UNIVERSITY OF CALIFORNIA, SAN DIEGO

How the Streptococcal M87 Protein Binds Human C4b-Binding Protein

A Thesis submitted in partial satisfaction of the requirements for the degree Master of Science

in

Chemistry

by

Matthew Allen McGowan

Committee in Charge:

Professor Partho Ghosh, Chair
Professor Neal Devaraj
Professor Victor Nizet

2017

The Thesis of Matthew Allen McGowan is approved and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2017

DEDICATION

This thesis is dedicated to Emily, my family, and my friends.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

GAS—Group A *Streptococcus*

HVR—hypervariable region

C4BP—C4b-binding protein

C4BPα1-2—the α1 and α2 domains of C4b-binding protein

CCP—complement control protein

His_6—hexahistidine amino acid sequence

pET28aNHisPP—pET28a plasmid containing a proceeding N-terminal hexahistidine tag and PreScission protease cleavage site

IPTG—isopropyl β-D-1-thiogalactopyranoside

PMSF—phenylmethanesulfonylfluoride

MWCO—molecular weight cut off

PBS—phosphate buffered saline

RAPD—rapid automated processing of data

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Partho Ghosh and Doctor Cosmo Buffalo for their mentorship, guidance, and patience throughout this project.

I would like to acknowledge Michael Ostertag for his work in developing an algorithm to quickly analyze M protein sequences, saving me an immeasurable amount of time.

Figures 2, 3, and 4 listed in the Appendix are reproductions of figures that appear in Conserved Patterns Hidden within Group A *Streptococcus* M Protein Hypervariability Recognize Human C4b-Binding Protein. Cosmo Buffalo, et al., Nature Microbiology, 2016. These figures are used with permission from Nature Publishing Group.

ABSTRACT OF THE THESIS

How the Streptococcal M87 Protein Binds Human C4b-Binding Protein

by

Matthew Allen McGowan

Master of Science in Chemistry

University of California, San Diego, 2017

Professor Partho Ghosh, Chair

The pathogenic bacterium *Streptococcus pyogenes*, or Group A *Streptococcus* (GAS), causes diseases throughout world. These diseases can range from mild to life threatening and leave the afflicted individuals susceptible to autoimmune response. Despite the widespread threat of GAS, no vaccine exists against it. The lack of a vaccine is due in large part to the antigenic variability of the N-terminus, or hypervariable region (HVR), of the GAS surface protein known as the M protein.

One promising approach to vaccine design is to study the interactions between the human complement inhibitor C4b-binding protein (C4BP) and the M protein, as C4BP recognizes many strains of GAS with differing HVRs. A previous study reported four crystal structures of M protein HVRs in complex with the binding domains of C4BP (C4BPα1-2), revealing conserved binding patterns between M proteins and C4BP that could be applied to about half of the GAS implicated in binding C4BP.

To continue this line of investigation, eight M protein HVRs whose sequences did not match the previous binding motifs (M14.5^{HVR}, M18.6 ^{HVR}, M58 ^{HVR}, M80 ^{HVR}, M75 ^{HVR}, M87 ^{HVR}, M112^{HVR}, PrtH^{HVR}) were cloned, expressed, and purified to obtain cocrystal complexes with C4BPα1-2. The 2.69 Å resolution limit structure of the M87^{HVR}-C4BPα1-2 complex was determined, which has further demonstrated the involvement of previously identified C4BP amino acids but has also identified unique interactions. The M87 binding pattern appears to occur in other M protein HVRs, giving further knowledge applicable to the design of a GAS vaccine.

Chapter 1: Introduction

The Medical Significance of Group A Streptococcus

Streptococcus pyogenes, often referred to as Group A *Streptococcus* (GAS), is a gram positive bacterium responsible for a variety of human diseases. In most cases, these diseases are mild and treatable with antibiotics (e.g. strep throat, impetigo, and scarlet fever). In rare cases, and most often in areas of the world where access to antibiotics and proper healthcare is limited, these infections can become invasive (e.g. necrotizing fasciitis, toxic shock syndrome, and sepsis) and present serious health risks with mortality rates of 20% or higher, even with proper treatment (Lamagni et al. 2009). GAS infection also leaves individuals vulnerable to autoimmune sequelae (post-streptococcal glomerulonephritis, rheumatic heart disease, rheumatic fever, and several CNS diseases) (Snider and Swedo 2003; Cunningham 2016).

Severe GAS infections are a global problem: a 2005 study reported 18.1 million individuals from various regions across the Earth were suffering from severe infections, with 1.78 million more cases developing annually (Carapetis et al. 2005). These infections lead to 517,000 deaths annually (Carapetis et al. 2005). Rheumatic heart disease is responsible for most of these cases (15.6 million cases and 233,000 deaths a year), and invasive GAS diseases are also prevalent (663,000 new cases annually, 163,000 deaths a year) (Carapetis et al. 2005). In Hong Kong and China, genes from the GAS M1T1 clone encoding antibiotic resistance and the M12 clone encoding scarlet fever-associated elements have been transferred between each other, leading to a resurgence of individuals suffering from scarlet fever (Davies et al. 2014; Ben Zakour et al. 2015).

The global prevalence of severe GAS infections, lack of antibiotic treatment in developing areas of the world, and the potential for further antibiotic resistance create a pressing need for a vaccine. Unfortunately, no vaccine has been developed to combat GAS infections. The lack of a vaccine is due, in large part, to the antigenic variability of the major antigen and surface-associated virulence factor of GAS, the M protein.

The Role of M protein in GAS

M proteins are bound to the cell wall of GAS and are a major virulence factor of *S. pyogenes*. M proteins form α-helical coiled-coil dimers which extend ~500 Å from the cell wall, allowing for interaction with host proteins (Fischetti 1989). The coiled coils have a heptad sequence periodicity, and maintain their structure through hydrophobic interactions between amino acids in the *a* and *d* positions of the heptad repeat (Fig. 1a). The M protein is composed of four regions, A-D, with sequence variability increasing from the C-terminus (D) to the N-terminus (A). The most N-terminal ~50 amino acids of M protein, which extend furthest into the host environment, are characterized by a high amount of sequence variability between strains (or *emm* types) of GAS, and are therefore known as the hypervariable region (HVR). The HVR is typically the region which interacts with host proteins, thus shielding GAS from opsonophagocytosis as well as preventing recognition by antibodies between strains (Metzgar and Zampolli 2011). More than 200 different HVRs have been identified, and it is this sequence that is used as a serological marker to identify strains of GAS.

The hypervariability of the M protein causes serial infections of GAS to be common and makes design of a vaccine against GAS difficult, as antibodies elicited for

one strain of GAS are often specific only to that strain (Dale et al. 2016). In addition to hypervariability, vaccine design is also stymied by the lack of an animal model that properly mimics disease in humans, the complex epidemiology of GAS, and the risk of autoimmune responses (Dale et al. 2011). Despite these obstacles, much progress has occurred in GAS vaccine design. A 2011 study resulted in a vaccine candidate recognizing 30 different HVRs when tested in rabbits (Dale et al. 2011). In addition to having bactericidal activity on the serotypes in the vaccine, there was also cross-coverage on several other non-targeted strains. This vaccine candidate, however, lacks epidemiological coverage. Though the vaccine would recognize 90% of invasive disease causing GAS in the U.S., other regions have coverage under 50% (Dale et al. 2011). To better combat GAS and overcome the hypervariability in the M protein, the ideal vaccine would elicit a highly cross-reactive antibody.

M Protein and its Interaction with C4b Binding Protein

Recent studies on the interactions between M proteins and human C4b-binding protein (C4BP) have implications for the development of a potential GAS vaccine that would offer broad coverage (Persson et al. 2006; Buffalo et al. 2016). C4BP, an inhibitor of the human innate immune system, prevents the formation of C3 convertase (C4bC3a in the classical and lectin pathways) by competitively binding C4b (Ricklin et al. 2010). As C3 convertase produces proteins that aid in and signal phagocytosis, C4BP functions to prevent damage to host tissue by controlling the production of these proteins. C4BP is composed of 7 α chains which bind C4b and one β chain which binds the anticoagulant protein S. These chains are formed by a series of linked complement control protein (CCP) domains (Ricklin et al. 2010) (Fig. 1b).

Group A *Streptococcus* is primarily eliminated via opsonophagocytosis. To combat this, many strains of GAS have developed the ability to bind C4BP. As C4BP has seven α chains which can bind C4b and disrupt the formation of C3 convertase, this recruitment leads to a decrease in phagocytic killing of GAS cells. The recruitment of C4BP to the cell surface has the added benefit of cloaking the cell from phagocytes by preventing the deposition of antibodies (Berggård et al. 2001).

C4BP appears to recognize a wide variety of GAS strains. In one study, ~88% of the GAS strains tested were found to interact with C4BP (Persson et al. 2006). This is remarkable considering C4BP is known to bind the M protein at the HVR (Berggård et al. 2001). The conserved binding to C4BP, despite such high variability in the M protein HVR, allows for a unique approach to vaccine design. By investigating the binding characteristics of these proteins, a vaccine eliciting a broadly neutralizing antibody could be developed. Though this vaccine would, at best, recognize only those M proteins which bind C4BP, it would be a great step forward in creating a comprehensive vaccine for GAS. Recent work studying the structures of four M protein HVRs ($M2^{HVR}$, $M22^{HVR}$, $M28^{HVR}$, $M49^{HVR}$) in complex with the first two domains of the C4BP α chain (C4BPα1-2) has shed light on the subtle binding patterns between M proteins and C4BP (Buffalo et al. 2016).

Previous M^{HVR}-C4BPα1-2 Complexes

Previous structural data of $M2^{HVR}$, $M22^{HVR}$, $M28^{HVR}$, and $M49^{HVR}$ in complex with C4BPα1-2 has led to the recognition of certain common characteristics in these binding interactions. Broadly, these cocrystal structures show two binding interfaces: a

major site formed by C4BPα2 and a minor site formed by C4BPα1 (Fig. 2 and 3). The interaction between M^{HVR} and C4BPα2 contains the most binding residues and has the largest buried surface area (~550 Å, slightly over twice as large as the interface with C4BPα1), thus appearing to be the main site of contact between the proteins. All M^{HVR} proteins studied interacted with the same five residues of C4BPα2, forming a "quadrilateral" (Fig. 2). This quadrilateral bound the M^{HVR}s via a hydrophobic interaction through C4BP Ile78 and Leu82, a hydrogen bond with the main chain nitrogen of C4BP His67, a hydrogen bond or salt bridge with C4BP Arg64, and a salt bridge (or in the case of M49 a hydrophobic interaction with side chain carbons) with C4BP Arg66 (Buffalo et al. 2016).

The C4BPα1 domain exhibited an 180° rotation from its unbound conformation when it associated with the M protein (Fig. 4). This rotation allowed Arg39 of C4BPα1 to form a salt bridge with a negative residue of the M protein (Fig. 3). The side chain carbons of Arg39, along with main chain residues of C4BPα1, formed a "hydrophobic nook" in which a bulky hydrophobic residue of the M^{HVR}s rested (Buffalo et al. 2016).

The binding patterns observed in these previous structures fell into two larger modes. One mode, M2/M49, had only one of the M protein α-helices of the dimeric coiled coil making contacts to and individual C4BP molecule, while the other mode, M22/M28, had both M protein α-helices of the coiled coil making contact to an individual molecule of C4BP (Fig. 2). The binding patterns were then compared to other M protein sequences. This comparison suggested 11 M proteins would bind in the fashion of M2 and M49, and 29 M proteins would bind in the fashion of M22 and M28 (Buffalo

et al. 2016). The assignment of these M proteins, however, still left ~50% of M proteins of the GAS strains demonstrated to bind C4BP uncategorized (Buffalo et al. 2016)

Chapter 2: Materials and Methods

Expression Constructs

Coding sequences for the first 100 amino acids of mature M proteins (M^{HVR}) were chemically synthesized and inserted into pUC18 (Genewiz) for M14.5 (amino acids 43-142), M18.6 (43-142), M58 (42-141), M75 (42-141), M80 (42-141), M87 (42-141), M112 (43-142), and PrtH (42-141). The coding sequences were then cloned into pET28a with sequences encoding a preceding N-terminal hexahistidine tag (His_6) and PreScission protease cleavage site (pET28aNHisPP). The M75 coding sequence contained an extra base, causing a frameshift in the sequence. Site-directed mutagenesis (Agilent QuikChange) was carried out according to the manufacturer's protocol in order to fix this error.

