Electronic Thesis and Dissertation Repository

11-24-2015 12:00 AM

Recombinase-based in vivo expression technology identifies a Streptococcus pyogenes bacteriocin important for niche adaptation in the nasopharynx

Brent D. Armstrong
The University of Western Ontario

Supervisor
Dr. John McCormick
The University of Western Ontario

Recommended Citation

RECOMBINASE-BASED *IN VIVO* EXPRESSION TECHNOLOGY IDENTIFIES A *STREPTOCOCCUS PYOGENES* BACTERIOCIN IMPORTANT FOR NICHE ADAPTATION IN THE NASOPHARYNX

(Thesis format: Integrated Article)

By

Brent Daniel <u>Armstrong</u>

Graduate Program in Microbiology and Immunology

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies
Western University
London, Ontario, Canada

ABSTRACT

Streptococcus pyogenes is a Gram-positive, human-specific bacterial pathogen with the ability to cause a wide range of diseases from strep throat to necrotizing fasciitis and toxic shock syndrome. In addition, *S. pyogenes* may also induce post-streptococcal sequelae including rheumatic fever, acute glomerulonephritis, and reactive arthritis. Although primarily recognized as a pathogen, *S. pyogenes* also colonizes the skin and throat often without causing disease, and while numerous surface adhesions are important to attach to these surfaces, additional factors important for colonization and persistence by *S. pyogenes* are poorly understood. In addition to host defence mechanisms, the upper respiratory tract also contains other endogenous microorganisms that compete for the same niche. In order to gain a more complete understanding as to how *S. pyogenes* is able to adapt within the nasopharyngeal environment, a recombinase-based *in vivo* expression technology (RIVET) system was developed to identify genes activated in the nasopharyngeal niche using a humanized murine model. Using RIVET, 82 unique clones were recovered revealing 22 'cryptic', 9 'typical', and 30 'antisense' *in vivo* induced promoters with single inserts. One 'typical' promoter (P_{blp}) that controlled a putative class IIb bacteriocin operon was further characterized. Testing this promoter as a single clone in the RIVET system confirmed its activation *in vivo*. However, *in vitro* testing failed to activate this bacteriocin, despite continued addition of the characterized pheromone, SilCR. A *blp* operon deletion was out-competed by the wild-type *S. pyogenes in vivo*, but not *in vitro*. Activation of the bacteriocin *in vivo* provides a further explanation for the ability of *S. pyogenes* to occupy the nasopharyngeal niche and begin the process of colonization. This work provides a new molecular tool for the *in vivo* analysis of *S. pyogenes*, and demonstrates a novel bacteriocin system important for niche adaptation by *S. pyogenes*.

KEYWORDS

Streptococcus pyogenes, RIVET, nasopharynx, bacteriocin, colonization

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ACRONYMS

°C	degrees Celsius
×	times
× g	times gravity
β-NAD	beta-nicotinamide adenine dinucleotide
Δ*blp*	*S. pyogenes* with *blp* operon deleted
Δ*upp*	*S. pyogenes* with *upp* deleted
μg	micrograms
μg mL^{-1}	microgram per milliliter
μL	microliter
μM	micromolar
ΦSpeC	bacteriophage streptococcal pyrogenic exotoxin C
5FU	5-fluorouracil
ADP-ribose	adenosine diphosphate ribose
AEBSF	4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride
APSGN	acute post-streptococcal glomerulonephritis
ARF	acute rheumatic fever
ASL	airway surface liquid
ATP	adenosine triphosphate
BHI	brain-heart infusion media
BLASTn	Basic Local Alignment Tool nucleotide
BLASTp	Basic Local Alignment Search Tool protein
bp	base pair
C3	complement protein 3
CcpA	catabolite control protein A
CBS	cystathionine β-synthase
CD(n)	cluster of differentiation (n)
cDNA	complementary deoxyribonucleic acid
CFU	colony forming units
cNT	complete nasal turbinate
CRISPR	clustered regularly interspaced short palindromic repeats

DC	dendritic cell
DFI	differential fluorescence induction
DNA	deoxyribonucleic acid
DNAses	deoxyribonucleases
DR	direct repeat
E-64	N-[N-(L-3- transcarboxyirane-2-carbonyl)-L-Leucyl]-agmatine
erm	erythromycin
FACS	fluorescence-activated cell sorting
FbaB	fibronectin-binding protein of group A streptococci type B
FCT	fibronectin-binding, collagen-binding T antigen
Fn	fibronectin
GAG	glycosaminoglycans
GCV	ganciclovir
gDNA	genomic deoxyribonucleic acid
GFP	green fluorescent protein
GlcNAc	N-acetyl-glucosamine
GTPase	guanosine triphosphate hydrolase
h	hour
HLA	human leukocyte antigen
HPK	histidine protein kinase
HSV	herpes simplex virus
HSV1-*tk*	herpes simplex virus *thymidine kinase*
IFNγ	interferon gamma
IgA	immunoglobulin A
IgG	immunoglobulin G
IL	interleukin
IP	induction peptide
ITP	invasive transcriptome profile
IVET	*in vivo* expression technology
IVIAT	*in vivo* induced antigen technology
IVIG	intravenous immunoglobulin

kb	kilobase pair
kDa	kilodaltons
LAB	lactic acid bacteria
Lcn G	lactococcin G
LTA	lipoteichoic acid
MalR	maltose repressor
Mbp	megabase pair
mg	milligram
mg mL^{-1}	milligram per milliliter
mga	multiple gene regulator of group A *Streptococcus*
MHC	major histocompatibility complex
min	minutes
mL	milliliters
mm	millimeters
mM	millimolar
mRNA	messenger ribonucleic acid
ms	millisecond
MSCRAMM	microbial surface components recognizing adhesive matrix molecules
NAD$^+$	Nicotinamide adenine dinucleotide (oxidized)
NADH	Nicotinamide adenine dinucleotide (reduced)
NAPlr	nephritis-associated plasmin receptor
ng mL^{-1}	nanogram per milliliter
nm	nanometer
OD	optical density
ORF	open reading frame
PAM	plasminogen-binding group A streptococcal M protein
PAMPs	pathogen-associated molecular patterns
PCR	polymerase chain reaction
PFBP	*S. pyogenes* fibronectin-binding protein
P$_{gyrA}$	*gyrase A* promoter
PIN	PilT N-terminus

Plg	plasminogen
PMN	polymorphonuclear leukocyte
pre-crRNA	pre-clustered, regularly interspaced short palindromic repeats ribonucleic acid
PRR	pattern recognition receptors
PrtF1	fibronectin-binding protein F1
PrtF2	fibronectin-binding protein F2
PTP	pharyngeal transcriptome profile
qRT-PCR	quantitative real time polymerase chain reaction
RD2	repeat domain 2
RHD	rheumatic heart disease
RIVET	recombinase-based *in vivo* expression technology
RNA	ribonucleic acid
RNA-seq	ribonucleic acid sequencing
rpm	revolutions per minute
RR	response regulator
SA-FF22	streptococcin A-FF22
SA-M57	streptococcin A-M57
SAg	superantigen
sec	seconds
SfbI	streptococcal fibronectin-binding protein I
SIC	streptococcal inhibitor of complement
sil	streptococcal invasion locus
ska/Ska	streptokinase gene/protein
SLO	streptolysin O
SLS	streptolysin S
SNP	single nucleotide polymorphism
SOF	serum opacity factor
speA	*streptococcal pyrogenic exotoxin A*
SpeB	streptococcal pyrogenic exotoxin B
speC	*streptococcal pyrogenic exotoxin C*

SPN	*S. pyogenes* nicotinamide adenine dinucleotide (NAD) glycohydrolase
sRNA	small ribonucleic acid
STM	signature-tagged mutagenesis
STSS	streptococcal toxic shock syndrome
TCR	T cell receptor
Tet	tetracycline
Th1	T-helper 1
THY	Todd Hewitt Yeast media
tk	thymidine kinase
TLR	Toll-like receptor
T_M	melting temperature
TNFα	tumour necrosis factor alpha
U	units
UFBD	upstream fibronectin binding domain
UMP	uracil monophosphate
UppP	undecaprenyl pyrophosphate phosphatase
UPRTase	uracil phosphoribosyl-transferase
UTR	untranslated region
V	volts
v/v	volume per volume
VCAM-1	vascular cell adhesion molecule-1
w/v	weight per volume

CHAPTER 1: INTRODUCTION

1.1 *Streptococcus pyogenes*

Streptococcus pyogenes (also commonly referred to as Group A *Streptococcus*) is a Gram-positive, human-specific bacterial pathogen. *S. pyogenes* is an aerotolerant anaerobic coccus that forms long chains of cells when actively dividing. A characteristic of *S. pyogenes* is the formation of zones of β-haemolysis around colonies when grown on blood agar plates (1). The genome of *S. pyogenes* is a single circular chromosome of ~1.9 Mbp with a low G+C content of ~39% and is well known for the abundance of exogenous genetic elements including prophage and prophage-like elements, and integrated conjugative elements. These elements occupy ~10% of the chromosome and provide the greatest amount of genetic difference between strains (2). Virulence factors are commonly found on these elements including the well-known superantigens (SAgs). Using the Lancefield classification of serologic typing that is based on surface carbohydrate production, *S. pyogenes* is the only member of the Group A streptococci. Serotyping is based on the surface expressed M protein, of which more than 100 serotypes exist (1, 3). More recently, this serotyping system has been replaced with the nucleotide sequence of the 5′ end of the *emm* gene, producing over 223 serotypes to date (4).

1.2 Global burden of disease caused by *S. pyogenes*

S. pyogenes has the ability to cause a wide spectrum of diseases and contributes to an immense burden of human illness on a global scale (5). These diseases can range from mild, non-invasive pharyngitis (strep throat) and impetigo, to far more severe invasive diseases such as necrotizing fasciitis and toxic shock syndrome. Further post-infection complications can also result from non-invasive diseases in the development of acute post-streptococcal glomerulonephritis (APSGN) and acute rheumatic fever (ARF), which may lead to rheumatic heart disease (RHD) and reactive arthritis (1).

Mild diseases by *S. pyogenes* cause the most number of cases, with pharyngitis contributing to more than 616 million cases per year and impetigo estimated to be at an incidence of 111 million people. Furthermore, approximately 1.78 million new severe

cases per year have been estimated at 663 000 invasive infections, 472 000 cases of APSGN, and 282 000 of RHD. The remaining 366 000 include stroke, endocarditis, or requiring prophylaxis due to RHD. Combined diseases lead to over 517 000 deaths per year world-wide. When broken down, APSGN causes the fewest number of deaths at 5000 per year, followed by invasive disease at 163 000. It is RHD that causes the most deaths at 233 000 per year, with the remaining 116 000 due to complications arising from diseases caused by RHD. There are an estimated 15.6 million current cases of RHD and this continues to be the greatest cause of death caused by *S. pyogenes* in the world (5).

Antibiotics have been routinely used to treat *S. pyogenes* infections for decades and, to date, there have been no documented strains of *S. pyogenes* that have developed resistance to β-lactam antibiotics (6). Unfortunately, resistance to second line antibiotics, such as macrolides, tetracyclines, and fluoroquinolones, seems to be on the rise, creating difficulties treating patients with allergies to penicillin and other β-lactams (6, 7). To combat this, development of an effective vaccine is ongoing, and as humans remain the only reservoir for *S. pyogenes*, it creates the possibility of severely reducing *S. pyogenes* disease (8).

1.3 Dissemination of *S. pyogenes* throughout the globe

Disease caused by *S. pyogenes* is not evenly distributed around the world. In developing countries, due to the lack of sanitary conditions and access to appropriate treatment, the rate of invasive disease and RHD is much higher than in developed countries (5). The dissemination of different strains also likely contributes to this imbalance. Different *emm* types are more prevalent in contrasting geographical areas, and specific *emm* serotypes are often more commonly associated with different types of disease (9). Accumulating over 100 data sets (38 081 *S. pyogenes* isolates), Steer *et al.* demonstrated the varying *emm* types occurring in Africa, Asia, Latin America, Middle East, Pacific Island countries/Indigenous Australians, and Established Market Economy countries (such as Canada, US, most of Europe, Japan, Australia, New Zealand) (9).

The top three *emm* types for each region divided by pharyngeal, skin, invasive, and combined diseases are shown in **Table 1.1**. These data, where available, demonstrate the

Table 1.1: Top *emm* types of the world divided by region and disease

Region	Disease and corresponding *emm* types												50% of isolates
	Pharyngeal			Skin			Invasive			Combined			
Africa	emm75	emm12	emm3	emm80	emm100	st62	-	-	-	emm12	emm75	st62	18
Asia	emm44	emm12	emm75	emm1	emm2	emm44	emm1	emm12	emm8	emm1	emm12	emm4	3
Latin America	emm1	emm12	emm3	emm83	emm53	emm33	emm1	emm12	emm75	emm1	emm12	emm2	6
Middle East	non-typable	emm1	emm12	-	-	-	emm3	emm28	non-typeable	non-typeable	emm12	emm1	5
Pacific Island/ Indigenous Australians	emm55	emm11	emm1	non-typeable	emm55	emm70	emm18	emm80	emm114	emm55	non-typeable	emm11	19
Established Market Economy Countries	emm12	emm1	emm4	emm53	emm81	emm28	emm1	emm28	emm3	emm1	emm12	emm28	5

Data from Steer *et al.*, 2009 (9)

diverse *emm* types between regions and disease, as well as the common strains that are important worldwide. Overall diversity of *S. pyogenes* can be demonstrated by the number of *emm* types that account for the first 50% of the total population. From the data accumulated by Steer *et al*., Africa and the Pacific Island counties/Indigenous Australians showed the greatest variability with the top 50% represented by 18 and 19 *emm* types isolated, respectively. On the other hand, Asia, Latin America, Middle East, and Established Economy Market Countries have a similar but much lower diversity with the top 50% representing only 3, 6, 5, and 5 *emm* types, respectively **(Table 1.1)** (9). The smaller proportions and therefore greater number of isolates represented, gives Africa and the Pacific Island counties/Indigenous Australians a far greater diversity compared to the other regions.

When looking at a global scale, *emm1* and *emm12* play an important role in causing disease world-wide. However, outbreaks of disease caused by *emm* types that are not typical to the area can also occur. For example, isolates from patients during an acute APSGN outbreak in the south province of Guizhou, China demonstrated that *emm60* and *emm63* predominated at 43.5% and 30.4%, respectively, with four other *emm* types present, all under 9% (10). ARF outbreaks have also occurred in Salt Lake City, Utah, United States in 1985 and 1998. In 1985, *emm18* dominated while in 1998 the most prominent were *emm4* and *emm12* (11). Thus, individual *emm* serotypes may dominate across large geographical regions, yet specific disease states and outbreaks can be linked to different serotypes.

1.4 Clinical diseases

1.4.1 Pharyngitis

Pharyngitis is an inflammatory response due to infection of the pharynx. *S. pyogenes* is the most common cause of bacterial pharyngitis and primarily affects school-age children ranging from 5-15 years of age (12, 13). Inflammation specifically involves the pharynx and the tonsils, often with the presence of pus, and includes symptoms such as fever, headache, nausea, and other flu-like symptoms. While normally associated with school-

age children, it can also be linked to all ages in crowded areas such as military facilities (1, 14). In addition to this, episodes become recurrent in a small percentage of people (15). Several theories exist to explain this phenomenon including antibiotic failure due to poor patience compliance, antibiotic degradation from nearby β-lactamase producing bacteria, tolerance, invasion of cells, and reinfection by *S. pyogenes* from repeated exposure or as a consequence from the eradication of competing bacteria (7, 16, 17). Tonsillectomy has shown to decrease infections caused by *S. pyogenes* with those suffering from recurrent pharyngitis (18, 19). Although pharyngitis caused by *S. pyogenes* is usually acute, uncomplicated, and self-limiting, if untreated it may lead to other consequences as described below (20).

1.4.2 Impetigo

Impetigo, or pyoderma, is a localized purulent infection of the skin that can occur in the bullous and nonbullous forms. While both forms can be caused by *Staphylococcus aureus*, only the more common nonbullous form can be caused by *S. pyogenes*. Nonbullous impetigo is most prevalent in warm, humid climates among children below school-age in economically disadvantaged families (21). Starting as a single red macule, nonbullous impetigo can rapidly become vesicles that rupture and dry, forming honey-coloured crusts. Although often spread by autoinoculation, it is self-limiting after several weeks if left untreated (22). While both *S. aureus* and *S. pyogenes* are sometimes found within the same skin lesions, typically *S. aureus* dominates the moderate climate cases and *S. pyogenes* predominates humid and warmer climates (23, 24). Interestingly, *S. pyogenes* strains that cause impetigo are not known to cause pharyngitis or ARF, but can be nephritic (these last two discussed below) (21). This implies a varying disease tropism for different strains of *S. pyogenes*, which is discussed in more detail in subsequent sections.

1.4.3 Scarlet fever

Typically occurring in children between the ages 4 and 7, scarlet fever is associated with pharyngitis, fever, a bright tongue with a strawberry appearance, and a sandpaper-like

rash starting on the face and neck and spreading to the trunk and other parts of the body (25). Scarlet fever used to be a major cause of child morbidity and mortality in the 19th and early 20th centuries. However, due to improved living conditions in industrialized nations and the discovery of antibiotics, incidents decreased and it became more of a mild disease (26, 27).

1.4.4 Necrotizing fasciitis

Necrotizing fasciitis, previously called *Streptococcus* gangrene, is a rare deep bacterial infection of the subcutaneous tissue, resulting in the destruction of the surrounding fascia and fat (28). This deep penetration can occur following a break in the epithelial layer caused by some form of trauma such as cuts, burns or even surgical procedures, but many incidences occur in the absence any known portal of entry (29, 30). Within the first day, flu-like symptoms appear as well as pain that is disproportionate to the injury. At a more advanced stage, swelling and dark patches will appear, eventually filling with fluid. The area will ultimately become necrotic, with separation of the dead tissue. It is also possible for patients to development streptococcal toxic shock syndrome, described below (29). Treatment of necrotizing fasciitis involves immediate surgical debridement of the affected tissues as well as antibiotic treatment and supportive therapy. Surgical debridement is performed until all dead tissue is removed, potentially leading to amputation (30).

1.4.5 Streptococcal toxic shock syndrome

Streptococcal toxic shock syndrome (STSS) is caused by a hyperactive immune system triggered by potent immunostimulatory toxins called superantigens (SAgs), leading to a massive cytokine response. In contrast to toxic shock syndrome caused by *S. aureus* where the infecting organism often remains localized, STSS typically occurs during the context of an invasive infection by *S. pyogenes* (31). Early symptoms may be flu-like and include fever, sore throat, swollen glands, rash, diarrhea, and vomiting. After 3-4 days, the fever remains persistent and hypotension and shock develop, which can lead to tachycardia and multi-organ failure. Coagulation can also occur leading to gangrene or also the development of necrotizing fasciitis from the invasive infection (32). Treatment

includes the use of antibiotics and fluids to maintain blood pressure, and may include additional developmental therapies such as the administration of intravenous immunoglobulin (IVIG) and activated protein C. IVIG would potentially contain anti-SAg antibodies, effectively neutralizing the SAgs, while activated protein C can help prevent coagulation from occurring (33-35).

1.4.6 Acute post-streptococcal glomerulonephritis

APSGN is a renal disorder first reported by von Plenciz over 200 years ago, and was a feared consequence following scarlet fever (36). Caused only by certain 'nephritogenic' *S. pyogenes* strains, APSGN typically affects children between the age of 2 and 12 following pharyngitis or a skin infection (37). Depending on the original infection, the latent period for APSGN is 1 to 4 weeks from a skin infection and 1 to 2 weeks for pharyngitis (1). Symptoms include edema, hypertension, and dark coloured urine from blood due to the impaired filtering properties of the glomerulus (1, 38). APSGN patients show spontaneous recovery over 95% of the time and generally only supportive therapy maintaining blood pressure and fluids is required. Very rarely are any long term effects seen after recovery in children; however, adults seem more likely to develop chronic symptoms afterward (1, 39).

1.4.7 Acute rheumatic fever and rheumatic heart disease

ARF is an autoimmune disease caused by the cross reaction of antibodies directed against *S. pyogenes* with human tissues (40). ARF most commonly involves inflammation of the joints, central nervous system, or the heart, typically arising 2 to 5 weeks after non-treated pharyngitis. Along with one of the major areas of swelling, additional symptoms include fever, malaise, tachycardia, and anemia (6, 41). Inflammation of the joints (arthritis) is the most common symptom and occurs in 75% of ARF patients. Affecting the large joints, painful episodes can last from a number of days to weeks, and treatment includes pain medication and anti-inflammatory drugs (41, 42). Occurring in approximately 10% of patients is a neurological manifestation of ARF known as Sydenham's chorea. The characteristic of the disease is the presence of involuntary, purposeless, and jerky movements of the face and extremities that disappear while sleeping, typically lasting 8 to

15 weeks (41, 42). Lastly, while not the most common, inflammation of the heart tissue is the most serious as permanent, long-term damage to the heart can occur. This damage is given the term rheumatic heart disease (RDH) and occurs most frequently to the mitral valve (41, 42). Healing of the valves results in scarring and fibrosis preventing proper function. Damaged valves show narrower openings and allow blood to leak back, which can lead to congestive heart failure. Treatment of RHD involves the use of anti-inflammatory and diuretic drugs with valve replacement in severe cases (42, 43).

1.5 Molecular mechanisms of colonization and disease

1.5.1 Barriers of the human host

The major recognized sites of colonization by *S. pyogenes* on the human host include the skin and the nasopharynx (44, 45). In order to attach and colonize, group A streptococci have to bypass or subvert a number of mechanical, chemical, and biological barriers. Healthy skin creates an intact barrier that microorganisms cannot penetrate, as well as the process of desquamation that must be overcome (46, 47). Along with this, skin has a low moisture content, low pH, and produces antimicrobial compounds such as fatty acids, lysozymes, and defensins (48, 49). Finally, *S. pyogenes* must compete with other bacteria for the same space, with *Staphylococcus epidermidis* a prime example, as it creates peptides toxic to *S. pyogenes* (47).

The oral cavity maintains the same physical barriers as the skin, and bacterial competition, but also contains saliva (50). Saliva has many different functions involved with eating, maintaining healthy teeth, and also has many antimicrobial properties (51). Saliva contains many components, including those of the innate and acquired immune system, that interact with the oral microbiota, controlling the composition of the microorganisms within the oral cavity (51). These include immunoglobulins, defensins, lysozyme, as well as cysteine and serine proteinase inhibitors (51, 52). For example, immunoglobulin A (IgA) from human saliva showed both interference with binding and opsonisation of *S. pyogenes in vitro* (53-55). Agglutinins have shown to cause co-aggregation with some strains, and subsequently these strains would be swallowed (56, 57). Additionally, the commensal bacteria *S. salivarius* produces a bacteriocin that is

known to be toxic to *S. pyogenes* (58).

The innate immune system also provides a significant barrier that *S. pyogenes* must overcome in order to colonize the host. Innate immunity utilizes a pattern recognition strategy based on a set of conserved molecular patterns, referred to as pathogen-associated molecular patterns (PAMPs) (59). Unique to microbes, PAMPs are recognized by pattern recognition receptors (PRRs), such as Toll-like receptors (TLR), on immune cells (60). These targets are generally indispensable to the organism and therefore rarely become modified due to mutation (61). With varying numbers between species, TLRs can detect multiple PAMPs with TLRs 1, 2, 4, 5, and 6 recognizing bacterial products exclusively (62). After recognition, PRRs signal to the host via signalling cascades, which ultimately results in the activation of gene expression and the production of a wide range of molecules such as cytokines, chemokines, cell adhesion molecules, and immunoreceptors (60).

TLRs are present on numerous innate cell types such as macrophages, neutrophils, and dendritic cells (DC). Distributed throughout the body in morphologically diverse forms, macrophages are capable of engulfing invading microbes, causing the release of cytokines. This in turn diverts blood flow to the affected tissue in order to attract cells capable of killing microbes including more macrophages as well as neutrophils and DCs (60-62). DCs, like macrophages, are also found in most parts of the body and kill microbes via phagocytosis. Coming into contact with pathogens causes these cells to activate, starting the production of cytokines (63). One important feature of both DCs and macrophages is the ability to present foreign antigen to naïve T cells, creating the bridge between the innate to the adaptive immune systems (61, 64). Neutrophils likewise combat invading microbes using phagocytosis, but also possess granules made up of proteolytic enzymes and antimicrobial proteins that combine with phagosomes in a process called degranulation (65, 66). Activated neutrophils also create neutrophil extracellular traps, generated by the release of DNA and granule components, to trap and destroy invading microbes (66).

The last major component of the innate immune system discussed here is the complement system, which can recognize invading pathogens through the classical, alternate, or lectin

pathway (67). Activated in different manners, the complement pathways begin a highly regulated cascade of enzymatic reactions that generate effector compounds able to intercept invading pathogens and induce a pro-inflammatory response (68, 69). Classical and lectin pathways are activated via bound antibodies and microbial carbohydrates, respectively. The alternate pathway, on the other hand, is autoactivated, with the hydrolyzation that begins the cascade always occurring at a slow rate (68, 70). These effector compounds bind to self and non-self carbohydrates, lipids, and proteins indiscriminately. Self-cells block further activation through the presence of specific regulatory proteins. Pathogens, lacking these proteins, further activate the complement system (68, 70). This leads to the production of C3b, an opsonin that coats the cell, inducing phagocytosis, and the assembly of the membrane attack complex, which causes pore formation and lysis (68).

1.5.2 Mechanisms of adhesion

Once the initial barriers have been crossed, *S. pyogenes* must adhere to host cells to begin colonization. To accomplish this, *S. pyogenes* utilize surface components called adhesins, which bind in a stereochemical manner to complementary molecules on the tissue surface called ligands or receptors (71). The process of adhesion and colonization has been studied for decades, but it still has not been completely characterized at the molecular level due to its complexity. *S. pyogenes* uses a multi-adhesin approach, which complicates research since removing a single adhesin may have little to no effect (50). While many different adhesins have been discovered, not all are present on each strain, and different adhesins have been found to be important depending on the tissue. Some of the best-characterized adhesins are described below.

1.5.2.1 Fibronectin binding proteins

Fibronectin (Fn) was first shown as a ligand by Simpson *et al.* with the adherence of *S. pyogenes* to polymorphonuclear leukocytes and oral epithelial cells (72, 73). Fn is a large glycoprotein dimer of two ~250 kDa fragments, each covalently linked near the C-terminus. Each monomer is made of 12 type I repeats, two type II repeats, and 15 to 17 type III repeats (74). Found as two forms, the soluble form occurs in various body fluids

such as plasma and the insoluble form occurs on cell surfaces and interstitial connective tissues (50, 75). Numerous adhesins have been found to bind Fn and have been grouped together as the Fn binding proteins.

First, the Fn binding protein F1 (PrtF1, also called SfbI) was found separately by Talay *et al*. and Hanski *et al.* (76, 77). Not present in all strains of *S. pyogenes,* complementation of PrtF1/SfbI to strains without it, and unable to adhere to Fn, were then able to bind Fn (76-78). Interestingly, Hanski *et al.* also showed that while adherence to epithelial cells was affected by the loss of PrtF1/SfbI, the loss of M protein had no effect (77). PrtF1/SfbI binds to Fn from two distinct regions; the upstream Fn-binding domain (UFBD) immediately followed by a region of 2 to 6 repeats of repeat domain 2 (RD2) (79-81). UFBD attaches to the N-terminal region of Fn composing of both the fibrin and collagen binding domains (70 kDa), while just the fibrin domain (30 kDa) binds most efficiently to RD2 and does not interact with the collagen domain (45 kDa) (79, 81).

Serum opacity factor (SOF, also called SfbII) was discovered based on its ability to turn human serum opaque. It was not until much later that it was found that SOF had Fn binding domain repeats located at the C-terminus. SOF was found on ~50% of *S. pyogenes* isolates, with 30% to 60% homology between serotypes (82-84). Unlike PrtF1/SbfI, SOF binds to the 30 kDa portion of Fn, utilizing the C-terminus repeats, with adherence maintained by as little as one repeat (85, 86). Knockout studies of the SOF gene showed that strains retained ~10% of the wild-type ability to bind Fn, and pre-treatment with anti-SOF antibodies reduced adherence to human epithelial HEp-2 cells by 50% (86, 87). SOF has also been shown to bind fibulin-1, a secreted glycoprotein associated with the extracellular matrix, through a domain different then that of Fn. It is hypothesized that SOF binds to Fn, fibulin-1, and gelatin forming a quaternary complex, but further research is required for confirmation (88).

Protein F2 (PrtF2) is a Fn-binding protein comprised of two distinct variants; *S. pyogenes* Fn-binding protein (PFBP) and Fn-binding protein of group A streptococci type B (FbaB). PFBP consists of a UFBD followed by three different repeat domains. Whole Fn, as well as the 70 kDa and 30 kDa N-terminal fragments, can bind to the PFBD domain (89-92). FbaB is shorter than PFBP due to a shortened central region, causing the

first half of the UFBD to be missing, followed by two to four different repeat domains (89, 92, 93). It remains unknown if the shorter UFBD domain affects Fn binding, however, a knockout of both variants showed a significant decrease in the ability to adhere to epithelial cells (91, 93). In addition, using an intraperitoneal mouse model, it was shown that FbaB is an important virulence factor, as a FbaB knockout showed a 50% reduction in lethality compared to the wild-type strain (93).

1.5.2.2 M protein

The best known surface structure of *S. pyogenes* is the M protein, encoded by the *emm* gene (94). The M protein is an α-helical coiled-coil, extending by ~50 nm from the surface of the cell, and ending with a short non-helical N-terminus (95, 96). Combined of many different sections, the common M protein framework is comprised of a signal sequence, a hypervariable terminus (A repeats), a less variable central domain (B repeats), and a highly conserved C-terminus (C repeats) (97-100). Known to be involved in virulence, it was first shown to be important with adherence in 1972 by Ellen *et al.* when an M protein deficient *S. pyogenes* adhered to human buccal epithelial cells in fewer numbers that the wild-type (101). This decrease could be repeated when the wild-type was pre-treated with anti-M protein antibodies, and this phenotype was shown numerous times by other groups (71, 101-104). Nevertheless, finding the M protein ligand has been complicated due to the multiple adhesins present, as well as the many different ligands to which the M protein binds.

One M protein ligand was found to be the soluble and insoluble forms of Fn (100, 105-107). This was best demonstrated when *L. lactis*, unable to bind Fn or epithelial cells, was given this ability when expressing M protein (100). While shown to bind to the B repeat region of the M protein, Fn has failed to bind to some M protein types, with M1, M3, and M5 showing definite binding (100, 106, 108, 109).

Another M protein ligand is the complement regulatory protein (CD46), a membrane protein that prevents autologous complement activation (110). Specifically, the C repeat of M protein was found to bind to the short consensus repeat regions 3 and 4 of CD46 (111). Once again, variability within M proteins was demonstrated when two of three *S.*

pyogenes strains attached to human CD46 expressed via mouse cells (104). Feito *et al.* also showed CD46 binding even with *emm* deleted, indicating the presence of other adhesins (112).

The M protein has also been shown to bind cell surface glycosaminoglycans (GAGs). Composed of repeating, unbranched disaccharide units, there are several different classes based on this composition. Binding to M protein were the dermatan sulphate and heparan sulphate classes (113, 114). Binding occurred predominantly via the C repeat region towards the C-terminus end (114).

Finally, there has also been contradictory evidence indicating that sialic acid and fucose glycoproteins, contained within mucin, both do and do not act as a ligand for M protein. Further evidence is required in order to determine whether or not either of these are potential ligands (115-117).

1.5.2.3 Pili

Another surface structure shown to be important for adhesion by *S. pyogenes* has been pili. While long known to be present on Gram-negative bacteria, it has only been in the last 11 years that pili have been found on Gram-positive bacteria, in particular *S. pyogenes* (118). Genes encoding *S. pyogenes* pili are located in the highly variable fibronectin-binding, collagen-binding T antigen (FTC) region of the genome, and there are nine different types (118, 119). These types are created through variations in gene sequence and overall operon structure. Regardless of these variations, the basic pili structure of *S. pyogenes* is composed of a single backbone protein polymerized to form long strands with an ancillary protein cap (118, 120-126). Many *in vitro* experiments demonstrated that the central region of the ancillary protein cap was used by *S. pyogenes* to adhere to host cells, although the corresponding ligand has never been discovered (120, 121, 124, 126, 127). Becherelli *et al.* also recovered fewer numbers of a cap protein knockout compared to wild-type using an intraperitoneal mouse model (126). On the other hand, Crotty Alexander *et al.* observed increased lethality of a mutant containing a knockout of the backbone protein compared to wild-type in three mouse models (subcutaneous, pneumonia, and sepsis). An explanation for this outcome is that pili does

indeed provide adhesion for colonization, but when single bacteria break away to invade the blood or deeper tissue, the presence of pili provides a pattern easily identified by phagocytes (127). Therefore, further work is required to determine the precise function of pili in *S. pyogenes*.

1.5.2.4 Hyaluronic acid capsule

The capsule of *S. pyogenes* is composed of hyaluronic acid, a high molecular weight disaccharide comprised of repeating units of glucuronic acid linked to N-acetylglucosamine (GlcNAc), and is a molecular mimic to the hyaluronic acid produced by humans (128, 129). Although produced at various levels by different strains, it provides protection from phagocytosis, demonstrated by the inability of the mutant lacking the capsule to survive in blood compared to the wild-type (13, 130-133). Further *in vitro* work demonstrated that hyaluronic acid utilizes CD44 as a ligand to bind to cells. Strangely, hyaluronic acid mutants were actually able to bind better than wild-type counterparts *in vitro*, presumably because hyaluronic acid masks other potential adhesins (128, 134). While this may appear as a disadvantage, the hyaluronic acid capsule may still provide a benefit against the human host. Utilizing intraperitoneal injection, pneumonia, and throat colonization models with mice, as well as baboon throat colonization model, the hyaluronic acid capsule mutant was cleared significantly faster or failed to cause disease when compared to the wild-type strain. The theory is that the hyaluronic acid capsule does indeed help with adhesion, and also provides a significant protection against phagocytosis to allow colonization (13, 128, 130-132).

1.5.2.5 Lipoteichoic acid

Lipoteichoic acid (LTA) is a cell wall polymer composed of an unbranched 1,3-linked glycerolphosphate backbone, linked to the bacterial membrane via a lipid anchor. The hydroxyl group of C2 is also modified with D-alanyl or glycosyl groups to varying degrees (135). LTA from *S. pyogenes* was found to stick to various cells (PMNs, buccal epithelial cells, HEp-2), and pre-treatment of host cells with LTA prevented adherence by *S. pyogenes*, thus demonstrating LTA as an adhesin (71, 136-141). Furthermore, utilizing an intranasal mouse model, treatment with LTA prior to *S. pyogenes* challenge

demonstrated a drastic reduction in colonization and the survival of all mice (141). The receptor for LTA on these various cells was determined to be Fn as *S. pyogenes* could no longer bind to Fn or Fn coated beads when pre-treated with LTA (73, 106). Specifically, it has been suggested that ester-linking fatty acids of LTA recognize a fatty acid binding site on Fn (142, 143).

1.5.2.6 Adhesin expression and tissue tropism

S. pyogenes possess a large number of adhesins that creates a complex system that is also difficult to study due to the many overlapping functions. It is also thought that differential distribution and regulation of these genes would allow *S. pyogenes* to produce an adhesin profile to allow for tissue-specific tropism (44). In support of this concept, research has shown that *S. pyogenes* strains do not all adhere to the same cells *in vitro*, indicating that they are expressing different genes.

Numerous groups showed no difference in adherence to buccal cells when the M protein was removed (108, 115, 144, 145). Conversely, Courtney *et al*. found differences when using HEp-2 epithelial cells and Wadstrom *et al*. found differences with human pharyngeal epithelial cells (115, 144, 145). Okada *et al*. was able to show that loss of PrtF1 only had an effect on the adherence to Langerhans cells, and the loss of M protein only showed an effect with keratinocytes (146). Similarly, Hanski *et al*. showed the importance of PrtF1 but not M protein with hamster trachea cells (77). Fibronectin binding protein 54, on the other hand, has been shown to be necessary for adhering to human buccal cells, but unnecessary for HEp-2 cells (147). Lastly, extracellular factor protein showed no change with Detroit 562, Ca9-22, and HEp-2 cells, however, did show a significant decrease in adherence to HaCaT cells (all epithelial cells) (148).

Based on data such as this, the theory of a two-step adhesion process arose. The first step involves a hydrophobic interaction between a lipid moiety of LTA and the fibronectin sector of the host cell. While this reaction is weak and reversible, it allows the repulsion between negatively charged *S. pyogenes* and host cell membranes to be overcome. The second step is then the attachment of an adhesin creating a tissue specific interaction that is almost irreversible with its receptor (50, 149, 150). Adhesins other than M protein

potentially included in the model include PrtF1, SOF, and PrtF2. Understanding the *in vivo* regulation of these adhesins and how the adhesion process occurs could greatly increase our knowledge of *S. pyogenes* and how it is able to colonize and potentially cause disease.

1.5.3 Secreted proteins

Along with surface structures, *S. pyogenes* secretes many proteins that aid in colonization and survival. A wide range of functions include degradation of DNA from neutrophil extracellular traps by the DNAse Sda1, inhibition of the complement system by streptococcal inhibitor of complement (SIC) and C5a peptidase, and inhibition of opsonophagocytosis via Mac protein (6, 151). *S. pyogenes* also secretes multiple proteases, a major one being SpyCEP, which degrades chemokines (152).

Streptococcal pyrogenic exotoxin B (SpeB) is a cysteine protease that has garnered considerable attention due to its predominant secretion and indiscriminate specificity (153). Some of its targets include immunoglobulins, C3b, chemokines, and fibronectin of the host, as well as self proteins such as M protein, C5a peptidase, Sda1, and streptokinase (153, 154). Contradictory evidence has hindered the exact role of SpeB in virulence, nevertheless, the ability to vaccinate against *S. pyogenes* infection utilizing anti-SpeB antibodies suggests an important role (155-157).

Also secreted are the pore forming toxins streptolysin S (SLS) and streptolysin O (SLO). SLS is a 2.8 kDa peptide that inserts into the membrane of targets such as erythrocytes, leukocytes, and platelets. A carrier molecule such as albumin, LTA, or RNA is also required to function (158). Targeting the same cells is SLO, an oxygen labile, 61 kDa protein that oligomerizes binding to target membrane cholesterol (159, 160). Conflicting results have made identifying their function in virulence uncertain, but it appears that they contribute to tissue destruction and the inhibition of neutrophil recruitment (159-163).

Finally are SAgs, potent immunostimulatory toxins that bind to the MHC class II of antigen presenting cells and the TCR of T cells, in a way that activates the T cell independent of the antigenic peptide (164). This results in an enormous activation of T

cells and an immense cytokine response, secreting predominantly IL-2, IFN-γ, TNF-α, IL-1β, and IL-6, with the ability to cause widespread tissue damage and organ dysfunction (165). Their purpose has been widely speculated since they cause the activation of the immune system, and possibly the death of the host, seemingly decreasing their own chances for survival. However, their importance was recently shown in a humanized mouse model of colonization when *S. pyogenes* failed to colonize efficiently once certain SAgs were knocked out. Restoration of the SAgs re-established the colonisation phenotype demonstrating their role in colonization (166).

1.5.4 Invasion

Another characteristic of *S. pyogenes* is the ability to invade different types of eukaryotic cells (167-169). *S. pyogenes* is able to invade host cells without causing any overt damage, yet can still be shed in the saliva, and can persist within cells in the absence of clinical disease (170). Internalization occurs through the interaction of integrins on the surface of epithelial cells with receptors on *S. pyogenes* via a bridge such as Fn, causing a remodelling of the cytoskeleton structure (171). Integrins are protein heterodimers made up of transmembrane α and β subunits with various combinations. This defines adhesion specificity, allowing for the interaction with extracellular matrix ligands such as Fn, vitronectin, laminin, and collagen (172). The best characterized receptors are PrtF1/SfbI and the M protein which interact with integrin α5β1 via whole Fn (173-176). Many interactions can induce clustering which leads to cellular signalling with different signalling cascades shown to be created between M protein and PrtF1/SfbI (171, 177). The M protein causes actin polymerization which leads to a cytoskeleton rearrangement leading to surrounding and engulfment by microvilli using a "zipper-like" method (177-179). Strains expressing PrtF1/SfbI, on the other hand, cause the formation of large invaginations using caveolae-mediated endocytosis without actin polymerization (180-182). The biological significance for *S. pyogenes* to enter cells has yet to be fully elucidated (16, 169).

1.5.5 Molecular basis of tissue tropism

S. pyogenes strains can be further subdivided according to the disease they cause, and

therefore the niche in which they prefer, based on a specific genetic profile, with some exceptions known to exist (44). As sequencing of many *emm* genes became available, specific profiles first began to form based on *emm* strains and the diseases caused. Based on the combination of four different subfamily regions encoding part of the peptidoglycan domain of the *emm* protein, *S. pyogenes* strains were divided into five patterns from A to E (183-185). Patterns A, B, and C are comprised of strains that cause pharyngitis, pattern D is comprised of strains that cause impetigo, and pattern E included strains that seem to have no preference and can cause both pharyngitis and impetigo (185). Pattern E strains may be particularly important as these isolates may allow for the transfer of genetic material, including mobile genetic elements, between pharyngeal and skin strains. A world wide sampling over six continents showed an overall distribution of 21% pattern A-C, 38% pattern D, and 37% pattern E (4% unidentifiable) (4). Continued sequencing has identified numerous other genes specific to each pattern.

An important genetic component of tissue tropism includes *mga*, formerly referred to as *mry* or *virA*, located upstream of *emm*. *mga* [multiple gene regulator of group A *Streptococcus* (186)] is a positive transcriptional regulator of the *emm* family genes along with *speB*, *scpA,* and *sof*, among others. Continual sequencing showed the divergence pattern of *mga* matching the pattern of the *emm* cluster downstream. When looking for genetic links to niche specialization, Bessen *et al.* found that *mga* could be divided into two alleles based on its sequences. First, *mga-1* alleles were found in A-C strains, while *mga-2* alleles were found in D and E (183, 185, 187). Comparison of *mga* within the same allele showed minimal nucleotide diversity with differences of only 3.3% and 2.5% within *mga-1* and *mga-2*, respectively. Comparison between alleles, on the other hand, showed a nucleotide diversity of 24.5%. Comparing amino acid diversity showed similar results of 2.9% and 1.5% within *mga-1* and *mga-2*, respectively, and 20.7% between the two alleles (187).

Another gene combination important with tissue tropism is that of plasminogen (Plg)-binding group A streptococcal M protein (PAM), one of the *emm*-like genes, and streptokinase (Ska). Host Plg binds to PAM, which interacts with Ska, yielding bacterial bound plasmin (188). Ska, a broad spectrum proteinase involved with dissolving blood

clots, is made up of three domains: alpha, beta, and gamma. When comparing sequences, beta showed the most divergence, causing *ska* to be divided into two major groups, with one sub group (*ska-1*, *-2a*, and *-2b*). Pattern A-C is represented mostly by *ska-1*; however, those that did have *ska-2* were all *ska-2a*. The vast majority of pattern E strains possessed *ska-1* and all A-C and E pattern strains lacked PAM. When looking at D pattern strains, the majority possessed *ska-2b* and also possessed PAM. The D strains that did possess *ska-1* lacked PAM. Since *emm* and *ska* are separated by 33-38 kb on the chromosome, it is less likely that they are associated due to tight linkage and is more likely due to epistasis from essential function for each other (188).

sof is another gene linked with tissue tropism, and while SOF is distinct from the M protein, it parallels the M protein by serological specificity. Located 10-16 kb upstream of the *emm* and *emm*-like genes, *sof* is regulated by *mga* (189, 190). Once thought to be a lipoproteinase, SOF has now shown to bind high density lipoprotein in human blood causing the release of lipid cargo, which in turn leads to serum opacity (191, 192). When comparing this to *emm* chromosomal patterns, *sof* is found in 100% of pattern E strains and very few other strains (44).

Also shown to be involved in tissue tropism is the FCT region, a section of the chromosome 300 kb downstream from *mga*. This region is highly recombinatorial and contains Fn- and collagen-binding proteins as well as T antigen loci (119). Observed under linkage disequilibrium with *mga* are the regulators of FCT; *rofA*, a positive transcriptional regulator, or *nra*, a negative transcriptional regulator. Strains of *S. pyogenes* contain only one or the other in most cases with *rofA* or *nra* occupying the same locus of the FCT region between strains (44, 187). Along with either *nra* or *rofA* includes many microbial surface components recognizing adhesive matrix molecules (MSCRAMMs) that bind to human fibronectin or collagen (*prtF1* and *prtF2*). This feature makes the proteins produced in the FCT region important for host tissue adherence (189). The majority of A-C pattern strains contain *rofA* along with *mga-1*, while D pattern strains harbour *nra* with *mga-2*. The E pattern strains contain *rofA*, similar to A-C strains, but possess *mga-2* similar to D pattern strains (44, 187). Along with this, the collagen binding protein encoded by *cpa* is present in the majority of pattern

D strains, but only about half of pattern E strains, and few pattern A-C strains. For *prtF1*, the distribution is the opposite being present in the majority of A-C and E strains and being present in few D strains. Finally, *prtF2* is distributed similar to that of *cpa* (189).

Another protein, *S. pyogenes* nicotinamide adenine dinucleotide (NAD) glycohydrolase (SPN), has also been implicated in playing a role in tropism. This enzyme cleaves the glycosidic bond of β–NAD$^+$, producing nicotinamide and adenosine diphosphate ribose (ADP-ribose) (193). Activity of SPN can partially predict the *emm* pattern type; an active SPN shows an E pattern 85.5% of the time, D pattern 10.9% of the time, and A-C only 3.6% of the time. When SPN is inactive, the D patterns predominates at 63.8%, the A-C pattern at 31%, and the E pattern at 5.2% (194).

Lastly, SAgs also have a molecular connection to particular *emm* pattern types. It had been assumed that SAgs would be spread equally between strains since they are contained on mobile genetic elements. However, when looking at the alleles of *streptococcal pyrogenic exotoxin A* or *C* (*speA*, *speC*) from sequenced *S. pyogenes* strains, non-random association was found. Patterns A-C contained *speC2* while pattern E contained *speC1*, despite both patterns being well represented of the nasopharyngeal reservoir. Pattern D also matched E containing *speC1*. *speA* consists of numerous alleles but was found to have arose from only two lineages termed I and II. This data showed *speA* to be the opposite of *speC* with E and A-C, with being of lineage I while D contained *speA* of lineage II (195).

Through the increasing ease at which genomes can be sequenced and the data accessed, the complex issue of tissue tropism with *S. pyogenes* is becoming more clear. While first grouped based solely on the *emm* gene sequence, an entire genomic sequence can now be associated with the disease causing phenotype, allowing for a more complete picture. Continually investigating the genome in order to understand how it functions will allow for a more complete phenotype associated with every genome.

1.5.6 *emm* types, diseases, and the associated non-suppurative sequelae

While each *emm* type pattern based on molecular evidence is associated with non-

invasive pharyngitis and/or impetigo, patterns have also been associated with post-streptococcal sequelae such as APSGN and ARF. Accepted dogma of *S. pyogenes* disease philosophy has been that ARF occurs following *S. pyogenes* pharyngitis (196). In early documented cases of ARF published in the 1880s, not a single patient was reported to have a skin disease before, during, or after an ARF attack (197). In the 20th century, ARF continued to be seen as a consequence of pharyngitis in the general population, but with increased occurrences to students at boarding schools and cadets in military academy. The first clear evidence of pharyngitis causing ARF was the use of randomized penicillin treatment for pharyngitis in military barracks in the 1950s. The risk of ARF within the penicillin treated group was 70% lower (197). It was also shown that not all pharyngitis strains cause ARF, with those causing ARF being termed rheumatogenic (42). Looking at specific molecular characteristics of the rheumatogenic strains, they fall under the A-C pattern and also lack the ability to produce lipoprotein lipase and SOF (42, 198). Based on these factors, rheumatogenic strains contain the M types 1, 3, 5, 6, 14, 18, 19, 24, 27, and 29. Falling under the A-C pattern further demonstrates the idea that ARF is preceded by pharyngitis (1, 4, 42). However, some differences do occur, such as in Polynesian populations of Hawaii, where ARF was caused by *S. pyogenes* with M types 71, 93, and 98 (D pattern) and M types 92, 103, and 112 (E pattern). Despite the decline of ARF in the industrial developed world, this disease remains endemic in the developing countries, and particularly among indigenous peoples (5).

