function in the context of sequence hypervariability a daunting task. Previous attempts to characterize binding focused on residues or regions that when mutated resulted in loss of function, and as a result are difficult to interpret. In particular, it remains unclear as to whether residues that effect binding are critical for the structural integrity of Opa proteins or directly contact the CEACAM receptors. In the belief that Opa fragments capable of binding to CEACAM must exist, I sought to exploit differences in receptor specificity to elicit a ‘gain of function’ based approach. I expressed individual surface exposed loops from Opa proteins that retained binding and indicated that CEACAM binding was mediated by a primary binding site localized to within the HV2 region for the Opa proteins tested. I also revealed evidence to suggest that the acquisition of heparin/vitronectin binding by some variants resulted in an associated loss in their ability to bind CEACAMs, although the CEACAM binding is masked rather than lost to HSPG, suggesting that these functions that contribute to different stages of the infection process, are mutually exclusive.
5.3 Results

5.3.1 Phage-display of Opa fragments to soluble CEACAM

It was unclear whether fragments of Opa proteins would be sufficient for adherence to CEACAM, as published data demonstrated that deletion of either HV1 or HV2 resulted in loss of CEACAM (Bos et al., 2002). While this seemed to suggest that a distinct combination of loops was required for an Opa protein to interact with CEACAM, other evidence suggested that an intact protein was not required. First, due to the natural competence for genetic transformation and the frequent recombination between conserved sequences within Opa genes, Opa loops (HV1 and HV2) seemed to be interchangeable in nature, with specific functions appearing to be linked to the presence of specific loops. This was evident in both N. meningitidis and N. gonorrhoeae (Malorny et al., 1998; Virji et al., 1999). For example, distinct N. gonorrhoeae clinical isolates FA1090 and MS11 have Opa variants that share an identical HV1 region but have unique HV2 regions (Gray-Owen et al., 1997b) and J. Cannon, personal communication). These variants can display differential binding specificities for CEACAMs (Gray-Owen et al., 1997b), suggesting the molecular requirements for adherence to certain CEACAM receptors may localize to specific loops. Even prior to the identification of CEACAMs as the host receptor for Opa, synthetic peptides from HV2 of OpaB of N. gonorrhoeae FA1090 were shown to be able to block gonococci from adhering to neutrophils (Naidts et al., 1991), suggesting that these fragments, were capable of binding CEACAMs expressed by these cells.
Fig. 5-1. A. Predicted secondary structure of Opa proteins. B. Multiple sequence alignment of Opa surface exposed loops. C. CEACAM binding of phage expressing Opa surface loops. Amino acid sequence identities are indicated by red (>90%) or blue (>50%).
To investigate the molecular basis of Opa-CEACAM binding, I initially focused on determining if individual Opa loops (SV, HV1, HV2 or CL) (Fig. 5-1A) have the potential to bind to CEACAMs. I hoped to use this information to then focus in on the specific sequence of CEACAM binding domains and their location within specific Opa loops. We initially cloned the surface exposed loops (SV, HV1, and HV2) as well as the highly conserved 4th loop (CL) and a periplasmic exposed loop (IL) from the MS11 Opa proteins (Fig. 5-1B) into a M13 pVIII phagemid to examine whether individual loops were capable of adhering to CEACAM.

We began our studies by focusing on the interaction between Opa and CEACAM5, since all but one of the *N. gonorrhoeae* MS11 Opa protein variants bind to this molecule. Recombinant M13 filamentous phage expressing either SV-, HV1-, HV2- or CL-derived loops from strain MS11 (Fig. 5-1A and B) fused to the pVIII coat protein were panned over purified monomeric CEACAM5 immobilized in the wells of an ELISA plate. Rather than selecting individual alleles, our first goal was to ascertain whether any binding function could be detected. As such, these phage populations were heterogeneous, containing all 11 allelic versions of each MS11 surface loops fused to pVIII coat protein. All three loops with variable sequence (SV, HV1 and HV2) were able to mediate some binding to CEACAM5. While HV2 seemed to have largest effect on binding, this may be indicative of a primary role with minor contributions from SV and HV1, or may be related to differential avidity of the isolated fragments (Fig. 5-1C).