Protein Purification

All plasmids containing M protein coding sequences were transformed into and expressed in *Escherichia coli* BL21-Gold (DE3) and purified as described previously with some exceptions (Buffalo et al. 2016). Specifically, after induction with 2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), bacteria were grown at 18 ℃ rather than room temperature for 16 h. The lysis buffer used was 300 mM NaCl, 100 mM NaP*i* buffer, pH 7.5, and once resuspended, bacteria were treated with deoxyribonuclease (200 µg/mL), lysozyme (200 µg/mL), and 0.5 mM phenylmethanesulfonylfluoride (PMSF). Ni-NTA purification was performed without deviation. Following elution from Ni-NTA agarose beads (Qiagen), M proteins were dialyzed (MWCO 3,500 Da, regenerated cellulose membrane) in 4 L of phosphate buffered saline (PBS) at 4 ℃ for 16 h, then

transferred to 2 L of cold PBS and dialyzed for a further 4 h at 4 ℃. M proteins were then purified by size exclusion chromatography (Superdex 75) using PBS at 4 ℃.

C4BPα1-2 was expressed and purified as described previously (Buffalo et al. 2016).

Crystallization

Purified M(14.5, M58, M80, M87)^{HVR} and C4BPα1-2 proteins were mixed (~7.5 mg total protein in a volume of 1-2 mL) at 1:1 molar ratios and dialyzed overnight in 10 mM Tris pH 8.0 at 4 ℃ (Slide-A-Lyzer Dialysis Cassette 3500 MCWO, 0.5-3 mL). The resulting protein solution was concentrated to ~15 mg/mL via ultrafiltration (Amicon Ultra centrifugal filter units MWCO 3,500), as determined by a Nanodrop 1000 measuring the absorbance at 280 nm and using the calculated molar extinction coefficient. This protein solution underwent crystallographic screening using a mosquito robot implementing hanging drop vapor diffusion.

For M87^{HVR}, the initial conditions leading to crystal formation (0.2 M KH_2PO_4, 20% w/v PEG 3,350, pH 4.8) were further refined by varying pH levels and concentration of the PEG 3,350. The optimal precipitant used for the M87^{HVR}-C4BPα1-2 complex was 0.1 M sodium citrate, pH 4.8, 0.2 M KH_2PO_4, 25% w/v PEG 3350 (Fig. 5). The crystals were then scooped into nylon microloops (Hampton CryoLoops) and transferred to precipitant containing 20% glycerol as a cryoprotectant, before being cryo-cooled in liquid N_2.

M87^{HVR}-C4BPα1-2 Cocrystal Structure Determination

The M87^{HVR}-C4BPα1-2 crystals were exposed to X-ray beams at the Advanced Photon Source, beamline 24-ID-E. Diffraction data were indexed, integrated, and scaled using the software Rapid Automated Processing of Data (RAPD) at NE-CAT (https://rapd.nec.aps.anl.gov/rapd) (Table 1).

The structure of M87^{HVR}-C4BPα1-2 was determined by molecular replacement with Phenix PhaserMR using the two C4BPα1-2 molecules from the structure of the M2 K65A/N66A-C4BPα1-2 complex as the search model (Adams et al. 2010; Buffalo et al. 2016). A molecular replacement solution with a log likelihood gain of 365 was identified. Electron density maps calculated from molecular replacement phases, as visualized by the software Coot, showed evidence for M87^{HVR} α-helices in the asymmetric unit, which contained one 2:2 complex. Side chains of M87^{HVR} were initially modeled as alanines. The M87^{HVR} model and theC4BPα1-2 model were subjected first to rigid body refinement. The register of M87^{HVR} was found by identifying electron density attributable to neighboring bulky hydrophobic residues (Tyr95, Phe91, and Trp92). Model building was done manually in Coot through inspection of σ_A-weighted $2mF_o - DF_c$ and $mF_o - DF_c$ omit electron density maps followed by refinement in Phenix using default settings as well as TLS parameterization (M87^{HVR}: 58-96, 97-136; C4BP: 0-13, 14-33, 34-73, 74-86, 87-96, 97-124) (Emsley et al. 2010). The model went through ~40 cycles of rebuilding and refinement. At the last stages of refinement, waters were modeled using default parameters. Two citrate ions were modeled into the electron density map, and Phenix ReadySet! was used to generate geometry restraints for these ions (Adams et al. 2010). Electron density was continuous for all of the main chain of C4BPα1-2 except for

residues 19 and 20, and for M87 from residues 45-124. The model had a MolProbity clashscore of 12.02 (95th percentile), no Ramachandran outliers (96.38% of amino acids in favored regions), 92.35% favored rotamers and 1.98% poor rotamers, and a MolProbity score of 2.06 (97th percentile) (Chen et al. 2010) (Table.2).

Algorithm to Determine M87-like Binding Patterns.

M proteins of similar binding modes to that of M87 were determined by matching a string of residues with similar charge characteristics (positive: lysine and arginine; negative: glutamic acid and aspartic acid), hydrophobicity (valine, isoleucine, leucine, tryptophan, tyrosine, and phenylalanine), or hydrogen bonding potential (acceptors: asparagine, aspartic acid, glutamine, glutamic acid, histidine, serine, threonine, tyrosine; donors: arginine, asparagine, glutamine, histidine, lysine, serine, threonine, tryptophan, tyrosine) relative to the positions seen in the M87 sequence (Appendix II). Any matching residue was given a score of 1, residues that could only hydrogen bond where a salt bridge was observed in the M87^{HVR}-C4BPα1-2 structure were scored as 0.5, and, if applicable, residues with opposite charge characteristics were given a score of -1. M protein peptide sequences from the CDC database were then parsed and given a total score based on the properties and relative position of the amino acids in the sequence. A cut off score of 9 (with a maximum score of 14, as M87 has 14 binding residues) was chosen, as M protein segments scoring 8.5 and below often did not match the heptad periodicity of M87 in the highest scoring alignment as predicted by COILS (Lupas, Van Dyke, and Stock 1991). M protein sequence segments with high scores found outside of the HVR were not considered.

Chapter 3: Results

Selection of M proteins

In order to expand the current knowledge of M protein-C4BP interactions, eight M protein hypervariable regions (M^{HVR}) were selected as crystallization candidates. M proteins were selected based on a residue pattern that suggested a unique binding mode from that observed in the four previously determined structures (Fig. 6). Though PrtH is predicted to have a similar binding mode to that observed in the M22/M28 pattern, it was selected based on its lack of a hydrophobic residue in the region of the sequence that is predicted to bind to the C4BPα1 hydrophobic nook (Fig. 6). Similarly, M112 is predicted to bind analogously to the M2/M49 mode, but has no charged residue to bind with Arg66 of C4BP, as was seen in all previously observed interactions. The remaining six M proteins (M14.5, M18.6, M58, M80, M75, M87) were selected from a group of different *emm* type *S. pyogenes* that bound C4BP in a whole cell binding assay, but were not predicted to bind in a mode similar to that previously observed (Fig. 6) (Persson et al. 2006; Buffalo et al. 2016).

Progress with Selected M proteins

An overview of the progress with these M^{HVR} proteins can be found in Figure 7. All these M^{HVR} proteins were successfully cloned into pET28aNHisPP vectors from pUC57 vectors and transformed into *Escherichia coli* BL21-Gold (DE3) for protein expression, with the exception of $M18.6^{HVR}$. A digestion product from pUC57 matching the approximate size of the M18.6 insert was observed in agarose gels (Fig. 8), but attempts to ligate this insert into pET28aNHisPP and transform the vector into *E. coli* BL21-Gold (DE3) did not yield a positive result. The other M^{HVR}s, once ligated into

pET28aNHisPP and transformed into *E. coli* BL21-Gold (DE3), were expressed and purified.

All M^{HVR} proteins were expressed at levels detectable by SDS-PAGE of bacterial lysates (Fig. 9). Purification was successful as assessed by SDS-PAGE (Fig. 9 and 10). Initially, expression was not observed with $M75^{HVR}$. After reviewing sequencing results of its coding sequence, a frameshift error caused by a base pair insertion near the end of the sequence was discovered. Site directed mutagenesis (Agilent QuikChange) was used to obtain the correct coding sequence. After transformation of *E. coli* BL21-Gold (DE3) with this corrected plasmid, protein expression upon induction was observed.

Once these M^{HVR} proteins were purified and concentrated, crystal screening was undertaken with $M14.5^{HVR}$, $M58^{HVR}$, $M80^{HVR}$, and $M87^{HVR}$. The $M80^{HVR}$-C4BPα1-2 complex did not form crystals. $M14.5^{HVR}$-C4BPα1-2 formed small crystals (~10 microns, in 0.1 M HEPES pH 7.5, 1 M sodium acetate trihydrate, 0.05M cadmium sulfate hydrate), which could not be replicated in subsequent optimization screens (Fig. 11a.). These crystals were never exposed to X-rays. $M58^{HVR}$-C4BPα1-2 formed crystals in several conditions. The first (2.0M $(NH_4)_2SO_4$, 0.1M Tris pH 8.3) gave rise to large (~300 microns) but fairly ill-defined crystals (Fig. 11b.). After exposing these crystals to X-ray beams, the resulting diffraction pattern suggested that the crystals were indeed protein, but the reflections were so few and faint that no usable data set could be obtained. Pursuing a different condition (0.2 M $LiNO_3$, 20% v/v PEG 3350, and 0.1 M sodium citrate, pH 5.2) led to crystals with a more well-defined morphology (spikes ~100 microns long, 10 microns wide), but the X-ray data obtained from these crystals were

again too weak to be of use (Fig. 11c). $M87^{HVR}$ was the only M protein to form a complex with C4BPα1-2 from which a structure could be determined.

Crystal Structure Analysis of $M87^{HVR}$-C4BPα1-2

The structure of the $M87^{HVR}$-C4BPα1-2 complex shares similarities with the four previously determined M^{HVR}-C4BPα1-2 complexes, but has unique features as well. $M87^{HVR}$ forms an α-helical coiled-coil dimer, with one C4BPα1-2 molecule bound on either side of the dimer (Fig. 12). The coiled coil in $M87^{HVR}$ is slightly unwound while in contact with C4BPα1-2, having a pitch of ~180 Å rather than the canonical 150 Å (Fig. 13). The N-terminal portion of $M87^{HVR}$ is in contact with C4BPα2, while the C-terminal portion of $M87^{HVR}$ is in contact with C4BPα1. This orientation is reflective of how intact C4BP approaches M protein attached to the cell wall of *S. pyogenes*. The binding mode is reminiscent of that seen with M22/M28, as both M protein α-helices are in contact with a single molecule of C4BPα1-2 (Buffalo et al. 2016) (Fig. 14, 15, 16, and 17).

The total buried surface area between the $M87^{HVR}$ dimer and the two C4BPα1-2 molecules is ~1600$Å^2$ (within the 1450-1690 $Å^2$ observed in previous structures) (Buffalo et al. 2016). Like previous structures, the interface is predominantly polar in character and has a modest surface complementarity of ~0.69, within the range of the other M^{HVR}-C4BPα1-2 interfaces (Buffalo et al. 2016). The C4BPα1 domain exhibits the characteristic 180˚ rotation seen in the other M protein-C4BP complexes. This rotation aids in the binding of $M87^{HVR}$ to C4BPα1-2 as it allows Arg39 of C4BPα1 to form a salt bridge with Asp96 of M87, the only non-hydrophobic binding interaction between M87 and the C4BPα1 domain (Fig. 18 and 19).