Driven by the epidemiological evidence from the northern Australian indigenous people is the idea that impetigo could also be a driving force behind ARF in impetigo endemic regions (199). The Australian indigenous people have the highest ARF rate in the world at ~550 per 100 000, yet rates of pharyngitis are low (200-202). Impetigo, on the other hand, is very common with frequencies of 50-70% every year (199, 201, 202). In addition, the classic rheumatic *emm* types are barely present, representing only 2% of the strains in one report (203), 6% of another (204), or were not present at all (201). This lead McDonald *et al.* to conclude that common and recurring skin infections may lead to an immune response against throat infections, and that ARF, at least in this area, may follow impetigo infection (202). The link between impetigo and ARF is interesting, but as Kaplan and Bisno described, the lack of antibodies fails to provide tangible proof of

active immunity, and has pointed out that other studies have shown the ability of rheumatogenic strains to persist within communities giving rise to the disease, despite being sometimes notoriously hard to isolate from ARF patients (196). Further work is required to be able to completely confirm or deny this theory.

Another question is the reason for the increased rate of ARF for the Australian indigenous people compared to the general population. Interestingly, other native populations also possess an increase incidence of ARF within their populations. The Māori and Pacific Islanders of New Zealand show a rate of ARF 10 to 20 times above the general New Zealand population, respectively (205). Looking at just the 5-14 year age group, those most likely to get the disease, the rate is a staggering 20 to 40 times the normal population. For the general population of all three groups, this leads to death rates per 100 000 people at 7.3 and 7.0 for the Māori and Pacific Islanders, respectively, and 3.7 for the general New Zealand population (206). A Hawaiian study comparing Caucasian to Hawaiian and other Polynesian residents of the rate of ARF showed rates ranging from 19.2 to 120.7 times greater than Caucasian people based on information from patients of 21 years and younger (207). Further research including all age groups demonstrated that the ethnic population still has a greater risk to ARF by 4.8 times. Most importantly, this was not due to socioeconomic status, therefore, showing a potential genetic link for an increased risk of ARF for certain populations (208).

APSGN, on the other hand, has followed infection of both the throat and skin (196). The most common strains that cause APSGN from skin infections include M types 2, 49, 60, and 61 of the E pattern, along with M59 of pattern D and M57 of pattern A-C. Those known to cause APSGN from throat infections are M1 and M12 from the A-C pattern and M4 and M25 of the E pattern (1, 4). Strains of pattern E seem to be the most common cause of APSGN from a molecular level. In Guizhou, China, an outbreak occurred with pattern E making up 73.9% of the strains, with the remaining patterns making up less than 9% each (10). Even in some instances when the rare M types were the ones causing disease, such as M types 48 and 73 in Trinidad and M type 63 in China, they were all still pattern E (10, 209).

1.5.7 Molecular mimicry

Autoimmune diseases, such as ARF and RHD, are caused when the immune system recognizes self antigens as foreign, which can then lead to inflammation and possibly even the destruction of specific tissues (210). This self-recognition can be induced due to what is known as molecular mimicry, defined as the sharing of epitopes between antigens on both the host and a pathogen (211). Occurring between disparate proteins or molecules, epitopes must be close enough in homology to share amino acids determinants or conformational shapes, but still distant enough in order to be recognized as foreign by the immune system (212). There have been defined antigens for mimicry such as identical amino acid sequences, homologous but non-identical amino acid sequences, and lastly, mimicry occurring on two dissimilar epitopes such as peptides and carbohydrates (211). Although it is rare to identify a single pathogen that initiates a specific autoimmune disease, ARF has been known to be caused by *S. pyogenes* for many decades (210). This has lead to considerable research involving molecular mimicry with *S. pyogenes* and host tissues.

Cavelti first discovered the presence of anti-cardiac antibodies in ARF patients in 1945, and was able to produce them in rats using dead streptococci, showing that damage could still occur to the heart without the need for active dividing bacteria (43). While present in those with non-supportive sequelae such as ARF and RHD, these antibodies can also be found in some patients with uncomplicated diseases such as pharyngitis, but at significantly lower titers (43, 213, 214). Most commonly affected is the mitral valve, which separates the left atrium, receiving oxygenated blood from the lungs, from the left ventricle, which pumps oxygenated blood through the aorta. The aortic, tricuspid, and pulmonary valves are less commonly affected, roughly in that order, with the pulmonary valve being rarely affected (41, 42).

It is thought that monoclonal antibodies created against the α-helical coiled-coil shape of M protein and GlcNAc of the group A carbohydrate are the cause of this disease. These antibodies are also able to bind to smooth muscle cells, cardiac myocytes, endothelial cells, valvular interstitial cells, basement membrane, and elastic regions (215, 216). Structures identified to contain the α-helical coiled-coil structures to which antibodies

bind to include laminin, keratin, and the intracellular proteins myosin, tropomyosin, and vimentin. Various experiments were able to show cross-reactivity by deriving antibodies from GlcNAc, M protein, laminin, myosin, and others that could bind to all reciprocal antigens (215-220).

The theory of ARF is that antibodies bind to elements on the mitral valve, causing damage through complement. This causes a disruption in the endothelial layer, exposing more antigenic sites. Endothelial cells, as a result, also express more vascular cell adhesion molecule-1 (VCAM-1). This leads to inflammation and the attraction of lymphocytes that are able to extravasate under the endothelial layer forming nodules, leading to a Th1 response with the cytokines IL-1, TNF-α, and IFN-γ expressed. Repeated bouts of ARF can leave scarring on the heart valve tissue, preventing proper function (211, 216). Surgery may also be required to correct severely damaged valves and, depending on the circumstances, can involve repair or complete replacement of the valve. It has been shown that valve repair should be utilized when possible due to better overall outcomes compared to valve replacement (221).

Other theories also exist to explain the cause of ARF and RHD. Root-Bernstein proposed that while the overall cause is still via molecular mimicry, due to the rarity of cases in relation to the number of *S. pyogenes* infections it must occur only when a dual infection happens along with the coxsackie virus, which can cause myocarditis (222). Dual infections of the cardiomyocytes allows an inflammatory response with binding of antibodies to laminin, collagen, and other exterior antigens, causing damage to expose myosin and other internal antigens (222). Tandon *et al*., on the other hand, suggested that the interaction between the N-terminus of the M protein and the CB3 region in collagen type IV creates an antibody response against collagen, as well as an inflammatory response, thus no specific molecular mimicry was involved (223).

APSGN, as previously described, is also associated with non-invasive diseases. As far back as the 18th century, APSGN following scarlet fever was well known and potentially dangerous. It was first postulated in 1903 by Clemens von Pirquet that APSGN was caused by antibody-driven immune reactions that harmed the body rather than being beneficial (37). Schick first demonstrated this idea around 1908 by comparing the

difference in time between the response to injections of a heterologous protein to the onset of APSGN after scarlet fever (224). Later in the 20th century it was realized that not only was scarlet fever caused by *S. pyogenes*, but also APSGN (37). Sera from patients with APSGN were shown to contain antibodies against type IV collagen, laminin, heparin sulfate proteoglycan, and galactosamine-containing proteoglycan of the basement membrane (225, 226). Along with these antibodies, complement protein 3 (C3) and other immunoglobulins were also found deposited on the glomeruli in APSGN patients (224).

Since Schick's work, it has been generally accepted that APSGN is an immune-mediated disease along with the idea that like rheumatogenic *S. pyogenes*, there are also nephritogenic strains. Despite this, the responsible antigen(s) from *S. pyogenes* still remains debatable (227). One of the most popular candidates is nephritis-associated plasmin receptor (NAPlr) (228). NAPlr is a glycolytic enzyme with plasmin binding activity, and shows adhesion to fibronectin, myosin, and actin (228, 229). In patients with ASPGN, anti-NAPlr antibodies were found in the sera and in glomerular biopsies with antibodies bound to glomerular endocapillary neutrophils, mesangial cells, endothelial cells and the glomerular basement membrane (228-232). There have been some discrepancies, however, as Batsford *et al.* was unable to find these same antibodies within ASPGN patients in either sera or biopsies and hypothesized that these differences could be do to dissimilarities in genetics and demographics of the groups used between researchers (227, 231). Despite this, some still believe the importance of NAPlr in the cause of the disease with the following theory. In the initial stages, before antibody production, NAPlr accumulates in the renal glomeruli on the mesangial matrix and glomerular basement membrane, entering in circulation from a throat or skin infection. Here it traps plasmin, in its active form, causing damage by degrading the glomerular basement membrane through its own action or through the further activation of promatrix metalloproteases. This also leads to inflammation due to the recruitment of neutrophils and macrophages (230, 232). Once antibodies are formed, immune complexes have the ability to pass through the damaged glomerular basement membrane where they accumulate in the subepithelial space creating "humps", a distinguishing characteristic of the disease. Finally, the activation of complement is the last step leading to the overt disease stage (232).

The other common antigen associated with APSGN is SpeB. Poon-King *et al.* isolated a plasmin-binding protein, that also bound with antibodies in APSGN sera, and N-terminal sequencing showed that it was SpeB (233). ASPGN patients showed high anti-SpeB (zymogen or protease form) titers compared against controls without the disease (227). Biopsies showed staining of anti-SpeB antibodies to mesangial areas and regions of peripheral capillary walls. Double staining showed some co-localization with NAPlr and localization with both C3 and IgG within the humps (227, 232). SpeB is also known for its plasmin binding activity, which like NAPlr, may cause inflammation leading to the penetration of leukocytes and antigen-antibody complexes (234).

Most interesting is the finding of Luo *et al.*, using a mouse model of APSGB. Injecting mice with a protease negative SpeB strain of *S. pyogenes* resulted in IgG deposits in the glomeruli, along the glomerular capillary walls, and in the mesangium of the kidney. Along with this, leukocyte infiltration and C3 deposition was found, along with proteinuria. When passively immunizing with a monoclonal anti-SpeB antibody, it was found to bind heat shock protein 70 and thioredoxin on mouse kidney endothelial cells. Antibody deposits and complement activation were also evident in the glomeruli along with proteinuria, showing glomerular damage. This provided not only evidence of molecular mimicry and associated auto antigens, but also the potential non-importance of protease activity from SpeB (235).

1.6 Identification of genes important for *in vivo* survival and pathogenesis

In order to fully understand the multifaceted process of *S. pyogenes* colonization at the molecular level, an approach is required that can identify numerous components simultaneously. More specifically is the identification of genes activated only by the environment created by the *in vivo* model. There are many different techniques that have been developed in order to identify genes in this way. Each has its own strengths and weaknesses, and sometimes are more suitable depending on the model and pathogen being used. Some strategies include signature-tagged mutagenesis (STM), microarray, RNA sequencing (RNA-seq), differential fluorescence induction (DFI), *in vivo* induced antigen technology (IVIAT), *in vivo* expression technology (IVET), and recombinase-based *in vivo* expression technology (RIVET), and each will be discussed below.

1.6.1 Signature-tagged mutagenesis

STM is a negative selection method that uses random integration of a transposon. First used in *Salmonella enterica* serovar Typhimurium (*S. typhimurium*), this method successfully identified known virulence genes, therefore, validating the method, as well as many new virulence genes (236). In STM, a transposon is flagged with a specific DNA tag to allow for specific PCR identification from a mix of multiple clones. In practice, a number of transposons, each with a different tag, are used to generate multiple transposon libraries, created from the organism of interest. These clones are separated, via plating, and individual clones are picked and maintained using microtiter plates. A portion of the library of clones is assembled as a pool containing one clone from each of the differently tagged transposon libraries. They are grown *in vitro*, and from here a portion is removed and pooled (the input pool), and DNA is extracted. The remaining pool is then exposed to the *in vivo* environment. The group of clones is recovered (the output pool), and DNA is extracted. PCR is then used do identify the clones present in both the input and output pools using primers specific for each transposon tag. The presence of a clone within the input pool but absent in the output pool identifies it as a gene dispensable *in vitro* but required *in vivo*. The clone can then be identified from the plates where it is stored for further analysis. This must then be repeated for the remaining clones (236-239).

This method has been successfully used in many different Gram-negative and positive bacteria including *Listeria monocytogenes* (240), *Neisseria meningitides* (241), *Burkholderia cenocepacia* (242), *S. aureus* (243), *Streptococcus agalactiae* (244), and also, importantly, *S. pyogenes* (239). While transposon libraries can be created with relative ease, it does require the storage of thousands of individual clones for the screening process, which can be labourious since STM is a negative selection screening system.

1.6.2 Microarray

One common technique used currently to evaluate global gene expression is the microarray (245). Along with gene expression profiling, microarrays can also be used for

(but not limited to) pathogen detection, genome comparison, single nucleotide polymorphism (SNP) detection, and alternate exon splicing. For a basic microarray, an array of known probes is bound to a solid support and an unknown target is then hybridized to the array where it can be analyzed and quantified (246). Many different microarray technologies exist varying in how the probe support is constructed and analyzed, but the most basic, and probably the most common, are the printed and the *in situ* microarrays (247).

Printed microarrays have probes created either by PCR or using oligonucleotide synthesis. Based on the different production methods, PCR probes typically range from 200 bp to 800 bp, while oligonucleotides probes typically range from 25 bp to 150 bp. The probes are then printed onto supports, such as glass, with the PCR probes first being denatured. To increase the accuracy of measurements, each gene has multiple probes associated with different regions of the gene (248). This leads to printed microarrays containing approximately 10 000 to 30 000 probes, but are considered low density (247). Next, the RNA target is extracted and converted to cDNA from two sources; the organism of interest obtained from the model system and the organism grown *in vitro* in laboratory conditions. The cDNA is also labelled with fluorescent dyes; the two different sources each represented by a different colour. From here equal amounts of cDNA from each condition are added to the slide to allow hybridization to the immobilized probes in a competitive manner. The chip is scanned to determine the expression of the genes by measuring the intensity of the fluorescent signal. The range of colour intensity represents the change of gene expression between the two samples (245, 249, 250).

In situ-synthesized microarrays use oligonucleotide probes that are printed directly onto quartz wafer chips used as supports. Similar to printed arrays, probes range from 25 bp to 150 bp but *in situ* arrays typically carry more that 10^6 probes and therefore are considered high density. These microarrays have high specificity due to the increased number of probes per gene, classically using 11 probes for every 600 bp. In addition, probes with mismatches are also included to ensure no mispairings are occurring. The RNA is extracted, converted, and labelled in the same manner, however, *in situ* microarrays do not use competitive hybridization, and instead, compare the two samples on two separate

identical microarrays (247, 250).

Microarrays have many associated benefits. They allow gene expression of an entire organism to be measured at once from a single RNA sample, eliminating the requirement of clone libraries (251). Also, the microarray can be performed with a relatively small amount of RNA obtained from the source. However, microarrays can only be constructed for organisms for which the genome sequence is known. Microarrays are also currently not cost effective, requiring specialized equipment for operation (251). Also important to consider is that the upregulation of mRNA still does not necessarily indicate the expression of protein, as other mechanisms may yet control translation of the gene product. Finally, microarrays only provide a snapshot time point from which the RNA was extracted. Therefore, an accurate depiction of the organism might not be shown, as certain genes may not be captured during their change in regulation.

1.6.3 RNA sequencing

RNA-seq uses deep sequencing in order to both map and sequence the transcriptome in question. Instead of capturing transcripts of interest via hybridization like microarrays, the transcripts are sequenced and mapped, without reference transcripts, creating a map of structure and expression level (249, 252). While different sequencing technologies are used, RNA-seq sample preparation is relatively the same. Briefly, RNA is converted to cDNA using random hexamers or poly-A primers. Followed by RNA degradation, the cDNA is made double stranded. The addition of adapters to both ends creates known sequences for PCR primers. The sample is then amplified via PCR and sequenced. The number of reads each gene receives allows the level of expression to be determined (252, 253). To date, the major RNA-seq systems include SOLiD sequencer, 454 FLX Pyrosequencer, and Illumina Genome Analyzer, each with their own technology for the PCR process.

RNA-seq provides many benefits despite being at an early stage. Most importantly, unlike hybridization technology, it can be used without a reference genome or transcript. Since it works at the level of the nucleotide by sequencing small fragments, RNA-seq can find anomalies such as SNPs, discern transcription boundaries, and identify where exons

splice. Finally, RNA-seq has much greater sensitivity for genes at both low and high expression, creating a more accurate overall readout (252). However, RNA-seq still remains very costly to use, with numerous sample reads required to ensure coverage of the genome and accuracy (249, 254). Challenges also currently exist within the bioinformatics programs used as long stretches of single nucleotides or repeats can be difficult to interpret accurately. Sample preparation can also bias the fragments towards either the 5' or 3' end, at the RNA or cDNA stage, respectively (252). Finally, the large amount of sequencing data requires appropriate bioinformatics resources and expertise. While user friendly software is being developed, using command line languages is still necessary to make use of these large data files (249).

1.6.4 Differential fluorescence induction

Another method used to identify genes induced in specific environments is DFI. This promoter trap approach utilizes green fluorescent protein (GFP) as a selectable marker and exploits fluorescence-activated cell sorting (FACS) to allow high-throughput screening of gene expression (255). A genomic library of fragments cloned upstream of a promoterless *gfp* is created and transformed into the organism of interest. If the fragment is an active promoter, *gfp* will be expressed, identifying the clone. The pool of clones is first grown under standard *in vitro* conditions allowing the activation of all constitutive promoters. Using FACS, these clones can be eliminated. The remaining pool of clones can then be grown under the condition of interest. Promoters specific for this condition will cause GFP expression, allowing these clones to be separated using FACS. This pool of clones can then be plated to identify single clones that can be sequenced to determine the active promoter (255). Other groups have utilized a method where the pool of clones is grown under the condition of interest first, identifying all active promoters, followed by growth under standard conditions to identify the promoters specific to the condition of interest (256-259).

DFI has shown versatility due to its ability to work for both Gram-positive and Gram-negative bacteria including *Streptococcus pneumoniae* (257), *S. aureus* (258), *S. typhimurium* (255), and *L. monocytogenes* (259). DFI is able to use a positive selection technology to screen a large number of clones rapidly by using the semi-automated

approach of flow cytometry. To date, DFI has never been used with *S. pyogenes*, however, successful creation of the system would require some modification. *S. pyogenes* seems to possess a high background rate when viewing with green fluorescence under a microscope (unpublished results). This may require the use of a red fluorescent gene, which is currently an active area of research in our laboratory.

1.6.5 *in vivo*-induced antigen technology

In vivo-induced antigen technology (IVIAT) is a technique that identifies the immunogenic antigens of a pathogen as opposed to genes or their promoters. Briefly, sera from infected individuals are pooled and incubated with the organism of interest grown in laboratory conditions. This allows for the absorption of the antibodies associated with *in vitro* growth from the collected sera. Next, a genomic library from the organism of interest is created in protein expression vectors that are expressed in *Escherichia coli*, followed by colony hybridization with the pool of remaining antibodies that has not been absorbed. Any antibodies that bind are interpreted as *in vivo* induced antigenic determinants. Following this, colonies that bind the *in vivo* enriched sera have plasmids extracted and sequenced to determine the gene upregulated *in vivo*. Confirmation can then be determined by repeating the system using individual clones and qRT-PCR utilizing model systems (238). IVIAT has been used successfully on a number of bacterial pathogens including *Mycobacterium tuberculosis*, *Vibrio cholerae* (260), and *S. pyogenes* (261).

IVIAT demonstrates the ability to identify *in vivo* induced genes utilizing a system that can probe the organism of interest using material obtained from the organisms intended site of infection, and not a model system. This has tremendous advantages as the organism is behaving in its natural state and therefore the results are more easily interpreted. Along with this, IVIAT allows the screening of many clones at once. However, the screening process tends to be laborious and the serum must be properly absorbed against the organism of interest in order to ensure that all antibodies against *in vitro* proteins are removed. Despite these efforts, there always seems to be a fair number of false positives as demonstrated by the groups utilizing this technique (261, 262).

1.7 Genes important for *in vivo* survival and pathogenesis of *S. pyogenes*

Most of these technologies have been utilized to identify *in vivo* induced genes required by *S. pyogenes* to grow under various conditions. These have helped to further improve knowledge about how *S. pyogenes* reacts with the human host from both pharyngitis and deep tissue infections. Below represents a summary of the important findings learned about this human pathogen utilizing these technologies.

The STM method has recovered a total of 29 genes shown to be required for *S. pyogenes in vivo* growth (239). Six of the genes identified have also been previously identified including *hasA*, *mga*, *smeZ*, *amrA*, and two *sil* genes, lending support to the validity of the other genes. Genes identified covered a range of categories including transport, regulators, cellular processes, and secreted proteins. Some of the interesting genes involved in virulence included transport. The first was a putative macrolide efflux pump (*mefE*) that was shown to be involved in biofilm formation but did not provide any antibiotic resistance (239, 263). Another was an ABC transporter, *salT*, which is part of the salivaricin bacteriocin locus. Interestingly, *S. pyogenes* does not produce the bacteriocin and so the function of the transporter remains unknown. Other putative transporters and cytosolic proteins of unknown function were also identified showing there are still many genes for which we have yet to learn the function (239, 263).

RNA-seq has been performed on *S. pyogenes* twice, both involved in the identification of small RNA (sRNA). Tesorero *et al.* was able to find a small RNA with a link to acid stress during growth and infection by identifying the mRNA involved (264). Another group, Deltcheva *et al.*, identified a novel sRNA that forms part of the complex involved in the processing of pre-crRNA (CRISPR RNA; clustered, regularly interspaced short palindromic repeats), short RNA elements that interfere with exogenous DNA elements (265).

Microarrays have probably given the most information and have been used in many models spanning *ex vivo* [human blood (266) and saliva (267)], *in vitro* (a range of temperatures) (268), and *in vivo* [mouse soft tissue infection models (269, 270) and a cynomolgus macaque model of pharyngitis (271)]. The first microarray chip for *S.*

pyogenes was created incorporating 92% of the known ORFs in strain SF370 and used to determine how temperature regulated gene expression. It was shown that 9% of the represented genes expressed a change when grown at 29°C, with a large number of those genes from the extracellular proteome (268). When using *ex vivo* experimentation, the strain and the technology changed. The microarray utilized shorter oligonucleotides of a higher density, creating much greater coverage of the genome, and was designed to cover six different *S. pyogenes* strains. Using this microarray technology, the two component system *sptR/S* was found to play a key role for persistence in saliva and was controlling genes involved with nutrient acquisition, response to oxidative stress, and evasion of innate and acquired immune responses (267, 272). When *S. pyogenes* was exposed to blood, the greatest change in gene expression was seen at the 30 minutes (min) period with 76% of the genome demonstrating a difference in expression. The streptokinase gene, *ska*, was upregulated along with adhesins such as *emm1*, collagen-binding proteins, and capsule; all of this indicating that *mga* was an important factor. *S. pyogenes* also changed the expression of many genes involved with metabolic functions, through the *covR/S* two-component system, in order to adapt to its new environment (266).

Numerous *in vivo* studies have also been performed to analyze gene expression at the genome level. From the mouse model of soft tissue infection, it was proposed that *S. pyogenes* goes through a three-step process in order to establish itself in a host: establishment, adaptation, and dissemination. These steps involve the orchestration of a number of genes at each stage. Establishment implicated the activation of genes involved in adherence and evasion of the immune system. Adaptation involved continued immune evasion as well as aggregate formation and rapid replication. Finally, the dissemination stage included nutrient acquisition along with tissue breakdown and shedding (269). Experimental pharyngitis using cynomolgus macaques also identified three separated phases of disease including colonization, acute, and asymptomatic. Colonization and inflammation was associated with the expression of SAgs and the different phases of diseases were associated with the regulators from two two-component systems (*covR/S* and *spy0680/spy0681*). *covR/S* was not expressed during the colonization phase, but was turned on during the acute phase, and repressed during the asymptomatic phase. *spy0680/spy0681* was turned on during the colonization and asymptomatic phases (271).

Realizing that two-component systems were important for the survival of the organism, the mouse abscess model was used with two-component system knockout mutants. The *spy0680/M5005_spy0681* knockout showed significantly larger abscess sizes compared to the wild-type, while the other strains showed no difference. This was not surprising as it is a known repressor and has been shown to be active during the asymptomatic phase of experimental pharyngitis in cynomolgus macaques (270).

Utilizing microarray data from nine strains isolated from patients before 1987, clusters were created composed of three invasive strains and six pharyngeal strains. Based on genome expression differences of ~10% they were divided into the pharyngeal transcriptome profile (PTP) and the invasive transcriptome profile (ITP). The distinction between the two genomes was found to be a seven bp insertion creating a truncated *covS*, the histidine kinase of the *covR/S* two-component system (273). Later work showed that SpyCEP, a protease that increases expression in the *covR/S* truncation, was able to cleave chemokines associated with neutrophil activation, explaining their absence from sites of infection (274). Important work also asked how *S. pyogenes* was changing from a metabolic perspective, and how this changed pathogenicity. It was found that catabolite control protein A (CcpA) controlled virulence factors as well as carbohydrate utilization genes, having a severe affect on mouse oropharynx colonization (275). Further work was done to examine the maltose repressor (MalR), a surface carbohydrate binding protein. A *malR* knockout showed a significant reduction in colonization, but no lack of invasive disease, demonstrating how carbon sources can determine pathogenesis (276). Future work continues to use the same type of gene chip covering research in many different areas with to increase our overall knowledge of *S. pyogenes*.

Finally, utilizing the IVIAT method, 16 genes were initially identified in *S. pyogenes*, after which three were identified to be truly upregulated *in vivo*, as determined by qRT-PCR analysis. These genes included *coaA*, a putative pantothenate kinase, *pbp1A*, a putative penicillin binding protein, and *tdcF*, a hypothetical protein (261). *coaA* catalyzes the first step in the biosynthetic pathway leading to coenzyme A, essential in the metabolism of fatty acids, carbohydrates, and amino acids (261). Penicillin-binding proteins are essential for cell morphology and are typically involved in peptidoglycan

synthesis, but *pbp1A* was also found to be important for the resistance of phagocytosis in *S. agalactiae* (261). *tdcF* encodes a hypothetical protein with 60% homology to a protein in *Vibrio vulnificus* that was found to have a translation initiation inhibitor function. These two genes may have a similar function found in *S. pyogenes* (261).

1.8 Rationale and Hypothesis

Each of the systems discussed have both strengths and weaknesses, and have been able to provide valuable information about *S. pyogenes* from different models and systems, including from the *in vivo* environment. One of the major drawbacks of the techniques mentioned is the limitation of the extracted data to a single time point. Utilizing RNA-seq, microarray, or DFI, data analysis shows gene expression only at the time of extraction, and therefore is unable to catch genes where expression might have increased and then fallen back to normal levels. STM allows the ability to identify genes necessary at different time points since once the gene is needed the clone will be removed from the pool. However, using this method requires the storage of a massive number of individual clones and *in vivo* studies also require large numbers of animals, making the method cost inhibitory. Finally, the limiting characteristic of IVIAT is due to the use of *E. coli* as a protein expression vector. Should expression of the protein be toxic, difficult to express, or expressed in low quantity, the protein would not be detected by the antiserum.

Another *in vivo*-induced potential model that is able to overcome some of the weaknesses of these systems is recombinase-based *in vivo* expression technology (RIVET). RIVET uses a method where a permanent change is inherited when a promoter is activated, which can then be screened for, and potentially allows all activated clones throughout the *in vivo* growth period to be identified. This strategy has never previously been used with *S. pyogenes* before, but has been performed on many Gram-positive organisms to date including *S. aureus* (277), *M. tuberculosis* (278), *Lactobacillus plantarum* (279), *Enterococcus faecalis* (280, 281), and *Streptococcus thermophilus* (282). In order to learn more about *S. pyogenes*, RIVET will be used for the first time in this organism.

It is hypothesized that RIVET will provide a novel tool that will be able to identify *S. pyogenes* genes that are important for colonization of the nasopharynx.

1.9 Major Objectives

In order to test this hypothesis there are two major objectives.

Objective 1: To construct and evaluate a functional RIVET system using *S. pyogenes* strain MGAS8232, utilizing a potential counter-selective gene.

Objective 2: To use the system within our established nasopharyngeal model to allow for the identification of *in vivo*-induced genes that are required for the colonization and adaptation of the nasopharyngeal niche, furthering our knowledge of the organism *S. pyogenes*.

1.10 References

1. **Cunningham MW**. 2000. Pathogenesis of group A streptococcal infections. Clin Microbiol Rev **13**:470.
2. **Beres SB**, **Musser JM**. 2007. Contribution of exogenous genetic elements to the group A *Streptococcus* metagenome. PLoS ONE **2**:e800.
3. **Ghosh P**. 2011. The nonideal coiled coil of M protein and its multifarious functions in pathogenesis. Adv Exp Med Biol **715**:197–211.
4. **McMillan DJ**, **Drèze P-A**, **Vu T**, **Bessen DE**, **Guglielmini J**, **Steer AC**, **Carapetis JR**, **Van Melderen L**, **Sriprakash KS**, **Smeesters PR**. 2013. Updated model of group A *Streptococcus* M proteins based on a comprehensive worldwide study. Clin Microbiol Infect **19**:E222–9.
5. **Carapetis JR**, **Steer AC**, **Mulholland EK**, **Weber M**. 2005. The global burden of group A streptococcal diseases. The Lancet Infectious Diseases **5**:685–694.
6. **Walker MJ**, **Barnett TC**, **McArthur JD**, **Cole JN**, **Gillen CM**, **Henningham A**, **Sriprakash KS**, **Sanderson-Smith ML**, **Nizet V**. 2014. Disease manifestations and pathogenic mechanisms of group a *Streptococcus*. Clin Microbiol Rev **27**:264–301.
7. **Shulman ST**, **Bisno AL**, **Clegg HW**, **Gerber MA**, **Kaplan EL**, **Lee G**, **Martin JM**, **Van Beneden C**. 2012. Clinical practice guideline for the diagnosis and management of group A streptococcal pharyngitis: 2012 update by the Infectious Diseases Society of America. CLIN INFECT DIS **55**:e86–e102.
8. **Dale JB**, **Fischetti VA**, **Carapetis JR**, **Steer AC**, **Sow S**, **Kumar R**, **Mayosi BM**, **Rubin FA**, **Mulholland K**, **Hombach JM**, **Schödel F**, **Henao-Restrepo AM**. 2013. Group A streptococcal vaccines: paving a path for accelerated development. Vaccine **31 Suppl 2**:B216–22.
9. **Steer AC**, **Law I**, **Matatolu L**, **Beall BW**, **Carapetis JR**. 2009. Global emm type distribution of group A streptococci: systematic review and implications for vaccine development. The Lancet Infectious Diseases **9**:611–616.
10. **Zheng M-H**, **Jiao Z-Q**, **Zhang L-J**, **Yu S-J**, **Tang G-P**, **Yan X-M**, **He L-H**, **Meng F-L**, **Zhao F**, **Zhang M-J**, **Xiao D**, **Yang Y-H**, **Nie W**, **Zhang J-Z**, **Wang Z-J**. 2009. Genetic analysis of group A streptococcus isolates recovered during acute glomerulonephritis outbreaks in Guizhou province of China. J Clin Microbiol **47**:715–720.
11. **Miner LJ**, **Petheram SJ**, **Daly JA**, **Korgenski EK**, **Selin KS**, **Firth SD**, **Veasy LG**, **Hill HR**, **Bale JF**, **Post-Streptococcal Syndrome Stydy Team**. 2004. Molecular characterization of *Streptococcus pyogenes* isolates collected during periods of increased acute rheumatic fever activity in Utah. The Pediatric Infectious Disease Journal **23**:56–61.
12. **Bisno AL**. 1996. Acute pharyngitis: etiology and diagnosis. PEDIATRICS **97**:949–954.
13. **Ashbaugh CD**, **Moser TJ**, **Shearer MH**, **White GL**, **Kennedy RC**, **Wessels MR**. 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.
14. **Balaji K**, **Thenmozhi R**, **Sundaravadivel M**, **Pandian SK**. 2012. Comparison of Bacterial Communities in the Throat Swabs from Healthy Subjects and

Pharyngitis Patients by Terminal Restriction Fragment Length Polymorphism. Appl Biochem Biotechnol **167**:1459–1473.

15. **Di Pierro F**, **Adami T**, **Rapacioli G**, **Giardini N**, **Streitberger C**. 2013. Clinical evaluation of the oral probiotic *Streptococcus salivarius* K12 in the prevention of recurrent pharyngitis and/or tonsillitis caused by *Streptococcus pyogenesin* adults. Expert Opin Biol Ther **13**:339–343.

16. **Osterlund A**, **Popa R**, **Nikkilä T**, **Scheynius A**, **Engstrand L**. 1997. Intracellular reservoir of *Streptococcus pyogenes in vivo*: a possible explanation for recurrent pharyngotonsillitis. Laryngoscope **107**:640–647.

17. **Pichichero ME**, **Green JL**, **Francis AB**, **Marsocci SM**, **Murphy AM**, **Hoeger W**, **Noriega C**, **Sorrento A**, **Gootnick J**. 1998. Recurrent group A streptococcal tonsillopharyngitis. The Pediatric Infectious Disease Journal **17**:809–815.

18. **Alho OP**, **Koivunen P**, **Penna T**, **Teppo H**, **Koskela M**, **Luotonen J**. 2007. Tonsillectomy versus watchful waiting in recurrent streptococcal pharyngitis in adults: randomised controlled trial. BMJ **334**:939–939.

19. **Orvidas LJ**, **St Sauver JL**, **Weaver AL**. 2006. Efficacy of tonsillectomy in treatment of recurrent group A beta-hemolytic streptococcal pharyngitis. Laryngoscope **116**:1946–1950.

20. **Dunne EM**, **Marshall JL**, **Baker CA**, **Manning J**, **Gonis G**, **Danchin MH**, **Smeesters PR**, **Satzke C**, **Steer AC**. 2013. Detection of group A streptococcal pharyngitis by quantitative PCR. BMC Infectious Diseases **13**:1–1.

21. **Bisno AL**, **Stevens DL**. 1996. Streptococcal infections of skin and soft tissues. N Engl J Med **334**:240–246.

22. **Cole C**, **Gazewood J**. 2007. Diagnosis and treatment of impetigo. Am Fam Physician **75**:859–864.

23. **Koning S**, **van der Sande R**, **Verhagen AP**, **van Suijlekom-Smit LWA**, **Morris AD**, **Butler CC**, **Berger M**, **van der Wouden JC**. 2012. Interventions for impetigo. Cochrane Database Syst Rev **1**:CD003261.

24. **Pereira LB**. 2014. Impetigo - review. An Bras Dermatol **89**:293–299.

25. **Wu P-C**, **Lo W-T**, **Chen S-J**, **Wang C-C**. 2014. Molecular characterization of group A streptococcal isolates causing scarlet fever and pharyngitis among young children: A retrospective study from a northern Taiwan medical center. Journal of Microbiology, Immunology and Infection **47**:304–310.

26. **Silva-Costa C**, **Carriço JA**, **Ramirez M**, **Melo-Cristino J**. 2014. Scarlet fever is caused by a limited number of *Streptococcus pyogenes* lineages and is associated with the exotoxin genes *ssa*, *speA* and *speC*. The Pediatric Infectious Disease Journal **33**:306–310.

27. **Ralph AP**, **Carapetis JR**. 2013. Group A streptococcal diseases and their global burden. Curr Top Microbiol Immunol **368**:1–27.

28. **Stevens DL**. 1992. Invasive group A streptococcus infections. CLIN INFECT DIS **14**:2–13.

29. **Taviloglu K**, **Yanar H**. 2007. Necrotizing fasciitis: strategies for diagnosis and management. World J Emerg Surg **2**:19.

30. **Misiakos EP**, **Bagias G**, **Patapis P**, **Sotiropoulos D**, **Kanavidis P**, **Machairas A**. 2014. Current concepts in the management of necrotizing fasciitis. Front Surg **1**:36.

31. **Brosnahan AJ**, **Schlievert PM**. 2011. Gram-positive bacterial superantigen

outside-in signaling causes toxic shock syndrome. FEBS Journal **278**:4649–4667.

32.   **McCormick JK**, **Yarwood JM**, **Schlievert PM**. 2001. Toxic shock syndrome and bacterial superantigens: an update. Annu Rev Microbiol **55**:77–104.

33.   **Stegmayr B**, **Björck S**, **Holm S**, **Nisell J**, **Rydvall A**, **Settergren B**. 1992. Septic shock induced by group A streptococcal infection: clinical and therapeutic aspects. Scand J Infect Dis **24**:589–597.

34.   **Kaul R**, **McGeer A**, **Norrby-Teglund A**, **Kotb M**, **Schwartz B**, **O'Rourke K**, **Talbot J**, **Low DE**. 1999. Intravenous immunoglobulin therapy for streptococcal toxic shock syndrome - a comparative observational study. Clin Infec Dis **28**:800–807.

35.   **Cone LA**, **Stone RA**, **Schlievert PM**, **Sneider RA**, **Rubin AM**, **Jesser K**, **Renker SW**. 2006. An early favorable outcome of streptococcal toxic shock syndrome may require a combination of antimicrobial and intravenous gamma globulin therapy together with activated protein C. Scand J Infect Dis **38**:960–963.

36.   **Yoshizawa N**. 2000. Acute glomerulonephritis. Internal medicine **39**:687–694.

37.   **Rodriguez-Iturbe B**, **Batsford S**. 2007. Pathogenesis of poststreptococcal glomerulonephritis a century after Clemens von Pirquet. Kidney Int **71**:1094–1104.

38.   **Madaio MP**, **Harrington JT**. 2001. The diagnosis of glomerular diseases: acute glomerulonephritis and the nephrotic syndrome. Arch Intern Med **161**:25–34.

39.   **Couser WG**. 1999. Glomerulonephritis. The Lancet **353**:1509–1515.

40.   **Guilherme L**, **Ramasawmy R**, **Kalil J**. 2007. Rheumatic fever and rheumatic heart disease: genetics and pathogenesis. Scand J Immunol **66**:199–207.

41.   **Lee JL**, **Naguwa SM**, **Cheema GS**, **Gershwin ME**. 2009. Acute rheumatic fever and its consequences: a persistent threat to developing nations in the 21st century. Autoimmun Rev **9**:117–123.

42.   **Mody GM**, **Mayosi BM**. 2003. Acute rheumatic fever. Rheumatology 1131–1141.

43.   **Seckeler MD**, **Hoke TR**. 2011. The worldwide epidemiology of acute rheumatic fever and rheumatic heart disease. Clin Epidemiol **3**:67–84.

44.   **Bessen DE**, **Lizano S**. 2010. Tissue tropisms in group A streptococcal infections. Future Microbiology **5**:623–638.

45.   **Bisno A**, **Brito M**, **Collins C**. 2003. Molecular basis of group A streptococcal virulence. The Lancet Infectious Diseases **3**:191–200.

46.   **Feingold DS**. 1986. Bacterial adherence, colonization, and pathogenicity. Arch Dermatol **122**:161–163.

47.   **Cogen AL**, **Nizet V**, **Gallo RL**. 2008. Skin microbiota: a source of disease or defence? British Journal of Dermatology **158**:442–455.

48.   **Heczko PB**, **Lutticken R**, **Hryniewicz W**, **Neugebauer M**, **Pulverer G**. 1979. Susceptibility of *Staphylococcus aureus* and group A, B, C, and G streptococci to free fatty acids. J Clin Microbiol **9**:333–335.

49.   **Chiller K**, **Selkin BA**, **Murakawa GJ**. 2001. Skin microflora and bacterial infections of the skin. J Investig Dermatol Symp Proc **6**:170–174.

50.   **Courtney HS**, **Hasty DL**, **Dale JB**. 2002. Molecular mechanisms of adhesion, colonization, and invasion of group A streptococci. Ann Med **34**:77–87.

51.   **Amerongen AVN**, **Veerman ECI**. 2002. Saliva-the defender of the oral cavity.

Oral Dis **8**:12–22.

52. **Dale BA**, **Tao R**, **Kimball JR**, **Jurevic RJ**. 2006. Oral antimicrobial peptides and biological control of caries. BMC Oral Health **6**:S13.

53. **Courtney HS**, **Hasty DL**. 1991. Aggregation of group A streptococci by human saliva and effect of saliva on streptococcal adherence to host cells. Infection and Immunity **59**:1661–1666.

54. **Fluckiger U**, **Jones KF**, **Fischetti VA**. 1998. Immunoglobulins to group A streptococcal surface molecules decrease adherence to and invasion of human pharyngeal cells. Infection and Immunity **66**:974–979.

55. **Brandt ER**, **Hayman WA**, **Currie B**, **Carapetis J**, **Jackson DC**, **Do KA**, **Good MF**. 1999. Functional analysis of IgA antibodies specific for a conserved epitope within the M protein of group A streptococci from Australian Aboriginal endemic communities. Int Immunol **11**:569–576.

56. **Maddocks SE**, **Wright CJ**, **Nobbs AH**, **Brittan JL**, **Franklin L**, **Strömberg N**, **Kadioglu A**, **Jepson MA**, **Jenkinson HF**. 2011. *Streptococcus pyogenes* antigen I/II-family polypeptide AspA shows differential ligand-binding properties and mediates biofilm formation. Molecular Microbiology **81**:1034–1049.

57. **Loimaranta V**, **Jakubovics NS**, **Hytönen J**, **Finne J**, **Jenkinson HF**, **Stromberg N**. 2005. Fluid- or surface-phase human salivary scavenger protein gp340 exposes different bacterial recognition properties. Infection and Immunity **73**:2245–2252.

58. **Dempster RP**, **Tagg JR**. 1982. The production of bacteriocin-like substances by the oral bacterium *Streptococcus salivarius*. Arch Oral Biol **27**:151–157.

59. **Schnare M**, **Barton GM**, **Holt AC**, **Takeda K**, **Akira S**, **Medzhitov R**. 2001. Toll-like receptors control activation of adaptive immune responses. Nat Immunol **2**:947–950.

60. **Mogensen TH**. 2009. Pathogen recognition and inflammatory signaling in innate immune defenses. Clin Microbiol Rev **22**:240–273.

61. **Beutler B**. 2004. Innate immunity: an overview. Molecular Immunology **40**:845–859.

62. **Janeway CA Jr.**, **Medzhitov R**. 2002. Innate immune recognition. Annu Rev Immunol **20**:197–216.

63. **Pulendran B**, **Ahmed R**. 2006. Innate immunity to immunological memory. Springer Science & Business Media.

64. **Pulendran B**, **Ahmed R**. 2006. Innate immunity to immunological memory **311**.

65. **Nauseef WM**. 2007. How human neutrophils kill and degrade microbes: an integrated view. Immunol Rev **219**:88–102.

66. **Brinkmann V**, **Reichard U**, **Goosmann C**, **Fauler B**, **Uhlemann Y**, **Weiss DS**, **Weinrauch Y**, **Zychlinsky A**. 2004. Neutrophil extracellular traps kill bacteria. Science **303**:1532–1535.

67. **Walport MJ**. 2001. Complement: first of two parts. N Engl J Med **344**:1058–1066.

68. **Zipfel PF**, **Skerka C**. 2009. Complement regulators and inhibitory proteins. Nature Reviews Immunology **9**:729–740.

69. **Gros P**, **Milder FJ**, **Janssen BJC**. 2008. Complement driven by conformational changes. Nature Reviews Immunology **8**:48–58.

70. **Thurman JM**, **Holers VM**. 2006. The central role of the alternative complement

pathway in human disease. J Immunol **176**:1305–1310.

71.   **Tylewska SK**, **Fischetti VA**, **Gibbons RJ**. 1988. Binding selectivity of *Streptococcus pyogenes* and M-protein to epithelial cells differs from that of lipoteichoic acid. Curr Microbiol **16**:209–216.

72.   **Simpson WA**, **Hasty DL**, **Mason JM**, **Beachey EH**. 1982. Fibronectin-mediated binding of group A streptococci to human polymorphonuclear leukocytes. Infection and Immunity **37**:805–810.

73.   **Simpson WA**, **Beachey EH**. 1983. Adherence of group A streptococci to fibronectin on oral epithelial cells. Infection and Immunity **39**:275–279.

74.   **Pankov R**. 2002. Fibronectin at a glance. Journal of Cell Science **115**:3861–3863.

75.   **Beachey EH**, **Simpson WA**. 1982. The adherence of group A streptococci to oropharyngeal cells: the lipoteichoic acid adhesin and fibronectin receptor. Infection **10**:107–111.

76.   **Talay SR**, **Valentin-Weigand P**, **Jerlström PG**, **Timmis KN**, **Chhatwal GS**. 1992. Fibronectin-binding protein of *Streptococcus pyogenes*: sequence of the binding domain involved in adherence of streptococci to epithelial cells. Infection and Immunity **60**:3837–3844.

77.   **Hanski E**, **Caparon M**. 1992. Protein F, a fibronectin-binding protein, is an adhesin of the group A streptococcus *Streptococcus pyogenes*. Proc Natl Acad Sci USA **89**:6172–6176.

78.   **Hanski E**, **Horwitz P**, **Caparon MG**. 1992. Expression of protein F, the fibronectin-binding protein of *Streptococcus pyogenes* JRS4, in heterologous streptococcal and enterococcal strains promotes their adherence to respiratory epithelial cells. Infection and Immunity **60**:5119.

79.   **Sela S**, **Aviv A**, **Tovi A**, **Burstein I**, **Caparon MG**, **Hanski E**. 1993. Protein F: an adhesin of *Streptococcus pyogenes* binds fibronectin via two distinct domains. Molecular Microbiology **10**:1049–1055.

80.   **Talay SR**, **Valentin-Weigand P**, **Timmis KN**, **Chhatwal GS**. 1994. Domain structure and conserved epitopes of Sfb protein, the fibronectin-binding adhesin of *Streptococcus pyogenes.* Molecular Microbiology **13**:531–539.

81.   **Ozeri V**, **Tovi A**, **Burstein I**, **Natanson-Yaron S**, **Caparon MG**, **Yamada KM**, **Akiyama SK**, **Vlodavsky I**, **Hanski E**. 1996. A two-domain mechanism for group A streptococcal adherence through protein F to the extracellular matrix. EMBO J **15**:989–998.

82.   **Kreikemeyer B**, **Talay SR**, **Chhatwal GS**. 1995. Characterization of a novel fibronectin-binding surface protein in group A streptococci. Molecular Microbiology **17**:137–145.

83.   **Rakonjac JV**, **Robbins JC**, **Fischetti VA**. 1995. DNA sequence of the serum opacity factor of group A streptococci: identification of a fibronectin-binding repeat domain. Infection and Immunity **63**:622–631.

84.   **Courtney HS**, **Dale JB**, **Hasty DL**. 2002. Mapping the fibrinogen-binding domain of serum opacity factor of group A streptococci. Curr Microbiol **44**:236–240.

85.   **Katerov V**, **Lindgren P-E**, **Totolian AA**, **Schalén C**. 2000. Streptococcal opacity factor: a family of bifunctional proteins with lipoproteinase and fibronectin-binding activities. Curr Microbiol **40**:149–156.

86.   **Oehmcke S**, **Podbielski A**, **Kreikemeyer B**. 2004. Function of the fibronectin-binding serum opacity factor of *Streptococcus pyogenes* in adherence to epithelial cells. Infection and Immunity **72**:4302–4308.

87.   **Jeng A**, **Sakota V**, **Li Z**, **Datta V**, **Beall B**, **Nizet V**. 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.

88.   **Courtney HS**, **Li Y**, **Twal WO**, **Argraves WS**. 2009. Serum opacity factor is a streptococcal receptor for the extracellular matrix protein Fibulin-1. J Biol Chem **284**:12966–12971.

89.   **Jaffe J**, **Natanson-Yaron S**, **Caparon MG**, **Hanski E**. 1996. Protein F2, a novel fibronectin-binding protein from *Streptococcus pyogenes*, possesses two binding domains. Molecular Microbiology **21**:373–384.

90.   **Rocha CL**, **Fischetti VA**. 1999. Identification and characterization of a novel fibronectin-binding protein on the surface of group A streptococci. Infection and Immunity **67**:2720–2728.

91.   **Kreikemeyer B**, **Oehmcke S**, **Nakata M**, **Hoffrogge R**, **Podbielski A**. 2004. *Streptococcus pyogenes* fibronectin-binding protein F2: expression profile, binding characteristics, and impact on eukaryotic cell interactions. J Biol Chem **279**:15850–15859.