5.3.2 Exploiting the diversity of MS11 Opa proteins

To understand the contribution of sequences capable of binding CEACAMs, these findings need to be translated into Opa protein’s native structural architecture. While
Fig. 5-2. A. Phylogenetic comparison of MS11 opa alleles with corresponding CEACAM receptor specificities. B. Strategy for generation of Opa chimeras.
Opa loops are interchangeable in nature, previous work has demonstrated that certain combinations of loops were incapable of binding to CEACAM. While, previous studies have knocked out function by mutations in highly conserved residues (de Jonge et al., 2003). I instead sought to gain and/or exchange function by changing variable sequences. As such, I used phylogenetic analysis to minimize sequence variations while maximizing functional differences (Fig. 5-2A). The contribution of the SV loop to binding appears to be limited to a minor role since Opa variants with the same sequence have differing CEACAM binding specificities and the deletion of the SV in a variety of Opa variants did not disrupt CEACAM binding. This allowed me to focus my search for binding determinants within the HV1 and HV2 loops. From this analysis, I selected 3 Opa variants, Opa50, Opa 55 and Opa58 that have distinct CEACAM receptor specificities, but share a significant level of sequence similarity. I then generated chimeras by exchanging the SV/HV1 and HV2/CL regions between pairs of these Opa variants (Fig.5-2B). For example, a chimera was generated containing the first two surface exposed loops (SV and HV1) of Opa55 with the HV2 and CL of Opa58 which will subsequently be referred to as Opa55-58.

5.3.4 Exchange of CEACAM receptor specificity

To examine the ability of these chimeric proteins to facilitate attachment to CEACAMs, I initially focused on the chimeras created from the two most closely related but functionally distinct MS11 Opa proteins: Opa55 (binds CEACAM5) and Opa58 (binds CEACAM1, CEACAM3, CEACAM5 and CEACAM6) (Fig. 5-3A). The functionality of such chimeras was not assured: chimeras generated by combining two variants that display identical binding specificities were non-functional (Bos et al., 2002). Both of the
Fig-5.3. A. Multiple sequence alignment of Opa55 and Opa58. B. Adherence and invasion into stably transfected Hela-CEACAM5 expressing cells. C. Adherence and invasion into stably transfected Hela-CEACAM1 expressing cells. Native and chimeric Opa proteins are labeled accordingly.
native Opa proteins I selected mediated attachment to host cell via CEACAM5, so I began by testing the ability of these chimeras to mediate bacterial attachment to and invade into transfected Hela cells that express CEACAM5 (Fig. 5-3B). The chimera Opa58-55 was not able to adhere or invade into Hela-CEACAM5 cells, suggesting that interaction between the loops sterically inhibited CEACAM binding, or that resultant protein lacked CEACAM binding sequences. Interestingly, the chimera containing the SV and HV1 of Opa55 in conjunction with the HV2 region of Opa58 (Opa55-58) was able to adhere to cells in a CEACAM5-dependant manner, although invasion was at a lower level than either native Opa (Fig. 5-3B). This result was promising since it confirmed that this chimera was expressed on the bacterial surface and retained CEACAM binding function. This is the first demonstration of a chimeric Opa protein that maintains the ability to interact with CEACAM, albeit with reduced CEACAM invasion potential to the wild type. It is unclear as to the reasons for the reduced invasion, although it might be attributed to effects such as steric interference between loops or even reduced surface expression of the chimeras. I then examined the ability of these mutant Opa proteins to adhere and invade via CEACAM1, as the native Opa55 protein is unable to bind CEACAM1 (Fig. 5-3C). While the Opa58-55 again was not able to bind to CEACAM1, the Opa55-58 chimera effectively adhered to and invaded into CEACAM1 expressing Hela cells (Fig. 5-3C). This suggests that CEACAM1 binding correlates with the Opa58 HV2 region.

**5.3.6 Gain of CEACAM receptor binding**

The change in binding specificity with exchange of HV2 prompted me to test whether introduction of the HV2 region of Opa58 into an Opa that does not adhere to CEACAM
Fig. 5-4. A. Multiple sequence alignment of Opa50 and Opa58. B. Adherence and invasion into stably transfected Hela-CEACAM5 expressing cells. C. Adherence and invasion into stably transfected Hela-CEACAM1 expressing cells. D. Adherence of soluble CEACAM1 and CEACAM5 to bacteria expressing Opa chimeras. Native and chimeric Opa proteins are labeled accordingly.
could promote a gain of CEACAM binding. As I have previously mentioned, Opa50 does not bind CEACAM but instead facilitates interactions with host cells by binding HSPG and/or vitronectin. Previous studies have demonstrated that Opa50 binding to HSPG containing syndecan receptors and vitronectin allows cross-linking of syndecans and the VN-specific integrin receptors, which causes bacterial uptake (van Putten and Paul, 1995; Duensing and van Putten, 1997). I tested the ability of chimeras generated from Opa50 and Opa58 (Fig. 5-4A) to adhere and invade into Hela cells expressing CEACAM1 or CEACAM5. The chimera containing the HV2 region of Opa58 was not able to adhere or invade to either CEACAM1 or CEACAM5 (Fig. 5-4B and C). However, surprisingly, the opposite chimera bound to both CEACAM1 and CEACAM5 (Fig. 5-4B and C). I confirmed this unexpected finding by demonstrating that this Opa58-50 bound specifically to soluble forms of CEACAM1 and CEACAM5 while Opa50 did not (Fig. 5-4D). Considering our past results, this made it interesting to consider that the HV2 region of Opa50 has the capacity to bind CEACAMs, but that the Opa50 HV1 region interferes with CEACAM binding by both the Opa50 and Opa58 HV2. The crystal structure of the neisserial outer membrane protein OpcA demonstrated that the heparin binding site was mediated by a variety of positively charged residues (Prince et al., 2002). Heparin binding in Opa50 has been mapped to the HV1 region, which contains a variety of positively charged residues that are required for HSPG binding. Deletion of the other three surface exposed loops did not affect the ability of Opa50 to adhere to heparin (Grant et al., 1999). We speculated that heparin binding and CEACAM binding are mutually exclusive because the heparin binding loop interfered with CEACAM binding. Replacement of the HSPG-specific HV1 loop of Opa50 with
Fig. 5-5. A. Multiple sequence alignment of Opa55 and Opa50. B. Adherence and invasion into stably transfected Hela-CEACAM5 expressing cells. C. Adherence and invasion into stably transfected Hela-CEACAM1 expressing cells. D. Multiple sequence alignment of SV and HV1 of Opa58 and Opa50, with mutations highlighted in green. E. Bacterial pull-down of soluble CEACAM5. Whole cell lysates were probed for both CEACAM and Opa. Native, chimeric and mutant Opa proteins are labeled accordingly.
that from Opa58 would have revealed CEACAM binding function.

