UNIVERSITY OF CALIFORNIA, SAN DIEGO

# Molecular analysis of *Streptococcus pyogenes* M1 protein

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

in

Chemistry

by

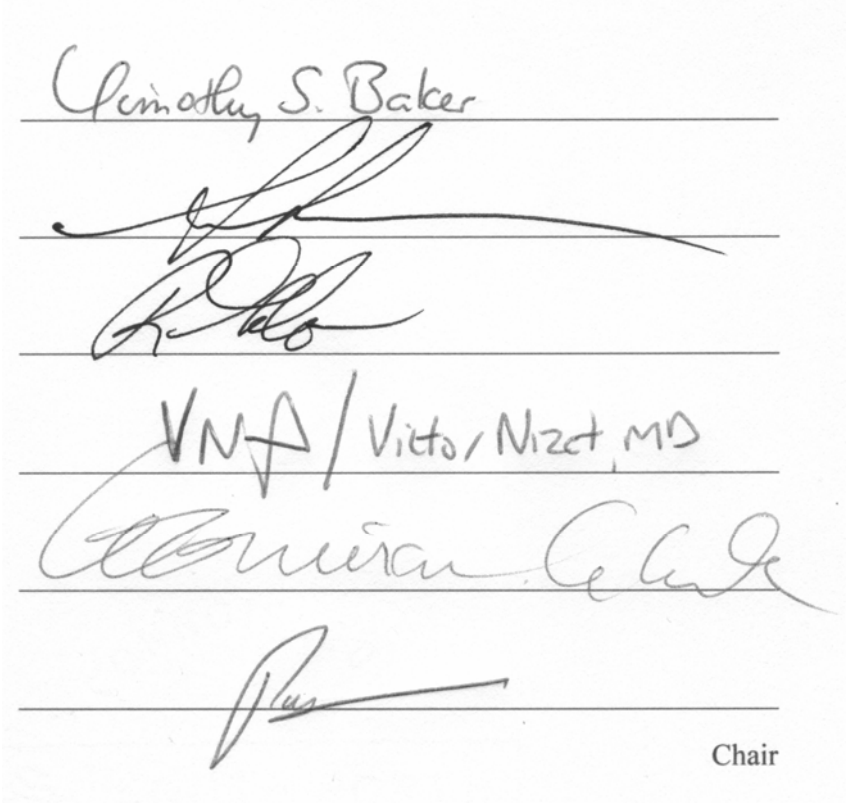
Case W. McNamara

Committee in charge

Professor Partho Ghosh, Chair
Professor Tim Baker
Professor Seth Cohen
Professor Richard Gallo
Professor Gouri Ghosh
Professor Victor Nizet

2006

This dissertation of Case W. McNamara is approved, and it is acceptable in quality and form for publication on microfilm:




University of California, San Diego

2006

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Partho Ghosh, for giving me the opportunity to learn and grow as a scientist in his distinguished lab. His passion, knowledge, positive attitude and patience supported me throughout my development. Importantly, he tolerated my amazing juggling skills, penchant for joking and spontaneous dancing (he caught me more than once).

My colleagues throughout the years have been amazing people who have taught me as much about science as they have life. When I first joined the lab Mike M., Sara, Lori and Larry embraced me and provided answers to all my numerous questions. I am especially appreciative for the guidance given by Jeff and Stephen. I am thankful for the many great relationships I had in graduate school. Mazen, Moni, Jonathan, Mike B., Rebecca, Jason, Loren (Sweet Lo), Devorah, Alicia and Jyothi have been wonderful friends and brilliant coworkers. I will especially miss interacting with Jason on a daily basis. Not only is he a better juggler than I am but he's a great friend. Also, I will forever be indebted to Jonathan for the incredible wealth of crystallography knowledge he provided me and his companionship on numerous synchrotron trips. I would be remiss not to give a special thanks to the Gouri Ghosh group who were always supportive, cooperative, friendly and happy.

I am grateful for my collaboration with the Victor Nizet Lab. Anand, George, Xavier, Sandy, Kelly, Jon, Arya, Laura, Heather and Aaron were inviting, caring and knowledgeable. I am especially thankful for my opportunity to interact and learn from Victor. He works very hard to create a family atmosphere and I think it's reflected by the incredible success of his lab. The generosity of Dr. Russ Doolittle with his time, knowledge and copious amounts of fibrinogen fragments D and DD are gratefully appreciated. Also I would like to thank Dr. Scott Pegan and Dr. Senyon Choe for the use of their static light scattering instrument.

Finally I want to thank my wife and family. Stacey, you have always supported my every ambition. Whether I said I was leaving lab in ten minutes to only leave an hour later, or you caught me daydreaming about research while I should have been focused on something else, you always showed understanding. Just one of the million things I love about you. Mom and dad, you supported me in every way possible. I am blessed to have parents that worked so hard to provide me with an easier path in my own life. This is as much my accomplishment as it is yours. Also I wanted to acknowledge and thank all the love and support I have received from my brother Jeff, sister Bree and Stacey's family – Doug, Padie, Pam and Kadee. Finally, a special shout out to Tiny Dancer, Ladybug, Tuna and Hermie. My pets may be oblivious to anything other than being feed or pet but I still love them.

# VITA

2000 B.S., Chemistry, California State University, San Marcos

2006 Ph.D., Chemistry, University of California, San Diego

## PUBLICATIONS

Lauth X, McNamara C.W., Myskowski S., Igwe E., Beall B., Norrby-Teglund A., Ghosh P., Gallo R., Nizet V. (2006) Group A Streptococcal M1 Protein Inhibits Cathelicidin Antimicrobial Peptide Killing. (Submitted to PLoS Medicine)

McNamara C.W., Nizet V., and Ghosh P. Fibrinogen binding in M1 protein is encoded in the imperfection of the coiled coil. (In preparation)

ABSTRACT OF THE DISSERTATION

Molecular analysis of *Streptococcus pyogenes* M1 protein

by

Case W. McNamara

Doctor of Philosophy in Chemistry

University of California, San Diego, 2006

Professor Partho Ghosh, Chair

*Streptococcus pyogenes*, or group A *Streptococcus* (GAS), is a common human pathogen that causes a wide spectrum of diseases ranging from mild conditions, such as pharyngitis and impetigo, to life-threatening disease, including streptococcal toxic shock syndrome and necrotizing fasciitis. Antiphagocytic M protein, a cell-wall anchored virulence factor, is expressed by all pathogenic strains of GAS. This molecule is believed to form an α-helical coiled coil and demonstrates extensive variability in the surface exposed N-terminus. Binding of host factors such as fibrinogen (Fg), IgG and C4BP have been identified as mechanisms by which M protein confers antiphagocytic properties to GAS.

The X-ray crystallographic structure of an N-terminal fragment of M1 protein is presented here, revealing a disordered N-terminus followed by ordering of the A-region into a dimeric α-helical coiled coil. The A-region is followed by the Fg-binding B-repeats, which have disrupted coiled-coil packing and causes the α-helices in this region to splay apart from one another. The disrupted coiled-coil packing leads to instability of the M1 oligomeric state. Static light scattering and circular dichroism show that the M1 fragment exists in a dynamic equilibrium between monomers and dimers. The oligomeric state is especially sensitive to temperature; however, binding assays demonstrate that fibrinogen binding is independent of temperature.

Structural analysis of the B-repeats showed that it contains many unfavorable residues in its *a* and *d* positions. These residues were mutated to optimal coiled-coil packing residues to determine the significance of M1 protein conserving non-canonical residues in these critical positions. Surprisingly, the optimized construct formed trimeric species and abrogated Fg-binding. The unexpected oligomeric state in the mutant constructs could be partially converted back to the monomer/dimer population observed in the native sequence by deleting a single amino acid just prior to the B-repeats. These species also did not bind Fg and indicated that the residues presiding in the core packing positions in the B-repeats may be critical in contacting Fg.

We propose the B-repeats maintain a splayed state in the mature protein presented on the GAS surface and that this conformation is required to expose the residues that would otherwise be involved in coiled-coil packing. Instead these residues form the Fg-binding site in M1 protein.

# I.

# Introduction

# Introduction

*Streptococcus pyogenes*, also known as group A *Streptococcus* (GAS), is a Gram-positive bacterial pathogen transmitted by direct contact with the saliva or nasal discharge from an infected individual. GAS infection causes a variety of diseases, varying from mild cases, such as pharyngitis and impetigo, to life-threatening conditions, including toxic shock and necrotizing fasciitis (Bisno, 1991). A resurgence of severe streptococcal diseases and rheumatic fever, an autoimmune disease caused by a delayed sequel to GAS infection, has appeared in the last 17 years (Stevens et al., 1989). Additionally, during the last decade, there has been a world-wide marked increase in the recognition and reporting of invasive GAS infections (Stevens, 1992).

M protein, a major virulence factor of all pathogenic GAS, is a fibril surface protein (Figure 1.1) and has been implicated as an adhesion, invasin, and an antiphagocytic molecule. The N-terminal half of M protein is highly variable whereas the C-terminal half is highly conserved and contains an LPXTG cell-wall anchoring motif (Fischetti, 1989). The high degree of antigenic variation observed in the N-terminus has resulted in the identification of more than 150 different serotypes. However, M serotypes 1, 3, 28, 12 and 89 account for nearly half of all GAS infections (CDC, 2005). A clone of the M1 serotype has persisted for more than 20

years as the leading cause of both mucosal and invasive GAS infection (Musser et al., 1995) making it an important serotype for study.

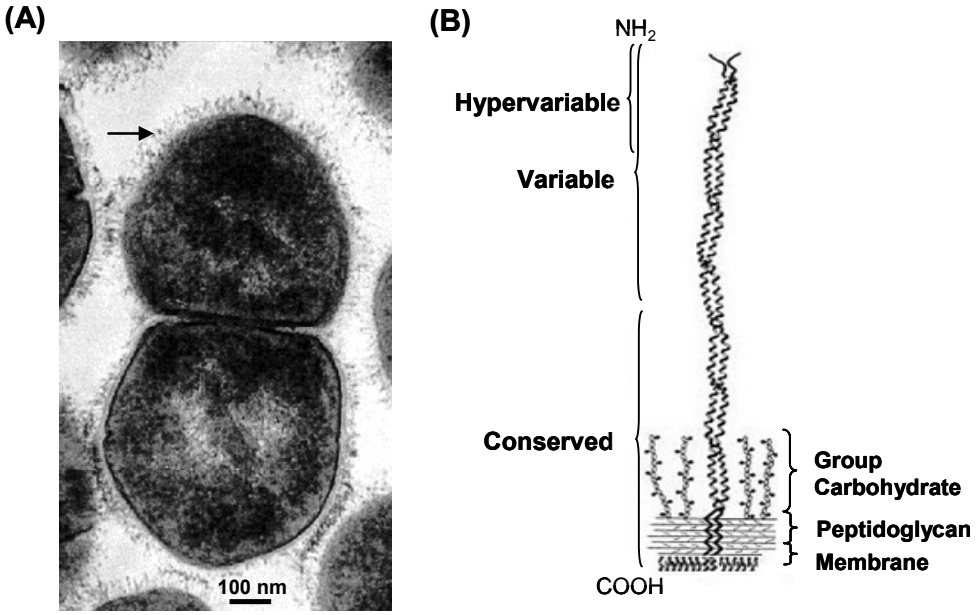



**Figure 1.1. Electron micrograph and schematic of M protein on the GAS surface.** (A) The cell wall and cell surface fibrils, consisting mainly of M protein (indicated by arrow), are well defined in an ultra-thin section of GAS imaged by electron microscopy. (B) Schematic of mature M protein covalently attached to the cell wall. In general, M proteins exhibit high amino acid variability in the N-terminus. Conversely, the C-terminal half is highly conserved. M proteins are covalently attached to the cell wall via the LPXTG motif.

(Images provided by M. Fazio and V.A. Fischetti, The Laboratory of Bacterial Pathogenesis and Immunology, Rockefeller University.)

Pathogenesis begins with the attachment of GAS to pharyngeal or dermal epithelial cells. The adhesion process is believed to be initiated by weak binding of GAS to host cells, which then leads to a second event that confers tissue specificity and high avidity (Hasty et al., 1992). A number of GAS adhesins have been identified (Hasty and Courtney, 1996), but evidence suggests that lipoteichoic acid (LTA)

accounts for 60% of adhesion to epithelial cells. LTA promotes adherence by associating with molecules on the streptococcal surface, including M protein (Simpson and Beachey, 1983). M protein has also been implicated as an adhesin that binds to HEp-2 cells, human laryngeal cancer cells (Courtney et al., 1997), and keratinocytes (Okada et al., 1994). Proposed mechanisms have suggested LTA enables weak association with the host cell and is followed by interaction of M protein and other adhesins to specific receptors that mediate strong adherence and permit colonization.

More recently, experimental evidence has shown that GAS not only adheres to, but can invade, epithelial cells. Intracellular invasion can occur at frequencies equal to or greater than classical intracellular bacterial pathogens, such as *Listeria* and *Salmonella* spp. (LaPenta et al., 1994). M proteins (Cue et al., 1998) and/or fibronectin-binding proteins (Molinari et al., 1997) are required for invasion. Specifically, M1 serotype M protein (Cue et al., 2001; Dombek et al., 1999) and SfbI/protein F1 (Ozeri et al., 1998), a fibronectin-binding protein, act as invasins and both are sufficient to mediate fibronectin-dependent uptake by epithelial cells. Additionally, growing evidence suggests that GAS has the capacity to survive inside phagocytic cells and this too relies on M and M-like proteins (Medina et al., 2003; Staali et al., 2003).

Survival of extracellular GAS from immune clearance is attributed to the antiphagocytic properties of M protein and the hyaluronic acid capsule (Moses et al., 1997). M protein, expressed on the surface of GAS, has been the best studied GAS

.

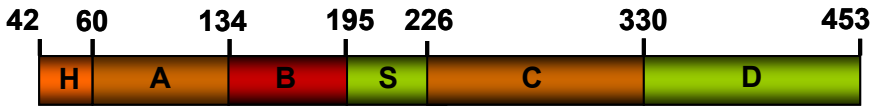



**Figure 1.2. Schematic of mature M1 protein.** The mature M1 protein of GAS is 48 kD and is composed of six distinct regions: hypervariable region (H), A-region (A), B-repeats (B), S-region (S), C-repeats (C), and the D-region (D).

virulence factor. Strains lacking M protein can not survive in blood and are readily killed by phagocytes (Lancefield, 1962). Acquisition of host factors that inhibit complement-mediated opsonization appears to be a major role of the M protein (Bisno, 1979). Recently it was demonstrated that binding of host C4-binding protein (C4BP), an inhibitor of complement activation, to the hypervariable N-terminus is an important mechanism in phagocytosis resistance (Berggard et al., 2001). However, the M1 serotype does not bind C4BP and the precise mechanism by which M1 protein contributes to bacterial virulence is not well understood.

Perhaps the most important role for M1 protein is to bind fibrinogen (Ringdahl et al., 2000). To date, only M proteins from serotypes M1, M3 (Schmidt et al., 1993), M5 (Ringdahl et al., 2000), and M6 (Horstmann et al., 1992) have been shown to directly bind fibrinogen. Binding of fibrinogen has been mapped to the B-repeats (Figure 1.2) in M1 protein and is suggested to have two potential roles. First, it may mask epitopes on the M protein surface or cell surface that would otherwise be recognized by antibodies. Second, M1-mediated deposition of fibrinogen on the GAS

surface may compete with streptococcal-associated C3bi for binding to a $\beta_2$-integrin inhibiting the required signal transduction for the phagocytic process.

M protein is thought to be a dimeric α-helical coiled coil (Phillips et al., 1981), which forms the basis for molecular mimicry and the post-infectious complication of rheumatic fever. Cross-reactive antibodies directed to M protein have been found to be cross-reactive to other mammalian coiled-coil molecules, including myosin (Quinn et al., 1998) and cardiac tropomyosin (Khanna et al., 1997; Mische et al., 1987). Despite the simplicity in design, the α-helical coiled-coil architecture is a powerful platform for introducing antigenic variation and diverse binding surfaces in streptococcal surface molecules.

Coiled-coil proteins have a characteristic heptad repeat that is defined by apolar residues occurring every first (position *a*) and fourth (position *d*) residue (Figure 1.3A). These residues orient to one face of the helix and create a hydrophobic stripe along its length. Oligomerization of the helices (Figure 1.2B) buries the hydrophobic surfaces and creates knob-into-hole packing between positions *a*, *a*', *d* and *d*' (Crick, 1953). As long as the apolar core packing residues (positions *a* and *d*) are conserved then a coiled-coil can tolerate high sequence variability in the other heptad positions. This phenomenon is especially true for positions *b*, *c* and *f*, and to a lesser extent *e* and *g*.

.

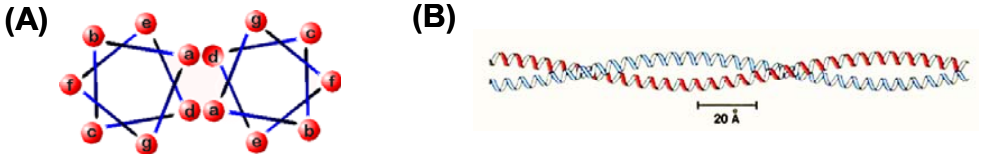



**Figure 1.3. Schematic of α–helical coiled-coil.**
(A) Heptad repeats dictate coiled-coil formation. Amino acids in positions *a* and *d* are usually apolar and salt bridging between adjacent *e-g'* and *e'-g* positions stabilize the structure. Image taken from http://cis.poly.edu/~jps/coilcoil.html. (B) Oligomerization is driven by the need to bury unfavorably exposed hydrophobic surfaces created by the *a* and *d* positions. The resulting coiled coil is characterized by the superhelical winding of the α–helices. Image taken from http://oregonstate.edu/instruction/bb450/lecturenoteskevin/proteinstructureloutline.html.

An intriguing characteristic of M1 protein is that it contains non-canonical coiled-coil regions. Relative to the flanking A- and S-region (Figure 1.2), the centrally located fibrinogen-binding B-repeats have a significantly decreased probability for coiled-coil formation. The preservation of a non-canonical coiled coil region in the middle of the molecule does not provide any obvious advantages for the protein or organism. Yet this molecule is a major virulence factor for the most common serotype found in GAS infections, particularly invasive disease. Thus, conservation of the non-canonical coiled-coil segment in M1 strains is likely to afford some selective advantage to the organism and contribute to the virulence potential of these strains.

The goal of the research presented in this dissertation is to increase the understanding of the molecular mechanisms of M1 protein in pathogenesis, in part, by solving the atomic structure by X-ray crystallography. Presented here is

characterization of M1 fragments harboring the B-repeats using biochemical and biophysical techniques (chapter 2), the crystal structure of an N-terminal fragment of M1 protein (chapter 3), and application of the structure to produce mutagenized B-repeats to probe for structure-function relationships in fibrinogen binding.

.

# References

Berggard, K., Johnsson, E., Morfeldt, E., Persson, J., Stalhammar-Carlemalm, M. and 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.

Bisno, A.L. (1979) Alternate Complement Pathway Activation by Group-a Streptococci - Role of M-Protein. *Infection and Immunity*, **26**, 1172-1176.

Bisno, A.L. (1991) Group-a Streptococcal Infections and Acute Rheumatic-Fever. *New England Journal of Medicine*, **325**, 783-793.

Courtney, H.S., Ofek, I. and Hasty, D.L. (1997) M protein mediated adhesion of M type 24 *Streptococcus pyogenes* stimulates release of interleukin-6 by HEp-2 tissue culture cells. *FEMS Microbiology Letters*, **151**, 65-70.

Crick, F.H.C. (1953) The packing of alpha helices: simple coiled-coils. *Acta Cryst.*, **6**, 685-697.

Cue, D., Dombek, P.E., Lam, H. and Cleary, P.P. (1998) Streptococcus pyogenes serotype M1 encodes mutliple pathways for entry into human epithelial cells. *Infect. Immun.*, **66**, 4593-4601.

Cue, D., Lam, H. and Cleary, P.P. (2001) Genetic dissection of the Streptococcus pyogenes M1 protein: regions involved in fibronectin binding and intracellular invasion. *Microbial Pathogenesis*, **31**, 231-242.

Dombek, P.E., Cue, D., Sedgewick, J., Lam, H., Ruschkowski, S., Finlay, B.B. and Cleary, P.P. (1999) High-frequency intracellular invasion of epithelial cells by serotype M1 group A streptococci: M1 protein-mediated invasion and cytoskeletal rearrangements. *Molecular Microbiology*, **31**, 859-870.

Fischetti, V.A. (1989) Streptococcal M-Protein - Molecular Design and Biological Behavior. *Clinical Microbiology Reviews*, **2**, 285-314.

.

Hasty, D.L. and Courtney, H.S. (1996) *Group A streptococcal adhesion: all of the theories are correct*. Plenum Press, New York.

Hasty, D.L., Ofek, I., Courtney, H.S. and Doyle, R.J. (1992) Multiple adhesins of streptococci. *Infection and Immunity*, **60**, 2147-2152.

Horstmann, R.D., Sievertsen, H.J., Leippe, M. and Fischetti, V.A. (1992) Role of Fibrinogen in Complement Inhibition by Streptococcal M-Protein. *Infection and Immunity*, **60**, 5036-5041.

Khanna, A.K., Nomura, Y., Fischetti, V.A. and Zabriskie, J.B. (1997) Antibodies in the sera of acute rheumatic fever patients bind to human cardiac tropomyosin. *J. Autoimmun.*, **10**, 99-106.

Lancefield, R.C. (1962) Current Knowledge of Type-Specific M Antigens of Group a Streptococci. *Journal of Immunology*, **89**, 307-&.

LaPenta, D., Rubens, C., Chi, E. and Cleary, P.P. (1994) Group A streptococci efficiently invade human respiratory epithelial cells. *Proc. Natl. Acad. Sci. USA*, **91**, 12115-12119.

Medina, E., Goldman, O., Toppel, A.W. and Chhatwal, G.S. (2003) Survival of Streptococcus pyogenes within host phagocytic cells: a pathogenic mechanism for persistence and systemic invasion. *J Infect Dis*, **187**, 597-603.

Mische, S.M., Manjula, B.N. and Fischetti, V.A. (1987) Relation of streptococcal M protein with human and rabbit tropomyosin: The complete amino acid sequence of human cardiac alpha tropomyosin, a highly conserved contractile protein. *Biochem Biophys Res Commun.*, **142**, 813-818.

Molinari, G., Talay, S.R., Valentin-Weigand, P., Rohde, M. and Chhatwal, G.S. (1997) The fibronectin-binding protein of *Streptococcus pyogenes*, SfbI, is involved in the internalization of group A streptococci by eptithelial cells. *Infect. Immun.*, **65**, 1357-1363.

Moses, A.E., Wessels, M.R., Zalcman, K., Alberti, S., Natanson-Yaron, S., Menes, T. and Hanski, E. (1997) Relative contributions of hyaluronic acid capsule and M

.

protein to virulence in a mucoid strain of the group A streptococcus. *Infect. Immun.*, **65**, 64-71.

Musser, J.M., Kapur, V., Szeto, J., Pan, X., Swanson, D.S. and Martin, D.R. (1995) Genetic diversity and relationships among Streptococcus pyogenes strains expressing serotype M1 protein: recent intercontinental spread of a subclone causing episodes of invasive disease. *Infect. Immun.*, **63**, 994-1003.

Okada, N., Pentland, P., Falk, P. and Caparon, M.G. (1994) M protein and protein F act as important determinants of cell-specific tropism of *Streptococcus pyogenes* in skin tissue. *J. Clin. Investig.*, **94**, 965-977.

Ozeri, B., Rosenshine, I., Mosher, D.F., Fassler, R. and Hanski, E. (1998) Roles of integrins and fibronectin in the entry of Streptococcus pyogenes into cells via protein F1. *Mol. Microbiol.*, **30**, 625-637.

Phillips, G.N., Jr., Flicker, P.F., Cohen, C., Manjula, B.N. and Fischetti, V.A. (1981) Streptococcal M protein: alpha-helical coiled-coil structure and arrangement on the cell surface. *Proc Natl Acad Sci U S A*, **78**, 4689-4693.

Quinn, A., Ward, K., Fischetti, V.A., Hemric, M. and Cunningham, M.W. (1998) Immunological relationship between the class I epitope of streptococcal M protein and myosin. *Infect. Immun.*, **66**, 4418-4424.

Ringdahl, U., Svensson, H.G., Kotarsky, H., Gustafsson, M., Weineisen, M. and Sjobring, U. (2000) A role for the fibrinogen-binding regions of streptococcal M proteins in phagocytosis resistance. *Mol Microbiol*, **37**, 1318-1326.

Schmidt, K.H., Mann, K., Cooney, J. and Kohler, W. (1993) Multiple binding of type 3 streptococcal M protein to human fibrinogen, albumin and fibronectin. *FEMS Immunol Med Microbiol.*, **7**, 135-143.

Simpson, W.A. and Beachey, E.H. (1983) Adherence of group A streptococci to fibronectin on oral epithelial cells. *Infection and Immunity*, **39**, 275-279.

Staali, L., Morgelin, M., Bjorck, L. and Tapper, H. (2003) Streptococcus pyogenes expressing M and M-like surface proteins are phagocytosed but survive inside human neutrophils. *Cellular Microbiology*, **5**, 253-265.

.

Stevens, D.L. (1992) Invasive Group a Streptococcus Infections. *Clinical Infectious Diseases*, **14**, 2-13.

Stevens, D.L., Tanner, M.H., Winship, J., Swarts, R., Ries, K.M., Schlievert, P.M. and Kaplan, E. (1989) Severe group A streptococcal infections associated with a toxic shock-like syndrome and scarlet fever toxin. *A. N. Engl. J. Med.*, **321**, 1-7.

.

# II.

# Characterization of M1 protein and M1 fragments

# Abstract

M protein, predicted to form a dimeric coiled coil, is an essential virulence factor of *S. pyogenes*, and functions by conferring survival against the microbicidial effects of macrophages and neutrophils. Association between M protein and human proteins, such as fibrinogen and immunoglobulin Fc, has been implicated in conferring survival by inhibiting phagocytic uptake, complement deposition, or both. A series of functional constructs were expressed in *E. coli* and characterized for X-ray crystallographic studies using biochemical and biophysical approaches. The B-repeats were shown to be the minimal region required to bind human fibrinogen, whereas both the hypervariable region and the S-region were required to bind human IgG. Biophysical studies included circular dichroism (CD) and static light scattering (SLS). The strong helical nature of the fragments was confirmed by CD. SLS demonstrated that full length M1 protein adopts a stable dimer, whereas the truncated, N-terminal fragments are destabilized and are in dynamic equilibrium between monomeric and dimeric states.

# Introduction

M protein is a major virulence factor of GAS and protects the bacterium from phagocytic cells of the immune system. M protein, covalently attached to the cell wall, is thought to form a dimeric coiled-coil structure (Phillips et al., 1981). The M1 variant of the M protein (48 kD) is expressed by invasive strains of GAS and is known to bind to a number of human proteins, including fibrinogen (Fg) (Kantor, 1965) and the Fc region of immunoglobulin G (IgG-Fc) (Heath and Cleary, 1989). The precise manner by which M1 protein affords protection against phagocytes is not well understood. To gain insight into the mechanism of action of the M1 protein, determination of the atomic structure by X-ray crystallography was pursued.

The mature M1 protein and various M1 fragments of the N-terminal half of the molecule were generated. Fibrinogen binding was previously mapped to the B-repeats (Ringdahl et al., 2000) and IgG-Fc binding was attributed to the adjacent S-region (Akesson et al., 1994) providing the means to assay functionality of the recombinant protein.

A small panel of functional fragments was cloned, expressed and purified from *E. coli*. Static light scattering and circular dichroism confirmed that M1 protein adopts a dimeric coiled-coil structure; however, the fragments showed destabilization of the oligomeric state in the absence of the highly conserved C-repeats and D-region.

.

# Experimental Procedures

### Cloning, expression and protein purification

Genomic DNA purified from GAS invasive strain 5448 of the M1 serotype was used as a PCR template to clone the mature form of the M1 protein, amino acid residues 42 to 453, from the *emm1* gene. The PCR product was shuttled into an intermediate vector pCRBlunt (Invitrogen) for sequencing. M1 fragments HC (residues 42-263), HB (residues 42-192), AC (residues 60-263), BC (residues 128-263) and B (128-192) were subcloned from M1-pCRBlunt for placement into a T7-polymerase based expression vector, pET28b (Novagen). These constructs (M1, HC, HB, AC, BC, and B) contain an artifactual residues at the N-terminus (MV) for expression and cloning purposes and at the C-terminus (LEHHHHHH) for purification.

Large scale expression was performed with *E. coli* strain BL21 (DE3) cells (Novagen) in 2-4 L Luria Broth at 37˚C and expression was induced with IPTG when the cells where in log phase growth ($OD_{600}$=0.65). Overexpression was permitted for 4-6 hours at 37ºC, then the cells were harvested by centrifugation and resuspended in 1/100 of the culture volume in lysis buffer (100 mM NaPi, pH 7.5, 250 mM NaCl, and 20 mM imidazole) with 0.5 mM PMSF. Cell lysis was accomplished by adding lysozyme (200 μg/mL) to the resuspended cells for a 20-minute incubation at room temperature (RT) followed by sonication. The lysate was clarified by centrifugation

.

and the filtered supernatant was added to Ni-NTA agarose resin (Qiagen). Bound protein was eluted with lysis buffer containing 250 mM imidazole, then the M1 protein-containing fractions were pooled and dialyzed against 20 mM Tris at pH 8.0. The sample was fractioned by anion-exchange chromatography (Poros HQ) in 20 mM Tris at pH 8.0, applying a linear salt gradient from 0 to 350 mM NaCl. M1 protein was further purified by gel filtration (Superdex 200) equilibrated with 10 mM Tris, pH 8.0, and 100 mM NaCl. After analysis by SDS-PAGE, the protein-containing fractions were pooled and concentrated to ~20 mg/mL and stored as flash-frozen aliquots at -80˚C.

Concentrations of proteins were determined spectroscopically by absorbance at 280 nm in 6 M guanidine hydrochloride buffered with 20 mM NaPi, pH 7.6, using the following extinction coefficients: 9530 $cm^{-1}M^{-1}$ (M1, HC, AC and HB) and 8250 $cm^{-1}M^{-1}$ (BC and B) (Gasteiger et al., 2005).

**Fibrinogen and IgG binding assays**

Binding assays were performed using 20 μL Ni-NTA agarose resin. The beads were pre-equilibrated with binding buffer (50 mM NaPi, pH 8.0, 150 mM NaCl and 20 mM imidazole), then incubated with 20 μg of the his-tagged constructs of M1 protein: M1, HC, HB, AC, BC, B. The beads were washed again with binding buffer then either 10-15 μg of human Fg-D, Fg-DD or IgG were introduced in a 50 μL volume to be incubated at the desired temperature for 30 minutes on an eppendorf rotating unit.

.

Beads were washed three times with binding buffer supplemented with 0.1% Triton X-100, and for IgG binding assays imidazole was also increased to 45 mM to further reduce non-specific binding. Samples were eluted by adding 10 μL 500 mM imidazole, recovered by syringe and the eluant was boiled with either 2x reducing or non-reducing sample buffer for analysis by 11 or 12.5% SDS-PAGE.

**Circular dichroism spectroscopy**

CD spectroscopy was performed with an AVIV CD spectrometer Model 202 (Aviv Instruments) equipped with a thermoelectric temperature control. The wavelength spectra were collected between 195 and 260 nm at 4˚C, 20˚C and 37˚C using a quartz cell of pathlength 1 cm. Samples contained 0.035 mg/mL M1 or HC in 7.5 mM NaPi, pH 7.5 and 50 mM NaF. Spectra were collected at 1.0 nm intervals with a 5 second averaging time per data point. The reference spectrum generated from a scan of the buffer without sample was subtracted from the spectra and the data was converted to mean residue ellipticity (MRE) for comparison of helicity. The MRE is expressed in millidegrees x square centimeter per decimole.

Experiments that followed the change in helicity versus temperature were performed by monitoring the CD signal at 222 nm (maximal wavelength absorbance). Samples were pre-chilled on ice and added to a cuvette that was maintained at 4˚C with stirring. Measurements were taken in 1˚C steps from 4-80˚C allowing for a 1 minute equilibration period before collecting the data. In order to test for thermal

.

reversibility, the signal was recorded for the sample as it was cooled at the same rate it was heated.

**Cross-linking with glutaraldehyde and $BS^3$**

Solutions of M1 or HC at 0.1 mg/mL were mixed in a total reaction volume of 150 μL in 20 mM NaPi, pH 7.6, and 150 mM NaCl. Glutaraldehyde was added to a final concentration of 0.33 mM or 1.0 mM and allowed to react at room temperature for periods of time ranging from 10 seconds up to 2 minutes. The reaction was quenched with the addition of 300 mM glycine at pH 9.0. The dilute samples were concentrated using TCA/DOC precipitation and analyzed by SDS-PAGE.

$BS^3$ (Pierce) cross-linking permitted protein concentrations at 0.5 mg/mL, which eliminated the need to perform TCA/DOC precipitation. A 10-fold molar excess of $BS^3$ (184 μM) was used in 15 μL volumes at various temperatures (4, 25, and 37˚C) with incubation periods ranging from 5 minutes up to 45 minutes. The reaction was quenched by adding 100 mM Tris at pH 7.5.

