UNIVERSITY OF CALIFORNIA, SAN DIEGO

Broad Recognition of Group A *Streptococcus* M Protein

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy

in

Chemistry

by

Cosmo Zephyrus Buffalo

Committee in charge:

Professor Partho Ghosh, Chair
Professor Timothy Baker
Professor Victor Nizet
Professor Joseph Noel
Professor Faik Tezcan

2015

The dissertation of Cosmo Zephyrus Buffalo is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2015

DEDICATION

I dedicate this dissertation to my family. Without everyone's love and support through the years, I would not have had the strength to follow my dreams. I love you.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

GAS – group A *Streptococcus*

HVR – hypervariable region

C4BP – C4b binding protein

C4BPα1-2 – C4b binding protein α-domains 1 & 2

MAC – membrane attack complex

CCP – complement control protein

SCR – short consensus repeats

IgG – immunoglobulin G

IgA – immunoglobulin A

HIV-1 – human immunodeficiency virus-1

Env – envelope protein

HA – hemagglutinin

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to begin by acknowledging professor Partho Ghosh for his guidance throughout my graduate career. His mentorship, honesty, and patience have truly been rewarding. As a mentor he allowed me the freedom to pursue my ideas and in the process gave me the motivation to succeed. He taught me that success comes in many forms but most importantly that it comes from being passionate about research. He challenged me, taught me the value of hard work, showed me the rewards of working smart, and passed on the importance of throwing everything, including the kitchen sink, at a problem. It has been and incredible process. This academic pursuit started with a vision, and with Partho's mentorship, that vision became a reality, culminating in a piece of work for which I am truly proud.

I would like to scknowledge my parents, Judith Linetty and Carl Stiefbold, for their unwavering support. At a young age my father, Carl, inspired in me a love of science and I never looked back. For this I incredibly thankful. I also want to acknowledge his powerful guidance and under-rated knowledge when it comes to all things science. I thank my mother, Judith for always making sure I kept it in line, for always telling me to follow my dreams, and giving me an honest perspective on life. I know I would not have been able to accomplish anything without both of my parents' enthusiasm and love. I would like to thank my brother, Om, for being my best friend, my favorite comedian, and for keeping me out of trouble. I would like to thank my sister Abra for teaching me not to take anything too seriously. I would like to thank

my sister Rachel for not teaching me to have a filter. I would like to thank my sister Lauren for teaching me to walk the path that I choose for myself.

I would like to recognize the current and former members of the Partho Ghosh lab for their help and support both inside and outside of lab. One of the most important things to lab productivity is having an environment filled with wonderful and motivated people. The Ghosh lab never disappointed. Of the lab members that made my graduate career such an amazing experience, I would most like to thank Adrian Bahn-Suh for his hard work and dedication to my project. It was a joy to be your mentor and you are a good friend.

I would like to recognize my amazing friends for helping me survive graduate school. I know I could not have gotten through this without your love and good spirits. I am blessed to be able to call you my friends and I appreciate you all very much. Ultimately, when you feel the love as I do, I no longer consider you simply my friends but my family as well.

Academically, I would like to thank my committee members for their suggestions and motivation throughout my graduate career.

Chapter 2, 3 and 4, in full, is material currently being prepared for submission for publication. I am the principle researcher/author on the paper: Broad recognition of group A *Streptococcus* M protein hypervariability by human C4BP. Cosmo Z. Buffalo, Adrian J. Bahn-Suh, Tapan Biswas, Victor N. Nizet, and Partho Ghosh

This work was supported by NIH grant T32 GM007240 (CZB), an AHA Predoctoral Fellowship (CZB), and NIH R01 AI096837 (PG and VN).

VITA

Education

2009-2015	University of California San Diego, San Diego, CA Doctor of Philosophy in Chemistry
2000-2005	Reed College, Portland, OR Bachelor of Arts

Publications (* marks primary authorship)

Feng X, Rodriguez-Contreras D, **Buffalo C**, Bouwer HGA, Kruvand E, Beverley SM, Landfear SM. Amplification of an alternate transporter gene suppresses the avirulent phenotype of glucose transporter null mutants in Leishmania mexicana. Molecular Microbiology. 2009;71:369-81.

Feng X, Feistel T, **Buffalo C**, McCormack A, Kruvand E, Rodriguez-Contreras D, Akopyants NS, Umasankar PK, David L, Jardim A, Beverley SM, Landfear SM. Remodeling of protein and mRNA expression in Leishmania mexicana induced by deletion of glucose transporter genes. Molecular and Biochemical Parasitology. 2011;175:39-48.

Macheboeuf P, **Buffalo C**, Fu C-y, Zinkernagel AS, Cole JN, Johnson JE, Nizet V, Ghosh P. Streptococcal M1 protein constructs a pathological host fibrinogen network. Nature. 2011;472:64-8. doi: 10.1038/nature09967.

Henningham A, Yamaguchi M, Aziz RK, Kuipers K, **Buffalo CZ**, Dahesh S, Choudhury B, Van Vleet J, Yamaguchi Y, Seymour LM, Ben Zakour NL, He L, Smith HV, Grimwood K, Beatson SA, Ghosh P, Walker MJ, Nizet V, Cole JN. Mutual exclusivity of hyaluronan and hyaluronidase in invasive group A Streptococcus. J Biol Chem. 2014;289(46):32303-15. doi: 10.1074/jbc.M114.602847.

Abstracts at Scientific Meetings

Cosmo Buffalo*, Adrian Bahn-Suh, Victor Nizet and Ghosh P. (2015) "Broad Recognition of Group A Streptococcus (GAS) M Protein: Implications for Vaccine Design." UC Systemwide Bioengeneering Symposium, Breakout Session on Infectious Disease Prevention and Therapies (Speaker).

ABSTRACT OF THE DISSERTATION

Broad Recognition of Group A *Streptococcus* M Protein

by

Cosmo Zephyrus Buffalo

Doctor of Philosophy in Chemistry

University of California, San Diego, 2015

Professor Partho Ghosh, Chair

The Gram-positive bacterium *Streptococcus pyogenes,* known as group A *Streptococcus* (GAS), is a widespread human pathogen responsible for numerous disease manifestations, ranging from mild infections to severe invasive diseases and major autoimmune sequelae. At present there is no vaccine against GAS, with one of the major impediments being the antigenic variability of the M protein, one of the primary surface-associated virulence factors of GAS. The M protein elicits protective

opsonizing antibodies, but these antibodies generally target the M protein hypervariable region (HVR) and accordingly the resulting immunity is specific to a particular M protein antigenic variant and does not extend to other M protein antigenic variants. With more than 200 distinct HVRs having been identified, the difficulty in developing a vaccine that offers universal coverage is apparent.

In addition to providing antigenic variation, the M protein inhibits phagocytic uptake and clearance of GAS by neutrophils and macrophages through the recruitment of human C4b-binding protein (C4BP), a negative regulator of the complement system. C4BP is bound by the M protein HVR. In stark contrast to the strict specificity of antibodies, C4BP recognizes a remarkably broad range of M protein HVRs (~88% in one study). This suggests that a structural understanding the C4BP-HVR interaction may be applicable to the design of broadly neutralizing antibodies.

To achieve the structural understanding necessary to address the obstacle of antigenic variability in GAS vaccine design, the co-crystal structures of multiple M protein HVRs ($M2^{HVR}$, $M22^{HVR}$, $M49^{HVR}$, and $M28^{HVR}$) in complex with the (C4BP) domains responsible for binding (C4BPα1-2) were determined. A comparative analysis of these co-crystal structures suggests a conserved C4BP binding mode. Based on the heptad position of interacting M protein residues, two distinct conserved M protein binding motifs were also identified. Structural observations were tested through a series of binding studies of M protein substitution mutants. These results identified M protein HVR residues necessary for C4BP binding as well as give evidence of a remarkable amount of tolerance at the HVR 'reading head'. Overall, this

analysis offers extensive structural insight into how C4BP recognizes a broad range of M protein HVRs, which potentially has direct implications in GAS vaccine design.

Chapter 1:

Introduction

INTRODUCTION

Group A Streptococcus-associated Morbidity and Mortality

The Gram-positive bacterium *Streptococcus pyogenes,* known as group A *Streptococcus* (GAS), is a widespread human pathogen responsible for a spectrum of diseases with a diversity of clinical manifestations, ranging from mild infections (e.g. pharyngitis and impetigo) to severe infections, which includes invasive disease (e.g. streptococcal toxic shock syndrome, bacteremia, cellulitis and necrotizing fasciitis) and major autoimmune sequelae (e.g. rheumatic heart disease, acute rheumatic fever and acute post-streptococcal glomerulonephritis) (Carapetis et al., 2005, Cunningham, 2000). Mild GAS infections have the ability to progress to severe infections. Major autoimmune sequelae typically occurs during the course of prolonged untreated infections. Progression to severe infections primarily affects developing nations where access to treatment and supportive healthcare is limited.

GAS infections are responsible for ~517,000 deaths each year due to severe infections (Carapetis et al., 2005). The prevalence of severe GAS infections is ~18 million cases with ~2 million new cases annually. The greatest disease burden is primarily due to rheumatic heart disease, which represents ~16 million cases, with ~280,000 new cases and 230,000 deaths annually. The burden of invasive disease is also high, with ~650,000 cases and an associated mortality of ~25%, representing ~163,000 deaths annually. In addition, there are ~700 million cases of non-invasive superficial infections annually (Carapetis et al.,

2005). Whereas antibiotic intervention is typically effective in treating non-invasive infections, severe invasive GAS infections may require aggressive supportive care and surgical debridement (Young et al., 2005). Currently, a safe and efficacious vaccine has yet to be developed for commercial use (Cole et al., 2008, Dale et al., 2013a).

Streptococcal M Protein

The M protein, the primary surface-associated virulence factor of GAS, plays a significant role in each of the aforementioned disease processes. This suggests that the M protein is a critical target in the development of effective treatments against GAS virulence, which includes vaccine development. M proteins are dimeric, α-helical, coiled-coils that extend ~500 Å from the GAS cell surface radially in the form of fimbriae (Fischetti, 1989) (Fig. 1.1.1a-b), and function by inhibiting phagocytic uptake and consequent killing of GAS by neutrophils and macrophages (McNamara et al., 2008, Ghosh, 2011, Carlsson et al., 2003). The M protein is synthesized in immature form with an amino-terminal signal sequence and a carboxy-terminal LPxTG motif for processing by sortase. The M protein signal sequence consists of ~40 residues, is well conserved amongst M types, and directs secretion of M protein to the bacterial division septum where it is removed proteolytically by peptidase (Carlsson et al., 2006). Upon secretion across the bacterial membrane, M protein is processed by sortase, which cleaves between the Thr and Gly residues of the LPxTG motif, and through

a transpeptidation reaction, covalently attaches the protein through the Thr to the peptidoglycan (Navarre and Schneewind, 1999).

The dimeric polypeptide chains of mature M protein comprise four blocks (A-D) (Fig. 1.1b), each differing in size and amino acid sequence. Sequence conservation steadily increases across these repeat regions from the N-terminus to the C-terminus of the protein (Fig. 1.1b). The coiled coil of the M protein is characterized by an amino acid heptad periodicity with non-polar residues present at the core *a* and *d* positions (Fig. 1.1c). This heptad periodicity is common of α-helical, coiled-coil proteins, such as myosin and tropomyosin. These latter two proteins are the human targets of GAS cross-reactive antibodies that are responsible for the onset of acute rheumatic fever and rheumatic heart disease. The predicted structure of the C-terminus, anchored in the cell wall, is highly conserved across GAS strains. Conversely, the N-terminus is defined by a hyperveriable (HVR) region, which extends into the environment and is composed of ~50 amino acids (Sandin et al., 2006b, Penfound et al., 2010b) (Fig 1b).

Antigenic variation at the HVR is a key component of GAS M protein's virulence as antibodies elicited from previous GAS exposures are not guaranteed to opsonize from M type to M-type. Such antigenic variation at the HVR also serves as the basis for the Lancefield serological classification of GAS. More than 200 distinct M types HVRs, as well as numerous subtypes, have been identified. The M protein HVR is the target of opsonizing antibodies (Fischetti, 1989, Sandin

et al., 2006a). Opsonizing antibodies evoked by the HVR are specific to the particular M type and offer no protection against GAS strains carrying other M types (Morfeldt et al., 2001). The antigenic variation is that even a single amino acid change is sufficient to cause a major antigenic change at the HVR (Persson et al., 2006). For this reason, individuals are susceptible to GAS serial infections. A common characteristic of the surface-associated proteins of many pathogenic microorganisms, sequence variability represents a major obstacle in the development of vaccines.

GAS M Protein-based Vaccine Design

Existing treatments of GAS include the administration of penicillin, macrolides or intravenous immunoglobulin. Penicillin, typically used to treat existing GAS infections, can also be administered prophylactically to prevent recurring GAS infections and subsequent sequelae in susceptible populations. Despite such efforts to control reccurring infections, GAS auto-immune sequelae continue to persist at endemic levels in developing countries (Carapetis et al., 2005). Penicillin treatment is not without shortfalls. Even without a documented case of GAS resistance to penicillin, studies suggest a 20-40% failure rate of penicillin against GAS pharyngitis (Pichichero and Casey, 2007). Macrolides are utilized as an alternative to penicillin, typically in the case of a penicillin allergy. However, there is concern that resistance may evolve and spread (Michos et al., 2009). Immunoglobulin is administered typically in the case of severe invasive

disease, such as streptococcal toxic shock syndrome and necrotizing fasciitis. Immunoglobulin is recommended as an adjunctive therapy as it neutralizes superantigens and promotes opsonophagocytosis (Pandey et al., 2009). Unfortunately, intravenous immunoglobulin treatment only offers short-term protection since no immunological memory is generated. While in most cases of GAS infection, penicillin or alternative antibiotics are effective in treating superficial GAS infections, the majority of GAS infections occur in geographical regions where access to such treatments is limited. A vaccine that is protective against all M types may be the only way to control or eliminate GAS disease globally.

GAS vaccine design is challenging. Concerning GAS vaccine development, many other obstacles exist in addition to sequence variability of the M protein HVR. Obstacles such as the complexity of the global epidemiology of GAS infections, the limited number of antigens that can be included in combination vaccines, the lack of a reliable animal model that mimics human disease, and the issues of autoimmunity and vaccine safety (Dale et al., 2013a) all need to be overcome for the eventual development of a GAS vaccine. An effective GAS vaccine is one that contains epitopes that represent all GAS strains, is highly immunogenic, and is a potent inducer of immune memory without cross-reacting with human tissue, and induces the production of both serum immunoglobulin G (IgG) and mucosal immunoglobulin A (IgA). Such a vaccine would prevent pharyngeal colonization, carriage, symptomatic and asymptomatic

infection, as well as impetigo, invasive disease, acute rheumatic fever, rheumatic heart disease, and acute post-streptococcal glomerulonephritis across a diversity of geographical regions. The diversity of M types in the global epidemiology of GAS infections makes vaccine design challenging, as prevalent M types differ greatly between not only developed and underdeveloped regions, but inter-regionally as well. Such M type diversity combined with the varying degrees of health care access and insufficient disease documentation makes design of a universal vaccine difficult. However, by addressing the many obstacles to GAS vaccine design, vaccine candidates have begun to emerge that are beginning to approach the level of protective coverage necessary for a commercially available vaccine.

Several GAS vaccine candidates are in various stages of clinical and pre-clinical trials. Given that the GAS virulence factor M protein is the major surface exposed antigen of GAS, M protein-based vaccines have been the subject of research and development for decades. Due to the antigenic variability at the M protein HVR, vaccine candidates have primarily come in two different forms: multivalent N-terminal (HVR) candidates and conserved region candidates (C-repeat region). Significant progress has been made recently in the design and clinical development of extensive multivalent M protein-based vaccines (Hu et al., 2002, McNeil et al., 2005, Dale et al., 2011) These N-terminal vaccine candidates have been shown to be opsonic, bactericidal, and protective. M Type specific vaccine candidates consist of the N-terminal HVRs of M proteins from

multiple different GAS M types fused together in tandem to form larger vaccine polypeptides. Based on direct surveillance data from the USA and Europe, the most current 30-valent M protein HVR-based vaccine candidate (Dale et al., 2011) contains protective M protein peptides from serotypes of GAS that covers 98% of all cases of pharyngitis in the US and Canada (Shulman et al., 2009), 90% of invasive disease in the US (O'Loughlin et al., 2007), and 78% of invasive disease in Europe (Luca-Harari et al., 2009). Pre-clinical studies have shown that the 30-valent HVR based vaccine candidate evokes antibodies against all 30 serotypes as well as provides cross-protection against some additional M types not included in the vaccine, albeit for reasons that are not entirely clear (Dale et al., 2011, Dale et al., 2013b). However, due to the scope of variability at the HVR and the complexity of the global epidemiology of GAS infections, the level of protection for the 30-valent vaccine candidate must still be improved upon.

An alternative strategy to multivalent M HVR vaccines in GAS vaccine design are vaccine candidates that contain conserved GAS antigens. These antigens are commonly shared by many or all M serotypes. Not surprisingly, these include the conserved regions of the M protein (Batzloff et al., 2004, Bessen and Fischetti, 1988, Bronze et al., 1992, Bauer et al., 2012) as well as a range of other protein antigens (Sabharwal et al., 2006, Dale et al., 1999, Kawabata et al., 2001, Courtney et al., 2003, Kapur et al., 1994, McCormick et al., 2000, Zingaretti et al., 2010, Liu et al., 2007, McMillan et al., 2004a, McMillan et al., 2004b, Lei et al., 2004). The M protein conserved region has been the second most developed

vaccine candidate behind the multivalent M HVR vaccines and has been shown to evoke cross-protective immunity against multiple serotypes (Bessen and Fischetti, 1988, Bronze et al., 1992, Batzloff et al., 2004). The conserved C-repeat M protein vaccine candidates J8 (12 amino acid minimal B cell epitope) and its parent peptide, J14, are the most recent M protein conserved region based vaccines (Batzloff et al., 2004). Though protective, these vaccine candidates have been shown to not be as protective as the multivalent M protein vaccines, as even the addition of J14 to the type specific multivalent vaccine did not enhance the overall level of protection of the 30-valent vaccine (Penfound et al., 2010a). Despite this result, if an appropriate conserved antigen can be identified it remains a promising approach to combine serotype specific and conserved antigens into a vaccine that would offer broad global coverage. It is the aim of this dissertation to address the lack of an effective conserved antigen by offering an alternative approach to vaccine development that takes advantage of the conserved binding of a host factor to the HVR despite the sequence variability seen between different M types.

GAS Resistance to Complement-mediated Opsonophagocytosis

In the absence of a vaccine against GAS and without access to antibiotics, the outcome of streptococcal infections is determined by the status of the host's. Once GAS colonizes its exclusively human host, the pathogen's survival is dependent on its ability to repel numerous immune defense mechanisms,

including complement and consequent phagocytosis. The complement system, part of the innate immune system, provides protection against pathogens without previous exposure or immunization and is one of the first lines of defense against bacterial colonization. The complement system comprises 30 soluble proteins and several membrane-associated complement receptors and inhibitors. In order to impair the complement-mediated phagocytic clearance of GAS, the M protein recruits specific host factors to the GAS cell surface that interfere with the deposition of opsonic antibodies and the activation of complement (Morfeldt et al., 2001, Berggard et al., 2001).

The key negative regulator of the complement system, C4b-binding protein (C4BP), is one of the most common of these recruited host factors (Persson et al., 2006). C4BP binding to the GAS cell surface results in diminished activation of phagocytes and diminished clearance of GAS, which may lead to the host being unable to limit the infection. The complement system cascade can be activated via three major routes: the classical, the alternative, and the lectin pathway. The classical pathway is initiated by the binding of antibodies to the bacterial cell surface; the lectin pathway is initiated by the binding of lectins to specific carbohydrate structures; and the alternate pathway is initiated by a 'tickover' mechanism (Ricklin et al., 2010). All three pathways converge at the level of C3 deposition, which is then hydrolyzed by C3 convertase (C4bC2a). Formation of C3 convertase generates chemoattractant anaphylatoxins and further amplifies deposition of C3 fragments on microbes, which opsonizes the microbial

target for phagocytosis. Complement-mediated formation of the lytic membrane attack complex (MAC) may result in direct lysis of gram-negative bacteria. Alternatively, gram-positive bacteria such as GAS are resistant to MAC-mediated cell lysis and are instead eliminated by phagocytes following opsonization with C3b and iC3b. C4BP is a negative regulatory protein of both the classical and lectin complement pathways.

Situated into a closely associated group of complement control regulators, the role of C4BP is to tightly regulate complement and to protect host tissue from complement-mediated destruction. Like most complement inhibitors, C4BP inhibits the C3 convertase of the complement system, preventing the proteolytic production of the major opsonin C3b by C4bC2a. C4BP does this by directly binding C4b, thereby competitively displacing C2a from the C4bC2a complex and thus increasing the rate of C4bC2a dissociation. C4BP is also a cofactor for complement factor I, a serine protease that cleaves and inactivates C4b (Gigli et al., 1979). C4BP recruited to the GAS surface by the M protein is still capable of regulating complement and causes a significant decrease in the production of C3b, inhibiting complement activation and ultimately diminishing opsonophagocytic killing of GAS (Blom et al., 2004, Thern et al., 1995, Accardo et al., 1996, Lindahl et al., 2000, Carlsson et al., 2003) (Fig.1). Additionally, C4BP recruitment by M protein also competes with the deposition of protective, opsonic antibodies on the GAS cell surface (Berggard et al., 2001) (Fig. 1.1). Recent *in vivo* findings using humanized C4BP transgenic mice have served to emphasize

the importance of binding complement inhibitors to GAS in the impairment of opsonophagocytosis (Ermert et al., 2015), which translates to enhanced virulence in a humanized whole animal model. Such a model may eventually prove invaluable in developing vaccines and therapeutics that rely on human complement activation (Ermert et al., 2015).

Recruitment of C4BP by the M Protein HVR

C4BP is a soluble glycoprotein (~200 mg/liter in plasma) (Dahlback et al., 1983). C4BP is the only circulating complement inhibitor with a oligomeric structure. The predominant isoform consists of seven α chains disulfide-bonded to a single β chain. Like many complement inhibitors, each of these chains is composed of multiple ~60-residue complement control protein (CCP) domains, also know as short consensus repeats (SCRs) or sushi domains (Barlow et al., 1993, Jenkins et al., 2006). C4BP is unique from other complement regulators in that it circulates in complex with vitamin-K-dependent protein S, providing the C4BP with the capacity to interact with negatively charged phospholipid membranes. Transmission electron microscopy indicated that up to six molecules of C4b bind to one molecule of C4BP at the distal (N-terminal) ends of the α chains (Ram et al., 2001). The first two CCP domains of the α chain (C4BPα1-2) are required to bind M protein HVRs (Accardo et al., 1996, Jenkins et al., 2006). These domains also interact with C4b (Blom et al., 2001, Blom et al., 1999), and evidence suggests, that upon binding to both C4b and M protein, C4BPα1-2

undergoes an intermodular reorientation (Jenkins et al., 2006, Blom et al., 2000). The M protein HVR and C4b binding sites are overlapping but not identical and show different modes of binding based on different sensitivities to salt concentration (Blom et al., 2000). Escape from complement attack by hijacking C4BP is emerging as a widespread immune evasion strategy, as many other pathogens also share the ability to sequester C4BP (Berggård et al., 2001, Jarva et al., 2005, Adams et al., 2010, Nordstrom et al., 2004, Prasadarao et al., 2002, Ram et al., 2001). Some of these pathogens, such as *Bordetella pertussis* and *Neisseria gonorrhoeae,* bind to the same portion of C4BP as the M protein (Berggård et al., 2001, Ram et al., 2001). The binding interaction between M protein HVR and a single C4BPα1-2 has a relatively modest binding affinity, with a K_d of 0.5 μM being reported for the $M4^{HVR}$-C4BP interaction (Jenkins et al., 2006). However, given that C4BP is oligomeric with seven α chains, the binding affinity of intact C4BP to the bacterial cell surface is much higher due to avidity, with a K_d in the picomolar range (Sanderson-Smith et al., 2014). Unlike C4BP binding to C4b, C4BP binding to M protein shows no dependence on pH or salt concentration (Blom et al., 2000), suggesting that forces other than electrostatics are critical in stabilizing the complex.

