

ABSTRACT

Title of Dissertation: Transcriptional Activation by the Group A Streptococcal Mga Regulator

Lara Louise Hause, Doctor of Philosophy, 2012

Dissertation Directed by: Associate Professor Kevin S. McIver
Department of Cell Biology and Molecular Genetics

Streptococcus pyogenes (Group A Streptococcus, GAS) is a Gram-positive obligate human pathogen that causes a range of diseases at many different tissue sites. The ability of this organism to colonize and persist within these various niches of the body correlates with broad changes in gene expression. Mga, the multiple gene regulator of GAS, is an important global transcriptional regulator of virulence genes that encode factors promoting adhesion, host cell invasion and immune evasion. Mga directly activates these genes by binding to specific promoter sites that range from 45 to 60 nucleotides in size based on DNaseI footprint analysis; however identified Mga binding sites share less than 50% DNA sequence similarity, making the identification of a consensus Mga binding site difficult. We have identified nucleotides necessary for Mga binding in the Mga-regulated *Pemm* promoter from the clinically relevant M1 MGAS5005 strain of GAS. Random and directed mutations were assessed for effects on transcription *in vivo* and DNA binding *in vitro*. This screen identified predominately Gs and Cs, in two clusters at the 3' and 5' end that suggest that Mga binds DNA as a dimer and reduced the *Pemm* binding site to 35 bp. However directed mutagenesis in other

binding sites found that these interactions were not necessarily conserved. These experiments also sought to establish a method to study genome-wide DNA binding and can successfully enrich for Mga-regulated genes. Protein-protein interactions with RNA polymerase are another key component to activate transcription. Functional *in vitro* transcription assays and *in vitro* co-purification assays were performed to determine if Mga interacts with either the α C terminal domain or domain 4 of σ . While Mga does appear to make protein-protein contacts with the holoenzyme, they do not occur through either domain alone. The dimerization of Mga through its EIIB domain was established by analytical ultracentrifugation. *In vitro* transcription assays linked phosphorylation by the phosphoenolpyruvate transferase system to the down regulation of Mga activity. By understanding how Mga interacts with essential elements of its promoters, this study seeks to define Mga's role in regulating virulence in this important human pathogen.

Transcriptional Activation by the Group A Streptococcal Mga Regulator

By

Lara Louise Hause

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2012

Advisory Committee:
Associate Professor Kevin McIver, Chair
Associate Professor Douglas Julin
Assistant Professor Vincent Lee
Professor Zvi Kelman
Professor David Fushman, Dean's Representative

© Copyright by
Lara Louise Hause
2012

Preface

It would not be called research if we knew what we were doing.

Acknowledgements

First, I would like to thank my advisor Dr Kevin McIver for his advice and support over the last five years. I would also like to thank my committee for their help as well. Dr. Dorothy Beckett assisted us with the AUC experiments. Dr. Daniel Nelson from IBBR and Dr. Daniel Kearns from University of Indiana assisted by donating PlyC and the bacterial 2 hybrid system, respectively.

I would like to thank the McIver lab for their support, both scientific and entertainment; Elise for her knowledge of proteins, Yoann for the coffee breaks, Kanika and Kathryn for their friendship. This past year Adam has been a great help in performing experiments, he is responsible for much of the TAP tagged protein work.

My parents and family have always given me their love and support, I wouldn't be here without them. I especially want to thank Trey for being awesome.

Table of Contents

PREFACE	II
ACKNOWLEDGEMENTS	III
LIST OF TABLES	VII
LIST OF FIGURES	VIII
LIST OF ABBREVIATIONS	IX
CHAPTER 1	1
LITERATURE REVIEW	1
<i>Historical Perspective</i>	<i>1</i>
<i>Classification</i>	<i>1</i>
General characteristics and growth requirements	1
Lancefield grouping	2
M and T grouping	2
Class determination.....	3
<i>Diseases</i>	<i>3</i>
<i>Throat</i>	<i>4</i>
<i>Pharyngitis</i>	<i>4</i>
<i>Scarlet Fever</i>	<i>5</i>
<i>Skin</i>	<i>5</i>
<i>Impetigo</i>	<i>5</i>
<i>Cellulitis</i>	<i>6</i>
<i>Erysipelas</i>	<i>6</i>
<i>Invasive</i>	<i>6</i>
<i>Puerperal Fever</i>	<i>6</i>
<i>Streptococcal Toxic Shock Syndrome</i>	<i>7</i>
<i>Bacteremia</i>	<i>7</i>
<i>Necrotizing Fasciitis</i>	<i>8</i>
<i>Post immune sequelae</i>	<i>8</i>
<i>Acute Rheumatic Fever</i>	<i>8</i>
<i>Acute Post Streptococcal Glomerulonephritis</i>	<i>9</i>
VACCINATION	10
VIRULENCE FACTORS OF GAS	10
CELL ASSOCIATED	11
M and M-like Proteins	11
MSCRAMMS	12
Streptococcal collagen-like protein.....	13
C5a protease.....	13
Capsule.....	13
Serum Opacity Factor	14
SECRETED	14
Streptococcal pyrogenic exotoxin	14
Streptococcal Inhibitor of Complement.....	15
REGULATION	15
<i>Two component systems</i>	<i>15</i>
CovRS	15
TrxRS.....	16
<i>Other Regulators</i>	<i>16</i>
CcpA	16
Rgg/RopB	17
RALPs.....	17
MGA	18
History/discovery	18
Regulon characteristics	18

Protein characteristics	19
DNA binding.....	20
Regulation.....	22
DNA BINDING PROTEINS.....	23
<i>Functions</i>	23
<i>DNA Binding Domains</i>	24
Helix-turn-Helix family	24
Helix-loop-helix.....	25
Zinc Finger.....	25
Leucine Zipper.....	26
Beta Ribbon Motif.....	27
RNA POLYMERASE.....	27
<i>The Holoenzyme</i>	27
<i>Steps of transcription</i>	28
<i>Activation and Repression</i>	29
PHOSPHOENOLPYRUVATE PHOSPHOTRANSFERASE SYSTEM (PTS).....	32
Introduction.....	32
Components	32
PTS and Virulence.....	34
SUMMARY	34
CHAPTER 2	35
MATERIALS AND METHODS	35
<i>Bacterial strains and media</i>	35
<i>DNA manipulations</i>	35
<i>Construction of Luciferase plasmids</i>	36
<i>Luciferase Assay</i>	36
<i>Expression and Purification Mga-His proteins from E. coli</i>	37
<i>Expression and Purification of Mga4-CBP from E. coli</i>	38
<i>Electrophoretic Mobility Shift assay (EMSA)</i>	38
<i>DNaseI Footprint Analysis</i>	39
<i>Methylation Protection and Interference Assay</i>	40
<i>Uracil and Missing Thymine Interference Assay</i>	41
<i>In vitro Transcription</i>	41
<i>In vitro Phosphorylation-Transcription</i>	42
<i>ChAP (Chromosome Affinity Purification)</i>	43
<i>Quantitative PCR</i>	44
<i>Sedimentation Equilibrium</i>	44
<i>DRACALA (Differential Radial Capillary Action of Ligand Assay)</i>	46
<i>Construction of Bacterial-Two-Hybrid plasmids</i>	46
<i>Bacterial-Two-Hybrid</i>	47
<i>Bacterial-Two-Hybrid Western Blots</i>	48
<i>Immunoblots</i>	49
<i>In vitro Co-Affinity Purification</i>	49
<i>Purification of RNA Polymerase</i>	50
<i>Purification of His₆-α, His₆-α-ΔCTD, His₆-αNTD-σ4, His₆-σ and His₆-$\sigma$$\Delta$4.....</i>	51
<i>Purification of σ</i>	51
<i>Creation of Mutant RNA Polymerases</i>	52
<i>Construction of Protein Expression vectors</i>	53
CHAPTER 3	57
NUCLEOTIDES CRITICAL FOR THE INTERACTION OF THE <i>STREPTOCOCCUS PYOGENES</i>	
MGA VIRULENCE REGULATOR WITH MGA-REGULATED PROMOTER SEQUENCES	57
INTRODUCTION	57

RESULTS	59
<i>Characterization of MBSs in the MIT1 MGAS5005.</i>	59
<i>Comparison of Mga binding sites between different promoter categories.</i>	62
<i>Biochemical analysis of the Pemm1 Mga binding site.</i>	62
<i>In vivo analysis of Pemm1 binding site mutants.</i>	65
<i>EMSA analysis of Mga binding to Pemm mutants</i>	67
<i>Conservation of critical Pemm1 nucleotides in other category A Mga-regulated promoters.</i>	69
DISCUSSION.....	73
CHAPTER 4.....	81
INTERACTION OF MGA WITH RNA POLYMERASE	81
INTRODUCTION	81
RESULTS.....	82
<i>Bacterial 2 Hybrid Assay for Protein-Protein Interactions</i>	82
<i>Creation of strains for in vivo protein-protein interaction studies</i>	85
<i>Creation of Mutant RNAP for in vitro transcription</i>	87
<i>In vitro Co-affinity Purification</i>	87
<i>In vitro Transcription of Mutant RNAP</i>	89
DISCUSSION	91
CHAPTER 5.....	93
INTERACTION OF MGA WITH OTHER PROMOTERS AND REGULATORY ELEMENTS	93
INTRODUCTION	93
RESULTS.....	94
<i>Dimerization of Mga4-His₆ and Δ139Mga4-His₆ in solution.</i>	94
<i>PTS Phosphorylation of Mga leads to inactivation in vitro.</i>	96
<i>Purification of Mga by TAP tagging</i>	97
<i>Purification of Mga1-HTH4-His₆.</i>	99
<i>Category B Binding Sites</i>	100
<i>Category C Binding Sites</i>	102
<i>Comparison of DRACALA to EMSA and Filter-binding.</i>	103
<i>Validation of the Chromosome Affinity Purification.</i>	105
DISCUSSION.....	106
CHAPTER 6.....	112
CONCLUSIONS	112
MODEL OF MGA INTERACTIONS AT THE PROMOTER	112
<i>Dimerization</i>	112
<i>Interactions with RNA Polymerase</i>	115
<i>PTS Regulation of Mga Activity</i>	117
<i>Mga's role as a regulator within GAS</i>	118
<i>Significance</i>	119
APPENDICES	1
TABLE OF <i>E. COLI</i> STRAINS	ERROR! BOOKMARK NOT DEFINED.
TABLE OF GAS STRAINS.....	ERROR! BOOKMARK NOT DEFINED.
TABLE OF PLASMIDS	ERROR! BOOKMARK NOT DEFINED.
TABLE OF PRIMERS	ERROR! BOOKMARK NOT DEFINED.
TABLE OF MUTAGENIC OLIGONUCLEOTIDES	ERROR! BOOKMARK NOT DEFINED.
TABLE OF QPCR PRIMERS	ERROR! BOOKMARK NOT DEFINED.
TABLE OF BINDING OLIGONUCLEOTIDES	ERROR! BOOKMARK NOT DEFINED.

List of Tables

Table 1: <i>E. coli</i> strains	Error! Bookmark not defined.
Table 2: GAS strains	Error! Bookmark not defined.
Table 3: Plasmids Used.....	Error! Bookmark not defined.
Table 4: Primers.....	11
Table 5: Mutagenic Oligonucleotides.....	Error! Bookmark not defined.
Table 6: qPCR Primers	Error! Bookmark not defined.
Table 7: Binding Oligonucleotides	Error! Bookmark not defined.

List of Figures

Figure 1: Regulation of Mga-associated virulence factors in GAS	11
Figure 2: The domains of the Mga protein	19
Figure 3: Architecture of Mga Regulated Promoters.....	21
Figure 4: Class I and Class II Activators	29
Figure 5: Determination of MIT1 <i>Pemm1</i> , <i>PscpA</i> and <i>Psic</i> Mga Binding Sites.....	60
Figure 6: Conservation of nucleotides in known Mga binding sites	61
Figure 7: Biochemical analyses of Mga binding to <i>Pemm</i>	64
Figure 8: Luciferase promoter reporter assays of <i>Pemm</i> site-directed mutations <i>in vivo</i>	66
Figure 9: DNA-binding activity of <i>Pemm</i> site-directed mutations	68
Figure 10: Role of functional <i>Pemm</i> nucleotides conserved in <i>PscpA</i> and <i>Psic</i>	71
Figure 11: Summary of <i>Pemm</i> nucleotides important for Mga binding and activity.....	72
Figure 12: Category A Minimum binding sites	75
Figure 13: Bacterial 2-Hybrid Assays.....	83
Figure 14: Creation of Mutant RNA Polymerases.....	86
Figure 15: <i>In vitro</i> Co-affinity Purification of Mga with RNAP	88
Figure 16: <i>In vitro</i> Transcription with mutant RNA polymerases	90
Figure 17: Dimerization of Mga4-His ₆ and Δ 139Mga4-His ₆	95
Figure 18: Effect of <i>in vitro</i> phosphorylation of Mga on transcription	97
Figure 19: Schematic of TAP tagged Mga	98
Figure 20: Schematic of HTH mutations	99
Figure 21: Purification of Mga1-HTH4-His ₆ and TAP Tagged Mga.....	100
Figure 22: DNA-binding and Transcriptional Activation of Category Promoters	101
Figure 23: EMSA comparing Mga1-His ₆ to MBP-Mga6 at <i>Pmga</i>	103
Figure 24: Mga-DNA binding by DRACALA	104
Figure 25: Validation of ChAP-Seq DNA	107
Figure 26: Model of Mga Binding to the <i>Pemm</i> MBS.....	113
Figure 27: Protein-protein contacts at the Promoter	115
Figure 28: PTS phosphorylation disrupts Mga dimerization.....	117

List of Abbreviations

aa	amino acid
ASPGM	acute post-streptococcal glomerulonephritis
ARF	acute rheumatic fever
bp	base pair
<i>B. subtilis</i>	<i>Bacillus subtilis</i>
BSA	bovine serum albumin
cAMP	cyclic adenosine monophosphate
CBP	calmodulin binding protein
ChAP	chromosome affinity purification
CovRS	control of virulence two component system
CTD	C-terminal domain
dH ₂ O	distilled water
dI-dC	deoxyinosinic-deoxycytidylic acid
DMS	dimethyl sulfate
DTT	dithiothreitol
ds	double stranded
<i>E. coli</i>	<i>Escherichia coli</i>
<i>Emm</i>	M protein
EMSA	electrophoretic mobility shift assay
Fba	fibronectin binding protein
FPLC	fast protein liquid chromatography
GAS	the group A streptococcus
gDNA	genomic DNA
GRAB	protein G-related alpha2-macroglobulin-binding protein
<i>grm</i>	gene regulated by <i>mga</i>
HLH	helix-loop-helix
HTH	helix-turn-helix
Ihk/Irr	isp-associated histidine kinase and response regulator
Kb	kilobase
kDA	kilodalton
LB	Luria-Bertani medium
LTA	lipotechoic acid
LZ	leucine zipper
M	molar
μ	micro
MBP	Maltose binding protein
MBS	Mga binding site
Mga	multigene regulator of GAS
Mrp	M-related protein
MSCRAMM	microbial surface component recognizing adhesive matrix molecules
NaDOC	sodium deoxycholate
NEB	New England Biolabs
Ng	nanogram

Ni-NTA	nickel-nitriloacetic acid
Nra	negative regulator of the group A streptococcus
Nt	not tested
OF	opacity factor
ORF	open reading frame
PCR	polymerase chain reaction
<i>pel</i>	pleiotropic effect locus
PNK	polynucleotide kinase
PRD	PTS regulatory domain
PTS	phosphoenolpyruvate transferase system
q	quantitative
RALPS	RofA-like proteins
RivR	RofA-like protein
RNA	ribonucleic acid
RNAP	RNA polymerase
RofA	regulator of protein F
RR	response regulator
SDS	sodium dodecyl sulfate
<i>S. pyogenes</i>	<i>Streptococcus pyogenes</i>
SclA	streptococcal collagen like protein
ScpA	C5a peptidase
Sic	streptococcal inhibitor of complement
Sof	serum opacity factor
STSS	streptococcal toxic shock syndrome
TBE	tris/borate/EDTA
TCS	two component system
THY	Todd-Hewitt yeast extract medium
TrxSR	two-component regulatory system X
UV	ultraviolet
wHTH	winged helix turn helix
WT	wild-type
ZYP	N-Z-amine/yeast extract/phosphate

Chapter 1

Literature Review

Historical Perspective

Recognizable descriptions of the diseases caused by the Group A Streptococcus (GAS, *Streptococcus pyogenes*) have existed since at least the 16th century. In 1874, Billroth is credited with using the name streptococcus, from the Greek *streptos*, for twisted or chain, and *kokkos*, meaning berry or seed, to describe the globular chain forming bacteria that had been identified by many investigators [1]. At this time Streptococci were named according to the disease they were isolated from, e.g. *Streptococcus scarlitinae* or *Streptococcus puerperalis*. In 1903, Schotmüller classified streptococci based on their hemolytic pattern on blood agar plates; α , β or γ [1]. In 1933, Rebecca Lancefield developed a method to classify the β -hemolytic streptococci serologically by the group-specific carbohydrate and later according to M-protein type [1]. The β -hemolytic strains that fell into the serotype A group of the Lancefield classifications system were also the majority of strains that were pathogenic in humans; thus they were renamed as *Streptococcus pyogenes*.

Classification

General characteristics and growth requirements

Streptococcus pyogenes, or group A streptococcus (GAS), is a Gram-positive, non-motile bacteria, which forms chains of varying lengths. GAS will appear on 5% blood agar plates as β -hemolytic colonies. GAS is both catalase and oxidase negative.

GAS is a fastidious bacterium that relies on the fermentation of sugars for growth. In the laboratory GAS is grown in Todd-Hewitt, a nutrient rich media that includes neopeptone extracts, dextrose as a carbon source, and a complex mixture of nutrients from beef-heart infusion [2]. Yeast extract (0.2%) is added to the Todd-Hewitt (THY) to further enhance growth [3]. Cultures are grown statically at 37°C with 5% CO₂ or under ambient conditions.

Lancefield grouping

The Lancefield grouping was developed by Rebecca Lancefield in 1933 to distinguish streptococcal species serologically. This grouping of streptococci is based on immunological differences in their cell-wall polysaccharides (group A, B, C G and F) or lipoteichoic acids (group D) [4]. The Lancefield test is a precipitin reaction that uses hot acid to extract the carbohydrate, which is then incubated with C-antigens to the surface carbohydrate from different streptococci [5].

M and T grouping

The streptococcal M protein (*emm*) has been used to further categorize GAS strains by M-serotype. Lancefield developed a method for characterizing GAS into serotypes by extracting the M protein from a given strain and comparing it against standardized typing sera [4]. The N-terminus of the protein contains a type-specific moiety that is recognized by the typing sera. However, due to difficulties with this technique, other methods of characterizing each GAS strain have been developed. A molecular technique developed by Beall and Facklam uses PCR to amplify the 5' hyper-

variable region of the *emm* gene. Sequence comparison is performed to identify the *emm*-type. More than 200 M-serotypes have been recognized [6].

GAS may also be characterized according to T antigen or pilin, which was first defined by trypsin resistance [6]. There are many fewer T types than M types, and a T-type may be found with several M-types. Strains with the same M-type may also have varying T-types.

Class determination

GAS serotypes are divided into two classes based on the reactivity of their M protein with an antibody directed against the C repeat region and the presence of serum opacity factor (SOF) [7]. This opacity factor (OF) typing strongly correlates with specific M-serotypes [8]. A Class I serotype has an M protein that has a surface exposed C repeat region and lacks SOF. A Class II serotype lacks the M protein repeat region while SOF is present.

Diseases

GAS is an obligate human pathogen capable of causing a wide-range of diseases within its host. These may be the benign self-limiting infections such as strep throat (pharyngitis); life threatening invasive diseases such as necrotizing fasciitis, and post-infection sequelae such as acute rheumatic fever. GAS contributes a great burden to global human health; several hundred million people will suffer from the mild streptococcal infections while approximately 500,000 people will die from the more invasive diseases each year [9].

Throat

The throat is one of the two primary sites of a GAS infection. These infections result when GAS colonizes the pharynx and associated structures.

Pharyngitis

The most common infection caused by GAS is streptococcal pharyngitis, better known as “strep throat”. About 616 million new cases of pharyngitis occur each year [9]. While pharyngitis can affect any age group, it is most common among children between 5 and 12 years of age [10]. Colonization may be asymptomatic, which increases carriage rates among the population [4]. Transmission of the bacteria occurs primarily through inhalation of aerosolized droplets or direct contact with respiratory secretions, which makes crowding an important factor in the spread of pharyngitis [10]. The infection is most contagious early in the acute stage infection and up to two weeks after acquiring the organism, unless treated with antibiotics. The usual incubation period is between 2 and 5 days and displays itself with a pronounced sore throat along with fever, headache, and general malaise. To diagnose a streptococcal pharyngeal infection, the current standard technique is to culture a throat swab on blood agar plates. After 24-48 hours, the presence of GAS is verified through the formation of β -hemolytic colonies. Rapid antigen diagnostic tests may also be used; while they produce more immediate results, they generally are less sensitive and can yield false positives [10]. Pharyngitis is readily treated with antibiotics, most commonly penicillin and erythromycin; however even without treatment a pharyngeal infection is generally self limiting and virtually all symptoms resolve within a week.

Scarlet Fever

Scarlet fever was once a more serious and deadly disease. Today in developed countries this disease is generally mild. A benign scarlet fever infection is typically associated with pharyngitis [4] and is additionally characterized by a fine, diffuse red rash [10]. During severe cases of septic and toxic forms of scarlet fever, high fever and delirium may occur. These severe cases typically occur along with more invasive symptoms and are caused by streptococcal pyrogenic exotoxins (SPEs) and can lead to systemic toxicity and death [11].

Skin

The second primary entry point of GAS is through the skin when GAS colonizes the squamous epithelium.

Impetigo

Streptococcal impetigo, also known as pyoderma, is a skin infection of the dermis and epidermis, which most commonly affects the face and the lower extremities [4]. The characteristics of impetigo are pus-filled, bacteria-rich blisters that form on the skin that ooze and develop a thick crust. Itching is the other primary symptom associated with impetigo. Impetigo is typically found among children between the ages 2-5 who live in unhealthy conditions, but may also manifest in adults following previous respiratory tract infections or skin diseases. As GAS has been shown to colonize and persist on the skin it can quickly enter and invade the skin to cause infection following a minor trauma [12]. Impetigo is contagious, but is easily treated with penicillin. While impetigo itself is mild, it is linked to the streptococcal secondary sequelae, acute glomerulonephritis.

Cellulitis

Cellulitis is a diffuse inflammation of the subcutaneous layers of the skin [13]. Common symptoms are pinkish skin color, swelling and pain. Skin irritations such as burns, animal bites, dry skin or rashes can dispose a person to cellulitis. Diabetics or the elderly, who may have poor blood circulation or a weakened immune system, are also at risk. Penicillin is generally used to treat the condition.

Erysipelas

Erysipelas is an acute infection that involves the superficial layers of skin and cutaneous lymphatics. The area of inflammation is raised, and clearly demarcated from unaffected skin [13]. Although historically the face was the site of infections, now it is most common on the lower extremities. The bacteria typically enter through the skin following a local trauma or abrasion. The infection is usually treated with penicillin and resolves within a few weeks.

Invasive

Invasive disease occurs when GAS leaves the primary points of infection and invades normally sterile tissues. These infections can be rapid and aggressive and result from a complex interaction between GAS and the human immune system.

Puerperal Fever

Puerperal fever, or childbed fever, while once a deadly disease in the 19th century, now is only seen sporadically and isolated cases. The development of aseptic techniques, such as those pioneered by Semmelweis, and antibiotics have drastically reduced the

number of cases, and now it is rarely fatal. Puerperal fever can be contracted during pregnancy, miscarriage or abortion and is characterized by marked infection of the genital tract and endometrial lining. These symptoms may be masked by abdominal pain and not immediately recognized following delivery. GAS can use the endometrial lining to gain entry to the surrounding structures and bloodstream [1]. Once the sepsis occurs, fever, leukocytosis and severe pain typically alert doctors to that the more severe disease is present.

Streptococcal Toxic Shock Syndrome

Streptococcal toxic shock syndrome (STSS) is characterized by the isolation of the bacteria from a normally sterile site followed within 24-48 hours by shock and organ failure. The severe symptoms are generally associated with GAS pyrogenic exotoxins (SPEs). These superantigens over stimulate the immune system, leading to the massive release of pro-inflammatory cytokines that then leads to shock and tissue damage [13]. Skin infections are the most common portal of entry for streptococcal TSS and it is often present alongside other deep-seated systemic infections. Since the 1980s, reports of STSS have become more frequent in North America and Europe. While the elderly is the most common age group for this disease, many patients between the ages of 20 and 50 with no underlying infections have also been afflicted. Despite modern treatments, 30% of patients die [11].

Bacteremia

Streptococcal bacteremia, or the presence of the bacteria in the bloodstream, is most common in the very young or elderly. Pharyngitis associated scarlet fever

predisposes children to bacteremia, while in the elderly, bacteremia is secondary to infections of the skin [11]. Bacteremia rates have been increasing in individuals between 14 and 40 years of age, in particular among intravenous drug users. Bacteremia is clinically characterized by fever, chills and shock, and is treated with antibiotic therapy [11].

Necrotizing Fasciitis

Necrotizing fasciitis is a deep-seated infection of the subcutaneous tissue that destroys the fascia and fat while leaving skin intact [11]. Streptococcal gangrene begins at the site of a trivial lesion, but within 24 hours there is aggressive development of heat, erythema and tenderness with rapid spreading. Unless appropriate intervention is taken, this may quickly become cutaneous gangrene, and inflammation may spread along the fascia [13]. Shock and organ failure also appear. In order to treat necrotizing fasciitis, surgical debridement is used to remove the affected areas. Even with aggressive treatment, a mortality rate exceeding 50% has been observed [4].

Post immune sequelae

The post-immune sequelae develop after the bacteria have been cleared from the body. These symptoms are an autoimmune reaction in where the antibodies against GAS instead attack healthy human tissue.

Acute Rheumatic Fever

Acute rheumatic fever (ARF) is a sequel that occurs 2-3 weeks after pharyngitis [4,14]. Five major clinical manifestations of the disease may occur according to the criteria set

by Jones: inflammation of the joints (arthritis), heart (carditis), central nervous system (chorea), skin (erythema marginatum), or subcutaneous nodules. ARF can lead to more severe manifestations such as congestive heart failure and Sydenham's chorea. ARF is an auto-immune disease and probably results as cross reaction between streptococcal components and host tissue [4]. While incidence of the disease has decreased in the developed world, it is still prevalent in developing countries, which makes it a leading cause of heart disease in children and young adults worldwide. As an episode of ARF predisposes patients to recurring attacks, and these recurring attacks are the major cause of death and disability from rheumatic heart disease, patients are given prophylactic antibiotic therapy [14].

Acute Post Streptococcal Glomerulonephritis

Acute post streptococcal glomerulonephritis (APSGN) is the acute inflammation of the renal glomeruli following a streptococcal infection. ASPGN symptoms develop 1 to 4 weeks after a streptococcal infection, and both pharyngeal and dermal strains can lead to glomerulonephritis [4]. Symptoms include edema, hypertension, hematuria, urinary sediment abnormalities, and decreased serum complement levels, with little fever. ASPGN is believed to develop when streptococcal antigen-antibody complexes deposit on the kidney glomeruli, leading to inflammation. Diagnosis is by urinalysis and treatment focuses on reducing blood pressure and edema. The mortality rate is very low; fewer than 0.5% of patients die from the initial disease and fewer than 2% die or progress to end stage renal disease [15].

Vaccination

Currently there are no vaccines against GAS infection. An effective vaccine would help protect millions from streptococcal infections each year, and are an important response to the increasing frequency of invasive disease. An ideal vaccine would target children between 5 and 15, contain a conserved GAS epitope, be highly immunogenic, induce both IgG and IgA and not provoke cross reactions with human tissues [16]. Vaccine development has focused on the N-terminal of the M protein since Lancefield had shown that antibodies against this region are both protective and bactericidal [17]. However for this strategy, even the newest 26-valent vaccine would be limited in the serotypes it can protect against, and some antibodies against M protein are cross-reactive with human tissue. Alternative vaccine strategies have focused on other antigens such as C5a peptidase, the group specific carbohydrate, and the pyrogenic exotoxins [16]. However, at this time none of these vaccine candidates have made it to clinical trial.

Virulence Factors of GAS

GAS utilizes many virulence factors that allow it to colonize the host, invade surrounding tissues, evade the immune system and disseminate throughout the body. Many of these virulence factors are surface associated, while others are secreted into the environment (Figure 1). Important cell associate factors include: M and M-like proteins, lipoteichoic acid, MSCRAMMs, streptococcal collagen like protein, C5a protease, capsule, serum opacity factor, streptolysin S, protein G-related α_2 -macroglobulin-binding protein (GRAB), *S. pyogenes* cell envelope protease (SpyCEP). Important secreted factors include: streptolysin O, streptococcal pyrogenic exotoxin B (SpeB), streptococcal

inhibitor of complement, streptokinase, *S. pyogenes* NAD-glycohydrolase, immunoglobulin G-degrading enzyme, the superantigens (pyrogenic exotoxins) and DNases. A description of the Mga-regulated virulence factors follows.

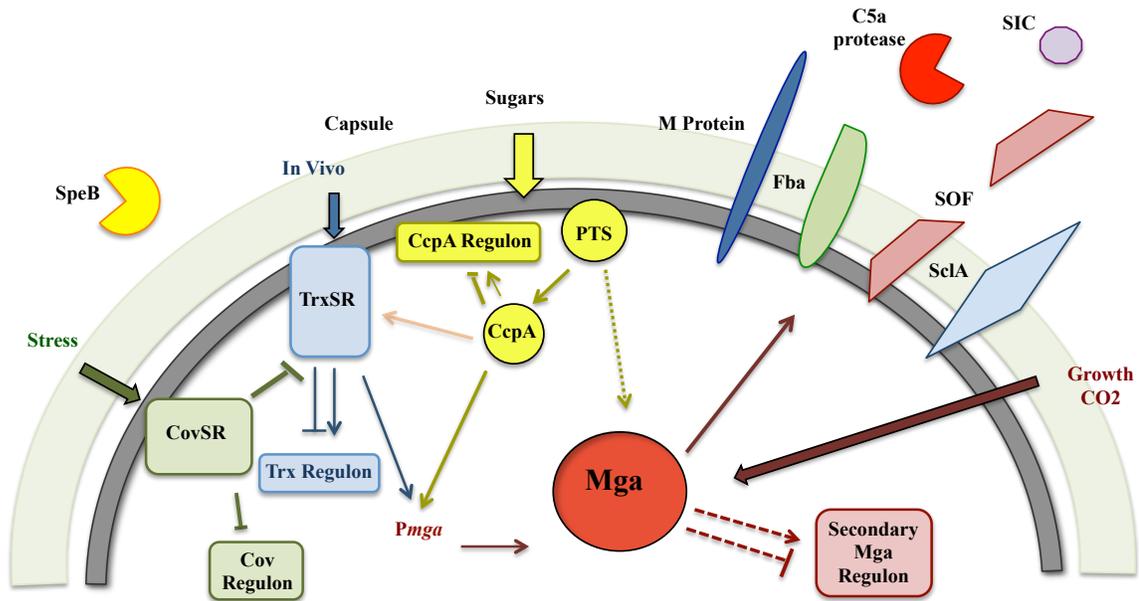


Figure 1 Regulation of Mga-associated virulence factors in GAS

Overview of the complexity of regulation of virulence genes within a GAS cell. Factors that regulate Mga include CovRS, TrxRS, CcpA, Hpr and RivRX. The directly Mga-regulated genes include M Protein, Fba, C5a protease, SIC, SOF and SclA. Capsule and SpeB are indirectly Mga-regulated in a few specific GAS strains.

Cell Associated

M and M-like Proteins

M protein is the major surface protein that is involved in adherence and immune evasion. The M-protein, encoded by *emm*, is part of a gene superfamily that encodes for many structurally similar, or M-like proteins [18]. These M-like or M-related proteins have a highly conserved domain structure within the cell wall associated region of the protein and include the genes *mrp*, *arp*, *emm*, *fcrA*, *sir*, *enn* and *sph*. A GAS strain may

have 1 to 3 of these genes, which are arranged on the chromosome in five different patterns, A-E [4]. Each serotype falls into one of these patterns, which are also strongly associated with tissue tropism. A-C are throat strains, D are skin strains while E infects both.

The M-protein is an α -helical coiled-coiled dimer that contains four different repeat regions, (A-D) and is anchored to the cell wall by a LPXTGX motif. The A region is hypervariable and confers serotypes specificity. The size of the protein can vary due to the number of repeats in the A and B regions [19].

M-protein functions to protect GAS from phagocytosis by the host. It performs this function by binding the complement regulatory protein H to interfere with opsonization of the cell and to fibrinogen [4]. The M-protein is also important for its ability to adhere to many different molecules, and different M proteins have different binding profiles. The A repeats may bind to plasminogen, IgA, IgG, human C4b-binding protein and factor H. The B repeats, which are semi-hypervariable; also bind to human serum albumin and fibrinogen. The M-protein may at times be released from the cell, where it acts a superantigen that activates T-cells and inflammatory responses during invasive diseases [20].

MSCRAMMS

The MSCRAMMs, microbial surface component recognizing adhesive matrix molecules, are many different proteins that allow for attachment and adherence to structures on the surface of eukaryotic cells or with the extracellular matrix. SfbX [21] and Fba [22] are two such proteins that are Mga-regulated. These are two of the many fibronectin-binding proteins that are encoded for by GAS.

Streptococcal collagen-like protein

Streptococcal collagen like protein, SclA, encoded for by *sclA*, is structurally similar to collagen and varies in length. SclA is important for adhesion during GAS pathogenesis; mutants have shown attenuation for virulence and a decrease in epithelial cell adherence [23]. SclA can inhibit the complement pathway by binding factor H [24]. SclA is capable of binding TAFI, a fibrolysis inhibitor, and modulate the inflammatory reactions through the recruitment of plasmin to the surface of the GAS cell [25].

C5a protease

Streptococcal C5a protease, *scpA*, is a surface-associated and anchored endopeptidase that cleaves the C5a chemotaxin of the complement system [26]. This action inhibits the recruitment of phagocytic cells to the site of infection, and enables GAS to evade the immune system [27].

Capsule

The capsule of GAS is composed of hyaluronic acid, and its production is encoded for by the *has* operon. Capsule is chemically identical to the hyaluronic acid found in human connective tissue, which helps GAS evade the immune system through molecular mimicry though not all strains are encapsulated [28]. Capsule has been shown to be indirectly Mga-regulated in a M1 and M6 strain [29].

Serum Opacity Factor

Serum opacity factor (SOF), *sof*, is only produced by Class II serotypes. SOF is an important adhesion that helps bind streptococcal cell to high-density lipoprotein via a fibronectin-mediated process [30]. SOF may be bound to the surface or released from the cell. Due its two functions, serum opacification and fibronectin/fibrinogen binding, it has been difficult to determine the role of SOF in virulence [31,32]. However, inactivation of *sof* has been shown to reduce virulence in an intraperitoneal mouse model of infection [31].

Secreted

Streptococcal pyrogenic exotoxin

Streptococcal pyrogenic exotoxin B, SpeB, is a cysteine protease with a somewhat ambiguous role in GAS. SpeB is indirectly Mga-regulated in M1, M2, M3 and M49 strains [29], however a plethora of other factors also affect the activity of this protein. SpeB is important for lethality in mouse models [33], resistance to phagocytosis [34] and fibrinogen cleavage [35], though there are studies were no effect on virulence were observed [36,37]. SpeB is thought to promote the prevention of complement activation though cleavage of streptococcal IgG-binding proteins, leading to survival of GAS [38,39] and cleaves the fibrinogen binding proteins that attach GAS to the host to promote the spread of infection [40,41].

Streptococcal Inhibitor of Complement

Streptococcal inhibitor of complement, *sic*, inhibits complement-mediated lysis by incorporation into the membrane attack complex [42]. Sic interacts with many components of host cells and the immune system to aid in GAS infections, these activities include: inactivation of LL-37 and human neutrophil α defense, which are two antibacterial peptides [43], aid in survival at mucosal surfaces [44] and alter the activities of lysozyme and secretory leukocyte proteinase inhibitor [45].

Regulation

In order to adapt to the difference niches within the host, GAS encodes for many different regulators, which include on average 13 two-component systems, and other regulators. Five of the two-component systems are involved in virulence gene regulation: CovRS, FasBCAX, Ihk/Irr, TrxRS and SptRS. Other important regulators include CcpA, the Rgg/RopBs, the RALPs (RofA-like proteins), MtsR, and Mga. To give an overview of the complexity of regulation within GAS, a description of the regulators that effect/are affected by Mga follows.

Two component systems

CovRS

CovRS, which stands for control of virulence, is the best described two-component system in GAS and extensive data has shown its importance in modulating gene expression during virulence. CovRS was originally identified by its role in capsule synthesis regulation, when it's name was CsrRS, but studies have since shown that this

system plays a wide spread role in regulation. CovRS directly or indirectly regulates ~15% of the genome, mainly through repression, and thus it is a major negative regulator [46]. CovRS is believed to respond to changing pH, temperature and osmolarity in the environment [47]. Spontaneous mutations that truncate CovS allow GAS to become more invasive [47–49] and can be selected for by passaging GAS through animals. Microarray analysis comparing wild-type and the *covS* mutant show marked differences in the transcription of many genes that are associated with virulence [50]. The regulation of virulence within GAS is complex, and as the master regulator within GAS, CovRS interacts with Mga in multiple, indirect ways by repressing the activity of CcpA, TrxRS and RivR, three other regulators that it directly represses [51–54].

TrxRS

TrxRS, two-component regulatory system X, is another two-component system that is directly linked to GAS virulence, and modulates Mga expression. *In vivo* murine studies have shown that a *trxR* mutant is attenuated for virulence and directly repressed by CovR [51]. TrxR was also shown to activate the core Mga regulon, and studies done in the lab have shown that TrxR can directly bind to *Pmga* [55].

Other Regulators

CcpA

CcpA, the carbon catabolite protein A, is the master regulator of carbon catabolite repression (CCR) that controls the use of carbon sources within a cell. When there is abundance of the preferred sugar glucose, CcpA complexes with P~S46-Hpr and binds to catabolite operons *cis*-acting catabolite responsive element (*cre*) sites to prevent gene

transcription [56]. In many bacteria, including GAS, CcpA also functions as a link between sugar metabolism and virulence [57,58] and can affect the expression of several virulence factors, such as repressing streptolysin S. Studies done in lab identified a *cre* site in the *Pmga* promoter and demonstrated that CcpA plays a role in activating *mga* transcription [52].

Rgg/RopB

Rgg or RopB for regulation of proteinase, is part of a family of transcriptional regulators that regulate the expression of extracellular products during stationary phase in a strain dependent manner [59]. Rgg is also a global transcriptional regulator which controls virulence, secondary metabolism and stress, and transcription factors, including repressing Mga expression [60]. Recently, members of the Rgg family have been shown to respond to small peptides, and one of these members, ComR, turns on competence genes in GAS, though sadly, successful transformation has not been demonstrated in the laboratory [61].

RALPs

The RALP (RofA-like protein) family contains four regulators that act at the transition between logarithmic and stationary phase growth, and modulates virulence genes. This family is composed of RofA, regulator of F, Nra, negative regulator of GAS, RALP3, and RivR, RALP iv. These regulators show strain specific activity; Nra represses pilus synthesis in an M49 strain but activates it in an M53 strain [62–64]. Furthermore, not all GAS strains will have each regulator, RALP3 has been identified in only a few of the sequence serotypes [65]. RofA and Nra [66,67] act to repress Mga,

while RivR directly binds to the *Pmga* promoter to activate *mga* expression [54]. This protein family is also notable, because similarly to Mga, a Phyre2 protein fold recognition search identifies putative PRD domains and an EIIB domain within the protein structure.

Mga

History/discovery

Mga, the multiple gene regulator of GAS, is a ubiquitous transcriptional regulator. The first evidence of Mga was found in a M12 strain lacking the expression of the M protein that was determined to have a deletion upstream of *emm* in a gene named *virR*, virulence regulator [68]. Separate experiments using Tn916 insertional mutagenesis of this upstream gene and its promoter led to decreased *emm* mRNA, providing evidence that the upstream region encoded a trans acting regulatory factor that was named Mry, M protein RNA yield [69]. VirR and Mry were determined to be homologous and renamed Mga [70]. This regulatory protein was further shown to positively control a core virulence regulon containing streptococcal C5a protease (ScpA), serum opacity factor (Sof) and type IIa IgG Fc receptor (FcRA) [71].