**Sedimentation equilibrium**

Sedimentation equilibrium experiments were carried out at 4˚C in a Beckman XL-A analytical ultracentrifuge with six-sector cells using three rotor speeds (7, 10 and 14 krpm). The samples were buffered in 10 mM Tris, pH 8.0, and 100 mM NaCl and analyzed at one concentration for M1 (2 mg/mL) and three concentrations for HC

.

(1.5, 2.0 and 3.0 mg/mL). After 16 hours of equilibration, the sample compartments were scanned at 280 nm. The data were analyzed by means of a nonlinear least squares fit program SEDFIT85 (Lebowitz et al., 2002) using values of partial specific volume of 0.73 mL/g for each construct and a buffer density of 1.007.

**Static light scattering**

Static light scattering (SLS) experiments were performed at 25˚C on either a 10 x 300 mm Superdex 200 HR (Pharmacia) or a 7.5 x 300 mm TSK G-4000 SW (TOSOHAAS) analytical size exclusion column in-line with a Mini-DAWN light scattering detector (Wyatt). The system was equilibrated with 20 mM Tris, pH 8.0, and 100 mM NaCl filtered to 0.1 μm and injections ranging from 160-800 μg of constructs M1, HC, HB and BC were performed. Data acquisition and analysis was carried out using ASTRA software (v 4.90.07) as described by the manufacturer using a $d\eta/dc$ of 0.185 for each construct. The system was calibrated using bovine serum albumin (Sigma).

.

# Results

### Overexpression and purification of M1 constructs from *E. coli*

Generation of M1 constructs for biochemical and biophysical characterizations, as well as crystallization trials, were designed with a C-terminal hexapeptide tag to assist in purification and binding assays. The boundaries of each construct were chosen to harbor the important Fg-binding B-repeats while systematically excluding other regions (Fig 2.1A). The two-step purification of metal chelation chromatography followed by gel filtration was sufficient for all the truncated M1 constructs and showed purity >95% by SDS-PAGE (Fig 2.1B, lanes 1-5). However, the full length M1 construct co-purified with a 30 kD degradation product. The degraded product was reactive to anti-pentahistidine antibodies indicating that it was a C-terminal fragment of M1 protein (data not shown). Inclusion of an anion exchange column as an intermediate purification step served to separate the major degraded M1 product (Fig. 2.1; M1: lane 2). To ensure highest purity of all constructs, the anion exchange step was included for all further protein purifications involving the M1 construct.

The elongated nature of this molecule was first evidenced in gel filtration. The expected mass of dimeric M1 protein is 96 kD but eluted as a 300 kD species based on globular protein standards (data not shown). The extended coiled-coil shape induced

.

aberrant gel filtration behavior for all the M1 constructs and was most pronounced in the longer constructs, M1 and HC.

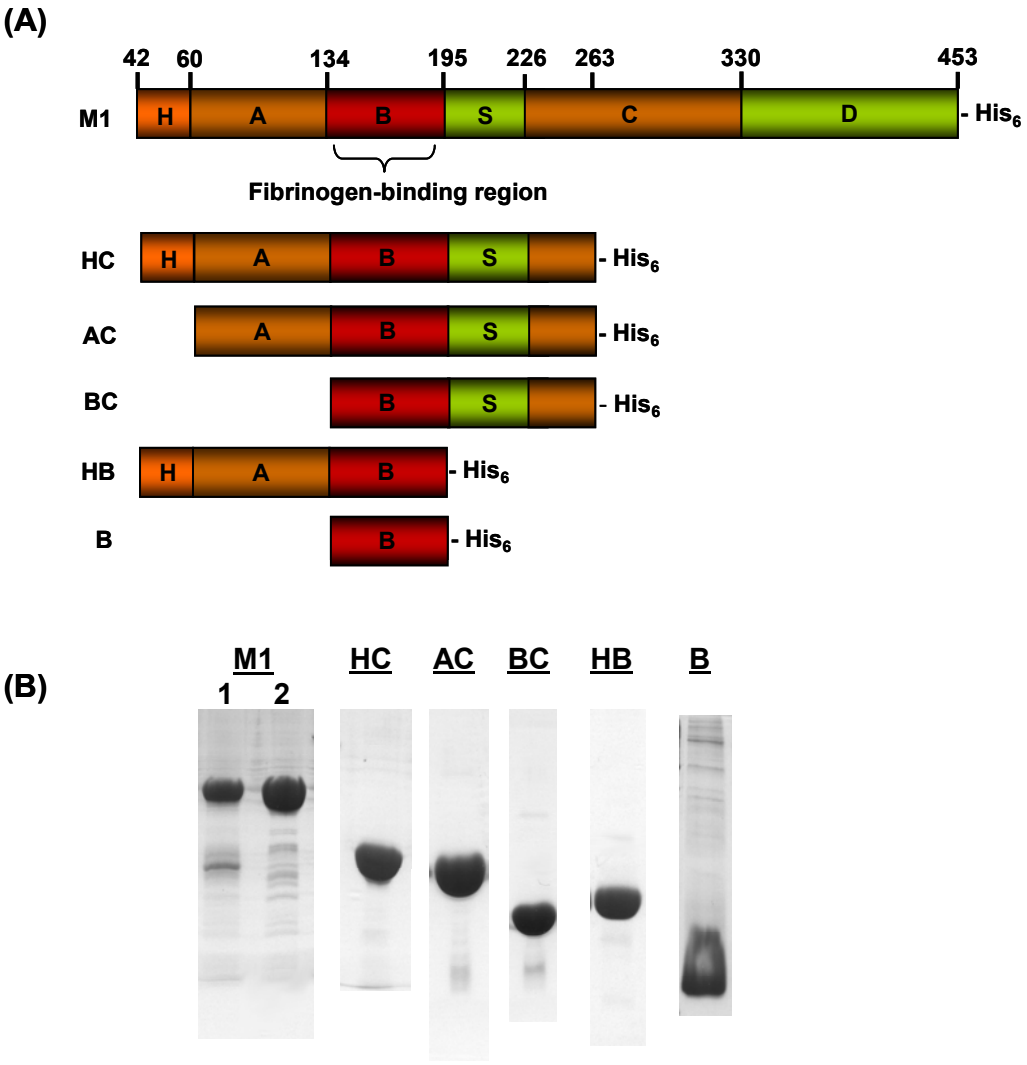



**Figure 2.1. M1 constructs expressed and purified from *E. coli*.**

(A) Schematic of GAS M1 protein (M1) and derived truncated constructs: HC, AC, BC, HB, and B. (B) Visualization of protein purity by SDS-PAGE for all purified constructs. M1 lane 1 represents M1 protein purified without, and lane 2 with, an anion exchange chromatography step.

The **B-repeats are sufficient to bind fibrinogen and fibrin**

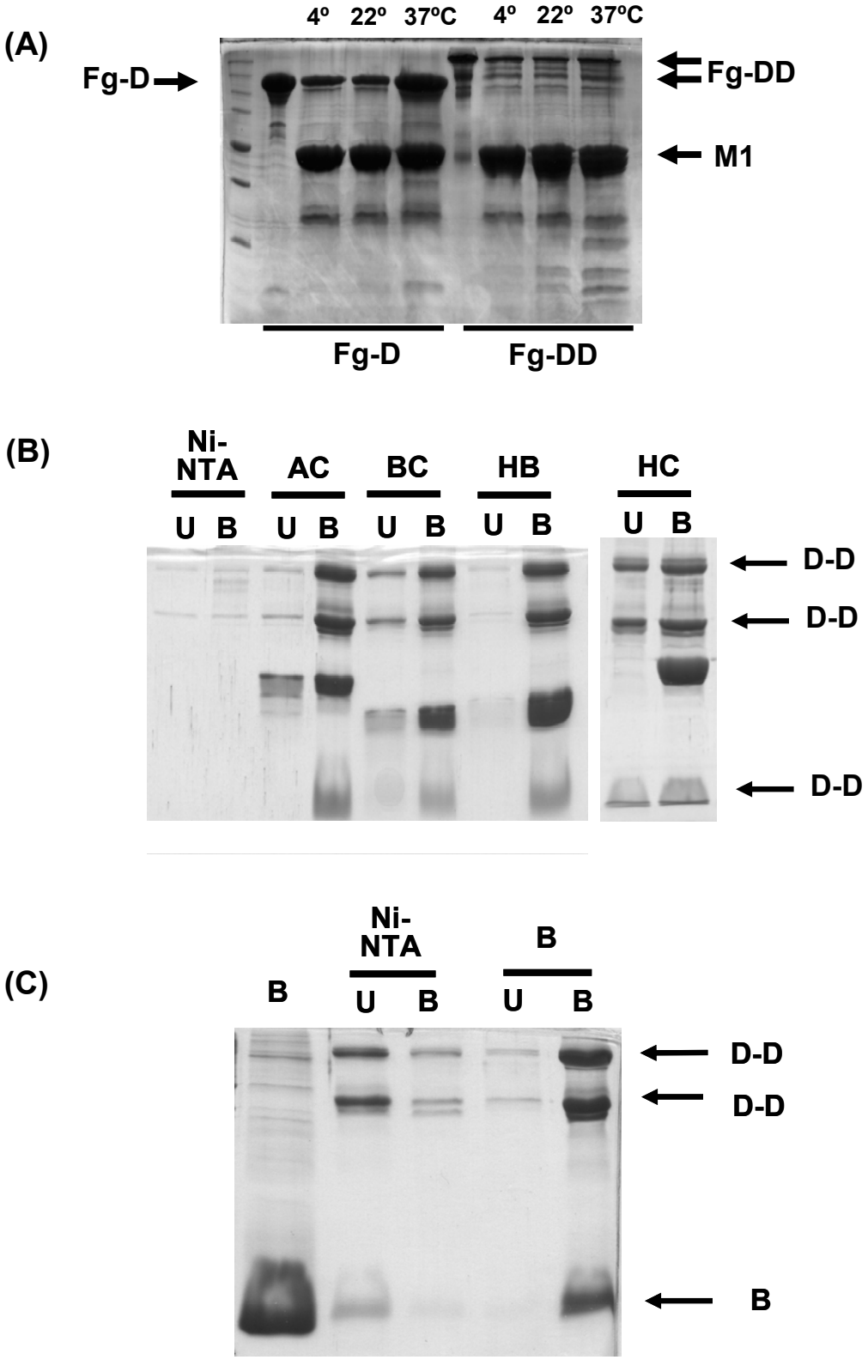
To confirm that functionally active M1 protein was purified from *E. coli*, the fibrinogen binding activity of each construct was tested. Fibrinogen fragment D (Fg-D), known to harbor the M1 protein binding site and fragment D-D (Fg-DD), a mimic of fibrin, were added to Ni-NTA resin pre-coated with M1 protein. Under non-reducing conditions both constructs run as a single band on SDS-PAGE gels; however, under reducing conditions the heterotrimeric polypeptide complex separates into 3 bands (α, β and γ chains). The samples were incubated at various temperatures, thoroughly washed, and eluted for visualization by SDS-PAGE. Binding was evident to both Fg fragments, suggesting M1 protein can not distinguish between Fg (Fig. 2.2A; lanes 3-5) and fibrin (Fig. 2.2A; lanes 7-9). Interestingly, binding exhibited temperature dependence with optimal binding occurring at physiological temperature (Fig. 2.2A; lane 5).

The truncated M1 constructs (HC, HB, AC, and BC) also retained functional binding to Fg (Fig. 2.2B; lanes 4, 6, 8 and 10). Unlike full length M1 protein, there was no temperature effect for Fg-DD binding to these constructs (data not shown) and confirmed that it was M1 protein, not Fg-D, that was influenced by temperature and facilitated increased binding. Most importantly, the B construct showed binding to Fg-DD, producing direct evidence that the B-repeats are sufficient to bind Fg (Fig. 2.2C; lane 5).

.

**Figure 2.2. All constructs bind fibrin(ogen) via the B-repeats**

Binding assays with human fibrinogen fragment D (Fg-D) or a fibrin mimic (Fg-DD) to Ni-NTA resin pre-coated with a specific M1 protein construct and visualized by non-reducing (A) or reducing (B and C) SDS-PAGE. (A) Elution of Fg-D and Fg-DD captured from resin coated with M1 protein demonstrates temperature-dependent binding. Lanes 2 and 6 are 5 μg purified Fg-D and Fg-DD, respectively. (B) Constructs AC, BC, HB and HC all show binding to Fg-DD and were eluted. (C) Construct B, comprising the entire B-repeat region in M1 protein, binds Fg-DD. Where labeled, lanes correspond to unbound/Ni-NTA flow through (U) and bound/Ni-NTA eluant (B). The construct that Ni-NTA was pre-coated with is labeled above the U/B lanes. Lanes labeled 'Ni-NTA' are controls where no protein was pre-coated.


(A)
4º 22º 37ºC 4º 22º 37ºC
Fg-D
Fg-DD
M1
Fg-D
Fg-DD
(B)
Ni-NTA
AC
BC
HB
HC
U B U B U B U B U B
D-D
D-D
D-D
(C)
Ni-NTA
B
B
U B U B
D-D
D-D
B

**The hypervariable and S-region are required for IgG-Fc binding**

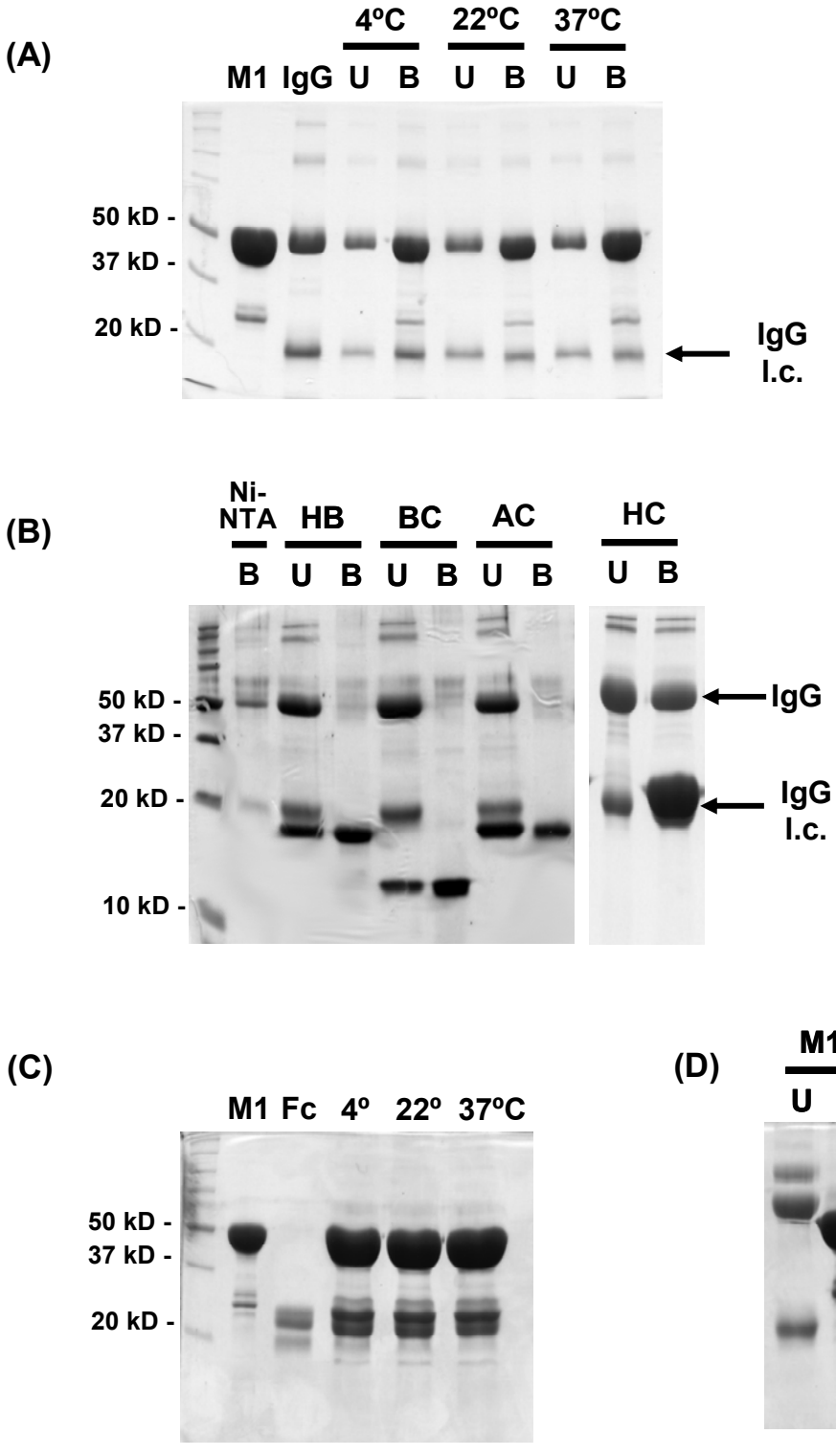
In order to establish that the purified constructs adopted the proper fold and retained functionality, human IgG binding was also examined. The experimental setup was similar to that used for Fg binding except that the pull-down fractions were analyzed by reducing SDS-PAGE. Under reducing conditions IgG separated into two bands on SDS-PAGE. The larger band at ~50 kD represented the heavy chain while the light chain (l.c.) migrated at ~20 kD. The heavy chain coincided with the mass of full length, monomeric M1 protein so the appearance of the light chain was used as a marker for IgG binding. Lanes 5, 7, and 9 in Figure 2.3A clearly show the IgG light chain co-eluting with M1 protein, confirming IgG binding. Unlike Fg-D, there was no temperature effect with M1 protein binding to IgG. Significant amounts of IgG were present in the flow through (Fig. 2.3A; lanes 4, 6 and 8) reflecting saturation and/or weak affinity of the IgG-binding site on M1 protein.

The other M1 constructs were also tested and it was expected that any construct with an S-region (i.e. HC, BC, and AC) would bind IgG while HB, which lacks the S-region, would not bind IgG (Akesson et al., 1994). Surprisingly, the only construct to demonstrate binding was HC (Fig. 2.3B; lane 10). Despite harboring an S-region, constructs BC and AC did not show binding as did HC (Fig. 2.3B; lane 10). The only difference between AC and HC is the presence of the non-

.

**Figure 2.3. The HVR and S-regions are required for IgG binding.**

Binding assays with human IgG to Ni-NTA resin pre-coated with a specific M1 protein construct and visualized by reducing SDS-PAGE. (A) M1 protein binds IgG in an temperature-independent manner. (B) The HC construct binds IgG, but constructs AC, BC, and HB do not. (C) Non-immune binding of M1 protein to human $IgG_1$-Fc. (D) M1 protein does not bind human IgA showing Ig specificity. Where labeled, lanes correspond to unbound/Ni-NTA flow through (U) and bound/Ni-NTA eluant (B). The construct pre-coated to the Ni-NTA resin is labeled above the U/B lanes. Lanes labeled 'Ni-NTA' are controls where no M1 construct was prebound to the resin.


(A)
4ºC
22ºC
37ºC
M1 IgG U B U B U B
50 kD -
37 kD -
20 kD -
IgG
l.c.
(B)
Ni-
NTA
HB
BC
AC
HC
B U B U B U B U B
50 kD -
37 kD -
20 kD -
10 kD -
IgG
IgG
l.c.
(C)
M1 Fc 4º 22º 37ºC
50 kD -
37 kD -
20 kD -
(D)
M1
U B

helical, 19-residue hypervariable region. The apparent requirement of the N-terminus was a novel, unexpected result.

Finally, the specificity of IgG binding to M1 protein was confirmed by showing that IgG-Fc is the minimal fragment required for binding (Fig. 2.3C; lanes 4, 5, and 6). As a control, M1 protein showed no affinity to human IgA (Fig. 2.3D), which was consistent with previous studies (Akerstrom et al., 1992; Raeder and Boyle, 1995).

**M1 constructs show strong alpha helical character in circular dichroism**

After establishing that the M1 constructs were functional, biophysical characterization was performed to compare the structural and thermal stability of the constructs. Circular dichroism was used to study the helical content and reversibility of thermal denaturation for constructs M1 and HC. Wavelength scans of each showed signatures representative of α-helix characterized by pronounced double minima at 208 and 222 nm (Fig. 2.4A). The overlapped spectra displayed very similar profiles but the M1 construct had significantly greater mean residue ellipticities compared to HC, suggesting a greater helical content in M1.

Thermal stability was investigated by monitoring the change in CD signal at 222 nm over a broad range of temperatures. Both melting curves (Fig. 2.4B, HC; 2.4C, M1) showed biphasic unfolding and the estimated midpoint temperature ($T_m$) for

.

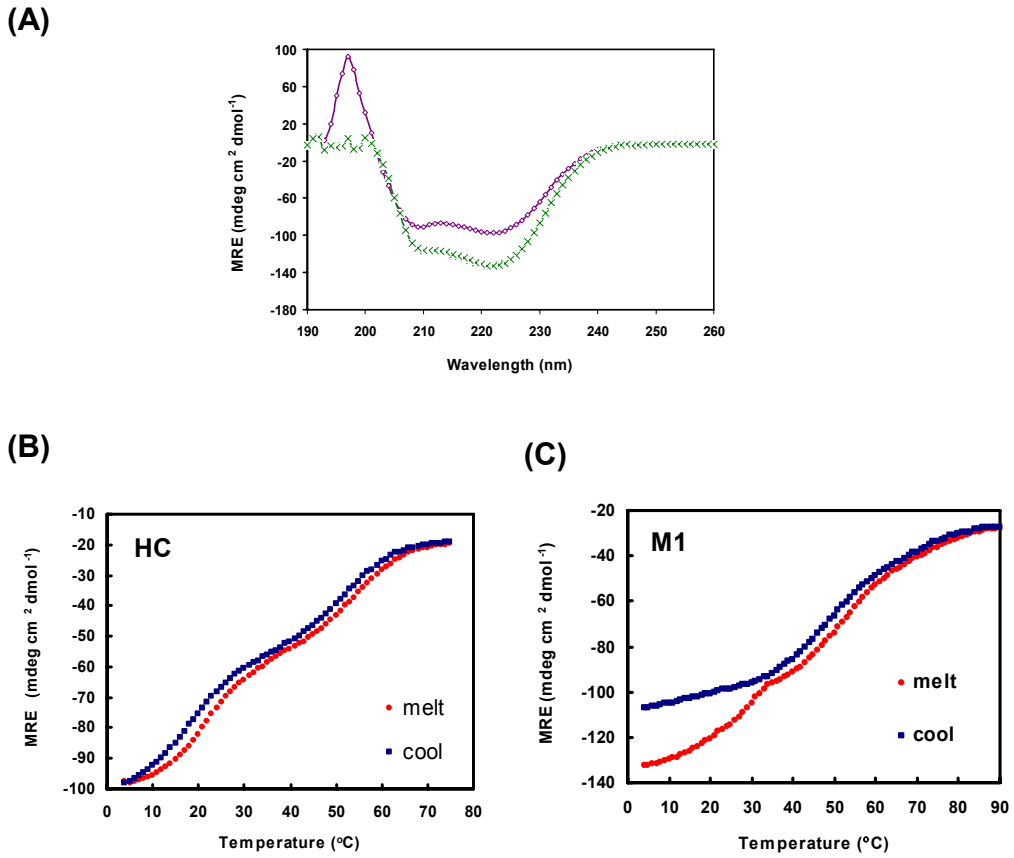



**Figure 2.4. Strong α-helical signal and reversible thermal denaturation of M1 constructs by circular dichroism**

(A) CD spectra for constructs M1 (green) and HC (purple) were collected from 260 – 190 nm and exhibited classic α-helical signatures. The change in signal at 222 nm was used to monitor the helical content during thermal denaturation of (B) HC and (C) M1. Biphasic unfolding gave mid-points of 22 ˚C and 52˚C for HC, and 27 ˚C and 57 ˚C for M1. The signal was converted to mean residue ellipticity (MRE) for comparison.

the first unfolding step in each was 22˚C and 27˚C for HC (Fig. 2.4B) and M1 (Fig. 2.4C), respectively. The $T_m$ of the second unfolding state of HC occurred at 52˚C and M1 at 57˚C. An even greater disparity was observed for temperatures that induced complete denaturation. Heating to 90˚C was required to denature M1 while 75˚C was

sufficient for HC, revealing a difference in overall stability. All these data suggest that M1 is more thermostable than the M1 fragment HC.

Interestingly, HC demonstrated complete reversible thermal denaturation of both transitions while M1 showed reversible thermal denaturation for only the second transition. It was found that if M1 was flash frozen in liquid nitrogen after thermal denaturation, then analyzed for thermal denaturation again that it still produced two-state unfolding (data not shown). This suggests that unfolding is reversible for M1 but may be a much slower event.

**Chemical cross-linking suggests M1 protein is trimeric *in vitro***

The M1 constructs proved to be functionally active in binding human IgG, IgG-Fc, and Fg fragments D and D-D. Further characterization of the constructs required confirmation that they maintained a static, dimeric conformation. Chemical cross-linking using common reagents, $BS^3$ and glutaraldehyde, were tested with full length M1 protein. Since M1 protein was shown to be dimeric (Phillips et al., 1981) it would serve as a control before testing the other fragments.

The high lysine content of M1 protein (13.6%) ensures that very favorable cross-linking conditions existed. Glutaraldehyde concentrations as well as the reaction incubation time had to be severely limited to control the production of cross-linked polypeptide visualized by SDS-PAGE. The cross-linking pattern for full length M1 protein consistently showed dimeric and trimeric species that formed simultaneously

.

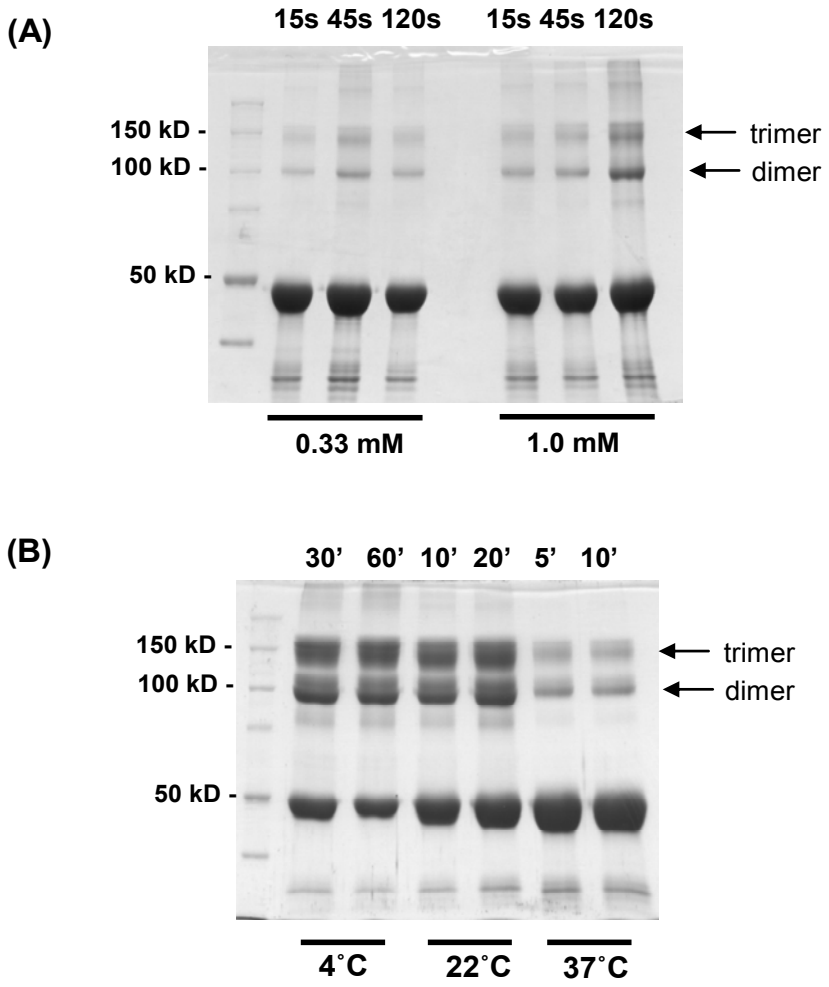



**Figure 2.5. Chemical cross-linking of construct M1 by glutaraldehyde and $BS^3$ yields trimers.** Visualization of cross-linked species by 10% SDS-PAGE showed bands corresponding to dimers and trimers. (A) Variant glutaraldehyde concentrations (0.33 mM or 1.0 mM) over a 2-minute period at room temperature showed no differences in the banding pattern. (B) $BS^3$ cross-linking, examined over a temperature range (4-37˚C), exhibited the same banding pattern as glutaraldehyde. Bands corresponding to dimers and trimers are explicitly indicated by the arrows.

(Fig. 2.5A). Even very short reaction times (<10 seconds) times failed to inhibit the formation of a band ascribed to trimeric M1 (data not shown).

A second reagent, $BS^3$, which has a decidedly slower rate of cross-linking, was also used. Again, the same pattern of dimeric and trimeric species was evident (Fig.

2.5B). Temperature had no significant affect on the cross-linking pattern, but did affect the rate of reaction. Comparison of lanes 4 and 7 in Fig 2.5B shows identical incubation periods at 25˚C and 37˚C, respectively. The RT sample formed significantly greater amounts of product whereas little product was observed in the sample treated at 37˚C.

The same banding pattern was also observed with the HC construct (data not shown) thereby demonstrating showing that it was not a construct-dependent result. Taken by themselves, these experiments suggested that the M1 protein exists as a trimer *in vitro*.

**Non-ideal behavior of M1 protein in sedimentation equilibrium experiments**

M1 protein was previously identified as a dimer by sedimentation equilibrium experiments (Phillips et al., 1981). These experiments were repeated using the full length M1 construct and the truncated HC construct. For unknown reasons the M1 construct generated unacceptably high residuals that were 100-fold greater than those observed for the HC construct. As a result, only the HC construct was analyzed, and because HC showed similar behavior to M1 in the cross-linking assays this seemed to represent a favorable alternative.

While the data appeared ideal for each the of HC runs (Fig. 2.6A), the molecular mass calculated at different centrifugation forces and protein concentrations gave highly variable results (Table 2.1). These data show that HC behaved as a dimer

.

only at the slowest centrifugation speed (7000 rpm) and highest protein concentration (3 mg/ml). An increase in the centrifugation speed to 10,000 rpm generated a 13 kD drop in calculated mass and corresponded to a species between monomer and dimer.

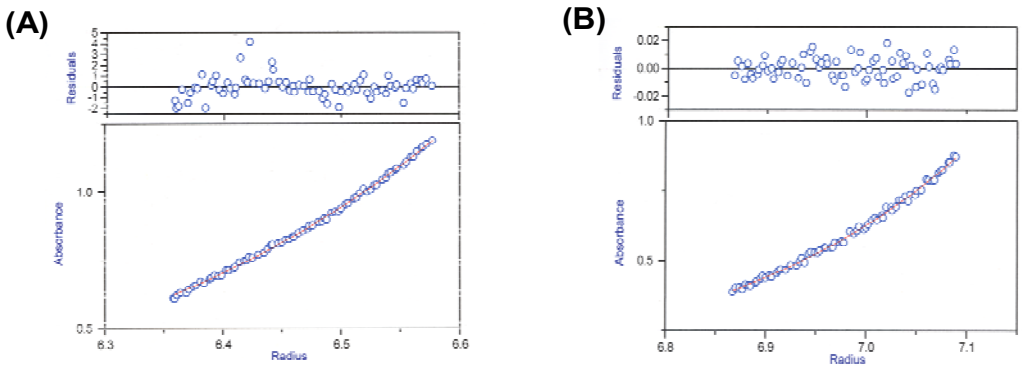



**Figure 2.6. Absorbance scans of M1 and HC in sedimentation equilibrium experiments.** Examples of data collected from experiments using (A) M1 and (B) HC constructs.

**Table 2.1. Summary of molar masses calculated by sedimentation equilibrium experiments.**

| Construct | mg/ml | rpm | Expected kD | Calculated kD |
|---|---|---|---|---|
| M1 | 3 | 7000 | 96 | 76 |
| HC | 1.5 | 7000 | 54 | 20.7 |
| HC | 1.5 | 10000 | 54 | 45.5 |
| HC | 1.5 | 14000 | 54 | 43.6 |
| HC | 2 | 7000 | 54 | 40.9 |
| HC | 2 | 10000 | 54 | 38.5 |
| HC | 2 | 14000 | 54 | 42 |
| HC | 3 | 7000 | 54 | 57.6 |
| HC | 3 | 10000 | 54 | 44 |
| HC | 3 | 14000 | 54 | 44.2 |

Intermediate masses were calculated for all the measurements recorded at 2 mg/ml. The data at 1.5 mg/ml appeared less reliable as a calculated mass below the actual mass of the HC polypeptide was given at 7000 rpm.

**M1 fragments form dynamic dimers**

Static light scattering used in conjunction with a gel filtration column proved to be a way to determine the oligomeric state of the M1 protein and the truncated constructs. Full length M1 protein was calculated by SLS to be 94 kD ± 4 kD (Fig. 2.7A) in excellent agreement with the derived mass of dimeric M1 (96 kD). However, constructs lacking the C-terminus (i.e. full set of C-repeats and D-region) showed oligomeric instability. SLS experiments with HB (Fig. 2.7B), BC, AC, and HC samples yielded molar masses intermediate to monomers and dimers and suggests they are all dynamic molecules (Table 2.2). This was supported by experiments in which increasing concentrations of HC and HB were measured by SLS. The increased concentration drove the equilibrium towards dimer formation and resulted in higher calculated masses (Table 2.2).