92.   **Ramachandran V**, **McArthur JD**, **Behm CE**, **Gutzeit C**, **Dowton M**, **Fagan PK**, **Towers R**, **Currie B**, **Sriprakash KS**, **Walker MJ**. 2004. Two distinct genotypes of *prtF2*, encoding a fibronectin binding protein, and evolution of the gene family in *Streptococcus pyogenes*. Journal of Bacteriology **186**:7601–7609.

93.   **Terao Y**, **Kawabata S**, **Nakata M**, **Nakagawa I**, **Hamada S**. 2002. Molecular characterization of a novel fibronectin-binding protein of *Streptococcus pyogenes* strains isolated from toxic shock-like syndrome patients. Journal of Biological Chemistry **277**:47428–47435.

94.   **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.

95.   **Pancholi V**, **Fischetti VA**. 1988. Isolation and characterization of the cell-associated region of group A streptococcal M6 protein. Journal of Bacteriology **170**:2618–2624.

96.   **Fischetti VA**, **Parry DA**, **Trus BL**, **Hollingshead SK**, **Scott JR**, **Manjula BN**. 1988. Conformational characteristics of the complete sequence of group A streptococcal M6 protein. Proteins **3**:60–69.

97.   **Smeesters PR**, **Mardulyn P**, **Vergison A**, **Leplae R**, **Van Melderen L**. 2008. Genetic diversity of group A streptococcus M protein: implications for typing and vaccine development. Vaccine **26**:5835–5842.

98.   **Smeesters PR**, **McMillan DJ**, **Sriprakash KS**. 2010. The streptococcal M protein: a highly versatile molecule. Trends in Microbiology **18**:275–282.

99.   **Akesson P**, **Schmidt KH**, **Cooney J**, **Bjorck L**. 1994. M1 protein and protein H: IgGFc- and albumin-binding streptococcal surface proteins encoded by adjacent genes. Biochem J **300 (Pt 3)**:877–886.

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

101. **Ellen RP**, **Gibbons RJ**. 1972. M protein-associated adherence of *Streptococcus pyogenes* to epithelial surfaces: prerequisite for virulence. Infection and Immunity **5**:826–830.
102. **Ellen RP**, **Gibbons RJ**. 1974. Parameters affecting the adherence and tissue tropisms of *Streptococcus pyogenes*. Infection and Immunity **9**:85–91.
103. **Wang JR**, **Stinson MW**. 1994. M protein mediates streptococcal adhesion to HEp-2 cells. Infection and Immunity **62**:442–448.
104. **Berkower C**, **Ravins M**, **Moses AE**, **Hanski E**. 1999. Expression of different group A streptococcal M proteins in an isogenic background demonstrates diversity in adherence to and invasion of eukaryotic cells. Molecular Microbiology **31**:1463–1475.
105. **Kantor FS**. 1965. Fibrinogen precipitation by streptococcal M protein: I. Identity of the reactants, and stoichiometry of the reaction. J Exp Med **121**:849–859.
106. **Courtney HS**, **Ofek I**, **Simpson WA**, **Hasty DL**, **Beachey EH**. 1986. Binding of *Streptococcus pyogenes* to soluble and insoluble fibronectin. Infection and Immunity **53**:454–459.
107. **Schmidt KH**, **Mann K**, **Cooney J**, **Köhler W**. 1993. Multiple binding of type 3 streptococcal M protein to human fibrinogen, albumin and fibronectin. FEMS Immunology & Medical Microbiology **7**:135–143.
108. **Caparon MG**, **Stephens DS**, **Olsen A**, **Scott JR**. 1991. Role of M protein in adherence of group A streptococci. Infection and Immunity **59**:1811–1817.
109. **Molinari G**, **Chhatwal GS**. 1999. Role played by the fibronectin-binding protein SfbI (Protein F1) of *Streptococcus pyogenes* in bacterial internalization by epithelial cells. J Infect Dis **179**:1049–1050.
110. **Seya T**, **Turner JR**, **Atkinson JP**. 1986. Purification and characterization of a membrane protein (gp45-70) that is a cofactor for cleavage of C3b and C4b. J Exp Med **163**:837–855.
111. **Giannakis E**, **Jokiranta TS**, **Ormsby RJ**, **Duthy TG**, **Male DA**, **Christiansen D**, **Fischetti VA**, **Bagley C**, **Loveland BE**, **Gordon DL**. 2002. Identification of the streptococcal M protein binding site on membrane cofactor protein (CD46). J Immunol **168**:4585–4592.
112. **Feito MJ**, **Sánchez A**, **Oliver MA**, **Pérez-Caballero D**, **Rodríguez de Córdoba S**, **Albertí S**, **Rojo JM**. 2007. Membrane cofactor protein (MCP, CD46) binding to clinical isolates of *Streptococcus pyogenes*: Binding to M type 18 strains is independent of Emm or Enn proteins. Molecular Immunology **44**:3571–3579.
113. **Kjellén L**, **Lindahl U**. 1991. Proteoglycans: structures and interactions. Annu Rev Biochem **60**:443–475.
114. **Frick I-M**, **Schmidtchen A**, **Sj bring U**. 2003. Interactions between M proteins of *Streptococcus pyogenes* and glycosaminoglycans promote bacterial adhesion to host cells. European Journal of Biochemistry **270**:2303–2311.
115. **Wadstrom T**, **Tylewska S**. 1982. Glycoconjugates as possible receptors for S*treptococcus pyogenes*. Curr Microbiol **7**:343–346.
116. **Wang JR**, **Stinson MW**. 1994. Streptococcal M6 protein binds to fucose-containing glycoproteins on cultured human epithelial cells. Infection and Immunity **62**:1268–1274.
117. **Ryan PA**, **Pancholi V**, **Fischetti VA**. 2001. Group A streptococci bind to mucin and human pharyngeal cells through sialic acid-containing receptors. Infection

and Immunity **69**:7402–7412.

118. **Mora M**, **Bensi G**, **Capo S**, **Falugi F**, **Zingaretti C**, **Manetti AGO**, **Maggi T**, **Taddei AR**, **Grandi G**, **Telford JL**. 2005. Group A streptococcus produce pilus-like structures containing protective antigens and Lancefield T antigens. Proc Natl Acad Sci USA **102**:15641–15646.

119. **Bessen DE**. 2002. Genomic localization of a T serotype locus to a recombinatorial zone encoding extracellular matrix-binding proteins in *Streptococcus pyogenes*. Infection and Immunity **70**:1159–1167.

120. **Abbot EL**, **Smith WD**, **Siou GPS**, **Chiriboga C**, **Smith RJ**, **Wilson JA**, **Hirst BH**, **Kehoe MA**. 2007. Pili mediate specific adhesion of *Streptococcus pyogenes* to human tonsil and skin. Cellular Microbiology **9**:1822–1833.

121. **Manetti AGO**, **Zingaretti C**, **Falugi F**, **Capo S**, **Bombaci M**, **Bagnoli F**, **Gambellini G**, **Bensi G**, **Mora M**, **Edwards AM**, **Musser JM**, **Graviss EA**, **Telford JL**, **Grandi G**, **Margarit I**. 2007. *Streptococcus pyogenes* pili promote pharyngeal cell adhesion and biofilm formation. Molecular Microbiology **64**:968–983.

122. **Zahner D**, **Scott JR**. 2008. SipA is required for pilus formation in *Streptococcus pyogenes* serotype M3. Journal of Bacteriology **190**:527–535.

123. **Quigley BR**, **Zähner D**, **Hatkoff M**, **Thanassi DG**, **Scott JR**. 2009. Linkage of T3 and Cpa pilins in the *Streptococcus pyogenes* M3 pilus. Molecular Microbiology **72**:1379–1394.

124. **Smith WD**, **Pointon JA**, **Abbot E**, **Kang HJ**, **Baker EN**, **Hirst BH**, **Wilson JA**, **Banfield MJ**, **Kehoe MA**. 2010. Roles of minor pilin subunits Spy0125 and Spy0130 in the serotype M1 *Streptococcus pyogenes* strain SF370. Journal of Bacteriology **192**:4651–4659.

125. **Nakata M**, **Kimura KR**, **Sumitomo T**, **Wada S**, **Sugauchi A**, **Oiki E**, **Higashino M**, **Kreikemeyer B**, **Podbielski A**, **Okahashi N**, **Hamada S**, **Isoda R**, **Terao Y**, **Kawabata S**. 2011. Assembly mechanism of FCT region type 1 pili in serotype M6 *Streptococcus pyogenes*. Journal of Biological Chemistry **286**:37566–37577.

126. **Becherelli M**, **Manetti AGO**, **Buccato S**, **Viciani E**, **Ciucchi L**, **Mollica G**, **Grandi G**, **Margarit I**. 2012. The ancillary protein 1 of *Streptococcus pyogenes* FCT-1 pili mediates cell adhesion and biofilm formation through heterophilic as well as homophilic interactions. Molecular Microbiology **83**:1035–1047.

127. **Crotty Alexander LE**, **Maisey HC**, **Timmer AM**, **Rooijakkers SHM**, **Gallo RL**, **Köckritz-Blickwede von M**, **Nizet V**. 2010. M1T1 group A streptococcal pili promote epithelial colonization but diminish systemic virulence through neutrophil extracellular entrapment. J Mol Med **88**:371–381.

128. **Cywes C**, **Stamenkovic I**, **Wessels MR**. 2000. CD44 as a receptor for colonization of the pharynx by group A *Streptococcus*. J Clin Invest **106**:995–1002.

129. **Crater DL**, **van de Rijn I**. 1995. Hyaluronic acid synthesis operon (*has*) expression in group A streptococci. J Biol Chem **270**:18452–18458.

130. **Wessels MR**, **Moses AE**, **Goldberg JB**, **DiCesare TJ**. 1991. Hyaluronic acid capsule is a virulence factor for mucoid group A streptococci. Proc Natl Acad Sci USA **88**:8317–8321.

131. **Wessels MR**, **Bronze MS**. 1994. Critical role of the group A streptococcal

capsule in pharyngeal colonization and infection in mice. Proc Natl Acad Sci USA **91**:12238–12242.

132. **Husmann LK**, **Yung DL**, **Hollingshead SK**, **Scott JR**. 1997. Role of putative virulence factors of *Streptococcus pyogenes* in mouse models of long-term throat colonization and pneumonia. Infection and Immunity **65**:1422–1430.

133. **Cole JN**, **Pence MA**, **Kockritz-Blickwede von M**, **Hollands A**, **Gallo RL**, **Walker MJ**, **Nizet V**. 2010. M protein and hyaluronic acid capsule are essential for *in vivo* selection of *covRS* mutations characteristic of invasive serotype M1T1 group A *Streptococcus*. MBio **1**:e00191–10–e00191–17.

134. **Schrager HM**, **Albertí 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. J Clin Invest **101**:1708–1716.

135. **Percy MG**, **Gründling A**. 2014. Lipoteichoic acid synthesis and function in Gram-positive bacteria. Annu Rev Microbiol **68**:81–100.

136. **Stewart FS**, **Martin WT**. 1962. Adsorption of a streptococcal red cell sensitising antigen to various tissues. The Journal of pathology and bacteriology **84**:251–253.

137. **Ofek I**, **Beachey EH**, **Jefferson W**, **Campbell GL**. 1975. Cell membrane-binding properties of group A streptococcal lipoteichoic acid. J Exp Med **141**:990–1003.

138. **Beachey EH**, **Ofek I**. 1976. Epithelial cell binding of group A streptococci by lipoteichoic acid on fimbriae denuded of M protein. J Exp Med **143**:759–771.

139. **Simpson WA**, **Ofek I**, **Sarasohn C**, **Morrison JC**, **Beachey EH**. 1980. Characteristics of the binding of streptococcal lipoteichoic acid to human oral epithelial cells. J Infect Dis **141**:457–462.

140. **Courtney H**, **Ofek I**, **Simpson WA**, **Beachey EH**. 1981. Characterization of lipoteichoic acid binding to polymorphonuclear leukocytes of human blood. Infection and Immunity **32**:625–631.

141. **Dale JB**, **Baird RW**, **Courtney HS**, **Hasty DL**, **Bronze MS**. 1994. Passive protection of mice against group A streptococcal pharyngeal infection by lipoteichoic acid. J Infect Dis **169**:319–323.

142. **Simpson WA**, **Ofek I**, **Beachey EH**. 1980. Fatty acid binding sites of serum albumin as membrane receptor analogs for streptococcal lipoteichoic acid. Infection and Immunity **29**:119–122.

143. **Courtney HS**, **Simpson WA**, **Beachey EH**. 1983. Binding of streptococcal lipoteichoic acid to fatty acid-binding sites on human plasma fibronectin. Journal of Bacteriology **153**:763–770.

144. **Courtney HS**, **Bronze MS**, **Dale JB**, **Hasty DL**. 1994. Analysis of the role of M24 protein in group A streptococcal adhesion and colonization by use of omega-interposon mutagenesis. Infection and Immunity **62**:4868–4873.

145. **Courtney HS**, **Hunolstein von C**, **Dale JB**, **Bronze MS**, **Beachey EH**, **Hasty DL**. 1992. Lipoteichoic acid and M protein: dual adhesins of group A streptococci. Microbial Pathogenesis **12**:199–208.

146. **Okada N**, **Pentland AP**, **Falk P**, **Caparon MG**. 1994. M protein and protein F act as important determinants of cell-specific tropism of *Streptococcus pyogenes* in skin tissue. Journal of Clinical Investigation **94**:965.

147. **Courtney HS**, **Dale JB**, **Hasty DI**. 1996. Differential effects of the streptococcal fibronectin-binding protein, FBP54, on adhesion of group A streptococci to human buccal cells and HEp-2 tissue culture cells. Infection and Immunity **64**:2415–2419.

148. **Linke C**, **Siemens N**, **Oehmcke S**, **Radjainia M**, **Law RHP**, **Whisstock JC**, **Baker EN**, **Kreikemeyer B**. 2012. The extracellular protein factor Epf from *Streptococcus pyogenes* is a cell surface adhesin that binds to cells through an N-terminal domain containing a carbohydrate-binding module. Journal of Biological Chemistry **287**:38178–38189.

149. **Hasty DL**, **Ofek I**, **Courtney HS**, **Doyle RJ**. 1992. Multiple adhesins of streptococci. Infection and Immunity **60**:2147.

150. **Courtney HS**, **Ofek I**, **Hasty DL**. 1997. M protein mediated adhesion of M type 24 *Streptococcus pyogenes* stimulates release of interleukin-6 by HEp-2 tissue culture cells. FEMS Microbiology Letters **151**:65–70.

151. **Lei B**, **DeLeo FR**, **Reid SD**, **Voyich JM**, **Magoun L**, **Liu M**, **Braughton KR**, **Ricklefs S**, **Hoe NP**, **Cole RL**, **Leong JM**, **Musser JM**. 2002. Opsonophagocytosis-inhibiting Mac protein of group A *Streptococcus*: identification and characteristics of two genetic complexes. Infection and Immunity **70**:6880–6890.

152. **Hidalgo-Grass C**, **Mishalian I**, **Dan-Goor M**, **Belotserkovsky I**, **Eran Y**, **Nizet V**, **Peled A**, **Hanski E**. 2006. A streptococcal protease that degrades CXC chemokines and impairs bacterial clearance from infected tissues. EMBO J **25**:4628–4637.

153. **Nelson DC**, **Garbe J**, **Collin M**. 2011. The cysteine proteinase SpeB from *Streptococcus pyogenes*– a potent modifier of immunologically important host and bacterial proteins. Biological Chemistry **392**:1077–1088.

154. **Hytönen J**, **Haataja S**, **Gerlach D**, **Podbielski A**, **Finne J**. 2001. The SpeB virulence factor of *Streptococcus pyogenes*, a multifunctional secreted and cell surface molecule with strepadhesin, laminin-binding and cysteine protease activity. Molecular Microbiology **39**:512–519.

155. **Kapur V**, **Maffei JT**, **Greer RS**, **Li LL**, **Adams GJ**, **Musser JM**. 1994. Vaccination with streptococcal extracellular cysteine protease (interleukin-1 beta convertase) protects mice against challenge with heterologous group A streptococci. Microbial Pathogenesis **16**:443–450.

156. **Kuo CF**, **Wu JJ**, **Lin KY**, **Tsai PJ**, **Lee SC**, **Jin YT**, **Lei HY**, **Lin YS**. 1998. Role of streptococcal pyrogenic exotoxin B in the mouse model of group A streptococcal infection. Infection and Immunity **66**:3931–3935.

157. **Tsao N**, **Cheng M-H**, **Yang H-C**, **Wang Y-C**, **Liu Y-L**, **Kuo C-F**. 2013. Determining antibody-binding site of streptococcal pyrogenic exotoxin B to protect mice from group A streptococcus infection. PLoS ONE **8**:e55028.

158. **Molloy EM**, **Cotter PD**, **Hill C**, **Mitchell DA**, **Ross RP**. 2011. Streptolysin S-like virulence factors: the continuing *sagA*. Nature Publishing Group **9**:670–681.

159. **Fontaine MC**, **Lee JJ**, **Kehoe MA**. 2003. Combined contributions of streptolysin O and streptolysin S to virulence of serotype M5 *Streptococcus pyogenes* strain Manfredo. Infection and Immunity **71**:3857–3865.

160. **Limbago B**, **Penumalli V**, **Weinrick B**, **Scott JR**. 2000. Role of streptolysin O in a mouse model of invasive group A streptococcal disease. Infection and

Immunity **68**:6384–6390.

161. **Sierig G**, **Cywes C**, **Wessels MR**, **Ashbaugh CD**. 2003. Cytotoxic effects of streptolysin O and streptolysin S enhance the virulence of poorly encapsulated group A streptococci. Infection and Immunity **71**:446–455.

162. **Datta V**, **Myskowski SM**, **Kwinn LA**, **Chiem DN**, **Varki N**, **Kansal RG**, **Kotb M**, **Nizet V**. 2005. Mutational analysis of the group A streptococcal operon encoding streptolysin S and its virulence role in invasive infection. Molecular Microbiology **56**:681–695.

163. **Lin A**, **Loughman JA**, **Zinselmeyer BH**, **Miller MJ**, **Caparon MG**. 2009. Streptolysin S inhibits neutrophil recruitment during the early stages of *Streptococcus pyogenes* infection. Infection and Immunity **77**:5190–5201.

164. **Rahman A-N**, **Bonsor DA**, **Herfst CA**, **Pollard F**, **Peirce M**, **Wyatt AW**, **Kasper KJ**, **Madrenas J**, **Sundberg EJ**, **McCormick JK**. 2011. The T cell receptor beta-chain second complementarity determining region loop (CDR2beta) governs T cell activation and Vbeta specificity by bacterial superantigens. Journal of Biological Chemistry **286**:4871–4881.

165. **Lintges M**, **van der Linden M**, **Hilgers RD**, **Arlt S**, **Lahham Al A**, **Reinert RR**, **Plücken S**, **Rink L**. 2010. Superantigen genes are more important than the *emm* type for the invasiveness of group A *Streptococcus* infection. J INFECT DIS **202**:20–28.

166. **Kasper KJ**, **Zeppa JJ**, **Wakabayashi AT**, **Xu SX**, **Mazzuca DM**, **Welch I**, **Baroja ML**, **Kotb M**, **Cairns E**, **Cleary PP**, **Haeryfar SMM**, **McCormick JK**. 2014. Bacterial superantigens promote acute nasopharyngeal infection by *Streptococcus pyogenes* in a human MHC class II-dependent manner. PLoS Pathog **10**:e1004155.

167. **LaPenta D**, **Rubens C**, **Chi E**, **Cleary PP**. 1994. Group A streptococci efficiently invade human respiratory epithelial cells. Proc Natl Acad Sci USA **91**:12115–12119.

168. **Greco R**, **De Martino L**, **Donnarumma G**, **Conte MP**, **Seganti L**, **Valenti P**. 1995. Invasion of cultured human cells by *Streptococcus pyogenes*. Research in Microbiology **146**:551–560.

169. **Osterlund A**, **Engstrand L**. 1995. Intracellular penetration and survival of *Streptococcus pyogenes* in respiratory epithelial cells *in vitro*. Acta Otolaryngol **115**:685–688.

170. **Hyland KA**, **Wang B**, **Cleary PP**. 2007. Protein F1 and *Streptococcus pyogenes* resistance to phagocytosis. Infection and Immunity **75**:3188–3191.

171. **Ozeri V**, **Rosenshine I**, **Ben-Ze'Ev A**, **Bokoch GM**, **Jou TS**, **Hanski E**. 2001. *De novo* formation of focal complex-like structures in host cells by invading streptococci. Molecular Microbiology **41**:561–573.

172. **Schaffner F**, **Ray A**, **Dontenwill M**. 2013. Integrin α5β1, the fibronectin receptor, as a pertinent therapeutic target in solid tumors. Cancers **5**:27–47.

173. **Molinari G**, **Talay SR**, **Valentin-Weigand P**, **Rohde M**, **Chhatwal GS**. 1997. The fibronectin-binding protein of *Streptococcus pyogenes*, SfbI, is involved in the internalization of group A streptococci by epithelial cells. Infection and Immunity **65**:1357–1363.

174. **Ozeri V**, **Rosenshine I**, **Mosher DF**, **Fässler R**, **Hanski E**. 1998. Roles of integrins and fibronectin in the entry of *Streptococcus pyogenes* into cells via

protein F1. Molecular Microbiology **30**:625–637.

175. **Cue D**, **Dombek PE**, **Lam H**, **Cleary PP**. 1998. *Streptococcus pyogenes* serotype M1 encodes multiple pathways for entry into human epithelial cells. Infection and Immunity **66**:4593–4601.

176. **Cue D**, **Southern SO**, **Southern PJ**, **Prabhakar J**, **Lorelli W**, **Smallheer JM**, **Mousa SA**, **Cleary PP**. 2000. A nonpeptide integrin antagonist can inhibit epithelial cell ingestion of *Streptococcus pyogenes* by blocking formation of integrin alpha 5beta 1-fibronectin-M1 protein complexes. Proc Natl Acad Sci USA **97**:2858–2863.

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

178. **Purushothaman SS**, **Wang B**, **Cleary PP**. 2003. M1 protein triggers a phosphoinositide cascade for group A *Streptococcus* invasion of epithelial cells. Infection and Immunity **71**:5823–5830.

179. **Nerlich A**, **Rohde M**, **Talay SR**, **Genth H**, **Just I**, **Chhatwal GS**. 2009. Invasion of endothelial cells by tissue-invasive M3 type group A streptococci requires Src kinase and activation of Rac1 by a phosphatidylinositol 3-kinase-independent mechanism. Journal of Biological Chemistry **284**:20319–20328.

180. **Molinari G**, **Rohde M**, **Guzman CA**, **Chhatwal GS**. 2000. Two distinct pathways for the invasion of *Streptococcus pyogenes* in non-phagocytic cells. Cellular Microbiology **2**:145–154.

181. **Rohde M**, **Müller E**, **Chhatwal GS**, **Talay SR**. 2003. Host cell caveolae act as an entry-port for group A streptococci. Cellular Microbiology **5**:323–342.

182. **Wang B**, **Li S**, **Dedhar S**, **Cleary PP**. 2007. Paxillin phosphorylation: bifurcation point downstream of integrin-linked kinase (ILK) in streptococcal invasion. Cellular Microbiology **9**:1519–1528.

183. **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.

184. **Hollingshead SK**, **Arnold J**, **Readdy TL**, **Bessen DE**. 1994. Molecular evolution of a multigene family in group A streptococci. Mol Biol Evol **11**:208–219.

185. **Bessen DE**, **Sotir CM**, **Readdy TL**, **Hollingshead SK**. 1996. Genetic correlates of throat and skin isolates of group A streptococci. J Infect Dis **173**:896–900.

186. **Scott JR**, **Cleary P**, **Caparon MG**, **Kehoe M**, **Heden L**, **Musser JM**, **Hollingshead S**, **Podbielski A**. 1995. New name for the positive regulator of the M protein of group A *Streptococcus*. Molecular Microbiology **17**:799.

187. **Bessen DE**, **Manoharan A**, **Luo F**, **Wertz JE**, **Robinson DA**. 2005. Evolution of transcription regulatory genes is linked to niche specialization in the bacterial pathogen *Streptococcus pyogenes*. Journal of Bacteriology **187**:4163–4172.

188. **Kalia A**, **Bessen DE**. 2003. Natural selection and evolution of streptococcal virulence genes involved in tissue-specific adaptations. Journal of Bacteriology **186**:110–121.

189. **Kratovac Z**, **Manoharan A**, **Luo F**, **Lizano S**, **Bessen DE**. 2007. Population genetics and linkage analysis of loci within the FCT region of *Streptococcus*

pyogenes. Journal of Bacteriology **189**:1299–1310.

190. **Kreikemeyer B**, **Martin DR**, **Chhatwal GS**. 1999. SfbII protein, a fibronectin binding surface protein of group A streptococci, is a serum opacity factor with high serotype-specific apolipoproteinase activity. FEMS Microbiology Letters **178**:305–311.

191. **Saravani GA**, **Martin DR**. 1990. Opacity factor from group A streptococci is an apoproteinase. FEMS Microbiology Letters **56**:35–39.

192. **Courtney HS**, **Zhang YM**, **Frank MW**. 2006. Serum opacity factor, a streptococcal virulence factor that binds to apolipoproteins A-I and A-II and disrupts high density lipoprotein structure. Journal of Biological … **281**:5515–5521.

193. **Meehl MA**, **Pinkner JS**, **Anderson PJ**, **Hultgren SJ**, **Caparon MG**. 2005. A novel endogenous inhibitor of the secreted streptococcal NAD-glycohydrolase. PLoS Pathog **1**:e35.

194. **Riddle DJ**, **Bessen DE**, **Caparon MG**. 2010. Variation in *Streptococcus pyogenes* NAD+ glycohydrolase is associated with tissue tropism. Journal of Bacteriology **192**:3735–3746.

195. **Bessen DE**, **Izzo MW**, **Fiorentino TR**, **Caringal RM**, **Hollingshead SK**, **Beall B**. 1999. Genetic linkage of exotoxin alleles and emm gene markers for tissue tropism in group A streptococci. J Infect Dis **179**:627–636.

196. **Kaplan EL**, **Bisno AL**. 2006. Antecedent streptococcal infection in acute rheumatic fever. CLIN INFECT DIS **43**:690–692.

197. **Parks T**, **Smeesters PR**, **Steer AC**. 2012. Streptococcal skin infection and rheumatic heart disease. Current Opinion in Infectious Diseases **25**:145–153.

198. **Carapetis JR**, **Currie BJ**, **Good MF**. 1996. Towards understanding the pathogenesis of rheumatic fever. Scand J Rheumatol **25**:127–131.

199. **McDonald M**, **Currie BJ**, **Carapetis JR**. 2004. Acute rheumatic fever: a chink in the chain that links the heart to the throat? The Lancet Infectious Diseases **4**:240–245.

200. **Brown A**, **Purton L**, **Schaeffer G**, **Wheaton G**, **White A**. 2003. Central Australian rheumatic heart disease control program. healthntgovau.

201. **Bessen DE**, **Carapetis JR**, **Beall B**, **Katz R**, **Hibble M**, **Currie BJ**, **Collingridge T**, **Izzo MW**, **Scaramuzzino DA**, **Sriprakash KS**. 2000. Contrasting molecular epidemiology of group A streptococci causing tropical and nontropical infections of the skin and throat. J Infect Dis **182**:1109–1116.

202. **McDonald MI**, **Towers RJ**, **Andrews RM**, **Benger N**, **Currie BJ**, **Carapetis JR**. 2006. Low rates of streptococcal pharyngitis and high rates of pyoderma in Australian aboriginal communities where acute rheumatic fever is hyperendemic. Clin Infec Dis **43**:683–689.

203. **Hartas J**, **Goodfellow AM**, **Currie BJ**, **Sriprakash KS**. 1995. Characterisation of group A streptococcal isolates from tropical Australia with high prevalence of rheumatic fever: probing for signature sequences to identify members of the family of serotype 5. Microbial Pathogenesis **18**:345–354.

204. **McDonald MI**, **Towers RJ**, **Fagan P**, **Carapetis JR**, **Currie BJ**. 2007. Molecular typing of *Streptococcus pyogenes* from remote Aboriginal communities where rheumatic fever is common and pyoderma is the predominant streptococcal infection. Epidemiol Infect **135**:1398–1405.

205. **Jaine R**, **Baker M**, **Venugopal K**. 2008. Epidemiology of acute rheumatic fever in New Zealand 1996-2005. J Paediatr Child Health **44**:564–571.

206. **Milne RJ**, **Lennon DR**, **Stewart JM**, **Vander Hoorn S**, **Scuffham PA**. 2012. Incidence of acute rheumatic fever in New Zealand children and youth. J Paediatr Child Health **48**:685–691.

207. **Kurahara D**, **Tokuda A**, **Grandinetti A**, **Najita J**, **Ho C**, **Yamamoto K**, **Reddy DV**, **Macpherson K**, **Iwamuro M**, **Yamaga K**. 2002. Ethnic differences in risk for pediatric rheumatic illness in a culturally diverse population. J Rheumatol **29**:379–383.

208. **Kurahara DK**, **Grandinetti A**, **Galario J**, **Reddy DV**, **Tokuda A**, **Langan S**, **Tanabe B**, **Yamamoto KS**, **Yamaga KM**. 2006. Ethnic differences for developing rheumatic fever in a low-income group living in Hawaii. Ethnicity & disease **16**:357.

209. **Reid HF**, **Bassett DC**, **Gaworzewska E**, **Colman G**, **Poon-King T**. 1990. Streptococcal serotypes newly associated with epidemic post-streptococcal acute glomerulonephritis. Journal of Medical Microbiology **32**:111–114.

210. **Cusick MF**, **Libbey JE**, **Fujinami RS**. 2011. Molecular mimicry as a mechanism of autoimmune disease. Clinic Rev Allerg Immunol **42**:102–111.

211. **Guilherme L**, **Kalil J**, **Cunningham M**. 2006. Molecular mimicry in the autoimmune pathogenesis of rheumatic heart disease. Autoimmunity **39**:31–39.

212. **Oldstone MBA**. 2014. Molecular mimicry: Its evolution from concept to mechanism as a cause of autoimmune diseases. Monoclonal Antibodies in Immunodiagnosis and Immunotherapy 140402093725003.

213. **Zabriskie JB**, **Hsu KC**, **Seegal BC**. 1970. Heart-reactive antibody associated with rheumatic fever: characterization and diagnostic significance. Clin Exp Immunol **7**:147–159.

214. **Kaplan MH**, **Svec KH**. 1964. Immunologic relation of streptococcal and tissue antigens III. Presence in human sera of streptococcal antibody cross-reactive with heart tissue. Association with streptococcal infection, rheumatic fever, and glomerulonephritis. J Exp Med **119**:651–666.

215. **Gulizia JM**, **Cunningham MW**, **McManus BM**. 1991. Immunoreactivity of anti-streptococcal monoclonal antibodies to human heart valves. Evidence for multiple cross-reactive epitopes. The American Journal of Pathology **138**:285–301.

216. **Galvin JE**, **Hemric ME**, **Ward K**, **Cunningham MW**. 2000. Cytotoxic mAb from rheumatic carditis recognizes heart valves and laminin. J Clin Invest **106**:217–224.

217. **Goldstein M**, **Halpern B**, **Robert L**. 1967. Immunological relationship between *Streptococcus* A polysaccharide and the structural glycoproteins of heart valve. Nature **213**:44–47.

218. **Quinn A**, **Ward K**, **Fischetti VA**, **Hemric M**, **Cunningham MW**. 1998. Immunological relationship between the class I epitope of streptococcal M protein and myosin. Infection and Immunity **66**:4418–4424.

219. **Faé KC**, **da Silva DD**, **Oshiro SE**, **Tanaka AC**, **Pomerantzeff PMA**, **Douay C**, **Charron D**, **Toubert A**, **Cunningham MW**, **Kalil J**, **Guilherme L**. 2006. Mimicry in recognition of cardiac myosin peptides by heart-intralesional T cell clones from rheumatic heart disease. J Immunol **176**:5662–5670.

220. **Ellis NMJ**, **Kurahara DK**, **Vohra H**, **Mascaro-Blanco A**, **Erdem G**, **Adderson EE**, **Veasy LG**, **Stoner JA**, **Tam E**, **Hill HR**, **Yamaga K**, **Cunningham MW**. 2010. Priming the immune system for heart disease: a perspective on group A streptococci. J INFECT DIS **202**:1059–1067.

221. **Wang Y-C**, **Tsai F-C**, **Chu J-J**, **Lin P-J**. 2008. Midterm outcomes of rheumatic mitral repair versus replacement. Int Heart J **49**:565–576.

222. **Root-Bernstein R**. 2014. Rethinking molecular mimicry in rheumatic heart disease and autoimmune myocarditis: laminin, collagen IV, CAR, and B1AR as initial targets of disease. Front Pediatr **2**:85.

223. **Tandon R**, **Sharma M**, **Chandrashekhar Y**, **Kotb M**, **Yacoub MH**, **Narula J**. 2013. Revisiting the pathogenesis of rheumatic fever and carditis. Nat Rev Cardiol **10**:171–177.

224. **Michael AF Jr**, **Drummond KN**, **Good RA**, **Vernier RL**. 1966. Acute poststreptococcal glomerulonephritis: immune deposit disease. Journal of Clinical Investigation **45**:237.

225. **Fillit H**, **Damle SP**, **Gregory JD**, **Volin C**, **Poon-King T**, **Zabriskie J**. 1985. Sera from patients with poststreptococcal glomerulonephritis contain antibodies to glomerular heparan sulfate proteoglycan. J Exp Med **161**:277–289.

226. **Kefalides NA**, **Pegg MT**, **Ohno N**, **Poon-King T**, **Zabriskie J**, **Fillit H**. 1986. Antibodies to basement membrane collagen and to laminin are present in sera from patients with poststreptococcal glomerulonephritis. J Exp Med **163**:588–602.

227. **Batsford SR**, **Mezzano S**, **Mihatsch M**, **Schiltz E**, **Rodríguez-Iturbe B**. 2005. Is the nephritogenic antigen in post-streptococcal glomerulonephritis pyrogenic exotoxin B (SpeB) or GAPDH? Kidney Int **68**:1120–1129.

228. **Yamakami K**, **Yoshizawa N**, **Wakabayashi K**, **Takeuchi A**, **Tadakuma T**, **Boyle MDP**. 2000. The potential role for nephritis-associated plasmin receptor in acute poststreptococcal glomerulonephritis. Methods **21**:185–197.

229. **Yoshizawa N**, **Yamakami K**, **Fujino M**, **Oda T**, **Tamura K**, **Matsumoto K**, **Sugisaki T**, **Boyle MDP**. 2004. Nephritis-associated plasmin receptor and acute poststreptococcal glomerulonephritis: characterization of the antigen and associated immune response. Journal of the American Society of Nephrology **15**:1785–1793.

230. **Oda T**, **Yamakami K**, **Omasu F**, **Suzuki S**, **Miura S**, **Sugisaki T**, **Yoshizawa N**. 2005. Glomerular plasmin-like activity in relation to nephritis-associated plasmin receptor in acute poststreptococcal glomerulonephritis. Journal of the American Society of Nephrology **16**:247–254.

231. **Takashi O**, **Yoshizawa N**, **Yamakami K**, **Tamura K**, **Kuroki A**, **Sugisaki T**, **Sawanobori E**, **Higashida K**, **Ohtomo Y**, **Hotta O**, **Kumagai H**, **Miura S**. 2010. Localization of nephritis-associated plasmin receptor in acute poststreptococcal glomerulonephritis. Human Pathology **41**:1276–1285.

232. **Oda T**, **Yoshizawa N**, **Yamakami K**, **Sakurai Y**, **Takechi H**, **Yamamoto K**, **Oshima N**, **Kumagai H**. 2012. The role of nephritis-associated plasmin receptor (NAPlr) in glomerulonephritis associated with streptococcal infection. Journal of Biomedicine and Biotechnology **2012**:1–9.

233. **Poon-King R**, **Bannan J**, **Viteri A**, **Cu G**, **Zabriskie JB**. 1993. Identification of an extracellular plasmin binding protein from nephritogenic streptococci. J Exp

Med **178**:759–763.

234. **Rodriguez-Iturbe B**, **Musser JM**. 2008. The current state of poststreptococcal glomerulonephritis. Journal of the American Society of Nephrology **19**:1855–1864.

235. **Luo YH**, **Kuo CF**, **Huang KJ**, **Wu JJ**, **Lei HY**, **Lin MT**, **Chuang WJ**, **Liu CC**, **Lin CF**, **Lin YS**. 2007. Streptococcal pyrogenic exotoxin B antibodies in a mouse model of glomerulonephritis. Kidney Int **72**:716–724.

236. **Hensel M**, **Shea JE**, **Gleeson C**, **Jones MD**, **Dalton E**, **Holden DW**. 1995. Simultaneous identification of bacterial virulence genes by negative selection. Science **269**:400–403.

237. **Chiang SL**, **Mekalanos JJ**, **Holden DW**. 1999. In vivo genetic analysis of bacterial virulence. Annu Rev Microbiol **53**:129–154.

238. **Rollins SM**, **Peppercorn A**, **Hang L**, **Hillman JD**, **Calderwood SB**, **Handfield M**, **Ryan ET**. 2004. *In vivo* induced antigen technology (IVIAT). Cellular Microbiology **7**:1–9.

239. **Kizy AE**, **Neely MN**. 2009. First *Streptococcus pyogenes* signature-tagged mutagenesis screen identifies novel virulence determinants. Infection and Immunity **77**:1854–1865.

240. **Cummins J**, **Casey PG**, **Joyce SA**, **Gahan CGM**. 2013. A mariner transposon-based signature-tagged mutagenesis system for the analysis of oral infection by *Listeria monocytogenes*. PLoS ONE **8**:e75437.

241. **Jamet A**, **Euphrasie D**, **Martin P**, **Nassif X**. 2013. Identification of genes involved in *Neisseria meningitidis* colonization. Infection and Immunity **81**:3375–3381.

242. **Hunt TA**, **Kooi C**, **Sokol PA**, **Valvano MA**. 2004. Identification of *Burkholderia cenocepacia* genes required for bacterial survival *in vivo*. Infection and Immunity **72**:4010–4022.

243. **Mei JM**, **Nourbakhsh F**, **Ford CW**, **Holden DW**. 1997. Identification of *Staphylococcus aureus* virulence genes in a murine model of bacteraemia using signature-tagged mutagenesis. Molecular Microbiology **26**:399–407.

244. **Jones AL**, **Knoll KM**, **Rubens CE**. 2000. Identification of *Streptococcus agalactiae* virulence genes in the neonatal rat sepsis model using signature-tagged mutagenesis. Molecular Microbiology **37**:1444–1455.

245. **Schena M**, **Shalon D**, **Davis RW**, **Brown PO**. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science **270**:467–470.

246. **Butte A**. 2002. The use and analysis of microarray data. Nat Rev Drug Discov **1**:951–960.

247. **Miller MB**, **Tang YW**. 2009. Basic concepts of microarrays and potential applications in clinical microbiology. Clin Microbiol Rev **22**:611–633.

248. **Kreil DP**, **Russell RR**, **Russell S**. 2006. Microarray oligonucleotide probes. Meth Enzymol **410**:73–98.

249. **Malone JH**, **Oliver B**. 2011. Microarrays, deep sequencing and the true measure of the transcriptome. BMC Biol **9**:34.

250. **Sasik R**, **Woelk CH**, **Corbeil J**. 2004. Microarray truths and consequences. Journal of molecular endocrinology **33**:1–9.

251. **Hinton J**, **Hautefort I**, **Eriksson S**, **Thompson A**, **Rhen M**. 2004. Benefits and

pitfalls of using microarrays to monitor bacterial gene expression during infection. Current Opinion in Microbiology **7**:277–282.

252. **Wang Z**, **Gerstein M**, **Snyder M**. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet **10**:57–63.

253. **Nagalakshmi U**, **Waern K**, **Snyder M**. 2010. RNA-seq: a method for comprehensive transcriptome analysis. Current Protocols in Molecular Biology 4.11. 1–4.11. 13.

254. **Auer PL**, **Doerge RW**. 2010. Statistical design and analysis of RNA sequencing data. Genetics **185**:405–416.

255. **Valdivia RH**, **Falkow S**. 1996. Bacterial genetics by flow cytometry: rapid isolation of *Salmonella typhimurium* acid-inducible promoters by differential fluorescence induction. Molecular Microbiology **22**:367–378.

256. **Bartilson M**, **Marra A**, **Christine J**, **Asundi JS**, **Schneider WP**, **Hromockyj AE**. 2001. Differential fluorescence induction reveals *Streptococcus pneumoniae* loci regulated by competence stimulatory peptide. Molecular Microbiology **39**:126–135.

257. **Marra A**, **Asundi J**, **Bartilson M**, **Lawson S**, **Fang F**, **Christine J**, **Wiesner C**, **Brigham D**, **Schneider WP**, **Hromockyj AE**. 2002. Differential fluorescence induction analysis of *Streptococcus pneumoniae* identifies genes involved in pathogenesis. Infection and Immunity **70**:1422–1433.

258. **Schneider WP**, **Ho SK**, **Christine J**, **Yao M**, **Marra A**, **Hromockyj AE**. 2002. Virulence gene identification by differential fluorescence induction analysis of *Staphylococcus aureus* gene expression during infection-simulating culture. Infection and Immunity **70**:1326–1333.

259. **Wilson RL**, **Tvinnereim AR**, **Jones BD**, **Harty JT**. 2001. Identification of *Listeria monocytogenes in vivo*-induced genes by fluorescence-activated cell sorting. Infection and Immunity **69**:5016–5024.

260. **Hang L**, **John M**, **Asaduzzaman M**, **Bridges EA**, **Vanderspurt C**, **Kirn TJ**, **Taylor RK**, **Hillman JD**, **Progulske-Fox A**, **Handfield M**, **Ryan ET**, **Calderwood SB**. 2003. Use of *in vivo*-induced antigen technology (IVIAT) to identify genes uniquely expressed during human infection with *Vibrio cholerae*. Proc Natl Acad Sci USA **100**:8508–8513.

261. **Salim KY**, **Cvitkovitch DG**, **Chang P**, **Bast DJ**, **Handfield M**, **Hillman JD**, **de Azavedo JCS**. 2005. Identification of group A *Streptococcus* antigenic determinants upregulated *in vivo*. Infection and Immunity **73**:6026–6038.

262. **Li S**, **Song J**, **Huang H**, **Chen W**, **Li M**, **Zhao Y**, **Cong Y**, **Zhu J**, **Rao X**, **Hu X**, **Hu F**. 2013. Identification of *in-vivo* induced genes of *Streptococcus suis* serotype 2 specially expressed in infected human. Microbial Pathogenesis **63**:8–15.

263. **Kizy AE**. 2008. A genetic exploration reveals novel *Streptococcus pyogenes* virulence factors important for *in vivo* survival and biofilm formation, p. 160. *In* Neely, MN (ed.). Digital Commons@Wayne State University, Detroit.

264. **Tesorero RA**, **Yu N**, **Wright JO**, **Svencionis JP**, **Cheng Q**, **Kim J-H**, **Cho KH**. 2013. Novel regulatory small RNAs in *Streptococcus pyogenes*. PLoS ONE **8**:e64021.

265. **Deltcheva E**, **Chylinski K**, **Sharma CM**, **Gonzales K**, **Chao Y**, **Pirzada ZA**, **Eckert MR**, **Vogel J**, **Charpentier E**. 2011. CRISPR RNA maturation by trans-

encoded small RNA and host factor RNase III. Nature **471**:602–607.

266. **Graham MR**, **Virtaneva K**, **Porcella SF**, **Barry WT**, **Gowen BB**, **Johnson CR**, **Wright FA**, **Musser JM**. 2005. Group A Streptococcus transcriptome dynamics during growth in human blood reveals bacterial adaptive and survival strategies. The American Journal of Pathology **166**:455.

267. **Shelburne SA**, **Sumby P**, **Sitkiewicz I**, **Granville C**, **DeLeo FR**, **Musser JM**. 2005. Central role of a bacterial two-component gene regulatory system of previously unknown function in pathogen persistence in human saliva. Proc Natl Acad Sci USA **102**:16037–16042.

268. **Smoot LM**, **Smoot JC**, **Graham MR**, **Somerville GA**, **Sturdevant DE**, **Migliaccio CA**, **Sylva GL**, **Musser JM**. 2001. Global differential gene expression in response to growth temperature alteration in group A *Streptococcus*. Proc Natl Acad Sci USA **98**:10416–10421.

269. **Graham MR**, **Virtaneva K**, **Porcella SF**, **Gardner DJ**, **Long RD**, **Welty DM**, **Barry WT**, **Johnson CA**, **Parkins LD**, **Wright FA**, **Musser JM**. 2006. Analysis of the transcriptome of group A *Streptococcus* in mouse soft tissue infection. The American Journal of Pathology **169**:927–942.

270. **Sitkiewicz I**, **Musser JM**. 2006. Expression microarray and mouse virulence analysis of four conserved two-component gene regulatory systems in group A *Streptococcus*. Infection and Immunity **74**:1339–1351.

271. **Virtaneva K**, **Porcella SF**, **Graham MR**, **Ireland RM**, **Johnson CA**, **Ricklefs SM**, **Babar I**, **Parkins LD**, **Romero RA**, **Corn GJ**, **Gardner DJ**, **Bailey JR**, **Parnell MJ**, **Musser JM**. 2005. Longitudinal analysis of the group A *Streptococcus* transcriptome in experimental pharyngitis in cynomolgus macaques. Proc Natl Acad Sci USA **102**:9014–9019.

272. **Shelburne SA**, **Granville C**, **Tokuyama M**, **Sitkiewicz I**, **Patel P**, **Musser JM**. 2005. Growth characteristics of and virulence factor production by group A *Streptococcus* during cultivation in human saliva. Infection and Immunity **73**:4723–4731.

273. **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 Pathog **2**:e5.

274. **Sumby P**, **Zhang S**, **Whitney AR**, **Falugi F**, **Grandi G**, **Graviss EA**, **DeLeo FR**, **Musser JM**. 2008. A chemokine-degrading extracellular protease made by group A *Streptococcus* alters pathogenesis by enhancing evasion of the innate immune response. Infection and Immunity **76**:978–985.

275. **Shelburne SA**, **Keith D**, **Horstmann N**, **Sumby P**, **Davenport MT**, **Graviss EA**, **Brennan RG**, **Musser JM**. 2008. A direct link between carbohydrate utilization and virulence in the major human pathogen group A *Streptococcus*. Proc Natl Acad Sci USA **105**:1698–1703.

276. **Shelburne SA**, **Sahasrobhajane P**, **Suber B**, **Keith DB**, **Davenport MT**, **Horstmann N**, **Kumaraswami M**, **Olsen RJ**, **Brennan RG**, **Musser JM**. 2011. Niche-specific contribution to streptococcal virulence of a MalR-regulated carbohydrate binding protein. Molecular Microbiology **81**:500–514.

277. **Lowe AM**, **Beattie DT**, **Deresiewicz RL**. 1998. Identification of novel staphylococcal virulence genes by *in vivo* expression technology. Molecular Microbiology **27**:967–976.

278. **Saviola B**, **Woolwine SC**, **Bishai WR**. 2003. Isolation of acid-inducible genes of *Mycobacterium tuberculosis* with the use of recombinase-based *in vivo* expression technology. Infection and Immunity **71**:1379–1388.

279. **Bron PA**, **Grangette C**, **Mercenier A**, **de Vos WM**, **Kleerebezem M**. 2004. Identification of *Lactobacillus plantarum* genes that are induced in the gastrointestinal tract of mice. Journal of Bacteriology **186**:5721–5729.

280. **Frank KL**, **Barnes AMT**, **Grindle SM**, **Manias DA**, **Schlievert PM**, **Dunny GM**. 2012. Use of recombinase-based *in vivo* expression technology to characterize *Enterococcus faecalis* gene expression during infection identifies *in vivo*-expressed antisense RNAs and implicates the protease Eep in pathogenesis. Infection and Immunity **80**:539–549.

281. **Holt JF**, **Kiedrowski MR**, **Frank KL**, **Du J**, **Guan C**, **Broderick NA**, **Dunny GM**, **Handelsman J**. 2014. *Enterococcus faecalis* 6-phosphogluconolactonase is required for both commensal and pathogenic interactions with *Manduca sexta*. Infection and Immunity.

282. **Junjua M**, **Galia W**, **Gaci N**, **Ophélie U**, **Genay M**, **Bachmann H**, **Kleerebezem M**, **Dary A**, **Roussel Y**. 2013. Development of the recombinase-based *in vivo* expression technology in *Streptococcus thermophilus* and validation using the lactose operon promoter. J Appl Microbiol.