Of note, the two C4BPα1-2 molecules bound to the $M87^{HVR}$ coiled coil differ slightly; these are named hereafter Interface 1 and Interface 2 (Fig. 14 and 16). In Interface 1, $M87^{HVR}$ is bound to five residues of C4BPα2 at six points of contact (Fig. 14 and 15), while in Interface 2, $M87^{HVR}$ is bound to six residues of C4BPα2 at six points of contact (Fig. 16 and 17). Both these sites, as seen in $M2^{HVR}$, $M22^{HVR}$, and $M28^{HVR}$, interact with Arg64 of C4BPα2 (Fig. 2, 14, and 16). Arg64 interacts with three separate residues of $M87^{HVR}$. Arg64 forms a salt bridge with its guanidinium group to Asp80, a hydrogen bond with its main chain nitrogen to Glu85, and a hydrogen bond to Tyr81. Lys63, which bridges the C4BPα1 and C4BPα2 domains, forms electrostatic interactions with Glu89 of $M87^{HVR}$. His67 of C4BP interacts in all M^{HVR}-C4BPα1-2 structures, and here its main chain nitrogen hydrogen bonds with Gln78 of $M87^{HVR}$ in Interface 1 and its imidazole group forms an electrostatic interaction with Glu75 in Interface 2 (Fig. 14 and 16).

Some interactions are unique to Interface 1 of the complex. Interactions with Arg66 of C4BPα2 are seen in all previously studied M^{HVR}-C4BPα1-2 complexes, but are observed only in Interface 1. Arg66 is hydrogen bonded to Ser82 of $M87^{HVR}$. Glu70 of C4BPα2 is involved in an electrostatic interaction with Lys66 of $M87^{HVR}$ (Fig. 14 and 15).

Interface 2 also has unique characteristics. Arg66, observed to interact in all previous HVR-C4BP structures and in Interface 1 of this structure, takes no part in Interface 2. The main chain carbonyls of C4BP Ile78 and Thr80 are hydrogen bonded to the amide nitrogen of $M87^{HVR}$ Gln73. Interestingly, this interaction takes place at a site

where only hydrophobic interactions have been observed in other complexes (Fig. 2 and 16). Interface 2 also shows a unique area of hydrophobic interaction between M87^{HVR} and C4BPα1-2 (Fig. 18 and 19). This hydrophobic pocket is in C4BPα1, very near the junction between domains. Here, the *a* and *d* residues Leu88 and Ile84, respectively, of the M87^{HVR} heptad repeat interact with C4BPα1 Val38 and Ile61, respectively (Fig. 18 and 19). Glu71 of M87 forms an electrostatic interaction with Gln75 of C4BP in Interface 2 (Fig 16 and 17).

The C4BPα1 domain binds M87^{HVR} with one electrostatic interaction, seen in both interfaces, between C4BPα1 Arg39 and M87^{HVR} Asp96 (Fig. 20 and 21). The bulky aromatic Trp92 of M87^{HVR} fits into the "hydrophobic nook" created by the alkyl chain of Arg39 and main chain residues of C4BPα1 (Fig. 20 and 21). This binding is similar to that seen with M2^{HVR}, whose Asp79 residue salt bridges to C4BP while M2^{HVR} Phe75 binds in the "hydrophobic nook". This site on C4BP has been identified in all structures to date.

The identified binding residues were used to predict which other M proteins would likely bind C4BP in a similar fashion. Six M proteins from strains of GAS shown to bind C4BP, but not matching any previously identified binding mode, appeared to fit the M87 mode (Fig. 22). In addition to the strains shown to bind C4BP, M proteins from 24 untested strains also fit this binding pattern (Fig. 23). Interestingly, M22 and several M proteins predicted in the M22/M28 binding motif also fit this predicted binding mode (Fig. 24). This match makes sense as both M87 and M22 bind with both helices to a single C4BP molecule, and both M87 and M22 bind His67, Arg64, and Arg39 in the

same relative location as described in the pattern (Fig. 24). M87 was most likely not previously categorized with M22 for two reasons. The first is that it does not have a hydrophobic residue that fits into the hydrophobic pocket of C4BPα2 as M22 does, as a glutamine is instead hydrogen bonded to the main chain carbonyls of residues in this pocket (Fig. 16 and 17). The second is that the M87 residue which binds Arg66 of C4BP, Ser82, is not in the same location of the binding pattern as the glutamate which binds Arg66 in the $M22^{HVR}$-C4BPα1-2 complex. Despite these differences, the overall binding patterns of M87 and M22 to C4BP are similar.

Chapter 4: Discussion

The identification of binding residues in the M87^{HVR}-C4BPα1-2 complex offers further information that may be invaluable towards the design of a comprehensive vaccine for GAS. The key binding residues of C4BP previously identified all take part in interactions on the M87^{HVR}-C4BPα1-2 interface, with the exception of Leu82 (Fig. 2, 14, and 16). Both M87^{HVR} helices are required to form the interface with C4BPα2—a binding mode similar to that observed in M22^{HVR} and M28^{HVR}. However, M87^{HVR} also has distinct binding interactions not seen in M22^{HVR} or M28^{HVR} which may give insight into designing a vaccine that will recognize M87 and similar M types. Notably, the differences observed between Interface 1 and Interface 2 of this complex support the idea of tolerance in the C4BP reading head, as even the same M protein may bind C4BP with multiple side chain conformations.

As seen before, The C4BPα2 domain hosts the majority of binding interactions in the complex and contains a majority of the buried surface area (Buffalo et al. 2016). The C4BPα1 domain also undergoes the characteristic 180˚ shift in order to bind to M87. These data further support the observations and predictions of previous work that the C4BPα2 domain is the major contributor of interactions with M proteins and that the C4BPα2 domain binds the M protein first, eventually leading to the domain rotation of C4BPα1 (Buffalo et al. 2016; Berggård et al. 2001). The hydrophobic interactions between C4BP Ile61 and Val31, and M87 Ile84 and Leu88, may further stabilize this orientation (Fig 15.1 and 15.2). The interaction of Arg39 and the hydrophobic nook it forms on C4BPα1 to a negative and/or hydrophobic residue of the M protein has now been observed consistently across five M^{HVR}-C4BPα1-2 structures. The conservation of

this interaction suggests the Arg39 hydrophobic nook is the most important site on the C4BPα1 domain for binding, although previous structural analysis showed that some M proteins (M22 and M28) do not require salt bridging to interact at this site.

The binding region of C4BPα2, previously identified as a "quadrilateral" in the M2/M49 and M22/M28 binding modes, shows several similarities and distinctions in the binding with M87. The C4BPα2 residues Arg64 and His67 are found to bind, albeit in several different modes, to all M protein structures to date. C4BP Arg64 is bound to M87 Asp 80 in Interface 1 and Interface 2, and forms one of only two salt bridges between the molecules (the other being between C4BP Arg39 and M87 Asp96). This interaction may be further stabilized by M87 Glu85 and Tyr81, which bind the main chain and side chain of Arg64, respectively. The retention of an interaction with Arg64 in all known structures suggests Arg64 has a great effect on the overall stability of the complex. Interface 1 and Interface 2 also exhibit interaction between C4BP His67, though the type of bond differs (Fig. 14 and 16). Both Arg64 and His67 have been subjected to mutagenesis experiments (R64Q and H67Q) in the M22 protein, and both mutants attenuated binding between M22 and C4BP (Blom et al. 2000). This study, along with the conservation of these residues in binding, suggests they are important in a broad range of interactions between C4BP and M proteins. A unique feature of this complex is seen in Interface 1. C4BP Ile78, which has been observed to bind all M proteins with hydrophobic interactions, was seen hydrogen bonded to M87 Gln73 with its main chain carbonyl. This interaction demonstrates the flexibility that residues on C4BP can exhibit when binding to different M proteins. As C4BP Arg66 is only involved in binding at interface 1 by making an

electrostatic interaction with M87 Ser82, its role in the structure does not seem to be as prominent as that of Arg64.

The subtle variations between M^{HVR}-C4BPα1-2 binding patterns, and even between different interfaces of a single complex, are a testament to the tolerance and flexibility of this interface. One may suspect the interactions found in Interface 1 and Interface 2 of $M87^{HVR}$, as well as those found in all previous structures, contribute most to the stability of the complex. Nonetheless, mutational analysis is required to better determine the significance of each binding interaction on the $M87^{HVR}$-C4BPα1-2 interface. As there are still 37 C4BP binding GAS strains whose M proteins do not fit any known binding mode, the study of their structure could add to the understanding of exactly what is needed to allow binding with C4BP, and which interactions are less significant.

The creation of a vaccine for GAS must overcome many challenges, such as cross-reactivity with native α-helical proteins like myosin and tropomyosin, broad epidemiological coverage, and immunogenic optimization. The most difficult part of this challenge is arguably the design of an antigen which will elicit an immune response against a broad range of the numerous strains of GAS. The foundation to such a design may very well come from research like that presented here, which reveals the intricacies of how C4BP binds to M proteins.

Appendix

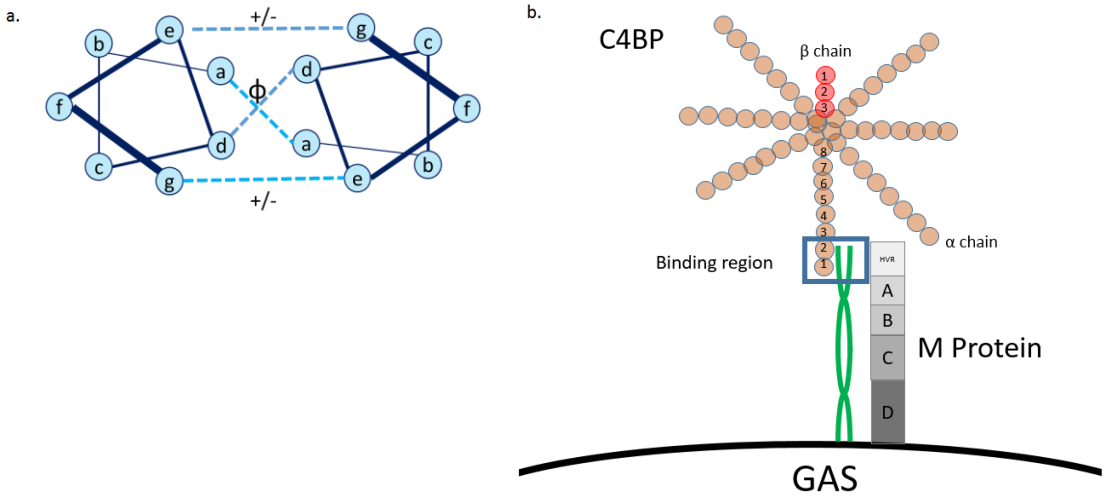

Figure 1: Schematic of the Heptad Repeat and Cartoon of M Protein Binding to C4BP.
a. A representation of the heptad repeat of a protein coiled coil. Residues at the *a* and *d* positions of the coil are usually hydrophobic (φ) and interact with each other. Occasionally, amino acids in the *e* and *g* positions of the heptad will further strengthen the coil via electrostatic interaction. **b.** A cartoon of C4BP bound to M protein. The boxed region represents the binding area; the α1 and α2 domains of C4BP bind the HVR of M protein. The column next to the M protein helix (green) displays its regions relative to the cell surface and C4BP.