In order to test this theory, I decided to examine the CEACAM binding ability of chimeras of Opa55 and Opa50 (Fig. 5-5A). Opa55 only binds to CEACAM5, so I initially tested the ability of these chimeras to adhere and invade into CEACAM5 expressing Hela cells. The chimera with HV1 of Opa55 in conjunction with the HV2 region of Opa50 did not appear to significantly adhere to CEACAM5 expressing cells but did seem to promote invasion into CEACAM5 expressing cells at levels higher than into Hela cells expressing no CEACAM, albeit at levels significantly lower than the native Opa55 protein (Fig. 5-5B). If the HV2 from Opa50 was responsible for CEACAM1 binding in the context of the Opa58-50 chimeras, then I would expect Opa55-50 to bind to CEACAM despite the fact that neither Opa50 nor Opa55 bind CEACAM1. Consistent with this model, Opa55-50 expressing bacteria effectively adhered and were engulfed in CEACAM1-expressing cells (Fig. 5-5C). This result is consistent with the core CEACAM binding domain being localized to HV2.

In order to test whether the HSPG-specific HV1 Prevented CEACAM binding by HV2, I attempted to rescue binding of the chimera Opa50-58. This chimera has the HV2 region of Opa58 that promoted binding to CEACAM1 and CEACAM5, combined with the HV1 from Opa50. I decided to make substitutions with the HV1 region of this chimera, hoping to rescue binding, by slowly converting this chimera to the native Opa58. I began making substitutions in the HV1, converting the residues from Opa50 to those that exist in Opa58. Surprisingly, a switching KNL to TDR, generated an allele that bound CEACAM5 (Fig. 5-5D). This charge reversal seemed to rescue binding.
suggesting that positively charged residues within the HSPG/VN binding sequence of Opa50 HV1 interferes with CEACAM binding.

5.4 Discussion

Despite Opa-CEACAM’s importance at multiple stages in the infection cycle, the extreme hypervariability of Opa proteins has hindered the identification of CEACAM binding domains. I initially investigated the molecular requirements for CEACAM binding by analyzing if Opa fragments could mediate attachment to CEACAM. The existence of such sequences was suggested by a study in the early 1990s that demonstrated peptide fragments of the HV2 region of OpaB from FA1090 were capable of blocking binding to neutrophils (Naids et al., 1991), however more recent data suggested that non-linear sequences combine to form a CEACAM-binding surface (Bos et al., 2002; de Jonge et al., 2003). Using phage-display of the surface exposed loops from the N. gonorrhoeae MS11Opa variants, it was clear that fragments of Opa proteins were sufficient to mediate attachment to CEACAM. Phage expressing loops containing Opa HV2 regions effectively bound to CEACAM, while the SV and HV1 regions seemed to mediate a minor interaction (Fig. 5-1C). However, these results suggested that the minimal binding requirements were localized to specific loops that could independently mediate binding.

The ability of single loops to independently bind CEACAMs provided crucial insight into understanding the basis of CEACAM binding. We felt that the requirement for particular combinations of loops being required for CEACAM binding (Bos et al., 2002; de Jonge et al., 2003) was difficult to reconcile with the fact that there appears to be a random shuffling of HV1 and HV2 domains in nature, and yet nearly all variants
bind CEACAM. Past studies mutagenized strictly conserved residues to look for a loss of CEACAM binding (de Jonge et al., 2003). This strategy cannot distinguish between an absence of CEACAM binding sequences versus indirect effects due to an inappropriate protein architecture. I initially focused on the two Opa variants that were closely related but were still functionally different, and attempted to exchange the binding specificity of these proteins by switching surface exposed loops in the context of the native Opa architecture. Some chimeras tested did fail to bind to CEACAMs. However, I was able to demonstrate that the introduction of the HV2 region of Opa58 into an Opa55 background maintained the ability to adhere to CEACAM5, but also gain the ability to bind to CEACAM1 (Fig. 5-3). This seemed to link CEACAM binding to HV2, as the specificity of Opa58 was transferred to Opa55 in the Opa55-58 chimeras.