CHAPTER 2: RECOMBINASE-BASED *IN VIVO* EXPRESSION TECHNOLOGY IN *STREPTOCOCCUS PYOGENES*

2.1 Introduction

2.1.1 *In vivo* expression technology

IVET was developed as a method to identify bacterial genes that are required in order to circumvent the immune system of the host and cause disease. The presumption was that environmental cues, differently represented in various host tissues, control the expression of specific genes (1). The first type of IVET system was developed utilizing a *purA* auxotrophic strain of *S. typhimurium,* which has greatly attenuated growth in minimal media and *in vivo*. Therefore, in order to survive *in vivo, purA* would have to be expressed. By cloning genome fragments prior to *purA*, *in vivo* induced promoters could be identified from clones that survived growth *in vivo*. Briefly, *S. typhimurium* genomic DNA was partially digested into various sizes and cloned into a suicide vector, creating a transcriptional fusion with both *purA* and *lacZY*, potentially controlling both genes. After transformation, the vectors were incorporated into the chromosome. The pool of clones was then passed through the chosen mouse model, recovered, and plated on rich media in order to separate constitutive promoters from those that were induced *in vivo*. Clones with constitutive promoters still expressed *lacZY* and were identifiable by a colourimetric assay. Those showing no colour, and therefore had no *lacZY* expression, were only active *in vivo*. These clones were also verified by lack of growth in minimal media, showing that *purA* was not expressed. In order to confirm *in vivo* activity, *in vitro* versus *in vivo lacZ* expression was measured by inoculating single clones into mice (1). This was later developed to be used by *Pseudomonas aeruginosa* utilizing a purine deficient strain. Clones were identified post-*in vivo* growth by finding smaller sized colonizes on media with a minimal amount of the required purine. This system was used to identify *in vivo*-induced genes in a mouse, and in human cystic fibrosis mucus (2, 3).

While this initial technology was successful, it was limited to the complementation of a purine auxotrophic strain. To expand the system, the *purA* gene was replaced with a chloramphenicol resistance marker. As before, a pool of plasmids containing random chromosomal fragments was created and cloned prior to the promoterless chloramphenicol gene. The pool of plasmids was then amplified and transferred into *S.*

typhimurium via conjugal transfer utilizing *E. coli*. Since the plasmid cannot replicate in *S. typhimurium*, only cells with chromosomal integrations could survive. The cells were then injected into mice, after which chloramphenicol was injected twice daily and added to drinking water. Once colonies were recovered, *in vivo* active promoters were those that grew *in vitro* showing no colour, and therefore did not express *lacZ* and could no longer resist chloramphenicol, due to the silencing of the *in vivo*-induced promoters. One noticeable problem was the cloning of multiple fragments into some plasmids. Therefore, potential clones were all verified via sequencing to ensure that only the fragment immediately to the 5' end of the *cat-lac* fusion was used for integration into the chromosome. In addition to a mouse model, a macrophage model was also used successfully as the *in vivo* system (4). Both of these IVET systems were used to identify *S. typhimurium* genes in various *in vivo* models (5), along with *Yersinia enterocolitica* (6, 7), demonstrating the versatility of the IVET model.

2.1.2 Recombinase-based *in vivo* expression technology

Over the years, IVET technology has been continually improved. A major limitation of the assay was sensitivity, as the system required continual expression of a gene to allow the survival of the clone, and therefore any promoters demonstrating any short bursts of expression would not be detected. In order to overcome this, a reporter system utilizing a heritable change was used to replace purine autotrophy (8, 9). The chromosomal library was cloned in front of a promoterless resolvase gene (*tnpR*), which recognizes *res* sites on DNA. TnpR mediates strand exchange between two *res* sites causing excision of the DNA in a non-reversible reaction, termed resolution. In order to make use of this system, a tetracycline marker was placed between two *res* sites, termed a cassette, and inserted into the chromosome of *V. cholera*. The chromosomal library was then cloned upstream of *tnpR* in plasmids that were then also integrated into the chromosome. Clones were first grown *in vitro,* followed by plating on tetracycline to eliminate those that contained constitutive promoters. Therefore, all colonies that were able to grow contained inactive promoters (or fragments not containing promoters) and were picked, pooled, and used to challenge mice. After recovery, patch plating was performed to recover tetracycline sensitive clones, and therefore those with *in vivo* active promoters (8, 9). This method

was a vast improvement over other IVET models due to the fact that it eliminated the *in vitro* promoters before the *in vivo* selection. This also showed the first discovery of antisense RNA transcripts utilizing IVET (9).

Lowe *at al*. performed the first IVET in a Gram-positive species, *S. aureus*, several years later (10). The chromosomal library was cloned in front of a promoterless *tnpR* and transformed into a strain containing an integrated kanamycin marker flanked by *res* sites. Without making use of a colourimetric assay, *in vitro* promoters were first eliminated by growth in the presence of kanamycin. Following a murine abscess model, bacteria were recovered and *in vivo* induced promoters were identified by kanamycin sensitivity via colony patching (10). Continual use and evolution of IVET beyond this point started to cause a change in the name of the system. IVET only referred to the earlier scheme in which the library fragments were cloned in front of a purine or antibiotic gene, while the system utilizing library fragments to express *tnpR* to excise a chromosomal fragment between *res* sites was now referred to as recombinase-based *in vivo* expression technology or RIVET (11-15).

There has been one additional major development in the RIVET system, which is now called second generation RIVET, that enhances post-*in vivo* screening. First performed in *M. tuberculosis*, a counter-selectable gene was included within the cassette (16). Counter-selection works by converting a non-toxic pro-drug into a toxic form, causing death. Meanwhile, the absence of the gene means the non-toxic pro-drug remains in its native form, and has no effect. In this case, the *sacB* gene causes sensitivity to sucrose, presumably due to the accumulation of levans (17). In order to identify promoters, a *M. tuberculosis* chromosomal library was cloned upstream of *tnpR*, and this was transformed into *Mycobacterium smegmantis* containing a *res-kan-sacB-res* cassette within the chromosome. Growth *in vitro* in the presence of kanamycin eliminated all *in vitro* expressing promoters and the remaining clones were then grown *in vivo*, recovered, and then grown on sucrose. All clones that contained *in vivo* induced promoters expressed *tnpR*, resolved the cassette, and were not affected by sucrose due to the loss of *sacB* (16). Second generation RIVET has been used with numerous Gram-positive species with *sacB* and uracil phosphoribosyl-transferase (*upp*) as popular counter-selection methods (16, 18-

20).

2.1.3 Utilization of the RIVET method in *S. pyogenes*

Based on the second generation RIVET system as described, a cassette was to be constructed for *S. pyogenes* consisting of two selection markers for pre- and post-*in vivo* selection **(Figure 2.1)**. However, the Cre recombination system was applied in this work due to its immediate availability and its previous successful use with *Streptococcus gordonii* in our laboratory. The Cre recombinase recognizes *loxP* sites to cause the permanent resolution event. The *cre* gene comes from the *E. coli* P1 bacteriophage and, together with the *loxP* site, is important in the viral life cycle. The P1 DNA must be circularized before being packaged, and so *loxP* sites at either end of the linear DNA strand are recombined via Cre to form a circular DNA strand (21). Therefore, similar to the *tnpR-res* system, *loxP* sites flanking selectable markers in the chromosome can be removed, or resolved, with the expression of *cre* within the cell. A single *loxP* site contains two 13-nucleotide inverted repeats separated by an 8-bp spacer (bold and underlined) making a 34 bp site (ATAACTTCGTATA**ATGTATGC**TATACGAAGTTAT) (21, 22). Two Cre molecules bind to each *loxP* site to perform the reaction, and if the *loxP* sites are oriented in the same direction, the DNA between the sites becomes excised via recombination. However, when the *loxP* sites are oriented in opposite directions, the DNA between the sites is simply inverted, and would remain within the chromosome. To perform the reaction it has been found that no energy co-factors are required, only buffers and simple salts (21).

In order to gain further understanding by how *S. pyogenes* is able to both colonize and compete within the nasopharynx, it is important to understand how *S. pyogenes* adapts and alters gene expression in this *in vivo* environment. Herein, a RIVET system was engineered and constructed, tested, and used for the first time with *S. pyogenes* in an established mouse pharyngeal model to provide clues as to how this important bacterial pathogen survives within this limited biological niche.

Figure 2.1: Overview of the RIVET design

A) A genetic RIVET cassette was inserted into the genome of *S. pyogenes*. The cassette contained an antibiotic resistance (tetracycline, *tet^R*) marker and a counter-selection gene (*thymidine kinase, tk*). B) Random genomic fragments from *S. pyogenes* were cloned upstream of a promoterless *cre* gene, and the plasmids (erythromycin resistant, *erm^R*) were transformed into *S. pyogenes* containing the cassette. C) The *S. pyogenes* promoter library was incubated *in vitro* under erythromycin and tetracycline selection. Erythromycin retains the plasmids and tetracycline eliminates any *in vitro* active promoters, as any active promoters will express *cre*, cause excision of the cassette, and the loss of tetracycline resistance. The pool of remaining clones was used in an *in vivo* model of colonization. Clones containing *in vivo* active promoters express *cre* and resolve the cassette. Clones were then recovered via patching for loss of *tet^R*. Initially, counter-selection, the conversion of a non-toxic pro-drug to a toxic form, was attempted, however it proved to be unreliable. Sequencing and bioinformatics were subsequently used to identify the *in vivo* induced fragments and the associated gene(s). False positives were ruled out by transforming single plasmids back into cassette-containing *S. pyogenes* and grown *in vitro* and *in vivo*.
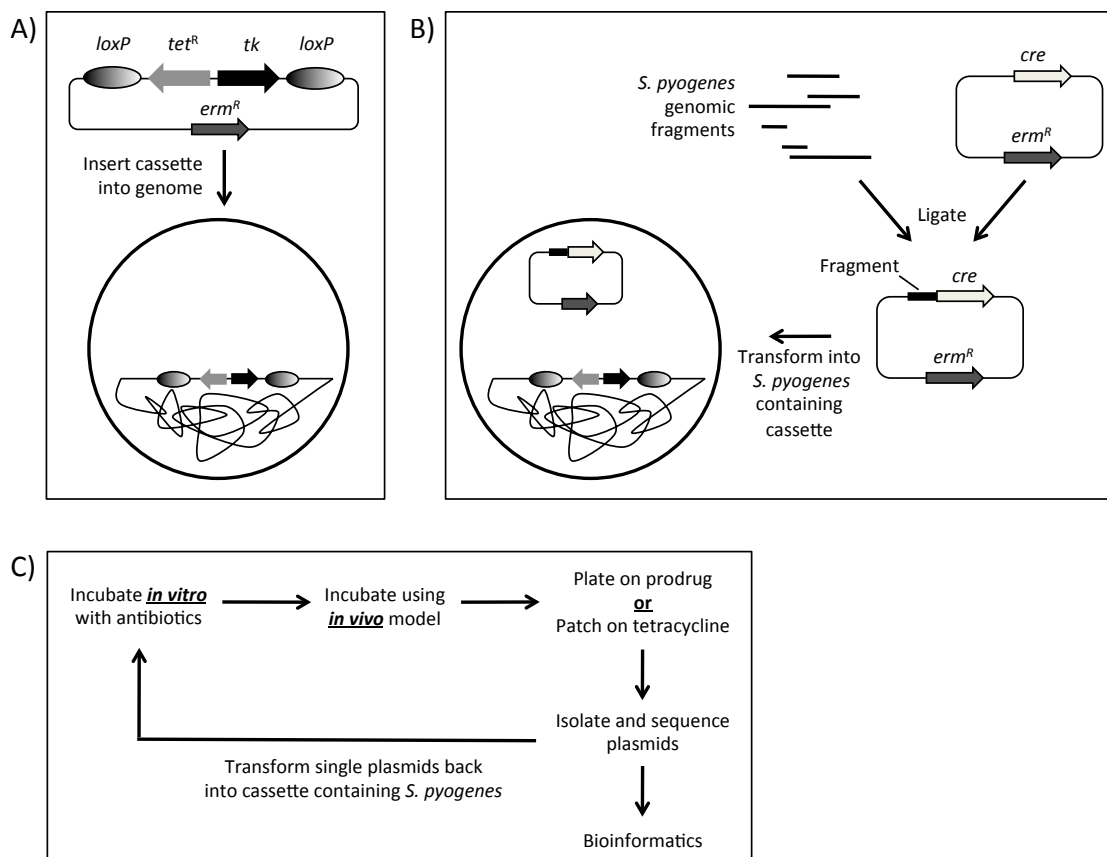
2.2 Materials and Methods

2.2.1 Bacteria and growth conditions

Bacterial strains used in this study are listed in **(Table 2.1)**. All cloning was carried out using *E. coli* XL1-Blue grown on Brain Heart Infusion (BHI) media (BD Biosciences, Franklin Lakes, NJ, USA) media supplemented with 1.5% agar (BD Biosciences, Franklin Lakes, NJ, USA) or Luria Burtani (LB) broth (BD Biosciences, Franklin Lakes, NJ, USA) broth shaking at 250 rpm. Erythromycin (erm) was supplemented at 150 μg mL^{-1} as required. *S. pyogenes* MGAS8232 is an M18 serotype isolated from a rheumatic fever patient in Utah in 1987 with a genome of 1.985 Mbp (NC_003485.1) (Smoot et al., 2002). This strain was grown statically in Todd Hewitt broth (BD Biosciences, Franklin Lakes, NJ, USA) with the addition of 1% yeast extract (THY) (EMD Millipore, Billerica, MA, USA) at 37°C. As appropriate, 1.5% agar (Invitrogen Life Sciences, Burlington, Ontario, Canada) and/or erm (1 μg mL^{-1}) and/or tetracycline (tet) (0.5 μg mL^{-1}) was added. *Bacillus subtilis* was grown statically overnight in BHI media at 37°C. All antibiotics were purchased from Acros Organics (Thermo Fischer Scientific, New Jersey, NJ, USA).

2.2.2 *S. pyogenes* total DNA extraction

Briefly, for total DNA extractions two mL of overnight culture was washed twice with one mL of 0.2 mM sodium acetate. The pellet was resuspended in 500 μL Tris EDTA Glucose buffer (10 mM Tris, 2 mM EDTA, 25% glucose), adding 5 μL 200 mg mL^{-1} lysozyme and 5 μL 10 U μL^{-1} mutanolysin. Cells were incubated for 1 hour (h) at 37°C. After centrifugation (21 000 × g), the pellet was resuspended in 500 μL lysis buffer (50 mM EDTA, 0.2% SDS) with 5 μL 20 mg mL^{-1} proteinase K, 5 μL 20 mg mL^{-1} RNase and incubated at 65°C for 2 h. Next, 50 μL 5 M potassium acetate was added, thoroughly mixed to precipitate proteins and centrifuged at 21 000 × g for 10 min. The supernatant was put into a new tube and mixed with 500 μL ice cold 95% ethanol to precipitate DNA. After one wash with 70% ethanol, DNA was dried and resuspended in 100 μL Qiagen Elution Buffer (Qiagen, Toronto, Ontario, Canada).

Table 2.1: Bacterial strains used in this study

Name	Genotype	Reference
E. coli Xl1-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F$^{'}$ *proAB lacIqZΔM15* Tn*10* (*tet^R*)]	Stratagene, USA
S. pyogenes MGAS8232	Isolated 1987, M18 serotype, 1.895 Mb, NC_003485.1	(23)
B. subtilis 168	4.216 Mb, NC_000964	(24)
S. pyogenes MGAS8232 Δ*upp*	MGAS8232 with *upp* removed from the chromosome	This study
S. pyogenes MGAS8232 Cas2	MGAS8232 containing the *loxP-tet^R-tk-loxP* cassette within the chromosome	This study

2.2.3 Polymerase chain reaction

All primers were designed using Primer3Plus (http://primer3plus.com/cgi-bin/dev/primer3plus.cgi) and are shown in **(Table 2.2)**. The provided melting temperature (T_m), calculated with sequences without the added restriction enzymes or other additional sequences, was used as the annealing temperature. Primers were ordered from Sigma-Aldrich (Oakville, Ontario, Canada) and resuspended in Milli'Q water at a 100 mM stock. All PCR was performed using the following cycle: 5 min at 95°C, [30 seconds (sec) 95°C denature, 1 min T_m°C anneal, 1 min per 0.8 kb of length 74°C extension] × 36 cycles, followed by 5 min at 74°C. All reactions utilized Pfu polymerase and Pfu buffer (20 mM Tris-HCl (pH 8.8), 10 mM KCl, 10 mM $(NH_4)SO_4$, 5 mM $MgSO_4$, 0.1 mg mL^{-1} bovine serum albumin, 0.1% Triton X-100) produced in house. Primers were used at a final concentration of 1 μM each and both Pfu and the DNA template were added at 1 μL per 100 μL of total reaction volume. All reactions were performed in a Peltier Thermocycler PTC-200 or a MJ Mini Gradient Thermal Cycler (Bio-Rad Laboratories Inc, Hercules, CA, USA).

2.2.4 DNA visualization

DNA was visualized on 0.8% w/v agarose (Invitrogen Life Sciences, Burlington, Ontario, Canada) gels. DNA was loaded using dye made in lab (5% w/v glycerol, 0.04% w/v bromophenol blue, 0.04% xylene cyanol) and electrophoresed for 1 h at 100 V. All gels were run with the 1 kb Plus DNA Ladder as a size standard (Invitrogen Life Sciences, Burlington, Ontario, Canada). Gels were stained utilizing ethidium bromide (0.1%) for 15 min and visualized under ultraviolet light.

2.2.5 Plasmid isolation, DNA digestion, and ligation

Plasmids were isolated from *E. coli* using the Qiagen Miniprep Kit (Qiagen, Toronto, Ontario, Canada) following the manufacturer's instructions. All plasmids used are listed in **(Table 2.3)**. Digestions were carried out utilizing restriction enzymes from New England Biolabs (Ipswich, Massachusetts, USA) or Roche (Mississauga, Ontario, Canada) following the manufacturer's instructions. All DNA was measured using a.

Table 2.2: Primers used in this study

Name	Sequence 5' to 3'[ab]	Restriction Site
upp Up For	CCC<u>TCTAGA</u>GTGGTGCCGATTTATTTGATG	*Xba*I
upp Up Rev	CCC<u>GGATCC</u>GCGAAATAACTTGACATTTTCC	*Bam*HI
upp Down For	CCC<u>CTGCAG</u>CTGGTGATCGTTTATTTGGGAC	*Pst*I
upp Down Rev	CCC<u>CTCGAG</u>CGGCCTTGATAAAGTTCATGG	*Xho*I
snPgyra for	CGC<u>GGATCC</u>GCAAAAGCTCATACGGTCTT	*Bam*HI
snPgyra rev	GCG<u>AAGCTT</u>CCATGGGATCTTGCATTTAAGGAATGCTC	*Nco*I
sacB For	CCC<u>GTCGAC</u>AGGAGGATGAAAGAAACGAACCAAAAGCCATA	*Sal*I
sacB Rev	CCC<u>CTGCAG</u>TTATTTGTTAACTGTTAATTGTCCTTGTTCAAGGATGC	*Pst*I
sacB RT for	GCGGATCAAAAATGACGATT	n/a
sacB RT rev	CCTTTCGCTTGAGGTACAGC	n/a
0442 RT for	ACTAGCGACGACGATTGCTT	n/a
0442 RT rev	ATAGCCAGTTGGCCACAAAC	n/a
upp RT for	CGGAGTGGTTCCTATCCTCA	n/a
upp RT rev	TTCACGCTCTTCCACATCAG	n/a
clpP RT for	CGTCGTCAAACCACATCAAC	n/a
clpP RT rev	ATCCCAATACCAGCACGAAG	n/a
tsf for	CGC<u>CTCGAG</u>ACTTGCTCAATTGAACCACG	*Xho*I
tsf rev loxp	GCG<u>ATCGAT</u>**ATAACTTCGTATAGCATACATTATACGAAGTTAT**CCGTTTTGACACAACAAAAGA	*Cla*I
tet for	CGC<u>GGATCC</u>AGATAAAAAGTTGATCTTTGTGAAAAC	*Bam*HI
tet rev	CGC<u>AAGCTT</u>TTAGAAATCCCTTTGAGAATGTTT	*Hin*dIII
tk pET for	GCG<u>CCATGG</u>CTTCATACCCATGTCA	*Nco*I
tk rev	CCC<u>TCTAGA</u>TTAGTTAGCTTCACCCATTTCACG	*Xba*I
pepO for loxp	GCG<u>TCTAGA</u>**ATAACTTCGTATAATGTATGCTATACGAAGTTAT**ACACCAATAAGGAAGCAAAAA	*Xba*I
pepO rev	GCG<u>CCGCGG</u>AGCCTAAATGATTGGTGGA	*Sac*II
tk RT for	GCTCCACCACCAGCTCTTAC	n/a
tk RT rev	GGTCGATGTGACGGTCTTCT	n/a
pepO RT for	ATTCTGAGCCTTCCTCACGA	n/a
pepO RT rev	CGAAGAAGGCAACGAAAAAG	n/a
tsf RT for	GGCGTTATGGACGCTAAAAA	n/a
tsf RT rev	TGCGTTTACCAATTCAACGA	n/a
cre F pTRK	GCG<u>CCTGCAG</u>AAGGAGGCACTCAAAATGTCCAATTTACTGACCGTA	*Pst*I
cre R pTRK	GCGC<u>CCCGGG</u>CTAATCGCCATCTTCCAGCAG	*Xma*I
cre F FOR pMSP	GCGC<u>TCTAGA</u>AAGGAGGCACTCAAAATCTCCAATTTACTGACCGTA	*Xba*I
cre F REV pMSP	GCGC<u>TCTAGA</u>AAGGAGGCACTCAAAATCTCCAATTTACTGACCGTA	*Xba*I
cre R REV pMSP	GCGC<u>CTGCAG</u>CTAATCGCCATCTTCCAGCAG	*Pst*I
cre R FOR pMSP	GCGC<u>CTGCAG</u>CTAATCGCCATCTTCCAGCAG	*Pst*I
M13 For	GTAAAACGACGGCCAGT	n/a
M13 Rev	CAGGAAACAGCTATGAC	n/a
a-restriction sites underlined; b-***loxP*** sites bolded		

Table 2.3: Plasmids used in this study

Name	Notes	Source/Reference
pTRKL2	erm^R low copy shuttle vector	(25)
pG$^+$host5	erm^R temperature sensitive shuttle vector	(26)
pMSP3535	erm^R shuttle vector with nisin inducible promoter (P$_{nis}$)	(27)
pMSP3535::*cre* reverse	*cre* cloned in reverse orientation of P$_{nis}$ as control	This study
pMSP3535::*cre* forward	*cre* cloned in proper orientation to P$_{nis}$	This study
pTRKL2::*cre*	*cre* cloned without a promoter to accept library fragments	This study
pG$^+$host5::Δ*upp*	pG$^+$host5 containing the construct used to remove *upp* from the *S. pyogenes* MGAS8232 genome	This study
pCAS4*tet*	pG$^+$host5 containing the *loxP-tet^R-tk-loxP* cassette for insertion into the genome of *S. pyogenes* MGAS8232	This study
pDG1515	tet^R vector designated for homologous recombination in *B. subtilis*	(28)
pUC57::*tk*	Codon optimized *tk*	GenScript, Corp.
pTRKL2::P$_{gyrA}$::*sacB*	pTRKL2 containing *sacB* under control of the MGAS8232 *gyrase A* promoter	This study
pTRKL2::P$_{gyrA}$::*upp*	pTRKL2 containing *upp* under control of the MGAS8232 *gyrase A* promoter	This study
pTRKL2::P$_{gyrA}$::*tk*	pTRKL2 containing *tk* under control of the MGAS8232 *gyrase A* promoter	This study

NanoDrop spectrophotometer (Thermo Scientific, Wilmington, Delaware, USA). Ligations were performed utilizing 0.5 μg or 1 μg of plasmid. Sticky end inserts were ligated in an insert:plasmid ratio of 1:1 and incubated for 1 h at 16°C. Blunt end ligations were incubated overnight at 4°C in a 5:1 ratio. Total reaction volumes were kept at 20 μL and used T4 DNA Ligase (New England Biolabs, Ipswich, Massachusetts, USA).

2.2.6 *E. coli* competent cells

To prepare competent *E. coli*, a 1% inoculation from an overnight *E. coli* culture was made into PSI broth (2% w/v Tryptone, 0.5% w/v yeast extract, 0.5% w/v magnesium sulphate, pH 7.6) and grown at 37°C with aeration until an OD_{600} of 0.5 was reached. Bacteria were cooled on ice for 15 min then centrifuged at $5000 \times g$. The pellet was resuspended $0.4\times$ the original volume in TfbI buffer (100 mM rubidium chloride, 50 mM manganese chloride, 30 mM potassium acetate, 10mM calcium chloride, 15% v/v glycerol, pH 5.8) and chilled for 15 min. The bacteria were centrifuged and resuspended in $0.04\times$ the original volume in TfbII buffer (75 mM calcium chloride, 10 mM rubidium chloride, 10 mM 3-[N-morpholino]propanesulfonic acid, 15% v/v glycerol, pH 6.5), aliquoted (200 μL) and placed immediately at -80°C.

2.2.7 *E. coli* transformation

For transformation, competent *E. coli* cells were thawed on ice, 10 μL of the ligation reaction was added, and cells were incubated on ice for 30 min. After heat shocking at 42°C for 45 sec, cells were left on ice for another 2 min, 900 μL of LB media was added, and cells are incubated at 37°C for 1 h with shaking at 250 rpm. Cells were plated onto BHI agar containing 150 μg mL^{-1} erm and incubated at 37°C until colonies formed.

2.2.8 Colony PCR to identify clones

In order to identify clones, individual colonies were numbered, picked using a sterile toothpick, and touched to the bottom of a PCR tube. A PCR master mix containing M13 primers (flanking the multiple cloning site of plasmids used) was added to each tube (20 μL) to identify clones containing inserts. Products were amplified and visualized on an

agarose gel to identify clones. All clones were verified with sequencing at London Regional Genomins Centre, London, Ontario, Canada.

2.2.9 *S. pyogenes* competent cells

To prepare competent *S. pyogenes* cells, THY broth containing 0.6% glycine was inoculated 1:50 with an overnight culture. After 2 h, hyaluronidase was added to a concentration of 1 mg mL^{-1}. Once the OD_{600} was between 0.25 to 0.3, bacteria were centrifuged at 7000 × g for 5 min. Bacteria were resuspended in 0.4× of the original volume in 15% glycerol. The bacteria were centrifuged and resuspended in 0.04× the original volume in 15% glycerol, aliquoted (200 μL) and placed immediately at -80°C.

2.2.10 *S. pyogenes* electroporation

To transform *S. pyogenes*, tubes of competent *S. pyogenes* cells were allowed to come to room temperature and 2 μg DNA was added to each tube of cells, mixed, and transferred to a 2mm electroporation cuvette. Using the BioRad GenePulser (Mississauga, Ontario, Canada), cuvettes were pulsed using 2100 V and a pulse length of 1.1 ms. Bacteria were then transferred to 10 mL THY and recovered at 37°C. After 6 h, bacteria were concentrated 10×, plated on THY containing the appropriate antibiotics, and incubated at the required temperature.

2.2.11 Construction of *S. pyogenes* MGAS8232 Δ*upp* via homologous recombination

Using the PCR primers listed in **Table 2.2**, two 500 bp portions of the chromosome immediately flanking *upp* were PCR amplified and purified using the QIAquick PCR Purification Kit (Qiagen, Toronto, Ontario, Canada) as per the manufactuer's instructions. The upstream PCR product, along with pG$^+$host5, was digested with *Xba*I and *Bam*HI, ligated, transformed into competent *E. coli*, and screened for positive clones. Once verified via sequencing, this intermediate clone was digested with *Pst*I and *Xho*I along with the downstream PCR product. These were then ligated and transformed into competent *E. coli*, to generate pG$^+$host5::Δ*upp*. The final clone was verified with sequencing.

Following electroporation of the temperature sensitive integration construct, pG⁺host5::Δ*upp*, cells were grown at 30°C in THY erm for 4 days, replacing media every 24 h. Next, cells were shifted to 40°C and grown for an additional 4 days in THY erm, changing media every 24 h. At this temperature, the plasmid no longer replicates, and cells that have integrated the plasmid will remain resistant to erm. The culture was then plated for single colonies. Clones were grown individually in THY erm, genomic DNA was extracted, and PCR was used to ensure integration of pG⁺host5::Δ*upp* into the chromosome. Once confirmed, clones were grown in liquid culture at 30°C for 4 days in THY, replacing media every 24 h. Two functioning origins of replication create an undesirable situation within the cell, forcing a recombination event to occur. The culture was then plated to obtain single colonies that are patched onto plates with and without antibiotics to isolate colonies that have lost the plasmid, all at 40°C. Individual clones were then screened by PCR for the correct deletion/insertion. Clones were then verified with sequencing.

2.2.12 Verification of *S. pyogenes* MGAS8232 Δ*upp*

In order to test for sensitivity to 5-fluorouracil (5-FU), MGAS8232 Δ*upp* and MGAS8232 wild-type were grown overnight in THY, serially diluted 10-fold, and plated on M9T1 (M9 salts, 1% Tryptone, 0.1% glucose, 1.5% agar) with and with out 30 μM 5-FU (EMD Millipore, Billerica, MA, USA) and grown overnight at 37°C. A sample from both MGAS8232 Δ*upp* and MGAS8232 wild-type overnight THY cultures were also taken for qRT-PCR analysis. For growth curve analysis MGAS8232 Δ*upp* and MGAS8232 wild-type was grown in triplicate overnight in THY. All samples were normalized to an OD_{600} 0.01 and plated in triplicate using 200 μL per well. Following this, 50 μL of mineral oil was placed in the top of each well and the plate was analyzed using a Bioscreen C Automated Microbiology Growth Curve Analysis System (Growth Curves USA, Piscataway, NJ, USA) at 37°C with constant 'medium' shaking, reading every 0.5 h over a 24 h period.

2.2.13 Construction of counter-selection plasmids

The *sacB* gene was PCR amplified from the genomic DNA of *B. subtilis* subs. *subtilis* str.

168, purified using the QIAquick PCR Purification Kit (Qiagen, Toronto, Ontario, Canada), and digested with *Sal*I and *Pst*I. pTRKL2::P$_{gyrA}$::*sacB* was created in a two step ligation process first cloning *sacB* into *Sal*I and *Pst*I digested pTRKL2. P$_{gyrA}$ was then PCR amplified from *S. pyogenes* MGAS8232, digested with *Bam*HI and *Xho*I, and ligated into the *Bam*HI and *Sal*I digested intermediate plasmid. Clones were verified with sequencing at each stage. The human herpes simplex virus-1 *thymidine kinase* gene was codon optimized for *S. pyogenes* and synthesized by GenScript Corp (Piscataway, NJ, USA). P$_{gyrA}$ was amplified from MGAS8232 DNA and digested with *Bam*HI and *Nco*I and *tk* was amplified from pUC57::*tk* (GenScript Corp) and digested with *Nco*I and *Xba*I. pTRKL2 was digested with *Xba*I and *Bam*HI, and in one ligation reaction, the promoter was fused to *tk* to create pTRKL2::P$_{gyrA}$::*tk*. The construct was verified with sequencing.

2.2.14 Testing counter-selection plasmids

To test the different counter-selection genes, each was first transformed into *S. pyogenes* MGAS8232 as described in **(2.2.10)**. *S. pyogenes* containing each plasmid construct, as well as pTRKL2 as a control, was grown overnight in THY erm, serially diluted 10 fold, plated on the respective media, and grown overnight at 37°C. *S. pyogenes* containing pTRKL2::P$_{gyrA}$::*sacB* was plated on THY erm with and with out 15% sucrose. Overnight growth from THY erm liquid culture was also used for qRT-PCR analysis. *S. pyogenes* containing pTRKL2::P$_{gyrA}$::*tk* was plated on BMEM (DMEM/F12, 1x vitamin solution, sodium bicarbonate, 1.5% agar) erm with and with out ganciclovir (Cedarlane Labs, Burlington, Ontario, Canada) at 1000 μg mL^{-1}. Media components were purchased as follows: M9 salts, Tryptone, and agar from BD Biosciences (Franklin Lakes, NJ, USA), glucose and sodium bicarbonate from Sigma-Aldrich (Oakville, Ontario, Canada), and DMEM/F12 and vitamin solution from Gibco Life Sciences (Burlington, Ontario, Canada)

2.2.15 *S. pyogenes* RNA extraction

S. pyogenes cells, from frozen or grown overnight in liquid media, were incubated with RNAprotect Cell Reagent (Qiagen, Toronto, Ontario, Canada) according to the manufacturers instructions. The cells were pelleted and resuspended in 500 μL Tris

EDTA Glucose buffer (10 mM Tris, 2 mM EDTA, 25% glucose), adding 5 μL 200 mg mL^{-1} lysozyme and 5 μL 10 U μL^{-1} mutanolysin. Cells were incubated for 1 h at 37°C. After centrifugation (21 000 × g), the pellet was resuspended in 500 μL lysis buffer (50 mM EDTA, 0.2% SDS) with 5 μL 20 mg mL^{-1} proteinase K and incubated at 65°C for 2 h. The RNeasy Kit (Qiagen, Toronto, Ontario, Canada) was used according to the manufacturers instructions with the lysed cells.

2.2.16 Quantitative reverse transcriptase polymerase chain reaction

cDNA was generated from the processed RNA with SuperScript II Reverse Transcriptase and Random Primers (both from Invitrogen, Burlington, Ontario, Canada) according to the manufacturers instructions. Primers were designed to maintain an amplicon length of ~200 bp and T$_m$ of ~60°C. Along with the *proS* control, all genes to be measured were first amplified from wild-type DNA and the samples were purified using a Qiagen QIAquick PCR Purification Kit (Qiagen, Toronto, Ontario, Canada), quantified using a NanoDrop spectrophotometer (Thermo Scientific, Wilmington, Delaware, USA), and serially diluted using Qiagen Elution Buffer to 10^{-10}. The qRT-PCR reactions were performed in triplicate using iQ SYBR Green Supermix (Bio-Rad, Mississauga, Ontario, Canada) with the appropriate primers and template. The primer set for each gene was added to reactions with the following: no template, DNA gradient of the gene being measured from 10^{-3} to 10^{-10}, and the cDNA from the sample being measured. Samples were run on a Rotor-Gene 6000 (Corbett Life Science, Kirkland, Quebec, Canada) and analyzed with the provided software. All cycles used an initial 95°C at 5 min followed by 40 cycles of 10 sec 95°C denature, 10 sec 60°C anneal, and 15 sec 72°C extension.

2.2.17 Construction of the *loxP-tet^R-tk-loxP* cassette (pCAS4*tet*)

Based on previous bioinformatic analysis of Rho-independent terminators (de Hoon *et al.*, 2005), a chromosomal location downstream of two opposing genes (*pepO* and *tsf*) with their own Rho-independent terminators was used to insert the cassette. Two regions of ~500 bp were designed, flanking the insertion site, to allow for homologous recombination of the cassette into the genome. The downstream recombination site was PCR amplified from the MGAS8232 genome and ligated into the *Xba*I and *Sac*II sites of

pG⁺host5. Next, the upstream recombination amplicon was inserted into the *Xho*I and *Cla*I sites of the intermediate clone. In order to incorporate the *loxP* sites, these sequences were included in the recombination site primers proximal to the cassette. Next, P$_{gyrA}$::*tk* was PCR amplified and cloned between the *loxP* sites utilizing *Bam*HI and *Xba*I. Lastly, *tet*^{R} was amplified from pDG1515, digested with *Hin*dIII and *Bam*HI, and cloned into the intermediate digested with *Cla*I and *Bam*HI to create the final cassette. Clones were verified with sequencing at each stage and for the final construct designated pCAS4*tet*. Each insert was also verified by individual excision from pCAS4*tet* and visualized by gel electrophoresis. The pCAS4*tet* was further tested for proper excision *ex vivo* using Cre recombinase (New England Biolabs, Ipswich, Massachusetts, USA) with supplied positive control and empty vector (negative control) according to the manufacturer's instructions. DNA was then visualized on an agarose gel.

2.2.18 Construction and verification of *S. pyogenes* MGAS8232 Cas2

Using the recombination sites, the cassette within pCAS4*tet* was inserted into the chromosome of MGAS8232 via homologous recombination using the same method outlined in **2.2.11**. Successful integration of the cassette was verified via sequencing. To test for spontaneous excision of the RIVET cassette, *S. pyogenes* MGAS8232 Cas2 (Cas2) was grown overnight in THY tet and subcultured (1%) into THY without any antibiotics. From here, daily subcultures would continue into THY without antibiotics. A sample was taken daily, serially diluted 10-fold, and plated onto THY agar with or without tet. Colony forming units (CFUs) were counted and compared for analysis.

To test excision of the RIVET cassette via Cre, pMSP3535, pMSP3535::*cre* reverse, and pMSP3535::*cre* **(Table 2.3)** were transformed into MGAS8232 Cas2 and grown overnight in THY erm/tet. Two samples were taken of each and resuspended in fresh THY erm with or without nisin at 100 ng mL^{-1}. After overnight growth, cells were serially diluted 10-fold, plated on THY agar containing erm or tet, and grown overnight. CFUs were counted and compared for analysis

2.2.19 Creation of the *S. pyogenes* promoter library and removal of *in vitro* activated promoters

Total genomic DNA (gDNA) from wild-type MGAS8232 was digested with various amounts of *Sau*3AI (0, 0.25, 0.5, 1, 2, 4 Units of enzyme) for 1 h. Digestions were used in 1:2 ratio ligations with pTRKL2::*cre* digested with *Bam*HI. Ligations were transformed into *E. coli*, and all colonies were scraped off plates, concentrated, and mixed plasmids were isolated using the Qiagen Miniprep Kit (Qiagen, Toronto, Ontario, Canada). These plasmids were then transformed into Cas2. Instead of plating, cells were grown in liquid in the presence of erm and tet for 24 h to remove *in vitro* active promoters. A small sample was taken, serially diluted 10-fold, and plated to determine CFUs, while the remaining cells were frozen at -80°C for further *in vivo* experiments.

2.2.20 Identification of *in vivo* induced promoters

Animal experiments were conducted in accordance with guidelines established by the Canadian Council on Animal Care and approved by the Animal Use Subcommittee at the University of Western Ontario **(Appendix 1)**. Using ~10^8 Cas2 CFUs from batches in which *in vitro* promoters were removed, the cells were warmed to room temperature for 30 min, and subsequently inoculated through the nasal route into C57BL/6 mice that expressed both human HLA-DR4 and HLA-DQ8 mice (29) according to the method of Kasper *et al.* (30). After 48 h, mice were sacrificed and the complete nasal passages were removed, homogenized, and plated as described (30). Colonies were enumerated on THY erm plates and the remaining cells were maintained at -80°C. From this frozen stock, ~150 CFU was plated onto THY agar plates and grown overnight. Colonies were patched onto THY agar plates with and without tet to screen for the loss of the cassette. Tet sensitive clones were subsequently grown, DNA was extracted, transformed into *E. coli* to purify plasmids, and inserts were sequenced. Homology searches of potential promoter regions were performed using the Basic Local Alignment Search Tool nucleotide (BLASTn) tool at the website of the National Center for Biotechnology Information (http://www-ncbi-nlm-nih-gov). Further analysis was conducted using the Comprehensive Microbial Resource at the J. Craig Venter Institute (http://cmr.jcvi.org/tigr-scripts/CMR/CmrHomePage.cgi).

2.3 Results

2.3.1 Selection of a suitable counter-selection method

In order to enhance the screening method of the RIVET system, a counter-selection strategy was first investigated. Counter-selection has been used in *S. pyogenes* successfully before as a method to screen for double cross over clones when creating mutants (31). Based on a strain with natural resistance to streptomycin from a mutated *rpsL* gene, the plasmid containing the genetic crossover contains a wild-type *rpsL* that would provide no protection against streptomycin. Since the wild-type sensitive allele is dominant over the mutant resistant allele, when both are present the strain is sensitive to streptomycin. When the sensitive gene is removed, the clone becomes resistant (32, 33). However, due to the lack of natural streptomycin resistance in MGAS8232, an alternative counter-selection was necessary. Three potential counter-selection methods were tested for their potential use: *sacB* (levansucrase), *uracil phosphoribosyltransferase* (*upp* [UPRTase]), and Herpes simplex virus 1 *thymidine kinase* (HSV1-*tk*).

The first potential counter-selection method attempted was *sacB*. The *sacB* gene encodes levansucrase, which catalyzes the hydrolysis of sucrose to glucose and fructose as well as the polymerization of fructosyl groups to form levan (34). It was shown previously that *B. subtilis* could be made sensitive to sucrose when expressing *sacB* with a mutation within the signal peptide. It was hypothesized that the lethality of a mutated levansucrase is caused by the detrimental accumulation of levans (34). Thought it could be made more efficient, the entire signal peptide was removed. Therefore, *sacB* was cloned from the *B. subtilis* 168 genome, without the signal peptide, and cloned into the pTRKL2 plasmid under the control of the *S. pyogenes gyrA* promoter (P$_{gyrA}$). After the clone was verified with sequencing, the clone pTRKL2::P$_{gyrA}$::*sacB* was transformed into *S. pyogenes* MGAS8232. Along with an empty vector control, pTRKL2::P$_{gyrA}$::*sacB* was tested on media containing up to 15% sucrose. The transcription of *sacB* was confirmed via qRT-PCR; however, even when grown on the highest concentration of sucrose, pTRKL2::P$_{gyrA}$::*sacB* did not cause any lethality **(Figure 2.2)**.

The second attempt to generate a counter-selection system in *S. pyogenes* was the use of

Figure 2.2: Evaluation of *sacB* as a counter-selection gene in *S. pyogenes*

A) Serial 10-fold dilutions of *S. pyogenes* MGAS8232 containing pTRKL2 (vector) or pTRKL2::P$_{gyrA}$::*sacB* grown on THY agar or THY agar including 15% sucrose. B) qRT-PCR analysis of *sacB* transcription from *S. pyogenes* containing pTRKL2 or pTRKL2::P$_{gyrA}$::*sacB* after overnight growth in THY. Expression is relative to the housekeeping gene *proS*. N=3, unpaired *t*-test, *** p<0.0001.

A)

Vector   P*gyrA*::*sacB*   Vector   P*gyrA*::*sacB*

10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}

no sucrose   15% sucrose

B)

***

Expression Relative to *proS*

2.0
1.5
1.0
0.5
0.0

Vector   P*gyrA*::*sacB*
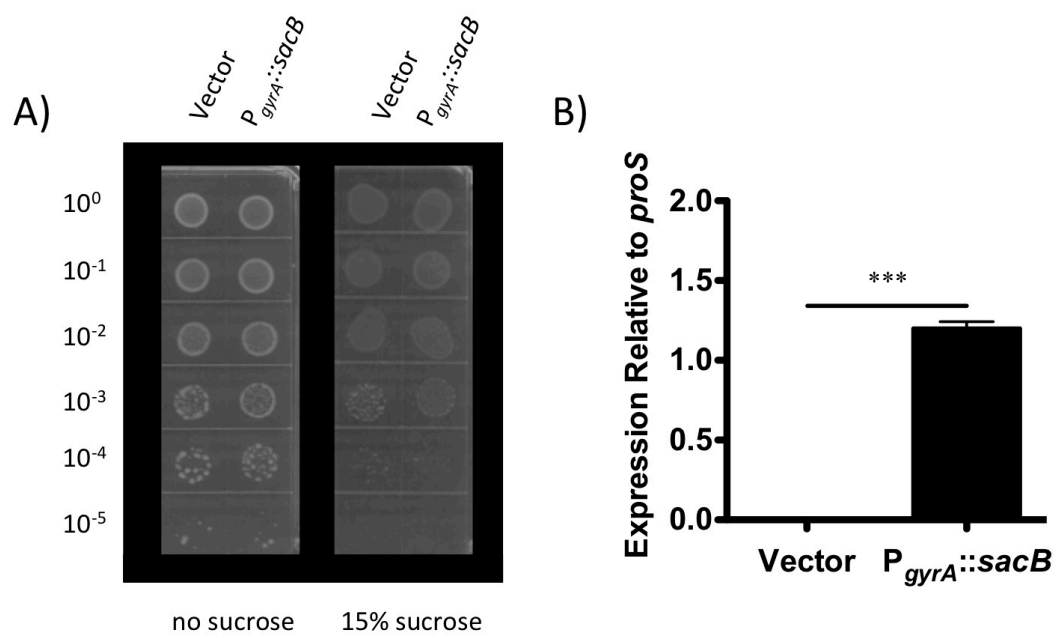
the *upp* gene. In the cell, UPRTase normally converts uracil to UMP, allowing the cell to use exogenous uracil. However, UPRTase can also convert 5-FU, a toxic pyrimidine, into 5-fluoro-UMP. This is then metabolized into 5-fluoro-deoxyuracil monophosphate which inhibits thymidylate synthetase, an enzyme required for growth (35). Since *upp* is endogenous to *S. pyogenes*, a knockout first had to be created. Utilizing molecular recombination techniques outlined in the Materials and Methods, MGAS8232 Δ*upp* was generated. Following this, growth on media with or without 5-FU, along with MGAS8232 wild-type as a control, demonstrated that 5-FU was no longer toxic to MGAS8232 Δ*upp* **(Figure 2.3A)**.

Further testing against MGAS8232 Δ*upp* was carried out to ensure that the mutation had no obvious detrimental effects. First evaluated was a growth curve against MGAS8232 wild-type to ensure that growth was not affected. Curiously, MGAS8232 Δ*upp* grew at a faster rate to a higher overall OD than MGAS8232 wild-type **(Figure 2.3B)**. Next, RNA was extracted from both MGAS8232 wild-type and MGAS8232 Δ*upp*, and the expression of the genes immediately upstream and downstream to *upp* were compared. While there was no difference in the expression of the upstream gene, *spyM18_0442,* the downstream gene, *clpP*, showed increased expression with MGAS8232 Δ*upp* **(Figure 2.3C)**. Therefore, MGAS8232 Δ*upp* contained differences that could have undesirable effects on *in vivo* growth, and for this reason, *upp* was not chosen for a counter-selectable marker.

The last counter-selection method evaluated was the *thymidine kinase* (*tk*) gene from the Human Herpes Simplex Virus-1 (HSV). The HSV-*tk* has relaxed specificity and can phosphorylate substrates other than thymidine. This allows for the addition of pro-drugs (e.g. ganciclovir) that can be phosphorylated by HSV-*tk*, but are not phosphorylated by native *tk*. Once the pro-drug is phosphorylated by HSV-*tk* to the mono state, it can be then recognized by cellular kinases and converted to the di- and tri-phosphorylated state, where it can be incorporated into replicating DNA, causing cessation of DNA replication, leading to cell death **(Figure 2.4A)**.