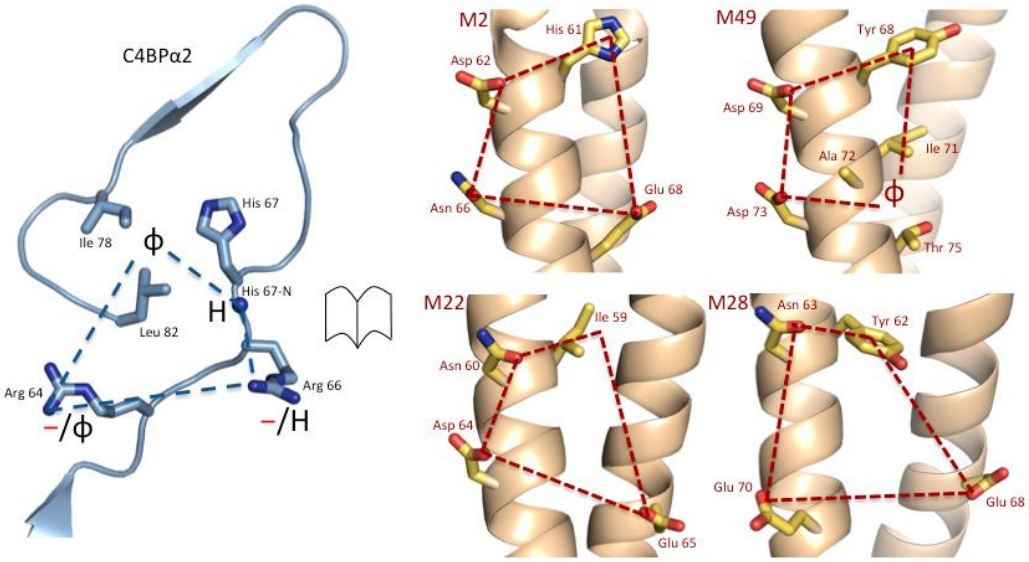

Figure 2: Binding Mode of M2, M49, M22, and M28 to C4BPα2. Reproduced from Buffalo, et al. 2016.

The C4BPα2 quadrilateral (blue dashed lines), with the C4BPα2 backbone shown in ribbon representation and key side chains shown as bonds. The chemical character of M protein residues that interact with the quadrilateral is depicted: φ, hydrophobic; —, negative; H, hydrogen bond forming. 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).

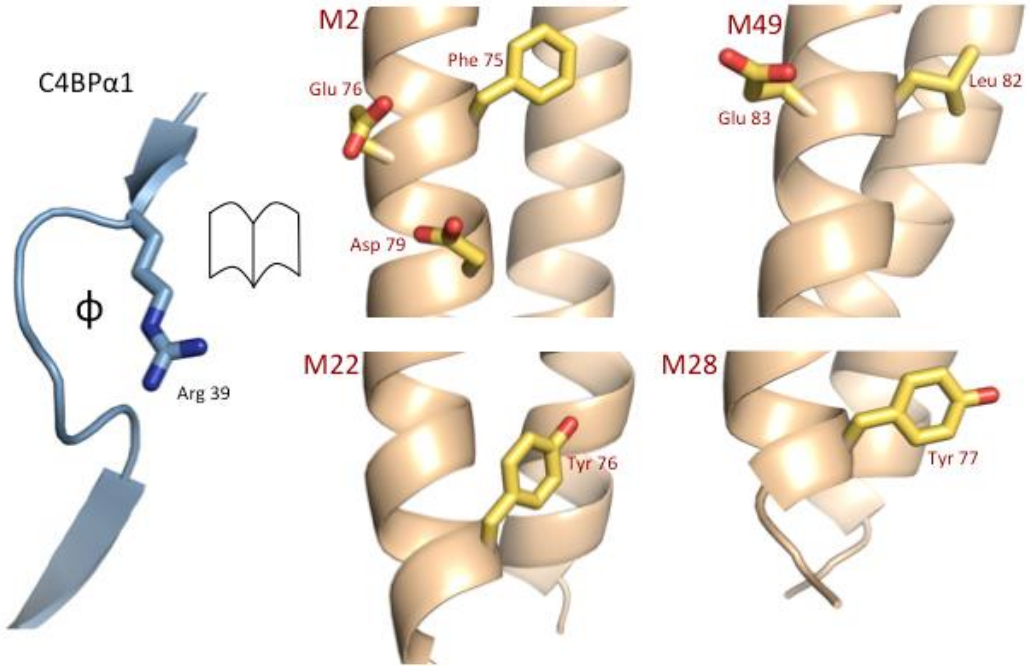

Figure 3: Binding Mode of M2, M49, M22, M28 to C4BPα1. Reproduced from Buffalo, et al. 2016.
The C4BPα1 Arg39 nook. M2, M49, M22, and M28 residues that interact with the C4BPα1 Arg39 nook shown in open-book representation

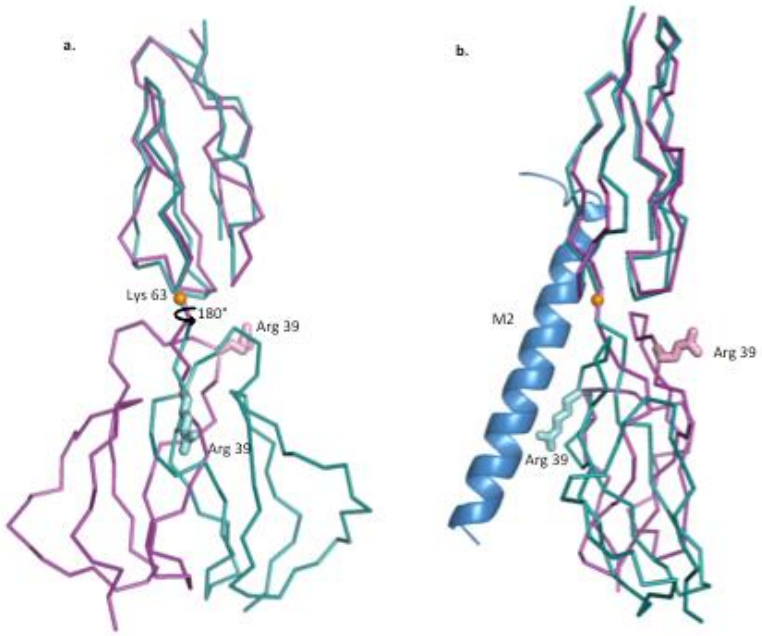

Figure 4: Rotation of C4BPα1-2. Reproduced from Buffalo, et al. 2016.
a. Superposition of free (magenta) and M protein-bound C4BPα1-2 (cyan) based on the C4BPα2 domain, depicted as Cα chain traces. C4BPα1 rotates 180° around Lys63 (left). The position of Arg39 is shown in bonds representation. **b.** 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.

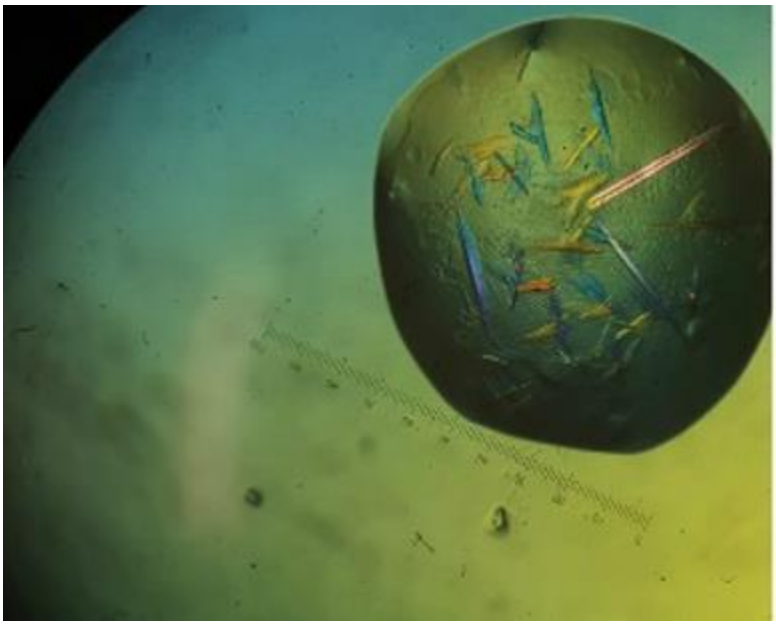

Figure 5: M87^{HVR}-C4BPα1-2 Cocrystal Complex.
The cocrystals of M87^{HVR}-C4BPα1-2 varied in size, but the largest were ~400 microns along the long axis and ~80 microns across the short axis.

```
PrtH      EGAKIDWQEEYKKLDEDNAKLVEVVETTSLENEKLKSENEENKKNLDKLS
M14.5     DRVSRSMSRDDLLNRAQNLEAKNHGLEHQNTKLSTENKTLQEQAEARQKE
M18.6     APLTRATADNKDELIKRANDYEIQNHQLTVENRKLKTDKEQLTKENDDLK
M58       DSSREVTNELTASMWKAQADSAKAKAKELEKQVEEYKKNYETLEKGYDDL
M75       EEERTFTELPYEARYKAWKSENDELRENYRRTLDKFNTEQGKTTRLEEQN
M80       HQLADAARREVLKGETVPAHLWYYQKEENDKLKSANEELETTLQKKEQEL
M87       ESPREVTNELAASVWKKKVEEAKEKASKLEKQLEEAQKDYSEIEGKLEQF
M112      DNNSSSVSVKNEVKLHNEIAALQEEKEKLLNELDNVKEEHKKDLDKVKED
```

Figure 6: Sequence Alignment for Selected M Proteins.
Residues at "d" position of the heptad repeat are highlighted in red. PrtH has no hydrophobic residue in the section of its sequence that is predicted to bind to the hydrophobic nook of C4BPα1 (highlighted green). M112 has no negative residue in the section of its sequence that is predicted to form an electrostatic interaction with C4BP (highlighted in blue).

Table 1: Data Collection and Refinement Statistics.

	$M87^{HVR}$-C4BPα1-2
DATA COLLECTION	
Space Group	P 21 21 2
Cell Dimension	
a, b, c (Å)	43.2
	73.8
	190.4
α, β, γ(°)	90, 90, 90
Wavelength (Å)	0.979
Resolution (Å)	58.33-2.69(2.79-2.69)
R_{merge}	0.09(0.95)
I/σ(I)	20.3(1.4)
Completeness (%)	98.7(90.1)
Redundancy	6.8(3.0)
$CC_{1/2}$	0.997(0.452)
REFINEMENT	
Resolution (Å)	2.69
No. reflections used in refinement	17414
No. reflections used in free set	1741
Rwork/Rfree	0.208/0.277
No. atoms	
Protein	3172
Ligand/ion	20
Water	22
B-factors	
Protein	64.8
Ligand/ion	109.6
water	44.3
R.m.s. Deviations	
Bond Lengths (Å)	0.01
Bond Angles (Å)	1.72

Table 2: Validation Statistics of the M87^{HVR}-C4BPα1-2 Complex.

Statistics	Raw Count	Percentage
Clashscore (all atoms)	12.34	
Poor rotamers	7	1.98%
Favored rotamers	326	92.35%
Ramachandran outliers	0	0.00%
Ramachandran favored	373	96.38%
MolProbity Score	2.07	
Cβ deviations >0.25Å	0/3247	0.00%
Bad bonds	14/4381	0.32%
Bad angles	0/16	0.00%
Cis Prolines	4/375	1.07%
CaBLAM outliers	3	0.79%
CA Geometry outliers	3	0.79%

M Protein	Restriction Digest from pUC57	Ligation into pET28a NhisPP	Transformed into BL21-gold (DE3) *E. coli*	Protein expressed	Protein Purified	Crystal Screen Created	Crystals Observed	Optimization Tray Created	Crystals Taken to Synchrotron	Crystal Structure Obtained
M14.5	✓	✓	✓	✓	✓	✓	✓	✓	X	X
M18.6	✓	X	X	X	X	X	X	X	X	X
M58	✓	✓	✓	✓	✓	✓	✓	✓	✓	X
M75	✓	✓	✓	✓	✓	X	X	X	X	X
M80	✓	✓	✓	✓	✓	✓	X	X	X	X
M87	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
M112	N/A	N/A	✓	✓	✓	X	X	X	X	X
PrtH	✓	✓	✓	✓	✓	X	X	X	X	X

Figure 7: Progress Toward Crystal Structures of Selected M Proteins.
Green check marks represent processes successfully completed. A red "X" represents a process that either failed or was not attempted for a given M protein. The M112^{HVR} did not need to be cloned, as it was provided cloned into pET28aNhisPP commercially.