C4BP binding has been mapped to the antigenic hypervariable region (HVR) of the M protein, the same M protein region responsible for eliciting type specific protective antibodies. In stark contrast to the binding mode of antibodies, C4BP has the ability to bind the HVR of a remarkably broad variety of M protein

types, representing ~88% of M types tested (Persson et al., 2006) (Fig. 2A). This binding is broadly specific, as not all M protein HVRs are seen to be bound by C4BP (Persson et al., 2006). Despite this conservation of C4BP binding, there is no apparent sequence similarity amongst the many C4BP binding HVRs (Persson et al., 2006), even though these HVRs have been found to bind the same site on C4BP (Morfeldt et al., 2001). M protein sequence variation is extensive and represents an apparent paradox because the M protein must retain the ability to specifically bind C4BP despite such variability. In order to achieve antigenic variation most proteins have a limited number of conserved residues to confer specific function while other residues vary significantly (Shakhnovich et al., 1996). In the case of M the protein, more than 200 variants have been identified for the ~50-residue HVR. This number is small compared to the total number of possible sequence variants. This suggests that, because antibodies that prevent binding of C4BP promote phagocytosis (Berggard et al., 2001, Carlsson et al., 2003), the limited sequence diversity among C4BP binding M-types may represent fitness selection during the evolution of antigenic variants that maintain the ability to bind C4BP.

Specific Aims

It has been proposed that all HVRs bind to the same region in C4BP and probably having similar structures (André et al., 2006, McNamara et al., 2008). However, until this present study, the molecular details and the basis for broad

specificity in the interaction between the M protein HVRs and C4BP remained undetermined. It was the goal of this dissertation to establish a molecular understanding of the HVR-C4BP interaction through co-crystal structure determination of multiple M protein HVRs in complex with the C4BP. A direct comparison of the different co-crystal structures will not only offer valuable insight into this pervasive bacterial virulence mechanism but also elucidate the mechanism of broad recognition of the M protein HVR by C4BP. This understanding of the broad recognition of the M protein HVR by C4BP may be applicable to the eventual development of therapeutic antibodies directed against the large set of C4BP-binding GAS M types. Based solely on the number of M protein HVRs that bind C4BP (~88% of those tested), a vaccine targeting the N-terminal HVR in such a broad manner as C4BP has the potential to offer a greater level of protective coverage than the multivalent and conserved region vaccine candidates currently in development.

C4BP-M^{HVR} interaction represents broad vaccine specificity that is lacking in a purely multivalent vaccine. In an attempt to address the obstacle of antigenic variability and offer an alternative approach to GAS vaccine design, I succeeded in determining the co-crystal structure of multiple M protein HVRs ($M2^{HVR}$, $M49^{HVR}$, $M22^{HVR}$, $M28^{HVR}$) in complex with the C4BP region (C4BPα1-2) responsible for broad recognition. A comparative analysis of these co-crystal structures suggests a conserved C4BP binding mode. A conserved 'quadrilateral' of binding interactions was identified at the C4BPα2 domain and composed of *(1)*

a hydrophobic pocket (C4BP His67, Ile78, and Leu82); *(2)* a hydrogen bonding group (mainchain nitrogen of C4BP His67); and two positively charged residues, C4BP *(3)* Arg64 and *(4)* Arg66. The HVR's contribute residues that form complementary quadrilaterals. A "hydrophobic nook' created by C4BP main chain atoms and the alkyl carbons of Arg39 is the major contact point on the C4BPα1 domain. Together these major points of contact represent the HVR 'reading head' of C4BP. To verify the proposed binding mode, I performed a series of alanine substitution mutations on the $M2^{HVR}$. These mutants were then evaluated for C4BP binding by coprecipitation assays. This analysis determined residues essential for binding, as well as identified alanine substitution mutations that showed an increase in binding capability. Surprising, the binding site is conformable, as a significant amount of tolerance between binding partners was identified. This suggests why C4BP is able to bind to so many HVRs with such broad specificity. In addition, based on the heptad position of M protein interacting residues within the coiled coil, I have identified two distinct conserved M protein binding modes (the M2/M49 mode and the M22/M28 Mode) responsible for binding C4BP. The heptad positions of the interacting residues are not identical between the two modes. Combined with the tolerance of the C4BP 'reading head' to recognize different residue binding partners, this fact helps to explain why HVR sequence similarity of C4BP binding M-types is not initially apparent from a pure sequence analysis standpoint (Persson et al., 2006). Overall,

our analysis offers extensive structural insight into how C4BP recognizes such a broad range of M protein HVRs.

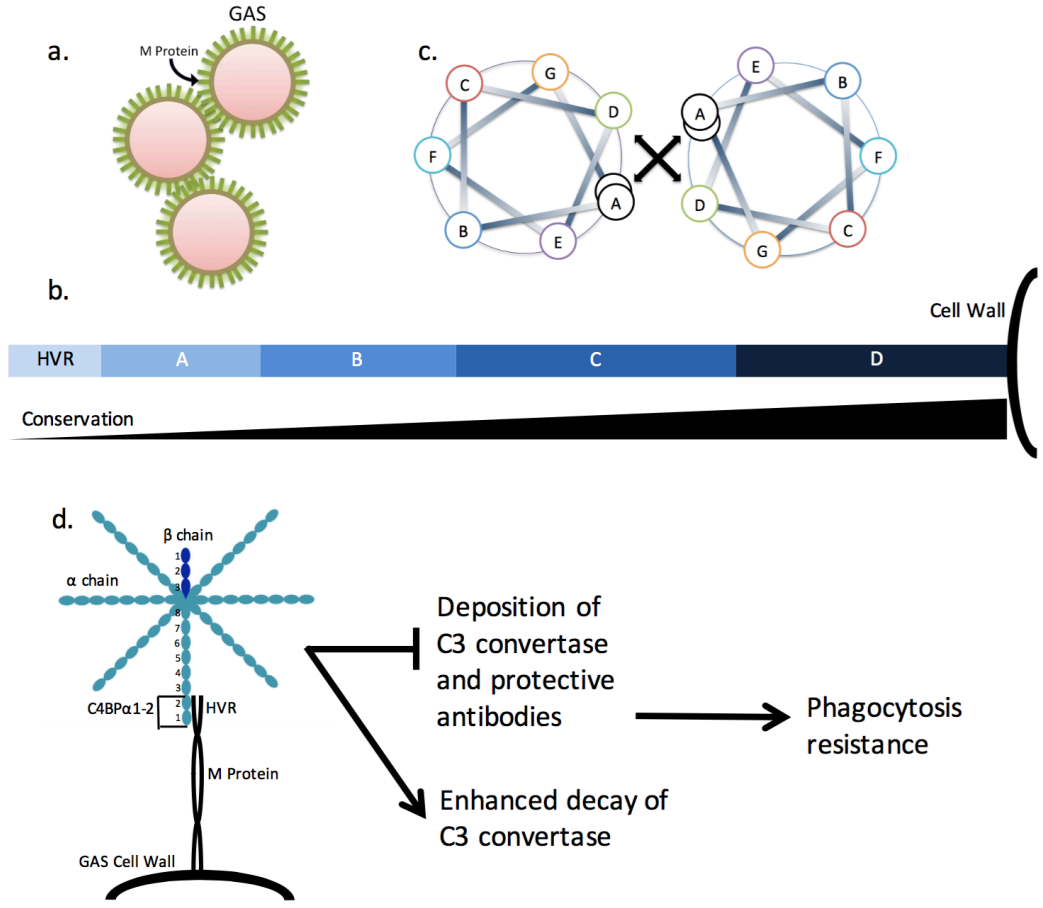


Figure 1.1. Schematic representation of GAS M protein and its binding to C4BP.

a. Schematic representation of whole cell GAS and the M protein extending radially from the GAS cell surface. **b.** Repeat regions of the GAS M protein as anchored to the GAS cell surface. Scale shows increase in conservation along the dimeric coiled-coil from the amino terminus to the carboxy terminus of the protein. **c.** Heptad repeat positions for the dimeric coiled coil of GAS M protein. Residues at the core *a* and *d* positions interact with one another typically through hydrophobic interactions. **d.** Schematic representation of the binding of human C4BP to the HVR of an M protein. The ~570 KDa C4BP molecule, which is composed of seven identical α chains and one β chain, binds to the surface-exposed amino-terminal HVR of the M protein dimeric coiled coil.

REFERENCES

ACCARDO, P., SANCHEZ-CORRAL, P., CRIADO, O., GARCIA, E. & RODRIGUEZ DE CORDOBA, S. 1996. Binding of human complement component C4b-binding protein (C4BP) to Streptococcus pyogenes involves the C4b-binding site. *J Immunol,* 157**,** 4935-9.

ADAMS, P. D., AFONINE, P. V., BUNKÓCZI, G., CHEN, V. B., DAVIS, I. W., ECHOLS, N., HEADD, J. J., HUNG, L. W., KAPRAL, G. J., GROSSE-KUNSTLEVE, R. W., MCCOY, A. J., MORIARTY, N. W., OEFFNER, R., READ, R. J., RICHARDSON, D. C., RICHARDSON, J. S., TERWILLIGER, T. C. & ZWART, P. H. 2010. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography,* 66**,** 213-221.

ANDRÉ, I., PERSSON, J., BLOM, A. M., NILSSON, H., DRAKENBERG, T., LINDAHL, G. & LINSE, S. 2006. Streptococcal M protein: Structural studies of the hypervariable region, free and bound to human C4BP. *Biochemistry,* 45**,** 4559-4568.

BARLOW, P. N., STEINKASSERER, A., NORMAN, D. G., KIEFFER, B., WILES, A. P., SIM, R. B. & CAMPBELL, I. D. 1993. Solution structure of a pair of complement modules by nuclear magnetic resonance. *J Mol Biol,* 232**,** 268-84.

BATZLOFF, M., YAN, H., DAVIES, M., HARTAS, J. & GOOD, M. 2004. Preclinical evaluation of a vaccine based on conserved region of M protein that prevents group A streptococcal infection. *Indian J Med Res,* 119 Suppl**,** 104-7.

BAUER, M. J., GEORGOUSAKIS, M. M., VU, T., HENNINGHAM, A., HOFMANN, A., RETTEL, M., HAFNER, L. M., SRIPRAKASH, K. S. & MCMILLAN, D. J. 2012. Evaluation of novel Streptococcus pyogenes vaccine candidates incorporating multiple conserved sequences from the C-repeat region of the M-protein. *Vaccine,* 30**,** 2197-205.

BERGGARD, K., JOHNSSON, E., MORFELDT, E., PERSSON, J., STALHAMMAR-CARLEMALM, M. & LINDAHL, G. 2001. Binding of human C4BP to the hypervariable region of M protein: a molecular mechanism of phagocytosis resistance in Streptococcus pyogenes. *Molecular Microbiology,* 42**,** 539-551.

BERGGÅRD, K., LINDAHL, G., DAHLBÄCK, B. & BLOM, A. M. 2001. Bordetella pertussis binds to human C4b-binding protein (C4BP) at a site similar to that used by the natural ligand C4b. *European Journal of Immunology,* 31**,** 2771-2780.

BESSEN, D. & FISCHETTI, V. A. 1988. Influence of intranasal immunization with synthetic peptides corresponding to conserved epitopes of M protein on mucosal colonization by group A streptococci. *Infect Immun,* 56**,** 2666-72.

BLOM, A. M., BERGGÂRD, K., WEBB, J. H., LINDAHL, G., VILLOUTREIX, B. O. & DAHLBÄCK, B. 2000. Human C4b-binding protein has overlapping, but not identical, binding sites for C4b and streptococcal M proteins. *Journal of immunology (Baltimore, Md. : 1950),* 164**,** 5328-5336.

BLOM, A. M., KASK, L. & DAHLBÄCK, B. 2001. Structural Requirements for the Complement Regulatory Activities of C4BP. *Journal of Biological Chemistry,* 276**,** 27136-27144.

BLOM, A. M., VILLOUTREIX, B. O. & DAHLBACK, B. 2004. Complement inhibitor C4b-binding protein-friend or foe in the innate immune system? *Mol Immunol,* 40**,** 1333-46.

BLOM, A. M., WEBB, J., VILLOUTREIX, B. O. & DAHLBACK, B. 1999. A cluster of positively charged amino acids in the C4BP alpha-chain is crucial for C4b binding and factor I cofactor function. *J Biol Chem,* 274**,** 19237-45.

BRONZE, M. S., COURTNEY, H. S. & DALE, J. B. 1992. Epitopes of group A streptococcal M protein that evoke cross-protective local immune responses. *J Immunol,* 148**,** 888-93.

CARAPETIS, J. R., STEER, A. C., MULHOLLAND, E. K. & WEBER, M. 2005. The global burden of group A streptococcal diseases. *Lancet Infectious Diseases.*

CARLSSON, F., BERGGÅRD, K., STÅLHAMMAR-CARLEMALM, M. & LINDAHL, G. 2003. Evasion of phagocytosis through cooperation between two ligand-binding regions in Streptococcus pyogenes M protein. *The Journal of experimental medicine,* 198**,** 1057-1068.

CARLSSON, F., STALHAMMAR-CARLEMALM, M., FLARDH, K., SANDIN, C., CARLEMALM, E. & LINDAHL, G. 2006. Signal sequence directs localized secretion of bacterial surface proteins. *Nature,* 442**,** 943-6.

COLE, J. N., HENNINGHAM, A., GILLEN, C. M., RAMACHANDRAN, V. & WALKER, M. J. 2008. Human pathogenic streptococcal proteomics and vaccine development. *Proteomics Clin Appl,* 2**,** 387-410.

COURTNEY, H. S., HASTY, D. L. & DALE, J. B. 2003. Serum opacity factor (SOF) of Streptococcus pyogenes evokes antibodies that opsonize homologous and heterologous SOF-positive serotypes of group A streptococci. *Infect Immun,* 71**,** 5097-103.

CUNNINGHAM, M. W. 2000. Pathogenesis of group A streptococcal infections. *Clinical microbiology reviews,* 13**,** 470-511.

DAHLBACK, B., SMITH, C. A. & MULLER-EBERHARD, H. J. 1983. Visualization of human C4b-binding protein and its complexes with vitamin K-dependent protein S and complement protein C4b. *Proc Natl Acad Sci U S A,* 80**,** 3461-5.

DALE, J. B., CHIANG, E. Y., LIU, S., COURTNEY, H. S. & HASTY, D. L. 1999. New protective antigen of group A streptococci. *J Clin Invest,* 103**,** 1261-8.

DALE, J. B., FISCHETTI, V. A., CARAPETIS, J. R., STEER, A. C., SOW, S., KUMAR, R., MAYOSI, B. M., RUBIN, F. A., MULHOLLAND, K., HOMBACH, J. M., SCHÖDEL, F. & HENAO-RESTREPO, A. M. 2013a. Group A streptococcal vaccines: Paving a path for accelerated development. *Vaccine.*

DALE, J. B., PENFOUND, T. A., CHIANG, E. Y. & WALTON, W. J. 2011. New 30-valent M protein-based vaccine evokes cross-opsonic antibodies against non-vaccine serotypes of group A streptococci. *Vaccine,* 29**,** 8175-8178.

DALE, J. B., PENFOUND, T. A., TAMBOURA, B., SOW, S. O., NATARO, J. P., TAPIA, M. & KOTLOFF, K. L. 2013b. Potential coverage of a multivalent M protein-based group A streptococcal vaccine. *Vaccine,* 31**,** 1576-1581.

ERMERT, D., SHAUGHNESSY, J., JOERIS, T., KAPLAN, J., PANG, C. J., KURT-JONES, E. A., RICE, P. A., RAM, S. & BLOM, A. M. 2015. Virulence of Group A Streptococci Is Enhanced by Human Complement Inhibitors. *PLoS Pathog,* 11**,** e1005043.

FISCHETTI, V. A. 1989. Streptococcal M protein: molecular design and biological behavior. *Clin Microbiol Rev,* 2**,** 285-314.

GHOSH, P. 2011. The nonideal coiled coil of M protein and its multifarious functions in pathogenesis. *Advances in Experimental Medicine and Biology,* 715**,** 197-211.

GIGLI, I., FUJITA, T. & NUSSENZWEIG, V. 1979. Modulation of the classical pathway C3 convertase by plasma proteins C4 binding protein and C3b inactivator. *Proc Natl Acad Sci U S A,* 76**,** 6596-600.

HENNINGHAM, A., GILLEN, C. M. & WALKER, M. J. 2013. Group a streptococcal vaccine candidates: potential for the development of a human vaccine. *Curr Top Microbiol Immunol,* 368**,** 207-42.

HU, M. C., WALLS, M. A., STROOP, S. D., REDDISH, M. A., BEALL, B. & DALE, J. B. 2002. Immunogenicity of a 26-valent group A streptococcal vaccine. *Infect Immun,* 70**,** 2171-7.

JARVA, H., RAM, S., VOGEL, U., BLOM, A. M. & MERI, S. 2005. Binding of the complement inhibitor C4bp to serogroup B Neisseria meningitidis. *J Immunol,* 174**,** 6299-307.

JENKINS, H. T., MARK, L., BALL, G., PERSSON, J., LINDAHL, G., UHRIN, D., BLOM, A. M. & BARLOW, P. N. 2006. Human C4b-binding protein, structural basis for interaction with streptococcal M protein, a major bacterial virulence factor. *The Journal of biological chemistry,* 281**,** 3690-3697.

JOHNSSON, E., BERGGÅRD, K., KOTARSKY, H., HELLWAGE, J., ZIPFEL, P. F., SJÖBRING, U. & LINDAHL, G. 1998. Role of the hypervariable region in streptococcal M proteins: binding of a human complement inhibitor. *Journal of immunology (Baltimore, Md. : 1950),* 161**,** 4894-4901.

JOHNSSON, E., THERN, A., DAHLBÄCK, B., HEDÉN, L. O., WIKSTRÖM, M. & LINDAHL, G. 1996. A highly variable region in members of the streptococcal M protein family binds the human complement regulator C4BP. *Journal of immunology (Baltimore, Md. : 1950),* 157**,** 3021-3029.

KAPUR, V., MAFFEI, J. T., GREER, R. S., LI, L. L., ADAMS, G. J. & MUSSER, J. M. 1994. Vaccination with streptococcal extracellular cysteine protease (interleukin-1 beta convertase) protects mice against challenge with heterologous group A streptococci. *Microb Pathog,* 16**,** 443-50.

KAWABATA, S., KUNITOMO, E., TERAO, Y., NAKAGAWA, I., KIKUCHI, K., TOTSUKA, K. & HAMADA, S. 2001. Systemic and mucosal immunizations with fibronectin-binding protein FBP54 induce protective immune responses against Streptococcus pyogenes challenge in mice. *Infect Immun,* 69**,** 924-30.

LEI, B., LIU, M., CHESNEY, G. L. & MUSSER, J. M. 2004. Identification of new candidate vaccine antigens made by Streptococcus pyogenes: purification and characterization of 16 putative extracellular lipoproteins. *J Infect Dis,* 189**,** 79-89.

LINDAHL, G., SJÖBRING, U. & JOHNSSON, E. 2000. Human complement regulators: A major target for pathogenic microorganisms. *Current Opinion in Immunology.*

LIU, M., ZHU, H., ZHANG, J. & LEI, B. 2007. Active and passive immunizations with the streptococcal esterase Sse protect mice against subcutaneous infection with group A streptococci. *Infect Immun,* 75**,** 3651-7.

LUCA-HARARI, B., DARENBERG, J., NEAL, S., SILJANDER, T., STRAKOVA, L., TANNA, A., CRETI, R., EKELUND, K., KOLIOU, M., TASSIOS, P. T., VAN DER LINDEN, M., STRAUT, M., VUOPIO-VARKILA, J., BOUVET, A., EFSTRATIOU, A., SCHALEN, C., HENRIQUES-NORMARK, B., STREP, E. S. G. & JASIR, A. 2009. Clinical and microbiological characteristics of severe Streptococcus pyogenes disease in Europe. *J Clin Microbiol,* 47**,** 1155-65.

MCCORMICK, J. K., TRIPP, T. J., OLMSTED, S. B., MATSUKA, Y. V., GAHR, P. J., OHLENDORF, D. H. & SCHLIEVERT, P. M. 2000. Development of streptococcal pyrogenic exotoxin C vaccine toxoids that are protective in the rabbit model of toxic shock syndrome. *J Immunol,* 165**,** 2306-12.

MCMILLAN, D. J., BATZLOFF, M. R., BROWNING, C. L., DAVIES, M. R., GOOD, M. F., SRIPRAKASH, K. S., JANULCZYK, R. & RASMUSSEN, M. 2004a. Identification and assessment of new vaccine candidates for group A streptococcal infections. *Vaccine,* 22**,** 2783-90.

MCMILLAN, D. J., DAVIES, M. R., BROWNING, C. L., GOOD, M. F. & SRIPRAKASH, K. S. 2004b. Prospecting for new group A streptococcal vaccine candidates. *Indian J Med Res,* 119 Suppl**,** 121-5.

MCNAMARA, C., ZINKERNAGEL, A. S., MACHEBOEUF, P., CUNNINGHAM, M. W., NIZET, V. & GHOSH, P. 2008. Coiled-coil irregularities and

instabilities in group A Streptococcus M1 are required for virulence. *Science (New York, N.Y.),* 319**,** 1405-1408.

MCNEIL, S. A., HALPERIN, S. A., LANGLEY, J. M., SMITH, B., WARREN, A., SHARRATT, G. P., BAXENDALE, D. M., REDDISH, M. A., HU, M. C., STROOP, S. D., LINDEN, J., FRIES, L. F., VINK, P. E. & DALE, J. B. 2005. Safety and immunogenicity of 26-valent group a streptococcus vaccine in healthy adult volunteers. *Clin Infect Dis,* 41**,** 1114-22.

MICHOS, A. G., BAKOULA, C. G., BRAOUDAKI, M., KOUTOUZI, F. I., ROMA, E. S., PANGALIS, A., NIKOLOPOULOU, G., KIRIKOU, E. & SYRIOPOULOU, V. P. 2009. Macrolide resistance in Streptococcus pyogenes: prevalence, resistance determinants, and emm types. *Diagn Microbiol Infect Dis,* 64**,** 295-9.

MORFELDT, E., BERGGÅRD, K., PERSSON, J., DRAKENBERG, T., JOHNSSON, E., LINDAHL, E., LINSE, S. & LINDAHL, G. 2001. Isolated hypervariable regions derived from streptococcal M proteins specifically bind human C4b-binding protein: implications for antigenic variation. *Journal of immunology (Baltimore, Md. : 1950),* 167**,** 3870-3877.

NAVARRE, W. W. & SCHNEEWIND, O. 1999. Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiol Mol Biol Rev,* 63**,** 174-229.

NORDSTROM, T., BLOM, A. M., FORSGREN, A. & RIESBECK, K. 2004. The emerging pathogen Moraxella catarrhalis interacts with complement inhibitor C4b binding protein through ubiquitous surface proteins A1 and A2. *J Immunol,* 173**,** 4598-606.

O'LOUGHLIN, R. E., ROBERSON, A., CIESLAK, P. R., LYNFIELD, R., GERSHMAN, K., CRAIG, A., ALBANESE, B. A., FARLEY, M. M., BARRETT, N. L., SPINA, N. L., BEALL, B., HARRISON, L. H., REINGOLD, A., VAN BENEDEN, C. & ACTIVE BACTERIAL CORE SURVEILLANCE, T. 2007. The epidemiology of invasive group A streptococcal infection and potential vaccine implications: United States, 2000-2004. *Clin Infect Dis,* 45**,** 853-62.

PANDEY, M., SEKULOSKI, S. & BATZLOFF, M. R. 2009. Novel strategies for controlling Streptococcus pyogenes infection and associated diseases: from potential peptide vaccines to antibody immunotherapy. *Immunol Cell Biol,* 87**,** 391-9.

PENFOUND, T. A., CHIANG, E. Y., AHMED, E. A. & DALE, J. B. 2010a. Protective efficacy of group A streptococcal vaccines containing type-specific and conserved M protein epitopes. *Vaccine,* 28**,** 5017-5022.

PENFOUND, T. A., CHIANG, E. Y., AHMED, E. A. & DALE, J. B. 2010b. Protective efficacy of group A streptococcal vaccines containing type-specific and conserved M protein epitopes. *Vaccine,* 28**,** 5017-22.

PERSSON, J., BEALL, B., LINSE, S. & LINDAHL, G. 2006. Extreme sequence divergence but conserved ligand-binding specificity in Streptococcus pyogenes M protein. *PLoS pathogens,* 2**,** e47.

PICHICHERO, M. E. & CASEY, J. R. 2007. Systematic review of factors contributing to penicillin treatment failure in Streptococcus pyogenes pharyngitis. *Otolaryngol Head Neck Surg,* 137**,** 851-857.

PRASADARAO, N. V., BLOM, A. M., VILLOUTREIX, B. O. & LINSANGAN, L. C. 2002. A novel interaction of outer membrane protein A with C4b binding protein mediates serum resistance of Escherichia coli K1. *J Immunol,* 169**,** 6352-60.