An important observation is that the calculated molar mass remains unchanged throughout the elution peak of the M1 fragments. This indicates that the same protein species is being measured for the duration of the experiment. Because the M1 fragments exist in equilibrium between dimeric and monomeric states this suggests the equilibrium is very fast. Thus only the average of the two species is measured. A

.

slower equilibrium would be expected to yield two distinct populations (i.e. monomers and dimers) that would elute separately by gel filtration.

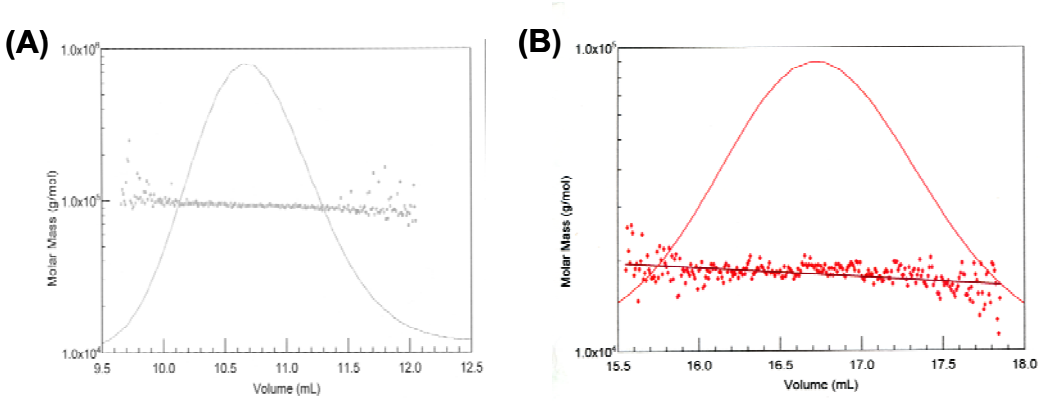



**Figure 2.7. Static light scattering confirms M1 protein is dimeric.** The calculated molar mass was recorded in 0.25-second intervals. (A) The protein absorbance peak of M1 (solid blue line) is shown with an overlay of the calculated molar mass (blue diamonds). (B) The protein absorbance peak of HB (solid red line) is shown with an overlay of the calculated molar mass (red diamonds). A linear, least-squares fit was applied to the data. The horizontal molar mass distribution in HB is representative of that observed for constructs HC and BC.

**Table 2.2. Summary of molar masses calculated by static light scattering.** The molar mass of M1, HC and HB were measured at increasing concentration.

| Construct | Injected Mass (ug) | Expected kD* | Calculated kD** |
|---|---|---|---|
| M1 | 160 | 96 | 103 (14%) |
| M1 | 800 | 96 | 94 (5%) |
| HC | 300 | 54 | 34 (4%) |
| HC | 750 | 54 | 41 (2%) |
| HB | 200 | 38 | 23 (4%) |
| HB | 400 | 38 | 25 (12%) |
| HB | 600 | 38 | 29 (3%) |
| BC | 500 | 32 | 24 (1%) |

* expected molar mass for a dimeric species

** experimental molar mass with percent error given in parentheses

# Discussion

A construct representing the mature M1 protein presented on the GAS surface and N-terminal fragments derived from the mature protein were characterized by biochemical and biophysical experiments. All the fragments maintained functionality despite demonstrating destabilization of their dimeric state. These initial results indicate that formation of a stable oligomeric state may not be necessary to maintain function.

The M1 construct behaved as expected. It bound IgG-Fc and Fg fragments D (Fg-D) and DD (Fg-DD), and also demonstrated the characteristic behavior associated with a dimeric, α-helical coiled-coil. Aberrant gel filtration profiles demonstrated the elongated shape of M1 protein and circular dichroism (CD) and static light scattering (SLS) experiments confirmed the strong α-helical content and dimerization, respectively.

In comparison to the full-length M1 construct, the M1 fragments demonstrated decreased α-helical content and suggested a dynamic equilibrium between monomer and dimer by SLS. Despite this apparent dynamic behavior, all the fragments maintained functional binding to Fg-D/DD. Additional studies designed to confirm the dynamic nature of the oligomeric state for the constructs led to inconclusive

.

results. Non-ideal behavior in sedimentation equilibrium experiments (SE) was observed in the mature and truncated forms of the M1 protein, and chemical cross-linking indicated *in vitro* formation of trimers.

CD melts showed biphasic unfolding of both the M1 and HC constructs. Interestingly, the transitions between folding states occurred slightly below 37˚C for each. Similar results using CD spectroscopy were previously reported for the M1 protein (Nilson et al., 1995), M4 (Cedervall et al., 1997) and H (an M-like protein from the M1 serotype; Akerstrom et al., 1992), proteins. At physiological temperature these molecules are believed to exist as monomers, and dimers are only observed at temperatures at or below RT. Evidence supporting these results was observed with Acanthamoeba myosin II. It was reported that a two-state mechanism ($T_m$ ~40˚C) is consistent with a model in which the dimeric rod unfolds with concomitant formation of two unfolded monomers (Zolkiewski et al., 1997).

Chemical cross-linking supported that M and M-like proteins are monomeric at 37˚C. Examination of the $BS^3$-mediated cross-linking of the M1 construct showed significant differences in the amount of cross-linking product found in 10-minute reactions of M1 protein cross-linked at 37˚C compared to that at RT. In comparison, relatively similar rates of formation were apparent in both the 4˚C and RT reactions. This indicated that fewer intermolecular contacts occurr at 37˚C possibly because of dimers disassembly into monomers, whereas at lower temperatures the M1 protein assembled into stable dimers permitting greater frequency of cross-linking.

.

Åkerström et al. (Akerstrom et al., 1992) also showed that the apparent monomeric behavior at physiological temperatures abolished IgG-binding in protein H, and suggested that binding requires a dimer. In contrast to protein H, M1 protein showed increased binding to Fg-D at physiological temperature. The same temperature dependence was not evident for the destabilized M1 fragments where binding was equivalent at all temperatures. These conclusions suggest the possibility that Fg binding to M1 protein is optimal when the α-helices adopt a non-supercoiled state.

The implications of the potential temperature effect on the oligomeric state of the M1 construct and M1 fragments were taken into consider during the crystallization trials described in Chapter 3.

.

# References

Akerstrom, B., Lindahl, G., Bjorck, L. and Lindqvist, A. (1992) Protein Arp and Protein-H from Group-a Streptococci - Ig Binding and Dimerization Are Regulated by Temperature. *Journal of Immunology*, **148**, 3238-3243.

Akesson, P., Schmidt, K.H., Cooney, J. and Bjorck, L. (1994) M1-Protein and Protein-H - Iggfc-Binding and Albumin-Binding Streptococcal Surface-Proteins Encoded by Adjacent Genes. *Biochemical Journal*, **300**, 877-886.

Cedervall, T., Johansson, M.U. and Åkerström, B. (1997) Coiled-coil structure of group A streptococcal M proteins. Different temperature stability of class A and C proteins by hydrophobic-nonhydrophobic amino acid substitutions at heptad positions a and d. *Biochemistry*, **36**, 4987-4994.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, A.P. and Bairoch, A. (2005) *Protein Identification and Analysis Tools on the ExPASy Server*. Humana Press.

Heath, D.G. and Cleary, P.P. (1989) Fc-Receptor and M-Protein Genes of Group-a Streptococci Are Products of Gene Duplication. *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 4741-4745.

Kantor, F.S. (1965) Fibrinogen Precipitation by Streptococcal M Protein .I. Identity of Reactants and Stoichiometry of Reaction. *Journal of Experimental Medicine*, **121**, 849-&.

Lebowitz, J., Lewis, M.S. and Schuck, P. (2002) Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Sci*, **11**, 2067-2079.

Nilson, B.H., Frick, I.M., Akesson, P., Forsen, S., Bjorck, L., Akerstrom, B. and Wikstrom, M. (1995) Structure and stability of protein H and the M1 protein from Streptococcus pyogenes. Implications for other surface proteins of gram-positive bacteria. *Biochemistry*, **34**, 13688-13698.

Phillips, G.N., Jr., Flicker, P.F., Cohen, C., Manjula, B.N. and Fischetti, V.A. (1981) Streptococcal M protein: alpha-helical coiled-coil structure and arrangement on the cell surface. *Proc Natl Acad Sci U S A*, **78**, 4689-4693.

Raeder, R. and Boyle, M.D.P. (1995) Analysis of Immunoglubulin G-binding-protein expression by invasive isolates of *Streptococcus pyogenes*. *Clin Diagn Lab Immunol.*, **2**, 484-486.

Ringdahl, U., Svensson, H.G., Kotarsky, H., Gustafsson, M., Weineisen, M. and Sjobring, U. (2000) A role for the fibrinogen-binding regions of streptococcal M proteins in phagocytosis resistance. *Mol Microbiol*, **37**, 1318-1326.

Zolkiewski, M., Redowicz, M.J., Korn, E.D., Hammer III, J.A. and Ginsburg, A. (1997) Two-state thermal unfolding of a long dimeric coiled-coil: the Acanthamoeba mysosin II rod. *Biochemistry*, **36**, 7876-7883.

.

# III.

# Crystallization and X-ray structure determination of M1 fragment, nHB

# Abstract

M and M-like proteins are elongated virulence factors, predicted to form α-helical coiled coils, which coat the streptococcal surface. The ability of these molecules to bind a variety of host ligands despite retaining the same fold demonstrates the power in their molecular design. The atomic structure of an N-terminal fragment of M1 protein (nHB; residues 42-194; 18 kD) was solved to 3.03 Å by X-ray crystallography. The crystallized M1 fragment is a dimeric, parallel, α-helical coiled, coil confirming previous experimental evidence. The C-terminal helices of the B-repeats show disrupted coiled-coil packing and splay apart from one another. This unexpected conformation may be adopted in the full length molecule and may contribute to diverse ligand binding.

# Introduction

M protein has long been assumed to form an α-helical coiled coil. Supporting evidence is that circular dichroism experiments reveal a strong helical content and analysis of the primary sequence for this family of proteins reveals a heptad repeat characteristic of α-helical coiled-coil folds. Further evidence comes from the visualization of the molecules on the streptococcal surface by electron microscopy, which shows that they are elongated, hair-like projections.

One important property of coiled-coil molecules is that the solvent accessible residues can be readily altered to a number of other amino acids without affecting the structure. This has two effects: antigenic variation and variable binding surfaces. The diversity of amino acids that can be displayed along the helices may also contribute to the ability of M proteins to bind diverse ligands (i.e. fibrinogen, IgG-Fc, Factor H, fibronectin, serum albumin) in the host. Another contributing factor to this diversity may also be encoded in structural imperfections of the coiled coil.

M1 protein contains a number of non-canonical residues (i.e. charged or polar) in coiled-coil positions *a* and *d* that are typically reserved for apolar residues. Thermodynamically it is favorable to bury the hydrophobic surfaces through oligomerization, but this requirement is lessened when polar and charged residues are introduced. The B-repeats (Fg-binding) and S-region (IgG-Fc-binding), two

.

functionally important regions of M1 protein, have a high frequency of non-canonical residues in the helical packing interface. The significance of maintaining non-canonical residues is not clear. However, it is speculated that interruption of the coiled coil may create greater diversity in binding surfaces and work to enable binding to different host ligands. Determination of the atomic structure of M1 protein by X-ray crystallography was pursued to better understand how M proteins function.

# Experimental Procedures

### Cloning, Expression, and Purification

The coding sequence for M1 protein residues 42–194 was amplified using PCR from *S. pyogenes* invasive strain 5448 of the M1 serotype and ligated into the NcoI and XhoI restriction sites of pET28b (Novagen). A terminator codon was included before the XhoI site to prevent addition of a C-terminal His tag (LEHHHHHH). The libraries of cysteine and methionine substitution mutations (Table 4.5) were introduced by site-directed mutagenesis (Stratagene) using strand overlap extension PCR (Higuchi et al., 1988). These constructs contain an artifactual sequence at the N-terminus (MV) for expression purposes.

.

Expression of tagless HB (nHB; 42–194) was induced in *E. coli* BL21 (DE3) using 1 mM isopropyl-b-D-thiogalactopyranoside (25ºC). The nHB construct and mutants were purified as previously described (Chen et al., 1995). In summary, bacteria were harvested by centrifugation and lysed by sonication (in 50mM tris buffer, pH 8.0, 150 mM NaCl, 10 mM βME, 5 mM EDTA, and 1 mM PMSF). The lysate was heated at 75˚C for 30 minutes, cooled on ice for 10 minutes, and clarified by centrifugation at 35000xg for 30 minutes (4˚C). Nucleic acids were removed by means of 0.5% polyethyleneimine precipitation; the supernatant was precipitated with 65% saturated ammonium sulfate, and resuspended and dialyzed in 20 mM tris buffer, pH 8.0 and 30 mM NaCl. The dialyzed sample was applied to an anion exchange column (Poros HQ) and eluted with a NaCl gradient. Pooled fractions were concentrated using Amicon ultrafiltration (MW cutoff 10 kD; Millipore) and further purified by size exclusion chromatography (Superdex 200) in 10 mM tris buffer, pH 8.0 and 100 mM NaCl. Purified nHB was concentrated to 20 mg/ml ($\varepsilon_{280}$=9530 $cm^{-1}M^{-1}$) and stored as flash-frozen aliquots at -80˚C.

Selenomethionine was incorporated into methionine mutants as described previously (Budisa et al., 1995) and purified as above, except for the addition of 1 mM dithiothreitol carried out in all the purification steps.

.

**Screening of heavy atom derivatives by native gel**

Identification of heavy atoms that bind nHB were quickly screened as described previously (Boggon and Shapiro, 2000). Briefly, 10 μg nHB was incubated with 5 mM heavy atom compounds in 10 μl of the crystallization solution for 1 hour at 4˚C. Non-reducing PAGE sample buffer (2x) was added and the samples were resolved by 10% PAGE.

**Crystallization and data collection**

Crystals were grown at 4ºC by the sitting drop vapor diffusion method using a 1:1 mixture of nHB and 32% 2-methyl-2,4-pentadiol, 275 mM ammonium acetate, and 100 mM citrate (pH 6.05). Constructs HC, AC and HB were crystallized by the same method but with the conditions and temperatures described in Table 4.1. Crystals were mounted in fiber loops and flash cooled in liquid $N_2$. Crystals of selenomethione mutants were first washed with mother liquor supplemented with fresh 1 mM TCEP before mounting. Single wavelength oscillation data (1.0˚) were collected from cryocooled crystals for HB at beamline 5.0.2 (Advanced Light Source, Berkeley, CA) to 3.52 Å (Table 4.2) and native nHB at beamline 19-ID (Advanced Photon Source, Argonne, IL) to 2.91 Å (Table 4.3). Inverse beam, three-wavelength oscillation data (1.0º) were collected from cryocooled selenomethionine crystals for nHB at beamline 19-ID to 3.20 Å and I148M-nHB at beamline 5.0.2 to 2.92 Å (Table 4.6). Data were processed using HKL2000 (Minor, 1997). The program novel_R (Diederichs and Darplus, 1997) was used to calculate the resolution limit of the data.

.

The crystals belong to space group $P2_12_12$ with cell dimensions a = 119 Å, b = 83 Å, and c = 93 Å; four protein molecules are contained in the asymmetric unit (65% solvent).

**Structure Determination and Refinement**

Selenomethionine sites were initially located with SOLVE (Terwilliger and Berendzen, 1999) and subsequent density modification was carried out by RESOLVE (Terwilliger, 2000). The molecular model was built with O (Jones et al., 1991). The initial model was refined against a 2.91 Å resolution data set (beamline 19-ID, Advanced Photon Source, Argonne, IL) using refinement with REFMAC5 (Table 4.6) (CCP4, 1994). The final round of refinement included the assignment of grouped B factors to the model using CNS1.1 (Brunger et al., 1998) followed by insertion of this model into REFMAC5 for solvent correction and anisotropic scaling. A random 5% of data were omitted from refinement for Rfree calculations. The model of nHB includes residues 57–192 for all chains (A, B, C, and D). All residues were checked and determined to within allowed regions of the Ramachandran plot.

.

# RESULTS

### Crystallization of M1 fragments at 4˚C and RT

Crystallization trials were performed against a 300-condition sparse matrix screen using vapor diffusion at RT. The M1 construct and M1 fragments HC, AC, HB, and BC were the only constructs tested. HC was found to crystallize in two different conditions with isopropanol and 2-methyl-2,4-pentanediol (MPD) acting as precipitants (Table 3.1). Optimization of the conditions yielded single crystals (Figure 4.1A and B) that demonstrated weak diffraction (>30 Å) from synchrotron radiation. Screening of various cryoprotectants had no improvement on diffraction, and a RT mount of the crystals yielded the same weak diffraction.

**Table 3.1. Summary of crystallization trials at RT and 4˚C.**

| Construct | T (˚C) | Condition | Morphology | Diffraction Limit |
|---|---|---|---|---|
| HC | RT | 45% isopropanol<br>0.1 M Na-citrate, pH 5.6 | hexagonal plates<br>(300 x 300 x 20 μm) | >30 Å |
| AC | RT | 32% MPD<br>0.1 M MES, pH 6.0 | rhombahedral crystals and hexagonal plates<br>(130 x 50 x 50 μm) | >30 Å |
| HB | 4 | 10% t-butanol<br>200 mM Na-citrate 0.1 M Mes, pH 6.9 | rod-like shape<br>(400 x 70 x 40 μm) | ~ 8 Å |

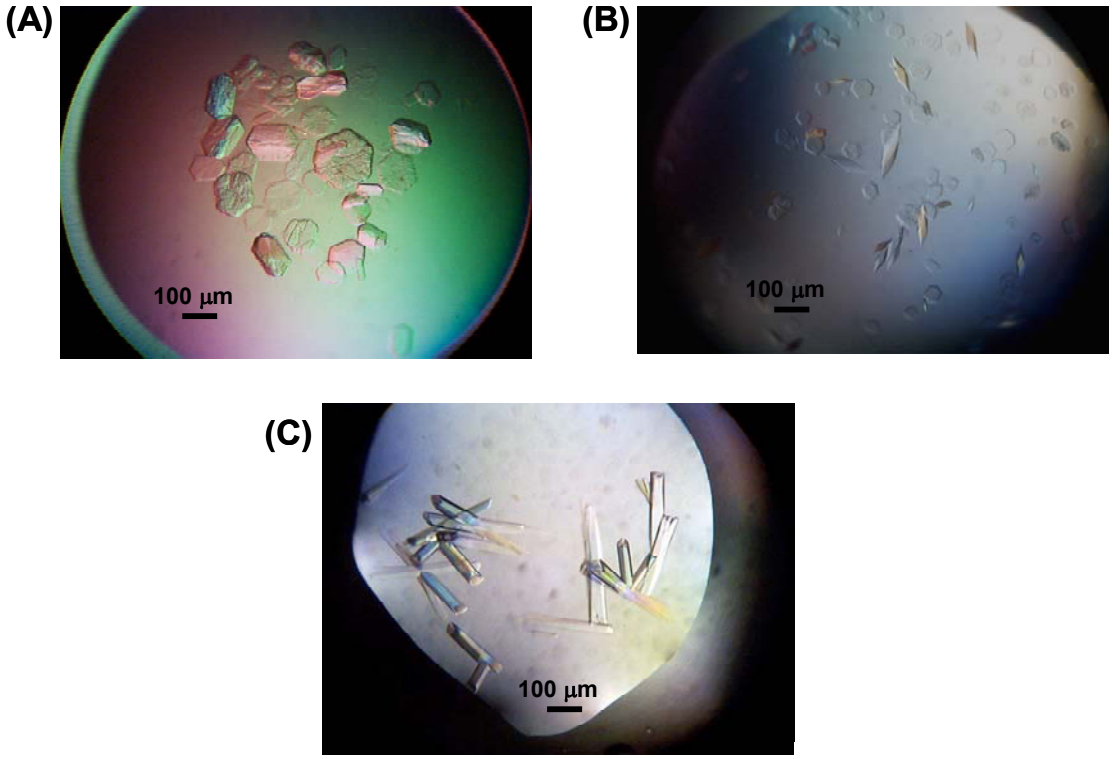
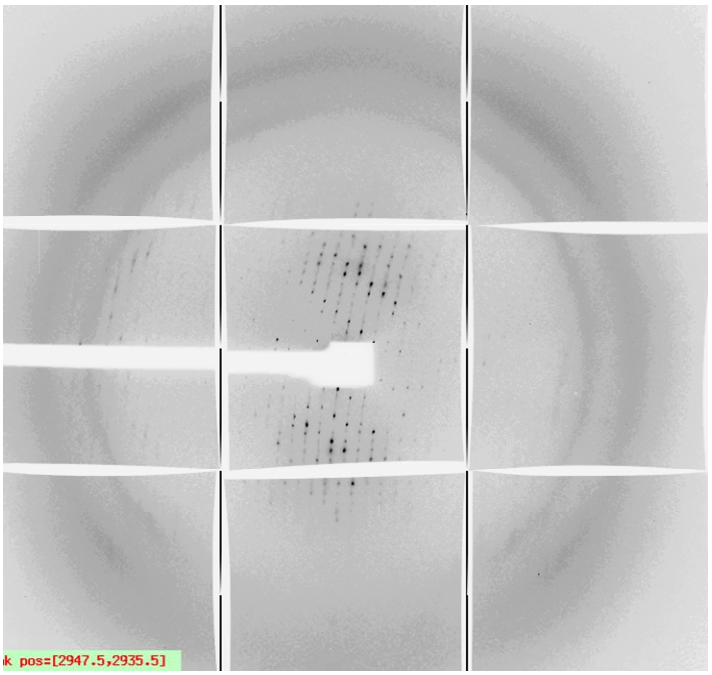



**Figure 3.1. Images of crystallized M1 fragments obtained by sitting drop vapor diffusion.** (A) Construct HC crystallized in 40% *t*-butanol (or 45% isopropanol) and 0.1 M citrate, pH 5.60, at RT. (B) Construct HC crystallized in 32% MPD and 0.1 M Mes, pH 6.5, at RT. (C) Construct HB crystallized in 10% *t*-butanol, 200 mM Na-citrate, and 0.1 M Mes, pH 6.9, at 4˚C.

Screening was repeated at 4˚C and a condition related to the previously described HC crystallization condition produced crystals of the HB construct (Table 3.1 and Figure 3.1C). The HB crystals proved to be very fragile and sensitive to osmotic changes, so cryoprotection was performed in small, stepwise dilutions of the crystals in a sitting drop format. No suitable reagent could be identified that reached cryoprotectant concentrations without cracking or dissolving the crystals.

(A)




(B)

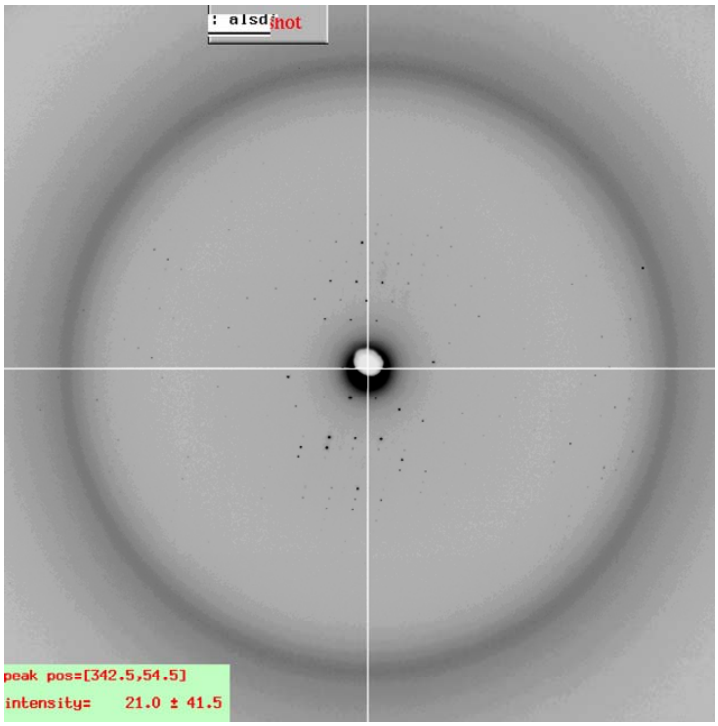



**Figure 3.2. Improved diffraction with optimized cryoprotection.** (A) Native crystals cryoprotected by stepwise additions of glycerol to 20% and PEG-750MME 10%. (B) Diffraction resulting from crystals grown in the presence of 5% glycerol and 2% PEG-750MME, and cryoprotected by stepwise addition to 20% glycerol and 10% PEG-750MME.

**Table 3.2. Summary of HKL2000 processed data of HB crystals from synchrotron radiation (ALS, BL5.0.2).**

| Resolution shells | | Intensity ( I ) | $\sigma_I$ | $I/\sigma_I$ | $R_{merge}$ | Completion ( % ) |
|---|---|---|---|---|---|---|
| 50.00 | 7.53 | 771.5 | 13.9 | 55.5 | 0.027 | 99.4 |
| 7.53 | 5.98 | 94.7 | 4.8 | 19.7 | 0.095 | 100.0 |
| 5.98 | 5.23 | 62.6 | 5.2 | 12.0 | 0.151 | 100.0 |
| 5.23 | 4.75 | 186.3 | 10.5 | 17.7 | 0.080 | 98.4 |
| 4.75 | 4.41 | 136.5 | 9.1 | 15.0 | 0.094 | 80.9 |
| 4.41 | 4.15 | 72.1 | 8.3 | 8.7 | 0.172 | 72.4 |
| 4.15 | 3.94 | 59.8 | 8.4 | 7.1 | 0.220 | 67.4 |
| 3.94 | 3.77 | 37.7 | 9.2 | 4.1 | 0.338 | 64.2 |
| 3.77 | 3.63 | 35.3 | 10.7 | 3.1 | 0.388 | 55.0 |
| 3.63 | 3.50 | 19.2 | 10.5 | 1.8 | 0.598 | 47.6 |
| Averages | | 179.9 | 8.9 | 20.2 | 0.077 | 78.9 |

Although it did not prevent crystal deterioration, a combination of cryoprotectants (20% glycerol and 10% PEG 750MME) proved to yield the best results. Damaged, but cryoprotected crystals, were tested for diffraction on a synchrotron source. Smeared reflections were strong to ~7-8 Å with very weak and anisotropic diffraction also detected in patches at ~4 Å (Figure 3.2A). Ultimately, diffraction was improved by introducing 3-5% glycerol and 1-2% PEG 750MME to the original crystallization condition. The presence of both cryoprotectants prior to their stepwise additions conferred a significant increase in crystal stability and eliminated cracking during cryoprotection. The result was a vast improvement in the diffraction quality and resolution (Figure 3.2B).

.

Data for this crystal were processed to 3.52 Å (2σ cut-off) with an overall $R_{merge}$= 7.7% (Table 3.2). Due to an indistinguishable screw axis, the HB crystals were indexed to both $P3_121$ and $P3_221$ with dimensions a = b = 120 Å, c = 92 Å and α = β = 90˚, γ = 120˚. The diffraction pattern showed strong anisotropy typical of fiber diffraction. The Matthew's coefficient ($V_m$ = 2.37, 48% solvent content), calculated using the CCP4 suite (CCP4, 1994), corresponded to four molecules (two dimers) per asymmetric unit.

**Improved crystallization condition using a tagless HB construct (nHB)**

Concurrent efforts to find a condition(s) yielding high resolution-diffracting crystals were sought during optimization of the HB crystals. The C-terminal hexahistidine tag (LEHHHHHH) to expose a new packing surface. The new construct failed to crystallize in the HB condition suggesting the hexahistidine tag was important in forming crystal contacts in those crystals. Recognizing the propensity for M1 fragments to crystallize in acidic pH and alcohol-based precipitants, a new condition was identified with 32% MPD as a precipitant at 4˚C (Figure 3.3A). A great advantage to this condition was the cryoprotection afforded by MPD at this concentration. To confirm the identity of the crystallized component, crystals were repeatedly washed with mother liquor and examined by SDS-PAGE (Figure 3.3B). Lane X shows a strong nHB band that migrates at the same mass as the purified nHB in the lane labeled "nHB."

.

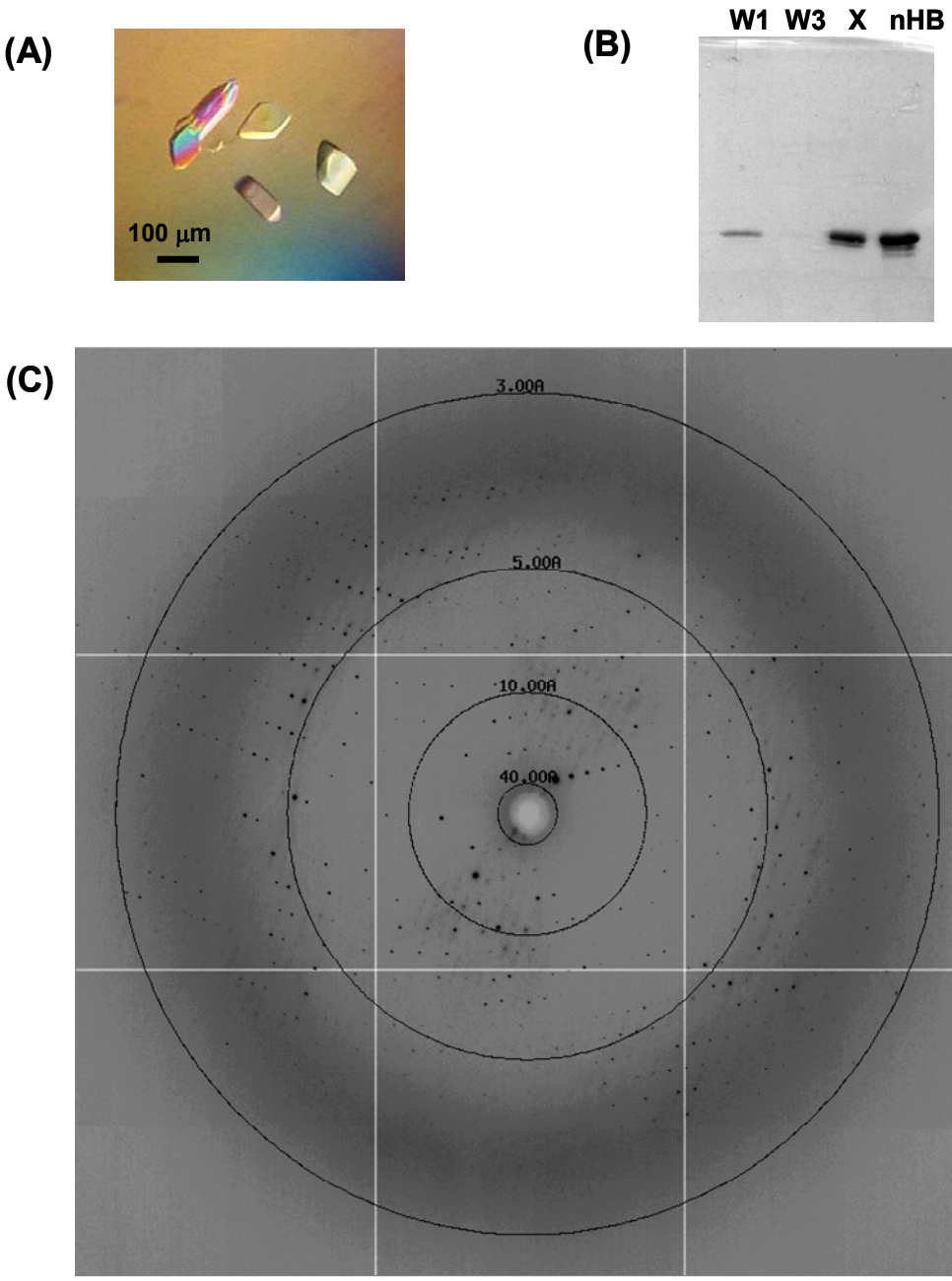



**Figure 3.3. Crystallization and diffraction of tagless HB (nHB).**

(A) Removal of the C-terminal hexahistidine tag (LEHHHHHH) lead to successfull crystallization of this construct in a new condition (32% MPD, 260-310 mM (NH4)OAc, 0.1 M Na-citrate, pH 6.05, at 4˚C. (B) Visualization of nHB crystals by 12.5% SDS-PAGE. Lanes correspond to first 10 μl wash of nHB crystal in mother liquor (W1), third and final wash (W3), crystal (X) and purified nHB used for crystallization trial (nHB). (C) Diffraction image collected from synchrotron radiation (APS, 19ID).

Data images revealed a strong anisotropy indicative of fiber diffraction (Figure 3.3C) as with HB and reflections were processed to 2.91 Å with $R_{merge}$ = 8.1% (Table 3.3). The nHB crystals were indexed to space group $P2_12_12$ with cell dimensions a = 119, b = 83, c = 93, and $\alpha = \beta = \gamma = 90^{\circ}$. Calculation of the Matthew's Coefficient ($V_m$) gave highest probability that 3 dimers ($V_m$ = 3.15; 42% solvent) were present in the asymmetric unit and a lesser probability for 2 dimers ($V_m$ = 2.10; 61% solvent).