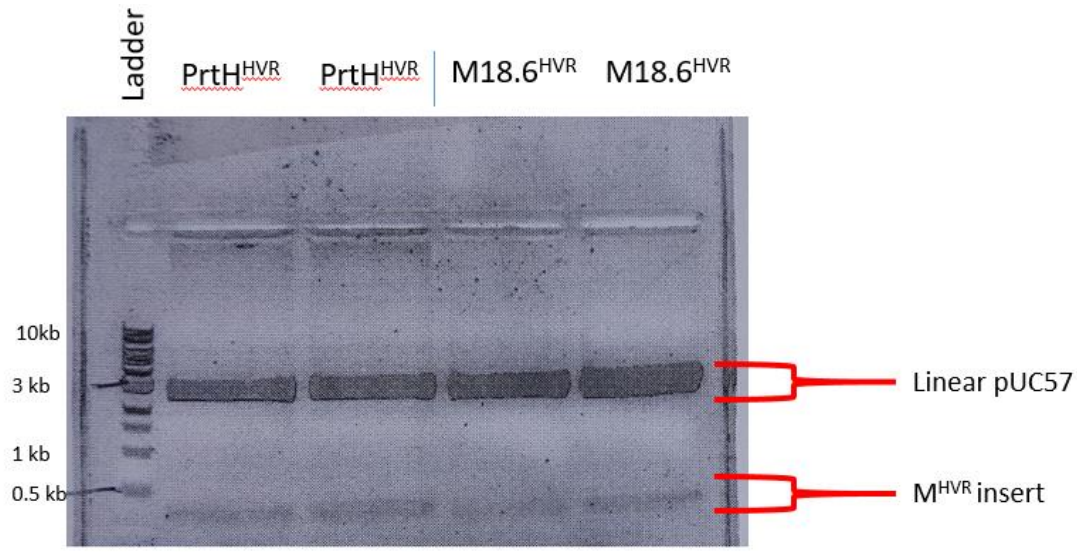

Figure 8: Restriction Digest of M^{HVR} Inserts.
Restriction digest with XhoI and BamHI (3 h, 37°C) shows excision of the $PrtH^{HVR}$ and $M18.6^{HVR}$ inserts. Restriction digests were run on 1% low melt agarose in TAE, and visualized with ethidium bromide (EtBr).

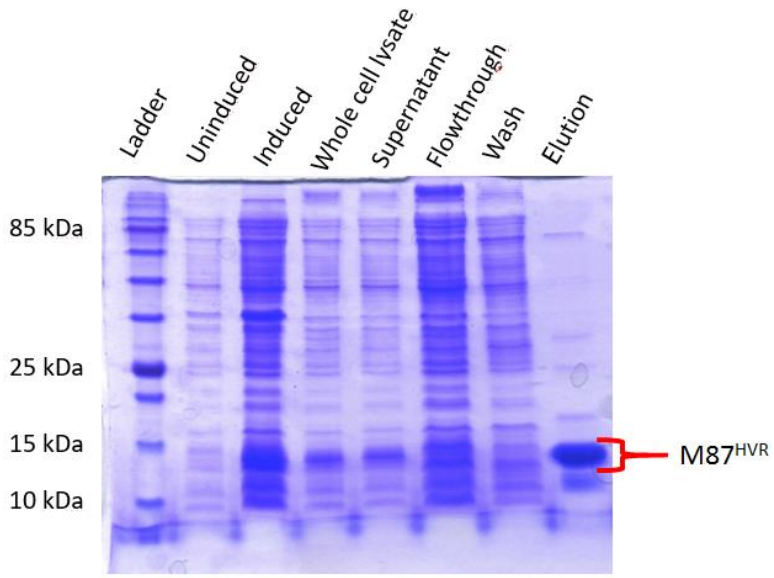

Figure 9: Expression Gel of $M87^{HVR}$.
SDS-PAGE shows $M87^{HVR}$ protein levels at different stages of expression and purification. A thick band at ~12 kDa after IPTG induction indicates successful expression. The flowthrough, wash, and elution fractions were taken during Ni^{2+}-NTA purification. The elution fraction contains contaminant proteins, indicating size exclusion is necessary.

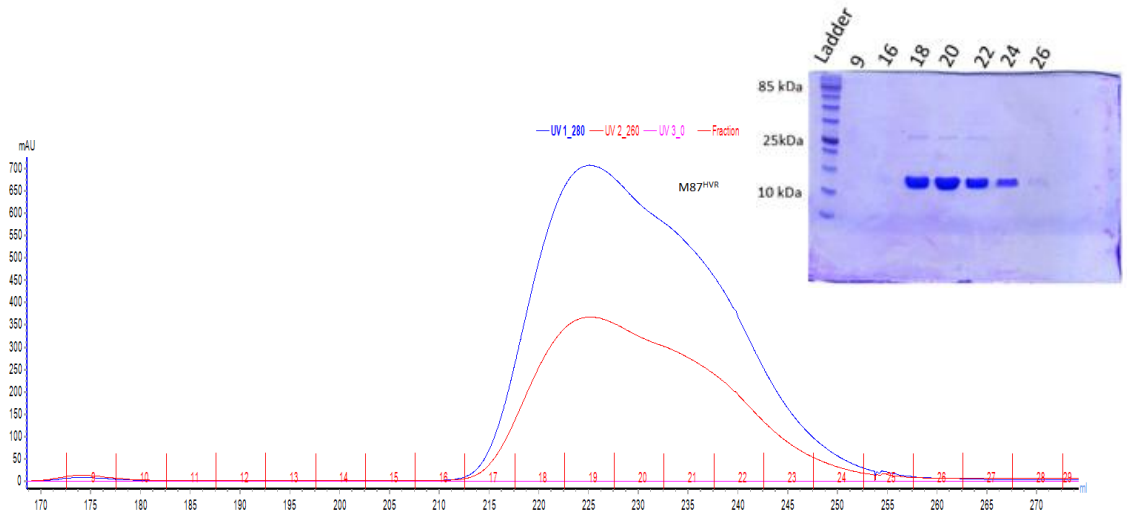

Figure 10: Size Exclusion of M87^{HVR}.
Chromatogram and corresponding SDS-PAGE for M87^{HVR}.

Figure 11: M58^{HVR}-C4BPα1-2 and M14.5^{HVR}-C4BPα1-2 Crystal Screen Drops.
a. A drop in the M14.5^{HVR}-C4BPα1-2 mosquito nanodrop screen. Precipitant: 0.1 M HEPES pH 7.5, 1 M sodium, acetate trihydrate, 0.05 M cadmiumsulfate hydrate. Size: less than 10 microns. These crystals were never exposed to X-ray beams. **b.** A well in the M58^{HVR}-C4BPα1-2 hanging drop crystal tray. Precipitant: 2.0 M $(NH_4)_2SO_4$, 0.1 M Tris pH 8.3. Size: ~300 microns. These crystals were taken to a synchrotron, and confirmed to be protein, but diffraction was weak and no usable data sets were obtained. **c.** A well in the M58^{HVR}-C4BPα1-2 hanging drop crystal tray. Precipitant: 0.2 M $LiNO_3$, 20% v/v PEG 3350, and 0.1 M sodium citrate pH 5.2. Size: ~70 microns long, ~10 microns wide. No usable data were obtained from these crystals.

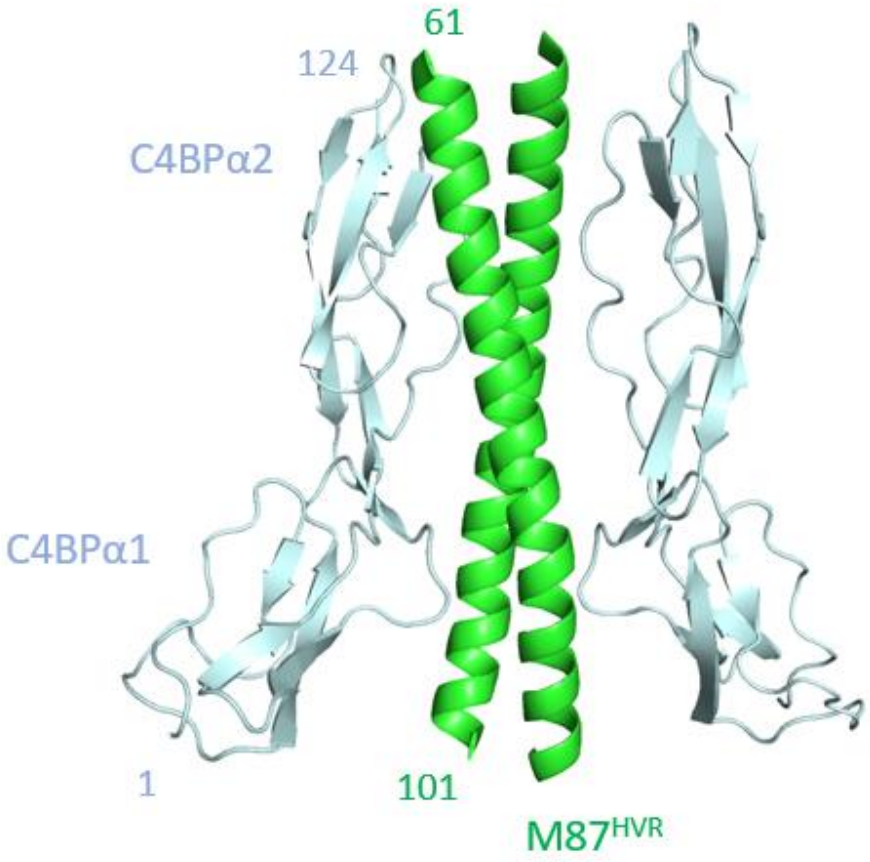

Figure 12: Structure of the M87^{HVR}-C4BPα1-2 Complex.
M87^{HVR} shown in green. C4BPα1-2 shown in cyan.

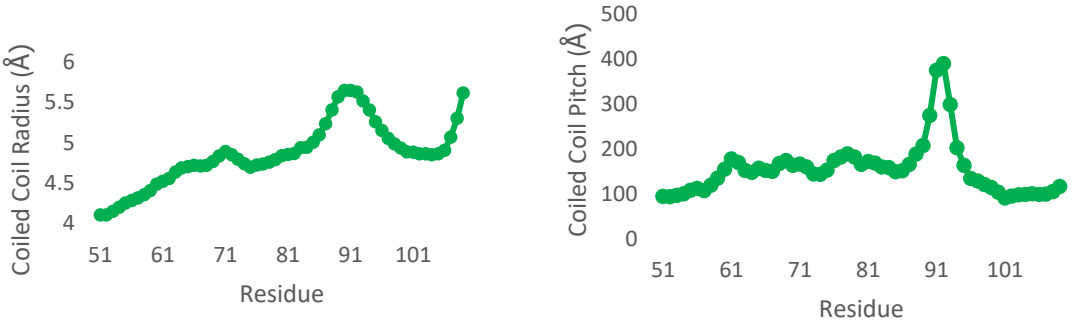

Figure 13: Radius and Pitch of the M87^{HVR} α-Helical Coiled Coil while Complexed with C4BPα1-2.
Radius and Pitch of the coiled coil increase where M87 binds the C4BPα1 domain of C4BP.

