MOLECULAR BASIS OF PNEUMOCOCCAL ADHERENCE AND COMPLEMENT EVASION: STRUCTURAL AND BIOCHEMICAL STUDIES OF PNEUMOCOCCAL VIRULENCE FACTOR, CBPA

By

David Otieno Achila

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Biochemistry and Molecular Biology - Doctor of Philosophy

2013

ABSTRACT

MOLECULAR BASIS OF PNEUMOCOCCAL ADHERENCE AND COMPLEMENT EVASION: STRUCTURAL AND BIOCHEMICAL STUDIES OF PNEUMOCOCCAL VIRULENCE FACTOR, CBPA

By

David Otieno Achila

Streptococcus pneumoniae is a human adapted pathogen of global importance as a leading cause of a wide spectrum of infections, such as pneumonia, bacteremia, meningitis, otitis media, and sinusitis. *S. pneumoniae* is highly resistant to host innate immunity during nasopharyngeal colonization and invasive infections. Because of pneumococcal strain variations, the vaccines currently available for pneumococcal infections are only effective against a subset of strains and do not provide universal protection. Promising novel antigens have been identified including CbpA, a surface protein whose gene is present in all characterized virulent *S. pneumoniae* strains. *S. pneumoniae* recruits FH, a negative regulator of complement system, to its surface and therefore evades complement-mediated clearance. CbpA binds to the human complement factor H (FH), but not to the FH proteins from other animal species that have been tested to date, including mouse and rabbit. In this study, we show for the first time that 9th domain of FH (FH CCP9) of the 20 FH domains is crucial and sufficient for its interaction with CbpA and that the interaction is very tight with a K_d of ~ 5 nM.

We also present a 1 Å crystal structure of the FH9 in complex with N-terminal α-helical domain of CbpA (CbpAN). We also found evidence of a hydrophobic core defined by four FH CCP9 residues and one CbpAN residue at the complex interface

which may be responsible for host-specificity. These findings will enhance the understanding of the host-specificity of this host-pathogen interaction and also inform creation of better animal models for studying pneumococcal diseases. The insights gained will also benefit design of universal protein-based vaccines and novel therapeutics for the treatment and prevention of otitis medium and other infections caused by pneumococcus.

To the late Wilfred and Ken Achila,

with gratitude and love

ACKNOWLEDGEMENTS

I am humbled by the huge number of people who have been by my side through my amazing journey at Michigan State University over the past few years. I sincerely want to thank the Lord, all my mentors, family, colleagues and friends for their support during this time. First, I am extraordinarily thankful to my mentor, Dr. Yan for believing in me and for his inspiration that made us achieve so much in a short time. As I step out I will not forget his logical approach to science and humor that made working with him an invigorating experience. I am very grateful to the amazing members of my dissertation committee: Dr. Wang, Dr. Burton, Dr. Vieille and Dr. Sheng Yang-He, for their charismatic and diligent guidance. I also want to acknowledge Dr. Robert Britton, Dr. Charles Hoogstraten, Dr. Jon, S., and Dr. Jon Kaguni for their contributions towards making me the scientist I am today.

In Dr. Yan's lab, I worked alongside a wonderful team (DR. Liu, Dr. Yue, Yan, Rahul, Yitchen and Bell) who taught me so much about science and humble ways of life, thanks to all of you. During my training at MSU, I worked with and mentored some outstanding undergraduate students (Bell, Corine, Marci, Anjelica, and Loren) who ended up contributing in some way to my work, thanks to all of you. I am very grateful to my dear parents, Nerea and the late Wilfred; and to the rest of my family members; Wilfred, Macaella, Nancy, Festus, George, Ken (the late), Ruth, Rose, Noel, Wiki, Eric, Sophy and Jim among others for supporting me throughout this long journey. Special thanks to my wonderful friends: Patrick, Joel, Kamba, Ruth, Akinyi, Obuya, Nikhil, Megha and many more whom I interacted with in one way or the other during this time.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

AILME	*Ailuropoda melanoleuca* (giant panda)
ALF	apolactoferrin
BgaA	β-galactosidase
CALJA	*Callithrix jacchus* (white-tufted-ear marmoset)
CBD	Choline binding domain
CbpA	Choline binding protein A
CbpAN	N-terminal domain of CbpA
CbpAR2	N-terminal R2 domain of CbpA
ChoP	Phosphorylcholine
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
Eno	enolase
FH	Factor H, Negative regulator of complement system
GAE	GTPase activating element
GAP	GTPase activating protein
GMPPNP	guanosine 5'-[β,γ-imido] triphosphate
HETGA	*Heterocephalus glaber* (naked mole rat)
hFH	Human FH
Hic	Factor H binding inhibitor of complement
Hyl	hyaluronate lyase
IgA	IgA1 protease
ITC	Isothermal Calorimetry
IPTG	isopropyl *β*-thiogalactoside
LOXAF	*Loxodonta africana* (African elephant)

LytA	autolysin A
MACFA	*Macaca fascicularis* (crab-eating macaque)
MACMU	*Macaca mulatta* (*Rhesus macaque*)
mFH	Mouse FH
Nan	neuraminidase
NOMLE	*Nomascus leucogenys* (Northern white-cheeked gibbon)
OTOGA	*Otolemur garnettii* (small-eared galago)
PavA	pneumococcal adhesion and virulence A
PbcA	Pneumococcal C3 binding protein
PiaA	pneumococcal iron acquisition A
PiuA	pneumococcal iron uptake
pIgR	Polymeric immunoglobulin receptor
Ply	pneumolysin
PONAB	*Pongo abelii* (*Sumatran orangutan*)
pppGpp	guanosine pentaphosphate
PsaA	pneumococcal surface antigen A
PspA	pneumococcal surface protein A
RA-GTPase	Ribosome associated GTPase
SC	Secretory component
SpsA	*S. pneumoniae* secretory IgA binding protein

CHAPTER 1

Background Information

1.0 Introduction

This study investigated the biochemistry and structural basis of the interactions of a pneumococcal surface protein, choline binding protein A (CbpA), with human Factor H and secretory component proteins. These interactions directly contribute to the pathogenesis of *Streptococcus pneumoniae* (*S. pneumoniae*).

1.1 *Streptococcus pneumoniae*

1.1.1 Pneumococcal diseases

Streptococcus pneumoniae is a Gram-positive bacterial human pathogen that commonly colonizes the upper airway and the nasopharynx of humans (Musher, 2009b). There are over 91 serotypes of *S. pneumoniae* (pneumococcus) identified so far based on the antigenic differences of their capsules. Colonization of upper airway by pneumococcus is asymptomatic, but it can cross mucosal barriers and gets into the normally sterile regions of the airway, the blood stream, and even the brain, resulting in rapid inflammation and disease. Pneumococcus can cause not only non-life-threatening infections (but with enormous economic burdens) such as otitis media (Rovers, 2008) but also life-threatening invasive infections such as pneumonia, meningitis and bacteremia, resulting in significant morbidity and mortality, and enormous economic burden.

According to a recent detailed analysis of the morbidity, mortality and economic burden of pneumococcal infections in the USA (Huang et al., 2011), it is estimated that, "in 2004, pneumococcal disease caused 4.0 million illness episodes, 22,000 deaths, 445,000 hospitalizations, 774,000 emergency department visits, 5.0 million outpatient

visits, and 4.1 million outpatient antibiotic prescriptions. Direct medical costs totaled $3.5 billion. Pneumonia (866,000 cases) accounted for 22% of all cases and 72% of pneumococcal costs. In contrast, acute otitis media and sinusitis (1.5 million cases each) comprised 75% of cases but only 16% of direct medical costs. Patients ≥65 years old, accounted for most serious cases and the majority of direct medical costs ($1.8 billion in healthcare costs annually). In this age group, pneumonia caused 242,000 hospitalizations, 1.4 million hospital days, 194,000 emergency department visits, 374,000 outpatient visits, and 16,000 deaths. However, if work loss and productivity are considered, the cost of pneumococcal disease among younger working adults (18–50) nearly equaled those ≥65." According to a recent estimation by Centers for Disease Control and Prevention (Prevention, 2012), pneumococcal infections account for estimated 175,000 cases of pneumonia, 50,000 cases of sepsis, and 3,000 to 6,000 cases of bacterial meningitis annually in the USA (Prevention, 2012). Pneumococcus is probably the number one killer of young children in the world, killing at least 1.2 million each year worldwide (Murdoch, 2008; Prevention, 2012; van der Poll and Opal, 2009).

1.1.2 Antibiotic therapy and vaccination

Pneumococcal infections are usually treated by administration of antibiotics (Musher, 2009a). Widespread use of antibiotics has, however, led to a rapid emergence of drug resistant isolates from the USA and other parts of the world (Musher, 2009a). Up to 50% of pneumococcal isolates in the US are resistant to penicillin (Farrell et al., 2007; Jacobs et al., 2008), the drug of first choice for treatment of pneumococcal disease (Musher, 2009a). Many of these isolates are also resistant to other antibiotics,

with some responding only to vancomycin, a drug of last resort in many clinics (Gamez and Hammerschmidt, 2012). Pneumococcal isolates with vancomycin resistance have also been reported earlier (Novak et al., 1999) and this situation reasserts the urgent need for novel therapeutics against pneumococcal infections. There is need to characterize the pneumococcal virulence factors in details with an intention of designing novel anti-virulence drugs that may help overcome the prevalent antibiotic resistance.

Vaccination has been successfully used to prevent pneumococcal infections for over three decades. Currently there are two licensed pneumococcal vaccines in the USA, a 23-valent polysaccharide vaccine (PPSV23; Pneumovax 23, Merck) and a 7-valent conjugate vaccine (PCV7; Prevnar®, Wyeth Pharmaceuticals Inc.), and several others are still in development (Moffitt and Malley, 2011; Pletz et al., 2008). The 23-valent pneumococcal vaccine is composed of capsular polysaccharides of 23 serotypes of pneumococci (1, 2, 3, 4, 5, 6B, 7F, 8, 9N, 9V, 10A, 11A, 12F, 14, 15B, 17F, 18C, 19A, 19F, 20, 22F, 23F and 33F), which cause about 90% of invasive pneumococcal disease in the USA (Prevention, 2012). Vaccination with this vaccine leads to production of serotype-specific antibodies, which coat pneumococci of corresponding serotypes and enhance phagocytosis by professional macrophages and polymorphonuclear leukocytes (Musher, 2009a; Pletz et al., 2008). The 23-valent polysaccharide vaccine has 60-80% efficacy in protecting adults with a healthy immune system, but is less effective in protecting immunocompromised patients and ineffective in protecting children less than 2-years old (Lynch and Zhanel, 2009), a most vulnerable population. The 7-valent conjugate vaccine, on the other hand, is a combination of polysaccharides from seven serotypes of pneumococci (4, 6B, 9V, 14, 18C, 19F, and

23F) conjugated to a non-toxic diphtheria toxin mutant protein CRM197 and covers about 90% of invasive pneumococcal diseases. These seven pneumococcal serotypes are commonly found in younger patients and healthy children, with an 80% coverage at the launch of the vaccine. This vaccine provokes a T-cell-dependent immune response and therefore antibody levels can be enhanced by repeat immunization with polysaccharide vaccines (Musher, 2009a; Tai, 2006). Although PCV7 vaccine has a level of coverage lower than that of PPSV23 vaccine, it is more effective against pneumococcal infections, with an efficacy of 97% against invasive pneumococcal infections of covered serotypes (Prevention, 2012).

However, the 7-valent conjugate vaccine has several significant limitations (Black et al., 2008; Kadioglu et al., 2008a). First, this vaccine has had a moderate impact on *S. pneumoniae* otitis media and carriage. Second, it covers only 7 out of at least antigenically distinct 91 serotypes, with coverage ranging from 22% in Bedouins, Israel to 88% in Hertfordshire, UK, the latter including potential cross-coverage within serogroups. Third, due to the limited serotype coverage, non-vaccine serotypes are steadily replacing the vaccine serotypes (serotype replacement) (Song et al., 2012; Weinberger et al., 2011). As a result, some clinically rare serotypes before vaccination now have major contributions. Fourth, it is technically challenging to include a larger number of different polysaccharides in a conjugate vaccine; the highest coverage of the current conjugated vaccine candidates is 13 serotypes. Finally, because of the complexity of conjugate vaccine production as well as probably the monopoly of the licensed conjugate vaccine, the cost of this vaccine remains high in comparison with other routine infant/child vaccines. The cost of a full regimen of the conjugate vaccine is

prohibitive for developing countries (Black et al., 2008). Therefore, there is an urgent need to develop a vaccine with a broader serotype coverage and effective for both invasive and non-invasive pneumococcal infections.

A promising alternative approach to the development of new-generation vaccines is the use of pneumococcal protein antigens (Barocchi et al., 2007; Gamez and Hammerschmidt, 2012; Moffitt and Malley, 2011). CbpA, a major pneumococcal virulence factor and a subject protein of this work, is an excellent vaccine candidate for development of protein antigen-based vaccines.

1.2 Complement System

The complement system is the first line of defense and plays a crucial role in innate immunity against invading microbes (Ricklin et al., 2010; Serruto et al., 2010; Zipfel et al., 2008). The human complement system consists of over 30 soluble and membrane bound proteins. The activation of complement yields proteins and products whose core functions are: (i) opsonize invading microbes for removal by phagocytosis, (ii) lyse the invading microbes by formation of membrane attack complex on their surfaces, and (iii) serve as a bridge between innate and adaptive immunity. The other important functions of activated complement include removal of immune complexes and clearance of cell debris and damaged cells.

The complement system is organized into three pathways, each with a cascade of enzymes, referred to as the classical, the lectin and the alternative pathways, each activated by a different mechanism (Ricklin et al., 2010; Serruto et al., 2010; Zipfel et al., 2008). The first two pathways are activated by specific mechanisms that recognize foreign invaders. The classical pathway is activated by antibody recognition of invading

microorganisms, i.e., formation of a particular antibody-antigen complex, such as a complex of natural IgM and phosphorylcholine on the cell wall of *S. pneumoniae*. Recognition of specific polysaccharide patterns on the surface of the bacteria triggers the lectin pathway. In contrast, the alternative pathway does not have a specific mechanism recognizing self from non-self and is always active, albeit at a low level, due to the slow, spontaneous hydrolysis of complement component C3. The alternative pathway is stimulated by invading microorganisms but suppressed by host proteins such as factor H (FH) and factor I so that autoimmunity is avoided. Upon activation, all the three pathways converge at the formation of C3 convertases, which amplify the cleavage of C3 into C3b (opsonin) and C3a (anaphylaxin) (Harboe and Mollnes, 2008; Ricklin et al., 2010).

Complement activation results in deposition of C3b on the bacterial surface followed by phagocytosis. Sustained complement activation, however, leads to the formation of C5 convertase which can bind and cleave C5 into C5a and C5b. C5b initiates the terminal pathway that brings together C5b, C6, C7, C8 and C9 to form a membrane attack complex, which directly kills bacteria (Serruto et al., 2010). Complement kills Gram-negative bacteria such as *Neisseria meningitidis* by both means, facilitation of phagocytosis and formation of the membrane attack complex, whereas it kills Gram-positive bacteria like *S. pneumoniae* only by promoting phagocytosis, because the thick peptidoglycan layer of Gram-positive bacteria prevents the membrane attack complex from penetrating into the membrane (Serruto et al., 2010).

S. pneumoniae activates the complement system mainly via the alternative and classic pathways (Musher, 2009b). The lectin pathway plays a less important role because mannose-binding lectin binds poorly to *S. pneumoniae* (Brown et al., 2002; Neth et al., 2000). The ability of *S. pneumoniae* to thrive and cause disease is dependent on its capability to avoid host phagocytosis promoted by the complement system (complement evasion). CbpA, the subject protein of this work, plays an important role in complement evasion.

The immune response is mostly accompanied by an inflammatory response and therefore may act as a double-edged sword (Zipfel et al., 2008). Over-activation of the complement system can lead to tissue damage, and therefore host cells need protection from the complement activity. The host cells are protected both by membrane-bound and fluid phase complement regulators such as complement factor H (FH) (Meri and Pangburn, 1990).

1.3 Factor H and its roles in complement regulation

1.3.1 Structure and function of FH

While the alternative pathway is the major pathway for complement activation, this pathway does not have a mechanism to distinguish self from non-self. One way to control autoimmunity is to produce proteins that inhibit the alternative complement activation pathway and prevent the deposition of C3b to host cells and tissues. FH protein plays both roles (Schmidt et al., 2008), and is produced mainly in the liver and also locally by a variety of cell types. FH is a large glycoprotein (~155 kDa) composed of 20 short consensus repeats (SCR) or complement control protein (CCP) modules, ~ 60 residues each, connected by linkers of 3-8 residues long. A multiple-sequence

alignment of the 20 CCPs of FH reveals four absolutely conserved Cys residues and one highly conserved Trp residue (Schmidt et al., 2008). The four Cys residues form two disulfide bonds, in the arrangement of Cys(I)-Cys(III) and Cys(II)-Cys(IV). Three-dimensional structures have been reported for all but six CCP modules. High resolution X-ray crystal structures are available for CCP1-4, CCP6-8 and CCP19-20 constructs, while solution NMR structures are available for CCP1-3, CCP5, CCP7, CCP12-13, CCP15-16 and CCP19-20 (Perkins et al., 2012). The CCP modules with no structure reported are CCP5, CCP 9-11, CCP14, and CCP17. All reported CCP structures have an ovoid shape stabilized by two disulfide bonds. The structures are composed of only β-strands and loops, with no helix. Adjacent modules in multi-module structures are arranged in a head-to-tail manner (Schmidt et al., 2008).

FH inhibits the alternative complement activation pathway in four ways (Schmidt et al., 2008): (i) inhibiting the formation of the precursor (C3bB) of the alternative-pathway C3 convertase (C3bBb) by competing with factor B for the binding of C3b, (ii) accelerating the disintegration of the C3 convertase by promoting the dissociation of Bb, (iii) functioning as a co-factor for factor I-mediated degradation and inactivation of C3b, and (iv) promoting the disintegration of the C5 convertase. All these FH actions are dependent on its binding C3b. The plasma concentration of FH is ~500 mg/L (3.2 μM), although it can vary widely in the range of 116-810 mg/L (0.8-5 μM), and is comparable to the concentration (~7 μM) of the precursor protein (C3) of C3b and significantly higher than the concentration (0.1 μM) of C3b (Perkins et al., 2012). The reported K_d value for binding of FH to C3b is ~1 μM. Two regions of FH, CCP1-4 and CCP19-20, can bind to C3b, with a K_d value of either region in the low micromolar range, ~12 μM

for CCP1-4 and ~4 μM for CCP19-20. Crystal structures have been reported for the complexes of CCP1-4 with C3b (Wu et al., 2009) and CCP19-20 with C3d (Herbert et al., 2012; Kajander et al., 2011; Morgan et al., 2011), which is a domain of C3b.

FH protect host cells and tissues from complement attack by binding to polyanionic markers such as heparin and glucosaminoglycans on the surface of host cells (Schmidt et al., 2008). FH recognition of host cell markers serves as a mechanism for distinguishing self from non-self for the alternative complement activation pathway. FH has two main binding sites for host polyanionic markers, one within CCP6-8 and the other within CCP19-20 (Perkins et al., 2012). The affinity of FH for host polyanionic markers is not known at present. Relevant reported K_d values include 9.2 nM for FH binding to commercial heparin (Yu et al., 2007), 9 μM for CCP19-20 binding to short heparin fragments with four disaccharide units (dp4) (Herbert et al., 2006), and <14 μM for CCP6-8 binding to longer heparin fragments with 10 disaccharide units (dp10) (Fernando et al., 2007). As FH has at least two heparin-binding sites, it is suggested that the affinity of FH for heparin and host cells is much higher than those of individual FH fragments, which is consistent with the high affinity of FH for the commercial heparin (Perkins et al., 2012).

1.3.2 FH recruitment as a mechanism of complement evasion

The fitness of pathogens is determined by many factors. One major factor is the ability of the pathogens to evade the elaborate machinery of the human immune system. The complement system plays a central role in innate immunity and hence a key target for immune evasion strategies of diverse microbes (Lambris et al., 2008;

Ricklin et al., 2010; Zipfel et al., 2008). As described earlier, complement activation involves cascades of reactions resembling blood coagulation, and therefore provides microbes many targets for developing means to escape complement attack. FH is an inhibitor of the alternative complement activation pathway and a protector of host cells from the friendly fires of the complement system. Taking advantage of these important FH functions, many microbes evolve FH-binding proteins and use them to recruit FH to avoid the host complement attack, just like host cells using FH to avoid the attack from its own complement.

Among the pathogens which recruit FH to their surfaces are the bacteria *Streptococcus pneumoniae, Escherichia coli, Borrelia burgdorferi, Staphylococcus aureas*; the virus HIV; the fungus *Candida albicans*; and the parasite *Echinococcus spp* (Ferreira et al.). The molecular and structural details of how microbes recruit FH to their surfaces have just started to emerge with the determination of the atomic structures of several complexes between FH and bacterial FH-binding proteins. One such example is the crystal structure of the complex of human CCP6-7 with the FH binding protein (fHbp) of *Neisseria meningitidis* (Schneider et al., 2009). The structure reveals that the meningococcal protein fhbp binds to CCP6-7 by mimicry of host glycosaminoglycans. Pneumococcus recruits FH via its surface protein CbpA, which enables it to resist complement mediated phagocytosis (Musher, 2009a; Rosenow et al., 1997). The FH-CbpA interaction is the focus of this work.

1.4 Molecular mechanism of pneumococcal pathogenesis

As described earlier, although pneumococcus is a commensal bacterium, it can cause severe diseases, especially in young children, the elderly and immunocompromised individuals (Musher, 2009a). Contributing factors to pneumococcal diseases include coexisting viral infection, allergy, damaged ciliated bronchial cells or increased mucus production. Like any bacterial pathogenesis, the key steps of pneumococcal pathogenesis include adherence, colonization, invasion, and damaging of host tissue and organ. Unlike many other bacterial pathogens, pneumococcus produces few toxins. Pneumococcal diseases are mainly caused by host intense inflammatory responses. The important biochemical events in the early stages of pneumococcal pathogenesis are illustrated in Fig.1 in this reference (Hammerschmidt, 2006), highlighting the roles of pneumococcal protein virulence factors, which is the focus of this work.

Pneumococci attach to host epithelial cells mainly via surface protein virulence factors. These protein adhesins include CbpA, PsaA, and PavA. Pneumococcal adherence is mainly due to interactions between pneumococcal and human surface proteins. CbpA binds to the secretory component of polymeric-immunoglobulin receptor (pIgR), PsaA most likely to the ectodomain of cadherin, and PavA to immobilized fibronectin. Pneumococci can become invasive by binding to pIgR, via CbpA, and hijacking it for transcytosis into the bloodstream (Brock et al., 2002), thus causing bacteremia. Pneumococcal invasion also involves degradation of host extracellular matrix by recruitment of host protease plasmin or plasminogen (the zymogen of plasmin) via the glycolytic enzymes glyceraldehyde-3-phosphate dehydrogenase

(GAPDH) and α-enolase. It is not known how the glycolytic enzymes get to the pneumococcal cell surface, as neither of the two enzymes has the classic secretion and membrane anchoring motifs. Once in the bloodstream, pneumococci can occasionally across the blood-brain barrier by binding to the receptor for platelet-activating factor (PAF) via phosphocholine (Cundell et al., 1995; Ring et al., 1998), causing bacterial meningitis. *S. pneumoniae* has also evolved mechanisms to stimulate host inflammatory responses, such as activation of Toll-like receptor 4 (TLR4) on antigen presenting cells by pneumolysin (Malley et al., 2003).

In addition to adherence and invasion, pneumococci must escape host immune attacks, both innate and adaptive immunities (Musher, 2009a). Innate immunity is the first line of host defense, and the main mechanism of host innate immunity to *S. pneumoniae* is phagocytosis by macrophages promoted by complement opsonization in early pneumococcal infection. If pneumococci survive the immediate attack of the host innate immune system, they encounter the attack of the host adaptive immune system. The main mechanism of host adaptive immunity to *S. pneumoniae* is generation of antibodies, which takes about a week. Antibodies bind to pneumococci and mark them for destruction by phagocytosis, and bind to pneumococcal toxins such as pneumolysin and neutralize their cytotoxic effects.

Complement alone is however ineffective in opsonic phagocytosis of pneumococcus in the absence of specific opsonic antibodies (Perry et al., 1993) and therefore in the absence of antibiotic treatment, pneumococcal infection of the nasopharynx is accompanied by acute inflammation manifested by steady growth of pneumococcus and intense tissue inflammation until anti-pneumococcus antibodies are produced (~7 days after infection)(Musher, 2009b). Although several bacterial factors have been implicated in evasion of phagocytic killing by pneumococcus (Kadioglu et al., 2008b), the precise mechanism of pneumococcal resistance to the innate immunity is not yet clearly known (Kadioglu and Andrew, 2004).

1.5 Pneumococcal Virulence Factors

S. pneumoniae produces many virulence factors (Kadioglu et al., 2008b), which can be divided into two groups based on their chemical compositions: protein and non-protein virulent factors. The main non-protein virulent factor is polysaccharide capsule. Some important pneumococcal virulence factors are described in this section, except CbpA, which is the subject protein of this work and described in the next section. A more extensive list of pneumococcal virulence factors, each with a brief functional description, is presented in Table 1.1.

1.5.1 Polysaccharide Capsule

Pneumococcal capsule is composed of repeating units of oligosaccharides, and nearly all clinical isolates of pneumococci produce a capsule, which forms the outermost layer of a pneumococcal cell (Musher, 2009a). Due to selection pressure from host immunity, capsule composition, structure and antigenicity vary among clinical isolates.

Ninety one capsular serotypes of pneumococcus have been identified to date. Capsule is a major pneumococcal virulence factor, and non-encapsulated strains are generally avirulent. It plays a critical role in preventing phagocytosis an immunologically naive host, although its mechanism is not clearly known at present (Kadioglu et al., 2008b; Musher, 2009b). Capsule also helps the pneumococcus to escape entrapment in the mucus and move to the epithelial surfaces in the early stage of pneumococcal pathogenesis (Kadioglu et al., 2008b). Capsular polysaccharides are the only active component of the 23-valent polysaccharide vaccine and the primary active component of the 7-valent conjugate vaccine (Moffitt and Malley, 2011; Musher, 2009b; Pletz et al., 2008).

1.5.2 Protein virulence factors

Like in other Gram-positive organisms, pneumococcus contains many protein virulence factors at its surface (Table 1.1). While some proteins are attached to pneumococcal surface without an apparent anchor, there are three major mechanisms that anchor these proteins: some (the LPXTG-proteins) are anchored to pneumococcal surface via a covalent bond through the action of sortases, and others via a noncovalent mechanism by binding to phosphocholine (choline binding proteins, CBPs) or by inserting into the outer leaflet of the membrane phospholipid bilayer via a diacylglyceride covalently attached to the proteins (the lipoproteins). These protein virulence factors play critical roles in and contribute substantially to pneumococcal pathogenesis (Gamez and Hammerschmidt, 2012; Jedrzejas, 2001; Kadioglu et al., 2008b; Musher, 2009b). Furthermore, they may serve as antigens for development of

protein-based pneumococcal vaccines (Barocchi et al., 2007; Gamez and Hammerschmidt, 2012; Moffitt and Malley, 2011). Consequently, studying these protein virulence factors is important not only to understanding pneumococcal pathogenesis but also to development of pneumococcal vaccines.

1.5.2.1 Pneumolysin

Pneumolysin (Ply) is a 53-kDa protein toxin released by all clinical isolates of *S. pneumoniae* and it amino acid sequence is highly conserved. It does not have a signal peptide for secretion and its mechanism of release remains controversial (Tai, 2006). It is generally believed that Ply is released upon autolysis, but it can be also released without the action of the main autolysin LytA (Balachandran et al., 2001). Ply is a potent virulent factor and belongs to the family of cholesterol-dependent, pore-forming toxins produced by many Gram-positive bacteria (Kadioglu et al., 2008b; Mitchell and Mitchell, 2010). It binds to membrane cholesterol and oligomerizes to form a large pore in the membrane of target cells. Each pore consists of about 40 subunits and has size of about 26 nm in diameter. Oligomerization and pore formation are responsible for the cytolytic activity of Ply. At sub-lytic concentrations, Ply also exerts a plethora of other activities, including modulation of the host inflammatory and immune responses. Ply plays an important role in pneumococcal colonization and invasion (Cockeran et al., 2002). Animal studies indicate that it contributes to bacteremia (Benton et al., 1995), pneumonia (Berry et al., 1989), and deafness but not inflammation associated with meningitis (Friedland et al., 1995; Wellmer et al., 2002; Winter et al., 1997).

1.5.2.2 Pneumococcal surface protein A (PspA)

PspA is a surface-exposed choline binding protein present in all pneumococcal strains isolated to date and vary significantly among pneumococcal strains both in amino sequence and in size (Crain et al., 1990; McDaniel et al., 1984; Waltman et al., 1990). It is composed of three domains: an N-terminal α-helical domain that protrudes from the bacterial cell surface, followed by a proline-rich domain, and then a C-terminal choline binding domain which anchors the protein to the pneumococcal surface (Yother and Briles, 1992). PspA has two main functions (Kadioglu et al., 2008b; Tai, 2006): (i) it inhibits opsonization by the active complement component, C3b, and helps pneumococcus to escape complement-mediated phagocytosis; and (ii) it binds to apolactoferrin and protects pneumococcus from the bactericidal activity of this human protein. Both functions are mapped to the N-terminal domain of PspA. A crystal structure of the complex of a lactoferrin-binding fragment of PspA and the N-lobe of human lactoferrin has been reported (Senkovich et al., 2007). The structure reveals that the lactoferrin-binding domain consists of four helices and its negatively charged surface binds to the highly cationic lactoferrin moiety of lactoferrin and therefore prevents the human protein from penetrating into the pneumococcal membrane. The importance of PspA to pneumococcal pathogenesis appears to be dependent on the serotypes of pneumococcal strains (Kadioglu et al., 2008b). PspA is required for *in vivo* growth of serotypes 3 and 4 strains but has no effect on a serotype 2 strain.

1.5.2.3 Pneumococcal adherence and virulence factor A (PavA)

PavA is a 64-kDa pneumococcal surface protein. The protein is located at the pneumococcal outer surface despite the lack of a signal peptide for secretion and any anchoring motif. PavA binds to immobilized fibronectin and therefore functions as an adhesin (Holmes et al., 2001; Kadioglu et al., 2010; Pracht et al., 2005; Schwarz-Linek et al., 2004). PavA is necessary for the survival of *S. pneumoniae* strain D39 in the nasopharynx tissue and mutation of PavA impairs translocation of pneumococci from the mucosal epithelial to the bloodstream. PavA plays an important role in pneumococcal colonization and invasion.

1.5.2.4 Pneumococcal surface adhesion A (PsaA)

PsaA, pneumococcal surface adhesin A, is one of the most extensively studied protein virulence factors (Rajam et al., 2008). It is a 34.6-kDa protein consisting of 309 amino acid residues, including 19-residue signal peptide. It is thought that PsaA is attached to bacterial cell membrane via a lipid component that is covalently joined to the protein. The protein is present in all isolated *S. pneumoniae* serotypes (Jedrzejas, 2001; Musher, 2009b; Rajam et al., 2008). PsaA functions as both a part of a Mn^{2+} permease and an adhesin. As part of the Mn^{2+} permease, it assists the transport of the metal into the cytoplasm of bacteria. Analysis of the crystal structure of PsaA (pdb: 1psz) suggests that PsaA can bind either Mn^{2+} or Zn^{2+} (Lawrence et al., 1998). However, the discovery that $PsaA^-$ pneumococcal cells grow optimally in the presence of Mn^{2+} suggests an essential role of PsaA in Mn^{2+} transport. PsaA also has a higher affinity for Mn^{2+} (3.3 nM) than for Zn^{2+} (23.1 nM), suggesting that even though it prefers to bind Mn^{2+}, Zn^{2+}

can compete at higher concentration (McDevitt et al., 2011). As an adhesin, PsaA enhances interaction between pneumococcus and E-catherin of the nasopharyngeal cells (Anderton et al., 2007). This interaction leads to up regulation of PsaA (Orihuela et al., 2004). PsaA therefore plays a crucial role in adherence and virulence of the pneumococcus. PsaA is immunogenic and highly conserved and shows little antigenic variation, and is therefore a good candidate for development of vaccines against pneumococcus. Indeed recombinant PsaA has been evaluated in a phase one clinical trial (Rajam et al., 2008).

Table 1.1 Pneumococcal virulence factors and their roles in colonization and disease.

Pneumococcal virulence factors and disease	**Main role in colonization**
Upper-airway colonization	
Capsule	Prevent entrapment in the nasal mucus, thereby allowing access to epithelial surfaces. Also inhibits effective opsonophagocytosis.
ChoP	Binds to PAFr on the epithelial surfaces of the human nasopharynx.
CbpA (also known as PspC)	Binds to human secretory component on a polymeric Ig receptor during the first stage of translocation across the epithelium.
Hyl	Breaks down hyaluronan-containing extracellular matrix components.
PavA	Binds to fibronectin.
Eno	Binds to plasminogen.
Competition in upper airway	
Bacteriocin (pneumocin)	Small antimicrobial peptide that targets members of the same species.
Respiratory-tract infection and pneumonia	
Ply	Cytolytic that also activates complement. An important determinant of virulence in the *in vivo* models of disease. Wide range of effects on host immune components at sub-lytic concentrations.
PspA	Prevents binding of C3 onto pneumococcal surface. Also binds lactoferrin.
LytA	Digests the cell wall, which results in release of Ply.
PsaA	Component of the ABC transport system, which is involved in resistance to oxidative stress
PiaA and PiuA	Component of the ABC transport system.
NanA and NanB	Aid colonization by revealing receptors for adherence, modifying the surfaces of competing bacteria that are within the same niche and/or modifying the function of host clearance glycoproteins.
IgA	Cleaves human IgA1

1.6 CbpA

CbpA is a major bacterial determinant that interacts specifically with host proteins pIgR, FH and C3 to enhance pneumococcal adherence, colonization and immune invasion (Hammerschmidt et al., 1997; Iannelli et al., 2004; Jarva et al., 2003a; Ogunniyi et al., 2001a; Rosenow et al., 1997; Zhang et al., 2000). The protein is also known by other names, PspC (pneumococcal surface protein C), SpsA (*S. pneumoniae* secretory IgA binding protein), Hic (Factor H binding inhibitor of complement), and PbcA (C3 binding protein). Hic usually refers to those proteins that contain an LPXTG motif instead of a choline-binding domain (CBD) for cell wall anchoring (Hammerschmidt, 2006).

CbpA is a large multifunctional protein with a conserved N-terminal signal peptide. A typical mature CbpA consists of five domains/regions (Hammerschmidt, 2006): an N-terminal domain (CbpAN), which binds FH; two repeated domains (R1 and R2), which bind pIgR or the secretory component (SC); a proline-rich domain; and a C-terminal CBD, which anchors the protein to pneumococcal surface (Rosenow et al., 1997). The R domains also bind to the C-terminal domain of vitronectin, and this binding is competitively inhibited by secretory IgA, suggesting that these human proteins binds to the same region or two overlapping regions of CbpA (Voss et al., 2013). The CbpA family of proteins can be classified into 11 groups due to polymorphism in the N-terminal region before the proline-rich domain (Iannelli et al., 2002).

All virulent *S. pneumoniae* strains contain at least a copy of *cbpA* gene (Iannelli et al., 2002) and it is essential for nasal colonization (Balachandran et al., 2002; Rosenow et al., 1997), lung infection (Balachandran et al., 2002; Rosenow et al., 1997),

sepsis (Iannelli et al., 2004), and meningitis in animal models (Orihuela et al., 2004). Mutation of *cbpA* in *S. pneumoniae* resulted in reduced binding activity to cytokine-activated epithelial cells and immobilized sialic acid *in vitro,* and a reduced nasopharyngeal colonization capability in animal models, suggesting that CbpA is responsible for the adherence of pneumococcus to eukaryotic cells (Rosenow et al., 1997). CbpA-deficient *S. pneumoniae* strain D39 displayed highly reduced ability to colonize the nasopharynx, indicating that CbpA plays an important role in pneumococcal colonization of the nasopharynx (Ogunniyi et al., 2007b). CbpA is also necessary for pneumococcal biofilm formation of pneumococcus *in vitro* (Munoz-Elias et al., 2008) and *in vivo* (Reid et al., 2009).

CbpA is highly immunogenic and has been shown to confer protective immunity against challenges with virulent pneumococcal strains in animal models (Balachandran et al., 2002; Brooks-Walter et al., 1999a; Ogunniyi et al., 2001a; Ricci et al., 2011; Rosenow et al., 1997). Biochemically, most, if not all, of biological functions of CbpA are dependent on its ability to interact with FH and pIgR/SC/SIgA. Nearly all tested pneumococcal clinical isolates bind to FH and pIgR/SC/SIgA (Hammerschmidt et al., 1997; Quin et al., 2005).

1.6.1 Binding of pIgR

The interaction between CbpA and the polymeric immunoglobulin receptor (pIgR) provides a mechanism for adherence of pneumococci to mucosal cells and pneumococcal invasion by hijacking the pIgR retrograde transcytosis machinery (Elm et al., 2004; Hammerschmidt et al., 1997; Zhang et al., 2000). pIgR is widely expressed

on mucosal epithelium and it transports polymeric immunoglobulin to mucosal surfaces (Mostov, 1999). This function of CbpA is supported by an observation that heterologous expression of CbpA on the surface of *Lactococcus lactis* enables them to adhere to host epithelial cells that are expressing human pIgR (Asmat et al., 2012).

Human pIgR is an 81-kDa transmembrane protein with an extracellular portion (SC) composed of five immunoglobulin (Ig)-like domains (D1-D5) of 104-114 amino acid residues each (Kaetzel, 2005). SC is stabilized by ten disulfide bonds, two disulfide bonds each in D1, D3 and D4, one in D2, and three in D5. The D3-D4 domains are responsible for binding to CbpA (Lu et al., 2003). Binding is attenuated in the presence of a reducing agent, suggesting that the disulfide bonds are important for stability of D3-D4 (Elm et al., 2004). pIgR can form a secretory IgA (sIgA) complex with IgA and is responsible for delivering IgA to mucosal secretions. sIgA is the most represented isotype of antibodies in mucosal secretions and binds to microbial pathogens and prevent their adherence to host epithelial cells (immune exclusion) (Corthesy, 2010; Mestecky, 1999).

CbpA binds pIgR via the homologous R1 and R2 domains (Hammerschmidt et al., 2000; Lu et al., 2003), which share 78% amino acid sequence identity.(Zhang et al., 2000). The solution NMR structure of R2 reveals that it consists of three α-helices adopting an antiparallel topology and forming a raft-like structure. The conserved residues in the loop between helix 1 and helix 2 form a ‘tyrosine fork’ structure (PDB ID. 1w9r) that is responsible for the interaction with pIgR (Luo et al., 2005). Despite the high sequence variation in the N-terminal region of CbpA, the SC/pIgR/SIgA-binding motif is highly conserved (Iannelli et al., 2002).