**Table 3.3. Summary of HKL2000 processed data of nHB crystals from synchrotron radiation (APS, 19ID).**

| Resolution shell | | Intensity ( I ) | $\sigma_I$ | $I/\sigma_I$ | $R_{merge}$ | Completion ( % ) |
|---|---|---|---|---|---|---|
| 50.00 | 6.27 | 3492.8 | 174.9 | 20.0 | 0.035 | 97.0 |
| 6.27 | 4.97 | 849.1 | 41.1 | 20.7 | 0.068 | 100.0 |
| 4.97 | 4.35 | 1401.5 | 59.8 | 23.4 | 0.062 | 99.9 |
| 4.35 | 3.95 | 765.9 | 36.3 | 21.1 | 0.075 | 100.0 |
| 3.95 | 3.67 | 462.1 | 33.8 | 13.7 | 0.118 | 100.0 |
| 3.67 | 3.45 | 225.3 | 31.1 | 7.2 | 0.224 | 100.0 |
| 3.45 | 3.28 | 143.6 | 31.5 | 4.6 | 0.333 | 100.0 |
| 3.28 | 3.13 | 138.8 | 32.4 | 4.3 | 0.334 | 99.4 |
| 3.13 | 3.01 | 109.4 | 33.4 | 3.3 | 0.383 | 98.4 |
| 3.01 | 2.91 | 78.0 | 35.6 | 2.2 | 0.462 | 96.5 |
| Averages | | 788.1 | 51.8 | 15.2 | 0.081 | 99.1 |

**Heavy atom derivative screen**

As a means to determine the X-ray phasing solution to the nHB data, formation of heavy atom derivatives was investiagated. A quick and efficient screening method was used to filter out heavy atoms that do not bind nHB. Samples of 10 μg nHB (pI ~ 5.6) were incubated 1 hour in the presence of 5-10 mM heavy atom compounds then visualized by native PAGE. Lanes demonstrating a gel shift were considered to have

.

**Table 3.4. Summary of attempts to produce heavy atom derivatives with native nHB crystals.**

| | Compound | gel shift | soaks | diffraction | site |
|---|---|---|---|---|---|
| 1 | cis-diammine dichloroplatinum (II) | yes | 1 - 33 days | 3.2 Å | no |
| 2 | mersalyl acid | no | - | - | - |
| 3 | thallium (III) chloride | yes | severe cracking | - | - |
| 4 | thalium (I) nitrate | no | - | - | - |
| 5 | mercury ethly phosphate | no | - | - | - |
| 6 | p-chloro-mercuriphenylsulfonic acid | no | - | - | - |
| 7 | triethyl lead acetate | no | - | - | - |
| 8 | iridium chloride hydrate | ppt | severe cracking | - | - |
| 9 | methyl mercury (II) chloride | no | - | - | - |
| 10 | europium (III) chloride | no | - | - | - |
| 11 | lead (III) nitrate | no | - | - | - |
| 12 | potassium hexachloroplatinate | ppt | 1 - 33 days | 3.1 Å | no |
| 13 | uranyl acetate dihydrate | no | - | - | - |
| 14 | samarium (III) nitrate | no | - | - | - |
| 15 | phenyl mercuric acetate | no | - | - | - |
| 16 | gold (III) chloride | ppt | 1 - 7 days | 8 Å | no |
| 17 | Bakers dimercurial | no | - | - | - |
| 18 | uranyl nitrate hexahydrate | no | - | - | - |
| 19 | potassium hexacyanoplatinate | no | - | - | - |
| 20 | samarium (III) chloride | ppt | severe cracking | - | - |
| 21 | mercury chloride | yes | 1 - 14 days | 3.0 Å | no |
| 22 | mercury acetate | yes | 1 - 14 days | 3.2 Å | no |
| 23 | mercury iodide | no | - | - | - |
| 24 | silver (I) nitrate | ppt | - | - | - |
| 25 | PCMB | no | 1 - 7 days | 3.2 Å | no |
| 26 | ytterbium (III) chloride | no | - | - | - |
| 27 | holmium (III) chloride | no | - | - | - |
| 28 | lutetium (III) acetate | no | - | - | - |
| 29 | lutetium (III)chloride | no | - | - | - |
| 30 | gadolinium (III) chloride | no | - | - | - |
| 31 | erbium (III) chloride | no | - | - | - |
| 32 | terbium (III) chloride | no | - | - | - |

specifically bound a heavy atom. The summarized results (Table 3.4) show that only four of the thirty-two compounds induced a detectable gel shift. Five others caused precipitation during the incubation or failed to enter the gel suggesting multiple, non-specific, binding events. All nine compounds were tested further and soaked with native nHB crystals for periods ranging from hours to weeks. Thallium, iridium and samarium compounds tested at concentrations from 0.1 – 5 mM destroyed the integrity of the crystal in all conditions. Mercury and platinum (II) did not compromise the crystal integrity but anomalous signal was never detected in the scaled reflections, proving binding did not occur. Finally, gold appeared to bind nHB but detrimentally altered the diffraction pattern quality. Reflections were smeared and diffraction was not observed beyond 8 Å. Unfortunately, heavy atom derivatization of native nHB via soaks proved to be unsuccessful.

**Screening of nHB mutants for a phasing solution**

The failure to produce useful heavy atom derivatives from soaks necessitated alternative experimental approaches for introducing heavy atoms. Selenomethionine-based single anomalous dispersion (Se-SAD) and multiple anomalous dispersion (Se-MAD) experiments with nHB showed weak anomalous scatter and uninterpretable electron density maps (data not shown). This was expected because nHB (~18 kD) contains only one internal methionine (A-region, residue 81) giving an unfavorable ratio of protein mass per selenium atom. In addition to low methionine content, M1 protein contains no cysteines. Mercury, a powerful heavy atom because of its large mass and the anomalous scatter it produces, forms covalent interactions with free

.

cysteines. The binding of one mercury atom per nHB polypeptide would be sufficient to provide phasing information by Hg-SAD/MAD experiments. Mutant constructs were designed to introduce a cysteine or a methionine using site-directed mutagenesis.

Prediction programs Coils (Lupas et al., 1991), PairCoil (Berger et al., 1995) and MultiCoil (Wolf et al., 1997) were used to assign the heptad register to the nHB residues. Cysteines were mutated into solvent exposed residues occurring in *b*, *c*, and *f* positions while leucines in *a* and *d* positions were targeted for methionine replacement (Figure 3.4A). The cysteine mutations showed variant behavior in the nHB crystallization condition (Table 3.5). Only mutants $K_{96}C$ and $T_{110}C$ were readily

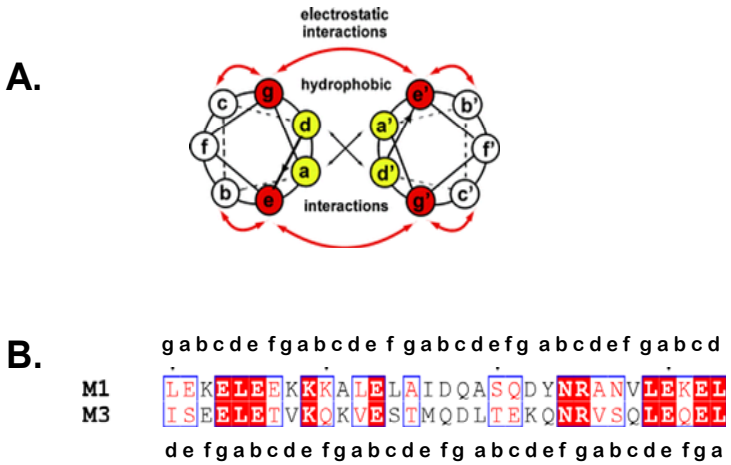



**Figure 3.4. Factors used in the rational design of cysteine and methionine residues.** (A) Helical wheel projection of coiled coil viewed from the N-terminus (taken from Pagel et al., 2005). Leucine residues predicted to occur at *a* and *d* positions were targeted for methionine substitutions. Solvent exposed positions *b*, *c*, and *f* were targeted for cysteine mutations for binding heavy atoms. (B) Alignment of B-repeats from M1 and M3 protein show conserved amino acids amid differing heptad registers.

**Table 3.5. Results of site-directed mutagenesis in nHB to introduce Cys or Met for heavy atom incorporation for phase determination.**

| Residue | Register | Expression | Crystallization | Diffraction |
|---|---|---|---|---|
| $K_{96}C$ | f | (+++) | (+++) | 3.0 Å |
| $D_{107}C$ | c | (++) | (++) | ~6 Å |
| $T_{110}C$ | f | (+++) | (+++) | 3.0 Å |
| $L_{112}C$ | a | (+++) | sperulites | N/A |
| $D_{120}C$ | b | (+++) | (+) | ~6 Å |
| $S_{127}C$ | b | (++) | (+) | ~6 Å |
| $L_{108}M$ | d | (++), degradation | spherulites | N/A |
| $L_{112}M$ | a | (++), degradation | spherulites | N/A |
| $L_{115}M$ | d | (++), degradation | spherulites | N/A |
| $L_{112,115}M$ | a/d | (++), degradation | spherulites | N/A |
| $I_{148}M$ | a | (+++) | (+++)* | 3.0 Å |
| $L_{161}M$ | g | (++) | spherulites | N/A |
| $L_{133,161}M$ | g/g | (++), degradation | spherulites | N/A |
| $I_{176}M$ | a | (+), degradation | spherulites | N/A |
| $I_{148,176}M$ | a/a | (+), degradation | spherulites | N/A |

(+++) = equivalent to native nHB
(++) = compromised in comparison to nHB, yet suitable
(+) = severe limitations in expression/crystal quality and optimization

* In the presence of 0.5-2% xylitol

optimized to produce crystals with diffraction of comparable quality to that found in native crystals. Crystals were soaked with fresh reducing agent then washed prior to heavy atom addition to assure free cysteine. Soaks with 0.1 – 2 mM mercury compounds caused cracking and dissolution of the crystals after 30 – 90 minutes. Poor quality diffraction patterns resulted and no anomalous signal could be measured. Harvesting of the crystals at time points just prior to cracking preserved the diffraction pattern but no heavy atom binding was detected. Cross-linking with glutaraldehyde

stabilized the crystal, yet detection of mercury binding remained elusive (data not shown).

The nHB construct was also sensitive to methionine mutations at leucine residues at positions 108, 112, 115, and 112 and 115 together. Many of the constructs showed strong degradation products and all mutations abrogated crystallization (Table 3.5). The challenges of finding a successful mutant lead to searches of methionine positions in other M proteins. The M3 protein was the only other M protein to have a methionine residue N-terminal of the C-repeats. Alignment of the M1 protein sequence, showed some amino acid conservation of their B-repeats (Figure 3.4B). Interestingly, the conserved residues did not share the same heptad register and the methionine residue (position *e*) in the M3 B-repeat aligned to isoleucine 148 (position *a*) in the M1 B-repeats. Since Met81 in M1 protein occurs at an *e* position and produced anomalous signal, showing the residue was ordered, mutations were directed to *e/g* positions occupied by hydrophobic residues. Leu133 and Leu161 (B-repeats, position *g*) were identified as the most similar to Met81 and engineered with a methionine mutation.

Of the generated mutants, only $I_{148}$M-nHB could be optimized in crystallization trials of additives to yield diffraction comparable to the native nHB crystals. Moderate crystal decay made Se-SAD experiments more desirable than those using Se-MAD. Up to 7 sites were found from SAD experiments using SHELX and SOLVE. Experimental electron density maps from RESOLVE using SAD data

.

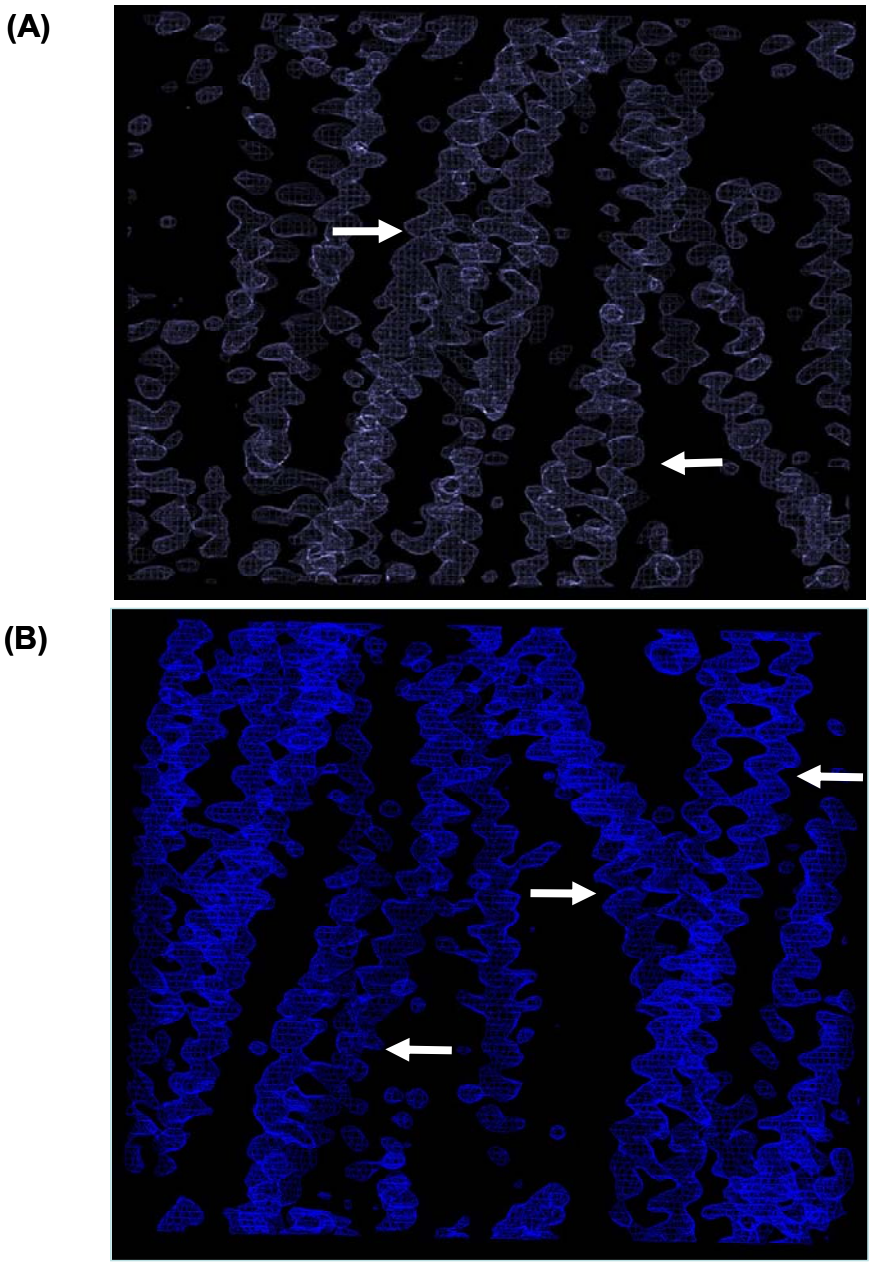



**Figure 3.5. Solve/Resolve experimental electron density maps of nHB constructs.** (A) Map using a Se-SAD data collection from $I_{148}$M-nHB crystals shows broken helical density (white arrows). (B) Map produced from a 4-λ Se-MAD data set collection from $I_{148}$M-nHB crystals yielded improved helical density with evidence of coiled-coiling (white arrows). The asymmetric unit was defined differently in each solution thus the views are not identical.

collection consistently produced interpretable, but interupted, helical density (Figure 3.5A). The number of sites found best corresponded with two dimers per asymmetric unit. Se-MAD showed little or no improvement over SAD experiments, with the exception of one crystal that withstood a four-wavelength Se-MAD data collection. Eight heavy atom sites were identified in SOLVE and RESOLVE produced an improved electron density map (Figure 3.5B).

**Model building**

Model building was performed in the program O using 14-residue poly-alanine helices to piece together the model fit it into the density. The amino acid register was deduced using the position and identity of the heavy atom sites (Se-Met81, Se-Met148), density for large hydrophobic side chains (Phe88, Tyr119, Tyr155, Tyr183) and accounting for the fact that side chains point towards the N-terminus in alpha helices (Christmas tree effect). Four poly-alanine helices were placed into density with side chain density for approximately 80% of the residues. The remaining invisible side chains that lacked density were still modeled using the most frequently occurring rotamers that avoided unfavorable interactions. The model was refined in REFMAC5 and the $2F_o - F_c$ and $F_o - F_c$ maps were used to improve the model before subsequent rounds of refinement. Omit maps (5-10% of residues removed from the model) were commonly employed for validation and to assure lack of model bias (Figure 3.6).

.

The model was then fit into the experimental electron density created from a native data set collected to 2.90Å. It was determined that individual, isotropic B factors could not be used in refinement because of the low observation to parameter ratio (17523 reflections:18196 parameters) and because it caused the $R_{free}$ to significantly increase.

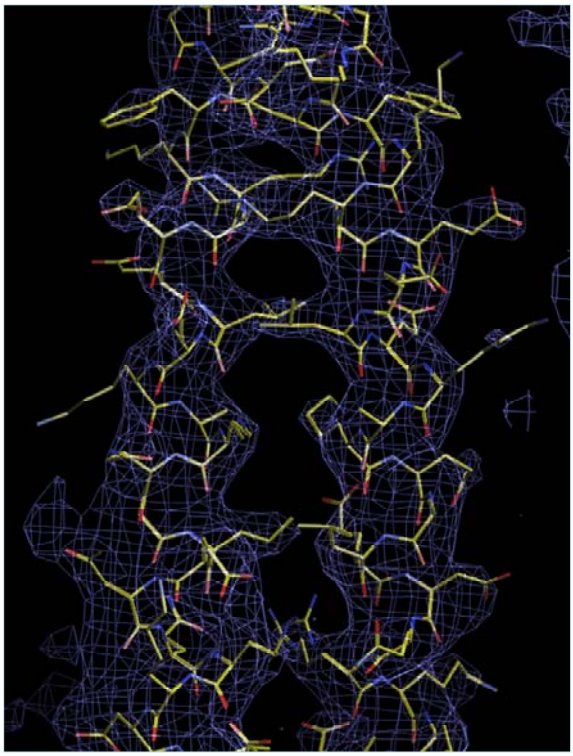


**Figure 3.6. Validating the model with $2F_o - F_c$ electron density omit maps.** Refinement was performed after residues 87 – 106 of chains A and B (~14% of the modeled residues) were omitted from the nHB model. The resultant density (1σ level) correlated extremely well to $2F_o - F_c$ maps generated with the entire model. The map also is representative of the lack of density for ~20% of the solvent-exposed side chains in the entire structure.

Refinement in REFMAC5 was restricted to assigning overall B factors to each molecule. Grouped B factors, which reduced the number of parameters by 3453

**Table 3.6.** X-ray data colleciton and refinement, nHB. Statics for the nHB model from the native crystal and Se-MAD data set. The statics for the highest resoltuion shell is given in parentheses after the cumulative value for all the shells.

| | Native | Se-MAD | | | |
|---|---|---|---|---|---|
| **Data Collection** | | | | | |
| Space group | $P2_12_12$ | $P2_12_12$ | | | |
| Cell dimensions | | | | | |
| *a, b, c* (Å) | 119.37, 83.40, 93.47 | 119.81, 83.63, 93.20 | | | |
| $\alpha, \beta, \gamma$ (°) | 90.00, 90.00, 90.00 | 90.00, 90.00, 90.00 | | | |
| | | *Peak* | *Inflection* | *Remote1* | *Remote 2* |
| Wavelength (Å) | 0.97934 | 0.9793 | 0.9794 | 0.9567 | 1.008 |
| Resolution | 50.00−2.91 | 50−2.81 | 50−2.90 | 50−2.90 | 50−2.78 |
| Highest shell | (3.00−2.91) | (3.00−2.91) | (2.91−2.81) | (3.00−2.90) | (2.88−2.78) |
| Rmerge | 8.1 (46.2) | 9.8 (52.3) | 9.1 (35.1) | 9.9 (39.3) | 9.9 (42.7) |
| I/σI | 15.2 (2.2) | 15.9 (1.9) | 11.9 (2.3) | 11.1 (1.9) | 9.7 (1.9) |
| Completeness (%) | 99.1 (96.5) | 99.5 (94.0) | 99.2 (96.1) | 99.0 (95.5) | 99.1 (93.2) |
| Redundancy | 5.3 (4.3) | 7.0 (5.1) | 3.6 (3.0) | 3.6 (2.8) | 6.5 (5.0) |
| **Refinement** | | | | | |
| Resolution (Å) | 50.00−3.03 (3.11−3.03) | | | | |
| No. unique refl. | 17523 | | | | |
| σ cutoff | None | | | | |
| *R cryst*, *R free* | 32.84, 33.23 (54.90, 55.4) | | | | |
| No. of atoms | | | | | |
| Protein | 4549 | | | | |
| Water | None | | | | |
| R.m.s. deviations | | | | | |
| Bond length (Å) | 0.009 | | | | |
| Bond angles (°) | 1.11 | | | | |
| *B*-factors | | | | | |
| Molecule A (main, side) | 84.9 (71.90, 97.1) | | | | |
| Molecule B (main, side) | 84.8 (69.2, 99.5) | | | | |
| Molecule C (main, side) | 102.5 (92.1, 112.3) | | | | |
| Molecule D (main, side) | 87.1 (74.4, 99.1) | | | | |

(17523 reflections: 14743 parameters), were assigned in CNS, and then resubmitted to REFMAC5 for rescaling the data. The resultant model, solved to a final resolution of 3.03Å, fit residues 57 – 192 of each chain with an $R/R_{free}$ of 32.8%/33.2% (Table 3.6).

**Overall structure**

The final model is comprised of two dimeric, parallel coiled coils (molecules AB and CD) that pack in an overlapping tail-to-tail orientation (Figure 3.7A). The packing is accommodated by the splaying of the coiled coils in the C-terminal B-repeats (residues 133 – 192) of each dimer molecule. The free helices preferentially form anti-parallel, coiled coils between molecule A of dimer 1 and molecule D of dimer 2, and molecule B of dimer 1 and molecule C of dimer 2. The intermediate A-region (residues 61 – 132) forms a canonical parallel coiled coil (Figure 3.7B). Superposition of dimer AB onto CD shows significant differences in the main chain geometry throughout this region. This difference in geometry accounted for the lack of non-crystallographic symmetry (NCS) and suggests that the A-region has dynamic properties. Finally, the non-helical N-terminus, referred to as the hypervariable region (HVR; residues 42 – 60), is largely missing from the electron density map. Residues 57 – 60 were modeled as an extended loop and it is assumed that the remaining residues are not ordered.

**Parallel coiled coil**

N-terminal residues 63-132, the A-region, form the apparent boundaries of the dimeric, parallel coiled coil in molecules AB and CD. Analysis by TWISTER

.

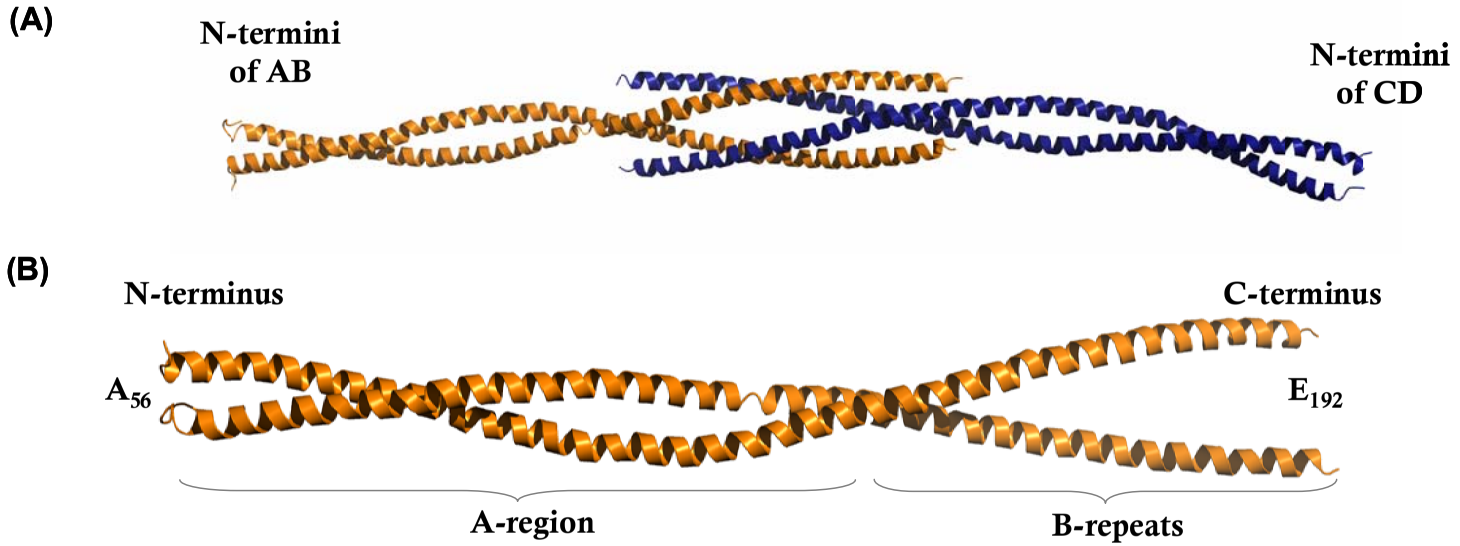



**Figure 3.7. Structure of HB fragment from M1 protein. (A) The asymmetric unit contains two nHB dimers that form a tail-to-tail tetrameric complex. (B) n**HB is a dimeric, coiled-coil protein measuring ~200 Å in length. The hypervariable region (residues 42-59) is non-helical with density only for residues 57-60. The A-region (residues 61 – 132) form a dimeric, coiled-coil structure. The B-repeats (residues 13-192) are splayed helices.

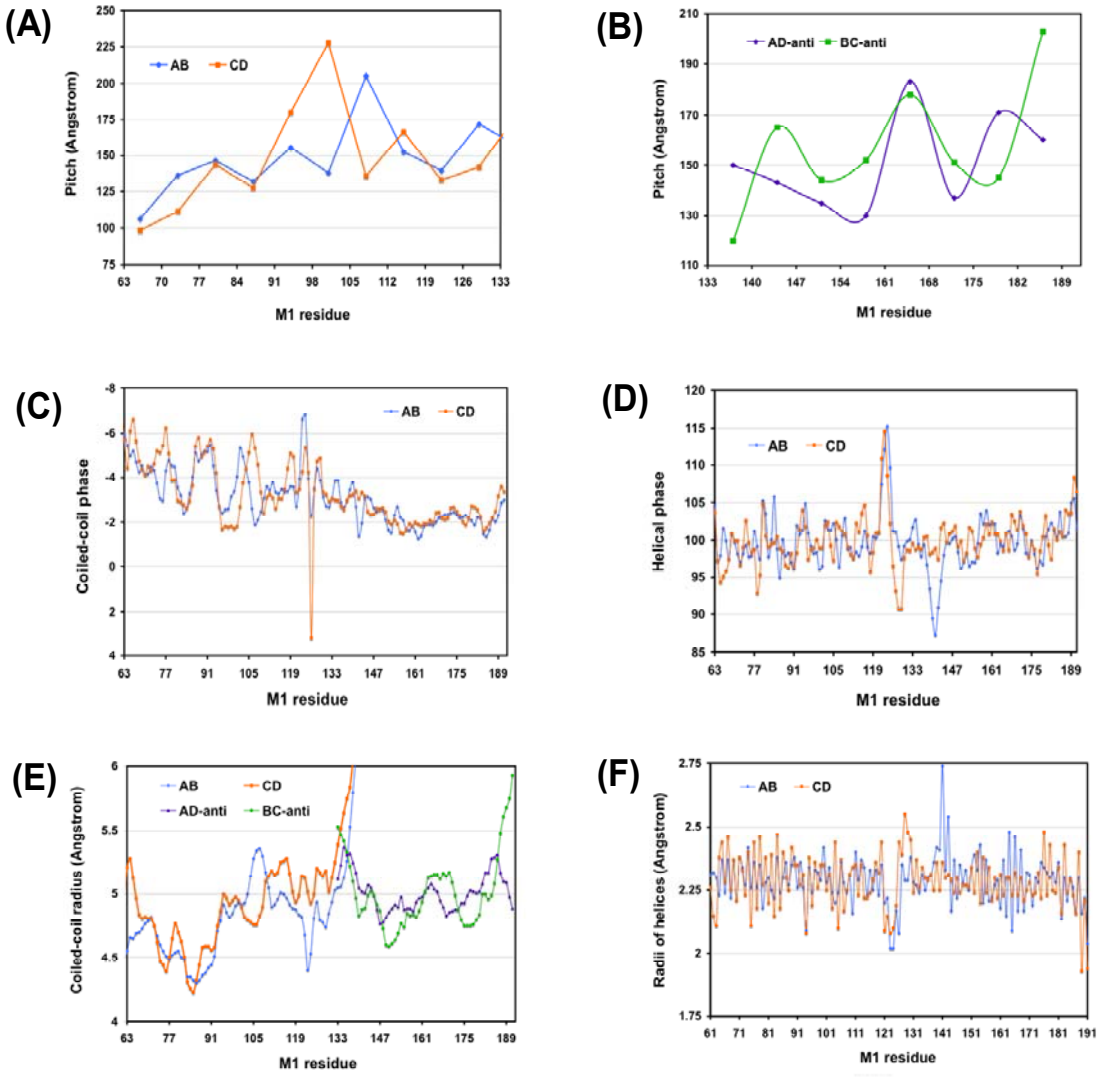



**Figure 3.8. Analysis of the coiled-coil structure and geometry using TWISTER.** The average coiled-coil pitch for each full heptad (abcdefg) in (A) the parallel, A-region interactions for dimer AB (blue) and dimer CD (orange), and (B) the anti-parallel B-repeat interactions between AD (purple) and BC (green). (C) The calculated coiled-coil phases for each residue of molecule AB (blue) and CD (orange). (D) A plot of the helical phases for each residue in molecule AB (blue) and CD (orange). (E) The coiled-coil radius for parallel interactions in dimers AB (blue) and CD (orange), and for anti-parallel interactions in dimers AD (purple) and BC (green). (F) Average helical radius of molecules A and B (blue) and molecules C and D (orange).

(Strelkov and Burkhard, 2002) shows that the average coiled-coil pitch for residues 63-132 is 146º and 147º for the AB and CD dimers, respectively (Figure 3.8A). Additionally the average coiled-coil radius for that same residue range is 4.8 Å for each dimer (Figure 4.8E). The coiled-coil radius and pitch are both consistent with values observed in other parallel, dimeric coiled-coils like GCN4 and tropomyosin. SOCKET (Walshaw and Woolfson, 2001), a program used to evaluate knobs-into-hole packing in coiled coils, showed stretches of canonical packing interspersed with several gaps within the residue range 63-119. Knobs-into-hole packing was not detected for residues 120-132 using the standard cut-off of 7.0 Å. Even use of a less restrictive cut-off of 7.4 Å did not detect knobs-into-hole packing.

**Non-canonical residues in the helical packing interface**

It is not uncommon for polar residues to occasionally occur in *a/d* positions within dimeric coiled coils, but the occurrence of charged residues in these positins is less frequent. Positioning of closely aposed residues generates unfavorable charge-charge repulsions. Also, the elongated sidechains of Glu, Lys, and Arg impose steric hindrance. Interestingly, charged residues occur in the *a* positions of consecutive heptad repeats within the A-region-- Lys98 and Arg105. To alleviate the potential charge repulsion and steric hindrance both residues adopt solvent accessible rotamers that orientate the side chains away from the helical interface and they interrupt complementary knobs-into-hole packing.

.

The density for Lys98 was absent at the distal end of the sidechain in each molecule suggesting it did not adopt a single rotamer. However, density for Arg105 in coils A, B, and D positioned the sidechain within hydrogen bond range of Gln104

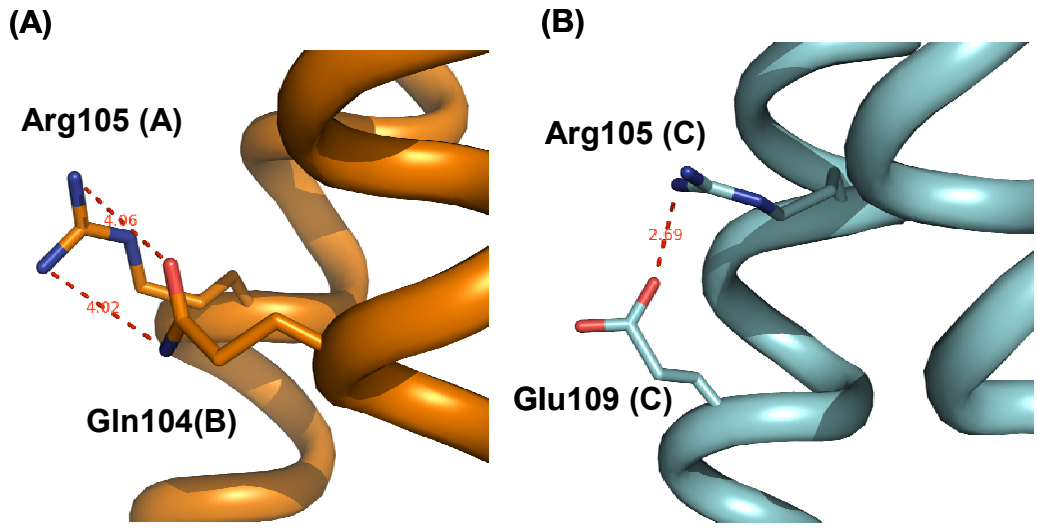



**Figure 3.9. Evidence that Arg105 can participate in inter- or intrahelix interactions to avoid packing into the coiled-coil interface.**

Arg105 occurs in position *a* of the coiled-coil heptad which is typically reserved for apolar residues. (A) In molecules A, B and D Arg105 points away from the helical interface and is positioned within hydrogen-bonding distance(4.02 and 4.06 Å) to Gln104 of the opposing molecule. (B) Arg105 in molecule C forms an intrahelix hydrogen bond (2.69 Å) with Glu109 (*i* + 4 interaction).