Regulon characteristics

Transcriptome analysis of the Mga regulon showed that it included over 10% of the genome during exponential growth by activating some genes while repressing others [72]. The core regulon is composed of a small number of activated genes involved in adhesion, internalization, and immune evasion and in a few serotypes, auto regulation.

The larger secondary regulon is composed of genes that have low levels of activation or repression, and are probably indirectly regulated. Some of these secondary genes are involved in sugar utilization, while others are part of metabolic operons. A number of secondary genes are also Mga regulated in a strain or serotype specific manner [72].

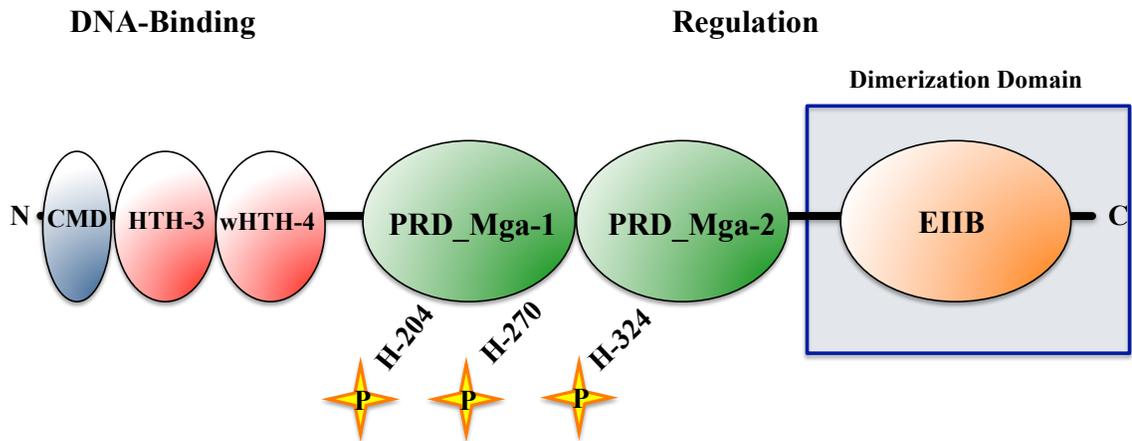


Figure 2 The domains of the Mga protein

The Mga protein has six described domains. At the N terminus is the CMD, a HTH and a wHTH, two PRD_Mga domains, and an EIIB-like domain. The dimerization domain is marked by the blue box. The phosphorylated histidines are marked with a star [73].

Protein characteristics

There are two alleles of Mga but each serotype has only one. The class 1 Mga allele is found in strains with the A-C chromosomal organization, which are SOF negative, while the class 2 are found with the D and E chromosomal arrangements and are SOF positive [74]. There is a 97% amino acid sequence identity within each allele and a 21% amino acid sequence difference between the two alleles. Most of this variation lies within the C-terminus. Studies so far have shown that these two alleles function in a similar manner [73] (Hondorp, 2012, unpublished). Mga is a 62-kDa protein that contains six predicted domains (Figure 2). At the N-terminus there is a conserved

Mga domain (CMD) with unknown functions, though mutations within this domain lead to transcriptional defects [75], a helix-turn-helix domain, HTH-3 and a winged helix-turn-helix domain, wHTH-4, that allow for direct transcriptional regulation. Mga contains two phosphotransferase regulatory domains that are located near the center of the amino acid sequence, PRD-Mga 1 and PRD-Mga 2, which link regulation of the protein to the sugar status of the cell. A EIIB^{GAT}-like domain is located at the C-terminus and is necessary for dimerization and transcriptional activation within a GAS cell [73].

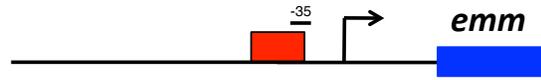
DNA binding

Mga directly activates its core primary genes through DNA binding and probably indirectly regulates the secondary genes of the regulon [29]. A binding site consensus sequence for Mga was formed by biochemical analysis of core Mga regulated promoters in the M6 strain JRS4. DNase I footprint analysis of the *emm* and *scpA* promoters determined a 45 base pair (bp) binding site centered at -54 relative to the start site of transcription; this is about twice as large as other known prokaryotic transcriptional regulators [76]. In addition, this binding site does not contain any internal symmetry, suggesting that only a single binding site is present. When the *Pmga* promoter was analyzed, two binding sites of 59 bps were identified, centered at -104 and -185 bps upstream of the start of transcription [77]. Analysis of the *PscIA* (streptococcal collagen like protein) promoter found two 45 bp binding sites based on the previously determined consensus sequence, one centered at -54 and the other at -175; however only the distal binding site was active *in vivo* [78]. While all of these sites have been shown to interact with Mga by DNaseI footprint or EMSA, alignments show less than 50% sequence identity. By using this “Mga binding site” (MBS) consensus sequence, additional

binding sites can be found in other GAS strains, however with each new discovered site added, the percent identity decreases. Three types of Mga regulated promoters have been categorized based on the location and number of MBSs (Figure 3).

Category A:

Pemm
(45 bp, centered at -54 bp)



PscpA
(44 bp, centered at -54 bp)



Category B:

PsclA
(~45 bp, centered at -175 bp)



Psof
(~45 bp, centered at -287 bp)



Category C:

Pmga
(Two 59 bp sites, -104 and -185 bp)

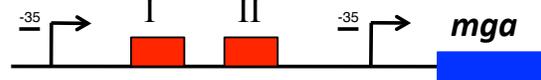


Figure 3 Architecture of Mga Regulated Promoters

There are three categories of Mga-regulated promoters divided according to the size, number and location of the Mga binding site [78].

The first type of promoter, category A, contains one binding site centered at -54 bps upstream and overlapping the promoter. The category B type of promoter contains a single binding site centered either -175 or -287 bps upstream from the start of transcription. The final type of promoter, category C, is *Pmga*, which has two 59 bp binding sites located upstream from the transcriptional start site [78].

Two HTH DNA binding domains within the N terminus of Mga were predicted from the amino acid sequence. Each was inactivated by mutagenesis and assessed for DNA binding activity. These experiments determined that the DNA binding activity of

all Mga regulated promoters was dependent on wHTH-4 while HTH-3 appears to be an accessory for binding at *Pmga* [78]. Sequence alignments show a 100% amino acid sequence conservation of wHTH-4, while HTH-3 is also identical between Mga of the same allele, but only 90% conserved between the alleles [79]. Furthermore Mga from one allele can also complement Mga from the other allele *in vivo* [80]. This data suggests that the great variation in Mga binding sites is not due to the protein but instead arises from the interactions with the nucleotides of the binding site.

Regulation

The signal that controls Mga expression is as of yet unknown, but experiments done in the lab have shown that the PTS, phosphoenolpyruvate transferase system, is an important part of this regulation. Mga contains two PRD domains that most closely resemble the mannose operon activator MtlR from *G. stearothermophilus*. Mga possesses three conserved histidines, 2 in PRD-1 and 1 in PRD-2 at residues 204, 270 and 324 (Hondorp, 2012, unpublished). By qRT-PCR of the Mga-regulated genes *arp* and *sof*, a double phosphometric mutant (H204D/H270D) Mga has greatly reduced activity. The double alanine mutation also has a small but significant decrease in Mga activity. By *in vitro* phosphorylation where the PTS system was reconstituted, a Mga triple alanine mutant (H204A/H270A/H324A) has a dramatic decrease in phosphorylation compared to wild-type.

A second key component of Mga regulation is the dimerization state of the protein. Immunoprecipitation experiments have shown that the inactive Mga D/D protein is also unable to form homodimers, though this protein binds DNA comparable to wild-type (Hondorp, 2012, unpublished). At the C-terminus Mga has an EIIB^{GAT}-like domain.

Removing this domain prevents dimerization of the protein, and transcriptional activation, although this protein also binds to DNA at wild-type levels (Hondorp, 2012). EIIB domains are phosphorylated on a cysteine residue, which serves as another potential regulatory site, however in Mga this residue is a glutamic acid, but the importance of this change is unknown. However it is clear that simply binding DNA is not sufficient for Mga activity, phosphorylation and dimerization are key elements to Mga's function.

DNA Binding Proteins

Functions

DNA binding proteins perform many functions in the cell and are essential for replication of the chromosome, repair of DNA, packaging of the chromosome, and regulation of gene expression. DNA-protein interactions can be general or specific to particular DNA sequences. The primary means by which DNA-binding proteins identify their target DNA is through sequence preference.

Initial hypotheses of how DNA and proteins interacted predicted that there would be a code that aligned amino acids to DNA sequences: experiments have shown that these hypotheses were far too simplistic. Many families of DNA binding domains have been identified that interact with a huge variety of potential DNA sequences. While there is no code for which side chains recognize which bases, these proteins do follow a general set of rules that allow for site-specific recognition [81]. While hydrophobic interactions do occur, hydrogen bonding is critical, between 1 and 2-dozen hydrogen bonds form along a protein-DNA interface. Structure is a key component of DNA recognition, both the folding of the protein itself and DNA sequence specific structures. The site-specific recognition occurs through contact with both the phosphodiester backbone and

the bases. These contacts occur mostly through interactions with the amino acid side chains and the bases within the major groove. Most major DNA binding motifs have an α -helical region that fits into the major groove of B-DNA, though β sheets and extended regions of polypeptide chain may play a critical role in some proteins. Purine contacts are particularly important as they are larger and have more hydrogen-bonding sites. Hydrogen bonds and/or salt bridges are generally formed with the phosphodiester oxygen in the DNA backbone. Finally, usually multiple binding domains are often necessary for site-specific recognition [81].

DNA Binding Domains

Helix-turn-Helix family

When the first crystal structures of the λ cro protein, *E. coli* CAP and the DNA-binding domain of the λ repressor were compared, all three proteins contained a similar helix-turn-helix (HTH) motif and was the first identified DNA binding domain [81]. The HTH family is widespread in both prokaryotes and eukaryotes, where it forms the basis of the homeodomain. At its simplest a HTH motif is composed of two short α helices which are oriented nearly perpendicular, connected by a three amino acid linker, or turn [82]. The first helix, or preceding helix, sits along the phosphodiester backbone, while the second helix, or recognition helix, fits inside the major groove to contact the bases. However, it would be a mistake to focus only on the recognition helix to understand the HTH's protein-DNA contacts, as both the preceding helix and polypeptides outside the domain play a role in recognition.

The HTH is not separate stable motif and depends on its surrounding domain to

give it shape and function [81]. There are many subfamilies within this group that are characterized by the structures that flank the HTH. This motif can be associated within different structures; in the CAP protein, the HTH is held by β -sheets, while in the λ repressor it is associated with α -helices, leading to many variations on the basic form. One common variation is the winged HTH (wHTH), where the HTH is followed by 1 or 2 β -hairpin turns [83]. The canonical wHTH has wings, which are extended loop structures, three β strands and three α helices, in the topological order Helix1, β -strand1-Helix2-turn-Helix3- β -strand2, Wing1- β -strand 3-Wing2 [83]. Helix 2 and 3 make up the preceding and recognition helices, respectively, and while helix 3 makes most of the site-specific DNA contacts the extended structures play a part in DNA recognition.

Helix-loop-helix

The helix-loop-helix (HLH) contains two domains, a dimerization domain and a DNA-binding domain [81] and has some similarities to the leucine zipper motif. The HLH dimer is a left-handed, four α -helical bundle with a loop connecting each dimer's α -helices [84]. The basic region of each dimer is inserted into the major groove where it contacts a DNA hexamer that commonly has the sequence CANNTG [84]. As HLH proteins may form both hetero and homodimers great control over gene activity can be exerted [81].

Zinc Finger

The zinc finger motif was first identified in the *Xenopus* transcription factor IIIA (TFIIIA) [81] and is part of a superfamily common in eukaryotes but relatively rare in prokaryotes [85]. The zinc finger is not only a DNA-binding domain; it is also a protein-

RNA and protein-protein interaction domain. In the presence of zinc, the two antiparallel β sheets and an α helix fold around the zinc ion. These fingers have relatively few fully conserved residues as most structural stability is due to the coordination of the zinc ion with the conserved hydrophobic core [85]. DNA-binding usually requires 2 to 4 tandem arranged fingers; if only 1 or 2 are present other secondary structures assist with recognition. The α helix of each finger sits in the major groove, and the protein wraps around the DNA as each successive α helix binds. Each finger docks in a similar manner and contacts an overlapping four base sub site, however most base contacts are formed with three bases on one strand on the DNA [85]. An enlarged major groove is another common feature of zinc-finger protein-DNA interactions.

Leucine Zipper

The leucine zipper (LZ) DNA-binding domain was first discovered as a conserved motif in eukaryotes [81] and the structure was first determined for yeast transcription factor GCN4 [84]. Like the HLH this motif is composed of two domains, a dimerization region and the DNA-binding motif that is characterized as a heptad repeat of leucines over 30 to 40 residues and a conserved repeat of hydrophobic residues that is located to the N terminus of the leucines. Biochemical evidence suggesting that the LZ forms a structure of two parallel α helices in a coiled-coiled arrangement that resembles a fork [84]. To bind DNA a relatively straight basic region of each dimer is positioned into the major groove to contact a half-site of 8 to 10 bps [84] where it makes contacts with the bases and phosphodiester backbone. The dimerization domain contributes to binding specificity by determining which LZ containing proteins will form stable dimers, which

also allows for fine-tuning to regulation, and guides the basic region into the major groove [84].

Beta Ribbon Motif

A small family of bacterial repressors which includes MetJ, Arc and Mnt, bind to DNA using antiparallel β sheets which is called the beta ribbon motif [81]. The domain contains a β sheet with two α helices; when MetJ dimerizes the β sheets align into an antiparallel β conformation that the α helices stabilize. Each β sheet enters the major groove and binds to a half site, resulting in a tetramer. Arc and Mnt appear to behave in a similar manner. Other regulators may contain this β -ribbon motif, but not enough is known to identify them based on sequence; Arc and Mnt were not confirmed to have this motif until structural studies were performed [81].

RNA Polymerase

The Holoenzyme

The first step in gene expression is to transcribe DNA into RNA, a process that is catalyzed by RNA polymerase (RNAP). The central function of this process makes RNAP the key target of transcriptional regulation in bacteria. The idea of this enzyme was first formulated in the 1950s alongside the discovery of mRNA, and in the early 1960s Audrey Stevens and Jerard Hurwitz created cell-free extracts from *E. coli* that produced RNA [86]. This enzyme extract possessed the ability to catalyze a new RNA chain on its own, but only from a DNA template. Further studies of this enzyme were undertaken to understand its function. When Richard Burgess and his colleagues passed the RNA polymerase holoenzyme over an anion exchange column, they identified 3

peaks, one which contained the holoenzyme, one that contained the core and one that was σ factor [87]. This core enzyme is composed of $\alpha_2\beta\beta'$, which retains its transcription function, but requires an additional σ factor to transcribe from intact DNA, and composes the holoenzyme of $\alpha_2\beta\beta'\sigma$. There is an additional ω subunit that is not necessary for transcription and Gram-positive bacteria have a δ subunit of undetermined function as well [88]. While core RNAP is sufficient for elongation and termination, the σ factor is necessary for transcriptional initiation [89]. The main, or housekeeping, σ factor is called σ^A , and contains four domains that are highly conserved among bacteria [90]. In *E. coli*, and other Gram-negative bacteria, σ^A is also known as σ^{70} , based upon the size of the protein, while in Gram-positive bacteria σ^A is only 43 kDa in size, the result of a ~245 amino acid deletion between domain 1 and 2 [90].

Steps of transcription

There are three steps transcription: initiation, elongation and termination. Initiation can be further divided into four steps as well: formation of a closed promoter complex, conversion from a closed promoter complex to an open promoter complex, polymerization of short nucleotides while the polymerase remains at the promoter, and promoter clearance, when the transcript becomes long enough to form a stable hybrid with the template strand. At this point the polymerase moves into elongation conformation, and dissociates from σ factor [91]. In order to initiate transcription, the holoenzyme must first recognize and bind to promoter DNA. This function is performed by σ factor, which binds to the core. σ factor is bound to the β' subunit, and this allows domain 2.4 and 4.2 of σ factor to be exposed to solvent and positioned to bind to the highly conserved elements of promoter DNA [92]. Domain 2.4 of σ recognizes the -10

hexamer with the consensus sequence TATAaT. Domain 4.2 of σ recognizes the -35 hexamer with the consensus sequence TTGaca. The space in between, called the discriminator, is also an important determinant of promoter strength, and 17 bp is ideal. The -10 sequence has been shown to be more important for recognition than the -35 and the presence of an extended -10 (TGnTATAaT) can compensate for a weak or absent -35 site [89]. The similarity to the consensus sequence and spacing of the -10 and -35 both contribute to promoter strength. σ factor plays other important roles as a target for transcription activators, it assists in melting the promoter near the transcription start site, it inhibits non specific interactions and serves to clear and release RNAP from the promoter [93].

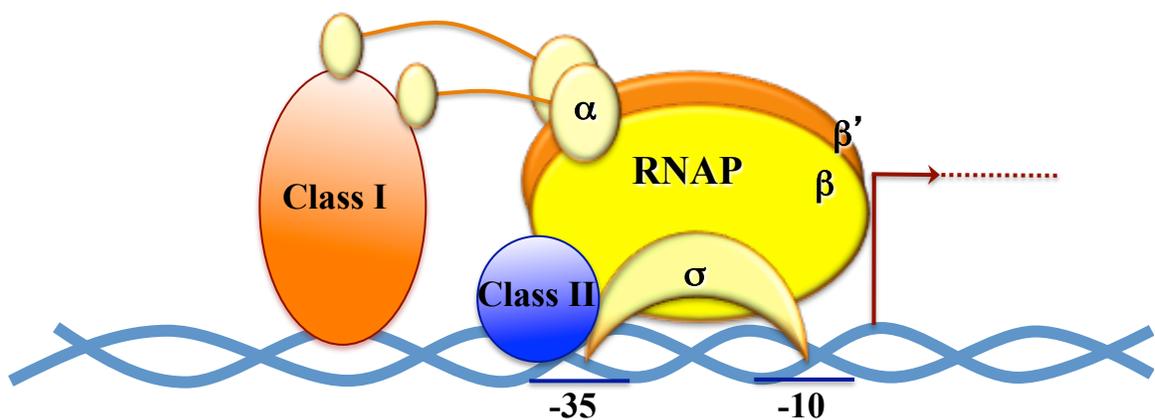


Figure 4 Class I and Class II Activators

Class I transcription factors (orange) are found bound to DNA upstream of the promoter and make protein-protein contacts with the C terminal domain of the α subunit. The Class II transcription factor (blue) binds adjacent to and overlapping the -35 site to stabilize σ factor at weak promoters.

Activation and Repression

Within a cell there is limited RNAP and σ factors available. In bacteria the most important target for gene regulation is the initiation of transcription and RNAP,

specifically through interactions with σ and the α subunit [94]. Bacteria control the initial of transcription by interactions with two main targets on the holoenzyme [95]. In Class I activation the transcription factor binds upstream of the promoter and interacts with the α -CTD (C terminal domain) in order to activate transcription (Figure 4). The α -CTD is a readily accessible target and its flexibility allows RNAP to interact with transcription regulators up to 100 bps upstream. It also has potential non-specific DNA binding activity that could be stabilized by protein-protein interactions to enhance gene expression. In Class II transcriptional activation, Class II factors bind to sites overlapping the -35 and interact with domain 4 of σ instead of the α -CTD [96]. Protein-protein interactions between the transcription factor and σ domain 4.2 stabilize DNA binding to overcome the limitations of a weak consensus -35 site [97]. As σ association with the core is necessary for promoter recognition, anti- σ factors are also used to control when a particular σ is available, allowing the cell to respond to changing environmental signals. While many bacteria have a variety of additional alternative σ factors to control gene expression, GAS has only one other σ factor, which is not connected to Mga [98].

Transcription factors may also act independently of RNA polymerase. The MerR family of transcription factors activate transcription by binding to the DNA ([99]. These promoters have a non-optimal 20 bp space between the -10 and the -35. The MerR proteins bind to this spacer, and through DNA distortion reposition the -10 and -35 and allow transcription to occur [100,101]. Overall, the cooperative effect of weak protein-protein interactions can lead to strong stimulation, or repression, of gene expression but only occur at the right promoter and promoter orientation.

Transcription factors may also act indirectly through the relief of repression. In the cell DNA is tightly compacted and when factors bound to the DNA such as HNS, histone for nucleoid structuring, IHF, integration host factor and FIS, factor for inversion stimulation, are disrupted, derepression of transcription occurs [102].

Repression of transcription can occur in several different manners. The simplest of these ways is where the repressor sterically hinders RNA polymerase from interacting with the promoter or interferes with its recruitment [95]. A repressor may bind to distal sites that create a DNA loop that then blocks RNA polymerase. Repressors can also act to modulate other transcription factors, for example in *E. coli* the protein CytR binds to CRP, preventing it from activating transcription [95].

The modulation of expression for most genes is regulated by multiple signals. Multiple transcription factors may activate or repress to regulate a given promoter. Four general mechanisms have been described for these interactions [95]. In the repositioning mechanism, the binding of a second activator can shift the first activator, which allows transcription to then occur. Alternatively, the second activator may bend the DNA, allowing the first activator to interact with RNA polymerase. Two transcription factors can make two independent contacts with RNA polymerase. This may be two class I activators interacting with each α subunit, or one class I and one class II activator stabilizing RNA polymerase through contacts with the α and σ subunits. In the cooperative binding method, the binding of one activator is dependent on binding a second activator. In the antirepression method, an antirepressor binds to the repressor, and allows another activator to then promote transcription.

Phosphoenolpyruvate phosphotransferase system (PTS)

Introduction

The phosphoenolpyruvate phosphotransferase system (PTS) is a widely conserved pathway by which bacteria transport and utilize sugars. As GAS does not possess a functional TCA cycle, the PTS is the primary means by which GAS ferments energy. Furthermore the PTS has been linked to virulence in GAS and other Gram-positive bacteria.

The PTS was first identified in *E. coli*, by Kundig, Ghosh and Roseman [103]. This system uses PEP to phosphorylate and transport various hexoses by a two-step reaction catalyzed by the general protein EI to a sugar specific EII protein using Hpr as an intermediate phosphor donor. PTS-regulation of the cell is accomplished through these changes in phosphorylation state.

Components

The PTS system widely conserved in bacteria and is composed of the general proteins EI and Hpr, and many sugar specific membrane bound EIIs. EI, which is encoded by *ptsI*, is a protein of about 63 kDa shows significant sequence identity between Gram-positive and Gram-negative bacteria [103]. EI has two domains, an N terminal phosphorylation domain, and a C terminal PEP binding domain, which is also necessary for dimerization. In the presence of Mg^{2+} , EI auto phosphorylates at the N-3 position of the imidazole ring of a conserved histidine (His-15 in GAS) [103]. Hpr, encoded by *ptsH*, is a protein of only 90 residues. In most enteric bacteria and firmicutes, the His-15 is phosphorylation on the N-1 residue of the imidazole ring. However in some

Gram-negative and low G+C Gram-positive bacteria, which includes GAS, Hpr can also be phosphorylated by HprK, Hpr kinase, on Ser-46 [103]. This phosphorylation site is not part of the phosphotransfer, but is a regulatory site which can reduce the transfer rate of P~EI to Hpr 100 fold. EIIs are the sugar specific transport and phosphorylation proteins of the PTS. The EII consists of an integral membrane domain that faces both the periplasmic and cytosolic space and may be composed of up to four separate proteins. There are four superfamilies with distinct evolutionary origins as determined by phylogeny that the various EIIs can be divided into: the glucose-fructose-lactose super family, the ascorbate-galacticol superfamily, the mannose family, and the dihydroxyacetone family [103].

Regulation of/by the PTS is accomplished by two parallel and overlapping pathways, carbon catabolite repression (CCR) by CcpA/Cre and by phosphorylation of phosphotransferase regulatory domain (PRD) containing proteins. During CCR, the Hpr kinase phosphorylates Hpr on serine residue 46, which makes it a substrate to bind CcpA [56]. The P~Ser-Hpr-CcpA complex then binds to *cre* sites to control the expression of secondary sugar operons. PRD containing transcriptional regulators, which many be activators or antiterminators, are very common in Gram-positive bacteria [104]. With the exception of CsiE from *E. coli*, all known PRD containing proteins have a duplication of the PRD where the N-terminal PRD is PRD-1 and the C-terminal PRD is PRD-2. A classic PRD containing protein has an essential histidine and a conserved arginine spaced 7 residues downstream, a strongly conserved glutamate at amino acid 57, and often a conserved histidine at residue 63. Each PRD contains 1 or 2 histidines that serve as targets of phosphorylation by P-His-Hpr. The phosphorylation effects the dimerization of

the proteins, and which histidine residue that is phosphorylation effects the activity of the protein; the phosphorylation of one PRD may activate, while the phosphorylation of the other PRD may repress [104].

PTS and Virulence

In Gram-positive bacteria experiments have shown a link between carbon metabolism and virulence. The importance of CcpA in GAS virulence has been established (Kinkel, 2008, Shelbourne, 2008). Our lab has also shown that a *ptsI* (EI) mutant has a hypervirulent phenotype at the site of infection in a murine skin model (Gera, unpublished). Phosphomimetics of the PRD domains in Mga lead to the down-regulation of Mga regulated genes *in vivo*, and are attenuated for virulence in skin model mouse infection (Hondorp, et al, in review). However, the PTS has not been directly shown to control Mga *in vivo*.

Summary

The experiments performed in the following studies were undertaken in order understand how Mga interactions at the promoter allow it to function as a transcription factor. We first dissected the protein-DNA interactions between Mga and a model category A promoter (*Pemm*) to understand how this process occurs, and then used this as a model for Mga interactions at other binding sites. To understand how Mga interacts with RNA polymerase, we studied protein-protein interactions between the α -CTD or domain 4 of σ factor. Studies were performed to determine the dimerization of the protein in solution. Finally *In vitro* phosphorylation-transcription assays were performed to link the PTS system to the regulation of the Mga protein.

Chapter 2

Materials And Methods

Bacterial strains and media

Bacterial strains and plasmids used in this study are shown in Table 1. GAS strain MGAS5005 (*covS*) is a well-characterized M1T1 invasive strain that has a sequenced genome available [105]. GA40634 is a M4 strain containing a different *mga* allele than MGAS5005. KSM547 is the Δ *mga* derivative of GA40634. One Shot® TOP10 Electrocomp™ *E. coli* (Stratagene) was used for site directed mutagenesis cloning. *E. coli* DH5 α was used for plasmid construction. *E. coli* C41 [DE3], a derivative of BL21[DE3], was used for protein expression [106]. *E. coli* BTH101 and DHM1 were used for the bacterial 2 hybrid assays. *E. coli* was grown in Luria-Bertani (LB) for plasmid construction. *E. coli* was grown in ZYP Auto-induction media [107]. GAS was cultured in Todd-Hewitt medium supplemented with 0.2% yeast extract (THY) and growth was assayed by absorbance using a Klett-Summerson photoelectric colorimeter with the A filter. Antibiotics were used at the following concentrations: ampicillin at 100 $\mu\text{g mL}^{-1}$ for *E. coli*; spectinomycin at 100 $\mu\text{g mL}^{-1}$ for *E. coli* and GAS; and kanamycin at 50 $\mu\text{g mL}^{-1}$ for *E. coli* and at 300 $\mu\text{g mL}^{-1}$ for GAS.

DNA manipulations

Plasmid DNA was isolated from *E. coli* using the Wizard® Plus SV Miniprep system (Promega). DNA fragments were gel purified from agarose using the Qiaquick Gel Extraction kit (Qiagen) or the Wizard® SV Gel and PCR Clean-Up system (Promega). PCR for cloning and generating probes was performed using *Taq* DNA

polymerase (New England Biolabs, NEB). PCR for site directed mutagenesis was performed using *Pfu Ultra* HF DNA polymerase (Stratagene). DNA sequencing was performed either using the SequiTherm Excel™ II DNA Sequencing kit (Epicentre, Inc.) or by Genewiz, Inc.

Construction of Luciferase plasmids

Pemm was amplified from genomic DNA using M1 *Pemm* L and M1 *Pemm* R. *PscpA* was amplified from genomic DNA using M1 *PscpA* Bam L and M1 *PscpA* Xho R. PCR products were ended filled using T4 polymerase (NEB) and blunt ligated into Zero Blunt® TOPO PCR Cloning Kit (Invitrogen). *Psic* was amplified from genomic DNA using the primers M1 *Psic* Luc BglII and M1 *Psic* Lux XhoI. *PsclA* was amplified from genomic DNA using SF370 *PsclA* BglII and SF370 *PsclA* XhoI, and *PsclA* without MBS1 was amplified from genomic DNA using SF370 *PsclA* w/o MBS1 BglII and SF370 *PsclA* XhoI. Mutagenic oligonucleotide pairs (Table 2) were synthesized to introduce point mutations into the Mga binding site using the QuikChange® Site Directed Mutagenesis Kit (Stratagene). Mutations were verified by DNA sequencing. pKSM720 (Table 1) was digested with BglII and XhoI and gel purified. Each insert was digested with BamHI and XhoI and gel purified. The inserts were ligated into pKSM720. Plasmids were verified by DNA sequencing and transformed into MGAS5005.

Luciferase Assay

Luciferase assays were performed as described previously [57]. MGAS5005 containing each luciferase plasmid were grown in 13 mLs Todd-Hewitt with spectinomycin at 37°C. Upon reaching Klett 20, 500 µL samples were taken

approximately every 15 Klett units until into stationary phase to assess activity across growth. At least three replicates were sampled at Klett 80 (mid-logarithmic phase) to compare percentage luciferase activity of each point mutation to wild-type. Samples were pelleted, supernatant was discarded and the samples were stored at -20°C overnight. The luciferase assay was performed using the Luciferase Assay System (Promega). The samples were resuspended in 1x lysis buffer (Promega) to normalize them to cell units according to the equation $4.5 = (x \text{ mL})(65 \text{ Klett units}/2)$. The luciferase assay was read using a Centro XS3 LB 960 luminometer (Berthold Technologies), into which 50 µL of Luciferin-D reagent was directly injected.

Expression and Purification Mga-His proteins from *E. coli*

Mga1-His₆ and Mga4-His₆ was purified as described previously [73]. *E. coli* C41(DE3) containing the plasmid pMga1-His or pKSM801 were grown in ZYP Auto-induction media for ~62 hours at 37°C. Cells were harvested by centrifugation at 4°C. The pellet was resuspended in NiNTA Lysis Buffer (50 mM NaH₂PO₄, 300 mM NaCl, 10 mM imidazole, pH 8.0), then incubated on ice with lysosyme for 30 minutes followed by sonication using a Branson sonifier 450 with a tapered microtip (setting 6, 50% duty cycle) pulsing 6 x 30 seconds with 1 minute breaks on ice. The lysate was spun for clarification at 12000 rpm 3-4 times, then passed through a 0.45 µM syringe filter. The lysate was loaded on a 750 µL NiNTA agarose column (Qiagen), washed 20, 50, 70 and 90 mM imidazole Wash Buffer and eluted with 250 mM Imidazole Elution buffer. Protein was detected by Coomassie staining. Fractions were dialyzed overnight at 4°C into 50 mM HEPES Citrate pH 7.5 with 50 mM EDTA. EDTA was washed out with 50 mM HEPES Citrate pH 7.5 and determined to be EDTA free by analysis with 4-(2-

pyridylazo)resorcinol (PAR). 100 μ L of flowthrough was mixed with 10 μ L 0.1 M ZnSO₄ and 1 μ L 10 mM PAR. When EDTA is removed, the solution changes from yellow to red. Protein concentration was analyzed by absorbance at 280 nm with the extinction coefficient of ϵ_{280} of 59650 M⁻¹ cm⁻¹ and Coomassie staining.

Expression and Purification of Mga4-CBP from *E. coli*

E. coli C41[DE3] containing pKSM289 were grown under the same conditions as pMga1-His and pKSM801. Pellets were resuspended in 1-4 volumes CaCl₂ Binding Buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 10 mM β -mercaptoethanol, 1.0 M MgAcetate, 1.0 mM imidazole, 2 mM CaCl₂) with 1x protease inhibitor (Roche). The cells were then incubated on ice with lysosyme for 30 minutes followed by sonication using a Branson sonifier 450 with a tapered microtip (setting 6, 50% duty cycle) pulsing 6 x 30 seconds with 2 minute breaks on ice. The lysate was spun for clarification at 12000 rpm 3-4 times, then passed through a 0.45 μ m syringe filter. The lysate was loaded onto a 1 mL Calmodulin Resin column and washed 2 times with 10 mLs CaCl₂ binding buffer. Mga4-CBP was eluted in 8 mLs CBP elution buffer (50 mM Tris-HCl, pH 8.0, 10 mM β -mercaptoethanol, 2 mM EGTA, 150 mM NaCl). Fractions were then dialyzed overnight against 4 L 50 mM HEPES/Citrate at 4°C, then washed and concentrated as described for Mga-His₆. The protein was assessed by Western blot and Coomassie staining.

Electrophoretic Mobility Shift assay (EMSA)

Electrophoretic mobility shift assays (EMSAs) were performed as described previously [76]. Briefly, 49-bp DNA probes were generated by annealing oligonucleotide pairs representing wild-type *Pemm1*, *PscpA1*, *Psic1*, and respective point mutations. Each

gel-purified oligonucleotide pair (12.5 μ M) was mixed with 10 mM Tris-HCl (pH 8.0) and 5 mM MgCl₂, heated to 85°C for 5 minutes, and allowed to anneal by slowly cooling to room temperature. Annealed oligonucleotides were end labeled with [γ -³²P] ATP using T4 polynucleotide kinase (NEB). Mga1- His₆ (2.5 μ M) was incubated with 0.1 nM each probe in band shift buffer [20 mM HEPES (pH 7.5), 1 mM EDTA, 0.6 mM dithiothreitol [DTT], 60 mM KCl, 5 mM MgCl₂, and 50 ng/ μ l poly(dI-dC)] for 20 minutes at room temperature. Loading dye (5% Ficoll, 0.1% bromophenol blue) (1/5 volume) was added, and each sample was separated on a 5% polyacrylamide gel at 140 V. The gels were then dried for 1 hour at 80°C, exposed to a phosphorimager plate, and scanned using a FLA-1500 phosphorimager (GE Healthcare).

DNaseI Footprint Analysis

Probes were generated by uniquely end labeling primers, and PCR amplifying with one labeled primer and one cold primer. Each PCR product was run across a 5% PAGE gel, extracted by the crush and soak method, and PCR purified using the QiaQuick PCR Purification kit (Qiagen). Binding reactions were set up as in EMSA. After reaching equilibrium the 1 μ L Turbo™ DnaseI (Ambion) was added to each reaction for 90 seconds. The reaction was precipitated with 150 μ L of DnaseI stop buffer (570 mM NH₄OAc, 50 μ g/mL tRNA, 80 % v/v ethanol). The reactions were then washed twice with 70% ethanol, dried under vacuum and resuspended in 5 μ L DNaseI gel loading dye (80% formamide, 1 x TBE, 0.1% xylene cyanol, 0.1% bromophenol blue). Reactions were separated on a 6% sequencing gel alongside a Sanger sequencing ladder. Gels were dried for 1 hour at 80°C, exposed to a phosphor imager plate and scanned using a FUJI-FLA-1500 or FLA-5000 phosphor imager.

Methylation Protection and Interference Assay

Binding reactions were set up as in EMSA to shift 50% of the probe with the following modifications. For the interference assay, probes were methylated prior to incubation. ~300,000 cpm of each probe was incubated with 100 μ L 2x DMS buffer (120 mM NaCl, 20 mM Tris-HCl, pH 8.0, 20 mM MgCl₂ and 2 mM EDTA) and dH₂O to a volume of 200 μ L. 1 μ L of DMS was added and the *Pemm*-R probe was incubated for 1 min 10 seconds, and the *Pemm*-L probe was incubated for 1 min 15 seconds, at room temperature to obtain approximately one methylation site per probe. The reaction was stopped with the additional of 50 μ L cold DMS stop buffer (1.5 M NaAcetate, pH 7.0, 1 M 2-mercapto-ethanol) followed by an ethanol precipitation. For the protection assay, after the binding reaction had been performed, 20 μ L 0.01% DMS was added to the reaction and incubated for 2 minutes. 1/10 volume of 250 mM dithiothreitol (DTT) was added and then the reaction separated on a 5% PAGE gel and exposed to film. Shifted and unbound probe were excised from the gel and extracted using the crush and soak method, followed by PCR purification. To reveal the modified As and Gs, the probes were then dried and resuspended in 30 μ L of 10 mM sodium phosphate, pH 6.8 and 1 mM EDTA. The probes were incubated for 15 minutes at 92°C, then 3 μ L 1 M NaOH was added for another 30 minutes. 320 μ L of 500 mM NaCl, 50 μ g/mL tRNA and 900 μ L ethanol were added to perform an ethanol precipitation. The probes washed once with 70% ethanol, dried and resuspended in DNaseI load dye. Reactions were run on a 6% sequencing gel alongside a Maxam-Gilbert sequencing ladder [108]. Gels were dried for 1 hour at 80°C, exposed overnight to a phosphor imager plate and scanned using a FUJI 1500 phosphoimager.

Uracil and Missing Thymine Interference Assay

Binding reactions were done as previously described with the following modifications: probes were generated in a reaction that had a 1/20 dUTP:dTTP ratio, so that one thymine was modified per binding site. For the missing thymine interference assay probes were digested with uracil-glycosylase (NEB) for 1 hour at 37°C, followed by PCR purification prior to incubation in the binding reaction. The binding reactions were set up to obtain 50 % shifted probe, run on a 5% PAGE gel and exposed to film. The bound and unbound fractions were extracted by crush and soak, followed by PCR purification. At this time the probes from the uracil interference assay were also digested with uracil glycosylase. The probes were then dried down and resuspended in 50 μ L 1M piperidine to generate strand breaks. The reaction was incubated at 90°C for 30 minutes, then placed on ice. 120 μ L n-Butanol, 50 μ L 1% SDS was added and the upper phase was extracted. This was repeated with 50 μ L n-butanol, and then the probes were dried. The probes were resuspended in 50 μ L dH₂O, redried, then resuspended in 10 μ L DnaseI gel loading dye. The reactions were separated on a 6% sequencing gel run at 1700 V for 1.5 hours, dried for 1 hour at 80°C, then exposed overnight to a phosphoimager plate.

***In vitro* Transcription**

In vitro transcription reactions were performed as follows [98]. 5-10 μ L of RNA polymerase was mixed with 1-3 μ L of σ per reaction and incubated on ice for at least 10 minutes. A 20 μ L reaction containing 4 μ L 5x Transcription buffer (330 mM Tris-Ac, pH 7.9, 10 mM MgAc, 0.1 mM DTT), 1 μ M DNA template, 0.5 μ L Rnase Inhibitor (NEB), 0-5 μ L 50 mM HEPES/Citrate, pH 7.5, 5-0 μ L 5 μ M Mga1-His₆ was incubated at RT for 20 minutes. 6-11 μ L holoenzyme was then added to the reaction and incubated at

37°C for 10 minutes. 1 μ L NTP mix (1 μ L each 10 mM ATP, GTP, CTP, [γ]³²P UTP and dH₂O) was added and incubated for 5 minutes at 37°C. 1 μ L cold UTP was added and incubated for 5 minutes at 37°C. The reaction was then stopped with 12.5 μ L Stop buffer (80% formamide, 12.5% 0.5% bromophenol blue, 20 mM EDTA, brought to volume in 1X TBE). Reactions were denatured for 5 minutes at 80°C then spun briefly. 30-35 μ L of the reaction or 6 μ L was then loaded on a 6% sequencing gel alongside a sequencing reaction to determine the size of transcripts.

***In vitro* Phosphorylation-Transcription**

In vitro phosphorylation of Mga-His₆ was performed as described by Hondorp, et al, (in review). 1 μ L 5 μ M His₆-EI in 50 mM HEPES, pH 7.5, 1 μ L 20 μ M His₆-Hpr in 50 mM HEPES, pH 7.5 with 1 μ L 60 μ M Mga4-His₆ was added to a 20 μ L reaction containing 10 mM MgCl₂/50 mM HEPES, pH 7.5. The phosphotransfer was initiated by adding 1 μ L [³²P]-PEP (~750,000 cpm in 50 mM HEPES, pH 7.5) and incubating for 20 minutes at 37°C. 5 μ L 5x cracking buffer was added to each reaction, then placed on ice. 20 μ L was loaded on a 10% SDS-PAGE gel and run at a constant 20 mAmps until the dye front reached the bottom of the gel. The gel was dried without heat for 5 minutes, then exposed to a phosphoimager cassette.

To perform the *in vitro* transcription assay, the phosphorylation reaction was modified to contain 1 μ M DNA template, 1 μ L 20 μ M Mga-His and 50 mM MgCl₂. The phosphorylation/DNA-binding was incubated for 20 minutes at 37°C, then the *in vitro* transcription assay was performed as described previously.