Successful invasion of epithelial cells by pneumococcus via CbpA-pIgR interaction needs the activities of Cdc42, phosphatidylinositol 3-kinase (PI3K) and Akt (Agarwal and Hammerschmidt, 2009). CbpA-pIgR mediated pneumococcal invasion also requires signaling of Src protein-tyrosine kinase, ERK1/2, JNK and focal adhesion kinase (Agarwal et al., 2010a; Agarwal and Hammerschmidt, 2009). Some of these signal transduction pathways depend on increase in cellular internal calcium concentration (Yordy and Muise-Helmericks, 2000). Infection of host epithelial cells by pneumococcus leads to an increase in intracellular calcium concentration and this intracellular calcium mobilization depend on CbpA-pIgR interaction (Asmat et al., 2011). The intracellular calcium mobilization results in diminished uptake of pneumococcus by epithelial cells, suggesting that it may be a host mechanism to reduce invasion by pneumococcus (Asmat et al., 2011).

1.6.2 Binding of FH

In order to cause infection, *S. pneumoniae* has to escape complement-mediated host phagocytosis. A major mechanism for pneumococcal evasion of host phagocytosis is recruitment of FH by *S. pneumoniae* via CbpA (Hammerschmidt, 2006). FH recruitment promotes pneumococcal adherence to and uptake of pneumococci by human cells (Agarwal et al., 2010b). FH binding has been mapped to the N-terminal domain of CbpA (CbpAN) (Hammerschmidt et al., 2007; Lu et al., 2006). However, which CCP module or modules of FH are involved in binding of CbpA remains unresolved. An earlier study indicated hFH domains CCP13–15 (Duthy et al., 2002), a later study CCP6–10 (Dave et al., 2004), and a more recent study two regions of hFH,

CCP8–11 and CCP19–20 (Hammerschmidt et al., 2007). No structural information is available on the CbpAN domain and its interaction with FH.

1.6.3 Host specificity of FH and pIgR binding

Even though human and mouse FH proteins share 61% sequence identity, pneumococcal D39 CbpA binds to human FH protein but not to FH homologues from mouse and other species (Frolet et al., 2010). This host specificity of pneumococcal interaction with FH was also observed in several other pneumococcal isolates from the blood and cerebrospinal fluid (Lu et al., 2008). Host specificity of FH binding has also been observed with *Borrelia bugdoferi* surface proteins, OspE and BBA68, which bind to human but not to mouse FH proteins (Kaetzel, 2001). Interaction of CbpA with pIgR was also found to be host-specific, with interaction only detected with human protein and not with proteins from mouse and other species (Lu et al., 2008). CbpA is therefore suggested to be a determinant for natural host adaptation of *S. pneumoniae* to humans (Lu et al., 2008). In this study we used an array of techniques to elucidate the structural and molecular basis of host-specific interaction between pneumococcus with FH.

1.6.4 CbpA in vaccine development

A good pneumococcal vaccine should cover most if not all of the over 90 serotypes. Unfortunately, the pneumococcal polysaccharide capsule-based vaccines currently in use cover only a few serotypes and it's technically very challenging to develop polysaccharide capsule-based vaccines with a broad coverage, as described earlier. Protein-based vaccines, however, have the potential for a broader spectrum of

pneumococcal serotypes (Gamez and Hammerschmidt, 2012). Several pneumococcal proteins, including PspA and CbpA, have since been shown to be immunogenic and are considered as potential targets for the development of protein-based vaccines (Barocchi et al., 2007; Gamez and Hammerschmidt, 2012; Moffitt and Malley, 2011; Tai, 2006). Some of the pneumococcal virulent factors have been tested for their ability to protect mouse models of both systemic and local infections against virulent pneumococcal challenge (Moffitt and Malley, 2011; Pletz et al., 2008).

CbpA is highly immunogenic (Holmlund et al., 2009; Linder et al., 2007; McCool et al., 2003; Moreno et al., 2012b; Simell et al., 2009; Zhang et al., 2006; Zhang et al., 2002). In particular, high levels of antibodies against CbpA can be detected in children (Holmlund et al., 2009; Simell et al., 2006; Zhang et al., 2006; Zhang et al., 2002), indicating that CbpA may be a good candidate antigen for development of protein-based pneumococcal vaccines for this age group. It has been suggested that anti-CbpA antibodies may protect children against subsequent pneumococcal colonization and acute otitis media (Simell et al., 2009; Zhang et al., 2006). Since the CbpA family of proteins can be divided into 11 groups based on their amino acid sequences (Iannelli et al., 2002), it is important to know the cross reactivity of the antibodies elicited by one CbpA protein. It has been shown that mouse antibodies elicited by a fragment of a group-3 CbpA can recognize most of the clinical isolates of *S. pneumoniae* (Moreno et al., 2012a). There is also a high degree of cross-reactivity between PspA- and CbpA-antibodies, suggesting that a limited number of proteins may be sufficient for development of pneumococcal protein-based vaccines (Linder et al., 2007). CbpA has been used as an antigen alone or in combination with other pneumococcal protein

virulence factors for development of protein-based pneumococcal vaccines (Brooks-Walter et al., 1999b; Cao et al., 2007; Cao et al., 2009; Daniels et al., 2010; Ferreira et al., 2009; Godfroid et al., 2011; Ogunniyi et al., 2007a; Ogunniyi et al., 2001b; Ricci et al., 2011). Immunization with CbpA alone is protective against pneumococcal challenge in mouse (Brooks-Walter et al., 1999b; Ogunniyi et al., 2001b; Ricci et al., 2011). CbpA also show synergistic effects on protecting mouse from pneumococcal challenge when it is combined with other protein virulence factors (Cao et al., 2007; Cao et al., 2009; Ogunniyi et al., 2007a; Ogunniyi et al., 2001b). It should be noted that most of these pneumococcal challenge studies used homologous strains. When both homologous and nonhomologous strains were used in a recent pneumococcal challenge study, immunization with an N-terminal CbpA fragment showed protection against a homologous pneumococcal strain challenge but no protection against nonhomologous pneumococcal strain challenges (Ricci et al., 2011), presumably due to the high strain variation in the amino acid sequence of the N-terminal fragment.

1.6 Scope and significance of this study

The evidence presented here highlights CbpA as an important pneumococcal virulence factor and a protective antigen. Its interactions with human immune proteins, FH and pIgR, are important for pneumococcus pathogenesis (Dave et al., 2001; Hammerschmidt et al., 1997; Jarva et al., 2003b; Lu et al., 2003; Ogunniyi et al., 2001a; Rosenow et al., 1997; Zhang et al., 2000) and here lies the significance of understanding the structural and molecular basis of these interactions. The atomic structure of the free R2 domain has been determined by solution NMR spectroscopy

(Luo et al., 2005), but how the R2 domain interacts with pIgR or SC is still largely unknown, as no structure has been reported for its complex with pIgR or SC. The boundaries of the N-terminal region of CbpA required for binding FH have not been defined. Which CCP module or modules of FH are required for binding CbpA remains controversial. The structure of the CbpAN domain has not been reported in either the free form or in complex with FH. The structural basis for host specific binding of FH is also not known. There is also no robust, cost-effective recombinant system for production of human FH or SC is available for biophysical studies of the interactions of CbpA with these proteins.

Here we used a structure prediction webserver to define the boundaries of the CbpAN domain and produced it in *E. coli*. We established an *E. coli* expression system for production of individual or combined CCP modules of FH at a level sufficient for biophysical studies and determined that a single CCP module of FH (the 9^{th} module, CCP9) is required and sufficient for tight binding of CbpAN by generating various individual or combined CCP modules of FH and measuring their affinities for CbpAN. We determined the solution structure of free CbpAN by NMR and the crystal structure of its complex with CCP9 by X-ray crystallography. By site-directed mutagenesis and computation analysis, we assessed the importance of the interface residues of both FH and CbpAN to the formation of the complex of the two proteins. Finally, based on the crystal structure of the complex of CbpAN and CCP9 and computational and site-directed mutagenesis studies, we elucidated the structural basis for host-specificity of FH binding. The insights gained from this study may inform redesign of CbpAN for

vaccine development and creation of better animal models for studying pneumococcal pathogenesis.

REFERENCES

REFERENCES

Agarwal, V., Asmat, T.M., Dierdorf, N.I., Hauck, C.R., and Hammerschmidt, S. (2010a). Polymeric immunoglobulin receptor-mediated invasion of Streptococcus pneumoniae into host cells requires a coordinate signaling of SRC family of protein-tyrosine kinases, ERK, and c-Jun N-terminal kinase. The Journal of biological chemistry *285*, 35615-35623.

Agarwal, V., Asmat, T.M., Luo, S., Jensch, I., Zipfel, P.F., and Hammerschmidt, S. (2010b). Complement regulator Factor H mediates a two-step uptake of *Streptococcus pneumoniae* by human cells. J Biol Chem *285*, 23486-23495.

Agarwal, V., and Hammerschmidt, S. (2009). Cdc42 and the phosphatidylinositol 3-kinase-Akt pathway are essential for PspC-mediated internalization of pneumococci by respiratory epithelial cells. The Journal of biological chemistry *284*, 19427-19436.

Anderton, J.M., Rajam, G., Romero-Steiner, S., Summer, S., Kowalczyk, A.P., Carlone, G.M., Sampson, J.S., and Ades, E.W. (2007). E-cadherin is a receptor for the common protein pneumococcal surface adhesin A (PsaA) of Streptococcus pneumoniae. Microb Pathog *42*, 225-236.

Asmat, T.M., Agarwal, V., Rath, S., Hildebrandt, J.P., and Hammerschmidt, S. (2011). Streptococcus pneumoniae infection of host epithelial cells via polymeric immunoglobulin receptor transiently induces calcium release from intracellular stores. The Journal of biological chemistry *286*, 17861-17869.

Asmat, T.M., Klingbeil, K., Jensch, I., Burchhardt, G., and Hammerschmidt, S. (2012). Heterologous expression of pneumococcal virulence factor PspC on the surface of Lactococcus lactis confers adhesive properties. Microbiology *158*, 771-780.

Balachandran, P., Brooks-Walter, A., Virolainen-Julkunen, A., Hollingshead, S.K., and Briles, D.E. (2002). Role of Pneumococcal Surface Protein C in Nasopharyngeal Carriage and Pneumonia and Its Ability To Elicit Protection against Carriage of Streptococcus pneumoniae. Infection and Immunity *70*, 2526-2534.

Balachandran, P., Hollingshead, S.K., Paton, J.C., and Briles, D.E. (2001). The autolylic enzyme LytA of *Streptococcus pneumoniae* is not responsible for releasing pneumolysin. J Bacteriol *183*, 3108-3116.

Barocchi, M.A., Censini, S., and Rappuoli, R. (2007). Vaccines in the era of genomics: the pneumococcal challenge. Vaccine *25*, 2963-2973.

Benton, K.A., Everson, M.P., and Briles, D.E. (1995). A pneumolysin-negative mutant of Streptococcus pneumoniae causes chronic bacteremia rather than acute sepsis in mice. Infect Immun *63*, 448-455.

Berry, A.M., Yother, J., Briles, D.E., Hansman, D., and Paton, J.C. (1989). Reduced virulence of a defined pneumolysin-negative mutant of Streptococcus pneumoniae. Infect Immun *57*, 2037-2042.

Black, S., Eskola, J., Whitney, C., and Shinefield, H. (2008). Pneumococcal conjugate vaccine and pneumococcal common protein vaccines. In Vaccines, S.A. Plotkin, P.A. Offit, and W.A. Orenstein, eds. (Philadelphia: Saunders).

Brock, S.C., McGraw, P.A., Wright, P.F., and Crowe, J.E., Jr. (2002). The human polymeric immunoglobulin receptor facilitates invasion of epithelial cells by Streptococcus pneumoniae in a strain-specific and cell type-specific manner. Infect Immun *70*, 5091-5095.

Brooks-Walter, A., Briles, D.E., and Hollingshead, S.K. (1999a). The pspC gene of Streptococcus pneumoniae encodes a polymorphic protein, PspC, which elicits cross-reactive antibodies to PspA and provides immunity to pneumococcal bacteremia. Infect Immun *67*, 6533-6542.

Brooks-Walter, A., Briles, D.E., and Hollingshead, S.K. (1999b). The *pspC* gene of *Streptococcus pneumoniae* encodes a polymorphic protein, PspC, which elicits cross-reactive antibodies to PspA and provides immunity to pneumococcal bacteremia. Infect Immun *67*, 6533-6542.

Brown, J.S., Hussell, T., Gilliland, S.M., Holden, D.W., Paton, J.C., Ehrenstein, M.R., Walport, M.J., and Botto, M. (2002). The classical pathway is the dominant complement pathway required for innate immunity to Streptococcus pneumoniae infection in mice. Proc Natl Acad Sci U S A *99*, 16969-16974.

Cao, J., Chen, D.P., Xu, W.C., Chen, T.M., Xu, S.X., Luo, J.Y., Zhao, Q., Liu, B.Z., Wang, D.S., Zhang, X.M.*, et al.* (2007). Enhanced protection against pneumococcal infection elicited by immunization with the combination of PspA, PspC, and ClpP. Vaccine *25*, 4996-5005.

Cao, J., Gong, Y., Li, D.R., Yin, N.L., Chen, T.M., Xu, W.C., Zhang, X.M., and Yin, Y.B. (2009). CD4(+) T lymphocytes mediated protection against invasive pneumococcal infection induced by mucosal immunization with ClpP and CbpA. Vaccine *27*, 2838-2844.

Cockeran, R., Anderson, R., and Feldman, C. (2002). The role of pneumolysin in the pathogenesis of Streptococcus pneumoniae infection. Current Opinion in Infectious Diseases *15*, 235-239.

Corthesy, B. (2010). Role of secretory immunoglobulin A and secretory component in the protection of mucosal surfaces. Future Microbiol *5*, 817-829.

Crain, M.J., Waltman, W.D., 2nd, Turner, J.S., Yother, J., Talkington, D.F., McDaniel, L.S., Gray, B.M., and Briles, D.E. (1990). Pneumococcal surface protein A (PspA) is serologically highly variable and is expressed by all clinically important capsular serotypes of Streptococcus pneumoniae. Infect Immun *58*, 3293-3299.

Cundell, D.R., Gerard, N.P., Gerard, C., Idanpaanheikkila, I., and Tuomanen, E.I. (1995). *Streptococcus pneumoniae* anchor to activated human cells by the receptor for platelet-activating-factor. Nature *377*, 435-438.

Daniels, C.C., Coan, P., King, J., Hale, J., Benton, K.A., Briles, D.E., and Hollingshead, S.K. (2010). The Proline-Rich Region of Pneumococcal Surface Proteins A and C Contains Surface-Accessible Epitopes Common to All Pneumococci and Elicits Antibody-Mediated Protection against Sepsis. Infect Immun *78*, 2163-2172.

Dave, S., Brooks-Walter, A., Pangburn, M.K., and McDaniel, L.S. (2001). PspC, a pneumococcal surface protein, binds human factor H. Infect Immun *69*, 3435-3437.

Dave, S., Carmicle, S., Hammerschmidt, S., Pangburn, M.K., and McDaniel, L.S. (2004). Dual roles of PspC, a surface protein of *Streptococcus pneumoniae*, in binding human secretory IgA and factor H. J Immunol *173*, 471-477.

Duthy, T.G., Ormsby, R.J., Giannakis, E., Ogunniyi, A.D., Stroeher, U.H., Paton, J.C., and Gordon, D.L. (2002). The human complement regulator factor H binds pneumococcal surface protein PspC via short consensus repeats 13 to 15. Infect Immun *70*, 5604-5611.

Elm, C., Braathen, R., Bergmann, S., Frank, R., Vaerman, J.P., Kaetzel, C.S., Chhatwal, G.S., Johansen, F.E., and Hammerschmidt, S. (2004). Ectodomains 3 and 4 of human polymeric Immunoglobulin receptor (hpIgR) mediate invasion of Streptococcus pneumoniae into the epithelium. The Journal of biological chemistry *279*, 6296-6304.

Farrell, D.J., Klugman, K.P., and Pichichero, M. (2007). Increased antimicrobial resistance among nonvaccine serotypes of Streptococcus pneumoniae in the pediatric population after the introduction of 7-valent pneumococcal vaccine in the United States. Pediatr Infect Dis J *26*, 123-128.

Fernando, A.N., Furtado, P.B., Clark, S.J., Gilbert, H.E., Day, A.J., Sim, R.B., and Perkins, S.J. (2007). Associative and structural properties of the region of complement factor H encompassing the Tyr402His disease-related polymorphism and its interactions with heparin. J Mol Biol *368*, 564-581.

Ferreira, D.M., Darrieux, M., Silva, D.A., Leite, L.C.C., Ferreira, J.M.C., Ho, P.L., Miyaji, E.N., and Oliveira, M.L.S. (2009). Characterization of Protective Mucosal and Systemic Immune Responses Elicited by Pneumococcal Surface Protein PspA and PspC Nasal Vaccines against a Respiratory Pneumococcal Challenge in Mice. Clinical and Vaccine Immunology *16*, 636-645.

Ferreira, V.P., Pangburn, M.K., and Cortes, C. Complement control protein factor H: the good, the bad, and the inadequate. Mol Immunol *47*, 2187-2197.

Friedland, I.R., Paris, M.M., Hickey, S., Shelton, S., Olsen, K., Paton, J.C., and McCracken, G.H. (1995). The limited role of pneumolysin in the pathogenesis of pneumococcal meningitis. J Infect Dis *172*, 805-809.

Frolet, C., Beniazza, M., Roux, L., Gallet, B., Noirclerc-Savoye, M., Vernet, T., and Di Guilmi, A.M. (2010). New adhesin functions of surface-exposed pneumococcal proteins. BMC Microbiol *10*, 190.

Gamez, G., and Hammerschmidt, S. (2012). Combat pneumococcal infections: adhesins as candidates for protein-based vaccine development. Curr Drug Targets *13*, 323-337.

Godfroid, F., Hermand, P., Verlant, V., Denoel, P., and Poolman, J.T. (2011). Preclinical Evaluation of the Pht Proteins as Potential Cross-Protective Pneumococcal Vaccine Antigens. Infect Immun *79*, 238-245.

Hammerschmidt, S. (2006). Adherence molecules of pathogenic pneumococci. Curr Opin Microbiol *9*, 12-20.

Hammerschmidt, S., Agarwal, V., Kunert, A., Haelbich, S., Skerka, C., and Zipfel, P.F. (2007). The host immune regulator factor H interacts via two contact sites with the PspC protein of *Streptococcus pneamoniae* and mediates adhesion to host epithelial cells. J Immunol *178*, 5848-5858.

Hammerschmidt, S., Talay, S.R., Brandtzaeg, P., and Chhatwal, G.S. (1997). SpsA, a novel pneumococcal surface protein with specific binding to secretory immunoglobulin A and secretory component. Mol Microbiol *25*, 1113-1124.

Hammerschmidt, S., Tillig, M.P., Wolff, S., Vaerman, J.P., and Chhatwal, G.S. (2000). Species-specific binding of human secretory component to SpsA protein of Streptococcus pneumoniae via a hexapeptide motif. Mol Microbiol *36*, 726-736.

Harboe, M., and Mollnes, T.E. (2008). The alternative complement pathway revisited. J Cell Mol Med *12*, 1074-1084.

Herbert, A.P., Kavanagh, D., Johansson, C., Morgan, H.P., Blaum, B.S., Hannan, J.P., Barlow, P.N., and Uhrin, D. (2012). Structural and functional characterization of the product of disease-related factor H gene conversion. Biochemistry *51*, 1874-1884.

Herbert, A.P., Uhrin, D., Lyon, M., Pangburn, M.K., and Barlow, P.N. (2006). Disease-associated sequence variations congregate in a polyanion recognition patch on human factor H revealed in three-dimensional structure. J Biol Chem *281*, 16512-16520.

Holmes, A.R., McNab, R., Millsap, K.W., Rohde, M., Hammerschmidt, S., Mawdsley, J.L., and Jenkinson, H.F. (2001). The pavA gene of Streptococcus pneumoniae encodes a fibronectin-binding protein that is essential for virulence. Mol Microbiol *41*, 1395-1408.

Holmlund, E., Quiambao, B., Ollgren, J., Jaakkola, T., Neyt, C., Poolman, J., Nohynek, H., and Kayhty, H. (2009). Antibodies to Pneumococcal Proteins PhtD, CbpA, and LytC in Filipino Pregnant Women and Their Infants in Relation to Pneumococcal Carriage. Clinical and Vaccine Immunology *16*, 916-923.

Huang, S.S., Johnson, K.M., Ray, G.T., Wroe, P., Lieu, T.A., Moore, M.R., Zell, E.R., Linder, J.A., Grijalva, C.G., Metlay, J.P*., et al.* (2011). Healthcare utilization and cost of pneumococcal disease in the United States. Vaccine *29*, 3398-3412.

Iannelli, F., Chiavolini, D., Ricci, S., Oggioni, M.R., and Pozzi, G. (2004). Pneumococcal Surface Protein C Contributes to Sepsis Caused by Streptococcus pneumoniae in Mice. Infection and Immunity *72*, 3077-3080.

Iannelli, F., Oggioni, M.R., and Pozzi, G. (2002). Allelic variation in the highly polymorphic locus pspC of Streptococcus pneumoniae. Gene *284*, 63-71.

Jacobs, M.R., Good, C.E., Beall, B., Bajaksouzian, S., Windau, A.R., and Whitney, C.G. (2008). Changes in serotypes and antimicrobial susceptibility of invasive Streptococcus pneumoniae strains in Cleveland: a quarter century of experience. J Clin Microbiol *46*, 982-990.

Jarva, H., Jokiranta, T.S., Wurzner, R., and Meri, S. (2003a). Complement resistance mechanisms of streptococci. Molecular Immunology *40*, 95-107.

Jarva, H., Jokiranta, T.S., Würzner, R., and Meri, S. (2003b). Complement resistance mechanisms of streptococci. Molecular Immunology *40*, 95-107.

Jedrzejas, M.J. (2001). Pneumococcal virulence factors: structure and function. Microbiol Mol Biol Rev *65*, 187-207 ; first page, table of contents.

Kadioglu, A., and Andrew, P.W. (2004). The innate immune response to pneumococcal lung infection: the untold story. Trends Immunol *25*, 143-149.

Kadioglu, A., Brewin, H., Hartel, T., Brittan, J.L., Klein, M., Hammerschmidt, S., and Jenkinson, H.F. (2010). Pneumococcal protein PavA is important for nasopharyngeal carriage and development of sepsis. Mol Oral Microbiol *25*, 50-60.

Kadioglu, A., Weiser, J.N., Paton, J.C., and Andrew, P.W. (2008a). The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease. Nature Rev Microbiol *6*, 288-301.

Kadioglu, A., Weiser, J.N., Paton, J.C., and Andrew, P.W. (2008b). The role of Streptococcus pneumoniae virulence factors in host respiratory colonization and disease. Nat Rev Microbiol *6*, 288-301.

Kaetzel, C.S. (2001). Polymeric Ig receptor: defender of the fort or Trojan horse? Curr Biol *11*, R35-38.

Kaetzel, C.S. (2005). The polymeric immunoglobulin receptor: bridging innate and adaptive immune responses at mucosal surfaces. Immunol Rev *206*, 83-99.

Kajander, T., Lehtinen, M.J., Hyvarinen, S., Bhattacharjee, A., Leung, E., Isenman, D.E., Meri, S., Goldman, A., and Jokiranta, T.S. (2011). Dual interaction of factor H with C3d and glycosaminoglycans in host-nonhost discrimination by complement. Proc Natl Acad Sci U S A *108*, 2897-2902.

Lambris, J.D., Ricklin, D., and Geisbrecht, B.V. (2008). Complement evasion by human pathogens. Nat Rev Microbiol *6*, 132-142.

Lawrence, M.C., Pilling, P.A., Epa, V.C., Berry, A.M., Ogunniyi, A.D., and Paton, J.C. (1998). The crystal structure of pneumococcal surface antigen PsaA reveals a metal-binding site and a novel structure for a putative ABC-type binding protein. Structure *6*, 1553-1561.

Linder, A., Hollingshead, S., Janulczyk, R., Christensson, B., and Akesson, P. (2007). Human antibody response towards the pneumococcal surface proteins PspA and PspC during invasive pneumococcal infection. Vaccine *25*, 341-345.

Lu, L., Lamm, M.E., Li, H., Corthesy, B., and Zhang, J.R. (2003). The human polymeric immunoglobulin receptor binds to Streptococcus pneumoniae via domains 3 and 4. The Journal of biological chemistry *278*, 48178-48187.

Lu, L., Ma, Y.Y., and Zhang, J.R. (2006). *Streptococcus pneumoniae* recruits complement factor H through the amino terminus of CbpA. J Biol Chem *281*, 15464-15474.

Lu, L., Ma, Z., Jokiranta, T.S., Whitney, A.R., DeLeo, F.R., and Zhang, J.-R. (2008). Species-Specific Interaction of Streptococcus pneumoniae with Human Complement Factor H. The Journal of Immunology *181*, 7138-7146.

Luo, R., Mann, B., Lewis, W.S., Rowe, A., Heath, R., Stewart, M.L., Hamburger, A.E., Sivakolundu, S., Lacy, E.R., Bjorkman, P.J.*, et al.* (2005). Solution structure of choline binding protein A, the major adhesin of Streptococcus pneumoniae. EMBO J *24*, 34-43.

Lynch, J.P., 3rd, and Zhanel, G.G. (2009). Streptococcus pneumoniae: epidemiology, risk factors, and strategies for prevention. Semin Respir Crit Care Med *30*, 189-209.

Malley, R., Henneke, P., Morse, S.C., Cieslewicz, M.J., Lipsitch, M., Thompson, C.M., Kurt-Jones, E., Paton, J.C., Wessels, M.R., and Golenbock, D.T. (2003). Recognition of pneumolysin by toll-like receptor 4 confers resistance to pneumococcal infection. Proc Natl Acad Sci U S A *100*, 1966-1971.

McCool, T.L., Cate, T.R., Tuomanen, E.I., Adrian, P., Mitchell, T.J., and Weiser, J.N. (2003). Serum immunoglobulin G response to candidate vaccine antigens during experimental human pneumococcal colonization. Infect Immun *71*, 5724-5732.

McDaniel, L.S., Scott, G., Kearney, J.F., and Briles, D.E. (1984). Monoclonal antibodies against protease-sensitive pneumococcal antigens can protect mice from fatal infection with Streptococcus pneumoniae. J Exp Med *160*, 386-397.

McDevitt, C.A., Ogunniyi, A.D., Valkov, E., Lawrence, M.C., Kobe, B., McEwan, A.G., and Paton, J.C. (2011). A molecular mechanism for bacterial susceptibility to zinc. PLoS Pathog *7*, e1002357.

Meri, S., and Pangburn, M.K. (1990). Discrimination between activators and nonactivators of the alternative pathway of complement: regulation via a sialic acid/polyanion binding site on factor H. Proc Natl Acad Sci U S A *87*, 3982-3986.

Mestecky, J., I. Moro, and B. J. Underdown. (1999). Mucosal immunoglobulins (San Diego: Academic Press).

Mitchell, A.M., and Mitchell, T.J. (2010). Streptococcus pneumoniae: virulence factors and variation. Clin Microbiol Infect *16*, 411-418.

Moffitt, K.L., and Malley, R. (2011). Next generation pneumococcal vaccines. Curr Opin Immunol *23*, 407-413.

Moreno, A.T., Oliveira, M.L., Ho, P.L., Vadesilho, C.F., Palma, G.M., Ferreira, J.M., Jr., Ferreira, D.M., Santos, S.R., Martinez, M.B., and Miyaji, E.N. (2012a). Cross-reactivity of antipneumococcal surface protein C (PspC) antibodies with different strains and evaluation of inhibition of human complement factor H and secretory IgA binding via PspC. Clin Vaccine Immunol *19*, 499-507.

Moreno, A.T., Oliveira, M.L.S., Ho, P.L., Vadesilho, C.F.M., Palma, G.M.P., Ferreira, J.M.C., Ferreira, D.M., Santos, S.R., Martinez, M.B., and Miyaji, E.N. (2012b). Cross-

Reactivity of Antipneumococcal Surface Protein C (PspC) Antibodies with Different Strains and Evaluation of Inhibition of Human Complement Factor H and Secretory IgA Binding via PspC. Clinical and Vaccine Immunology *19*, 499-507.

Morgan, H.P., Schmidt, C.Q., Guariento, M., Blaum, B.S., Gillespie, D., Herbert, A.P., Kavanagh, D., Mertens, H.D., Svergun, D.I., Johansson, C.M.*, et al.* (2011). Structural basis for engagement by complement factor H of C3b on a self surface. Nat Struct Mol Biol *18*, 463-470.

Mostov, K.E.a.C.S.K. (1999). Immunoglobulin transport and the polymeric immunoglobulin receptor, pIgR, 2nd edn (San Diego.: Academic Press).

Munoz-Elias, E.J., Marcano, J., and Camilli, A. (2008). Isolation of Streptococcus pneumoniae biofilm mutants and their characterization during nasopharyngeal colonization. Infect Immun *76*, 5049-5061.

Murdoch, D.R. (2008). Raising the profile of pneumococcal disease. Intern Med J *38*, 381-383.

Musher, D.M. (2009a). Streptococcus pneumoniae. In Principles and Practice of Infectious Diseases, G.L. Mandell, J.E. Bennett, and R.D. Dolin, eds. (Philadelphia: Elsevier Churchill Livingstone), pp. 2623-2642.

Musher, D.M. (2009b). Streptococcus pneumoniae (Philadelphia: Elsevier Churchill Livingstone).

Neth, O., Jack, D.L., Dodds, A.W., Holzel, H., Klein, N.J., and Turner, M.W. (2000). Mannose-binding lectin binds to a range of clinically relevant microorganisms and promotes complement deposition. Infect Immun *68*, 688-693.

Novak, R., Henriques, B., Charpentier, E., Normark, S., and Tuomanen, E. (1999). Emergence of vancomycin tolerance in Streptococcus pneumoniae. Nature *399*, 590-593.

Ogunniyi, A.D., Grabowicz, M., Briles, D.E., Cook, J., and Paton, J.C. (2007a). Development of a vaccine against invasive pneumococcal disease based on combinations of virulence proteins of *Streptococcus pneumoniae*. Infect Immun *75*, 350-357.

Ogunniyi, A.D., LeMessurier, K.S., Graham, R.M.A., Watt, J.M., Briles, D.E., Stroeher, U.H., and Paton, J.C. (2007b). Contributions of Pneumolysin, Pneumococcal Surface Protein A (PspA), and PspC to Pathogenicity of Streptococcus pneumoniae D39 in a Mouse Model. Infection and Immunity *75*, 1843-1851.

Ogunniyi, A.D., Woodrow, M.C., Poolman, J.T., and Paton, J.C. (2001a). Protection against Streptococcus pneumoniae elicited by immunization with pneumolysin and CbpA. Infect Immun *69*, 5997-6003.

Ogunniyi, A.D., Woodrow, M.C., Poolman, J.T., and Paton, J.C. (2001b). Protection against Streptococcus pneumoniae elicited by immunization with pneumolysin and CbpA. Infect Immun *69*, 5997-6003.

Orihuela, C.J., Gao, G., Francis, K.P., Yu, J., and Tuomanen, E.I. (2004). Tissue-specific contributions of pneumococcal virulence factors to pathogenesis. J Infect Dis *190*, 1661-1669.

Perkins, S.J., Nan, R., Li, K., Khan, S., and Miller, A. (2012). Complement factor H-ligand interactions: self-association, multivalency and dissociation constants. Immunobiology *217*, 281-297.

Perry, F.E., Elson, C.J., Greenham, L.W., and Catterall, J.R. (1993). Interference with the oxidative response of neutrophils by Streptococcus pneumoniae. Thorax *48*, 364-369.

Pletz, M.W., Maus, U., Krug, N., Welte, T., and Lode, H. (2008). Pneumococcal vaccines: mechanism of action, impact on epidemiology and adaption of the species. Int J Antimicrob Agents *32*, 199-206.

Pracht, D., Elm, C., Gerber, J., Bergmann, S., Rohde, M., Seiler, M., Kim, K.S., Jenkinson, H.F., Nau, R., and Hammerschmidt, S. (2005). PavA of Streptococcus pneumoniae modulates adherence, invasion, and meningeal inflammation. Infect Immun *73*, 2680-2689.

Prevention, C.f.D.C.a. (2012). Pneumococcal disease (Washington DC: Public Health Foundation).

Quin, L.R., Carmicle, S., Dave, S., Pangburn, M.K., Evenhuis, J.P., and McDaniel, L.S. (2005). In vivo binding of complement regulator factor H by Streptococcus pneumoniae. J Infect Dis *192*, 1996-2003.

Rajam, G., Anderton, J.M., Carlone, G.M., Sampson, J.S., and Ades, E.W. (2008). Pneumococcal surface adhesin A (PsaA): a review. Crit Rev Microbiol *34*, 131-142.

Reid, S.D., Hong, W., Dew, K.E., Winn, D.R., Pang, B., Watt, J., Glover, D.T., Hollingshead, S.K., and Swords, W.E. (2009). Streptococcus pneumoniae forms surface-attached communities in the middle ear of experimentally infected chinchillas. J Infect Dis *199*, 786-794.

Ricci, S., Janulczyk, R., Gerlini, A., Braione, V., Colomba, L., Iannelli, F., Chiavolini, D., Oggioni, M.R., Bjorck, L., and Pozzi, G. (2011). The factor H-binding fragment of PspC

as a vaccine antigen for the induction of protective humoral immunity against experimental pneumococcal sepsis. Vaccine *29*, 8241-8249.

Ricklin, D., Hajishengallis, G., Yang, K., and Lambris, J.D. (2010). Complement: a key system for immune surveillance and homeostasis. Nat Immunol *11*, 785-797.

Ring, A., Weiser, J.N., and Tuomanen, E.I. (1998). Pneumococcal trafficking across the blood-brain barrier. Molecular analysis of a novel bidirectional pathway. J Clin Invest *102*, 347-360.

Rosenow, C., Ryan, P., Weiser, J.N., Johnson, S., Fontan, P., Ortqvist, A., and Masure, H.R. (1997). Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of Streptococcus pneumoniae. Mol Microbiol *25*, 819-829.

Rovers, M.M. (2008). The burden of otitis media. Vaccine *26 Suppl 7*, G2-4.

Schmidt, C.Q., Herbert, A.P., Hocking, H.G., Uhrin, D., and Barlow, P.N. (2008). Translational mini-review series on complement factor H: structural and functional correlations for factor H. Clin Exp Immunol *151*, 14-24.

Schneider, M.C., Prosser, B.E., Caesar, J.J., Kugelberg, E., Li, S., Zhang, Q., Quoraishi, S., Lovett, J.E., Deane, J.E., Sim, R.B.*, et al.* (2009). Neisseria meningitidis recruits factor H using protein mimicry of host carbohydrates. Nature *458*, 890-893.

Schwarz-Linek, U., Hook, M., and Potts, J.R. (2004). The molecular basis of fibronectin-mediated bacterial adherence to host cells. Mol Microbiol *52*, 631-641.

Senkovich, O., Cook, W.J., Mirza, S., Hollingshead, S.K., Protasevich, II, Briles, D.E., and Chattopadhyay, D. (2007). Structure of a complex of human lactoferrin N-lobe with pneumococcal surface protein A provides insight into microbial defense mechanism. J Mol Biol *370*, 701-713.

Serruto, D., Rappuoli, R., Scarselli, M., Gros, P., and van Strijp, J.A. (2010). Molecular mechanisms of complement evasion: learning from staphylococci and meningococci. Nat Rev Microbiol *8*, 393-399.

Simell, B., Ahokas, P., Lahdenkari, M., Poolman, J., Henckaerts, I., Kilpi, T.M., and Kayhty, H. (2009). Pneumococcal carriage and acute otitis media induce serum antibodies to pneumococcal surface proteins CbpA and PhtD in children. Vaccine *27*, 4615-4621.

Simell, B., Jaakkola, T., Lahdenkari, M., Briles, D., Hollingshead, S., Kilpi, T.M., and Kayhty, H. (2006). Serum antibodies to pneumococcal neuraminidase NanA in relation to pneumococcal carriage and acute otitis media. Clinical and Vaccine Immunology *13*, 1177-1179.

Song, J.H., Dagan, R., Klugman, K.P., and Fritzell, B. (2012). The relationship between pneumococcal serotypes and antibiotic resistance. Vaccine *30*, 2728-2737.

Tai, S.S. (2006). Streptococcus pneumoniae protein vaccine candidates: properties, activities and animal studies. Crit Rev Microbiol *32*, 139-153.

van der Poll, T., and Opal, S.M. (2009). Pathogenesis, treatment, and prevention of pneumococcal pneumonia. Lancet *374*, 1543-1556.

Voss, S., Hallstrom, T., Saleh, M., Burchhardt, G., Pribyl, T., Singh, B., Riesbeck, K., Zipfel, P.F., and Hammerschmidt, S. (2013). The Choline-binding Protein PspC of Streptococcus pneumoniae Interacts with the C-terminal Heparin-binding Domain of Vitronectin. The Journal of biological chemistry *288*, 15614-15627.

Waltman, W.D., McDaniel, L.S., Gray, B.M., and Briles, D.E. (1990). Variation in the molecular weight of PspA (pneumococcal surface protein A) among Streptococcus pneumoniae. Microb Pathog *8*, 61-69.

Weinberger, D.M., Malley, R., and Lipsitch, M. (2011). Serotype replacement in disease after pneumococcal vaccination. Lancet *378*, 1962-1973.

Wellmer, A., Zysk, G., Gerber, J., Kunst, T., Von Mering, M., Bunkowski, S., Eiffert, H., and Nau, R. (2002). Decreased virulence of a pneumolysin-deficient strain of Streptococcus pneumoniae in murine meningitis. Infect Immun *70*, 6504-6508.

Winter, A.J., Comis, S.D., Osborne, M.P., Tarlow, M.J., Stephen, J., Andrew, P.W., Hill, J., and Mitchell, T.J. (1997). A role for pneumolysin but not neuraminidase in the hearing loss and cochlear damage induced by experimental pneumococcal meningitis in guinea pigs. Infect Immun *65*, 4411-4418.

Wu, J., Wu, Y.Q., Ricklin, D., Janssen, B.J., Lambris, J.D., and Gros, P. (2009). Structure of complement fragment C3b-factor H and implications for host protection by complement regulators. Nat Immunol *10*, 728-733.

Yordy, J.S., and Muise-Helmericks, R.C. (2000). Signal transduction and the Ets family of transcription factors. Oncogene *19*, 6503-6513.

Yother, J., and Briles, D.E. (1992). Structural properties and evolutionary relationships of PspA, a surface protein of Streptococcus pneumoniae, as revealed by sequence analysis. J Bacteriol *174*, 601-609.