(position *g*) in the opposing molecule. Potential interhelix bonds were identified for distances ranging between 3.82 – 4.06 Å for participating atoms of Gln104 and Arg105 (Figure 3.9A). For coil C, however, clear density reveals that the rotamer conformation of Arg105 positions itself to uniquely form an intrahelix hydrogen bond to Glu109 (position *e*) at a distance of 2.69 Å (Figure 3.9B). Intrahelix salt bridges of

this nature (i to i+4) have been found to contribute to monomeric α-helical stability (Spek et al., 1998).

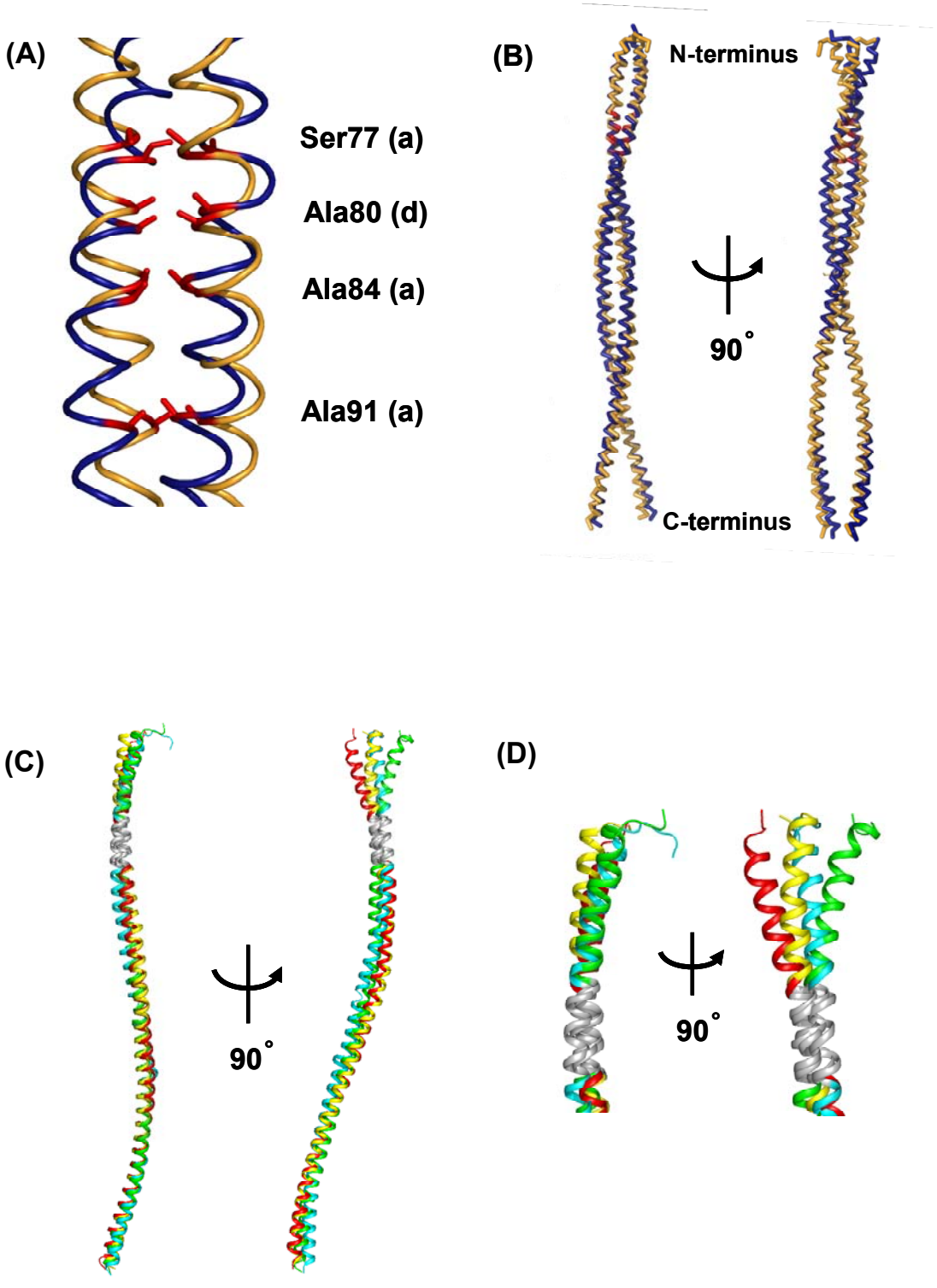
Non-canonical residues, Lys98 and Arg105, induce local unwinding of the coiled-coil superhelix in dimer CD between residues 95-102. Thisdoubles the coiled-coil pitch from 150 to ~300 Å (Figure 3.8A). Dimer AB shows a more moderate rise in coiled-coil pitch over that same range, and instead undergoes a more pronounced superhelical unwinding at residues 106-109 (Figure 3.8A). This is also accompanied by a local increase of the coiled-coil radius from 5.0 to 5.4Å (Figure 3.8E). This disrupts packing of this segment and only Leu108 (coil A; *d* position) is able to maintain an isolated knob-into-hole packing. Since the coiled-coil phase and α-helical phase (Figure 3.8C and D) show no significant changes over these same regions, it can be concluded that there is no local unwinding of the α-helix. Thus M1 protein relies on local superhelical unwinding alone to compensate for these non-canonical residues.

**An alanine cluster at the helical interface increases flexibility of the molecule**

The other prominent gap in the knobs-into-hole packing pattern coincides with a cluster of alanines located at residues 80 (position *d*), 84 (position *a*) and 91 (position *a*) (Figure 3.10A). This stretch is characterized by a significant reduction of the coiled-coil radius from ~5.0 Å to 4.25 Å (Figure 3.8E) and includes Ser77 (position *a*). The disruption in complementary knobs-into-hole packing increases the

.

**Figure 3.10. An alanine-rich core spanning 14 residues contributes to flexibility at the N-terminus.** (A) Enlarged view of alanine-rich core from Ala80 to Ala91. The small bulk of Ser77 (position *a*) also facilitates narrowing of the coiled-coil. Residues are colored red and labeled with residue, number and heptad position. (B) Superposition of dimers AB (orange) and CD (blue) show structural differences in the backbone atoms, with the greatest divergence at the N-terminus. Alanines from residues 80 – 91 are highlighted in red. (C) Superposition of molecules A (yellow), B (cyan), C (green) and D (red) show that the greatest variation precede the alanine-rich core (colored in gray). (D) Close-up view of the top portion of panel (C).


(A)
Ser77 (a)
Ala80 (d)
Ala84 (a)
Ala91 (a)
(B)
N-terminus
90°
C-terminus
(C)
90°
(D)
90°

structural heterogeneity of the A-region N-terminal to the Ala-cluster. Successive alanines over three *a/d* positions and/or alanines that work to disperse optimal residues are known to destabilize coiled coils (Lu and Hodges, 2004).

Superposition of the AB and CD dimers highlights a pronounced change in the coiled-coil axis N-terminal to the alanine cluster (Figure 3.10B). The differences are easier to appreciate when the individual chains are superimposed (Figure 3.10C and D). Analysis with Dyndom (Hayward and Berendsen, 1998) confirmed that the alanine cluster acts to create local hinge regions. Comparison of the A and D chains show residues 88-89 are involved in helix bending; whereas residues 78-81 are the site for bending in the B and C chains. These bends correspond to the boundaries of the Ala-cluster (Ser77 – Ala91).

**The A-region is stabilized by two different canonical coiled-coil stretches**

The destabilized stretches in the A-region are counteracted by canonical coiled-coil segments. Three consecutive leucines (Leu108, Leu112, and Leu115) in *a/d* positions work synergistically to optimize core packing and lend increased stability to residues 108 – 119 (Lu and Hodges, 2004). This stabilized segment flanks the disruptive Lys98 and Arg105 heptads and restores knob-into-hole packing. Also the first two and a half heptads of the A-region show canonical packing. Starting at Asn63 and ending at Ala80, the start of the Ala-cluster, this stretch helps to ‘clamp’ the N-terminus of the dimer.

.

**Asymmetric packing of the coiled coils is initiated by a break in the secondary structure**

In both dimers breaks in the secondary structure were observed from residue 123 to 127 (molecules B and C only) and these lead to asymmetric packing between the coils. Coils A and D retained the correct heptad register whiereas coils B and C undergo a phase shift equivalent to a one residue deletion. The resulting heptad register changed from abcedfg→bcdefga causing Trp128 to change from position *c* to *d* (Figure 3.11). The packing of a bulky tryphtophan to a core position further disrupts the coiled-coil and marks the beginning of a steady increase in coiled-coil radius and producing the splaying of the B-repeat region.

**Anti-parallel coiled coil**

The boundaries for the C-terminal splaying and anti-parallel helical packing for molecules AD and BC occur in the region encompassing residues 133 – 192 (Figure 3.12A and B), and coincide with the entire length of the B-repeats. With very similar coiled-coil geometry, the anti-parallel and parallel interactions display indistinguishable superhelical winding (Figure 3.8B) and coiled-coil radii (Figure 3.8E). Anti-parallel dimers AD (anti-AD) and BC (anti-BC) maintain an average radius of 5.02 Å and coiled-coil pitch values of 152 Å and 157 Å, respectively (Figure 3.8B). Both are typical values for coiled coils. Comparison of the buried surface area in AREAIMOL in the CCP4 suite revealed that the anti-parallel region (30.3%) gave consistent results to that of the A-region of the M1 protein (30.2%), GCN4 (30.3%) and tropomyosin (30.3%).

.

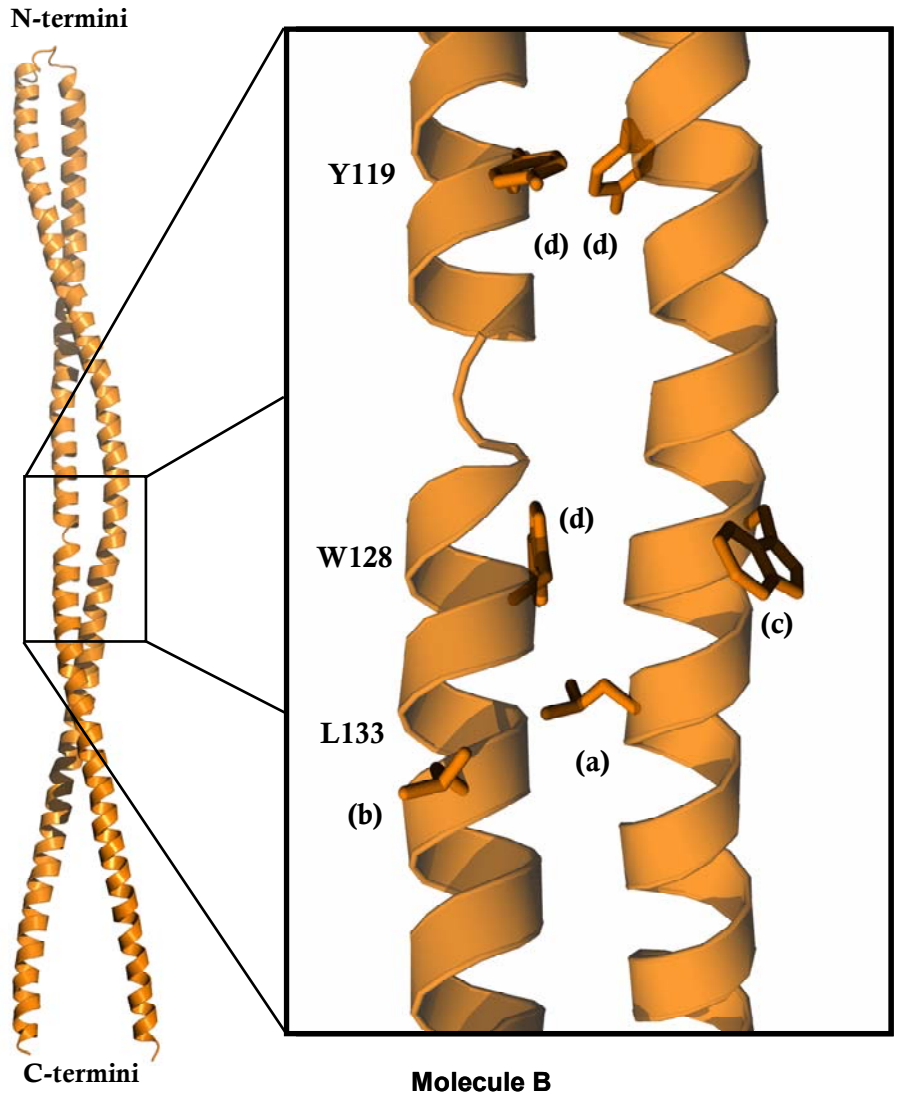



**Figure 3.11. Asymmetric packing induced at the end of the A-region may contribute to splaying of B-repeats.** The final two heptad repeats (residues 119 – 125 and 126 – 132) of the A-region produce heterogeneous secondary structure leading to a change in the coiled-coil register for one of the helices. Following the helical break (residues 124 – 126) in molecule B W128 changes from a solvent exposed *c* position to a sterically unfavorable *d* position. Conversely, L133 changes from an optimal *a* position to a less favorable *b* position.

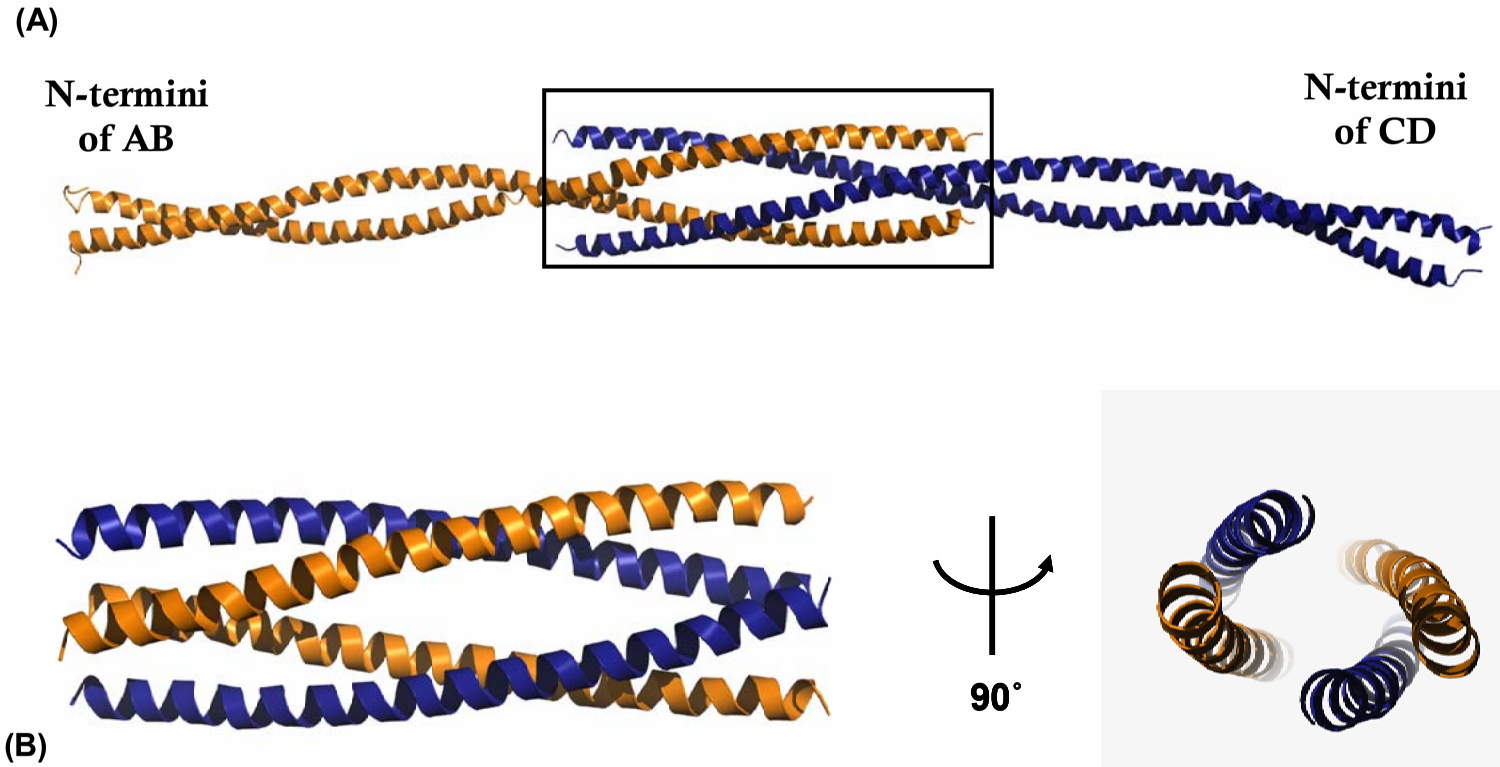



**Figure 3.12. Packing of both nHB dimers in the asymmetric unit.** (A) Each asymmetric unit contains two dimers (AB and CD) packed tail-to-tail via anti-parallel coiled-coils between B-repeats. (B) Clsoe-up view of anti-parallel interactions highlighted by the box in part A. Rotation along the vertical axis by 90˚ shows the supercoiling of splayed helices in AB with splayed helices of CD.

Even though the antiparallel helices retained similar geometric parameters observed in other coiled coils there was very little knobs-into-hole packing identified by SOCKET. Additionally assignment of an uninterrupted heptad register to these residues was not possible. TWISTER and SOCKET produced conflicting assignments adding further evidence that the anti-parallel interactions are abnormal. Relying on SOCKET, the apparent coiled-coil boundaries in anti–AD and –CB extended from residues 141 – 179. Knobs-into-hole packing was consistently observed for Glu162 (position *a*), Leu165 (position *d*), Lys169 (position *a*), and Leu172 (position *d*)., but was sporadic for the rest of the antiparallel interaction. Interestingly, a salt bridge between Glu162 and Lys169 occurs between the antiparallel coils and enables these non-canonical residues to promote canonical coiled-coil packing for a short stretch.

Studies of coiled-coil orientation have shown that electrostatic interactions between e-e' and g-g' can promote antiparallel coiled-coils (Monera et al., 1994). The absence of e-e' and g-g' salt bridges in the antiparallel helices of the B-repeats show further evidence that theydo not form a coiled-coil interaction. Despite the presence of a small region of canonical coiled-coil packing the remainder of the antiparallel coils exhibited no other stabilizing interactions.

Interestingly, the asymmetry introduced at the end of the A-region was resolved in the anti-parallel arrangement. This may be due to that fact that coils A and D (observed no change in the Trp128 register) and coils B and C (observed a +1 change in Trp128 register) form anti-parallel interactions with each other.

.

**Comparison of nHB model to prediction programs**

Residues 42 – 194 of the M1 protein were assigned heptad registers using Coils (Figure 3.13A) and PairCoil (3.13B) prediction servers. While both are consistent with the model PairCoil did a better job of predicting the low probability of coiled-coil packing in the B-repeat segment of M1. The 28-residue window for Coils appeared to be the most accurate prediction of the three generated windows (14-, 21-, and 28-residues). The most pronounced drops in coiled-coil probability (<40%) for both programs occurs at the end of the A-region, which nicely corresponds to disruption of the coiled coil and asymmetric packing in the model.

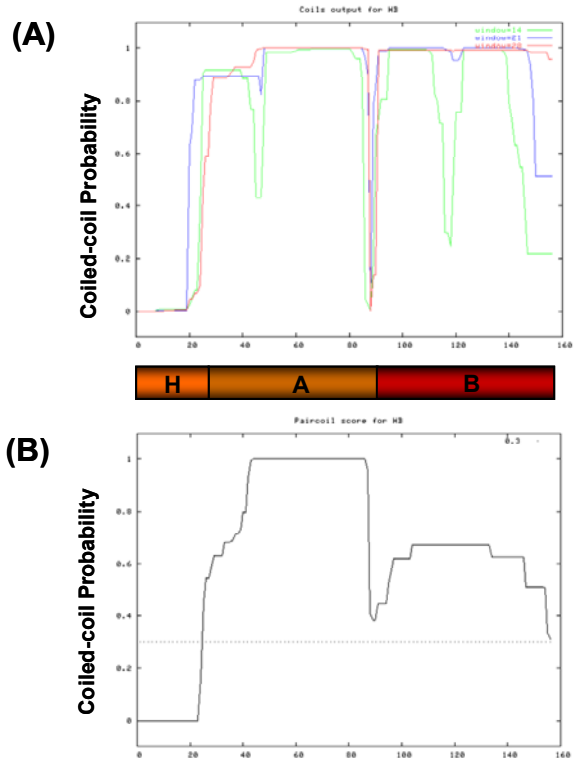



**Figure 3.13. Prediction of coiled-coil probability in mature M1 protein.** (A) Coils and (B) PairCoil server output predictions. Three different residue windows were used in Coils to predict coiled-coil probability: 14-residue (green), 21-residue (blue) and 28-residue (red). The PairCoil prediction most closely matched the experimentally determined structure.

# Discussion

The crystal structure of an N-terminal fragment of M1 protein (nHB) revealed three destabilized segments that caused disruption of coiled-coil packing in the A-region and B-repeats: a grouping of alanines in core packing positions *a* and *d* occurs between residues 80 -91, basic residues are present in consecutive *a* positions (Lys98 and Arg105), and asymmetric packing of the helices beginning at Trp128. Each of these segments contributes to instability of nHB and explains why nHB forms a dynamic oligomeric state between monomer and dimer. The model of nHB also showed similar features to myosin and tropomyosin and may provide insights to the molecular basis of rheumatic fever.

The α-helices in the nHB model adopt distinct tertiary structures in the A-region and B-repeats. The A-region predictably folds into an α-helical coiled coil whereas the α-helices of the B-repeats splay apart from one another. Analysis of the coiled-coil packing in M1 protein shows that the A-region has alternating stabilizing and destabilizing segments. The flexibility and dynamic nature of nHB was evident in the atomic temperature factors and the lack of non-crystallographic symmetry (NCS) between symmetry-related dimers in the asymmetric unit. The average temperature factors for the mainchain atoms were ~70 $Å^2$. NCS was not detected between the A-regions and B-repeats of the AB and CD dimers.

.

The splayed helices of dimer AB and dimer CD preferentially form dimeric, anti-parallel interactions and create a dimer of dimers. Precedents for antiparallel associations of symmetry-related dimers have been observed for two other dimeric coiled coils. These include the APC tumor suppressor (Day and Alber, 2000) and the C-terminal fragment of α-tropomyosin (Li et al., 2002). However, the interactions in these molecules were both characterized as α-helical bundles and do not share the same coiled-coil geometry present in the M1 crystal. Despite the novel observation, biochemical data for nHB does not support tetramer formation *in vitro*. As presented in chapter III, HB and the other M1 fragments freely exchange between monomer and dimer states. Since the B-regions have limited knobs-into-hole packing to stabilize the antiparallel coiled coil, the tetrameric species was determined to be a very weak interaction captured by the crystallization of nHB.

A major factor in the splaying of the B-repeats are the amino acids encoded at the *a* and *d* positions of the coiled coil. For tropomyosin, splaying of a 22-residue segment at the C-terminus is attributed to the occurrence of bulky residues in successive core positions (Tyr to *a*, and Ile to *d*). The breakdown of knobs-into-holes packing leading up to the splaying of the B-repeats occurs by a related, but different, pattern in M1 protein. The bulky Tyr119 at position *a* is followed by Ala122, and these two residues preclude stable coiled-coil packing. The subsequent *a/d* positions are occupied by a less favorable polar (Thr126) and an unfavored charged (Asp129) residue. The break in secondary structure (123-127) in chains B and C induce

.

asymmetric packing and places another bulky residue (Trp128) to core packing position *d*, this in turn triggers a dramatic increase in the coiled-coil radius. This 10-residue stretch of destabilizing residues ensures disruption of coiled-coil packing and lead to helical splaying. Continued appearance of non-canonical residues throughout the predicted *a/d* positions of the B-repeats decreases the coiled-coil probability for this region and maintains the splayed state.

Human cardiac α-tropomyosin, a highly conserved contractile protein between vertebrate species, can share up to 50% sequence identity to segments of M protein and may have relevance in the pathogenesis of rheumatic fever (Mische et al., 1987). Comparison of the nHB model to crystal structures of tropomyosin reveal similar structural features; specifically, both proteins include alanine clusters at core positions that cause multiple bending of the coils (Brown et al., 2001) and each has demonstrated helical splaying (Li et al., 2002).

Unlike tropomyosin, M1 protein has three regions C-terminal to the splayed helices that could promote stability: the S-region (195 – 225), he highly conserved C-repeats (226 – 329), and the D-region (330 – 401). The beginning of the S-region (195 – 204) is predicted to have the same probability as the B-repeats to form coiled-coils. However, the residues in core positions become increasingly canonical. This continues through the C-repeats and into the D-region up to residue 401 (402-453 are non-helical and are involved in cell wall spanning and attachment). While small breaks in the heptad register and/or non-canonical residues can be tolerated in isolated

.

regions of coiled coils, the magnitude of the coiled-coil break in the nHB model seems to be unlikely remedied by the simple addition of a flanking, canonical coiled-coil. The C-repeats most likely restrict the splaying but there is no evidence to show that it could enforce coiled-coil packing.

Another feature common to tropomyosin and the M1 protein is the alanine cluster. The core of tropomyosin contains many alanine clusters that result in multiple bending conformations, which serves a functional role. The A-region of M1 protein contains a cluster from residues 81 – 91 and induces two distinct conformations in the nHB N-termini. A similar phenomenon is observed in the coiled-coiled region of myosin (Elliott and Offer, 1978).

An important question is what functional advantage exists for the B-repeats to maintain a non-canonical coiled coil? One hypothesis is the proposed role that M and M-like proteins play in autoaggregating GAS. Autoaggregation is important in the survival of GAS protects the bacteria from phagocytosis (Frick et al., 2000). A similar mechanism has been reported with hemagluttinin from *Bordetella pertussis* (Menozzi et al., 1994). For the M1 serotype, protein H and M1 protein promote aggregation through homophilic interactions. Due to the restricted orientation of these proteins on the GAS surface, antiparallel interactions seem the likely mode of interaction in aggregates. The N-terminus of M1 protein has been shown to have a dynamic oligomeric state *in vitro* between monomer and dimer when expressed as truncated fragments. So the possibility remains that surface-associated M1 protein may also

.

have a dynamic N-terminus and facilitate antiparallel interactions. However, the crystal structure and biochemistry of M1 fragments support that the parallel coiled coil is more favorable than any potential antiparallel interaction.

The only known function for the B-repeats is to bind fibrinogen, and this represents is a critical event in GAS survival (Ringdahl et al., 2000). The non-canonical nature of the B-repeats is proposed to be is necessary to create this particular mode of function.

.

# References


Berger, B., Wilson, D.B., Wolf, E., Tonchev, T., Milla, M. and Kim, P.S. (1995) Predicting coiled coils by use of pairwise residue correlations. *Proc Natl Acad Sci U S A*, **92**, 8259-8263.

Boggon, T.J. and Shapiro, L. (2000) Screening for phasing atoms in protein crystallography. *Structure*, **8**, R143-R149.

Brown, J.H., Kim, K.H., Jun, G., Greenfield, N.J., Dominguez, R., Volkmann, N., Hitchcock-DeGregori, S.E. and Cohen, C. (2001) Deciphering the design of the tropomyosin molecule. *Proc Natl Acad Sci U S A*, **98**, 8496-8501.

Brunger, A.T., Adams, P.D., Clore, B.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Cryst.*, **D54**, 905-921.

Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J. and Huber, R. (1995) High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in *Escherichia coli*. . *Eur. J. Biochem.*, **230**, 788-796.

CCP4. (1994) The CCP4 suite: programs for protein crystallography. *Acta Cryst.*, **D50**.

Chen, J., Wharton, S.A., Weissehorn, W., Calder, L.J., Hughson, F.M., Skehel, J.J. and Wiley, D.C. (1995) A soluble domain of the membrane-anchoring chain of influenza virus hemagglutinin ($HA_2$) folds in *Escherichia coli* into the low-pH-induced conformation. *Proc. Natl. Acad. Sci. USA*, **92**, 12205-12209.

Day, C.L. and Alber, T. (2000) Crystal structure of the amino-terminal coiled-coil domain of the APC tumor suppressor. *J Mol Biol*, **301**, 147-156.

Diederichs, K. and Darplus, A. (1997) *Nature Structural Biology*, **4**, 269-275.

Elliott, A. and Offer, G. (1978) Shape and flexibility of the myosin molecule. *J Mol Biol*, **123**, 505-519.

Frick, I.M., Morgelin, M. and Bjorck, L. (2000) Virulent aggregates of Streptococcus pyogenes are generated by homophilic protein-protein interactions. *Molecular Microbiology*, **37**, 1232-1247.


.

Hayward, S. and Berendsen, H.J.C. (1998) Systematic analysis of domain motions in proteins from conformational change; new results on citrate synthase and T4 lysozyme. *Proteins*, **30**, 144-154.

Higuchi, R., Krummel, B. and Saiki, R.K. (1988) A general method of *in vitro* preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. *Nucleic Acids Res.*, **16**, 7351-7367.

Jones, T.A., Zou, J.-Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for the building of protein models in electron density. *Acta Cryst.*, **A47**, 110-119.

Li, Y., Mui, S., Brown, J.H., Strand, J., Reshetnikova, L., Tobacman, L.S. and Cohen, C. (2002) The crystal structure of the C-terminal fragment of striated-muscle alpha-tropomyosin reveals a key troponin T recognition site. *Proc Natl Acad Sci U S A*, **99**, 7378-7383.

Lu, S.M. and Hodges, R.S. (2004) Defining the minimum size of a hydrophobic cluster in two-stranded alpha-helical coiled-coils: effects on protein stability. *Protein Sci*, **13**, 714-726.

Lupas, A., Van Dyke, M. and Stock, J. (1991) Predicting coiled coils from protein sequences. *Science*, **252**, 1162-1164.

Menozzi, F.D., Boucher, P.E., Riveau, G., Gantiez, C. and Locht, C. (1994) Surface-associated filamentous hemagluttinin induces autoaggregation of Bordetella pertussis. *Infect. Immun.*, **62**, 4261-4269.

Minor, Z.O.a.W. (1997) *Processing of X-ray Diffraction Data Collected in Oscillation Mode* Academic Press New York.

Mische, S.M., Manjula, B.N. and Fischetti, V.A. (1987) Relation of streptococcal M protein with human and rabbit tropomyosin: The complete amino acid sequence of human cardiac alpha tropomyosin, a highly conserved contractile protein. *Biochem Biophys Res Commun.*, **142**, 813-818.

Monera, O.D., Kay, C.M. and Hodges, R.S. (1994) Electrostatic interactions control the parallel and antiparallel orientation of alpha-helical chains in two-stranded alpha-helical coiled-coils. *Biochemistry*, **33**.

Ringdahl, U., Svensson, H.G., Kotarsky, H., Gustafsson, M., Weineisen, M. and Sjobring, U. (2000) A role for the fibrinogen-binding regions of streptococcal M proteins in phagocytosis resistance. *Mol Microbiol*, **37**, 1318-1326.

Spek, E.J., Bui, A.H., Lu, M. and Kallenbach, N.R. (1998) Surface salt bridges stabilize the GCN4 leucine zipper. *Protein Sci*, **7**, 2431-2437.

.

Strelkov, S.V. and Burkhard, P. (2002) Analysis of alpha-helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation. *J Struct Biol*, **137**, 54-64.

Terwilliger, T.C. (2000) Maximum likelihood density modification. *Acta Cryst.*, **D56**, 965-972.

Terwilliger, T.C. and Berendzen, J. (1999) Automated MAD and MIR structure solution. *Acta Cryst.*, **D55**, 849-861.

Walshaw, J. and Woolfson, D.N. (2001) Socket: a program for identifying and analysing coiled-coil motifs within protein structures. *J Mol Biol*, **307**, 1427-1450.

Wolf, E., Kim, P.S. and Berger, B. (1997) MultiCoil: a program for predicting two- and three-stranded coiled coils. *Protein Sci*, **6**, 1179-1189.

.

# IV.

# Characterization of fibrinogen binding to the B-repeats

# Abstract

Fibrinogen (Fg) is a major component of plasma, and the ability of GAS to bind Fg has been linked to bacterial survival. The B-repeats of M1 protein are responsible for binding Fg in the M1 serotype. Stoichiometric studies showed that M1 protein and the M1 fragment HB bind fibrinogen fragment D (Fg-D) in a 2:2 complex (i.e. one M1 dimer binds two molecules of Fg-D). Destabilization of the M1 oligomeric state has been partially attributed to the B-repeats, which contain non-canonical residues in coiled-coil packing positions. Mutation of the non-canonical residues to optimal residues for dimeric coiled coil formation results in a trimeric M1 species. Deletion of a single amino acid prior to the optimized B-repeats corrects the heptad register and helps to form monomeric/dimeric species. These species do not bind Fg and show that the *a* and *d* positions in the B-repeats are critical in forming specific interactions with Fg.