In order to engineer this human viral gene to be expressed in a prokaryote, the cDNA was codon optimized for *S. pyogenes* and synthesized. Once produced, *tk* was cloned under the control of P$_{gyrA}$ in pTRKL2, transformed into MGAS8232 and tested with the pro-

Figure 2.3: Evaluation of *upp* as a counter-selection gene in *S. pyogenes* and MGAS8232 Δ*upp* fitness

Serial 10-fold dilutions of *S pyogenes* MGAS8232 wild-type (top dilutions) or MGAS8232 Δ*upp* (bottom dilutions) were grown on A) M9 tryptose agar or M9 tryptose agar containing 30 µM 5-FU. B) Growth curves comparing MGAS8232 wild-type and Δ*upp*. Overnight cultures were subcultured to OD_{600} 0.01, and grown for 24 h, with OD measured every 0.5 h. N=3. C) qRT-PCR transcriptional analysis of *spyM18_0442*, *upp*, and *clpP* from MGAS8232 wild-type and Δ*upp* after overnight growth in THY. N=3, unpaired *t*-test, ns=0.8490 (*spyM18_0442*), *** p=0.0004 (*upp*), * p=0.0464 (*clpP*).
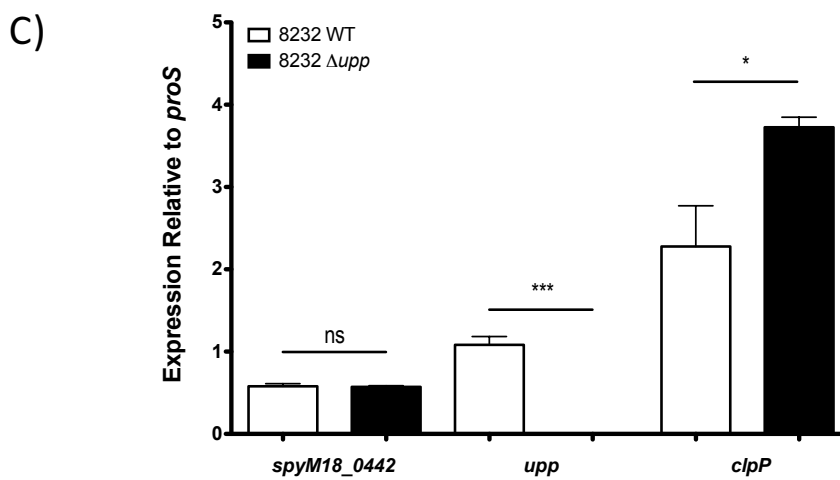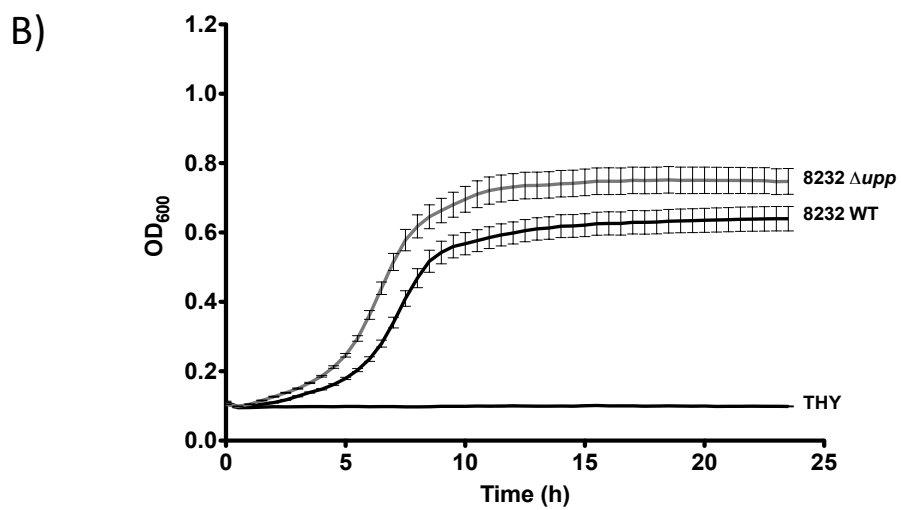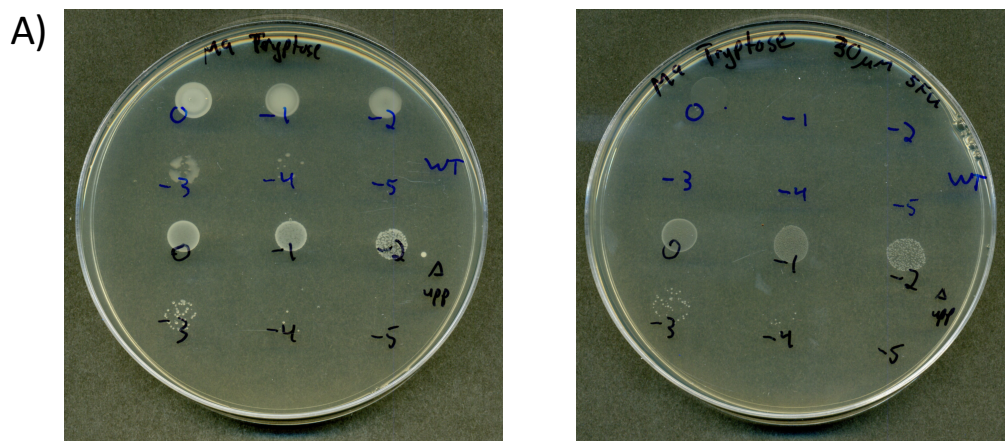
Figure 2.4: Ganciclovir method of counter-selection and initial test in MGAS8232

A) Ganciclovir (GCV) and incorporation via HSV1-*tk***.** Due to its lack of specificity, HSV1-*tk* is able to mono-phosphorylate GCV (inset) to GCV-MP. The kinases of *S. pyogenes* then have the ability to di- and tri-phosphorylate GCV-MP. GVC-TP can then be incorporated into DNA, preventing further incorporation of nucleotides, stopping DNA replication, and leading to cell death. B) Evaluation of HSV-*tk* and GCV as a counter-selection pair in *S. pyogenes*. Serial 10-fold dilutions of *S. pyogenes* MGAS8232 containing pTRKL2 (vector) or pTRKL2::P*gyrA*::*tk* grown on BMEM agar and BMEM agar containing 1000 µg/ml GCV.

A)

B)

0 µg/ml   1000 µg/ml GCV
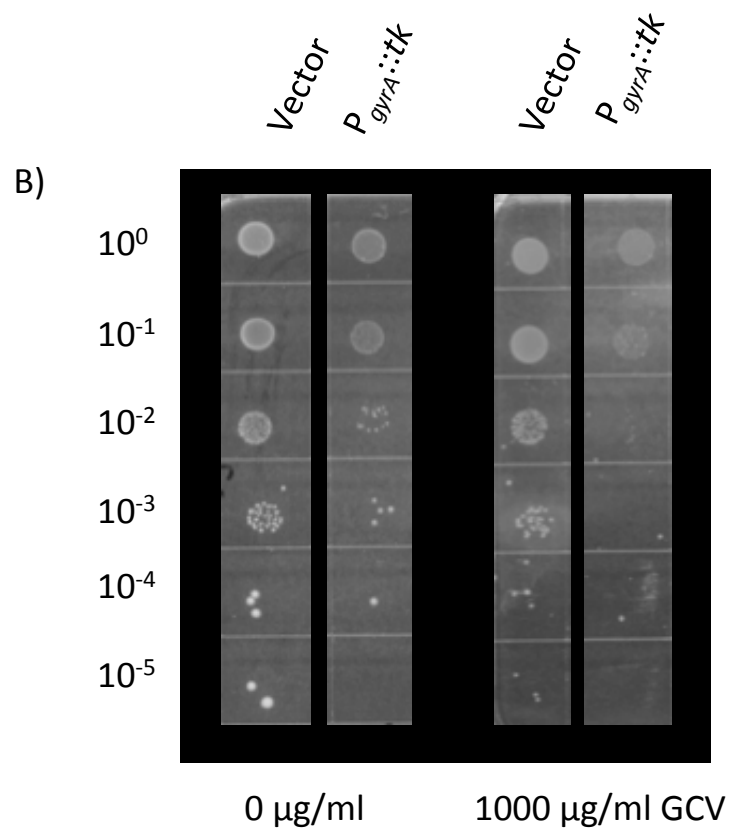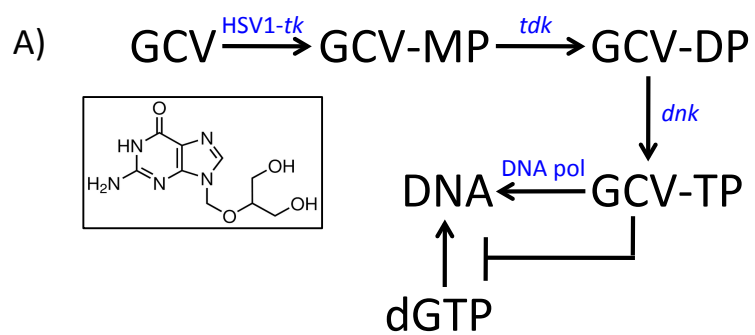
drug ganciclovir. At the highest concentration tested, a counter-selection phenotype was demonstrated in which ~2 logs of growth were lost **(Figure 2.4B)**. This then provided a potential counter-selection method that could be built into the RIVET design for *S. pyogenes*.

2.3.2 Construction of pCAS4*tet*

To engineer a RIVET system, a suitable location for the cassette to be inserted into the *S. pyogenes* MGAS8232 genome had to be found prior to construction could begin. An appropriate location would not compromise growth or the expression of nearby genes. It was decided to identify two genes that terminated facing each other to avoid disrupting any promoters. Additionally, two genes were selected each with predicted transcriptional terminators so that the cassette could be placed between, and therefore not disrupt any transcriptional units. Using the work of Hoon *et al*., it was possible to find many promising sites, and ultimately the region between *pepO* and *tsf* was selected as an appropriate location (36). Using 500 bp of chromosomal DNA to allow for recombination, the cassette could be inserted between the transcriptional terminators. These two chromosomal fragments were PCR amplified, each containing the necessary *loxP* site on one PCR primer, and each fragment was cloned into pG$^+$host5. Between the *loxP* sites *tet* and P$_{gyA}$::*tk,* were then cloned, forming the cassette **(Figure 2.5A)**. Once completed, pCAS4*tet*, was verified via sequencing and restriction digestion analysis **(Figure 2.5B)**. Finally, pCAS4*tet* was tested for its ability to excise the cassette using the Cre protein. Using empty pG$^+$host5, a positive control (linear stranded DNA with *loxP* sites), and pCAS4*tet*, all were incubated with and without Cre protein and analyzed by agarose gel electrophoresis **(Figure 2.5C)**. The negative control, pG$^+$host5, showed no changes with the presence of Cre, however the positive control showed the addition of bands, indicating that Cre had caused recombination of the *loxP* sites contained within the DNA. Lastly, in pCAS4*tet,* when incubated with Cre the top band of DNA seems to have been split into two (arrows), as well as the addition two faint bands of approximately 3.5 kb and 5.5 kb (arrows) **(Figure 2.5C)**. This indicated recombination of the *loxP* sites by Cre, and therefore, potential resolution of the cassette.

Figure 2.5: Visual verification of pCAS4*tet* construction and excision of the cassette via Cre

A) Plasmid map showing the completed pCAS4*tet* indicating restriction enzymes used. The *loxP* sites are enlarged to show their position.   B) pCAS4*tet* was verified via restriction enzyme analysis. Each component was removed from the construct as indicated by restriction enzyme digestion and ran on a 0.8% agarose gel. C) Verification of excision of the cassette from pCAS4*tet ex vivo*. The negative control (pG$^+$host5), the positive control (provided with Cre protein), and pCAS4*tet* were incubated with or without Cre and ran on a 0.8% agarose gel. Arrows indicate recombination products of pCAS4*tet*.

A)

B)

Ladder
1. pCAS4*tet* uncut
2. Cassette
3. *tetracycline*
4. *thymidine kinase*
5. *pepO* recombination site
6. *tsf* recombination site
Ladder
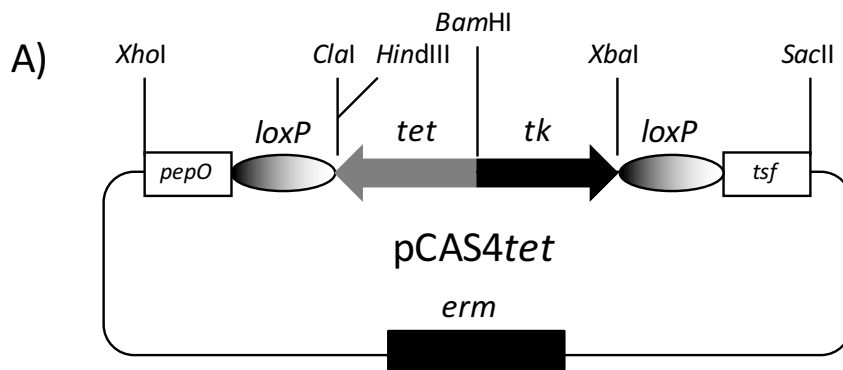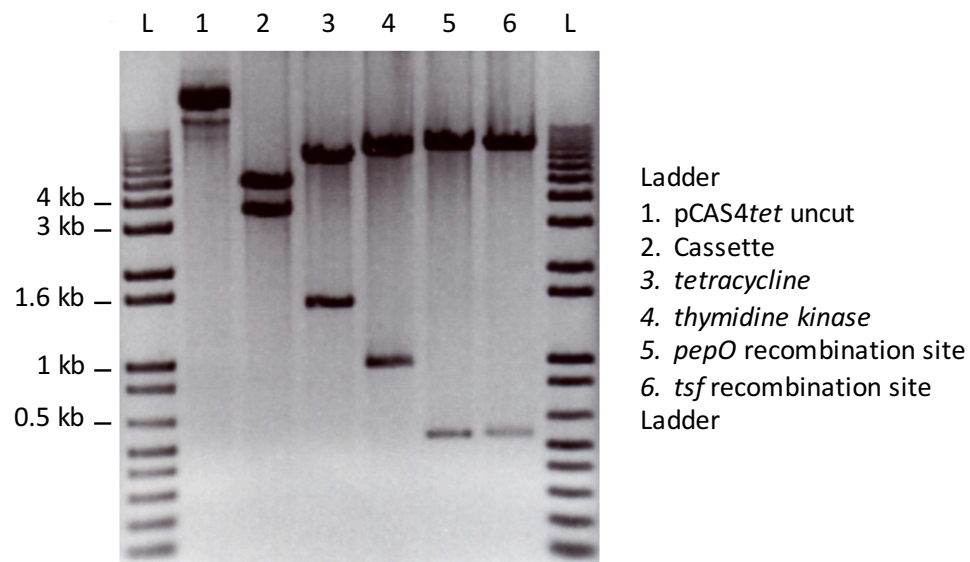
C)
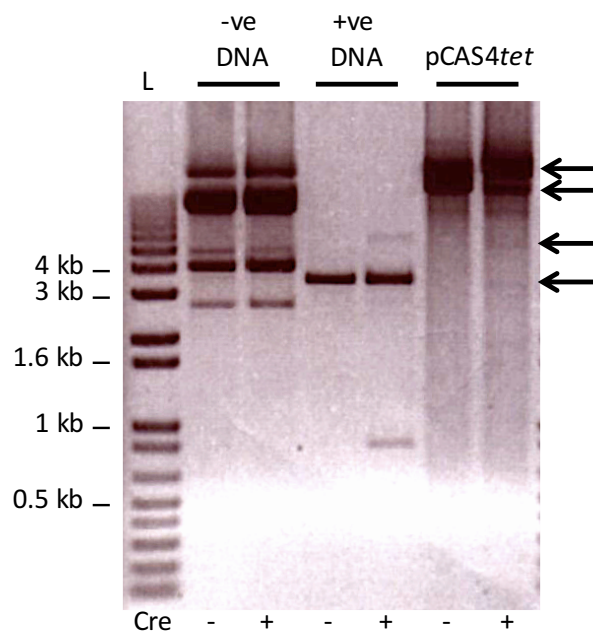
2.3.3 Creation and verification of the cassette containing strain Cas2

Next, pCAS4*tet* was transformed into MGAS8232 and the cassette was inserted into the chromosome, via recombination events similar to the construction of MGAS8232 Δ*upp,* to create Cas2. Using qRT-PCR, the expression of *pepO* and *tsf* was compared between Cas2 and MGAS8232 wild-type grown *in vitro*. These results showed no significant difference in gene expression between the two strains **(Figure 2.6A)**. Secondly, Cas2 was subcultured for 7 days *in vitro* without tet to determine if any spontaneous resolution of the cassette occurred. There appeared to be little noticeable spontaneous loss of the cassette **(Figure 2.6B)**.

Further testing was also performed to ensure that the cassette could be excised from the chromosome properly. In order to analyze this, a plasmid containing *cre* under the control of a nisin inducible promoter was transformed into Cas2. Included were two controls: *cre* in the reverse orientation, and the empty vector. Overnight cultures were given fresh media with erm, with or without nisin, and plated on THY erm, as a control, and THY tet, the next day **(Figure 2.7)**. As expected, all strains were able to grow on THY erm regardless of the presence or absence of nisin. On the THY tet plate, the empty vector and *cre* in reverse controls both showed the similar amount of growth with or without nisin. This indicated that no cassette excision was occurring. With *cre* in the forward orientation, loss of growth was evident on the THY tet plate both in the presence and absence of nisin. Excision of the cassette in the absence of nisin was most likely occurring because of small amounts of *cre* expression due to the leakiness of the nisin promoter. In contrast, those grown overnight in nisin showed no growth on THY tet, indicating that the cassette was properly excised under the control of Cre.

Cas2 was also tested for counter-selection, and gave an initial positive result, however with further testing, the method was unfortunately deemed unreliable in Cas2 (data not shown). While qRT-PCR was able to verify HSV-*tk* transcription **(Figure 2.6A)**, Western blot analysis failed to indicate the presence of protein from whole cell MGAS8232 and Cas2 extracts (data not shown). A positive control confirmed that antibodies were functional.

Figure 2.6: Assessing Cas2 suitability

A) Measuring transcription of genes flanking the cassette. qRT-PCR analysis of *pepO*, *tk*, and *tsf* transcription after overnight growth comparing MGAS8232 wild-type and Cas2 relative to *proS* housekeeping gene. N=3, unpaired *t*-test, ns=0.1439 (*pepO*), ***p<0.0001 (*tk*), ns=0.0838 (*tsf*). B) Cassette stability in Cas2. Cas2 was grown overnight in THY containing tet, and subcultured daily in THY. Each day 10-fold dilutions of the overnight culture were plated on THY agar and THY agar tet. CFUs were counted the next day with the THY agar count set as 100%.

A)

B)
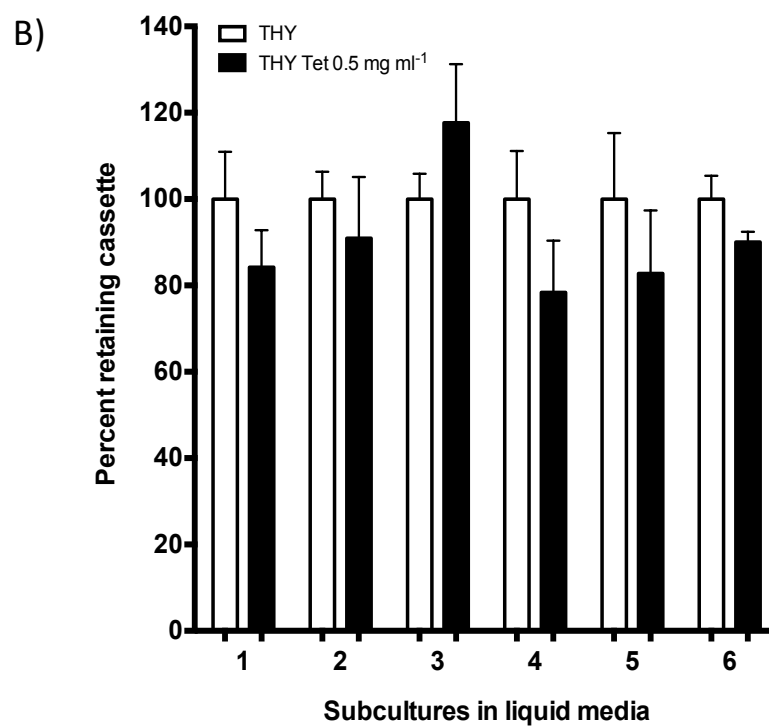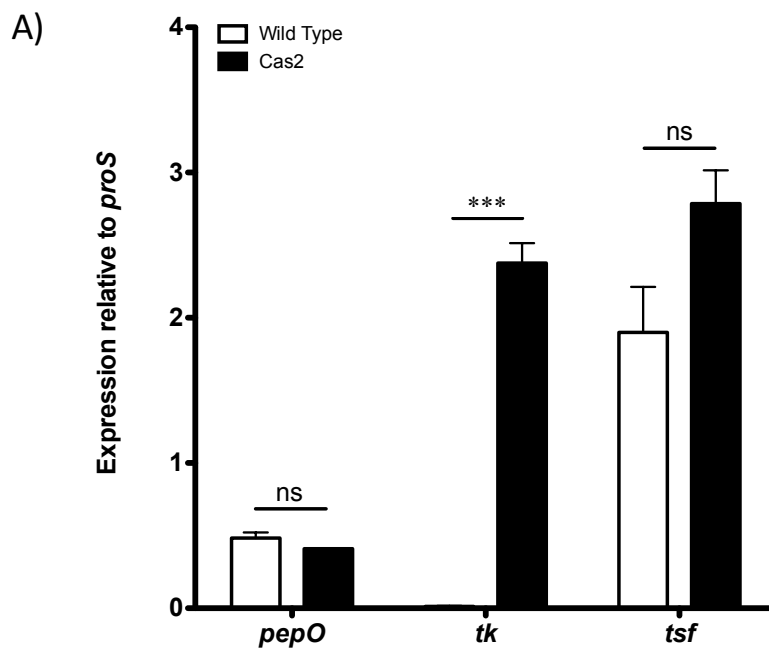
Figure 2.7: Evaluation of cassette resolution in *S. pyogenes* Cas2

Using the nisin inducible vector pMSP3535, Cas2 was transformed with vector, *cre* reverse, or *cre* forward. Two sets of overnight cultures were given fresh THY with erm, and half were also given nisin. Following overnight growth, serial 10-fold dilutions were plated on A) THY agar with erm and B) THY agar with tet.
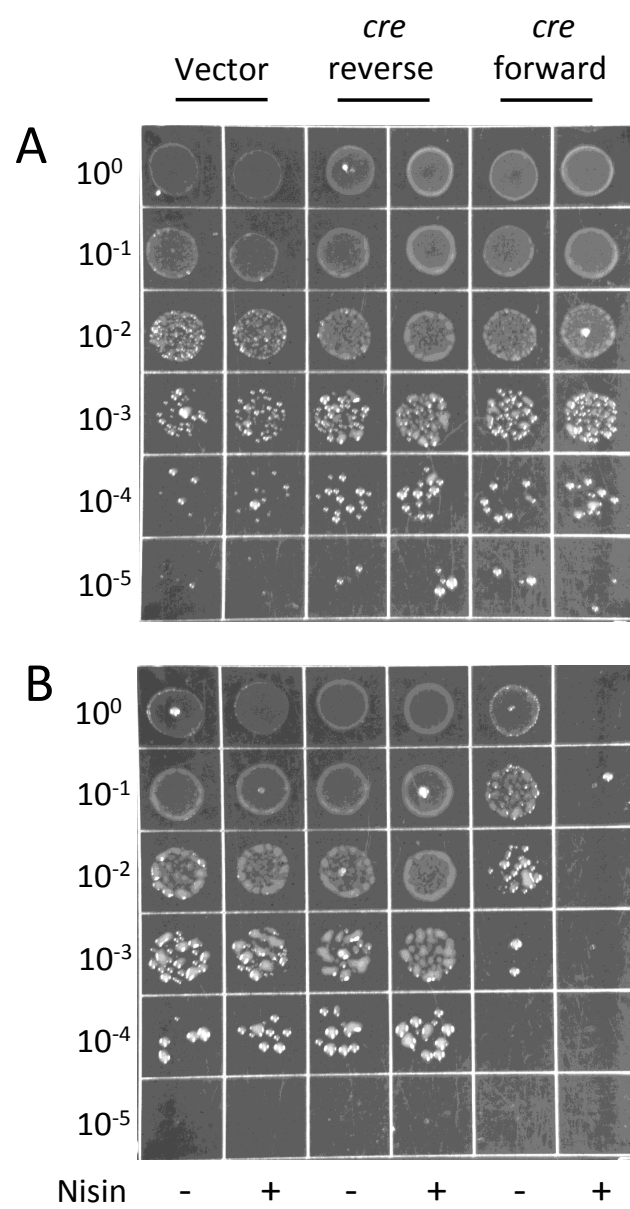
2.3.4 Creation of a genomic library and removal of *in vitro* active promoters

In order to create a *S. pyogenes* promoter library, genomic DNA was extracted from MGAS8232 and digested with various amounts (0, 0.25, 0.5, 1, 2, 4 Units) of the restriction enzyme *Sau*3AI for 1 h. Verified on an agarose gel, **Figure 2.8A** demonstrates successful digestion. The DNA was then purified, pooled, ligated into the pTRKL2::*cre* plasmid in a 1:2 ratio (vector:insert) and transformed into *E. coli*. In order to ensure fragment variability, random clones were selected and PCR amplified to determine the insert length, with one representative gel shown in **Figure 2.8B**. All *E. coli* clones were then pooled and plasmids were extracted. The plasmids, containing library fragments, were transformed into Cas2, and grown overnight in the presence of erm and tet in order to eliminate the clones expressing *in vitro* promoters. The surviving clones were then concentrated and frozen.

2.3.5 Identification of promoter regions induced *in vivo* using an acute nasopharyngeal infection model

For the *in vivo* identification of promoters, 10^8 cells were taken from the frozen library stocks and inoculated into the nasal passages of each of four HLA-DR4/DQ8 mice. The complete nasal turbinates (cNTs) were recovered, and after 48 h, counter-selection was attempted, but proved to be unreliable (data not shown). Colonies were therefore patched to identify clones for which the cassette had been resolved. In total, 171 plasmids were isolated and sequenced, leading to 82 unique sequences after duplicates were removed **(Appendix 2).** Sequences were analyzed using BLASTn to determine which portion of the MGAS8232 genome was contained within the plasmid. The inserts were divided into four different categories based on their position in the chromosome. The first were 'typical' promoters in which the fragment was upstream of the 5' end of an annotated ORF. Second were 'cryptic' promoters, where the fragment was contained completely within an ORF and facing the same direction. Third were 'antisense' promoters, which were also contained completely within an ORF, but were facing the opposite direction. While not typical of obvious ORFs, these cryptic and antisense promoters could be involved in the regulation of the genome as sRNA molecules. Lastly, some plasmids also contained multiple inserts, which likely arose from cloning artifacts.

**Figure 2.8: Confirmation of *S. pyogenes* MGAS8232 gDNA digestion and cloned
fragment size variability**

A) Digestion of *S. pyogenes* genomic DNA. MGAS8232 gDNA was digested with 0,
0.25, 0.5, 1, 2, and 4 Units of *Sau*3AI for 1 h and analyzed on a 0.8% agarose gel. B)
Verification of fragment variability in the genomic library. Plasmids were extracted from
a random number of transformed bacteria and fragment size was measured using PCR
with flanking primers. Included are no template as a negative control (-) and *cre* only
positive control (+). Representative PCRs are shown on a 0.8% agarose gel.

A)

Ladder
1. 0 U *Sau*3AI
2. 0.25 U *Sau*3AI
3. 0.5 U *Sau*3AI
4. 1 U *Sau*3AI
5. 2 U *Sau*3AI
6. 4 U *Sau*3AI
Ladder
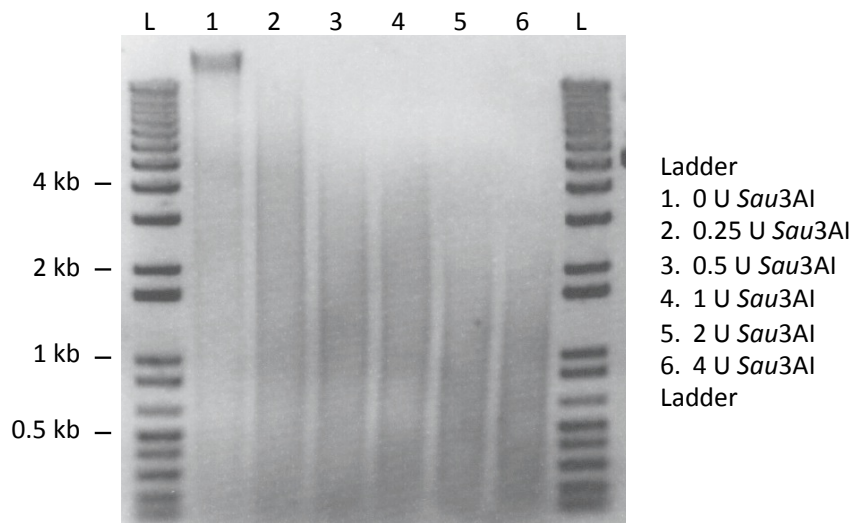
B)
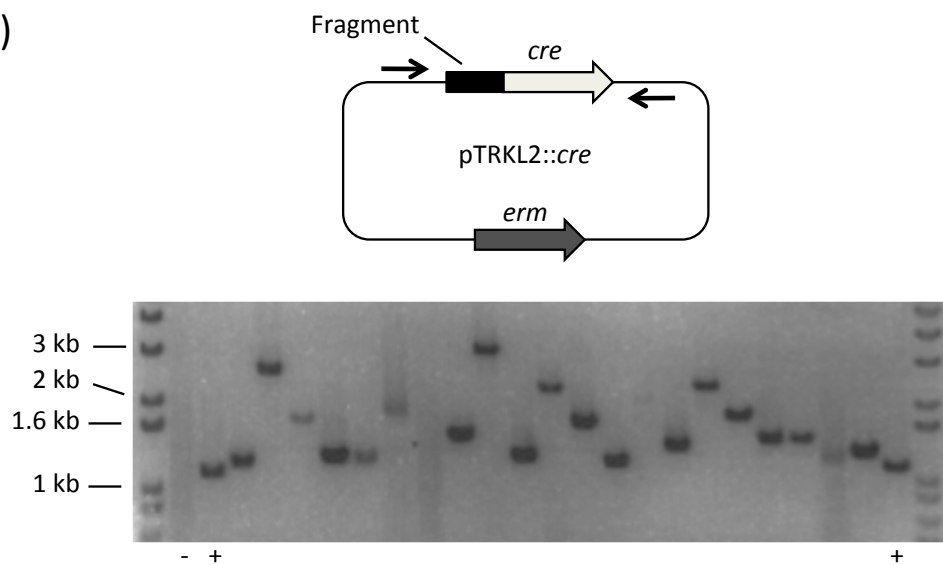
Since it would be difficult to determine which insert acted as the promoter in each case, they were removed from further analysis.

To further classify the identified promoters, they were divided by sense (including 'typical' and 'cryptic') and antisense, and using the gene they were associated with, further divided based on function using the J. Craig Venter Institute Comprehensive Microbial Resource **(Figure 2.9)**. This allowed us to determine how the *in vivo* induced loci were distributed according to their potential function. The majority of processes associated with sense clones include amino acid biosynthesis, degradation of proteins, peptides, and glycopeptides, cell envelope, cell processes, and energy metabolism. While the highest category for antisense clones was hypothetical proteins, the highest categories with functions were cellular processes and DNA metabolism. Some interesting differences between the two categories are the presence of amino acid biosynthesis and regulatory functions, found only in the sense clones, while only the antisense clones contain the purines, pyrimidines, nucleosides, and nucleotides category.

2.3.6 Analysis of 'typical' promoter

From the total number of clones, nine 'typical' promoters were found **(Table 2.4)**. Function was predicted using Basic Local Alignment Search Tool protein (BLASTp) analysis to similar known proteins. Of particular interest were clones IVI100 and IVI156, representing a minor structural protein for ΦSpeC and a bacteriocin-like protein, respectively. Previous work in our lab has shown ΦSpeC to spontaneously excise from the chromosome *in vitro*, however, this has never been further explored *in vivo*. Additionally, a role for bacteriocins in niche adaptation by *S. pyogenes* has not been previously studied. These two genes possibly provide new areas of research related to colonization of the host. In order to test the validity of these promoters, the promoter-containing plasmids of IVI100 and IVI156 were transformed back into Cas2 and re-evaluated both *in vitro* and *in vivo*, to establish if the promoters activated as previously determined.

Figure 2.9: Separating RIVET clones based on function

Potential *in vivo*-induced promoter elements were divided between sense and antisense, and using the J. Craig Venter Institute Comprehensive Microbial Resource. The sequences were further divided based on the predicted function of the associated gene.

Sense

Regulatory functions 6.1%
Protein synthesis 3.0%
Cell envelope 12.1%
Degradation of proteins, peptides, and glycopeptides 12.1%
Cellular process 12.1%
DNA metabolism 3.0%
Amino acid biosynthesis 12.1%
Energy metabolism 12.1%
Unknown function 6.1%
Unclassified 3.0%
Transport and binding proteins 9.1%
Mobile and extrachromosomal element functions 9.1%

Antisense

Purines, pyrimidines, nucleosides, and nucleotides 9.1%
Cell envelope 9.1%
Hypothetical proteins 18.2%
Cellular processes 13.6%
Biosynthesis of cofactors, prosthetic groups and carries: pyridoxine 4.5%
DNA metabolism 13.6%
Unknown function 4.5%
Unclassified 4.5%
Transport and binding proteins 4.5%
Energy metabolism 9.1%
Mobile and extrachromosomal elemental functions 9.1%

Table 2.4: Potential 'typical' promoters identified

Clone	Corresponding Gene	Description
IVI49	*spyM18*_2257	chromosome segregation protein
IVI53	*spyM18*_1223	amino acid ABC transporter, periplasmic amino-acid binding protein
IVI60	*spyM18*_0197	hypothetical protein
IVI72	*spyM18*_0784	tagatose-6-phosphate aldose/ketose isomerase
IVI84	*spyM18*_2004	conserved hypothetical protein
IVI87	*spyM18*_1912	pyruvate formate-lyase
IVI100	*spyM18*_0771	hypothetical protein, phage associated
IVI156	*spyM18*_0544	putative BlpM homologue
IVI176	*spyM18*_0414	conserved hypothetical protein

2.3.7 Clone IVI156 contained an *in vivo*-induced promoter

Cas2 containing either IVI100 or IVI156 was grown *in vitro* overnight in THY erm and then plated on THY erm agar and THY erm/tet agar. For a control, Cas2 containing the promoterless pTRKL2::*cre* was used. As with the initial RIVET screen, both IVI100 and IVI156 showed no promoter activity *in vitro*, as indicated by the maintenance of tet resistance in the Cas2 **(Figure 2.10)**. The promoterless control also maintained tet resistance. Next, multiple cultures of each clone were grown overnight *in vitro*, combined, enumerated, and frozen. From the frozen stock, 10^8 of the negative control, as well as IVI100 and IVI156 were each used to inoculate HLA-DR4/DQ8 mice. The recovered clones were then plated and compared as they were for the *in vitro* portion **(Figure 2.10)**. As expected, the negative control showed no loss of the cassette. Clone IVI100 appeared to show some loss of growth on THY tet, although this was not statistically different, potentially indicating partial loss of the cassette *in vivo*. This may indicate activation of this promoter only occured in a portion of the cells *in vivo*. Clone IVI156, on the other hand, showed complete loss of tet resistance, indicating full *in vivo* activation.

2.4 Discussion

S. pyogenes is a prominent human pathogen that is responsible for over 700 million global infections each year. In order to accomplish this*, S. pyogenes* must be able to colonize its host efficiently. Since humans remain the only natural host, *S. pyogenes* has evolved ways in which to efficiently colonize via the skin or pharynx. Colonization is a complicated process involving many facets from both the bacterium and the host. In order to further understand this process, we took a broad approach and constructed a versatile RIVET system to identify genes in *S. pyogenes* that were induced specifically within the pharyngeal environment using a mouse model.

Following the method for the second generation RIVET, a counter-selection method was initially included in order to make post-*in vivo* selection more efficient. Three different methods were attempted with varying success. The *sacB* gene provided no counter-

Figure 2.10: Verification of promoter activity from clones IVI100 and IVI156

The two clones, IVI100 and IVI156,along with pTRKL2::*cre* as a negative control, were transformed individually into Cas2, grown *in vitro*, and inoculated into HLA-DR4/DQ8 mice (*in vivo*). From each condition, serial 10-fold dilutions were plated on THY agar with or without tet and enumerated. N=6 for *cre* and IVI157, N=7 for IVI100.
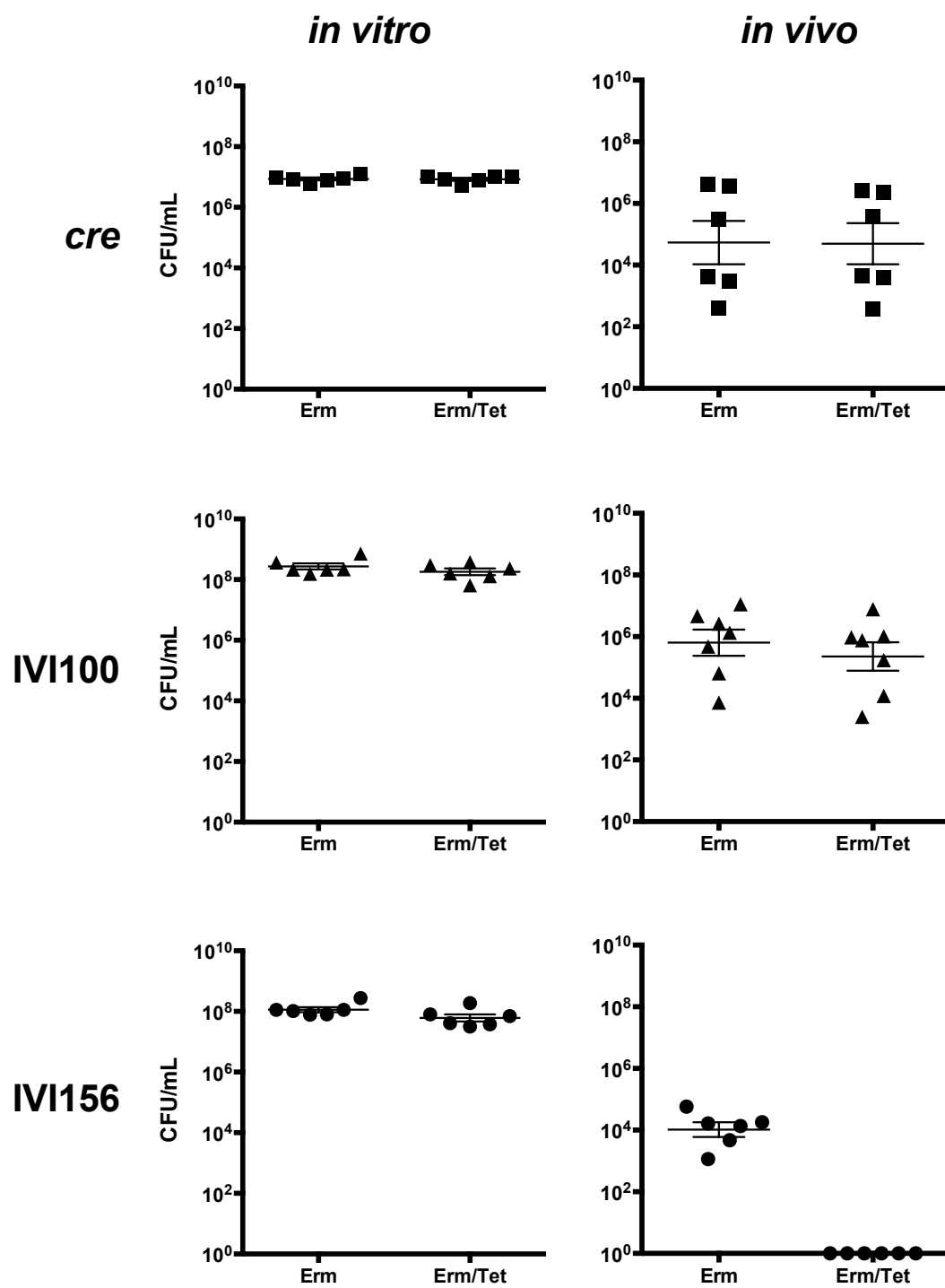
selection phenotype in the presence of sucrose **(Figure 2.2A)**. Although it was shown that the *sacB* gene was transcribed under our experimental conditions, we cannot rule out potential problems with translation or the protein product itself **(Figure 2.2B)**. This problem could have been related to the complete removal of the signal peptide. While the *upp*/5-FU combination was quite robust as a counter-selection system, the polar effects caused by the knockout were undesirable for use *in vivo* **(Figure 2.3)**. Additionally, use of *upp* could not be used broadly in *S. pyogenes,* as each strain would require mutation of the endogenous *upp* gene. One gene that did seem to have some initial success was that of HSV1-*tk* in conjunction with GCV **(Figure 2.4B)**. Unfortunately, continued use produced inconsistent results post-*in vivo*. Similar to *sacB*, transcription of HSV-*tk* was shown **(Figure 2.6A)**, however, the presence of the protein itself was not (unpublished data). This could be due a weak antibody or poor translation of the gene. There is also the possibility that *S. pyogenes* is able to break down GCV and metabolize it, hence the high concentration initially required. Whatever the reason, should counter-selection be explored in the future with *S. pyogenes*, there are still several different methods that could be attempted. Although more laborious, patching colonies onto THY tet to identify clones that had resolved the cassette was successful, therefore this method was used *in lieu* of counter-selection.

Once Cas2 had been created along with the library, *in vitro* promoters were removed and the remaining fragments were exposed to the *in vivo* model. From this pool, 82 unique plasmids were recovered representing *in vivo* induced fragments **(Appendix 2)**. These potential promoters controlled genes that covered a broad range of categories with some noticeable differences between the sense and antisense groups **(Figure 2.9)**. In particular, the antisense fragments created an interesting group as these potential promoters would likely be involved in negative regulation of the corresponding gene product.

Compared to eukaryotes, the bacterial chromosome is quite compact and was once thought to be relatively simple, with protein being the only regulatory mechanisms (37-39). It has only been in the past 12 years that regulatory RNAs are commonly being found and shown to contain important regulatory roles (40). Regulatory RNAs generally fall into one of three categories: an element contained within the 5' untranslated region

(UTR) of the mRNA, *trans*-encoded sRNA that target one or more genes that are located elsewhere on the chromosome, or *cis*-encoded sRNA that are encoded on the opposite strand of the gene they target (40, 41). Multiple studies have recently found putative sRNAs within *S. pyogenes*, supporting the idea that they represent important regulatory mechanisms (42-46). Three of these, *fasX*, *pel*, and *rivX,* have been studied extensively with *fasX* and *rivX* representing *trans*-encoded sRNAs and *pel* representing a *cis*-encoded sRNA. Using microarray analysis, *fasX* (*fibronectin/fibrinogen-binding/haemolytic-activity/streptokinase-regulator-X*) was shown to have many mRNA targets after exposure to human plasma (47). By binding to the 5'UTR of the target mRNA, *fasX* was able to improve mRNA stability and translation for some targets and reduce it for others. It was best known for its ability to increase the expression of streptokinase and streptolysin S while it negatively affected the adhesion genes Fbp54, Mrp, and pili (47-49). The pleiotropic effect locus (*pel*) is a 459 bp transcript found to have a positive regulatory affect on the M protein, streptokinase, SpeB, and streptolysin S. Once the region was sequenced it was discovered that *sagA*, the structural gene of streptolysin S, was within *pel,* while the remaining eight genes of the operon were further downstream. Contained within different reading frames from each other, the termination sequence for *pel* is contained within *sagA* and *sagB* (50-52). While it is unknown how *pel* specifically regulates virulence factor expression, it has been demonstrated that it works at the RNA level by showing continual function even when translation has been prevented (52). Also shown to function without translation, *rivX* is an sRNA that provides a link between the regulators CovRS and Mga. The two-component system CovRS, along with some other unknown environmental factor(s), is able to suppress *rivX*, which is able to enhance Mga expression either directly, or though another regulator. This then has an affect on numerous virulence factors such as *speB* and *emm* (53, 54). This demonstrates the complicated systems used in order to control the expression of bacterial genes.

The RIVET system seems to have uncovered some potentially interesting *cis*-encoded sRNAs. Three of these sRNAs potentially control the virulence factors Mga, SpeA and MF3. Mga is an important regulator controlling greater than 10% of the genome through both activation and repression (55). At its core, Mga controls genes encoding adhesion, invasion, and immune evasion, and therefore is vital during the early stages of

colonization (56, 57). Showing peak expression during exponential growth, Mga causes the self-autoinduction as well as the induction of M and M family proteins, SOF, C5a peptidase, and SIC, among others (56, 58, 59). Virtaneva *et al.* demonstrated this expression pattern along with SpeA, SMEZ, and two different DNases (*spd3* and *sdaD2*) using cynomolgus macaques in an *in vivo* model (60). Continual growth then leads to early and late stationary phases representing a state of persistence, and potentially spread, and continual change of gene expression. This further leads to the repression of *mga* and its associated genes as well as *speA*, *spd3* and *sdaD3* (59, 60). Therefore, it is possible that *S. pyogenes* makes use of sRNA for quick repression of *mga*, *speA*, and *mf3* along with others.

From the group of sense clones, nine appeared to represent typical promoters, driving the expression of a gene(s) **(Table 2.4)**. The corresponding proteins were searched using BLASTp to identify a potential function and three appeared to be involved with metabolism. This is not surprising since *S. pyogenes* is able to grow in multiple niches within the human body including the pharynx, the skin, and also within deeper tissues. Each environment represents its own set of challenges, one in particular being that of nutrient availability. In order to survive, *S. pyogenes* must alter the expression of genes within its chromosome accordingly to make use of the nutrients available.

Within the upper respiratory tract, *S. pyogenes* is most prevalent within the oropharynx, but can also be found within the nasopharynx and the nasal cavity (61). The oropharynx represents the area immediately behind the oral cavity, starting with the soft palate and the base of the tongue, and ending at the posterior pharyngeal wall (62). Running up the pharynx above and beyond the soft palate until reaching the nasal septum is the nasopharynx, past which is the nasal cavity (63, 64). There are few sources from which *S. pyogenes* can obtain nutrients within the pharynx and nasal cavity with relatively the same provisions between niches. Common nutrients include those obtained from the dead and dying cells of the host and other foreign cells, along with molecules from the interstitial fluid of the host. Another major nutritional component in each niche is the fluid created for that area: nasal fluid is present in the nasal cavity, airway surface liquid (ASL) lines the nasopharynx and the posterior of the oropharynx, and saliva coats the oral

cavity and the anterior of the oropharynx (61, 65). These fluids provide many functions including lubrication, protection, and pH balance. The fluids are relatively similar with major differences including water content and the concentration of certain solutes. Nasal cavity fluid and ASL consist of 90-95% water, while saliva consists roughly of 99% water. Some of the major solutes contained within these fluids include proteins, mucin, DNA, urea, and various electrolytes. In addition, saliva can also contain food debris that may provide several nutrients (61, 65-67).

One of the typical promoters involved with metabolism, IVI87, controls the protein pyruvate formate-lyase, which is involved in carbohydrate utilization. The preferred carbon course of *S. pyogenes* is glucose, and when it is found in excess a state of homofermentative glycolysis is maintained. Through glycolysis, glucose is converted to pyruvate, which is then converted to lactate, and excreted (68). However, within the upper respiratory tract there exists a very limited availability of free carbohydrates such as glucose (65). *S. pyogenes* starts to compensate for this by changing to a process of mixed-acid fermentation, allowing for the continued production of adenosine triphosphate (ATP) via the recycling of NAD^+. Mixed-acid fermentation starts with the activation of pyruvate formate-lyase, demonstrated by clone IVI87, to begin converting pyruvate to Acetyl-CoA and formate (69-71). Acetyl-CoA can then be used in two pathways: conversion to acetate creating ATP, or conversion to ethanol in the process of converting NADH back to NAD^+ (72). However, in order to sustain itself within the upper respiratory tract, *S. pyogenes* must also find an alternate carbon source (70, 73, 74).

One important carbon source used by bacteria in the upper respiratory tract, including various types of streptococci, is that of mucin (75-78). Mucin is a large repeating network of interconnecting molecules, reaching up to 1000 kDa, consisting of a protein backbone containing carbohydrate side chains linked via serine or threonine (65, 79). These carbohydrate side chains include fucose, galactose, N-acetylgalactosamine, N-acetylglucosamine, and sialic acid (80). *S. pyogenes* can utilize galactose through the *lac* operon and some strains, but not all, contain fucosidases, sialidases, and other genes that assist in the degradation of N-glycans (81). *S. pyogenes* also possess the enzymes EndoS and hyaluronidase that release N-acetylglucosamine from the Fc region of IgG and

hyaluronate, respectively (65, 82). RIVET also indicated the expression of *spyM18_0784,* via clone IVI72, a tagatose-6-phosphate aldose/ketose isomerase, involved in the pathway for the metabolism of N-acetylgalactosamine (83). This could indicate the use of a N-acetylgalactosamine as a carbon source within the pharynx by *S. pyogenes*.

The last identified activated gene that may be involved with metabolism was *spyM18_1223*, the binding protein for a polar amino acid ABC transporter. The binding protein captures the substrate, delivering it to the membrane domains that transfer it across the cytoplasmic membrane (84, 85). Detection of *spyM18_1223* by other groups has also supported *in vivo* induction. When grown in THY, *spyM18_1223* was shown to be below antibody detection level (86). However, when grown in blood antibody binding was detected and, more importantly, *spyM18_1223* expression was found to have increased two-fold when analyzed from *S. pyogenes* positive pharyngitis patients (87, 88). This suggested that *spyM18_1223* was only expressed under certain metabolic environments, including one that was replicated within the mouse nares.