Yu, J., Wiita, P., Kawaguchi, R., Honda, J., Jorgensen, A., Zhang, K., Fischetti, V.A., and Sun, H. (2007). Biochemical analysis of a common human polymorphism associated with age-related macular degeneration. Biochemistry *46*, 8451-8461.

Zhang, J.R., Mostov, K.E., Lamm, M.E., Nanno, M., Shimida, S., Ohwaki, M., and Tuomanen, E. (2000). The polymeric immunoglobulin receptor translocates pneumococci across human nasopharyngeal epithelial cells. Cell *102*, 827-837.

Zhang, Q.B., Bernatoniene, J., Bagrade, L., Pollard, A.J., Mitchel, T.J., Paton, J.C., and Finn, A. (2006). Serum and mucosal antibody responses to pneumococcal protein antigens in children: relationships with carriage status. Eur J Immunol *36*, 46-57.

Zhang, Q.B., Choo, S., and Finn, A. (2002). Immune responses to novel pneumococcal proteins pneumolysin, PspA, PsaA, and CbpA in adenoidal B cells from children. Infect Immun *70*, 5363-5369.

Zipfel, P.F., Hallstrom, T., Hammerschmidt, S., and Skerka, C. (2008). The complement fitness factor H: role in human diseases and for immune escape of pathogens, like pneumococci. Vaccine *26 Suppl 8*, I67-74.

CHAPTER 2

Structural Basis for Host Specificity of Factor H Recruitment

by *Streptococcus pneumoniae*

Author Contributions

David Achila developed protein production and purification protocols, purified all proteins, and did all ITC experiments; Aizhuo Liu performed NMR experiments, NMR data analysis and structure determination, crystallization screening and optimization, and X-ray diffraction data collection; Rahul Banerjee carried out molecular dynamics simulations and MM-GBSA analysis; Erik Martinez-Hackert phased the X-ray data using the NMR structure determined by Aizhuo Liu as the search model and built and refined the crystal structure; Yue Li made the overexpression systems and site-directed mutants; Jing-Ren Zhang and Honggao Yan conceived the project; Honggao Yan supervised this work and coordinated the writing of a manuscript to be submitted to EMBO Journal in August, 2013.

ABSTRACT

Many human pathogens have strict host specificity, which affects not only their epidemiology but also development of animal models and vaccines. Complement factor H (FH) is recruited to pneumococcal cell surface in a human-specific manner via the pneumococcal protein virulence factor, choline binding protein A (CbpA). FH recruitment enables *Streptococcus pneumoniae* to evade surveillance by the human complement system and contributes to pneumococcal host specificity. The molecular determinants of host specificity of complement evasion are unknown. Here we report the structural determinants for the host-specific FH recruitment by the pneumococcus. The results reveal how FH recruitment can serve as a mechanism for both pneumococcal complement evasion and adherence and suggest new approaches for development of better animal models for pneumococcal infection and redesign of the virulence factor for pneumococcal vaccine development.

2.1 Introduction

The complement system is a major component of vertebrate innate immunity and one of the most important frontline defenders against microbial invaders (Ricklin et al., 2010; Serruto et al., 2010). Uncontrolled or inappropriate complement activation, however, can be damaging to the host. A major player in regulating complement activation and preventing complement from damaging host tissues is complement factor H (FH)(Ferreira et al., 2010; Zipfel and Skerka, 2009), a 155 kDa protein with 20 homologous domains termed complement control protein (CCP) modules, also known as short consensus repeats (SCR). FH recruitment is a mechanism widely exploited by microbial pathogens to evade complement attack (Ferreira et al., 2010; Serruto et al., 2010). Recent studies revealed that human pathogens, *Streptococcus pneumonia*e, *Neisseria meningitides* and *Neisseria gonorrhoeae*, recruit FH in a human-specific manner, indicating FH recruitment as a contributing factor to the host specificity of these pathogens (Granoff et al., 2009; Lu et al., 2008; Ngampasutadol et al., 2008). Conversely, the broad host range of the zoonotic bacterial pathogen *Borrelia burgdorferi* sensu lato is attributed partly to its many FH-binding proteins (Kurtenbach et al., 2002; Stevenson, 2002).

S. pneumoniae is a major cause of human acute otitis media, pneumonia, and meningitis, leading to high morbidity and mortality, particularly in children and the elderly, and enormous economic burden in both developed and developing countries (Huang et al., 2011; O'Brien et al., 2009). Under natural conditions, *S. pneumoniae* displays strict host specificity to humans (Musher, 2009). Animal studies showed that the complement system is essential for immunity to *S. pneumoniae* (Gross et al., 1978;

Hosea et al., 1980; Kerr et al., 2005). Clinical surveys indicated that patients with complement deficiencies have increased susceptibility to recurrent *S. pneumoniae* infections (Alper et al., 1970; Carneiro-Sampaio and Coutinho, 2007; Sampson et al., 1982).

Two most critical aspects of pneumococcal pathogenesis are adherence and immune evasion (Hammerschmidt, 2006). The former allows pneumococci to colonize the human host and the latter enables pneumococci to evade the attack of the human immune system. CbpA protein, also termed PspC or SpsA, is a major pneumococcal virulence factor and plays an important role in both pneumococcal adherence and immune evasion (Hammerschmidt, 2006). CbpA is a multidomain protein attached to pneumococcal cell wall via its C-terminal phosphorylcholine-binding domain. Its middle homologous R1 and R2 domains mediate pneumococcal adherence and transcytosis across mucosal respiratory epithelial cells by binding to human polymeric immunoglobulin receptor (pIgR) (Brock et al., 2002; Hammerschmidt et al., 2000; Lu et al., 2003; Luo et al., 2005; Zhang et al., 2000). CbpA's N-terminal domain (CbpAN) promotes complement evasion by binding to human FH (hFH) (Hammerschmidt et al., 2007; Lu et al., 2006). Interestingly, CbpA binds only to human pIgR and FH, but not to their counterparts in other animal species tested thus far, suggesting that these CbpA-host interactions contribute to the strict host specificity of *S. pneumoniae* (Hammerschmidt et al., 2000; Lu et al., 2008). Animal model studies demonstrated that CbpA-deficient pneumococcal strains have attenuated capacity to colonize and cause infections (Balachandran et al., 2002; Iannelli et al., 2004; Rosenow et al., 1997). CbpA is one of a few pneumococcal protein virulence factors that can stimulate production of

antibodies in humans (McCool et al., 2002; McCool et al., 2003) and offers protection against challenge by virulent pneumococci in animal models (Balachandran et al., 2002; Brooks-Walter et al., 1999; Cao et al., 2007; Ogunniyi et al., 2007; Ogunniyi et al., 2001; Rosenow et al., 1997).

While it is well established that CbpA binds to hFH via its N-terminal domain (Hammerschmidt et al., 2007; Lu et al., 2006), the hFH region involved in binding of CbpA has not been definitively established. Previous studies indicated that CbpA binds to hFH domains CCP6–10 (Dave et al., 2004) or CCP13–15 (Duthy et al., 2002), but a more recent study indicated that CbpA interacts with two regions of hFH, CCP8–11 and CCP19–20 (Hammerschmidt et al., 2007). Furthermore, the structural basis for binding and host specificity is not known. Here it has been shown that a single hFH domain is sufficient for recognition by CbpAN. We present the solution NMR structure of CbpAN and the crystal structure of the hFH domain in complex with CbpAN, and by computational analysis and reciprocal site-directed mutagenesis, i.e., substitution of hFH residues with mouse FH (mFH) residues and vice versa, we demonstrate that a “hydrophobic lock” on the hFH domain is a critical structural determinant for host specific FH recruitment.

2.2 Methods

2.2.1 Construction of the CbpAN overexpression system and protein purification

The domain boundaries of CbpAN of the TIGR4 strain were defined based on the secondary structure prediction using the SOPMA server (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) (Geourjon and Deleage, 1995). The DNA fragment encoding CbpAN was amplified by PCR from the genomic DNA of the *S. pneumoniae* strain TIGR4 with primers CbpANf and CbpANr (Table 2.1). The amplified DNA fragment was cloned into the expression vector pET17bHR (lab-made, derived from pET17b) by digestion with the restriction enzymes *BamHI* and *NdeI* and ligation. The cloned DNA fragment was sequenced to ensure the correct coding sequence. The overexpression plasmid construct was transformed into the *Escherichia coli* strain BL21(DE3)pLysS for the production of the His-tagged CbpAN. The expression strain was cultured in LB media containing 20 μg/ml chloramphenicol and 100 μg/ml ampicillin with vigorous shaking at 37 °C. The culture was placed on ice when its OD_{600} reached 0.8. The production of CbpAN was then induced with 0.5 mM IPTG and the culture was incubated with shaking overnight at 16 °C.

The cells were harvested by centrifugation and resuspended in buffer A containing 50 mM sodium phosphate and 300 mM NaCl, pH 8.0, and lysed by sonication. The lysate was clarified by centrifugation at 16,000 g for 20 min before loading onto a Ni-NTA agarose column pre-equilibrated with buffer A. After washing with 10 mM imidazole in buffer A, the column was eluted with a linear 10-250 mM imidazole gradient. The protein-containing fractions were analyzed by SDS-PAGE. Fractions containing pure CbpAN were pooled and concentrated using an Amicon concentrator.

When desired, the His-tag was cleaved with TEV protease (Sun et al., 2011). The cleaved His-tag and the uncleaved protein were separated by a second passage through a Ni-NTA agarose column. The flow-through was collected, concentrated and loaded onto a Sephadex G-75 gel filtration column. The fractions containing pure CbpAN were pooled, concentrated, and dialyzed first against 5 mM sodium phosphate, pH 8.0, and against water for 12 h. The dialyzed protein solution was lyophilized and stored at -80 °C. Isotopically labeled (^{15}N- or ^{15}N/^{13}C-) CbpAN was produced and purified in the same way except that LB media were replaced with M9 media containing $^{15}NH_4Cl$ as the sole nitrogen source or $^{15}NH_4Cl$ and ^{13}C-glucose as the sole nitrogen and the sole carbon source, respectively.

2.2.2 Construction of the overexpression systems for FH domains and protein purification

Synthetic DNA coding for hFH and mouse FH (mFH) with codons optimized for expression in *E. coli* were synthesized by Biomatik. DNA fragments encoding various CCPs were cloned into the expression vector pET17bHMHT (lab-made, derived from pET17b (Novagen)) using the same recombinant DNA methods as described earlier. The primers for PCR cloning were hFH9f and hFH9r for hFH CCP9, hFH8f and hFH10r for hFH CCP8-10, hFH9f and hFH10r for hFH CCP9-10, hFH13f and hFH15r for hFH CCP13-15, and hFH19f and hFH20r for hFH CCP19-20 (Table 2.1). The correct coding sequences were confirmed by DNA sequencing. The overexpression plasmid constructs were transformed into the *E. coli* strain SHuffle T7 Express *LysY* (NEB). The fusion proteins produced by the overexpression systems contained two His-tags, one before

the maltose binding protein (MBP) and another between the MBP and the CCPs, and a thrombin cleavage site between the second His-tag and the CCPs. The procedure for protein production and cell harvesting was the same as described earlier.

The harvested *E. coli* cells were resuspended in buffer B containing 20 mM Tris-HCl and 150 mM NaCl, pH 7.5, and lysed by sonication. The lysate was clarified by centrifugation and loaded onto a Ni-NTA agarose column pre-equilibrated with buffer B. After washing with buffer B containing 20 mM imidazole, the column was eluted with a linear 20-250 mM imidazole gradient. The protein-containing fractions were analyzed by SDS-PAGE, pooled and concentrated. An equimolar amount of lab-made His-tagged *E. coli* DsbC (Wang, 2002) was added to the concentrated fusion protein solution for disulfide bond reshuffling. After an overnight dialysis against buffer B at room temperature, a $CaCl_2$ stock solution was added to a final concentration of 0.125 mM, followed by the addition of 0.75 U thrombin/mg of the fusion protein to cleave off the MBP tag. After 3 h of incubation at room temperature, phenylmethylsulfonyl fluoride (PMSF) was added to a final concentration of 0.5 mM to inactivate thrombin and therefore terminate the cleavage reaction. The uncleaved fusion protein and the cleaved MBP were removed by chromatography on a second Ni-NTA agarose column eluted with a linear 0-100 mM imidazole gradient in buffer B. The fractions containing CCPs were concentrated and further purified by gel filtration chromatography on a Sephadex G-75 column. The purified protein solutions were dialyzed first against 5 mM sodium phosphate, pH 8.0, then against 2 mM sodium phosphate, pH 8.0, and lyophilized. The lyophilized proteins were stored at -80 $^{\circ}$C.

Table 2.1 Primers for PCR cloning and mutagenesis of FH and CbpAN.

Name	Sequence
CbpANf	5’- G GAA TTC CAT ATG GAT TCA GAA CGA GAT AAG GCA AGG-3’
CbpANr	5’- CG GGA TCC TCA CTT TTC AAA CTT AGA CAC AGC TTC-3’
hFH8f	5’- G GAA TTC CAT ATG ACG TGT TCG AAA AGT TCC ATT G-3’
hFH9f	5’- G GAA TTC CAT ATG TCT TGT GAT ATC CCG GTT TTC ATG-3’
hFH9r	5’- G GGA TCC TCA ACG TTC GTA GCA AAT CGG CAG ATC-3’
hFH10r	5’- G GGA TCC TCA TTG GAC CTG TTC TTT GCA GAT CGG-3’
hFH13f	5’- G GAA TTC CAT ATG AAA TGT AAA AGC TCT AAC CTG ATT ATC-3’
hFH15r	5’- G GGA TCC TCA CAG ACC TTC GCA CTG CGG CGG-3’
hFH19f	5’- G GAA TTC CAT ATG AAA TGT GGT CCG CCG CCG CCG-3’
hFH20r	5’- G GGA TCC TCA GCG TTT CGC ACA GGT CGG-3’
hFHE523Af	5’- GAA TGC CAT GAT GGC TAC GCA TCA AAT ACG GGC TCG -3’
hFHE523Ar	5’- CGA GCC CGT ATT TGA TGC GTA GCC ATC ATG GCA TTC-3’
hFHD542Af	5’- GC TAT AAC GGT TGG TCT GCT CTG CCG ATT TGC TAC G-3’
hFHD542Ar	5’- CGT AGC AAA TCG GCA GAG CAG ACC AAC CGT TAT AGC-3’
hFHY547Af	5’- GG TCT GAT CTG CCG ATT TGC GCC GAA CGT TGA GGA TCC-3’
hFHY547Ar	5’- GGA TCC TCA ACG TTC GGC GCA AAT CGG CAG ATC AGA CC-3’
hFHM497Ef	5’- GT GAT ATC CCG GTT TTC GAG AAC GCT CGC ACC AAA AAT G-3’
hFHM497Er	5’- CAT TTT TGG TGC GAG CGT TCT CGA AAA CCG GGA TAT CAC-3’
hFHL543Tf	5’- C TAT AAC GGT TGG TCT GAT ACC CCG ATT TGC TAC GAA CGT-3’
hFHL543Tr	5’- CAC GTT CGT AGC AAA TCG GGG TAT CAG ACC AAC CGT TAT AG-3’
hFHI545Sf	5’- GG TCT GAT CTG CCG AGT TGC TAC GAA CGT-3’
hFHI545Sr	5’- ACGTTCGTAGCAACTCGGCAGATCAGACC-3’
hFHL543T+I545Sf	5’- GG TCT GAT ACC CCG AGT TGC TAC GAA CGT-3’
hFHL543T+I545Sr	5’- ACGTTCGTAGCAACTCGGGGTATCAGACC-3’
hFHN525E+G527Kf	5’- GC CAT GAT GGC TAC GAA TCA GAA ACG AAA TCG ACC ACG GGT AGC ATT GTG-3’

Table 2.1 (Cont'd)

hFHN525E+G527Kr	5'- CAC AAT GCT ACC CGT GGT CGA TTT CGT TTC TGA TTC GTA GCC ATC ATG GC-3'
mFH9f	5'- G GA ATT CAT ATG TCT TGC GAC ATG CCG GT-3'
mFH9r	5'- G GGA TCC TCA ACG TTC GTA GCA AGA CGG G-3'
mFHE497Mf	5'-CT TGC GAC ATG CCG GTT TTC ATG AAC TCT ATC ACC AAA AAC AC -3'
mFHE497Mr	5'- GT GTT TTT GGT GAT AGA GTT CAT GAA AAC CGG CAT GTC GCA AG -3'
mFH9T543L+S545lf	5'- CC TAC TAC GGT TGG TCT GAC CTC CCG ATT TGC TAC GAA CGT TGA GGA TCC -3'
mFH9T543L+S545lr	5'- GGA TCC TCA ACG TTC GTA GCA AAT CGG GAG GTC AGA CCA ACC GTA GTA GG -3'
mFHE525N+K527Gf	5'- C CTG GTT GGT TTC GAA AAC AAC TAC GGT CAC ACC AAA GGT TCT ATC AC -3'
mFHE525N+K527Gr	5'- GTG ATA GAA CCT TTG GTG TGA CCG TAG TTG TTT TCG AAA CCA ACC AGG -3'
mFHK527Gf	5'- GTT GGT TTC GAA AAC GAA TAC GGT CAC ACC AAA GGT TCT ATC ACC-3'
mFHK527Gr	5'- GGT GAT AGA ACC TTT GGT GTG ACC GTA TTC GTT TTC GAA ACC AAC-3'
mFHY537Nf	5'- C ACC TGC ACC TAC AAC GGT TGG TCT GAC -3'
mFHY537Nr	5'- GTC AGA CCA ACC GTT GTA GGT GCA GGT G-3'

2.2.3 Site-directed mutagenesis

FH mutants were generated using the PCR-based QuickChange method according to the protocol developed by Stratagene and the primers listed in Table 2.1. Mutations were selected and correct coding sequences verified by DNA sequencing. Multiple mutations were generated in a sequential fashion.

2.2.4 Isothermal titration calorimetry (ITC)

ITC measurements were performed at 25 °C on a VP-ITC isothermal titration calorimeter (MicroCal) according to the protocol of Velazquez-Campoy and Freire (Velazquez-Campoy and Freire, 2006). The lyophilized proteins were dissolved in buffer

C containing 50 mM HEPES and 50 mM KCl, pH7.5, and dialyzed against the same buffer. The concentrations of the dialyzed protein solutions were determined by measuring OD_{280} and using extinction coefficients calculated by the method of Gill and von Hippel (Gill and Vonhippel, 1989) using the ProtParam tool of the ExPASy web server.(Gasteiger et al., 2005) The sample cell was loaded with FH protein and the syringe with CbpAN. The binding enthalpies were obtained by injecting CbpAN into the sample cell under stirring conditions. A typical ITC experiment consisted of 25 injections 10 µl each at 6 min intervals. The ITC data were analyzed using Origin 5.0.

2.2.5 Solution structure determination

All protein NMR samples were prepared by dissolving the lyophilized cleaved CbpAN in a pH 7.0 buffer composed of 50 mM phosphate, (pH 7.0) 50 mM NaCl, 100 µM NaN_3, and 50 µM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS, as the internal NMR reference) in either 95% H_2O:5% D_2O or 100% D_2O. For the solution structure determination of free CbpAN, protein concentration was typically ~1.0 mM. For the NMR assignment of CbpAN in complex with hFH CCP9, the samples were prepared by mixing 1.0 mM labeled CbpAN with 1.1 mM unlabeled hFH CCP9. For assaying the binding of CbpAN to different hFH CCPs, each NMR sample consisted of 0.2 mM ^{15}N-CbpAN and 0.5-1.0 mM unlabeled hFH CCPs.

All NMR data were acquired at 20 $^{\circ}$C on a Bruker AVANCE 900 MHz NMR spectrometer equipped with the TCI cryo-probe head. The raw NMR data were processed using NMRPipe (Delaglio et al., 1995) and analyzed with NMRView

(Johnson and Blevins, 1994). Sequential resonance assignments were accomplished by the analysis of standard 3D double- and triple-resonance NMR data (Cavanagh et al., 2006). Backbone resonance assignment was achieved by the analysis of a set of standard triple-resonance NMR data acquired with $^{13}C/^{15}N$-labeled CbpAN in the buffer containing 5% D_2O, including HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH, HNCO, and HN(CA)CO. The assignment was confirmed by the analysis of a 3D ^{15}N-edited NOESY spectrum acquired with a ^{15}N-labeled CbpAN sample. The resonance assignment of aliphatic side chains was accomplished through the analysis of 3D HCCH-TOCSY data, and the assignment of aromatic side chains through the analysis of HBCB (CG)CDHD and HBCB(CG)(CD)CEHE NMR data. The same strategy was used for the sequential assignment of labeled CbpAN in complex with unlabeled hFH CCP9. 3D ^{13}C-dispersed NOESY spectra were recorded using a protein sample in the buffer made in 100% D_2O. The mixing times for all NOESY spectra were set to 70 ms.

The solution structure of free CbpAN was calculated using CYANA(Güntert, 2004) (version 2.1) with inputs of inter-proton distance restraints derived from NOEs and backbone torsion-angle restraints derived from chemical shifts using TALOS(Shen et al., 2009). The top 20 conformers generated from 50 initial random ones by the CYANA calculations (with the lowest values of the target function) were subjected to energy minimization in explicit solvent using the program AMBER 10 (Case et al., 2008) and the ff99SB force field. Each of 20 conformers was neutralized with a Cl^- ion and solvated in a rectangular TIP3P water box with its boundaries at least 10 Å away from the protein atoms. The system was minimized in two stages. In the first stage, the water

molecules and Cl^- ion were minimized by 15,000 steps of steepest descent followed 5,000 steps of conjugated gradient with the conformation of the protein molecule fixed. In the second stage, the whole system was minimized with 25,000 steps of steepest descent followed by 15,000 steps of conjugated gradient with the protein molecule restrained only by the distance and torsion-angle restraints derived from the NMR data. The statistics of the AMBER-refined structure are very similar to those of the original structure obtained by the CYANA calculations except that the number of hydrogen bonds increased dramatically, from 38 before to 61 after energy minimization.

2.2.6 Crystal structure determination

The complex of CbpAN and hFH CCP9 was made by mixing the two proteins with a molar ratio of 1.1 to 1.0. The mixed protein solution was concentrated and further purified by FPLC using a HiLoad 26/60 gel filtration column. The complex emerged as a sharp peak well separated from the free CbpAN. The fractions containing the protein complex were pooled, concentrated to ~10 mg/ml, and dialyzed against a buffer containing 10 mM Tris-HCl, pH 8.0. The dialyzed protein solution was concentrated up to ~20 mg/ml and centrifuged to remove any potential insoluble material. Crystallization screening was performed with a Cryphon robot (Art Robbins Instrument) using QIAGEN screening kits and the sitting drop method. Crystallization conditions were optimized manually by the hanging drop vapor diffusion method at 20 $^{\circ}$C. The optimized well solution was composed of 100 mM sodium acetate, pH 4.6, and 18-20% (w/v) PEG 3350. Crystals appeared overnight and reached typically the size of $0.25 \times 0.25 \times 0.35$ mm within 3 days. The crystals were then transferred into reservoir buffers containing

sequentially higher concentrations of glycerol (5–25%, v/v) and flash frozen in liquid nitrogen. X-ray data from frozen crystals were collected at the Advanced Photon Source, LS-CAT beamline 21-ID. Crystals diffracted to ~1.10 Å Bragg spacings (for processing and refinement statistics, see Table 2.5). Two datasets were used in the structure determination, including a 1.08-Å native dataset and a 2.29-Å dataset from a NaI soaked crystal. Of the 20 hFH CCPs, no structure has been reported for six CCPs, including CCP9. Phasing with the crystal structures of other CCPs was not successful. An MR solution was found for the 2.29 Å NaI dataset using Phaser (McCoy et al., 2007) with the CbpAN NMR structure as the search model. Subsequent model building and refinement with ARP/wARP (Cohen et al., 2008) and REFMAC5 (Murshudov et al., 2011) using the 1.08 Å native data set yielded essentially a complete model, including the missing hFH CCP9. Electron density was clear and continuous for both proteins and the only disordered regions were found at the N- and C-termini. The model was further refined by simulated annealing using PHENIX (Adams et al., 2010). Small errors were corrected manually with Coot (Emsley and Cowtan, 2004). Water molecules were added during the final rounds of refinement using PHENIX (Adams et al., 2010). Figures were prepared using PyMOL (DeLano, 2002).

2.2.7 Molecular dynamics (MD) simulations and MM-GBSA analysis

All MD simulations were performed using AMBER 12 (Case et al., 2012) and the ff12SB force field. The initial coordinate for the MD simulation of the complex of hFH CCP9 and CbpAN was taken from the crystal structure of the complex. The initial coordinates for the MD simulation of the complex of mFH CCP9 and CbpAN was

obtained by homology modeling. A model of the bound mFH CCP9 was built using the SWISS-MODEL server (Arnold et al., 2006; Bordoli et al., 2009; Schwede et al., 2003) with the structure of the bound hFH CCP9 as the template. The sequence identity between hFH and mFH CCP9 is 65%. The model of the complex of mFH CCP9 and CbpAN was assembled with the crystal structure of the bound CbpAN and the modeled structure of mFH CCP9. The complexes of FH mutant proteins were generated by assuming that the mutations have no significant effects on the structures of the proteins. Each complex was solvated in a rectangular TIP3P water box, with the boundaries at least 10 Å away from the protein atoms, and neutralized with appropriate number of Na^+ ions. The system was energy minimized with a combination of steepest descent and conjugated gradient methods, 5,000 steps each, first on the water molecules with all heavy atoms of the proteins restrained and then on the whole system without any positional restraint. The minimized system was heated from 0 to 300 K in 500,000 steps under constant volume conditions (NVT) and equilibrated for 4 ns under isobaric, isothermal (NPT) conditions at 300 K. The system was simulated further under NPT conditions at 300 K for 40 ns. Temperature was controlled by Langevin dynamics with a collision frequency of 1.0 ps^{-1}. The time step was 2 fs, with all covalent bonds to hydrogen atoms restrained using the SHAKE algorithm (Ryckaert et al., 1977). Long-range electrostatic interactions were evaluated using the particle mesh Ewald method (Essmann et al., 1995) with a nonbonded cutoff of 10 Å. All MD trajectories are stable based on the time evolution of RMSD.

MM-GBSA analysis and energy decomposition were performed with a single trajectory approach using the python script MMPBSA.py (Miller et al., 2012) provided

with AmberTools 12 (Case et al., 2012). A generalized Born implicit solvent model with 0.1 molar salt concentration was used for the calculations. The calculations were performed on 500 snapshots for each simulation and averaged to estimate the global free energy change of the binding and the contribution from each FH9 and CbpAN residue towards the global free energy change. The entropic contribution to the free energy was not considered in the calculations as in most recent studies (Case et al., 2012), because we were interested in only the mutational effects on binding, i.e., the relative free energies, and entropy calculation is not as accurate as the enthalpy calculation.

2.3 Results

2.3.1 Purification of functional FH and CbpA proteins

CbpAN and FH proteins were purified to homogeneity by a combination of affinity and gel filtration chromatographic protocols (Fig. 2.1). Optimal conditions for protein expression were first determined by trying out a variety of *E. coli* strains and different growth conditions. Soluble functional CbpAN and FH proteins were better expressed at 16 $^{\circ}$C overnight in BL 21 (DE3) pLysS and Shuffle pLysY respectively. FH proteins were expressed as MBP-fusion proteins and to improve population of active protein, a DsbC facilitated *in vitro* disulfide bond reshuffling step was incorporated before thrombin cleavage. A second His-tag was also engineered into the expression vector between MBP fusion protein and thrombin cleavage site to improve the binding of FH proteins to Ni^{+}-NTA column. Protein yield of ~70 mg/L and 5 mg/L were realized for CbpAN and FH9 respectively. We therefore established a novel technique for isolation of functional

FH proteins from *E. coli*, which was likely hindering biophysical characterization of this important protein that is hijacked by several pathogens to evade the host immune system. The secondary structure of CbpAN predicted by the SABLE webserver (http://sable.cchmc.org/) (Fig. 2.2) is consistent with that of solution NMR structure solved later (Fig. 2.6a).

2.3.2 Localization of the CbpA-binding domain in hFH

To identify which domain or domains of hFH bind to CbpA, we produced several truncated hFH proteins, CCP6-8, CCP7-9, CCP8-10, CCP10-12, CCP13-15, CCP16-18, and CCP19-20 and measured their binding activity by the biophysical methods ITC and NMR. Of these truncated hFH proteins, only CCP7-9 and CCP8-10 bound to CbpAN. The binding was tight with a K_d of ~1.5 nM (Tables 2.2 and 2.3). To check whether all three domains of CCP8-10 are needed for tight binding of CbpAN, we deleted the N-terminal CCP8 first and then the C-terminal CCP10 and measured their effects on binding CbpA. The K_d values of CCP9-10 and CCP9, were similar to that of CCP8-10, all under 5 nM (Fig. 2.3 and Tables 2.3). We also produced CCP10 and it showed no binding activity based on ITC measurements. These results together indicated that a single CCP domain is sufficient for binding of CbpAN and neither CCP8 nor CCP10 contributes significantly to binding of CbpA.

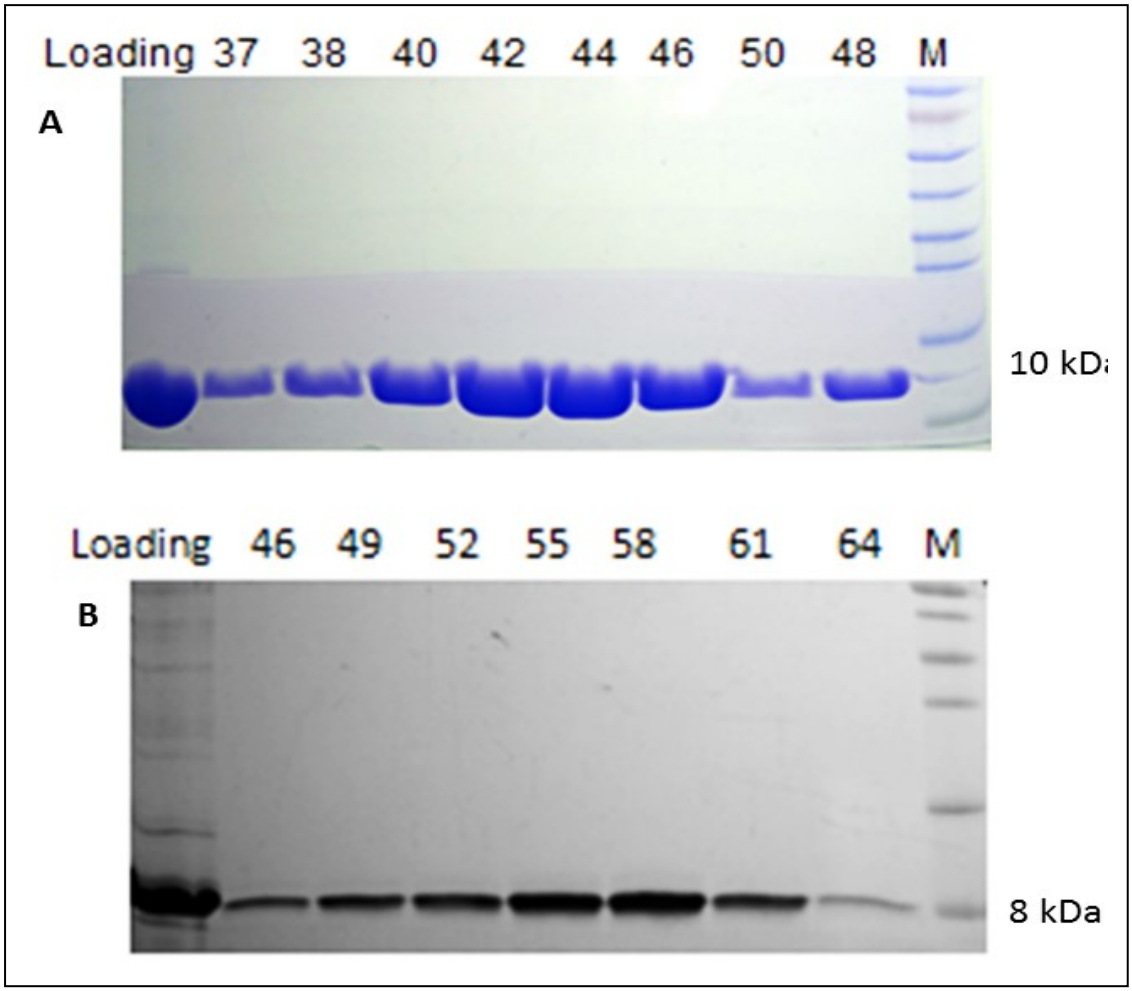


Figure 2.1 15 % Comassie stained gels of CbpAN (A) and FH CCP9 (B) gel filtration fractions. M lane is for molecular protein marker and the rest of the lanes are for the fraction numbers indicated.

Because CCP13–15 (Duthy et al., 2002) and CCP19–20 (Hammerschmidt et al., 2007) were previously reported to bind to CbpA but our ITC measurements could not detect the binding activity of either of these truncated hFH proteins, we further investigated these CCP domains by examining their effects on the chemical shifts of ^{15}N-CbpAN. The chemical shifts of CbpAN and its complex with hFH CCP9 were assigned by multidimensional NMR spectroscopy (Fig. 2.4A-C). Backbone amide

resonances were sequentially assigned for all residues of CbpAN in complex with hFH CCP9 and all but four residues of free CbpAN. The unassigned residues were Lys 95, Lys 96, Arg 97, and His 98, which are located between helix I and II (data not shown) and whose amide NMR signals were very weak. Addition of hFH CCP9 caused significant changes in the chemical shifts of many residues of CbpAN (Fig. 2.4A and Fig. 2.5). Significant chemical shift changes were mostly congregated in the region spanning residues 85-111 (Fig. 2.5), the C-terminal segment of helix I and the N-terminal segment of helix II. In contrast, addition of 0.5 mM CCP13-15 (Fig. 2.4B) or 1.0 mM CCP19-20 (Fig. 2.4C) caused essentially no change in the chemical shifts of CbpAN. The NMR chemical shift perturbation experiments indicated that neither CCP13-15 nor CCP19-20 binds to CbpAN even at millimolar concentrations. Chemical shift changes were calculated using equation: $\Delta\delta = \{0.5[\Delta\delta(^{1}H^{N})^{2} + (0.2\Delta\delta(^{15}N))^{2}\}^{1/2}$

Table 2.2 Determination of the region of FH protein that interacts with CbpAN by ITC.

FH Proteins	Domains	Interaction
FH6-8	6-8	-
FH7-9	7-9	+
FH8-11	8-11	+
FH10-12	10-12	-
FH13-15	13-15	-
FH16-18	16-18	-
FH19-20	19-20	-

(+) indicates interaction while (–) denotes no interaction

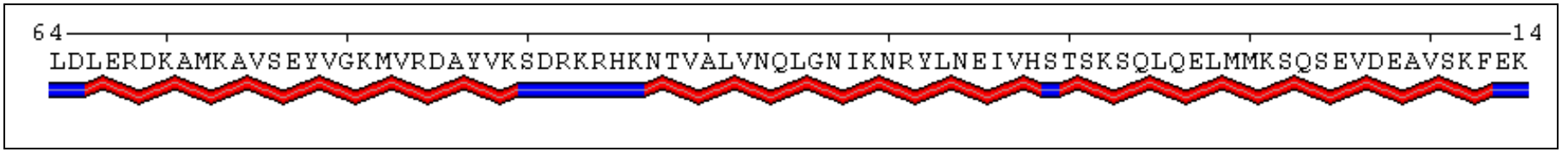


Figure 2.2 Prediction of secondary structure of CbpAN (TIGR4) by Sable webserver. Wavy red tubes represent the 3 predicted α-helices of CbpAN while blue tubes represent loops.

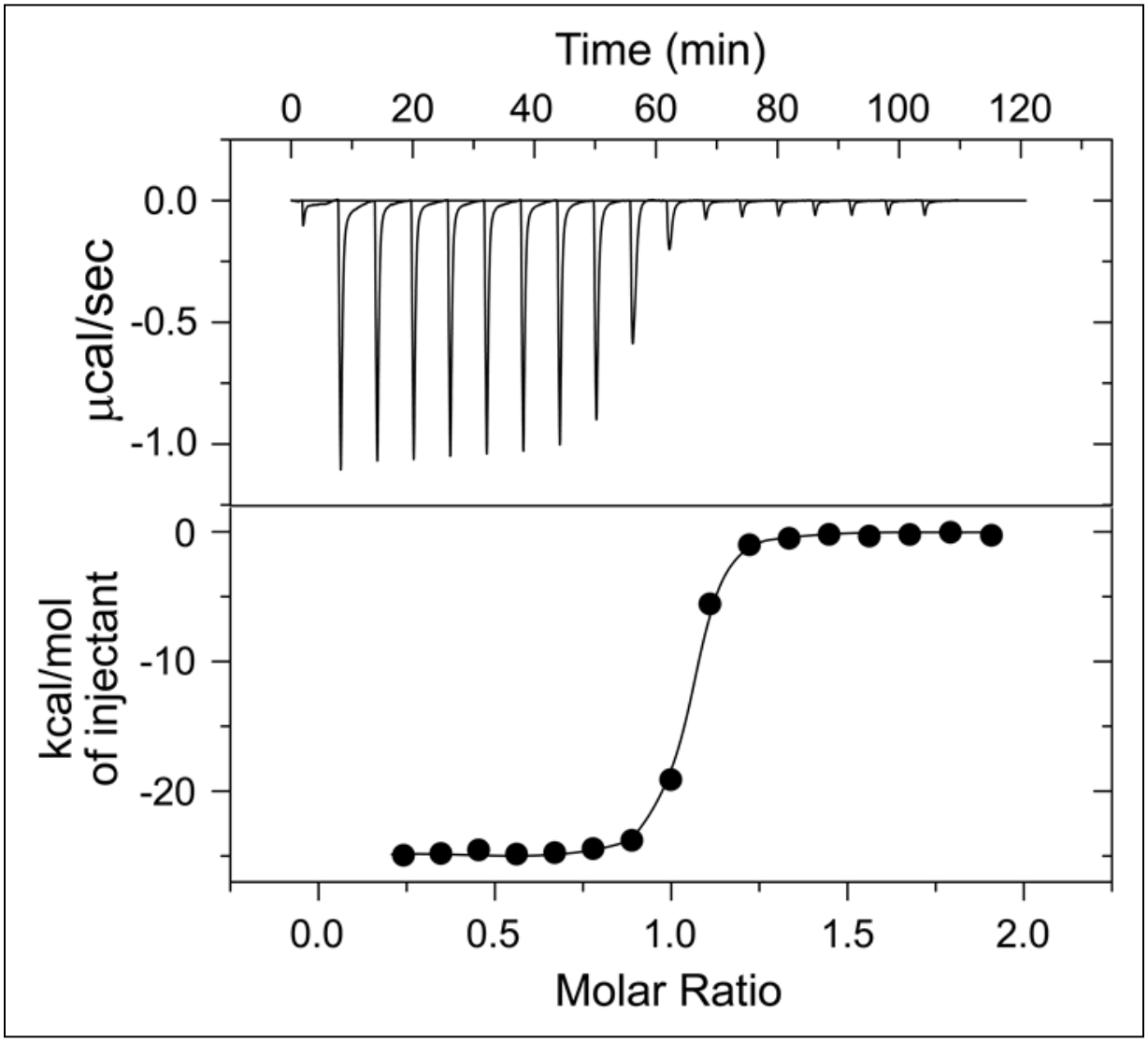


Figure 2.3 ITC analysis of the binding of hFH CCP9 and CbpAN.