ChAP (Chromosome Affinity Purification)

ChAP assays were performed as described by Anbalagan et al with modifications [109]. The strain KSM547, the M4 GA40634 isogenic *mga*- strain containing either pLZ12-Spc or pKSM808 was grown overnight in 10 mLs THY supplemented with kanamycin 300 and spectinomycin 100. A 1/20 dilution of the overnight culture was used to inoculate 75 mLs THY supplemented with spectinomycin to mid-logarithmic phase, Klett 75-80 at 37°C. The cells were then incubated on ice for 10 minutes. Proteins were crosslinked to DNA with 1% formaldehyde for an additional 30 minutes with gentle stirring every five minutes. The crosslinking was stopped by adding 1M glycine to a final concentration 125 mM and incubated on ice for another 5 minutes. The cells were pelleted by centrifugation and then washed twice with PBS, pH 7.3. The pellet was resuspended in 500 µL of IP Buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl). 2.5 µL PlyC/5 mLs cells was added to lyse the cells and incubate on ice for 20 minutes. An additional 500 µL IP buffer and 1x protease inhibitor (Roche, in IP buffer) was then added. DNA was sonicated with a target size of ~400 bp. The cells were spun for 20 minutes at 15 000 x g and the supernatant was collected. A 20 µL sample was removed for Western analysis. In order to perform the affinity purification 50 µL of the NiNTA agarose slurry (Qiagen) was washed three times with IP buffer. The NiNTA agarose was added to the supernatant and incubate from 15 min at 4°C with rocking. The slurry was spun briefly at 15 000 rpm and the supernatant was removed. 500 µL of IP Wash Buffer (IP buffer with 10 mM imidazole) was gently added and inverted to mix. The slurry was then washed in the same manner with 20 mM and 50 mM imidazole IP Wash Buffer. 250 µL 250 mM Imidazole Elution Buffer was then added to the slurry and incubate 5

min at 4°C with rocking. The slurry was spun at 15 000 rpm for 5 minutes and the supernatant was collected. A 10 µL sample was removed for Western analysis. The formaldehyde induced crosslinks were then reversed by incubating the supernatant at 65°C overnight (6 hours). 150 µL TE containing glycogen (0.27 mg/mL) and proteinase K (100 µg/mL) was added and incubated for 2 hours at 37°C to digest proteins. The supernatant was extracted once with equal volume phenol-chloroform and washed with an equal volume isopropanol and incubate overnight at -20°C. The DNA was spun for 10 minutes at 15 000 rpm to remove the isopropanol and washed once with 100 µL 70% ethanol. The DNA was dried under vacuum without heat and resuspended in 50 µL dH₂O. The DNA was then assessed by quantitative PCR. Samples were submitted for library formation and Illumina sequencing at IBBR.

Quantitative PCR

DNA collected from the ChAP assay was analyzed by qPCR for enrichment of Mga specific DNA binding sites in the cells containing Mga versus the empty vector control. 5 ng of DNA was added to a Sybr Green Master mix (Applied Biosystems) containing 5 µg of each specified real-time primer. The real-time RT-PCR experiments were completed using a Lightcycler 480 (Roche) and binding sites were detected in the relative quantification mode. Samples were compared to *mga- gyrA* gene levels, with the levels presented representing ratios of the values in the mutant/values in the wild-type.

Sedimentation Equilibrium

Sedimentation measurements were performed in an XLI analytical ultracentrifuge (Beckman Coulter) using cells equipped with 2-hole (3 mm or 1.2 cm path length)

charcoal-filled epon centrepieces. Mga4-His₆ or Δ139Mga4-His₆ was first dialysed overnight into 50 mM HEPES/Citrate containing 100 mM NaCl at 4°C. Full-length Mga4-His₆ prepared at 7.5, 20 and 30 μM was centrifuged at 14, 16, 18, 20 and 22 K r.p.m. and truncated Δ139Mga4-His₆ prepared at 7.5 and 30 μM was centrifuged at 18, 20 and 22 K r.p.m. The data were first analysed WinNonLin for a single species model in [110] to obtain the reduced buoyant molecular mass, *s*, from which the molecular weight was calculated using the following equation:

$$\sigma = \frac{(1 - \bar{v}\rho)}{RT} \omega^2$$

where *M* is the molecular weight, *v* is the partial specific volume obtained using SEDNTERP (<http://www.rasmb.bbri.org>) 0.7437 cm³ g⁻¹ for Mga4-His₆ and 0.7445 cm³ g⁻¹ for Δ139Mga4-His₆, *ρ* is the buffer density, *R* is the gas constant, *T* is the temperature in Kelvin and *ω* is the angular velocity. The data for the full-length protein, which yielded a molecular weight higher than that expected for the monomer, was further subjected to analysis using monomer–oligomer models. The nine data sets were globally analyzed using a monomer-dimer model to obtain an association constant, *K_a*, using the equation:

$$c_t(r) = \delta + c_m(r_o) e^{\sigma_m \left(\frac{r^2 - r_o^2}{2} \right)} + K_a (c_m(r_o))^2 e^{2\sigma_m \left(\frac{r^2 - r_o^2}{2} \right)}$$

where *c_t* is the total concentration at position *r*, *δ* is the baseline offset, *c_m(r_o)* is the concentration at the reference radial position *r_o*, and *σ_m* is the reduced molecular weight of the monomer. The association constants, which are obtained in absorbance units from the analysis, are reported as molar equilibrium dissociation constants or *K_{DIM}*. The

quality of each analysis was assessed by the square root of variance of the fit and the distribution of residuals.

DRACALA (Differential Radial Capillary Action of Ligand Assay)

Binding reactions were performed as for EMSA. 5 μ L was spotted in triplicate on a nitrocellulose membrane and allowed to dry. Blots were exposed overnight to a phosphorimager plate and scanned using a FUJI-phosphoimager. Densitometry was performed using Multigauge. The equations

$$F_B = (I_{\text{Inner}} - I_{\text{Background}})/(I_{\text{Total}})$$

and

$$I_{\text{Background}} = A_{\text{Inner}} \times ((I_{\text{Total}} - I_{\text{Inner}})/(A_{\text{Total}} - A_{\text{Inner}}))$$

where F_B is the fraction bound, I is intensity and A is area were used to calculate the amount of DNA bound for each protein concentration. The data was then plotted using the GraphPad Prism and used to calculate K_d [111].

Construction of Bacterial-Two-Hybrid plasmids

To create a N-terminal T18-tagged σ , α and δ , the plasmid pT18C-link was digested with BamHI and EcoRI, then gel purified. *rpoD* (σ) was amplified using the primers T18C-rpoD-L and T18C-rpoD-R, *rpoA* (α) was amplified using the primers T18C-rpoA-L and T18C-rpoA-R, and *rpoE* (δ) was amplified using the primers T18C-rpoE-BamHI and T18C-rpoE-EcoRI from MGAS5005 gDNA, digested with BamHI and EcoRI, then ligated with pT18C-link, to create the plasmids **pKSM223** (pT18C- σ), **pKSM224** (pT18C- α) and **pKSM237** (pT18C- δ).

To create A C-terminal T18 or T25-tagged σ , α , δ , Mga1 and Mga1- Δ 139, the plasmid pT18N-link and pT25N-link were digested with HindIII and BamHI, then gel purified. *rpoD* was amplified using the primers T18N-rpoD-HindIII and T18N-rpoD-BamHI, *rpoA* was amplified using the primers T18N-rpoA-L and T18N-rpoA-R, and *rpoE* was amplified using the primers T18N-rpoE-HindIII and T18N-rpoE-BamHI from MGAS5005 gDNA the digested with HindIII and BamHI. *mga* was amplified from MGAS5005 gDNA using the primers T25N-Mga-L and T25N-Mga-R, then bluntly ligated into pCRII-Blunt-TOPO to create T25N-Mga-TOPO. *mga* was then digested with HindIII and BamHI and gel purified. *mga1- Δ 139* was amplified from MGAS5005 gDNA using the primers T25N-Mga- Δ 139-HindIII and T25N-Mga- Δ 139-BamHI, then digested with BamHI and HindIII. These were then ligated into pT18N-link or pT25N-link to create the plasmids **pKSM225** (pT18N- σ), **pKSM228** (pT18N- α), **pKSM229** (pT25N- α), **pKSM230** (pT25N- δ), **pKSM233** (pT18N- δ), **pKSM226** (pT25N-Mga), **pKSM227** (pT18N-Mga) and **pKSM236** (pT25N-Mga- Δ 139).

To create a C-terminal T18-his- σ and T18-his- α , pT18N-link was digested with EcoRI and HindIII and gel purified. *His-rpoA* was amplified from pKSM234 using the primers T18N-rpoA-EcoRI and T18N-his-rpoA-HindIII, and *his-rpoD* was amplified from pKSM246 using the primers T18N-his-rpoD-EcoRI and T18N-his-rpoA-HindIII, then digested with EcoRI and HindIII, and ligated into pT18N-link to create **pKSM277** (pT18N-his- α) and **pKSM278** (pT18N-his- σ).

Bacterial-Two-Hybrid

The selected plasmids for analysis were co-transformed (5 ng each for the negative controls, 100 ng each for the experimentals) into BTH101 and outgrown for 1

hour at 37°C. The bacteria were plated on MacConkey agar supplemented with 1% maltose and 1 mM IPTG and placed at 30°C overnight. Following overnight incubation, colonies were patched on fresh plates, placed at 30°C and monitored for a color change of white to pink until the negative controls reverted.

Alternatively plasmids were plated on LB agar supplemented with 1 mM IPTG, and 2 mM X-gal, placed at 30°C, and monitored for a color change of white to blue until the negative controls were reverted. After an overnight incubation, single colonies were used to inoculate LB media and grown to an OD₆₀₀ of 1.5-1.7. 100 µL of cells were washed three times in saline, then spotted on A+M minimal media (3.6 µM FeCl₃, 40 µM MgCl₂, 0.1 mM MnCl₂, 10 mM NH₄Cl, 75 µM Na₂SO₄, 0.5 mM KH₂PO₄, 1.2 mM NH₄NO₃, 1 mM MgSO₄, 0.8% glucose, 0.0001% thiamine, 0.2% caseine hydrolysate, 0.008% x-gal, 100 µg/mL Amp, 25 µg/mL Kan, 1 mM IPTG, 1.5% agar) and monitored for a color change of white to blue at 30°C.

Bacterial-Two-Hybrid Western Blots

A single colony each of pMgal-His, pKSM277 (pT18N-His₆-α) and pKSM278 (pT18N-His₆-σ) in C41[DE3] was used to inoculate 30 mL ZYP-5052. One set of flasks was placed at 30°C and grown for ~48 hours. One set of flasks was grown for ~8 hours at 37°C, then grown ~14 hours at RT. For total protein 100 µL was collected from each flask, the media was removed, then the pellet was resuspended in 1x cracking buffer. To isolate soluble proteins, 750 µL was collected from each flask and pelleted. The pellet was resuspended in 150 µL B-Per Reagent (Pierce), vortexed for 1 minute, spun at 15000 rpm for 5 minutes, then the supernatant collected. 12.5 µL of each whole

cell sample and 25 μ L of each soluble fraction was then run on a 10% SDS-PAGE gel, and probed for α -his by Western blotting.

Immunoblots

Protein samples were run on 10 or 12% SDS-PAGE gels with 4% stacking gel for approximately 50 minutes at 180 V. Proteins were transferred to nitrocellulose membranes using the Mini-Protean apparatus (Bio-Rad) in 1x transfer buffer (25 mM Tris base, 0.2 M glycine, 20% methanol). Membranes were blocked for 1 hour at room temperature in blocking solution (5% (w/v) dried milk in PBS-tween). For the detection of His-tagged proteins, blots were incubated with a 1:1 000 dilution of α -His antibody (Roche), 0.4 mg/mL of α -CBP antibody (Genscript) for CBP-tagged proteins, and 1:1 000 of the polyclonal anti-rabbit Mga4 antibody, for two hours at room temperature, followed by three 5 minute washes with PBS-tween. Blots were incubated with 1:20 000 α -mouse-HRP (His-tagged proteins), 1:20 000 α -rabbit-HRP (for CBP-tagged proteins and Mga) for 1 hour in blocking solution followed by three washes. Blots were visualized using the SuperSignal West Femto Substrate (Thermo Scientific) and a LAS-3000 CCD camera (FujiFilm).

***In vitro* Co-Affinity Purification**

12 μ L RNAP + 2 μ L σ was pre-incubated on ice for 10 minutes. A 40 μ L reaction of 10 μ L of \sim 20 μ M Mga4-CBP and in 50 mM HEPES, pH 7.9, 75 mM MgAc was incubated for 20 minutes at RT with 10 μ L Mga1-His₆. When assessing the pulldowns in the presence of a MBS, 1 μ M template DNA was also included at this time. Holoenzyme was added and the reaction was incubated for 20 minutes at RT. 20 μ L NiNTA agarose

was added, then the reaction was incubated for 10 minutes at RT with gentle rocking. The reactions were washed 2 times 100 μ L 20 mM Imidazole, 50 mM HEPES, pH 7.9, 10 mM MgAc by centrifugation. 40 μ L 250 mM Imidazole, 50 mM HEPES, pH 7.9, 10 mM MgAc was added, then the reactions was incubated at RT fo 5 minutes with gentle rocking before the final elution was collected. The co-purified proteins were assessed by immunoblots probing for α -CBP and α -His.

Purification of RNA Polymerase

300 mLs to THY broth was inoculated with a 1/20 dilution of the strain JRS4-Polhis and grown to late logarithmic phase (~Klett 130) [98]. The cells were then harvested by centrifugation, and frozen at -80°C. The cells were resuspended in 1 mL of ice cold Lysis Buffer P (300 mM NaCl, 50 mM Na₂HPO₄ and 5% glycerol). 2.5 μ L PlyC/10 mLs cell culture was added, then incubated on ice for 20 minutes. 100 μ L 25x protease inhibitor (Roche) (in Lysis Buffer P) was then added. DNA was sheared by sonication using a Branson sonifier with the settings microtip 5, 50% duty cycle for 20 seconds. The lysate was then spun for clarification and passed through a 0.45 μ m filter and passed over a 750 μ L column of washed NiNTA agarose. The agarose was washed with 20 mLs Lysis buffer P containing 20 mM imidazole and eluted with 8 mL of Lysis buffer P containing 400 mM imidazole. Fractions were immediately concentrated by centrifugation and dialyzed for 2 hours against 1x Transcription buffer (33 mM Tris-Ac, pH 7.9, 10 mM MgAc, 0.1 mM DTT) at 4°C. For use in the *in vitro* AP assay, RNAP was dialyzed against 50 mM HEPES, 10 mM MgAc for 2 hours at 4°C. After checking for the β and β' band by Coomassie staining, 5% glycerol was added and RNAP was aliquoted and frozen at -80°C. This purification should be completed in 1 day for best

results. Aliquots were freshly thawed for each use, and the protein stays active for ~1 month under these conditions.

Purification of His₆- α , His₆- α - Δ CTD, His₆- α NTD- σ 4, His₆- σ and His₆- σ Δ 4

1 L of ZYP-5052 inoculated with C41[DE3](pKSM234) or C41[DE3](pKSM553) was grown for ~16 hours at 37°C, or with C41[DE3](pKSM235) for ~48 hours at 37°C. C41[DE3](pKSM246) and C41[DE3](pKSM279) were grown for 8 hours at 37°C, then RT for ~16 hours. Cells were harvested by centrifugation. Pellets were then treated as described for the Mga1-His₆ protein purification. The lysate was passed over a 750 μ L NiNTA agarose column, washed with 20 mLs 20 mM imidazole NiNTA wash buffer and 10 mLs 50 mM imidazole NiNTA wash buffer, and eluted into 7 fractions with 8 mLs NiNTA elution buffer. Fractions were checked for purity by Coomassie staining. Fractions were then dialyzed overnight into 1x Transcription buffer at 4°C. After any precipitate was removed, fractions were aliquoted and stored at -80°C.

Purification of σ

σ factor was purified from *E. coli* following the protocol of Burgess [112]. An overnight culture of BL21[DE3](pLysS)(pEU7534) containing the GAS σ factor was used to inoculate 2x 500 mL flasks of LB with Amp 100 μ g/ μ L and grown at 37°C to OD₆₀₀ of 0.8. The culture was induced with 1 mM of IPTG for 30 minutes. Rifampicin was added at 150 μ g/mL and then the culture was grown for an additional 3.5 hours. Cells were harvested by centrifugation and stored at -80°C. The pellet was resuspended in 30 mLs Buffer A (50 mM Tris-HCl, pH 7.9, 5% glycerol, 0.1 mM EDTA, 50 mM NaCl, 0.1 mM DTT) with 1x protease inhibitor, then sonicated as previously described. 2

mLs of 20% sodium deoxycholate (NaDOC) was added, mixed, then incubated for 10 minutes at 4°C. The lysate was then centrifuged at 13000 rpm for 10 minutes at 4°C and the supernatant was discarded. The pellet was resuspended in 20 mL buffer A + 2% NaDOC and centrifuged for an additional 10 minutes and the supernatant was discarded. The pellet was resuspended in 40 mL buffer A + 0.4% Sarkosyl and incubated for 20°C for 30 minutes. After centrifugation then supernatant was collected and diluted to 400 mLs by adding buffer A in increments of 80 mLs, with 10-15 minutes between each addition, while at 4°C. The supernatant was dialyzed twice for 8 hours against 4 Ls of buffer A at 4°C. The supernatant was spun and passed through a 0.45 µm filter to remove any precipitate and then loaded onto a HiTrap Q FF anion exchange column by FPLC. The column was washed for 15 minutes (4mL/min) with buffer A, then eluted for 60 minutes with a linear gradient from buffer A to buffer B (Buffer A + 1 M NaCl) and collected in 5 mL fractions. 20 µL of fractions with the highest OD₂₈₀ peak were then analyzed by Coomassie staining. Fractions with a strong and pure 55 kDa band were pooled and dialyzed against 1 L storage buffer (buffer A + 45% glycerol) overnight at 4°C.

Creation of Mutant RNA Polymerases

RNAP, his₆-α and his₆-α-ΔCTD were purified as previously described with the following modifications. After concentrating to ~1 mL, RNAP was dialyzed into Fold/Refold buffer (Transcription Buffer + 5% glycerol) at 4°C for 2 hours, changing twice. RNAP was then divided in 2 fractions and dialyzed into Fold/Refold buffer with 0.1 M GuaHCl for 2 hours at, 1.0 M GuaHCl for 2 hours and 6.0 M GuaHCl for 30 minutes, at 4°C. To the fraction α, 100 µL of his₆-α and to fraction Δ, 100 µL of his₆-α-

Δ CTD was added in excess to RNAP, then immediately dialyzed into 1 L Fold/Refold buffer for 2 hours, changing 1 time at 4°C. The samples were spun in a 100 kDa centricon (Ambion) to remove unincorporated subunits, and Fold/Refold buffer was added to maintain volume. Coomassie and Western blotting were performed to assess purity and the presence of the β' , α , α - Δ -CTD components. 50 μ L aliquots were made and stored at -80°C.

Construction of Protein Expression vectors

In order to overexpress and purify the N-terminal his tagged α , α - Δ CTD, α - Δ 1/3CTD and α - Δ 2/3CTD, the plasmid pProEX-htb was digested BamHI and XbaI, then gel purified. The *rpoA* (α subunit) was amplified using the primers RpoA-His-Tag-L and RpoA-His-Tag-R, the α - Δ CTD was amplified using the primers RpoA-His-Tag-L and RpoACTD-His-Tag R, α - Δ 1/3CTD was amplified using the primers RpoA-His-Tag-L and RpoA-Trunc1-XbaI, and α - Δ 2/3CTD was amplified using RpoA-His-Tag-L and RpoA-Trunc2-XbaI from MGAS5005 gDNA, digested with BamHI and XbaI, then ligated into pProEX-htb to create the plasmids **pKSM234** (pProEX-htb- α), **pKSM235** (pProEX-htb- α - Δ CTD), **pKSM282** (pProEX-htb- α - Δ 1/3CTD) and **pKSM283** (pProEX-htb- α - Δ 2/3CTD).

In order to over express and purify the C-terminal his tagged α , α - Δ CTD, α - Δ 1/3CTD and α - Δ 2/3CTD, the plasmid pET21a was digested with NdeI and HindIII, then gel purified. The full length α subunit was amplified using the primer RpoA-pET21A-HindIII and RpoA-pET21A-NdeI, the α - Δ CTD was amplified using the primers RpoA-pET21a-NdeI and RpoACTD-pET21a-HindIII, α - Δ 1/3CTD was amplified using the primers RpoA-pET21a-NdeI and RpoA-Trunc1-HindIII and α - Δ 2/3CTD was

amplified using the primers RpoA-pET21a-NdeI and RpoATrunc2-HindIII from MGAS5005 gDNA, then digested with NdeI and HindIII and ligated into pET21A to create the plasmids **pKSM291** (pET21a- α - Δ CTD), **pKSM292** (pET21a- α), **pKSM296** = (pET21a- α - Δ 1/3CTD) and **pKSM297** (pET21a- α - Δ 2/3CTD2/3).

To over express and purify a N-terminal his-tagged σ or σ - Δ domain4, the plasmid pProEX-htb was digested with BamHI and XbaI, then gel purified. σ was amplified using the primers Sigma-his-BamHI and Sigma-His-XbaI, and σ - Δ domain4 was amplified using the primers Sigma-his-BamHI and Sigma-hisdelta4+stop-XbaI from MGAS5005 gDNA, then digested with BamHI and XhoI and ligated into pProEX-htb to create the plasmids **pKSM246** (pProEX-htb- σ) and **pKSM279** (pProEX-htb- σ - Δ domain4).

To overexpress and purify a N-terminal his tagged α NTD- σ domain4, the plasmid pProEX-htb was digested with BamHI and XbaI. α NTD was amplified from gDNA using the primers RpoA-His-tag-L and RpoA-NTD-R. σ domain4 was amplified from gDNA using the primers σ domain4 overlap and Sigma-his-XbaI. The fragments were joined by Splicing by Overlapping Extension-PCR (PCR-SOE) using the primer RpoA-His-tag-L and Sigma-his-XbaI, digested with BamHI and XbaI and ligated into pProEX-htb to create the plasmid **pKSM553**.

In order to create a M1 Mga HTH-3/4 or M1 Mga HTH-4, protein for purification, M1 Mga HTH 4a and M1 Mga HTH 4b were used to introduce the alanines into the recognition helix in pKSM805 or pKSM874 by Site directed mutagenesis. Mutations were confirmed by sequencing. pET21a and the SDM template plasmid were digested with NdeI and XhoI. The 1.5 kb band containing the mutation and the 4.5 kb

band from pET21a were gel purified and ligated to create the plasmid **pKSM249** (Mga1-HTH3/4-His₆) and **pKSM250** (Mga1-HTH4-His₆).

In order to express and purify the C-terminus of Mga (PRD1, PRD2 and the EIIB domains) or the N terminus of Mga (CMD, HTH-3 and wHTH-4) pET21a was digested with NdeI and XhoI and gel purified. The C-terminus of Mga was amplified from MGAS5005 gDNA using the primers Mga-XhoI and M1-C-Mga-NdeI, digested with NdeI and XhoI and ligated into pET21a to create the plasmid **pKSM264** (pET21a-CMga). The N-terminus of Mga was amplified using the primers Mga1-NdeI and MgaN180-XhoI, digested with NdeI and XhoI, the ligated into pET21a to create the plasmid **pKSM265** (pET21a-N180Mga).

To over express and purify a C-terminal CBP-His tagged Mga1, Mga4, Mga4-Δ139 and Mga4-Δ29, the plasmid pCal-C was digested with NcoI and BamHI, then gel purified. Mga1 was amplified using the primers Mga1-CBP-NcoI and Mga1-CBP-BglII from SF370 gDNA, Mga4 was amplified using the primers Mga4-CBP-NcoI and Mga4-CBP-BglII, Mga4-Δ139 was amplified using the primers Mga4139-CBP-NcoI and Mga4139-CBP-BglII, Mga4-Δ29 was amplified using the primers Mga4139-CBP-NcoI and Mga429-CBP-BglII, then digested with NcoI and BglII and ligated into pCal-C to create the plasmids **KSM288** (pCal-C-Mga1), **pKSM289** (pCal-C-Mga4), **pKSM550** (pCal-C-Mga4-Δ139) and **pKSM299** (pCal-C-Mga4-Δ29). The plasmid pET21a was then digested with NdeI and XhoI and gel purified. Mga4 was amplified from pKSM289, Mga4-Δ139-CBP was amplified from pKSM550 and Mga4-Δ29-CBP was amplified from pKSM299 using the primers Mga4-CBP-NdeI and Mga4-CBP-XhoI, digested with NdeI and XhoI and ligated into pET21a to create the plasmids **pKSM298** (pET21a-

Mga4-CBP), **pKSM551** (pET21a-Mga4- Δ 139-CBP) and **pKSM552** (pET21a-Mga4- Δ 29-CBP).

To over express and purify a C-terminal his tagged RofA and RivR, pET21a was digested with NdeI and XhoI. RofA was amplified using the primers M1-RofA-NdeI and M1-RofA-XhoI and rivR was amplified using the primers M1-RivR-NdeI and M1-RivR-XhoI from MGAS5005 gDNA, digested with NdeI and XhoI and ligated into pET21a to create the plasmids **pKSM269** (pET21a-RofA) and **pKSM270** (pET21a-RivR). In order to overexpress and purify a N-terminal His-MBP tagged RofA and RivR, the plasmid pVL847 was digested with NdeI and XhoI. RofA was amplified using the primers RofA-NdeI and RofA-MBP-his-XhoI and RivR was amplified using the primers RivR-NdeI and RivR-MBP-his-XhoI from MGAS5005 gDNA, digested with NdeI and XhoI and ligated into pVL847 to create the plasmids **pKSM286** (pVL847-RofA) and **pKSM287** (pVL847-RivR).

Chapter 3

Nucleotides Critical for the Interaction of the *Streptococcus pyogenes* Mga Virulence Regulator with Mga-Regulated Promoter Sequences

Introduction

Regulation of gene expression in response to changing stimuli allow bacteria to rapidly adapt to their constantly changing environment. Control of transcription is often mediated by direct interactions between target gene promoters and specialized DNA-binding proteins that either enhance (activate) or inhibit (repress) RNA polymerase-mediated initiation [96]. Transcription factors possess DNA-binding domains that allow them to recognize and specifically bind to a conserved DNA sequence (binding site) within their target promoters. A conserved family of DNA binding motifs found within many prokaryotic transcription factors, as well as in eukaryotic cells, is the helix-turn-helix (HTH) domain [113]. The second helix in the HTH fold is often called the “recognition” helix because it forms the principal DNA-protein interface by inserting into the major groove of the DNA to interact with specific nucleotides; however, DNA contacts may vary across the fold [113]. HTH domains can be quite diverse in structure, with the winged HTH (wHTH) possessing an additional C-terminal β -strand hairpin [113]. In order to differentially regulate gene expression, DNA-binding proteins must be able to discriminate specific sequences. These sequences often contain a dyad symmetry reflecting that dimers and other multimers of the DNA-binding protein interact with the DNA [81].

Mga, the multiple gene activator of GAS, regulates expression of approximately

10% of the genome [72]. The core regulon is composed of a small number of key virulence factors that Mga activates through binding to their promoter DNA, including genes encoding M protein (*emm*), M-like proteins (*arp* and *mrp*), C5a peptidase (*scpA*), and the streptococcal inhibitor of complement (*sic*) [72].

Based on studies primarily done in the serotype M6 strain JRS4, three categories of Mga-regulated promoters (categories A, B, and C) were proposed based on the number of binding sites and their position relative to the start of transcription [78]. Category A promoters (*Pemm* and *PscpA*) were defined using DNase I footprinting; these promoters are composed of a single 45-bp binding site centered at -54 from the start of transcription overlapping the -35 hexamer [76]. A category B promoter (*PscIA* and *Psof*) was defined by sequence alignment, electrophoretic mobility shift assay (EMSA) analysis, and *in vitro* transcription. These promoters have a single 45-bp binding site that is located further upstream (-168) from the start of transcription [78,114]. A category C promoter (*Pmga*), defined by DNase I footprinting, is composed of two 59-bp binding sites located far upstream (-100 and -181) from the start of transcription [77]. Based on the positions of putative binding sites, category A appears to be the most common pattern among Mga-regulated promoters in sequenced GAS strains. Interestingly, sequence alignments of these binding sites exhibit very low sequence identity, making it difficult to determine how Mga interacts with its promoters. In this study, we dissect the protein-DNA interactions between Mga and a model category A promoter (*Pemm*) to understand how this process occurs and test whether these findings can be applied to other Mga-regulated promoter binding sites in GAS.

Results

Characterization of MBSs in the MIT1 MGAS5005.

Published biochemical analyses of Mga binding sites (MBSs) (EMSA, DNase I footprinting) have focused on a single serotype M6 GAS strain JRS4 [76,77]. To determine whether Mga-promoter interactions were conserved in other GAS serotypes, Mga binding sites for *Pemm* and *PscpA*, two category A promoters [78], were characterized in the invasive MIT1 strain MGAS5005. The Mga-regulated *sic* gene is found exclusively in M1 GAS and possesses a predicted category A promoter (*Psic*) based on sequence alignment with M6 sequences; therefore, direct DNA binding studies on *Psic* were also performed. Each promoter was amplified from the MGAS5005 genome and was cloned in front of a promoterless firefly luciferase (*luc*) gene in the reporter plasmid pKSM720 [52] for analysis in wild-type MGAS5005 and in the isogenic *mga*-inactivated strain KSM165L.5005 (Figure 5A). Luciferase activity was assessed at mid-logarithmic phase (80 Klett units) at a point associated with maximal Mga activity. The *Pemm-luc* promoter showed the highest luciferase activity (1.6×10^5 relative luciferase units [RLU]), *Psic-luc* showed intermediate activity (8.7×10^3 RLU), and *PscpA-luc* exhibited the lowest activity (2.7×10^2 RLU) at this time point (data not shown). All three promoters showed significantly reduced luciferase activity in the *mga* inactivated KSM165L.5005 compared to the wild-type (Figure 5A), confirming the Mga-dependent transcriptional activation of *Pemm*, *PscpA*, and *Psic* in the MIT1 background.

EMSAs using 0.1 nM double-stranded 49-mer oligonucleotide probes of each promoter binding site and various amounts of purified Mga1-His₆ protein found that Mga bound to each with maximal binding at 2.5 μ M protein (data not shown). At this level of

protein, Mga shifted 62% of the *Pemm1*-MBS 49-mer probe, 60% of the *Psic1*-MBS 49-mer probe, and 35% of the *PscpA1*-MBS 49-mer probe (Figure 5B). To delineate the nucleotides bound by Mga, DNase I footprint assays were performed on both strands of each of the three MGAS5005 promoters using increasing amounts of purified Mga1-His₆ (Figure 5C, antisense; data not shown, sense). In each case, Mga protected a 45-bp region of DNA immediately upstream of the -35 region (Figure 5D) that correlated exactly with the binding sites predicted by sequence alignments to the established M6 category A binding sites (Figure 6A)



Figure 6 Conservation of nucleotides in known Mga binding sites

(A) ClustalW nucleotide alignment of category A, B, and C Mga binding sites identified by DNase I footprinting (*Pemm*, *PscpA*, and *Pmga*) or overlapping EMSA (*Psc1A*) in the serotype M6 strain JRS4 [76]. Conserved nucleotides in all sequences are in red, conserved nucleotides in n-1 sequences are orange, conserved nucleotides in n-2 sequences are brown. Gaps introduced to maximize sequence alignment are indicated by dashes. (B) ClustalW nucleotide alignment of category A *Pemm* and *PscpA* Mga binding sites from M6 JRS4 and M1T1 MGAS5005 GAS. (C) Sequence of *Pemm1* 49-mer double-stranded oligonucleotide probe encompassing the 45-bp Mga binding site [73] used in this study (Table 7). Nucleotides are numbered from 5' to 3'.

Comparison of Mga binding sites between different promoter categories.

The goal of this study was to identify the nucleotides within a Mga binding site that are important for interacting with Mga, resulting in functional activation of transcription. A sequence alignment using a modified ClustalW of the published Mga binding sites (*Pemm6*, *PscpA*, *PsclA*, *Pmga1*, and *Pmga2*) from M6 JRS4 [76,77] with the MIT1 MGAS5005 sites (*Pemm1*, *PscpA*, and *Psic*) representing all three categories of Mga-regulated promoters, exhibits only 13.4% nucleotide identity (Figure 6B). This variability across the different types of binding sites has made it difficult to define a “core DNA-binding sequence.” However, Mga binding sites from comparable promoters found in other GAS serotypes exhibit much higher nucleotide similarity, as seen with *Pemm* and *PscpA* from MIT1 and M6 GAS, which shows a nucleotide identity of 49.1% (Figure 6B, asterisks). Because *Pemm* is conserved in many GAS serotypes, is strongly regulated by Mga, and shows one of the highest transcript levels of any GAS gene *in vivo* (6), the 45-bp MIT1 *Pemm1* from strain MGAS5005 was chosen as the paradigm Mga binding site for the studies described here (Figure 6C, shaded region). Conserved nucleotides found to be important for Mga binding and activation in *Pemm* were then tested in other category A Mga-regulated promoters (*PscpA* and *Psic*).

Biochemical analysis of the *Pemm1* Mga binding site.

Biochemical assays were performed to assess the role of each thymine, adenine, and guanine of the *Pemm* binding site for Mga interaction. The methyl group of thymine, nitrogen-3 of adenine and nitrogen-7 of guanine have all been identified as points of contact between protein and DNA [115,116]. Therefore, biochemical assays that specifically disrupt these potential sites of Mga interaction were chosen. In each assay,

Mga1-His₆ was incubated with a randomly modified 226-bp M1 *Pemm* PCR probe so that 50% of the probe was bound, and separated by EMSA. Strand scissions were then induced in the bound and free DNA fractions to reveal the modified nucleotides, followed by separation on a 6% acrylamide sequencing gel. Nucleotides important for DNA binding are those found in the free DNA lane but are diminished or missing in the bound DNA lane.

Uracil interference assays were used to target the thymines in the binding site by randomly replacing them with uracil during the PCR amplification of the probe using a dTTP/dUTP ratio that gave one substitution per binding site (Figure 7A, bottom gel). Nucleotides in the binding site were numbered 5' to 3' using the *Pemm1* 49-mer Mga binding probe as a reference (Figure 6C). On the sense strand, thymine 39 (T39) was reduced (64% of free) in the bound fraction, while on the antisense strand T13 was also diminished (49%) in the bound fraction (Figure 7A, bottom gel, and C).

In missing thymine interference assays, incorporated uracils were cleaved by a uracil DNA deglycosylase, leaving only the sugar phosphate backbone prior to incubation with protein (Figure 7A, top gel). On the sense strand, T11 (28%) and T39 (45%) were identified as being important for binding (Figure 7C), and on the antisense strand, T13 (48%), T33 (12%), T34 (7%), and T35 (7%) were also reduced in the bound fraction (Figure 7A, top gel, and C).

Methylation protection assays were performed to identify those guanines or adenines protected from methylation by Mga binding (Figure 7B, top gel). Guanines are methylated on the nitrogen-7 position in the major groove of the DNA helix, while adenines are methylated on the nitrogen-3 located in the minor groove. Guanines G9

(8%), G10 (14%), and G19 (77%) were identified on the sense strand (Figure 7C), and G12 (39%) and A39 (57%) were identified on the antisense strand (Figure 7B, top gel, and C). Methylation interference assays were performed to determine at which guanines and adenines would prior methylation prevent Mga binding (Figure 7B, bottom gel). Nucleotides G9 (11%), G10 (11%), G18 (75%), G19 (62%), G40 (21%), and G41 (16%) were identified on the sense strand (data not shown), and A11 (8%), G12 (23%), and A39 (33%) were identified on the antisense strand (Figure 7B, bottom gel, and C). A summary of all the biochemical results is provided using the *Pemm1* sequence (Figure 7D).

***In vivo* analysis of *Pemm1* binding site mutants.**

Luciferase assays were performed to study the effect on transcriptional activity of a *Pemm1-luc* reporter by directed mutagenesis of selected conserved nucleotides based on the alignment of the M1 and M6 *Pemm* and *PscpA* Mga binding site (Figure 6B). In addition, mutations were introduced into all cytosine nucleotides on the sense strand (C3, C12, C23, C29, C38, and C42) as well as any nucleotides identified as important for binding in the biochemical assays above yet not already targeted. *Pemm-luc* plasmids containing each mutant promoter, a wild-type *Pemm-luc* plasmid, and a promoterless *luc* control plasmid were transformed into wild-type strain MGAS5005. Samples were taken at mid-logarithmic phase (80 Klett units), a time of maximal Mga-regulated expression, in order to quantify activity. The wild-type *Pemm* promoter was set at 100% relative luciferase activity, and the activity of each mutated promoter was calculated as a percentage of the activity of the wild-type (Figure 8A). Strains with the C3A, G10A, A13C, G18A, A33C, and G41A mutations had expression greater than 75% and were

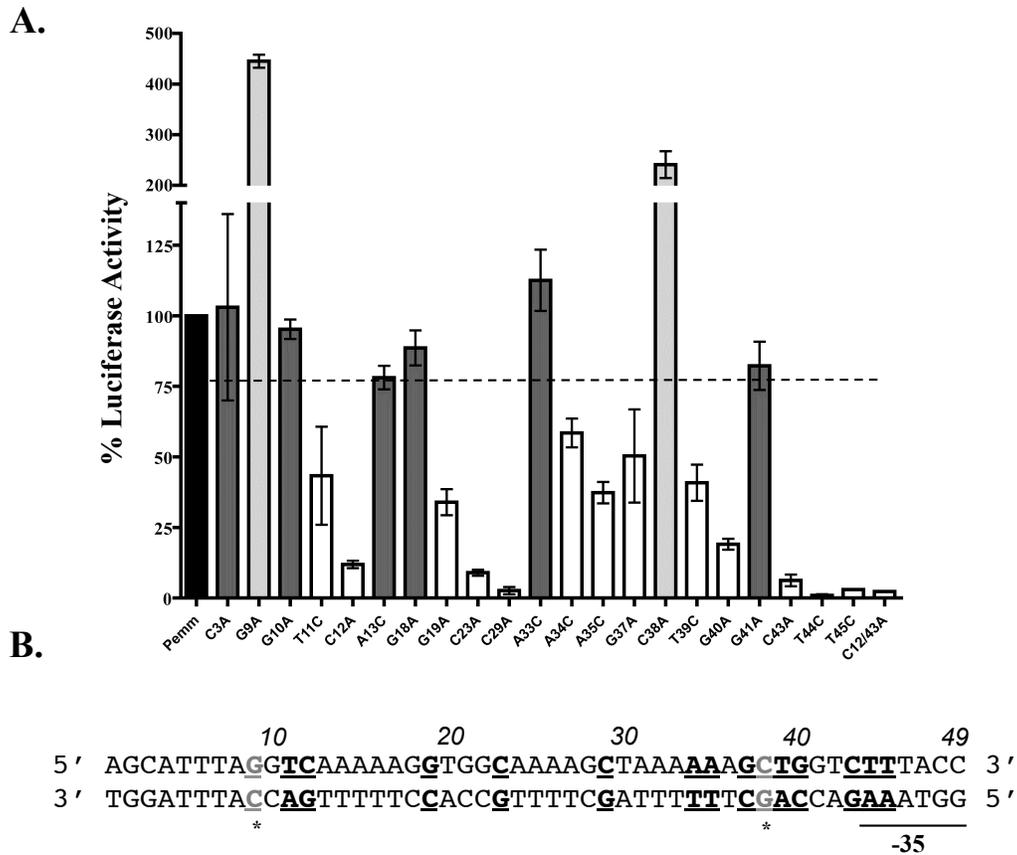


Figure 8 Luciferase promoter reporter assays of *Pemm* site-directed mutations *in vivo*

The 226-bp *Pemm1* probe for the wild-type strain and mutant strains with each point mutation was cloned into the firefly luciferase reporter pKSM720 and transformed into MIT1 MGAS5005 GAS for *in vivo* analysis. (A) Quantification of the relative luciferase activity (RLU) of each *Pemm* point mutation was compared to that of the wild-type and shown as percent luciferase activity. Mutants showing less than 75% wild-type activity (white), 75 to 100% of wild-type activity (dark gray), and greater than wild-type activity (light gray) are indicated. The broken line denotes 75% of wild-type luciferase activity. (B) Schematic diagram showing the locations of all nucleotides within the *Pemm1* 49-mer identified as important for *Pemm* activity (bold and underlined). The nucleotides leading to increased promoter activity (G9A and C38A) are shown as gray underlined letters with an asterisk.

considered to have wild-type activity (Figure 8A, dark gray bars). Strains with the single mutations T11C, C12A, G19A, C23A, C29A, A34C, A35C, G37A, T39C, G40A, C43A, T44C, and T45C and the double mutation C12/43A (C-to-A mutations at positions 12 and 43) had a significant decrease (less than 75%) in luciferase activity (Figure 8A, white

bars). Strains with two different mutations, G9A and C38A, had increased luciferase activity, which increased transcriptional activity to 445% and 241% of the wild-type, respectively (Figure 8A, light gray bars). These two mutated plasmids were also transformed into the *mga*-inactivated KSM165L-5005 strain. Luciferase assays with this strain showed that these mutations caused the same amount of activity as the wild-type *Pemm* promoter in the absence of Mga and that the increase in transcriptional activation with each *PemmI* mutant is Mga dependent (data not shown). A summary of all the *in vivo* reporter results is provided using the *PemmI* sequence (Figure 8B).