The next typical promoter activated in the mouse colonization model controlled *spyM18_2257* or *parB*. Binding to *parS* sites present at the *oriC* of the chromosome, ParB is thought to bind to other proteins and aid in chromosome segregation (89). In other streptococcal strains ParB has shown to be expressed as both an operon with the upstream gene *htrA*, as well as an individual gene with its own promoter (90, 91). In this case, Clone IVI49 identified the activation of *parB* individually *in vivo*. Since chromosome segregation would be occurring *in vitro*, it is possible that *parB* was only being expressed in the operon form *in vitro* but switched to the individual form *in vivo* to be picked up by RIVET. This change in expression could occur due to changes in the nutrients available and the appearance of other means of stress.

For three promoters, *spyM18_0197, spyM18_0414*, and *spyM18_2004,* very little is known of the controlled gene(s) other than the domains present on the potential protein that is produced. The protein product of *spyM18_0414* contains a single cystathionine β-synthase (CBS) pair (domain), which can be found in all three kingdoms of life. CBS domains can be found in proteins of a wide range of functions including channels, kinases, and metabolic enzymes (92, 93). The function of this particular protein remains

unknown. The second protein, encoded by *spyM18_0197*, contains a transglutaminase domain. Transglutaminases are enzymes responsible for a variety of post-translational modifications to proteins, principally introducing a cross-link between lysine and glutamine residues (94). Some pathogens highjack the transglutaminase, such as *S. aureus,* which uses the host transglutaminase of plasma factor XIIIa to cross-link itself to fibrin (95). Other bacteria produce toxins that are transglutaminases that target host Rho proteins, modifying them post-translationally and impairing their GTPase activity (96). Finally, *Francisella tularensis* demonstrated attenuated growth in macrophages and mice when a protein with a transglutaminase domain was deleted (97). How exactly *S. pyogenes* uses *spyM18_0197* has yet to be determined, but it has been shown to be useful for the survival of other organisms. The third protein was *spyM18_2004,* containing a PilT N-terminus (PIN) domain. PIN domains are found in all three kingdoms of life and function in signalling and riboexonuclease activity (40, 98, 99). Once again, it is unknown how exactly this protein is used by *S. pyogenes*.

Lastly, two of the typical promoters had been selected for further study, the first being *spyM18_0771,* a structural component of ΦSpeC. The exact function of *spyM18_0771* is unknown, however it is similar to Gp58, a minor structural protein from a *Lactococcus delbrueckii* bacteriophage. While there was no activation from this promoter *in vitro*, there only appeared to be a subpopulation of the *in vivo* group in which the promoter may have been activated **(Figure 2.10)**. Unpublished observations in our lab have shown the spontaneous excision of ΦSpeC *in vitro* at the genetic level using PCR, and this bacteriophage element can be induced using mitomycin C. Other groups have also shown the production of active ΦSpeC upon co-culture with human pharyngeal cells (100). We believe that during *in vitro* growth, the phage DNA may be recombining in and out of the chromosome, but not progressing any further in the lytic phase. *In vivo*, an unidentified signal may induce the phage, including structural components, allowing for the detection of the *spyM18_0771* promoter. Why only a certain number of cells seem to express the phage *in vivo* could be due to microenvironments within the mouse cNT.

The second promoter selected for further testing was upstream of *spyM18_0544*, which showed full *in vivo* activation **(Figure 2.10)**. The *spyM18_0544* gene encodes a putative

bacteriocin-like peptide with 50% similarity to the *blpM* bacteriocin gene of *S. pneumoniae*. While bacteriocins, in particular lantibiotics, have been characterized from *S. pyogenes*, there are no reports to our knowledge of functional Class II bacteriocins from this organism. As a human pathogen, virulence factors involved with the interaction with the human body garner much more attention. However, when interacting with the human host, *S. pyogenes* will encounter many other species of bacteria, some of which are vying for the same space. The outcome of these interactions will also likely be critical because failure to compete with the natural microbiota will also prevent colonization from occurring. Further work on this mechanism is required to determine how *S. pyogenes* might make use of this tool when colonizing.

The development of the RIVET system for the first time in *S. pyogenes* was a successful endeavour that led to the identification of numerous *in vivo*-induced genes. RIVET was able to identify not only just genes, but also potentially sRNA under the control of an inducible promoter. Although further validation is necessary, these promoters help demonstrate the complex nature at which the chromosome and its elements are regulated and how multifaceted the colonization process is at the molecular level. RIVET was able to identify promoters from several different categories with many including metabolism and virulence factors. One very interesting promoter identified was IVI156, controlling a bacteriocin-like protein. A bacteriocin would make an important niche adaptation tool for *S. pyogenes* as it may allow for the elimination of competing bacteria. Further research will determine how *S. pyogenes* uses it as part of its colonization process.

2.5 References

1.   **Mahan MJ**, **Slauch JM**, **Mekalanos JJ**. 1993. Selection of bacterial virulence genes that are specifically induced in host tissues. Science **259**:686–688.
2.   **Wang J**, **Mushegian A**, **Lory S**, **Jin S**. 1996. Large-scale isolation of candidate virulence genes of *Pseudomonas aeruginosa* by *in vivo* selection. Proc Natl Acad Sci USA **93**:10434–10439.
3.   **Wang J**, **Lory S**, **Ramphal R**, **Jin S**. 1996. Isolation and characterization of *Pseudomonas aeruginosa* genes inducible by respiratory mucus derived from cystic fibrosis patients. Molecular Microbiology **22**:1005–1012.
4.   **Mahan MJ**, **Tobias JW**, **Slauch JM**, **Hanna PC**, **Collier RJ**, **Mekalanos JJ**. 1995. Antibiotic-based selection for bacterial genes that are specifically induced during infection of a host. Proc Natl Acad Sci USA **92**:669–673.
5.   **Heithoff DM**, **Conner CP**, **Hanna PC**, **Julio SM**, **Hentschel U**, **Mahan MJ**. 1997. Bacterial infection as assessed by *in vivo* gene expression. Proc Natl Acad Sci USA **94**:934–939.
6.   **Young GM**, **Miller VL**. 1997. Identification of novel chromosomal loci affecting *Yersinia enterocolitica* pathogenesis. Molecular Microbiology **25**:319–328.
7.   **Gort AS**, **Miller VL**. 2000. Identification and characterization of *Yersinia enterocolitica* genes induced during systemic infection. Infection and Immunity **68**:6633–6642.
8.   **Camilli A**, **Beattie DT**, **Mekalanos JJ**. 1994. Use of genetic recombination as a reporter of gene expression. Proc Natl Acad Sci USA **91**:2634–2638.
9.   **Camilli A**, **Mekalanos JJ**. 1995. Use of recombinase gene fusions to identify *Vibrio cholerae* genes induced during infection. Molecular Microbiology **18**:671–683.
10.   **Lowe AM**, **Beattie DT**, **Deresiewicz RL**. 1998. Identification of novel staphylococcal virulence genes by *in vivo* expression technology. Molecular Microbiology **27**:967–976.
11.   **Angelichio MJ**, **Camilli A**. 2002. *In vivo* expression technology. Infection and Immunity **70**:6518–6523.
12.   **Bron PA**, **Grangette C**, **Mercenier A**, **de Vos WM**, **Kleerebezem M**. 2004. Identification of *Lactobacillus plantarum* genes that are induced in the gastrointestinal tract of mice. Journal of Bacteriology **186**:5721–5729.
13.   **Veal-Carr WL**, **Stibitz S**. 2004. Demonstration of differential virulence gene promoter activation *in vivo* in *Bordetella pertussis* using RIVET. Molecular Microbiology **55**:788–798.
14.   **Castillo AR**, **Woodruff AJ**, **Connolly LE**, **Sause WE**, **Ottemann KM**. 2008. Recombination-based *in vivo* expression technology identifies *Helicobacter pylori* genes important for host colonization. Infection and Immunity **76**:5632–5644.
15.   **Junjua M**, **Galia W**, **Gaci N**, **Ophélie U**, **Genay M**, **Bachmann H**, **Kleerebezem M**, **Dary A**, **Roussel Y**. 2013. Development of the recombinase-based *in vivo* expression technology in *Streptococcus thermophilus* and validation using the lactose operon promoter. J Appl Microbiol.
16.   **Saviola B**, **Woolwine SC**, **Bishai WR**. 2003. Isolation of acid-inducible genes of

Mycobacterium tuberculosis with the use of recombinase-based *in vivo* expression technology. Infection and Immunity **71**:1379–1388.

17. **Pelicic V**, **Reyrat JM**, **Gicquel B**. 1996. Expression of the *Bacillus subtilis sacB* gene confers sucrose sensitivity on mycobacteria. Journal of Bacteriology **178**:1197–1199.

18. **Osorio CG**, **Crawford JA**, **Michalski J**, **Martinez-Wilson H**, **Kaper JB**, **Camilli A**. 2005. Second-generation recombination-based *in vivo* expression technology for large-scale screening for *Vibrio cholerae* genes induced during infection of the mouse small intestine. Infection and Immunity **73**:972–980.

19. **Frank KL**, **Barnes AMT**, **Grindle SM**, **Manias DA**, **Schlievert PM**, **Dunny GM**. 2012. Use of recombinase-based *in vivo* expression technology to characterize *Enterococcus faecalis* gene expression during infection identifies *in vivo*-expressed antisense RNAs and implicates the protease Eep in pathogenesis. Infection and Immunity **80**:539–549.

20. **Holt JF**, **Kiedrowski MR**, **Frank KL**, **Du J**, **Guan C**, **Broderick NA**, **Dunny GM**, **Handelsman J**. 2014. *Enterococcus faecalis* 6-phosphogluconolactonase is required for both commensal and pathogenic interactions with *Manduca sexta*. Infection and Immunity.

21. **Abremski K**, **Hoess R**. 1984. Bacteriophage P1 site-specific recombination. J Biol Chem **259**:1509–1514.

22. **Hoess RH**, **Abremski K**. 1985. Mechanism of strand cleavage and exchange in the Cre-*lox* site-specific recombination system. Journal of Molecular Biology **181**:351–362.

23. **Smoot J**, **Barbian K**, **Van Gompel J**, **Smoot L**, **Chaussee MS**, **Sylva GL**, **Sturdevant DE**, **Ricklefs SM**, **Porcella SF**, **Parkins LD**, **Beres SB**, **Campbell DS**, **Smith TM**, **Zhang Q**, **Kapur V**, **Daly JA**, **Veasy LG**, **Musser JM**. 2002. Genome sequence and comparative microarray analysis of serotype M18 group A *Streptococcus* strains associated with acute rheumatic fever outbreaks. Proc Natl Acad Sci USA **99**:4668–4673.

24. **Kunst F**, **Ogasawara N**, **Moszer I**, **Albertini AM**, **Alloni G**, **Azevedo V**, **Bertero MG**, **Bessières P**, **Bolotin A**, **Borchert S**, **Borriss R**, **Boursier L**, **Brans A**, **Braun M**, **Brignell SC**, **Bron S**, **Brouillet S**, **Bruschi CV**, **Caldwell B**, **Capuano V**, **Carter NM**, **Choi SK**, **Cordani JJ**, **Connerton IF**, **Cummings NJ**, **Daniel RA**, **Denziot F**, **Devine KM**, **Düsterhöft A**, **Ehrlich SD**, **Emmerson PT**, **Entian KD**, **Errington J**, **Fabret C**, **Ferrari E**, **Foulger D**, **Fritz C**, **Fujita M**, **Fujita Y**, **Fuma S**, **Galizzi A**, **Galleron N**, **Ghim SY**, **Glaser P**, **Goffeau A**, **Golightly EJ**, **Grandi G**, **Guiseppi G**, **Guy BJ**, **Haga K**, **Haiech J**, **Harwood CR**, **Hènaut A**, **Hilbert H**, **Holsappel S**, **Hosono S**, **Hullo MF**, **Itaya M**, **Jones L**, **Joris B**, **Karamata D**, **Kasahara Y**, **Klaerr-Blanchard M**, **Klein C**, **Kobayashi Y**, **Koetter P**, **Koningstein G**, **Krogh S**, **Kumano M**, **Kurita K**, **Lapidus A**, **Lardinois S**, **Lauber J**, **Lazarevic V**, **Lee SM**, **Levine A**, **Liu H**, **Masuda S**, **Mauël C**, **Medigue C**, **Medina N**, **Mellado RP**, **Mizuno M**, **Moestl D**, **Nakai S**, **Noback M**, **Noone D**, **O'Reilly M**, **Ogawa K**, **Ogiwara A**, **Oudega B**, **Park SH**, **Parro V**, **Pohl TM**, **Portelle D**, **Porwollik S**, **Prescott AM**, **Presecan E**, **Pujic P**, **Purnelle B**, **Rapoport G**, **Rey M**, **Reynolds S**, **Rieger M**, **Rivolta C**, **Rocha E**, **Roche B**, **Rose M**, **Sadaie Y**, **Sato T**, **Scanlan E**, **Schleich S**, **Schroeter R**, **Scoffone F**, **Sekiguchi J**, **Sekowska A**, **Seror SJ**, **Serror P**,

Shin BS, **Soldo B**, **Sorokin A**, **Tacconi E**, **Takagi T**, **Takahashi H**, **Takemaru K**, **Takeuchi M**, **Tamakoshi A**, **Tanaka T**, **Terpstra P**, **Togoni A**, **Tosato V**, **Uchiyama S**, **Vandebol M**, **Vannier F**, **Vassarotti A**, **Viari A**, **Wambutt R**, **Wedler H**, **Weitzenegger T**, **Winters P**, **Wipat A**, **Yamamoto H**, **Yamane K**, **Yasumoto K**, **Yata K**, **Yoshida K**, **Yoshikawa HF**, **Zumstein E**, **Yoshikawa H**, **Danchin A**. 1997. The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis.* Nature **390**:249–256.

25. **O'Sullivan DJ**, **Klaenhammer TR**. 1993. High- and low-copy-number *Lactococcus shuttle* cloning vectors with features for clone screening. Gene **137**:227–231.

26. **Biswas I**, **Gruss A**, **Ehrlich SD**, **Maguin E**. 1993. High-efficiency gene inactivation and replacement system for Gram-positive bacteria. Journal of Bacteriology **175**:3628.

27. **Bryan EM**, **Bae T**, **Kleerebezem M**, **Dunny GM**. 2000. Improved vectors for nisin-controlled expression in Gram-positive bacteria. Plasmid **44**:183–190.

28. **Guérout-Fleury AM**, **Shazand K**, **Frandsen N**, **Stragier P**. 1995. Antibiotic-resistance cassettes for *Bacillus subtilis.* Gene **167**:335–336.

29. **Nooh MM**, **El-Gengehi N**, **Kansal R**, **David CS**, **Kotb M**. 2007. HLA transgenic mice provide evidence for a direct and dominant role of HLA class II variation in modulating the severity of streptococcal sepsis. J Immunol **178**:3076–3083.

30. **Kasper KJ**, **Zeppa JJ**, **Wakabayashi AT**, **Xu SX**, **Mazzuca DM**, **Welch I**, **Baroja ML**, **Kotb M**, **Cairns E**, **Cleary PP**, **Haeryfar SMM**, **McCormick JK**. 2014. Bacterial superantigens promote acute nasopharyngeal infection by *Streptococcus pyogenes* in a human MHC class II-dependent manner. PLoS Pathog **10**:e1004155.

31. **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**:1591–1601.

32. **Lederberg J**. 1951. Streptomycin resistance; a genetically recessive mutation. Journal of Bacteriology **61**:549–550.

33. **Reyrat JM**, **Pelicic V**, **Gicquel B**, **Rappuoli R**. 1998. Counterselectable markers: untapped tools for bacterial genetics and pathogenesis. Infection and Immunity **66**:4011.

34. **Bramucci MG**, **Nagarajan V**. 1996. Direct selection of cloned DNA in *Bacillus subtilis* based on sucrose-induced lethality. Applied and Environmental Microbiology **62**:3948–3953.

35. **Fabret C**, **Ehrlich SD**, **Noirot P**. 2002. A new mutation delivery system for genome-scale approaches in *Bacillus subtilis*. Molecular Microbiology **46**:25–36.

36. **de Hoon MJL**, **Makita Y**, **Nakai K**, **Miyano S**. 2005. Prediction of transcriptional terminators in *Bacillus subtilis* and related species. PLoS Comp Biol **1**:e25.

37. **Rasmussen S**, **Nielsen HB**, **Jarmer H**. 2009. The transcriptionally active regions in the genome of *Bacillus subtilis*. Molecular Microbiology **73**:1043–1057.

38. **Georg J**, **Hess WR**. 2011. *cis*-Antisense RNA, another level of gene regulation in bacteria. Microbiol Mol Biol Rev **75**:286–300.

39. **Toledo-Arana A**, **Solano C**. 2010. Deciphering the physiological blueprint of a

bacterial cell. Bioessays **32**:461–467.

40. **Sesto N**, **Wurtzel O**, **Archambaud C**, **Sorek R**, **Cossart P**. 2012. The excludon: a new concept in bacterial antisense RNA-mediated gene regulation. Nature Publishing Group **11**:75–82.

41. **Chhatwal GS**. 2013. Host-pathogen interactions in streptococcal diseases.

42. **Perez N**, **Treviño J**, **Liu Z**, **Ho SCM**, **Babitzke P**, **Sumby P**. 2009. A genome-wide analysis of small regulatory RNAs in the human pathogen group A *Streptococcus*. PLoS ONE **4**:e7668.

43. **Mraheil MA**, **Billion A**, **Kuenne C**, **Pischimarov J**, **Kreikemeyer B**, **Engelmann S**, **Hartke A**, **Giard J-C**, **Rupnik M**, **Vorwerk S**, **Beier M**, **Retey J**, **Hartsch T**, **Jacob A**, **Cemič F**, **Hemberger J**, **Chakraborty T**, **Hain T**. 2010. Comparative genome-wide analysis of small RNAs of major Gram-positive pathogens: from identification to application. Microbial Biotechnology **3**:658–676.

44. **Deltcheva E**, **Chylinski K**, **Sharma CM**, **Gonzales K**, **Chao Y**, **Pirzada ZA**, **Eckert MR**, **Vogel J**, **Charpentier E**. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature **471**:602–607.

45. **Patenge N**, **Billion A**, **Raasch P**, **Normann J**, **Wisniewska-Kucper A**, **Retey J**, **Boisguérin V**, **Hartsch T**, **Hain T**, **Kreikemeyer B**. 2012. Identification of novel growth phase- and media-dependent small non-coding RNAs in *Streptococcus pyogenes* M49 using intergenic tiling arrays. BMC Genomics **13**:550.

46. **Tesorero RA**, **Yu N**, **Wright JO**, **Svencionis JP**, **Cheng Q**, **Kim J-H**, **Cho KH**. 2013. Novel regulatory small RNAs in *Streptococcus pyogenes*. PLoS ONE **8**:e64021.

47. **Liu Z**, **Treviño J**, **Ramirez-Peña E**, **Sumby P**. 2012. The small regulatory RNA FasX controls pilus expression and adherence in the human bacterial pathogen group A *Streptococcus*. Molecular Microbiology **86**:140–154.

48. **Kreikemeyer B**, **Boyle MD**, **Buttaro BA**, **Heinemann M**, **Podbielski A**. 2001. Group A streptococcal growth phase-associated virulence factor regulation by a novel operon (Fas) with homologies to two-component-type regulators requires a small RNA molecule. Molecular Microbiology **39**:392–406.

49. **Ramirez-Peña E**, **Treviño J**, **Liu Z**, **Perez N**, **Sumby P**. 2010. The group A *Streptococcus* small regulatory RNA FasX enhances streptokinase activity by increasing the stability of the *ska* mRNA transcript. Molecular Microbiology **78**:1332–1347.

50. **Li Z**, **Sledjeski DD**, **Kreikemeyer B**, **Podbielski A**, **Boyle MD**. 1999. Identification of *pel*, a *Streptococcus pyogenes* locus that affects both surface and secreted proteins. Journal of Bacteriology **181**:6019–6027.

51. **Nizet V**, **Beall B**, **Bast DJ**, **Datta V**, **Kilburn L**, **Low DE**, **De Azavedo JC**. 2000. Genetic locus for streptolysin S production by group A *Streptococcus*. Infection and Immunity **68**:4245–4254.

52. **Mangold M**, **Siller M**, **Roppenser B**, **Vlaminckx BJM**, **Penfound TA**, **Klein R**, **Novak R**, **Novick RP**, **Charpentier E**. 2004. Synthesis of group A streptococcal virulence factors is controlled by a regulatory RNA molecule. Molecular Microbiology **53**:1515–1527.

53. **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.

54. **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.

55. **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.

56. **Hondorp ER**, **McIver KS**. 2007. The Mga virulence regulon: infection where the grass is greener. Molecular Microbiology **66**:1056–1065.

57. **Collin M**, **Schuch R**. 2009. Bacterial sensing and signaling.

58. **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.

59. **Kreikemeyer B**, **McIver KS**, **Podbielski A**. 2003. Virulence factor regulation and regulatory networks in *Streptococcus pyogenes* and their impact on pathogen–host interactions. Trends in Microbiology **11**:224–232.

60. **Virtaneva K**, **Porcella SF**, **Graham MR**, **Ireland RM**, **Johnson CA**, **Ricklefs SM**, **Babar I**, **Parkins LD**, **Romero RA**, **Corn GJ**, **Gardner DJ**, **Bailey JR**, **Parnell MJ**, **Musser JM**. 2005. Longitudinal analysis of the group A *Streptococcus* transcriptome in experimental pharyngitis in cynomolgus macaques. Proc Natl Acad Sci USA **102**:9014–9019.

61. **Wilson M**. 2008. Bacteriology of humans: an ecological perspective. John Wiley & Sons.

62. **Shah S**, **Garritano FG**. 2015. Pediatric oral anatomy. Operative Techniques in Otolaryngology - Head and Neck Surgery **26**:2–7.

63. **Harkema J**, **Carey S**, **Wagner J**. 2006. The nose revisited: a brief review of the comparative structure, function, and toxicologic pathology of the nasal epithelium. Toxicologic Pathology **34**:252–269.

64. **Miller AJ**. 2002. Oral and pharyngeal reflexes in the mammalian nervous system: their diverse range in complexity and the pivotal role of the tongue. Crit Rev Oral Biol Med **13**:409–425.

65. **Wilson M**. 2005. Microbial inhabitants of humans. Cambridge University Press.

66. **Humphrey SP**, **Williamson RT**. 2001. A review of saliva: normal composition, flow, and function. J Prosthet Dent **85**:162–169.

67. **Dodds MWJ**, **Johnson DA**, **Yeh C-K**. 2005. Health benefits of saliva: a review. Journal of Dentistry **33**:223–233.

68. **Levering J**, **Musters MWJM**, **Bekker M**, **Bellomo D**, **Fiedler T**, **de Vos WM**, **Hugenholtz J**, **Kreikemeyer B**, **Kummer U**, **Teusink B**. 2012. Role of phosphate in the central metabolism of two lactic acid bacteria - a comparative systems biology approach. FEBS Journal **279**:1274–1290.

69. **Neijssel OM**, **Snoep JL**, **Teixeira de Mattos MJ**. 1997. Regulation of energy source metabolism in streptococci. Soc Appl Bacteriol Symp Ser **26**:12S–19S.

70. **Wood DN**, **Weinstein KE**, **Podbielski A**, **Kreikemeyer B**, **Gaughan JP**, **Valentine S**, **Buttaro BA**. 2009. Generation of metabolically diverse strains of *Streptococcus pyogenes* during survival in stationary phase. Journal of Bacteriology **191**:6242–6252.

71. **Anbalagan S**, **Dmitriev A**, **McShan WM**, **Dunman PM**, **Chaussee MS**. 2012. Growth phase-dependent modulation of Rgg binding specificity in *Streptococcus pyogenes*. Journal of Bacteriology **194**:3961–3971.

72. **Yamamoto Y**, **Sato Y**, **Takahashi-Abbe S**, **Takahashi N**, **Kizaki H**. 2000. Characterization of the *Streptococcus mutans* pyruvate formate-lyase (PFL)-activating enzyme gene by complementary reconstitution of the *in vitro* PFL-reactivating system. Infection and Immunity **68**:4773–4777.

73. **Shelburne SA**, **Sumby P**, **Sitkiewicz I**, **Okorafor N**, **Granville C**, **Patel P**, **Voyich J**, **Hull R**, **DeLeo FR**, **Musser JM**. 2006. Maltodextrin utilization plays a key role in the ability of group A *Streptococcus* to colonize the oropharynx. Infection and Immunity **74**:4605–4614.

74. **Shelburne SA**, **Keith D**, **Horstmann N**, **Sumby P**, **Davenport MT**, **Graviss EA**, **Brennan RG**, **Musser JM**. 2008. A direct link between carbohydrate utilization and virulence in the major human pathogen group A *Streptococcus*. Proc Natl Acad Sci USA **105**:1698–1703.

75. **Mothey D**, **Buttaro BA**, **Piggot PJ**. 2013. Mucin can enhance growth, biofilm formation, and survival of *Streptococcus mutans*. FEMS Microbiology Letters **350**:161–167.

76. **Byers HL**, **Homer KA**, **Beighton D**. 1996. Utilization of sialic acid by viridans streptococci. J Dent Res **75**:1564–1571.

77. **Yesilkaya H**, **Manco S**, **Kadioglu A**, **Terra VS**, **Andrew PW**. 2008. The ability to utilize mucin affects the regulation of virulence gene expression in *Streptococcus pneumoniae*. FEMS Microbiology Letters **278**:231–235.

78. **Van der Hoeven JS**, **Van Den Kieboom C**. 1990. Utilization of mucin by oral *Streptococcus* species. Antonie Van Leeuwenhoek **57**:165–172.

79. **Parham P**. 2014. The immune system. Garland Science.

80. **Slomiany BL**, **Murty VL**, **Piotrowski J**, **Slomiany A**. 1996. Salivary mucins in oral mucosal defense. Gen Pharmacol **27**:761–771.

81. **Stepper J**, **Dabin J**, **Eklof JM**, **Thongpoo P**, **Kongsaeree P**, **Taylor EJ**, **Turkenburg JP**, **Brumer H**, **Davies GJ**. 2013. Structure and activity of the *Streptococcus pyogenes* family GH1 6-phospho-β-glucosidase SPy1599. Acta Crystallogr D Biol Crystallogr **69**:16–23.

82. **Collin M**, **Olsen A**. 2001. EndoS, a novel secreted protein from *Streptococcus pyogenes* with endoglycosidase activity on human IgG. EMBO J **20**:3046–3055.

83. **Hu Z**, **Patel IR**, **Mukherjee A**. 2013. Genetic analysis of the roles of *agaA*, *agaI*, and *agaS* genes in the N-acetyl-D-galactosamine and D-galactosamine catabolic pathways in *Escherichia coli* strains O157:H7 and C. BMC Microbiol **13**:94.

84. **Beek ter J**, **Guskov A**, **Slotboom DJ**. 2014. Structural diversity of ABC transporters. The Journal of General Physiology **143**:419–435.

85. **Eitinger T**, **Rodionov DA**, **Grote M**, **Schneider E**. 2010. Canonical and ECF-type ATP-binding cassette importers in prokaryotes: diversity in modular organization and cellular functions. FEMS Microbiology Reviews **35**:3–67.

86. **Severin A**, **Nickbarg E**, **Wooters J**, **Quazi SA**, **Matsuka YV**, **Murphy E**, **Moutsatsos IK**, **Zagursky RJ**, **Olmsted SB**. 2007. Proteomic analysis and identification of *Streptococcus pyogenes* surface-associated proteins. Journal of Bacteriology **189**:1514–1522.

87. **Lei B**, **Liu M**, **Chesney GL**, **Musser JM**. 2004. Identification of new candidate vaccine antigens made by *Streptococcus pyogenes*: purification and characterization of 16 putative extracellular lipoproteins. J Infect Dis **189**:79–89.

88. **Livezey J**, **Perez L**, **Suciu D**, **Yu X**, **Robinson B**, **Bush D**, **Merrill G**. 2011. Analysis of group A *Streptococcus* gene expression in humans with pharyngitis using a microarray. Journal of Medical Microbiology **60**:1725–1733.

89. **Pinho MG**, **Kjos M**, **Veening J-W**. 2013. How to get (a)round: mechanisms controlling growth and division of coccoid bacteria. Nature Publishing Group **11**:601–614.

90. **Ahn SJ**, **Lemos JAC**, **Burne RA**. 2005. Role of HtrA in Growth and competence of *Streptococcus mutans* UA159. Journal of Bacteriology **187**:3028–3038.

91. **Minnen A**, **Attaiech L**, **Thon M**, **Gruber S**, **Veening J-W**. 2011. SMC is recruited to *oriC* by ParB and promotes chromosome segregation in *Streptococcus pneumoniae*. Molecular Microbiology **81**:676–688.

92. **Bateman A**. 1997. The structure of a domain common to archaebacteria and the homocystinuria disease protein. Trends Biochem Sci **22**:12–13.

93. **Ignoul S**. 2005. CBS domains: structure, function, and pathology in human proteins. AJP: Cell Physiology **289**:C1369–C1378.

94. **Milani A**, **Vecchietti D**, **Rusmini R**, **Bertoni G**. 2012. TgpA, a protein with a eukaryotic-like transglutaminase domain, plays a critical role in the viability of *Pseudomonas aeruginosa*. PLoS ONE **7**:e50323.

95. **El-Etr SH**, **Subbian S**, **Cirillo SLG**, **Cirillo JD**. 2004. Identification of two *Mycobacterium marinum* loci that affect interactions with macrophages. Infection and Immunity **72**:6902–6913.

96. **Lemonnier M**, **Landraud L**, **Lemichez E**. 2007. Rho GTPase-activating bacterial toxins: from bacterial virulence regulation to eukaryotic cell biology. FEMS Microbiology Reviews **31**:515–534.

97. **Brotcke A**, **Weiss DS**, **Kim CC**, **Chain P**, **Malfatti S**, **Garcia E**, **Monack DM**. 2006. Identification of MglA-regulated genes reveals novel virulence factors in *Francisella tularensis*. Infection and Immunity **74**:6642–6655.

98. **Arcus VL**, **Backbro K**, **Roos A**, **Daniel EL**, **Baker EN**. 2004. Distant structural homology leads to the functional characterization of an archaeal PIN domain as an exonuclease. Journal of Biological Chemistry **279**:16471–16478.

99. **Anantharaman V**, **Aravind L**. 2006. The NYN domains: novel predicted RNAses with a PIN domain-like fold. RNA Biol **3**:18–27.

100. **Broudy TB**, **Pancholi V**, **Fischetti VA**. 2001. Induction of lysogenic bacteriophage and phage-associated toxin from group A streptococci during coculture with human pharyngeal cells. Infection and Immunity **69**:1440–1443.

CHAPTER 3: BACTERIOCINS AS A COLONIZATION TOOL FOR *STREPTOCOCCUS PYOGENES*

3.1 Introduction

3.1.1 Competition for the ideal niche

In nature, almost all microbes are found growing and interacting in complex systems, however our knowledge of these networks still remains incomplete largely due to the inability to culture the majority of microbial species (1). Living in symbiosis with the human body are bacteria, archaea, viruses, fungi, and protists. These microbes growing within us can be mutual (beneficial to both), commensal (beneficial to one, the other is unaffected), or parasitic (one growing at the expense of the other) (2). The adult human body is host to approximately 10^{14} cells within various niches, each home to multiple species of bacteria (1-4). These microbes help maintain a healthy existence by promoting differentiation of host tissues, liberating nutrients from unusable dietary substances, stimulating the immune system, and importantly, protecting from invasion by pathogens (5). Discussed earlier, the body defends against pathogens utilizing physical barriers, antimicrobial peptides, and the complement and immune systems. In addition to this, the microbiota creates an unfavourable environment or competes directly with pathogens, a process referred to as "colonization resistance" (5, 6). However, sometimes a disruption can occur within this homeostatic balance allowing an exogenous organism to colonize a particular niche, which can be detrimental when this invader is a pathogen (7, 8).

Favourable niches on the human body where *S. pyogenes* is able to grow and survive are the skin and pharynx, each having a unique environment and microbiota. In order to live within a favourable habitat, bacteria possess mechanisms, or tools, in order to provide a competitive advantage over other organisms. These include secondary metabolites such as antibiotics, glycolipids, toxins, non-ribosomal antimicrobial peptides, and bacteriocins (9-11). While *S. pyogenes* maintains an arsenal of virulence factors that target host defences, the microbiota provide a different challenge. In Chapter 2, the use of the RIVET system showed the activation of the bacteriocin gene *spyM18_0544* when *S. pyogenes* MGAS8232 was exposed to the nasopharynx of the mouse model. Use of a bacteriocin may provide the ability for *S. pyogenes* to overcome colonization resistance presented by the microbiota of the host. Further work shown here attempted to determine

the role of this bacteriocin as a necessary tool for colonization by *S. pyogenes*.

3.1.1.1 The skin

The human skin has a large microbiota consisting of bacteria, viruses, and eukaryotic organisms. It can be an inhospitable environment for many microbes due to the low pH, lack of moisture, a temperature below that of 37°C, and the presence of antimicrobial compounds such as fatty acids, peptides, and enzymes. In addition to this, the continual sloughing of dead skin cells on the surface can also prevent bacteria from colonizing (2, 9, 12). The main organisms making up the microbiota on the skin consist of the genera *Corynebacterium, Staphylococcus, Propionibacterium, Micrococcus, Malassezia, Brevibacterium, Dermabacter,* and *Acinetobacter* (2, 11). These microbiota utilize many methods to prevent pathogens such as *S. pyogenes* from colonizing. For example, *S. epidermidis* produces antimicrobial peptides and *Propionibacterium acnes* produces fatty acids, both that are harmful to *S. pyogenes* (13, 14). *S. pyogenes* can be found on the skin, but it is often in a transient stage, and rarely part of the natural microbiota. In the transient stage, *S. pyogenes* is unable to grow or multiply, therefore, leaving them unable to colonize asymptomatically. However, in some instances *S. pyogenes* is able to disseminate in a local area to cause an infection (2, 8, 12, 15).

3.1.1.2 The oral cavity

The human oral cavity is a unique environment containing different areas such as the teeth, gums, cheeks, tongue, hard and soft palate, tonsils, pharynx, and esophagus, each having its own microbiota (9, 16). It is a moist environment, kept at a relatively constant temperature between 34°C and 36°C, with most areas maintaining a pH close to neutral (7, 17). The human mouth contains Archaea, amoeba, fungi, and bacteria with the most common phyla in the pharynx being Firmicutes, Bacteroidetes, Proteobacteria, Actinobacteria, and Fusobacteria (2, 3, 18, 19). Within the mouth, *S. pyogenes* colonizes the tonsils within the nasopharynx and oropharynx and can be found asymptomatically ranging from 5% to 10% of healthy adults and ranging as high as 20% in school children during the winter months (20). Within the oral cavity, *S. pyogenes* faces multiple mechanisms of colonization resistance. Lactobacilli exist within the oral tract and the

lactic acid produced by these bacteria has been shown to degrade the LTA of *S. pyogenes* (7, 15). In addition, lactobacilli work indirectly by decreasing production of IL-17 and IL-23 through the induction of TLR2/TLR4, preventing an inflammatory response and decreasing the cytotoxic response and growth of *S. pyogenes* (21). Streptococcal species also play an important role as they make up a large proportion, if not the majority, of the oral microbiota (16-19). Multiple strains of *Streptococcus oralis* and *S. salivarius* were able to show antagonistic properties against *S. pyogenes*. Planktonic *S. oralis* and *S. salivarius* prevented the growth of *S. pyogenes* when grown together, possibly indicating nutrient competition, while pre-treatment with either individually also reduced *S. pyogenes* cell adhesion (8, 22). Similar to lactobacilli, *S. salivarius* strains also caused immune modulation via a decrease in IL-6, TNF-α, and IL-1β, reducing the inflammatory response (22). Finally, *S. salivarius* also possesses a bacteriocin with potent antibacterial activity against *S. pyogenes* (8, 22). As demonstrated, the resident bacteria provide a difficult obstacle to overcome for successful colonization by *S. pyogenes*. Therefore, the production of an antimicrobial directly affecting the host microbiota, such as a bacteriocin, could allow *S. pyogenes* to circumvent their defences and more easily colonize.

3.1.2 Bacteriocins

Bacteriocins are ribosomally synthesized polypeptides that possess antibacterial activity. Produced by both Gram-positive and Gram-negative bacteria, bacteriocins can be active against single species or across genera (15, 21, 23, 24). Importantly, bacteria also possess mechanisms making them immune to their own bacteriocins. Bacteriocins range in terms of size, target, mode of action, and mechanism of immunity, with activity between nano and picomolar concentrations (24-26). It is estimated that upwards of 99% of the Bacteria and Archaea have at least one bacteriocin, but extensive genome analyses would be required to make this more accurate (24). Research on this area, particularly with Gram-positives, has become more popular due to the realization that they are far more prevalent than previously thought, and because they are proving to have useful applications in several industries.

3.1.2.1 Potential Uses

Since bacteriocins are potent antimicrobials they have potential commercial uses in various fields. Bacteriocins with the broadest spectrum of activity, in most cases, are hampered for commercial productivity due to poor solubility (27, 28). In the food industry, consumers demand products with minimal processing, lacking chemical preservatives, yet still maintain a long shelf life. Food-safe bacteria can produce bacteriocins, typically lactic acid bacteria (LAB), which provide an alternative to undesirable chemicals. These bacteriocins can potentially prevent the growth of both pathogens and spoilage organisms (24, 29, 30). Bacteriocins can be introduced either through the addition of starter cultures within food products, the addition of purified bacteriocins, or using food products from the first method as an ingredient in another (30, 31). Bacteriocins have been approved in numerous countries for use in dairy foods, meats, fermented and non-fermented vegetables, juices, canned vegetables, and others (32).

Bacteriocins may also have potential as alternatives to antibiotics in order to combat the growing problem of antibiotic resistant strains. The potential advantages of bacteriocins includes 1) the current lack of widespread immunity, 2) the potential specificity prevents the widespread destruction of the bacterial communities within the body, and 3) bacteriocins possess a higher potency than antibiotics with little toxicity to human cells (28). The bacteriocin nisin A has been demonstrated to be effective against methicillin resistant *S. aureus*, and importantly, shows effectiveness against biofilms, normally resistant to vancomycin (33). *E. faecalis* resistant to vancomycin has also shown susceptibility to bacteriocins (34). Furthermore, with animals in the food industry, bacteriocins have been able to reduce the incidence of post weaning diarrhea in piglets, therefore, improving growth, in addition to vastly reducing the population of *Campylobacter jejuni* in chickens, and in another instance, increasing the overall growth rate of chickens while reducing the number of *Bacteroides* and *Enterobacteriacae* (32).

Despite having little toxicity to normal eukaryotic cells, some bacteriocins have shown to have anti-neoplastic potential (32). Numerous bacteriocins including microcins, pyocins, pediocins, and colicins have all shown activity against cancer cells derived from humans

and various animals (35). Interestingly, bacteriocins still seem to have a specific cell target range they affect much like their bacterial target. While their activity against bacterial targets is understood in some cases, how they are able to inhibit eukaryotic cells has not been established (8, 22, 35, 36).

3.1.2.2 Bacteriocin Classification

Gram-positive bacteriocins are divided into groups in a system first created by Klaenhammer (37). Over the years this classification system has been changed and debated as the number of bacteriocins has increased, as well as the knowledge of their form and function. Based on the beginning of Klaenhammer's system, the most agreed upon division consists of three groups: (I) Lantibiotics are bacteriocins that derive their name from the addition of unusual amino acids such as lanthionine, β-methyllanthionine, and dehydrated residues. Produced as prepropeptides, they undergo post-translational modification by specific enzymes to modify these residues. A final enzyme cleaves the signal peptide during secretion of the mature peptide. Smaller than 5 kDa, lantibiotics are peptides that are active against the target cell membrane. Various groups have subdivided lantibiotics based on structure, enzymes used for modification, modified residues, and/or conserved sequences, however, there does not appear to be any sort of consensus. The most well-known and researched lantibiotic is nisin, produced by *L. lactis*. (II) Class II bacteriocins are small heat-stable, membrane active, non-lanthionine peptides that are less than 10 kDa. The vast majority are produced in an immature form containing a signal sequence, which is cleaved at the double glycine (Gly-Gly) processing site upon secretion, producing the active form. Class II bacteriocins are predicted to form a β-sheet structure, with amphiphilic helices containing differing amounts of hydrophobicity, and an overall moderate to high heat stability (100-121°C). Class II bacteriocins can also be further subdivided based on structure. Class IIa contain the pediocin-like bacteriocins, such as pediocin PA-1, that include a group of *Listeria*-active peptides that possess an N-terminal consensus sequence of -Tyr-Gly-Asn-Gly-Val-Xaa-Cys-. Class IIb bacteriocins are made up of two peptides for activity (labelled α and β), with lactococcin G (Lcn G) as the first found and best studied. Class IIc bacteriocins are peptides that are made circular by the covalent linking of the N- and C- termini and Class IId bacteriocins consists of

single, linear, non-pediocin peptides including those that are leaderless, *sec*-dependent or otherwise, without an assigned class. Further subdividing has been considered due to the different peptides involved. The Class III bacteriocins consist of large heat-labile proteins, greater that 30 kDa, and are divided into lytic proteins (Class IIIa) and non-lytic proteins (Class IIIb). Examples include lysostaphin and helveticin J, respectively (25, 32, 38-40). Some groups have suggested the reclassification of class III bacteriocins to bacteriolysins since they are deemed lytic enzymes rather than peptides (24, 38). There has also been some debate of the existence of peptides containing non-proteinaceous moieties and therefore the addition of a Class IV. This includes antimicrobial peptides requiring post-translational cysteine S-glycosylation for function, but some groups argue that they are S-linked glycopeptides and not bacteriocins (41-43). Further research is required in order to determine whether or not this could represent another class of bacteriocins. While each class has a unique method of action, immunity, and control, specifically Class IIb will be specifically looked at further since the putative bacteriocin genes identified via RIVET were consistent with a Class IIb bacteriocin.

3.1.2.3 Class IIb mechanism of action

All known Gram-positive bacteriocins function by disrupting the membrane of their target (32, 44). Class IIb bacteriocins are typically effective against a narrow range of targets, many staying within the same species (45-47). This specificity seems to be determined predominately by residues within the N-terminus of the peptides, with the β-peptide seeming especially important (48, 49). The best-known Class IIb bacteriocin is Lcn G of *L. lactis*. Lcn G is made up of two distinct amphiphilic helical peptides (Lcn-α and Lcn-β) that must be present in a ratio of approximately 1:1 for activity (28, 50-53). The presence of GxxxG motifs on both peptides allows the formation of a parallel transmembrane helical-helical structure with the residues 1 to 13 of Lcn-α aligning with the residues 14 to 24 of Lcn-β. This structure is perpendicular to the membrane, creating a pore (32-34, 50, 53-56). The pore allows the flux of potassium, sodium, and other monovalent cations, with the exception of protons. This causes depletion in ATP pools as the cell attempts to maintain osmotic balance, eventually leading to the dissipation of membrane potential and cell death (51, 52). Recently, a probable receptor for Lcn G on

L. lactis target cells was found as undecaprenyl pyrophosphate phosphatase (UppP). UppP is a membrane spanning protein that was expressed in cells sensitive to Lcn G. When UppP was no longer expressed, normally sensitive *L. lactis* cells become resistant. However, the exact involvement of UppP in Lcn G pore formation has yet to be elucidated (57).

3.1.2.4 Class IIb immunity

Self-immunity to Class IIb bacteriocins such as Lcn G is achieved using a single immunity gene (58). These immunity proteins lack a signal sequence and have been shown to be present in the membrane and the cytoplasm (35, 59, 60). Through the use of hybrid experiments it was shown that it is not the toxicity itself that triggers the immunity, but recognition of the bacteriocin. Based on the knowledge of Class IIa bacteriocins, it is speculated that Lcn G binds to its receptor, presumably UppP, creating a conformational change and inducing membrane leakage. The immunity protein is thought to recognize the bacteriocin-receptor complex where it can bind to prevent further leakage (37, 57, 61, 62).

3.1.2.5 Class IIb expression and control

Two-peptide bacteriocins can be found constitutively expressed by some strains while others are transcriptionally regulated by quorum-sensing systems. Quorum sensing systems are composed of an induction peptide (IP), a histidine protein kinase (HPK), and a response regulator (RR), with the last two collectively known as a two-component system (24, 38, 43, 45). The IPs, also known as pheromones, are small, unmodified peptides ranging from 19 to 24 amino acids containing an N-terminal secretion signal that is recognized by the same export machinery as the bacteriocin it controls (63). Pheromones allow the cell to measure its density within the environment and respond accordingly by triggering the expression of genes through the two-component system (64). During initial growth, the pheromone, its required transporter, and two-component system are expressed at a low constitutive level. When the pheromone reaches the threshold level, it triggers activity of the two-component system (autoinduction) through binding to the outer cell membrane bound HPK (45, 64, 65). Inducing an

autophosphorylation event on a conserved histidine residue on the cytoplasmic side of the protein, this phosphate is then transferred to the cytoplasmic RR. This changes the conformation of the RR allowing for a direct interaction with the binding sites of associated promoters, consisting of two repeated sequences separated by a spacer, leading to massive expression of the bacteriocin and associated genes, as well as more pheromone, two-component system, and transporter (45, 65). Transcription occurs for a certain amount of time until, due to an unknown mechanism, the genes are down-regulated causing expression to revert to basal levels (64).

3.1.3 Bacteriocins of *S. pyogenes*

Identified by creating a zone of inhibition of an agar plate, it was first discovered in 1949 that a small number of *S. pyogenes* possessed antimicrobial activity (31, 66). It was not until many decades later this inhibitory zone was identified as the lantibiotic streptococcin A-FF22 (SA-FF22) (41, 67, 68). SA-FF22 showed no activity against Gram-negative strains, but was active against a wide range of Gram-positive strains including *S. agalactiae*, *Streptococcus equi*, *Streptococcus cremoris*, *L. lactis*, *Bacillus cereus*, and *B. subtilis* (69, 70). Production of SA-FF22 requires the chromosomally encoded operon *scnKRAA'MTFEG,* which consists of a two-component system (*scnKR*), the lantibiotic (*scnA*), modification and transport enzymes (*scnMT*), and an immunity complex (*scnFEG*). Presently there is no known function of *scnA'*. Production of the lantibiotic begins as a prepropeptide of 51 amino acids and after modification and cleavage of the leader sequence, export results in a final 26 amino acid peptide. A similar lantibiotic was also isolated from M49 strains of *S. pyogenes* showing only minor differences in amino acid sequences (27, 69, 71).

S. pyogenes also produces another less well-characterized lantibiotic called streptin. Streptin is active against a wide range of Gram-positive species such as *L. lactis*, *S. pneumoniae*, *S. salivarius*, *Streptococcus sanguis*, *Actinomyces* species, and *Clostridium* species, among others. Streptin showed no activity against a wide range of Gram-negative species with *Prevotella intermedia* being the only exception (27, 51, 52, 57, 69, 72, 73). Production of streptin is required by the operon *srtIRKATCBFEG* located on the chromosome. While the process is similar of SA-FF22, streptin utilizes three genes for

modification and secretion (*srtTCB*) and also possesses an additional gene involved in immunity (*srtI*) (7, 73). While there has also been some thought that the streptin operon is also involved in streptolysin S formation, other groups have shown no link, indicating that more research in that area is required (7, 58, 73, 74).

S. pyogenes can also produce a Class III bacteriocin termed streptococcin A-M57 (SA-M57) that is encoded on the 3.3 kb plasmid pDN571. Represented by the gene *scnM57*, the only other gene appearing on the plasmid is *repA* for replication, indicating that all other required genes must be chromosomally encoded. It is believed that SA-M57 is produced as a prepeptide and secreted by a Sec-dependent transport system to yield a mature peptide of 16.9 kDa. An immunity protein has not yet been identified. While it has not been tested with a large number of organisms for activity, SA-M57 has shown activity against *S. epidermidis*, *M. luteus*, *L. lactis*, and *Actinomyces* species, to name a few. There has been no activity against any Gram-negative species to date (9, 75-77).

While bacteriocins are not highly characterized within the *S. pyogenes* species, they do seem to be actively present and could potentially provide an additional advantage when colonizing a niche. In Chapter 2, *spyM18_0544* was identified as being induced in *S. pyogenes* MGAS8232 by an unidentified signal while colonizing the upper respiratory tract of humanized mice. This gene represents a putative bacteriocin, which is likely part of a larger operon including the necessary components. Bacteriocins represent an interesting antimicrobial that is possessed by many bacteria and may help in colonization. We hypothesized that *spyM18_0544* (herein referred to as *blpM*), and the downstream operon, represents a functional bacteriocin system that *S. pyogenes* uses for colonization of the nasopharynx.