Table 2.3 Thermodynamic binding parameters of human and mouse FH domains.

Protein	K_d (nM)	ΔG (kcal/mol)	ΔH (kcal/mol)	TΔS (kcal/mol)
hFH CCP				
CCP8-10	1.49 ± 0.5	−12.02 ± 0.17	−18.45 ± 0.46	−6.44 ± 0.43
CCP9-10	3.4 ± 1	−11.54 ± 0.15	−19.63 ± 0.67	−7.92 ± 0.66
CCP9	4.72 ± 0.4	−11.35 ± 0.05	−21.12 ± 0.67	−9.86 ± 0.62
CCP10	ND			
CCP13-15	ND			
CCP19-20	ND			
hFH CCP9 Mutant Proteins				
M497E	206 ± 20	−9.11 ± 0.05	−15.53 ± 0.32	−6.4 ± 0.29
E523A	47± 4	−9.99 ± 0.04	−15.5 ± 0.6	−5.5 ± 0.66
D542A	1640 ± 40	−7.88 ± 0.01	−18.35 ± 2.6	−10.5 ± 2.6
L543T	3580 ± 1800	−7.42 ± 0.24	−36.0 ± 0	−28.4 ± 0.48
I545S	12300 ± 590	−6.69 ± 0.03	−12.8 ± 0.8	−5.75 ± 0.44
Y547A	38000 ± 1200	−6.02 ± 0.03	−2.04 ± 0.2	−3.68 ± 0.05
M497E+L543T	17400 ± 850	−6.49 ± 0.02	−2.44 ± 0.3	−4.06 ± 0.29
M497E+L543T+I545S	ND			
mFH CCP9 and Mutant Protein				
CCP9	ND			
E497M+T543L+S545I	973 ± 21	−8.19 ± 0.01	−12.17 ± 0.13	−4.02 ± 0.18
E497M+T543L+S545I +Y538N	1080 ± 130	−8.13 ± 0.07	−13.2 ± 0	−5. 69 ± 0.24
E497M+T543L+S545I +E525N+K527G	5.78 ± 0.6	−11.23 ± 0.06	−14.5 ± 0	−3.41 ± 0.17

*ND, not detectable.

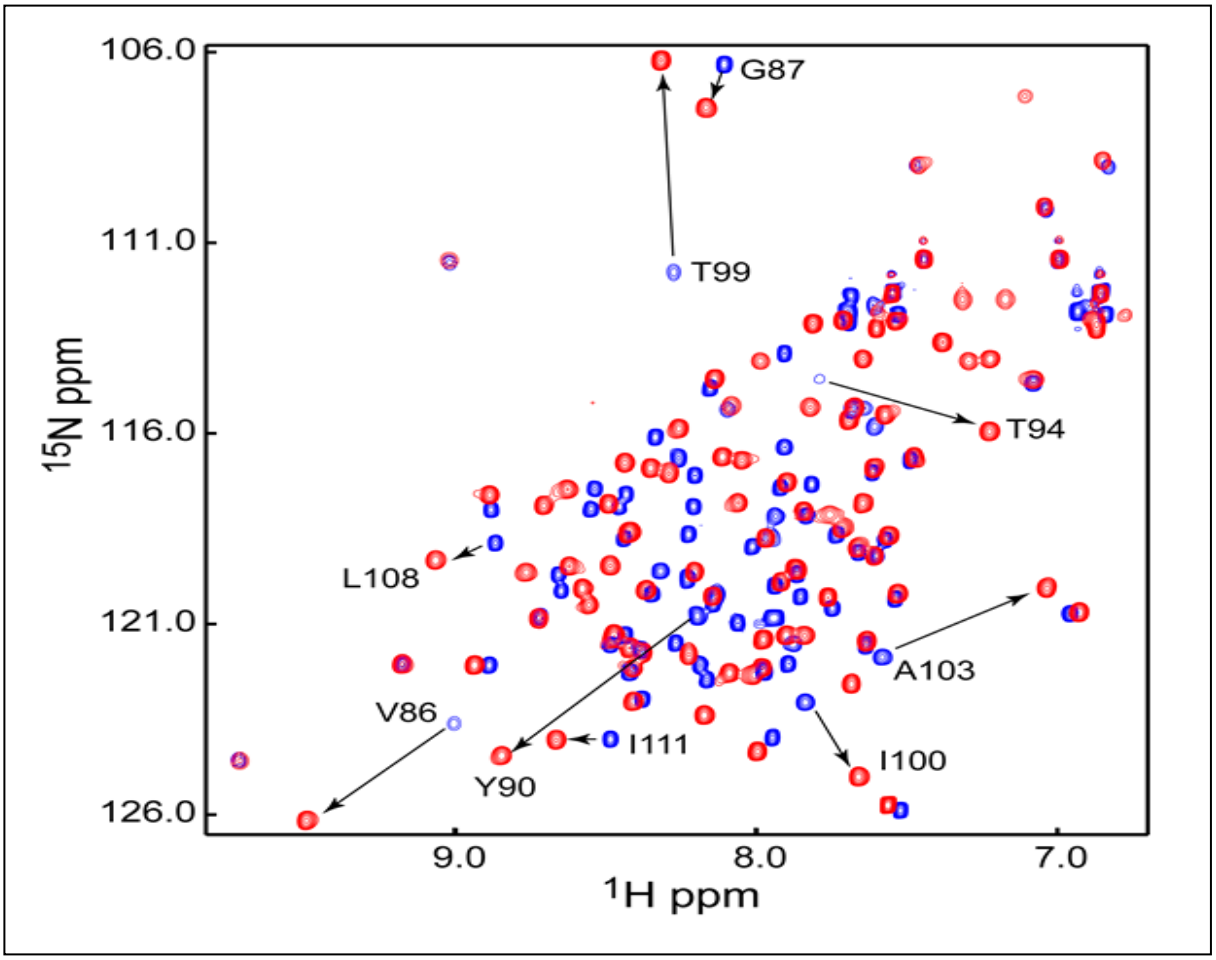


Figure 2.4A Overlaid NMR spectra of free CbpAN (blue) or CbpAN + FH CCP9 (red). The shifted peaks are indicated by arrows.

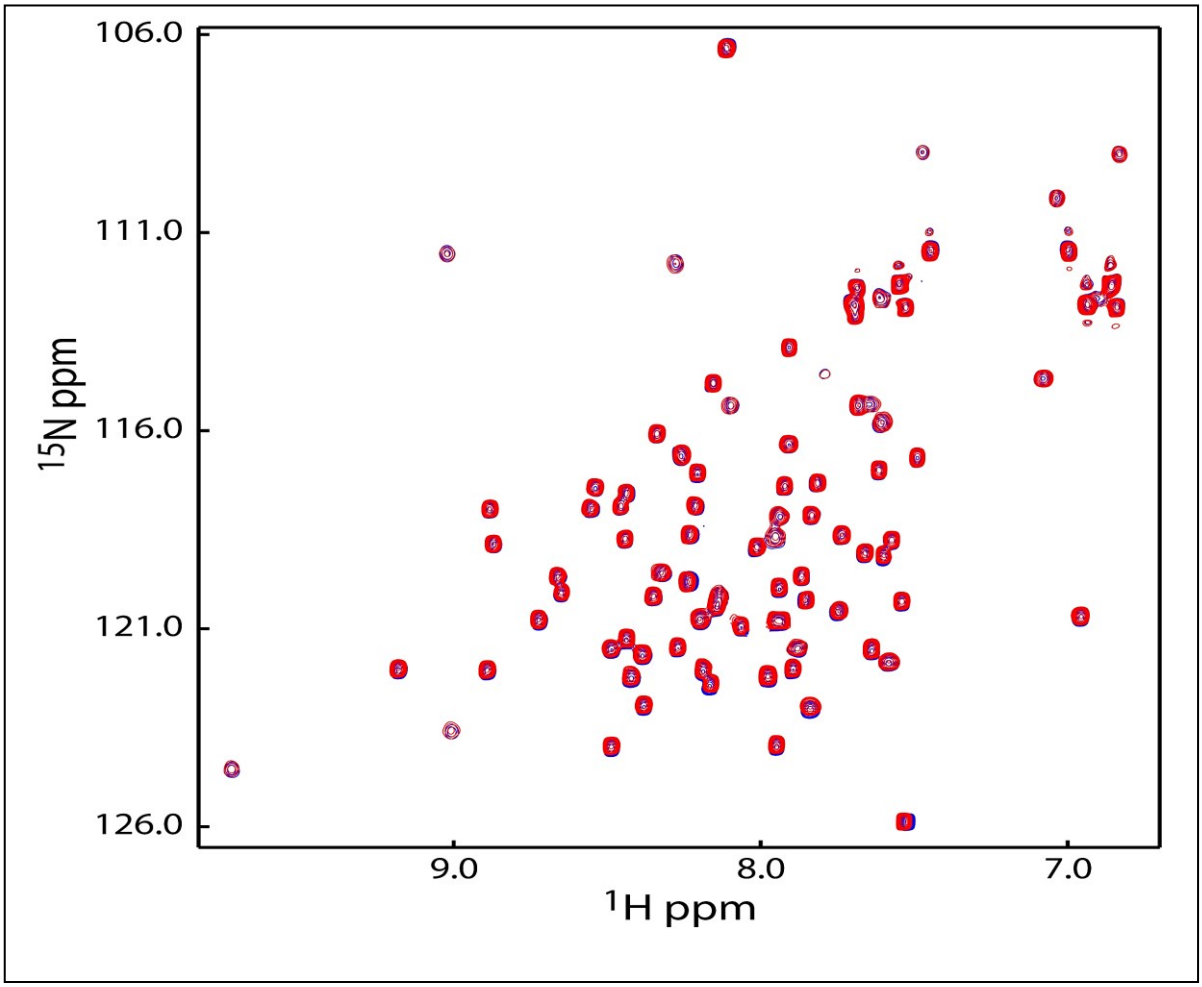


Figure 2.4B Overlaid NMR spectra of free CbpAN (blue) or CbpAN + FH CCP13-15 (red). None of the peaks shifted suggesting no binding between the two proteins.

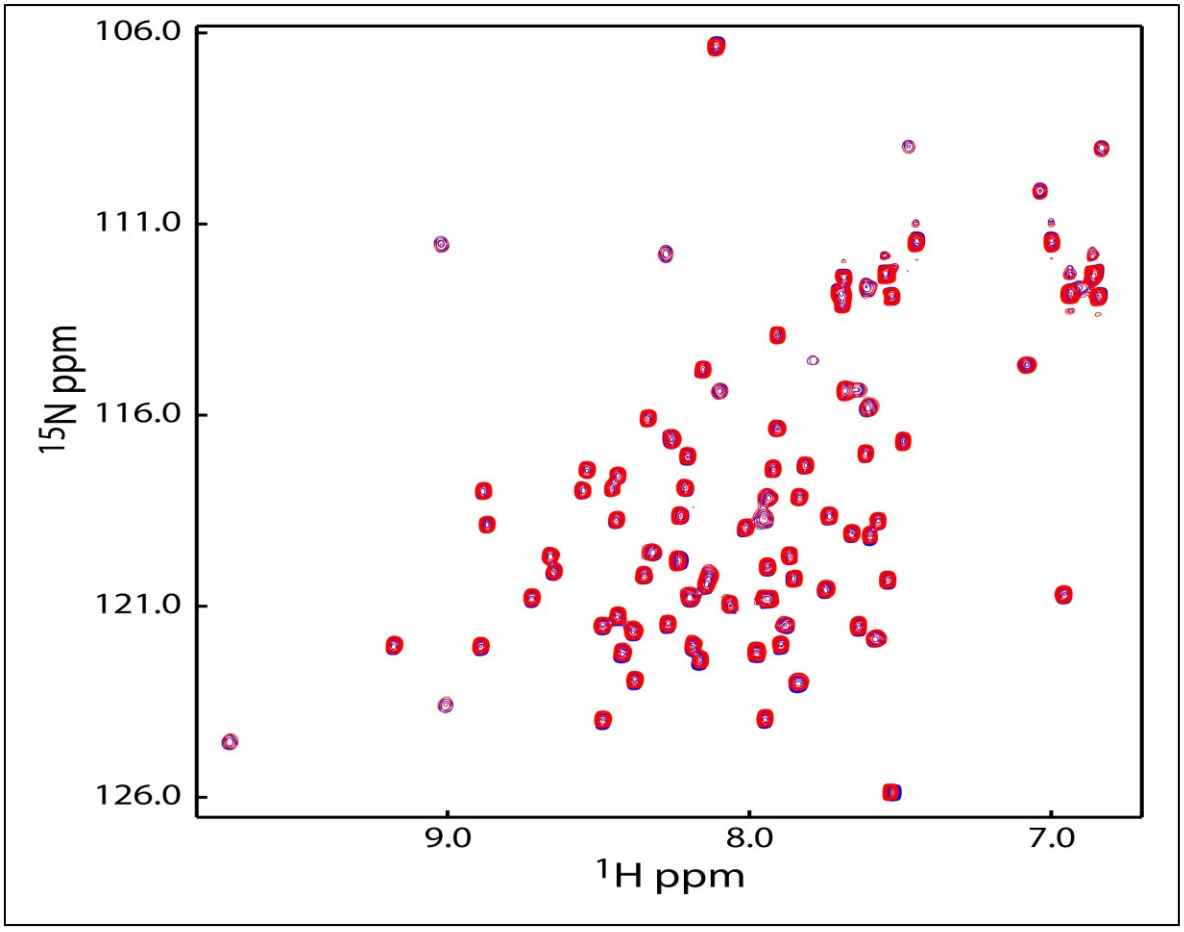


Figure 2.4C Overlaid NMR spectra of CbpAN (blue) or CbpAN + FH CCP19-20 (red). None of the peaks shifted suggesting no binding between the two proteins.

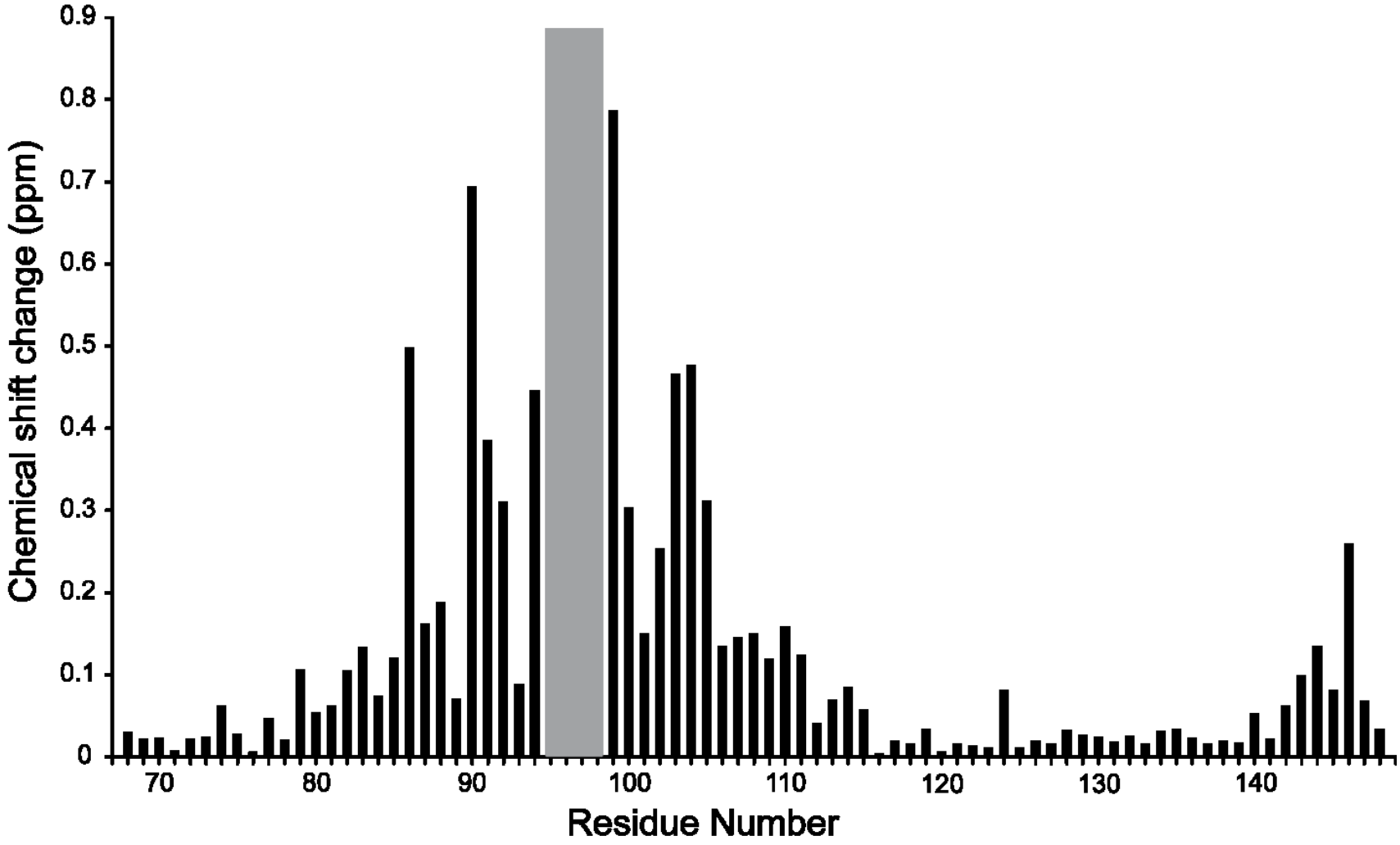


Figure 2.5 Perturbation of the chemical shifts of the backbone amides of CbpAN by the binding of hFH CCP9. The loop region with missing assignment for the free protein is indicated by a gray bar.

2.3.3 The solution NMR structure of free CbpAN and crystal structure of its complex with CCP9

To elucidate the structural basis for the binding specificity, we determined the NMR solution structure of CbpAN and the crystal structure of its complex with hFH CCP9. Over 98% of proton resonances were sequentially assigned, with only the chemical shifts of the amide protons of Lys 95, Lys 96, Arg 97 and His 98 and α-protons of Lys 138 missing. The NMR structure of free CbpAN was determined with 1,678 distance and 150 dihedral angle restraints. The solution structure is of high quality, with an average RMSD (to the mean structure) of 0.53 ± 0.10 Å for backbone atoms and 1.04 ± 0.09 Å for all heavy atoms and 97.7% of backbone torsion angles in the most favorable region and the rest (2.3%) in either the favorite or generally allowed region (see Table 2.4 for more NMR restraint and structure refinement statistics). CbpAN has a compact structure consisting of a bundle of three helices with an up-down-up antiparallel topology (Fig. 2.6a). The first helix spans residues Asp 68-Lys 92, the second helix His 98-His 121, and the third helix residues Glu 125-Glu 147. The pIgR-binding R1/R2 domains of CbpA also consist of three helices, but the packing of the helices of CbpAN is clearly different from that of the R2 domains, which form a flat, raft-like structure (Luo et al., 2005).

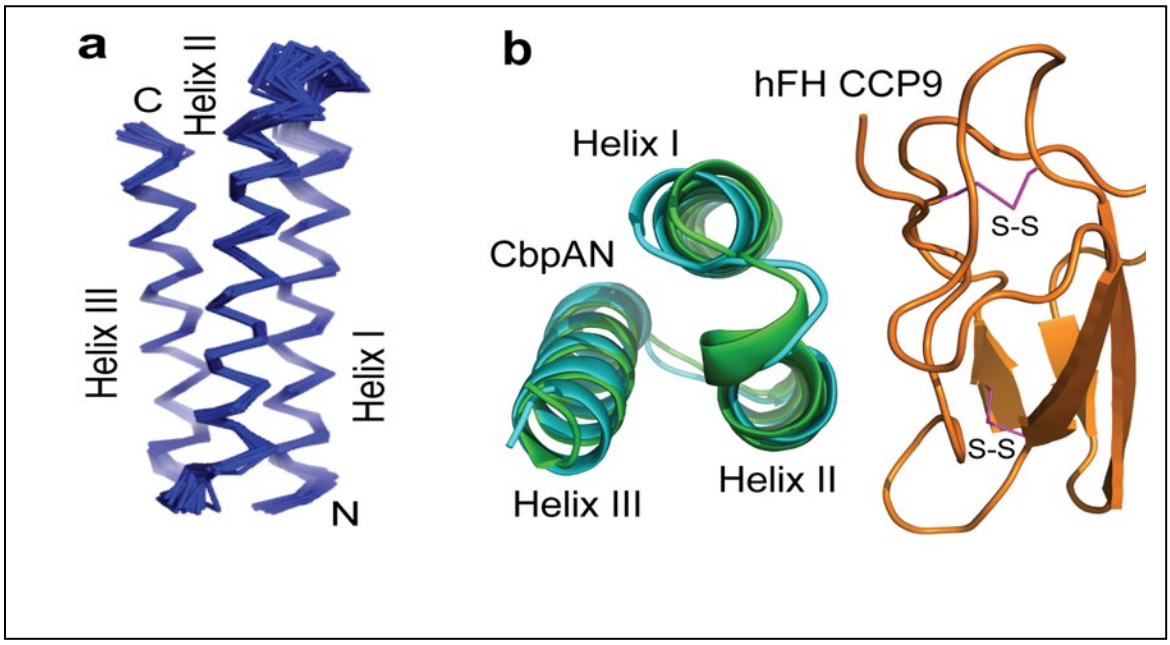


Figure 2.6 Structural analysis of CbpAN and its complex with hFH CCP9. (a) NMR structure of free CbpAN, the C^{α} traces of the top 20 conformers are drawn and (b) Crystal structure of the complex of CbpAN (green) and hFH CCP9 (orange) superposed with the NMR structure of free CbpAN (cyan). The side-chains that form the two disulfide bonds (S-S) are drawn in purple lines

The crystals of the complex of CbpAN and hFH CCP9 were obtained by co-crystallization and diffracted to 1.08-Å resolution. The crystal structure of the complex was determined by molecular replacement using the NMR structure of CbpAN as the searching model, because the structure of CCP9 has not been reported and attempts to solve the complex structure by molecular replacement using the structures of other CCP domains were not successful. The final refined structure has an R_{free} value of 17.03% and R_{work} value of 15.44%. Ramachandran analysis using MolProbity (Chen et al., 2010) revealed that 99.3 % of residues reside in geometrically favored regions (see

Table 2.5 for more crystallographic and refinement statistics). The crystal structure of bound CbpAN is very similar to the NMR structure of the free protein with a C^{α} RMSD of 0.545 Å (Fig. 2.6b), indicating that binding of CCP9 does not cause significant conformational changes in CbpAN. CCP9 has a typical CCP fold stabilized by two disulfide bonds that are formed by four invariant Cys residues (Fig. 2.6b), one between Cys 491 and Cys 535 and the other between Cys 518 and Cys 546.

Table 2.4 NMR and refinement statistics for structure determination of free CbpAN*

NMR distance and dihedral constraints	
Distance constraints	
Total NOE	1,678
Intra-residue	482
Inter-residue	1,196
Sequential ($\|i - j\| = 1$)	350
Medium-range ($\|i - j\| < 4$)	525
Long-range ($\|i - j\| > 5$)	321
Hydrogen bonds	38 (61)**
Total dihedral angle restraints	150
ϕ	75
ψ	75
Structure statistics	
Residual CYANA target function	0.029 ± 0.017
Violations from experimental restraints in the 20 CYANA structures with lowest target functions (Å)	
Number of distance constraint violations (> 0.1 Å)	9
Number of dihedral angle constraint violations (> 2.5º)	0
Number of van der Waals violations (> 0.1 Å)	13
Maximum distance constraint violations (Å)	0.44
Maximum van der Waals violations (Å)	0.30
Deviations from idealized geometry	
Bond lengths (Å)	0.005 (0.013)
Bond angles (Å)	0.6 (1.8)
Impropers (Å)	none
Average pairwise r.m.s. deviation (Å)***	
Heavy	0.99 ± 0.11 (1.04 ± 0.09)
Backbone	0.50 ± 0.10 (0.53 ± 0.10)

*The corresponding numbers after energy minimization are listed in parentheses.
**Hydrogen bonding occurs among at least 11 out the 20 structures.
***Pairwise r.m.s. deviation to the mean structure was calculated for 20 structures.

Table 2.5. Data collection and refinement statistics for crystal structure determination of the complex of hFH CCP9 and CbpAN*

	Native	NaI
Data collection		
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions		
a, *b*, *c* (Å)	41.59, 48.98, 71.38	41.56, 48.98, 70.98
α, β, γ (Å)	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution (Å)	50.00-1.08 (1.12-1.08)	50.00-2.29 (2.33 -2.29)
R_{sym} or R_{merge}	0.044 (0.359)	0.097 (0.221)
$I/\sigma I$	35.18 (5.23)	23.43 (11.90)
Completeness (%)	99.6 (99.7)	98.8 (76.2)
Redundancy	6.9 (5.9)	13.5 (11.8)
Refinement		
Resolution (Å)	19.96-1.08	
No. reflections	63060	
R_{work} / R_{free}	0.1544/0.1703	
No. atoms	1484	
Protein	1229	
Water	255	
B-factors		
Protein	12.49	
Water	25.43	
R.m.s. deviations		
Bond lengths (Å)	0.011	
Bond angles (°)	1.243	

*Values in parentheses are for highest-resolution shell.

Based on structural alignment, CCP9 is most similar to CCP4 (PDB ID: 2WII) with a C^{α} RMSD of 1.14 Å for 54 residues and CCP8 (PDB ID: 2V8E) with a C^{α} RMSD of 1.34 Å for 59 residues (Fig. 2.7a). Significant deviations can be seen only in a short deletion in CCP4 and a short insertion in CCP8 (Fig. 2.7a). The sequence identities of CCP9 to CCP4 and CCP8 both are only 30.2% with many interface residues not conserved, suggesting that the domain specificity of FH binding to CbpAN is determined by interface residues unique to CCP9.

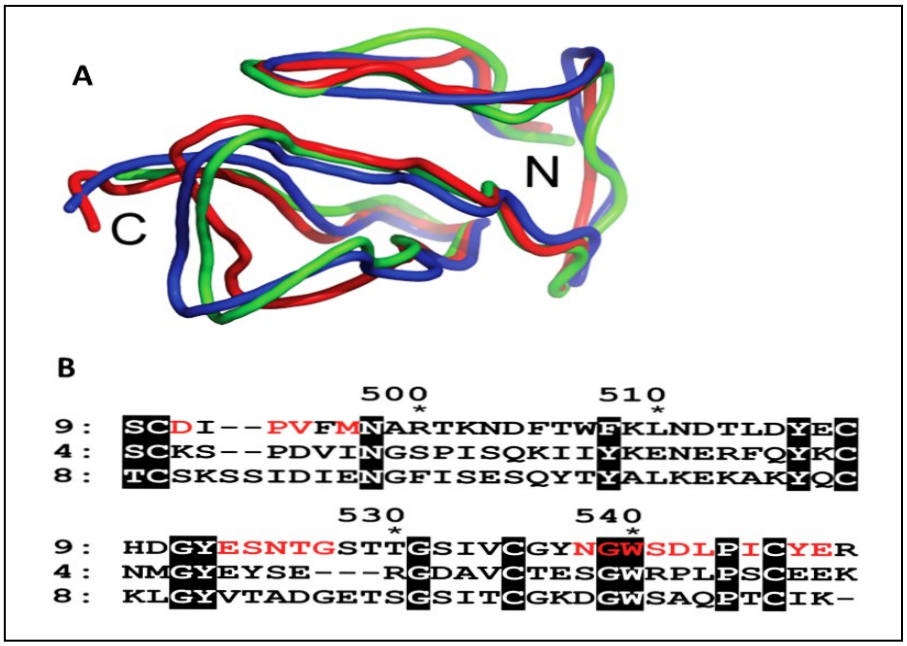


Figure 2.7 Structural and sequence alignment of hFH CCP4, CCP8 and CCP9. (A), Superposed crystal structures of hFH CCP9 (blue), CCP4 (red), and CCP8 (green) and (B), Structure-based amino acid sequence alignment of hFH CCP9, CCP4, and CCP8. The conserved residues between the three FH CCPs are shaded in black and FH CCP9 residues at the interface of binding CbpAN are colored in red. The numbering is that of FH CCP9.

The interface between CbpAN and hFH CCP9 contains 17 residues from CbpAN and 18 residues from CCP9. All interface residues of CbpAN are from a 22 residue region, residues 86-117, spanning the C-terminal part of the first helix and almost the entire second helix. This region matches almost exactly with the region of CbpAN showing the most significant chemical shift changes upon the addition of CCP9 (Fig. 2.5). The interface residues of CCP9 are dispersed in amino acid sequence, from three segments, the first segment at the N-terminus, the second in the middle, and the third at the C-terminus (Fig. 2.8). Of the 18 interface residues of CCP9, only eight residues are conserved in the sequence alignment of the CCP9 domains of 18 mammalian FH

proteins (alignment not shown. the aigned FH CCP9 are: human, bovine, sheep, rat, pig, mouse, horse, rabbit, dog, *Nomascus leucogenys* (Northern white-cheeked gibbon), *Pongo abelii* (*Sumatran orangutan*), *Macaca fascicularis* (crab-eating macaque), *Macaca mulatta* (*Rhesus macaque*), *Callithrix jacchus* (white-tufted-ear marmoset), *Loxodonta africana* (African elephant), *Heterocephalus glaber* (naked mole rat), *Otolemur garnettii* (small-eared galago) and *Ailuropoda melanoleuca* (giant panda)).

Table 2.6 Intermolecular hydrogen bonds observed in the MD simulation of the complex of hFH CCP9 and CbpAN**.**

hFH CCP9…CbpA	Occurrence (%)	Distance* (Å)	Crystal Structure (Å)
Val495-O…HO-Tyr90	96.3	2.72 ± 0.11	2.68
Glu523-OE2…NZ-Lys95	76.0	2.77 ± 0.09	2.88
Trp540-O…ND2-Asn106	63.0	2.85 ± 0.08	2.88
Asp542-OD1…NE2-His98	94.5	2.79 ± 0.09	2.71
Tyr547-OH…OG-Ser93	91.8	2.78 ± 0.10	2.63
Tyr547-OH…ND1-His98	86.1	2.82 ± 0.09	2.63

*Average distance and standard deviations.

CbpAN and CCP9 are held together by both polar and nonpolar interactions (Fig. 2.8). The polar interactions include six hydrogen bonds, of which four are to the side chains of hFH residues: one to the carboxylate of Glu 523, another to the carboxylate of Asp 542, and two to the hydroxyl of Tyr 547. Of the nonpolar interactions, the most notable are those of a cluster of four hydrophobic residues of hFH, Val 495, Met 497, Leu 543, and Ile 545, with their side-chains holding the phenol group of CbpA Tyr 90 like a lock (henceforth termed “hydrophobic lock”). The phenol group of CbpA Tyr 90 is

like a key inserted into the lock with its tip (the phenol hydroxyl) latched to CCP9 via a hydrogen bond to the backbone oxygen of hFH Val 495. Interestingly, of the four residues that constitute the hydrophobic lock, no residue is conserved among 18 mammalian proteins FH CCP9 we aligned. A complete hydrophobic lock is found only in some primate FHs, including human, chimpanzee (has identical amino acid sequence to that hFH CCP9), *Nomascus leucogenys*, and *Pongo abelii* FHs. Only one residue, Val 495, is found in FH CCP9s of mouse or rat, two animals most frequently used as model animals for studying pneumococcal infections (alignment not shown).

2.3.4 The contributions of hFH residues to binding of CbpAN and structural determinants for host specificity of FH recruitment

To investigate the relative importance of the interactions between CCP9 and CbpAN to the stability of the complex, we performed MD simulations followed by energy calculations by MM-GBSA analysis. Based on the frequency of the occurrence of hydrogen bonds in the MD trajectory (Table 2.6), of the six hydrogen bonds between CCP9 and CbpAN, the most stable hydrogen bond is between the oxygen of hFH Val 495 and the hydroxyl of CbpA Tyr 90, indicating that the lock-key interaction is strong. The MM-GBSA computational analysis suggested that, of the 18 interface residues of hFH, the most important residues for binding to CbpA are Tyr 547 and the four hydrophobic residues that constitute the hydrophobic lock (Fig. 2.8). The large energetic contribution of hFH Tyr 547 is due to its two stable hydrogen bonds to CbpA (Fig. 2.8 and Table 2.6) and many van der Waals interactions, with over 80% of its solvent-accessible area buried upon formation of the complex.

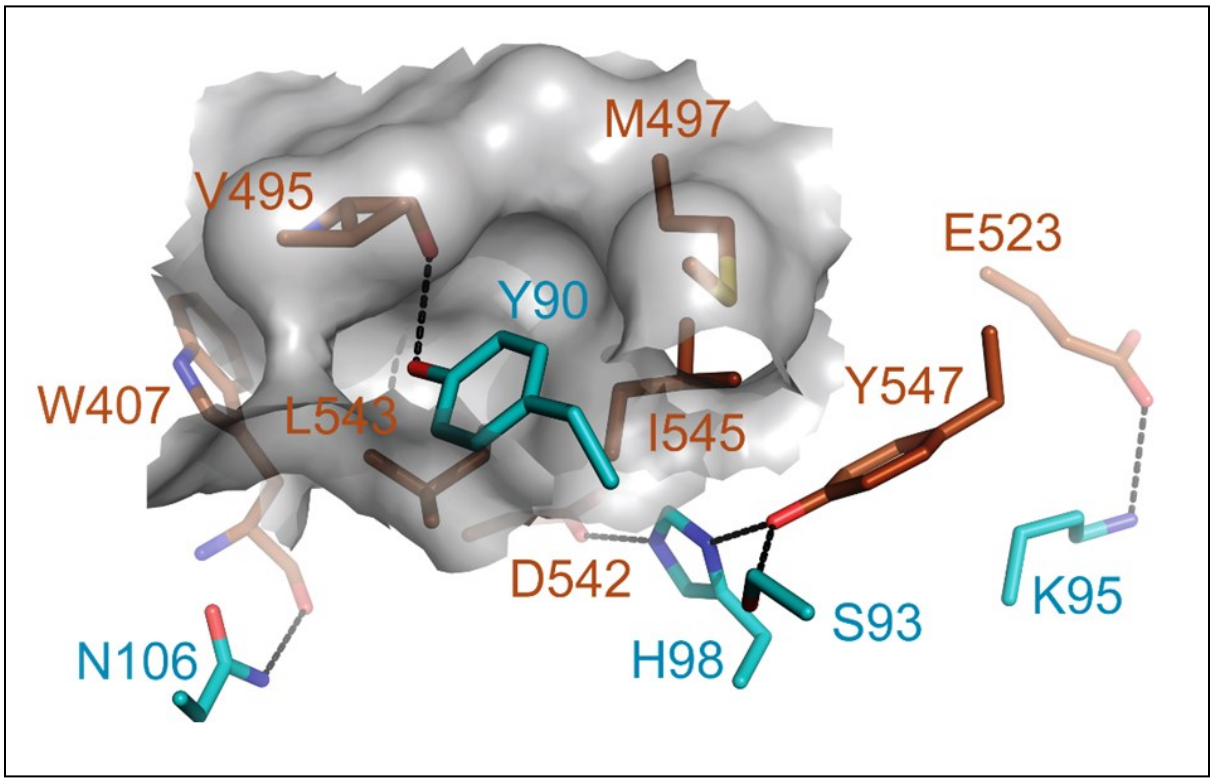


Figure 2.8 Interaction interface of FH CCP9(orange) and CbpAN(cyan). Hydrogen bonds are indicated by dash lines. The hydrophobic lock (a cluster of hydrophobic residues, Val 495, Met 497, Leu 543 and Ile 545) of hFH CCP9 is drawn in a surface representation.

Table 2.7 Binding free energies of wild type (WT) and mutant (MT) hFH and mFH CCP9 proteins calculated by MM-GBSA (kcal/mol).

System	ΔE_{vdw}	ΔE_{ele}	ΔG_{GB}	ΔG_{SA}	ΔG_{cal}
WT hCCP9	−70.1 ± 9.1	−392.2 ± 50.5	426.2 ± 53.0	−11.1 ± 1.3	−47.2 ± 7.1
WT mCCP9	−54.3 ± 21.4	−304.5± 119.1	355.0 ± 138.0	−9.5 ± 3.7	−13.4 ± 7.0
MT hCCP9	−67.0 ± 9.1	−436.2 ± 53.2	490.8 ± 45.6	−11.2 ± 1.1	−23.7 ± 8.0
MT mCCP9	−60.3 ± 14.4	−267.78±72.5	304.5 ± 78.3	−10.0 ± 2.4	−33.6 ± 9.5

The contribution of Val 495 to CbpA binding also includes that of the hydrogen bond from its backbone oxygen to the hydroxyl of CbpA Tyr 90.

To validate the computational analysis, we replaced the side chains of the three CCP9 residues involved in intermolecular hydrogen bonding, Glu 523, Asp 542, and Tyr 547, with Ala. ITC analysis of the three mutant proteins showed that of these three residues, Tyr 547 is indeed the most important residue for binding CbpAN followed by Asp 542 and Glu 523 (Table 2.3). Although Asp 542 does not appear as important for binding of CbpAN according to the computational analysis, the hydrogen bond between this residue and CbpAN is stable and consistent with the energetic contribution of Asp 542 (Table 2.6).

These conserved residues across FH CCP9 from different species include all residues involved in intermolecular hydrogen bonding. The exceptions are three of the four residues that constitute the hydrophobic lock, Met 497, Leu 543 and Ile 545, which are replaced by a negatively charged residue (Glu 497) and two polar residues (Thr 543 and Ser 545) in mFH. Both computational and mutagenesis analyses indicated that all these replacements have significant effects on binding of CbpAN. MM-GBSA analysis suggested that the three replacements together have large effects on binding of CbpAN (23.5 kcal/mol, Table 2.7), not only diminishing the favorable interactions of these replaced residues but also affecting the energetic contributions of other residues, including Val 495, Trp 540 and Tyr 547 (Fig. 2.9b). Individually, the effect of the I545S replacement on hFH CCP9 binding energy is the largest (4.14 kcal/mol), followed by that of the L543T replacement (2.68 kcal/mol) and then that of the L543T replacement (1.88 kcal/mol). In the context of mFH CCP9, the M497E replacement also introduces a large unfavorable effect on binding of CbpAN (Fig. 2.9b).