When I attempted to confirm the role of HV2 from Opa58 by introducing this region into Opa50, a protein unable to bind to CEACAMs, the chimeras failed to bind to either CEACAM1 or CEACAM5. While this was not totally surprising given the potential for interference between loops, the opposite chimera containing the SV and HV1 of Opa58 with the HV2 of Opa50, unexpectedly promoted binding to both CEACAM1 and CEACAM5 (Fig. 5-4). While, at first these findings appeared to contradict each other, Opa50 is unique in with features that could contribute to this observation. Opa50 binds to HSPG and VN, through a positively charged motif within the HV1 region and is one of only a handful of Opa variants that are not able to interact with CEACAM receptors. This led us to speculate that the HSPG-binding HV1 region of Opa50 may interfere with an inherent ability of the Opa50 HV2 region to adhere to CEACAMs. To confirm
inhibition of CEACAM binding by HV1 of Opa50, I created chimeras of Opa50 and Opa55 to see if the HV1 region of Opa55 could also promote binding to CEACAM. The introduction of the HV1 region of Opa55 into the Opa50 background allowed weak interactions with CEACAM5, but effectively bound to CEACAM1 (Fig. 5-5B and C). Neither of the native proteins were able to bind to CEACAM1, consistent with this being a masked function of Opa50 HV2. If the HV1 region of Opa50 was inhibiting the ability of the HV2 region to adhere to CEACAM in the context of the native protein framework, then I hypothesized that the chimera Opa50-58 containing the heparin binding region, was unable to bind to CEACAM for the same reason as Opa50. Changing residues in the Opa50 HV1 region to corresponding residues from the Opa58 HV1 would eventually establish binding since the conversion would eventually result in a complete Opa58 configuration. Surprisingly, one exchange resulted in a rescue of CEACAM binding (Fig. 5-5D): the change from KNL to TDR was sufficient to establish binding function. Combined, my results provide strong evidence that the binding is mediated by the HV2 region, and that HV1 can influence binding in the context of the native protein. Indeed, HSPG and CEACAM binding appear to be mutually exclusive functions.

These results clearly link the determinants for CEACAM binding to the HV2 region of the Opa proteins tested, yet the HV2 sequence of Opa proteins often display no observable consensus sequence despite a close phylogenetic relationship. Each Opa protein interacts with a different combination of CEACAMs on host cells and likely mediates these interactions with different binding affinities, suggesting that distinct CEACAM binding determinants may exist in different Opa variants. In the absence of conserved sequences, it seems likely that Opa proteins may have a propensity for
specific amino acids residues within the HV2 region, and that there is a selection for mutants that incorporate mutations that provide antigenic variation, while retaining the ability to interact with one or more CEACAMs. This is undoubtedly facilitated by the natural competence of *Neisseria* sp. for genetic transformation allowing the HV1 and HV2 regions to be continuously shuffled. As such, *Neisseria* sp. are able to efficiently avoid host adaptive immune response, in part due to their ability to maintain a large repertoire of antigenically variable Opa proteins, yet functionally related Opa variants.
5.5 Materials and Methods

Bacterial strains, plasmids and culture conditions.

M13-based phagemid pG8ASET (Accession no.AF130864) has been previously described (Zhang et al., 1999). Cloning vector ptrc99a (Accession no. U13872) has also been previously described (Amann et al., 1988). E. coli strains Top10 cells (Invitrogen) and TG-1 (Maxim Biotech) E. coli were grown at 37°C on LB medium with standard antibiotic supplementation and in LB medium with 0.01M ITPG (Sigma Aldrich) as required.

General DNA methods. The preparation of plasmid DNA and the purification of PCR products were performed using kits manufactured by QIAGEN (Santa Clarita, CA). Sequencing was done by ACGT Corporation (Toronto). Mult-Alin (Corpet, 1988) was used to generate sequence alignments.

DNA cloning. Opa loops were amplified by PCR using primers (Table. 5-1) and cloned into the phagemid pG8ASET (Zhang et al., 1999). Phagemid clones were tested for expression for recombinant pVIII expression by detection with an anti-E-tag antibody (Abcam). ptrc99a plasmids containing the opa gene from N. gonorrhoeae MS11 has been were generated by E.M. Kupsch and described previously (Gray-Owen et al., 1997b;Kupsch et al., 1996). To clone chimeric opa genes, ptrc99a Opa-expressing clones were digested with ClaI and fragments were purified and religated to generate hybrid opa gene. Mutagenesis was performed using QuikChange® Site-Directed Mutagenesis Kit (Strategene).