RAM, S., CULLINANE, M., BLOM, A. M., GULATI, S., MCQUILLEN, D. P., MONKS, B. G., O'CONNELL, C., BODEN, R., ELKINS, C., PANGBURN, M. K., DAHLBACK, B. & RICE, P. A. 2001. Binding of C4b-binding protein to porin: a molecular mechanism of serum resistance of Neisseria gonorrhoeae. *J Exp Med,* 193**,** 281-95.

RICKLIN, D., HAJISHENGALLIS, G., YANG, K. & LAMBRIS, J. D. 2010. Complement: a key system for immune surveillance and homeostasis. *Nat Immunol,* 11**,** 785-97.

SABHARWAL, H., MICHON, F., NELSON, D., DONG, W., FUCHS, K., MANJARREZ, R. C., SARKAR, A., UITZ, C., VITERI-JACKSON, A., SUAREZ, R. S., BLAKE, M. & ZABRISKIE, J. B. 2006. Group A streptococcus (GAS) carbohydrate as an immunogen for protection against GAS infection. *J Infect Dis,* 193**,** 129-35.

SANDERSON-SMITH, M., DE OLIVEIRA, D. M., GUGLIELMINI, J., MCMILLAN, D. J., VU, T., HOLIEN, J. K., HENNINGHAM, A., STEER, A. C., BESSEN, D. E., DALE, J. B., CURTIS, N., BEALL, B. W., WALKER, M. J., PARKER, M. W., CARAPETIS, J. R., VAN MELDEREN, L., SRIPRAKASH, K. S., SMEESTERS, P. R. & GROUP, M. P. S. 2014. A systematic and functional classification

of Streptococcus pyogenes that serves as a new tool for molecular typing and vaccine development. *J Infect Dis,* 210**,** 1325-38.

SANDIN, C., CARLSSON, F. & LINDAHL, G. 2006a. Binding of human plasma proteins to Streptococcus pyogenes M protein determines the location of opsonic and non-opsonic epitopes. *Molecular Microbiology,* 59**,** 20-30.

SANDIN, C., CARLSSON, F. & LINDAHL, G. 2006b. Binding of human plasma proteins to Streptococcus pyogenes M protein determines the location of opsonic and non-opsonic epitopes. *Mol Microbiol,* 59**,** 20-30.

SHAKHNOVICH, E., ABKEVICH, V. & PTITSYN, O. 1996. Conserved residues and the mechanism of protein folding. *Nature,* 379**,** 96-8.

SHULMAN, S. T., TANZ, R. R., DALE, J. B., BEALL, B., KABAT, W., KABAT, K., CEDERLUND, E., PATEL, D., RIPPE, J., LI, Z., SAKOTA, V. & NORTH AMERICAN STREPTOCOCCAL PHARYNGITIS SURVEILLANCE, G. 2009. Seven-year surveillance of north american pediatric group a streptococcal pharyngitis isolates. *Clin Infect Dis,* 49**,** 78-84.

STEER, A. C., LAW, I., MATATOLU, L., BEALL, B. W. & CARAPETIS, J. R. 2009. Global emm type distribution of group A streptococci: systematic review and implications for vaccine development. *Lancet Infect Dis,* 9**,** 611-6.

THERN, A., STENBERG, L., DAHLBÄCK, B. & LINDAHL, G. 1995. Ig-binding surface proteins of Streptococcus pyogenes also bind human C4b-binding protein (C4BP), a regulatory component of the complement system. *Journal of immunology (Baltimore, Md. : 1950),* 154**,** 375-386.

WANG, C. H., CHIANG-NI, C., KUO, H. T., ZHENG, P. X., TSOU, C. C., WANG, S., TSAI, P. J., CHUANG, W. J., LIN, Y. S., LIU, C. C. & WU, J. J. 2013. Peroxide responsive regulator PerR of group A streptococcus is required for the expression of phage-associated DNase Sda1 under oxidative stress. *PLoS ONE,* 8.

YOUNG, M. H., ARONOFF, D. M. & ENGLEBERG, N. C. 2005. Necrotizing fasciitis: pathogenesis and treatment. *Expert Rev Anti Infect Ther,* 3**,** 279-94.

ZINGARETTI, C., FALUGI, F., NARDI-DEI, V., PIETROCOLA, G., MARIANI, M., LIBERATORI, S., GALLOTTA, M., TONTINI, M., TANI, C., SPEZIALE, P.,

GRANDI, G. & MARGARIT, I. 2010. Streptococcus pyogenes SpyCEP: a chemokine-inactivating protease with unique structural and biochemical features. *FASEB J,* 24**,** 2839-48.

Chapter 2:

Methodology for M^{HVR}-C4BPα1-2 Complex Formation and Crystallization

METHODOLOGY OF M^{HVR}-C4BPα1-2 COMPLEX FORMATION AND CRYSTALLIZATION

M Type Selection Criteria

The M proteins from the GAS strains M2 (AP2), M4 (Arp4) and M22 (Sir22) have been studied extensively with regards to their ability to bind to C4BP and consequent resistance to complement-mediated opsonophagocytosis (Morfeldt et al., 2001, Berggard et al., 2001, Carlsson et al., 2003, Johnsson et al., 1998, Johnsson et al., 1996, Persson et al., 2006). For this reason, I chose to study the M proteins from these specific M types through crystal structure determination of the M protein HVR in complex with C4BPα1-2. When M4 and M22 proteins proved to be difficult to co-crystallize in complex with C4BPα1-2, M28 and M49, which had also been shown to bind C4BP (Persson et al., 2006), were added to the pool of possible crystallization candidates given their epidemiological relevance (Steer et al., 2009) and their availability through Dr. Victor Nizet at UCSD.

DNA Manipulation

The coding sequences of mature M2 (42-367), M4 (42-356), M22 (42-335), M28 (42-363), and M49 (42-359) proteins were cloned from GAS strains M2 (AP2), M4 (Arp4), M22 (Sir22), M28 (CDC reference strain 4041-05), and M49 (NZ131) (Morfeldt et al., 2001, Persson et al., 2006), respectively, into a

modified version of pET28a vector (Novagen) containing an N-terminal His_6-tag followed by a PreScission™ protease (GE Healthcare) cleavage site. Truncation-constructs of these proteins consisting of only the N-terminal 79, 86 or 100 residues were generated by site-directed mutagenesis through the insertion of amber stop codons at the appropriate sites. As a strategy to find the ideal length of the HVR and flanking region, 100 amino acids was chosen as a starting length. Eventually the 100 amino acid M^{HVR}s that were difficult to crystalize (M4, M22 and M49) were truncated to an ideal length (79, 79, and 86 amino acids respectively). Due to this truncation strategy, I was ultimately able to co-crystalize the M^{HVR}s of M4, M22, and M49 in complex with C4BP. However, the $M4^{HVR}$ cocrystal never produced usable diffraction data for crystallographic studies and was eventually abandoned as a structural target. Specific mutations were introduced into the M2 protein also using site-directed mutagenesis. Site directed mutagenesis was performed according to the QuickChange™ Manual except that 50 μL reactions were set up for polymerase chain reaction (PCR) instead of 12.5 μL reactions.

The coding sequence of CCP1-2 domains of the human C4BP (C4BPα1-2) (Jenkins et al., 2006) (a kind gift from G. Lindahl) was cloned into the modified pET28a vector described above. To obtain selenomethionine-substituted protein to be used in crystallographic phase determination (discussed more thoroughly in chapter 3), methionine residues were introduced in the coding sequence of C4BPα1-2 at positions 29, 46, and/or 71 by site-directed mutagenesis.

Protein Expression and Purification

M proteins were expressed in *Escherichia coli* BL21 (DE3) and purified using a previously described procedure (McNamara et al., 2008) with minor modifications. Specifically, bacteria were lysed with a C-5 Emulsiflex (Avestin Inc., Ottawa, Canada), and for M2 (wild-type and variants) purification, imidazole was not included in the lysis and wash buffers and ion exchange chromatography was omitted.

C4BPα1-2 was expressed in *E. coli* Rosetta 2 (Novagen) cells. It was purified and refolded as previously described (André et al., 2006), except for use of a C-5 Emulsiflex for lysis. Where needed, the N-terminal His_6-tags of M proteins and C4BPα1-2 were removed by PreScission™ protease cleavage according to manufacturer's instructions. After cleavage and reverse Ni^{2+}-NTA purification, M protein and C4BPα1-2 constructs were purified by size-exclusion chromatography (Superdex 200) in a buffer containing 150 mM NaCl, 50 mM Tris, pH 8.5. Proteins were concentrated to ~20 mg/mL by ultrafiltration; protein concentrations were determined by measuring absorbance at 280 nm and calculated molar extinction coefficients. Aliquots were flash-frozen in liquid N_2 and stored at -80 °C.

Selenomethionine (SeMet) substituted proteins C4BPα1-2 (L29M/L46M), C4BPα1-2 (L29M/L71M), and C4BPα1-2 (L46M/L71M) were purified as described previously (Wang et al., 2013).

Confirmation of Binding Interactions

Binding between $Chis_6$-C4BPα1-2 and the mature M types of M2, M4 and M22 was initially performed by a co-precipitation assay. For this assay, 40 μg of $CHis_6$-C4BPα1-2 protein was mixed with 120 μg of mature M2 protein (wild-type or mutant variant) in 50 μL of phosphate buffered saline (PBS) at 37 °C for 30 min. Fifty μL of Ni^{2+}-NTA agarose beads were equilibrated in PBS, then added to the protein mix in a 1:1 beads:PBS (100 μL) slurry and incubated for 30 min at 37 °C under agitation. The unbound fraction was removed and the beads were washed three times with 0.5 mL of PBS supplemented with 15 mM imidazole and eluted with 40 μL PBS supplemented with 500 mM imidazole. Bound and unbound proteins were resolved in non-reducing SDS-PAGE and visualized by Coomassie-staining (Fig 2.1).

After the binding interactions were confirmed for mature M proteins, the M^{HVR} constructs for M2, M4 and M22 were tested for their ability to bind C4BP. Given that the size of $Chis_6$-C4BPα1-2 (15,075 Da) was similar to that of the 100 amino acid HVR versions of M2, M4, and M22 (~13,000 Da), the binding interaction was difficult to resolve by a co-precipitation assay. For that reason, binding and complex formation was tested by size-exclusion chromatography. Complex formation was observed when I mixed the two proteins at a 1:1 molar ratio at high concentration (20 mg/mL complex = ~0.75 μM) and resolved via Superdex 200 size exclusion chromatography. Complex formation was observed

as a higher molecular weight peak compared to M^{HVR} or C4BPα1-2 protein run individually over the same column (Fig. 2.2). The co-precipitation and size-exclusion chromatography assays were not performed on the M28 and M49 proteins for the following reason. Understanding that these interactions were consistent amongst the M2, M4, and M22 proteins, I went directly to crystallization trials as soon as I was able to obtain the $M28^{HVR}$ and $M49^{HVR}$ proteins by cloning and purification.

Crystallization

$M2^{HVR}$ (residues 42-141), $M2^{HVR}$ (K65A/N66A), $M22^{HVR}$ (42-120), $M28^{HVR}$ (42-141), and $M49^{HVR}$ (42-127) proteins were mixed with C4BPα1-2 at a 1:1 molar ratio (final concentration of complex ~5 mg/mL) and dialyzed overnight at 4 °C in 10 mM Tris, pH 8. After dialysis, the samples were concentrated by ultrafiltration to ~20 mg/mL. Crystallization was performed by the hanging drop vapor-diffusion method.

The native $M2^{HVR}$-C4BPα1-2, $M2^{HVR}$ (K65A/N66A)-C4BPα1-2, and $M28^{HVR}$-C4BPα1-2 complexes and the SeMet-labeled $M2^{HVR}$-C4BPα1-2 (L29M/L46M) and $M2^{HVR}$-C4BPα1-2 (L46M/L71M) complexes were co-crystallized at 20 °C by mixing 1 μL of complex (~20 mg/mL in 10 mM Tris, pH 8.0) with 1 μL of the reservoir solution containing 1.5 M $(NH_4)_2SO_4$, 0.1 M Bis-Tris Propane, pH 7.0 (Fig. 2.3). Crystal formation occurred typically in ~1 week.

These crystals were transferred to reservoir solution supplemented with 20% ethylene glycol for cryo-preservation, mounted in fiber loops, and flash-frozen in liquid N_2. Crystals containing SeMet-labeled protein were treated similarly except that they were bathed in reservoir solution supplemented with freshly prepared 1 mM TCEP prior to cryopreservation.

The $M22^{HVR}$-C4BPα1-2 complex was co-crystallized at 20 °C by mixing 0.1 μL of the complex (~20 mg/mL in 10 mM Tris, pH 8.0) with 0.1 μL of the reservoir solution containing 2 M ammonium sulfate, 2% PEG 400, and 100 mM HEPES pH 7.5 (Fig. 2.3). Crystal formation occurred typically in ~1 week. These crystals were transferred to reservoir solution supplemented with 20% glycerol before being flash-frozen in liquid N_2.

SeMet-labeled $M49^{HVR}$-C4BPα1-2 (L29M/L46M) complex was co-crystallized at 20 °C by mixing 1 μL of the complex at (~20 mg/mL in 10 mM Tris, pH 8.0) with 1 μL of the reservoir solution containing 1.6 M Na/K PO_4, pH 6.9 (Fig. 2.3). Crystal formation occurred typically in ~1 week. These crystals were transferred to reservoir solution supplemented with 20% glycerol before being flash-frozen in liquid N_2.

Upon crystallization of the $M2^{HVR}$-C4BPα1-2 complex, co-crystallization was confirmed by non-reducing SDS-PAGE analysis. Crystals from multiple (N=10) drops were collected and washed three times by a process of resuspension in 1 mL of reservoir solution followed by centrifugation and aspiration of wash

buffer. Crystals were then resuspended in 20 uL of 1x SDS-PAGE running buffer and visualized by Coomossie stained non-reducing SDS-PAGE (Figure 2.4).

Acknowledgements

The text of this chapter, in full, is material currently being prepared for submission for publication. I am the principle researcher/author on the paper: Broad recognition of group A *Streptococcus* M protein hypervariability by human C4BP. Cosmo Z. Buffalo, Adrian J. Bahn-Suh, Tapan Biswas, Victor N. Nizet, and Partho Ghosh.

Adrian J Bahn-Suh contributed to this chapter by aiding significantly in the cloning, mutagenesis, protein expression and protein purification aspects of the research presented here.

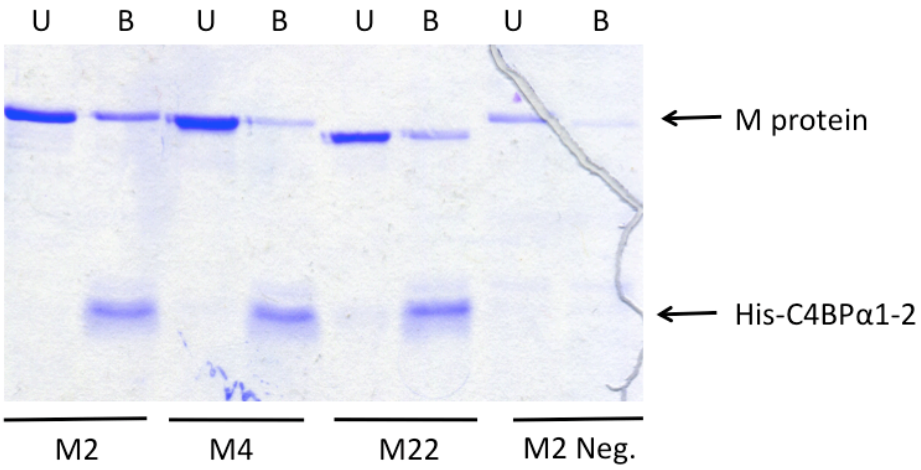


Figure 2.1: CHis$_6$-C4BPα1-2 binds mature M2, M4, & M22 *in vitro.*

Association of His-tagged C4BPα1-2 with mature M2, M4 and M22 at 37 °C was assessed by an Ni^{2+}-NTA agarose coprecipitation assay and visualized by non-reducing, Coomassie-stained SDS-PAGE. The bound fractions (B) follow the unbound fractions (U) for all three M types. In all three cases, the His- C4BPα1-2 is able to coprecipitate the M protein. M2 Negative (Neg.) control contained no M2 His-tagged C4BPα1-2 in the reaction.

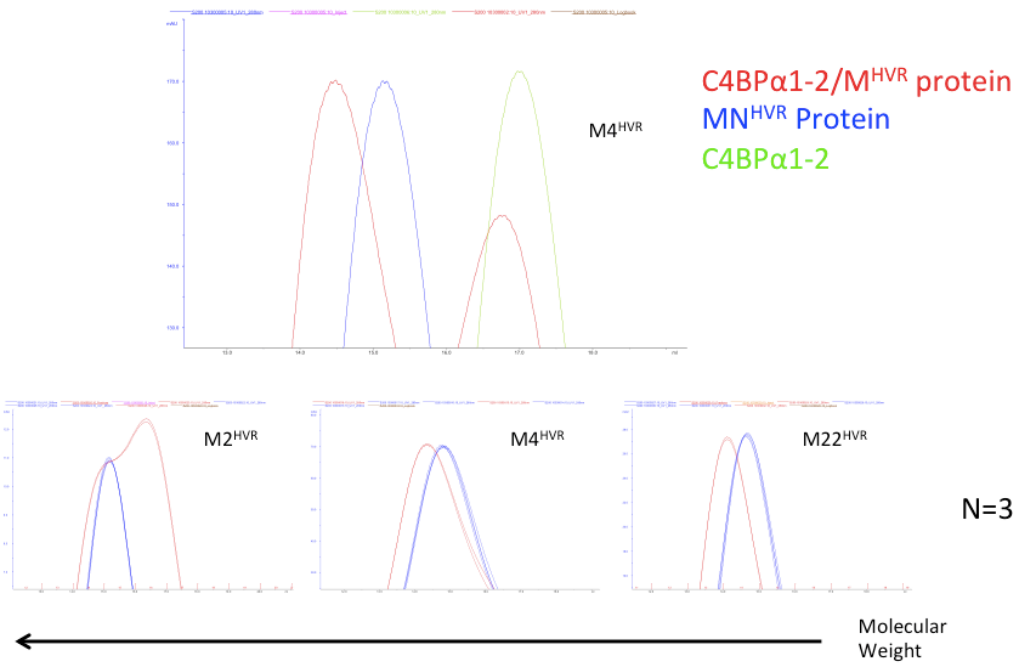


Figure 2.2: Association of M^{HVR} with C4BPα1-2 as assessed by gel shift size exclusion chromatography.

The upper chromatagram shows how the $M4^{HVR}$– C4BPα1-2 complex (red) runs at a higher molecular weight (MW) than the $M4^{HVR}$ (blue) and C4BPα1-2 (green) run independently. The lower chromatograms exclude the visualization of C4BPα1-2 and only show the separation between the three isolated M^{HVR}s and the the cooresponding complex. All three complexes run at a higher molecular weight than the HVR by itself. The molar absorptivity of the the $M2^{HVR}$ protein is extremely low. The lower molecular weight peak that dominates is unbound C4BPα1-2, and the front of the peak represents the complex for which there is a gel shift.

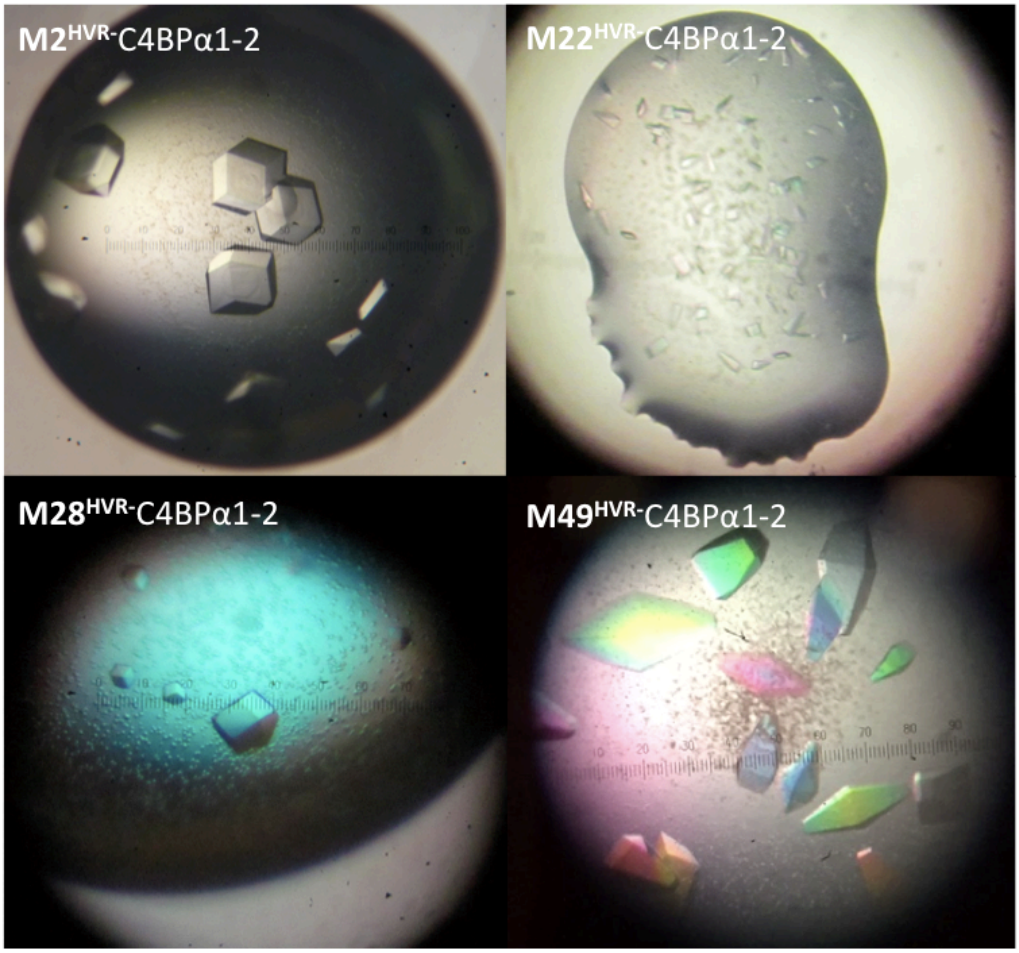


Figure 2.3: Co-crystallization of M^{HVR-}C4BPα1-2 complexes.

Photos of the cocrytal morphology of the of the M^{HVR-}C4BPα1-2 complexes. The $M2^{HVR-}$C4BPα1-2 co-crystal size was ~300 microns in diameter. The M^{HVR-}C4BPα1-2 complex co-crystallized in the P 43 3 2 space group. The The $M2^{HVR-}$C4BPα1-2 crystal showed no birefringence because it crystallized in a cubic space group. The $M22^{HVR-}$C4BPα1-2 co-crystal was ~50 µm in diameter. The $M22^{HVR-}$C4BPα1-2 crystal crystallized in the P 21 21 21 space group. The $M28^{HVR-}$C4BPα1-2 co-crystal was ~100 µm in diameter. The $M28^{HVR-}$C4BPα1-2 complex co-crystallized in a P 43 3 2 space group. The $M49^{HVR-}$C4BPα1-2 crystal at its largest measured ~400 µm along its long axis and ~250 µm across its short axis. The $M28^{HVR-}$C4BPα1-2 crystal crystallized in the P 43 21 2 space group. Again, this crystal showed no birefringence because it crysallized in a cubic space group.

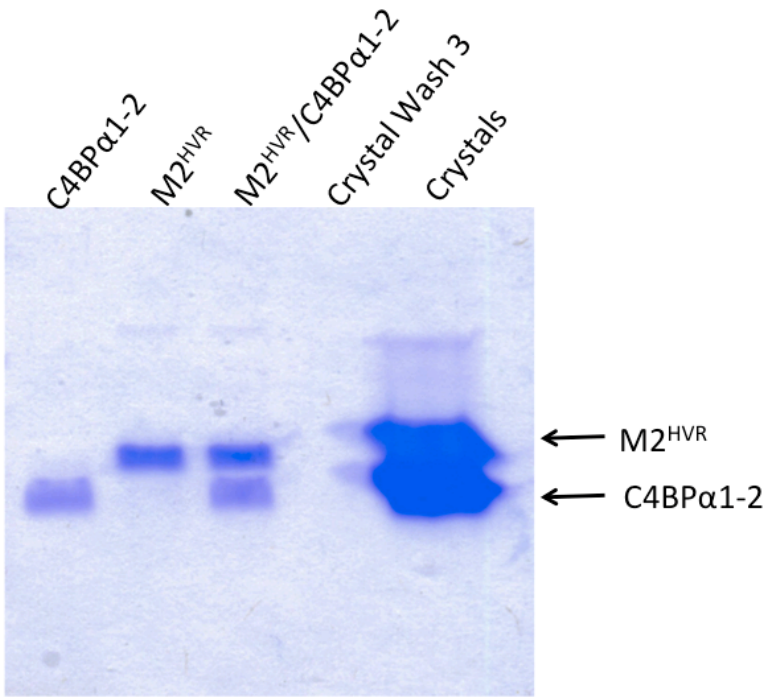


Figure 2.4: Conformation of $M2^{HVR}$-C4BPα1-2 co-crystallization.