EMSA analysis of Mga binding to *Pemm* mutants

EMSA analysis was performed in order to determine the effect on Mga binding of the nucleotides identified by either the biochemical binding assays or luciferase reporter assays. In each assay, 2.5 μ M Mga1-His₆ was incubated with either 0.1 nM concentration of the *PemmI* MBS 49-mer probe or a mutated probe at the ratio of protein to probe previously determined to be saturating with the probe (Figure 9A and data not shown). All mutant *PemmI* probes were constructed so that guanines and cytosines were mutated to adenines, whereas the adenines and thymines were mutated to cytosine. Following EMSA, densitometry was performed to measure the amount of total probe bound, and each mutated probe was then compared to the wild-type to calculate the percentage shift (Figure 9B). Since the EMSA was saturating for the wild-type, this was set at 100%. Mga shifted wild-type amounts (<75%) of the *PemmI* mutants A13C, G18A, C23A, A33C, G41A, T44C, and T45C MBS 49-mer probes (Figure 9B, dark gray bars). Mga shifted significantly less (>75%) of the G9A, G10A, T11C, C12A, G19A, C29A, A34C, A35C, G37A, T39C, G40A, and C43A *PemmI* mutants and the double

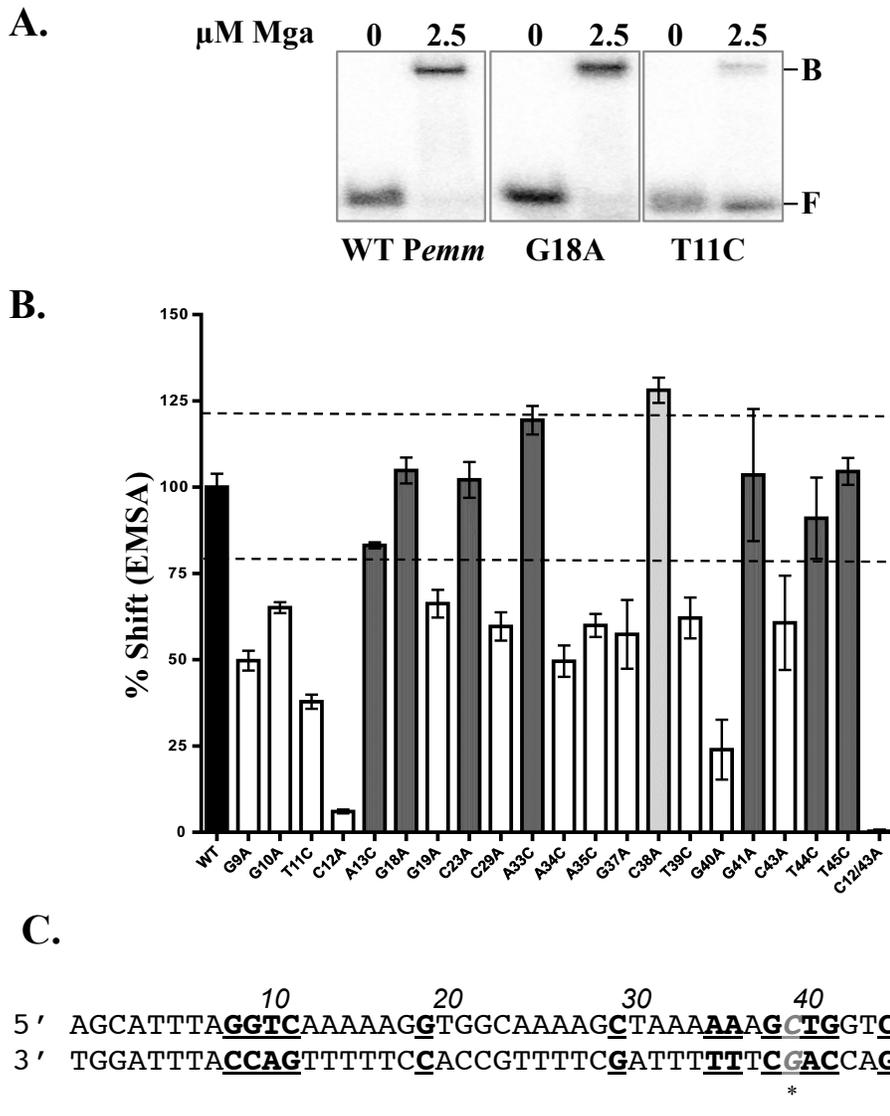


Figure 9 DNA-binding activity of *Pemm* site-directed mutations

EMSA analysis of strains with *Pemm1* 49-mer point mutations compared to the wild-type strain was performed to identify nucleotides important for Mga binding. Each probe (0.1 nM) was incubated with 2.5 μ M Mga1-His₆ (except the C38A 49-mer was incubated with 1.25 μ M Mga1-His₆). (A) Representative EMSA results for *Pemm1* wild-type (WT), G18A, and T11C 49-mer probes with Mga1-His₆ (2.5 μ M) or without Mga1-His₆ (0) are shown. The positions of free (F) and bound (B) bands are indicated to the right of the gel. (B) Quantification of EMSA results for each *Pemm1* 49-mer point mutant (shown below the bars) as determined by densitometry and shown as percent shifted by Mga1-His₆ compared to wild-type *Pemm1*. Mutants that shift less than the wild-type (white), comparable to the wild-type (dark gray), and greater than the wild-type (light gray) are indicated. The broken line denotes 80% and 120% of wild-type binding. (C) Schematic diagram showing the locations of all nucleotides within the *Pemm1* 49-mer identified as important for DNA binding (bold and underlined). The C38A mutation leading to increased binding is shown in gray italic underlined letters with an asterisk.

mutant C12/43A MBS 49-mers (Figure 9B, white bars). The *Pemm1* C38A MBS 49-mer was found to have a wild-type shift when incubated with 2.5 μ M protein/0.1 nM probe (data not shown); however, when incubated with 1.25 μ M protein/0.1 nM probe, *Pemm1* C38A MBS 49-mer bound 128% of the probe compared to the wild-type (Figure 9B, light gray bars). A summary of all the DNA-binding results is provided using the *Pemm* sequence (Figure 9C).

Conservation of critical *Pemm1* nucleotides in other category A Mga-regulated promoters

A goal of this study was to use our in-depth analysis of *Pemm1* to determine whether these results could be used to predict important nucleotides in other category A binding sites. To test this, directed mutations were subsequently made in *PscpA* (C5a peptidase gene promoter) and *Psic* (secreted inhibitor of complement gene promoter) MIT1 Mga binding sites. Three conserved nucleotides were chosen for analysis, C12A, G40A, C43A, and a double mutation C12/43A, that had exhibited both binding and activation defects in *Pemm*, and were located at either end of the binding site. Luciferase reporter assays using wild-type and mutant *PscpA-luc* and *Psic-luc* alleles were performed as described above (Figure 10A to C). The C12A mutation showed widely variable impacts in the various promoters, with 12% of wild-type activity in *Pemm1*, yet 16,265% of wild-type activity in *PscpA* and wild-type levels in *Psic*. The G40A mutation had decreased luciferase expression in *PscpA* similar to *Pemm*, but dramatically increased expression (1,179% of wild-type) in *Psic*. Only the C43A single mutation and the C12/43A double mutation resulted in a comparable decrease in promoter activity from all three promoters compared to their wild-type allele. EMSA analysis was

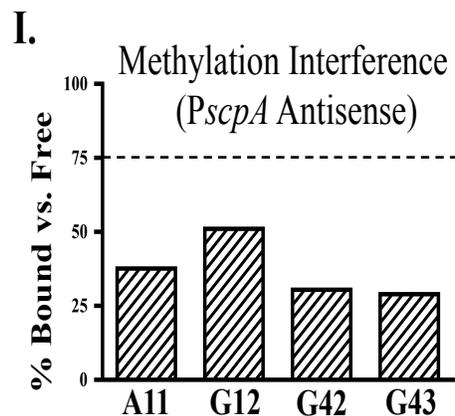
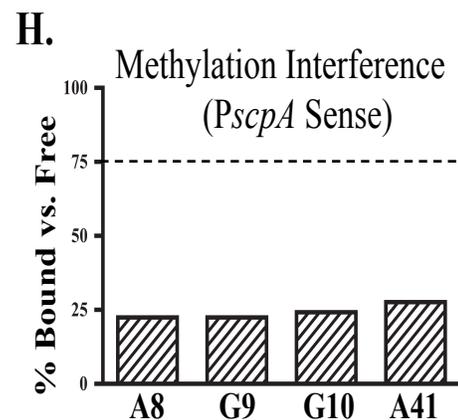
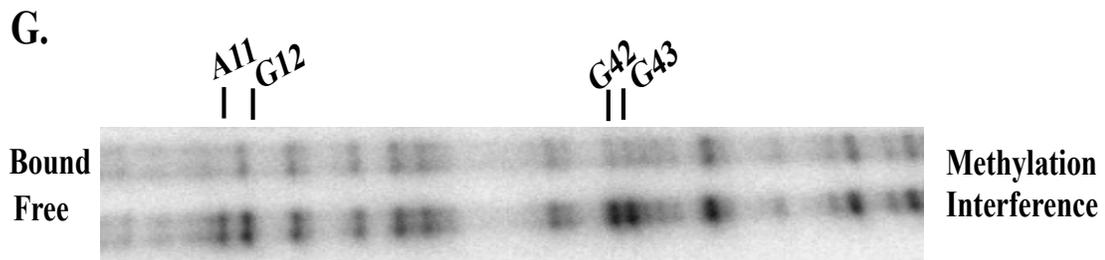
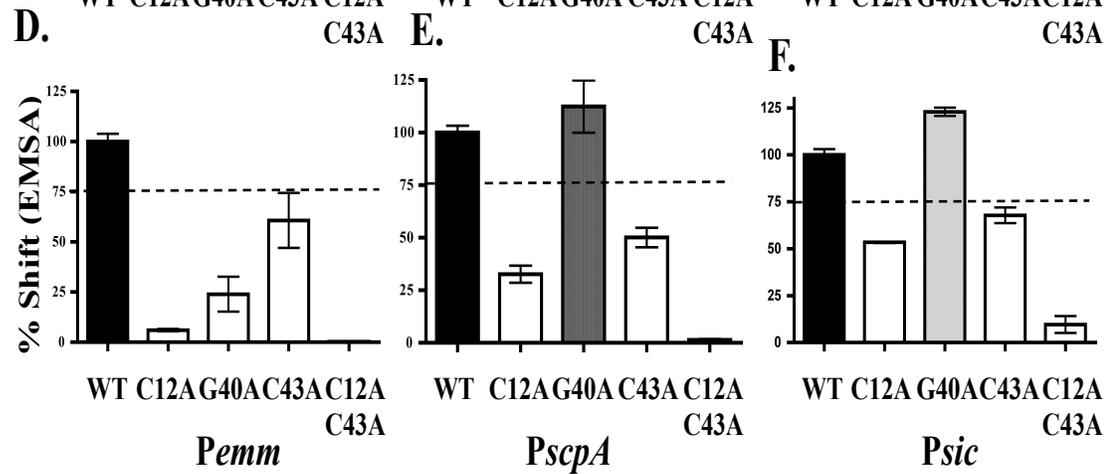
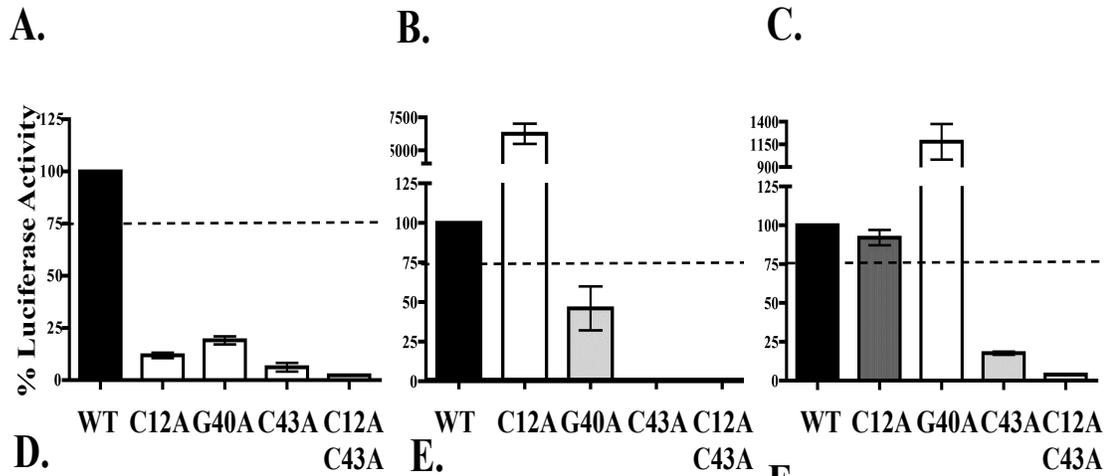


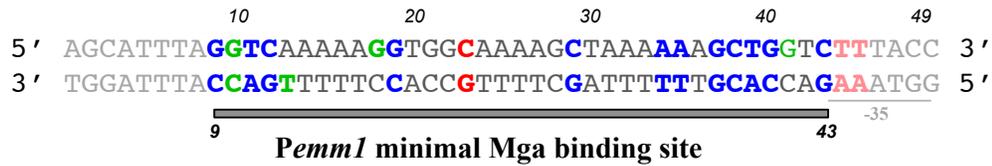
Figure 10 Role of functional *Pemm* nucleotides conserved in *PscpA* and *Psic*

Three nucleotides conserved in *Pemm*, *PscpA*, and *Psic* (C12, G40, and C43) and found to be important for both activity and DNA binding in *Pemm1* were chosen for analysis. (A to F) Single mutations (C12A, G40A, and C43A) and the C12A C43A double mutation in *Pemm* (A), *PscpA* (B), and *Psic* (C) were assayed for promoter activity by luciferase reporter assay (A to C) and DNA binding by EMSA (D to F) for each promoter. Quantification of RLU is shown as percent luciferase activity compared to the respective wild-type promoter. Mutants showing less than 75% wild-type activity (light gray bars), 75 to 100% of wild-type activity (dark gray bars), and greater than wild-type activity (white bars) are indicated. Quantification of EMSA *Pemm* (D), *PscpA* (E), and *Psic* (F) is shown as percent shifted by Mga1-His₆ compared to the respective wild-type. Mutants that shift less than the wild-type (light gray bars), comparable to the wild-type (dark gray bars), and greater than the wild-type (white bars) are indicated. (G) Methylation Interference on the *PscpA* Antisense strand. (H and I) Quantitation of methylation interference assays on the sense (H) and antisense (I) strands of *PscpA*. Nucleotides exhibiting a reduction in percentage bound versus free presented from 5' to 3'. The values are averages of two independent experiments. The broken line in all panels denotes 75% of either wild-type binding or luciferase activity.

performed on the same mutations introduced into a *PscpA* MBS 49-mer and a *Psic* MBS 49-mer (Figure 10D to F). The strain with the C12A mutation shifted less than the wild-type did for all three binding sites, despite the fact that normal (*Psic*) and even increased (*PscpA*) expression was observed in the cognate luciferase reporter assays. The G40A mutation resulted in normal wild-type binding in *PscpA* that did not correlate with luciferase results. However, the *Psic* G40A mutant showed 123% of wild-type binding that mirrored the increased *Psic* G40A luciferase expression. Finally, the C43A and C12/43A probes had a decrease in the amount of protein shifted for all three promoters that correlated directly with reduced luciferase activity. Overall, C43 appears to play a conserved role in both binding and transcriptional activation in all category A Mga-regulated promoters tested. In contrast, C12 and G40 impacted Mga activation and binding differently between the three.

Given the observed variability in the importance of *Pemm1* residues conserved in other category A promoters for Mga binding, we performed a methylation interference assay on *PscpA* as previously described for *Pemm1* (Figure 10G). On the sense strand,

A.



B.

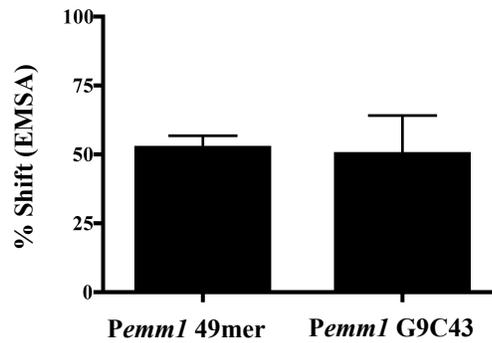


Figure 11 Summary of *Pemm* nucleotides important for Mga binding and activity

(A) Schematic diagram summarizing nucleotides important for Mga-dependent activation (red), Mga binding (green), or both (blue) identified in this study. Based on these results, a proposed minimal *Pemm1* Mga binding site of 35 bp from C9 to C43 is indicated by bar below and bold sequences, with nucleotides not essential for binding and activation in faded font. (B) Quantification of EMSA comparing Mga1-His₆ binding to *Pemm1* 49-mer and *Pemm1* G9C43 probes. Data are presented as a percent shift of the total probe.

the nucleotides A8 (64%), G9 (22%), G10 (24%), and A41 (28%) were identified as important for Mga1-His₆ binding (Figure 10H), whereas on the antisense strand, A11 (38%), G12 (51%), G42 (30%), and G43 (29%) were critical (Figure 10I). G9 and G10 (sense) and A11 and G12 (antisense) were identified in both *Pemm1* and *PscpA*, suggesting that they play comparable roles. However, the identified A8 (sense) and G42 (antisense) in *PscpA* are irrelevant thymines (T8 and T42) in *Pemm1*. Nucleotides at position 41 were identified as important for binding in both *Pemm1* (G41) and *PscpA* (A41) but were different residues. While G40 was important in *Pemm1* (Figure 10D), it was not identified by methylation interference in *PscpA* and gave an opposite EMSA

result when mutated (Figure 10E). Finally, G43 on the sense strand was identified in *PscpA*, but not *Pemm*; however, the cognate sense strand C43A mutation resulted in decreased binding in both promoters. These data further support the conclusion that while Mga does utilize conserved residues for binding at different category A promoters, overall binding occurs in a promoter-specific context.

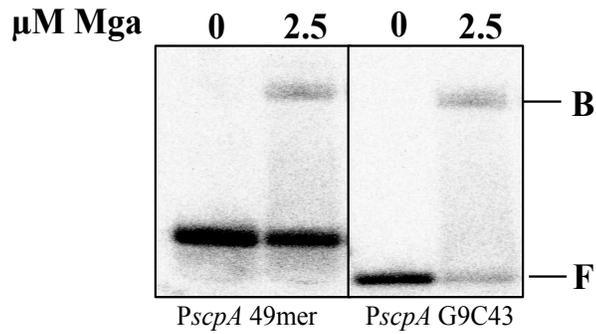
Overall, we propose that the minimal nucleotides within *Pemm1* critical for proper interaction with Mga should encompass the bases required for both binding and activation (Figure 11A, gray bar), resulting in a smaller 35-bp binding region from G9 to C43. In support of this hypothesis, EMSA analyses comparing this minimal *Pemm1* G9C43 35mer probe to the larger *Pemm1* 49-mer probe using Mga1-His₆ revealed that they had essentially identical binding profiles (Figure 11B). EMSA analysis was then performed to determine if the *PscpA* and *Psic* possessed the same minimum binding site. Oligonucleotide probes containing *PscpA* G9 to C43 and *Psic* G9 to C43 were generated. The *PscpA* G9C43 34mer shifted 48.1% of the *PscpA1* MBS 49mer (Figure 12A) while the *Psic* G9C43 34mer shifted 35.2% compared to the *Psic1* MBS 49mer (Figure 12B) and therefore are not minimum binding sites.

As the *Pemm* binding site indicated that Mga might bind as a dimer, the binding site was divided to investigate half-site binding. Oligonucleotide probes containing 0.1 nM *Pemm* 3-20, *Pemm* 28-47 and the *Pemm1* MBS 49mer, and 0.05 nM each *Pemm* 3-20 and *Pemm* 28-47 were incubated with 2.5 μ M Mga1-His₆. After an overnight exposure no shift was detected for either half site alone or in combination (Figure 12C).

Discussion

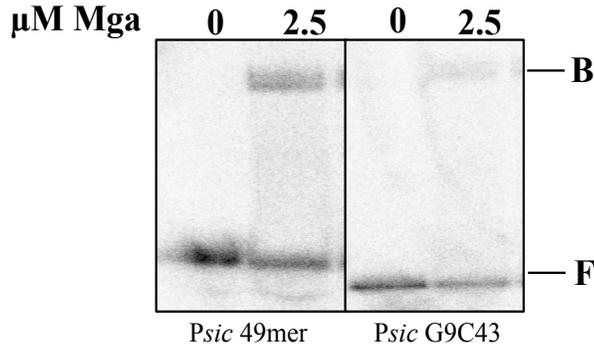
This study identified 34 separate nucleotides within the 45-bp *Pemm1* Mga

A. *PscpA1*



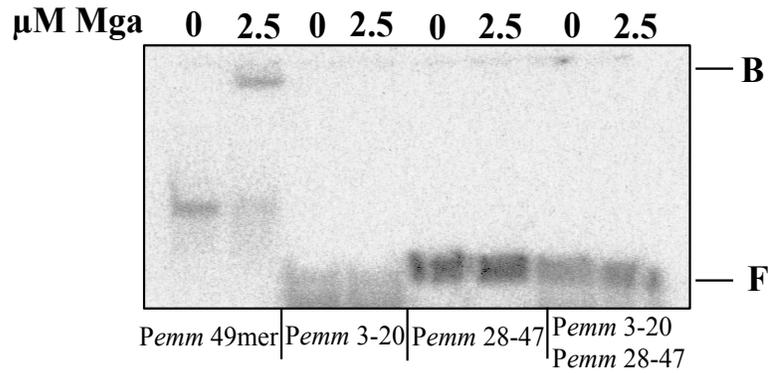
10 20 30 40 49
 5' GGATAAGGAGGGTCACAACTAAGCAACTCTTAAAAAGCTGACCTTTACT 3'
 3' CCTATTCTCCAGTGTGGATTTCGTTGAGAATTTTCGACTGGAAATGA 5'
 -35

B. *Psic1*



10 20 30 40 49
 5' GTAAGGAGAGGGTCACAACTAAACAACCTCTTAAAAAGCTGACCTTTACT 5'
 3' CATTCTCTCCAGTGTGGATTGTTGAGAATTTTCGACTGGAAATGA 3'
 -35

C. *Pemm1*



10 20 30 40 49
 5' AGCATTTAGGTCAAAAAGGTGGCAAAA GCTAAAAAGCTGGTCTTTACC 3'
 3' TGGATTTACCAGTTTTTCCACCGTTTTTCGATTTTTTCGACCAGAAATGG 5'
 -35

Figure 12 Category A Minimum binding sites

EMSA analyses of the *PscpA* and *Psic* minimum binding site and the *Pemm1* half sites. 0.1 nM of each oligonucleotide probed was incubated with 2.5 μ M Mga1-His₆. Nucleotides in bold indicate the minimum binding site. (A) EMSA analysis of the *PscpA1* G9C43 34mer versus the *PscpA1* MBS 49mer probe. (B) EMSA analysis of the *Psic1* G9C43 34mer versus the *Psic1* MBS 49mer. (C) EMSA analysis of the *Pemm1* 3-20 and *Pemm1* 28-47 half-sites.

binding site established by DNase I footprinting (Figure 5) that contribute to either DNA binding, transcriptional activation, or both (Figure 11A, colored nucleotides). Some nucleotides (C23, T44, and T45) and their complementary antisense bases contribute only to Mga-dependent transcriptional activation (Figure 11A, red nucleotides). Nucleotides G10 and G18 (sense strand) and C10 and T13 (antisense strand) show a contribution to binding by at least one biochemical method yet have only minor effects on transcriptional activation (Figure 11A, green nucleotides). The nucleotides G9, T11, C12, G19, C29, A34, A35, G37, C38, T39, G40, and C43 (sense strand), along with their complementary bases (antisense strand), had effects on both binding and transcriptional activation (Figure 11A, blue nucleotides). Therefore, the most common phenotype reflected in this study showed mutations that both reduced binding and activity.

The nucleotides identified within *Pemm1* necessary for both binding and activation are biased toward guanines and cytosines (66.7%) compared to the overall G+C content (37.5%) found within the initial 45-bp binding site. Interestingly, most of the bases that are not required for Mga binding or activation are found as runs of 4 to 6 adenines (sense strand) that could be functioning to orient Mga to the DNA, as spacer regions between the points of direct contact, or introducing curvature [117]. The methylation protection and interference assays (Figure 7) predominately identified guanine residues located within the major groove of the DNA helix as important for Mga

binding. Specifically, direct interactions are suggested to occur in the major groove at G9, G10, G18, G19, G40, and G41 (sense strand) and G12 (antisense strand). The predominant DNA-binding domain of Mga (wHTH-4) is a winged helix-turn-helix domain that would be expected to use its recognition helix to contact nucleotides in the major groove [72,79]. Furthermore, the charged residues within the Mga recognition helix are lysine (positions 5 and 9) and arginine (position 6), which have been shown in other wHTH proteins to form hydrogen bonds primarily with guanines at N7 or O6 in the major groove [79]. This corresponds with our guanine methylation assays, since they target N7 in the major groove. Some minor groove interactions were identified at A11 and A39 on the antisense strand; however, these can result from DNA interactions with the C-terminal β -strand “wing” of the wHTH domain [113,116]. Nucleotides G18 and G19 are subject to hypercleavage by DNase I footprinting upon Mga binding (Fig. 2.1C, *Pemm1*, asterisks) and may indicate a location of DNA bending. It is possible that methylation of G18 and G19 may actually prevent this flexibility and indirectly lead to the observed reduction in Mga binding and activation. Interestingly, the methylation interference assay performed on *PscpA* did not show a potential bend, which may suggest that the flexibility of the DNA affects Mga’s ability to activate transcription.

Most of the critical nucleotides in *Pemm* are found clustered at the 5’ and 3’ ends of the binding site with a few dispersed between the ends (Figure 11A). Combined with the large size of the DNase I-protected region (45 bp), this suggests that Mga might interact with DNA as a dimer despite the lack of any apparent dyad symmetry. Recently, we were able to show that Mga can form dimers in solution and that this self-interaction occurs *in vivo* [73]. Interestingly, although the dimerization of the protein is necessary for

transcriptional activation, it does not change the affinity with which Mga interacts with DNA promoter targets. A newly available crystal structure for the *Enterococcus faecalis* Mga-like regulator EF_3013 (Protein Data Bank [PDB] accession no. 3SQN) showed that this ortholog also formed a homodimer in the absence of bound DNA and possessed an amino-terminal wHTH DNA-binding domain in each monomer. Using the PyMol molecular visualization system (www.pymol.org), the wHTH recognition helices in each dimer were estimated to be approximately 95 Å to 100 Å apart (data not shown), corresponding to about 30 nucleotides ($3.4 \text{ Å} \times 30 = 102 \text{ Å}$). Although this is slightly smaller than the 35 bp predicted for the minimal Mga binding site (Figure 11A), it is based on an orthologous protein, and it does support the hypothesis that Mga and related regulators might interact with target DNA at two distinct sites within the binding region. Further studies will be necessary to confirm the stoichiometry of Mga molecules in this interaction but initial studies suggest that Mga cannot bind independently to these potential “half sites” in *Pemm1*.

As discussed above, the majority of mutations (24/34) demonstrated both reduced Mga binding *in vitro* and reduced activation *in vivo* (Figure 8, 2.5, and 2.7A). This supports a model where less Mga bound to a promoter leads to less transcriptional activation of that promoter. Even when the mutation led to an increase in Mga binding to *Pemm* (C38 mutant), the resulting *Pemm-luc* activity was also increased over that of the wild-type. However, the G9 mutant presented with decreased Mga binding yet showed an increase in Mga-dependent transcriptional activation (Figure 8 and 2.5). It is possible that while Mga has less affinity for this mutation, it may still be positioned to interact with RNA polymerase, and the lower binding affinity may enhance promoter clearance,

leading to an increase in activity. Regardless, this suggests that the exact role of G9 in Mga binding and activation is more complex and will require further investigation.

Interestingly, there does not always appear to be a direct correlation between the amount of DNA bound versus the amount of transcriptional activation. For example, the binding site with the C12A mutation shifted 6.03% and had 11.2% of wild-type RLU, while the binding site with the C23A mutation shifted 59.65% and had 2.62% of wild-type RLU. The location of the mutation could potentially change the orientation of one dimer to another, affecting how Mga interacts with RNAP and influences transcription. In this case, the effect on transcription would be cumulative with the effect on binding. The T44C and T45C mutations both decreased transcription levels *in vivo* without altering Mga binding (Figure 8 and 2.5), which was predicted, as both residues are part of the -35 hexamer recognized by RNA polymerase. Since these nucleotides are also outside the 35-bp minimal Mga binding site (Figure 11A), it suggests that they are protected from DNase I digestion but do not contribute to direct protein-DNA contacts. The C23A mutation also showed a decrease in transcription but no effect on binding ability. As with the G9 mutation discussed above, future studies will focus on how much and where Mga binds DNA contributes to transcriptional activation. Combination mutants of the up transcriptional mutations G9A and C38A with a strong down mutation such as C12A or C23A could also be used to dissect how different mutations combine to affect both binding and transcriptional activity and whether one mutation can compensate for another.

The MIT1 *Pemm1* binding site was chosen for analysis as a possible paradigm for how Mga binds to DNA at other similar Mga-regulated promoters. To test this

possibility, we made directed mutations in the category A binding sites for *PscpA* and *Psic* using conserved nucleotides found to be essential for Mga-*Pemm1* interactions (Figure 10). Interestingly, the phenotypes varied considerably between *PscpA*, *Psic*, and *Pemm* for both Mga binding *in vitro* and promoter activation *in vivo*. A C43A mutation and a C12/43A double mutation had the same effect at each of the three category A promoters, suggesting that Mga may interact at this nucleotide in a conserved manner at each target. However, this was not the case for the other two conserved nucleotides. A C12A mutation resulted in a decrease in binding at all three promoters, but *in vivo* activity varied considerably (Figure 10). A G40A mutation had the greatest variation between promoters with wild-type binding and reduced activation in *PscpA* compared to increased binding and activation in *Psic*. Methylation interference assays performed on the *PscpA* binding site further demonstrate that Mga interactions with its promoters are only partially conserved. Of the 7 nucleotides identified in *PscpA*, 3 were unique to this promoter. Interestingly, *PscpA* has an inverted trinucleotide repeat of GGT. This pattern is only partially conserved in *Pemm1*; the repeat is present at the 5' end on the binding site, but the sequence differs at the 3' end.

A true minimum-binding site will contain the entire sequence necessary for DNA binding and have the same shift as wild-type. EMSA analysis found that the *Pemm1* G9C43 35mer did contain all sequence necessary for a wild-type shift, but the *Psic* G9C34 34mer and the *PscpA* G9C43 34mer did not. The methylation interference assay of *PscpA* indicated that A8 was important for binding, which is not encompassed by the *Pemm* minimum site. This nucleotide is also present in *Psic*. Due to variation in the binding sequence, the *PscpA* and *Psic* G9C43 is only 34 bp and including the A8 would

make these sites the same size as the *Pemm* G9C43 35mer. This further indicates that even within promoters of the same category, Mga interacts with each binding site in a unique manner.

It can be said that all of the conserved *Pemml* nucleotides did have some importance for Mga interactions. However, these results show that *Pemml* serves only as a general model for identifying important Mga contacts in other category A promoters. As Mga appears to interact differently with each of its promoters, detailed analysis of these interactions would need to be determined for each individual promoter.

Chapter 4

Interaction of Mga with RNA Polymerase

Introduction

Bacterial transcription factors that bind DNA commonly enhance transcription by directly stabilizing or recruiting RNA polymerase to the promoter by protein-protein interactions. Most of these can be simply categorized as Class I or Class II transcription factors, although other interactions may occur. Class I proteins interaction with the α -CTD (C terminal domain); the classic example is CRP, cyclic AMP receptor protein, at the *lac* promoter [95]. Class II transcription factors stabilize interactions through protein-protein interactions with domain 4 of σ factor, and includes the transcription factors PhoB and AraC [97]. These factors may act independently or in concert, and some factors do not interact with the holoenzyme itself.

The Category A Mga binding site is centered at -54 from the start of transcription and overlapping the -35 hexamer [78], which suggests that Mga is positioned to interact directly with RNA polymerase to activate transcription. If Mga acts as a class I transcription factor, the α -CTD will be necessary to stabilize its binding at these promoter and obtain Mga-dependent transcriptional activation[96]. Alternatively, Mga may function as a class II transcription factor, and domain 4 of σ factor would be necessary for the same result. The following studies present preliminary evidence that Mga does not interact solely with either the α -CTD or domain 4 of σ factor, which suggests that at the Category A promoters, Mga may interact with jointly with the α -CTD and σ domain4, or with a different subunit of the holoenzyme.

Results

Bacterial 2 Hybrid Assay for Protein-Protein Interactions

In order to study protein-protein interactions *in vivo*, a bacterial two-hybrid system was used. In this system the two catalytic subunits of the *Bordetella pertussis* adenylate cyclase, T18 and T25, are fused to the proteins of interest [118]. These plasmids are then co-transformed into an *E.coli* strain lacking adenylate cyclase (*cya*). If the T18 and T25 are brought into close contact through protein-protein interactions, cAMP is made, which activates CRP, and thus the *lac* and *mal* operons. *E. coli* that can now use lactose or maltose can be identified by a color change of white to blue or from white to pink.

A comparison of the Δcya *E. coli* strains DHM1 and BTH101 on LB-X-gal-IPTG and MacConkey Maltose plates showed that the BTH101 strain on MacConkey Maltose-IPTG gave the most consistent phenotype. On the MacConkey Maltose plates, positive colonies were characterized by a deep magenta color, and the media remained red. Negative colonies were clear or yellow, and the surrounding media turned a yellowish color due to a pH increase that accompanies the utilization of peptones for energy. To assay for protein-protein interactions, equal amounts of the plasmids pT18N-link/pT25N-link, pT18N-Mga/pT25N-Mga, pT18N-his₆- α /pT25N-Mga and pT18N-his₆- σ /pT25N-Mga were transformed into BTH101, plated on MacConkey Maltose-IPTG to obtain countable colonies and placed at 30°C. Colonies were considered positive for protein-protein interactions if they turned magenta while the negative controls remained yellow, over 3-5 days. Mga-Mga colonies displayed a strong magenta color within 24 hours of plating, as

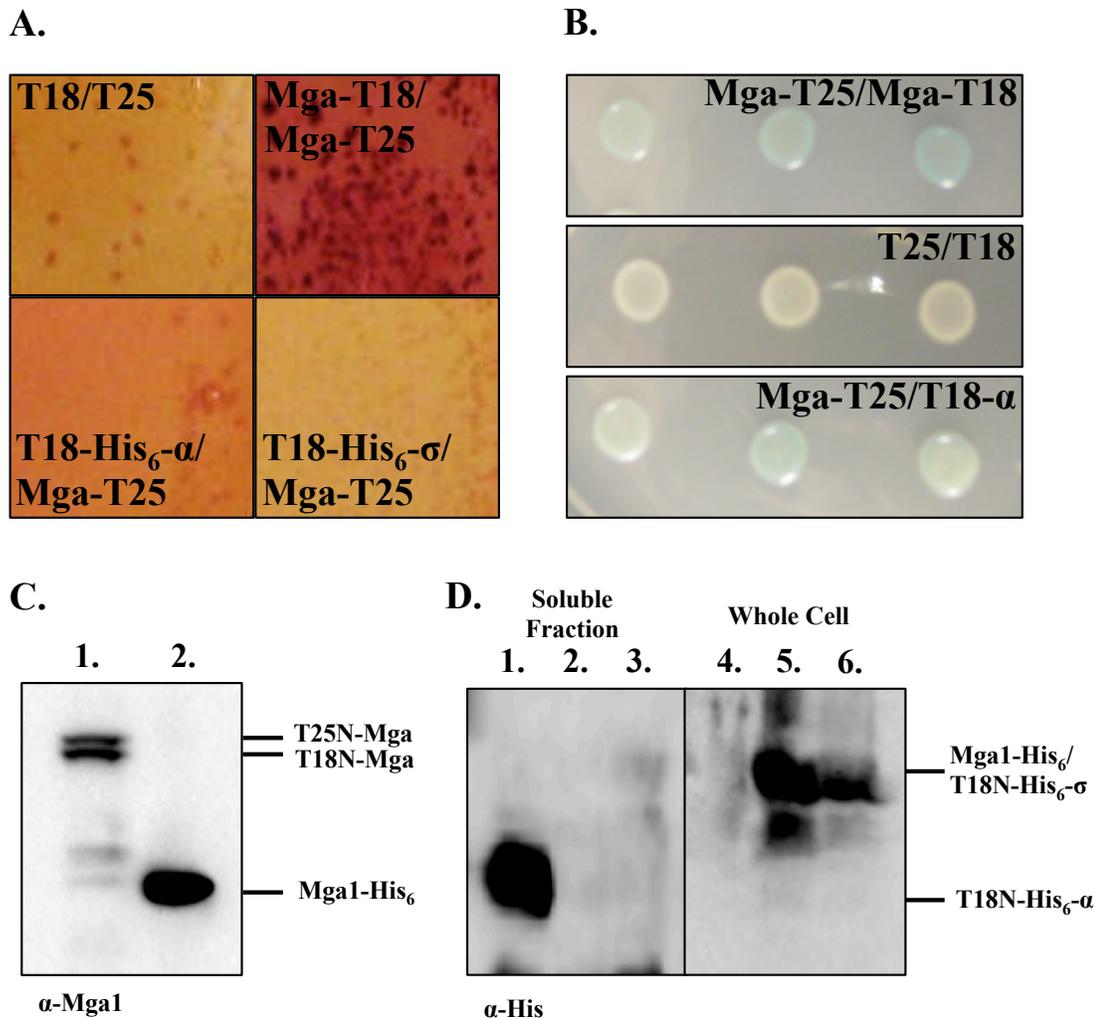


Figure 13 Bacterial 2-Hybrid Assays

(A) pT18N-link/pT25N-link, pT18N-Mga/ pT25N-Mga, pT15N-his₆- α /pT25N-Mga and pT18N-His₆- σ /pT25N-Mga in the strain BTH101 on MacConkey Maltose media grown at 30°C for ~ 4 days. Negative colonies are pale pink to yellow on a yellow background while positive colonies appear as magenta on a red background. (B) pT25N-Mga/pT18N-Mga, pT25N-link/pT18C-link, pT18N- α /pT25N-Mga in the strain DHM1 plated on A+B minimal media at 30 °C. (C) Western blot showing the full-length Mga-T18 and Mga-25 protein expressed in BTH101 at ~52 hours growth in LB at 30°C, with the Mga1-His₆ protein for a size comparison. (D) Expression of pT18N-His₆- α and pT18N-His₆- σ grown in ZYP-5052 for ~52 hours at 30°C. 1, 2 and 3 are the soluble fraction of T18-His₆- α , T18-His₆- σ and Mga1-His₆ grown under the same conditions, 4, 5 and 6 are the whole cell sample of T18-His₆- α , T18-His₆- σ and Mga1-His₆.

expected as Mga is known to dimerize (Figure 13A) [73]. In this assay the Mga- α and Mga- σ colonies turned yellow, or were negative for protein-protein interactions.

While initial studies found a weak positive interaction between Mga and α when transformed into DHM1, grown to OD₆₀₀ of 1.6 in LB at 37°C, spotted on A+B minimal media and placed at 30°C overnight, this result could not be replicated (Figure 13B). The assay was also performed after swapping the location of the tag, but these results were also negative. The δ subunit was tested for interactions, but remained again the result was negative.