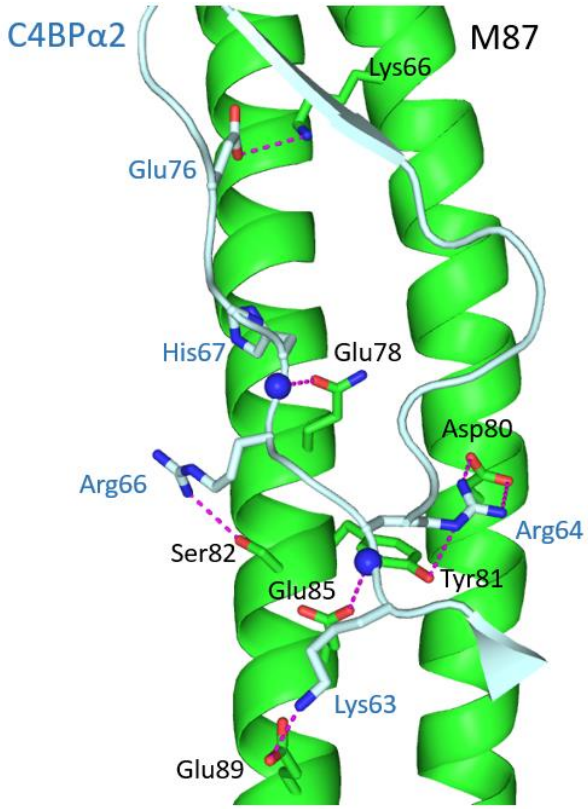

Figure 14: Interface 1 between M87^{HVR} and C4BPα2.
Residues of C4BPα2 (cyan) are labeled in cyan. Residues of M87^{HVR} (green) are labeled in black.

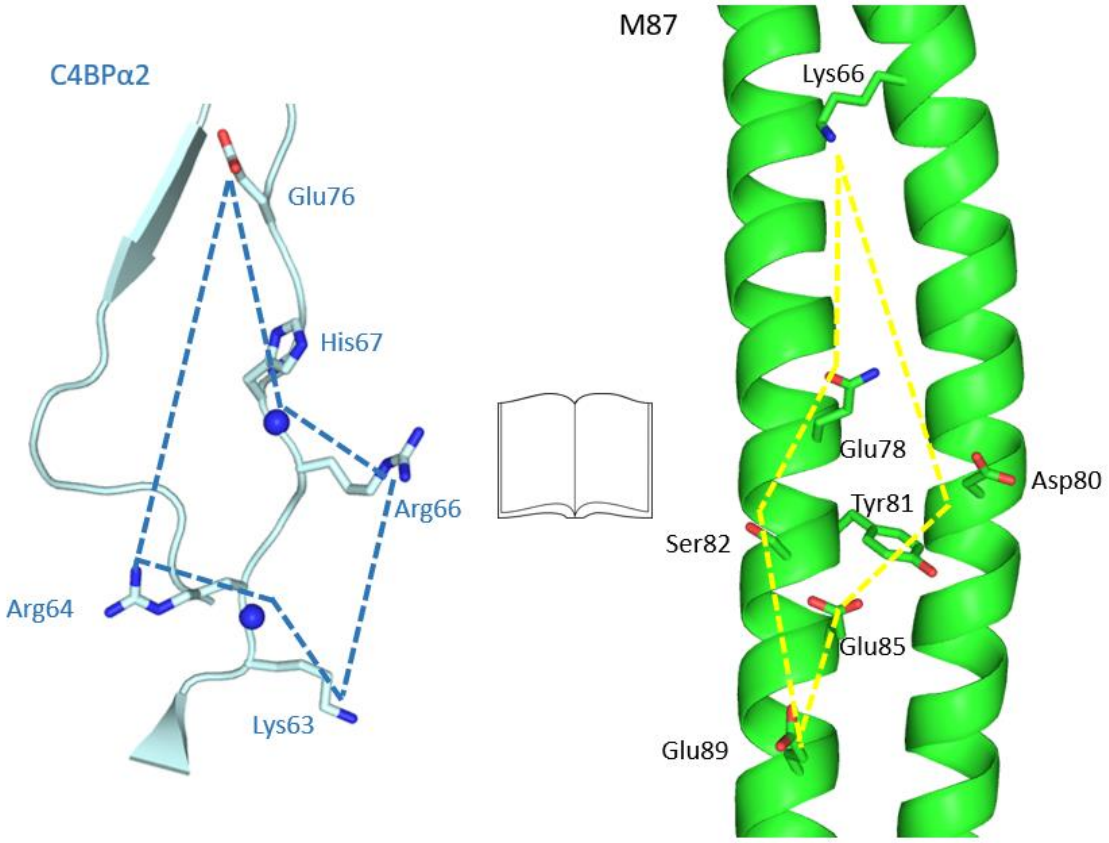

Figure 15: Interface 1 between M87^{HVR} and C4BPα2 Displayed in Open Book Format.
The C4BPα2 domain shows six points of contact with M87^{HVR}. Residues of C4BPα2 (cyan) are labeled in cyan. Residues of M87^{HVR} (green) are labeled in black.

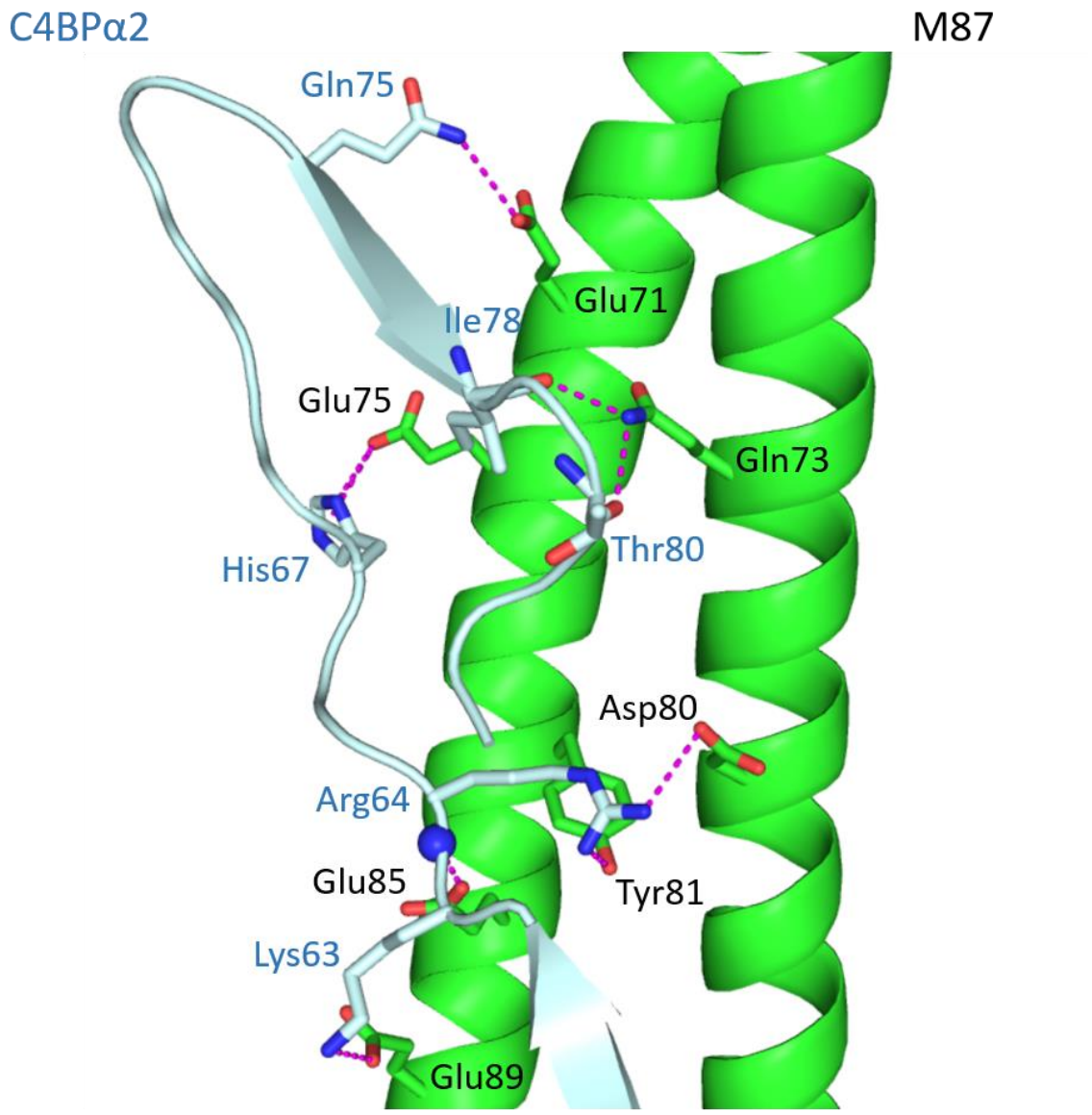

Figure 16: Interface 2 between C4BPα2 and M87^{HVR}.
Residues of C4BPα2 (cyan) are labeled in cyan. Residues of M87^{HVR} (green) are labeled in black.

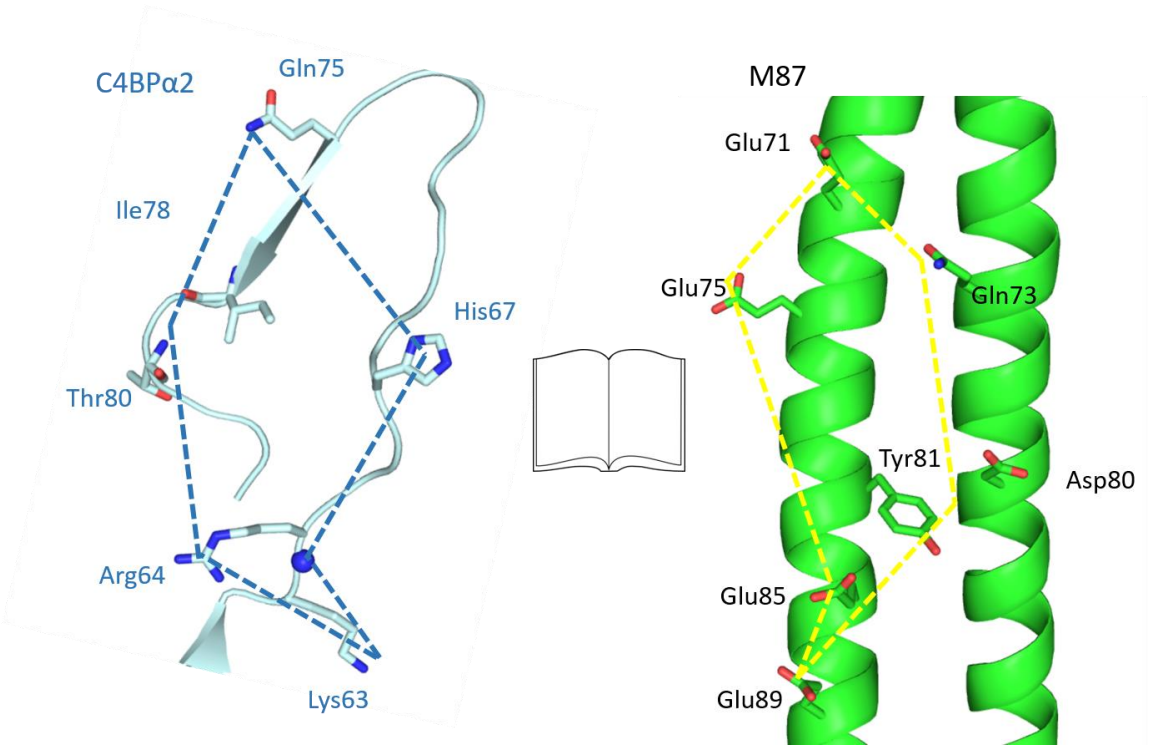

Figure 17: Interface 2 between M87^{HVR} and C4BPα2 Displayed in Open Book Format.
The C4BPα2 domain shows six points of contact with M87^{HVR}. Residues of C4BPα2 (cyan) are labeled in cyan. Residues of M87^{HVR} (green) are labeled in black.