Crystals of $M2^{HVR}$-C4BPα1-2 contain both $M2^{HVR}$ and C4BPα1-2, as determined by dissolving a crystal and examining the contents by SDS-PAGE. The first three lanes show C4BPα1-2, $M2^{HVR}$, and C4BPα1-2 mixed with $M2^{HVR}$ as standards. The next lane shows that the final wash of the crystal contains no proteins, and the last lane shows the dissolved crystal.

References

ANDRÉ, I., PERSSON, J., BLOM, A. M., NILSSON, H., DRAKENBERG, T., LINDAHL, G. & LINSE, S. 2006. Streptococcal M protein: Structural studies of the hypervariable region, free and bound to human C4BP. *Biochemistry,* 45**,** 4559-4568.

BERGGARD, K., JOHNSSON, E., MORFELDT, E., PERSSON, J., STALHAMMAR-CARLEMALM, M. & LINDAHL, G. 2001. Binding of human C4BP to the hypervariable region of M protein: a molecular mechanism of phagocytosis resistance in Streptococcus pyogenes. *Molecular Microbiology,* 42**,** 539-551.

CARLSSON, F., BERGGÅRD, K., STÅLHAMMAR-CARLEMALM, M. & LINDAHL, G. 2003. Evasion of phagocytosis through cooperation between two ligand-binding regions in Streptococcus pyogenes M protein. *The Journal of experimental medicine,* 198**,** 1057-1068.

JENKINS, H. T., MARK, L., BALL, G., PERSSON, J., LINDAHL, G., UHRIN, D., BLOM, A. M. & BARLOW, P. N. 2006. Human C4b-binding protein, structural basis for interaction with streptococcal M protein, a major bacterial virulence factor. *The Journal of biological chemistry,* 281**,** 3690-3697.

JOHNSSON, E., BERGGÅRD, K., KOTARSKY, H., HELLWAGE, J., ZIPFEL, P. F., SJÖBRING, U. & LINDAHL, G. 1998. Role of the hypervariable region in streptococcal M proteins: binding of a human complement inhibitor. *Journal of immunology (Baltimore, Md. : 1950),* 161**,** 4894-4901.

JOHNSSON, E., THERN, A., DAHLBÄCK, B., HEDÉN, L. O., WIKSTRÖM, M. & LINDAHL, G. 1996. A highly variable region in members of the streptococcal M protein family binds the human complement regulator C4BP. *Journal of immunology (Baltimore, Md. : 1950),* 157**,** 3021-3029.

MCNAMARA, C., ZINKERNAGEL, A. S., MACHEBOEUF, P., CUNNINGHAM, M. W., NIZET, V. & GHOSH, P. 2008. Coiled-coil irregularities and instabilities in group A Streptococcus M1 are required for virulence. *Science (New York, N.Y.),* 319**,** 1405-1408.

MORFELDT, E., BERGGÅRD, K., PERSSON, J., DRAKENBERG, T., JOHNSSON, E., LINDAHL, E., LINSE, S. & LINDAHL, G. 2001. Isolated hypervariable regions derived from streptococcal M proteins specifically bind human C4b-binding protein: implications for antigenic

variation. *Journal of immunology (Baltimore, Md. : 1950),* 167**,** 3870-3877.

PERSSON, J., BEALL, B., LINSE, S. & LINDAHL, G. 2006. Extreme sequence divergence but conserved ligand-binding specificity in Streptococcus pyogenes M protein. *PLoS pathogens,* 2**,** e47.

STEER, A. C., LAW, I., MATATOLU, L., BEALL, B. W. & CARAPETIS, J. R. 2009. Global emm type distribution of group A streptococci: systematic review and implications for vaccine development. *Lancet Infect Dis,* 9**,** 611-6.

WANG, C. H., CHIANG-NI, C., KUO, H. T., ZHENG, P. X., TSOU, C. C., WANG, S., TSAI, P. J., CHUANG, W. J., LIN, Y. S., LIU, C. C. & WU, J. J. 2013. Peroxide responsive regulator PerR of group A streptococcus is required for the expression of phage-associated DNase Sda1 under oxidative stress. *PLoS ONE,* 8Fi

Chapter 3:

M^{HVR}-C4BPα1-2 Co-crystal Phasing Strategies, Structure Refinement and Verification

M^{HVR}-C4BPα1-2 CO-CRYSTAL PHASING STRATEGIES, STRUCTURE REFINEMENT AND VERIFICATION.

Phasing Strategies

This chapter outlines the phase determination strategy I adopted. The first M^{HVR}-C4BPα1-2 complex I crystallized was the $M2^{HVR}$-C4BPα1-2 complex. A solution state structure of C4BPα1-2 had previously been determined (Jenkins et al., 2006). Based on this fact, a molecular replacement solution was pursued first. After collection of a native data set, a series of molecular replacement runs were performed in Phenix (Adams et al., 2010) using the program Phaser. Using the complete solution state structure of C4BPα1-2 as a search model, I was not able to find a solution using molecular replacement. Upon eventual crystal structure determination of the $M2^{HVR}$-C4BPα1-2 complex, it was determined that a molecular replacement solution was not possible due to a ~180° rotation of C4BPα1 with respect to C4BPα2 centered at the interdomain residue of Lys63. This domain shift had previously been predicted but the extent of the shift was not known (Jenkins et al., 2006). At the time, I addressed this predicted domain shift by using individual C4BPα domains of C4BPα1-2 as the search model. This approach again did not yield a molecular replacement solution even when the individual domains were used as a search model in combination with a modified M1 HVR, for which we have the structure (McNamara et al., 2008). This modified M1 HVR had all residues except core *a* and *d* heptad positions

substituted with alanines. GCN4 was also used as a search model in combination with the C4BPα1-2 individual domains. Despite all search models attempted, no solution was found.

After molecular replacement proved to be an unsuccessful phasing strategy for the $M2^{HVR}$-C4BPα1-2 complex, alternative phasing strategies were pursued. The second strategy investigated was isomorphous replacement. To obtain an isomorphous crystal containing incorporated heavy metals, I began by soaking native $M2^{HVR}$-C4BPα1-2 crystals in reservoir solutions supplemented with various heavy metal compounds at various concentrations and various soak times. However, despite all metal, soak time and concentration combinations used, heavy atom positions could not be located within the crystal upon performing analysis of the isomorphous diffraction data. To aid in heavy atom binding to the protein crystal, cysteine substitution mutations were introduced into the $M2^{HVR}$ at the C-terminal region of our construct (not within the N-terminal 50 amino acid HVR sequence necessary for C4BP binding). To maximize my chances of metal binding, I made three separate single-site cysteine mutants, K101C, Q116C and Q134C. Both the K101C and Q116C were at the *f* heptad position, the residue most distal from the core of the coiled coil. The Q134C mutation was at the more central *g* position. After a series of soaks similar to those mentioned above, a solution was not obtained for any of the three cysteine mutants upon analysis of our isomorphous diffraction data.

The anomalous dispersion phasing strategy was the last strategy available and the one that proved ultimately successful. However, I encountered many hurdles along the way. Initially, it was more practical to introduce a selenomethionine-label into the $M2^{HVR}$ protein due to the fact the protein was easily expressed and purified in *E. coli* with robust yield. Despite C4BPα1-2 containing one native methionine, the fact that the protein needs refolding and has lower purification yields suggested that it was a less than ideal target for selenomethionine incorporation. It was determined by calculating the anomalous signal contribution of selenomethionine that two methionine residues would be sufficient to acquire the necessary anomalous signal (~3.5% anomalous signal). UHowever, $M2^{HVR}$ has no native methionines. Due to the evolutionary tendency for leucine to be substituted with methionine, methionine substitutions at leucines present at core *a* or *d* positions were introduced. These substitutions were introduced into the C-terminal region of our construct (not within the N-terminal 50 amino acid HVR sequence necessary for C4BP binding). Using a combination of four different *a* and *d* position sites Leu99, L111, L114 and L134, a series of double mutants were produced (1) L99M, L114M, (2) L111M, L134M, (3) L99M, L134M, and (4) L111M, L114M. Upon crystallization and data collection of these $M2^{HVR}$ substitution mutants in complex with C4BPα1-2 crystals, no anomalous signal was detected upon analysis of the data. This suggested that the region the selenomethionines substitutions were introduced into was not ordered in the crystal. Upon eventual crystal structure determination of the $M2^{HVR}$-

C4BPα1-2, it was shown that this is indeed the case. Such disorder might also suggest a reason why the isomorphous replacement strategy was also unsuccessful in phase determination. At the time, to address this concern, I decided to substitute methionines for leucines at core *a* and *d* positions of the heptad located in the 50 amino acids of the N-terminal HVR region. Unfortunately, these mutant proteins did not crystallize, suggesting that the disruption made by introduction of the selenomethionines at these positions was sufficient to interfere with $M2^{HVR}$ binding to C4BPα1-2.

After no luck with incorporation of selenomethionine into the $M2^{HVR}$, selenomethionine was incorporated into C4BPα1-2. C4BPα1-2 has one native methionine at residue 14. The addition of two more methionines was predicted to provide sufficient anomalous signal (~4.3%) for structure determination of our complex. Again, in a series of substitution mutants, leucines were substituted with methionine at two of three C4BPα1-2 residue (positions 29, 46, or 71), producing three unique double substitution mutants. Residues 29 and 46 are in domain C4BPα1 and residue 71 is in domain C4BPα2. After successful refolding of C4BPα1-2 and co-crystallization, all diffraction data for the three crystals provided sufficient anomalous signal to obtain a solution. Ultimately this strategy yielded the $M2^{HVR}$-C4BPα1-2 crystal structure. Subsequently, this crystal structure allowed for a molecular replacement solution of the $M28^{HVR}$-C4BPα1-2 co-crystal structure using only the C4BPα1-2 protein as a search model from the $M2^{HVR}$-C4BPα1-2 structure. The $M22^{HVR}$-C4BPα1-2 co-crystal structure

eventually was solved using portions of the M28^{HVR}-C4BPα1-2 co-crystal structure as a search model. The M49^{HVR}-C4BPα1-2 structure also had to be solved by anomalous dispersion in a similar manner to that of M2^{HVR}-C4BPα1-2.

Data Collection

All data were collected with crystals under a liquid N_2 cryostream (~190 °C). Diffraction data for native M2^{HVR}-C4BPα1-2 were collected at the Stanford Synchrotron Radiation Lightsource (SSRL) beamline 9-2. For phasing purposes, single wavelength anomalous dispersion (SAD) data were collected from SeMet-labeled M2^{HVR}-C4BPα1-2 (L29M/L46M) and M2^{HVR}-C4BPα1-2 (L46M/L71M) at the Advanced Photon Source (APS) beamline 19-ID. Diffraction data for native M22^{HVR}-C4BPα1-2 were collected at APS beamline 24-ID-C and native diffraction data for M28^{HVR}-C4BPα1-2 were collected at the Advanced Lightsource (ALS) beamline 8.2.1. For phasing purposes, SAD data were collected from SeMet-labeled M49^{HVR}-C4BPα1-2 (L29M/L46M) at APS beamline 24-ID-E. M2^{HVR}-C4BPα1-2, M2^{HVR}-C4BPα1-2 (L29M/L46M), M2^{HVR}-C4BPα1-2 (L46M/L71M) and M28^{HVR}-C4BPα1-2 data were indexed, integrated, and scaled using HKL2000 (Otwinowski and Minor, 1997). M22^{HVR}-C4BPα1-2 and M49^{HVR}-C4BPα1-2 (L29M/L46M) data were indexed, integrated, and scaled using XDS (Kabsch, 2010).

Structure Determination and Refinement

For structure determination of $M2^{HVR}$-C4BPα1-2, Se sites from SeMet-labeled $M2^{HVR}$-C4BPα1-2 (L29M/L46M) and $M2^{HVR}$-C4BPα1-2 (L46M/L71M) SAD data were located and phases calculated for each data set using the program Autosol within Phenix (Adams et al., 2010). The two sets of phases were then combined using the Reflection File Editor program in Phenix. From the combined phase set, four Se sites, three at substituted methionines and one at the native Met 14, were identified per asymmetric unit, which contained one $M2^{HVR}$ and one C4BPα1-2 molecule.

Model building was performed with the program Coot (Emsley and Cowtan, 2004), and refinement with the program Phenix. Using the existing NMR structure as a starting point, the C4BPα1-2 molecule was manually built into density by inspection of SAD phased maps. Refinement was performed with Phenix using default parameters, and TLS parameterization ($M2^{HVR}$: 53-57; 58-86, C4BP: 0-59; 60-124) was performed at later stages of refinement. About 25 iterative cycles of building and refinement, with each refinement step consisting of 3-5 rounds and using experimental phases as constraints were performed. The $M2^{HVR}$ was then built into density by inspection of σ_A-weighted $2mF_o - DF_c$ and $mF_o - DF_c$ maps. The amino acid register for the coiled-coil of the $M2^{HVR}$ was assigned from the well-defined density of large side chains (i.e., His20, Phe75, His85). After the $M2^{HVR}$ model was built, ~10 iterative cycles of building and refinement were performed, as above. The structure was then refined against

higher resolution (2.56 Å) native data by an additional ~50 iterative cycles of building and refinement using the calculated phases (Table 3.1). Individual B-factors were refined isotropically throughout. Water molecules were added in the final stages of refinement using Phenix with default parameters (3σ peak height in σ_A-weighted $mF_o - DF_c$ maps). Continuous electron density was evident for the entire main chain of C4BPα1-2 and for residues 53-86 of the $M2^{HVR}$. Not all residue side chains were visible in C4BPα1-2 density, particularly those in some of the larger loops. The final model was validated with MolProbity (Chen et al., 2010), which indicated residues in the Ramachandran plot with 90.3% preferred regions, 8.4% in allowed regions and 1.3% in disallowed regions. The MolProbity clashscore for the $M2^{HVR}$-C4BPα1-2 complex was 18.23 (81^{st} percentile), with an overall MolProbilty score of 3.03 (44^{th} percentile).

The structure of $M2^{HVR}$ (K65A/N66A)-C4BPα1-2 was determined by refining $M2^{HVR}$-C4BPα1-2 structure with K65A/N66A substitutions against the $M2^{HVR}$ (K65A/N66A)-C4BPα1-2 crystal data (Table 3.1). Identical R_{free} flags were maintained for the two structures. Approximately 15 cycles of building and refinement were performed iteratively with the programs Coot and Phenix Refine using default parameters, and with TLS parameterization ($M2^{HVR}$: 53-57; 58-86, C4BP: 0-59; 60-124). The final model was validated with MolProbity, which indicated residues in the Ramachandran plot with 92.3% in preferred regions, 5.8% in allowed regions and 1.9% in disallowed regions. The MolProbity clashscore for the $M22^{HVR}$ (K65A/N66A)-C4BPα1-2 complex was 7.34 (98th

percentile), with an overall MolProbilty score of 2.64 (50th percentile).

The structure of M28^{HVR}-C4BPα1-2 was determined by molecular replacement using the program Phaser (within Phenix) with C4BPα1-2 of the M2^{HVR}-C4BPα1-2 structure as the search model (Table 3.1). A molecular replacement solution was found with a Log-Likelihood Gain of 378.998. The resultant electron density map displayed an alpha helix in the asymmetric unit that accounts for the M28^{HVR} molecule bound to C4BPα1-2. The program Coot was used for model building. The M28^{HVR} helix was build into density; well-defined density of large hydrophobic residues (i.e. Tyr62, Tyr76, Tyr77) were used to determine the register of the coiled-coil. The model was then subjected to cycles of rigid body refinement followed by inspection of σ_A-weighted $2mF_o - DF_c$ and $mF_o - DF_c$ omit electron density maps and rebuilding. Approximately 30 cycles of building and refinement were performed iteratively with the programs Coot and Phenix Refine using default parameters, with TLS parameterization (M28^{HVR}: 55-60, 61-65, 66-70, 71-75, 76-80; C4BP: 0-6, 7-14, 15-21, 22-28, 29-40, 41-46, 47-53, 54-58, 59-68, 69-77, 78-89, 90-97, 98-105, 106-118, 119-124) at the later stages of refinement. Individual B-factors were refined isotropically. Continuous electron density was evident for the entire main chain of C4BPα1-2 and residues 54-80 of the M28^{HVR}. The final model was validated with MolProbity, which indicated residues in the Ramachandran plot with 79.1% preferred regions, 16.8% in allowed regions and 4.1% in disallowed regions. The overall MolProbity clashscore for the M28^{HVR}-C4BPα1-2 complex was 57.74 (32nd percentile) with

and overall MolProbilty score of 3.11 (24th percentile).

The structure of the M22^{HVR}-C4BPα1-2 complex was determined by molecular replacement (Table 3.1). A crystallographic model of a complex containing a dimer of M28^{HVR} bound to one C4BPα1-2 molecule was used as a search model. The initial solution placed two copies of the search model in the P 21 21 21 asymmetric unit. The Mathews probability calculation suggests 4 C4BPα1-2 molecules and 2 M22^{HVR} dimers in the asymmetric unit at a solvent concentration of ~50%, suggesting the need to build two additional C4BPα1-2 molecules. After refinement of the initial model, the four additional domains of the two C4BPα1-2 molecules were placed stepwise and individually into density between rounds of iterative refinement. Compared to the bound form of C4BPα1-2 observed in other structures, a different orientation of C4BPα1 domain C4BPα1-2 was observed. This observation indicates why a straightforward molecular replacement search with two C4BPα1-2 molecules bound to the M28^{HVR} dimer was unsuccessful. Strong density of the M28^{HVR} dimer was present throughout the building and refinement process. The complete model without substituting in M22^{HVR} side chains for those of M28 was subsequently refined with Phenix. The program Coot was used for model building. The M22^{HVR} helix sidechains were subsequently built into density using the well-defined density of large hydrophobic residues (i.e. Tyr66, Tyr67) to determine the register of the coiled-coil. The model was then subjected to cycles of rigid body refinement followed by inspection of σ_A-weighted $2mF_o - DF_c$ and $mF_o - DF_c$ omit electron

density maps and rebuilding.Aproximately 30 cycles of building and refinement were performed iteratively with the programs Coot and Phenix Refine using default parameters, and with TLS parameterization ($M22^{HVR}$ chain A: 52-80, chain C: 52-79, chain E: 52-80, chain G: 52-79; C4BP chain B: 1-13, 14-27, 28-59, 60-73, 74-86, 87-102, 103-109, 110-115, 116-124, Chain D: 0-59, 60-124, chain F: 1-59, 60-124, chain H: 0-13, 14-33, 34-47, 48-59, 60-74, 75-86, 87-109, 110-124) at the later stages of refinement. Individual B-factors were refined isotropically. Water molecules were added in the final stages of refinement using Phenix with default parameters (3σ peak height in σ_A-weighted $mF_o - DF_c$ maps). Continuous electron density was evident for the entire main chain of C4BPα1-2 and for residues 52-79(80) of the $M22^{HVR}$. The final model was validated with MolProbity (Chen et al., 2010), which indicated residues in the Ramachandran plot with 92.6% preferred regions, 4.2% in allowed regions and 3.2% in disallowed regions. The overall MolProbity clashscore for the $M22^{HVR}$-C4BPα1-2 complex was 9.81 (97^{th} percentile), with an overall MolProbilty score of 2.49 (78^{th} percentile).

For structure determination of $M49^{HVR}$-C4BPα1-2, Se sites from SeMet-labeled $M49^{HVR}$-C4BPα1-2 (L29M/L46M) SAD data were located and phases calculated using the program Autosol within Phenix. Six Se sites were identified per asymmetric unit, which was found to contain an $M49^{HVR}$ dimer and two C4BPα1-2 molecules. This is consistent with the total of two SeMet substitutions introduced into C4BPα1-2.

Model building was performed with the program Coot, and refinement with the program Phenix. Using the C4BPα1-2 molecule from the M2^{HVR}-C4BP structure, C4BPα1-2 was built into density by inspection of SAD phased maps. Refinement was performed in Phenix using default parameters, with of TLS parameterization (M49^{HVR} chain A: 56-60, 61-126, chain C: 56-126, C4BP chain B: 0-10, 11-62, 63-124, Chain D: 0-13, 14-27, 28-33, 34-44, 45-53, 54-62, 63-73, 74-86, 87-102, 103-124) at the later stages of refinement. About 20 iterative cycles of building and refinement were performed with each refinement step consisting of 3-5 rounds and using experimental phases as constraints. The M49^{HVR} was then built into density by inspection of σ_A-weighted $2mF_o - DF_c$ and $mF_o - DF_c$ maps. The amino acid register for the coiled-coil of the M49^{HVR} was assigned from the well-defined density of large side chains (i.e., His20, Phe75, His85). After the M49^{HVR} model was built, ~10 iterative cycles of building and refinement were performed, as above (Table 3.1). Individual B-factors were refined isotropically throughout. Waters were added in the final stages of refinement using Phenix with default parameters (3σ peak height in σ_A-weighted $mF_o - DF_c$ maps). Continuous electron density was evident for residues 56-124(126) of the M49^{HVR} and most of the main chain of C4BPα1-2 except for some of the larger loops. The final model was validated with MolProbity, which indicated residues in the Ramachandran plot with 88.5.0% in preferred regions, 8.4% in allowed regions and 3.1% in disallowed regions. The overall MolProbity clashscore for the M49^{HVR}-C4BPα1-2 complex was 15.91 (91st percentile) with an

overall MolProbilty score of 3.23 (38th percentile).

Structure Verification Co-Precipitation Assays

Forty μg of CHis$_6$-C4BPα1-2 protein was mixed with 120 μg of mature M2 protein (wild-type or mutant variant) in 50 μL of PBS at 37 °C for 30 min. Fifty μL of Ni^{2+}-NTA agarose beads were equilibrated in PBS, then added to the protein mix in a 1:1 beads:PBS (100 μL) slurry and incubated for 30 min at 37 °C under agitation. The beads were washed three times with 0.5 mL of PBS supplemented with 15 mM imidazole and eluted with 40 μL PBS supplemented with 500 mM imidazole. Input control and bound proteins were resolved in non-reducing SDS-PAGE and visualized by Coomassie-staining.

Acknowledgements

The text of this chapter, in full, is material currently being prepared for submission for publication. I am the principle researcher/author on the paper: Broad recognition of group A *Streptococcus* M protein hypervariability by human C4BP. Cosmo Z. Buffalo, Adrian J. Bahn-Suh, Tapan Biswas, Victor N. Nizet, and Partho Ghosh.

Tapan Biswas contributed to this chapter by aiding in the structural determination of the M22^{HVR}-C4BPα1-2 co-crystal structure by molecular replacement.

Table 3.1. Data collection, phasing and refinement statistics for native and SAD (SeMet) structures.