3.2 Materials and Methods

3.2.1 Bacterial strains

All strains used are shown in **Table 3.1** and grown as previously described in **2.2.1**.

Table 3.1: Bacterial strains used in this study

Name	Genotype	Reference
E. coli XL1-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F′ *proAB lacIqZ∆M15* Tn*10* (*tet^R*)]	Stratagene, USA
S. pyogenes MGAS8232	isolated 1987, M18 serotype, 1.895 Mb, NC_003485.1	(81)
S. pyogenes ∆*blp*	MGAS8232 with the *blp* operon deleted from the chromosome	This study
S. pyogenes Cas2	MGAS8232 containing the *loxP-tet^R-tk-loxP* cassette within the chromosome	This study

3.2.2 Reagents

The silCR (DIFKLVIDHISMKARKK) (78, 79) and control (FAKDIISKLVIRMHDKK) peptides were synthesized by Biomatik Corporation (Cambridge, Ontario, Canada) and resuspended in Milli'Q water to 10 mg mL^{-1}. The control peptide was created by scrambling the amino acids present in silCR (http://users.umassmed.edu/ian.york/Scramble.shtml). Inhibitors N-[N-(L-3-transcarboxyirane-2-carbonyl)-L-Leucyl]-agmatine (E-64) and 4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride (AEBSF) (Sigma-Aldrich, Oakville, Ontario, Canada) and prepared in Milli'Q water at 1 mM and 1 M, respectively.

3.2.3 Bioinformatics

The genome of *S. pyogenes* MGAS8232 has been sequenced and annotated (NC_003485.1) (80, 81). ORFs were found in DNA sequences using ORF Finder (http://www.ncbi.nlm.nih.gov/projects/gorf/). Both DNA and protein sequences were aligned using ClustalW Multiple Sequence Alignment (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Proteins were compared using BLASTp from NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Promoter elements were found using SoftBerry prediction of bacterial promoters (http://linux1.softberry.com/berry.phtml?topic=bprom&group=programs&subgroup=gfindb).

3.2.4 *in vitro* activation of the *blpM* promoter

MGAS8232 Cas2 containing IVI156 was grown overnight in THY media containing erm and tet as previously described. Cells were subcultured into fresh THY erm along with the silCR peptide (10 μg mL^{-1}) or the control peptide (10 μg mL^{-1}), and with and without the protease inhibitor E-64 (10 μM) or AEBSF (1 mM). After 24 h, 1 mL of culture was taken, serially diluted 10-fold, and plated on THY erm, with and without tet. For qRT-PCR, a sample was taken at 6 h and processed as described in **2.2.15** for RNA extraction and **2.2.16** for qRT-PCR

3.2.5 Creation of the Δ*blp* mutant

All PCR primers and plasmids used are listed in **Table 3.2**. PCR was used to amplify a 500 bp portion of the chromosome upstream of *blpM*, as well as a 500 bp portion of the chromosome downstream of *spyM18_0547*. Products were purified utilizing a QIAquick PCR Purification Kit (Qiagen, Toronto, Ontario, Canada) according to the manufacturer's instructions. The upstream portion, along with pG⁺host5, was digested using *Pst*I and *Sal*I, purified, and ligated using the same method found in **2.2.5**. After confirming the intermediate clone via sequencing, it was then digested with *Sal*I and *Bam*HI, along with the downstream fragment. The fragments were purified, ligated, and transformed as previously described. The final clone, pG⁺host5::Δ*blp* **(Table 3.2)** was confirmed via sequencing and transformed into MGAS8232 to undergo chromosomal recombination as described in **2.2.10**.

3.2.6 *in vitro* and *in vivo* competition experiments

Wild-type MGAS8232 and MGAS8232 Δ*blp* were grown overnight in THY, concentrated, and frozen at -80°C. A small portion was serially diluted 10 fold and plated to calculate CFUs. MGAS8232 and MGAS8232 Δ*blp* were warmed to room temperature for 30 min and mixed 1:1 based on CFU counts. For *in vitro* experiments, a 1% inoculation was grown in 10 mL of THY broth and grown for 24 h, which was chosen to mimic the actual growth period *in vivo*. For *in vivo* experiments, 10^8 CFUs from the 1:1 mixture were used to inoculate mice as described in **2.2.20**. Following harvesting of the nasal passages at 48h, a sample was serially diluted 10 fold and plated to calculate CFUs and the remainder was frozen at -80°C. Fifty colonies from each *in vivo* inoculation were plated on THY, numbered, and 20 colonies were picked based on a random number generator (www.randomizer.org/form.htm). Each colony was grown overnight in THY, total DNA was extracted, and PCR was used to determine if the colony was wild-type or Δ*blp* based on the presence or absence of the *blp* operon using primers listed in **Table 3.2.**

Table 3.2: Primers and plasmids used in this study

Primer Name	Sequence 5' to 3'[a]	Restriction Site
blp up for	GCG<u>CTGCAG</u>TCATCTTAGTGACTTTTACTGTCGG	*Pst*I
blp up rev	GCG<u>GTCGAC</u>AAGAACAGTATCCTGGTTTCAAAA	*Sal*I
blp down for	GCG<u>GTCGAC</u>CCTCACTAAATATATAAATGACTATCTATCACTG	*Sal*I
blp down rev	GCG<u>GGATCC</u>TAGAATCACTGCAGAAATAATAACAATACC	*Bam*HI

Plasmid Name	Notes	Source/Reference
pG⁺host5	erm^R temerature sensitive shuttle vector	(82)
pG⁺host5:: Δ*blp*	pG⁺host5 containing the construct used to delete the bacteriocin and immunity protein from the *S. pyogenes* MGAS8232 genome	This study

[a]-restriction sites underlined

3.3 Results

3.3.1 Bioinformatic analysis of the *blp* operon

Following the sequencing of the *in vivo*-induced promoter within clone IVI156, downstream sequences in MGAS8232 were analyzed. IVI156 contained a nucleotide sequence immediately upstream of the gene *spyM18_0544*. Previously found in *S. pyogenes* JS95, this genes has been labelled as *blpM*[H] (9, 83) **(Figure 3.1)** due to its similarity to *blpM*, a gene encoding the first peptide of a Class IIb bacteriocin system in *S. pneumonia* **(Figure 3.2)**. The *blpM*[H] gene was 228 bp, and encoded a theoretical peptide of 75 amino acids **(Figure 3.2)**. As mentioned earlier, Class II bacteriocins utilize double-glycine type leader sequences for export (27, 84) and based on the location of the double glycine sequence in *blpM*[H], a mature BlpM[H] peptide of 52 amino acids with a leader sequence of 23 amino acids was predicted **(Figure 3.2)**.

As a putative Class IIb bacteriocin, a second peptide would be required for activity (27, 85, 86). Immediately downstream of *blpM*[H] was *spyM18_0545*, also annotated as a bacteriocin peptide. Utilizing the nomenclature derived from de Saiziew *et al.* and Hidalgo-Grass *et al.*, this second gene was designated as *blpN*[H] **(Figure 3.1)** (9, 27). *blpN*[H] had a predicted ORF of 201 bp encoding a theoretical peptide of 66 amino acids with similarity to BlpN of *S. pneumoniae*. Once again, based on the location of the double glycine sequence, BlpN[H] was predicted to consist of a mature peptide of 48 amino acids with a leader sequence of 18 amino acids **(Figure 3.2)**.

Also present on both BlpM[H] and BlpN[H] are four and five GxxxG motifs, respectively, which allow for helix-helix interactions between the two peptides to create a membrane penetrating structure (27, 55). Since the Class IIb bacteriocins require both peptides for activity, this bacteriocin will herein be referred to as BlpMN[H]. The use of BLASTp indicated that BlpMN[H] was widespread in virtually all fully sequenced strains of *S. pyogenes*, with similarities of 97% and above. However, two notable exceptions were found in MGAS5005 and SF370. While BlpN[H] was still 97% homologous or above for both strains, BlpM[H] was truncated in MGAS5005 and had less than 30% similarity in

Figure 3.1: The *blp* operon and *sil* locus of *S. pyogenes* MGAS8232

The sequence contained within clone IVI156 is indicated above the gene map. The *blp* operon consisted of a Class IIb bacteriocin (*blpMN^H)* along with the potential immunity genes *spyM18_0546* and *spyM18_0547*. The *sil* locus is composed of a two-component system (*silAB*), the signaling peptides *silC* and *silCR*, and an ABC transport system (*silDE*). *silD* is truncated due to a missing nucleotide and is split into two genes, the second being *spyM18_0541*

silA
428226

silB

silC
silCR
spyM18_0541

silD

silE
433733
clone IVI156
434717

blpM^H
blpN^H
spyM18_0546
435443
spyM18_0547
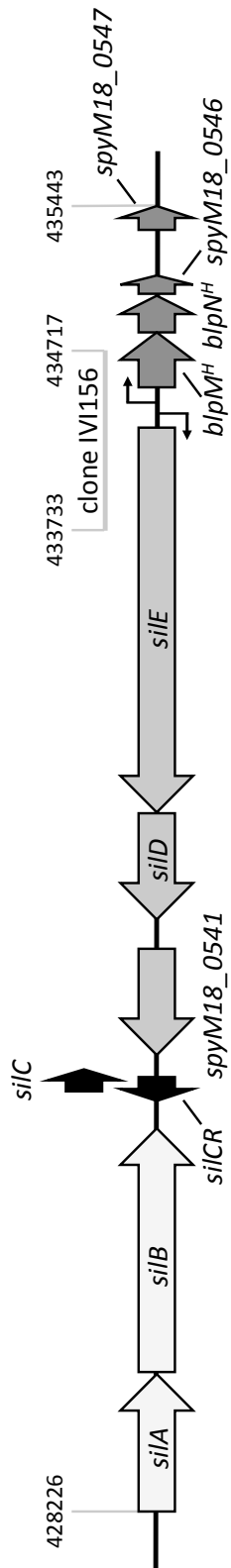
Figure 3.2: BlpMN^H and BlpMN sequences

Leader sequences are separated from mature peptides at the double glycine (indicated by arrow). **G**xxx**G** highlights this motif found in Class IIb bacteriocins.

```
                                    leader                            mature peptide
                                              ↓

S. pyogenes BlpM^H    MEIKKLETFHQMTIEKLAKVEGG  KNNWQANVSGVIAAGSAGAAIGFPVCGV-------ACGYIGAKTAITLWAGVTGATGGF
S. pneumoniae BlpM    MNTKMMEQFHEMDIAMLSSIEGG  KNNWQTNVFEGGSAAFGGWGLGTAICAASGVGAPFMGACGYIGAKFGVALWAGVTGATGGF
                      * * +* **+* * *+ +****   ***** **  +* *  +*  +*  ********* + *************

S. pyogenes BlpN^H    MTTMKELTINDMASISGG       NAPGDAVIGGLGGLASGLKFCKLPHPVLAGGCVVGFTVGGAYLGYTAN
S. pneumoniae BlpN    MNTYCNINETMLSEVYGG       NSGGAAVVAALGCAAGGVKYGKILGPWGAAIGGIGAVVCGYLAYTATS
                      * *  + ++ + *** *+ *     ** * **++ ** * **++ *+  +*  *  ** ***
```

SF370. Finally, further comparison of BlpMN^H indicated that it was also present within several other streptococcal species including *Streptococcus dysgalactiae*, *S. equi*, and *Streptococcus uberis*.

Bacteriocin systems also require an immunity gene, which is typically found downstream from the bacteriocin structural genes. Downstream from *blpN^H* was a potential ORF annotated as *spyM18_0546*, predicted to encode a 33 amino acid peptide **(Figure 3.1)**. Using BLASTp, the peptide was found in some strains of *S. pyogenes* (e.g. *SPYJRS4_0399*) and *S. dysgalactiae* subsp. *equisimilis*, (*SDE12394_02605*), however, it was found to be of 69 amino acids in length. Alignment was only found for the first 25 amino acids of the N-terminus between *spyM18_0546*, *SPYJRS4_0399*, and *SDE12394_02605* **(Figure 3.3)**. Comparison of the nucleotide sequence showed one less adenine nucleotide within the transcription region of *spyM18_0546*, alternating the reading frame, and potentially creating a premature stop codon **(Figure 3.3)**. This peptide could potentially be an immunity protein, although its function remains unknown.

Further downstream was *spyM18_0547*, theoretically encoding a peptide of 42 amino acids that could also potentially function as a bacteriocin immunity protein **(Figure 3.1)**. BLASTp analysis showed *spyM18_0547* to be well conserved within 30 *S. pyogenes* strains, such as HSC5 (*L897_02190*). However, *spyM18_0547* may actually be truncated, as the most common predicted size is a 64 amino acid peptide found in over 130 strains, including common studied strains such as SF370 (*spy_0486*). Comparison of the peptide and nucleotide sequences of *spyM18_0547*, *L897_02190*, and *spy_0486* showed the loss of a single adenine base pair within the coding sequence of *spy_0486*, upstream from the start codon of *spyM18_0547* and *L897_02190* **(Figure 3.4)**. This peptide potentially produces an immunity protein as well, though the correct sequence and the corresponding sequence is once again unknown. The predicted *blp* operon DNA sequence and matching peptides is represented in **Figure 3.5**.

3.3.2 Bioinformatics of the regulation and transport of BlpMN^H

In addition to the bacteriocin structural and immunity genes, Class IIb bacteriocins also require dedicated transporters, and additionally, often encode proteins involved in

Figure 3.3: Comparison of *spyM18_0546* DNA and protein sequences from various strains

spyM18_0546 is shown to have a premature stop codon (starting at base pair 100) due to the loss of a single adenine nucleotide (should be base pair 77) when compared to the same peptide from *S. pyogenes* JS94 (*SPYJRS4_0399*) and *S. dysgalactiae* subsp. *equisimilis* ATCC 12394 (*SDE12394_02605*). Base pair numbers are indicated on the right.

```
spyM18_0546      ATGATCTATACGATTTTTATTATTATGAATGCGTCTATAATGTTTTTAATCATTGTTGAT  60
                  M  I  Y  T  I  F  I  I  M  N  A  S  I  M  F  L  I  I  V  D
SPYJRS4_0399     ATGATCTATACGATTTTTATTATTATGAATGCGTCTATAATGTTTTTAATCATTGTTGAT  60
                  M  I  Y  T  I  F  I  I  M  N  A  S  I  M  F  L  I  I  V  D
SDE12394_02605   ATGATCTATACGATTTTTATTATTATGCATGCGTCTATAATGTTTTTAATGATTGTTGAT  60
                  M  I  Y  T  I  F  I  I  M  H  A  S  I  M  F  L  M  I  V  D


spyM18_0546      GCCATTCATAAAAAAA-TGTTGATAAATCTATACCATATTTAGTGTTTTTAATTTTCAAT 119
                  A  I  H  K  K     M  L  I  N  L  Y  H  I  *  -  -  -  -  -  -
SPYJRS4_0399     GCCATTCATAAAAAAAATGTTGATAAATCTATACCATATTTAGTGTTTTTAATTTTCAAT 120
                  A  I  H  K  K  N  V  D  K  S  I  P  Y  L  V  F  L  I  F  N
SDE12394_02605   GCCATTCATAAAAAAAATGTTGATAAATCTATACCATATTTAGTGTTTTTAATTTTCAAT 120
                  A  I  H  K  K  N  V  D  K  S  I  P  Y  L  V  F  L  I  F  N


spyM18_0546      CAATTATCCCTTTCGGAAAAATCGGTAAACATGTTATCTATCGTAGCTTTGGTAGCGTCA 179
                  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -
SPYJRS4_0399     CAATTATCCCTTTCGGAAAAATCGGTAAACATGTTATCTATCGTAGCTTTGGTAGCGTCA 180
                  Q  L  S  L  S  E  K  S  V  N  M  L  S  I  V  A  L  V  A  S
SDE12394_02605   CAATTATCCCTTTCGGAAAAATCGGTAAACATGTTATCTATCGTAGCTTTGGTAGCGTCA 180
                  Q  L  S  L  S  E  K  S  V  N  M  L  S  I  V  A  L  V  A  S


spyM18_0546      TTATTTGTTTTTATCGATAAAAGAAAATAG 209
                  -  -  -  -  -  -  -  -  -  -  -
SPYJRS4_0399     TTATTTGTTTTTATCGATAAAAGAAAATAG 210
                  L  F  V  F  I  D  K  R  K  *
SDE12394_02605   TTATTTGTTTTTATCGATAAAAGAAAATAG 210
                  L  F  V  F  I  D  K  R  K  *
```

Figure 3.4: Comparison of *spyM18_0547* DNA and protein sequences from various *S. pyogenes* strains

spyM18_0547 and *L897_02190* are annotated with a start codon further downstream in comparison to others such as *spy_0846*. In addition, *spy_0846* is missing a single adenine nucleotide (base pair 18) within its coding region. Base pair numbers are indicated on the right.

```
spyM18_0547    ATGGAAGTGCTAAAAAAATGGTCTGACTATCATAAATGGAGGCGGAAAAGACGGAGCAAA  60
                - - - - - - - - - - - - - - - - - - - - -
L897_02190     ATGGAAGTGCTAAAAAAATGGTCTGGCTATCATAAATGGAGGCGGAAAAGACGGAGCAAA  60
                - - - - - - - - - - - - - - - - - - - - -
spy_0486       ATGGAAGTGCTAAAAAA-TGGTCTGGCTATCATAAATGGAGGCGGAAAAGACGGAGCAAA  59
                M  E  V  L  K  N    G  L  A  I  I  N  G  G  G  K  D  G  A  N


spyM18_0547    TATCTTTTTGACAGGGATGGCAGGTGCTGCGCAGGGAGTGACAGTTTGTGCTCAGACAGG 120
                 -  -  M  T  G  M  A  G  A  A  Q  G  V  T  V  C  A  Q  T  G
L897_02190     TATCTTTTTGACAGGGATGGCAGGAGCTGCGCAGGGAGTGACAGTTTGTGCTCAGACAGG 120
                 -  -  M  T  G  M  A  G  A  A  Q  G  V  T  V  C  A  Q  T  G
spy_0486       TATCTTTTTGACAGGGATGGCAGGTGCTGCGCAGGGAGTGACAGTTTGTGCTCAGACAGG 119
                 I  F  L  T  G  M  A  G  A  A  Q  G  V  T  V  C  A  Q  T  G


spyM18_0547    CGTATTTATCCCTTGGCAAGGTTATATTTTGTGTGGAGCAGCTGGTGCTGCTACAAATAT 180
                 V  F  I  P  W  Q  G  Y  I  L  C  G  A  A  G  A  A  T  N  I
L897_02190     CGTATTTATCCCTTGGCAAGGTTATATTTTGTGTGGAGCAGCTGGTGCTGCTACAAATAT 180
                 V  F  I  P  W  Q  G  Y  I  L  C  G  A  A  G  A  A  T  N  I
spy_0486       CGTATTTATCCCTTGGCAAGGTTATATTTTGTGTGGAGCAGCTGGTGCTGCTACAAATAT 179
                 V  F  I  P  W  Q  G  Y  I  L  C  G  A  A  G  A  A  T  N  I


spyM18_0547    TATCTGGCCTCACTAA 196
                 I  W  P  H  *
L897_02190     TATCTGGCCTCACTAA 196
                 I  W  P  H  *
spy_0486       TATCTGGCCTCACTAA 195
                 I  W  P  H  *
```

Figure 3.5: The *blp* operon sequence in *S. pyogenes* MGAS8232

The DNA and predicted amino acid sequences encoded within the *blp* operon of MGAS8232. The nucleotide and translation products for the *blp* locus are given below for the indicated region with corresponding nucleotide numbers as annotated in the *S. pyogenes* MGAS8232 genome sequence. Directions of the ORFs are indicated by arrows.

```
                      ←silE
AACAAATGTTTTTCGATAGGATATCATTTGTAAAAGACCTTTCTATTAAATACTAAGCTA          434260
 V  F  T  K  R  Y  S  I  M

TTATGACATATATATAGGGAGTTTTTATCAAAAATCTTAAAAGGTATTTATAAAAGCATG          434320

AATGGTATCATTTAAACTTCTTAATTAATTAACTATTGTCACAGCGTAAGTGTAGAGAGG          434380

ATAAAAAGACATTTCATGATGAAAAATAGCCTTTTAAGATTTTGAAACCAGGATACTGTT          434440

CTTTATGATATTCTGTACTAGCTCGAGGTTGCAACGAGACAAATTTAAGAATTATGAAAA          434500

                spyM18_0544(blpMᴴ)→
GAAGTTATTCGATATGGAGATAAAGAAACTGGAAACATTTCACCAAATGACTATAGAAAA          434560
            M  E  I  K  K  L  E  T  F  H  Q  M  T  I  E  K

GCTTGCCAAGGTTGAGGGTGGTAAGAATAATTGGCAGGCAAATGTCAGTGGGGTTATCGC          434620
 L  A  K  V  E  G  G  K  N  N  W  Q  A  N  V  S  G  V  I  A

TGCAGGTAGTGCTGGAGCTGCTATAGGTTTTCCAGTTTGCGGTGTAGCCTGTGGTTATAT          434680
 A  G  S  A  G  A  A  I  G  F  P  V  C  G  V  A  C  G  Y  I

TGGAGCTAAAACAGCTATTACTCTTTGGGCTGGTGTAACAGGAGCTACAGGCGGCTTCTA          434740
 G  A  K  T  A  I  T  L  W  A  G  V  T  G  A  T  G  G  F  *
                spyM18_0545(blpNᴴ)→
GAGGAAAGAAAGGATGACAACAATGAAGGAATTAACAATTAATGACATGGCTAGTATAAG          434800
            M  T  T  M  K  E  L  T  I  N  D  M  A  S  I  S

TGGAGGAAACGCTCCAGGAGATGCTGTAATTGGTGGTCTTGGAGGCTTAGCTTCTGGATT          434860
 G  G  N  A  P  G  D  A  V  I  G  G  L  G  G  L  A  S  G  L

GAAATTTTGTAAACTGCCACATCCAGTATTGGCAGGTGGATGTGTTGTAGGATTTACAGT          434920
 K  F  C  K  L  P  H  P  V  L  A  G  G  C  V  V  G  F  T  V
                            spyM18_0546→
TGGTGGAGCATATTTGGGATATACAGCAAATTAAATGATCTATACGATTTTTATTATTAT          434980
 G  G  A  Y  L  G  Y  T  A  N  *  M  I  Y  T  I  F  I  I  M

GAATGCGTCTATAATGTTTTTAATCATTGTTGATGCCATTCATAAAAAAATGTTGATAAA          435040
 N  A  S  I  M  F  L  I  I  V  D  A  I  H  K  K  M  L  I  N

TCTATACCATATTTAGTGTTTTTAATTTTCAATCAATTATCCCTTTCGGAAAAATCGGTA          435100
 L  Y  H  I  *

AACATGTTATCTATCGTAGCTTTGGTAGCGTCATTATTTGTTTTTATCGATAAAAGAAAA          435160

TAGTGATTATGCCAATGAATTTAGATATCACTCTTAGCTTTTTAAAATACAAAAGAGTGA          435220

AATTAAAATGAAAAAGACTTCAAAAATATGGAAGTGCTAAAAAAATGGTCTGACTATCAT          435280

                        spyM18_0547→
AAATGGAGGCGGAAAAGACGGAGCAAATATCTTTTTGACAGGGATGGCAGGTGCTGCGCA          435340
                        L  T  G  M  A  G  A  A  Q

GGGAGTGACAGTTTGTGCTCAGACAGGCGTATTTATCCCTTGGCAAGGTTATATTTTGTG          435400
 G  V  T  V  C  A  Q  T  G  V  F  I  P  W  Q  G  Y  I  L  C

TGGAGCAGCTGGTGCTGCTACAAATATTATCTGGCCTCACTAAATATATAAATGACTATC          435460
 G  A  A  G  A  A  T  N  I  I  W  P  H  *

TATCACTGATATATAAGTAGTAGAGGAGTGGTATGAAAATTTTAGATATTATTTTTAACC          435520
```

regulation of the system. Upstream of *blpM*[H] in MGAS8232 was a group of genes that correspond to the streptococcal invasion locus (*sil*) (9). The *sil* locus is comprised of an ABC transporter (*silDE*), a pheromone (*silCR*), a peptide (*silC*), and a two-component system (*silAB*) **(Figure 3.1)**. While its exact purpose is unknown, the *sil* locus has been shown to be important for invasive infections (9, 63, 64, 76) and can regulate BlpMN[H] and its associated immunity protein (7).

MGAS8232 encodes a full-length *silE* predicted to produce a 717 amino acid ATP-binding cassette transporter according to BLASTp results. However, *silD* and *spyM18_0541* appear to be truncated portions of a complete 454 amino acid SilD, which works in conjunction with SilE to make a complete ABC-binding cassette transporter. Analysis of the nucleotide sequence revealed that a missing adenine caused early termination of MGAS8232 *silD* and *spyM18_0541* appears to encode the C-terminus of *silD* (7, 9, 11) **(Figure 3.6)**.

In MGAS8232, further upstream of the putative bacteriocin transporters were *silC* and *silCR*, overlapping each other by 91 bp (7, 9), creating peptides of 39 and 41 amino acids, respectively **(Figure 3.1)**. SilCR contains pheromone characteristics such as a double-glycine type secretion signal, creating a mature peptide of 17 amino acids, and an RKK motif at the C-terminus (9, 11). On the other hand, the peptide SilC appears to lack any secretion signal or leader sequence indicating it may function in the cytoplasm. BLASTp analysis of SilCR indicated 100% identity with other *S. pyogenes* strains encoding the *sil* locus, except JS95, which contains a single mutation changing the start codon from ATG to ATA (9, 78). While *silC* was not annotated on the MGAS8232 genome, the correct DNA sequence was present and 100% identitcal to *silC* in *S. pyogenes* strains IB7 and JS95, and *S. dysgalactiae* subsp. *equisimilis* strains GGS1800 and N3.

Finally, the two-component histidine kinase and response regulator, encoded by *silB* and *silA*, respectively, were encoded adjacent to the *silCR*/*silC* genes **(Figure 3.1)**. Once again, BLASTp analysis indicated nearly 100% identity between MGAS8232 and the *S. pyogenes* strains that encode the *sil* locus. One exception was *silB* of JS95 that contains a short range of 11 different amino acids, seemingly unique to JS95.

Figure 3.6: The DNA and protein sequence of *silD* and *spyM18_0541*

Comparison of *silD* and *spyM18_0541* of MGAS8232 with *silD* of JS95. The loss of an adenine nucleotide (base pair 479 in MGAS8232) creates a premature stop codon in MGAS8232 *silD* (starting at base pair 580). ORF *spyM18_0541* is the C-terminus of *silD* starting at base pair 747 and finishing at the proper stop codon of *silD*. Large stretches of identical sequence were removed (indicated by ……). Base pair number is indicated on the right.

```
8232 silD      ATGAATCCAAACCTTTTTAAAAGTGCGGAATTTTATCATCGACGTCATCATAACTTTGCA……  60
                 M  N  P  N  L  F  K  S  A  E  F  Y  H  R  R  H  H  N  F  A ……
JS95 silD      ATGAATCCAAACCTTTTTAAAAGTGCGGAATTTTATCATCGACGTCATCATAACTTTGCA……  60
                 M  N  P  N  L  F  K  S  A  E  F  Y  H  R  R  H  H  N  F  A ……


8232 silD      ……ACGTTTAATCATTTTCTCAAACAAACCCAAGATATAATCATTGGGTTTGCTAAAAAAA-T  479
               ……T  F  N  H  F  L  K  Q  T  Q  D  I  I  I  G  F  A  K  K  M
JS95 silD      ……ACGTTTAATCATTTTCTCAAACAAACCCAAGATATAATCATTGGGTTTGCTAAAAAAAAT  480
               ……T  F  N  H  F  L  K  Q  T  Q  D  I  I  I  G  F  A  K  K  N


8232 silD      GCTGAAGTCAATAATCAAGCAAGTCTTGCTAATCATGCCATTTCAGTTATTACTAGTCAG  539
                 L  K  S  I  I  K  Q  V  L  L  I  M  P  F  Q  L  L  L  V  S
JS95 silD      GCTGAAGTCAATAATCAAGCAAGTCTTGCTAATCATGCCATTTCAGTTATTACTAGTCAG  540
                 A  E  V  N  N  Q  A  S  L  A  N  H  A  I  S  V  I  T  S  Q


8232 silD      CTTGACGAACTTCATCAAAAAATTACAGATTATGAAGAGCTAAAGCGAGCTATTAATAAT……  599
                 L  T  N  F  I  K  K  L  Q  I  M  K  S  *  -  -  -  -  -  - ……
JS95 silD      CTTGACGAACTTCATCAAAAAATTACAGATTATGAAGAGCTAAAGCGAGCTATTAATAAT……  600
                 L  D  E  L  H  Q  K  I  T  D  Y  E  E  L  K  R  A  I  N  N ……


spyM18_0541 ……CAAAATATCTCAGAACTTGAATCTTCAATGGCAAACCTTAGTATCCAACGTGCTAGTACA  779
            ……  -  -  -  -  -  -  -  -  -  -  M  A  N  L  S  I  Q  R  A  S  T
JS95 silD   ……CAAAATATCTCAGAACTTGAATCTTCAATGGCAAACCTTAGTATCCAACGTGCTAGTACA  780
            ……  Q  N  I  S  E  L  E  S  S  M  A  N  L  S  I  Q  R  A  S  T


spyM18_0541    GGAAATTTATCTCTTCCAGACACTAGTCATAGGATCAAAATAGATATTTTAAAAACACAA……  839
                 G  N  L  S  L  P  D  T  S  H  R  I  K  I  D  I  L  K  T  Q ……
JS95 silD      GGAAATTTATCTCTTCCAGACACTAGTCATAGGATCAAAATAGATATTTTAAAAACACAA……  840
                 G  N  L  S  L  P  D  T  S  H  R  I  K  I  D  I  L  K  T  Q ……


spyM18_0541 ……TATTTCAATTACTATAAAGATAAGATACTCAACAGTTTTAATTAA 1364
            ……  Y  F  N  Y  Y  K  D  K  I  L  N  S  F  N  *
JS95 silD   ……TATTTCAATTACTATAAAGATAAGATACTCAACAGTTTTAATTAA 1365
            ……  Y  F  N  Y  Y  K  D  K  I  L  N  S  F  N  *
```

Regulation of the *sil* locus is thought to occur through the expression of *silC* and *silCR*. Expression of *silC* causes suppression of two operons to very low levels: *silE/D/CR* and *blpM^H/N^H/spy_0586* (7, 80). However, once SilCR reaches its threshold level, it activates SilA via SilB, to initiate expression of the *blpM^H/N^H/spy_0586* and *silE/D/CR* operons, causing suppression of *silC*. Once SilCR levels decline back to normal levels, suppression by SilC will once again occur (7, 83). Interestingly, *silAB* seems to show little to no response to SilCR, but has demonstrated activity when exposed to decreased temperature or H_2O_2 (7, 27, 76). *silC* is also under the control of its own promoter, though how either *silAB* or *silC* is regulated is still unknown.

It was then proposed that the promoter for *blpM^H*, identified *in vivo* by the RIVET system, was under control of SilAB from the *sil* locus. Examination of the DNA region between *silE* and *blpM^H* showed the -35 and -10 regions of both promoters, predicted using online software **(Figure 3.7)**. Also present were the direct repeats for both *blpM^H* (DR2) and *silE* (DR1), the DNA region to which the response regulator SilA binds in order to regulate expression, originally found by Belotserkovsky *et al.* in strain *S. pyogenes* JS95 (27). DR1 and DR2 contain a repeat consensus sequence of ACCTTT[T/C]A[T/A]G **(Figure 3.7)**. When comparing direct repeat sequences between strains, it was found that the spacer for DR2 in MGAS8232 was only 10 bp, while this spacer was 11 bp in JS95 **(Figure 3.8)**. Comparing the repeat sequences of numerous *S. pyogenes* strains showed that both spacer sizes are present for DR2; with the 10 bp spacer present in strains MGAS8232, SF370, MGAS5005, and Manfredo, with the 11 bp spacer present in strains JS95, HSC5, and MGAS315 **(Figure 3.8)**. The spacer for DR1 appears to be 11 bp for all strains presented.

3.3.3 *in vitro* activation of *blpM*

Once the components of the *sil* locus and *blp* operon were identified, the ability of SilCR to induce the system was first tested. Cas2 transformed with the IVI156 plasmid was grown in THY erm containing either the synthesized SilCR peptide, or a scrambled peptide control. Cells were plated at various time points and activation of the *blp* operon promoter was monitored though the loss of tet resistance. Despite testing at various

Figure 3.7: Promoter regions of *blpM^H* and *silE* from *S. pyogenes* MGAS8232

The underlined and italicized sequences show the predicted -35 and -10 regions of *blpM^H* and *silE*. A boxed letter indicates transcription start site with an arrow pointing in the direction of transcription. Direct repeat pairs for the *silE/D/CR* (DR1) and *blp* (DR2) operons are bolded and labeled as left (L) and right (R) on either side of the corresponding spacer region according to the direction of transcription. Consensus sequence of the repeat sequence is shown below with DR1 sequences shown in reverse complement to match DR2. Final consensus sequence: ACCTTT[T/C]A[T/A]G.

```
                                                              ↰
          ←silE
TTTTTCGATAGGATATCATTTGTAAAAGACCTTTCTATTAAATACTAAGCTATTATGACA
                                                          ‾‾‾‾‾‾‾‾
                                                            −10
                          DR1R          spacer          DR1L
TATATATAGGGAGTTTTTATCAAAAATCTTAAAAGGTATTTATAAAAGCATGAATGGTAT
                    ‾‾‾‾‾‾
                     −35

CATTTAAACTTCTTAATTAATTAACTATTGTCACAGCGTAAGTGTAGAGAGGATAAAAA

      DR2L          spacer          DR2R
GACATTTCATGATGAAAAATAGCCTTTTAAGATTTTGAAACCAGGATACTGTTCTTTAT
                                        ‾‾‾‾‾‾               ‾‾‾‾‾‾
                                         −35                  −10
     ↱
GATATTCTGTACTAGCTCGAGGTTGCAACGAGACAAATTTAAGAATTATGAAAAGAAGT
‾‾‾

        blpMᴴ→
TATTCGATATGGAGATAAAGAAACTGGAAACATTTCACCAAATGACTATAGAAAAGCTT


                    Consensus
              DR1L  ACCATTCATG
              DR1R  ACCTTTTAAG
              DR2L  ACATTTCATG
              DR2R  GCCTTTTAAG
                5'−ACCTTTYAWG−3'
```

Figure 3.8: Spacer comparison between *S. pyogenes* strains

Direct repeat pairs for the *silE/D/CR* (DR1) and *blp* (DR2) operons are bolded and labeled as left (L) and right (R) on either side of the corresponding spacer region according to the direction of transcription.

```
                     DR1R         spacer       DR1L
MGAS8232   ATCAAAAATCTTAAAAAGGTATTTATAAAAGCCATGAATGGTATCATTTAAACTTCTT-ATTAATTAACTATTGTCACAG
JS95       ATCAAAAATCTTAAAAAGGTATTTATAAAAGCCATGAATGGTATCATTTAAACTTCTTAATTAATTAACTAATTGTCACAG
M1         ATCAAAAATCTTAAAAAGGTATTTATAAAAGCCATGAATGGTATCATTTAAACTTCTTAATTAATTAACTAATTGTCACAG
MGAS5005   ATCAAAAATCTTAAAAAGGTATTTATAAAAGCCATGAATGGTATCATTTAAACTTCTTAATTAATTAACTAATTGTCACAG
HSC5       ATCAAAAATCTTAAAAAGGTATTTATAAAAGCCATGAATGGTATCATTTAAACTTCTTAATTAATTAACTAATTGTCACAG
MGAS315    ATCAAAAATCTTAAAAAGGTATTTATAAAAGCCATGAATGGTATTATTTAAACTTCTTA---ATTAACTATTGTCACAG
Manfredo   ATCAAAAATCTTAAAAAGGTATTTATAAAAGCCATGAATGGTATCATTTAAACTTCTTAATTAACTATTATCACAG
           ****************************************************** ***********.*******
```

```
                     DR2L         spacer       DR2R
MGAS8232   CGTAAGTGTAGAGAGGATAAAAAGACATTTCATGATGAAAAA-TAGCCTTTTAAGATTTTGAAACCAGGATACTGTTC
JS95       CGTAAGTGTAGAGAGGATAAAAAGACATTTCATGATGAAAAA-TAGCCTTTTAAGATTTTGAAGCCATGATGATACTGTTC
M1         CGTAAGTGTAGAGAGGATAAAAAGACATTTCATGATGAAAAA-TAGCCTTTTAAGATTTTGAAACCATGATACTGTTC
MGAS5005   CGTAAGTGTAGAGAGGATAAAAAGACATTTCATGATGAAAAA-TAGCCTTTTAAGATTTTGAAACCATGATACTGTTC
HSC5       CGTAAGTGTAGAGAGGATAAAAAGACATTTCATGATGAAAAACGGCCTTTTAAGATTTTGAAGCCATGATGATACTGTTC
MGAS315    TGTAAGTGTAGAGAGGATAAAAAGACATTTCATGATGAAAAACAGCCTTTTAAGATTTTGAAACCATGATACTGTTC
Manfredo   CGTAAGTGTAGAGAGGATAAAAAGACATTTCATGATGAAAAA-CAGCCTTTTAAGATTTTGAAACCATGATACTGTTC
           ******************************************  *** **********.*.***********
```

concentrations, SilCR failed to produce any detectable activation after 24 h (unpublished data). Reasoning that proteases could potentially be destroying the peptides *in vitro* the experiment was subsequently attempted with the addition of the cysteine protease inhibitor E-64, yet after 24 h all colonies remained tet resistant. Next, a serine protease inhibitor, AEBSF, was attempted after it was found that blocking the serine protease HtrA in *S. pneumoniae* caused an increase in *blpMN* (29, 86). However, treatment of cells with AEBSF had no effect (unpublished data). Numerous other methods rather than the synthetic peptides were attempted including HLA-DR4/DQ8 mouse cNT homogenate, growth with human peripheral blood mononuclear cells, human neutrophils, and human serum, in an attempt to create the required environment, yet all failed to activate *blpM^H in vitro* (unpublished data). Thus, despite all the different methods attempted, the *blp* operon could not be activated *in vitro*.

3.3.4 MGAS8232 wild-type and Δ*blp* competitive *in vivo* analysis

In order to test the impact of BlpMN^H *in vivo*, a knockout was created to delete the entire *blp* operon, designated MGAS8232 Δ*blp*. Used to inoculate HLA-DR4/DQ8 mice as before, Δ*blp* was mixed 1:1 with MGAS8232 wild-type in a competition model. The recovered bacteria were plated and 20 colonies were randomly picked and identified as Δ*blp* or MGAS8232 via PCR. As a control, the experiment was also repeated with mixtures grown *in vitro* for 24 h to mimic the cell densities that are reached *in vivo*. Growth *in vitro* appeared to show a slight shift to the wild-type, but still showing a wild-type:Δ*blp* ratio percent averaging 56.7:43.3 **(Figure 3.9)**, demonstrating that the *blp* operon is likely not important for *in vitro* growth. Surprisingly, when wild-type:Δ*blp* was grown *in vivo*, the four of the seven mice returned only or predominantly wild-type, while the remaining returned ratios of wild-type:Δ*blp* in roughly equal numbers **(Figure 3.9)**

While not the expected result, it seemed to show that lack of the *blp* operon provided a disadvantage *in vivo* against wild-type and that activation of the operon may not always occur or else cells *in vivo* may exist in different locations.

Figure 3.9: Competition between wild-type *S. pyogenes* MGAS8232 and MGAS8232 Δ*blp*

Wild-type and Δ*blp* were mixed 1:1 and either grown overnight *in vitro* (N=7) or using the nasopharyngeal infection model in HLA-DR4/DQ8 mice (N=7). Cells were plated on THY agar and gDNA was isolated from random colonies and subjected to PCR analysis to identify cells as either wild-type or Δ*blp*.
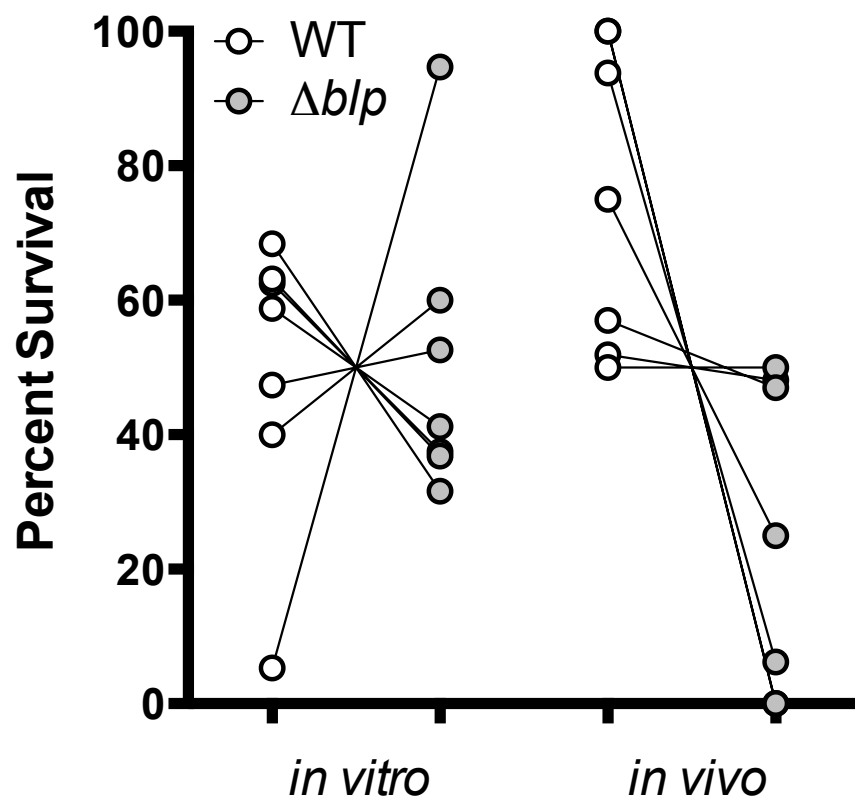
3.4 Discussion

S. pyogenes is an important human pathogen that utilizes the upper respiratory tract and the skin as the most common reservoirs of its host. Research on this organism generally focuses on the virulence factors *S. pyogenes* possess in order to overcome the physical and immunological defenses of the human body. However, another important part of the human body is the resident bacteria. These bacteria create a competitive environment that potentially makes it difficult for invading bacteria to survive (31). Therefore, a tool such as a bacteriocin could be advantageous for pathogens in order to colonize. While the bacteriocins used by pathogens seems to receive little attention, *E. coli*, *S. pneumoniae*, *Haemophilus influenzae*, and *Streptococcus mutans* have all demonstrated an improved ability to colonize when compared to bacteriocin-negative controls (64, 87). *S. pyogenes* has only been shown to express bacteriocins in a handful of strains and they have received little notice. Using an upper respiratory tract model, this research is the first of our knowledge to demonstrate the *in vivo* activation of a Class IIb bacteriocin in *S. pyogenes* and the colonization advantage it may provide.

Based on our results, it appeared that MGAS8232 possessed the bacteriocin BlpMN^{H}, followed immediately downstream of *blpN^{H}* by the two ORFs *spyM18_0546* and *spyM18_0547*, one likely being the required immunity protein **(Figure 3.1)**. Based on sequence analysis, BlpMN^{H} possessed the requirements of a Class IIb bacteriocin including GxxxG motifs and a double glycine for separation of the leader sequence from the mature peptide **(Figure 3.2)**. More recent work in the McCormick laboratory has been able to demonstrate a functional *ex vivo* BlpMN^{H} bacteriocin, showing *L. lactis* with sensitivity to the combined recombinant BlpM^{H} and BlpN^{H} peptides. Furthermore, *L. lactis* was resistant to the antimicrobial activity of BlpMN^{H} when expressing a fragment that included both *spyM18_0546* and *spyM18_0547* **(**McCormick laboratory, unpublished data**)**. Which ORF encodes the immunity gene has yet to be determined. Furthermore, this also provides data towards identifying the targets of BlpMN^{H}. Class IIb bacteriocins are typically effective against a narrow range of targets, many staying within the same species (45-47, 66). Susceptibility of *L. lactis* may suggest a wider range of targets than expected. Determination of further BlpMN^{H} targets could be determined simply by

testing strains for susceptibility. Likely candidates would be other streptococcal strains and in particular, those found within the oral cavity such as *S. salivarius*, *S. mitis*, *S. oralis*, *S. gordonii*, *S. mutans*, or *S. pneumoniae* (41).

Upstream of *blpM*[H] are the genes that make up the *sil* locus **(Figure 3.1)**. While *blpMN*[H] seems to be widely found among virtually all *S. pyogenes* strains, *sil* only seems to be encoded in ~12-25% of the isolates tested (70, 88). Additionally, the locus does not appear to be functional in all isolates. Sequencing *sil*[+] strains for deletions, frameshifts, and mutations in the *silCR*, *silC*, and *silD*, Plainvert *et al*. found that only 56% of strains appeared to have a functional *sil* locus based on sequence (69). Furthermore, there seemed to be no link between the presence of *sil* and invasive status (27, 69). Plainvert *et al*. also found functional *sil* on both invasive and non-invasive strains at 30% and 28%, respectively (27, 69). Other groups, including Jing *et al*. and Billal *et al*. also found *sil*[+] strains in both invasive and non-invasive strains (89, 90). In fact, the presence of *sil* seemed to be more related to *emm* type, with common types being *emm4*, *emm87*, *emm90*, *emm94*, and *emm118* (69, 88). The *emm* type also gave a general indication as to whether or not *sil* was functional within the strain. For example, *emm87* appeared to be functional based on sequence nearly 100% of the time, while *emm14* (JS95) nearly always had ATA as a *silCR* start codon, and *emm4* and *emm18* are known to have a truncated *silD* (69, 79, 88).

Control of *sil* has demonstrated to be controlled through a regulatory circuit consisting of *silC* and *silCR* (7). When off, SilC represses the *blp* operon as well as an operon consisting of the transporters (*silDE*) and *silCR*. However, threshold level of SilCR allows these operons to be turned on, with the expression of *silCR* interfering with *silC* due to the large portion of overlapping sequence (7). Indeed, the effect has been identified several times with invasive *in vivo* models. Starting with JS95 and HSC5, neither expressing *silCR* naturally, loss of *silC* demonstrated the loss of virulence potential (9, 76) and the addition of exogenous SilCR showed a reduction in virulence and lesion size (78). While *in vitro*, Δ*silC* showed a reduction in biofilm formation (80) and the addition of SilCR demonstrated a reduced transcription of *spyCEP* (83). In addition, using microarray, it was found that 46 genes were altered with the addition of

SilCR (27). Nearly half (18 genes) were involved with metabolism, seven had a function involved with gene regulation, and interestingly, *sagA* (SLS) transcription was also increased. Furthermore, *sagA* transcription was also increased by the addition of SilCR in a *sil⁻* strain, demonstrating that SilCR activates more than just the *sil* locus and *blp* operon (27, 86) Exactly how SilC is able to repress its targets, and any other genes it may control, has yet to be elucidated.

Upstream of *silC* is the AlgR/AgrA/LytR family transcription regulator *silAB* that is thought to activate the bacteriocin genes as well as the associated transporter genes. The response regulator, SilA, binds via the LytR-type domain as a dimer to the DNA site made up of two nine bp direct repeats, [T/A][A/C][C/A]GTTN[A/G][T/G], separated by a spacer that seems to vary in length (7, 27, 91). Based on the sequence found by Belotserkovsky *et al*. in strain JS95, which included an additional tenth nucleotide, the *S. pyogenes* MGAS8232 direct repeat followed the consensus almost exactly at ACCTTT[C/T]A[T/A]G **(Figure 3.7)** (27). Belotserkovsky *et al*. had concluded that the *S. pyogenes* pattern was made up of ten nucleotides since modification of the tenth nucleotide (guanine) showed a detrimental effect to the promoter (27).