Introduction of the hydrophobic lock in mFH CCP9 by replacing Glu 497, Thr 543 and Ser 545 with the corresponding hFH residues significantly increased the binding energy (20.2 kcal/mol, Table 2.7), approaching the magnitude of the loss of binding energy when the hydrophobic lock of hFH CCP9 was destroyed by replacing the hFH residues with the corresponding mFH residues.

Experimentally, when Met 497, Leu 543 and Ile 545 of hFH CCP9 were individually replaced with the corresponding residues in mFH by mutagenesis, the three mutant proteins, M497E, L543T, and I545S, all had a significantly reduced affinity for CbpAN (Table 2.3). The K_d values of these mutant proteins increased by 40, 560 and 2,600-fold, respectively, relative to that of the wild-type CCP9, corresponding to a decrease in binding energy of 1.36, 4.66 and 5.33 kcal/mol, respectively. The extents of effects of the mutations on binding of CbpAN were very similar to those calculated by MM-GBSA analysis. When hFH CCP9 was altered with all the above three replacements, binding of CbpAN was no longer detectable by ITC.

To further validate that the hydrophobic lock is the critical structural determinant of host specificity of FH binding, we produced mFH CCP9. The binding of the wild-type mFH CCP9 to CbpAN was not detectable by ITC. When Glu 497, Thr543, and Ser 545 of mFH CCP9 were replaced with the corresponding hFH residues, the reconstructed hydrophobic lock dramatically enhanced the CbpA-binding activity of mFH CCP9, resulting in a K_d of 973 nM, though the binding activity was still significantly lower than that of hFH CCP9, which has a K_d of 4.72 nM. MM-GBSA analysis of the “humanized” mFH CCP9 suggested that Glu 525, Lys 527, and Tyr 538 might be the reasons for

reduced affinity (Fig. 2.9c): the contribution of Tyr 538 to binding is less than that of hFH Asn 538, the interactions of Lys 527 are unfavorable whereas that of hFH Gly 527 is neutral, and the combined contributions of Glu 523, Glu 525 and Lys 527 are significantly less than those of hFH Glu 523, Asn 525, and Gly 527.

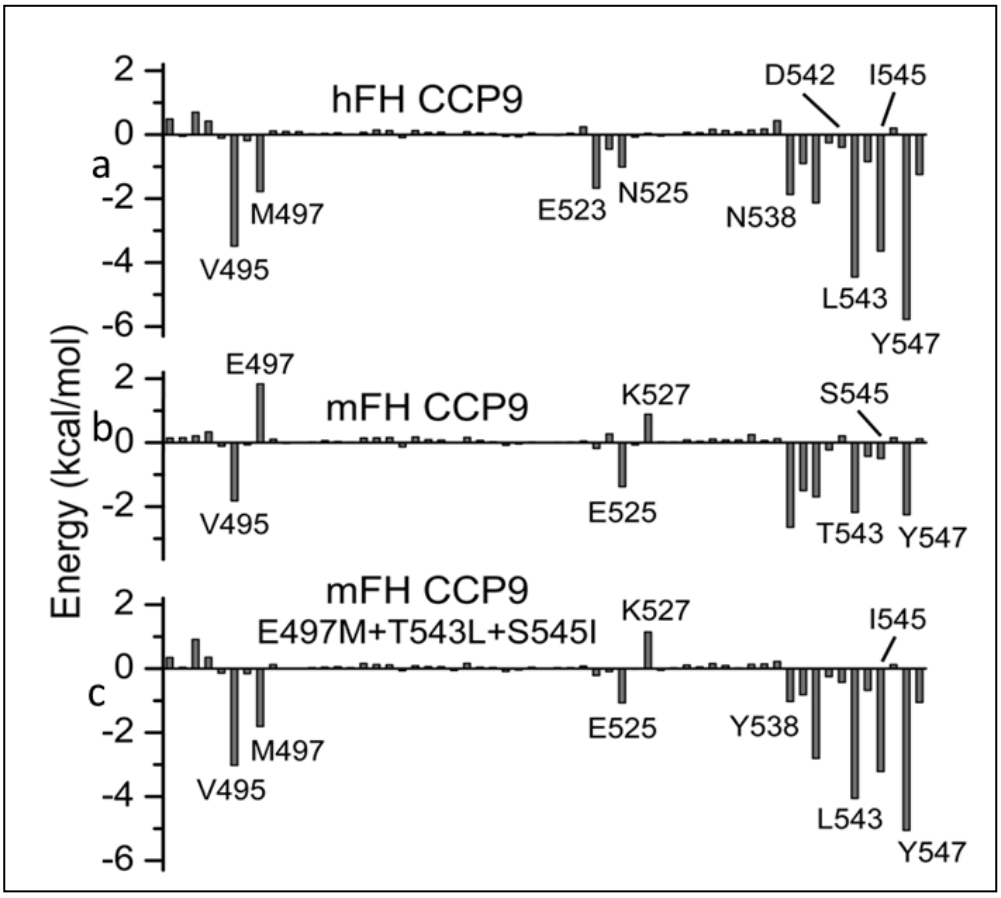


Figure 2.9 Binding energy decompositions obtained by MM-GBSA analysis for the wild-type hFH CCP9 (a), the wild-type mFH CCP9 (b) and the mFH CCP9 triple mutant protein with Glu 497 replaced by Met, Thr 543 by Leu, and Ser 545 by Ile (c).

To test these predictions, we made two additional mutants with the mFH triple mutant (E497M+T543L+S545I) as a template, a quadruple mutant with the addition of the mutation Y538N and a quintuple mutant with the addition of the double mutations E525N+K527G. The K_d value of the quadruple mutant protein was similar to that of the

triple mutant protein (Table 2.3, 1,080 vs. 973 nM), indicating that the Y538N mutation does not enhance the binding activity of mFH CCP9. In contrast, the addition of the double mutations E525N+K527G lowered the K_d value to near that of hFH CCP9 (5.78 vs. 4.72 nM), suggesting that the hydrophobic lock and these two additional residues are the key structural features recognized by CbpA.

2.4 Discussion

In this study, using truncated FH proteins, binding measurements by ITC, NMR, and structural analysis by X-ray crystallography, we has been demonstrated that a single hFH CCP domain is sufficient for tight binding of pneumococcal protein virulence factor CbpA. It has been shown that CCP6 of hFH is required for binding of a meningococcal protein virulence factor known as fHbp, but whether a single CCP6 domain of hFH is sufficient for tight binding of the meningococcal protein is not known (Schneider et al., 2009). Intermolecular interactions in the two systems are also significantly different: hydrophobic interactions are more important than polar interactions for CCP9 binding to CbpA, whereas polar interactions such as hydrogen bonds and salt bridges are mainly responsible for CCP6-7 binding to fHbp (Schneider et al., 2009). Antivirulence is a novel strategy currently pursued for development of therapeutic agents against many bacterial pathogens (Cegelski et al., 2008; Clatworthy et al., 2007). The structure-function information reported here will be useful for development of agents against the major pneumococcal virulence factor CbpA as a research tool for studying pneumococcal pathogenesis and therapeutic applications.

We have further demonstrated that CCP19-20 are essentially not involved in binding of CbpA, which is consistent with studies showing that binding of glycosaminoglycans to CCP19-20 does not affect binding of FH to pneumococci (Agarwal et al., 2010; Duthy et al., 2002). The function of CCP19-20 is to recognize glycosaminoglycans on the surface of host cells and to prevent the complement system from attacking host tissues (Ferreira et al., 2010). The ability of FH to bind CbpA on the pneumococcal cell wall and glycosaminoglycans on the surface of human epithelial cells

at the same time allows pneumococci connect to human cells via FH. Thus, FH recruitment also promotes pneumococcal adherence to and uptake of pneumococci by human cells (Agarwal et al., 2010).

Because *S. pneumonia* is a human-specific pathogen, no animal is an ideal model for studying pneumococcal pathogenesis and vaccine development. Mouse is the most frequently used animal model for these purposes, probably due to the ease of manipulation and affordability, but there are fundamental differences in interactions of *S. pneumonia* with human and mouse. One approach to overcome these differences is to "humanize" the mouse by replacement of mouse genes with human counterparts (Sacca et al., 2010), and "humanized" mice have emerged as powerful tools for studying infectious diseases (Legrand et al., 2009). Our findings regarding the host specificity of FH recruitment offer a novel approach to the development of a more biologically relevant mouse model for studying pneumococcal pathogenesis and vaccine development. Our results indicate that out of more than 1,200 amino acids, only a few point mutations are needed to humanize mFH with respect to binding of CbpA. Such an approach may be preferable to a complete replacement of the mFH gene with the hFH gene, as it is not known whether hFH can regulate the complement system and serve other functions in mouse like mFH.

Limited serotype coverage is a major drawback of current capsular polysaccharide-based pneumococcal vaccines, as *S. pneumoniae* has over 90 different serotypes (Gamez and Hammerschmidt, 2012). While the current vaccines are enormously beneficial and valuable, vaccination with these vaccines has resulted in the rise of pneumococcal strains that are not covered by the vaccines (Weinberger et al.,

2011). Consequently, there is an urgent need for development of broadly neutralizing, protein-based vaccines (Gamez and Hammerschmidt, 2012). Because of its high immunogenicity, CbpAN is considered a promising candidate antigen for vaccine development (Ricci et al., 2011). Its high sequence variations, however, limit cross-strain protection. In addition, because of high serum levels of hFH (Ferreira et al., 2010), tight FH binding may also affect the efficacy of CbpAN as a vaccine antigen, as suggested for the meningococcal FH binding protein (Johnson et al., 2012; Schneider et al., 2009). The structural information reported here can be used to redesign CbpAN to enhance its efficacy as a vaccine antigen by removing its strain-specific structural features and weakening its binding of hFH.

REFERENCES

REFERENCES

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W.*, et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213-221.

Agarwal, V., Asmat, T.M., Luo, S., Jensch, I., Zipfel, P.F., and Hammerschmidt, S. (2010). Complement regulator Factor H mediates a two-step uptake of *Streptococcus pneumoniae* by human cells. J Biol Chem *285*, 23486-23495.

Alper, C.A., Abramson, N., Johnston, R.B., Jandl, J.H., and Rosen, F.S. (1970). INCREASED SUSCEPTIBILITY TO INFECTION ASSOCIATED WITH ABNORMALITIES OF COMPLEMENT-MEDIATED FUNCTIONS AND OF THIRD COMPONENT OF COMPLEMENT (C3). N Engl J Med *282*, 349-&.

Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics *22*, 195-201.

Balachandran, P., Brooks-Walter, A., Virolainen-Julkunen, A., Hollingshead, S.K., and Briles, D.E. (2002). Role of pneumococcal surface protein C in nasopharyngeal carriage and pneumonia and its ability to elicit protection against carriage of Streptococcus pneumoniae. Infect Immun *70*, 2526-2534.

Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J., and Schwede, T. (2009). Protein structure homology modeling using SWISS-MODEL workspace. Nature Protocols *4*, 1-13.

Brock, S.C., McGraw, P.A., Wright, P.F., and Crowe, J.E. (2002). The human polymeric immunoglobulin receptor facilitates invasion of epithelial cells by *Streptococcus pneumoniae* in a strain-specific and cell type-specific manner. Infect Immun *70*, 5091-5095.

Brooks-Walter, A., Briles, D.E., and Hollingshead, S.K. (1999). The pspC gene of Streptococcus pneumoniae encodes a polymorphic protein, PspC, which elicits cross-reactive antibodies to PspA and provides immunity to pneumococcal bacteremia. Infect Immun *67*, 6533-6542.

Cao, J., Chen, D.P., Xu, W.C., Chen, T.M., Xu, S.X., Luo, J.Y., Zhao, Q., Liu, B.Z., Wang, D.S., Zhang, X.M.*, et al.* (2007). Enhanced protection against pneumococcal infection elicited by immunization with the combination of PspA, PspC, and ClpP. Vaccine *25*, 4996-5005.

Carneiro-Sampaio, M., and Coutinho, A. (2007). Immunity to microbes: Lessons from primary immunodeficiencies. Infect Immun *75*, 1545-1555.

Case, D.A., Darden, T.A., Cheatham, I., Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Crowley, M., Walker, R.C., Zhang, W.*, et al.* (2008). AMBER 10 (San Francisco: University of California).

Case, D.A., Darden, T.A., Cheatham, I., Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Walker, R.C., Zhang, W., Merz, K.M.*, et al.* (2012). AMBER 12 (San Francisco: University of California).

Cavanagh, J., Fairbrother, W.J., Palmer, A.G.I., and Skelton, N.J. (2006). Protein NMR Spectroscopy, 2 edn (New York: Academic Press).

Cegelski, L., Marshall, G.R., Eldridge, G.R., and Hultgren, S.J. (2008). The biology and future prospects of antivirulence therapies. Nature Rev Microbiol *6*, 17-27.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr *66*, 12-21.

Clatworthy, A.E., Pierson, E., and Hung, D.T. (2007). Targeting virulence: a new paradigm for antimicrobial therapy. Nat Chem Biol *3*, 541-548.

Cohen, S.X., Ben Jelloul, M., Long, F., Vagin, A., Knipscheer, P., Lebbink, J., Sixma, T.K., Lamzin, V.S., Murshudov, G.N., and Perrakis, A. (2008). ARP/wARP and molecular replacement: the next generation. Acta Crystallogr D Biol Crystallogr *64*, 49-60.

Dave, S., Carmicle, S., Hammerschmidt, S., Pangburn, M.K., and McDaniel, L.S. (2004). Dual roles of PspC, a surface protein of *Streptococcus pneumoniae*, in binding human secretory IgA and factor H. J Immunol *173*, 471-477.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR *6*, 277-293.

DeLano, W.L. (2002). The PyMOL Molecular Graphics System (Palo Alto, CA, USA: DeLano Scientific).

Duthy, T.G., Ormsby, R.J., Giannakis, E., Ogunniyi, A.D., Stroeher, U.H., Paton, J.C., and Gordon, D.L. (2002). The human complement regulator factor H binds pneumococcal surface protein PspC via short consensus repeats 13 to 15. Infect Immun *70*, 5604-5611.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.

Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., and Pedersen, L.G. (1995). A smooth particle mesh Ewald method. J Chem Phys *103*, 8577-8593.

Ferreira, V.P., Pangburn, M.K., and Cortes, C. (2010). Complement control protein factor H: The good, the bad, and the inadequate. Mol Immunol *47*, 2187-2197.

Gamez, G., and Hammerschmidt, S. (2012). Combat Pneumococcal Infections: Adhesins as Candidates for Protein-Based Vaccine Development. Curr Drug Targets *13*, 323-337.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In The Proteomics Protocols Handbook, J.M. Walker, ed. (Totowa, NJ: Humana Press), pp. 571-607.

Geourjon, C., and Deleage, G. (1995). SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci *11*, 681-684.

Gill, S.C., and Vonhippel, P.H. (1989). Calculation of protein extinction coefficients from amino-acid sequence data. Anal Biochem *182*, 319-326.

Granoff, D.M., Welsch, J.A., and Ram, S. (2009). Binding of Complement Factor H (fH) to *Neisseria meningitidis* Is Specific for Human fH and Inhibits Complement Activation by Rat and Rabbit Sera. Infect Immun *77*, 764-769.

Gross, G.N., Rehm, S.R., and Pierce, A.K. (1978). EFFECT OF COMPLEMENT DEPLETION ON LUNG CLEARANCE OF BACTERIA. J Clin Invest *62*, 373-378.

Güntert, P. (2004). Automated NMR structure calculation with CYANA. Methods Mol Biol *278*, 353–378.

Hammerschmidt, S. (2006). Adherence molecules of pathogenic pneumococci. Curr Opin Microbiol *9*, 12-20.

Hammerschmidt, S., Agarwal, V., Kunert, A., Haelbich, S., Skerka, C., and Zipfel, P.F. (2007). The host immune regulator factor H interacts via two contact sites with the PspC protein of Streptococcus pneamoniae and mediates adhesion to host epithelial cells. J Immunol *178*, 5848-5858.

Hammerschmidt, S., Tillig, M.P., Wolff, S., Vaerman, J.P., and Chhatwal, G.S. (2000). Species-specific binding of human secretory component to SpsA protein of *Streptococcus pneumoniae* via a hexapeptide motif. Mol Microbiol *36*, 726-736.

Hosea, S.W., Brown, E.J., and Frank, M.M. (1980). THE CRITICAL ROLE OF COMPLEMENT IN EXPERIMENTAL PNEUMOCOCCAL SEPSIS. J Infect Dis *142*, 903-909.

Huang, S.S., Johnson, K.M., Ray, G.T., Wroe, P., Lieu, T.A., Moore, M.R., Zell, E.R., Linder, J.A., Grijalva, C.G., Metlay, J.P.*, et al.* (2011). Healthcare utilization and cost of pneumococcal disease in the United States. Vaccine *29*, 3398-3412.

Iannelli, F., Chiavolini, D., Ricci, S., Oggioni, M.R., and Pozzi, G. (2004). Pneumococcal surface protein C contributes to sepsis caused by Streptococcus pneumoniae in mice. Infect Immun *72*, 3077-3080.

Johnson, B.A., and Blevins, R.A. (1994). NMRView: A computer program for the visualization and analysis of NMR data. J Biomol NMR *4*, 603-614.

Johnson, S., Tan, L., van der Veen, S., Caesar, J., De Jorge, E.G., Harding, R.J., Bai, X.L., Exley, R.M., Ward, P.N., Ruivo, N.*, et al.* (2012). Design and Evaluation of Meningococcal Vaccines through Structure-Based Modification of Host and Pathogen Molecules. Plos Pathogens *8*, e1002981.

Kerr, A.R., Paterson, G.K., Riboldi-Tunnicliffe, A., and Mitchell, T.J. (2005). Innate immune defense against pneumococcal pneumonia requires pulmonary complement component C3. Infect Immun *73*, 4245-4252.

Kurtenbach, K., De Michelis, S., Etti, S., Schafer, S.M., Sewell, H.S., Brade, V., and Kraiczy, P. (2002). Host association of *Borrelia burgdorferi* sensu lato - the key role of host complement. Trends Microbiol *10*, 74-79.

Legrand, N., Ploss, A., Balling, R., Becker, P.D., Borsotti, C., Brezillon, N., Debarry, J., de Jong, Y., Deng, H.K., Di Santo, J.P.*, et al.* (2009). Humanized Mice for Modeling Human Infectious Disease: Challenges, Progress, and Outlook. Cell Host & Microbe *6*, 5-9.

Lu, L., Lamm, M.E., Li, H.M., Corthesy, B., and Zhang, J.R. (2003). The human polymeric immunoglobulin receptor binds to *Streptococcus pneumoniae* via domains 3 and 4. J Biol Chem *278*, 48178-48187.

Lu, L., Ma, Y.Y., and Zhang, J.R. (2006). *Streptococcus pneumoniae* recruits complement factor H through the amino terminus of CbpA. J Biol Chem *281*, 15464-15474.

Lu, L., Ma, Z., Jokiranta, T.S., Whitney, A.R., DeLeo, F.R., and Zhang, J.R. (2008). Species-Specific Interaction of *Streptococcus pneumoniae* with Human Complement Factor H. J Immunol *181*, 7138-7146.

Luo, R.S., Mann, B., Lewis, W.S., Rowe, A., Heath, R., Stewart, M.L., Hamburger, A.E., Sivakolundu, S., Lacy, E.R., Bjorkman, P.J.*, et al.* (2005). Solution structure of choline binding protein A, the major adhesin of *Streptococcus pneumoniae*. EMBO J *24*, 34-43.

McCool, T.L., Cate, T.R., Moy, G., and Weiser, J.N. (2002). The immune response to pneumococcal proteins during experimental human carriage. J Exp Med *195*, 359-365.

McCool, T.L., Cate, T.R., Tuomanen, E.I., Adrian, P., Mitchell, T.J., and Weiser, J.N. (2003). Serum immunoglobulin G response to candidate vaccine antigens during experimental human pneumococcal colonization. Infect Immun *71*, 5724-5732.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658-674.

Miller, B.R., McGee, T.D., Swails, J.M., Homeyer, N., Gohlke, H., and Roitberg, A.E. (2012). MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. J Chem Theory Comput *8*, 3314-3321.

Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr *67*, 355-367.

Musher, D.M. (2009). Streptococcus pneumoniae. In Principles and Practice of Infectious Diseases, G.L. Mandell, J.E. Bennett, and R.D. Dolin, eds. (Philadelphia: Elsevier Churchill Livingstone), pp. 2623-2642.

Ngampasutadol, J., Ram, S., Gulati, S., Agarwal, S., Li, C., Visintin, A., Monks, B., Madico, G., and Rice, P.A. (2008). Human factor H interacts selectively with Neisseria gonorrhoeae and results in species-specific complement evasion. Journal of Immunology *180*, 3426-3435.

O'Brien, K.L., Wolfson, L.J., Watt, J.P., Henkle, E., Deloria-Knoll, M., McCall, N., Lee, E., Mulholland, K., Levine, O.S., Cherian, T.*, et al.* (2009). Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. Lancet *374*, 893-902.

Ogunniyi, A.D., Grabowicz, M., Briles, D.E., Cook, J., and Paton, J.C. (2007). Development of a vaccine against invasive pneumococcal disease based on combinations of virulence proteins of *Streptococcus pneumoniae*. Infect Immun *75*, 350-357.

Ogunniyi, A.D., Woodrow, M.C., Poolman, J.T., and Paton, J.C. (2001). Protection against Streptococcus pneumoniae elicited by immunization with pneumolysin and CbpA. Infect Immun *69*, 5997-6003.

Ricci, S., Janulczyk, R., Gerlini, A., Braione, V., Colomba, L., Iannelli, F., Chiavolini, D., Oggioni, M.R., Bjorck, L., and Pozzi, G. (2011). The factor H-binding fragment of PspC as a vaccine antigen for the induction of protective humoral immunity against experimental pneumococcal sepsis. Vaccine *29*, 8241-8249.

Ricklin, D., Hajishengallis, G., Yang, K., and Lambris, J.D. (2010). Complement: a key system for immune surveillance and homeostasis. Nat Immunol *11*, 785-797.

Rosenow, C., Ryan, P., Weiser, J.N., Johnson, S., Fontan, P., Ortqvist, A., and Masure, H.R. (1997). Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of Streptococcus pneumoniae. Mol Microbiol *25*, 819-829.

Ryckaert, J.P., Ciccotti, G., and Berendsen, H.J.C. (1977). Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. J Comput Phys *23*, 327-341.

Sacca, R., Engle, S.J., Qin, W.N., Stock, J.L., and McNeish, J.D. (2010). Genetically Engineered Mouse Models in Drug Discovery Research. In Mouse Models for Drug Discovery: Methods and Protocols, G. Proetzel, and M.V. Wiles, eds., pp. 37-54.

Sampson, H.A., Walchner, A.M., and Baker, P.J. (1982). RECURRENT PYOGENIC INFECTIONS IN INDIVIDUALS WITH ABSENCE OF THE 2ND COMPONENT OF COMPLEMENT. J Clin Immunol *2*, 39-45.

Schneider, M.C., Prosser, B.E., Caesar, J.J., Kugelberg, E., Li, S., Zhang, Q., Quoraishi, S., Lovett, J.E., Deane, J.E., Sim, R.B.*, et al.* (2009). Neisseria meningitidis recruits factor H using protein mimicry of host carbohydrates. Nature *458*, 890-893.

Schwede, T., Kopp, J., Guex, N., and Peitsch, M.C. (2003). SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res *31*, 3381-3385.

Serruto, D., Rappuoli, R., Scarselli, M., Gros, P., and van Strijp, J.A.G. (2010). Molecular mechanisms of complement evasion: learning from staphylococci and meningococci. Nature Rev Microbiol *8*, 393-399.

Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS+ : a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR *44*, 213-223.

Stevenson, B. (2002). Differential Binding of Host Complement Inhibitor Factor H by *Borrelia burgdorferi* Erp Surface Proteins: a Possible Mechanism Underlying the Expansive Host Range of Lyme Disease Spirochetes. Infect Immun *70*, 491-497.

Sun, P., Tropea, J.E., and Waugh, D.S. (2011). Enhancing the Solubility of Recombinant Proteins in *Escherichia coli* by Using Hexahistidine-Tagged Maltose-Binding Protein as a Fusion Partner. Methods Mol Biol *705*, 259-274.

Velazquez-Campoy, A., and Freire, E. (2006). Isothermal titration calorimetry to determine association constants for high-affinity ligands. Nat Protocols *1*, 186-191.

Wang, C.C. (2002). Protein disulfide isomerase as an enzyme and a chaperone in protein folding. Methods Enzymol *348*, 66-75.

Weinberger, D.M., Malley, R., and Lipsitch, M. (2011). Serotype replacement in disease after pneumococcal vaccination. Lancet *378*, 1962-1973.

Zhang, J.R., Mostov, K.E., Lamm, M.E., Nanno, M., Shimida, S., Ohwaki, M., and Tuomanen, E. (2000). The polymeric immunoglobulin receptor translocates pneumococci across human nasopharyngeal epithelial cells. Cell *102*, 827-837.

Zipfel, P.F., and Skerka, C. (2009). Complement regulators and inhibitory proteins. Nat Rev Immunol *9*, 729-740.

CHAPTER 3

Determination of Structural Requirements of CbpA for Binding to Factor H

Abstract

CbpA is a multifunctional pneumococcal surface protein required for adherence, colonization and immune evasion. It interacts in a species-specific manner with human factor H (FH) and polymeric Ig receptor (pIgR). All isolated virulent strains of pneumococcus bind FH. We recently located the regions of FH and CbpA that are required for their interaction. Multiple sequence analysis revealed that the 82-residue N-terminal domain of CbpA (CbpAN) that is required for its interaction with FH is hypervariable. Based on insights from high resolution structures of apo- and FH-bound CbpAN (Chapter 2) and CbpAN sequence analysis, we investigated how specific CbpAN residues contribute to FH binding. Several conserved interface and non-interface residue positions that are likely crucial for FH recruitment function were identified. Our results suggest that pneumococcus likely uses a common FH binding strategy and antigenic variance in CbpAN to challenge host immune response.

3.1 Introduction

The complement system is a crucial part of host immunity against *S. pneumoniae,* a leading cause of human diseases such as otitis media, pneumonia, meningitis, and septicemia (Musher, 2009; Walport, 2001). Activation of the complement system leads to deposition of an opsonin, C3b, on the pneumococcal surface and eventual removal by neutrophils (Yuste et al., 2008). Survival of *S. pneumoniae* therefore depends on its ability to inhibit the complement system and one way it does that is by recruiting a negative regulator of the complement system, FH, to its surface in order to evade complement-mediated phagocytosis (Rosenow et al., 1997). Recruitment of FH also helps the pneumococcus to adhere and colonize the upper airway, which is a crucial step in its pathogenesis. All virulent pneumococcal strains bind FH but with varied affinity (Quin et al., 2006). Carriage strains tend to recruit more FH than systemic or mucosal isolates (Quin et al., 2006). The recruitment of human FH is accomplished via protein-protein interaction between the N-terminal domain of CbpA and FH (Lu et al., 2006).

CbpA is immunogenic and a good candidate for development of universal protein-based vaccines against pneumococcus (Brooks-Walter et al., 1999; Ogunniyi et al., 2007). Understanding the key features of the CbpA N-terminal domain responsible for FH recruitment may facilitate design of novel therapeutics or protein-based vaccines against *S. pneumoniae*. The amino acid sequence of CbpA is highly variable (Iannelli et al., 2002) and as a result, CbpA proteins have been classified into 6 subtypes. The C-terminal choline-binding and the middle proline-rich domains are fairly conserved, whereas the N-terminal α-helical domain exhibits high variation in amino acid sequence

(Iannelli et al., 2002). Since FH binding activity resides in this highly variable N-terminal domain of CbpA, this study sought to define the structural features required for their interaction.

Previous studies looking at the effects of variability of the *cbpA* gene locus on function either included the whole protein sequence or the entire N-terminal α-helical region comprising the FH binding domain as well as the R1 and R2 domains associated with binding to polymeric Ig receptor, pIgR, (Iannelli et al., 2002) and these studies may not have clearly defined the key features of CbpA required for binding FH. In chapter 2, it was demonstrated that only 82 N-terminal residues of CbpA are sufficient for interaction with FH (Chapter 2). Here, based on insights from high resolution structures of apo- and FH-bound CbpAN (Chapter 2), the contributions of specific CbpAN residues to FH binding were investigated. Several conserved interface residue and non-interface residue positions that are likely crucial for FH recruitment function were identified.

3.2 Materials and Methods

3.2.1 Purification of CbpA-N (TIGR4) and its variant proteins

To construct overexpression systems for CbpAN domains of various pneumococcal strains, DNA fragments encoding them were amplified by PCR from the genomic DNA preparations of respective *S. pneumoniae* strains using primers listed in Table 3.1, which incorporated flanking BamHI and NdeI restriction sites. The amplified DNA fragments were cloned into the BamHI and NdeI restriction sites of the vector pET17bHR. For the poorly soluble CbpAN variants, the DNA fragment encoding them were subcloned into the vector pET17bHMR, which produces an N-terminally His_6-tagged maltose-binding protein (MBP) fusion protein and that can be cleaved by TEV protease. Both pET17bHR and pET17bHMR were derived from the pET17b (Novagen) plasmid, and the only difference is that the latter has the MBP-coding DNA between the His_6-tag and the TEV protease cleavage site. Correct coding sequences of the cloned DNA fragments were confirmed by DNA sequencing at the Michigan State University Genomic Core Facility.

The His_6-tagged CbpAN was over-produced in the *E. coli* strain BL21(DE3) pLysS. The cells were induced with 0.5 mM IPTG when the culture reach an OD_{600} of 0.7 and incubated at 16 $^{\circ}$C overnight. After harvesting, the cells were resuspended in a buffer containing 50 mM sodium phosphate and 300 mM NaCl, pH 8.0, and lysed by sonication. The lysate was clarified by centrifugation at 16,000 g for 20 min before loading onto a Ni-NTA column pre-equilibrated with the same buffer. After washing with 10 mM imidazole in the phosphate buffer, the column was eluted using a 10-250 mM

imidazole gradient. The protein-containing fractions were analyzed by electrophoresis on a 15 % SDS-PA gel. Fractions containing pure CbpAN were pooled and concentrated using an Amicon stirred cell. When desired, the His-tag was cleaved by incubating the protein overnight with TEV protease (Kapust et al., 2001) followed by a second passage through Ni-NTA column, where the flow-through was collected, and further purification by gel filtration chromatography on a Sephadex G-75 column. The protein-containing fractions were concentrated, dialyzed against 5 mM sodium phosphate (pH 8.0) and then against distilled water for 12 h, and lyophilized. MBP-fused CbpA variants were purified by the same procedure for just His-tagged proteins.

CbpAN (TIGR4) mutants were generated by a PCR-based method with pET17bHRCbpAN (TIGR4) as the template, following a QuickChange® site-directed mutagenesis protocol by Stratagene. The primers used to generate the mutants are listed in Table 3.2. The presence of desired mutations and the lack of any undesired mutation were confirmed by DNA sequencing. All CbpAN mutant proteins were purified following same protocol as used for the wild-type CbpAN.

3.2.2 CbpAN sequence and Consurf analysis

Multiple sequence alignment analysis of CbpANs from different *S. pneumoniae* strains was done using the MUSCLE 3.7 (Edgar, 2004) tool in the Phylogeny fr webserver. Nineteen unique CbpA sequences were derived from the NCBI Gene Bank and annotated in a conventional way. A bioinformatics tool, Consurf, was also used to project the sequence conservation of CbpAN to its three-dimensional structure. The information derived from the Consurf analysis may identify patches of the protein

surface consisting of highly conserved residues and such patches may be of functional significance (Glaser et al., 2003; Landau et al., 2005). The Consurf analysis was performed by loading the coordinate of the NMR structure of CbpAN (Chapter 2) and the aligned 19 CbpAN amino acidsequences onto the Consurf webserver (http://consurf.tau.ac.il/index_proteins.php) (Glaser et al., 2003; Landau et al., 2005) and the default parameter set. The conservation of the CbpA surface derived from the Consurf analysis was drawn with the program PyMOL (http://www.pymol.org/).

3.2.3 Isothermal Titration Calorimetry

FH CCP9 was purified as described in Chapter 2. Protein concentrations were estimated by UV light absorbance at 280 nm based extinction coefficient values obtained using a service (ProParam) of the Expasy web server (http://web.expasy.org/protparam/). ITC measurements and data analysis were performed as described in Chpater 2.

Table 3.1 Primers used to clone CbpAN variants.

Primer	Nucleotide sequence
TIGR4f	5’- G GAA TTC CAT ATG GAT TCA GAA CGA GAT AAG GCA AGG-3’
TIGR4r	5’- CG GGA TCC TCA CTT TTC AAA CTT AGA CAC AGC TTC-3’
CbpAND39f	5’- G GAA TTC CAT ATG GCA AAG ACA GAA CAT AGG AAA GC-3’
CbpAND39r	5’- CG GGA TCC TCA TTT TTT AAA CTT CTC AAA AGC TGC GTC-3’
CbpANST858f	5’- G GAA TTC CAT ATG AAA GCT GCT AAA CAA GTC GAT G-3’
CbpANST858r	5’- CGG GAT CCA TTA TTT TTT AAA CTG CTC AAA AGC TGC GTC TAA C-3’
CbpANST860f	5’- G GAA TTC CAT ATG GAA GAT ATA AAA AAA ATG TTG AGT GAG ATC C-3’
CbpANST860r	5’- CGG GAT CCA TTA TTT TTT AAA CTT CTC AAA AGC TGC GTC TAA C-3’.
CbpANST863f	5’- G GAA TTC CAT ATG CAA GAT ATA TCG AAG AAG TAT GCT GAT G-3’.
CbpANST863r	5’- CGG GAT CCA TTA TTT TTT AAA CTG CTC AAA AGC TGC GTC TAA C-3’
CbpANST865f	5’- G GAA TTC CAT ATG ATA GAA CGA GAT AAG GCA AAG ACA GCG-3’
CbpANST865r	5’- CGG GAT CCA TTA CTT TTT ATA TTT AGA CAC AGC TTC ATC TAG-3’
CbpANST866f	5’- G GAA TTC CAT ATG GAA CAT AGG AAA GCT GCT GAA CAA TTC G-3’
CbpANST866r	5’- CGG GAT CCA TTA TTT TTT AAA CTG CTC AAA AAC TGC GGT TAA C-3’
CbpANST869f	5’- G GAA TTC CAT ATG AAA GCT GCT GAA CAA GTC GAT G-3’
CbpANST869r	5’- CGG GAT CCA TTA TTT GTT AAA CTG CTC AAA AGC TGC GTC TAT C-3’
CbpANST872f	5’- G GAA TTC CAT ATG CAA GAT ATA TCG AAG AAG TAT GCT GAT G-3’
CbpANST872r	5’- CGG GAT CCA TTA TTT TTT AAA CTG CTC AAA AGC TGC GGT TAA C-3’

Table 3.2 Primers used to generate CbpAN (TIGR4) mutants.

Primer	Nucleotide sequence
K95Af	5'-GAG AGC TAT GCA AAA TCA ACT GCA AAG CGA CAT ACA ATT ACT GTA GC-3'
K95Ar	5'-GCT ACA GTA ATT GTA TGT CGC TTT GCA GTT GAT TTT GCA TAG CTC TC -3'
Y115Af	5'-G TTG AAC AAC ATT AAG AAC GAG GCT TTG AAT AAA ATA GTT GAA TCA AC-3'
Y115Ar	5'-GTT GAT TCA ACT ATT TTA TTC AAA GCC TCG TTC TTA ATG TTG TTC AAC-3'
R136Af	5'-CAG ATA CTG ATG ATG GAG AGT GCA TCA AAA GTA GAT GAA GCT GTG -3'
R136Ar	5'-CAC AGC TTC ATC TAC TTT TGA TGC ACT CTC CAT CAT CAG TAT CTG -3'
Y90Af	5'-GTG GGT GAG AGC GCT GCA AAA TCA ACT AAA AAG CG -3'
Y90Ar	5'- CGC TTT TTA GTT GAT TTT GCA GCG CTC TCA CCC AC -3'
N106Af	5'-CT GTA GCT CTA GTT GCC GAG TTG AAC AAC ATT AAG -3'
N106Ar	5'- CTT AAT GTT GTT CAA CTC GGC AAC TAG AGC TAC AG-3'
H98Af	5'-GCA AAA TCA ACT AAA AAG CGA GCT ACA ATT ACT GTA GCT C -3'
H98Ar	5'-GAG CTA CAG TAA TTG TAG CTC GCT TTT TAG TTG ATT TTG C -3'
V102Af	5'-CGA CAT ACA ATT ACT GCA GCT CTA GTT AAC GAG -3'
V102Ar	5'-CTC GTT AAC TAG AGC TGC AGT AAT TGT ATG TCG -3'
V105Af	5'- CAT ACA ATT ACT GTA GCT CTA GCT AAC GAG TTG AAC AAC ATT AAG-3'
V105Ar	5'- CTT AAT GTT GTT CAA CTC GTT AGC TAG AGC TAC AGT AAT TGT ATG-3'
K112Af	5'- CTA GTT AAC GAG TTG AAC AAC ATT GCG AAC GAG TAT TTG AAT AAA ATA G-3'
K112Ar	5'- CTA TTT TAT TCA AAT ACT CGT TCT TAA TGT TGT TCA ACT CGT TAA CTA G -3'

3.3 Results

3.3.1 Approaches

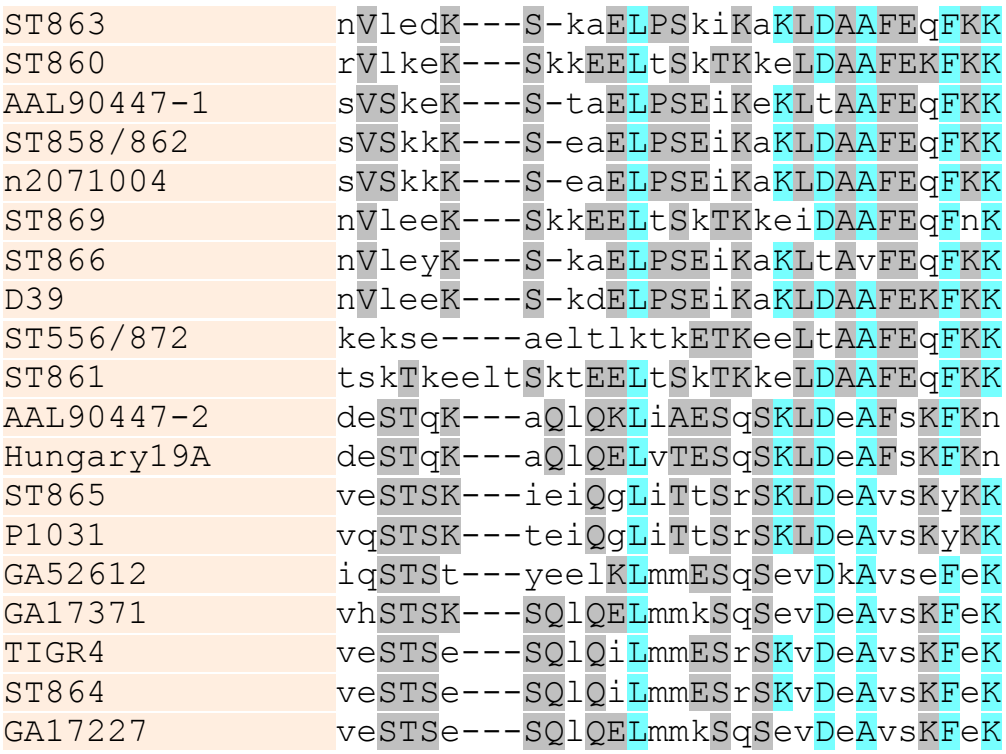
To map the interaction interface of CbpAN, two approaches were used: (i) site-directed mutagenesis and (ii) using natural variations found in different pneumococcal strains. Site-directed mutants of CbpAN (TIGR4) were designed initially based on amino acid sequence conservation and 2D-NMR titration data and later on based on the crystal structure of the complex of CbpAN and FH CCP9. With respect to CbpA variants, TIGR4 and D39 are considered reference strains and all others were selected from pneumococcal strains isolated from patients with otitis media. The interaction between FH9 and various CbpAN proteins was measured by ITC (Fig. 3.6).