	M2-C4BP	M2-C4BP (SeMet L46M, L71M)	M2-C4BP (SeMet L29M, L46M)	M28-C4BP	M22-C4BP	M49-C4BP (L29M, L46M)	M2 (K65A, N66A)
Data collection							
Space group	P 43 3 2	P 43 3 2	P 43 3 2	P 43 3 2	P 21 21 21	P 43 21 2	P 43 3 2
Cell dimensions							
a, *b*, *c* (Å)	148.3 148.3 148.3	148.7 148.7 148.7	148.3 148.3 148.3	133.7 133.7 133.7	68.08 80.35 152.9	78.1 78.1 345.3	148.6 148.6 148.6
α, β, γ(°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Wavelength	0.984 Å	0.979 Å	0.979 Å	0.979 Å	0.979 Å	0.979 Å	0.979 Å
Resolution (Å)	50-2.56(2.60-2.56)	50-2.90(2.95-2.90	50-3.00(3.05-3.00)	50-3.02(3.07-3.02)	80.35-2.54(2.68-2.54)	86.3-2.74(2.89-2.72)	50-2.29(2.37-2.29)
R_{merge}	0.18(1.00)	0.18(1.00)	0.15(1.00)	0.14(1.00)	0.14(1.00)	0.23(1.00)	0.13(1.00)
$I / \int I$	67.3(4.17)	21.3(2.17)	22.4(3.44)	13.2(0.81)	10.7(1.2)	11.3(1.00)	46.4(3.22)
Completeness (%)	100(100)	99.9(100)	100(100)	99.8(100)	99.8(99.9)	99.8(99.1)	100(100)
Redundancy	42.8(43.1)	40.6(40.7)	22.9(22.6)	9.5(9.7)	6.6(6.7)	9.4(9.6)	40.5(32.6)
cc1/2	1.00(0.86)	0.99(0.86)	0.99(0.86)	0.99(0.48)	1.00(0.60)	1.00(0.61)	0.99(0.86)
Refinement							
Resolution (Å)	2.56			3.02	2.54	2.74	2.29
No. reflections	18514			7480	28328	28699	47375
R_{work} / R_{free}	0.21/0.22			0.24/0.28	0.21/0.27	0.25/0.32	0.20/0.22
No. atoms							
Protein	1335			1197	4838	2953	1328
Ligand/ion	0			0	0	25	0
Water	74			0	79	286	74
B-factors							
Protein	78.2			102.1	74.0	117.4	52.8
Ligand/ion						115.8	
Water	103.8				54.6	85.0	48.4
R.m.s deviations							
Bond lengths (Å)	0.01			0.01	0.01	0.01	0.01
Bond angles(°)	1.35			1.41	1.48	1.16	1.35

REFERENCES

ADAMS, P. D., AFONINE, P. V., BUNKÓCZI, G., CHEN, V. B., DAVIS, I. W., ECHOLS, N., HEADD, J. J., HUNG, L. W., KAPRAL, G. J., GROSSE-KUNSTLEVE, R. W., MCCOY, A. J., MORIARTY, N. W., OEFFNER, R., READ, R. J., RICHARDSON, D. C., RICHARDSON, J. S., TERWILLIGER, T. C. & ZWART, P. H. 2010. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography,* 66**,** 213-221.

CHEN, V. B., ARENDALL, W. B., HEADD, J. J., KEEDY, D. A., IMMORMINO, R. M., KAPRAL, G. J., MURRAY, L. W., RICHARDSON, J. S. & RICHARDSON, D. C. 2010. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography,* 66**,** 12-21.

EMSLEY, P. & COWTAN, K. 2004. Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography,* 60**,** 2126-2132.

JENKINS, H. T., MARK, L., BALL, G., PERSSON, J., LINDAHL, G., UHRIN, D., BLOM, A. M. & BARLOW, P. N. 2006. Human C4b-binding protein, structural basis for interaction with streptococcal M protein, a major bacterial virulence factor. *The Journal of biological chemistry,* 281**,** 3690-3697.

KABSCH, W. 2010. Xds. *Acta Crystallogr D Biol Crystallogr,* 66**,** 125-32.

MCNAMARA, C., ZINKERNAGEL, A. S., MACHEBOEUF, P., CUNNINGHAM, M. W., NIZET, V. & GHOSH, P. 2008. Coiled-coil irregularities and instabilities in group A Streptococcus M1 are required for virulence. *Science (New York, N.Y.),* 319**,** 1405-1408.

OTWINOWSKI, Z. & MINOR, W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology, 307-326.*

Chapter 4:

Crystallographic Results and Structural Analysis

CRYSTALLOGRAPHIC RESULTS AND STRUCTURAL ANALYSIS

To understand the broad specificity of the M protein HVR-CBP interaction, I determined cocrystal structures of four M protein HVRs (M2, M22, M28, and M49) (Figs. 4.1, 4.2) bound to the first two domains of the C4BP α chain (Fig. 1.1). As expected, the M protein HVRs studied here have little sequence identity to one another (Fig. 4.1). Despite the lack of HVR sequence relationship, the structures are astonishingly similar. The HVRs form parallel, dimeric, α-helical, coiled coils, with two C4BPα1-2 molecules bound to each M protein dimer (Figs. 4.2, 4.3). The portions of the HVR's that contact C4BPα1-2 have canonical coiled-coil structures, except for M2 which is unwound, with an average pitch of ~240 Å rather than the canonical 150 Å (Fig. 4.4). C4BPα1 is proximal to the C-terminal portion of the HVR and C4BPα2 to the N-terminal portion, in agreement with the approach of intact C4BP to the streptococcal surface (Fig 1.1).

The C4BPα1 and α2 domains are relatively unchanged from their free structures as determined by NMR (average rmsd ~1.5 and ~1.0 Å for domains 1 and 2, respectively), except that domain 1 is rotated 180° with respect to domain 2 (Fig. 4.5). This rotation is consistent with mutagenesis (Blom et al., 2000) and structural evidence (André et al., 2006), and is discussed further below. The intermolecular interface is substantial, with a total of ~1450-1690 $Å^2$ of surface

area being buried. However, most of this surface area is polar in character, and the fit is far from hand-in-glove (surface complementarities 0.56-0.73). These observations suggest a modest binding affinity, consistent with 0.5 μM K_d (André et al., 2006) observed for the interaction between C4BPα1-2 and $M4^{HVR}$. This binding is greatly strengthened (i.e., picomolar K_d) through avidity between intact, multi-armed C4BP and surface-localized M protein (Sanderson-Smith et al., 2014).

Most significantly, the structures reveal that a uniform set of C4BP residues recognize the sequence diverse HVRs. The structures also reveal that the HVRs interact with C4BP in two similar but distinct binding modes, with M2 and M49 representative of one group and M22 and M28 of a second. This difference is appreciated at a broad level by noting that for M2 and M49, a single M protein α-helix contacts a single C4BPα1-2 molecule, whereas for M22 and M28, the M protein helices are rotated slightly such that a single M protein α-helix contacts two C4BPα1-2 molecules (Figs. 4.2 and 4.7).

Most of the HVR-interacting residues come from C4BPα2 and take the form of a 'quadrilateral' that is composed of: *(1)* a hydrophobic pocket (C4BP His67, Ile78, and Leu82); *(2)* a hydrogen bonding group (mainchain nitrogen of C4BP His67); and two positively charged residues, C4BP *(3)* Arg64 and *(4)* Arg66 (Fig 4.7a). The segment that holds this 'quadrilateral' is structurally unvarying in the four structures, being stabilized by a disulfide bond at Cys65 and limited in conformation by Pro68. The HVR's contribute residues that form

complementary quadrilaterals (Fig 4.7), but with the heptad positions of these residues differing between the M2/M49 and M28/M22 modes (Fig. 4.9). In all four structures, a hydrophobic residue (usually an aromatic) fits into the *(1)* hydrophobic pocket and a residue immediately following hydrogen bonds to the *(2)* mainchain nitrogen of His67. The contacts to C4BP *(3)* Arg64 and *(4)* Arg66 are predominantly electrostatic (usually salt bridge), but in the case of M49, a polar residue is absent and instead Arg64 makes hydrophobic contacts, extending its alkyl chains across several M49 residues (Fig 4.7a-b).

Conserved contacts from C4BPα1 are fewer in number than that of C4BPα2. The key C4BPα1 residue is Arg39, which forms electrostatic contacts as well as hydrophobic ones (like Arg64) through its alkyl chain, which in conjunction with mainchain atoms of C4BPα1 forms a 'hydrophobic nook' (Fig 4.8). All four HVRs have hydrophobic residues that interact with the 'hydrophobic nook', with M2, M49, and M22 having these residues in the *e* heptad position and M28 in the spatially proximal *b* heptad position. M2 and M49 also have negatively charged residues that interact with C4BP Arg39, whereas M22 and M28 do not (Fig. 4.7).

The importance of C4BP Arg39 provides an explanation for the aforementioned 180° rotation of C4BPα1 (around a hinge at Lys63, Fig. 4.5). In free C4BP, Arg39 is positioned on the opposite side of the HVR-binding site located on C4BPα2 and thus not in a position to interact with the HVR. This rotation was seen in all four structures, except in one of the two C4BPα1-2

molecules bound to M22, where a crystal contact appears to prevent this rotation (Fig 4.2, 4.6, 4.17). A similar rotation appears to be necessary for the interaction of C4BP with C4b (Blom et al., 2000). However, the purpose of C4BP having different free and bound conformations is unclear.

The identification of C4BP-interacting residues in the four HVRs made it clear that the M2/M49 and M22/M28 binding modes are conserved in a larger group of HVRs (Fig. 4.9). Analysis showed that 11 other M protein HVRs could be assigned to the M2/M49 mode (Fig. 4.10) and 29 to the M22/M28 mode (Fig. 4.11). Thus, the structural evidence enabled identification of hidden sequence conservation in a large number of HVRs. Of note, a further 46 HVRs could not be assigned to either mode (Fig. 4.12), suggesting that HVRs interact with C4BP in other ways not identified by the four structures presented here.

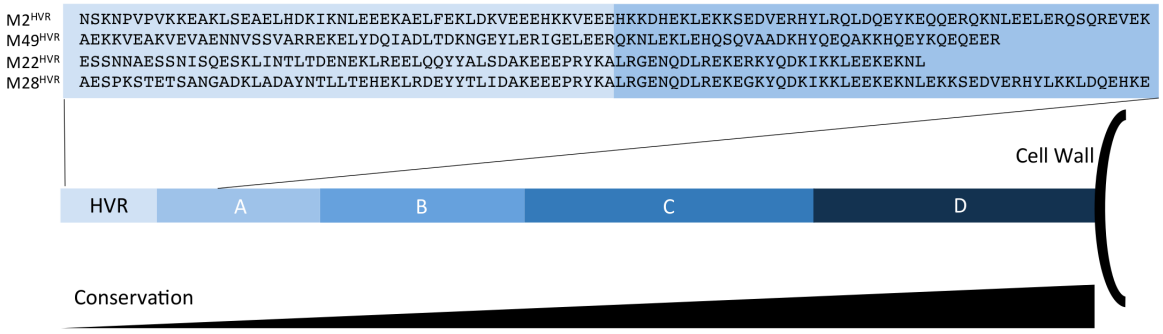
I next sought to understand the contribution of individual M protein residues to the interaction with C4BP, and turned to M2, since its complex with C4BPα1-2 was the first to be structurally characterized. Alanine substitution mutants were created in the M2 residues mentioned above that make conserved contacts (i.e., with the C4BPα2 quadrilateral and C4BPα1 Arg39) along with two that make contacts that are unique to M2 (Lys65 and Glu83) (Fig 4.14). The $M2^{HVR}$-C4BPα1-2 interaction was assessed through a Ni^{2+}-NTA agarose coprecipitation assay using His-tagged C4BPα1-2 (Fig 4.13). Of the single-site substitutions, the aromatic F75A showed a decrease in binding. This residue fits into the hydrophobic pocket of the nook in C4BPα1. Two double-site

substitutions also diminished binding. The first, D62A/E68A, removes two of the polar contacts to the C4BPα2 quadrilateral and the second, E76A/D79A, removes the two salt bridges to C4BPα1 Arg39 (Fig. 4.14).

Surprisingly, two of the single-site mutations, K65A and N66A, increased binding, as did the K65A/N66A double mutant (Fig. 4.13). To understand whether substitutions at these sites reordered the binding site, I determined the structure of the complex of the $M2^{HVR}$ K65A/N66A mutant with C4BPα1-2 and found that the structures of wild-type and the double-substitution complex were nearly identical (rmsd 0.15 Å) (Fig. 4.19). This result suggests that these two residues are poorly tolerated in the binding site. Lys65 forms a hydrogen bond to the mainchain oxygen of C4BP Arg64, which is unique to M2 and not seen in the other HVRs structurally characterized here (Figs. 4.15-18). But this positively charged residue is also sandwiched between two other positively charged key residues, C4BP *(3)* Arg64 and *(4)* Arg66, providing an explanation for why its substitution by Ala leads to enhanced binding. The other residue, M2 Asn66, forms a hydrogen bond to the C4BPα2 quadrilateral residue *(4)* Arg66. However, these residues have the highest B-factors in the binding site (Fig. 4.20), and in other M proteins that belong to the M2/M49 class, the equivalent of Asn66 is almost always Asp or Glu (Figs. 4.7, 4.10 and 4.15). Consistent with this, I substituted Asn66 with Asp and found that this increased the binding strength (Fig 4.13). Thus, it appears that C4BP Arg66 prefers a salt bridge (e.g. N66D) or no interaction (i.e., N66A) to a hydrogen bond, presumably because the salt bridge

provides sufficient binding energy whereas the hydrogen bond does not, and having no interaction relieves the entropic cost of ordering the sidechain of the Arg.

This large set of data clarified that in the M2/M49 mode, the *(1)* hydrophobic pocket prefers an aromatic residue, *(3)* Arg64 prefers electrostatic interactions but other potential cases of hydrophobic interactions as seen in M49 exist (e.g., M84 and M118), and Arg39 prefers to interact with two negatively charged residues as seen for M2. In the M22/M28 mode, the *(1)* hydrophobic pocket does not have a clear preference for aromatic residues but does prefer hydrophobic ones, *(3)* Arg64 prefers electrostatic interactions, and Arg39 does not require interaction with negatively charged residues as seem in the M2/M49 mode. In both the M2/M49 and the M22/M28 modes, the (4) Arg66 prefers to interact with negatively charged residues, and the 'hydrophobic nook' does not have a clear preference for aromatic resides given the sequence analysis data. This analysis suggests the clear identification of a C4BP 'reading head' that is comprised of the five major interactions of C4BPα1-2. This 'reading head' shows a great deal of tolerance in binding partners that may be responsible for its ability to bind many different M protein HVRs with such broad specificity despite no apparent sequence similarity. This 'reading head' is responsible for binding ~88% of tested M-types tested (Persson et al., 2006) and could be directly applied to the eventual design of broadly neutralizing antibodies.




```
M2          -----NSKNPVPVKKEAKLSEA--ELHDKIKNLEEEKAELFEKLDKVEEEHKKVEEE
M22         -----ESSNNAESSNISQESKLINTLTDENEKLREELQQYYALSDAKEEEPRYKA--
M28         AES--PKSTETSANGADKLADAYNTLLTEHEKLRD---EYYTLIDAKEEEPRYKA--
M49         AEKKVEAKVEVAENNVSSVARREKELYDQIADLTDKNGEYLERIGELEER-------
                   . . .  . :    * : .* :  :    . **
```

Figure 4.1. Schematic of M protein domains.

M proteins are hypervariable at their N-termini, with conservation increasing towards their C-termini. The sequences of M2, M49, M22, and M28 cocrystallized with C4BPα1-2 are depicted, with the HVR in light blue and the A region in darker blue. A multiple sequence alignment of the M2, M49, M22, and M28 HVRs is shown, as carried out using MUSCLE (Edgar, 2004a, Edgar, 2004b).

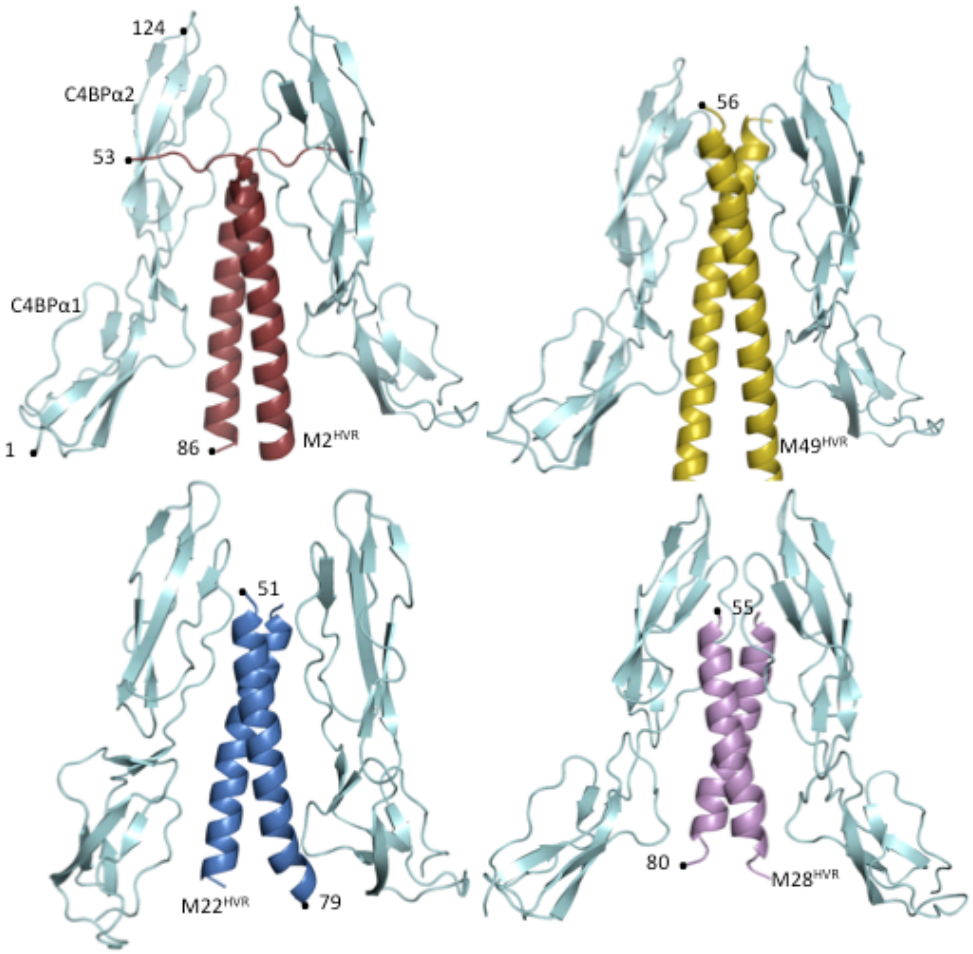


Figure 4.2. Structures of M^{HVR}-C4BPα1-2 complexes.

C4BPα1-2 (cyan) in complex with $M2^{HVR}$ (red), $M49^{HVR}$ (yellow), $M22^{HVR}$ (blue), and $M28^{HVR}$ (magenta).

Figure 4.3 Superposition of M^{HVR}-C4BPα1-2 complexes.

Superposition is based on one of the two C4BPα1-2 molecules of the complex (shown at right in cyan). $M2^{HVR}$ is red and its second C4BPα1-2 molecule green; $M49^{HVR}$ is yellow and its second C4BPα1-2 orange; $M22^{HVR}$ is blue and its second C4BPα1-2 is omitted due to a crystal contact restricting its orientation; and $M28^{HVR}$ is magenta and its second C4BPα1-2 pink.

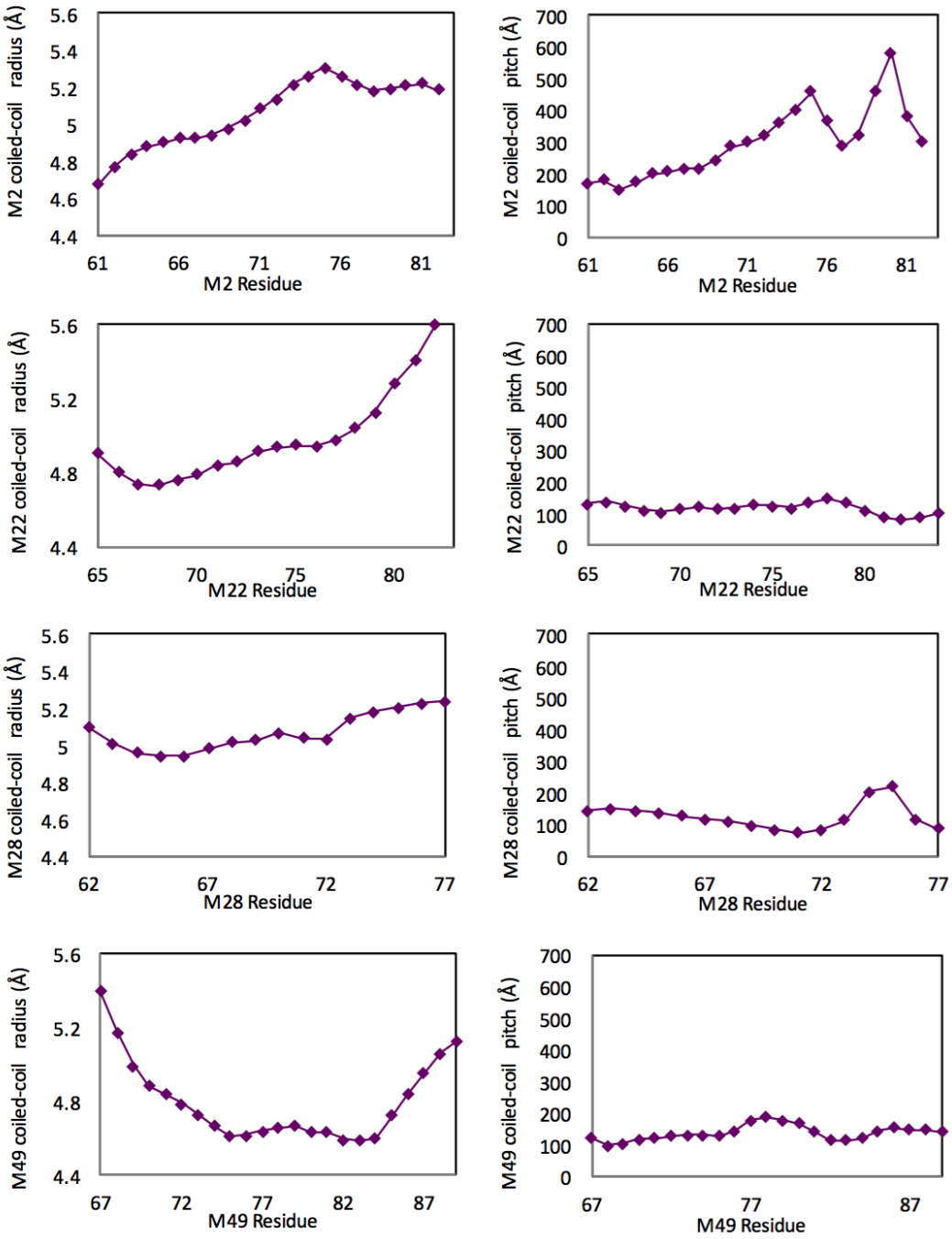


Figure 4.4. Coiled coil parameters.

Radius and pitch of the α-helical coiled coils of M protein HVRs at the interface with C4BPα1-2. Coiled coil parameters were calculated using the program TWISTER (Strelkov and Burkhard, 2002).

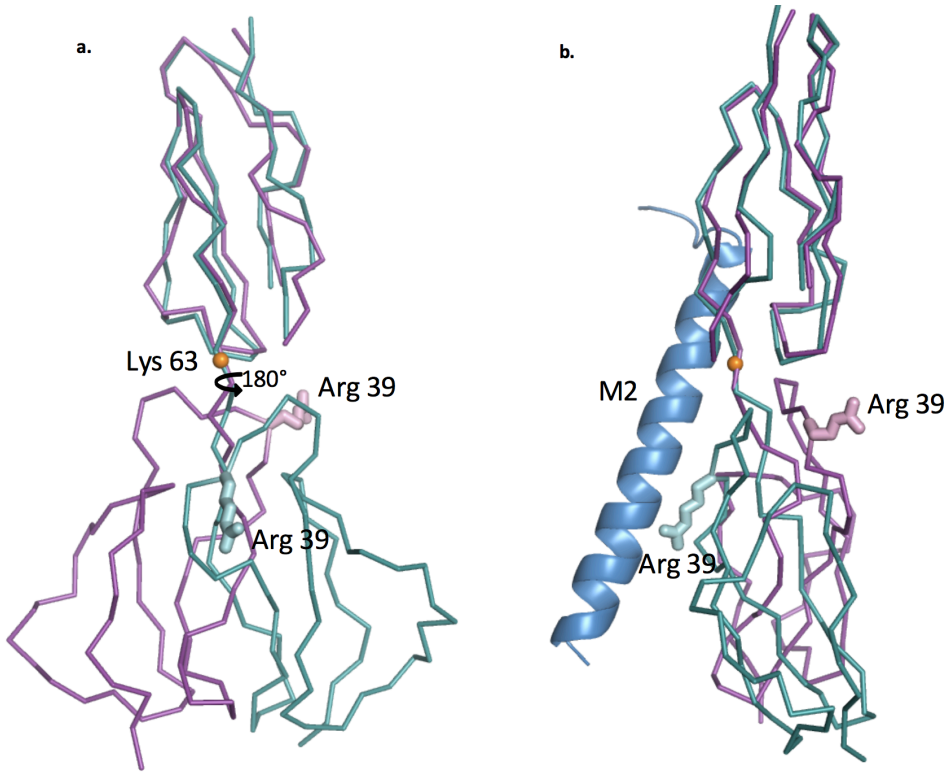


Figure 4.5. Rotation of C4BPα1-2.

a. Superposition of free (magenta) and M protein-bound C4BPα1-2 (cyan) based on the C4BPα2 domain, depicted as Cα chain traces. C4BPα1 rotates 180° around Lys63 (left). The position of Arg39 is shown in bonds representation. **b.** 90° rotation view of the superposition shown in panel a, with one α-helix of the $M2^{HVR}$, which interacts with Arg39, shown as a blue ribbon.