Phage-display. Recombinant phage was generated using methods previously described. Briefly, E. coli expressing pG8ASET phagemid vectors were infected with M13KO7
helper phage (NEB) at an MOI of 50 and incubated for 15 minutes at 37°C. Infected cells were added to 5ml of 0.5% agar, mixed and poured onto LB plates with ampicillin (100µg/ml). After incubation overnight at 37°C, top agar was removed and added to 2ml of LB with vigorous shaking for 3 hours. Samples were centrifuged at ~13000g for 10 minutes at 4°C and then filter-sterilized using 0.2µm filters. Phage was then titered by infection of TG-1 *E.coli* for 15 minutes at 37°C, serially diluted and plated on LB with ampicillin (100µg/ml). Phage stocks were then standardized based on their titers and panned over Nunc plates™ that were first coated with soluble CEACAM5 (10µg/ml) and blocked with 10% BSA. Wells were washed 3 times with PBS and eluted off using 0.1M glycine pH 2.1 and neutralized with 2.0M Tris (pH 8.7). Phage were then used to infect TG-1 *E.coli* plated on LB with ampicillin (100µg/ml) and counted.

**Measurement of bacterial attachment and invasion to Hela-CEACAM cells.**

Overnight cultures of *E. coli* expressing *N. gonorrhoeae* Opa proteins were grown in LB amp (100µg/ml) with IPTG. Cells were washed with RPMI. Stably transfected Hela cells expression CEACAM1, CEACAM5 or no CEACAM (NEO- empty vector), were seed in 24-well plates to a confluence of ~2 x 10⁵ cells. Bacteria were added at an MOI of 100, centrifuged at 500rpm at room temperature for 5 minutes then incubated at 37°C with 5% CO₂ for 2 hours. For invasive bacteria, gentamicin was added at 50µg/ml for 2 additional hours. Cells were washed 5 times with RPMI. Cells were lysed in 1.0% saponin for 15 minutes at 37°C with 5% CO₂. Lysates were then diluted in PBS and plated on LB agar with Amp 100µg/ml. Recovered bacteria were then counted.

**Soluble CEACAM binding assays.** Overnight cultures of *E. coli* expressing, *N. gonorrhoeae* Opa proteins were grown in LB amp (100µg/ml) and IPTG (0.01M). Cells
were washed with PBS/Mg/Ca (10 mM MgCl$_2$ and 5 mM CaCl$_2$). For ELISAs, Nunc
plates$^\text{TM}$ were coated with heat-killed bacteria and dried at 37°C overnight. Wells were
blocked using PBS containing 4% skim milk, followed by incubation with soluble
CEACAM5 or CEACAM1. CEACAM was detected by anti-CEACAM antibodies
(Dako) followed by goat anti-rabbit conjugated to horseradish peroxidase (HRP). HRP-
based colorimetric detection system (Sure Blue$^\text{TM}$) was used to measure CEACAM
binding by detection at 650 nm.

**Pull down assays and western blot.** *E.coli* expressing Opa variants at 1x10$^8$ cells/ml
were washed 2 times in PBS/Mg/Ca. 500µl of 2µg/ml CEACAM5 (Dako) in
PBS/Mg/Ca was added to cells and the cells were incubate at 4°C for 3hrs
spinning/shaking. Bacteria were then washed 3x with PBS/Mg/Ca at 4°C, lysed in 50µl
of SDS lysis buffer with BME and then whole cell lysates were boiled and resolved by
SDS-PAGE in 12.5 % (wt/vol) polyacrylamide separating gels, transferred to
Immobilon-P membranes (Millipore, Bedford, MA) and probed with the appropriate
antibody. Opa was detected using monoclonal antibody (MAb) 4b12c11, CEACAMs
were detected using CEA-specific rabbit polyclonal antisera (DAKO, Glostrup,
Denmark). The secondary antibodies were goat anti-mouse IgG or anti-rabbit IgG
conjugated to horseradish peroxidase (Jackson ImmunoResearch, West Grove, PA).
Antigen-antibody complexes were visualized by Chemiluminescence, achieved via the
use of the Western Lightning Chemiluminescence Reagent Plus (New England Nuclear,
Boston, MA).
<table>
<thead>
<tr>
<th>Primers</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SV-F</td>
<td>CACCCATGGTTCGATAGCCTACGCCGCC</td>
</tr>
<tr>
<td>SV-R</td>
<td>CACCCATGGCGAGATGGAATGCGTACGGAT</td>
</tr>
<tr>
<td>HV1-F</td>
<td>CACCCATGTTTCGAATAGCGCAGATTATGCC</td>
</tr>
<tr>
<td>HV1-R</td>
<td>CACCCATGGCGAGAGAAACGCGTGGA</td>
</tr>
<tr>
<td>HV2-F</td>
<td>CACCCATGTTTCGCGCGTCGCTAC</td>
</tr>
<tr>
<td>HV2-R</td>
<td>CACCCATGCGACTGCGCCAGGCCAC</td>
</tr>
<tr>
<td>CL-F</td>
<td>CACCCATGTTTCGACGCCGACGTACCGCTAC</td>
</tr>
<tr>
<td>CL-R</td>
<td>CACCCATGGCGAGTACGCACGCCCAATGGGC</td>
</tr>
<tr>
<td>IL-F</td>
<td>CACCCATGTTGCCCCTCTTTCTCTCGGCTT</td>
</tr>
<tr>
<td>IL-R</td>
<td>CACCCATGGCGAGCGCGATAGGGTTTGAA</td>
</tr>
</tbody>
</table>