Western blots were performed to determine if full-length, soluble proteins were expressed in the two-hybrid assay. LB-IPTG-X-gal was inoculated with pT18N-link/pT25N-link or pT18N-Mga/pT25N-Mga, and grown at 30°C. Samples were taken every ~12 hours for ~52 hours. When probed with α -Mga1 antibody, a doublet corresponding to Mga-T18 and Mga-T25 was detected faintly at ~24 hours and showed an intense band at ~52 hours (Figure 13C). Antibodies against the T18 and T25 domains of adenylate cyclase were tested but these antibodies had a strong cross-reaction in DHM1, BTH101 and C41 *E. coli* backgrounds. pT18N-His₆- α and His₆- σ were then constructed. As the α -His antibody also had cross reaction to BTH101, pT18N-His₆- α and pT18N-His₆- σ were transformed into C41[DE3] and grown in ZYP-5052 at 30°C and assessed for protein production and solubility. Full-length proteins were detected in the whole cell lysate, but only His₆- α was detected in the soluble fraction (Figure 13D). While the α subunit could be available to interact with Mga during the assay, the negative result for Mga- σ factor was due to an artifact of the assay, and was inconclusive.

Creation of strains for *in vivo* protein-protein interaction studies

To study protein-protein interactions between Mga and σ factor *in vivo* in GAS, a replicating plasmid expressing a N terminal His₆- σ or a His₆- σ - Δ 4 under the control of *PrpsL* (pKSM284 and pKSM285) was transformed into the M4 GA40634 GAS strain. Each strain was grown until mid-logarithmic phase and the lysate was passed over a NiNTA agarose column to verify that each protein was expressed in GAS (Figure 14A).

To study the interaction between Mga and α , the His₆- α and His₆- α - Δ CTD were also cloned into a replicating plasmid under *PrpsL*. Transformations of >250 μ g of plasmid failed to produce colonies. The plasmid pMSP3535-H3, which contains a nisin inducible promoter was then used to determine if controlling the expression of the protein would overcome the problem of transformation. However, this plasmid was also untransformable.

In order to overcome this limitation the suicide plasmid pCIV2, which is kanamycin resistant but cannot replicate in GAS, was used. ~1 kb of identical upstream DNA was PCR SOEed to His₆- α - Δ CTD under the expression of the artificial promoter *Pami*, which created the plasmid pKSM281. After transformation, homologous recombination allows for a strain with either a His₆- α and α - Δ CTD or a His₆- α - Δ CTD and the wild-type α . This strategy was used to maintain a full-length copy of α in the genome. The presence of the plasmid and the 5' junction with gDNA was detected by colony PCR. Due to cross reaction with the α -His antibody, Western blotting was inconclusive. To determine if the N-terminal His tag was an available target and which variant of α was his tagged, a GA40634.pKSM281 strain was grown to mid-logarithmic phase, then purified over a NiNTA column. Bands corresponding to the β and β' subunit

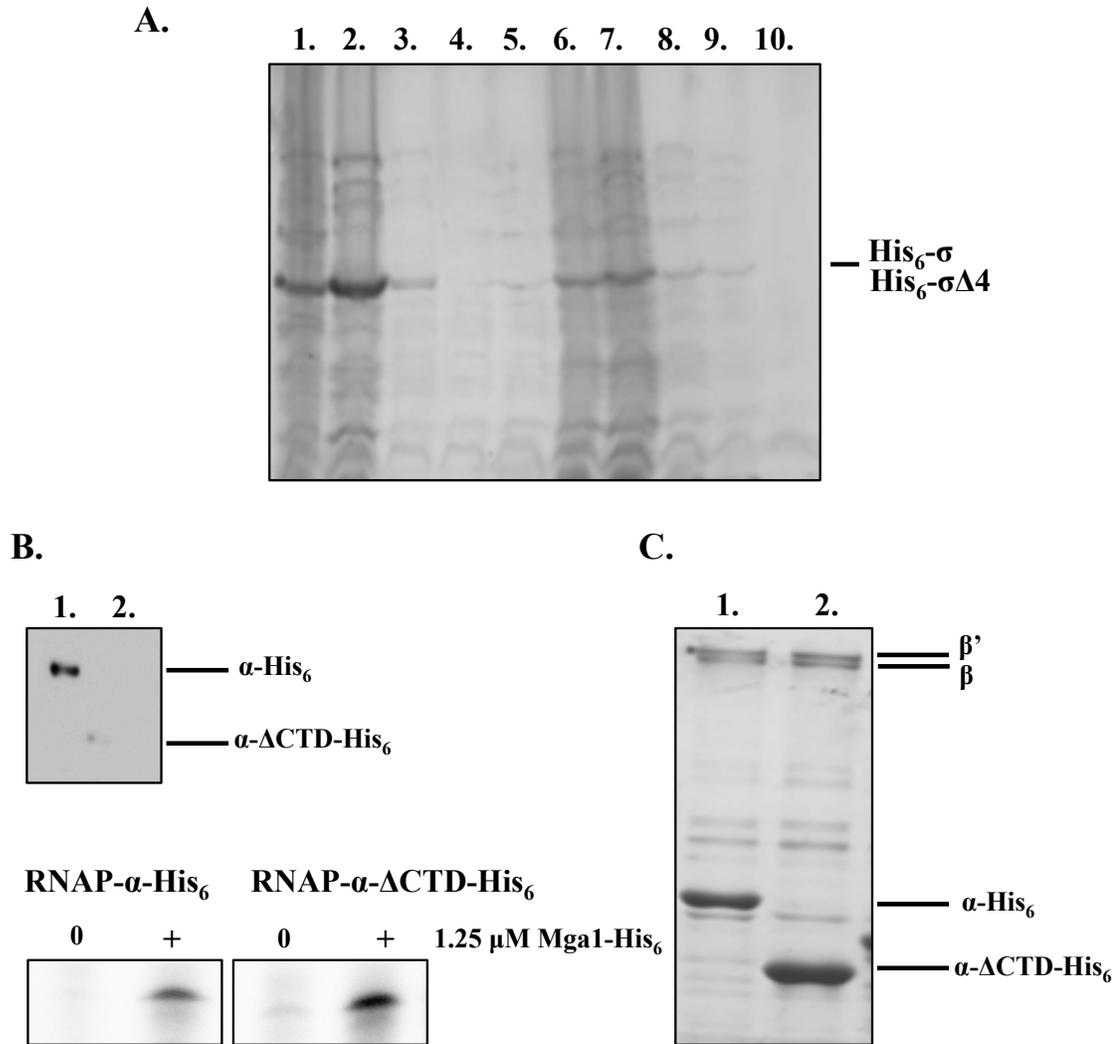


Figure 14 Creation of Mutant RNA Polymerases

(A) Coomassie gel depicting His₆-σ purified from GAS 1. Pellet 2. Eluent. 3. Wash 1. 4. Wash 2. 5. Fractions or His₆-σ-Δ4 6. Pellet 7. Eluent 8. Wash 1 9. Wash 2 and 10. Fractions. (B) Western blot for RNAP purified from 1. GA40634.pKSM295 (α-His₆) or 2. GA40634.pKSM295 (α-ΔCTD-His₆). *In vitro* transcription using the *Pemm* 232 template of the RNAPs purified from GA40634.pKSM295 and GA40634.pKSM294. (C) Coomassie blue staining of RNAP containing 1. α-His₆, or 2. α-ΔCTD-His₆ after denaturing and refolding with a >20 fold molar excess of the His tagged subunits.

were detected in the lysate by Coomassie staining, but were lost in the eluted fraction, suggesting that the His-tag was buried within the core (Data not shown).

An accessible C-terminal α-His₆ and α-ΔCTD-His₆ were then cloned into pCIV2 under the control of *Pami* and transformed into M4 GA40634, creating the strains

GA40634.pKSM295 and GA40634.pKSM294. Colony PCR detected the plasmid and the 5' junction. Active RNAP was successfully purified by over NiNTA agarose from each (Figure 14B) and His-tagged proteins of ~35 and 27 kDa could be detected by Western blotting.

Creation of Mutant RNAP for *in vitro* transcription

Mutant RNAPs were created to determine if Mga interacts with domain 4 of σ or the CTD of α . Core RNAP were purified from the strain JRS4-PolHis [98] while each His-tagged α and σ were purified from *E. coli*. Before each assay, the core was incubated with a σ factor. In order to integrate the α constructs, core RNAP was denatured with 6 M Guanidinium chloride, then spiked with a ~20 fold molar excess of either His₆- α or His₆- α - Δ CTD, then these new subunits were used as the nucleus of core formation. Excess α was removed by ultrafiltration and the polymerases were assessed by Coomassie staining and Western blot (Figure 14C).

***In vitro* Co-affinity Purification**

In vitro co-affinity purification assays were performed to study protein-protein interactions with an individual subunits or the intact holoenzyme, either in solution, or while bound to the promoter. 10 μ L of Mga4-CBP was incubated with 10 μ L of 20 μ M Mga1-His₆, His₆- α , His₆- α - Δ CTD, His₆- σ or His₆- σ - Δ 4, and purified with NiNTA agarose. The eluent, wash and purified fractions were run for Western blotting and then probed with both α -His and α -CBP. The His-tagged proteins were present in the initial and purified fractions. Mga4-CBP was found in the initial fraction of all samples, but only in the purified fraction with Mga1-His₆ (Figure 15A).

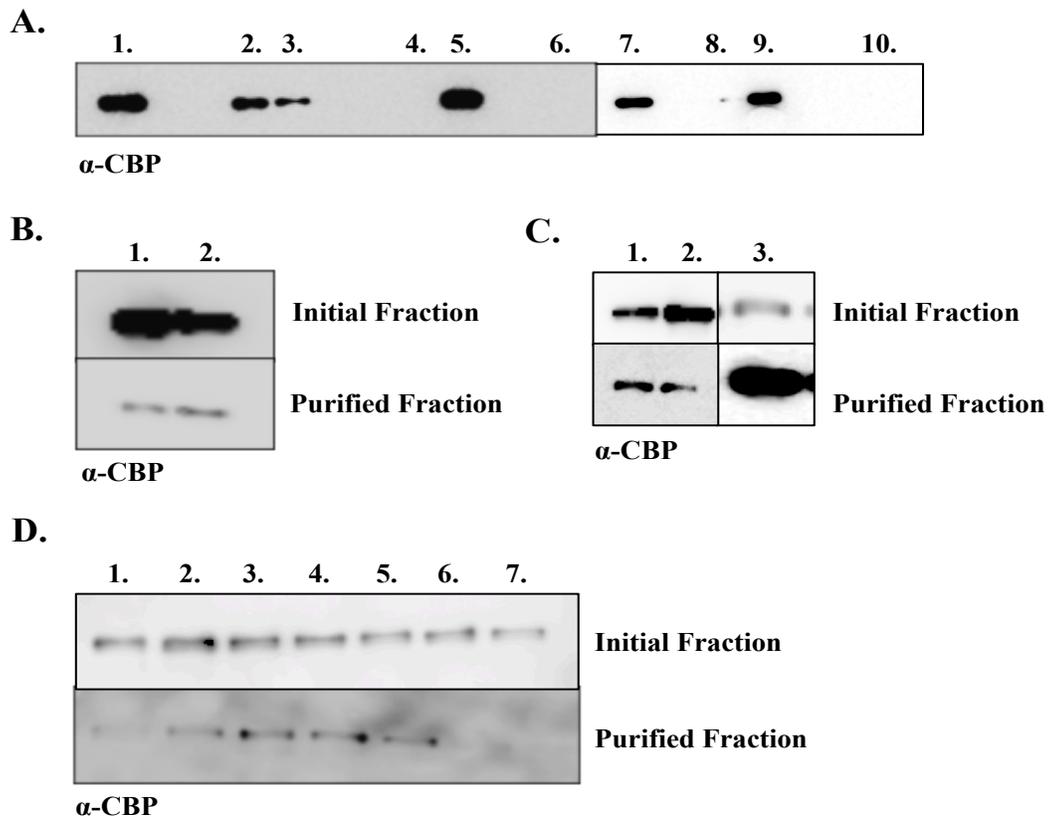


Figure 15 *In vitro* Co-affinity Purification of Mga with RNAP

(A) Western blot probing for Mga4-CBP co-purified with Mga1-His₆ 1. Initial Fraction 2. Purified Fraction, with His₆- α 3. Initial Fraction 4. Purified Fraction, with His₆- α - Δ CTD 5. Initial Fraction, 6. Purified Fraction, with His₆- σ 7. Final Fraction 8. Purified Fraction, with His₆- σ - Δ 4 9. Initial Fraction, 10. Purified Fraction. (B) Western blot probing for Mga4-CBP co-purified with 1. Mga1-His₆, and 2. RNAP wild-type holoenzyme in solution. (C) Western blot probing for Mga4-CBP co-purified with 1. Mga1-His₆ bound to *Pemm* 232 2. RNAP bound to *Pemm* 232 and 3. RNAP bound to MBS +10 *PrpsL*. D. Western blot probing for Mga4-CBP co-purified with 1. Mga1-His₆, 2. RNAP wild-type holoenzyme 3. RNAP with His₆- σ - Δ 4 4. RNAP with His₆- α 5. RNAP with His₆- α - Δ CTD 6. His-MBP and 7. with NiNTA agarose solely.

Mga4-CBP was then incubated with the holoenzyme either in the presence of a Mga binding site. When incubated in solution, Mga4-CBP co-purified with the holoenzyme (Figure 15B). Mga4-CBP was detected when incubated with the

holoenzyme and either the *Pemm* 232 or the MBS +10 *PrpsL* 1201 template, indicating that it can co-purify when with a MBS (Figure 15C).

Mga4-CBP was then incubated in solution with the wild-type RNAP- σ , RNAP-His₆- σ - Δ 4, RNAP-His₆- α , RNAP-His₆- α - Δ CTD, and Mga1-His₆. Mga4-CBP was also incubated with His-MBP and without any His-tagged protein to serve as controls for non-specific interactions. Mga4-CBP was detected in all starting fractions. Mga4-CBP co-purified with Mga and each holoenzyme, but no Mga4-CBP was detected co-purifying with His-MBP or through non-specific interactions with the NiNTA agarose (Figure 1D).

***In vitro* Transcription of Mutant RNAP**

In vitro transcription assays were performed to determine if the deletion mutants of α or σ resulted in a loss of Mga-dependent transcriptional activation. The *PrpsL* +30 1201 PCR product and the *Pemm* 232 PCR product served as the control and Mga-dependent promoter templates, respectively. When wild-type holoenzyme was incubated with *PrpsL*, a band of ~ 257 bp was detected (Figure 16A). The addition of 1.25 μ M Mga1-His₆ had no effect on this reaction. Wild-type holoenzyme incubated with the *Pemm* 232 transcript produced a band of ~232 bp, showing basal levels of transcription. When the *Pemm* 232 template was incubated 1.25 μ M Mga1-His₆, a band of ~232 bp had a strong increase in intensity compared to the basal level of transcription of the *Pemm* promoter in the absence of Mga.

To determine if Mga interacts as a Class II transcription factor, RNAP core was incubated with either His₆- σ or His₆- σ - Δ 4 to generate holoenzymes. Three times as much His₆- σ - Δ 4 than His₆- σ needed to be added to core RNAP to activate transcription and overall this holoenzyme was much less efficient than the wild-type. The buffer for this

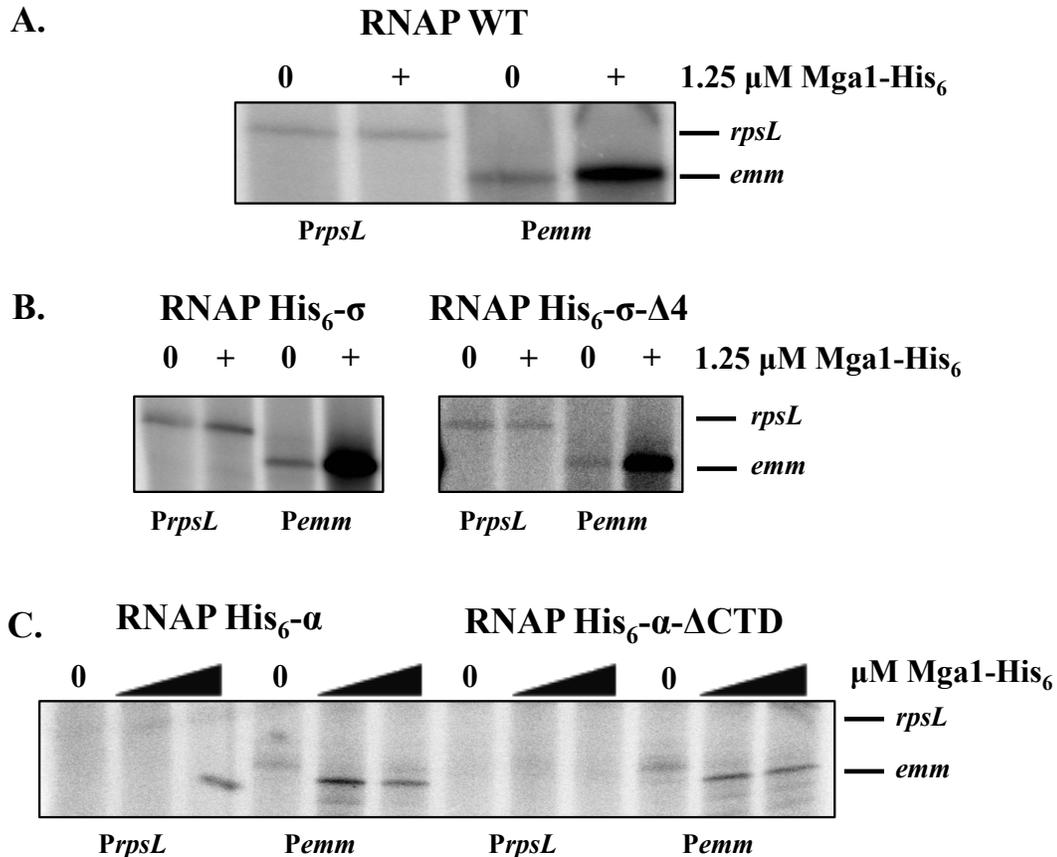


Figure 16 *In vitro* Transcription with mutant RNA polymerases

(A) *In vitro* transcription with 1.25 μ M Mga1-His₆ on the *PrpsL* +30 1201 or the *Pemm* 232 templates using wild-type holoenzyme. (B) Transcription assay using 1.25 μ M Mga1-His₆ on the *PrpsL* +30 1201 or *Pemm* 232 template using wild-type core and either His₆- σ or His₆- σ - Δ 4. (C) Transcription assay using 1.25 or 2.5 μ M Mga1-His₆ on the *PrpsL* +30 1201 or *Pemm* 232 template with either the refolded RNAP His₆- α or the RNAP His₆- α - Δ CTD with wild-type σ factor.

reaction was also adjusted to decrease the concentration of HEPES from 50 to 10 mM. These conditions were then used for both His₆- σ and His₆- σ - Δ 4 holoenzymes. RNAP-His₆- σ had no Mga-dependent transcription on the *PrpsL* +30 1201 template, but showed a strong increase in transcript when Mga1-His₆ was added to the *Pemm* 232 template. The RNAP-His₆- σ - Δ 4 holoenzyme also had an increase in of the 232 bp transcript when Mga1-His₆ was added to the *Pemm* 232 template, but no Mga-dependent transcription was observed on the *PrpsL* +30 1201 template (Figure 16B).

When *in vitro* transcription was performed using the RNAP-His₆- α and RNAP-His₆- α - Δ CTD, no transcript was detected from the *PrpsL* +30 1201 templates. However, when incubated with *Pemm* 232, each transcript showed an increase in the intensity upon the addition of Mga1-His₆ over basal levels of *Pemm* transcription (Figure 16C).

Discussion

These studies provide preliminary evidence that Mga does not activate transcription by functioning solely as a Class I or Class II transcription factor. The bacterial two-hybrid system was unable to detect protein-protein interactions between Mga and the α subunit. However RNAP is a 3-dimensional, multisubunit protein that interacts with DNA, and therefore studying the Mga- α interaction outside of this context may miss these interactions. The protein may also not be properly folded. The bacterial two-hybrid assay also was unable to detect an interaction with σ factor. While the T18-His₆- σ protein was produced, it does not appear to be in the soluble fraction. Therefore protein-protein interactions may be missed if *E. coli* places this protein in inclusion bodies where it cannot interact with its binding partner. Overall the two-hybrid system is an insufficient tool for studying interactions of these protein complexes.

Mga was able to activate transcription from *Pemm* when domain 4 of σ factor or the CTD of α was removed, indicating that these domains are not the sole point of interaction for Mga to activate transcription. However, Mga does appear to co-purify with the holoenzyme *in vitro*, suggesting that the protein-protein contacts are present. Interestingly, these interactions action can be detected in solution, in the absence of DNA, which may suggest that Mga recruits the polymerase to the promoter. Mga is

suspected to bind to DNA as a dimer; it is possible that one Mga monomer contacts the α -CTD and the other Mga monomer contacts domain 4, and that either of these protein-protein contacts is sufficient to activate transcription *in vitro*. To determine if Mga simultaneously interacts with these domains, *in vitro* transcription and co-affinity purification with the RNAP double mutant and $\Delta 139$ -Mga will be performed. The RNAP double mutant should have no Mga-dependent transcription or should not co-purify with the Mga4-CBP protein if Mga contacts both domains.

The amino acids involved in these contacts would be identified by first narrowing the region of interaction in each domain, combined with alanine scanning mutagenesis. Amino acids will also be targeted based on known interactions with other transcription factors. If Mga makes protein-protein contacts elsewhere in the holoenzyme, identifying the amino acid contact points will be more difficult. Directed mutations will be made based on other known interactions. These mutations will be assessed for their ability to activate transcription and co-purify with Mga *in vitro*. The His₆- σ and His₆- σ - $\Delta 4$ proteins will be purified from GAS and will be assessed for activity by *in vitro* transcription. Co-affinity purification *in vivo* will then be performed to confirm the biological relevance of the *in vitro* interactions.

Chapter 5

Interaction of Mga with other Promoters and Regulatory Elements

Introduction

This studies discusses preliminary work done to better understand how Mga functions as a whole within GAS. Studies have so far focused on identifying the Mga binding sites [76–78,114], identifying the key binding domains [79,119], more recently studying how the domains of the protein contribute to its regulation [73], and how Mga interacts with just one of its binding sites [120]. These studies look at various aspects of Mga binding, multimerization and activation in isolation, but ultimately the goal is to fit this work, along with previous studies into a broad understanding of Mga's function both at the promoter and cell-wide.

One aspect of these studies is to better understand the role of each domain within the protein. Mga contains two DNA binding domains, HTH-3 and wHTH-4. Directed mutations of each of these domains observed that the wHTH-4 is the essential DNA-binding domain and that HTH-3 plays a role in *Pmga* binding [79], but the role of HTH-3 overall is not known. The EIIB^{GAT} domain at the C terminus is responsible for Mga dimerization [73]. Though the $\Delta 139$ Mga EIIB truncation binds DNA with the same affinity as wild-type, it does not activate transcription *in vivo*. Mga contains two PRD domains and the histidines H204/H270 of PRD1 are phosphorylated *in vitro* by the PTS (Hondorp, et al, in review). While PTS phosphorylation *in vitro* and decrease in

expression of Mga-regulated genes had been observed *in vivo* with phosphomimetic mutants, this has not been directly linked.

When the known Mga-binding sites from all three categories of Mga-regulated promoters (*Pemm6*, *PscpA*, *PsclA*, *Pmga1*, and *Pmga2*, *Pemm1*, *PscpA*, and *Psic*) [76,77] were aligned using ClustalW, only 13.4% sequence identity was observed [120]. However after dissecting the *Pemm1* binding site, it appears that Mga interacts with each binding site in a distinct manner.

Previous microarray studies found that the Mga regulon is composed of ~10% of the GAS genome. Interestingly, as Mga had been previously known as only a transcriptional activator, in M1 SF370 103 genes and in M4 GA40634 118 genes were repressed. Initial studies scanned the mannose (*ptsA-D*) operon, glucose (*ptsG*), maltose (*malE-G*) and iron uptake (*siuADBG*) for the consensus Mga binding site, but none were identified [72]. However the EMSA analysis found no binding to any of these promoters, which suggests that, the Mga-regulation occurs indirectly. The Mga consensus sequence is weak, a consensus of *Pemm6* and *PscpA6* did not identify the Mga binding site in *Pmga6* [76] indicating that Mga binding sites may easily be overlooked. Initial ChAP, chromosome affinity purification, studies were performed to better understand the full binding profile for Mga, and how it controls its regulon.

Results

Dimerization of Mga4-His₆ and Δ 139Mga4-His₆ in solution

Previous gel filtration experiments found that Mga4-His₆ forms oligomers in solution with increasing amounts of NaCl, while the Δ 139Mga4-His₆ protein remains monomeric [73]. To verify this oligomerization, sedimentation equilibrium by analytical

ultracentrifugation (AUC) experiments were performed on Mga4-His₆ and Δ 139Mga4-His₆ in 50 mM HEPES/Citrate pH 7.5 with of 100 mM NaCl. Mga4-His₆ was prepared at 7.5, 20 and 30 μ M and centrifuged at 14, 16, 18, 20 and 22 Krpm (Figure 17A). Analysis of the data indicates an average molecular weight of 89.9 kDa, (confidence interval 83.5-96.3 kDa), significantly larger than the value of 63.2 kDa predicted. When fit to a monomer-dimer model, the equilibrium dissociation constant for dimerization (K_{dim}) was found to be 9.41 μ M (confidence interval 5.99-14.5 μ M)[73]. Analysis of the data using

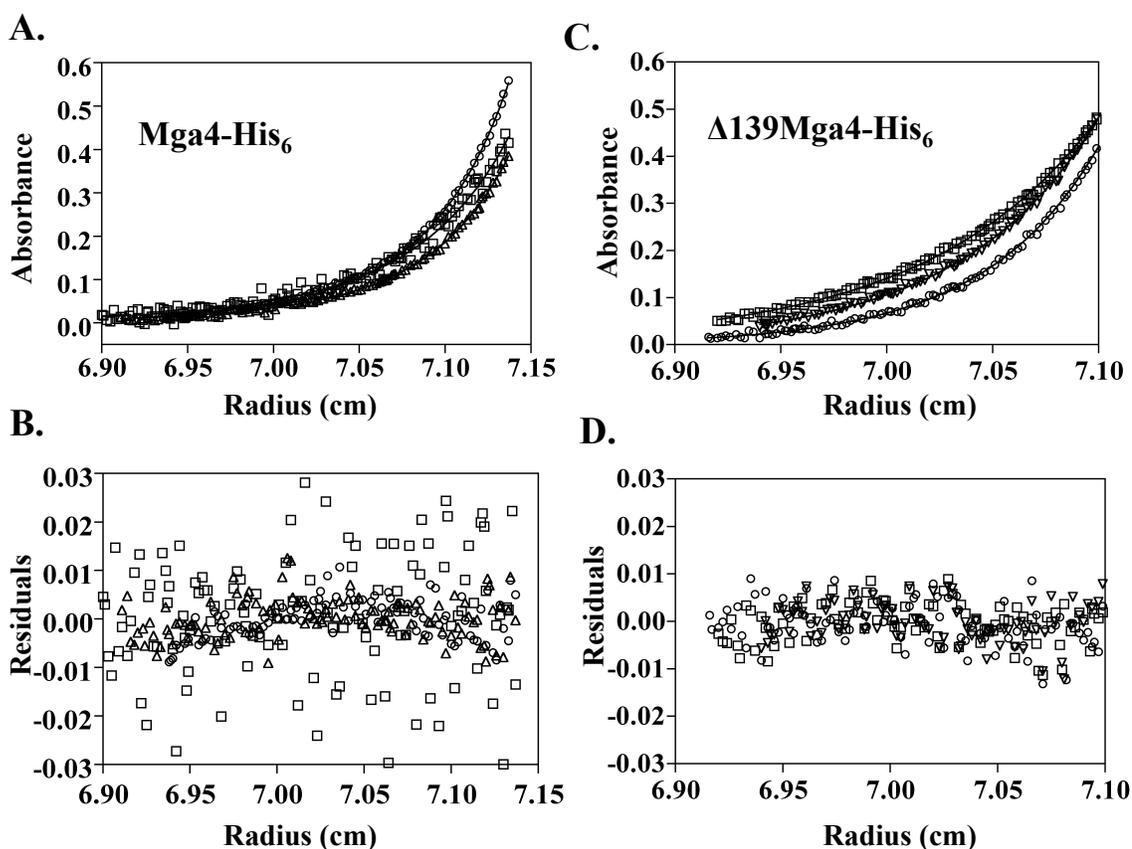


Figure 17 Dimerization of Mga4-His₆ and Δ 139Mga4-His₆

Sedimentation equilibrium of Mga4-His₆ and Δ 139Mga4-His₆ in 100 mM NaCl, 50 mM HEPES/Citrate, pH 7.5 [74]. (A) 7.5 (circle) 20 (square) and 30 (triangle) μ M Mga4-His₆ spun at 16 Krpm. Fifteen data sets were fit to a monomer-dimer model (solid lines) resulting in the residuals depicted (B). (C) 7.5 μ M of Δ 139Mga4-His₆ centrifuged at 18 (square) 20 (inverted triangle) and 22 (circle) Krpm. Six data sets were fit to a model for a single homogenous species (solid lines) resulting in the residuals depicted (D).

monomer-trimer or monomer-tetramer models indicated no improvement in the fit based on the magnitude of the square root of the variance of the fit and the residuals of the fit (data not shown). $\Delta 139\text{Mga4-His}_6$ was prepared at 7.5 and 30 μM and centrifuged at 18, 20 and 22 Krpm (Figure 17C). For the $\Delta 139\text{Mga4-His}_6$ protein the best fit of the data indicated a monomeric species in the presence of 100 mM NaCl with a molecular weight of 50.7 kDa, which matches the predicted weight of 47.2 kDa [73]. Interestingly, the AUC experiments performed with Mga4-His₆ in the absence of 100 mM NaCl indicated a K_{dim} of 6.13 μM , (confidence intervals 3.48-10.7 μM) presence of 100 mM NaCl (data not shown), suggesting that in this system salt did not have a significant effect.

PTS Phosphorylation of Mga leads to inactivation *in vitro*

Previously we found that Mga was phosphorylated *in vitro* via the PTS, and *in vivo* that the Mga phosphomimetic mutant has a decrease in the inactivation of Mga-regulated genes (Hondorp, in review). In order to directly link phosphorylation of the protein to the downregulation of its activity, first the PTS system was reconstituted *in vitro* and used to phosphorylate Mga. As expected, when EI, Hpr and Mga were incubated in the presence of PEP, a band that corresponded to P~Mga was detected (Figure 18A).

To link this phosphorylation to down-regulation of Mga-dependent transcription, the phosphorylation reaction was performed with template DNA, but in the presence or absence of PEP. Then the *in vitro* transcription was performed as described previously. In the absence of PEP, a strong increase in *emm* transcript was observed when Mga was incubated with *Pemm* compared to the Mga minus lane, while no difference was seen in the *rpsL* transcript levels. When PEP was added to the reaction, the *rpsL* transcript again

showed no difference in the presence or absence of Mga in the reaction. However, at the *Pemm* promoter, in the presence of PEP and the components of the PTS system, no Mga-dependent transcription was observed (Figure 18B). This was observed for both Mga4-His₆ and Mga1 His₆.

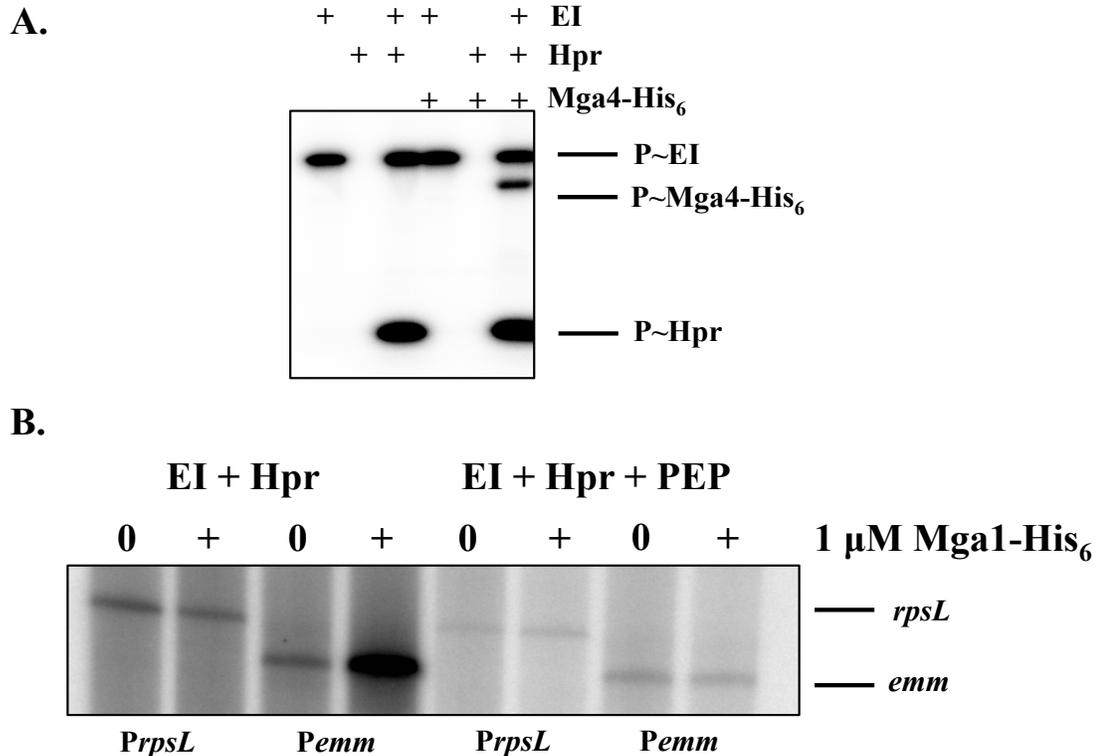


Figure 18 Effect of *in vitro* phosphorylation of Mga on transcription

(A) The reconstitution of the PTS phosphorelay *in vitro* (Hondorp, 2012, in review). In the presence of PEP, EI phosphorylates Hpr, which passes the phosphate to the PRDs of Mga. (B) *In vitro* transcription assay performed with the phosphorylated Mga1-His₆ leads to the downregulation of transcription.

Purification of Mga by TAP tagging

To enhance the purity of Mga for x-ray crystallography, a TAP, tandem affinity purification, tagged protein was created. Previous work found that Mga from an M4 GAS strain gave better yield during protein purification than the Mga from a M1 GAS strain, and so the M4 Mga (Mga4) protein was used [73]. Mga4, was first cloned into

pCAL-C to introduce a C-terminal, thrombin cleavable CBP, calmodulin binding protein, tag. This Mga4-CBP was then cloned into pET21a to add a C-terminal 6x His tag (Figure 19). Mga1, $\Delta 29$ Mga4, which has a size difference but no change in activity, and $\Delta 139$ Mga4 CBP-His₆, the dimerization mutant, proteins were also created. The wild-type Mga4-CBP-His₆ was then grown in C41[DE3] in ZYP-5052 for ~ 60 hours as previously described [73]. The protein was purified first over the calmodulin column, followed by NiNTA agarose, as well as the reverse (Figure 21B). Between the calmodulin and NiNTA columns, the protein was initially dialyzed overnight against 4 L NiNTA lysis buffer. At this point Mga remained bound to NiNTA agarose.

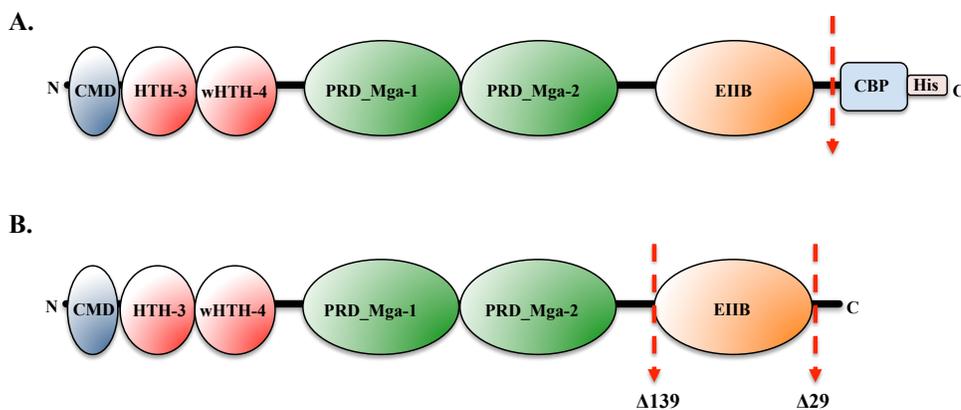


Figure 19 Schematic of TAP tagged Mga

(A) Full-length Mga with both a CBP and His₆ tag at the C terminal end of the protein for tandem affinity purification. The red arrow indicates the thrombin cleavage site. (B) Red lines indicate the location of the $\Delta 29$ and $\Delta 139$ truncations. $\Delta 29$ removes the disordered region at the C-terminus of the protein. $\Delta 139$ removes the EIIB dimerization domain.

Mga4-CBP-His₆ was then dialyzed against 1 L NiNTA lysis buffer, with four buffer exchanges in order to remove the EGTA from the CaCl₂ Elution buffer. This allowed Mga to be eluted from the NiNTA column. When passed over the NiNTA column, the affinity of Mga4-CBP-His₆ for the NiNTA agarose was reduced compared to Mga4-His₆, all protein eluted into 70 mM imidazole wash, as opposed to the 250 mM

imidazole elution buffer. Proteins were dialyzed into 50 mM HEPES/Citrate, pH 7.5, and concentrated to ~500 μ L. For each direction of purification, multiple bands were detected along with the expected 65 kDa band, so that Mga4-CBP-His₆ was approximately half the protein present (Figure 21B).

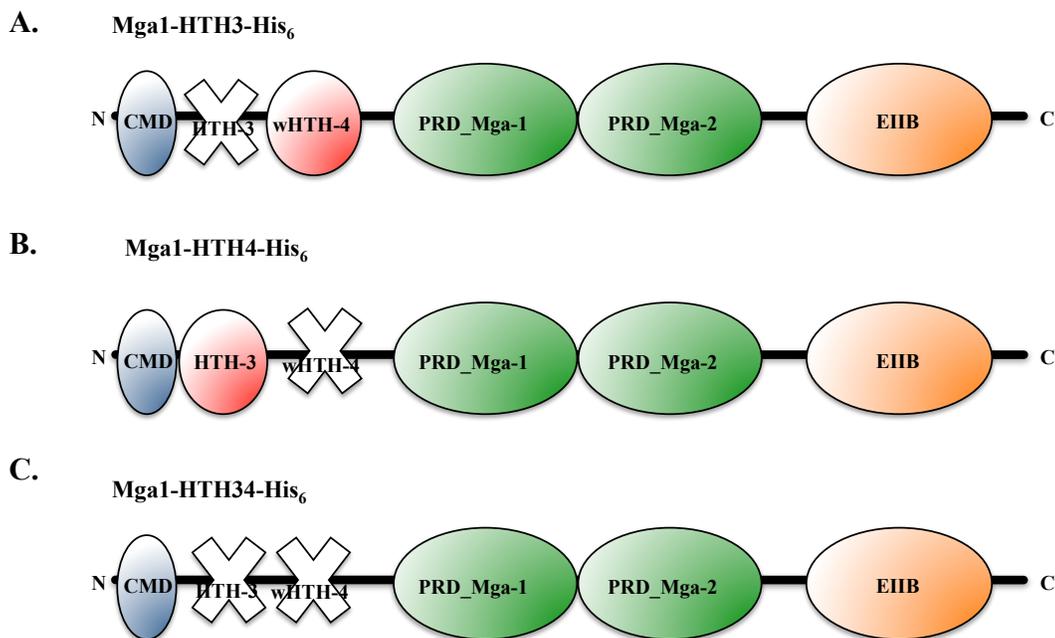


Figure 20 Schematic of HTH mutations

(A) Mga1-HTH3-His₆ contains point mutations within the recognition helix of HTH-3 in order to study the binding ability of wHTH-4. (B) Mga1-HTH4-His₆ contains point mutations within the recognition helix of wHTH-4 on order to study the binding ability of HTH-3. (C) Mga1-HTH34-His₆ has a mutation binding domain and does not bind to DNA.

Purification of Mga1-HTH4-His₆

To study the contribution of HTH-3 and wHTH-4, the essential binding domain, to DNA-protein interactions mutant proteins were created for expression in *E. coli*. The Mga1-HTH3-His₆ protein contains a mutation in the recognition helix of HTH3, Mga1-HTH4-His₆ contains a mutation in the recognition helix of HTH4 and the Mga1-HTH34-His₆ contains both mutations (Figure 20). Expression of each protein was monitored

across growth in ZYP-5052 for ~56 hours at 37°C. Western blots of the soluble fraction showed that Mga1-HTH4-His₆ and Mga1-HTH34-His₆ were expressed as the wild-type protein, but Mga1-HTH3-His₆ was insoluble, even when grown at RT (Figure 21A). The Mga1-HTH4-His₆ protein was subsequently purified in the same manner as Mga1-His₆.