3.3.2 Multiple sequence alignment

Nineteen CbpAN amino acid sequences were retrieved and aligned using MUSCLE 3.7 webserver. The multiple sequence alignment (MSA) revealed a large sequence variation, with some sharing only ~28 % sequence identity. The highly conserved residues in or near loop 1 reinforce the importance of this region for binding FH and suggest that in spite of the high sequence variation, FH binding likely takes place through a common strategy. The conserved residues around loop 1 were targeted for mutagenic analysis to test their importance to FH binding (Table 3.4).

Figure 3.1 Multiple sequence alignment of CbpAN sequences. Residues colored cyan are conserved, grey are semi conserved while white is least conserved.

Figure 3.1 cont'd



3.3.3 Consurf analysis

In an attempt to identify sequence and structural features that may be important for FH binding, Consurf analysis was performed on CbpAN variants from 19 pneumococcal strains. The analysis results are presented in grades 1 through 9 with 1 being for the least conserved residue position while 9 is the most conserved (Figure 3.2). Grade 5 is considered intermediate. The results showed that there are 50/79 residues with > 50% (intermediate) evolutionary conservation and 13/79 positions with >89% (high) conservation. The sequence conservation is projected to the three-dimensional structure of the protein. As shown in Fig. 3.3, the most conserved region of CbpAN is around loop 1, suggesting that these residues serve either structural or functional purposes. Indeed, the crystal structure of the complex of CbpAN and FH CCP9 reveals that loop 1 is part of the interface in binding FH (Chapter 2). The conservation of CbpAN residues are shown in Figs. 3.2 and 3.3. The Consurf analysis shows clearly that even though CbpAN sequence is not generally conserved, there are core conserved residue positions that may support the important function of recruiting FH to the surface of pneumococcus.

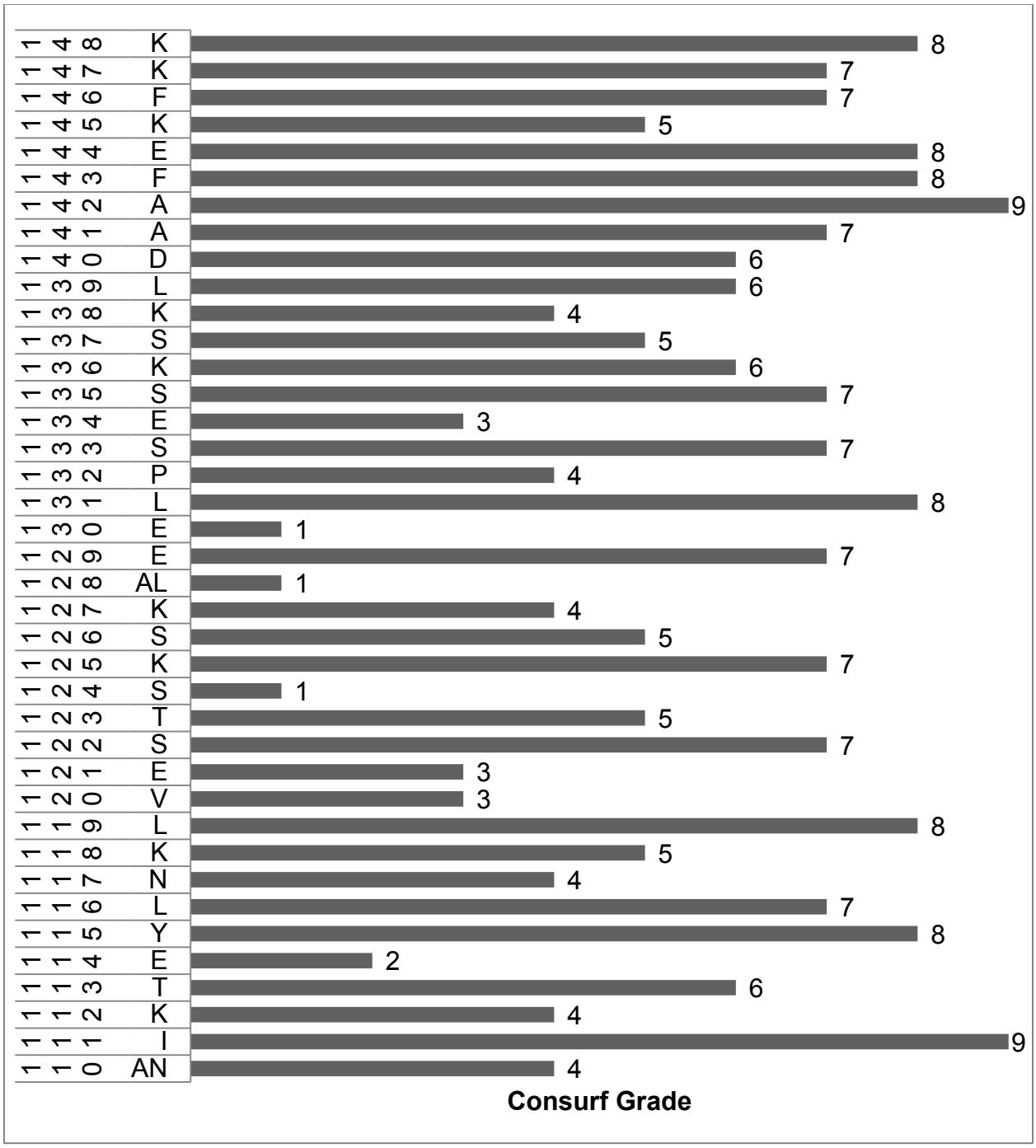
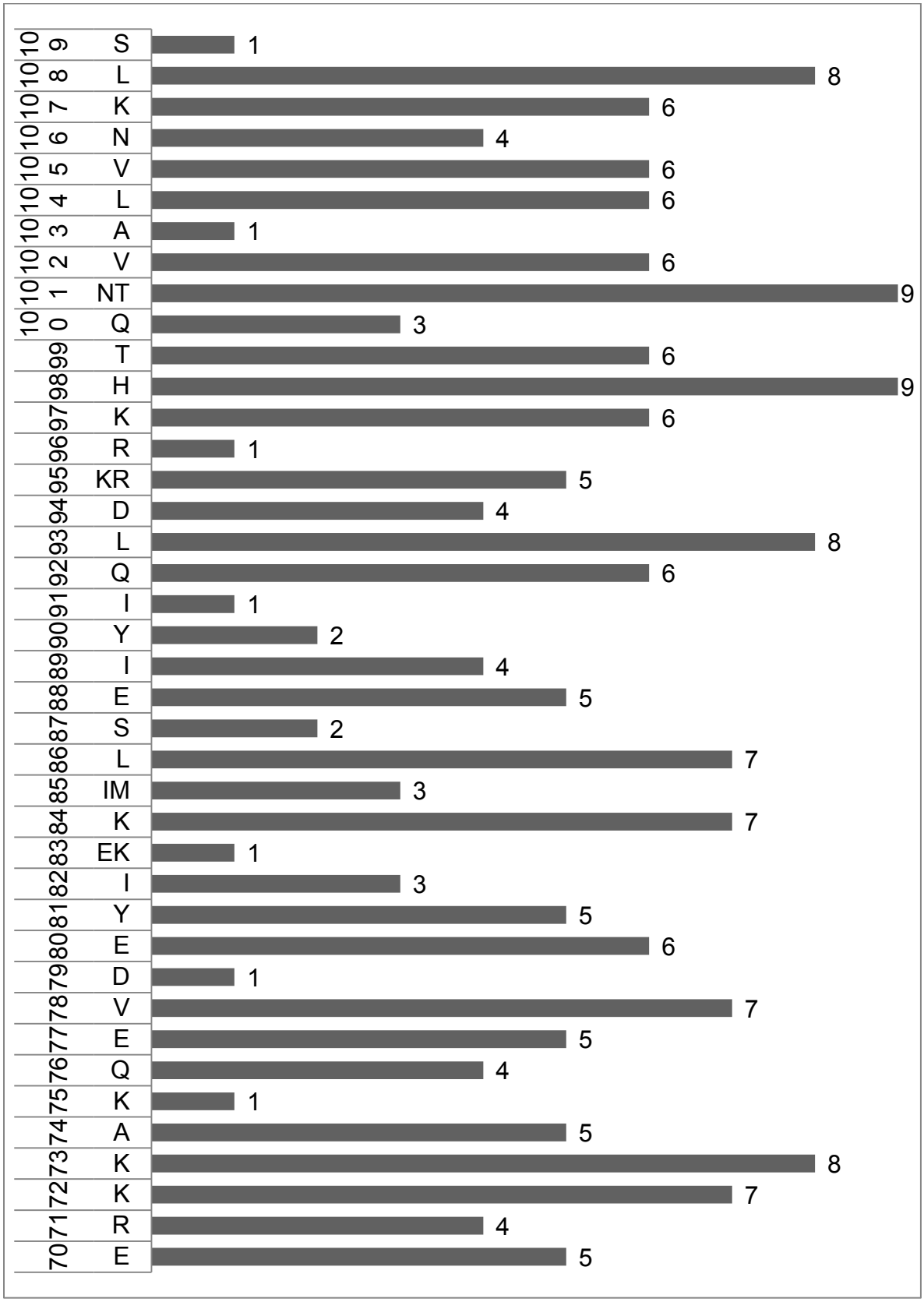


Figure 3.2 Degree of CbpAN sequence conservation from Consurf analysis. CbpAN residues and their numbers are indicated in the Y axis and for some positions, most frequent two similar residues are indicated.

Figure 3.2 (cont'd)



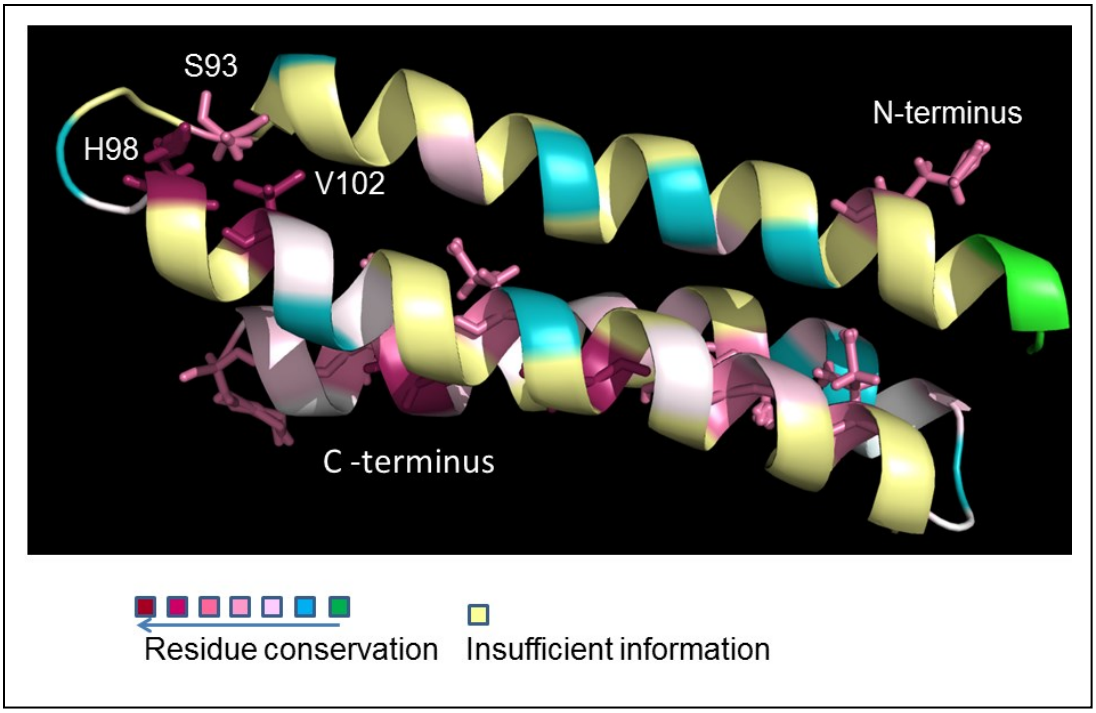


Figure 3.3 Pymol structure showing conservation scores generated by Consurf analysis and rendered on the structure by Pymol software. The conserved residues (S93, H98 and V102) around loop 1 of CbpAN are indicated.

Phylogenetic trees generated from Consurf and Phylogenic fr (Fig. 3.4) webserver (http://www.phylogeny.fr/)(Anisimova and Gascuel, 2006; Castresana, 2000; Chevenet et al., 2006; Guindon and Gascuel, 2003) analysis of CbpAN sequences were essentially the same. CbpAN sequences analyzed are broadly clustered in 5 clades and representative members of 4 of the 5 clades were successfully isolated and tested for their FH9 binding properties (Table 3.4). However, such phylogenetic trees do not necessarily reflect an evolutionary relationship, because sequence divergence may be due to horizontal gene transfer between pneumococci (Iannelli et al., 2002).

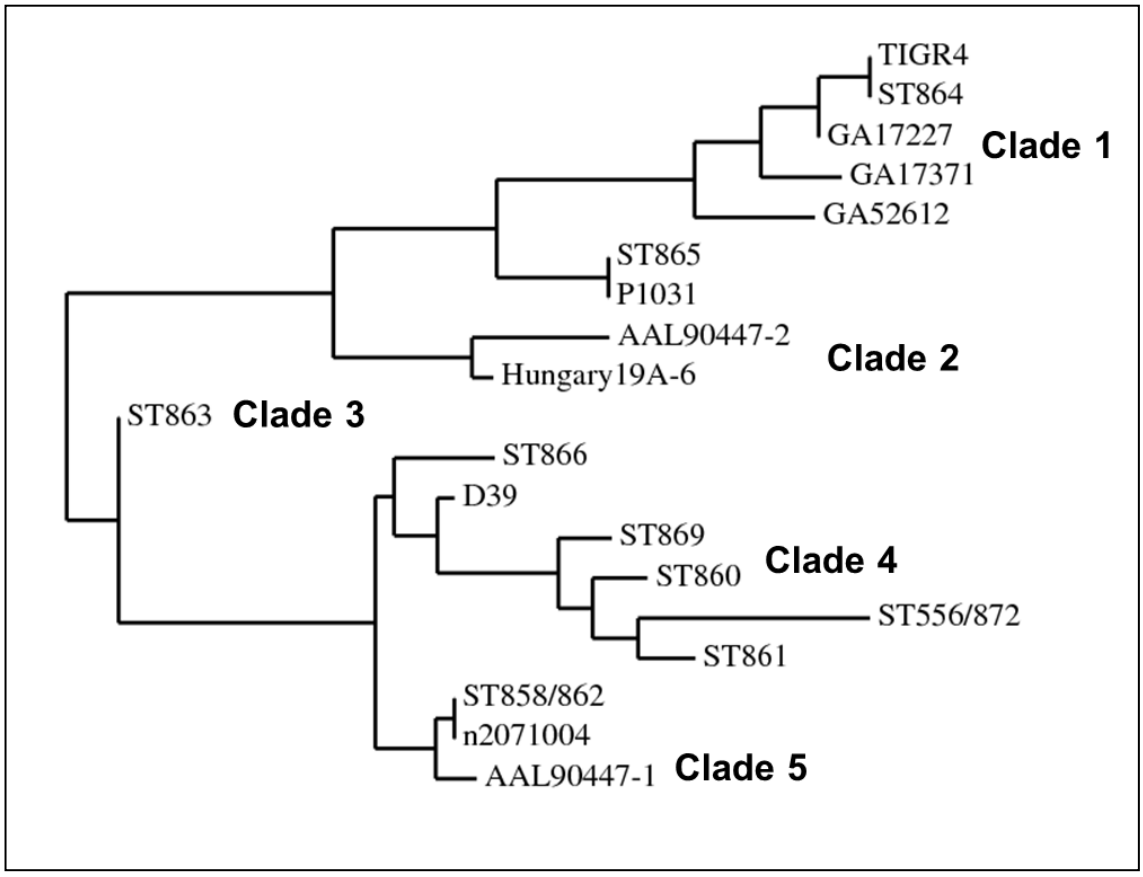


Figure 3.4 Phylogenetic tree showing evolutionary relationship among CbpAN sequences.

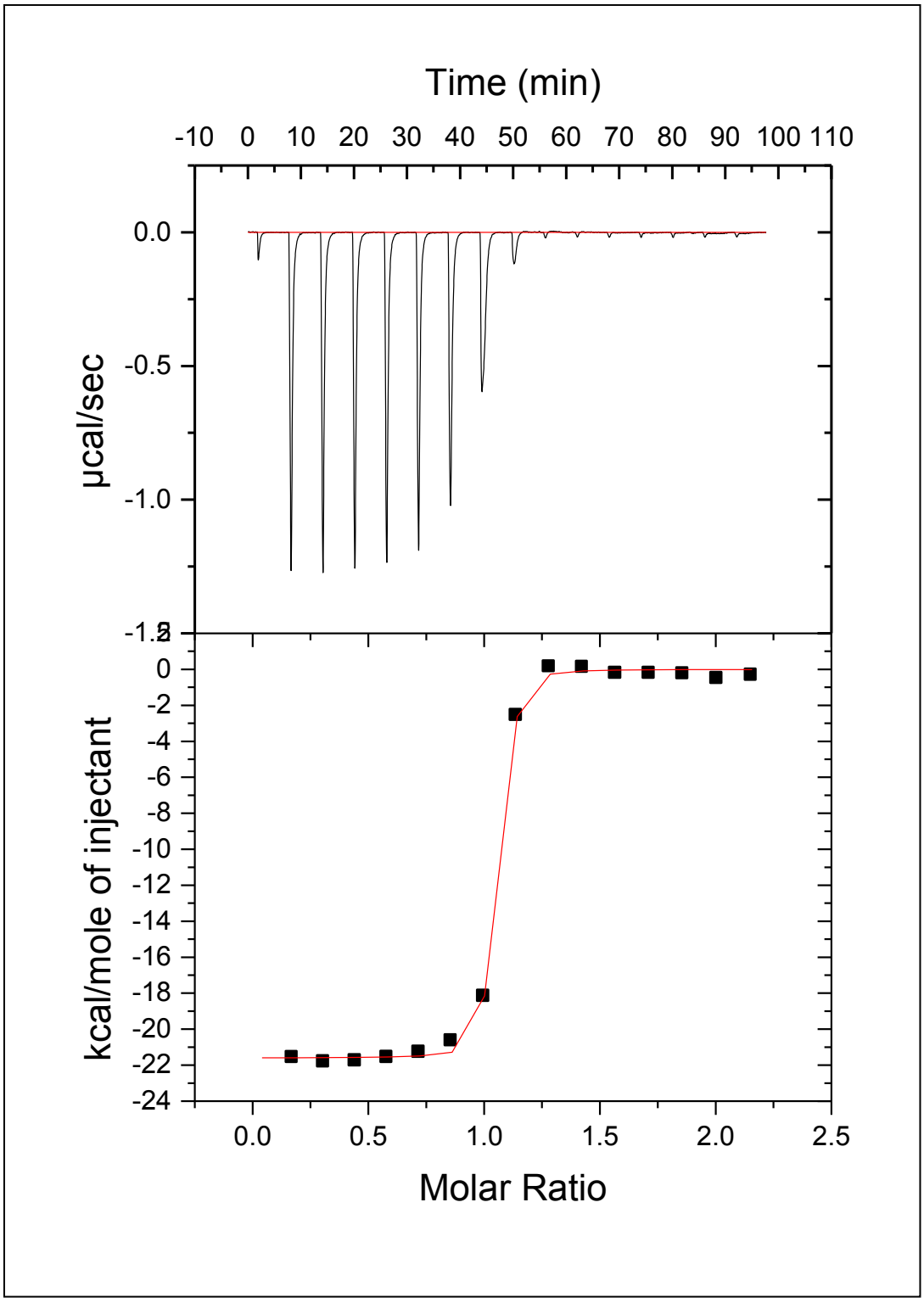


Figure 3.5 ITC analysis of CbpAN (TIGR4) R136A mutant binding to FH CCP9.

3.3.4 FH binding properties of CbpAN variants

To study the sequence and structural features of CbpA for binding to FH CCP9, natural CbpAN variants were produced and their binding properties were measured by ITC. Some of the CbpAN variants were poorly soluble when expressed as His-tagged proteins in *E. coli* and therefore were produced as MBP-fusion proteins, in which case MBP-CbpAN (TIGR4) was used as a reference protein. The variants selected represented different phylogenetic groups (Table 3.3 and Fig. 3.4). Of the seven CbpAN variants tested, only CbpAN from strain ST872 showed a dramatic decrease (~50 fold) in binding affinity. This may be due to low CbpAN sequence similarity of TIGR4 and ST872, which stands at 29%. The pneumococcal strains tested here are invasive and were originally isolated from the blood (strains D39, ST858, ST863, ST869 and ST872) or cerebrospinal fluid (ST866) (Lu et al., 2008).

Table 3.3 Binding of FH CCP9 by CbpAN variants

CbpAN Variants	K_d (nM)	ΔG° (kcal/mol)	ΔH(kcal/mol)	$T\Delta S$(kcal/mol)
CbpAN				
TIGR4	4.72 ± 0.4	-11.35	-21.1 ± 0.67	-9.86 ± 0.62
ST858	27.1 ± 1.13	-10.31	-16.6 ± 0.8	-6.89 ± 0.16
ST869	24 ± 1.0	-10.38	-17.0 ± 1.05	-7.05 ± 0.38
ST872	254 ± 15	-9.0	-5.44 ± 0.49	3.56 ± 0.53
MBP-CbpAN				
TIGR4	6.12 ± 0.66	-11.09	-25.3 ± 1.61	-14.6 ± 1.63
ST863	2.38 ± 0.5	-11.75	-15.8 ± 0.91	-4.08 ± 0.93
ST866	4.99 ±	-11.31	-10.7 ± 1.96	-1.99 ± 0.61
D39	12.3 ± 0.63	-10.78	-21.8 ± 0.60	-10.93 ± 0.60

3.3.5 FH binding properties of site-directed CbpAN mutant proteins

The crystal structure of CbpAN (TIGR4) in complex with FH9 has revealed an interface maintained by both polar and non-polar interactions (Fig. 3.5). Residue H98 of CbpAN forms two hydrogen bonds, one each with D542 and Y547 of FH CCP9. The amino group of K95 is involved in a solvent exposed salt bridge with E523 of FH CCP9, while the backbone carbonyl of S92 of CbpAN is hydrogen bonded with the hydroxyl of Y547 of FH CCP9 and the side-chain amide of N106 of CbpAN forms a hydrogen bond contact with the backbone carbonyl of W540 of FH CCP9.

Several other fairly conserved residues at the interface also contribute to the interaction between CbpAN and FH CCP9. They include Y90 (key), which is surrounded by 4 non-polar residues (lock) of FH CCP9, V102, V105, and K115. The “hydrophobic lock” comprising the 4 hydrophobic residues has been shown to be important for host specificity of *S. pneumoniae* (Chapter 2).

Site-directed mutants were designed based on the multiple sequence alignment, NMR chemical shift mapping, and the crystal structure of the complex of CbpAN and FH CCP9. The results are summarized in Table 3.4. Of the nine mutations, all but three residues caused a significant decrease in affinity for FH CCP9. The mutations which had no significant effect on binding of FH CCP9 were those of N106, Y115 and R136. Of the six mutations that had significant effects on binding of FH CCP9, H98A, K95A, K112A, V102A, V105A and Y90A, those of H98 and V102 had most dramatic effects, elevating the K_d value by 3,100- and 170-fold, respectively. The results indicated that the CbpAN residues H98, K95, K112, V102, V105 and Y90 are all important for binding FH CCP9, and that H98 is the most important residue in this regard.

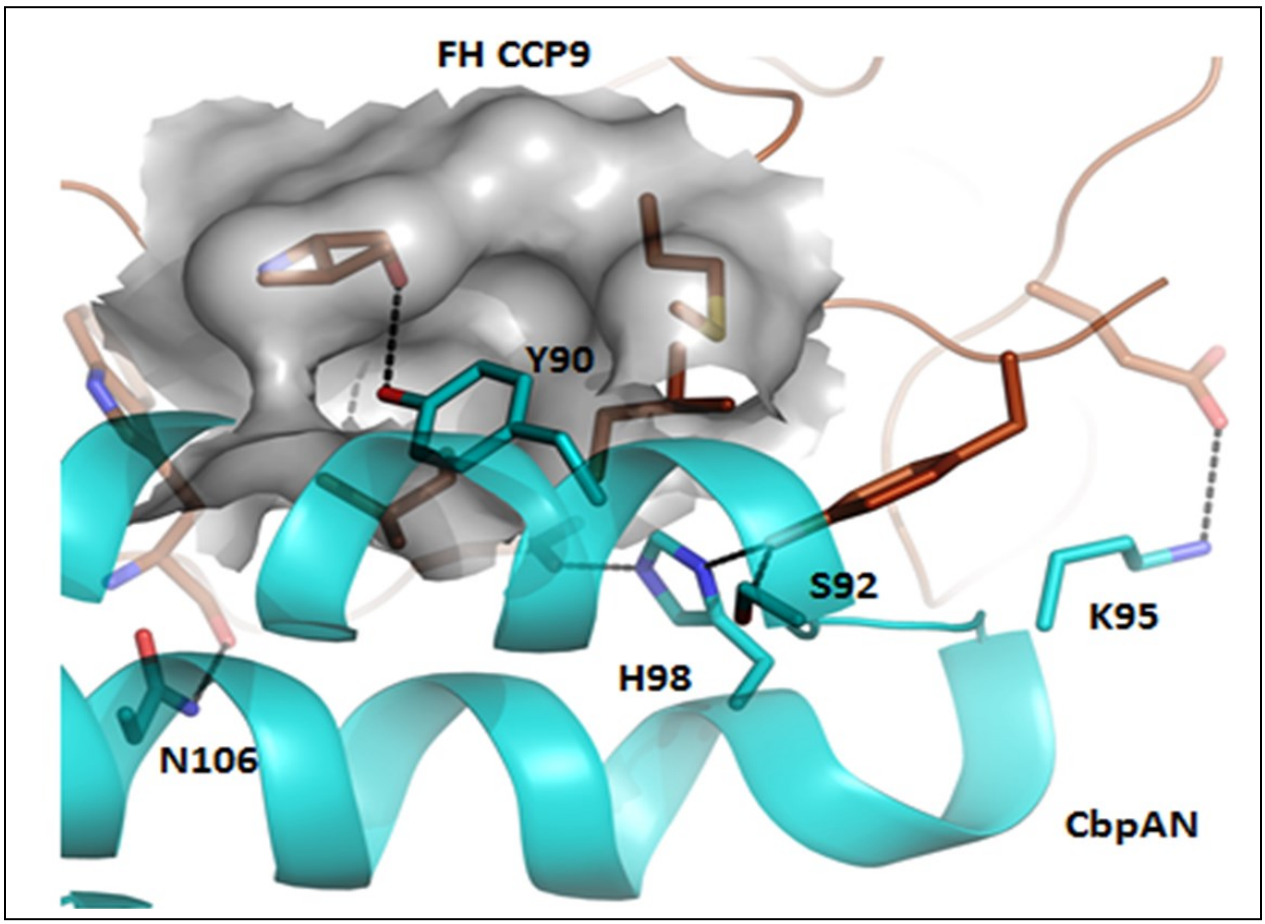


Figure 3.6 Pymol figure showing the non-polar and polar contacts between CbpAN (cyan) and FH CCP9 (gold).

Table 3.4 Binding of FH CCP9 by CbpAN (TIGR4) and its mutants

CbpAN mutants	K_d (nM)	ΔG °(kcal/mol)	ΔH(kcal/mol)	TΔS(kcal/mol)
Wt	4.72 ± 0.2	-11.4	-21.1 ± 0.67	-9.86 ± 0.62
Y90A	73.0 + 7.25	-10.1	-13.0 ± 0.3	-3.07 ± 0.3
K95A	48.8 ± 4.05	-10.4	-19.1 ± 0.5	-8.73 ± 0.6
H98A	14800 ± 365	-7.94	-8.8 ± 0.62	-1.8 ± 0.8
V102A	790 ± 71	-9.68	-21.0 ± 0.5	-12.6 ± 0.48
V105A	93.8 ± 8.75	-10.9	-16.5 ± 0	-6.93 ± 0.04
N106A	9.72 ± 1.26	-10.9	-15.0 ± 7.5	-4.0 ± 0.6
K112A	39.9 ± 0.95	-10.1	-20.1 ± 0.2	-9.94 ± 0.1
Y115A	8.51 ± 0.77	-11.0	-23.0 ± 1.9	-11.9 ± 1.2
R136A	5.09 ± 0.09	-11.3	-21.3 ± .35	-9.91 ± 0.38

3.4 Discussion

S. pneumoniae recruits negative complement regulator, FH, via protein-protein interaction with CbpA to evade the host immune system (Rosenow et al., 1997). CbpA binds FH through its N-terminal domain (CbpAN) (Agarwal et al., 2010a; Agarwal et al., 2010b; Balachandran et al., 2002; Dave et al., 2001; Duthy et al., 2002; Hammerschmidt et al., 1997; Lu et al., 2003; Lu et al., 2006; Rosenow et al., 1997). All virulent strains of pneumococcus tested to date bind to FH (Quin et al., 2006) and that implies that the sequence variation in CbpAN does not abolish its binding to FH. Given the sequence hypervariation in the FH binding domain across the tested strains, it is intriguing that a fairly similar FH binding affinity is maintained among them (Table 3.3).

This finding suggests that there may be a signature structural feature in CbpAN for binding to FH.

Here, an attempt was made to zoom into the FH binding domain of CbpA to uncover those essential structural features of CbpAN for binding to FH. Two approaches were used to achieve this. First, from CbpAN sequence analysis, energy calculations, and analysis of the interface of the crystal structure of FH-CbpA complex, specific CbpAN residues were picked and mutated to assess their importance in binding to FH CCP9. Second, by looking at evolutionary sequence relationships between several CbpAN variants and correlating that with the binding affinities obtained by ITC when they were titrated with FH CCP9, important CbpAN residues were identified. Both approaches converged on the fact that indeed a signature structural feature for binding FH CCP9 may exist with key interface residues such as H98 and non-interface residues such as V105 playing leading roles.

Most of the CbpAN residues that this study found to be crucial for its interaction with FH CCP9, like the conserved H98, fall within the putative “HTITVALVNELN” FH binding motif identified earlier by peptide spot array analysis (Lu et al., 2006)(Table 3.4). The consistency in the results of these two independent studies confirms the important region of CbpAN for binding to FH. Put together with sequence analysis data which indicate that H98, T99, N/T101, V102, L104, and V105 residues of CbpAN are highly conserved (Fig. 3.2), we have confirmed a conserved FH binding motif in CbpAN. Among the CbpAN variants tested for their binding activity to FH CCP9, only ST872 showed a significantly reduced K_d. The sequence similarity between CbpANs from

TIGR4 and ST872 is only 29 % but the conserved residues in the binding motif that are important for interaction with FH CCP9 such as H98, V102, and L108 are present in both of them. A103, N106 and E107 are however replaced by D, K and K respectively in ST872 and this may introduce unfavorable interactions with FH CCP9. One way to investigate further the weaker binding of CbpAN from ST872 to FH CCP9 is by modeling its structure based on CbpAN (TIGR4) structure to verify the orientation of the important interface residues.

K95 of CbpAN forms a solvent-exposed salt bridge with E523 of FH CCP9 and its mutation to Ala weakens CbpAN interaction with FH CCP9 about 10-fold while mutation H98A weakens the interaction with FH9 more than 3,000 fold. Reciprocal mutation of Y547 on FH CCP9 also greatly reduces the binding affinity of the two proteins (Chapter 2). Mutation of the other four CbpAN residues (K112, V102, V105 and Y90) to Ala had significant effects on binding affinity. Y90 showed a large chemical shift on 2D NMR spectrum upon FH9 binding commensurate with its critical role in host specificity. Interestingly, mutation of N106 that is involved in backbone hydrogen bond with W540, to Ala essentially had no impact on binding. Although Y115 and R136 are fairly conserved residues, when mutated to Ala, no significant change in binding affinity was observed, consistent with their location outside the interaction interface (Table 3.4).

Even though the identified CbpAN residues appear to be conserved and important for FH binding, more CbpAN variants, especially from clade 2 (Fig. 3.4) should be analyzed to validate the findings of this study. Since FH is highly abundant in blood and binds tightly to CbpAN (Chapter 2), the discovery of residues whose mutation lowers the binding affinity may help in designing effective CbpA-based vaccines with

appropriate affinity that will not be mopped up by the abundant plasma FH. More importantly, the finding that some residues that are important for binding FH are structurally conserved in CbpA raises hope that a universal vaccine or therapeutic agents against pneumococcus can be designed.

REFERENCES

REFERENCES

Agarwal, V., Asmat, T.M., Dierdorf, N.I., Hauck, C.R., and Hammerschmidt, S. (2010a). Polymeric immunoglobulin receptor-mediated invasion of Streptococcus pneumoniae into host cells requires a coordinate signaling of SRC family of protein-tyrosine kinases, ERK, and c-Jun N-terminal kinase. J Biol Chem *285*, 35615-35623.

Agarwal, V., Asmat, T.M., Luo, S., Jensch, I., Zipfel, P.F., and Hammerschmidt, S. (2010b). Complement regulator Factor H mediates a two-step uptake of Streptococcus pneumoniae by human cells. J Biol Chem *285*, 23486-23495.

Anisimova, M., and Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst Biol *55*, 539-552.

Balachandran, P., Brooks-Walter, A., Virolainen-Julkunen, A., Hollingshead, S.K., and Briles, D.E. (2002). Role of Pneumococcal Surface Protein C in Nasopharyngeal Carriage and Pneumonia and Its Ability To Elicit Protection against Carriage of Streptococcus pneumoniae. Infection and Immunity *70*, 2526-2534.

Brooks-Walter, A., Briles, D.E., and Hollingshead, S.K. (1999). The pspC gene of Streptococcus pneumoniae encodes a polymorphic protein, PspC, which elicits cross-reactive antibodies to PspA and provides immunity to pneumococcal bacteremia. Infect Immun *67*, 6533-6542.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol *17*, 540-552.

Chevenet, F., Brun, C., Banuls, A.L., Jacq, B., and Christen, R. (2006). TreeDyn: towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics *7*, 439.

Dave, S., Brooks-Walter, A., Pangburn, M.K., and McDaniel, L.S. (2001). PspC, a pneumococcal surface protein, binds human factor H. Infect Immun *69*, 3435-3437.

Dave, S., Carmicle, S., Hammerschmidt, S., Pangburn, M.K., and McDaniel, L.S. (2004). Dual roles of PspC, a surface protein of Streptococcus pneumoniae, in binding human secretory IgA and factor H. J Immunol *173*, 471-477.

Duthy, T.G., Ormsby, R.J., Giannakis, E., Ogunniyi, A.D., Stroeher, U.H., Paton, J.C., and Gordon, D.L. (2002). The Human Complement Regulator Factor H Binds Pneumococcal Surface Protein PspC via Short Consensus Repeats 13 to 15. Infection and Immunity *70*, 5604-5611.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res *32*, 1792-1797.

Glaser, F., Pupko, T., Paz, I., Bell, R.E., Bechor-Shental, D., Martz, E., and Ben-Tal, N. (2003). ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics *19*, 163-164.

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol *52*, 696-704.

Hammerschmidt, S., Talay, S.R., Brandtzaeg, P., and Chhatwal, G.S. (1997). SpsA, a novel pneumococcal surface protein with specific binding to secretory immunoglobulin A and secretory component. Mol Microbiol *25*, 1113-1124.

Iannelli, F., Oggioni, M.R., and Pozzi, G. (2002). Allelic variation in the highly polymorphic locus pspC of Streptococcus pneumoniae. Gene *284*, 63-71.

Jarva, H., Janulczyk, R., Hellwage, J., Zipfel, P.F., Bjorck, L., and Meri, S. (2002). Streptococcus pneumoniae evades complement attack and opsonophagocytosis by expressing the pspC locus-encoded Hic protein that binds to short consensus repeats 8-11 of factor H. J Immunol *168*, 1886-1894.

Kapust, R.B., Tozser, J., Fox, J.D., Anderson, D.E., Cherry, S., Copeland, T.D., and Waugh, D.S. (2001). Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. Protein Eng *14*, 993-1000.

Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T., and Ben-Tal, N. (2005). ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucleic Acids Res *33*, W299-302.

Lu, L., Lamm, M.E., Li, H., Corthesy, B., and Zhang, J.R. (2003). The human polymeric immunoglobulin receptor binds to Streptococcus pneumoniae via domains 3 and 4. J Biol Chem *278*, 48178-48187.

Lu, L., Ma, Y., and Zhang, J.R. (2006). Streptococcus pneumoniae recruits complement factor H through the amino terminus of CbpA. J Biol Chem *281*, 15464-15474.

Lu, L., Ma, Z., Jokiranta, T.S., Whitney, A.R., DeLeo, F.R., and Zhang, J.R. (2008). Species-specific interaction of Streptococcus pneumoniae with human complement factor H. J Immunol *181*, 7138-7146.

Musher, D.M. (2009). Streptococcus pneumoniae (Philadelphia: Elsevier Churchill Livingstone).

Ogunniyi, A.D., Grabowicz, M., Briles, D.E., Cook, J., and Paton, J.C. (2007). Development of a Vaccine against Invasive Pneumococcal Disease Based on Combinations of Virulence Proteins of Streptococcus pneumoniae. Infection and Immunity *75*, 350-357.

Quin, L.R., Onwubiko, C., Carmicle, S., and McDaniel, L.S. (2006). Interaction of clinical isolates of Streptococcus pneumoniae with human complement factor H. FEMS Microbiol Lett *264*, 98-103.