**Table 5-1.** Primers for amplification of Opa surface loops
Chapter 6: Summary and Future directions.
6.1 Summary

The number of pathogens that have been recognized to specifically target CEACAMs on host cells is steadily growing, and interactions with CEACAMs have been demonstrated in a variety of host tissues, cell types in vitro (Chen and Gotschlich, 1996; Virji et al., 1996a; Gray-Owen et al., 1997a; Gray-Owen et al., 1997b; Boulton and Gray-Owen, 2002; Hill et al., 2001; Hill and Virji, 2003; Berger et al., 2004; Barnich et al., 2007) and, more recently in vivo (Bookwalter et al., 2008). My studies have focused on understanding the molecular basis of bacterial attachment to CEACAMs by *M. catarrhalis* and *Neisseria* sp. I set out to characterize the molecular determinants for CEACAM binding within UspA1 and Opa and then compare and contrast CEACAM binding by *M. catarrhalis* and *Neisseria* sp. It is now evident *M. catarrhalis* and *Neisseria* sp. use two structurally and phylogenetically unrelated adhesins to bind CEACAMs on host cells; UspA1 and Opa proteins, respectively. Amazingly these pathogens also have a homologue of each other’s CEACAM-binding adhesins, the Opa-like protein A (OlpA) protein of *M. catarrhalis* (Chapter 1) and the *Neisseria* sp. NadA proteins are structural homologues of Opa and UspA1, respectively, yet neither bind CEACAMs. It is remarkable that these pathogens have convergently acquired CEACAM binding in distinct adhesins, despite encoding structurally similar proteins.

Once it was clear that these two different adhesins mediated CEACAM binding, I sought to understand the molecular determinants responsible for these specific protein-protein interactions. I identified sequences within UspA1 required for adherence to CEACAM (Chapter 2) and also sequences that facilitate binding of Opa to CEACAMs on host cells (Chapter 4). In both cases, the binding determinants appeared to be
variable: UspA proteins exist as modular structures comprised of cassettes with sequences linked to differential functions, including CEACAM binding (Chapter 3), while Opa proteins possess antigenically variable loops that retain CEACAM binding (Chapter 4). The CEACAM binding regions of these two adhesins exist in completely divergent protein architectures, with the CEACAM binding regions in UspA1 existing in the context of an α-helical coiled coil structure while Opa sequences are localized to surface exposed loops that are presumably flexible in nature.

UspA1 proteins also do not appear to have different receptor specificities, as they either contain a highly conserved sequence cassette that mediate CEACAM binding or lacks this sequence (Chapter 3). In contrast, the variable Opa sequences each bind various combinations of CEACAMs, possibly with differential binding affinities. In addition, *M. catarrhalis* strains contain a single UspA1-encoding loci, while gonococci typically encode up to 11 independently regulated opa alleles encoded by separate loci spread across their chromosomes.

During my exhaustive efforts to identify sequence motifs that were conserved among CEACAM-binding adhesins, I was surprised to find a significant level of sequence identity between the CEACAM-binding regions of Opa58 and UspA1 (Fig.6-1). The level of identity was striking, especially given that my comparison of Opa variants has failed to yield any linear sequence conservation. It seems remarkable that sequence conservation exists in the context of these two distinct protein architectures, while it doesn’t appear to exist within the Opa family. The residues that differ between UspA1 and Opa58 HV2 CEACAM binding regions appear to be localized to the tip of
Fig. 6-1. Comparison of predicted structures and CEACAM binding sequences of Opa58 and UspA1.
the Opa58 HV2 loop, as opposed to the residues that exist in the stalk region of UspA1. The architectural context in which these two binding regions exist suggest that a linear sequence determinant rather than a complex binding pocket most likely mediates CEACAM binding.