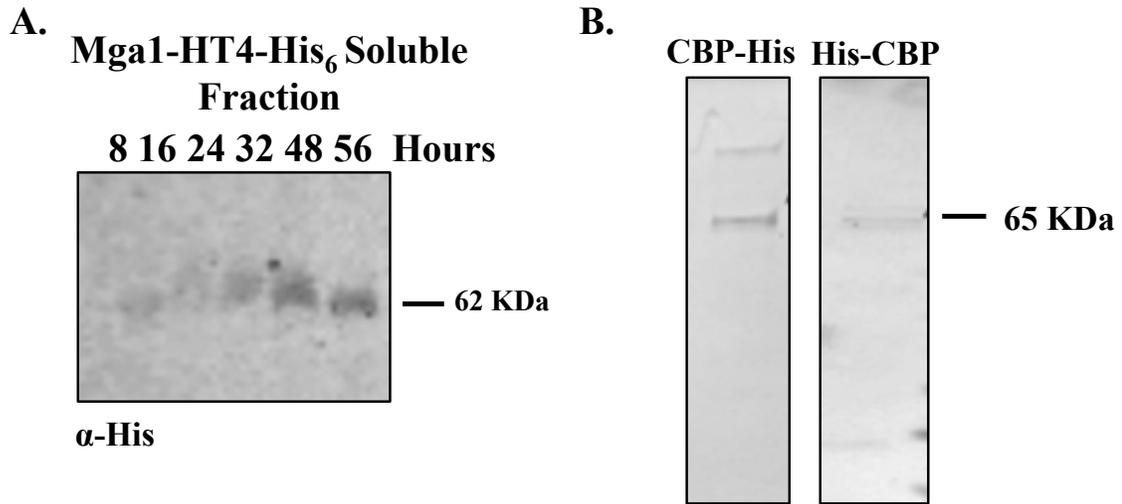


Figure 21 Purification of Mga1-HTH4-His₆ and TAP Tagged Mga

(A) Western blot probing with α-His for Mga1-HTH4-His₆ in the soluble fraction across growth found that the protein expressed similarly to wild-type. (B) Coomassie gel of the final purified and concentrated Mga4-CBP-His₆ after purification from the CBP column to the NiNTA column, or from NiNTA to the CBP column shows the predicted 65 KDa band, as well as several other bands of near equal intensity.

Category B Binding Sites

In order to study Mga interactions at a category B binding site, luciferase assays were performed to characterize the transcriptional activity of PsclA. The M1 PsclA binding site was cloned to drive luciferase expression on a plasmid and transformed into MGAS5005. Samples of the *PsclA-luc* and promoterless *luc* plasmids were taken in

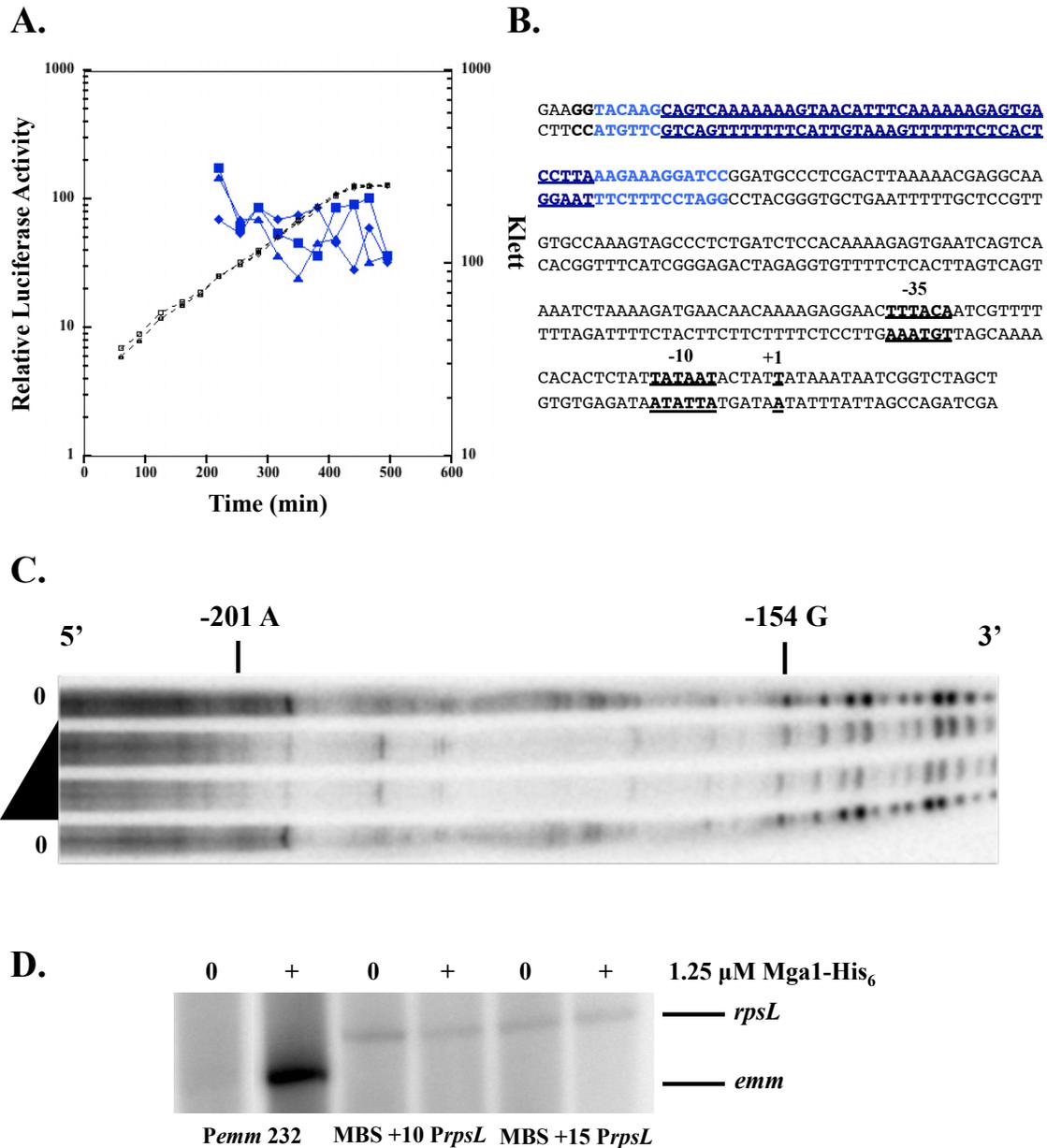


Figure 22 DNA-binding and Transcriptional Activation of Category B Promoters

(A) Luciferase assay of *PsclA-luc* shows low levels of activity across exponential growth in the strain MGAS5005. The dashed line represents growth, the blue line represents luciferase activity. (B) The *PsclA* promoter. The predicted MBS is underlined in dark blue [78], the actual MBS is extended by the light blue nucleotides. The -35, -10 and +1 transcription start are highlighted in bold. (C) DNaseI footprint on the sense strand of *PsclA* protects the nucleotides from -201 to -154 relative to the start of transcription. (D). *In vitro* transcription assay for Mga-dependent transcription when the *Pemm* MBS is placed either 10 or 15 bp upstream of the *PrpsL* -35 hexamer. Mga-dependent transcription is only observed at the wild-type *Pemm* 232 template.

triplicate every ~15 Klett over exponential growth. *PscIA-luc* showed low levels of expression across growth, ~ 1-2 x 10² relative luciferase units (RLU) (Figure 22A).

The *PscIA* MBS site had been determined by sequence alignments and then confirmed by EMSA analysis and *in vitro* transcription but was not directly footprinted [78]. The *PscIA* site was predicted to be 44 bp and encompass the nucleotides -153 to -197 relative to the +1 transcription start site (Figure 22B). To determine the *PscIA* binding site DNase I footprints were performed. In this assay a footprint of 59 bp was detected that protected the region from nucleotides -201 to -154 (Figure 22C).

Both Category B and C promoters activate transcription from a distal binding site. In order to understand the mechanism of activation at a distance an *in vitro* template was constructed where the *Pemm* MBS was placed either 10 or 15 bp upstream from the *PrpsL* -35 hexamer to create the templates MBS +10 *PrpsL* 1201 and MBS +15 *PrpsL* 1201. The *Pemm* MBS was used as it produces the strongest Mga-dependent transcriptional activation, and because the exact binding site has been defined, while *PscIA* has very low levels of activity and where Mga binds to DNA is not as clear. *In vitro* transcription assays performed with each of these templates did not have an increase in *PrpsL* transcript levels upon the addition of Mga1-His₆ to the reaction (Figure 22D).

Category C Binding Sites

EMSA analysis was performed to determine if Mga1-His₆ interacted with *Pmga* in the same manner as MBP-Mga6, which is the M6 strain Mga protein originally used for DNA-binding studies [76]. Increasing amounts of each protein was incubated with ~2.5 ng of the *Pmga*₂₃₁₅ probe, which contains both *Pmga* MBSs, as previously

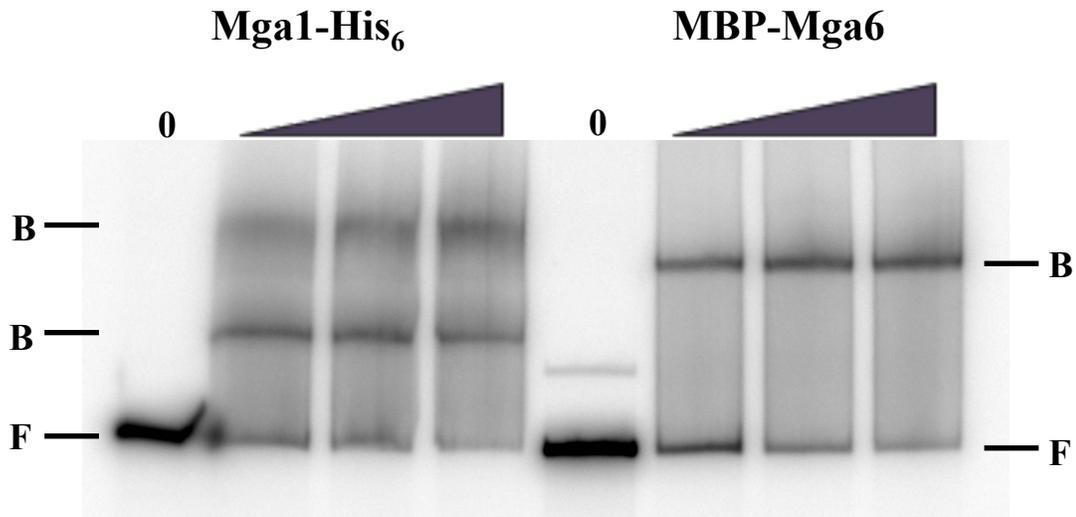


Figure 23 EMSA comparing Mga1-His₆ to MBP-Mga6 at *Pmga*

~2.5 ng of the *Pmga*₂₃₁₅ probe was incubated with 2, 3 and 3 μ M Mga1-His₆ or 8.3, 10.45 and 16.6 μ g/ μ L of MBP-Mga6. Two distinct bands were present when *Pmga* was incubated with Mga1-His₆ while only 1 band was observed with MBP-Mga6.

described. While MBP-Mga6 shifted a single band, at each concentration, the Mga1-His₆ protein shifted the probe as a doublet (Figure 23).

Comparison of DRACALA to EMSA and Filter-binding

DRACALA, differential radial capillary action of ligand assay, is a method for assessing protein-DNA binding in which bound ligand remains stuck on a nitrocellulose membrane while unbound ligand can move away by capillary action [111]. Densitometry can then be used to measure free versus unbound ligand, and subsequently obtain K_d values. DRACALA was performed to compare the K_d calculated by this method to the previously published K_d that was determined for Mga by filter-binding [73]. Mga4-His₆ was incubated with 0.1 nM *Parp* MBS 49mer and Mga1-His₆ was incubated with 0.1 nM of *Pemml* MBS 49mer. The binding reaction was performed as in the EMSA analyses, using 0.01, 0.1, 0.5, 1, 5, 10 and 20 μ M of either Mga-His₆ protein. 5 μ L of each reaction was spotted in triplicate on nitrocellulose, exposed to a phosphoimager cassette and then

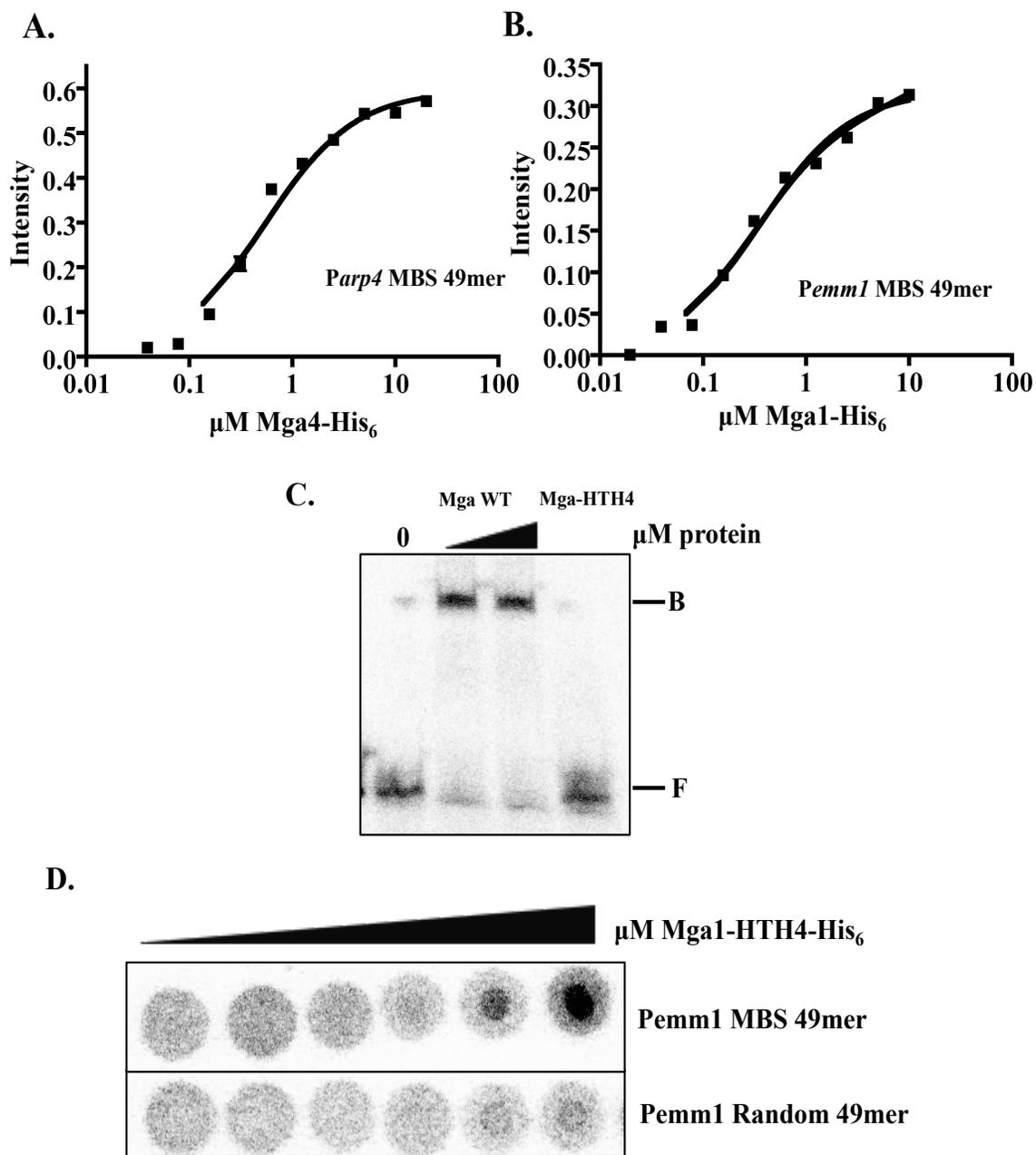


Figure 24 Mga-DNA binding by DRACALA

(A) DRACALA of Mga4-His₆ with the *Parp* MBS 49mer calculated a K_d of 286 nM. (B) DRACALA of Mga1-His₆ with the *Pemm1* MBS 49mer calculated a K_d of 396 nM. DRACALA values are ~10 fold greater than those obtained by filter-binding. (C) EMSA comparing the DNA binding ability of the HTH-3 to wild-type. 0.1 nM *Pemm1* MBS 49mer was incubated with 2.5 and 5 μM of Mga1-His₆ or 5 μM of Mga1-HTH4-His₆. (D) DRACALA of Mga1-HTH4-His₆ protein. 0.01, 0.1, 0.5, 1, 5, and 10 μM Mga1-HTH4-His₆ was incubated with either 0.1 nM of the *Pemm1* MBS 49mer or the *Pemm1* Random 49mer. The Mga1-HTH4-His₆ protein demonstrates DNA binding by DRACALA but not by EMSA.

densitometry was performed to determine the bound versus free probe. The K_d for Mga4-His₆ binding to *Parp* was 286 nM, while the previously determined K_d was ~30-50 nM (Figure 24A). The K_d for Mga1-His₆ binding to *Pemm* was 396 nM while previously a K_d of ~50-65 nM was determined (Figure 24B). Each measurement by DRACALA was ~10 fold greater than previously found.

To study the contribution of the HTH-3 to DNA binding the Mga1-HTH4-His₆ protein was tested for DNA binding activity by EMSA and DRACALA. For EMSA, 0.1 nM of the *Pemm1* MBS 49mer was incubated with 2.5 and 5 μ M of Mga1-His₆ and 5 μ M of Mga1-HTH4-His₆ (Figure 24C). For DRACALA, 0.01, 0.1, 0.5, 1, 5 and 10 μ M Mga1-HTH4-His₆ was incubated as for the EMSA with either 0.1 nM the *Pemm1* MBS 49mer or the *Pemm1* Random 49mer. As described previously, the protein with only HTH-3 was unable to bind DNA in the EMSA [121], but did show some binding activity in the DRACALA assay (Figure 24D).

Validation of the Chromosome Affinity Purification

To identify the genome wide binding profile of Mga, ChAP assays were performed. The GAS strains KSM547 (pLZ12-Spc) (*mga-/mga-*) and KSM547 (pKSM808) (*mga-/mga+*) were grown to mid-logarithmic phase (Klett 75-80). Following crosslinking with 1% formaldehyde, cells were lysed. The DNA was sheared by both a sonication using the Covaris system and then the Mga was purified over NiNTA agarose. Crosslinks were reversed overnight and the DNA was purified. The affinity purification was assessed by collecting samples after cell lysis, and after the protein was eluted from the NiNTA agarose. Western blots for these samples were probed with a α -his antibody. A His-tagged protein of ~62 kDa, corresponding to Mga,

was detected in the whole cell lysis and final elution of the strain containing pKSM808 (Figure 25A). No proteins were detected in the *mga*- strains.

The DNA was analyzed by qRT-PCR for enrichment of Mga binding sites in the Mga + strain over the Mga- strain. *gyrA* was used as an internal control in the reaction. Primers for *Parp* and *Pmrp*, two directly Mga-regulated genes were chosen as a positive control. On average of 3 replicates, *Parp* was enriched 3.32 fold +/- 0.19, and *Pmrp* was enriched 7.87 +/- 0.97 (Figure 25B). *PrpsL*, a non-Mga regulated gene, and *PmalR*, a Mga-regulated gene that does not bind to Mga, were used to determine if promoter DNA was being enriched over genomic DNA. Neither promoter was significantly enriched. As the purified DNA was successfully enriched for MBSs, it was sent for library formation and sequencing.

DNA was first submitted for quality analysis by Bioanalyzer and was found to be ~40 bp in size (Figure 25C), which is insufficient to separate unused linkers during the formation of the sequencing library. gDNA was extracted from samples directly before crosslinking, after crosslinking and after overnight incubation at 65°C and assessed on a 1% agarose gel. The gDNA was degraded during the crosslinking step, leading to overshearing of the DNA fragments for library preparation. Due to this degradation, sequencing libraries could not be generated from the enriched DNA.

Discussion

Preliminary studies were performed to study further interactions of Mga at the promoter, from the view of the DNA and the Mga protein. Dimerization is a necessary attribute of Mga's role as a transcriptional activator. Sedimentation equilibrium was performed on the Mga4-His₆ and the Δ 139Mga4-His₆ to determine the equilibrium

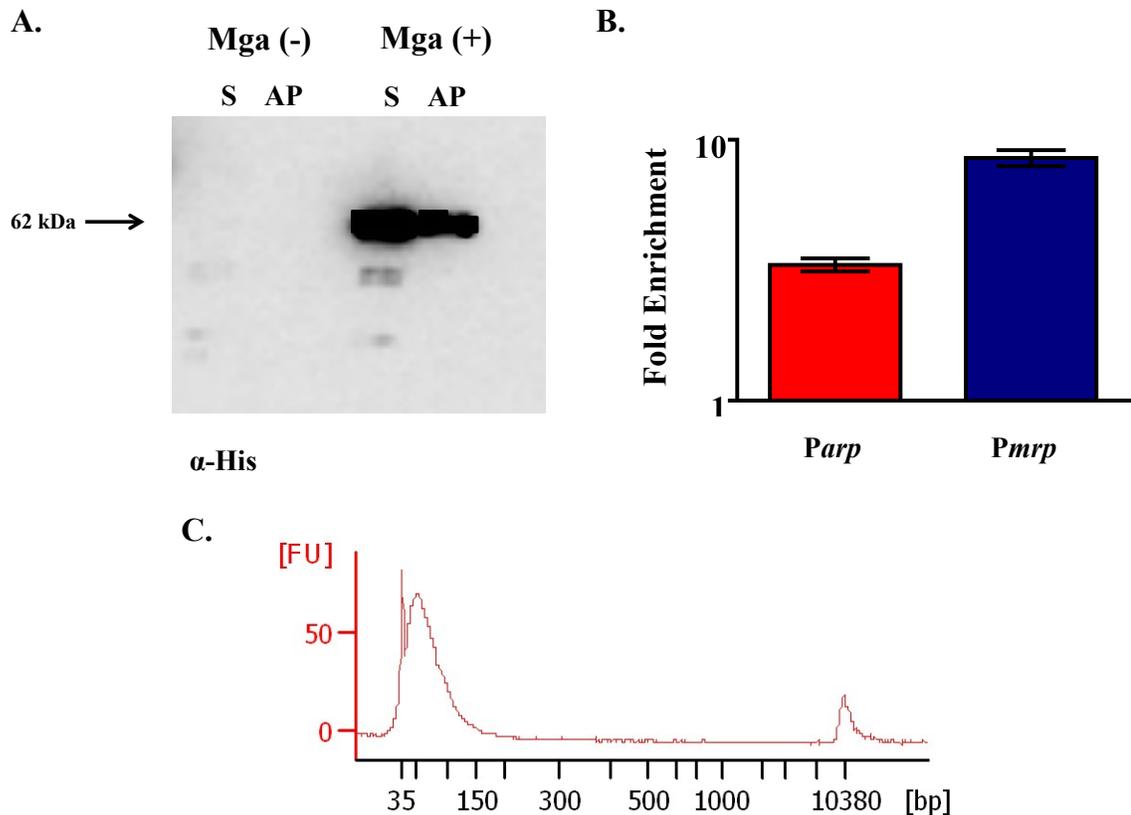


Figure 25 Validation of ChAP-Seq DNA

(A) Western blot for the presence of Mga after cell lysis (S) and after purification by NiNTA agarose (AP) detects a band in the Mga + strain but not the Mga- strain. (B) qPCR data showing a significant enrichment of DNA containing the MBS at *Parp* and *Pmrp* over the Mga- strain relative to *gyrA*. (C) High Sensitivity DNA Assay found that the DNA fragments in ChAP sample DNA are ~40 bp in size.

dimerization coefficient of each and assess the effect of salt concentration on the dimerization. Interestingly, while the previous gel filtration experiments found that salt concentration effected the formation of dimers, this effect was not seen by AUC. Future studies will be needed to fully understand the mechanism of dimerization. These experiments also set up for future sedimentation equilibrium experiments to determine if Mga binds DNA as a dimer.

The *in vitro* phosphorylation-transcription assay demonstrates the first direct link between PTS phosphorylation of Mga and the subsequent inhibition of Mga-dependent

gene regulation, supporting the hypothesis that Mga down-regulates its genes in response to the carbon status of the cell through the PTS. Further assays using the Mga H204A/H270A protein will be performed to demonstrate that this effect is specific to the phosphorylated histidines of Mga's PRDs. This Mga A/A protein should be insensitive to phosphorylation. Ideally this phosphorylation would also be detected within a GAS cell. Interestingly, while only a small portion of Mga appears to be phosphorylated *in vitro* by the PTS system, Mga-dependent transcription was completely abrogated. This suggests that the bulk of the protein may be misfolded or aggregated, and only a fraction of Mga is actually active.

A Mga-CBP-His₆ protein was created in order to enhance the purity of the protein for X-ray crystallography by tandem affinity purification. Initial purifications found that Mga was not the predominated band in in the final fraction when purified in either direction. Much of the protein also appears to be lost during purification. The increased purity of the Mga-His₆ protein is partly due to increased washes combined with increasing amounts of imidazole to remove impurities [73]. However, the Mga-CBP-His₆ protein elutes at a much lower imidazole concentration. The CBP tag also introduced a thrombin cleavage site. If the CBP-His₆ tag can be successfully removed, than further optimizing the purification would be highly beneficial to crystallizing the protein.

Mutations were created in the HTH-3 and wHTH-4 to study the contribution of each domain to DNA-binding in isolation. Unfortunately the M1 Mga HTH-3 mutant protein was insoluble in *E. coli*. As the M4 strain Mga protein is somewhat less finicky, future studies may focus on the M4 version or alternatively the HTH-3 mutant may be

purified from GAS. The wHTH-4 protein was expressed similarly to the wild-type Mga1-His₆ protein, and so it was purified and tested for DNA binding.

While the majority of work so far has focused on Category A promoters, and *Pemm* specifically, understanding how Mga interacts with Category B and C promoters was also an important goal. The *Pmga* promoter had previously been characterized for expression across growth [57], but no Category B promoters had been tested. A *PsclA-luc* plasmid was constructed and monitored for expression. *PsclA* was found to have very low luciferase activity across growth, comparable to what was seen for *PscpA* [120]. The Category B promoters *PsclA* and *Psof-sfbX* were both identified by sequence alignments and confirmed by EMSA and *in vitro* transcription. This study performed the first DNaseI footprints on the *PsclA* promoter. Surprisingly the footprint protected a region of 59 bp, as opposed to the predicted 44, which resembles the *Pmga* binding sites. As *PsclA* is centered -173 bp upstream from the start of transcription, the additional protected region may be a result of DNA bending to interact with the downstream promoter. This may also be the explanation for the large size of the two *Pmga* MBSs. Footprints will be performed on other Category B promoters to confirm this phenotype.

A template that placed the *Pemm* MBS binding site 10 and 15 bp upstream of the *PrpsL* -35 hexamer was created to shift the binding site 1 and 1.5 helical turns away from the promoter. At each promoter this adjustment was sufficient to disrupt Mga's ability to activate transcription. Future studies will be done to determine how far from the -35 the binding site can be moved before Mga-dependent transcription is disrupted. MBSs at Category B promoters will also be investigated to determine how the position of the binding site affects transcription. Mga may activate transcription at the Category B

promoters through a different mechanism than at the Category A promoters, so the RNAP mutants will also be assessed at these promoters.

To investigate the Category C promoter, EMSA using *Pmga*₂₃₁₅, which contains both MBSs was performed. While the MBP-Mga6 protein shifted in a single band, the Mga1-His₆ protein shifted as a doublet. Each Mga protein has an identical wHTH-4, which is the essential binding helix, and an identical HTH-3, that acts as an accessory at *Pmga* [79] so this could be an artifact of the large size of the MBP tag. The Mga-His₆ protein will be used in future studies that would include DNaseI footprinting, and competition assays with probes containing either or both *Pmga* MBS site.

DRACALA assays were performed using the Mga4-His₆ protein with the *Parp* MBS 49mer and the Mga1-His₆ protein with the *Pemm* MBS 49mer to compare the K_d from this technique to the previously published values obtained by filter-binding. For each protein/probe combination the K_d determined by DRACALA was approximately 10 fold larger than that determined by the filter-binding technique. The reason for this difference has yet to be determined. Mga appears to readily precipitate onto the nitrocellulose membrane during filter binding assays, making it difficult to obtain clear measurements. While the DRACALA method has its limitations, it offers the advantage of being a quick and straightforward method of assessing many binding sites.

As previously seen, the wHTH-4 mutant had no DNA-binding activity when tested by EMSA [79], however binding activity was detected by DRACALA. EMSA extends the DNA binding assay over time and so with weak binding, this interaction may be lost. The DRACALA assay is much quicker, and appears to be more sensitive.

Future studies with the HTH mutants or weak binding sites should be performed with DRACALA in order to not overlook weak or transient interactions.

These initial studies demonstrate that the ChAP technique can successfully enrich for DNA containing Mga binding sites. This technique could also be applied to other GAS transcription factors. However, at this time the DNA is too small to be successfully incorporated into the sequencing libraries. This degradation may be due to the many DNases present or a variation in strain used and further studies will be needed to overcome this problem. If the experiment finds new targets for DNA-binding, these sequences will be first compared against genes of the known Mga-regulon in both the GA40634 strain used and in the clinically relevant MGAS5005 strain. DRACALA will be used to test probes containing these sequences for DNA-binding activity. Probes that bind to Mga will then be aligned with known Mga binding sites. Further binding studies will also be performed to compare the contribution of Mga's two DNA binding domains to these potential new sequences.

Chapter 6

Conclusions

The Group A streptococcus is an obligate human pathogen that causes a wide range of diseases within diverse niches of the human body. Mga, the multigene regulator of GAS, is a key global transcription factor that regulates many of the virulence genes necessary for infection. While Mga's broad function as the activator of virulence genes such as *emm*, *scpA* and *sclA* is known, the specifics of these interactions have not been well understood. Overall these studies have focused on the mechanism by which Mga functions at the promoter and together this data has allowed us to build a model of Mga at the promoter.

Model of Mga interactions at the promoter

Dimerization

The key components that allow Mga to activate transcription are DNA-binding, dimerization, protein-protein contacts with RNA polymerase and PTS phosphorylation. DNA-binding and dimerization are tightly linked (Figure 26). The distribution of nucleotides within the *PemmI* binding site, in particular the GGTC and GCTG motifs support a model in which two Mga proteins are bound [120]. The lack of apparent half-site binding (this study) and the cooperativity observed in the binding curve [73] also support the idea that Mga DNA binds as a dimer. That Mga does form multimers *in vivo* and dimers in solution [73] has also been shown and is further supported by the crystal structure of the *E. faecalis* Mga homolog EF3013 (Protein Data Bank [PDB] accession no. 3SQN). The distance between the wHTHs in this protein are ~100 Å or 30 bp. While

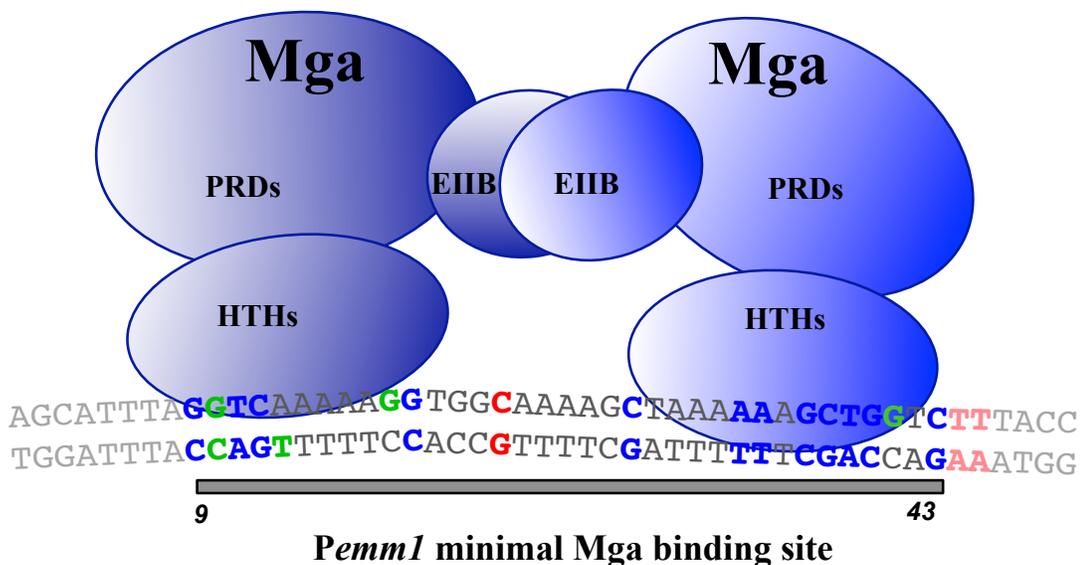


Figure 26 Model of Mga Binding to the Pemm MBS

Mga forms a dimer in solution through its EIIB domain. The arrangement of the domains is based on the *E. faecalis* Mga homolog EF3013. The wHTH4 forms specific contacts with GGTC at the 5' end and with the GCTG at the 3' end. The 3' dimer also makes contacts with C29 and C43. As the distance between the wHTHs is ~100 Å or 30 bp, a bend in Pemm at G18 and G19 fits the binding site to the protein.

this is slightly smaller than the Pemm1 minimum MBS, the differences between Mga and this homolog, and DNA-bending can further fit a Mga dimer onto DNA.

To demonstrate the validity of this model, definitively showing the dimerization of Mga at a MBS will be the most important future experiment. In order to do so, sedimentation equilibrium studies will be performed, first using Mga with Pemm, and later using Mga with a binding site from each category. SAXS, small angle X-ray scattering, is another technique that can be used to study this interaction. The current effort to crystallize Mga within the presence of DNA will also shed light on this interaction. $\Delta 139$ Mga, which binds to DNA with a similar K_d as wild-type but does not dimerize will be another key component to understand Mga-DNA interactions. DNaseI footprints will also be performed to determine if this protein binds monomerically.

While overall Mga binding sites have low sequence identity we propose that Mga does make specific contacts with DNA using wHTH-4. In the *Pemm1* binding site the specific contacts are with 5' GGTC and 3' GCTG. The 3' dimer makes additional contacts with C29 and C43. Similar motifs are found in the other binding sites that suggest that Mga recognizes a small specific sequence. *PscpA* and *Psic* have a 5' GGTC motif, as does *Pmga2*. *PsclA* and *Pmga1* each have a 1 nucleotide change: to AGTC for *PsclA* and and GTTC *Pmga1*. *PscpA*, *Psic*, *PsclA* and *Pmga1* also have a 3' GGTC motif that is inverted and on the opposite strand. The distance and orientation of these motifs likely plays an important role in affinity and activity of each binding site. Each half-site may also vary in affinity for a Mga monomer, so that while Mga binds as dimer, some breathing occurs between the interactions. One hypothesis is that the variation of the known Mga binding sites is allows for fine control of the genes expressed. This hypothesis is illustrated by comparing *Pemm*, *PscpA* and *Psic*. The luciferase assays found that *Pemm* is highly expressed, *PscpA* is lowly expressed and *Psic* falls in between. These finding are quite interesting in that *PscpA* and *Psic* are extremely similar in sequence, but have more than ~10 fold difference in expression. *Pemm* and *Psic* also shown similar DNA-binding abilities, but again *Pemm* expression is more than 10 fold greater than *Psic*.

As these studies have shown that predictions using the *Pemm* MBS as model do not necessarily predict the function of nucleotides in other binding site, each site will need to be screened individually. To rapidly screen the binding sites, biochemical interference assays will be used. Methylation interference and protection assays will be most informative, in particular to look for the GGT motif. Directed mutagenesis will be

performed to understand the roles of DNA-bending, spacing and orientation of the GGTC motifs within the MBS play in DNA-binding affinity and ultimately transcriptional activation.

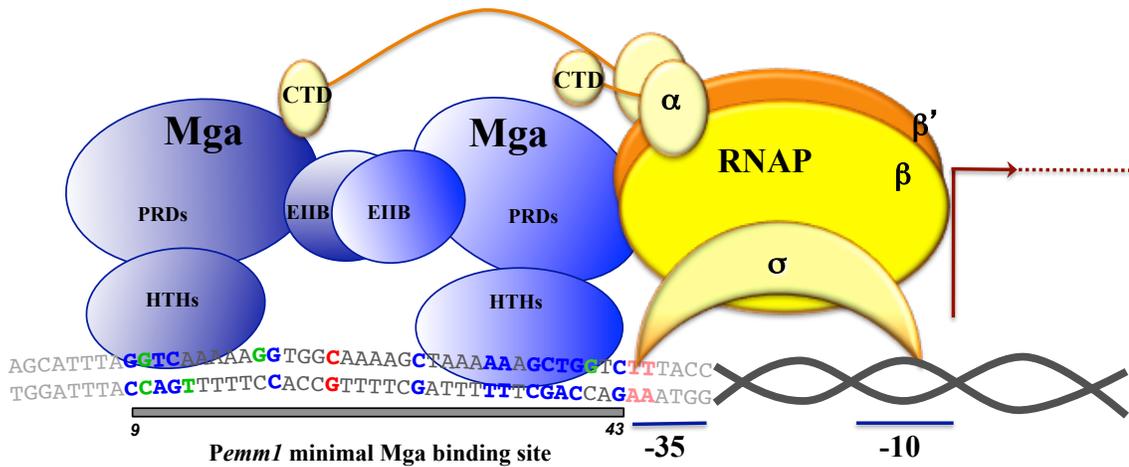


Figure 27 Protein-protein contacts with RNA Polymerase

We predict that the Mga dimer forms protein-protein contacts with RNAP through the α -CTD and domain-4 of σ concurrently, either in solution where Mga recruits RNAP to the promoter, or when an already bound Mga stabilizes the polymerase at the promoter. These protein-protein contacts allow RNAP to more quickly find the promoter and stabilize it through the formation of the elongation conformation.

Interactions with RNA Polymerase

Protein-protein contacts between Mga and the holoenzyme are another essential component for Mga-dependent transcriptional activation. While Mga does interact with the holoenzyme in solution, neither the α -CTD nor the domain 4 of σ factor are the sole point of contact (this study). However at bacterial promoters multiple transcription factors may be present to interact with RNAP [95]. As Mga likely binds as a dimer and this dimer is necessary for activation, we propose a model where the Mga dimer forms protein-protein contacts with RNAP through the α -CTD and domain-4 of σ concurrently. These interactions occur in solution where Mga recruits RNAP to the promoter, or an

already bound Mga stabilizes the polymerase at the promoter (Figure 27). These protein-protein contacts allow RNAP to both more quickly find the promoter and then stabilize it through the formation of the elongation conformation. Most point mutations had a simple connection between DNA-binding affinity and transcriptional activation: less binding, less activity, more binding, more activity. These are simple to place in the model, as more/less Mga binding leads to more/less RNAP recruitment leads to more/less transcription. However mutations were identified with less binding and more activation (*Pemm1* G9A, *PscpA* C12A), wild type binding with a decrease activity (*Pemm1* C23A, *PscpA*) and mutations with a decrease in binding but no change in activity (*Pemm1* G10A, *Pemm1* G18A, *PscpA* C12A). The G9A mutation decreases binding but increases transcription possibly through increasing promoter clearance. The orientation and spacing of the Mga dimer may explain the phenotype of these other mutations. *Pemm* contains a strong DNA bend at A17, G18 and G19 that is not present in *PscpA*, while the GGT motif at the 3' end of *PscpA* is shifted 3' from the CGT that appears to be important in *Pemm*. If Mga is binding as dimer, then these difference in the space between the “half-sites” both linearly along the DNA and in 3-dimensional space could have important implications for making the optimal contacts with RNAP. Future studies of these interactions will include DNaseI footprinting of Mga with RNAP polymerase with the mutants from both the protein and DNA standpoint. The $\Delta 139$ Mga mutant is of particular interest in observing how Mga makes contacts with RNAP; DNaseI footprint, methylation interference assays as well as *in vitro* co-affinity purification assays will be performed to understand how activation occurs.