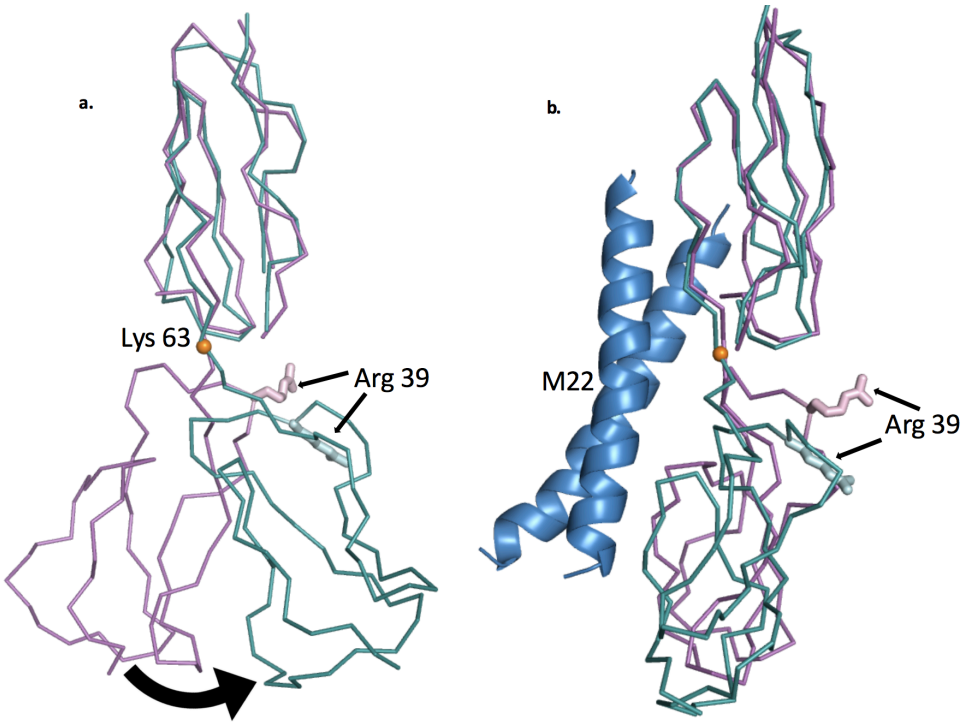


Figure 4.6. Tilt of C4BPα1-2 in M22 secondary binding mode.

a. Superposition of free (magenta) and M22 protein-bound C4BPα1-2 (cyan) based on the C4BPα2 domain, depicted as Cα chain traces. C4BPα1 tilts at Lys63 (left). The position of Arg39 is shown in bonds representation. **b.** 90° rotation view of the superposition shown in panel a, with both α-helices of the $M2^{HVR}$, which does not interact with Arg39, shown as a blue ribbon.

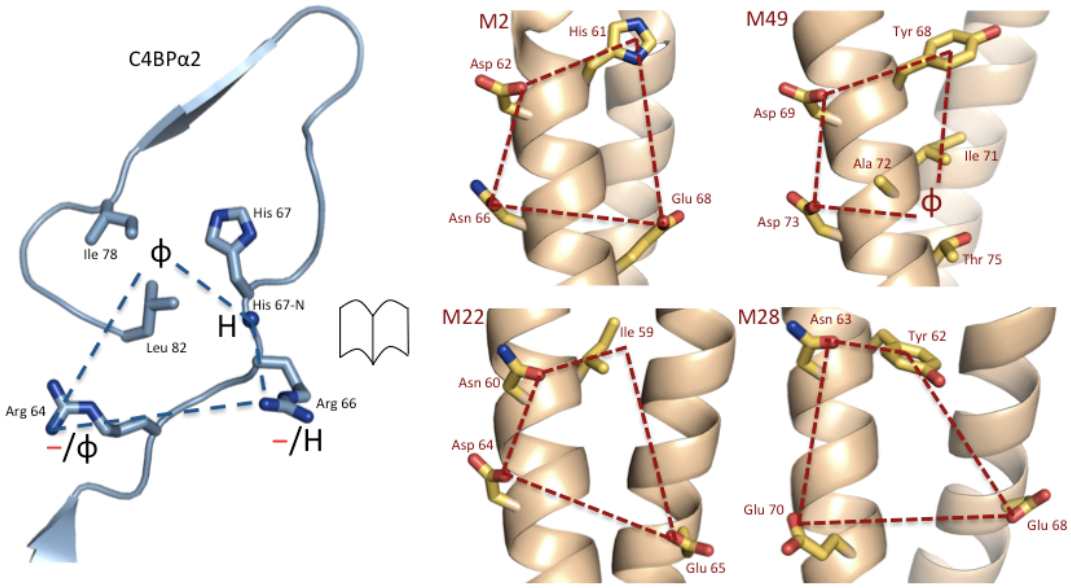


Figure 4.7. C4BPα2 binding mode.

a. The C4BPα2 quadrilateral (blue dashed lines), with the C4BPα2 backbone shown in ribbon representation and key sidechains shown as bonds (here and in following panels). The chemical character of M protein residues that interact with the quadrilateral is depicted: ϕ, hydrophobic; —, negative; H, hydrogen bond forming. **b.** M2, M49, M22, and M28 residues that interact with the C4BPα2 quadrilateral, shown in open-book representation with respect to C4BPα2. The M protein residues form the complementary quadrilateral (red dashed lines).

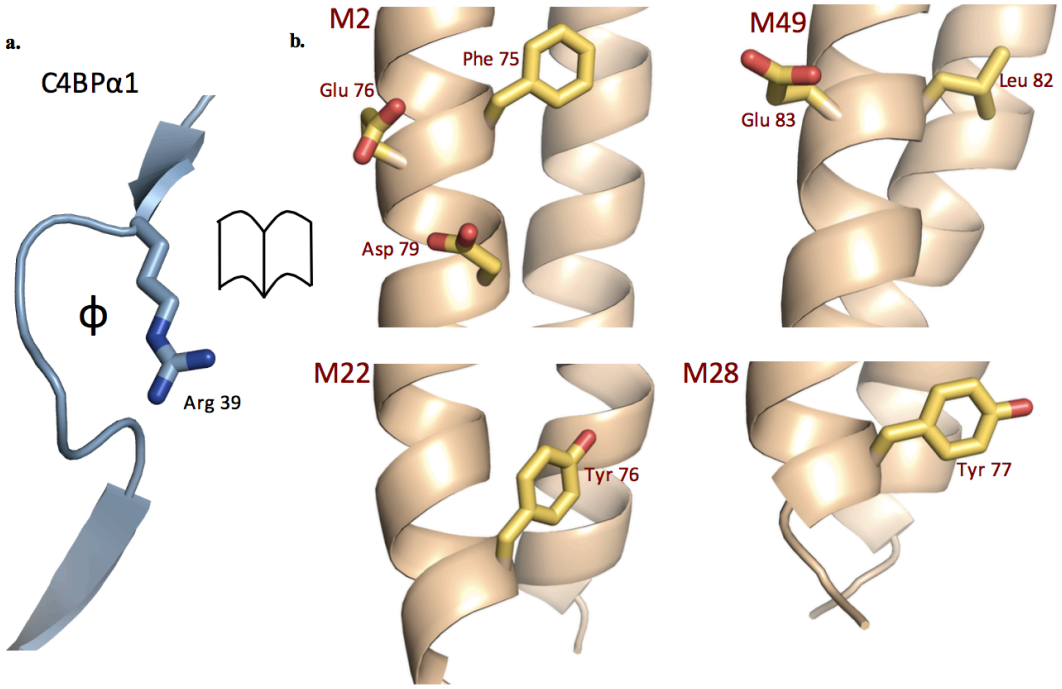


Figure 4.8. C4BPα1 binding mode.

a. The C4BPα1 Arg39 nook. ϕ denotes a hydrophobic pocket. **b.** M2, M49, M22, and M28 residues that interact with the C4BPα1 Arg39 nook shown in open-book representation.

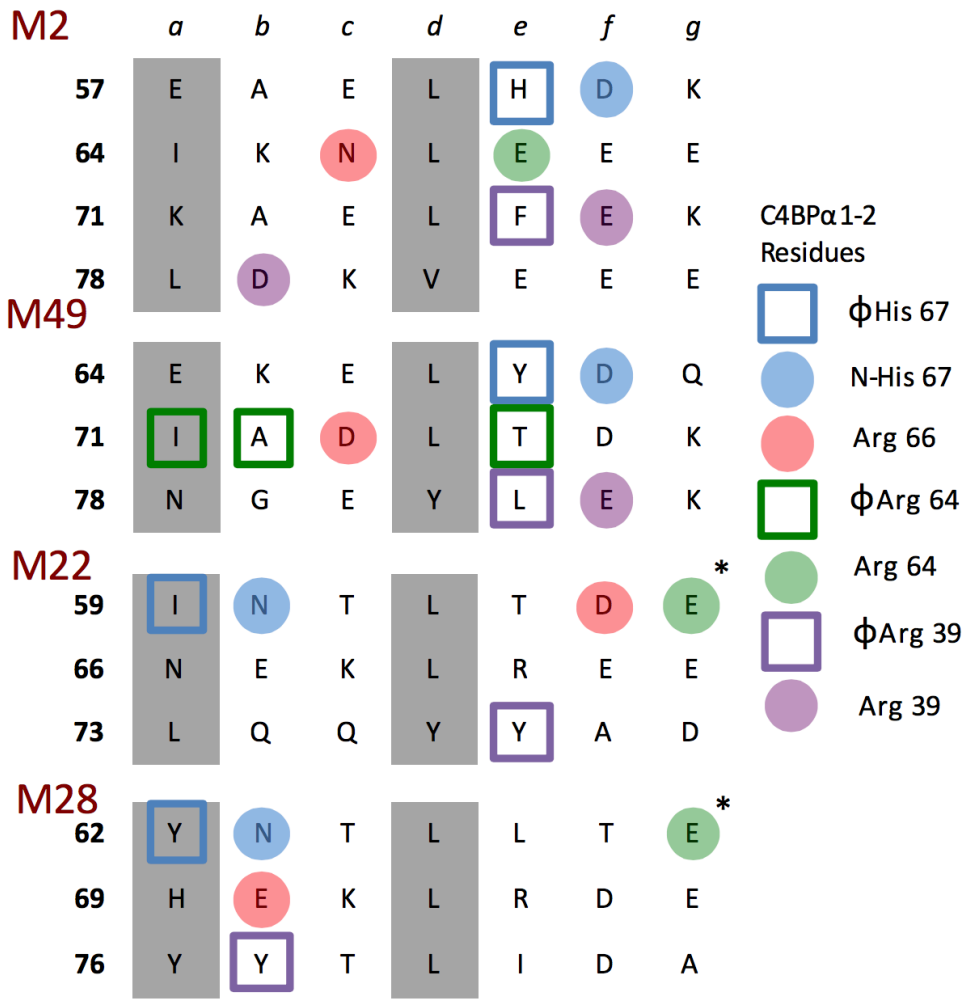


Figure 4.9. C4BP-binding modes of M proteins.

Heptad register of M2, M49 and M28 HVRs (*a* and *d* residues in grey). M protein residues interacting with C4BPα1-2 residues are highlighted according to their corresponding C4BPα1-2 interaction.

```
          abcdefgabcdefgabcdefgabcdefgabcdefg
M2        KKEAKLSEAELHDKIKNLEEEKAELFEKLDKVEEE
M49       VSSVARREKELYDQIADLTDKNGEYLERIGELEER
M114      KEATKLSEAELYNKIQELEEGKAELFDKLEKVEEE
M73       KEAKKLNEAELYNKIQELEEGKAELFDKLEKVEEE
M77       EGVSVGSDASLHNRITDLEEEREKLLNKLDKVEEE
M84       ASVKKNNEEELHNKIADLLDQNEEYLNKIDELKEG
M89       SVSVKDNEKELHNKIADLEEERGEHLDKIDELKEE
M97       GPVPRSLWLREYDKNQELTKKLTEFEEKLLQN
M102.1    SSVPVKKAAELYDKIKELEEGREELLNDLDKVKED
M106      QKQNVSSNGRIYEIYDELQTKYDELQTKHEELLGE
M112      SSVSVKNEVKLHNEIAALQEEKEKLLNELDVKEEH
M118      ADSNASSVAKLYNQIADLTDKNGEYLERIEELEER
M124      ATKSKLSEAELHDKIKNLEEEKAELFEKLDKVEEE
```

Figure 4.10. Sequence alignment of C4BP-binding M protein HVRs of the M2/M49 binding mode.

Residues that contact or are predicted to contact C4BP are in red, and the heptad register is indicated above. Residues falling at core d positions of the heptad register are highlighted in blue.

```
          abcdefgabcdefgabcdefgabcdefgabcdefg
M22       ISQESKLINTLTDENEKLREELQQYYALSDAKEEE
M28       ADKLADAYNTLLTEHEKLRDEYYTLIDAKEEEPRY
M4        AWNWPKEYNALLKENEELKVEREKYLSYADDKEKD
M4.1      AWNWPKEYNALLKENEEFKVEREKYLSYADDKEKD
PrtH      NAKLVEVVETTSLENEKLKSENEENKKNLDKLSKD
M9        LSVPKTEYDKLYDDYDKLQEKSAEYLERIGELEER
M8        NEQLINELNNLIEENNDLKDKLARNLDLLDNTREK
M11       TNVSADLYNSLWDENKTLREKQEEYITKIQNEETK
M15.1     WKLTIEEYNKLLDENEKLKEKNEEYLEKIGEQEER
M25       AKAAEAKVDKLEKQLEGYKKLEEDYFNLEKR
M42       KVKLEVLYNSLWEENKTLREEQEEYIAKIDKLDEK
M44       GSVSLELYDKLSDENDILREKQDEYLTKIDGLDKE
M55       LYQERQRLQDLKSKFQDLKNRSEGYIQQYYDEEKN
M59       ELTLQQKYDALTNENKSLRRERDNYLNYLYEKEEL
M60       ISKERELINTLVDENNKLMEERARHLDLIDNI
M61       GSVSLELYDKLSDENDILREKQDEYLTKIDGLDKE
M63.4     KLTLEHKYNALTNENKSLIRREKDKYLYEKEELEK
M66       QNTWEKRYQKLSDDHTLLQDAIEEISSENEKLKSE
M67       GGVRLDLYDKLSKENDILREKQDEYLTKIDELGEK
M69.1     GGVSLDLYSKLLNENDILRDKQDDYLTKIDELTEK
M76       SNVSINLYNELQAEHDKLQTKHEELLAEHDALKEK
M78       SITNEQLIDKLVEENNDLKEERAKYLDLLDNREKD
M81       ENVPKQQYNALWEENEDLRGRERKYIAKLEKEEIQ
M85       TSVSADLYNSLWDENKTLREKQGEYITKIQNEETK
88.1      ISNNERLINELTDENNELKDKLARSLDLLDNTREK
M92       SGSVSTPYNNLLNEYDDLLAKHGELLSEYDALKEK
M96       LDQFGRDYDELQKKYDKLDKENKEYASQLGK
M109      ADNLAKEYNTLLTENEKLREELQQYYALIDAKEEE
M110      HEELWKEYDILKEKLDKDQEEREKIELNYLK
111.1     VTAPAHFWENQRREIEKLKGEIDQLKLLLGKS
M117      GSVSIDRYNELSGEYNKLLDQNGNLLDENEILREK
```

Figure 4.11. Sequence alignment of C4BP-binding M protein HVRs of the M2/M49 binding mode.

Residues that contact or are predicted to contact C4BP are in red, and the heptad register is indicated above. Residues falling at core d positions of the heptad register are highlighted in blue.

```
M14.5      DRVSRSMSRDDLLNRAQNLEAKNHGLEHQNTKLSTENKTLQEQAEARQKE
M18         KDELIKRANGYEIQNHQLTVENKKLKIDKEQLTKENDDLKTEKDQLEQRS
M29.2     RVYITRRMTKEDVEKIANDLDTENHGLKQQNEQLSTEKQGLEEQNKQLST
M38         EGEPREVSEELVNSNPVLLNKKIAKLKEELANKEQESKESKEAIDALNNI
M54      EVLTRRQSQDPKYVTQRISDLEVKNHDLENKNEKLTSENQNLKNKTTELE
M62         EEAGASRTITSSENISKLYDENSKLIEERADLLGKLEEKEDKLESVERQY
M90     EGKAAAVSRSNSEQNNSEQNNLEKRYRKLSDDYTVLQEAIEGISSENEKL
M94         EEASNNGQLTLQHKNNALTSENESLRREKEELEKKNKELDSQVAGLIGVV
M99         DGERVPKNNRLSKKYSELSEKYGALSEKYGALLDKQGALLDKQEELEKEN
M105     EVNTRSRAQDAGYQKGRADKLETENHGLKFQNEKLQNQNNDLKTQTATLT
M18.6    APLTRATADNKDELIKRANDYEIQNHQLTVENRKLKTDKEQLTKENDDLK
M32      KAVTRGTVSDPETARQTIDKYDIKNHQLTQENEKLTKEKEELTQENEKLT
M36      KALTRSTASNSETARQTINDYEIKNHNLTQENEKLTEQNKELTSEKEKLT
M46        AAVTRHMSTEQLKQRVREYDIENHKLKTDKARLEAEKGQLETKKNELEAK
M71      RAITRATSDDPAKLKQMVEGYELENHTLKNDKEKLTTENSALTTEKNRLT
M100.1   RVTTRSQAQDAAGLKEKADQYEVRNHELEHNNEKLKTENSDLKTENSKLT
M115     KAVTRSTASDPEKARQTINEYEVKNHKLTQDNERLAQEKKGLTQNNERLT
M58     DSSREVTNELTASMWKAQADSAKAKAKELEKQVEEYKKNYETLEKGYDDL
M79.1    DSRDITGTLPATMWKQKAEEAKAKASNLEKQLEEARKDYSQIEEKLEQFG
M87     ESPREVTNELAASVWKKKVEEAKEKASKLEKQLEEAQKDYSEIEGKLEQF
M103    DSPRDVTSDLTTSMWKKKAEEAEAKASKFEKQLEDYKKAQKDYYEIEEKL
M104      EGVNRHNSEQNTWEKRYRELSEDHALLEATIDDISLENEKLKSENKKNLE
M33         EEHEKVTQAREAVIREMQQRGTNFGPLLASTMRDNHNLKETLDKTKKEID
M41.2       EGNARLAQAQEEALRDVLNNTPHNQLRDAYAGAFRRNNELEKIIQEKERE
M43.2       EEHPDVVAARESVLNNVRVPGTLWLRQKEENDKLKSEKKGLETELQEKEQ
M52         DQPVDHHRYTEANDAVLQGRTVSARALLHEINKNGQLRSENEELKADLQK
M64         DRLHPGYTAANNAARNEFLVPAGAVLHEREKNDELRLKNEELKADLQKKE
M68.4       EEVKKAEEVKKAEESESKSAAKMWENMYKELDRDYSLLEKTVENMSLENM
M70         EEHESVTRAREAAIREMMRQGRGDFAPLLANAIRDNNNLTETLDKTKKEI
M72         NRADDARREVLRGQFVEAELWHHQIQENDQLKLEKEELKSDLQKKEQELK
M74         FTVTRSMTRDYLAKVVQDFDTKNHELETHNSELSATNQTLQGQVEAEQKK
M75         EEERTFTELPYEARYKAWKSENDELRENYRRTLDKFNTEQGKTTRLEEQN
M80         HQLADAARREVLKGETVPAHLWYYQKEENDKLKSANEELETTLQKKEQEL
M82         DSSSRDITEAGVSKFWKSKFDAEQNRANELEKKLSGYEKDYKTLEQEYEN
M83.1       DNPRYTDAHNAVTQGRTVPLQNLLHEMDKNGKLRSENEELKADLQKKEQE
M86         DNVGRVDVDKIREEALHQAIGGMTNVQLRNTLAGSFRMNDELKKAIQEKE
M91         ADDHPGAVAARNDVLSGFSVPGNVWYRQHQEIGKLKSEKEELETELQEKE
M98         DRYTDAHNAVTQGRTVPLRNLLLEMDKNSKLRSENEELQAGLQEKERELE
M101        ADHPSYTAAKDEVLSKFSVPGHVWAHEREKNDKLSSENEGLKAGLQEKEQ
M107        AEAQAQAQAEAKAEAKAPAPAKAPAKAQTREKQLLLLEEYRKLEEGYFNL
M108        KEHESVTRAREAAIRQMMQQGGRDFAPLLADTIRDNNNLRETLDETKKEI
M116.2      DHPLYTAANNAVRNGLSPSDRAVLAEIDKNDKLRLENKELKAGLQEKEQE
M119.2      DQPNHPRYTDANNAVRNGLSPRDRAVLAEIDKNDKLRLENEKLKAGLEEL
M120        DDNPRYTAAQDEVLRELPGQAQAFSRAFLYERQKNGELRLENEGLKTALQ
M121        DQPNHPGYTEANNAVLNGYSVPLRYWAHEREKNDKLSSENEELKAGLQKK
M123        AENHPLAESARRQVLGESTVPASAWYYQKEENDKLKSENEGLKTDLQGKE
```

Figure 4.12. C4BP-binding M protein HVRs that cannot be classified as belonging to either M2/M49 or M22/M28 classes.

Residues falling at core d positions of the heptad register are highlighted in blue.

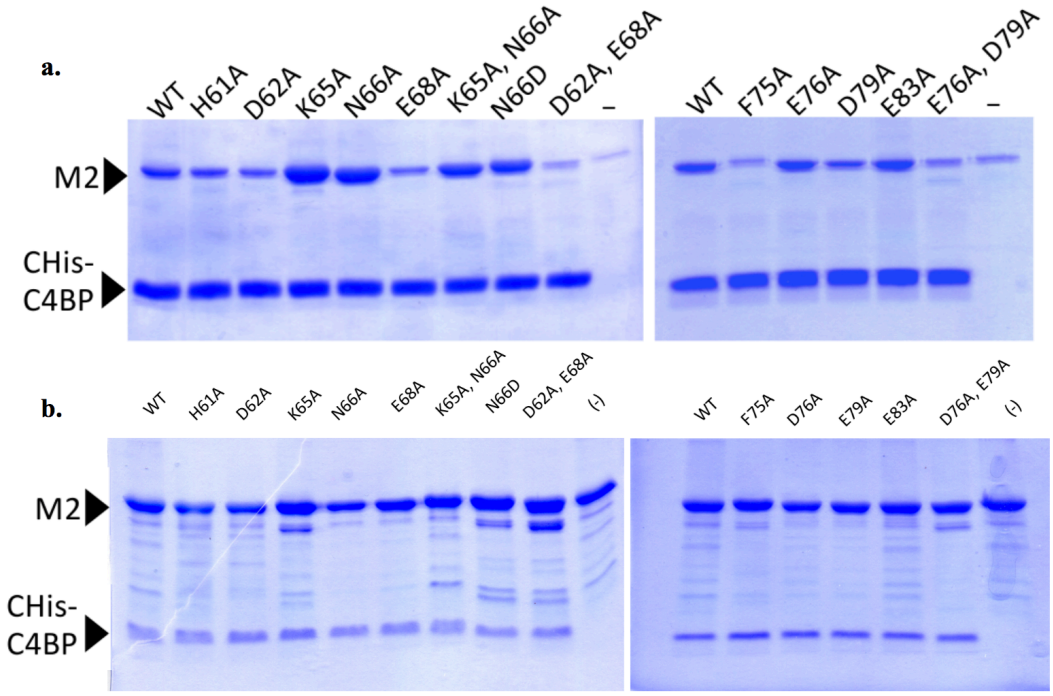


Figure 4.13. Mutational analysis of C4BP-M2 interactions.

a. Association of His-tagged C4BPα1-2 with wild-type and mutant M2 HVR at 37 °C, as assessed by a Ni^{2+}-NTA agarose coprecipitation assay and visualized by non-reducing, Coomassie-stained SDS-PAGE. **b.** Experimental inputs for His-tagged C4BPα1-2 with wild-type and mutant M2 HVR coprecipitation assay. Inputs visualized by non-reducing, Coomassie-stained SDS-PAGE.