Identification of a consensus CEACAM binding domain within Opa proteins has remained elusive. Why has it been so difficult to establish a consensus for CEACAM binding? It appears that the requirement for binding to CEACAMs are dynamic: apparently unrelated sequences appear to mediate binding. This presumably explains why different variants bind to different combinations of CEACAMs, however variants with the same specificity are not necessarily related. For example, Opa52 and Opa58 both bind CEACAM1, CEACAM3, CEACAM5 and CEACAM6, yet display little sequence similarity. While some CEACAM residues are required for binding are conserved, other residues are important for binding to a subset of Opa variants suggesting that primary contact points may change. Constant evolution of Opa sequences to avoid the adaptive immune response must be balanced by selection for CEACAM binding. Considering that CEACAM specificity of individual Opa variants, it is interesting many of the cell types encountered during bacterial infection have multiple different CEACAMs on the surface, perhaps indicating a redundancy exploited by the bacteria to generate functionally identical proteins that appear to be distinct. However it must be considered that the distinct functions assigned to each CEACAM has the potential to differentially affect the cell response.

Differences in pathogenesis may determine why *Neisseria* sp. have adapted to contain an array of antigenically Opa proteins with the ability to bind to CEACAM,
while \textit{M. catarrhalis} have been able to retain a highly conserved adhesin sequence. In this context, it is important to consider that all \textit{Neisseria} sp. bind CEACAMs, while our studies revealed clinical isolates of \textit{M. catarrhalis} that lacked CEACAM binding sequences and were unable to adhere to this receptor. Whether this correlates with different cell tropism and/or pathology remains to be explained. As an example, it is intriguing to speculate that colonization of the nasopharynx may require CEACAM binding, but that movement into lungs could either be promoted by (or allow) the loss of CEACAM binding.

\section*{6.2 Future directions.}

\textit{M. catarrhalis:}

The expanding role for CEACAMs in a variety of distinct infections makes future studies into the molecular basis of CEACAM binding enticing. From my studies there are many obvious avenues to pursue for future research. Fragments of UspA1 protein containing the CEACAM binding determinants are sufficient to interact with CEACAMs (Hill \textit{et al.}, 2005). What remains unclear is the molecular basis of this interaction and the minimal sequence required for binding. Identification of the minimal sequence requirements and subsequent attempts to co-crystallize this with the CEACAM N-domains would provide concrete information that would define the molecular basis of this interaction and aid in development of bacterial-based CEACAM therapeutics.

The specific contribution of CEACAM during \textit{M. catarrhalis} infection also remains unclear. Recognizing that CEACAM-specific antibodies protect chinchillas from \textit{H. influenzae} otitis media provides an exciting avenue to investigate the role of UspA1-CEACAM interactions during infection \textit{in vivo} (Bookwalter \textit{et al.}, 2008). How
CEACAM binding affects *H. influenzae* tropism within the upper respiratory tract, whether CEACAM binding allows bacterial entry into and/or transcytosis across mucosal epithelia and whether CEACAM binding influences the host cellular response to infection must be explained. If these animals also prove susceptible to *M. catarrhalis* infections, they will undoubtedly prove to be an important tool in understanding the infection process by this opportunistic pathogen. In particular, I observed that some *M. catarrhalis* clinical isolates express UspA1 proteins that mediate interactions with CEACAM, while others do not. *In vivo* analysis of the ability of these *M. catarrhalis* strains to colonize and establish infections in chinchillas or other animal models could provide insight in delineating the contribution of CEACAM binding during *M. catarrhalis* infection and disease. One might expect, for example, that CEACAM binding is essential for colonization of the upper respiratory tract but that this function would be dispensable for infections deeper in the respiratory tract if mucus clearance is compromised. An *in vivo* model for *M. catarrhalis* infection would also allow us to investigate the role of the *M. catarrhalis* OlpA protein during infection and perhaps elucidate its function. OlpA is constitutively expressed and highly conserved amongst *M. catarrhalis* clinical isolates, yet we have generated isogenic mutants that do not express OlpA that appear normal *in vitro*. I am intrigued by the possibility that their phenotype will emerge during an *in vivo* infection.

Neisserial Opa mediated engagement of CEACAMs results in a variety of effects in different human cells types, inducing invasion into non-phagocytic cells, opsin-independent engulfment by professional phagocytes, and inhibitory signaling within various leukocytes. Opa proteins were the first bacterial adhesins recognized to bind
CEACAMs and remain the paradigm for understanding the effect of CEACAM binding on host cellular response. However it must be considered that the context in which CEACAM binding regions are presented may induce differential effects. For example, does the large fibrillar UspA1 protein allow transcellular transeptosis across polarized epithelia and/or T-cell immune suppression in a manner reminiscent of the neisserial Opa proteins?