PTS Regulation of Mga Activity

The PTS phosphorylation of Mga is the last component that allows Mga to activate transcription identified so far. This phosphorylation results in the loss of Mga-dependent transcriptin *in vitro* and specifically this phosphorylation disrupts dimerization as shown by co-IP studies using a Mga phosphomimetic (Hondorp, 2012, in review). In EF3013, the phosphorylated histidines are not present. However the PRD-1 of each monomer, which contains the phosphorylation site, are separated and do not form a dimer

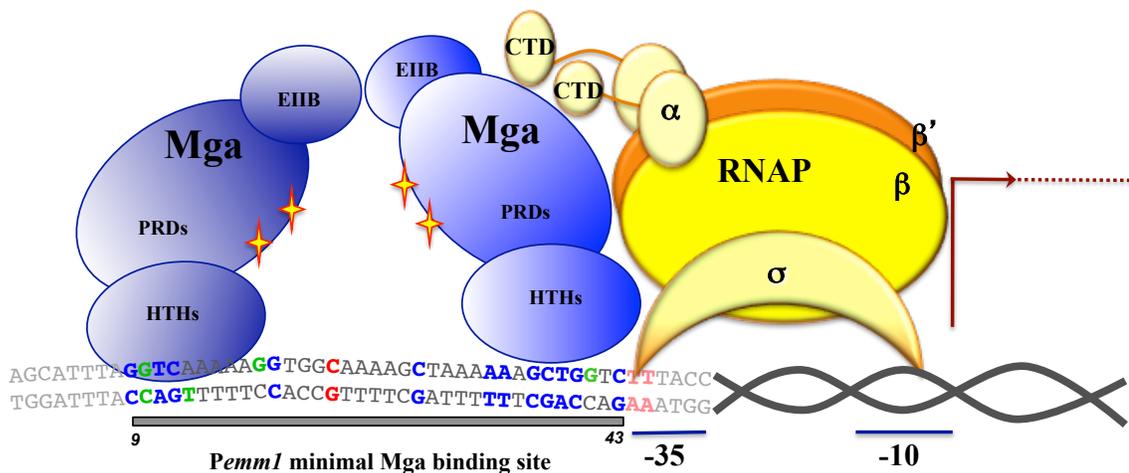


Figure 28 Effect of PTS phosphorylation on Mga dimerization

The PTS phosphorylation of Mga disrupts dimerization, and transcriptional activation, but not DNA binding (Hondorp, 2012, in review). We predict that the PRD-1 phosphorylation changes the structure of the protein so that the EIIB domains no longer dimerize. The phosphorylated protein still binds to DNA, but as a monomer instead of a dimer. This also disrupts protein-protein interactions with RNA polymerase, which turns off Mga-dependent transcription.

interface. Our model therefore proposed that this phosphorylation changes the structure of the protein so that the EIIB domains no longer interact. As a result Mga may bind DNA as a monomer and the protein-protein interactions with RNAP are disrupted.

In the future, the *in vitro* phosphorylation-transcriptions experiment will be performed using the Mga A/A protein. In this protein, the histidines that are

phosphorylated have been mutated to alanines, and therefore this protein should not be down-regulated by the PTS system. The *in vitro* phosphorylation of Mga will also be used to confirm the loss of dimerization phenotype by co-affinity purification. Changes to DNA binding will also be studied by DNaseI footprinting. Changes to protein-protein interactions with RNAP will be studied by co-affinity purification as well as DNaseI footprinting with Mga~P.

Mga's role as a regulator within GAS

Another important goal of these studies is refine Mga's role as a regulator within a GAS cell. Genome-wide binding profiling by ChAP-Seq will create a complete picture of where Mga binds in the genome, and could help answer the question of how it regulates ~10% of the GAS genome. Identifying under what conditions Mga is phosphorylated and dephosphorylated *in vivo* would also help explain how this regulation occurs.

While the wHTH-4 is the essential DNA-binding domain, HTH-3 may still play an important role. Future studies will look at the contribution to binding of each DNA-binding domain at both known MBSs and any further identified MBSs.

The spacing between the binding site and the promoter could another component of how Mga differentially regulates gene expression. Transcription activation will be studied at a distance will be studied by changing both the distance between a MBS. with particular attention to the orientation along the DNA helix, and the promoter, and the flexibility of the DNA in between.

Significance

While these studies have narrowly focused on the function of Mga as a transcriptional activator, Mga belongs to a larger family of PRD containing transcriptional regulators. Phyre2 analysis of RivR, another virulence regulator within GAS, also detects PRDs within its structure. Many other streptococcal species contain Mga orthologues, which include DmgB (*dmgB*) or MgrC, in *S. dysgalactiae* and MgrA (*mgrA*) in *S. pneumoniae*. Outside of the streptococci, the virulence regulator AtxA of *Bacillus anthracis* contains PRD domains, and can also dimerize through its C-terminal EIIB domain [122]. Understanding how Mga functions will aid in understanding this class of PTS-sensing virulence regulators.

Appendices

Table 1: *E. coli* Strains

<i>E. coli</i> Strains	Description	Reference
DH5 α	(<i>hsdr17 recA1 gyra endA1 relA1</i>)	[123]
C41[DE3]	F- <i>ompt hsdSB(rb - mb -) gal dcm (de3)</i>	[106]
OneShotTop10 Electrocompetent	(F- <i>mcra</i> δ (<i>mrr-hsdRms-mcrbc</i>) ϕ 80 <i>lacZ</i> δ <i>m15</i> δ <i>lacX74 recA1 arad139</i> δ (<i>ara-leu</i>) 7697 <i>galU galk rpsL (strr) endA1 nupg</i> λ -)	Stratagene
DHM1	F-, <i>cya-854, recA1, endA1, gyra96 (nal r), thi1, hsdR17, spot1, rfbD1, glnV44(as)</i>	Gift, D. Kearns
BTH101	F-, <i>cya-99, arad139, gale15, galk16, rpsL1 (str r), hsdR2, mcra1, mcrb1</i>	Gift, D. Kearns

Table 2: GAS strains used in these studies

GAS Strains	Description	Reference
MGAS5005	<i>emm1, Δcovs</i>	[124]
KSM165-L.5005	<i>emm1 Mga::pKSM165-L.5005 Kn^R</i>	[73]
GA40634	<i>emm4</i>	Georgia Emerging Infections Program (GaEIP)
KSM547	<i>emm4 mga::pKSM547 Kn^R</i>	[72]
JRS4-PolHis	<i>str rpoC::pPolHisPyo sp^R kn^R</i>	[98]
GA40634.pKSM294	pKSM294 Kn ^R	This Study
GA40634.pKSM295	pKSM295 Kn ^R	This Study

Table 3: Table of Plasmids

Plasmid Name	Description	Reference
pCal-C	Expression vector C-terminal CBP tag	Stragagene
pCR-Blunt-II-TOPO	pUC ori fl ori Kn ^R Ampr LacZa	Invitrogen
pCIV2	GAS suicide vector, Ω Kn ^R Cassette	[125]
pET21a	Expression Vector C terminal 6x His tag	Novagen
pMSP3535-H3	Nisin Inducible promoter, GAS replicating Plasmid	[126]
pProEX-htb	Expression Vector N-terminal 6x His tag	Sigma(Genosys)
pT18N-link	C-terminal T18 expressing two-hybrid vector with linker	Gift, D. Kearns
pT18C-link	N-terminal T18 expressing two-hybrid vector with linker	Gift, D. Kearns
pT25N-link	C terminal T25 expressing two-hybrid vector with linker	Gift, D. Kearns
pJRS525	GAS replicating vector Sp ^R	[127]
pLZ12-Spc	GAS replicating vector Sp ^R	[128]
pVL847	Expression vector N terminal 10x His-MBP	[129]
pMga1-His	M1 Mga-his under PT7 promoter in pET21a for purification of c-terminal tagged protein in <i>E. coli</i>	[73]
pKSM720	Sp ^R promoterless luciferase plasmid	[57]
pKSM210	<i>Pemm1</i> -luciferase reporter plasmid in pKSM720 backbone	[120]
pKSM211	<i>PscpA1</i> -luciferase reporter plasmid in pKSM720 backbone	[120]
pKSM212	<i>Pemm1</i> -luciferase reporter with the C43A mutation	[120]
pKSM213	<i>Pemm1</i> -luciferase reporter with the A35C mutation	[120]
pKSM214	<i>Pemm1</i> -luciferase reporter with the G40A mutation	[120]
pKSM215	<i>Pemm1</i> -luciferase reporter with the G37A mutation	[120]
pKSM216	<i>Pemm1</i> -luciferase reporter with the C38A mutation	[120]
pKSM217	<i>Pemm1</i> -luciferase reporter with the T44C mutation	[120]
pKSM218	<i>Pemm1</i> -luciferase reporter with the T45C mutation	[120]
pKSM219	<i>Pemm1</i> -luciferase reporter with the C12A mutation	[120]
pKSM220	<i>Pemm1</i> -luciferase reporter with the C23A mutation	[120]
pKSM221	<i>Pemm1</i> -luciferase reporter with the C29A mutation	[120]
Plasmid Name	Description	Reference
pKSM222	<i>Pemm1</i> -luciferase reporter with the T11C mutation	[120]
pKSM223	rpoD in pT18C-link	This Study

Name	Description	Reference
pKSM224	rpoA in pT18C-link	This Study
pKSM225	rpoD in pT18N-link	This Study
pKSM226	Mga in pT25N-link	This Study
pKSM227	Mga in pT18N-link	This Study
pKSM228	rpoA in pT18N-link	This Study
pKSM229	rpoA in pT25N-link	This Study
pKSM230	rpoD in pT25N-link	This Study
pKSM231	M1 Sf370 <i>PscIA</i> in pKSM720	This Study
pKSM232	<i>Pemml</i> -luciferase reporter with the C3A mutation	[120]
pKSM233	rpoE in pT18N-link	This Study
pKSM234	α in pProEX-htb	This Study
pKSM235	α - Δ CTD in pProEX-htb	This Study
pKSM236	Mga1- Δ 139 in pT25N-link	This Study
pKSM237	rpoE in pT18C-link	This Study
pKSM238	M1 Sf370 <i>PscIA</i> minus MBS1	This Study
pKSM239	M1 <i>PscpA</i> delta attenuator in pKSM720	This Study
pKSM240	<i>Pemml</i> -luciferase reporter with the G9A mutation	[120]
pKSM241	<i>Pemml</i> -luciferase reporter with the G10A mutation	[120]
pKSM242	<i>Pemml</i> -luciferase reporter with the T39C mutation	[120]
pKSM243	<i>PscpAI</i> -luciferase reporter with the C12A mutation	[120]
pKSM244	<i>Pemml</i> -luciferase reporter with the C12/43A mutation	[120]
pKSM245	<i>PsicI</i> -luciferase reporter plasmid in pKSM720 backbone	[120]
pKSM246	σ in pProEX-htb	This Study
pKSM247	His ₆ - α - Δ CTD in pKSM683	This Study
pKSM683	pKSM201-PrpsL	This Study
pKSM248	His ₆ - α in pKSM683	This Study
pKSM249	Mga1 HTH3/4 in pET21a	This Study
pKSM250	Mga1 HTH-4 in pET21a	This Study
pKSM251	His ₆ - α - Δ CTD in pMSP3535-H3	This Study
pKSM252	His ₆ - α in pMSP3535-H3	This Study
pKSM253	pKSM874 modified to M1Mga HTH3	This Study
pKSM254	pKSM874 modified to M1 HTH3/4	This Study
pKSM255	pKSM874 modified to M1 HTH4	This Study
pKSM256	<i>PscpAI</i> -luciferase reporter with the C43A mutation	[120]
pKSM257	<i>PscpAI</i> -luciferase reporter with the C12/43A mutation	[120]
pKSM258	<i>Pemml</i> -luciferase reporter with the G41A mutation	[120]
pKSM259	<i>Pemml</i> -luciferase reporter with the A13C mutation	[120]
pKSM260	<i>Pemml</i> -luciferase reporter with the G18A mutation	[120]
pKSM261	<i>Pemml</i> -luciferase reporter with the G19A mutation	[120]

Name	Description	Reference
pKSM262	<i>Pemml</i> -luciferase reporter with the A33C mutation	[120]
pKSM263	<i>Pemml</i> -luciferase reporter with the A34C mutation	[120]
pKSM264	C terminus of M1 Mga in pET21a	This Study
pKSM265	N180 of M1 Mga in pET21a	This Study
pKSM266	<i>Pmga</i> -mga1-HTH4 in pJRS525	This Study
pKSM267	<i>Pmga</i> -mga1-HTH3 in pJRS525	This Study
pKSM268	<i>Pmga</i> -mga1-HTH3/4 in pJRS525	This Study
pKSM269	M1 RofA in pET21a	This Study
pKSM270	M1 RivR in pET21a	This Study
pKSM271	<i>PsicI</i> -luciferase reporter with the G40A mutation	[120]
pKSM272	<i>PsicI</i> -luciferase reporter with the C12A mutation	[120]
pKSM273	<i>PsicI</i> -luciferase reporter with the C43A mutation	[120]
pKSM274	<i>PsicI</i> -luciferase reporter with the C12/43A mutation	[120]
pKSM275	<i>PspcA1</i> -luciferase reporter with the G40A mutation	[120]
pKSM276	M4 RALP3 in pET21a (adam)	This Study
pKSM277	His ₆ - α in pT18N-link	This Study
pKSM278	His ₆ - σ in pT18N-link	This Study
pKSM279	σ - Δ domain4 in pProEX-htb	This Study
pKSM280	M4 Mga E413A in pKSM808 (adam)	This Study
pKSM281	His ₆ - α suicide in pCIV2	This Study
pKSM282	rpoA- Δ 1/3CTD in pProEX-htb	This Study
pKSM283	rpoA- Δ 2/3CTD in pProEX-htb	This Study
pKSM284	His ₆ - σ in pKSM683	This Study
pKSM285	His ₆ - σ - Δ domain4 in pKSM683	This Study
pKSM286	His-MBP-RofA1 in pVL847	This Study
pKSM287	His-MBP-RivR1 in pVL847	This Study
pKSM288	Mga1 in pCal-C	This Study
pKSM289	Mga4 in pCal-C	This Study
pKSM290	Promoter upstream α - <i>Pami</i> -his ₆ - α - Δ CTD in pCIV2	This Study
pKSM291	α - Δ CTD in pET21a	This Study
pKSM292	α in pET21a	This Study
pKSM293	Promoter upstream α - <i>Pami</i> -his ₆ - α - Δ 1/3CTD in pCIV2	This Study
pKSM294	<i>Pami</i> - α - Δ CTD-his ₆ -downstream in pCIV2	This Study
pKSM295	<i>Pami</i> - α -his ₆ -downstream in pCIV2	This Study
pKSM296	α - Δ 1/3CTD in pET21a	This Study
pKSM297	α - Δ 2/3CTD in pET21a	This Study
pKSM298	Mga4-CBP in pET21a	This Study
pKSM299	Mga4 Δ 29 in pCal-C	This Study
pKSM415	<i>PrpsL</i> in pBluescript	This Study
pKSM420	<i>Psof</i> in pBluescript	This Study
pKSM550	Mga4 Δ 139 in pCal-C	This Study
pKSM551	Mga4 Δ 139-CBP in pET21a	This Study
pKSM552	Mga4 Δ 29-CBP in pET21a	This Study

Plasmid Name	Description	Reference
pKSM553	α NTD- σ domain4 in pProEX-htb	This Study
pKSM683	<i>PrpsL</i> in pKSM201	This Study
pKSM801	<i>Mga4</i> in pET21a	This Study
pKSM808	<i>Pmga-mga4-his₆</i> in pLZ12-Spc	This Study
pKSM802	<i>Mga4</i> Δ 29 in pET21a	This Study
pKSM874	WT <i>Mga1-his₆</i> under <i>PrpsL</i> with native GAS RBS	This Study
TOPO-Pemm	<i>Pemml</i> in pCR-Blunt-II-TOPO	[120]
TOPO-Pemm C3A	<i>Pemml</i> with the C3A mutation in pCR-Blunt-II-TOPO	[120]
TOPO-Pemm G9A	<i>Pemml</i> with the G9A mutation in pCR-Blunt-II-TOPO	[120]
TOPO-Pemm G10A	<i>Pemml</i> with the G10A mutation in pCR-Blunt-II-TOPO	[120]
TOPO-Pemm T11C	<i>Pemml</i> with the T11C mutation in pCR-Blunt-II-TOPO	[120]
TOPO-Pemm C12A	<i>Pemml</i> with the C12A mutation in pCR-Blunt-II-TOPO	[120]
TOPO-Pemm A13C	<i>Pemml</i> with the A13C mutation in pCR-Blunt-II-TOPO	[120]
TOPO-Pemm G18A	<i>Pemml</i> with the G18A mutation in pCR-Blunt-II-TOPO	[120]
TOPO-Pemm G19A	<i>Pemml</i> with the G19A mutation in pCR-Blunt-II-TOPO	[120]
TOPO-Pemm C23A	<i>Pemml</i> with the C23A mutation in pCR-Blunt-II-TOPO	[120]
TOPO-Pemm C29A	<i>Pemml</i> with the C29A mutation in pCR-Blunt-II-TOPO	[120]
TOPO-Pemm A33C	<i>Pemml</i> with the A33C mutation in pCR-Blunt-II-TOPO	[120]
TOPO-Pemm A34C	<i>Pemml</i> with the A33C mutation in pCR-Blunt-II-TOPO	[120]
TOPO-Pemm A35C	<i>Pemml</i> with the A35C mutation in pCR-Blunt-II-TOPO	[120]
TOPO-Pemm G37A	<i>Pemml</i> with the G37A mutation in pCR-Blunt-II-TOPO	[120]
TOPO-Pemm C38A	<i>Pemml</i> with the C38A mutation in pCR-Blunt-II-TOPO	[120]
TOPO-Pemm T39C	<i>Pemml</i> with the T39C mutation in pCR-Blunt-II-TOPO	[120]
TOPO-Pemm G40A	<i>Pemml</i> with the G40A mutation in pCR-Blunt-II-TOPO	[120]
TOPO-Pemm G41A	<i>Pemml</i> with the G41A mutation in pCR-Blunt-II-TOPO	[120]
TOPO-Pemm C43A	<i>Pemml</i> with the C43A mutation in pCR-Blunt-II-TOPO	[120]
TOPO-Pemm T44C	<i>Pemml</i> with the T44C mutation in pCR-Blunt-II-TOPO	[120]
TOPO-Pemm T45C	<i>Pemml</i> with the T45C mutation in pCR-Blunt-II-TOPO	[120]
TOPO-Pemm C12/43A	<i>Pemml</i> with the C12/43A mutation in pCR-Blunt-II-TOPO	[120]

Name	Description	Reference
TOPO-<i>PscpA</i>	<i>PscpA1</i> in pCR-Blunt-II-TOPO	[120]
TOPO-<i>PscpA</i> C12A	<i>PscpA1</i> with the C12A mutation in pCR-Blunt-II-TOPO	[120]
TOPO-<i>PscpA</i> G40A	<i>PscpA1</i> with the G40A mutation in pCR-Blunt-II-TOPO	[120]
TOPO-<i>PscpA</i> C43A	<i>PscpA1</i> with the C43A mutation in pCR-Blunt-II-TOPO	[120]
TOPO-<i>PscpA</i> C12/43A	<i>PscpA1</i> with the C12/43A mutation in pCR-Blunt-II-TOPO	[120]

Plasmids Created and Used in these studies.

Table 4: Table of Primers

Target	Name	Sequence 5'-3'	Reference
GA40634	M4 RpoA outside L	ATTACATTCGTATTTTAGTAGGTGACCG	This Study
	M4 RpoA outside R	TTCCCGCTAGGAAAAATCCT	This Study
	Downstream RpoA Overlap L	CACCACCACTGATATAGGAGGACAAAATGGCTTA	This Study
	Downstream RpoA R	GAAAACCATAATCGCCTTATTCTCAG	This Study
	Downstream RpoA XmaI	tttCCC GG GTTAAACTAATTCATAATTGC	This Study
Mga1	Mga1 CBP BglII	tttAGATCTAGTTGTGGAGGG	This Study
	Mga1 CBP NcoI	tttCCATGGATGTATGTAAGT	This Study
	Mga XhoI	gCTCGAGTGTCTAAGTTGTGGAGGG	
	T25N Mga HindIII	AAGCTTAATGTATGTAAGTAAGTTGTTT	This Study
	T25N Mga R	GGATCCTTAGTTGTGGAGGG	This Study
	T25N Mga139 HindIII	AAGCTTAATGTATGTAAGTAAGTTGTTTCAAGTCAACAG	This Study
	T25N Mga139 BamHI	GGATCCTTTTCCCAAGTGATGAAAAAGGC GTAGATCAATT	This Study
	M1 C Mga NdeI	cccCATATGATGAACTGTGAGCGGCTAC	This Study
	Mga1 NdeI	gggggCATATGTATGTATGTAAGTAAGTTG	[73]
	MgaN180 XhoI	gggggCTCGAGCTTTAAGTTTAGG	[73]
Mga4	Mga4 CBP BglII	tttAGATCTTGATGATGTTGCTTG	This Study
	Mga4 CBP NcoI	tttCCATGGATGAAGTTAATGC	This Study
	Mga4CBP pET NdeI	tttCATATGATGCATGTAAGTAAATTG	This Study
	Mga4CBP pET XhoI	tttCTCGAGAAGTGCCCCG	This Study
	Mga4 139 CBP NcoI	tttCCATGGATGCATGTAAGTA	This Study
	Mga4 139 CBP BglII	tttAGATCTTTCCAGGTAATAAAGAAACAATA	This Study
	Mga4 29 CBP BglII	tttAGATCTTTTCGAGGTTT	This Study
	Pami	Pami in vitro L	GTACACAAGGGATATCTGCAGAATT
Pami in vitro R		AAAATACAGGTTTTTCGGTCGTTGG	This Study
Pami Sew L		TCATAGCCGAATAGC	This Study
Pami	Pami Sew R	CCAACCATTATATCAC	This Study
	Pami Sew Inside L	GGATATCTGCAGAATTTCG	This Study
	Pami Sew Inside XbaI	tttTCTAGAGGATATCTGCAGAATTTCGCC	This Study
PaphA3	PaphA3 +30 in vitro L	ACTATGTTATACGCCAACTTTCAA	This Study
	PaphA3 In vitro template XhoI	cccCTCGAGGCAGCGGTATTTTTCGATCAGTTTTT	This Study
pCal-C	pCALC-L	CCTGCCACCATACCCACG	This Study
	pCALC-R	CCCGTTTAGAGGCCCA	This Study
pCIV2	pCIV2 Seq A	GGATCCCCTGCGGTGT	This Study
	pCIV2 Seq B	CATTAGGCACCCAGGC	This Study
Pemm	M1 SF370 Pemm L	GGATCCTCCACAACCTTAGACAGC	[120]
	M1 SF370 Pemm R	CTCGAGCGTGTATTATTTTAGCCA	[120]
	M1 Pemm Luc L	gggGGATCCTCCACAACCTTAGACAGC	[120]
	M1 Pemm Luc R	gggCTCGAGCGTGTATTATTTTAGCCA	[120]
	M1 FPR Pemm L	CCCAGTCACGACGTTGTAAAA	[120]
	M1 FPR Pemm R	CCCTCATTTTCAGGGTTAACTCTAA	[120]
	Pemm 35 In vitro	CATTAATAGCATTTAGGTCAAAAA	This Study
Target	Name	Sequence 5'3'	Reference

	<i>Pemm in vitro</i> SOE L	TAAACCTATTCATTGTTTTAAAAATATCT C	This Study
	<i>Pemm in vitro</i> SOE R	GGTAAAGACCAGCTTTTTTAGCTTTT	This Study
	<i>Pemm in vitro</i> 232 R	CTCCAGCGGTTCCATCCTCT	This Study
	M1 <i>Pemm in vitro</i> 132 R	AATTCGAGCTCCCATCTGAA	This Study
	M1 <i>Pemm in vitro</i> 164 R	ATTTTACCAACAGTACCGGAATG	This Study
	M1 <i>Pemm in vitro</i> 114 R	AACTCTCCTGCATCCTGCA	This Study
pKSM720	720 conf L	ACGACGTTGTAAAACGACGGC	[120]
	720 conf R	AGCCTTATGCAGTTGCTCTCC	[120]
Pmga	Oyl15	CAGTCACGATCACGCAAT	[77]
	Oyr 25	AATTGACTGAAGTATGATAGAAT	[77]
pProEX-htb	pProL	GTGAGCGGATAACAATTT	[55]
Target	Name	Sequence 5'-3'	Reference
pProEX-htb	pProR	AAAATCTTCTCTCATCCG	[55]
PscpA	M1 FPL <i>PscpA</i> L	AGTCCGTAATACGACTCACTTAAGGCCT	[120]
	M1 FPL <i>PscpA</i> R	GCAAACAGGGGTTATTTGCATATGATAC A	[120]
	M1 FPR <i>PscpA</i> L New	TAACGCCAGGGTTTTCCAG	[120]
	M1 FPR <i>PscpA</i> R New	CTTGCTTTTGTGATAATGATTAAATGT	[120]
	M1 <i>PscpA</i> Bam L	gcGGATCCTATGTCTAAAAGAATGAG	[120]
	M1 <i>PscpA</i> Xho R	gcCTCGAGGATGAGAGACTTTGTCTT	[120]
Psic	M1 <i>Psic</i> Luc BglII L	cacAGATCTCAGCAGTTGTAAAACGCAAA G	[120]
	M1 <i>Psic</i> Luc XhoI L	gggCTCGAGTAGTATTCTCTCCTTAATAAA TT	[120]
	M1 FP <i>Psic</i> L	CGCAAAGAAGAAAATAAGCTATC	[120]
	M1 FP <i>Psic</i> R	TGCAGGAATTCCTCGAGTAGTAT	[120]
PscIA	<i>PscIA</i> FPL L	GAAGATCTAACAAACAAGTAAAG	This Study
	<i>PscIA</i> FPL R	ACTTTTTGTGGAGATCAGA	This Study
	<i>PscIA</i> FPR L	AGGGCTACTTTGGCACTTGC	This Study
	<i>PscIA</i> FPR R	CGGCCAGTCCGTAATACGACT	This Study
	<i>PscIA</i> BglII	cccAGATCTAACAAACAAGTAAAGAAGAA ACCTA	This Study
	<i>PscIA</i> w/o MBS BglII	cccAGATCTAAGAAAGGATCCGGATG	This Study
	<i>PscIA</i> XhoI	cccCTCGAGTGGTAGCTAGACCTGATTATT TATA	This Study
Psof	<i>Psof</i> L3	TTTGGTCTCAGACGGCGCCA	[78]
PrpsL	<i>PrpsL</i> +30 <i>in vitro</i> L	CATAAGCAATTGCATCAAAGG	This Study
	<i>PrpsL</i> 35 <i>in vitro</i>	TTCTATTTGACATGAAGTGCCG	This Study
	<i>PrpsL</i> <i>In vitro</i> SOE +10	TGGTCTTTACCAAATTCTATTTGACATG AAGT	This Study
Target	Name	Sequence 5'-3'	Reference

	<i>PrpsL In vitro</i> SOE +15	TGGTCTTTACCAAGGAAAAATTCTATTTG ACAT	This Study
	<i>PrpsL In vitro</i> 1201 R	AACAGCTATGACCATGATTACGCCAA	This Study
RivR	M1 RivR NdeI	gttCATATGTTGGATTATTAT	This Study
	M1 RivR XhoI	gttCTCGAGAGAAGGAACT	This Study
	RivR1 HisMBP XhoI	cccCTCGAGTTAAGAAGGAACTCTCCAAA GTTCTTCTAAACGTTT	This Study
RofA	M1 RofA NdeI	gttCATATGTTGATAGAAAAATACTTGGAA T	This Study
	M1 RofA XhoI	tttCTCGAGTGTTAATTGCTTGGTTAAATCA GCTTGGAAATTT	This Study
	M1 RofA HisMBP XhoI R	cccCTCGAGTTATGTTAATTGCTT	This Study
RpoA	T18C rpoA BamHI L	cccGGATCCATGATTGAGTTTGAAAA	This Study
	T18C rpoA EcoRI R	cccGAATTCTTATTTATCGTTTTTTAGTCCG AGAC	This Study
	T18N rpoA L	cccAAGCTTTATGATTGAGTTTG	This Study
	T18N rpoA R	cccGGATCCATTTTATCGTTTTT	This Study
	T18N his rpoA EcoRI	cccGAATTCTTATCGTTTTTTAGTCCGAGAC CT	This Study
	T18N his rpoA HindIII	cccAAGCTTTCATCACCATCACCA	This Study
	RpoA his Tag L	cccGGATCCATGATTGAGTTTGAAAAACCA ATAATAA	This Study
	RpoA His Tag R	cccCTAGATTATTTATCGTTTTTTAGTCCG AGACC	This Study
	RpoACTD His Tag R	cccCTAGATTAATCGTTCACTTTTTCAGTT TCTTTC	This Study
	rpoA pet21a HindIII	tttAAGCTTTTTATCGTTTTTTAGTCCGAGA CCTAA	This Study
	RpoA pet21a NdeI	tttCATATGATGATTGAGTTTGAAAAACCAA TAATAA	This Study
	RpoACTD pet21a HindIII	tttAAGCTTATCGTTCACTTTTTCAGTTTCTT TCAT	This Study
	RpoACTD Sew Overlap L	TAATGGTTGCGGTCCGTATAATCTGT	This Study
	RpoACTD Sew 2R	ATTAATCGTTCACTTTTTCAGTTTCTTT	This Study
	RpoA Upstream Sew 2L	ATGAAGGTAAGACCATCGGTAA	This Study
	RpoA Upstream Overlap Sew R	GCAGATATCCCTTGTGTACTATTTGT	This Study
	RpoANTD R	ATCGTTCACTTTTTCAGTTTCTTT	This Study
	RpoA Pet Overlap L	GATATAATGGTTGGTTTAACTTTAAGAAG GAGA	This Study
	RpoA Pet R	TCAGTGGTGGTGGTGGTGGT	This Study
	RpoA Trunc1 HindIII	tttAAGCTTAAGATTACGGACTTTCATCA	This Study
	RpoA Trunc2 HindIII	tttAAGCTTACGTTTTAAACAGTTATATGAG CG	This Study
	RpoA Trunc2 XmaI2	tttCCCGGGTTAACGTTTTAAACAGTTAT	This Study
Target	Name	Sequence 5'-3'	Reference

	RpoAhis Trunc1 XmaI	tttCCCGGGTTAAAGATTACG	This Study
RpoD	T18C-rpoD-L	cccGGATCCATGACAAAACAA	This Study
	T18C-rpoD R	cccGAATTCTTAGTCCTCTATAAAGTCT	This Study
	T18N rpoD HindIII	cccAAGCTTTATGACAAAACAAAAGAAA TAACAAC	This Study
	T18N rpoD BamHI	cccGGATCCATGTCTCTATAAAGTCTCTT AATTGTT	This Study
	T18N his rpoD EcoRI	cccGAATTCGAGTCCTCTATAAA	This Study
	Sigma his BamHI	tttGGATCCATGACAAAACAAAAGAAATA ACAAC	This Study
	Sigma His XbaI	cttTCTAGATTAGTCCTCTATAAAGTCTC	This Study
	Sigma hisdelta4 +stop XhoI	tttCTCGAGTTAAAGAACAACACGCGTCGT	This Study
	RpoA-RpoD4 Overlap	GTGAACGATCGTGAGCAATTG	This Study
	RpoDhis RBS EcoRI	tttGAATTCAGGAAACAGACCATG	This Study
	RpoDd4his RBS XbaI	tttTCTAGAGTTAAAGAACAACACGCGT	This Study
	RpoDhis RBS XbaI	tttTCTAGATTAGTCCTCTATAAAGTCTCT	This Study
rpoE	T18N rpoE HindIII	cccAAGCTTAATGGTAGAGAATGATAAAAT AAGGAGAACTG	This Study
	T18N rpoE BamHI	cccGGATCCTGGAGAACTGGTTCTTCATCT TCTTCATCTTCTT	This Study
	T18C rpoE BamHI	cccGGATCCATGGTAGAGAATGATAAAAT	This Study
	T18C rpoE EcoRI	cccGAATTCTTAGAGAACTGGTTCTTCA	This Study
pVL847	pVL847 Seq L	GAAGTCTTACGAGGAAGAGTTGG	This Study
	pLV847 Seq R	GCCCTTTCGTCTTCAAGAATT	This Study

Table 5 Primers

Lower case is clamp; Italics are restriction sites.

Table 6: Table of Mutagenic Oligonucleotides

Target	Name	Sequence 5'-3'	Reference
Pemml-TOPO	<i>Pemml</i> C3A SDM L	TCAAAAACAGATTCATCATTAATAG <u>A</u> ATTTAGGTCAA AAAGGTGGCAAAAG	[120]
	<i>Pemml</i> C3A SDM R	CTTTTGCCACCTTTTTGACCTAAAT <u>T</u> CTATTAATGAT GAATCTGTTTTTGA	[120]
	<i>Pemml</i> G9A SDM L	ACTCAAAAACAGATTCATCATTAATAGCATTTA <u>A</u> GTC AAAAAGGTGGCAAAA	[120]
	<i>Pemml</i> G9A SDM R	TTTTTGCCACCTTTTTGAC <u>T</u> TAAATGCTATTAATGATG AATCTGTTTTTGAGT	[120]
	<i>Pemml</i> G10A SDM L	CAGATTCATCATTAATAGCATTTAG <u>A</u> TCAAAAGGTG GCAAAAGCTAAAAA	[120]
	<i>Pemml</i> G10A SDM R	TTTTTAGCTTTTTGCCACCTTTTTG <u>A</u> TCTAAATGCTAT TAATGATGAATCTG	[120]
	<i>Pemml</i> T11C SDM L	GATTCATCATTAATAGCATTTAGG <u>C</u> CAAAAAGGTGGC AAAAGCTAAAAA	[120]
	<i>Pemml</i> T11C SDM R	TTTTTAGCTTTTTGCCACCTTTTTG <u>G</u> CCATAATGCTAT TAATGATGAATC	[120]
	<i>Pemml</i> C12A SDM L	GATTCATCATTAATAGCATTTAGGT <u>A</u> AAAAAGGTGGC AAAAGCTAAAAAAG	[120]
	<i>Pemml</i> C12A SDM R	CTTTTTTAGCTTTTTGCCACCTTTTT <u>T</u> ACCTAAATGCT ATTAATGATGAATC	[120]
	<i>Pemml</i> A13C SDM L	TTCATCATTAATAGCATTTAGGT <u>C</u> AAAAGGTGGCAA AAGCTAAAAAAG	[120]
	<i>Pemml</i> A13C SDM R	CTTTTTTAGCTTTTTGCCACCTTTTT <u>G</u> GACCTAAATGCT ATTAATGATGAA	[120]
	<i>Pemml</i> G18A SDM L	ACAGATTCATCATTAATAGCATTTAGGTCAAAAA <u>A</u> GT GGCAAAAGCTAAAAA	[120]
	<i>Pemml</i> G18A SDM R	TTTTTTAGCTTTTTGCCAC <u>T</u> TTTTTGACCTAAATGCTA TTAATGATGAATCTGT	[120]
	<i>Pemml</i> G19A SDM L	TTAATAGCATTTAGGTCAAAAAG <u>A</u> TGGCAAAAGCTAA AAAAGCTGG	[120]
	<i>Pemml</i> G19A SDM R	CCAGCTTTTTTAGCTTTTTGCCA <u>T</u> CTTTTTGACCTAAA TGCTATTAA	[120]
	<i>Pemml</i> C23A SDM L	GCATTTAGGTCAAAAAGGTGG <u>A</u> AAAAGCTAAAAAGC TGGTCT	[120]
	<i>Pemml</i> C23A SDM R	AGACCAGCTTTTTTAGCTTTTT <u>T</u> CCACCTTTTTGACCT AAATGC	[120]
	<i>Pemml</i> C29A SDM L	GGTCAAAAAGGTGGCAAAAG <u>A</u> TAAAAAGCTGGTCTT TACC	[120]
Pemml-TOPO	<i>Pemml</i> C29A SDM R	GGTAAAGACCAGCTTTTTT <u>A</u> TCTTTTTGCCACCTTTTT GACC	[120]
	<i>Pemml</i> A33C SDM L	TAGGTCAAAAAGGTGGCAAAAGCTAA <u>C</u> AAAGCTGGTC TTTAC	[120]
	<i>Pemml</i> A33C SDM R	GTAAGACCAGCTTT <u>G</u> TTAGCTTTTTGCCACCTTTTTG ACCTA	[120]
	<i>Pemml</i> A34C SDM L	AGGTCAAAAAGGTGGCAAAAGCTAA <u>C</u> AAGCTGGTCT TTAC	[120]
	<i>Pemml</i> A34C SDM R	GTAAGACCAGCTT <u>G</u> TTTAGCTTTTTGCCACCTTTTTG ACCT	[120]
Target	Name	Sequence 5'-3'	Reference

	<i>Pemml</i> A35C SDM L	AGGTCAAAAAGGTGGCAAAAGCTAAAA <u>C</u> AGCTGGTCT TTACC	[120]
	<i>Pemml</i> A35C SDM R	GGTAAAGACCAGCT <u>G</u> TTTTAGCTTTTGCCACTTTTT GACCT	[120]
	<i>Pemml</i> G37A SDM L	TCAAAAAGGTGGCAAAAGCTAAAAAA <u>A</u> CTGGTCTTTA CTTTTTGG	[120]
	<i>Pemml</i> G37A SDM R	CCAAAAGGTAAAGACCAG <u>T</u> TTTTTTAGCTTTTGCCAC CTTTTTGA	[120]
	<i>Pemml</i> C38A SDM L	GGTGGCAAAAGCTAAAAAAG <u>A</u> TGGTCTTTACTTTTG GCTT	[120]
	<i>Pemml</i> C38A SDM R	AAGCCAAAAGGTAAAGACC <u>A</u> TCTTTTTTAGCTTTTGC CACC	[120]
	<i>Pemml</i> T39C SDM L	GTGGCAAAAGCTAAAAAAGC <u>C</u> GGTCTTTACTTTTGG CTTT	[120]
	<i>Pemml</i> T39C SDM R	AAAGCCAAAAGGTAAAGACC <u>G</u> GCTTTTTTAGCTTTT CCAC	[120]
	<i>Pemml</i> G40A SDM L	GGTGGCAAAAGCTAAAAAAGCT <u>A</u> GTCTTTACTTTTG GCTTTTAT	[120]
	<i>Pemml</i> G40A SDM R	ATAAAAGCCAAAAGGTAAAGAC <u>T</u> AGCTTTTTTAGCTT TTGCCACC	[120]
	<i>Pemml</i> G41A SDM L	GGTGGCAAAAGCTAAAAAAGCT <u>G</u> ATCTTTACTTTTG GCTTTTATTA	[120]
	<i>Pemml</i> G41A SDM R	TAATAAAAGCCAAAAGGTAAAGAT <u>C</u> AGCTTTTTTAGC TTTTGCCACC	[120]
	<i>Pemml</i> C43A SDM L	GTGGCAAAAGCTAAAAAAGCTGGT <u>A</u> TTTACTTTTGG CTTTTATTTT	[120]
	<i>Pemml</i> C43A SDM R	AAATAATAAAAGCCAAAAGGTAA <u>A</u> TACCAGCTTTTTT AGCTTTTGCCAC	[120]
	<i>Pemml</i> T44C SDM L	GGCAAAAGCTAAAAAAGCTGGT <u>C</u> TTACTTTTGGCT TTTATTATTT	[120]
Pemml- TOPO	<i>Pemml</i> T44C SDM R	AAATAATAAAAGCCAAAAGGTAA <u>G</u> GACCAGCTTTTTT AGCTTTTGCC	[120]
	<i>Pemml</i> T45C SDM L	GGCAAAAGCTAAAAAAGCTGGT <u>C</u> TACCTTTTGGCT TTTATTATTAC	[120]
	<i>Pemml</i> T45C SDM R	AATAATAAAAGCCAAAAGGT <u>A</u> GAGACCAGCTTTTTTA GCTTTTGCC	[120]
PscpA1- TOPO	<i>PscpA1</i> C12A SDM L	TCTAAAAGAATGTGGATAAGGAGGT <u>A</u> ACAAACTAAGC AACTCTTAA	[120]
	<i>PscpA1</i> C12A SDM R	TTTAAGAGTTGCTTAGTTTGT <u>T</u> ACCTCCTTATCCTCA TTCTTTTAGA	[120]
	<i>PscpA1</i> G40A SDM L	CTAAGCAACTCTTAAAAAGCT <u>A</u> ACCTTTACTAATAAT CATC	[120]
	<i>PscpA1</i> G40A SDM R	GATGATTATTAGTAAAGG <u>T</u> TAGCTTTTTTAAGAGTTGC TTAG	[120]
	<i>PscpA1</i> C43A SDM L	ACAAACTAAGCAACTCTTAAAAAGCTGAC <u>A</u> TTTACTA ATAATCATCTTTGTTTTATAAT	[120]
	<i>PscpA1</i> C43A SDM R	ATTATAAAACAAAGATGATTATTAGTAA <u>A</u> TGTCAGCT TTTTAAGAGTTGCTTAGTTTGT	[120]
pKSM245	<i>Psic1</i> C12A SDM L	AATGAGGTAAAGGAGAGGT <u>A</u> ACAAACTAAACAACCTC	[120]
	<i>Psic1</i> C12A SDM R	GAGTTGTTTTAGTTTTGT <u>T</u> ACCTCCTTAAACCTCATT	[120]
Target	Name	Sequence 5'-3'	Reference

	<i>PsiI</i> G40A SDM L	TAAACAACCTCTTAAAAAGCT <u>A</u> ACCTTTACTAATAATC GTC	[120]
	<i>PsiI</i> G40A SDM R	GACGATTATTAGTAAAGGT <u>T</u> AGCTTTTTTAAGAGTTGT TTA	[120]
	<i>PsiI</i> C43A SDM L	CAACTCTTAAAAAGCTGAC <u>A</u> TTTACTAATAATCGTCT TTG	[120]
	<i>PsiI</i> C43A SDM R	CAAAGACGATTATTAGTAA <u>T</u> GTCAGCTTTTTAAGAG TTG	[120]
Mga4	M4 mga HTH4 a	GCTGAAGGCTGTTTGT <u>CAGC</u> <u>G</u> <u>C</u> AGCTACCCTCAAAC GCC	[79]
	M4 Mga hth4 b	GCTGAAGAGCTCTTTGTGAGC <u>G</u> <u>C</u> <u>T</u> <u>G</u> CAACACTCAAG CGTCTCATC	[79]
	M4 Mga HTH3 A	TATGCAGTTCATGAAAGAAGTAGGTGGAATTAC <u>C</u> <u>G</u> <u>C</u> <u>G</u> <u>C</u> AGACGGCTATATTAATATTTG	This study
	M4 Mga HTH3 B	CAAATATTAATATAGCCGT <u>C</u> <u>G</u> <u>C</u> <u>G</u> <u>C</u> GTAATTCCAC CTACTTCTTTCATGAACTGCATA	This study

Mutations are bolded and underlined.