Rosenow, C., Ryan, P., Weiser, J.N., Johnson, S., Fontan, P., Ortqvist, A., and Masure, H.R. (1997). Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of Streptococcus pneumoniae. Mol Microbiol *25*, 819-829.

Velazquez-Campoy, A., and Freire, E. (2006). Isothermal titration calorimetry to determine association constants for high-affinity ligands. Nat Protoc *1*, 186-191.

Walport, M.J. (2001). Complement. First of two parts. N Engl J Med *344*, 1058-1066.

Yuste, J., Khandavilli, S., Ansari, N., Muttardi, K., Ismail, L., Hyams, C., Weiser, J., Mitchell, T., and Brown, J.S. (2010). The effects of PspC on complement-mediated immunity to Streptococcus pneumoniae vary with strain background and capsular serotype. Infect Immun *78*, 283-292.

Yuste, J., Sen, A., Truedsson, L., Jonsson, G., Tay, L.S., Hyams, C., Baxendale, H.E., Goldblatt, F., Botto, M., and Brown, J.S. (2008). Impaired opsonization with C3b and phagocytosis of Streptococcus pneumoniae in sera from subjects with defects in the classical complement pathway. Infect Immun *76*, 3761-3770.

CHAPTER 4

Preliminary Study of Human Secretory Component and Future Directions

4.1 Introduction

The interaction between CbpA protein and human secretory component (SC) enhances adherence and translocation of pneumococcus across the epithelium (Dave et al., 2004; Zhang et al., 2000). Mature CbpA protein consists of about 663 amino acids with a theoretical mass of 75 kDa (Rosenow et al., 1997). The N-terminal region of CbpA adjacent to FH binding domain consists of six α-helical structures organized into two direct repeats (R1 and R2), which share 89% amino acid identity (Rosenow et al., 1997). The interaction between R1 and R2 of CbpA and the secretory component (SC) of polymeric Ig receptor (pIgR) has been directly linked to the ability of pneumococci to adhere and invade human nasopharyngeal cells (Zhang et al., 2000). The receptor for pneumococcal CbpA on host mucosal cells was identified as pIgR (Zhang et al., 2000), an integral membrane protein required for transportation of polymeric IgM and IgA across mucosal epithelial cells (Mostov, 1999). PIgR has an N-terminal extracellular ligand-binding domain (SC), a single transmembrane region and a short C-terminal cytoplasmic portion (Mostov and Blobel, 1982). SC is composed of five Ig-like domains (D1-D5), each 104-114 amino acids long (Mostov and Blobel, 1982). CbpA-deficient pneumococcus exhibits decreased adherence to human cells and attenuated nasopharyngeal colonization in rats (Rosenow et al., 1997).

CbpA interacts with the SC of pIgR via its two “YRNYPT” hexapeptide motifs, one each in R1 and R2 (Hammerschmidt et al., 2000). An NMR structure of R2 of CbpA(TIGR4) revealed that it is composed of 3-α-helical raft-like structures with the hexapeptide responsible for SC binding located in the loop 1 region between the first and second helices (Luo et al., 2005). Deletional analysis of pIgR located the CbpA

binding site on pIgR to the D3 and D4 of the extracellular SC domain (Lu et al., 2003). Interaction between CbpA and SC, secretory IgA (sIgA) and pIgR was found to be species specific, with interaction only detected with human but not with counterparts from mouse and other species (Lu et al., 2008; Zhang et al., 2000). To understand pneumococcal pathogenesis it is necessary to investigate the molecular and structural basis of the host specific interaction between CbpAN and SC. Even though the CbpAR2 structure (Luo et al., 2005) provides valuable insights on the interaction between pneumococcus and pIgR, a comprehensive picture of this interaction can only be obtained by mapping the interaction interface of both SC and CbpA, which can be achieved by solving the structure of the SC-CbpA complex and subsequent mutagenic analysis of interface residues of both proteins. This is the overall goal of this study.

To achieve this goal, it is desirable to produce functional SC protein in sufficient quantities for biochemical and biophysical characterization. In this report, I highlight preliminary efforts towards establishment of an *Escherichia coli*-based system for production of functional SC proteins. Using ITC, I demonstrated that MBP-fusion SC proteins produced in *E. coli* are functionally active and bind to CbpAR2 with high affinity. This result was confirmed by 2D-NMR titration where labeled CbpAR2 was titrated with unlabeled SC3-4 fusion protein. (Dr. Liu, Preliminary data)

4.2 Materials and Methods

4.2.1 Construction of the CbpAR2 overexpression system and protein purification

The DNA fragment encoding CbpAR2 was amplified by PCR from the genomic DNA of *S. pneumoniae* strain TIGR4 with primers CbpAR2f and CbpAR2r (Table 4.1). The amplified DNA fragment was cloned into the expression vector pET17bHR (lab-made, derived from pET17b (Novagen) by digestion with the restriction enzymes *BamHI* and *NdeI* and ligation. The cloned DNA fragment was sequenced to ensure the correct coding sequence. The overexpression plasmid construct was transformed into the *E. coli* strain BL21 (DE3) for the production of the His-tagged CbpAR2. The expression system was cultured in LB media containing 20 μg/ml chloramphenicol and 100 μg/ml ampicillin with vigorous shaking at 37 °C overnight. The cells were harvested by centrifugation and resuspended in buffer A containing 50 mM sodium phosphate, pH 8.0 and 300 mM NaCl, and lysed by sonication. The lysate was clarified by centrifugation at 16,000 g for 20 min before loading onto a Ni-NTA agarose column pre-equilibrated with buffer A. After washing with 10 mM imidazole in buffer A, the column was eluted with a linear gradient of 10-250 mM imidazole in buffer A. The protein-containing fractions were analyzed by SDS-PAGE and the fractions containing pure CbpAR2 were pooled and concentrated using a stirred cell Amicon concentrator. When desired, the His-tag was cleaved with TEV protease (Sun et al., 2011). The cleaved His-tag and the uncleaved protein were removed by a second passage through a Ni-NTA agarose column. The flow-through was collected, concentrated, and loaded onto a Sephadex G-75 gel filtration column. The fractions containing pure CbpAR2 were pooled, concentrated, and dialyzed first against 5 mM sodium phosphate, pH 8.0, and against water for 12 h. The dialyzed protein solution was lyophilized and stored at -80 °C.

Isotopically labeled ^{15}N- CbpAR2 was produced and purified in the same way except that LB medium was replaced with M9 medium containing $^{15}NH_4Cl$ as the sole nitrogen source.

4.2.2 Construction of the overexpression systems for SC domains and protein purification

DNA fragments encoding various full length SC (SC1-5) and domains 3 and 4 of SC (SC3-4) were cloned into the expression vector pET17bHMT (lab-made, derived from pET17b (Novagen)) using the same methods as described in section 4.2.1. The primers for PCR cloning were SC1-5f and SC1-5r for SC1-5, and SC3-4f and SC3-4r (Table 4.1). The correct coding sequences were confirmed by DNA sequencing. The overexpression plasmids were transformed into the *E. coli* strain SHuffle T7 Express *LysY* (New England Biosciences, NEB) for SC3-4 and *E. coli* strain BL21 (DE3) C41 pLysS for SC1-5. The fusion proteins produced by the overexpression systems contained one His-tag before the maltose binding protein (MBP) followed by a thrombin cleavage and the SC construct. The procedure for protein production and cell harvesting was the same as described earlier. The harvested *E. coli* cells were resuspended in buffer B containing 20 mM Tris-HCl and 150 mM NaCl, pH 7.5, and lysed by sonication. The lysate was clarified by centrifugation and loaded onto a Ni-NTA agarose column pre-equilibrated with buffer B. After washing with buffer B containing 20 mM imidazole, the column was eluted with a linear 20-250 mM imidazole gradient. The protein-containing fractions were analyzed by SDS-PAGE, pooled and concentrated. The fractions containing SC fusion proteins were concentrated and further purified by gel filtration chromatography on a P-100 column. The purified protein solutions were

dialyzed first against 5 mM sodium phosphate, pH 8.0, then against 2 mM sodium phosphate, pH 8.0, and lyophilized. The lyophilized proteins were stored at -80 °C.

4.2.3 Isothermal Titration Calorimetry

FH and CbpA proteins were purified as described above and protein concentration was estimated using absorbance at λ = 280 nm. ITC measurements were done as described previously (Velazquez-Campoy and Freire, 2006). Before titration, purified H_6MBPSC3-4 and CbpAR2 proteins were dialyzed against the same buffer (50 mM HEPES, pH7.5, 50 mM KCl). 40 μM H_6MBPSC3-4 was loaded into the sample cell and 800 μM CbpAR2 was loaded into the syringe. Titration was performed at 25 ^{0}C by injecting H_6CbpAR2 in 10 μL increments into H_6MBPSC3-4 in the cell with reference power of 5 cal/s under high feedback mode. A total of 25 sequential injections were done with inter-injection time of 360 s. Same protein concentrations and ITC conditions were used to test the interaction between H_6MBPSC1-5 and CbpAR2. Dissociation constant (K_d) and binding stoichiometry (N) were extracted from analysis of data with Origin version 5.0.

Table 4.1 Primers for cloning SC3-4, SC1-5 and CbpAR2.

Primer	Sequence
CbpAR1f	5'- G GAA TTC CAT ATG ACA GAA CCA GGA GAA AAG GTA GCA G -3'
CbpAR2f	5'- G GAA TTC CAT ATG GAA AAA AAG GTA GCA GAA GCT G-3'
CbpAR2r	5'- CG GGA TCC TCA TTT TTC TTT AAC TTT ATC TTC TTC-3'
SC3f	5'- CG GGA TC CAT ATG GAG CTG GTT TAT GAA GAC CTG AGG-3'
SC4r	5'- G GAA TTC TCA TTC GAT AAT CTT GAT CTC CAC GGT G-3'
SC1f	5'- CG GGA TC CAT ATG AAG AGT CCC ATA TTT GGT CCC GAG-3'
SC5r	5'- G GAA TTC TCA TTC AAC TGC CAC ATA GAC GGC TGC AG-3'

4.3 Preliminary results

4.3.1 Purification of CbpAR2 and SC proteins

Labeled and unlabeled CbpAR2 proteins were purified to homogeneity by a combination of affinity and size exclusion chromatography (Fig. 4.1). Recombinant CbpAR2 is stable and its solution NMR structure is available (PDB 1w9r) (Luo et al., 2005). SC proteins are however difficult to purify because SC1-5 fusion protein does not bind well to several columns including Ni^{2+}, ion exchange (DEAE, MonoQ) and hydrophobic column. SC3-4 fusion protein partially binds to Ni^{2+} column but the fusion protein is only partially cleaved upon incubation with thrombin protease. SC3-4 was later cloned into a pET17bHMHT vector encoding two His tags, one at the N-terminus and another between MBP fusion protein and thrombin cleavage site and expressed in *E. coli* SHuffle pLysY to improve its retention on Ni^{2+}-column as well as thrombin cleavage. On a pilot scale, this new vector improved both binding to Ni^{2+}-column and cleavage by thrombin for SC3-4 but attenuated expression of SC1-5. However the SC3-4 protein purified from the double His-tag vector, PET17bHMHT appears to be only slightly active (<10%) for binding CbpAR2.

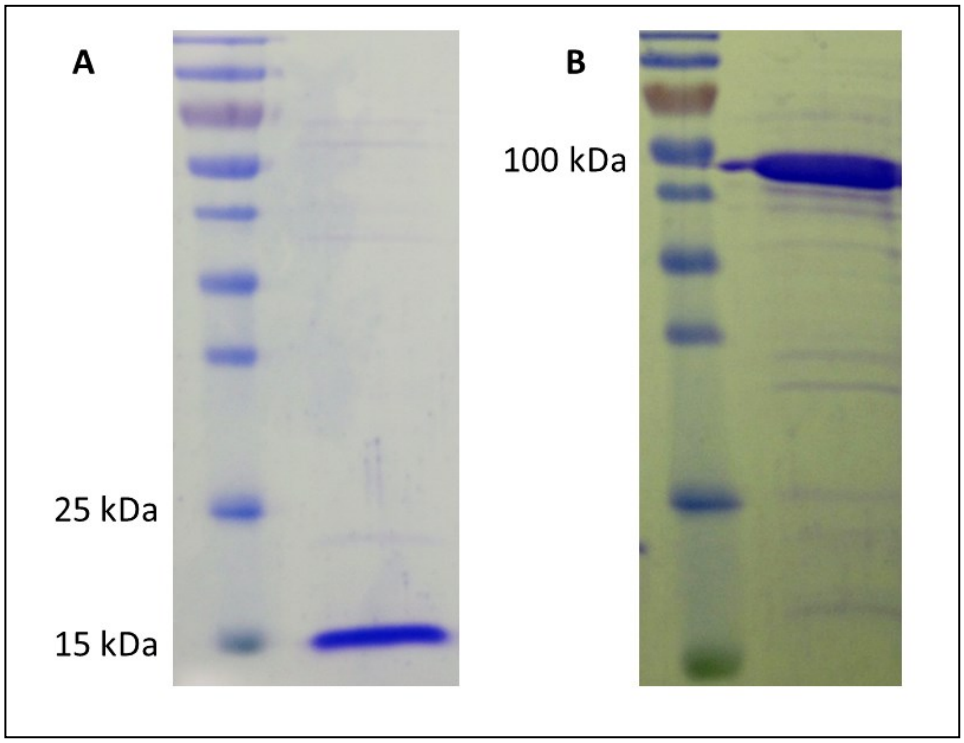


Figure 4.1 Comassie stained 12 % SDS-PA gels of purified CbpAR2 (A) and H_6MBPSC3-4 (B) proteins. Lane 1 is for protein marker and lane 2 is for purified protein in (A) and (B).

4.3.2 SC binds tightly to CbpAR2

SC and CbpAR2 proteins were purified as described above. The interaction between H_6CbpAR2 and H_6MBPSC3-4 was tested by ITC and confirmed by 2D-NMR titration assays. MBP- fusion proteins of SC were used in both assays. For ITC experiment, 800 μM CbpAR2 was injected into a cell loaded with 40 μM H_6MBPSC3-4. For 2D-NMR titration assay, the spectrum of ^{15}N –H_6CbpAR2 was first collected alone then in complex with H_6MBPSC3-4. CbpAR2 resonance assignment had already been reported (Luo et al., 2005). As a control, an ITC assay where H_6MBP was titrated into

H_6CbpAR2 was done to ensure that MBP was not contributing to the interaction between the two proteins. Tight interaction was detected between H_6CbpAR2 and H_6MBPSC3-4 with a K_d of ~ 11 nM. The preliminary 2D NMR data also pointed to a tight interaction between the two proteins (Dr. Liu, data not shown). The ITC experiment as well as purification of SC proteins needs to be optimized and this is where the focus of this study is going to be in the near future.

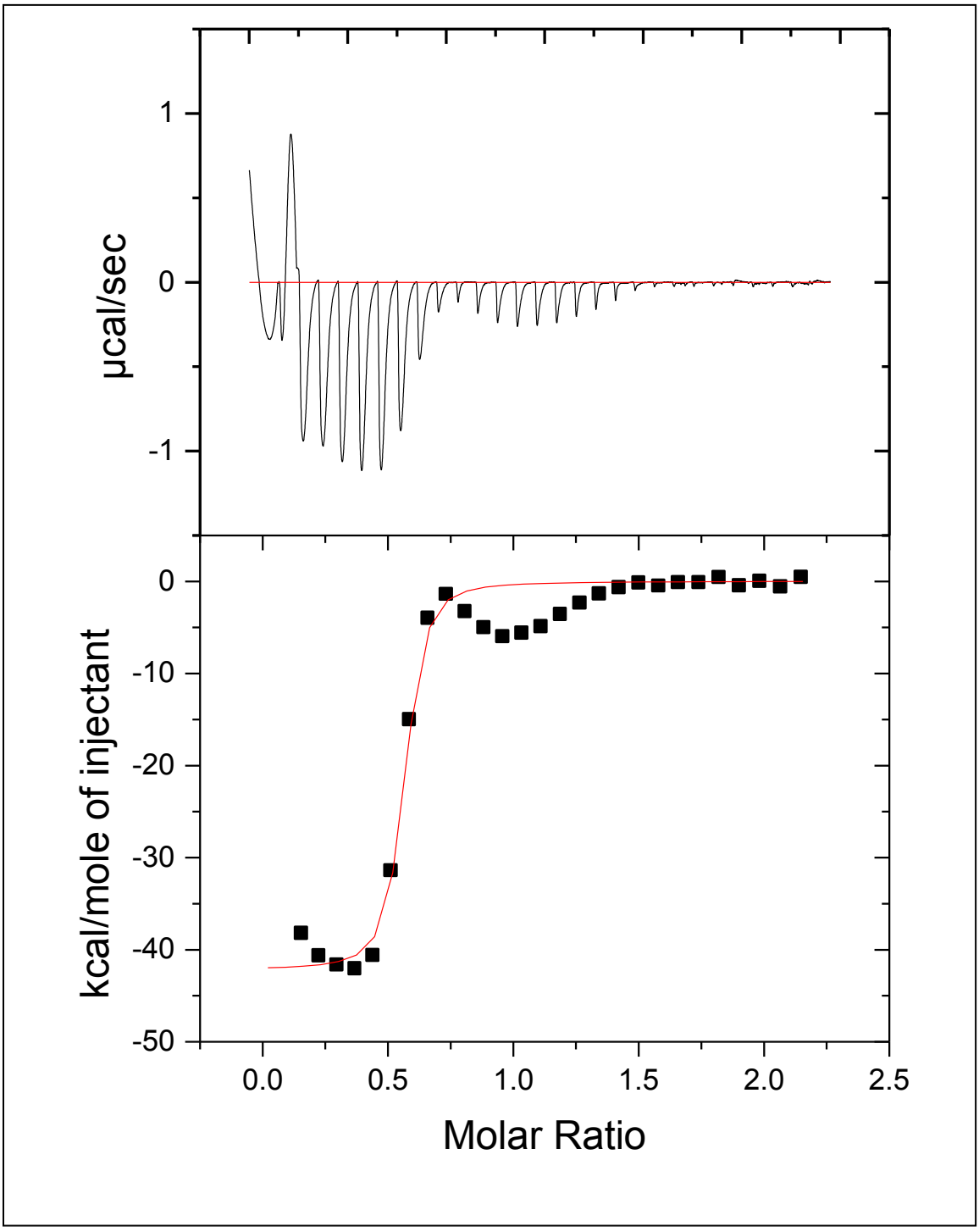


Figure 4.2 ITC analysis of the interaction between H_6CbpAR2 and H_6MBPSC3-4.

4.4 Discussion

The interaction between CbpA and human SC proteins enhances adherence and translocation of pneumococcus across the epithelium (Dave et al., 2004; Zhang et al., 2000). This interaction is localized to the R1 and R2 repeats of CbpA and the SC3-4 domains of SC (Dave et al., 2004; Zhang et al., 2000). CbpA is immunogenic and a good candidate for development of protein-based anti-pneumococcus vaccine (Briles et al., 2000; Brooks-Walter et al., 1999). CbpA R2 protein has been structurally characterized (Luo et al., 2005) but no high resolution structure is available for any of the 5 domains of SC protein and this has hindered detailed mapping of the interaction interface of the two proteins. In the current study, we initiated structural and biochemical characterization of the interaction between CbpA and SC.

One major hindrance in this study has been isolation of functional SC proteins. There is no reliable *E.coli*-based method for isolating SC proteins in sufficient quantities and quality for biophysical characterization. Previously, SC has been purified from milk or bile (Hammerschmidt et al., 1997) or expressed and isolated from CHO cells producing human SC (Luo et al., 2005) in functional form. The protein obtained from these sources, even though fully active, are in limited amounts and may not be enough support intensive biophysical studies. We therefore attempted to create a reliable *E. coli* based system for producing soluble SC proteins. Good yields have been obtained for MBP-fusion proteins of SC3-4 (Fig 4.1) and SC1-5. However, fusion proteins are only partially active (N value of 0.5 indicating 50% activity) (Fig 4.2) Our efforts to purify free SC proteins were met with major challenges, including: (i) the fusion proteins either do not or bind poorly to most chromatographic columns; (ii) thrombin cleavage efficiency of

fusion protein is low; and (iii) upon cleavage of MBP, the proteins tend to gradually precipitate out of solution even at 4 °C. In future, optimization of SC purification should explore enhancing reshuffling of disulfide bonds *in vitro* by incubation with DsbC isomerase before or during thrombin cleavage, and adjustment of buffer conditions to establish optimum conditions for the stability of SC proteins.

4.5 Future Directions

S. pneumoniae is still a leading cause of high mortality and morbidity in the first world and developing countries, especially among the children, the elderly and immunocompromised populations (Musher, 2009). More antibiotic-resistant pneumococcal strains continue to be isolated and the two licensed capsule-based vaccines are effective only against a subset of the over 90 pneumococcal strains typified to date (Gamez and Hammerschmidt, 2012). It is also feared that the intensive use of these capsule-based vaccines will lead to emergence of replacement virulent strains that are not currently covered by vaccines. Several pneumococcal virulence factors have been discovered and partially characterized to establish their contributions to the pathogenesis of pneumococcus (Jedrzejas, 2004). The pneumococcus uses multiple strategies to adhere, colonize, and invade human organs as well as to evade host immune system. It achieves its potency by, (i) secreting proteases which inactivate host immune system, (ii) producing extracellular virulence factors such as CbpA, which recruits host regulators of complement or disguise host factors and (iii) expressing complement inhibitors (Serruto et al., 2010). Several surface-exposed pneumococcal virulence factors have since been described but very few have been characterized in details, biochemically or structurally (Kadioglu et al., 2008).

A lot has been learnt on the extracellular pneumococcal virulence factors over the past three decades (Kadioglu et al., 2008). One of the main goals for their characterization in terms of structure, protein-protein interactions, and functional studies has been to elucidate the immune evasion mechanisms of the pneumococcus and to learn about its pathogenesis. The consensus as at now is that the pneumococcus, just like other human pathogens, recruits regulators of complement activity (RCAs) such as factor H (FH), factor H like -1 (FHL-1) and C4 binding protein (C4BP) to its surface to escape complement attack (Hammerschmidt et al., 2007; Kadioglu et al., 2008). The recruited RCAs remain functional and can prevent complement activity on the surface of the pathogens. The pneumococcus recruits FH through interaction with CbpA. The molecular details of these interactions are just starting to emerge with determination of high resolution structures of the bacterial virulence factors, host factors and their complexes. For example, while *Neisseria meningitidis* use protein mimicry of host carbohydrates to recruit FH via a tight interaction (K_d ~ 5 nM) of its FH binding protein (fHbp) and FH CCP 6-7 (Schneider et al., 2009), *Borrelia burgdorferi* binds FH CCP19-20 in similar but not identical way as host endothelial cells via its surface protein OspE (Bhattacharjee et al., 2013) and our results (Chapters 2 and 3) show that CbpAN bind to FH CCP9 with a similar affinity (~ 4.7 nM) but through a completely different interface composed of both polar and non-polar interactions. Structural and sequence alignment of FH CCP8, CCP4 and CCP9 demonstrated that only FH CCP9 has the requisite interface residues for binding to CbpAN. “Hydrophobic locking” discovered here in the interaction between CbpAN and FH CCP9 has never been described before for any other host-pathogen interaction.

The mechanisms used by *Borrelia burgdorferi, N. meningitidis* and *S. pneumoniae* to recruit FH are different and because virulence factors from other pathogens interacts with different regions of FH (Ferreira et al.), it is not likely that RCA recruitment mechanism is conserved among them. This observation suggests that the hope of discovering a universal mode of inhibiting pathogen-RCA interaction may not be easily achieved. It is becoming clear that, even though a common strategy to divert human arsenals by pathogens is by recruitment of RCAs, the pathogens use diverse interfaces to interact with different regions of RCAs and for this reason, it will need characterization of more virulence factors in order to establish the rules of thumb, if any, for this apparently important immune evasion strategy. Biochemical characterization a lone may results in controversial data on the interaction sites of virulent factors on FH. A good example of this is in determination of FH region that interacts with CbpA where different research groups came up with contrasting results (Hammerschmidt, 2006). Using a combination of biochemical, computational and structural studies we authoritatively identified the regions and key residues of FH and CbpA required for their interaction (Chapters 2 and 3). The same tools also allowed us to map the interaction interfaces of both proteins. Therefore, application of this kind of approach holds the promise for speeding up the characterization of host-pathogen interactions, which is crucial for the pathogenesis of diverse microbes. Determination of more structures of complexes between FH and other virulent factors should be pursued in future with an aim to understand the array of FH recruitment strategies utilized by diverse pathogens. These types of studies will greatly benefit from a novel FH expression and isolation protocol that we established and optimized (Chapter 2). Through this protocol,

functional relevant domains of FH can be rapidly isolated in sufficient quantities and used in both biochemical and biophysical characterization of their interaction with virulent factors from pathogens. The information derived from these types of studies may in future stimulate design of broad-based protein vaccines and novel therapeutics aimed at disrupting the interaction between pathogens and FH.

The findings in this study will inform generation of mouse models for studying pneumococcal diseases by introducing the five host-specificity determinant residues in to mFH CCP9 thereby reconstituting CbpA binding activity while at the same time retaining the other functions of mFH. The discovery here that CbpA variants retain a common FH binding feature increases hope of designing broad-based small molecules for disruption of this interaction or a universal vaccine against the pneumococcus. Despite the significance of the findings highlighted in this dissertation, a few challenges still remain on the path of applying them in the development of vaccines or therapeutic agents against pneumococcus. Of particular concern is the tight interaction between FH and CbpA, considering that FH is highly abundant in plasma. The studies on CbpA in Chapter 3 can give a clue on how to design CbpA antigens with moderate FH binding affinity but this itself faces the huddle of antigenic variance in the CbpAN domain. The tight interaction between FH and CbpA also poses a challenge at designing an appropriate small molecule that can effectively disrupt the interaction. In conclusion we have determined the requisite regions of FH and CbpA for their interaction, mapped their interaction interfaces, determined the structural basis for host-specific interaction between FH and CbpA, and provided clues into essential features of CbpA for binding FH protein. We have also established an *E. coli*-based system for isolation of functional

truncated FH proteins, which should catalyze the on-going effort to characterize the interaction between FH and various human pathogens. To this end our laboratory has already embarked on studying interactions between FH and virulent factors from diverse human pathogens.

REFERENCES

REFERENCES

Bhattacharjee, A., Oeemig, J.S., Kolodziejczyk, R., Meri, T., Kajander, T., Lehtinen, M.J., Iwai, H., Jokiranta, T.S., and Goldman, A. (2013). Structural Basis for Complement Evasion by Lyme Disease Pathogen Borrelia burgdorferi. The Journal of biological chemistry *288*, 18685-18695.

Briles, D.E., Hollingshead, S., Brooks-Walter, A., Nabors, G.S., Ferguson, L., Schilling, M., Gravenstein, S., Braun, P., King, J., and Swift, A. (2000). The potential to use PspA and other pneumococcal proteins to elicit protection against pneumococcal infection. Vaccine *18*, 1707-1711.

Brooks-Walter, A., Briles, D.E., and Hollingshead, S.K. (1999). The pspC gene of Streptococcus pneumoniae encodes a polymorphic protein, PspC, which elicits cross-reactive antibodies to PspA and provides immunity to pneumococcal bacteremia. Infect Immun *67*, 6533-6542.

Dave, S., Carmicle, S., Hammerschmidt, S., Pangburn, M.K., and McDaniel, L.S. (2004). Dual roles of PspC, a surface protein of Streptococcus pneumoniae, in binding human secretory IgA and factor H. J Immunol *173*, 471-477.

Ferreira, V.P., Pangburn, M.K., and Cortes, C. Complement control protein factor H: the good, the bad, and the inadequate. Mol Immunol *47*, 2187-2197.

Gamez, G., and Hammerschmidt, S. (2012). Combat pneumococcal infections: adhesins as candidates for protein-based vaccine development. Curr Drug Targets *13*, 323-337.

Hammerschmidt, S. (2006). Adherence molecules of pathogenic pneumococci. Curr Opin Microbiol *9*, 12-20.

Hammerschmidt, S., Agarwal, V., Kunert, A., Haelbich, S., Skerka, C., and Zipfel, P.F. (2007). The Host Immune Regulator Factor H Interacts via Two Contact Sites with the PspC Protein of Streptococcus pneumoniae and Mediates Adhesion to Host Epithelial Cells. The Journal of Immunology *178*, 5848-5858.

Hammerschmidt, S., Talay, S.R., Brandtzaeg, P., and Chhatwal, G.S. (1997). SpsA, a novel pneumococcal surface protein with specific binding to secretory immunoglobulin A and secretory component. Mol Microbiol *25*, 1113-1124.

Hammerschmidt, S., Tillig, M.P., Wolff, S., Vaerman, J.P., and Chhatwal, G.S. (2000). Species-specific binding of human secretory component to SpsA protein of Streptococcus pneumoniae via a hexapeptide motif. Mol Microbiol *36*, 726-736.
Jedrzejas, M.J. (2004). Extracellular virulence factors of Streptococcus pneumoniae. Front Biosci *9*, 891-914.

Kadioglu, A., Weiser, J.N., Paton, J.C., and Andrew, P.W. (2008). The role of Streptococcus pneumoniae virulence factors in host respiratory colonization and disease. Nat Rev Microbiol *6*, 288-301.

Lu, L., Lamm, M.E., Li, H., Corthesy, B., and Zhang, J.R. (2003). The human polymeric immunoglobulin receptor binds to Streptococcus pneumoniae via domains 3 and 4. The Journal of biological chemistry *278*, 48178-48187.

Lu, L., Ma, Z., Jokiranta, T.S., Whitney, A.R., DeLeo, F.R., and Zhang, J.-R. (2008). Species-Specific Interaction of Streptococcus pneumoniae with Human Complement Factor H. The Journal of Immunology *181*, 7138-7146.

Luo, R., Mann, B., Lewis, W.S., Rowe, A., Heath, R., Stewart, M.L., Hamburger, A.E., Sivakolundu, S., Lacy, E.R., Bjorkman, P.J.*, et al.* (2005). Solution structure of choline binding protein A, the major adhesin of Streptococcus pneumoniae. EMBO J *24*, 34-43.

Mostov, K.E., and Blobel, G. (1982). A transmembrane precursor of secretory component. The receptor for transcellular transport of polymeric immunoglobulins. The Journal of biological chemistry *257*, 11816-11821.

Mostov, K.E.a.C.S.K. (1999). Immunoglobulin transport and the polymeric immunoglobulin receptor, pIgR, 2nd edn (San Diego.: Academic Press).
Musher, D.M. (2009). Streptococcus pneumoniae (Philadelphia: Elsevier Churchill Livingstone).

Rosenow, C., Ryan, P., Weiser, J.N., Johnson, S., Fontan, P., Ortqvist, A., and Masure, H.R. (1997). Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of Streptococcus pneumoniae. Mol Microbiol *25*, 819-829.

Schneider, M.C., Prosser, B.E., Caesar, J.J., Kugelberg, E., Li, S., Zhang, Q., Quoraishi, S., Lovett, J.E., Deane, J.E., Sim, R.B.*, et al.* (2009). Neisseria meningitidis recruits factor H using protein mimicry of host carbohydrates. Nature *458*, 890-893.

Serruto, D., Rappuoli, R., Scarselli, M., Gros, P., and van Strijp, J.A. (2010). Molecular mechanisms of complement evasion: learning from staphylococci and meningococci. Nat Rev Microbiol *8*, 393-399.

Sun, P., Tropea, J.E., and Waugh, D.S. (2011). Enhancing the Solubility of Recombinant Proteins in Escherichia coli by Using Hexahistidine-Tagged Maltose-Binding Protein as a Fusion Partner. Methods Mol Biol *705*, 259-274.

Velazquez-Campoy, A., and Freire, E. (2006). Isothermal titration calorimetry to determine association constants for high-affinity ligands. Nat Protoc *1*, 186-191.

Zhang, J.R., Mostov, K.E., Lamm, M.E., Nanno, M., Shimida, S., Ohwaki, M., and Tuomanen, E. (2000). The polymeric immunoglobulin receptor translocates pneumococci across human nasopharyngeal epithelial cells. Cell *102*, 827-837.

APPENDIX

Additional Studies: Biochemical characterization of the ribosome assembly GTPase RbgA in *Bacillus subtilis*

Research described in this chapter was originally published in Journal of Biological Chemistry.
David Achila, Megha Gulati, Nikhil Jain, and Robert A. Britton (2012) Biochemical characterization of ribosome assembly GTPase RbgA in *Bacillus subtilis.*
J. Biol. Chem. 2012, 287:8417-8423.

Abstract

The ribosome biogenesis GTPase A protein (RbgA) is involved in the assembly of the large ribosomal subunit in *Bacillus subtilis* and homologs of RbgA are implicated in the biogenesis of mitochondrial, chloroplast, and cytoplasmic ribosomes in archaea and eukaryotes. The precise function of how RbgA contributes to ribosome assembly is not understood. Defects in RbgA give rise to a large ribosomal subunit that is immature and migrates at 45 S in sucrose density gradients. The study reports a detailed biochemical analysis of RbgA and its interaction with the ribosome. It was found that RbgA, like most other GTPases, exhibits a very low k_{cat} (14 h^{-1}) and has a high K_m (90 μM). Homology modeling of the RbgA switch I region using the K-loop GTPase MnmE as a template suggested that RbgA requires K^+ ions for GTPase activity, which was confirmed experimentally. Interaction with 50 S subunits, but not 45 S intermediates, increases GTPase activity ~55 fold. Stable association with 50 S subunits and 45 S intermediates was nucleotide dependent and GDP did not support strong interaction with either of the subunits. GTP and guanosine 5′-(β,γ-imido)triphosphate (GMPPNP) were sufficient to promote association with the 45 S intermediate while only GMPPNP was able to support binding to the 50 S subunit, presumably due to the stimulation of GTP hydrolysis. These results support a model in which RbgA promotes a late step in ribosome biogenesis and that one role of GTP hydrolysis is to stimulate dissociation of RbgA from the ribosome.

A1.1 INTRODUCTION

Ribosome assembly is a complex, tightly regulated process that involves the coordinated assembly of 3 RNA molecules and over 50 proteins (Nierhaus, 1991; Nomura and Erdmann, 1970). Although more than 150 accessory proteins required for the assembly of ribosomes in eukaryotes have been identified (Dez and Tollervey, 2004), relatively few proteins dedicated to the assembly of ribosomes have been discovered in bacteria (Britton, 2009; Culver, 2003; Wilson and Nierhaus, 2007). Several ribosome associated GTPases (RA-GTPase) such as RbgA (YlqF), Era, YqeH, YphC (EngA), CgtEA, YloQ (YjeQ, RsgA) and YsxC have been implicated in the assembly of either the 50 S or 30 S ribosomal subunits in bacteria (Britton, 2009; Campbell et al., 2005; Gollop and March, 1991; Jiang et al., 2006; Matsuo et al., 2006; Schaefer et al., 2006; Uicker et al., 2006; Uicker et al., 2007), however their precise functions in ribosome assembly has remained elusive. The RA-GTPase RbgA is required for a late maturation step of the 50 S subunit in *B. subtilis* (Matsuo et al., 2006; Uicker et al., 2006), and its homologs have been shown to be required for the assembly of the large ribosomal subunit in eukaryotes (Bassler et al., 2001; Jensen et al., 2003; Kallstrom et al., 2003; Saveanu et al., 2001).

Depletion of RbgA in *B. subtilis* stalls 50 S subunit assembly, resulting in the accumulation of a large ribosomal intermediate that migrates more slowly (45 S) through a sucrose gradient and lacks three ribosomal proteins, L16, L27, and L36 (Matsuo et al., 2006; Uicker et al., 2006; Uicker et al., 2007). RbgA can interact with the 45 S intermediate and 50 S subunits, but the latter interaction was only observed in the presence of a non-hydrolysable analog of GTP (Matsuo et al., 2006; Uicker et al., 2006;

Uicker et al., 2007). It was also observed that intact 50 S stimulates the GTPase activity of RbgA (Matsuo et al., 2006). These observations led to proposal of a model in which RbgA promotes a late step in ribosome assembly and that GTPase activation of RbgA occurs upon correct assembly of the 50 S subunit followed by dissociation of RbgA from the subunit (Matsuo et al., 2006; Uicker et al., 2006). In this way RbgA would serve as a checkpoint to ensure proper formation of the 50 S subunit. However, it was later reported that the GTPase activity of RbgA is maximally stimulated by the 45 S intermediate (referred to as pre-50 S in this model) and “free” 50 S subunits but not by mature 50 S subunits (those isolated by dissociation of subunits from 70 S ribosomes) (Matsuo Y, 2007). These additional results led to a revised model in which RbgA promotes a GTPase dependent structural rearrangement of the 45 S complex that then allows L16 and L27 to subsequently bind. GDP-bound RbgA remains associated with the ribosome until some undetermined signal triggers the release of RbgA (Matsuo Y, 2007). The later model posits that structural differences exists between “free” 50 S and mature 50 S subunits to explain the differential activation of RbgA by free 50 S subunits.

To address these two conflicting models, a study was undertaken to conduct a thorough biochemical analysis of RbgA and explore its interaction with the ribosome in details. These results demonstrate that the GTPase activity of RbgA was maximally stimulated by both mature and free 50 S subunits, by more than ten times over the stimulation observed with the 45 S intermediate. Interaction assays of RbgA with the different ribosomal subunits show that GDP-bound RbgA does not stably associate with the ribosome and suggests that the GTPase activity of RbgA promotes dissociation from the ribosome. This work discusses the specific findings in the context of two

conflicting models and further clarifies the role of GTPase activity in RbgA function during ribosome assembly.

A 1.2 Material and Methods

A1.2.1 Growth Conditions and Strain Construction

All experiments were performed at 37 °C in Luria-Bertani (LB) medium. When necessary, antibiotics were added at the following concentrations: chloramphenicol (5 μg ml^{-1}), ampicillin (100 μg ml^{-1}). All *B. subtilis* strains used in this study were derived from the wild-type strain JH642 (RB247). *B. subtilis* RB301 (Pspank–*rbgA*) and RB418 (Pspank–*infB*) strains were constructed as previously described (Uicker et al., 2006). RB301 and RB418 accumulate 45 S and 50 S subunits, respectively, when grown in the absence of IPTG and were used to purify 45 S and free 50 S ribosomal subunits. *E. coli* BL21 (DE3) transformed with plasmids containing full-length *rbgA* placed under the control of the T7 promoter was used to overexpress RbgA proteins (Uicker et al., 2006).