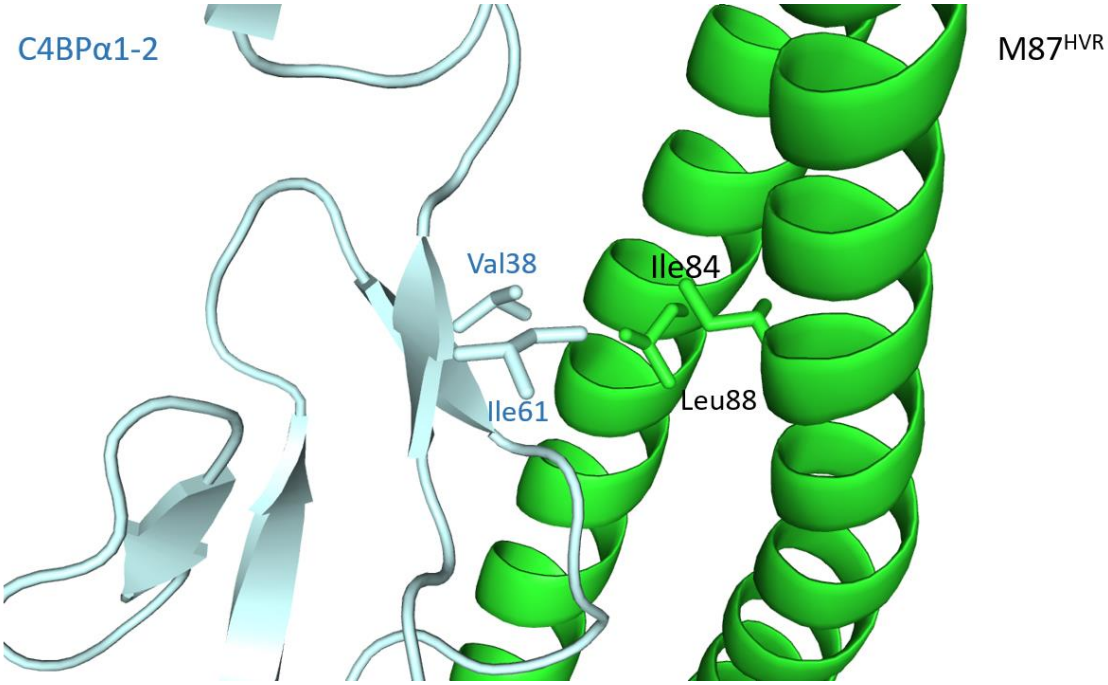

Figure 18: Hydrophobic Binding Site of C4BPα1, Interface 2.
One side of the M87^{HVR}-C4BPα1-2 binding interface exhibits a hydrophobic patch. This patch involves the M87^{HVR} residues Ile84 and Leu88, *d* and *a* residues of the heptad repeat, and C4BPα1-2 residues Val38 and Ile61. Residues of C4BPα2 (cyan) are labeled in cyan. Residues of M87^{HVR} (green) are labeled in black.

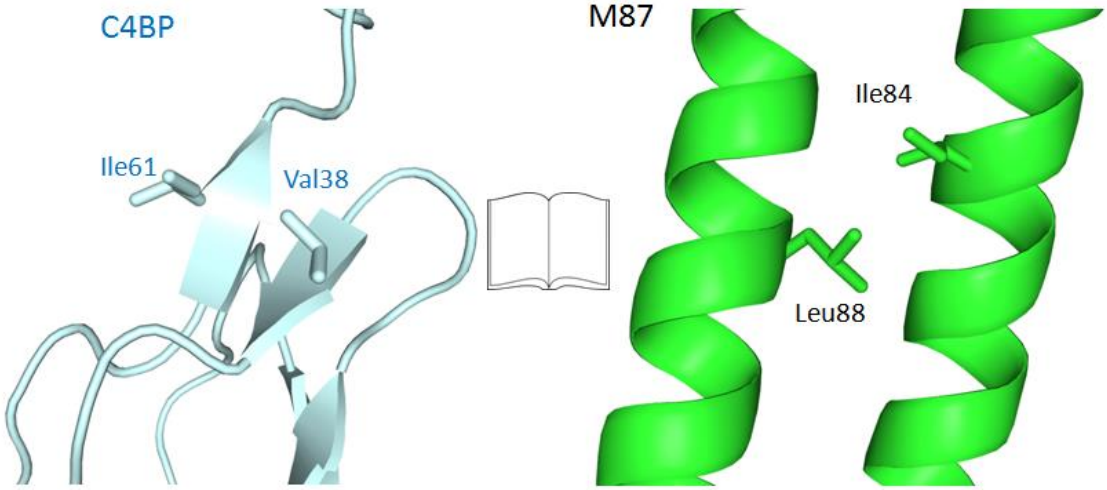

Figure 19: Hydrophobic Binding Site of C4BPα1, Interface 2, Displayed in Open Book Format.
Residues of C4BPα2 (cyan) are labeled in cyan. Residues of M87^{HVR} (green) are labeled in black.

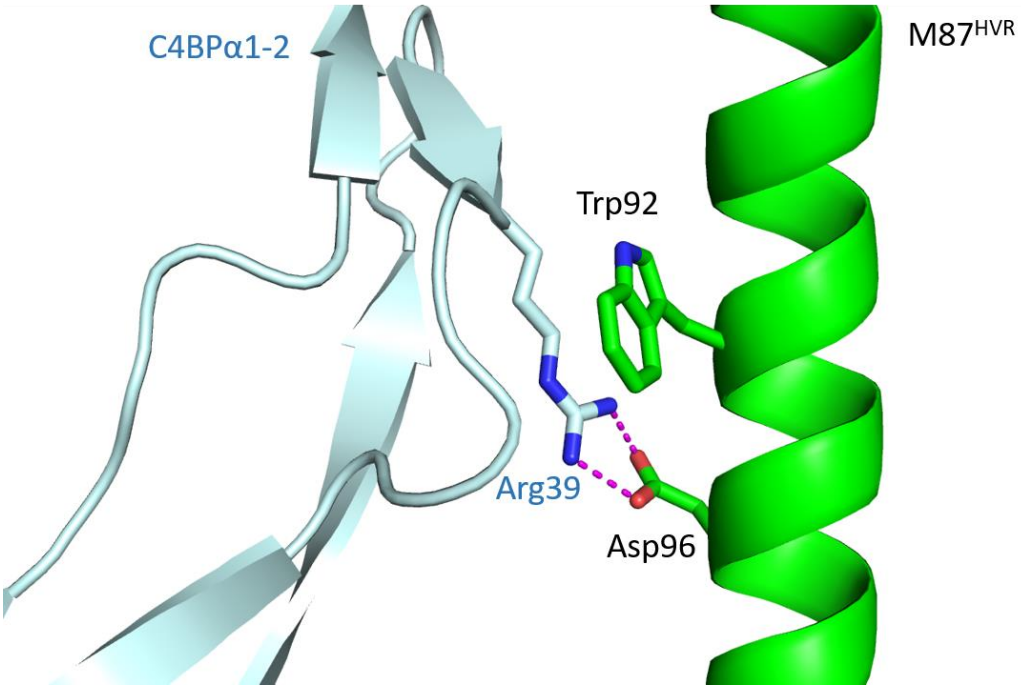

Figure 20: C4BPα1 "Hydrophobic Nook".
Arg39 of C4BPα1 salt bridges with Asp96 of M87. Trp92 of M87^{HVR} rests in the hydrophobic nook formed by Arg39 and main chain atoms of C4BPα1. Residues of C4BPα2 (cyan) are labeled in cyan. Residues of M87^{HVR} (green) are labeled in black.

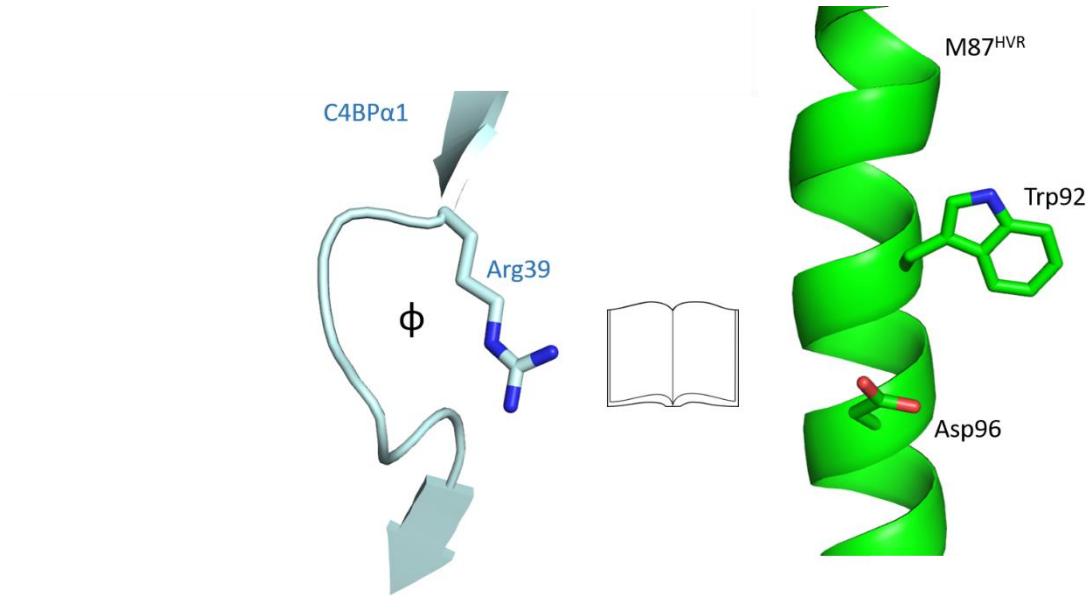

Figure 21: C4BPα1 "Hyrdophobic Nook" Displayed in Open Book Format.
Residues of C4BPα2 (cyan) are labeled in cyan. Residues of M87^{HVR} (green) are labeled in black.

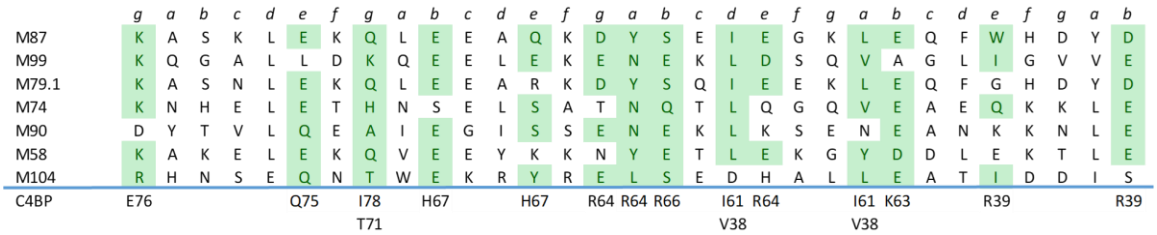

Figure 22: Sequence Alignment of M Protein HVRs, from C4BP Binding GAS Strains, to the M87 Binding Mode.

All M proteins listed are from GAS strains which have been tested to bind C4BP in Persson, et al. 2006. Residues which bind or are predicted to bind are highlighted in green. The heptad register is listed on top. The corresponding residues of C4BP which bind to M87 are listed at the bottom.