Table 7: Table of qPCR Primers

Target	Name	Sequence 5'-3'	Reference
<i>gyrA</i>	M4 <i>gyrA</i> RT-L	AGTGTCATTGTGGCAAGAGC	[72]
	M4 <i>gyrA</i> RT-R	CACACCGAGTTCATTCATCC	[72]
<i>Parp</i>	M4 <i>Parp</i> qPCR L	GGAAGCCCTTCCTCTTTT	This Study
	M4 <i>Parp</i> qPCR R	GCGGTAAAAGGTAAAGACCAG	This Study
<i>PmalR</i>	M4 <i>PmalR</i> qPCR L	AACCTGATCCACATCCCACT	This Study
	M4 <i>PmalR</i> qPCR R	AGCTTGAAATCATGGCAAAAA	This Study
<i>Pmrp</i>	M4 <i>Pmrp</i> qPCR L	TAGGATTTTCAGACGTCATGGT	This Study
	M4 <i>Pmrp</i> qPCR R	AGCCAAAAGGTAAAGGTCAGT	This Study
<i>PrpsL</i>	M4 <i>PrpsL</i> qPCR L	GCAATTGCATCAAAGGAAAAA	This Study
	M4 <i>PrpsL</i> qPCR R	GCAACAATTGTCAGCACGTC	This Study

qPCR Primers used for validation of ChAP DNA.

Table 8: Table of Binding Oligonucleotides

Name	Sequence 5'-3'	Reference
<i>Pemml</i> MBS 49mer	AGCATTTAGGTCAAAAAGGTGGCAAAAGCTAAAAAAGCTGGTCTTTACC TCGTAAATCCAGTTTTTCCACCGTTTTTCGATTTTTTTCGACCAGAAATGG	[73]
<i>Pemml</i> G9C43 35mer	GGTCAAAAAGGTGGCAAAAGCTAAAAAAGCTGGTCCAGTTTTTCCACCGTTTTTCGATTTTTTTCGACCAG	[120]
<i>Pemml</i> 3-20	CATTTAGGTCAAAAAGGT GTAAATCCAGTTTTTCCA	This Study
<i>Pemml</i> 28-47	GCTAAAAAAGCTGGTCTTTA CGATTTTTTTCGACCAGAAAT	This Study
<i>Pemml</i> G9A MBS 49mer	AGCATTTAAGTCAAAAAGGTGGCAAAAGCTAAAAAAGCTGGTCTTTACC TCGTAAATTCAGTTTTTCCACCGTTTTTCGATTTTTTTCGACCAGAAATGG	[120]
<i>Pemml</i> G10A MBS 49mer	AGCATTTAGATCAAAAAGGTGGCAAAAGCTAAAAAAGCTGGTCTTTACC TCGTAAATCTAGTTTTTCCACCGTTTTTCGATTTTTTTCGACCAGAAATGG	[120]
<i>Pemml</i> T11C MBS 49mer	AGCATTTAGGC ^C CAAAAAGGTGGCAAAAGCTAAAAAAGCTGGTCTTTACC TCGTAAATCCGGTTTTTCCACCGTTTTTCGATTTTTTTCGACCAGAAATGG	[120]
<i>Pemml</i> C12A MBS 49mer	AGCATTTAGGTAAAAAAGGTGGCAAAAGCTAAAAAAGCTGGTCTTTACC TCGTAAATCCATTTTTTCCACCGTTTTTCGATTTTTTTCGACCAGAAATGG	[120]
<i>Pemml</i> A13C MBS 49mer	AGCATTTAGGTCCAAAAGGTGGCAAAAGCTAAAAAAGCTGGTCTTTACC TCGTAAATCCAGGTTTTTCCACCGTTTTTCGATTTTTTTCGACCAGAAATGG	[120]
<i>Pemml</i> G18A MBS 49mer	AGCATTTAGGTCAAAAAGGTGGCAAAAGCTAAAAAAGCTGGTCTTTACC TCGTAAATCCAGTTTTTTCACCGTTTTTCGATTTTTTTCGACCAGAAATGG	[120]
<i>Pemml</i> G19A MBS 49mer	AGCATTTAGGTCAAAAAGATGGCAAAAGCTAAAAAAGCTGGTCTTTACC TCGTAAATCCAGTTTTTTCACCGTTTTTCGATTTTTTTCGACCAGAAATGG	[120]
<i>Pemml</i> C23A MBS 49mer	AGCATTTAGGTCAAAAAGGTGGAAAAAGCTAAAAAAGCTGGTCTTTACC TCGTAAATCCAGTTTTTCCACCTTTTTTCGATTTTTTTCGACCAGAAATGG	[120]
<i>Pemml</i> C29A MBS 49mer	AGCATTTAGGTCAAAAAGGTGGCAAAAAGATAAAAAAGCTGGTCTTTACC TCGTAAATCCAGTTTTTCCACCGTTTTCTATTTTTTTCGACCAGAAATGG	[120]
<i>Pemml</i> A33C MBS 49mer	AGCATTTAGGTCAAAAAGGTGGCAAAAAGCTAAACAAGCTGGTCTTTACC TCGTAAATCCAGTTTTTCCACCGTTTTTCGATTTGTTTCGACCAGAAATGG	[120]
<i>Pemml</i> A34C MBS 49mer	AGCATTTAGGTCAAAAAGGTGGCAAAAAGCTAAAACAAGCTGGTCTTTACC TCGTAAATCCAGTTTTTCCACCGTTTTTCGATTTGTTTCGACCAGAAATGG	[120]
<i>Pemml</i> A35C MBS 49mer	AGCATTTAGGTCAAAAAGGTGGCAAAAAGCTAAAAAGCTGGTCTTTACC TCGTAAATCCAGTTTTTCCACCGTTTTTCGATTTGTTTCGACCAGAAATGG	[120]
<i>Pemml</i> G37A MBS 49mer	AGCATTTAGGTCAAAAAGGTGGCAAAAAGCTAAAAAAAGCTGGTCTTTACC TCGTAAATCCAGTTTTTCCACCGTTTTTCGATTTTTTTCGACCAGAAATGG	[120]
<i>Pemml</i> C38A MBS 49mer	AGCATTTAGGTCAAAAAGGTGGCAAAAAGCTAAAAAAGATGGTCTTTACC TCGTAAATCCAGTTTTTCCACCGTTTTTCGATTTTTTTCGACCAGAAATGG	[120]
<i>Pemml</i> T39C MBS 49mer	AGCATTTAGGTCAAAAAGGTGGCAAAAAGCTAAAAAAGC ^C GGTCTTTACC TCGTAAATCCAGTTTTTCCACCGTTTTTCGATTTTTTTCGACCAGAAATGG	[120]
<i>Pemml</i> G40A MBS 49mer	AGCATTTAGGTCAAAAAGGTGGCAAAAAGCTAAAAAAGCTAGTCTTTACC TCGTAAATCCAGTTTTTCCACCGTTTTTCGATTTTTTTCGATCAGAAATGG	[119]
<i>Pemml</i> G41A MBS 49mer	AGCATTTAGGTCAAAAAGGTGGCAAAAAGCTAAAAAAGCTGATCTTTACC TCGTAAATCCAGTTTTTCCACCGTTTTTCGATTTTTTTCGACTAGAAATGG	[120]
<i>Pemml</i> C43A MBS 49mer	AGCATTTAGGTCAAAAAGGTGGCAAAAAGCTAAAAAAGCTGGTATTTACC TCGTAAATCCAGTTTTTCCACCGTTTTTCGATTTTTTTCGACCATAAATGG	[120]
<i>Pemml</i> T44C MBS 49mer	AGCATTTAGGTCAAAAAGGTGGCAAAAAGCTAAAAAAGCTGGTCTTTACC TCGTAAATCCAGTTTTTCCACCGTTTTTCGATTTTTTTCGACCAGGAATGG	[120]
<i>Pemml</i> T45C MBS 49mer	AGCATTTAGGTCAAAAAGGTGGCAAAAAGCTAAAAAAGCTGGTCTCTACC TCGTAAATCCAGTTTTTCCACCGTTTTTCGATTTTTTTCGACCAGAGATGG	[120]
Name	Sequence 5'-3'	Reference

<i>Pem1</i> C12/43A MBS 49mer	AGCATTTAGGTAAAAAGGTGGCAAAGCTAAAAAGCTGGT A TTTACC TCGTAAATCCATTTTTTCCACCGTTTTTCGATTTTTTCGACC A TAAATGG	[120]
<i>PsepAI</i> MBS 49mer	GGATAAGGAGGTCACAACTAAGCAACTCTTAAAAAGCTGACCTTTACT CCTATTCCTCCAGTGTGGATTTCGTTGAGAATTTTTTCGACTGGAAATGA	[120]
<i>PsepAI</i> G9C43 34mer	GGTCACAACTAAGCAACTCTTAAAAAGCTGACC CCTCCAGTGTGGATTTCGTTGAGAATTTTTTCGACTGGA	This Study
<i>PsepAI</i> C12A MBS 49mer	GGATAAGGAGGTAACAACTAAGCAACTCTTAAAAAGCTGACCTTTACT CCTATTCCTCCAT T GTGGATTTCGTTGAGAATTTTTTCGACTGGAAATGA	[120]
<i>PsepAI</i> G40A MBS 49mer	GGATAAGGAGGTCACAACTAAGCAACTCTTAAAAAGCT A ACCTTTACT CCTATTCCTCCAGTGTGGATTTCGTTGAGAATTTTTTCGAT T TGGAAATGA	[120]
<i>PsepAI</i> C43A MBS 49mer	GGATAAGGAGGTCACAACTAAGCAACTCTTAAAAAGCTGAC A TTTACT CCTATTCCTCCAGTGTGGATTTCGTTGAGAATTTTTTCGACTG T AATGA	[120]
<i>PsepAI</i> C12/43A MBS 49mer	GGATAAGGAGGTAACAACTAAGCAACTCTTAAAAAGCTGAC A TTTACT CCTATTCCTCCAT T GTGGATTTCGTTGAGAATTTTTTCGACTG T AATGA	[120]
<i>PsicI</i> MBS 49mer	GTAAGGAGAGGTCACAACTAAACAACCTCTTAAAAAGCTGACCTTTACT CATTCTCTCCAGTGTGGATTTCGTTGAGAATTTTTTCGACTGGAAATGA	[120]
<i>PsicI</i> C12A MBS 49mer	GTAAGGAGAGGTAACAACTAAACAACCTCTTAAAAAGCTGACCTTTACT CATTCTCTCCAT T GTGGATTTCGTTGAGAATTTTTTCGACTGGAAATGA	[120]
<i>PsicI</i> G40A MBS 49mer	GTAAGGAGAGGTCACAACTAAACAACCTCTTAAAAAGCT A ACCTTTACT CATTCTCTCCAGTGTGGATTTCGTTGAGAATTTTTTCGAT T TGGAAATGA	[120]
<i>PsicI</i> C43A MBS 49mer	GTAAGGAGAGGTCACAACTAAACAACCTCTTAAAAAGCTGAC A TTTACT CATTCTCTCCAGTGTGGATTTCGTTGAGAATTTTTTCGACTG T AATGA	[120]
<i>PsicI</i> C12/43A MBS 49mer	GTAAGGAGAGGTAACAACTAAACAACCTCTTAAAAAGCTGAC A TTTACT CATTCTCTCCAT T GTGGATTTCGTTGAGAATTTTTTCGACTG T AATGA	[120]
MBS Random 49mer	TTTAGAAACAAAGGCATCAGTCGACCTGAAGCTATTTAGAAAAAGGGTC AAATCTTTGTTCCGTAGTCACGTGGACTTCGATAAATCTTTTCCAG	[120]

Mutations are underlined and bold.

References

1. Denny Jr. FW (2000) History of hemolytic streptococci and associated diseases. In: Stevens DL, Kaplan EL, editors. *Streptococcal Infections: Clinical Aspects, Microbiology, and Molecular Pathogenesis*. New York: Oxford University Press. pp. 1–18.
2. Todd EW, Hewitt LF (1932) A new culture medium for the production of antigenic streptococcal haemolysin. *Journal of Pathology and Bacteriology* 35: 973–974.
3. Vincent WF, Lisiewski KJ (1969) Improved growth medium for group A streptococci. *Applied and Environmental Microbiology* 18: 954–955.
4. Cunningham MW (2000) Pathogenesis of group A streptococcal infections. *Clinical Microbiology Reviews* 13: 470–511.
5. Lancefield RC (1933) A serological differentiation of human and other groups of hemolytic streptococci. *Journal of Experimental Medicine* 57: 571–595.
6. Bessen DE (2009) Population biology of the human restricted pathogen, *Streptococcus pyogenes*. *Infection, Genetics, and Evolution* 9: 581–593.
7. Bessen DE, Jones KF, Fischetti VA (1989) Evidence for two distinct classes of streptococcal M protein and their relationship to rheumatic fever. *The Journal of Experimental Medicine* 169: 269–283.
8. Johnson DR, Kaplan EL (1993) A review of the correlation of T-agglutination patterns and M-protein typing and opacity factor production in the identification of group A streptococci. *Journal of Medical Microbiology* 38: 311–315.

9. Carapetis JR, Steer AC, Mulholland EK, Weber M (2005) The global burden of group A streptococcal diseases. *Lancet Infectious Diseases* 5: 685–694.
10. Shulman ST, Tanz RR, Gerber MA (2000) Streptococcal Pharyngitis. In: Stevens DL, Kaplan EL, editors. *Streptococcal Infections: Clinical Aspects, Microbiology, and Molecular Pathogenesis*. New York: Oxford University Press. pp. 76–97.
11. Stevens DL (2000) Life-threatening streptococcal infections: scarlet fever, necrotizing fasciitis, myositis, bacteremia, and streptococcal toxic shock syndrome. In: Stevens DL, Kaplan EL, editors. *Streptococcal Infections: Clinical Aspects, Microbiology, and Molecular Pathogenesis*. New York: Oxford University Press. pp. 163–179.
12. Anthony BF (2000) Streptococcal pyoderma. In: Stevens DL, Kaplan EL, editors. *Streptococcal Infections: Clinical Aspects, Microbiology, and Molecular Pathogenesis*. New York: Oxford University Press. pp. 144–151.
13. Bisno AL, Stevens DL (1996) Streptococcal infections of skin and soft tissues. *New England Journal of Medicine* 334: 240–245.
14. Markowitz M, Kaplan EL (2000) Rheumatic Fever. In: Stevens DL, Kaplan EL, editors. *Streptococcal Infections: Clinical Aspects, Microbiology, and Molecular Pathogenesis*. New York Oxford: Oxford University Press. pp. 133–143.
15. Holm SE, Nordstrand A, Stevens DL, Norgreen M (2000) Acute post-streptococcal glomerulonephritis. In: Stevens DL, Kaplan EL, editors. *Streptococcal Infections: Clinical Aspects, Microbiology, and Molecular Pathogenesis*. New York: Oxford University Press. pp. 152–162.

16. Steer AC, Batzloff MR, Mulholland K, Carapetis JR (2009) Group A streptococcal vaccines: facts versus fantasy. *Current Opinion in Infectious Diseases* 22: 544–552.
17. Lancefield RC (1962) Current knowledge of type-specific M antigens of group A streptococci. *Journal of Immunology* 89: 307–313.
18. Hollingshead SK, Readdy TL, Yung DL, Bessen DE (1993) Structural heterogeneity of the *emm* gene cluster in group A streptococci. *Molecular Microbiology* 8: 707–717.
19. Ghosh P (2011) The Nonideal Coiled Coil of M protein and its Multifarious Functions in Pathogenesis. In: Linke D, Goldman A, editors. *Bacterial Adhesion*. Dordrecht: Springer Netherlands, Vol. 715. pp. 197–211.
20. Pålman L, Olin A (2008) Soluble M1 protein of *Streptococcus pyogenes* triggers potent T cell activation. *Cellular Microbiology* 10: 404–414.
21. Jeng A, Sakota V, Li Z, Datta V, Beall B, et al. (2003) Molecular genetic analysis of a group A streptococcus operon encoding serum opacity factor and a novel fibronectin-binding protein, SfbX. *Journal of Bacteriology* 185: 1208–1217.
22. Terao Y, Kawabata S, Kunitomo E, Murakami J, Nakagawa I, et al. (2001) Fba, a novel fibronectin-binding protein from *Streptococcus pyogenes*, promotes bacterial entry into epithelial cells, and the *fba* gene is positively transcribed under the Mga regulator. *Molecular Microbiology* 42: 75–86.
23. Lukomski S, Nakashima K, Abdi I, Cipriano VJ, Ireland RM, et al. (2000) Identification and characterization of the *scl* gene encoding a group A

- streptococcus extracellular protein virulence factor with similarity to human collagen. *Infection and Immunity* 68: 6542–6553.
24. Caswell CC, Lukomska E, Seo NS, Höök M, Lukomski S (2007) Scl1-dependent internalization of group A *Streptococcus* via direct interactions with the $\alpha 2\beta(1)$ integrin enhances pathogen survival and re-emergence. *Molecular Microbiology* 64: 1319–1331.
 25. Pählman LI, Marx PF, Mörgelin M, Lukomski S, Meijers JCM, et al. (2007) Thrombin-activatable fibrinolysis inhibitor binds to *Streptococcus pyogenes* by interacting with collagen-like proteins A and B. *The Journal of Biological Chemistry* 282: 24873–24881. doi:10.1074/jbc.M610015200.
 26. Cleary PP, Prahbu U, Dale JB, Wexler DE, Handley J (1992) Streptococcal C5a peptidase is a highly specific endopeptidase. *Infection and Immunity* 60: 5219–5223.
 27. Ji Y, McLandsborough L, Kondagunta A, Cleary PP (1996) C5a peptidase alters clearance and trafficking of group A streptococci by infected mice. *Infection and Immunity* 64: 503–510.
 28. Schragger HM, Alberti S, Cywes C, Dougherty GJ, Wessels MR (1998) Hyaluronic acid capsule modulates M protein-mediated adherence and acts as a ligand for attachment of group A streptococcus to CD44 on human keratinocytes. *Journal of Clinical Investigation* 101: 1708–1716.
 29. Hondorp ER, McIver KS (2007) The Mga virulence regulon: infection where the grass is greener. *Molecular Microbiology* 66: 1056–1065.

30. Courtney HS, Hasty DL, Dale JB (2006) Anti-phagocytic mechanisms of *Streptococcus pyogenes*: binding of fibrinogen to M-related protein. *Molecular Microbiology* 59: 936–947.
31. Courtney HS, Hasty DL, Dale JB (2003) Serum opacity factor (SOF) of *Streptococcus pyogenes* evokes antibodies that opsonize homologous and heterologous SOF-positive serotypes of group A streptococci. *Infection and Immunity* 71: 5097–5103.
32. Katerov V, Lindgren PE, Totolian AA, Schalen C (2000) Streptococcal opacity factor: a family of bifunctional protein with lipoproteinase and fibronectin-binding activities. *Current Microbiology* 40: 149–156.
33. Lukomski S, Sreevatsan S (1997) Inactivation of *Streptococcus pyogenes* extracellular cysteine protease significantly decrease mouse lethality of serotype M3 and M49 strains. *Journal of Clinical Investigation* 99: 2574–2580.
34. Lukomski S, Montgomery CA, Rurangirwa J, Geske RS, Barrish JP, et al. (1999) Extracellular cysteine protease produced by *Streptococcus pyogenes* participates in the pathogenesis of invasive skin infection and dissemination in mice. *Infection and Immunity* 67: 1779–1788.
35. Matsuka YV, Pillai S, Gubba S, Musser JM, Olmsted SB (1999) Fibrinogen cleavage by the *Streptococcus pyogenes* extracellular cysteine protease and generation of antibodies that inhibit enzyme proteolytic activity. *Infection and Immunity* 67: 4326–4333.
36. Ashbaugh CD, Moser TJ, Shearer MH, White GL, Kennedy RC, et al. (2000) Bacterial determinants of persistent throat colonization and the associated immune

- response in a primate model of human group A streptococcal pharyngeal infection. *Cellular Microbiology* 2: 283–292.
37. Ashbaugh CD, Warren HB, Carey VJ, Wessels MR (1998) Molecular analysis of the role of the group A streptococcal cysteine protease, hyaluronic acid capsule, and M protein in a murine model of human invasive soft-tissue infection. *Journal of Clinical Investigation* 102: 550–560.
 38. Collin M, Svensson MD, Sjöholm AG, Jensenius JC, Sjöbring U, et al. (2002) EndoS and SpeB from *Streptococcus pyogenes* inhibit immunoglobulin-mediated opsonophagocytosis. *Infect* 70: 6646–6651.
 39. Eriksson A, Norgren M (2003) Cleavage of antigen-bound immunoglobulin G by SpeB contributes to streptococcal persistence in opsonizing blood. *Infection and Immunity* 71: 211–217.
 40. Nyberg P, Rasmussen M, Von Pawel-Rammingen U, Björck L (2004) SpeB modulates fibronectin-dependent internalization of *Streptococcus pyogenes* by efficient proteolysis of cell-wall-anchored protein F1. *Microbiology* 150: 1559–1569.
 41. Wei L, Pandiripally V, Gregory E, Clymer M, Cue D (2005) Impact of the SpeB protease on binding of the complement regulatory proteins factor H and factor H-Like protein 1 by *Streptococcus pyogenes*. *Infection and Immunity* 73: 2040–2050.
 42. Fernie-King BA, Seilly DJ, Willers C, Wurznner R, Davies A, et al. (2001) Streptococcal inhibitor of complement (SIC) inhibits the membrane attack complex by preventing uptake of C567 onto cell membranes. *Immunology* 103: 390–398.

43. Frick IM, Åkesson P, Rasmussen M, Schmidtchen A, Björck L (2003) SIC, a secreted protein of *Streptococcus pyogenes* that inactivates antibacterial peptides. *Journal of Biological Chemistry* 278: 16561–16566.
44. Hoe NP, Ireland RM, DeLeo FR, Gowen BB, Dorward DW, et al. (2002) Insight into the molecular basis of pathogen abundance: group A streptococcus inhibitor of complement inhibits bacterial adherence and internalization into human cells. *Proceedings of the National Academy of Sciences* 99: 7646–7651.
45. Fernie-King BA, Seilly DJ, Davies A, Lachmann PJ (2002) Streptococcal inhibitor of complement inhibits two additional components of the mucosal innate immune system: secretory leukocyte proteinase inhibitor and lysozyme. *Infection and Immunity* 70: 4908–4916.
46. Graham MR, Smoot LM, Migliaccio CA, Virtaneva K, Sturdevant DE, et al. (2002) Virulence control in group A streptococcus by a two-component gene regulatory system: global expression profiling and in vivo infection modeling. *Proceedings of the National Academy of Sciences* 99: 13855–13860.
47. Dalton TL, Scott JR (2004) CovS inactivates CovR and is required for growth under conditions of general stress in *Streptococcus pyogenes*. *Journal of Bacteriology* 186: 3928–3937.
48. Sumby P, Whitney AR, Graviss EA, DeLeo FR, Musser JM (2006) Genome-wide analysis of group a streptococci reveals a mutation that modulates global phenotype and disease specificity. *PLoS Pathogens* 2: e5.

49. Walker MJ, Hollands A, Sanderson-Smith ML, Cole JN, Kirk JK, et al. (2007) DNase Sda1 provides selection pressure for a switch to invasive group A streptococcal infection. *Nature Medicine* 13: 981–985.
50. Kansal RG, Datta V, Aziz RK, Abdeltawab NF, Rowe S, et al. (2010) Dissection of the molecular basis for hypervirulence of an in vivo-selected phenotype of the widely disseminated M1T1 strain of group A *Streptococcus* bacteria. *Journal of Infectious Disease* 201: 855–865.
51. Leday TV, Gold KM, Kinkel TL, Roberts SA, Scott JR, et al. (2008) TrxR, a new CovR-repressed response regulator that activates the Mga virulence regulon in group A *Streptococcus*. *Infection and Immunity* 76: 4659–4668.
52. Almengor AC, Kinkel TL, Day SJ, McIver KS (2007) The catabolite control protein CcpA binds to *Pmga* and influences expression of the virulence regulator Mga in the group A streptococcus. *Journal of Bacteriology* 189: 8405–8416.
53. Roberts SA, Churchward GG, Scott JR (2007) Unraveling the regulatory network in *Streptococcus pyogenes*: the global response regulator CovR represses rivR directly. *Journal of Bacteriology* 189: 1459–1463.
54. Roberts SA, Scott JR (2007) RivR and the small RNA RivX: the missing links between the CovR regulatory cascade and the Mga regulon. *Molecular Microbiology* 66: 1506–1522.
55. Gold KM (2011) PhD Dissertation. Characterization of the TrxSR Two-Component Signal Transduction System of *Streptococcus Pyogenes* and its Role in Virulence Regulation University of Maryland, College Park.

56. Stulke J, Hillen W (1999) Carbon catabolite repression in bacteria. *Current Opinion in Microbiology* 2: 195–201.
57. Kinkel TL, McIver KS (2008) CcpA-mediated repression of streptolysin S expression and virulence in the group A streptococcus. *Infection and Immunity* 76: 3451–3463.
58. Shelburne 3rd SA, Keith D, Horstmann N, Sumbly P, Davenport MT, et al. (2008) A direct link between carbohydrate utilization and virulence in the major human pathogen group A Streptococcus. *Proceedings of the National Academy of Sciences* 105: 1698–1703.
59. Lyon WR, Gibson CM, Caparon MG (1998) A role for trigger factor and an rgg-like regulator in the transcription, secretion and processing of the cysteine proteinase of *Streptococcus pyogenes*. *EMBO Journal* 17: 6263–6275.
60. Chaussee MS, Sylva GL, Sturdevant DE, Smoot LM, Graham MR, et al. (2002) Rgg influences the expression of multiple regulatory loci to coregulate virulence factor expression in *Streptococcus pyogenes*. *Infection and Immunity* 70: 762–770.
61. Mashburn-Warren L, Morrison D a, Federle MJ (2012) The cryptic competence pathway in *Streptococcus pyogenes* is controlled by a peptide pheromone. *Journal of Bacteriology* 194: 4589–4600.
62. Kreikemeyer B, Nakata M, Koller T, Hildisch H, Kourakos V, et al. (2007) The *Streptococcus pyogenes* serotype M49 Nra-Ralp3 transcriptional regulatory network and its control of virulence factor expression from the novel eno ralp3 epf sagA pathogenicity region. *Infection and Immunity* 75: 5698–5710.

63. Kreikemeyer B, Beckert S, Braun-Kiewnick A, Podbielski A (2002) Group A streptococcal RofA-type global regulators exhibit a strain-specific genomic presence and regulation pattern. *Microbiology* 148: 1501–1511.
64. Luo F, Lizano S, Bessen DE (2008) Heterogeneity in the polarity of Nra regulatory effects on Streptococcal pilus gene transcription and virulence. *Infection and Immunity* 76: 2490–2497.
65. Kwinn LA, Khosravi A, Aziz RK, Timmer AM, Doran KS, et al. (2007) Genetic characterization and virulence role of the RALP3/LSA locus upstream of the streptolysin S operon in invasive MIT1 Group A Streptococcus. *Journal of Bacteriology* 189: 1322–1329.
66. Podbielski A, Woischnik M, Leonard BA, Schmidt KH (1999) Characterization of *nra*, a global negative regulator gene in group A streptococci. *Molecular Microbiology* 31: 1051–1064.
67. Beckert S, Kreikemeyer B, Podbielski A (2001) Group A streptococcal *rofA* gene is involved in the control of several virulence genes and eukaryotic cell attachment and internalization. *Infection and Immunity* 69: 534–537.
68. Spanier JG, Jones SJ, Cleary P (1984) Small DNA deletions creating avirulence in *Streptococcus pyogenes*. *Science* 225: 935–938.
69. Caparon MG, Scott JR (1987) Identification of a gene that regulates expression of M protein, the major virulence determinant of group A streptococci. *Proceedings of the National Academy of Sciences* 84: 8677–8681.

70. Scott JR, Cleary P, Caparon MG, Kehoe M, Heden L, et al. (1995) New name for the positive regulator of the M protein of group A streptococcus. *Molecular Microbiology* 17: 799.
71. McLandsborough LA, Cleary PP (1995) Insertional inactivation of *virR* in *Streptococcus pyogenes* M49 demonstrates that VirR functions as a positive regulator of ScpA, FcRA, OF, and M protein. *FEMS Microbiology Letters* 128: 45–51.
72. Ribardo DA, McIver KS (2006) Defining the Mga regulon: comparative transcriptome analysis reveals both direct and indirect regulation by Mga in the group A streptococcus. *Molecular Microbiology* 62: 491–508.
73. Hondorp ER, Hou SC, Hempstead AD, Hause LL, Beckett DM, et al. (2012) Characterization of the group A streptococcus Mga virulence regulator reveals a role for the C-terminal region in oligomerization and transcriptional activation. *Molecular Microbiology* 83: 953–967.
74. Haanes EJ, Heath DG, Cleary PP (1992) Architecture of the vir regulons of group A streptococci parallels opacity factor phenotype and M protein class. *Journal of Bacteriology* 174: 4967–4976.
75. Vahling CM, McIver KS (2005) Identification of residues responsible for the defective virulence gene regulator Mga produced by a natural mutant of *Streptococcus pyogenes*. *Journal of Bacteriology* 187: 5955–5966.
76. McIver KS, Heath AS, Green BD, Scott JR (1995) Specific binding of the activator Mga to promoter sequences of the *emm* and *scpA* genes in the group A streptococcus. *Journal of Bacteriology* 177: 6619–6624.

77. McIver KS, Thurman AS, Scott JR (1999) Regulation of *mga* transcription in the group A streptococcus: specific binding of Mga within its own promoter and evidence for a negative regulator. *Journal of Bacteriology* 181: 5373–5383.
78. Almengor AC, McIver KS (2004) Transcriptional activation of *sclA* by Mga requires a distal binding site in *Streptococcus pyogenes*. *Journal of Bacteriology* 186: 7847–7857.
79. McIver KS, Myles RL (2002) Two DNA-binding domains of Mga are required for virulence gene activation in the group A streptococcus. *Molecular Microbiology* 43: 1591–1602.
80. Andersson G, McIver K, Heden LO, Scott JR (1996) Complementation of divergent *mga* genes in group A streptococcus. *Gene* 175: 77–81.
81. Pabo CO, Sauer RT (1992) Transcription factors: structural families and principles of DNA recognition. *Annual Reviews in Biochemistry* 61: 1053–1095.
82. Wintjens R, Rooman M (1996) Structural classification of HTH DNA-binding domains and protein-DNA interaction modes. *Journal of Molecular Biology* 262: 294–313.
83. Huffman JL, Brennan RG (2002) Prokaryotic transcription regulators: more than just the helix-turn-helix motif. *Current Opinion in Structural Biology* 12: 98–106.
84. Ellenberger T (1994) Getting a grip on DNA recognition: structures of the basic region leucine zipper, and the basic region helix-loop-helix DNA-binding domains. *Current Opinion in Structural Biology* 4: 12–21.
85. Wolfe S, Nekludova L, Pabo C (2000) Dna recognition by Cys2His2Zinc finger proteins. *Annual review of Biophysics and Biomolecular Structure* 3: 183–212.

86. Darnell J (2011) RNA: Life's Indispensable Molecule. 1st ed. Cold Spring Harbor, N. Y.: Cold Spring Harbor Laboratory. p.
87. Burgess RR (1969) Separation and characterization of the subunits of ribonucleic acid polymerase. *The Journal of Biological Chemistry* 244: 6168–6176.
88. Haldenwang WG (1995) The sigma factors of *Bacillus subtilis*. *Microbiological Reviews* 59: 1–30.
89. Helmann JD (2009) RNA polymerase: a nexus of gene regulation. *Methods* 47: 1–5.
90. Lonetto M, Gribskov M, Gross C (1992) The sigma 70 family: sequence conservation and evolutionary relationships. *Journal of Bacteriology* 174: 3943–3849.
91. Weaver RF (2008) *Molecular Biology*. 4th ed. Boston: McGraw Hill Higher Education. p.
92. Burgess RR, Anthony L (2001) How sigma docks to RNA polymerase and what sigma does. *Current Opinion in Microbiology* 4: 126–131.
93. Ghosh T, Bose D, Zhang X (2010) Mechanisms for activating bacterial RNA polymerase. *FEMS Microbiology Reviews* 34: 611–627.
94. Dove S, Joung J, Hoch J (1997) Activation of prokaryotic transcription through arbitrary protein-protein contacts. *Nature* 386: 627–630.
95. Browning DF, Busby SJ (2004) The regulation of bacterial transcription initiation. *Nature Reviews Microbiology* 2: 57–65.
96. Ishihama A (1993) Protein-protein communication within the transcription apparatus. *Journal of Bacteriology* 175: 2483–2489.

97. Dove S, Darst S, Hochschild A (2003) Region 4 of σ as a target for transcription regulation. *Molecular Microbiology* 48: 863–874.
98. Opdyke JA, Scott JR, Moran CP (2001) A secondary RNA polymerase sigma factor from *Streptococcus pyogenes*. *Molecular Microbiology* 42: 495–502.
99. Severinov K (2000) RNA polymerase structure-function: insights into points of transcriptional regulation. *Current Opinion in Microbiology* 3: 118–125.
100. Chiu ML, Viollier PH, Katoh T, Ramsden JJ, Thompson CJ (2001) Ligand-induced changes in the *Streptomyces lividans* TipAL protein imply an alternative mechanism of transcriptional activation for MerR-like proteins. *Biochemistry* 40: 12950–12958.
101. Outten CE, Outten FW, O’Halloran TV (1999) DNA distortion mechanism for transcriptional activation by ZntR, a Zn(II)-responsive MerR homologue in *Escherichia coli*. *The Journal of Biological Chemistry* 274: 37517–37524.
102. Busby SJW, Savery NJ (2007) Transcription activation at bacterial promoters. *Encyclopedia of Life Sciences*.
103. Deutscher J, Francke C, Postma PW (2006) How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiology and Molecular Biology Reviews* 70: 939–1031.
104. Stulke J, Arnaud M, Rapoport G, Martin-Verstraete I (1998) PRD--a protein domain involved in PTS-dependent induction and carbon catabolite repression of catabolic operons in bacteria. *Molecular Microbiology* 28: 865–874.
105. Sumby P, Porcella SF, Madrigal AG, Barbian KD, Virtaneva K, et al. (2005) Evolutionary origin and emergence of a highly successful clone of serotype M1

- group A streptococcus involved multiple horizontal gene transfer events. *Journal of Infectious Diseases* 192: 771–782.
106. Miroux B, Walker JE (1996) Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of Molecular Biology* 260: 289–298.
 107. Studier FW (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification* 41: 207–234.
 108. Maxam AM, Gilbert W (1977) A new method for sequencing DNA. *Proceedings of the National Academy of Sciences* 74: 560–564.
 109. Anbalagan S, McShan WM, Dunman PM, Chaussee MS (2011) Identification of Rgg binding sites in the *Streptococcus pyogenes* chromosome. *Journal of Bacteriology* 193: 4933–4942.
 110. Johnson ML, Correia JJ, Yphantis DA, Halvorson HR (1981) Analysis of data from the analytical ultracentrifuge by nonlinear least-squares techniques. *Biophysical Journal* 36: 575–588.
 111. Roelofs KG, Wang J, Sintim HO, Lee VT (2011) Differential radial capillary action of ligand assay for high-throughput detection of protein-metabolite interactions. *Proceedings of the National Academy of Sciences* 108: 15528–15533.
 112. Burgess RR (1996) Purification of overproduced *Escherichia coli* RNA polymerase sigma factors by solubilizing inclusion bodies and refolding from Sarkosyl. *Methods in Enzymology* 273: 145–149.

113. Aravind L, Anantharaman V, Balaji S, Babu MM, Iyer LM (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiology Reviews* 29: 231–262.
114. Almengor AC, Walters MS, McIver KS (2006) Mga is sufficient to activate transcription *in vitro* of *sof/sfbX* and other Mga-regulated virulence genes in the group A streptococcus. *Journal of Bacteriology* 188: 2038–2047.
115. Ivarie R (1987) Thymine methyls and DNA-protein interactions. *Nucleic Acids Research* 15: 9975–9983.
116. Xiong Y, Sundaralingam M (2001) Protein–Nucleic acid Interaction: major groove recognition determinants. *Encyclopedia of Life Sciences*.
117. Haran TE, Mohanty U (2009) The unique structure of A-tracts and intrinsic DNA bending. *Quarterly Reviews of Biophysics* 42: 41–81.
118. Karimova G, Ullmann A, Ladant D (2000) A bacterial two-hybrid system that exploits a cAMP signaling cascade in *Escherichia coli*. *Methods in Enzymology* 328: 59–73.
119. Vahling CA, McIver KS (2006) Functional Domains in the Multigene Regulator of the Group A Streptococcus Dallas: University of Texas Southwestern Medical Center.
120. Hause LL, McIver KS (2012) Nucleotides critical for the interaction of the *S. pyogenes* Mga virulence regulator with Mga-regulated promoter sequences. *Journal of Bacteriology* 194:

121. Vahling CM, McIver KS (2006) Domains required for transcriptional activation show conservation in the Mga family of virulence gene regulators. *Journal of Bacteriology* 188: 863–873.
122. Hammerstrom TG, Roh JH, Nikonowicz EP, Koehler TM (2011) *Bacillus anthracis* virulence regulator AtxA: oligomeric state, function and CO(2) signalling. *Molecular Microbiology* 82: 634–647.
123. Hanahan D, Meselson M (1983) Plasmid screening at high colony density. *Methods in Enzymology* 100: 333–342.
124. Lukomski S, Nakashima K, Abdi I, Cipriano VJ, Shelvin BJ, et al. (2001) Identification and characterization of a second extracellular collagen-like protein made by group A streptococcus: control of production at the level of translation. *Infection and Immunity* 69: 1729–1738.
125. Okada N, Geist RT, Caparon MG (1993) Positive transcriptional control of *mry* regulates virulence in the group A streptococcus. *Molecular Microbiology* 7: 893–903.
126. Eichenbaum Z, Federle MJ, Marra D, de Vos WM, Kuipers OP, et al. (1998) Use of the lactococcal *nisA* promoter to regulate gene expression in Gram-positive bacteria: comparison of induction level and promoter strength. *Applied and Environmental Microbiology* 64: 2763–2769.
127. McIver KS, Scott JR (1997) Role of Mga in growth phase regulation of virulence genes of the group A streptococcus. *Journal of Bacteriology* 179: 5178–5187.

128. Husmann LK, Scott JR, Lindahl G, Stenberg L (1995) Expression of the Arp protein, a member of the M protein family, is not sufficient to inhibit phagocytosis of *Streptococcus pyogenes*. *Infection and Immunity* 63: 345–348.
129. Lee VT, Matewish JM, Kessler JL, Hyodo M, Hayakawa Y, et al. (2007) A cyclic-di-GMP receptor required for bacterial exopolysaccharide production. *Molecular Microbiology* 65: 1474–1484.