A1.2.2 Purification of RbgA proteins

RbgA protein with a histidine tag (His_6) at the C-terminus was isolated as previously described (Uicker et al., 2006). Briefly, *E. coli* BL21(DE3) cells transformed with plasmid containing full-length *rbgA* under IPTG inducible T7 promoter (Uicker et al., 2006) were grown to OD_{600} of 0.5 at 37 °C in LB medium containing 100 μg/ml ampicillin and induced by addition of 1 mM isopropyl *β*-thiogalactoside (IPTG). The cells

were harvested by centrifugation after 3 h and resuspended in a binding buffer (20 mM sodium phosphate, pH 7.5, 0.5 M NaCl and 20 mM imidazole). The cells were lysed by three consecutive passes through a French press at between 1,400 to 1,600 psi and clarified by centrifugation at 16,000 *g* for 20 min. RbgA-His_6 was isolated from the cell lysate by affinity chromatography using HisTrapHP (nickel-nitrilotriacetic acid resin; GE Healthcare) in a Bio-Rad BioLogic LP chromatography system. Cell lysate was injected into the column pre-equilibrated with binding buffer and allowed to equilibrate for 5 min then washed with 5 column volumes of wash buffer (20 mM sodium phosphate, pH 7.5, 0.5 M NaCl and 60 mM imidazole) followed by step elution with binding buffer containing 250 mM imidazole. Elution was monitored by UV absorption at 280 nm and RbgA containing fractions were verified by SDS-polyacrylamide gel electrophoresis (SDS-PAGE), concentrated and desalted by exchanging buffer to desalting buffer (50 mM Tris-HCl pH 7.5, 750 mM KCl, 5 mM $MgCl_2$, 20 mM imidazole, 2 mM DTT and 10% glycerol) and then to storage buffer (50 mM Tris-HCl, pH 7.5, 750 mM KCl, 5 mM $MgCl_2$, 2 mM DTT and 10% glycerol). Purity of isolated RbgA was verified by SDS-PAGE before storage at -20 ºC. K59A, P129R, S134A and F180A mutations were introduced into *rbgA* using a QuikChange IIXL kit (Stratagene).

A1.2.3 Preparation of ribosomal particles

Mature 50 S, free 50 S and 45 S large ribosomal subunits were prepared by sucrose density centrifugation (Charollais et al., 2003; Lin et al., 2004). 50 S and 45 S complexes were isolated from lysates of RB418 and RB301 cells respectively. Mature

50 S subunits were isolated by subjecting the RB247 lysate to low Mg^{2+} buffer (buffer B containing 1 mM Mg^{2+}), thereby dissociating the 70 S ribosome into 50 S and 30 S subunits. The cells were grown to OD_{600} of 0.5 at 37 °C in LB medium with or without IPTG. RB301 and RB418 cells were grown without IPTG for several generations to deplete the cells of RbgA and initiation factor 2, respectively, until a doubling time of ~ 150 min was reached. Chloramphenicol (Sigma-Aldrich) was added to a final concentration of 100 μg ml^{-1} 5 min prior to harvesting. Cells were harvested by centrifugation at 5,000 *g* for 10 min and resuspended in lysis buffer (10 mM Tris-HCl (pH 7.5), 60 mM KCl, 10 mM $MgCl_2$, 0.5% Tween 20, 1 mM DTT, 1x Complete EDTA-free protease inhibitors (Roche Applied Science) and 10 U ml^{-1} RNase-free DNase (Roche Applied Science). Cells were lysed by three consecutive passes through a French press at 1,400 to 1,600 psi, clarified by centrifugation at 16,000 *g* for 20 min. Clarified cell lysates were loaded on top of 10-25% sucrose density gradients equilibrated in buffer B (10 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 50 mM NH_4Cl) and centrifuged using a SureSpin 630 rotor (Sorvall) for 4.5 h at 30,000 rpm. Gradients were then fractionated on Bio-Rad BioLogic LP chromatography system by monitoring UV absorbance at 254 nm. Fractions corresponding to ribosomal subunits of interest were pooled, concentrated using 100 kDa cutoff filters (Millipore), and stored in buffer A (10 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 60 mM KCl and 1 mM DTT) at -80 °C. To purify mature 50S subunits, RB247 cells were grown in LB medium at 37 °C to an OD_{600} of

0.5, harvested, and resuspended in lysis buffer containing 1 mM Mg^{2+} concentration. The lysate was incubated on ice for 30 min and centrifuged for 14 h at 20,600 rpm through a 25-40% sucrose gradient prepared in buffer B containing 1 mM Mg^{2+} (10 mM Tris-HCl, pH 7.6, 1 mM $MgCl_2$, 50 mM NH_4Cl). 50 S subunit peak fractions were pooled, concentrated, and stored in buffer A.

A1.2.4. Characterization of GTPase activity of RbgA

The GTP hydrolysis activity of RbgA was determined by incubating RbgA with GTP for 30 min and measuring the released free phosphate by a Malachite green/ammonium molybdate colorimetric assay (BioAssays Systems) (Lanzetta et al., 1979). Intrinsic GTPase activity was assayed in triplicate at a constant RbgA protein concentration of 2 μM and a range of GTP concentrations (8 to 650 μM) (USB corp.). GTPase reaction was started by addition of protein to GTP solutions at 37 °C in reaction buffer (50 mM Tris-HCl, pH 7.5, 200 mM KCl, 5 mM $MgCl_2$, and 1 mM DTT) for 30 min and terminated by addition of Malachite reagent. It was determined that under these conditions GTP hydrolysis was in the linear range of the assay and less than 10% of the substrate had been consumed. Released phosphate was detected by monitoring color formation at 620 nm using a 96-well plate reader (Tecan Sunrise^{TM}). Experiments were repeated for at least three times and values for K_m and k_{cat} were calculated by fitting the data to the Michaelis-Menten equation with nonlinear regression curve fitting using GraphPad Prism (Graph-Pad Software Inc.; version 5.0).

A1.2.5 Stimulation of GTPase activity of RbgA by ribosomal subunits

To probe the effect of ribosomal subunits on the GTPase activity of RbgA, 100 nM of purified ribosomal subunits (mature 50 S subunits from dissociated 70 S, free 50 S subunits from initiation factor 2 depleted cells, and 45 S subunits from RbgA-depleted cells) were individually incubated with 100 nM RbgA for 15 min at 37 °C in a reaction buffer (50 mM Tris-HCl, pH 7.5, 200 mM KCl, 5 mM $MgCl_2$, and 1 mM DTT) and added to 8-650 μM GTP to start the reaction. It was predetermined that, under these conditions, the values were in a linear range of the assay. The activity of RbgA-subunit complex was determined by standard malachite green assay as described above for RbgA proteins alone. Three biological replicates (independent RbgA and ribosome preparations) were performed, each with three technical replicates.

A1.2.6 Importance of K^+ for GTPase activity of RbgA.

For testing with the GTPase activity of RbgA in the presence of K^+, the protein was purified as described above and stored in storage buffer (50 mM Tris-HCl, pH 7.5, 750 mM KCl, 5 mM $MgCl_2$, 2 mM DTT, 10% glycerol). To test the effects of Na^+, the protein was exchanged and stored in NaCl buffer (50 mM Tris-HCl, pH 7.5, 750 mM NaCl, 5 mM $MgCl_2$, 2 mM DTT and 10% glycerol). The intrinsic activity of RbgA was assayed by incubation of 2 μM of RbgA with 200 μM GTP for 15 min at 37 °C. Stimulation of GTPase activity was tested by incubation of 100 nM of free 50 S subunits

with 100 nM RbgA and 200 μM GTP for 15 min at 37 °C. The activity was determined by standard malachite green assay as described above.

A1.2.7 Homology modeling of the K-loop of RbgA

Homology modeling of the K-loop of RbgA was carried out with MODELLER (Sali et al., 1995) using MnmE as a template. For the remainder of RbgA, the existing crystal structure was used (PDB code 1puj). A total of 20 initial models were constructed and the best model was chosen based on the lowest energy (molpdf) and the lowest discrete objective protein energy (DOPE) function. A ligand-based superimposition of RbgA (Protein Data Bank code-1puj) and MnmE (code 2gj8) was carried out using LigAlign script (Heifets and Lilien) in PyMOL (Schrödinger, Version 1.2r3pre). The figure was generated by using Chimera (Pettersen et al., 2004).

A1.2.8 Interaction between RbgA and ribosomal particles

RbgA protein and ribosomal particles were purified as described above. The RbgA subunit binding assay was performed as previously described (Matsuo et al., 2006). The protocol was modified, however, to include a15-min incubation with a high salt buffer C (10 mM Tris-HCl (pH 7.5), 60 mM KCl, 500 mM NH_4Cl, 10 mM $MgCl_2$ and 1 mM DTT) to further test the specificity of RbgA-subunit interaction. Briefly, 60 pmol of RbgA was preincubated with 1.5 mM of different guanine nucleotides for 15 min followed by the addition of 10 pmol purified subunits. The binding was allowed to proceed for 15 min at 37 °C, followed by centrifugation once through Microcon 100

columns (Millipore) with a cutoff of 100 kDa. The RbgA-subunit complexes were washed once with buffer B and twice with buffer C before elution. Ribosome bound RbgA was detected by separating the complexes on NuPAGE 12% BisTris/SDS gels (Invitrogen), followed by Western blot analysis, using rabbit anti-RbgA and HRP-conjugated goat anti-rabbit polyclonal antibodies. RbgA specific bands were visualized using a Western Lightning chemiluminescent detection system (PerkinElmer Life Sciences).

A1.3 Results

A1.3.1 RbgA has a low intrinsic GTPase activity.

Most ribosome associated-GTPases have low intrinsic GTPase activity in the absence of an effector protein or molecule. To characterize the intrinsic GTPase activity of RbgA, RbgA-His_6 was purified and measured GTP hydrolysis using a malachite green assay. Previously, it was shown that the C-terminal fusion of six histidine residues to RbgA yields a functional protein that can support wild-type growth of *B. subtilis* as well as interaction with the ribosome (Uicker et al., 2006). In addition, a steady state kinetic analysis of RbgA GTPase activity was performed using the RbgA-His_6 protein (Fig. A1.1). The wild-type RbgA had a k_{cat} of 14 h^{-1}, a K_m of 90 μM, and a k_{cat}/K_m of 47 $M^{-1}s^{-1}$ (Table A1.1).

To confirm that the low rate of hydrolysis observed was due to RbgA and not a contaminant, we created three mutants that altered important residues in the P-loop (P129R, S134A) or the G4 region that specifies binding for guanine nucleotides (K59A)

and tested their ability to hydrolyze GTP (Fig. A1.1 and Table A1.1). All three mutants displayed lower GTPase activity and a reduced k_{cat}/K_m, indicating that the GTPase activity observed for wild-type RbgA was not due to a contaminant.

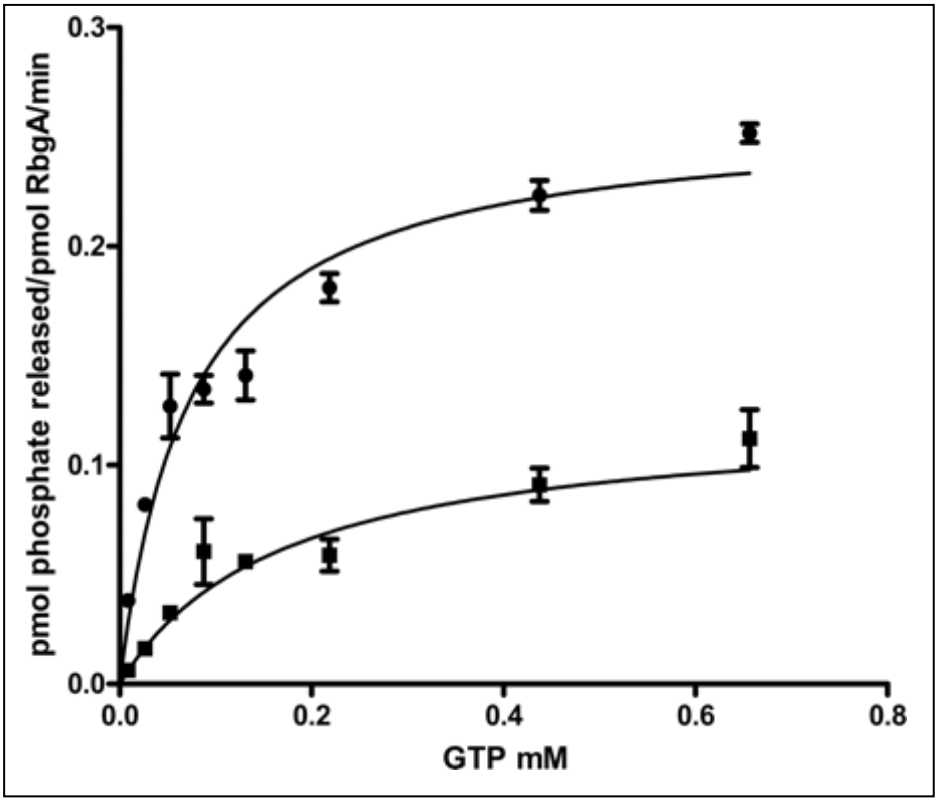


Figure A1.1 Kinetic analysis of GTP hydrolysis rates of RbgA proteins. Representative GTP hydrolysis curves for wild-type RbgA (●) and the P-loop variant S134A (■) are shown.

Table A1.1 Intrinsic GTPase activity of RbgA mutants.

RbgA Proteins	k_{cat} (h^{-1})	K_m (μM)	k_{cat}/K_m (M^{-1} s^{-1})
Wild Type	15 ± 0.4	72 ± 5	58
F180A	18 ± 0.8	65 ± 9	77
P129R	8 ± 0.8	290 ± 58	8
K59A	10 ± 1	286 ± 60	10
S134A	7 ± 0.6	146 ± 32	13

A1.3.2. GTPase activity of RbgA is maximally stimulated by mature or free 50 S ribosomal subunits.

To better understand the role of GTPase activity in the function of RbgA, GTP hydrolysis of RbgA was characterized in the presence of 50 S subunits and 45 S intermediates. Three types of large ribosomal subunits were tested that included: mature 50 S subunits isolated by dissociating subunits from 70 S ribosomes, free 50 S subunits isolated by depleting initiation factor 2 (IF-2) and 45 S intermediates purified from RbgA depleted cells. Previous work indicated that 45 S intermediates and free 50 S subunits had a higher stimulatory activity than mature 50 S subunits (Matsuo Y, 2007). Ribosomal subunits were mixed with RbgA in a ratio of 1:1 and the GTPase activity of RbgA was monitored using a malachite green assay. Both mature and free 50 S subunits were capable of stimulating the GTPase activity of RbgA ~ 55-fold (Fig. A1.2 and Table A1.1).

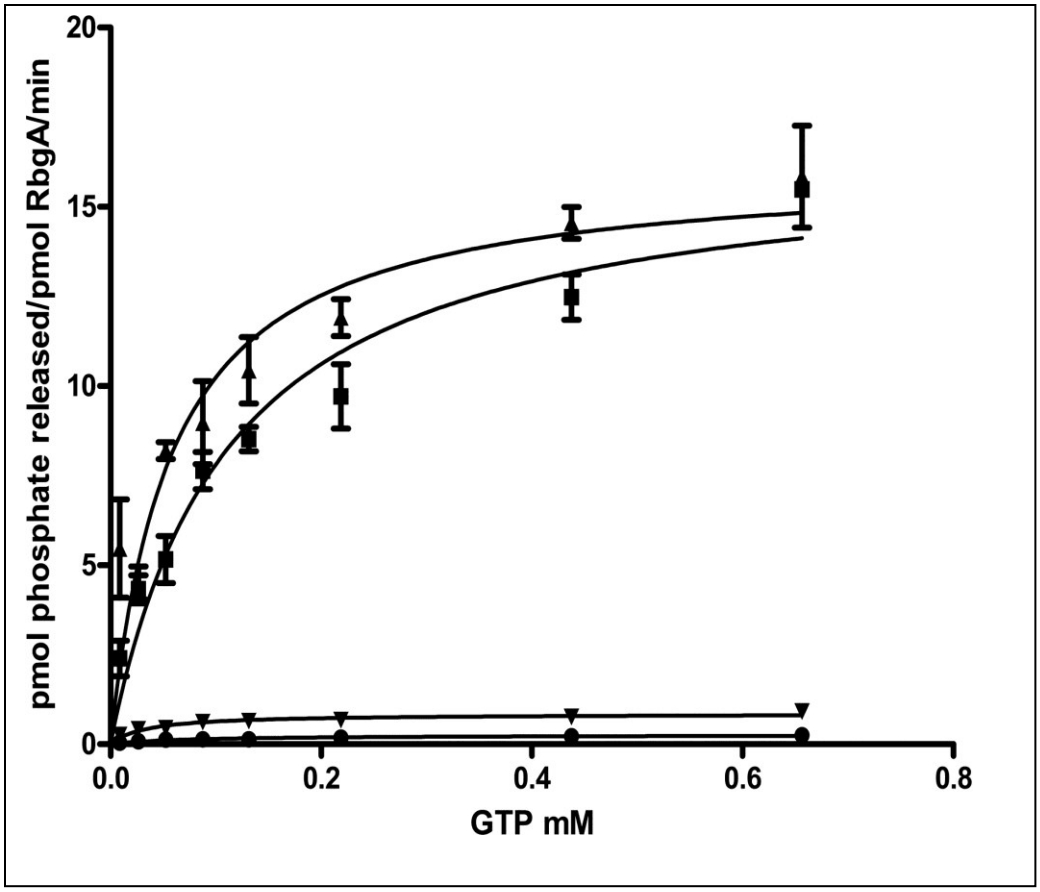


Figure A1.2 Stimulation of GTPase activity of RbgA by ribosomal particles. The representative curves are of GTP hydrolysis rates with reaction mixtures containing RbgA only (●), RbgA and the mature 50 S subunit (■), RbgA and the free 50 S subunit (▲), and RbgA and the 45 S intermediate (▼).

In contrast, the 45 S intermediate isolated from RbgA-depleted cells showed a markedly reduced ability to stimulate the GTPase activity of RbgA (5-fold) over intrinsic GTPase activity. Michaelis-Menten analysis of the GTPase activity of RbgA indicated that an increase in GTPase activity with 50 S subunits was primarily governed by an increase in k_{cat} (Table A1.1). Conversely, interaction with the 45 S intermediate decreased the K_m of RbgA for GTP to 4 μM, indicating a tighter affinity for GTP in the presence of the 45 S complex while k_{cat} was largely unaffected. The results showed that GTPase activity of

RbgA is maximally stimulated by the 50 S subunits (free and mature), in contrast to previously published results (Matsuo Y, 2007).

Table A1.2 Kinetic parameters of RbgA in the presence or absence of ribosomal subunits.

Ribosome	k_{cat} (h^{-1})	K_m (μM)	k_{cat}/K_m (M^{-1} s^{-1})
none	14 ± 2	90 ± 13	43
50S free	755 ± 207	32 ± 6	6533
50S mature	807 ± 158	62 ± 16	3616
45S intermediate	69 ± 16	4 ± 3	4791

A1.3.3 Potassium is required for optimal GTPase activity.

Recent work has indicated that three translation associated GTPases (MnmE, FeoB and YqeH) utilize a unique mechanism for GTP hydrolysis in which a K^+ ion participates in the activation of GTP hydrolysis (Anand et al.; Ash et al.; Scrima and Wittinghofer, 2006). The structures of these GTPases demonstrate that a K^+ ion functions as a GTPase Activating Element (GAE), analogous to GTPase activating proteins (GAP) in eukaryotic GTPases. This K^+ ion occupies a position, usually occupied by an arginine residue, and referred to as “arginine finger” in GAPs of Ras related GTPases (Scheffzek et al., 1997). The positive charge provided by arginine or K^+ at this position stabilizes the transition state by neutralizing negative charges that build up during the transition state. In GTPases that require K^+, the K^+ ion is held in position by a region of the GTP binding domain that has been designated the “K-loop”

(Scrima and Wittinghofer, 2006). The K-loop is located upstream of switch I and interacts with K^+ through the peptide backbone of amino acids and shields the K^+ ion from the bulk solvent. An additional side-chain interaction is provided by a conserved asparagine residue situated within the P-loop.

Inspection of the primary sequence of RbgA indicated a possible K-loop type sequence; however this part of the protein is not resolved in any of the crystal structures of RbgA to date. Therefore, we modeled this K-loop in RbgA using transition state structure of MnmE (Scrima and Wittinghofer, 2006) (Fig. A1.3A). Out of 20 homology models generated, the most energetically favorable model shows a high degree of similarity to the K-loop of MnmE. Moreover, the asparagine residue from P loop (GIPNVGKS in RbgA), that makes a direct contact with the K^+ in other K^+ activated GTPases, is also oriented similar to the P loop asparagine in the RbgA crystal structure (Fig. A1.3B). This suggests that RbgA may also utilize K^+ in the activation of GTPase activity similar to MnmE, FeoB and YqeH.

Next, this study investigated whether K^+ was required for maximal stimulation of RbgA GTPase activity. The intrinsic GTPase activity and stimulation of RbgA GTPase was monitored in the presence of 250 mM NaCl or KCl. The intrinsic GTPase activity of RbgA is reduced 133-fold when NaCl is substituted for KCl, demonstrating that K^+ is required for optimal RbgA GTPase activity (Table A1.2). The stimulation of RbgA GTPase activity by the ribosome is also affected by Na^+ but to a lesser degree, showing

a 22-fold reduction in GTPase activity when co-incubated with 50 S subunits compared to K^+ conditions. These results show that K^+ is required for maximal GTPase activity.

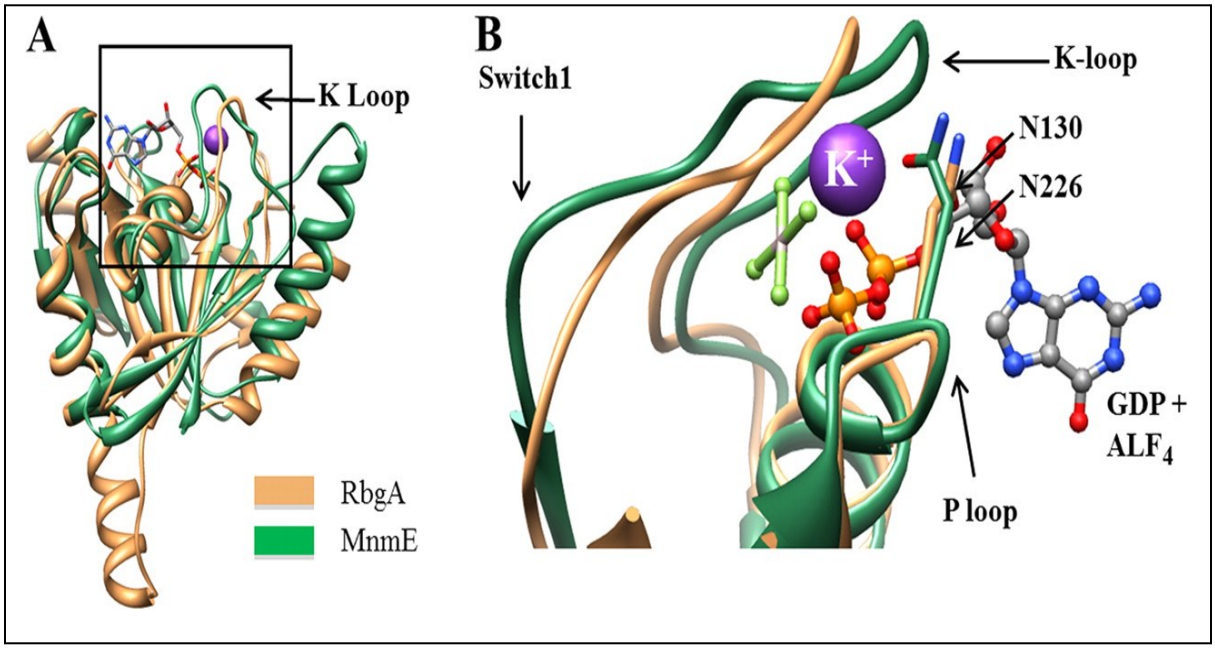


Figure A1.3. Superimposition of MnmE and homology model of RbgA **(***A) and* an enlarged view of the catalytic pocket with GDP and bound potassium is shown in which Asn-130 from the P-loop of RbgA (brown) coordinates the bound K^+ ion (indicated by a purple sphere) in a similar fashion as Asn-226 from the P-loop of MnmE (green) (indicated by arrows) (B).

A 1.3.4 RbgA interacts with the 50 S ribosomes and 45 S intermediates in a nucleotide specific manner.

Previous work showed that RbgA interacts with the large ribosomal subunit when incubated with a non-hydrolyzable analog of GTP (Matsuo et al., 2006). To explore the interaction of RbgA with these subunits in detail, an interaction of RbgA with both purified 50 S subunits and 45 S intermediate was characterized in the presence of GTP,

GDP, and GMPPNP. Saturating levels of guanine nucleotides were used to ensure all RbgA molecules were bound to nucleotide. 45 S intermediates or 50 S subunits were incubated with RbgA for 15 min at 37 °C and were spun through a 100 kDa cutoff microcon filter to remove unbound RbgA. RbgA bound to the subunits was retained on top of the filter and was detected using RbgA specific polyclonal antibodies. It was found that RbgA associates most stably with the 45 S subunit in the GTP and GMPPNP bound forms and that binding is greatly reduced in the presence of GDP (Fig. A1.4).

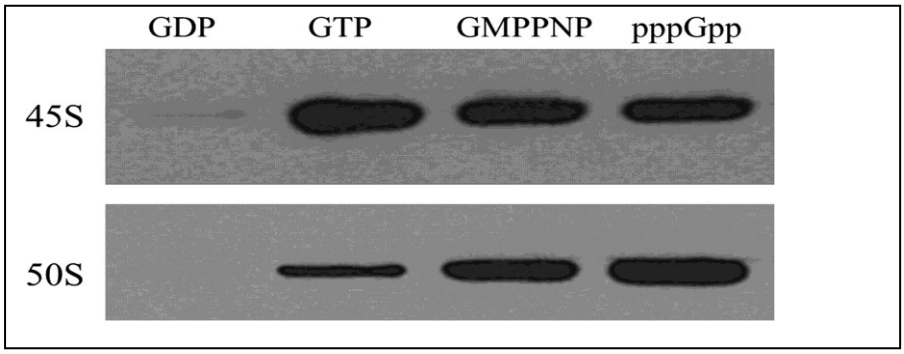


Figure A1.4 Western blot analysis of the interaction between RbgA and ribosome in presence of different guanine nucleotides.

In the presence of the 50 S subunit, RbgA binding is maximally enhanced in the presence of GMPPNP but is greatly reduced in the presence of GTP and GDP. The difference between GTP and GMPPNP is likely due to GTP hydrolysis in the presence of the 50 S ribosome, resulting in dissociation from the ribosome. The alarmone, pppGpp, was also tested and it affected the RbgA interaction with the ribosome. It was

found that the addition of pppGpp enhanced an interaction of RbgA with both the 45 S intermediate and 50 S subunits.

A1.4 DISCUSSION

Previous genetic and biochemical studies aimed at elucidating the role of RbgA in large ribosomal subunit assembly have yielded two conflicting models. One model posits that RbgA couples a final maturation step of the 50 S subunit with the activation of GTP hydrolysis, which signals RbgA to leave the ribosome (Matsuo et al., 2006; Uicker et al., 2006). This maturation step involves the incorporation of ribosomal proteins L16, L27, and L36. However, a revised model was proposed suggesting that RbgA uses GTP hydrolysis to induce a conformational change in the 45 S subunit that is independent of the incorporation of L16, L27, and L36 (Matsuo Y, 2007). The revised model is based upon the finding that the 45 S subunit can maximally stimulate RbgA GTPase activity *in vitro*, although one major complication of the model is that free 50 S subunits, but not mature 50 S subunits, can also stimulate GTP hydrolysis to the same level as the 45 S intermediate. Because both of these models rely on limited biochemical evidence, a detailed analysis of RbgA interaction with the ribosome was undertaken to clarify the role of GTP hydrolysis in RbgA function.

Several lines of evidence presented here support the model in which RbgA utilizes GTPase activity to dissociate from the ribosome. First, GTPase stimulation by 50 S subunits, free or mature, was ten times more than that observed with 45 S intermediates. Second, RbgA interaction with 50 S subunits was nucleotide dependent, with reduced levels of binding observed in the presence of GTP and GDP compared

with the non-hydrolyzable analog GMPPNP. In contrast, the 45 S intermediate displayed a similar binding in the presence of GTP and GMPPNP with greatly reduced binding in the presence of GDP. Decreased binding of RbgA to the 50 S subunit in the presence of GTP, but not GMPPNP, is consistent with GTP hydrolysis governing dissociation of RbgA from the ribosome. Third, a steady state kinetic analysis of RbgA in the presence of the different ribosomal subunits showed that the K_{cat} of RbgA was dramatically increased only in the presence of the 50 S subunit. This indicates that the 50 S, but not the 45 S intermediate, serves a GAP-like function for RbgA. Additionally, K_m was decreased only in presence of 45 S but not 50 S, suggesting a cooperative binding of 45 S and GTP to RbgA. These data support a model in which the GTPase activity of RbgA is required to dissociate RbgA from the 50 S subunit. It is possible that RbgA either serves to monitor correct ribosome formation with its GTPase activity or it couples the hydrolysis of GTP to a conformational change in the subunit while subsequently dissociating from the subunit.

An important question to address is why a maximal stimulation of RbgA GTPase activity was observed in the presence of 50 S subunits (free or mature) while previous work shows maximal stimulation with the 45 S intermediate and free 50 S subunits but not mature subunits (Matsuo Y, 2007); a steady-state kinetic analysis of RbgA indicates a potential answer to this question. The K_m for RbgA in the absence of a ribosome is ~90 μM, and given the slow forward reaction in the absence or presence of the ribosome, the K_m can be a good estimate for the K_d of RbgA for GTP. In the presence

of 50 S subunits the K_m is only marginally reduced (62 μM for mature 50 S and 32 μM for free 50 S subunits). In the presence of the 45 S intermediate the K_m of RbgA is reduced to 4 μM, an indication that interaction with the 45 S intermediate fosters tighter association with GTP. In previous work, it was reported that GTPase assays were performed using a concentration of 10 μM GTP, which was well below the K_m of RbgA for GTP in the presence of the 50 S subunit (Matsuo Y, 2007). These assay conditions would result in most of the RbgA protein being unbound to GTP and should underestimate the activity of RbgA in the presence of 50 S subunits. Since GTP concentration in the cell is in the range of 0.5-5 mM (Lopez et al., 1979; Neidhardt et al., 1990), the assay from this work likely reflects a more physiologically relevant condition. Finally, the results demonstrate that RbgA is a member of a growing family of GTPases that utilize K^+ as part of the mechanism of GTP hydrolysis.

Previous work on the biochemistry of RbgA was performed with NH_4Cl (Matsuo Y, 2007). While NH_4^+ supports an hydrolysis of the GTP in YqeH, it is reduced 2-fold in its activity indicating that NH_4^+ is not sufficient to replace K^+ for full activity (Anand et al.). Future experiments on the biochemistry of RA-GTPases thus warrant investigation of K^+ as a possible GTPase activating element. The results presented here support a model in which the guanine nucleotide state of RbgA regulates the association of RbgA with the ribosome. GDP significantly inhibits RbgA interaction with either the 45 S intermediate or the 50 S subunit, and the interaction with the 50 S subunit is only

stabilized in the presence of non-hydrolyzable analogs of GTP. These results indicate that GTP hydrolysis promotes RbgA dissociation from the 50 S subunit, similar to other translation factor GTPases such as EF-Tu. EF-Tu couples proper accommodation of the aa-tRNA to GTPase activity, resulting in release of EF-Tu from the ribosome (Schmeing et al., 2009).

It is therefore proposed that RbgA promotes a late step in large subunit assembly that allows the stable incorporation of ribosomal proteins L16, L27, and L36. While the nature of this rearrangement is still unknown, once properly achieved, it is proposed that RbgA “senses” the correct assembly of the ribosome which in turn promotes GTP hydrolysis and subsequent RbgA dissociation (refer to figure 5 in the published manuscript). Alternatively, RbgA could act directly on the 45 S intermediate independently of incorporation of L16, L27, or L36. Finally, it was previously proposed that regulating a late assembly process would allow RbgA to act as a checkpoint governing the release of active 50 S subunits into the translation active pool of ribosomes in the cell. Previous work with BipA also showed that pppGpp can regulate GTPase association with ribosomal subunits (deLivron and Robinson, 2008). It is intriguing that alarmone pppGpp enhances association of RbgA with both 45 S and 50 S subunits, given the fact that RbgA likely binds to the subunit interface and would block its association with the 30 S subunit. Under times of translational stress, it is postulated that pppGpp can block the additional maturation of 45 S intermediates and prevent other newly formed 50 S subunits from participating in translation

REFERENCES

REFERENCES

Anand, B., Surana, P., and Prakash, B. Deciphering the catalytic machinery in 30S ribosome assembly GTPase YqeH. PLoS One *5*, e9944.

Ash, M.R., Maher, M.J., Guss, J.M., and Jormakka, M. The initiation of GTP hydrolysis by the G-domain of FeoB: insights from a transition-state complex structure. PLoS One *6*, e23355.

Bassler, J., Grandi, P., Gadal, O., Lessmann, T., Petfalski, E., Tollervey, D., Lechner, J., and Hurt, E. (2001). Identification of a 60S preribosomal particle that is closely linked to nuclear export. Mol Cell *8*, 517-529.

Britton, R.A. (2009). Role of GTPases in bacterial ribosome assembly. Annu Rev Microbiol *63*, 155-176.

Campbell, T.L., Daigle, D.M., and Brown, E.D. (2005). Characterization of the Bacillus subtilis GTPase YloQ and its role in ribosome function. Biochem J *389*, 843-852.

Charollais, J., Pflieger, D., Vinh, J., Dreyfus, M., and Iost, I. (2003). The DEAD-box RNA helicase SrmB is involved in the assembly of 50S ribosomal subunits in Escherichia coli. Mol Microbiol *48*, 1253-1265.

Culver, G.M. (2003). Assembly of the 30S ribosomal subunit. Biopolymers *68*, 234-249.
deLivron, M.A., and Robinson, V.L. (2008). Salmonella enterica serovar Typhimurium BipA exhibits two distinct ribosome binding modes. J Bacteriol *190*, 5944-5952.

Dez, C., and Tollervey, D. (2004). Ribosome synthesis meets the cell cycle. Curr Opin Microbiol *7*, 631-637.

Gollop, N., and March, P.E. (1991). A GTP-binding protein (Era) has an essential role in growth rate and cell cycle control in *Escherichia coli*. J Bacteriol *173*, 2265-2270.

Heifets, A., and Lilien, R.H. LigAlign: flexible ligand-based active site alignment and analysis. J Mol Graph Model *29*, 93-101.

Jensen, B.C., Wang, Q., Kifer, C.T., and Parsons, M. (2003). The NOG1 GTP-binding protein is required for biogenesis of the 60S ribosomal subunit. J Biol Chem *278*, 32204-32211.

Jiang, M., Datta, K., Walker, A., Strahler, J., Bagamasbad, P., Andrews, P.C., and Maddock, J.R. (2006). The Escherichia coli GTPase CgtAE is involved in late steps of large ribosome assembly. J Bacteriol *188*, 6757-6770.

Kallstrom, G., Hedges, J., and Johnson, A. (2003). The putative GTPases Nog1p and Lsg1p are required for 60S ribosomal subunit biogenesis and are localized to the nucleus and cytoplasm, respectively. Mol Cell Biol *23*, 4344-4355.

Lanzetta, P.A., Alvarez, L.J., Reinach, P.S., and Candia, O.A. (1979). An improved assay for nanomole amounts of inorganic phosphate. Anal Biochem *100*, 95-97.

Lin, B., Thayer, D.A., and Maddock, J.R. (2004). The Caulobacter crescentus CgtAC protein cosediments with the free 50S ribosomal subunit. J Bacteriol *186*, 481-489.

Lopez, J.M., Marks, C.L., and Freese, E. (1979). The decrease of guanine nucleotides initiates sporulation of Bacillus subtilis. Biochim Biophys Acta *587*, 238-252.

Matsuo, Y., Morimoto, T., Kuwano, M., Loh, P.C., Oshima, T., and Ogasawara, N. (2006). The GTP-binding protein YlqF participates in the late step of 50 S ribosomal subunit assembly in Bacillus subtilis. J Biol Chem *281*, 8110-8117.

Matsuo Y, O.T., Loh PC, Morimoto T, Ogasawara N (2007). Isolation and characterization of a dominant negative mutant of Bacillus subtilis GTP-binding protein, YlqF, essential for biogenesis and maintenance of the 50 S ribosomal subunit. J Biol Chem *282*, 25270-25277.

Neidhardt, F.C., Ingraham, J.L., and Schaechter, M. (1990). Physiology of the Bacterial Cell: A Molecular Approach (Sinauer Associates).

Nierhaus, K.H. (1991). The assembly of prokaryotic ribosomes. Biochimie *73*, 739-755.
Nomura, M., and Erdmann, V.A. (1970). Reconstitution of 50S ribosomal subunits from dissociated molecular components. Nature *228*, 744-748.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem *25*, 1605-1612.

Sali, A., Potterton, L., Yuan, F., van Vlijmen, H., and Karplus, M. (1995). Evaluation of comparative protein modeling by MODELLER. Proteins *23*, 318-326.

Saveanu, C., Bienvenu, D., Namane, A., Gleizes, P.E., Gas, N., Jacquier, A., and Fromont-Racine, M. (2001). Nog2p, a putative GTPase associated with pre-60S subunits and required for late 60S maturation steps. Embo J *20*, 6475-6484.

Schaefer, L., Uicker, W.C., Wicker-Planquart, C., Foucher, A.E., Jault, J.M., and Britton, R.A. (2006). Multiple GTPases participate in the assembly of the large ribosomal subunit in Bacillus subtilis. J Bacteriol.

Scheffzek, K., Ahmadian, M.R., Kabsch, W., Wiesmuller, L., Lautwein, A., Schmitz, F., and Wittinghofer, A. (1997). The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. Science *277*, 333-338.

Schmeing, T.M., Voorhees, R.M., Kelley, A.C., Gao, Y.G., Murphy, F.V.t., Weir, J.R., and Ramakrishnan, V. (2009). The crystal structure of the ribosome bound to EF-Tu and aminoacyl-tRNA. Science *326*, 688-694.

Schrödinger, L. (Version 1.2r3pre). The PyMOL pp. Molecular Graphics System.
Scrima, A., and Wittinghofer, A. (2006). Dimerisation-dependent GTPase reaction of MnmE: how potassium acts as GTPase-activating element. EMBO J *25*, 2940-2951.

Uicker, W.C., Schaefer, L., and Britton, R.A. (2006). The essential GTPase RbgA (YlqF) is required for 50S ribosome assembly in Bacillus subtilis. Mol Microbiol *59*, 528-540.

Uicker, W.C., Schaefer, L., Koenigsknecht, M., and Britton, R.A. (2007). The essential GTPase YqeH is required for proper ribosome assembly in Bacillus subtilis. J Bacteriol *189*, 2926-2929.

Wilson, D.N., and Nierhaus, K.H. (2007). The weird and wonderful world of bacterial ribosome regulation. Crit Rev Biochem Mol Biol *42*, 187-219.