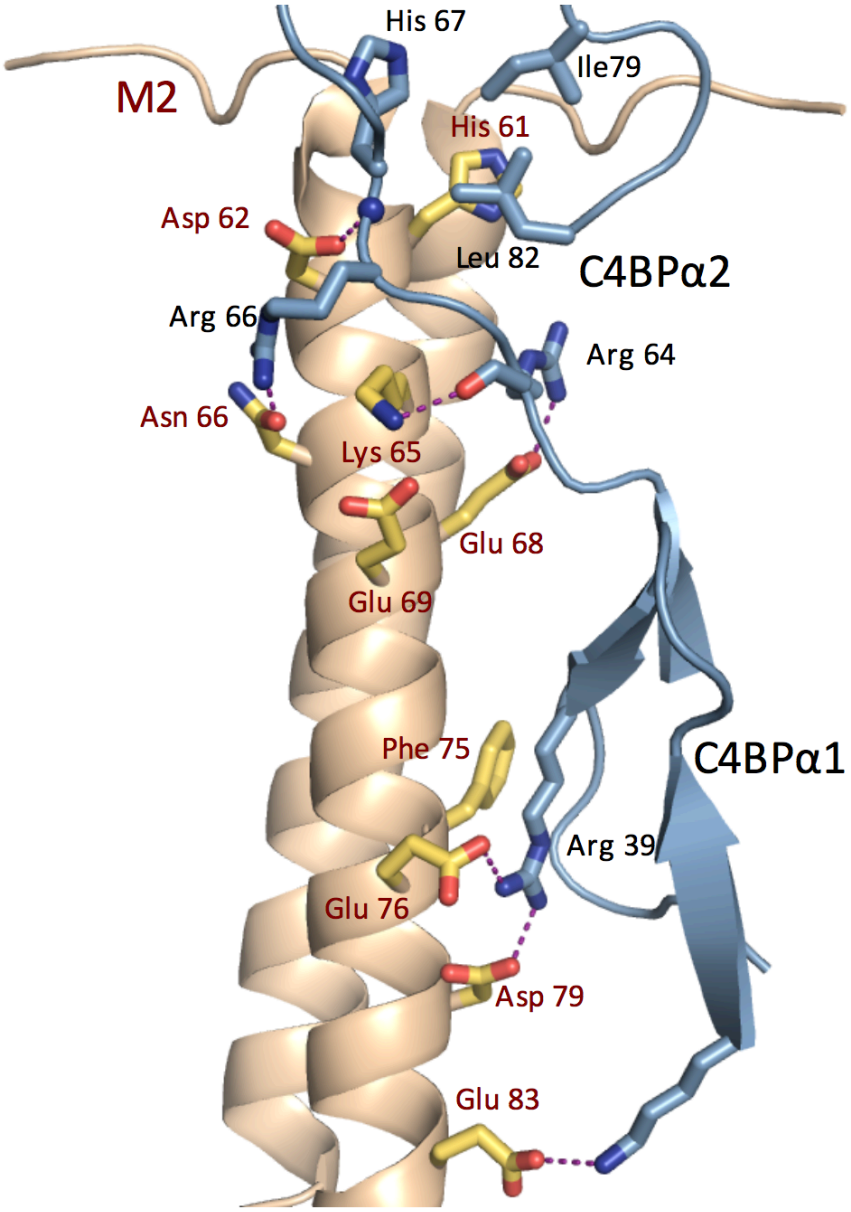


Figure 4.14. Structure of M2^{HVR}-C4BPα1-2

Structure of the M2^{HVR} (gray ribbon representation with key sidechains in bonds representation, for which carbons are yellow, oxygens red, and nitrogens blue) bound to C4BPα1-2 (cyan ribbon representation, with key sidechains in bonds representation, for which carbons are cyan, oxygens red, and nitrogens blue). Hydrogen bonds and salt bridges depicted by dashed magenta lines.

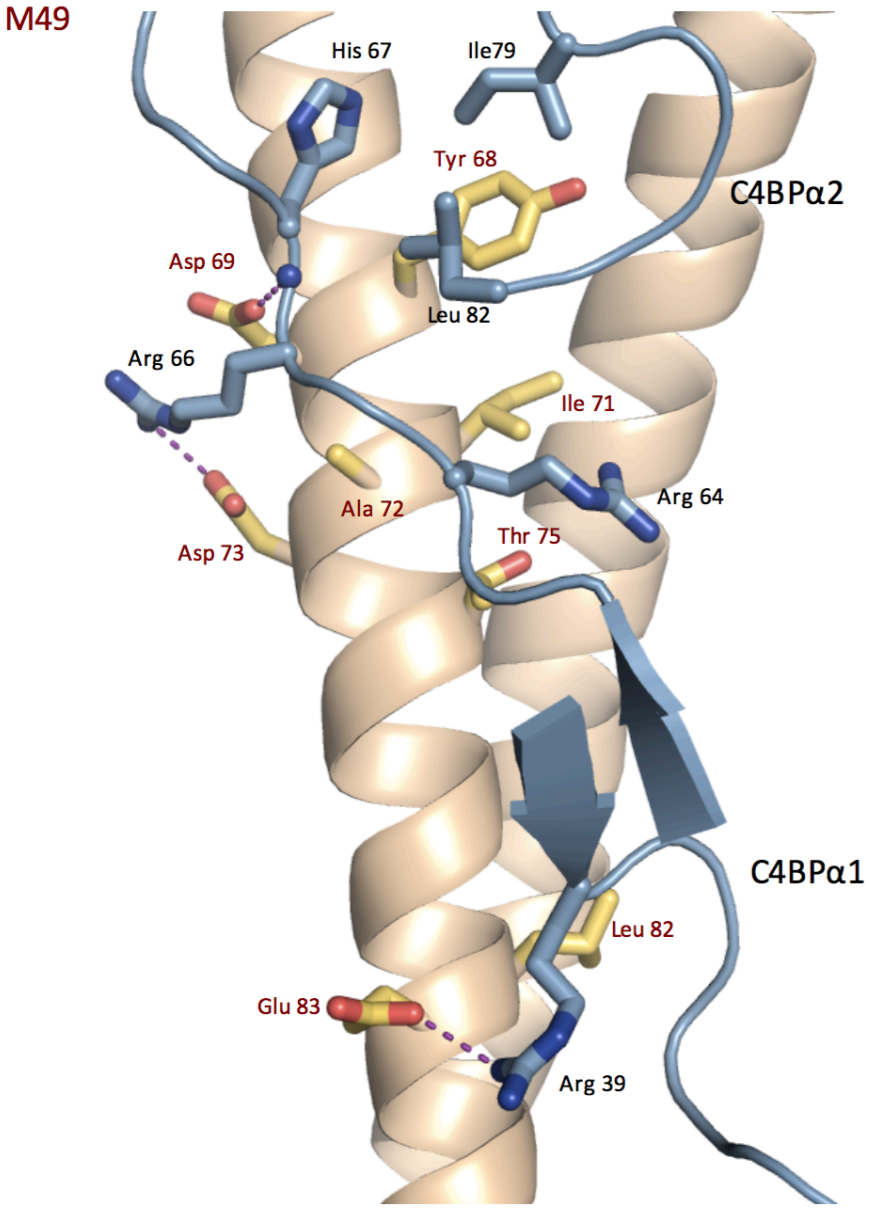


Figure 4.15. Structure of M49^{HVR}-C4BPα1-2.

The depiction is as in Figure 4.14.

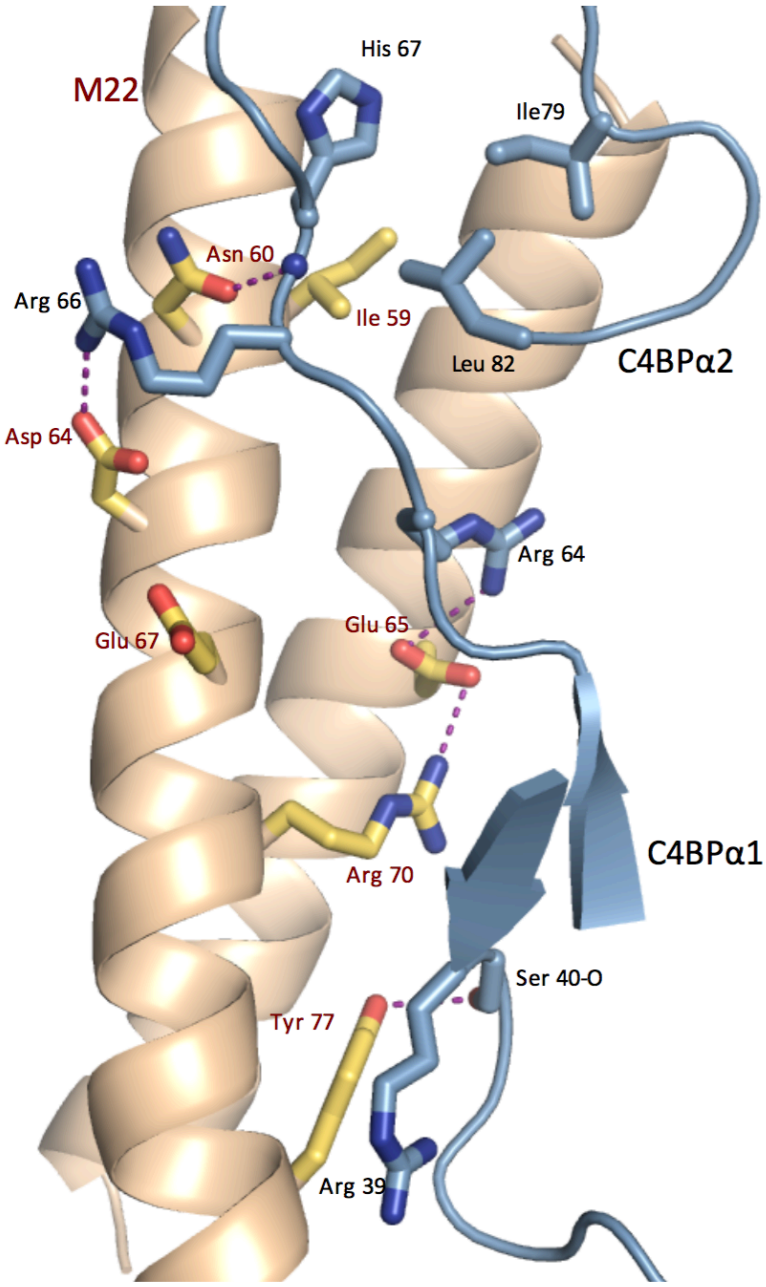


Figure 4.16. Structure of M22^{HVR}-C4BPα1-2a.

The depiction is as in Figure 4.14.

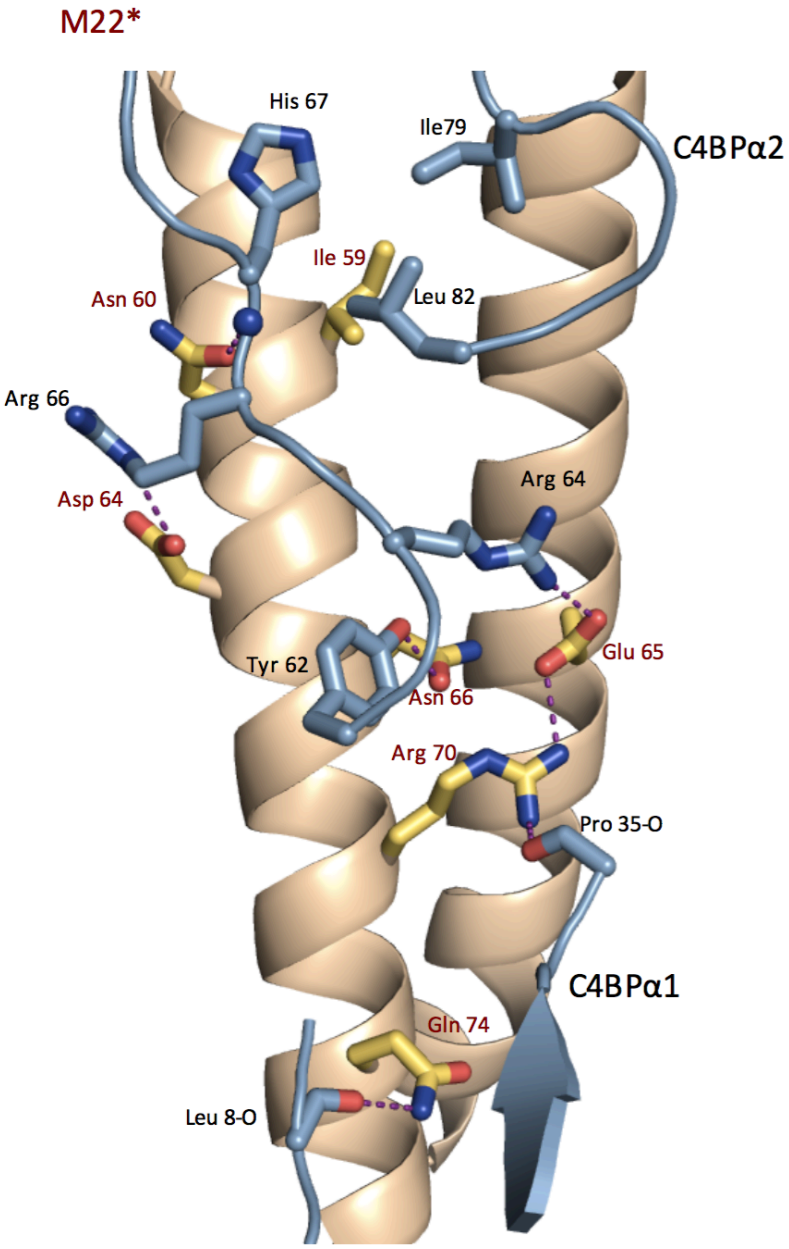


Figure 4.17. Structure of M22*^{HVR}-C4BPα1-2a.

The depiction is as in Figure 4.14.

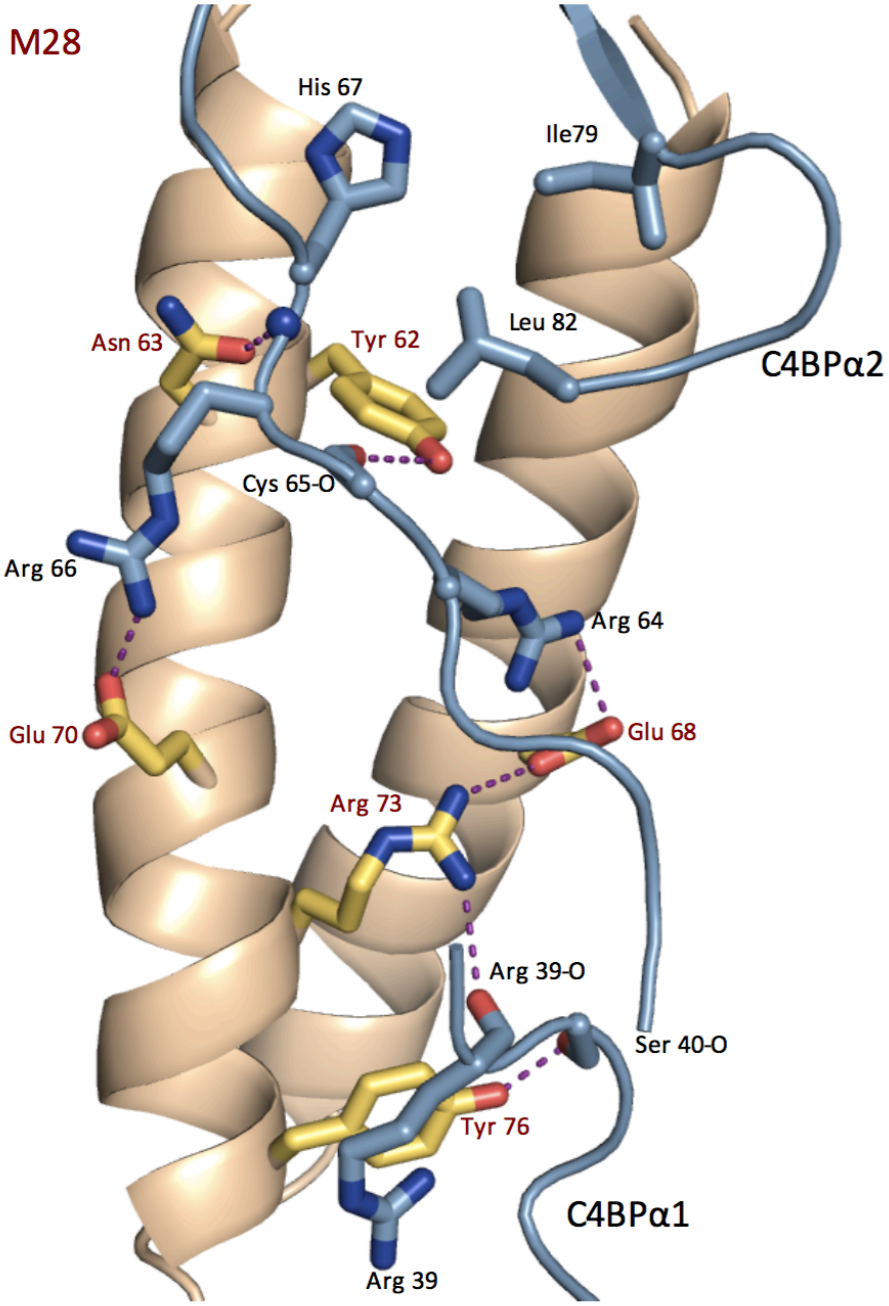


Figure 4.18. Structure of M28^{HVR}-C4BPα1-2a.

The depiction is as in Figure 4.14.

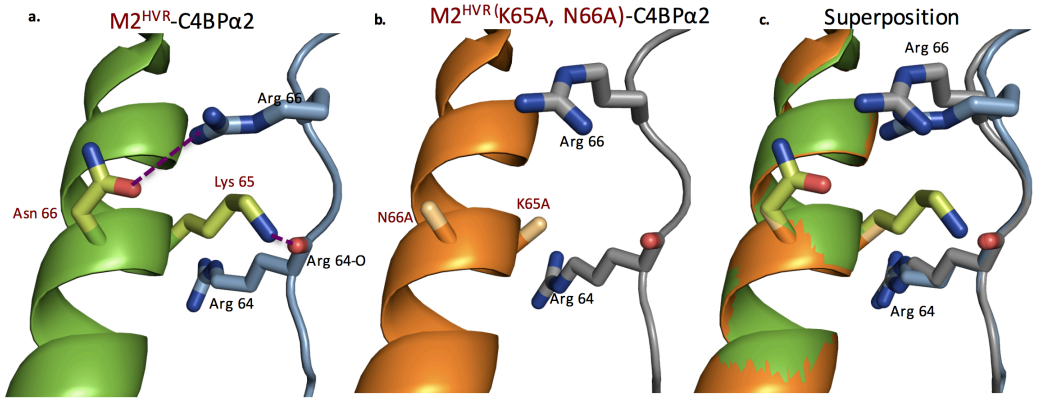


Figure 4.19. Interactions of $M2^{HVR}$ and $M2^{HVR}$(K65A, N66A) with C4BPα2.

a. $M2^{HVR}$ is depicted as a green ribbon, with Lys65 and Asn66 in bonds representation. C4BPα2 is in cyan ribbon representation, with Arg64, the mainchain carbonyl of Arg64, and Arg66 in bonds representation. Hydrogen bonds depicted as red dashed lines. **b.** $M2^{HVR}$(K65A, N66A) is depicted as a gold ribbon, with Ala65 and Ala66 in bonds representation. C4BPα2 is in gray ribbon representation, with the same groups as in panel a shown. **c.** Superposition of the structures shown in panels a and b.

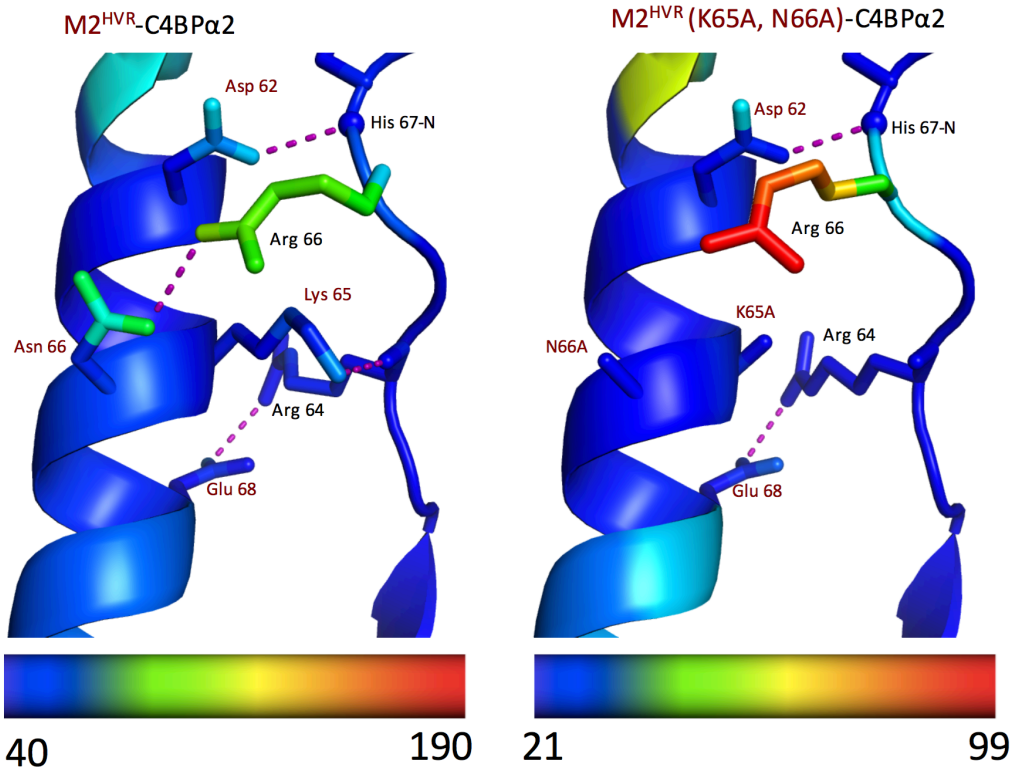


Figure 4.20. B-factors of M2^{HVR} and M2^{HVR}(K65A, N66A) with C4BPα2.

B-factors of M2^{HVR} (left) and M2^{HVR}(K65A, N66A) (right) as represented by color spectrum. Representations center at M2 Lys65 Asn66 and corresponding alanine substitution mutants highlighting differences in B-factors at these and surrounding residues. All resides are shown in bonds representation with the Cα backbone shown in ribbon representation.

REFERENCES

ANDRÉ, I., PERSSON, J., BLOM, A. M., NILSSON, H., DRAKENBERG, T., LINDAHL, G. & LINSE, S. 2006. Streptococcal M protein: Structural studies of the hypervariable region, free and bound to human C4BP. *Biochemistry,* 45**,** 4559-4568.

BLOM, A. M., BERGGÅRD, K., WEBB, J. H., LINDAHL, G., VILLOUTREIX, B. O. & DAHLBÄCK, B. 2000. Human C4b-binding protein has overlapping, but not identical, binding sites for C4b and streptococcal M proteins. *Journal of immunology (Baltimore, Md. : 1950),* 164**,** 5328-5336.

EDGAR, R. C. 2004a. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics,* 5**,** 113.

EDGAR, R. C. 2004b. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res,* 32**,** 1792-7.

PERSSON, J., BEALL, B., LINSE, S. & LINDAHL, G. 2006. Extreme sequence divergence but conserved ligand-binding specificity in Streptococcus pyogenes M protein. *PLoS pathogens,* 2**,** e47.

SANDERSON-SMITH, M., DE OLIVEIRA, D. M., GUGLIELMINI, J., MCMILLAN, D. J., VU, T., HOLIEN, J. K., HENNINGHAM, A., STEER, A. C., BESSEN, D. E., DALE, J. B., CURTIS, N., BEALL, B. W., WALKER, M. J., PARKER, M. W., CARAPETIS, J. R., VAN MELDEREN, L., SRIPRAKASH, K. S., SMEESTERS, P. R. & GROUP, M. P. S. 2014. A systematic and functional classification of Streptococcus pyogenes that serves as a new tool for molecular typing and vaccine development. *J Infect Dis,* 210**,** 1325-38.

Chapter 5

Discussion

DISCUSSION

Vaccination is historically one of the most important medical interventions for the prevention of infectious diseases. Unfortunately, due to the antigenic variability of the surface-exposed M protein, the development of a broadly protective GAS vaccine has been elusive. In this dissertation I have presented a novel approach to GAS vaccine development that takes advantage of the broad specificity of C4BP for the HVR of GAS M protein. By comparative structural analysis of the co-crystal structures of four M^{HVR}-C4BP complexes, I identified a structurally conserved C4BP 'reading head' responsible for recognizing a broad range of M protein HVRs. This 'reading head' is made up of a C4BPα2 'quadrilateral' of contacts (Fig. 4.7) and a C4BPα1 'hydrophobic nook' (Fig. 4.8). The thorough comparative analysis of multiple M^{HVR}-C4BP interactions presented here could potentially be applicable to the eventual design of broadly neutralizing antibodies.