**Neisseria Opa proteins:**

Opa proteins adhere to CEACAMs despite lacking sequence conservation and have proven a difficult system in which to characterize the molecular requirements of CEACAM binding. I demonstrated that fragments of Opa proteins contain the sequence determinants sufficient for binding to CEACAM and fragments from different Opa proteins can interact with varying combinations of CEACAMs receptors. I also demonstrated that the context in which these loops exist may influence their binding in that the HSPG-specific Ops50 HV1 hinders CEACAM binding by region. With respect to the experiments in Chapter 5, the most obvious next experiment would be to confirm that Opa50 has inherent CEACAM binding ability. The mutations demonstrated to alleviate HV1 mediated inhibition in the Opa50-58 chimeras should be made in the context of Opa50, both to reveal the inherent CEACAM binding by Opa50 and confirm inhibitory role of the HV1 region.

In establishing that HV2 of each of the proteins tested appear to facilitate CEACAM binding, I can further test the molecular determinants for specific combinations of receptors by exchanging residues in the HV2 region of various Opa
proteins such as Opa55 and Opa58 to identify residues that may allow for adherence to different CEACAMs.

Additionally it would be useful to characterize binding by measuring binding affinities and specificities for a variety of isolated and recombinantly expressed Opa HV loops. The shortening of fragments should yield minimal fragments, and while these fragments may have unique sequences, it may allow us to establish the composition of physical properties that exist for adherence to differing combinations of CEACAMs. However, I believe that to truly understand the molecular requirements for Opa-CEACAM, we may need to produce a variety of co-crystal structures of Opa-CEACAMs and/or Opa fragments with CEACAM N-domains.

**CEACAM therapeutics:**

While understanding how a variety of different pathogens exploit CEACAMs during infections is obviously important, the isolation of CEACAM binding fragments presents an exciting therapeutic tool for a variety of conditions. CEACAM receptors have important physiological roles in contact-dependent regulation of growth and proliferation, and are up regulated in a variety of cancers. The existence of various bacterial-derived CEACAM binding fragments capable of interacting with different CEACAMs could prove useful in design of a variety of different therapeutics that target specific CEACAMs to elicit favorable cell responses.

I have previously mentioned that both CEACAM5 and CEACAM6 are up-regulated in a variety of cancers. Peptide fragments for CEACAM5 that prevent intercellular CEACAM5-CEACAM5 binding effectively block the transformative function of CEACAM5, yet immune responses generated to the recombinant peptide
Fig. 6-2. Inhibition of proliferation of purified CD4+ T-cell. CD4+ T-cells were stimulated with anti-CD3/CD28 and IL-2, and then incubated with phage expressing Opa loops, helper phage, anti-CEACAM, Isotype Ab or with neither antibody/phage. Proliferation was measured by counting at various 24 hour time points up to 96 hours after stimulation (In collaboration with Hannah Lee).
would be anti-self. CEACAM5-specific antibodies have also been used to target tumors with radioactive isotopes (J. Shively, personal communication) however anti-idiotypic immunity has prevented further development. Identification of multiple different bacterial peptides that can target tumors that over-express CEACAM5 have important therapeutic value. The ability to use peptides as a delivery system that can specifically target cancer cells is an exciting and potentially important advance in directed cancer therapeutics.

CEACAM1 also has been shown to induce immunosuppressive effects in CD4+ T-cells (Boulton and Gray-Owen, 2002). By developing bacterial based therapeutics that specifically target CEACAM1, coupled with the understanding the molecular nature of CEACAM-mediated inhibition of proliferation and activation of lymphocytes, we may ultimately be able to design immune-modulatory drugs with the potential to benefit patients with auto-immune disease. This is illustrated by a preliminary experiment done through collaboration with Hannah Lee, where phage expressing HV1 and HV2 from Opa 52 and Opa57 were tested for their ability to prevent CD4+ T-cell activation and proliferation. To our surprise, these experiments showed great promise, as phage expressing recombinant Opa loops suppressed T-cell activation to a significant level in comparison to helper phage, a level indistinguishable for the direct ligation of CEACAM1 by our “gold standard” antisera (Fig. 6-2).

My studies have yielded important information regarding the molecular basis of CEACAM binding by diverse human-restricted bacterial pathogens. I began these studies with antigenically and structurally distinct bacterial adhesins and now have shown that the CEACAM binding determinant lie within linear amino acid sequences in
each of these adhesins. The sequences identified have the potential to impact multiple infectious diseases, autoimmunity and cancers and future studies will hopefully focus on the development of these important CEACAM-based therapeutics.
Reference List


Helminen, M.E., Maciver, I., Latimer, J.L., Cope, L.D., McCracken, G.H., Jr., and Hansen, E.J. (1993) A major outer membrane protein of *Moraxella catarrhalis* is a target
for antibodies that enhance pulmonary clearance of the pathogen in an animal model.  

*JID* **170**: 867-872.

*Microbiology and Molecular Biology Reviews* **68**: 692-744.


*EMBO J* **19**: 5989-5999.


*Cancer Res* **55**: 190-197.


Ref Type: Journal (Full)