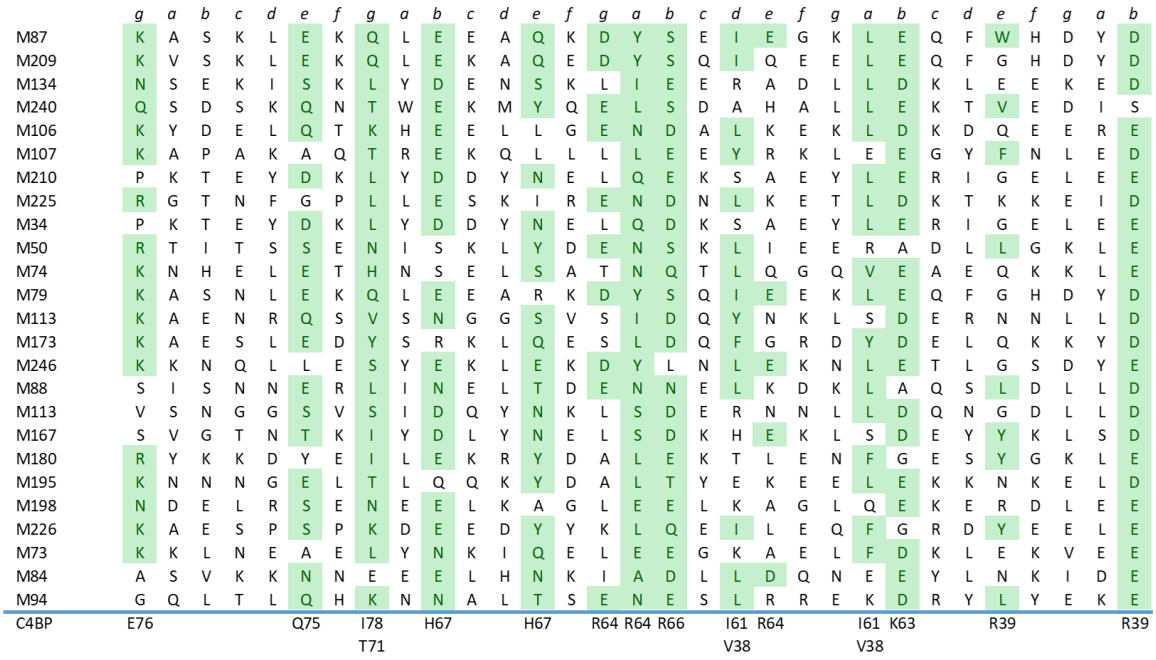

Figure 23: Sequence Alignment of M Protein HVRs, from GAS Strains Untested in C4BP binding, to the M87 Binding Mode.

All M proteins listed have not been tested in a C4BP binding experiment. Residues which bind or are predicted to bind are highlighted in green. The heptad register is listed on top. The corresponding residues of C4BP which bind to M87 are listed at the bottom.

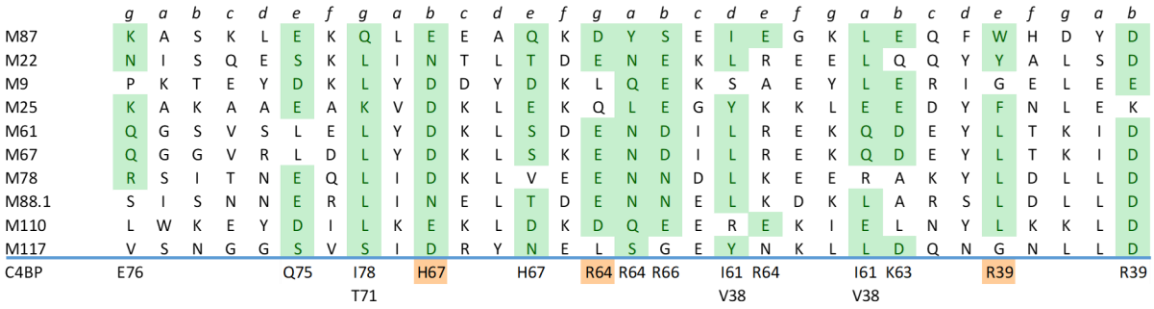

Figure 24: Sequence Alignment of M22-like Protein HVRs to the M87 Binding Mode.
All M proteins listed have been predicted to bind like M22. Residues which bind or are predicted to bind are highlighted in green. The heptad register is listed on top. The corresponding residues of C4BP which bind to M87 are listed at the bottom. Residues of C4BP highlighted in orange bind both M87 and M22 at the same location.

Figures 2, 3, and 4 listed in the Appendix are reproductions of figures that appear in Conserved Patterns Hidden within Group A *Streptococcus* M Protein Hypervariability Recognize Human C4b-Binding Protein. Cosmo Buffalo, et al., Nature Microbiology, 2016. These figures are used with permission from Nature Publishing Group.

References

Adams, Paul D., Pavel V. Afonine, Gábor Bunkóczi, Vincent B. Chen, Ian W. Davis, Nathaniel Echols, Jeffrey J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart. 2010. "*PHENIX* : A Comprehensive Python-Based System for Macromolecular Structure Solution." *Acta Crystallographica Section D Biological Crystallography* 66 (2). International Union of Crystallography: 213–21. doi:10.1107/S0907444909052925.

Berggård, K, E Johnsson, E Morfeldt, J Persson, M Stålhammar-Carlemalm, and G Lindahl. 2001. "Binding of Human C4BP to the Hypervariable Region of M Protein: A Molecular Mechanism of Phagocytosis Resistance in Streptococcus Pyogenes." *Molecular Microbiology* 42 (2): 539–51. http://www.ncbi.nlm.nih.gov/pubmed/11703674.

Blom, A M, K Berggård, J H Webb, G Lindahl, B O Villoutreix, and B Dahlbäck. 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 (10): 5328–36. http://www.ncbi.nlm.nih.gov/pubmed/10799895.

Buffalo, Cosmo Z., Adrian J. Bahn-Suh, Sophia P. Hirakis, Tapan Biswas, Rommie E. Amaro, Victor Nizet, and Partho Ghosh. 2016. "Conserved Patterns Hidden within Group A Streptococcus M Protein Hypervariability Recognize Human C4b-Binding Protein." *Nature Microbiology* 1 (11): 16155. doi:10.1038/nmicrobiol.2016.155.

Carapetis, Jonathan R, Andrew C Steer, E Kim Mulholland, and Martin Weber. 2005. "The Global Burden of Group A Streptococcal Diseases." *The Lancet Infectious Diseases* 5 (11): 685–94. doi:10.1016/S1473-3099(05)70267-X.

Chen, Vincent B., W. Bryan Arendall, Jeffrey J. Headd, Daniel A. Keedy, Robert M. Immormino, Gary J. Kapral, Laura W. Murray, Jane S. Richardson, and David C. Richardson. 2010. "*MolProbity* : All-Atom Structure Validation for Macromolecular Crystallography." *Acta Crystallographica Section D Biological Crystallography* 66 (1): 12–21. doi:10.1107/S0907444909042073.

Cunningham, Madeleine W. 2016. *Post-Streptococcal Autoimmune Sequelae: Rheumatic Fever and Beyond*. *Streptococcus Pyogenes : Basic Biology to Clinical Manifestations*. University of Oklahoma Health Sciences Center. http://www.ncbi.nlm.nih.gov/pubmed/26866235.

Dale, James B., Michael R. Batzloff, P. Patrick Cleary, Harry S. Courtney, Michael F. Good, Guido Grandi, Scott Halperin, Immaculada Y. Margarit, Shelly McNeil, Manisha Pandey, Pierre R. Smeesters, and Andrew C. Steer. 2016. *Current Approaches to Group A Streptococcal Vaccine Development*. *Streptococcus Pyogenes : Basic Biology to Clinical Manifestations*. http://www.ncbi.nlm.nih.gov/pubmed/26866216.

Dale, James B., Thomas A. Penfound, Edna Y. Chiang, and William J. Walton. 2011.

"New 30-Valent M Protein-Based Vaccine Evokes Cross-Opsonic Antibodies against Non-Vaccine Serotypes of Group A Streptococci." *Vaccine* 29 (46): 8175–78. doi:10.1016/j.vaccine.2011.09.005.

Davies, Mark R, Matthew T Holden, Paul Coupland, Jonathan H K Chen, Carola Venturini, Timothy C Barnett, Nouri L Ben Zakour, Herman Tse, Gordon Dougan, Kwok-Yung Yuen, and Mark J Walker. 2014. "Emergence of Scarlet Fever Streptococcus Pyogenes emm12 Clones in Hong Kong Is Associated with Toxin Acquisition and Multidrug Resistance." *Nature Genetics* 47 (1): 84–87. doi:10.1038/ng.3147.

Emsley, P., B. Lohkamp, W. G. Scott, and K. Cowtan. 2010. "Features and Development of Coot." *Acta Crystallographica Section D: Biological Crystallography* 66 (4). International Union of Crystallography: 486–501. doi:10.1107/S0907444910007493.

Fischetti, V A. 1989. "Streptococcal M Protein: Molecular Design and Biological Behavior." *Clinical Microbiology Reviews* 2 (3). American Society for Microbiology (ASM): 285–314. http://www.ncbi.nlm.nih.gov/pubmed/2670192.

Jenkins, Huw T, Linda Mark, Graeme Ball, Jenny Persson, Gunnar Lindahl, Dusan Uhrin, Anna M Blom, and Paul N Barlow. 2006. "Human C4b-Binding Protein, Structural Basis for Interaction with Streptococcal M Protein, a Major Bacterial Virulence Factor." *The Journal of Biological Chemistry* 281 (6). American Society for Biochemistry and Molecular Biology: 3690–97. doi:10.1074/jbc.M511563200.

Lamagni, Theresa L., Shona Neal, Catherine Keshishian, David Powell, Nicola Potz, Richard Pebody, Robert George, Georgia Duckworth, Jaana Vuopio-Varkila, and Androulla Efstratiou. 2009. "Predictors of Death after Severe Streptococcus Pyogenes Infection." *Emerging Infectious Diseases* 15 (8): 1304–7. doi:10.3201/eid1508.090264.

Lupas, A., M. Van Dyke, and J. Stock. 1991. "Predicting Coiled Coils from Protein Sequences." *Science* 252 (5009): 1162–64. doi:10.1126/science.252.5009.1162.

Metzgar, David, and Antonella Zampolli. 2011. "The M Protein of Group A Streptococcus Is a Key Virulence Factor and a Clinically Relevant Strain Identification Marker." *Virulence* 2 (5): 402–12. doi:10.4161/viru.2.5.16342.

Persson, Jenny, Bernard Beall, Sara Linse, Gunnar Lindahl, and PF Zipfel. 2006. "Extreme Sequence Divergence but Conserved Ligand-Binding Specificity in Streptococcus Pyogenes M Protein." *PLoS Pathogens* 2 (5). Securacopy: e47. doi:10.1371/journal.ppat.0020047.

Ricklin, Daniel, George Hajishengallis, Kun Yang, and John D Lambris. 2010. "Complement: A Key System for Immune Surveillance and Homeostasis." *Nature Immunology* 11 (9): 785–97. doi:10.1038/ni.1923.

Snider, Lisa A, and Susan E Swedo. 2003. “Post-Streptococcal Autoimmune Disorders of the Central Nervous System.” *Current Opinion in Neurology* 16 (3): 359–65. doi:10.1097/01.wco.0000073938.19076.31.

Zakour, Nouri L. Ben, Mark R. Davies, Yuanhai You, Jonathan H. K. Chen, Brian M. Forde, Mitchell Stanton-Cook, Ruifu Yang, Yujun Cui, Timothy C. Barnett, Carola Venturini, Cheryl-lynn Y. Ong, Herman Tse, Gordon Dougan, Jianzhong Zhang, Kwok-Yung Yuen, Scott A. Beatson, and Mark J. Walker. 2015. “Transfer of Scarlet Fever-Associated Elements into the Group A Streptococcus M1T1 Clone.” *Scientific Reports* 5 (1). Nature Publishing Group: 15877. doi:10.1038/srep15877.