.

# Introduction

Survival of GAS against immune clearance is afforded, in part, by the acquisition of inhibitors that block the deposition of complement Cb3 on the streptococcal surface. M and M-like proteins, which are the major proteinaceous components on the cell surface, bind various ligands, including Fg, Ig-Fc, Factor H and C4BP. Each of these host ligands prevent activation of the alternative complement pathway. M protein from the M1 serotype (M1 protein) has been implicated in binding fibrinogen (Ringdahl et al., 2000) and IgG (Akesson et al., 1994). The binding sites for these ligands reside in the surface exposed, N-terminal half of the molecule.

The mode of binding between M1 protein and Fg has proved to be challenging to characterize. The crystal structure of an N-terminal fragment of M1 protein shows splaying of the helices in the region responsible for Fg binding (the B-repeats) suggesting this is a destabilized domain. Moreover, M1 fragments lacking the highly conserved C-terminus are in equilibrium between monomeric and dimeric states.

In order to test whether the splaying/dissociation of the B-repeats is necessary to accommodate Fg binding, the core packing positions were mutated to promote optimal coiled-coil interactions. These mutants preferentially formed trimeric species and abrogated binding to Fg and IgG. Deletion of a single amino acid towards the

.

end of the A-region optimized the heptad pattern and produced a population of monomeric/dimeric mutants which also had no functionality. All this suggests the residues in the *a* and *d* positions contribute in interactions with Fg..

# Experimental Procedures

### Cloning

The predicted *a* and *d* positions in the B-repeats (M1 protein residues 134 – 194) were mutated by custom gene synthesis (optHB+, residues 42 – 202; GenScript) to valine and leucine residues, respectively. OptHB (optHB+ residues 42 – 194) was subcloned by PCR from the optimized gene fragment into the NcoI and XhoI restriction sites of pET28b (Novagen). Deletion mutants (optHBΔW128 and optHBΔL133) were generated using the QuikChange protocol (Stratagene).

A two-step strand overlap PCR method was utilized to produce optM1 (M1 protein sequence 42 – 453 with optimized residues in the B-repeats). PCR was carried out for the mutated N-terminal fragment of M1 protein (optHB+; residues 42 – 202) with an N-terminal primer containing a NcoI restriction site and a C-terminal primer containing overlapping residues (197 – 205) to the native C-terminal M1

.

protein fragment. PCR for the native C-terminal M1 fragment (residues 197 – 453) was carried out using a N-terminal primer with complementary sequence to residues 197 – 205 of the mutated N-terminal fragment and a C-terminal primer with a XhoI restriction site. Upon strand overlap PCR, the construct optM1 (42 – 453) was cloned into the NcoI and XhoI restriction sites of pET28b (Novagen). The optM1-pET28b vector was used as the template in deletion mutations of Trp128 (optM1ΔW128) and Leu133 (optM1ΔL133) using the QuikChange protocol (Stratagene). These constructs contain artifactual sequences at the N-terminus (MV) for expression and cloning purposes and at the C-terminus (LEHHHHHH) for purification.

**Expression and purification**

Large scale expression was performed with *E. coli* strain BL21 (DE3) cells (Novagen) in 1-4 L Luria Broth at 37˚C and expression was induced with IPTG when the cells where in log phase growth ($OD_{600}$=0.65). Overexpression was permitted for 4 hours at 37ºC, then harvested by centrifugation and resuspended in 1/100 of the culture volume in lysis buffer (100 mM NaPi, pH 7.5, 250 mM NaCl, and 20 mM imidazole) with 0.5 mM PMSF. Cell lysis was accomplished by adding lysozyme (200 μg/mL) to the resuspended cells for a 20-minute, RT incubation followed by sonication. The lysate was clarified by centrifugation and the filtered supernatant was added to Ni-NTA agarose resin (Qiagen). Bound protein was eluted with lysis buffer containing 250 mM imidazole, and then the M1 protein-containing fractions were pooled and diafiltrated with 20 mM tris buffer, pH 8.0, and 100 mM

.

NaCl. Samples were concentrated to ~20 mg/mL and stored as flash-frozen aliquots at -80˚C.

Concentrations of proteins were determined spectroscopically by absorbance at 280 nm in 6 M guanidine hydrochloride buffered with 20 mM NaPi, pH 7.6, using the following extinction coefficients: 6990 $cm^{-1}M^{-1}$ (optHB, optM1, optHBΔL133, and optM1ΔL133) and 1490 $cm^{-1}M^{-1}$ (optHBΔW128 and optM1ΔW128)(Gasteiger et al., 2005).

**Circular dichroism spectroscopy**

CD spectroscopy was performed on an AVIV CD spectrometer Model 202 (Aviv Instruments) equipped with a thermoelectric temperature control. The wavelength spectra were collected between 200 and 260 nm at 4˚C, 20˚C and 37˚C using a quartz cell of pathlength 1 cm. Samples contained 1.15 μM and 1.50 μM of HB and M1, respectively, in 7.5 mM NaPi, pH 7.5 and 50 mM NaF. Spectra were collected at 1.0 nm intervals with a 5 second averaging time per data point. The reference spectrum generated from a scan of the buffer without sample was subtracted from the spectra and the data was converted to mean residue ellipticity (MRE) for comparison of helicity. The MRE is expressed in millidegrees x square centimeter per decimole.

.

Experiments to follow the change in helicity versus temperature were performed by monitoring the CD signal at 222 nm (maximal wavelength absorbance). Samples were pre-chilled on ice and added to a cuvette maintained at 4˚C with stirring. Measurements were taken in 1˚C steps from 4-80˚C allowing for a 1 minute equilibration period before collecting the data. In order to test for thermal reversibility, the signal was recorded for the sample as it was cooled at the same rate it was heated.

**Qualitative stoichiometry determination by densitometric analysis of SDS-PAGE**

Fibrinogen fragment D (Fg-D) was added in a slight molar excess (4.6 – 4.9 nmol) to 75 μg HB (4 nmol) and 190 μg M1 (4 nmol) in 20 mM tris buffer, pH 8.0, and 100 mM NaCl then incubated at room temperature for 1 hour. The complex was purified by gel filtration (Superdex 200) and 5 – 20 μl of the fraction corresponding to the first half of the peak was analyzed by non-reducing SDS-PAGE. Standards comprised of 1:1 or 1:2 molar ratios between Fg-D and the corresponding M1 construct (HB or M1) were included in amounts of 1 μg, 2.5 μg, 5 μg, and 10 μg total protein per lane. Intensities of the resolved bands were analyzed by the Kodak EDAS 120 imaging system and quantitated using the Kodak 1D software.

**Static light scattering**

Static light scattering (SLS) experiments were performed at 25˚C on either a 7.5 x 300 mm TSK G-4000 SW (TOSOHAAS) analytical size exclusion column in-

.

line with a Mini-DAWN light scattering detector (Wyatt). The system was equilibrated with 20 mM Tris, pH 8.0, and 100 mM NaCl filtered to 0.1 μm and injections ranging from 300-600 μg of constructs optHB, optHBΔW128, optHBΔL133, optM1, optM1ΔW128 and optM1ΔL133 were performed. Data acquisition and analysis was carried out using ASTRA software (v 4.90.07) as described by the manufacturer using a $d\eta/dc$ of 0.185 for each construct. The system was calibrated using bovine serum albumin (Sigma).

**Fibrinogen and IgG binding assays**

Binding assays were performed using 20 μL Ni-NTA agarose resin. The beads were pre-equilibrated with binding buffer (50 mM NaPi, pH 8.0, 150 mM NaCl and 20 mM imidazole), then incubated with 20 μg of the desired his-tagged construct (optHB, optHBΔW128, optHBΔL133, optM1, optM1ΔW128 or optM1ΔL133). The beads were washed again with binding buffer then either 10 μg of human Fg-D or IgG were introduced in a 50 μL volume to be incubated at room temperature for 30 minutes on an eppendorf rotating unit. Beads were washed three times with binding buffer supplemented with 0.1% Triton X-100 and for IgG binding assays imidazole was also increased to 45 mM to further reduce non-specific binding. Samples were eluted by adding 10 μL 500 mM imidazole, recovered by syringe and the eluant was boiled with either 2x non-reducing sample buffer for analysis by SDS-PAGE.

.

# Results

**Construct HB is significantly less structured at physiological temperatures**

The crystal structure of nHB showed nearly complete helical content for residues 61 – 190. The only exception was a small break in the B and C chains of residues 123-127. The construct crystallized as a dimer at 4˚C, so the helical content of HB was investigated at RT and physiological temperature using circular dichroism (CD). CD spectra were collected for HB and M1 from 200 nm – 260 nm. At 4˚C, each construct gave a strong helical signature (Figure 4.1A). Interestingly, the ratio of signal between 222 nm and 208 nm was noticeably different for each construct (M1 = 1.37, HB = 1.00) and became more pronounced in HB at physiological temperatures (0.88). The ratio between these wavelengths ($[\theta]_{222\ nm}/[\theta]_{208\ nm}$) can be used to determine whether the coiled coil exists in a dimeric or monomeric state. A value greater than 1.0 is characteristic for a coiled coil whereas values below 1.0 are indicative of monomeric α-helices (Litowski and Hodges, 2001). The ratio of 0.88 for HB confirms that it is predominately in the monomeric state at 37˚C.

The helical content of the constructs was calculated using the program *k2d* (Andrade et al., 1993) (Table 4.1). M1 showed a 20% decrease in α-helical content when heated from 4˚C to 37˚C, whereas HB exhibited a decrease from 98% to 61%. Predictably, the decrease in α-helical structure was almost entirely replaced by the formation of random coils.

.

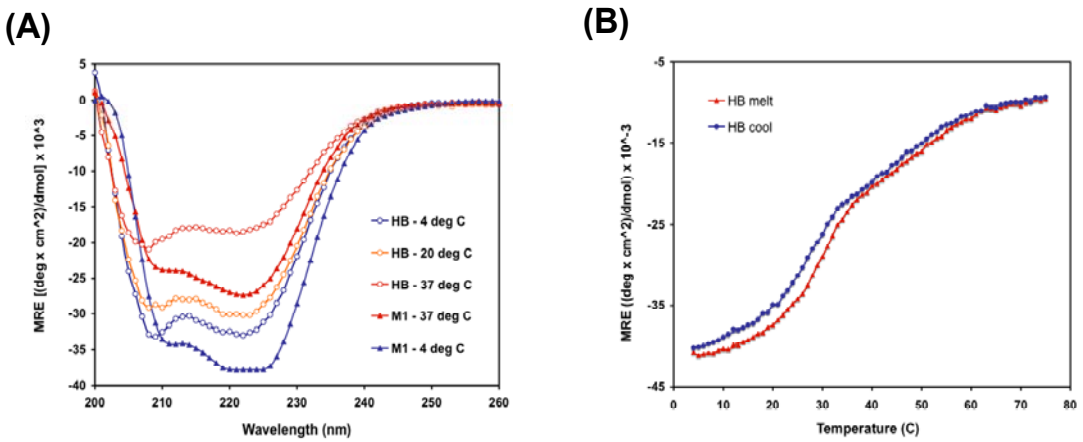



**Figure 4.1. Circular dichroism demonstrates the decreased thermal stability of HB.** (A) Wavelength scans of HB (1.11 μM; open circles) and M1 (1.67 μM; solid triangles) at 4 (blue), 20 (orange), and 37˚C (red). (B) The change in signal at 222 nm was used to monitor the helical content of HB during thermal denaturation. Biphasic unfolding gave mid-points of 27˚C and 48˚C, and $T_m$ = 35˚C. The signal was converted to mean residue ellipticity (MRE) in all spectra for comparison.

**Table 4.1. The estimation of protein secondary structure of constructs HB and M1 from circular dichroism** using *k2d*.

| | HB | | | M1 | | |
|---|---|---|---|---|---|---|
| **Temp.** | **alpha** | **beta** | **random** | **alpha** | **beta** | **random** |
| **4˚C** | 0.98 | 0.00 | 0.02 | 1.00 | 0.00 | 0.00 |
| **20˚C** | 0.91 | 0.00 | 0.09 | n/a | n/a | n/a |
| **37˚C** | 0.61 | 0.07 | 0.32 | 0.80 | 0.00 | 0.20 |
| | **-0.37** | **+0.07** | **+0.30** | **-0.20** | **+0.00** | **+0.20** |

The structural stability of HB was further investigated using a thermal denaturation melt. Like all the other M1 fragments (Chaper 3, Figure 3.4), HB

follows a two-state unfolding path and thermal denaturation was shown to be completely reversible (Figure 4.1B). The midpoint of the first unfolding state was ~27˚C and ~48˚C for the second. The transition between these states was estimated to occur near 35˚C. M1 protein demonstrated greater thermal stability in the second transition (~55˚C) These values continued to support that HB is destabilized at physiological temperature.

These data suggest that results for experiments of HB performed at RT may not apply to those at physiological temperatures. However, it has been shown that Fg binding to HB is not inhibited at physiological temperature (data not shown) and occurs in a temperature-independent manner.

**M1 protein binds fibrinogen fragment D in a 2:2 stoichiometry**

The binding stoichiometry between M1 protein and Fg was determined using two techniques: densitometric analysis of purified M1:Fg-D complexes and static light scattering. Fibrinogen fragment D (Fg-D) was incubated in a molar excess with either the M1 or HB construct at RT, then purified by gel filtration. Fractions representing the complex were examined by SDS-PAGE along with dilutions of standards corresponding to 1:1 or 2:1 molar ratios of Fg-D to the particular M1 construct (Figure 4.2A and B). The imaging software Kodak 1D was used to quantify the band intensities in the gel (Table 4.2). Comparison of the ratio of intensities between Fg-D and HB in the purified complex (1.33±0.08) correlated very well to the

.

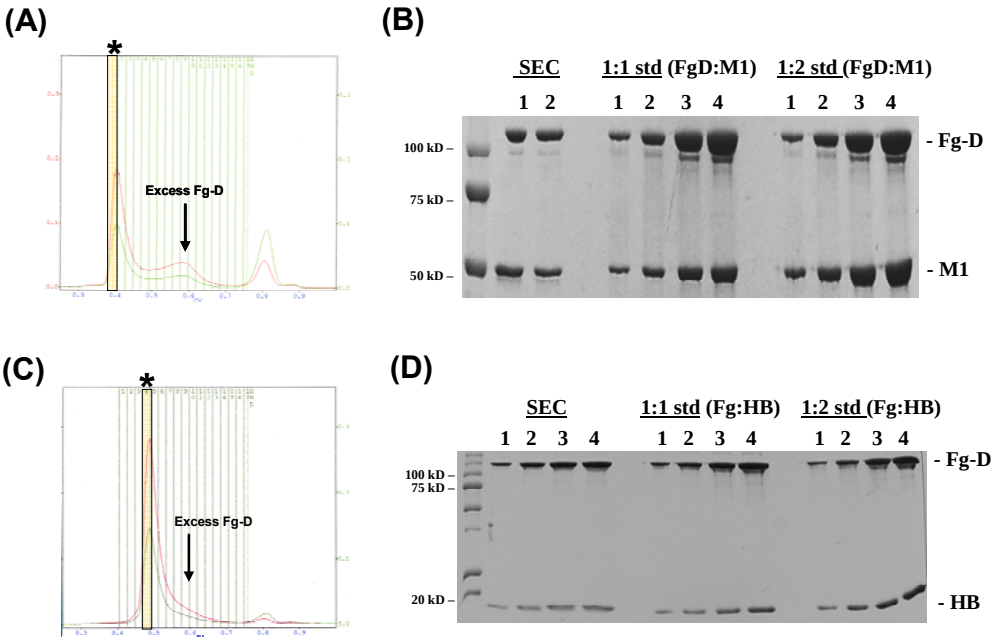



**Figure 4.2. Densitometric analysis of purified Fg-D:HB complex and standards by SDS-PAGE.** (A) M1 and Fg-D were incubated in a 1.0:1.2 molar ratio, then purified by gel filtration. The highlighted lane (*) was collected and (B) resolved by 10.0% SDS-PAGE along with dilutions of Fg-D:M1 standards in 1:1 or 1:2 molar ratio. The experiment was repeated using HB instead of M1. (C) HB and Fg-D were incubated in a 1.0:1.2 molar ratio, then purified by gel filtration. The highlighted lane (*) was collected and (D) resolved by 12.5% SDS-PAGE along with dilutions of Fg-D:HB standards in 1:1 or 1:2 molar ratio.

**Table 4.2. Summary of results from densitometric analysis for Figure 4.2.**

| | HB | | | | M1 | | |
|---|---|---|---|---|---|---|---|
| | SEC | 1:1 std | 1:2 std | | SEC | 1:1 std | 1:2 std |
| 1 | 1.27 | 1.33 | 0.85 | 1 | 3.33 | 3.1 | 1.22 |
| 2 | 1.39 | 1.27 | 0.87 | 2 | 3.15 | 2.91 | 1.16 |
| 3 | n/a | 1.39 | 0.91 | 3 | 2.83 | 2.83 | 1.23 |
| 4 | n/a | 1.42 | 0.94 | 4 | 2.77 | 2.89 | 1.22 |
| avg | 1.33±0.08 | 1.35±0.06 | 0.89±0.04 | avg | 3.02±0.26 | 2.93±0.12 | 1.21±0.03 |

standards representing a 1:1 stoichiometry (1.35±0.06). A 1:1 stoichiometry was also observed with the purified complex of Fg-D and M1 (complex = 3.02±0.26 and 1:1 standards = 2.93±0.12). Because M1 protein forms a stable dimer it was concluded that the absolute binding stoichiometry was 2:2. That is one M1 dimer binds two Fg-D molecules. Due to the dynamic oligomeric state of HB the absolute stoichiometry could not be deduced from this technique.

Static light scattering (SLS) of the Fg-D:HB complex was performed to determine the absolute stoichiometry between binding partners. Previous SLS studies had shown HB to exist in a dynamic equilibrium between monomeric and dimeric states (Chapter 3, Table 3.2). Analysis of the complex by SLS (Figure 4.3) gave a molar mass of 205.4 kD (0.6% error). This corresponded excellently to a 2:2 stoichiometry where a 210 kD product would result from the binding of two Fg-D molecules (86 kD each) to dimeric HB (19 kD per monomer). The Fg-D:M1 complex eluted in the void volume and co-eluted with protein aggregates, so the precise molecular mass of the complex could not be calculated.

Interestingly, binding of Fg-D stabilized the dimeric state in HB. This is significant because HB has been characterized as a destabilized fragment. Stabilization might occur a single molecule of Fg-D contacts the two HB chains and ‘freezes’ the dimeric state. Binding of human serum albumin to the C-repeats of M1

.

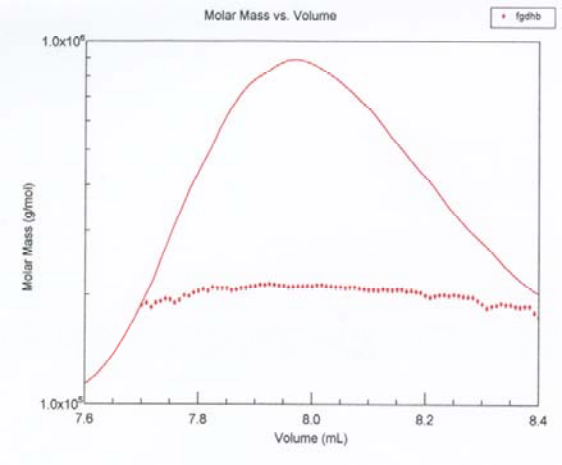



**Figure 4.3. Static light scattering shows 2:2 binding stoichiometry between HB and Fg-D.** Fibrinogen fragment D was incubated in 1.2x molar excess to M1 protein fragment HB (residues 42 – 194) and injected onto a gel filtration column for analysis by SLS. The solid red line represents the UV absorbance and the dotted line is the molar mass calculated at 0.5 second intervals.

protein (Gubbe et al., 1997) have also been shown by CD to promote increased structural stability.

## Optimization of the core residues in the B-repeats to induce coiled-coil packing

The role of oligomerization in the B-repeats for Fg binding was investigated by optimizing the core-packing residues to promote dimerization (Figure 4.4). Identical heptad registers were predicted by Coils (Lupas et al., 1991), PairCoil (Berger et al., 1995) and Multicoil (Wolf et al., 1997) for the entire B-repeats (residues 134-194) and they corresponded to the observed *a* and *d* positions in the crystal structure. Valines were mutated into *a* positions, which are preferential to apolar β–branched sidechains (Zhu et al., 1993), and leucines were mutated into the *d* positions, mimicking leucine zippers (Landschultz et al., 1988). The PairCoils

.

algorithm was found to be better than the Coils algorithm in predicting the coiled-coil probability. Only PairCoils predicted the B-repeats to have low probability to form a coiled-coil packing. MultiCoil, based on the same algorithm as in PairCoil, is useful because it can differentiate between trimeric and dimeric coiled-coils, and was used to confirm the probability that the Val/Leu optimized construct would form a dimeric M1 species (Figure 4.5A, HB; B, M1; C, optHB; D, optM1). A dramatic increase in coiled-coil probability from ~60% to 100% was predicted for the entire length of the optimized B-repeats.

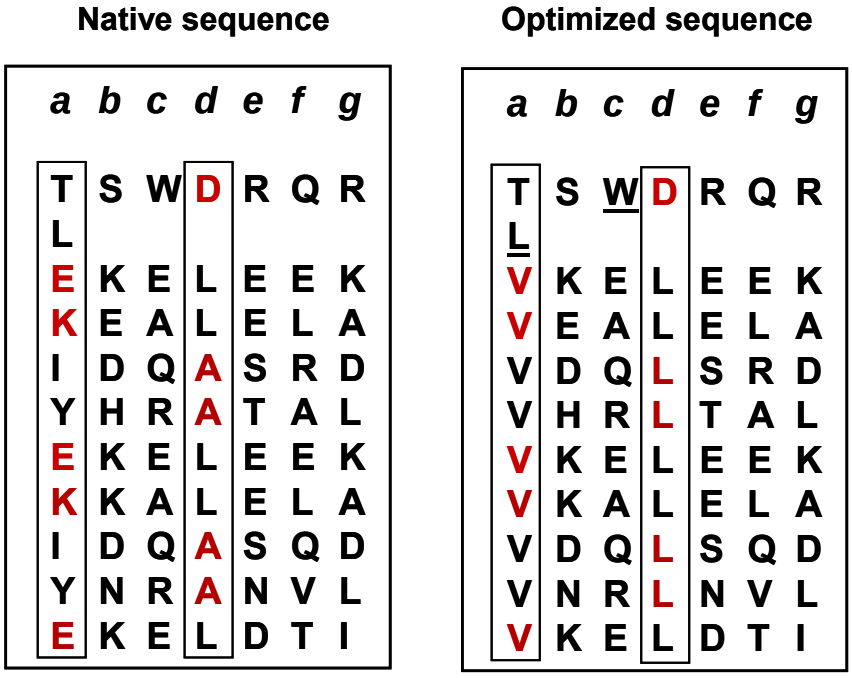



**Figure 4.4. Optimization of predicted core packing residues in the B-repeats.** (A) Predicted heptad register for residues 126 - 196. Residues colored in red are non-canonical. (B) Amino acid optimization in the mutant. Valines were mutated in all *a* positions and leucines to the *d* positions. Underlined residues (Trp128 and Leu133) were targeted in deletion mutants to improve the heptad register.

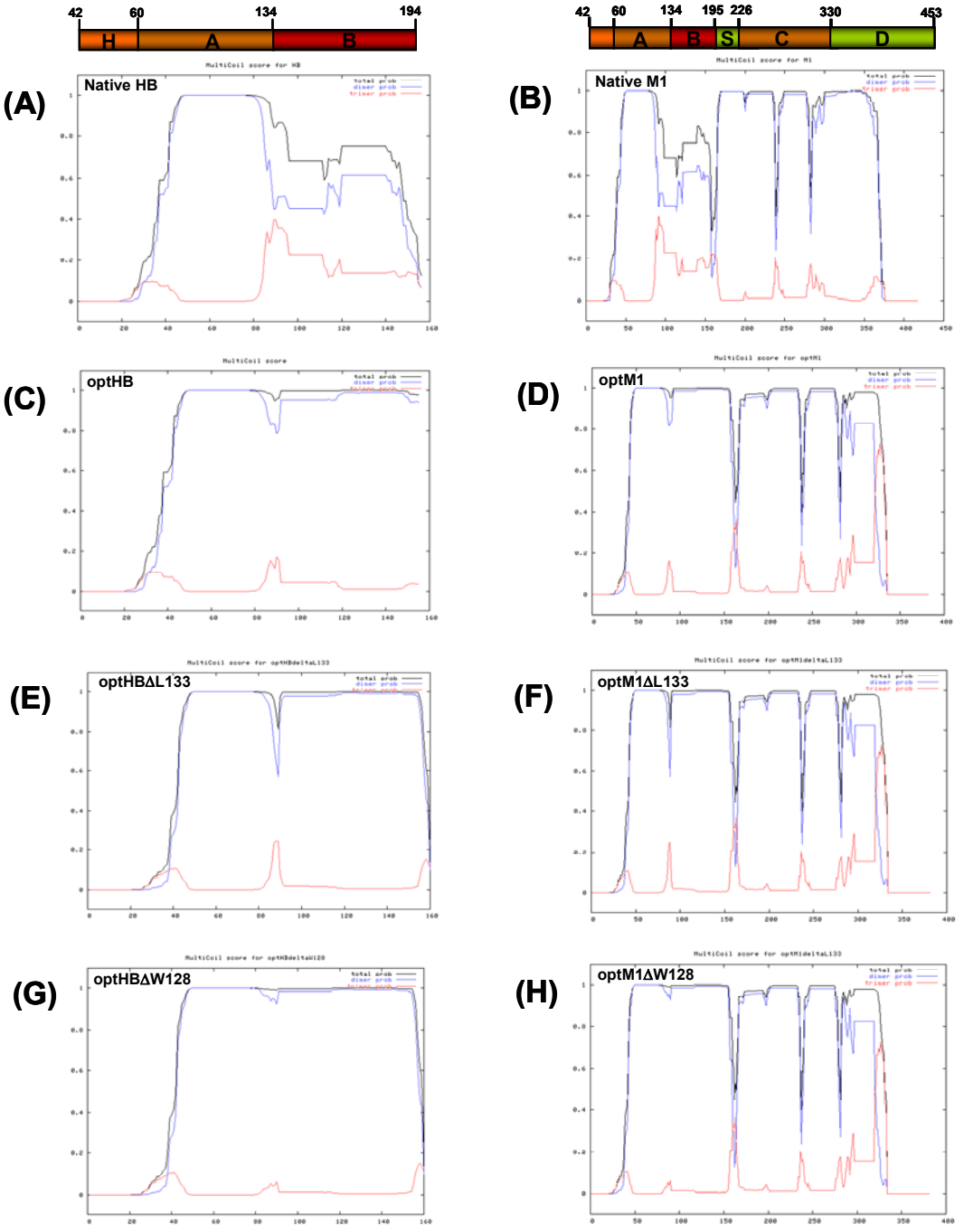



**Figure 4.5. MultiCoil prediction of coiled-coil probability in each mutant.** Schematic diagram of HB and M1 are provided at top for reference. The y-axis represents probability with 1.0 being 100%. Predictions for (A) native HB, (B) native M1, (C) optHB, (D) optM1, (E) optHBΔL133, (F) optM1 Δ L133, (G) optHB Δ W128, and (H) optM1 Δ W128.

A small, but sharp, break in the coiled-coil probability was evident at the end of the A-region and corresponded to the asymmetric packing (beginning at Trp128) observed in the crystal structure. The crystal structure also showed that the heptad register was disrupted by a single amino acid insert between residues 128 – 133.. In studies of inserts between heptads it was found that the insertion of one amino acid is the least tolerated and most destabilizing (Hicks et al., 2002). Deletion mutants were designed to eliminate either Leu133 or Trp128 to further optimize this region (Figure 4.5E, optHBΔL133; F, optM1ΔL133; G, optHBΔW128; H, optM1ΔW128).

**The optimized constructs preferentially form trimers**

Expression levels and purity of the mutants were consistent with that of the native constructs (Figure 4.6A). Analysis of the constructs by gel filtration showed distinct changes in the elution times. OptHB was a single species that eluted before the HB construct, suggesting it had a greater mass (Figure 4.6B; optHB, green trace; HB, red trace). Interestingly, the optHB deletion mutants optHBΔL133 and optHBΔW128 ran as a doublet (Figure 4.6B, blue trace). The first peak eluted at the same time as optHB while the elution time of the second peak corresponded with native HB. However, purification of the optimized M1 constructs indicated an unexpected change in the oligomeric state.

The optM1 constructs all eluted before M1 suggesting that these constructs formed larger species (Figure 4.6C; optM1, blue; optM1ΔL133/optM1ΔW128, green;

.

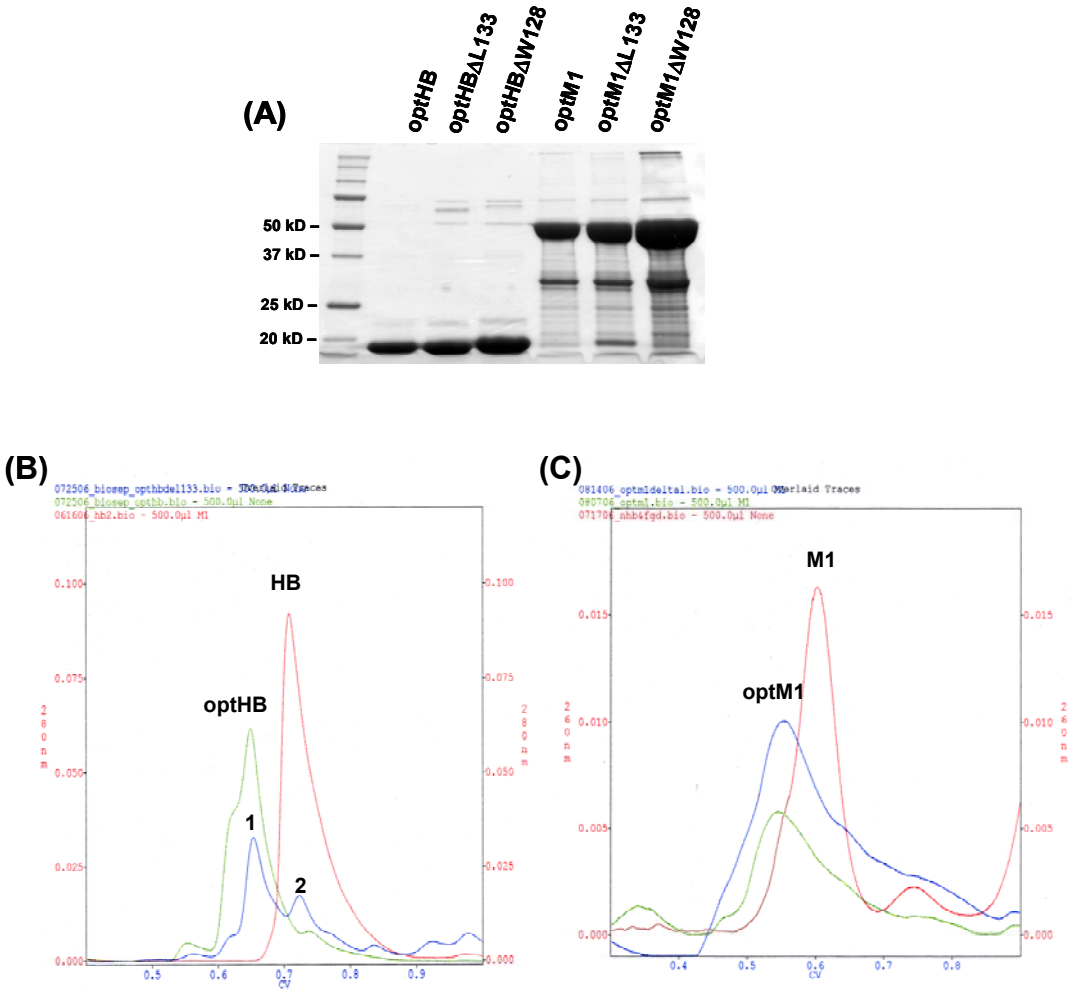



**Figure 4.6. Optimized constructs of HB and M1 exhibit increased mass by gel filtration.** (A) Ni-NTA purified constructs visualized by 10% SDS-PAGE. The optM1 constructs showed degradation similar to the mature M1 construct, including co-purification of a 30 kD proteolytic fragment. (B) The optimized HB construct (optHB; green) demonstrated a significant shift compared to the native HB construct (red). Optimized constructs with a single amino acid deletion (optHBΔL133 and optHBΔW128) exhibited a doublet peak (blue). The dominant peak (1) had the same elution time as the optHB construct while the elution time of the secondary peak (2) correlated with native HB. (C) All three optimized M1 constructs (optM1, blue; optM1ΔL133, green; optM1ΔW128, not shown) eluted at identical times and demonstrated a greater mass than native M1 (red).