All four structures show a domain shift of C4BPα1 with regards to C4BPα2 when bound to M protein compared to the unbound state (Fig. 4.5). This shift confirms the previously predicted domain movement that was based on chemical shift perturbations observed between the M-protein bound and unbound C4BPα1-2 NMR spectra (Jenkins et al., 2006). Our structures show that this domain shift is a ~180° rotation around Lys63. This rotation positions the

C4BPα1 'hydrophobic nook' such that it contacts a hydrophobic residue of the M protein HVR (Fig. 4.8). The importance of Lys63 had previously been demonstrated in the M22-C4BP interaction, as a K63Q substitution mutation showed a reduction in binding (Blom et al., 2000). A similar domain shift was also predicted for C4b binding as well since the same K63Q mutation also attenuated C4b binding. Without the rotation, Arg39 is spatially distal to the other residues shown to be involved in C4b binding (i.e. Arg64 and His67) (Blom et al., 2000). These data suggest that there are shared characteristics between C4b and M protein binding despite the binding sites not being identical (Blom et al., 2000).

The structures I present here might suggest that contacts by C4BPα2 are the primary contributors to binding as all the structures show a majority of residue contacts (Fig. 4.7) and buried surface area is at the M-C4BPα2 interface. However, previous studies (Jenkins et al., 2006) as well as our mutagenesis results (Fig. 4.13) have shown that C4BPα1 also plays an important role in binding. There is no evidence to suggest that the C4BPα1 domain rotation occurs in the unbound state (Jenkins et al., 2006), implying that an initial binding interaction may occur at C4BPα2, which initiates the rotation of C4BPα1 which eventually comes to rest when the Arg39 'nook' comes into contact with the hydrophobic residue of the HVR. The fact that one of the C4BPα1-2 molecules bound to the $M22^{HVR}$ dimer in the $M22^{HVR}$-C4BPα1-2 co-crystal structure is blocked from rotating due to crystal contacts (Fig. 4.5 and 4.17) also suggests that

this rotation only comes after an initial binding event centered at C4BPα2. It also suggests that for M22-C4BP binding the C4BPα1 interaction is the more secondary of the two C4BPα1-2 domain interactions as a C4BPα1 alternate binding interaction is tolerated in the crystal structure (Fig. 4.17).

Previous binding studies have been performed on the M22 protein, for which we have a crystal structure, which implicate many of the residues we observe in our binding mode. In particular, C4BPα2 Arg64, Arg66 and His67 and C4BPα1 Arg39 have all been substituted with Gln and showed changes in binding affinity (Blom et al., 2000). Two of these mutants (R64Q and H67Q) showed attenuation in binding whereas the other two (R66Q and R39Q) showed an increase in binding, albeit for reasons that are not entirely clear. The potential flaw of these binding studies is that the substituted glutamine might introduce unpredictable gain of function interactions with residues not normally involved with M^{HVR}-C4BPα1-2 binding. For this reason, I chose alanine substitution mutations in my $M2^{HVR}$ binding studies as they represent a side chain deletion. The binding interaction studies I performed on the $M2^{HVR}$ showed that an alanine substitution at Asn66, the residue that interacts with C4BP Arg66 in the 'quadrilateral', showed an increase in binding (Fig. 4.13). This suggests that not all quadrilateral interactions are ideal, yet they can be tolerated without fully disrupting binding. As antigenic variation is a competing force in GAS evolution, such tolerance may be necessary for C4BP to bind to such an array of M-types with broad specificity. My mutagenesis analysis also shows that single $M2^{HVR}$

disruptions in the quadrilateral are not sufficient to attenuate binding, as binding is maintained despite alanine substitutions at any of the four points of contact (Fig. 4.13). This observation, along with the case of M49, where a polar residue is absent to interact with Arg64, and instead Arg64 makes hydrophobic contacts (Fig. 4.7), further supports the idea of tolerance in the 'reading head'. Regarding Arg39, it should be noted that in the M22 structure we observed two distinct binding modes at C4BPα1 in the crystal structure (Fig. 4.16 and 4.17), suggesting that Arg39 may not be as important in this particular M-C4BPα1-2 interaction. This correlates with the R39Q substitution that increases binding to M22 (Blom et al., 2000). Our mutagenesis binding data of the M2-C4BP interaction suggest that the polar Arg39 makes a key contribution to binding. Combined, these data suggests that (1) glutamine may satisfy the same hydrophobic requirements as an arginine in the 'hydrophobic nook' for the M22/M28 binding mode, which doesn't show salt bridging to Arg39, and (2) the polar properties of arginine are not necessary in the M22/M28 binding mode compared to that of the M2/M49 binding mode where salt bridging with Arg39 is necessary (Fig 4.13). The data also support the idea of C4BP tolerance in recognizing different M protein binding modes, and just how unique each individual binding mode may be.

Two different M protein binding modes were identified, M2/M49 and M22/M28. The most striking characteristic that distinguishes them is the fact the M2/M49 mode is able to satisfy the complimentary 'quadrilateral' of contacts with residues present on a single M protein of the coiled-coil dimer, whereas the

M22/M28 mode requires residues from both M proteins of the dimer. This structural difference results in different heptad positions for the interacting residues of the separate modes. This fact may have contributed to the inability to identify conserved binding modes across the many different M protein HVRs by sequence analysis alone (Persson et al., 2006). In addition, a further 46 M protein HVRs were not assignable to either mode. Because the C4BP 'reading head' is common to both M2/M49 and M22/M28 mode recognition, I suggest that other M protein HVRs will also be bound in a similar manner. To identify new, potentially unique binding modes, further structural investigation into these 46 M protein HVRs is required. Such structural analysis may offer additional insight into the tolerance of the C4BP 'reading head' for the many HVRs that bind C4BP.

Acknowledgements

The text of this chapter, in full, is material currently being prepared for submission for publication. I am the principle researcher/author on the paper: Broad recognition of group A *Streptococcus* M protein hypervariability by human C4BP. Cosmo Z. Buffalo, Adrian J. Bahn-Suh, Tapan Biswas, Victor N. Nizet, and Partho Ghosh.

Partho Ghosh contributed significantly to the analysis of the data presented here.

REFERENCES

BLOM, A. M., BERGGÂRD, K., WEBB, J. H., LINDAHL, G., VILLOUTREIX, B. O. & DAHLBÄCK, B. 2000. Human C4b-binding protein has overlapping, but not identical, binding sites for C4b and streptococcal M proteins. *Journal of immunology (Baltimore, Md. : 1950),* 164**,** 5328-5336.

JENKINS, H. T., MARK, L., BALL, G., PERSSON, J., LINDAHL, G., UHRIN, D., BLOM, A. M. & BARLOW, P. N. 2006. Human C4b-binding protein, structural basis for interaction with streptococcal M protein, a major bacterial virulence factor. *The Journal of biological chemistry,* 281**,** 3690-3697.

PERSSON, J., BEALL, B., LINSE, S. & LINDAHL, G. 2006. Extreme sequence divergence but conserved ligand-binding specificity in Streptococcus pyogenes M protein. *PLoS pathogens,* 2**,** e47.

Chapter 6

Future Directions

FUTURE DIRECTIONS

The current era of vaccine development has many tools at its disposal to identify new and novel vaccine candidates. Historically, vaccines were typically made of crude mixtures of inactivated or attenuated disease agents. However, over the past couple decades, important technological and computational advances have enabled considerable progress in the design and discovery of immunogenic, recombinant protein, vaccine antigens. Genomic science gave rise to the field of reverse vaccinology. Reverse vaccinolgy starts with genomic information and attempts to identify gene targets applicable to vaccine development in silico, without the need of cultivating the pathogen. Reverse vaccinology, aided by whole genome sequencing, DNA micro arrays, proteomics, and bioinformatics have allowed rapid discovery and identification of putative virulence factors, cell surface associated proteins and potential vaccine candidates from an array of bacterial species including GAS. Despite the advent of such high throughput techniques mentioned above, the development of a broadly protective GAS vaccine has remained elusive.

We have taken a structural biology approach to GAS vaccine design. Major advances in structural biology have yielded molecular insights into the immunogenic determinants defining protective antigens, enabling their rational optimization. A structure-based design could potentially allow for the

modification of antigens to make them better immunogens. Examples have been described where, on the basis of structural information, the sequence of an antigen was altered to make it a stronger immunogen. For example, the surface of factor H-binding protein of *Meningococcus* was engineered to contain non-overlapping epitopes from three meningococcal antigenic variants, which resulted in a single molecule that was able to induce protective antibodies against all sequence variants (Scarselli et al., 2011).

Not unique to GAS, antigenic variation continues to present a huge obstacle in vaccine design, even with the most comprehensive structural understanding of the proteins involved. For example, the prospect of a human immunodeficiency virus-1 (HIV-1) or an influenza vaccine that offers long term protective immunity has been incredibly challenging, as these viruses have evolved a multitude of mechanisms to evade humoral immunity. Understanding the underlying reasons for the difficulty in eliciting protective immunity is necessary, and ultimately, a new and novel approach to vaccine design may be required. With regards to all three of these diseases, greater understanding has begun to take shape in the form of direct structural analysis of the proteins involved in eliciting neutralizing antibodies and the identification of conserved regions between the multitude of different strains.

In the case of HIV-1, the error prone DNA-dependent RNA polymerase activity of HIV-1 reverse transcriptase and HIV's ability to undergo RNA recombination give the virus the ability to generate extraordinary diversity

(Korber et al., 2000). This has proven to be prohibitive in the development of an effective vaccine as, like GAS, most of the antibodies elicited against HIV-1 are strain specific and not directed to the highly conserved regions of the virus. Compounding further the problem of genetic variation in HIV-1 vaccine design is the constant genetic mutation of the virus within the infected host long after initial infection such that within a single individual, HIV-1 still exists as a transient target to the immune system. The surface protein which displays this extensive antigenic variability is the viral envelope protein (Env).

Env is the major surface exposed antigen of HIV-1 and responsible for viral attachment to CD4 on host immune cells to initiate infection. Beyond just antigenic variation, Env has also developed an array of biochemical features to evade neutralization, such as heavy glycosylation with host derived carbohydrates, which the immune system recognizes as 'self', effectively masking the virus. In order to get around such evasive mechanisms, vaccine researchers have used structural information to locate the conserved regions of Env (Burton et al., 2004, Nabel et al., 2011). As a consequence, such structural knowledge has allowed researchers to artificially alter the surface of the Env protein by targeted removal of specific regions of the protein so that only the region of interest, the conserved CD4 binding site, is presented to the immune system (Kwong et al., 1998, Zhou et al., 2007). Such a modified version of HIV-1 is non-infectious and only exposes highly conserved regions that are targets for broadly neutralizing antibodies.

Using the modified Env protein as an antigenic target has allowed for the isolation of B cells from people infected with HIV that recognize this region specifically. The B cell library compiled has led to the identification of certain B cells that produce antibodies that recognize the conserved Env region specifically and neutralize over 90% of circulating strains by preventing viral binding to the target cell (Wu et al., 2010, Zhou et al., 2010, Wu et al., 2011). Such antibodies, combined with an array of neutralizing antibodies identified by deep sequencing (Zhou et al., 2010) suggest that broadly neutralizing antibodies could be elicited in humans and offer protective efficacy (Kwong et al., 2011).

Similar to HIV-1 virus, the problem of frequent viral antigenic change significantly obstructs the development of an influenza vaccine that offers long-term protective immunity. Despite the genetic diversity being considerably less that that of HIV-1, current influenza vaccines need to be reformulated against circulating strains annually. Given such vaccine limitations, a universal influenza vaccine capable of conferring broad cross-protection against multiple subtypes of influenza is a desirable goal. Like HIV-1 Env and GAS M protein, the influenza surface-associated protein hemagglutinin (HA) is the main antigen required for protective immunity. Historically, vaccines were directed against the variable globular head domain of HA. Yet since the discovery of antibodies specific to a highly conserved stalk region of HA, the goal of eliciting broadly protective antibodies against this region has been a primary target for development of an influenza vaccine (Wong and Webby, 2013).

Like HIV-1, multiple monoclonal antibodies with broadly neutralizing efficacy were discovered in B-cells libraries compiled from individuals infected with influenza. These antibodies were directed against the conserved HA stalk region (Ekiert et al., 2009, Ekiert et al., 2011, Corti et al., 2011, Russell, 2011, Pica et al., 2012). Similar to the conserved region of the HIV-1 Env protein, the HA stalk domain is mostly shielded from the immune system by the variable and immunodominant head domain during natural infection or conventional influenza vaccination. In order to elicit the desired immune response, augmented exposure of the stalk domain to the host immune system through antigen design is crucial. Using structural information and similar to the targeted removal of specifc regions of of the HIV-1 Env protein in order to expose the conserved regions, a current stalk-oriented antigen design strategy includes truncating HA so that it lacks the antigenic globular head domain (Steel et al., 2010). Employing such a direct structure-based approach to influenza and HIV-1 vaccine design forces an immune response that does not otherwise occur in nature, and thus the prospect of penetrating the defenses that antigeniccally variant pathogens have evolved becomes more real.

When in comes to the rational design of broadly neutralizing antibodies to GAS, similar approaches have been taken to that of HIV-1 and influenza. The isolation and vaccination with the M protein conserved C region is an example of such a similar approach. However, these conserved regions generally have lower immunogenicity than the variable regions (Penfound et al., 2010, Bontjer et al.,

2013, Jang and Seong, 2014). Thus, this thesis attempts to expand further on the structural understanding of the M protein and target the antigenically variant regions of the M protein with regards to GAS vaccine design instead of avoid them. With the work presented in this thesis, I suggest that the rational structural design of broadly neutralizing antibodies against GAS is possible by direct molecular mimicry of the M^{HVR}-C4BPα1-2 interaction. Using the information from our structural analysis of the M^{HVR}-C4BPα1-2 interaction, it may be possible to design a vaccine against GAS with broad protection despite the hurdle of substantial antigenic variation. Using the continually improving structural and computational tools available, such as molecular modeling and molecular dynamics simulations, optimizing such a broadly neutralizing antibody becomes all the more possible.

Of the 90 C4BP binding M HVRs we analyzed, 46 did not fit into one of the two separate C4BP motifs that we identified (M2/49 or M22/28). Of the 46 C4BP binding HVRs yet to be assigned to a binding mode, some show regular heptad repeats similar to that of the two identified binding modes. It is a possibility that with such regularity, other binding modes may be identified by co-crystal structural determination. However, many show irregular heptad repeats and extensive sequence variability. This lack of heptad periodicity in many C4BP binding HVRs suggests that the non-ideality of the coiled coil may play a critical role for specific M types in recognizing the 'reading head' of C4BP. In order to

investigate this possibility, further M^{HVR}-C4BPα1-2 co-crystal structures will need to be determined for M types with such non-ideality of the coiled coil.

With further understanding of the M^{HVR}-C4BP interaction based on new co-crystal structure analysis, the ability to rationally design broadly neutralizing antibodies targeting C4BP binding HVRs will only be enhanced. However, even with the current knowledge based on the four structures I have presented here, the beginning stages of design are possible by translating the HVR 'reading head' of C4BP on the variable loops of an IgG. These rationally designed broadly neutralizing antibodies could then be used as a form of passive immunization to be administered to elicit opsonization. Factors such as electrostatics, which may be responsible for bringing C4BP and M proteins into proximity for interaction, may also need to be calculated into the design. Also, if such an approach is to be successful, the tolerance of the HVR 'reading head' of C4BP will have to be accounted for, as this is responsible for C4BP recognizing a broad range of HVRs. It should be noted that since C4BP binds human C4b and M proteins by different modes, the risk that such a rationally designed IgG would bind C4b and regulate complement are slim (Blom et al., 2000). Eventually, this IgG could be tested for its ability to bind M protein through one of the many different analytical tools available.

Of potentially greater value would be a vaccine antigen that could elicit the natural production of antibodies that mimic the HVR 'reading head' of C4BP. Using tools such as gene targeted phage display to generate random epitopes that

are subsequently tested for optimal binding to our rationally designed IgG, we could identify possible antigens that could elicit broadly neutralizing antibodies. Upon vaccination, such an antigen could potentially offer sustained immunity over the course of one's lifetime as compared to the passive immunity mentioned above. Such an antigen could be administered as a vaccine independently or in combination with other GAS antigens. As C4BP does not recognize all M protein HVRs, the addition of such a peptide to an already broadly neutralizing, multivalent vaccine could potentially offer expanded coverage against most if not all M types. Using the many structural and computational tools available, this antigen could be optimized for increased immunogenicity. Once designed, such an antigen would be tested in murine mouse models to determine if it could elicit an immune response that was opsonic, bactericidal, and protective. Ultimately, this antigen could prove to be an invaluable addition to the incredibly challenging field of GAS vaccine design.

REFERENCES

BONTJER, I., MELCHERS, M., TONG, T., VAN MONTFORT, T., EGGINK, D., MONTEFIORI, D., OLSON, W. C., MOORE, J. P., BINLEY, J. M., BERKHOUT, B. & SANDERS, R. W. 2013. Comparative Immunogenicity of Evolved V1V2-Deleted HIV-1 Envelope Glycoprotein Trimers. *PLoS One,* 8**,** e67484.

BLOM, A. M., BERGGÅRD, K., WEBB, J. H., LINDAHL, G., VILLOUTREIX, B. O. & DAHLBÄCK, B. 2000. Human C4b-binding protein has overlapping, but not identical, binding sites for C4b and streptococcal M proteins. *Journal of immunology (Baltimore, Md. : 1950),* 164**,** 5328-5336.

BURTON, D. R., DESROSIERS, R. C., DOMS, R. W., KOFF, W. C., KWONG, P. D., MOORE, J. P., NABEL, G. J., SODROSKI, J., WILSON, I. A. & WYATT, R. T. 2004. HIV vaccine design and the neutralizing antibody problem. *Nat Immunol,* 5**,** 233-6

CORTI, D., VOSS, J., GAMBLIN, S. J., CODONI, G., MACAGNO, A., JARROSSAY, D., VACHIERI, S. G., PINNA, D., MINOLA, A., VANZETTA, F., SILACCI, C., FERNANDEZ-RODRIGUEZ, B. M., AGATIC, G., BIANCHI, S., GIACCHETTO-SASSELLI, I., CALDER, L., SALLUSTO, F., COLLINS, P., HAIRE, L. F., TEMPERTON, N., LANGEDIJK, J. P., SKEHEL, J. J. & LANZAVECCHIA, A. 2011. A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins. *Science,* 333**,** 850-6.KORBER, B., MULDOON, M., THEILER, J., GAO, F., GUPTA, R., LAPEDES, A., HAHN, B. H., WOLINSKY, S. & BHATTACHARYA, T. 2000. Timing the ancestor of the HIV-1 pandemic strains. *Science,* 288**,** 1789-96.

EKIERT, D. C., BHABHA, G., ELSLIGER, M. A., FRIESEN, R. H., JONGENEELEN, M., THROSBY, M., GOUDSMIT, J. & WILSON, I. A. 2009. Antibody recognition of a highly conserved influenza virus epitope. *Science,* 324**,** 246-51.

EKIERT, D. C., FRIESEN, R. H., BHABHA, G., KWAKS, T., JONGENEELEN, M., YU, W., OPHORST, C., COX, F., KORSE, H. J., BRANDENBURG, B., VOGELS, R., BRAKENHOFF, J. P., KOMPIER, R., KOLDIJK, M. H., CORNELISSEN, L. A., POON, L. L., PEIRIS, M., KOUDSTAAL, W., WILSON, I. A. & GOUDSMIT, J.

2011. A highly conserved neutralizing epitope on group 2 influenza A viruses. *Science,* 333**,** 843-50.

JANG, Y. H. & SEONG, B. L. 2014. Options and obstacles for designing a universal influenza vaccine. *Viruses,* 6**,** 3159-80.

KWONG, P. D., MASCOLA, J. R. & NABEL, G. J. 2011. Rational design of vaccines to elicit broadly neutralizing antibodies to HIV-1. *Cold Spring Harb Perspect Med,* 1**,** a007278.

KWONG, P. D., WYATT, R., ROBINSON, J., SWEET, R. W., SODROSKI, J. & HENDRICKSON, W. A. 1998. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature,* 393**,** 648-59.

NABEL, G. J., KWONG, P. D. & MASCOLA, J. R. 2011. Progress in the rational design of an AIDS vaccine. *Philos Trans R Soc Lond B Biol Sci,* 366**,** 2759-65.

PENFOUND, T. A., CHIANG, E. Y., AHMED, E. A. & DALE, J. B. 2010. Protective efficacy of group A streptococcal vaccines containing type-specific and conserved M protein epitopes. *Vaccine,* 28**,** 5017-5022.

PICA, N., HAI, R., KRAMMER, F., WANG, T. T., MAAMARY, J., EGGINK, D., TAN, G. S., KRAUSE, J. C., MORAN, T., STEIN, C. R., BANACH, D., WRAMMERT, J., BELSHE, R. B., GARCIA-SASTRE, A. & PALESE, P. 2012. Hemagglutinin stalk antibodies elicited by the 2009 pandemic influenza virus as a mechanism for the extinction of seasonal H1N1 viruses. *Proc Natl Acad Sci U S A,* 109**,** 2573-8.

RUSSELL, C. J. 2011. Stalking influenza diversity with a universal antibody. *N Engl J Med,* 365**,** 1541-2.

SCARSELLI, M., ARICO, B., BRUNELLI, B., SAVINO, S., DI MARCELLO, F., PALUMBO, E., VEGGI, D., CIUCCHI, L., CARTOCCI, E., BOTTOMLEY, M. J., MALITO, E., LO SURDO, P., COMANDUCCI, M., GIULIANI, M. M., CANTINI, F., DRAGONETTI, S., COLAPRICO, A., DORO, F., GIANNETTI, P., PALLAORO, M., BROGIONI, B., TONTINI, M., HILLERINGMANN, M., NARDI-DEI, V., BANCI, L., PIZZA, M. & RAPPUOLI, R. 2011. Rational design of a meningococcal antigen inducing broad protective immunity. *Sci Transl Med,* 3**,** 91ra62.

STEEL, J., LOWEN, A. C., WANG, T. T., YONDOLA, M., GAO, Q., HAYE, K., GARCIA-SASTRE, A. & PALESE, P. 2010. Influenza virus vaccine based on the conserved hemagglutinin stalk domain. *MBio,* 1.

WONG, S. S. & WEBBY, R. J. 2013. Traditional and new influenza vaccines. *Clin Microbiol Rev,* 26**,** 476-92.

WU, X., YANG, Z. Y., LI, Y., HOGERKORP, C. M., SCHIEF, W. R., SEAMAN, M. S., ZHOU, T., SCHMIDT, S. D., WU, L., XU, L., LONGO, N. S., MCKEE, K., O'DELL, S., LOUDER, M. K., WYCUFF, D. L., FENG, Y., NASON, M., DORIA-ROSE, N., CONNORS, M., KWONG, P. D., ROEDERER, M., WYATT, R. T., NABEL, G. J. & MASCOLA, J. R. 2010. Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science,* 329**,** 856-61.

WU, X., ZHOU, T., ZHU, J., ZHANG, B., GEORGIEV, I., WANG, C., CHEN, X., LONGO, N. S., LOUDER, M., MCKEE, K., O'DELL, S., PERFETTO, S., SCHMIDT, S. D., SHI, W., WU, L., YANG, Y., YANG, Z. Y., YANG, Z., ZHANG, Z., BONSIGNORI, M., CRUMP, J. A., KAPIGA, S. H., SAM, N. E., HAYNES, B. F., SIMEK, M., BURTON, D. R., KOFF, W. C., DORIA-ROSE, N. A., CONNORS, M., PROGRAM, N. C. S., MULLIKIN, J. C., NABEL, G. J., ROEDERER, M., SHAPIRO, L., KWONG, P. D. & MASCOLA, J. R. 2011. Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science,* 333**,** 1593-602.

ZHOU, T., GEORGIEV, I., WU, X., YANG, Z. Y., DAI, K., FINZI, A., KWON, Y. D., SCHEID, J. F., SHI, W., XU, L., YANG, Y., ZHU, J., NUSSENZWEIG, M. C., SODROSKI, J., SHAPIRO, L., NABEL, G. J., MASCOLA, J. R. & KWONG, P. D. 2010. Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science,* 329**,** 811-7.

ZHOU, T., XU, L., DEY, B., HESSELL, A. J., VAN RYK, D., XIANG, S. H., YANG, X., ZHANG, M. Y., ZWICK, M. B., ARTHOS, J., BURTON, D. R., DIMITROV, D. S., SODROSKI, J., WYATT, R., NABEL, G. J. & KWONG, P. D. 2007. Structural definition of a conserved neutralization epitope on HIV-1 gp120. *Nature,* 445**,** 732-7.