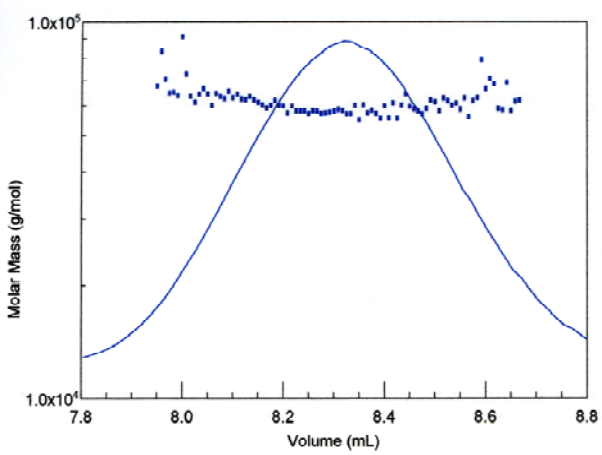



**Figure 4.7. Static light scattering proves optHB is trimeric.**

The solid blue line represents the UV absorbance of optHB and the dotted line is the molar mass calculated at 0.5 second intervals.

| Construct | Expected kD* | Calculated kD** |
|---|---|---|
| optHB | 38 | 60.5 (7%) |
| optHBΔW128 | 38 | 61 (9%), 31.5 (10%) |
| optHBΔL133 | 38 | 61 (11%), 29 (10%) |
| optM1 | 96 | 135 (8%) |
| optM1ΔW128 | 96 | 128.1 (10%) |
| optM1ΔL133 | 96 | 131 (12%) |

* mass calculated for a dimer
** The experimental molar mass is given with the percent error in parentheses

protein. Therefore, static light scattering was employed to provide a more reliable measure of mass.

Analysis of optHB (Figure 4.7) and the three optM1 constructs gave conclusive evidence that these constructs form trimers (Table 4.3). The deletion

.

mutants were particularly interesting because they separated into two species by gel filtration. The largest peak corresponded to a trimer whereas the secondary peak behaved similar to HB and yielded a mass intermediate to monomeric and dimeric optHB (Table 4.3). The latter peak behaved like native HB, which was found to form a dynamic equilibrium between monomeric and dimeric states. This was surprising because the probability for dimeric coiled-coil formation was nearly 100% and the mutations were expected to stabilize the dimeric state.

**Optimization of the core residues in the B-repeats abrogates functionality**

Fg-D and IgG-binding to the optimized constructs were assayed to determine if the oligomeric state affected function. As in previous binding assays, the M1 construct was bound to Ni-NTA resin then incubated with the ligand. The resin was washed then eluted for examination by SDS-PAGE. All six constructs (optHB, optHBΔW128, optHB, ΔL133 optM1, optM1ΔW128, and optM1ΔL133) showed no binding to Fg-D, but the positive control, HB, showed strong binding (Figure 4.8A).

An important experiment was to assay the secondary peaks from gel filtration of optHBΔW128 and optHBΔL133. These peaks corresponded to dynamic dimers, which was similar behavior exhibited by HB. These peaks were purified by gel filtration and assayed for Fg-D binding. As before, no binding was observed (data not shown). In spite of the negative result, this experiment was informative. If these species remained as monomers/dimers as identified by static light scattering, then this

.

result indicated that the mutated core residues are in part responsible for interacting with Fg-D.

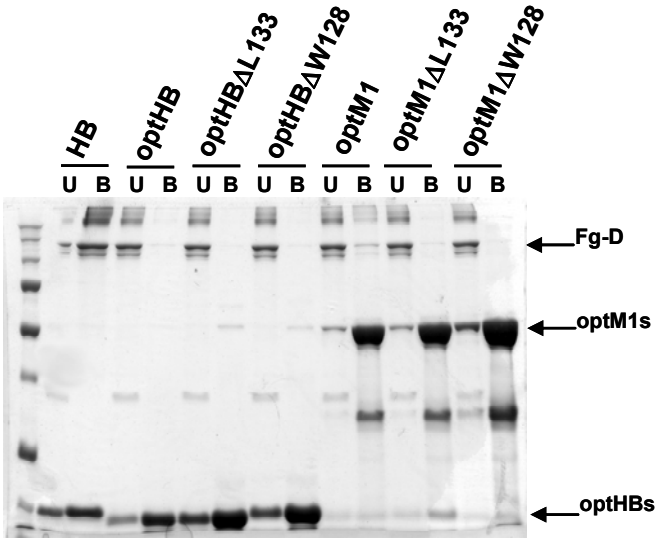



**Figure 4.8. Optimized constructs result in loss-of-function.** Human fibrinogen fragment D (Fg-D) was incubated with Ni-NTA resin pre-coated with specific M1 protein constructs, washed, eluted and visualized by non-reducing SDS-PAGE. Optimized HB and M1 constructs showed no interaction with Fg-D, whereas HB showed strong binding. Lanes correspond to unbound/flow-through (U) and bound/eluted protein (B).

# Discussion

The majority of core packing positions *a* and *d* in the Fg-binding B-repeats are occupied by non-canonical residues. In M1 fragments these residues destabilize the B-repeats and also the entire coiled-coil structure. In spite of a destabilized oligomeric state and partial thermal denaturation at physiological temperatures, this fragment retains a 2:2 binding stoichiometry to fibrinogen fragment D (Fg-D). The stabilization of the dimeric state in HB by Fg-D binding was intriguing because only intact, full length M1 protein has been shown to form stable dimeric states. The stoichiometry of binding also gave insights to the molecular mechanism of vascular leakage in patients with streptococcal toxic shock syndrome (STSS).

The formation of an M1 fragment in complex with Fg *in vivo* has been reported in patients with STSS (Herwald et al., 2004). Molecular pathogenesis of this disease is initiated by the release of M1 protein from the bacterial surface by a host protease. A 29 kD, N-terminal fragment (similar to construct HC) binds Fg and creates a precipitate that ultimately triggers vascular leakage. The molecular mechanism of the precipitation can be explained by the stoichiometry of binding between M1 protein and Fg. Since a 2:2 complex is formed (i.e. dimeric M1 protein binds two Fg molecules) and each Fg molecule has two fragment D domains, then each Fg is able to bind two M1 dimers. The stoichiometry and dual binding sites for M1 protein in Fg leads to polymerization of M1 protein and Fg. The stabilization of

.

the M1 fragments into dimers from binding Fg is an important event and facilitates the polymerization.

The increase stability of HB caused by Fg-D binding is most likely a consequence of the ligand making contacts with both helices of the B-repeats. Fg binding acts to clamp the B-repeats and hence stabilizes the dimer. In the absence of supporting evidence, the molecular mechanism of Fg binding to the B-repeats remained unclear. It was unknown if the B-repeats remained splayed in the intact M1 protein and, if so, was the splaying necessary to accommodate fibrinogen binding? This problem was approached by redesigning the core packing residues in the B-repeats to assure formation of dimeric coiled-coil packing. Although MultiCoil predicted that the mutant B-repeats would form a dimeric coiled coil, these constructs were instead found to form trimeric species. The trimeric species was unexpected because MultiCoil calculated 0% probability that this region would form a trimeric coiled coil. It was deduced that the most likely species was an antiparallel, three-stranded coiled coil because the MultiCoil algorithm was not designed to account for antiparallel interactions.

The propensity for a trimeric state in the mutants was interesting and suggested that canonical residues in the coiled-coil packing positions in the B-repeats altered the coiled-coil oligomeric state for the entire M1 molecule. Therefore, M1 protein requires non-canonical residues in core-packing positions to form dimeric species. Further manipulation of the optimized constructs included elimination of one

.

amino acid (either Trp128 or Leu133) in positions where a natural single amino acid inserts disrupt the M1 heptad register. In the optimized HB construct (optHB), deletions of either of these residues led to the formation of two species: a trimer species and a second species in dynamic equilibrium between the monomeric and dimeric state. While it was unclear to why the latter species did not form a stable dimer, it did present an opportunity to test Fg binding in a mutant construct that was not trimeric. Surprisingly this species failed to bind Fg-D, implicating that the *a* and *d* positions of the B-repeats are critical for Fg binding. The involvement of residues in the *a* and *d* positions supported the notion that the B-repeats must adopt an open or 'unzipped' state in the intact M1 molecule.

It is hypothesized that the M1 coiled coil conserves non-canonical core packing residues in the B-repeats to disrupt coiled-coil packing. The disruption leads to dissociation of the helices in the B-repeats, which is necessary to expose residues in the *a* and *d* position for Fg binding. This is a fascinating mechanism because this suggests that function is dependent on localized destabilization of the M1 coiled coil.

.

# References

Akesson, P., Schmidt, K.H., Cooney, J. and Bjorck, L. (1994) M1-Protein and Protein-H - Iggfc-Binding and Albumin-Binding Streptococcal Surface-Proteins Encoded by Adjacent Genes. *Biochemical Journal*, **300**, 877-886.

Andrade, M.A., Chacon, P., Merelo, J.J. and Moran, F. (1993) Evaluation of secondary structure of proteins from UV circular dichroism using an unsupervised learning neural network. *Prot Eng.*, **6**, 383-390.

Berger, B., Wilson, D.B., Wolf, E., Tonchev, T., Milla, M. and Kim, P.S. (1995) Predicting coiled coils by use of pairwise residue correlations. *Proc Natl Acad Sci U S A*, **92**, 8259-8263.

Brown, J.H., Kim, K.H., Jun, G., Greenfield, N.J., Dominguez, R., Volkmann, N., Hitchcock-DeGregori, S.E. and Cohen, C. (2001) Deciphering the design of the tropomyosin molecule. *Proc Natl Acad Sci U S A*, **98**, 8496-8501.

Cunningham, M.W., Antone, S.M., Smart, M., Liu, R. and Kosanke, S. (1997) Molecular analysis of human cardiac myosin-cross-reactive B- and T-cell epitiopes of the group A Streptococcal M5 protein. *Infect. Immun.*, **65**, 3913-3923.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, A.P. and Bairoch, A. (2005) *Protein Identification and Analysis Tools on the ExPASy Server*. Humana Press.

Gubbe, K., Misselwitz, R., Welfe, K., Werner, R., Schmidt, K.H. and Welfe, H. (1997) C repeats of the streptococcal M1 protein achieve the human serum albumin binding ability by flanking regions which stabilize the coiled-coil conformation. *Biochemistry*, **36**, 8107-8113.

Herwald, H., Cramer, H., Morgelin, M., Russell, W., Sollenberg, U., Norrby-Teglund, A., Flodgaard, H., Lindbom, L. and Bjorck, L. (2004) M protein, a classical bacterial virulence determinant, forms complexes with fibrinogen that induce vascular leakage. *Cell*, **116**, 367-379.

.

Hicks, M.R., Walshaw, J. and Woolfson, D.N. (2002) Investigating the tolerance of coiled-coil peptides to nonheptad sequence inserts. *J. Structure Bio*, **137**, 73-81.

Landschultz, W.H., Johnson, P.F. and McKnight, S.L. (1988) The leucine zipper: a hypothetical structure common to a new class of DNA-binding proteins. *Science*, **240**, 1759-1764.

Li, Y., Brown, J.H., Reshetnikova, L., Blazsek, A., Farkas, L., Nyitray, L. and Cohen, C. (2003) Visualization of an unstable coiled coil from the scallop myosin rod. *Nature*, **424**, 341-345.

Litowski, J.R. and Hodges, R.S. (2001) Designing heterodimeric two-stranded alpha-helical coiled-coils: the effect of chain length on protein folding, stability and specificity. *J. Peptide Res.*, **58**, 477-492.

Lupas, A., Van Dyke, M. and Stock, J. (1991) Predicting coiled coils from protein sequences. *Science*, **252**, 1162-1164.

Ringdahl, U., Svensson, H.G., Kotarsky, H., Gustafsson, M., Weineisen, M. and Sjobring, U. (2000) A role for the fibrinogen-binding regions of streptococcal M proteins in phagocytosis resistance. *Mol Microbiol*, **37**, 1318-1326.

Spraggon, G., Everse, S.J. and Doolittle, R.F. (1997) Crystal structures of fragment D and human fibrinogen and its crosslinked counterpart from fibrin. *Nature*, **389**, 455-462.

Wolf, E., Kim, P.S. and Berger, B. (1997) MultiCoil: a program for predicting two- and three-stranded coiled coils. *Protein Sci*, **6**, 1179-1189.

Zhu, B.Y., Zhou, N.E., Kay, C.M. and Hodges, R.S. (1993) Packing and hydrophobicity effects on protein folding and stability: effects of beta-branched amino-acids, valine and isoleucine, on the formation and stability of 2-stranded alpha-helical coiled coils leucine zippers. *Protein Sci*, **2**, 383-394.

.

# V.

# Future Directions

# Discussion

The findings presented in this thesis raise new questions pertaining to the function of the M1 protein from GAS. These questions are centered on the molecular mechanism of rheumatic fever, and elucidation of the precise M1 residues of the B-repeats involved in Fg binding using X-ray crystallographic and computational techniques.

The crystal structure of an N-terminal fragment of the M1 protein has been described in Chapter III. It shows that the M1 protein has features that are shared with myosin and tropomyosin, two host, coiled-coil proteins that have both been shown to be cross-reactive with antibodies directed to GAS M proteins (Khanna et al., 1997; Quinn et al., 1998). A cluster of alanines at residues 80, 84 and 91 of M1 protein are similar to alanine clusters observed in the crystal structure of tropomyosin (Brown et al., 2001). Additionally, M1 protein forms only one salt bridge between the *e* and *g* positions between opposing coiled-coil helices. The absence of e-g' interactions is atypical of coiled coils; however, the crystal structure of myosin shows that it also does not form salt bridges between the e-g' positions (Li et al., 2003). It is unknown if the described structural similarities between the M1 protein and tropomyosin/myosin lead to a common epitope resulting in rheumatic fever.

A direct means for testing the molecular mimicry hypothesis would be to directly visualize interactions between M1 and cross-reactive antibodies. To identify such antibodies, the M1 fragment that has been crystallized (HB) would first need to be screened against a panel of antibodies that are also known to be cross-reactive against tropomyosin or myosin. Similar studies have already been performed with the M5 protein and human cardiac myosin (Cunningham et al., 1997). Antibodies identified to be cross-reactive to the M1 fragment and tropomyosin/myosin may then be selectively proteolyzed to enable purification of Fab fragments, which then can be introduced into co-crystallization trials with the HB fragment. The crystal structure of such a complex would give atomic resolution of the epitope and provide insight to the mechanism of rheumatic fever.

The second crystallization experiment to be pursued in the future is directed at showing that the *a* and *d* positions of the B-repeats participate in Fg binding. Experiments described in Chapter IV show that mutagenesis of the *a* and *d* positions in the B-repeats to canonical coiled-coil residues (valine and leucine, respectively) abrogates Fg binding. It was later found that the mutant B-repeats formed a trimeric coiled coil. However, deletion of Trp128 or Leu133 in the mutant HB constructs lead to the formation of an unstable dimeric species. The unstable dimeric species was determined to behave similarly to the native HB construct and existed in a dynamic equilibrium between monomeric and dimeric states. Interestingly, this species also did not bind Fg and suggested that the *a* and *d* positions are important in Fg binding. Co-crystallization of the native HB construct with fibrinogen fragment D (Fg-D)

.

would provide the best method to identify the precise M1 residues involved in Fg interactions. Additionally, the precise M1 binding site in Fg-D would be defined from a crystal structure. M1 residues identified to form interactions with Fg would then be individually mutated and assayed in Fg binding to characterize their importance in Fg binding and verify the crystal structure.

Computational modeling provides an indirect method that may be an alternative technique if co-crystallization techniques are not successful for the M1 protein/Fg-D complex. The crystal structures for the M1 B-repeats and Fg-D (Spraggon et al., 1997) have been solved and could be input to a protein docking algorithm such as Rosetta++. Due to the numerous docking solutions output by this method, verification of the correct complex would be highly dependent on site-directed mutagenesis of M1 residues predicted to interact with Fg.

The goal of the proposed research is to continue the elucidation of important interactions of the M1 protein with host ligands at atomic resolution. Identification of critical M1 residues in pathogen-host interactions may, in turn, be applicable to other GAS M proteins. Thus, creating a greater understanding of GAS pathogenesis.

.

# References

Brown, J.H., Kim, K.H., Jun, G., Greenfield, N.J., Dominguez, R., Volkmann, N., Hitchcock-DeGregori, S.E. and Cohen, C. (2001) Deciphering the design of the tropomyosin molecule. *Proc Natl Acad Sci U S A*, **98**, 8496-8501.

Cunningham, M.W., Antone, S.M., Smart, M., Liu, R. and Kosanke, S. (1997) Molecular analysis of human cardiac myosin-cross-reactive B- and T-cell epitiopes of the group A Streptococcal M5 protein. *Infect. Immun.*, **65**, 3913-3923.

Khanna, A.K., Nomura, Y., Fischetti, V.A. and Zabriskie, J.B. (1997) Antibodies in the sera of acute rheumatic fever patients bind to human cardiac tropomyosin. *J. Autoimmun.*, **10**, 99-106.

Li, Y., Brown, J.H., Reshetnikova, L., Blazsek, A., Farkas, L., Nyitray, L. and Cohen, C. (2003) Visualization of an unstable coiled coil from the scallop myosin rod. *Nature*, **424**, 341-345.

Quinn, A., Ward, K., Fischetti, V.A., Hemric, M. and Cunningham, M.W. (1998) Immunological relationship between the class I epitope of streptococcal M protein and myosin. *Infect. Immun.*, **66**, 4418-4424.

Spraggon, G., Everse, S.J. and Doolittle, R.F. (1997) Crystal structures of fragment D and human fibrinogen and its crosslinked counterpart from fibrin. *Nature*, **389**, 455-462.

.

# VI.

# Appendix A.

# Summary of TRAP expression and characterization

# Abstract

*Plasmodium* surface protein TRAP is an integral component of the invasion machinery utilized by malaria sporozoites to promote entry into hepatocytes. Identifying the hepatocyte ligands recognized by TRAP is critical step for understanding molecular mechanisms of adhesion and invasion. In an attempt to elucidate the hepatocyte receptor(s) a biochemical approach dependent on producing functional TRAP constructs was undertaken. Expression and purification of constructs comprising critical receptor-binding regions, A-domain and TSR, of the ecto-domain produced unstable protein that readily aggregated. Stabilizing factors such as small molecule additives (i.e. glycerol, EDTA, divalent cations), binding ligands (heparin), detergents, fusion proteins and alternative expression systems failed to produce stable, soluble protein. The lack of functional TRAP made it unfeasible to pursue biochemical determination of the TRAP receptor(s) on hepatocytes.

# Introduction

*Plasmodium*, the causative agent of malaria, belongs to the phyla Apicomplexa, which are characterized by their distinctly polarized and elongated extracellular stages called zoites and the specialized secretory vesicles they contain (micronemes and rhoptries) at the apical end. These vesicles, which secrete their contents at the anterior tip of the zoite, play a key role in attachment and invasion of target cells. They also help to mediate a unique behavior of Apicomplexan zoites called gliding motility, which is a substrate-dependent type of locomotion that occurs without change in the zoite shape.

*Plasmodium* sporozoites are first formed within oocysts in the *Anopheles* mosquito midgut. After being released by mature oocysts into the body cavity of the insect, the sickle-shaped sporozoites migrate to and invade the salivary glands, followed by injection into a susceptible vertebrate host during a subsequent blood meal. Sporozoites rapidly invade hepatocytes, where parasite development generates thousands of merozoites. In turn, these invade and multiply within erythrocytes, establishing malaria infection.

Amazingly as few as two to ten *Plasmodium* sporozoites are necessary to establish malaria infection within the host (Ungureanu et al., 1977) underscoring the incredible efficiency of invasion by sporozoites into hepatocytes. It has been

.

proposed that specific interactions between adhesive parasite proteins and hepatocyte receptor proteins mediate invasion because sporozoites invade hepatocytes within 1 hour after injection (Verhave and Meis, 1984) and hepatocytes are the only cell type in which sporozoites can develop into exo-erythocytic forms (EEF). However, the specific mechanism of *Plasmodium* sporozoite navigation and invasion into hepatocytes has remained elusive.

Two sporozoite proteins, circumsporozoite (CS) and thrombospondin related anonymous protein (TRAP), play integral roles in hepatocyte adhesion and invasion. CS protein is the major sporozoite surface protein and likely binds a glycosylphosphatidylinositol (GPI)-anchored protein (Cerami et al., 1992)and contains a thrombospondin type I repeat (TSR; (Lawler and Hynes, 1986). In brief, TSR domains act as adhesive modules in numerous proteins (Adams and Tucker, 2000) primarily binding heparan sulfates (HS), which are chains of disaccharide repeats covalently linked to protein (HS proteoglycans, HSPGs) and are ubiquitously present on mammalian cells(Esko and Selleck, 2002). Recombinant CS protein used *in situ* and *in vivo* binds only to the liver and clearly demonstrates a high degree of cell type specificity (Cerami et al., 1992).Specifically CS protein binds with high affinity to regions of the HS chains that are highly sulfated (Pinzon-Ortiz et al., 2001). This is particularly significant because liver HSPG are more highly sulfated than HS found on other cell types (Lyon et al., 1994) thus giving insight into sporozoite specificity. However, while CS is important for sporozoite attachment it is not essential for sporozoite internalization (Pinzon-Ortiz et al., 2001). Conversely,

.

TRAP has been shown to be essential for cell invasion, *in vivo* infectivity and sporozoite gliding (Sultan et al., 1997).

TRAP is a transmembrane protein stored in micronemes, becomes surface-exposed at the sporozoite anterior tip, and is released onto the substrate during gliding locomotion (Kappe et al., 1999). The cytosolic tail is believed to bind a cytoplasmic motor system (Kappe et al., 1999) while the extracellular portion of TRAP contains two adhesive domains, an A-domain of von Willebrand factor (Girma et al., 1987) and a single TSR, followed by a repeat region of variable length and sequence depending on the plasmodial species (Templeton and Kaslow, 1997). Combinations of A-domain(s) and TSRs are found in micronemal transmembrane proteins described in many apicomplexan genera. Moreover, there is substantial evidence that these proteins have similar functions to TRAP (Carruthers and Sibley, 1999; Kappe et al., 1999; Templeton et al., 2000; Yuda et al., 1999).

The A-domain of TRAP is homologous with the A1 domain of the von Willebrand factor (vWF) family of proteins and shows similarities to the I domain of integrins (Lawler and Hynes, 1986). The structural homology suggests that functional homology may also follow. Thus the TRAP A-domain is expected to bind cell-surface or cell-matrix molecules (Ueda et al., 1994; Wengelnik et al., 1999). A distinct feature of A-domains is the metal-ion-dependent adhesion site (MIDAS) motif. The primary sequence of the MIDAS motif is Asp-X-Ser-X-Ser (DXSXS) and

.

as the name suggests plays a role in bivalent cation binding and receptor recognition (Lee et al., 1995; Michishita et al., 1993).

A cooperative role between the TRAP A-domain and TSR has been suggested based on mutagenesis experiments that introduced loss-of-function mutations in the A-domain and TSR. Mutations directed to the MIDAS motif or to key basic residues in the TSR resulted in a significant decrease of sporozoite infectivity *in vivo* (Matuschewski et al., 2002).

Both the A-domain and TSR play significant roles for invasion of *Plasmodium* sporozoites into hepatocytes. The work presented here is an investigation aimed at identifying the hepatocyte receptor(s) for the A-domain and TSR of TRAP using biochemical methods.

.

# Experimental Procedures

**Cloning and expression of TRAP constructs in *E. coli*.** All constructs were generated from a DNA vector harboring the gene encoding *Plasmodium berghei* TRAP (a generous gift from Dr. Victor Nussenzweig). Three general construct designs were employed: the A-domain and TSR (ATSR: amino acid residues 25-281), the A-domain only (A: residues 25-230), and the A-domain lacking $Cys_{33}$ and $Cys_{229}$ (Afree: residues 40-226). Constructs were cloned into a T7-polymerase based expression vector, pET15b (Novagen), enabling the addition of a N-terminal hexahistidine tag accompanied by a thrombin cleavage site. Additionally, these same constructs were also expressed as MBP chimeras using pETMalc-H, a derivative of Novagen's pMal vector system (Pryor and Leiting, 1997). The pETMalc-H vector produces an N-terminal fusion to *E. coli* maltose binding protein (MBP). Hexahistidine tags are present at the N- and C-terminus of MBP for purification, and a thrombin/cleavage site exists in a linker region between fusions for removal of MBP.

Large scale expression was performed with *E. coli* strain BL21 (DE3) cells (Novagen) in 2-4 L Luria Broth at 37˚C. Expression was induced with 0.5 mM IPTG when the cells where in log phase growth ($OD_{600}$=0.65). Overexpression was carried out for 18-20 hours at 25ºC, then cell pellets were harvested by centrifugation and stored at -80˚C.

.

**Purification of TRAP inclusion bodies.** Bacterial pellets were resusupended in 20 mL Sucrose solution (25% sucrose, 50 mM Tris, pH 8.0, 150 mM NaCl, 5 mM EDTA, 10 mM DTT) using a homogenizer. Cell lysis was achieved by adding lysozyme (Sigma) to a final concentration of 0.3 mg/mL and stirring for 20 minutes at room temperature followed by sonication. Inclusion bodies were separated from soluble material by centrifuging at 30,000xg. The supernatant was discarded and the crude inclusion bodies were resuspended by a homogenizer into a Trition solution (0.5-2.0% Triton X-100, 50 mM Tris, pH 8.0, and 100 mM NaCl). The inclusion bodies were pelleted at 30,000xg and the wash step was repeated. A third wash was performed using high salt (50 mM Tris, pH 8.0, and 1M NaCl) followed by a urea wash (50 mM Tris, pH 8.0, 150 mM NaCl and 1M urea). The final wash was a Tris buffer (100 mM Tris, pH 8.0 and 150 mM NaCl) and the resultant pellet was resuspended in 3-5 mL 6 M guanidine HCl, 50 mM Tris, pH 8.0, 1 mM EDTA, and 10 mM DTT using the homogenizer. The solubilized inclusion bodies were spun at 100,000xg for 1 hour and the supernatant was stored at -80˚C. Quantification of protein concentration was performed using UV spectroscopy (A and Afree, $\varepsilon = 20460$ $M^{-1}cm^{-1}$; ATSR, $\varepsilon = 32620$ $M^{-1}cm^{-1}$).

**Purification and refolding of ATSR inclusion bodies.** Denatured ATSR was added to a final concentration of 1 μM to the refolding buffer (50 mM $NaP_i$ buffer, pH 7.9, 500 mM guanidine HCl, 300 mM NaCl, 3 mM $MgCl_2$, and 0.2 mM LDAO) and incubated at RT with gentle agitation on a rotating platform for 24-36 hours. For

.

purification by gel filtration, large sample volumes >10 mL were directly concentrated by Pellicon and/or Amicon ultrafiltraton devices using 10,000 MWCO membranes and injected onto a Superdex 200 (Pharmacia) column equilibrated in 50 mM Tris, pH 8.0, 500 mM NaCl, and 5 mM β-ME. Fractions were analyzed by SDS-PAGE and the purest fractions were pooled and concentrated using dialysis against storage buffer containing 30% PEG 30K.

**Binding assay and purification of refolded ATSR by heparin column.** Refolded ATSR required dialysis into a low salt buffer before proceeding to the heparin column, HE20 (Poros). To minimize loss of protein, rapid exchanges (2 hour intervals) against 50 mM $NaP_i$ buffer, pH 7.5, 30 mM NaCl, and 10 mM β-ME were performed twice. The sample was filtered (0.22 μm) to remove precipitate and loaded onto the column. An elution gradient of 0% to 50% 3M NaCl was used to fractionate the sample. Fractions were TCA precipitated and analyzed by SDS-PAGE.

**Far Western Blot**. HepG2 cells (ATCC) cultured in Eagle's minimum essential media supplemented with 10% FBS, 2 mM L-glutamine and Earle's BSS to contain 1.5 g/L sodium bicarbonate, 0.1 mM non-essential amino acids, and 1.0 mM sodium pyruvate, were washed and harvested to yield 1 million cells per mL buffer (50 mM Tris, pH 7.6, 150 mM NaCl and 1 mM EDTA). SDS-PAGE (2x) sample buffer was added to the cells and repeated pipetting with 22-guage syringe was used to shear the

.

DNA and reduce sample viscosity. A total of 250,000 cells were loaded per well onto a 10% SDS-PAGE that also included pre-stained markers, ExoU (negative control), and ExoU-his (positive control). The gel was transferred to PVDF membrane and blocked with 2% casein in TBS from 1 hour to overnight then washed with TBS-Tween 20. TRAP was added to a concentration of 1-10 μg/mL and incubated with the blot for 1-2 hours. Mouse α-penta His antibody in TBS was used to bind TRAP followed by a goat α-mouse antibody conjugated to HRP to visualize bands by luminescence on film.

**Purification of MBP fusions to TRAP constructs**. Cell pellets from overnight expression of MBP-TRAP were resuspended in $1/10^{th}$ the culture volume in 50 mM $NaP_i$ buffer, pH 8.0, 500 mM NaCl, and 5 mM β-ME and lysed by sonication. Insoluble cell debris was removed by centrifuging at 30,000xg. The supernatant was filtered (0.22 μm) and incubated with Ni-NTA agarose resin that was pre-equilibrated with 50 mM $NaP_i$ buffer, pH 8.0, and 500 mM NaCl. Bound protein was eluted with buffer supplemented with 250 mM imidazole. Protein concentration was assayed by UV spectroscopy (MBP-A and Afree, $\varepsilon = 82200\ M^{-1}cm^{-1}$; MBP-ATSR, $\varepsilon = 97040\ M^{-1}cm^{-1}$).

**Cloning and expression trials of TRAP-FC in transient Sf9 cells.** Expression of ATSR and A were attempted using a constitutive secreted protein expression system in *Spodoptera frugiperda* (Sf9) cells (Invitrogen). The expression vector, pMib,

.

which contains a N-terminal Honeybee melittin secretion signal, was engineered to also include a C-terminal fusion to human $IgG_1$-Fc, pMib-Fc. Sf9 at $2x10^6$ cells were transfected with pMib-TRAP-Fc (2 μg) and 20 μL Cellfectin reagent in Grace's insect essential media supplemented with 10% FBS. After 2 days of Blasticidin S-HCl selection (80 μg/mL), a FBS-free media was introduced and resistant cell lines were checked for protein expression in the medium over a 3-day period. One hundred μL aliquots were taken from each well and TCA precipitated. The samples were run on an 11% SDS-PAGE then transferred to PVDF membrane. The inclusion of $IgG_1$-Fc eliminated the need for a primary antibody. Rabbit α-human $IgG_1$-Fc conjugated to HRP was incubated with the blot for detection with luminescence using film.

.

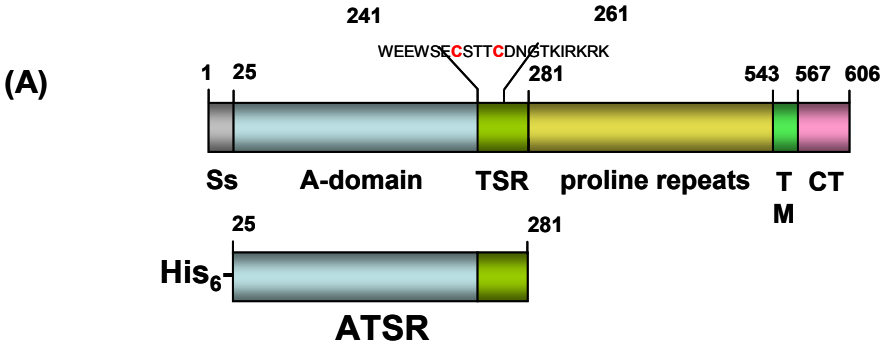
# Results

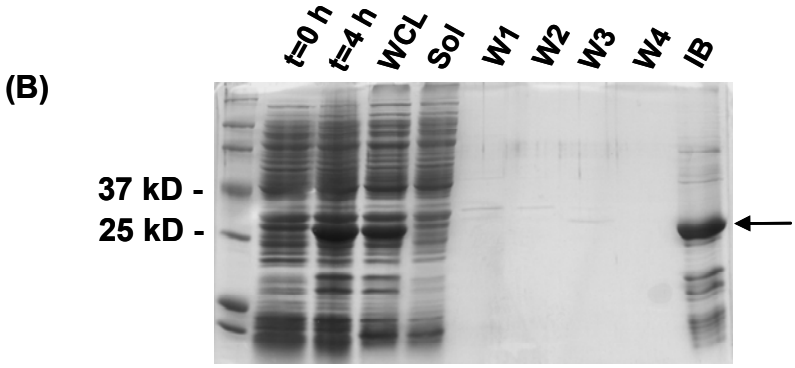
### Purification of TRAP inclusion bodies from E. coli

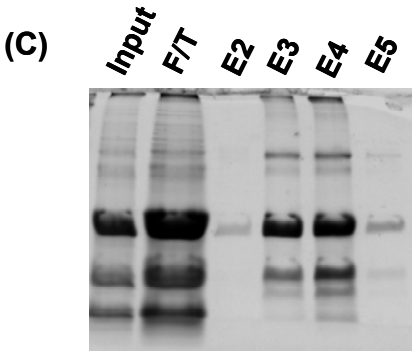
Expression of TRAP was first investigated in *E. coli* expression vectors. The A-domain and thrombospondin-related (TSR) region of TRAP, both shown to facilitate adhesion to host hepatocytes, were included in a construct containing an N-terminal hexahistidine tag (Fig. 6.1A). Overexpression of this construct, ATSR, resulted in no soluble products when visualized by SDS-PAGE (Fig. 6.1B, lane Sol); however, an overexpression band was evident in the insoluble fraction (Fig. 6.1B, lane IB). Purification of the inclusion bodies yielded 10 mg ATSR per liter of culture with an estimated final purify of 80%. To further increase the purity of ATSR before refolding, purification by Ni-NTA under denaturing conditions was attempted. Inexplicably a majority of the ATSR was found in the flow through (Fig. 2.1C, lane F/T) with only a small fraction of ATSR binding to the resin. Unfortunately the most prominent impurities also had affinity to the Ni-NTA matrix resulting in ATSR enriched with impurities. Interestingly the same results were observed with ATSR under denaturing conditions using cation exchange (data not shown). With no other feasible options to further assist in the purification of denatured ATSR refolding was attempted with the product in hand.

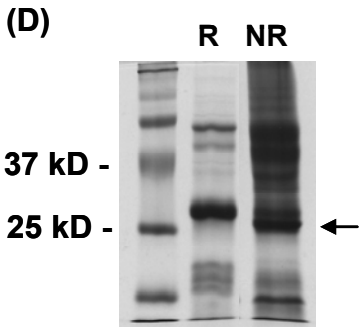
.

**Figure 6.1. Expression, purification and refolding of ATSR from *E. coli*.** (A) Schematic of *P. berghei* TRAP and recombinant TRAP construct, ATSR. (B) SDS-PAGE visualization of ATSR expressed as inclusion bodies in *E. coli* BL21 (DE3). Lanes correspond to total bacterial proteins prior to IPTG expression (t=0), 4 hours after induction (t=4), whole cell lysate (WCL), soluble fraction (Sol), inclusion body washes including 1.0% Triton X-100 (W1), Tris (W2), 1 M NaCl (W3), 1M urea (W4) and ATSR inclusion bodies resuspended in Gmd•HCl. (C) Ni-NTA purification under denaturing conditions and visualized by SDS-PAGE. The purity of ATSR before purification (Input), column flow through (F/T), and acidic pH elutions (E2-5). (D) Comparison of refolded ATSR under reducing (R) and non-reducing (NR) conditions and analyzed by SDS-PAGE. Arrow indicates the doublet band.


(A)
241
261
WEEWSECSTTCDNGTKIRKRK
1
25
281
543
567
606
Ss
A-domain
TSR
proline repeats
T
M
CT
25
281
His6
ATSR


(B)
t=0 h
t=4 h
WCL
Sol
W1
W2
W3
W4
IB
37 kD -
25 kD -


(C)
Input
F/T
E2
E3
E4
E5


(D)
R
NR
37 kD -
25 kD -

**Refolding ATSR inclusion bodies**

A moderately successful refolding condition had been previously identified by a postdoctoral researcher in the lab, Tamal Raha, and was employed as a starting point for refolding. In addition a published protocol for refolding TRAP with a functional TSR region was also incorporated (McCormick et al., 1999). The percent recovery of refolded ATSR from the input was approximately 15%. Reduced and non-reduced samples of the refolded protein were analyzed by SDS-PAGE to examine for the expected formation of disulfide bonds in the A-domain. Non-reduced samples ran as a doublet compared to the singlet band observed in the reduced samples (Fig. 6.1D; lane NR).

This demonstrated that there was a population of refolded species that formed intramolecular disulfide linkages and under non-reducing conditions kept the polypeptide in a more compact conformation permitting quicker migration by SDS-PAGE. It was assumed that the presence of the disulfide bond suggested that the refolded ATSR had assumed the correct conformation. To assure that any remaining reactive, solvent accessible cysteine residues were not available to participate in unfavorable intermolecular disulfide linkages iodoacetamine (IAA) was added to the refolded protein.

As a last purification step, ATSR was exchanged into a high salt buffer by gel filtration. Based on gel filtration standards the molecular mass of the eluted species was determined to be 115 kD (Fig. 6.2A). Since the expected mass of an ATSR

.

monomer is 31.1 kD the observed peak corresponded at best to a tetrameric species. Importantly, gel filtration appeared to have very little success in purifying refolded ATSR from the numerous impurities indicating other purification steps would be needed to obtain pure ATSR. It is also worth noting that the eluted peak had a higher absorbance at 260 nm than 280 nm, which is atypical of protein samples.

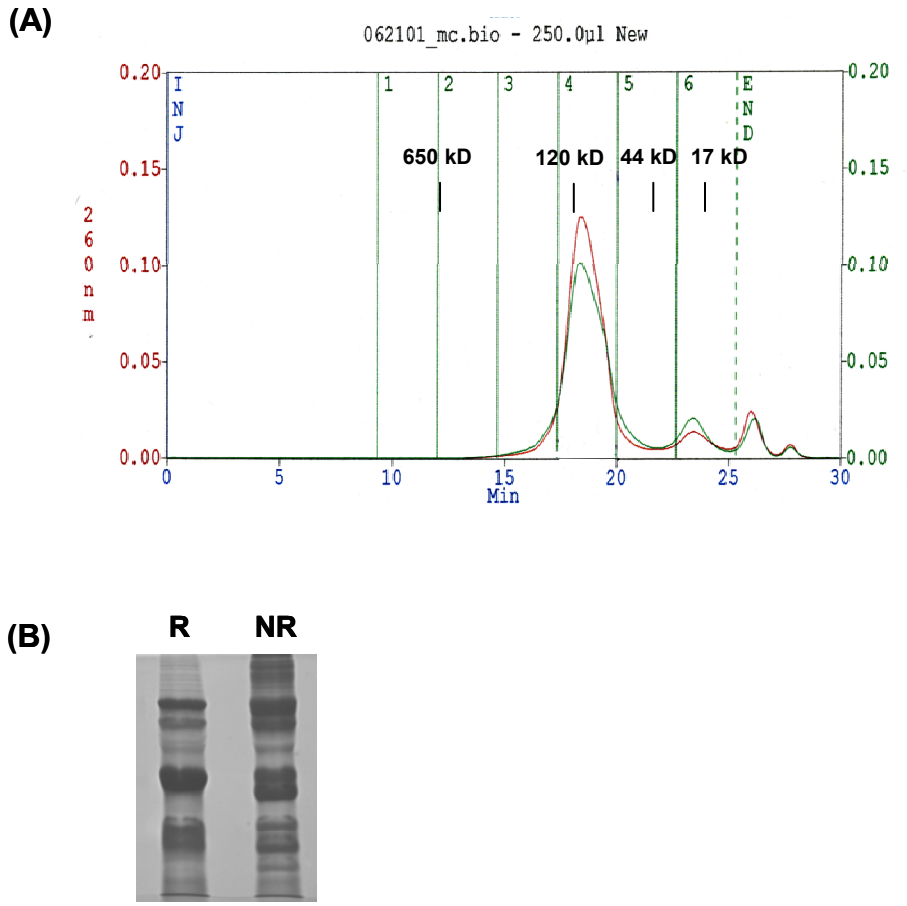



**Figure 6.2. Gel filtration analysis of refolded ATSR.** (A) Gel filtration profile of refolded ATSR. Elution times for gel filtration standards are included. (B) The major peak from gel filtration, fraction 4, was visualized by SDS-PAGE under reducing (R) and non-reducing (NR) conditions.

**Refolded ATSR precipitates from solution**

Following gel filtration, ATSR was recovered, concentrated to 0.1 to 0.5 mg/ml and stored at 4˚C. Protein concentrations above 1 mg/ml were avoided because the protein was observed to immediately begin precipitating from solution. Despite maintaining a relatively dilute protein concentration the ATSR construct was very unstable and remained in the tetrameric state for <10 hours before shifting entirely to a population of high molecular mass aggregates that also precipitated from solution. Analysis of the precipitate by SDS-PAGE confirmed that it was almost entirely composed of ATSR (data not shown).

Further investigations were performed to determine the optimal storage buffer. It was concluded that salt concentration was the most critical factor and concentrations of ≥500 mM NaCl helped to temporarily stabilize the tetrameric state in TRAP and delay aggregation to ~10 hours. Low salt concentrations (<250 mM) always produced aggregated TRAP by gel filtration. Additives such as glycerol, 1 M urea, magnesium salts, and reducing agents in the presence of low or high salt had no effect on preserving the tetrameric state of ATSR or even delaying the formation of aggregates. Heparin, a known ligand to TSR, also did not induce stability (data not shown).

**ATSR binds heparin**

McCormick et al. (McCormick et al., 1999) demonstrated that heparin, an anionic sugar moiety associated with the extracellular matrix of mammalian cells, is a

.

ligand for TRAP. The binding can be attributed to the TSR motif which is included in the ATSR construct. To assess the quality of the refolded ATSR it was subjected to a heparin affinity column (Fig. 6.3A). The results show that ATSR has a high affinity for the heparin matrix as the column flow through showed no detectable ATSR (Fig. 6.3B, lane 2) and elution of ATSR using a salt gradient was observed only at concentrations of >500 mM NaCl. Analysis of the eluted fractions by SDS-PAGE revealed that the heparin column does remove some of the impurities from the refolded ATSR preparation.

However, binding of ATSR was dependent on first dialyzing the refolded protein into a low salt buffer. Even at concentrations below 0.1 mg/ml, ATSR precipitated during a 4-hr dialysis against the heparin binding buffer necessitating filtration of the dialyzed sample prior to injection onto the column. This precipitation resulted in a >80% loss of the starting material.

.

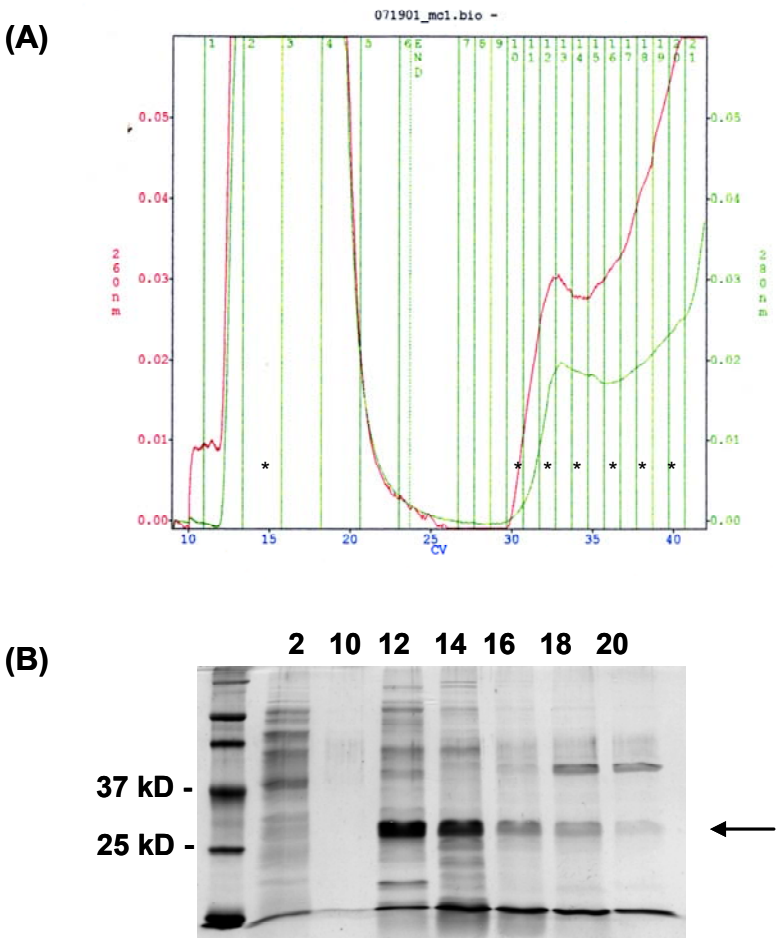



**Figure 6.3. ATSR shows good affinity to heparin.**

**(A) Binding of ATSR to** immobilized heparin was tested using a salt gradient to disrupt binding. (B) SDS-PAGE visualization of heparin column fractions. Fraction 2 corresponds to the flow-through while fractions 10 (~480 mM NaCl), 12 (~630 mM NaCl) and 14 (~780 mM NaCl) represent the major elution peak. ATSR was observed to elute with 630 mM NaCl and was present in fractions 12 and 14 (arrow).

**Attempting to identify the hepatocyte receptor for TRAP**

Refolded ATSR was utilized directly after gel filtration to assure that the construct was in the tetrameric state and not aggregated. To identify potential ligands a far western blot was performed in which ATSR was incubated with HepG2 lysates that had been transferred onto a membrane. A control far western was run in parallel in which the TRAP incubation step was excluded. In both the control and experimental blot only a single band at ~50 kD was evident (Fig. 6.4; second lane of each blot, HL). This was deduced to be non-specific binding of the primary antibody to a molecule in the HepG2 lysate. The absence of an unique band in the experimental blot suggested that the TRAP construct was either too dilute or not properly folded to permit functional binding. Extension of the incubation times and/or a 10-fold increase in the concentration of ATSR did not produce any additional bands. As a result new protein expression strategies were attempted to generate functional TRAP constructs.

.

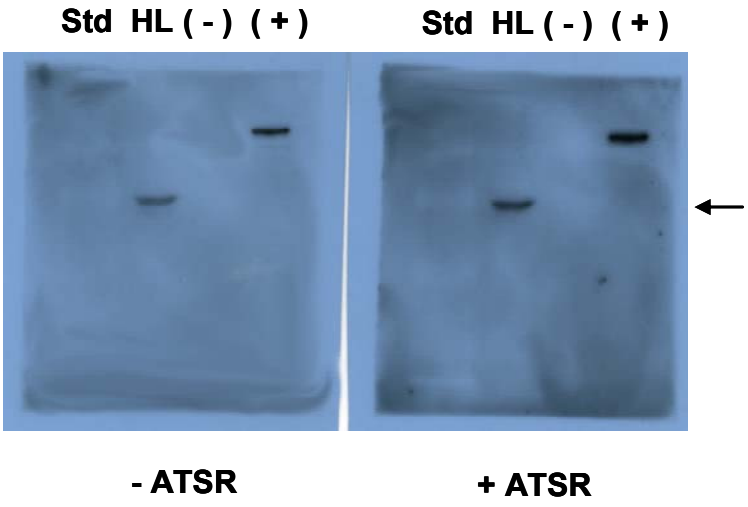



**Figure 6.4: Overlay blot of refolded ATSR to HepG2 lysates shows no binding.** Whole cell lysates of HegG2 cells were transferred to PVDF and either were (+ATSR) or were not (-ATSR) incubated with 10 ug/ml ATSR. Anti-pentahis antibody was used to detect bound ATSR followed by a secondary antibody conjugated to HRP. No unique bands were observed in the experimental blot versus the control blot. A band correlating to non-specific binding of the primary antibody is indicated by the arrow. Lanes correspond to molecular weight standards (Std), HepG2 whole cell lysate (HL), negative control protein ExoU ( - ), and positive control protein ExoU-his ( + ). The detected band in the HepG2 lystate is indicated by the arrow.

**MBP fusions produce soluble aggregates**

To enhance solubility of TRAP constructs they were engineered into an expression vector that resulted in an C-terminal fusion partner to *E. coli* maltose-binding protein (MBP) fusions. The ATSR construct (MBP-ATSR) and two additional constructs encoding residues 25-234 of the A-domain (MBP-A) and residues 40-226 of the A-domain (MBP-Afree) were chosen for expression trials.

.

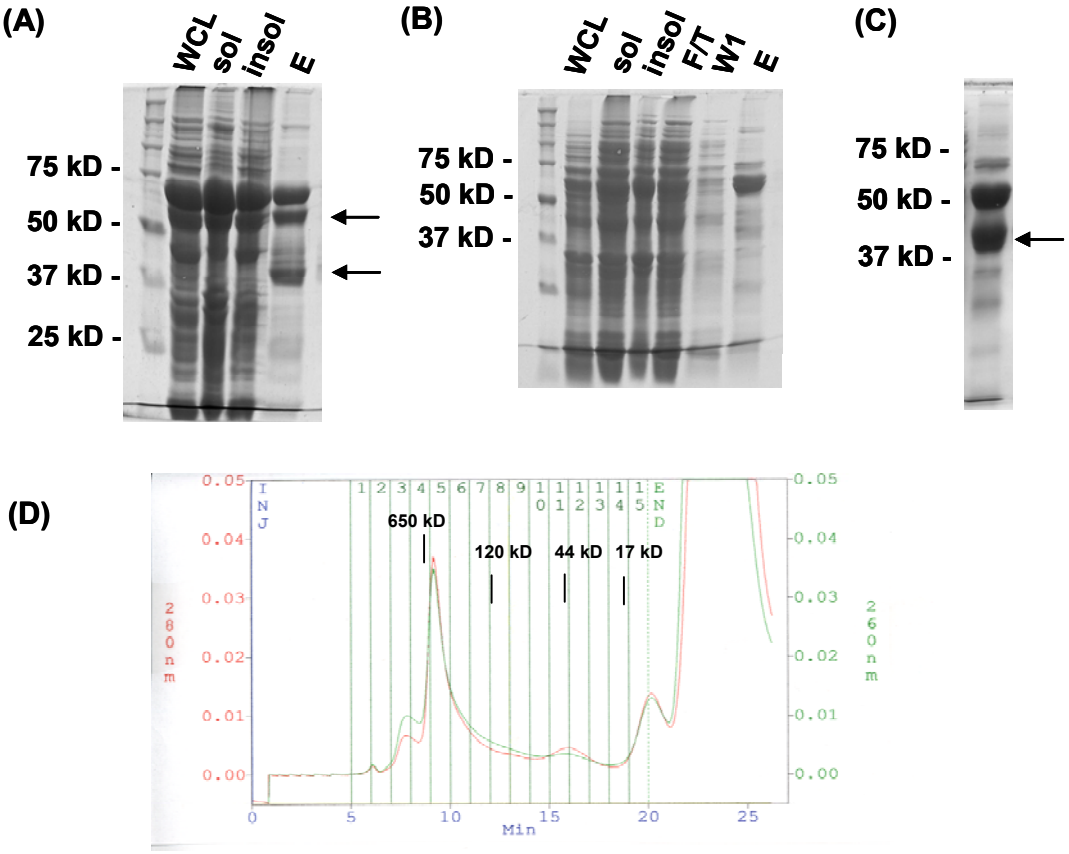



**Figure 6.5. MBP-TRAP fusions produce soluble protein in *E. coli*.** SDS-PAGE visualization of expression tests for (A) MBP-ATSR and (B) MBP-A (25-234 of A-domain) and (C) showing soluble protein captured by Ni-NTA resin for MBP-Afree (40-226 of A-domain). Degradation products of MBP-ATSR and MBP-Afree are indicated by arrows. Lanes correspond to *E. coli* whole cell lysate (WCL), soluble fraction (Sol), insoluble fraction (Insol), Ni-NTA flow through (F/T), Ni-NTA wash (W1) and boiled beads (E). (D) Gel filtration profile of purified MBP-A running near the void volume. MBP-ATSR and MBP-Afree also elute as high molecular weight species (data not shown). Elution times of gel filtration standards are included.

Overexpression of MBP-ATSR was evident in induced fractions and analysis of the solubility of the construct showed that approximately 50% of the overexpressed product was soluble (Fig. 6.5A, lane Sol). Purification by Ni-NTA showed that three

prominent products were captured and eluted from the column (Fig. 6.5A, lane E). The largest species (Band 1) migrated just below the 75 kD marker and corresponded to the expected size of ATSR (30 kD) fused to MBP (41 kD). Band 2 appeared to represent a degraded fusion product while Band 3 migrated at the expected mass of MBP alone.

MBP-A and MBP-Afree behaved similar to MBP-ATSR, except MBP-A showed no significant degradation (Fig. 6.5B; lane 7, E). The Afree construct purified as a doublet (Fig. 6.5C) and the lower band corresponded to the mass of free MBP.

Unlike refolded ATSR, these construct did not show any time-dependent precipitation and appeared relatively stable. While these constructs showed identical affinity to the heparin binding column as refolded ATSR (data not shown), gel filtration revealed that all the recovered protein existed as a soluble aggregate (Fig. 6.5D). This construct was also used unsuccessfully in overlay assays and yielded identical results to those observed for refolded ATSR (data not shown).

**Lack of expression of TRAP-Fc in insect cells**

An alternative expression system was pursued to circumvent the problems observed in *E. coli* expression. Expression of ATSR and the A-domain alone were pursued in insect cells using a C-terminal fusion partner (Fc-$IgG_1$). Expression and secretion of Fc-$IgG_1$ control into the media was evident after one day and increased

.

over a three-day expression test by immunoblotting (Fig. 6.6, lanes Fc 1-3d). However, no protein expression was observed in the media for the TRAP-Fc products by the third day (Fig. 6.6, A-Fc,lanes 4-6 and ATSR, lanes 7-9). The insect cells were not harvested and immunoblotted for TRAP expression, so it is unknown whether TRAP was expressed did occur but was not secreted.

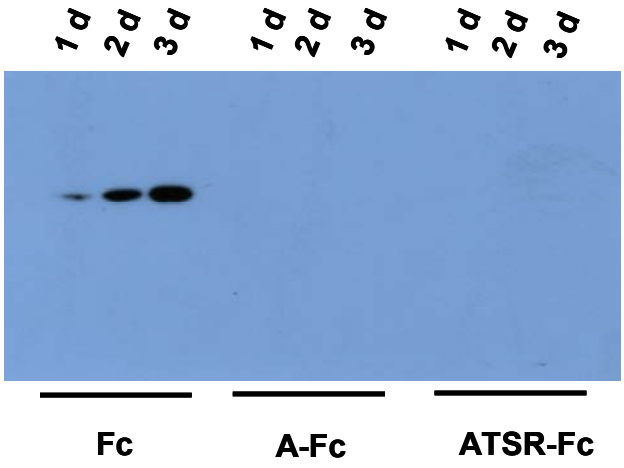



**Figure 6.6. TRAP-Fc constructs fail to express in transfected Sf9 cells.** Detection of secreted A-Fc and ATSR-Fc from transfected Sf9 cells using anti-$IgG_1$-Fc antibody conjugated to HRP was unsuccessful over a 3-day period. Transfection with the vector alone gave detectable expression after 1 day and showed steady accumulation over the 3 day period.

.

# Discussion

*Plasmodium berghei* TRAP constructs expressed in *E. coli* or Sf9 cells for biochemical assays did not demonstrate stability or apparent functionality. *E. coli* expression lead to insoluble protein, whereas MBP fusions produced soluble aggregates. Moreover, refolded inclusion bodies befell a similar fate and completely precipitated from solution. Finally, no expression of TRAP-Fc was detected from transient transfection of Sf9 cells.

Purification and refolding of TRAP constructs A, Afree, and ATSR had already been established by a previous lab member, Tamal Raha. All three constructs could be purified to modest levels and refolded, but quickly precipitated from solution. Despite these constructs exclusively forming inclusion bodies when overexpressed in *E. coli*, precedence for successfully refolding an A-domain had been established. The structure of von Willebrand Factor A1, an homolog to TRAP A-domain, was solved from refolded protein (Emsley et al., 1998). Obtaining functional TRAP appeared to be the function of optimizing refolding and storage conditions.

In the absence of an assay to ascertain functional TRAP, three characteristics of refolded protein were relied upon to confirm successful refolding: disulfide bond formation, elution as a monomer as assayed by gel filtration and heparin binding. In testing for the formation of disulfide linkages in the A-domain, a distinct gel shift on

.

non-reducing SDS-PAGE was seen for approximately half of the refolded protein population for A and ATSR. This was viewed as an encouraging sign that refolding was successful because structures of crystallized homologs showed a conserved disulfide linkage (Bienkowska et al., 1997; Emsley et al., 1998; Qu and Leahy, 1995). However, a total of four cysteines (residues 33, 200, 206, and 229) are present in the TRAP A-domain making it uncertain whether disulfide formation occurred between the appropriate residues. To address this concern, a construct (Afree, residues 40-226) excluding Cys33 and Cys229 (the cysteines most likely participating in the linkage) was designed. Ultimately this construct behaved similarly to the other constructs leading to aggregate formation. Persistent aggregation problems did not necessarily mean that functional TRAP was not produced, but it did indicate that if it were produced it had a short lifespan.

All the refolded constructs consistently migrated as tetramers by gel filtration. This was a surprising result since other A-domains appeared to be monomeric. In the absence of a high salt buffer, TRAP would immediately aggregate and elute in the void volume in gel filtration. Gel filtration experiments also demonstrated that high salt concentrations ($\geq$500 mM NaCl) were necessary to delay aggregation. Inclusion of heparin, glycerol, magnesium, or EDTA in gel filtration buffers had no effect on aggregation properties of TRAP. It was therefore assumed that the TRAP A-domain preferentially formed tetramers.

.

It was previously shown that heparin is a ligand of TSR (Muller et al., 1993) and the A-domain (McCormick et al., 1999). As a final test of functionality, heparin binding was characterized using immobilized heparin. Again, constructs A, Afree, and ATSR showed strong binding, necessitating 600-700 mM NaCl to disrupt binding and indicating a moderate degree of specificity. The intrinsically basic pI (8.63) posed a slight caveat as non-specific interactions would be expected between a protein predicted to have a positive charge, or one with basic surface patches, and a negatively charged matrix. In fact, TRAP always eluted with a higher $A_{260}$ to $A_{280}$ ratio which might have been indicative of non-specific binding of DNA to basic patches on the TRAP surface.

Characterization of the constructs as tetrameric complexes containing disulfide bonds and showing specificity to heparin was the best evidence that functional TRAP had been produced. The identical behavior of all three constructs combined with the observations by Matuschewski et al. (Matuschewski et al., 2002) that the A-domain and TSR work cooperatively during *in vivo* infection suggested ATSR was the best construct on which to focus attention. Overlay assays, or far westerns, utilizing ATSR as the primary agent to probe for hepatocyte ligands blotted to a membrane failed to show any binding to HepG2 lysates. Blotting purified membrane fractions of HepG2 cells also yielded no novel bands between the control and experimental lanes.

.

Failures for the assay could reside in misfolded TRAP that did not adopt the native state thus failing to recognize the hepatocyte receptor(s). The sensitivity of the assay is another possible factor and radio-labeled TRAP might have proved to yield greater sensitivity. Later assays did employ a polyclonal α-TRAP antibody instead of the α-pentahis antibody but this did not produce different results.

MBP fusions were also pursued as a method to prevent aggregation. While MBP did prevent precipitation, the fusions were found to be as soluble aggregates. Interestingly, these constructs also bound the heparin column at identical salt conditions indicating that heparin binding is not impaired by aggregation. Thus, binding of heparin does not correspond to well behaved protein and proved to be unhelpful in identifying functional TRAP.

Finally, expression in an insect cell system using TRAP-Fc constructs failed to show any expression. The same results were seen in *P. pastoris* and baculovirus expression by a previous lab member highlighting the challenges of expressing TRAP from *P. berghei*.

.

# References

Adams, J.C. and Tucker, R.P. (2000) The thrombospondin type 1 repeat (TSR) superfamily: Diverse proteins with related roles in neuronal development. *Developmental Dynamics*, **218**, 280-299.

Bienkowska, J., Cruz, M., Atiemo, A., Handin, R. and Liddington, R. (1997) The von Willebrand Factor A3 domain does not contain a Metal Ion-dependent adhesion motif. *J Biol Chem*, **272**, 25162-25167.

Carruthers, V.B. and Sibley, L.D. (1999) Mobilization of intracellular calcium stimulates microneme discharge in Toxoplasma gondii. *Mol Microbiol*, **31**, 421-428.

Cerami, C., Frevert, U., Sinnis, P., Takacs, B., Clavijo, P., Santos, M.J. and Nussenzweig, V. (1992) The basolateral domain of the hepatocyte plasma membrane bears receptors for the circumsporozoite protein of Plasmodium falciparum sporozoites. *Cell*, **70**, 1021-1033.

Emsley, J., Cruz, M., Handin, R. and Liddington, R. (1998) Crystal structure of the von Willebrand Factor A1 domain and implications for the binding of platelet glycoprotein Ib. *J Biol Chem*, **273**, 10396-10401.

Esko, J.D. and Selleck, S.B. (2002) Order out of chaos: Assembly of ligand binding sites in heparan sulfate. *Annual Review Of Biochemistry*, **71**, 435-471.

Girma, J.P., Meyer, D., Verweij, C.L., Pannekoek, H. and Sixma, J.J. (1987) Structure-function relationship of human von Willebrand factor. *Blood*, **70**, 605-611.

Kappe, S., Bruderer, T., Gantt, S., Fujioka, H., Nussenzweig, V. and Menard, R. (1999) Conservation of a gliding motility and cell invasion machinery in Apicomplexan parasites. *J Cell Biol*, **147**, 937-944.

.

Lawler, J. and Hynes, R.O. (1986) The structure of human thrombospondin, an adhesive glycoprotein with multiple calcium-binding sites and homologies with several different proteins. *J Cell Biol*, **103**, 1635-1648.

Lee, J.O., Rieu, P., Arnaout, M.A. and Liddington, R. (1995) Crystal structure of the A domain from the alpha subunit of integrin CR3 (CD11b/CD18). *Cell*, **80**, 631-638.

Lyon, M., Deakin, J.A. and Gallagher, J.T. (1994) Liver heparan sulfate structure. A novel molecular design. *J Biol Chem*, **269**, 11208-11215.

Matuschewski, K., Nunes, A.C., Nussenzweig, V. and Menard, R. (2002) Plasmodium sporozoite invasion into insect and mammalian cells is directed by the same dual binding system. *Embo J*, **21**, 1597-1606.

McCormick, C.J., Tuckwell, D.S., Crisanti, A., Humphries, M.J. and Hollingdale, M.R. (1999) Identification of heparin as a ligand for the A-domain of *Plasmodium falciparum* thrombospondin-related adhesion protein. *Mol Biochem Parasitol*, **100**, 111-124.

Michishita, M., Videm, V. and Arnaout, M.A. (1993) A novel divalent cation-binding site in the A domain of the beta 2 integrin CR3 (CD11b/CD18) is essential for ligand binding. *Cell*, **72**, 857-867.

Muller, H.M., Scarselli, E. and Crisanti, A. (1993) Thrombospondin related anonymous protein (TRAP) of Plasmodium falciparum in parasite-host cell interactions. *Parassitologia*, **35 Suppl**, 69-72.

Pinzon-Ortiz, C., Friedman, J., Esko, J. and Sinnis, P. (2001) The binding of the circumsporozoite protein to cell surface heparan sulfate proteoglycans is required for Plasmodium sporozoite attachment to target cells. *Journal Of Biological Chemistry*, **276**, 26784-26791.

Pryor, K.D. and Leiting, B. (1997) High-level expression of soluble protein in Escherichia coli using a His6-tag and maltose-binding-protein double-affinity fusion system. *Protein Expr Purif*, **10**, 309-319.

.

Qu, A. and Leahy, D.J. (1995) Crystal structure of the I-domain from the CD11a/CD18 (LFA-1,$a_L$b2) integrin. *Proc Natl Acad Sci U S A*, **92**, 10277-10281.

Sultan, A.A., Thathy, V., Frevert, U., Robson, K.J., Crisanti, A., Nussenzweig, V., Nussenzweig, R.S. and Menard, R. (1997) TRAP is necessary for gliding motility and infectivity of plasmodium sporozoites. *Cell*, **90**, 511-522.

Templeton, T.J. and Kaslow, D.C. (1997) Cloning and cross-species comparison of the thrombospondin-related anonymous protein (TRAP) gene from Plasmodium knowlesi, Plasmodium vivax and Plasmodium gallinaceum. *Mol Biochem Parasitol*, **84**, 13-24.

Templeton, T.J., Kaslow, D.C. and Fidock, D.A. (2000) Developmental arrest of the human malaria parasite Plasmodium falciparum within the mosquito midgut via CTRP gene disruption. *Mol Microbiol*, **36**, 1-9.

Ueda, T., Rieu, P., Brayer, J. and Arnaout, M.A. (1994) Identification of the complement iC3b binding site in the beta 2 integrin CR3 (CD11b/CD18). *Proc Natl Acad Sci U S A*, **91**, 10680-10684.

Ungureanu, E., Killick-Kendrick, R., Garnham, P.C., Branzei, P., Romanescu, C. and Shute, P.G. (1977) Prepatent periods of a tropical strain of Plasmodium vivax after inoculations of tenfold dilutions of sporozoites. *Trans R Soc Trop Med Hyg*, **70**, 482-483.

Verhave, J.P. and Meis, J.F. (1984) The biology of tissue forms and other asexual stages in mammalian plasmodia. *Experientia*, **40**, 1317-1329.

Wengelnik, K., Spaccapelo, R., Naitza, S., Robson, K.J., Janse, C.J., Bistoni, F., Waters, A.P. and Crisanti, A. (1999) The A-domain and the thrombospondin-related motif of Plasmodium falciparum TRAP are implicated in the invasion process of mosquito salivary glands. *Embo J*, **18**, 5195-5204.

Yuda, M., Sakaida, H. and Chinzei, Y. (1999) Targeted disruption of the plasmodium berghei CTRP gene reveals its essential role in malaria infection of the vector mosquito. *J Exp Med*, **190**, 1711-1716.

.