Also shown to be important for proper activation is the AT-rich spacer length between repeats, typically ranging from 10-13 bp (92, 93). The length of this spacer has shown to be important, as changes such as losing a single nucleotide has shown to have detrimental effects to expression (27, 55, 94-96). Belotserkovsky *et al.* showed *S. pyogenes* JS95 to have a spacer of 11 bp and that expression was lost when modified to a length of 10 or 12 bp (27). As shown in **Figure 3.8**, the spacer region of DR2 is 10 or 11 bp depending on the strain, with MGAS8232 having a 10 bp spacer. Based on the evidence from Belotserkovsky *et al*., this spacer would prevent SilA from binding to DR2, and therefore prevent transcription of the *blp* operon.

Assembling the *sil* locus system within MGAS8232 leads to two major shortfalls. The truncated *silD* implies that neither SilCR nor BlpMN^H could be secreted. However, the initial identification of P*blpM*, and subsequent repetition experiments *in vivo*, indicated that the promoter portion identified in pIVI156 was activated *in vivo* **(Figure 2.10)**. If BlpMN^H activation did occur by SilA binding, SilCR secretion would have been required,

indicating that one or both of the *silD* truncations **(Figure 3.1)** could form with SilE to create a working transporter, or alternatively, a different transporter system was used. However, the failure of *blpMN^H* activation *in vitro* through the addition of exogenous SilCR would suggest that activation did not occur through the *sil* locus. In addition to SilCR, other potential environmental stimuli also failed to activate the *blp* operon including HLA-DR4/DQ8 mouse cNT homogenate, human peripheral blood mononuclear cells, human neutrophils, and human serum (unpublished data).

While *blpMN^H* could not be activated *in vitro*, we tested its impact using the *in vivo* model known to activate it. A surprising almost 'all or none' type of response was recovered after inoculation by a 1:1 mix with roughly half going either way. In one half, the ratio recovered between wild-type and Δ*blp* was nearly even, suggesting that *blpMN^H* was not activated, and therefore, not important for colonization **(Figure 3.9)**. The other half seemed to demonstrate an activation of the *blp* operon since there was little to no Δ*blp* strain recovered **(Figure 3.9)**. It is thought that the MGAS8232 wild-type present was exposed to the correct environmental stimuli to cause expression of the *blp* operon, eliminating Δ*blp* due to the lack of an immunity protein. Why stimulation occurred only in half of the mice is unknown and continued work will attempt to identify the environmental condition activating the *blp* operon. However, this *in vivo* work does suggest that while *blpMN^H* is unlikely to be regulated by the *sil* locus due to SNPs, the lack of Δ*blp* in some instances would suggest that *blpMN^H* is able to regulate and transport through another system.

Colonizing the pharynx, *S. pyogenes* must make a 'space' for itself in order to obtain nutrients and grow against a plethora of other bacteria. Specifically, on the human body as a host, free nutrients can be limited, causing many different bacteria to attempt to colonize the same niche. Therefore, bacteriocins provide a powerful tool that could be used by *S. pyogenes*, as well as other pathogens, to aid in the colonization process. This research represents the first attempt to determine the use of the Class IIb bacteriocin BlpMN^H by *S. pyogenes* when colonizing a nasopharyngeal model. While it seemed that *blpMN^H* was not under control of *sil*, it could explain why *blpMN^H* remains within nearly all strains while functional *sil* seems to be in such a small percentage. Although the

mechanism that causes the activation of $blpMN^H$ in MGAS8232 still needs to be established, it appears to be functional, providing a potential mechanism to help outcompete bacteria also striving for the pharynx of the human host.

3.5 References

1. **Wintermute EH**, **Silver PA**. 2010. Dynamics in the mixed microbial concourse. Genes Dev **24**:2603–2614.
2. **Wilson M**. 2005. Microbial inhabitants of humans. Cambridge University Press.
3. **Lemon KP**, **Klepac-Ceraj V**, **Schiffer HK**, **Brodie EL**, **Lynch SV**, **Kolter R**. 2010. Comparative analyses of the bacterial microbiota of the human nostril and oropharynx. MBio **1**:1–9.
4. **Dethlefsen L**, **McFall-Ngai M**, **Relman DA**. 2007. An ecological and evolutionary perspective on human-microbe mutualism and disease. Nature **449**:811–818.
5. **Costello EK**, **Stagaman K**, **Dethlefsen L**, **Bohannan BJM**, **Relman DA**. 2012. The application of ecological theory toward an understanding of the human microbiome. Science **336**:1255–1262.
6. **McFarland LV**. 2000. Normal flora: diversity and functions. Microbial ecology in health and disease **12**:193–207.
7. **Eran Y**, **Getter Y**, **Baruch M**, **Belotserkovsky I**, **Padalon G**, **Mishalian I**, **Podbielski A**, **Kreikemeyer B**, **Hanski E**. 2007. Transcriptional regulation of the *sil* locus by the SilCR signalling peptide and its implications on group A *Streptococcus* virulence. Molecular Microbiology **63**:1209–1222.
8. **Fiedler T**, **Riani C**, **Koczan D**, **Standar K**, **Kreikemeyer B**, **Podbielski A**. 2013. Protective mechanisms of respiratory tract streptococci against *Streptococcus pyogenes* biofilm formation and epithelial cell infection. Applied and Environmental Microbiology **79**:1265–1276.
9. **Hidalgo-Grass C**, **Ravins M**, **Dan-Goor M**, **Jaffe J**, **Moses AE**, **Hanski E**. 2002. A locus of group A *Streptococcus* involved in invasive disease and DNA transfer. Molecular Microbiology **46**:87–99.
10. **Hibbing ME**, **Fuqua C**, **Parsek MR**, **Peterson SB**. 2010. Bacterial competition: surviving and thriving in the microbial jungle. Nature Publishing Group **8**:15–25.
11. **Michael-Gayego A**, **Dan-Goor M**, **Jaffe J**, **Hidalgo-Grass C**, **Moses AE**. 2013. Characterization of *sil* in invasive group A and G streptococci: antibodies against bacterial pheromone peptide SilCR result in severe infection. Infection and Immunity **81**:4121–4127.
12. **Chiller K**, **Selkin BA**, **Murakawa GJ**. 2001. Skin microflora and bacterial infections of the skin. J Investig Dermatol Symp Proc **6**:170–174.
13. **Hentges DJ**. 1993. The anaerobic microflora of the human body. CLIN INFECT DIS **16**:S175–S180.
14. **Cogen AL**, **Nizet V**, **Gallo RL**. 2008. Skin microbiota: a source of disease or defence? British Journal of Dermatology **158**:442–455.
15. **Maudsdotter L**, **Jonsson H**, **Roos S**, **Jonsson A-B**. 2011. Lactobacilli reduce cell cytotoxicity caused by *Streptococcus pyogenes* by producing lactic acid that degrades the toxic component lipoteichoic acid. Antimicrobial Agents and Chemotherapy **55**:1622–1628.
16. **Dewhirst FE**, **Chen T**, **Izard J**, **Paster BJ**, **Tanner AC**, **Yu W-H**, **Lakshmanan A**, **Wade WG**. 2010. The human oral microbiome. Journal of

Bacteriology **192**:5002–5017.

17. **Marcotte H**, **Lavoie MC**. 1998. Oral microbial ecology and the role of salivary immunoglobulin A. Microbiol Mol Biol Rev **62**:71–109.

18. **Gao Z**, **Kang Y**, **Yu J**, **Ren L**. 2014. Human pharyngeal microbiome may play a protective role in respiratory tract infections. Genomics, Proteomics & Bioinformatics **12**:144–150.

19. **Aas JA**, **Paster BJ**, **Stokes LN**, **Olsen I**, **Dewhirst FE**. 2005. Defining the normal bacterial flora of the oral cavity. J Clin Microbiol **43**:5721–5732.

20. **Wilson M**. 2008. Bacteriology of humans: an ecological perspective. John Wiley & Sons.

21. **Rizzo A**, **Losacco A**, **Carratelli CR**, **Di Domenico M**, **Bevilacqua N**. 2013. *Lactobacillus plantarum* reduces *Streptococcus pyogenes* virulence by modulating the IL-17, IL-23 and Toll-like receptor 2/4 expressions in human epithelial cells. International Immunopharmacology **17**:453–461.

22. **Guglielmetti S**, **Taverniti V**, **Minuzzo M**, **Arioli S**, **Stuknyte M**, **Karp M**, **Mora D**. 2010. Oral bacteria as potential probiotics for the pharyngeal mucosa. Applied and Environmental Microbiology **76**:3948–3958.

23. **Cintas LM**, **Casaus MP**, **Herranz C**, **Nes IF**, **Hernandez PE**. 2001. Review: bacteriocins of lactic acid bacteria. Food Science and Technology International.

24. **Cotter PD**, **Hill C**, **Ross RP**. 2005. Bacteriocins: developing innate immunity for food. Nat Rev Microbiol **3**:777–788.

25. **Riley MA**, **Chavan MA**. 2006. Bacteriocins: ecology and evolution. Springer.

26. **Nes IF**, **Yoon S-S**, **Diep DB**. 2007. Ribosomally synthesiszed antimicrobial peptides (bacteriocins) in lactic acid bacteria: a review. Food Science and Biotechnology.

27. **Belotserkovsky I**, **Baruch M**, **Peer A**, **Dov E**, **Ravins M**, **Mishalian I**, **Persky M**, **Smith Y**, **Hanski E**. 2009. Functional analysis of the quorum-sensing streptococcal invasion locus (*sil*). PLoS Pathog **5**:e1000651.

28. **Cotter PD**, **Ross RP**, **Hill C**. 2013. Bacteriocins - a viable alternative to antibiotics? Nature Publishing Group **11**:95–105.

29. **Dawid S**, **Sebert ME**, **Weiser JN**. 2009. Bacteriocin activity of *Streptococcus pneumoniae* is controlled by the serine protease HtrA via posttranscriptional regulation. Journal of Bacteriology **191**:1509–1518.

30. **Schillinger U**, **Holzapfel R**, **Holzapfel WH**. 1996. Potential of antagonistic microorganisms and bacteriocins for the biological preservation of foods. Trends in Food Science & Technology **7**:1–7.

31. **Pennisi E**. 2005. A mouthful of microbes. Science **307**:1899–1901.

32. **Yang S-C**, **Lin C-H**, **Sung CT**, **Fang J-Y**. 2014. Antibacterial activities of bacteriocins: application in foods and pharmaceuticals. Front Microbiol **5**:241.

33. **Okuda K-I**, **Zendo T**, **Sugimoto S**, **Iwase T**, **Tajima A**, **Yamada S**, **Sonomoto K**, **Mizunoe Y**. 2013. Effects of bacteriocins on methicillin-resistant *Staphylococcus aureus* biofilm. Antimicrobial Agents and Chemotherapy **57**:5572–5579.

34. **Galvin M**, **Hill C**, **Ross RP**. 1999. Lacticin 3147 displays activity in buffer

against gram-positive bacterial pathogens which appear insensitive in standard plate assays. Lett Appl Microbiol **28**:355–358.

35. **Cornut G**, **Fortin C**, **Soulières D**. 2008. Antineoplastic properties of bacteriocins. American Journal of Clinical Oncology **31**:399–404.

36. **Lancaster LE**, **Wintermeyer W**, **Rodnina MV**. 2007. Colicins and their potential in cancer treatment. Blood Cells, Molecules, and Diseases **38**:15–18.

37. **Klaenhammer TR**. 1993. Genetics of bacteriocins produced by lactic acid bacteria. FEMS Microbiology Reviews **12**:39–85.

38. **Perez RH**, **Zendo T**, **Sonomoto K**. 2014. Novel bacteriocins from lactic acid bacteria (LAB): various structures and applications. Microb Cell Fact **13 Suppl 1**:S3.

39. **Zacharof MP**, **Lovittb RW**. 2012. Bacteriocins produced by lactic acid bacteria. Procedia APCBEE.

40. **Willey JM**, **van der Donk WA**. 2007. Lantibiotics: peptides of diverse structure and function. Annu Rev Microbiol **61**:477–501.

41. **Tunkel AR**, **Sepkowitz KA**. 2002. Infections caused by viridans streptococci in patients with neutropenia. CLIN INFECT DIS **34**:1524–1529.

42. **Oman TJ**, **Boettcher J**, **Wang H**, **Okalibe XN**, **van der Donk WA**. 2011. Sublancin is not a lantibiotic but an S-linked glycopeptide. Nature chemical biology **7**:78–80.

43. **Stepper J**, **Shastri S**, **Loo TS**, **Preston JC**, **Novak P**, **Man P**, **Moore CH**, **Havlíček V**, **Patchett ML**, **Norris GE**. 2011. Cysteine S-glycosylation, a new post-translational modification found in glycopeptide bacteriocins. FEBS Lett **585**:645–650.

44. **Nishie M**, **Nagao J-I**, **Sonomoto K**. 2012. Antibacterial peptides "bacteriocins": an overview of their diverse characteristics and applications. Biocontrol Sci **17**:1–16.

45. **Eijsink VGH**, **Axelsson L**, **Diep DB**, **Håvarstein LS**, **Holo H**, **Nes IF**. 2002. Production of class II bacteriocins by lactic acid bacteria; an example of biological warfare and communication. Antonie Van Leeuwenhoek **81**:639–654.

46. **Riley MA**, **Wertz JE**. 2002. Bacteriocins: evolution, ecology, and application. **56**:117–137.

47. **Nes IF**, **Holo H**. 2000. Class II antimicrobial peptides from lactic acid bacteria. Biopolymers **55**:50–61.

48. **Oppegård C**, **Fimland G**, **Thorbaek L**, **Nissen-Meyer J**. 2007. Analysis of the two-peptide bacteriocins lactococcin G and enterocin 1071 by site-directed mutagenesis. Applied and Environmental Microbiology **73**:2931–2938.

49. **Zendo T**, **Koga S**, **Shigeri Y**, **Nakayama J**, **Sonomoto K**. 2006. Lactococcin Q, a novel two-peptide bacteriocin produced by *Lactococcus lactis* QU 4. Applied and Environmental Microbiology **72**:3383–3389.

50. **Nissen-Meyer J**, **Holo H**, **Håvarstein LS**, **Sletten K**, **Nes IF**. 1992. A novel lactococcal bacteriocin whose activity depends on the complementary action of two peptides. Journal of Bacteriology **174**:5686–5692.

51. **Moll G**, **Ubbink-Kok T**, **Hildeng-Hauge H**, **Nissen-Meyer J**, **Nes IF**,

Konings WN, Driessen AJ. 1996. Lactococcin G is a potassium ion-conducting, two-component bacteriocin. Journal of Bacteriology 178:600–605.

52.   Moll G, Hildeng-Hauge H, Nissen-Meyer J, Nes IF, Konings WN, Driessen AJ. 1998. Mechanistic properties of the two-component bacteriocin lactococcin G. Journal of Bacteriology 180:96–99.

53.   Rogne P, Fimland G, Nissen-Meyer J, Kristiansen PE. 2008. Three-dimensional structure of the two peptides that constitute the two-peptide bacteriocin lactococcin G. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1784:543–554.

54.   Hauge HH, Nissen-Meyer J, Nes IF, Eijsink VG. 1998. Amphiphilic alpha-helices are important structural motifs in the alpha and beta peptides that constitute the bacteriocin lactococcin G-enhancement of helix formation upon alpha-beta interaction. Eur J Biochem 251:565–572.

55.   Oppegård C, Schmidt J, Kristiansen PE, Nissen-Meyer J. 2008. Mutational analysis of putative helix−helix interacting GxxxG-motifs and tryptophan residues in the two-peptide bacteriocin lactococcin G. Biochemistry 47:5242–5249.

56.   Oppegård C, Rogne P, Kristiansen PE, Nissen-Meyer J. 2010. Structure analysis of the two-peptide bacteriocin lactococcin G by introducing D-amino acid residues. Microbiology 156:1883–1889.

57.   Kjos M, Oppegård C, Diep DB, Nes IF, Veening J-W, Nissen-Meyer J, Kristensen T. 2014. Sensitivity to the two-peptide bacteriocin lactococcin G is dependent on UppP, an enzyme involved in cell-wall synthesis. Molecular Microbiology 92:1177–1187.

58.   Håvarstein LS, Diep DB, Nes IF. 1995. A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. Molecular Microbiology 16:229–240.

59.   Venema K, Haverkort RE, Abee T, Haandrikman AJ, Leenhouts KJ, de Leij L, Venema G, Kok J. 1994. Mode of action of LciA, the lactococcin A immunity protein. Molecular Microbiology 14:521–532.

60.   Quadri LE, Sailer M, Terebiznik MR, Roy KL, Vederas JC, Stiles ME. 1995. Characterization of the protein conferring immunity to the antimicrobial peptide carnobacteriocin B2 and expression of carnobacteriocins B2 and BM1. Journal of Bacteriology 177:1144–1151.

61.   Diep DB, Skaugen M, Salehian Z, Holo H, Nes IF. 2007. Common mechanisms of target cell recognition and immunity for class II bacteriocins. Proc Natl Acad Sci USA 104:2384–2389.

62.   Oppegård C, Emanuelsen L, Thorbek L, Fimland G, Nissen-Meyer J. 2010. The lactococcin G immunity protein recognizes specific regions in both peptides constituting the two-peptide bacteriocin lactococcin G. Applied and Environmental Microbiology 76:1267–1273.

63.   Kleerebezem M, Quadri LE. 2001. Peptide pheromone-dependent regulation of antimicrobial peptide production in Gram-positive bacteria: a case of multicellular behavior. Peptides 22:1579–1596.

64.   Straume D, Kjos M, Nes IF, Diep DB. 2007. Quorum-sensing based bacteriocin production is down-regulated by N-terminally truncated species

of gene activators. Mol Genet Genomics **278**:283–293.

65. **Nes IF**, **Diep DB**, **Håvarstein LS**, **Brurberg MB**, **Eijsink V**, **Holo H**. 1996. Biosynthesis of bacteriocins in lactic acid bacteria. Antonie Van Leeuwenhoek **70**:113–128.

66. **Sherwood NP**, **Russell BE**, **Jay AR**, **Bowman K**. 1949. Studies on streptococci. III. New antibiotic substances produced by beta hemolytic streptococci. J Infect Dis **84**:88–91.

67. **Tagg JR**, **Wannamaker LW**. 1976. Genetic basis of streptococcin A-FF22 production. Antimicrobial Agents and Chemotherapy **10**:299–306.

68. **Hynes WL**, **Ferretti JJ**, **Tagg JR**. 1993. Cloning of the gene encoding streptococcin A-FF22, a novel lantibiotic produced by *Streptococcus pyogenes*, and determination of its nucleotide sequence. Applied and Environmental Microbiology **59**:1969–1971.

69. **Plainvert C**, **Dinis M**, **Ravins M**, **Hanski E**, **Touak G**, **Dmytruk N**, **Fouet A**, **Poyart C**. 2014. Molecular epidemiology of *sil* locus in clinical *Streptococcus pyogenes* strains. J Clin Microbiol **52**:2003–2010.

70. **Tagg JR**, **Read RS**, **McGiven AR**. 1973. Bacteriocin of a group A *Streptococcus*: partial purification and properties. Antimicrobial Agents and Chemotherapy **4**:214–221.

71. **Hynes WL**, **Friend VL**, **Ferretti JJ**. 1994. Duplication of the lantibiotic structural gene in M-type 49 group A *Streptococcus* strains producing streptococcin A-M49. Applied and Environmental Microbiology **60**:4207–4209.

72. **Hynes WL**, **Tagg JR**. 1985. Production of broad-spectrum bacteriocin-like activity by group A streptococci of particular M-types. Zentralbl Bakteriol Mikrobiol Hyg A **259**:155–164.

73. **Wescombe PA**, **Tagg JR**. 2003. Purification and characterization of streptin, a type A1 lantibiotic produced by *Streptococcus pyogenes*. Applied and Environmental Microbiology **69**:2737–2747.

74. **Karaya K**, **Shimizu T**, **Taketo A**. 2001. New gene cluster for lantibiotic streptin possibly involved in streptolysin S formation. J Biochem **129**:769–775.

75. **Simpson WJ**, **Tagg JR**. 1983. M-type 57 group A *Streptococcus* bacteriocin. Can J Microbiol **29**:1445–1451.

76. **Kizy AE**, **Neely MN**. 2009. First *Streptococcus pyogenes* signature-tagged mutagenesis screen identifies novel virulence determinants. Infection and Immunity **77**:1854–1865.

77. **Heng NCK**, **Burtenshaw GA**, **Jack RW**, **Tagg JR**. 2004. Sequence analysis of pDN571, a plasmid encoding novel bacteriocin production in M-type 57 *Streptococcus pyogenes*. Plasmid **52**:225–229.

78. **Hidalgo-Grass C**, **Dan-Goor M**, **Maly A**, **Eran Y**, **Kwinn LA**, **Nizet V**, **Ravins M**, **Jaffe J**, **Peyser A**, **Moses AE**, **Hanski E**. 2004. Effect of a bacterial pheromone peptide on host chemokine degradation in group A streptococcal necrotising soft-tissue infections. Lancet **363**:696–703.

79. **Salim KY**, **de Azavedo JC**, **Bast DJ**, **Cvitkovitch DG**. 2008. Regulation of *sagA*, *siaA* and *scpC* by SilCR, a putative signaling peptide of *Streptococcus pyogenes*. FEMS Microbiology Letters **289**:119–125.

80. **Lembke C**, **Podbielski A**, **Hidalgo-Grass C**, **Jonas L**, **Hanski E**, **Kreikemeyer B**. 2006. Characterization of biofilm formation by clinically relevant serotypes of group A streptococci. Applied and Environmental Microbiology **72**:2864–2875.

81. **Smoot J**, **Barbian K**, **Van Gompel J**, **Smoot L**, **Chaussee MS**, **Sylva GL**, **Sturdevant DE**, **Ricklefs SM**, **Porcella SF**, **Parkins LD**, **Beres SB**, **Campbell DS**, **Smith TM**, **Zhang Q**, **Kapur V**, **Daly JA**, **Veasy LG**, **Musser JM**. 2002. Genome sequence and comparative microarray analysis of serotype M18 group A *Streptococcus* strains associated with acute rheumatic fever outbreaks. Proc Natl Acad Sci USA **99**:4668–4673.

82. **Biswas I**, **Gruss A**, **Ehrlich SD**, **Maguin E**. 1993. High-efficiency gene inactivation and replacement system for Gram-positive bacteria. Journal of Bacteriology **175**:3628.

83. **Hidalgo-Grass C**, **Mishalian I**, **Dan-Goor M**, **Belotserkovsky I**, **Eran Y**, **Nizet V**, **Peled A**, **Hanski E**. 2006. A streptococcal protease that degrades CXC chemokines and impairs bacterial clearance from infected tissues. EMBO J **25**:4628–4637.

84. **van Belkum MJ**, **Worobo RW**, **Stiles ME**. 1997. Double-glycine-type leader peptides direct secretion of bacteriocins by ABC transporters: colicin V secretion in *Lactococcus lactis*. Molecular Microbiology **23**:1293–1301.

85. **Nissen-Meyer J**, **Oppegård C**, **Rogne P**, **Haugen HS**, **Kristiansen PE**. 2010. Structure and mode-of-action of the two-peptide (class-IIb) bacteriocins. Probiotics Antimicrob Proteins **2**:52–60.

86. **Salim KY**, **Cvitkovitch DG**, **Chang P**, **Bast DJ**, **Handfield M**, **Hillman JD**, **de Azavedo JCS**. 2005. Identification of group A *Streptococcus* antigenic determinants upregulated *in vivo*. Infection and Immunity **73**:6026–6038.

87. **Dawid S**, **Roche AM**, **Weiser JN**. 2007. The *blp* bacteriocins of *Streptococcus pneumoniae* mediate intraspecies competition both *in vitro* and *in vivo*. Infection and Immunity **75**:443–451.

88. **Bidet P**, **Courroux C**, **Salgueiro C**, **Carol A**, **Mariani-Kurkdjian P**, **Bonacorsi S**, **Bingen E**. 2007. Molecular epidemiology of the *sil* streptococcal invasive locus in group A streptococci causing invasive infections in french children. J Clin Microbiol **45**:2002–2004.

89. **Jing HB**. 2006. Epidemiological analysis of group A streptococci recovered from patients in China. Journal of Medical Microbiology **55**:1101–1107.

90. **Billal DS**, **Hotomi M**, **Shimada J**, **Fujihara K**, **Ubukata K**, **Sugita R**, **Yamanaka N**. 2008. Prevalence of streptococcus invasive locus (*sil*) and its relationship with macrolide resistance among group A *Streptococcus* strains. J Clin Microbiol **46**:1563–1564.

91. **Nikolskaya AN**, **Galperin MY**. 2002. A novel type of conserved DNA-binding domain in the transcriptional regulators of the AlgR/AgrA/LytR family. Nucleic Acids Research **30**:2453–2459.

92. **Diep DB**, **Håvarstein LS**, **Nes IF**. 1996. Characterization of the locus responsible for the bacteriocin production in *Lactobacillus plantarum* C11. Journal of Bacteriology **178**:4472–4483.

93. **de Saizieu A**, **Gardès C**, **Flint N**, **Wagner C**, **Kamber M**, **Mitchell TJ**,

       Keck W, **Amrein KE**, **Lange R**. 2000. Microarray-based identification of a novel *Streptococcus pneumoniae* regulon controlled by an autoinduced peptide. Journal of Bacteriology **182**:4696–4703.

94.   **Risøen PA**, **Johnsborg O**, **Diep DB**, **Hamoen L**, **Venema G**, **Nes IF**. 2001. Regulation of bacteriocin production in *Lactobacillus plantarum* depends on a conserved promoter arrangement with consensus binding sequence. Mol Genet Genomics **265**:198–206.

95.   **Knutsen E**, **Ween O**, **Håvarstein LS**. 2004. Two separate quorum-sensing systems upregulate transcription of the same ABC transporter in *Streptococcus pneumoniae*. Journal of Bacteriology **186**:3078–3085.

96.   **Blomqvist T**, **Steinmoen H**, **Håvarstein LS**. 2006. Pheromone-induced expression of recombinant proteins in *Streptococcus thermophilus*. Arch Microbiol **186**:465–473.

CHAPTER 4: CONCLUSIONS

S. pyogenes is a well known pathogen that is responsible for over 500 000 deaths worldwide each year. While much is known about *S. pyogenes* pathogenesis, how this organism is able to successfully attach, colonize, and compete within the human pharynx remains incomplete (1, 2). This is due to the fact that a large number of factors exhibit overlapping functions, creating a complex system that is difficult to dissect. In order to further solve this problem, information needs to be gathered at a genomic level to evaluate how *S. pyogenes* alters gene regulation within specific environments. Numerous methods exist, each having their own advantages and setbacks. In this research, RIVET was successfully used for the first time with *S. pyogenes*, to identify genetic elements that were specifically induced during the entire colonization process of the nasopharyngeal humanized mouse model.

4.1 Chapter 2 Conclusions

RIVET was chosen due to its ability to detect gene induction activity throughout the entire infection period, as opposed to alternative methods that capture results at specific time points. In addition, since each recovered RIVET clone is sequenced, even those from microenvironments within the model can still be identified. Part of the second-generation RIVET system was the use of counter-selection, and while this was attempted with the use of HSV-*tk* and ganciclovir, it unfortunately proved too unreliable to pursue. Despite this set back, patching to identify tet sensitive clones was still successfully performed to identify 9 potential 'typical' promoters, demonstrating a functional system **(Table 2.4)**.

In the past decade there has been a surge in the increase of sRNA found in bacterial chromosomes with increasing questions about their role in biological systems and precise mechanisms of action (3). Several methods have been used in order to identify sRNA in bacteria including bioinformatics, microarrays, RNA-seq, and RIVET. sRNA may be better identified in RIVET since they are not locked to a specific algorithm as with bioinformatics and may experience fewer overall artifacts due to the absence of a cDNA synthesis step as in microarrays and RNA-seq (3-5). Potentially, sRNA for three interesting virulence factors, *mga*, *speA*, and *mf3*, was found utilizing the RIVET system.

These virulence factors have shown states of repression during the late growth phases of animal models, presumed to be transitioning to the persistent state (6, 7). It is possible that *S. pyogenes* makes use of sRNA as a mechanism of tight repression of these genes, as well as others, in response to precise *in vivo* conditions. Many other potential sRNA candidates were also identified by the RIVET system and have yet to be explored more thoroughly to identify possible roles in colonization.

When considering 'typical' promoters that appeared to control the expression of genes, three of these, pyruvate formate-lyase, tagatose-6-phosphate aldose/ketose isomerase, and a polar amino acid ABC transporter are all suggestive of genes being active in an environment where nutrients are scarce, such as the nasopharynx (8-13). The functions of the remaining genes identified by RIVET are less obvious due to the limited amount of available information and their potential role in the colonization process. However, further verification with clone IVI156 revealed a promoter for the bacteriocin-like gene *spyM18_0544* (*blpM^H*) with consistently induced *in vivo* activity **(Figure 2.10).** This lead to further work to identify how *blpM^H* was used by *S. pyogenes* in the colonization process.

4.2.1 Future Work

The promoters identified through RIVET represent potential genetic elements contributing to the overall adaptation of *S. pyogenes* when colonizing the nasopharynx in order to survive. Confirming true *in vivo* activation through further testing via the RIVET method or qRT-PCR is the first step in deciphering their role in the colonization process. In addition to this, further investigational analysis of remaining recovered clones could be performed in order to identify more *in vivo* induced promoters.

Continued development of RIVET within *S. pyogenes* could also successfully make use of other methods to facilitate a post-*in vivo* selection screening method in the same vein as counter-selection. One potential method is the inclusion of a fluorescent marker under the control of a constitutive promoter within the cassette. This would allow the use of FACS to identify those cells without the cassette after the *in vivo* selection. Other groups, for example, have taken a cassette containing a single antibiotic resistance marker and

inserted it after the first codon of a second resistant marker causing it to be split, losing expression of the gene. Resolution of the cassette allowed for the reassembly of the split gene and the activation of a new antibiotic resistance (14, 15). Lastly, the recessive nature and streptomycin resistance of *rpsL* could be exploited. This would require the addition of the wild-type MGAS8232 *rpsL* into the cassette, while the chromosomal *rpsL* would be mutated, as found in *S. pyogenes* JRS4, to create a base streptomycin resistance. When both *rpsL* versions are present in the same cell the resistant version is recessive and, therefore, only loss of the cassette allows for streptomycin resistance (16, 17). The successful addition of one of these systems would allow for a far more efficient screening process and a greatly improved RIVET system in *S. pyogenes*.

4.2 Chapter 3 Conclusions

Found to be a Class IIb bacteriocin, $blpM^H$ would require $blpN^H$ for bacteriocidal activity and the associated immunity protein for self-survival **(Figure 3.1)**. To our knowledge, this is the first time $blpMN^H$ has been shown to be activated *in vivo* under biological conditions. Within MGAS8232, we also believe $blpMN^H$ can be under the control of a regulatory system other than *sil*. This stems from the inability of wild-type MGAS8232 to respond to the SilCR pheromone *in vitro* or growth in many *ex vivo* conditions, yet can eliminate MGAS8232 Δ*blp* from the correct *in vivo* environment **(Figure 3.9)**.

S. pyogenes must be able to outcompete endogenous bacteria already present whithin the human body (13, 26-28). In spite of this, few species of *S. pyogenes* are known to express bacteriocins, and none have been shown to express Class IIb, or make use of a bacteriocin to promote *in vivo* survival (18-25). Bacteriocins and their use against pathogens is an increasing area of research due to its potential as an antibiotic alternative (29). Yet, their use by the pathogens themselves as another virulence factor for gaining access to its ideal niche does not appear to be an area of active research. Deciphering how *S. pyogenes* may make use of bacteriocins for niche adaptation is a fascinating area of potential research and could have impacts on our knowledge of other human pathogens. Since we are still at the beginning stages, many questions still remain concerning regulation, immunity, and secretion of $blpMN^H$ during the colonization process of *S. pyogenes*. Continued research will ultimately increase our overall

knowledge of how *S. pyogenes*, and possibly other pathogens, interact with the microbiota of the human body when attempting to colonize and cause infection.

4.2.1 Future Work

A necessary part of the bacteriocin system is an immunity protein. Recent work from our group using functional experiments has identified an immunity phenotype within a fragment containing both *spyM18_0546* and *spyM18_0547*. Similar experiments can be repeated with smaller portions of this fragment to identify the coding region responsible for immunity function. The immunity protein, along with *blpMN^H*, likely makes up the entire operon regulated by the promoter identified by IVI156 with the RIVET system.

Further work is also critical in determining the method of secretion for BlpMN^H. It is first necessary to determine if the present *silE, silD*, and *spyM18_0541* are responsible for the secretion of BlpMN^H. It may be possible to express both bacteriocin and transporter from a plasmid under the control of an inducible promoter (such as pMSP3535) within a heterologous host such as *L. lactis*. One possibility is that the truncated *silD* is not required, as the secondary transport protein has been shown to be unnecessary for at least one Class IIb bacteriocin, and possibly others (30). Should *silE* not be the transporter, further searching through the use of bioinformatics may be used to identify a potential source for further testing.

Finally, the lack of induction via SilCR would suggest regulation through a different system. However, attempting to decipher how *blpMN^H* and its immunity protein are regulated may be difficult. Should regulation occur by the binding of a protein to the DNA region upstream of *blpMN^H*, whether as part of a two-component system or another regulatory system, pull down assays with *in vivo* grown MGAS8232 cell lysates, for example, could be performed using the known DNA sequence attached to beads as bait. Mass spectrometry could then be used to identify any separated proteins. Identification of this regulator would be necessary to determine the activation process of *blpMN^H* within the nasopharynx.

4.3 Overall Conclusions

Utilizing the RIVET method in order to identity promoters involved with the colonization process by *S. pyogenes* has laid the foundation for numerous areas of research to pursue. Adjustments to the RIVET system used with *S. pyogenes* could improve its efficiency and applicability for analysis in other environments such as human blood or saliva, or in models of necrotizing fasciitis or other invasive diseases. A fascinating potential colonization-promoting mechanism used by *S. pyogenes* identified via RIVET is through the use of bacteriocins. Interest in bacteriocins has increased significantly in the past ten years, especially in LAB, due to the potential use of bacteriocins as an antibiotic alternative in both clinical and food industry settings (31). However, the bacteriocins of pathogens, used as virulence factors, have not yet received much attention. The few studies represented have demonstrated the importance of bacteriocins with *S. pneumoniae* (32) and *S. mutans* (33) in order to colonize their preferred niche, validating their importance. Identifying the activation of bacteriocins, through this work, by *S. pyogenes* during colonization also contributes to this idea. Ascertaining when these bacteriocins are expressed, along with the bacteria they target, may provide important information as to how pathogens such as *S. pyogenes* cause disease and may aid in identifying improved ways to inhibit or prevent their colonization and reduce the global burden.
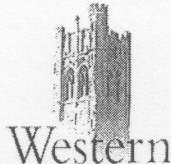
4.4 References

1. **Cunningham MW**. 2000. Pathogenesis of group A streptococcal infections. Clin Microbiol Rev **13**:470.
2. **Walker MJ**, **Barnett TC**, **McArthur JD**, **Cole JN**, **Gillen CM**, **Henningham A**, **Sriprakash KS**, **Sanderson-Smith ML**, **Nizet V**. 2014. Disease manifestations and pathogenic mechanisms of group a *Streptococcus*. Clin Microbiol Rev **27**:264–301.
3. **Thomason MK**, **Storz G**. 2010. Bacterial antisense RNAs: how many are there, and what are they doing? Annu Rev Genet **44**:167–188.
4. **Silby MW**, **Rainey PB**, **Levy SB**. 2004. IVET experiments in *Pseudomonas fluorescens* reveal cryptic promoters at loci associated with recognizable overlapping genes. Microbiology (Reading, Engl) **150**:518–520.
5. **Holt JF**, **Kiedrowski MR**, **Frank KL**, **Du J**, **Guan C**, **Broderick NA**, **Dunny GM**, **Handelsman J**. 2014. *Enterococcus faecalis* 6-phosphogluconolactonase is required for both commensal and pathogenic interactions with *Manduca sexta*. Infection and Immunity.
6. **Kreikemeyer B**, **McIver KS**, **Podbielski A**. 2003. Virulence factor regulation and regulatory networks in *Streptococcus pyogenes* and their impact on pathogen–host interactions. Trends in Microbiology **11**:224–232.
7. **Virtaneva K**, **Porcella SF**, **Graham MR**, **Ireland RM**, **Johnson CA**, **Ricklefs SM**, **Babar I**, **Parkins LD**, **Romero RA**, **Corn GJ**, **Gardner DJ**, **Bailey JR**, **Parnell MJ**, **Musser JM**. 2005. Longitudinal analysis of the group A *Streptococcus* transcriptome in experimental pharyngitis in cynomolgus macaques. Proc Natl Acad Sci USA **102**:9014–9019.
8. **Yamamoto Y**, **Sato Y**, **Takahashi-Abbe S**, **Takahashi N**, **Kizaki H**. 2000. Characterization of the *Streptococcus mutans* pyruvate formate-lyase (PFL)-activating enzyme gene by complementary reconstitution of the *in vitro* PFL-reactivating system. Infection and Immunity **68**:4773–4777.
9. **Hu Z**, **Patel IR**, **Mukherjee A**. 2013. Genetic analysis of the roles of *agaA*, *agaI*, and *agaS* genes in the N-acetyl-D-galactosamine and D-galactosamine catabolic pathways in *Escherichia coli* strains O157:H7 and C. BMC Microbiol **13**:94.
10. **Lei B**, **Liu M**, **Chesney GL**, **Musser JM**. 2004. Identification of new candidate vaccine antigens made by *Streptococcus pyogenes*: purification and characterization of 16 putative extracellular lipoproteins. J Infect Dis **189**:79–89.
11. **Livezey J**, **Perez L**, **Suciu D**, **Yu X**, **Robinson B**, **Bush D**, **Merrill G**. 2011. Analysis of group A *Streptococcus* gene expression in humans with pharyngitis using a microarray. Journal of Medical Microbiology **60**:1725–1733.
12. **Wilson M**. 2008. Bacteriology of humans: an ecological perspective. John Wiley & Sons.
13. **Wilson M**. 2005. Microbial inhabitants of humans. Cambridge University Press.
14. **Sawicki JA**, **Monks B**, **Morris RJ**. 1998. Cell-specific ecdysone-inducible expression of FLP recombinase in mammalian cells. Biotech **25**:868–70– 872–5.
15. **Gersbach CA**, **Gaj T**, **Gordley RM**, **Barbas CF**. 2010. Directed evolution of recombinase specificity by split gene reassembly. Nucleic Acids Research **38**:4198–4206.
16. **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**:1591–1601.

17. **Reyrat JM**, **Pelicic V**, **Gicquel B**, **Rappuoli R**. 1998. Counterselectable markers: untapped tools for bacterial genetics and pathogenesis. Infection and Immunity **66**:4011.

18. **Sherwood NP**, **Russell BE**, **Jay AR**, **Bowman K**. 1949. Studies on streptococci. III. New antibiotic substances produced by beta hemolytic streptococci. J Infect Dis **84**:88–91.

19. **Tagg JR**, **Wannamaker LW**. 1976. Genetic basis of streptococcin A-FF22 production. Antimicrobial Agents and Chemotherapy **10**:299–306.

20. **Hynes WL**, **Ferretti JJ**, **Tagg JR**. 1993. Cloning of the gene encoding streptococcin A-FF22, a novel lantibiotic produced by *Streptococcus pyogenes*, and determination of its nucleotide sequence. Applied and Environmental Microbiology **59**:1969–1971.

21. **Tagg JR**, **Read RS**, **McGiven AR**. 1973. Bacteriocin of a group A *Streptococcus*: partial purification and properties. Antimicrobial Agents and Chemotherapy **4**:214–221.

22. **Hynes WL**, **Friend VL**, **Ferretti JJ**. 1994. Duplication of the lantibiotic structural gene in M-type 49 group A *Streptococcus* strains producing streptococcin A-M49. Applied and Environmental Microbiology **60**:4207–4209.

23. **Hynes WL**, **Tagg JR**. 1985. Production of broad-spectrum bacteriocin-like activity by group A streptococci of particular M-types. Zentralbl Bakteriol Mikrobiol Hyg A **259**:155–164.

24. **Wescombe PA**, **Tagg JR**. 2003. Purification and characterization of streptin, a type A1 lantibiotic produced by *Streptococcus pyogenes*. Applied and Environmental Microbiology **69**:2737–2747.

25. **Simpson WJ**, **Tagg JR**. 1983. M-type 57 group A *Streptococcus* bacteriocin. Can J Microbiol **29**:1445–1451.

26. **Lemon KP**, **Klepac-Ceraj V**, **Schiffer HK**, **Brodie EL**, **Lynch SV**, **Kolter R**. 2010. Comparative analyses of the bacterial microbiota of the human nostril and oropharynx. MBio **1**:1–9.

27. **Gao Z**, **Kang Y**, **Yu J**, **Ren L**. 2014. Human pharyngeal microbiome may play a protective role in respiratory tract infections. Genomics, Proteomics & Bioinformatics **12**:144–150.

28. **Aas JA**, **Paster BJ**, **Stokes LN**, **Olsen I**, **Dewhirst FE**. 2005. Defining the normal bacterial flora of the oral cavity. J Clin Microbiol **43**:5721–5732.

29. **Cavera VL**, **Arthur TD**, **Kashtanov D**, **Chikindas ML**. 2015. Bacteriocins and their position in the next wave of conventional antibiotics. Int J Antimicrob Agents.

30. **Vaughan A**, **Eijsink VGH**, **van Sinderen D**. 2003. Functional characterization of a composite bacteriocin locus from malt isolate *Lactobacillus sakei* 5. Applied and Environmental Microbiology **69**:7194–7203.

31. **Perez RH**, **Zendo T**, **Sonomoto K**. 2014. Novel bacteriocins from lactic acid bacteria (LAB): various structures and applications. Microb Cell Fact **13 Suppl 1**:S3.

32. **Dawid S**, **Roche AM**, **Weiser JN**. 2007. The *blp* bacteriocins of *Streptococcus pneumoniae* mediate intraspecies competition both *in vitro* and *in vivo*. Infection

and Immunity **75**:443–451.

33. **Hillman JD**, **Dzuback AL**, **Andrews SW**. 1987. Colonization of the human oral cavity by a *Streptococcus mutans* mutant producing increased bacteriocin. J Dent Res **66**:1092–1094.

APPENDICES

Appendix 1: Animal ethics approval

Western

April 27, 2009

This is the Original Approval for this protocol
A Full Protocol submission will be required in 2013

Dear Dr. McCormick:

Your Animal Use Protocol form entitled:
Bacterial SAgs Play a Key Role in the Pathogenesis of Streptococcus pyogenes
Funding Agency CIHR – Grant #RDAF-107845

has been approved by the University Council on Animal Care. This approval is valid from **April 27, 2009 to April 30, 2010**. The protocol number for this project is **2009-038**.

1. This number must be indicated when ordering animals for this project.
2. Animals for other projects may not be ordered under this number.
3. If no number appears please contact this office when grant approval is received.
   If the application for funding is not successful and you wish to proceed with the project, request that an internal scientific peer review be performed by the Animal Use Subcommittee office.
4. Purchases of animals other than through this system must be cleared through the ACVS office. Health certificates will be required.

ANIMALS APPROVED FOR 4 Years

SPECIES & SECT D.5.1 GROUP ID#	STRAIN &/or OTHER SPECIES DETAIL For Rodents, Also Provide Vendor Stock #	AGE or WEIGHT & SEX	4-YEAR TOTAL ANIMAL NUMBER
Mouse ID# 2,5,8,B1	C57Bl/6 HLA-DR4	8-11 week old/Male	72+4 breeding mate
Mouse ID# 3,6,9,B2	C57Bl/6 HLA-DR4/DQ8	8-11 week old/Male	72+4 breeding mate
Mouse ID# 11	C57Bl/6 HLA-DQ6	8-11 week old/Male	24+4 breeding mate
Mouse ID# 1,4,7,10	C57Bl/6	8-11 week old/Male	96
Mouse ID# 2,5,8,B1	C57Bl/6 HLA-DR4	8-11 week old/Female	72+4 breeding mate
Mouse ID# 3,6,9,B2	C57Bl/6 HLA-DR4/DQ8	8-11 week old/Female	72+4 breeding mate
Mouse ID# 11	C57Bl/6 HLA-DQ6	8-11 week old/Female	24+4 breeding mate
Mouse ID# 1,4,7,10	C578l/6	8-11 week old/Female	96

REQUIREMENTS/COMMENTS
Please ensure that individual(s) performing procedures on live animals, as described in this protocol, are familiar with the contents of this document.

The holder of this Animal Use Protocol is responsible to ensure that all associated safety components (biosafety, radiation safety, general laboratory safety) comply with institutional safety standards and have received all necessary approvals. Please consult directly with your institutional safety officers.

c.c. Approved Protocol   - J. McCormick, D. Mazzuca Siroen, W. Lagerwerf
     Approval Letter   - J. McCormick, D. Mazzuca Siroen, W. Lagerwerf

Appendix 2: Clones recovered using RIVET

Clone	Size (bp)	Designation	Region in Genome	Gene designation	Features
IVI1	96	antisense	1741745 to 1741840	*spyM18_2077*	positive regulatory protein Mga
IVI3	129	cryptic	1713341 to 1713213	*spyM18_2050*	putative PTS system, enzyme II, A component
IVI4	244	cryptic	570003 to 570246	*spyM18_0710*	putative ATP-dependent DNA helicase
IVI7a	214	cryptic	912892 to 912679	*spyM18_1098*	putative thiamine biosynthesis lipoprotein precursor
IVI7b	178	antisense	1152397 to 1152574	*spyM18_1380*	putative uridine kinase
IVI11	71	antisense	214459 to 214389	*spyM18_0239*	hypothetical protein
IVI14a	853	antisense	640829 to 639977	*spyM18_0802*	SagC, hypothetical protein
IVI14b	108	cryptic	873659 to 873766	*spyM18_1051*	putative D,D-carboxypeptidase, penicillin-binding protein
IVI14c	57	cryptic	147131 to 147075	*spyM18_0155*	hypothetical protein
IVI15	64	cryptic	1725608 to 1725671	*spyM18_2064*	mitogenic exotoxin Z precursor
IVI16	113	intergenic	751419 to 751307	*spyM18_0916* / *spyM18_0921*	546 bp at 5' side: putative peptidoglycan hydrolase / 655 bp at 3' side: N-terminal fragment of Sib38 homolgue
IVI17a	312	antisense	1546687 to 1546998	*spyM18_1868*	hypothetical protein
IVI17b	162	antisense	678000 to 677839	*spyM18_0836*	putative ATP-dependent exonuclease, subunit A
IVI19	232	antisense	1491691 to 1491460	*spyM18_1806*	hypothetical phage protein
IVI22	320	cryptic	1271790 to 1271471	*spyM18_1541*	putative undecaprenyl-PP-MurNAc-pentapeptide-UDPGlcNAc GlcNAc
IVI23a	151	cryptic	1601103 to 1601253	*spyM18_1926*	conserved hypothetical protein
IVI23b	44	cryptic	1010750 to 1010707	*spyM18_1202*	conserved hypothetical protein
IVI24	514	cryptic	136555 to 137068	*spyM18_0144*	V-type Na+ -ATPase subunit I
IVI26a	296	cryptic	1073889 to 1073594	*spyM18_1498*	hypothetical phage protein
IVI26b	298	cryptic	1482683 to 1482388	*spyM18_1793*	hypothetical phage protein
IVI27	417	antisense	721697 to 721281	*spyM18_0890*	putative pyrimidine regulatory protein
IVI28a	107	antisense	677377 to 677271	*spyM18_0836*	putative ATP-dependent exonuclease, subunit A
IVI28b	78	cryptic	986273 to 986197	*spyM18_1177*	putative sugar ABC transporter (permease protein)
IVI29a	107	antisense	591092 to 590986	*spyM18_0742*	hypothetical phage protein

IVI29b	105	antisense	1231877 to 1231981	*spyM18_1488*	hypothetical phage protein
IVI29c	38	cryptic	539410 to 539447	*spyM18_0676*	putative cell division protein
IVI30	161	intergenic	1640515 to 1640355	*spyM18_1973*	73 bp at 5' side: conserved hypothetical protein
				spyM18_1974	117 bp at 3' side: putative type I site-specific deoxyribonuclease
IVI34a	161	intergenic	1640515 to 1640355	*spyM18_1973*	73 bp at 5' side: conserved hypothetical protein
				spyM18_1974	117 bp at 3' side: putative type I site-specific deoxyribonuclease
IVI34b	51	cryptic	1714518 to 1714568	*spyM18_2050*	putative PTS system, enzyme II, A component
IVI35	78	cryptic	1464307 to 1464230	*spyM18_1763*	hypothetical phage protein
IVI37	107	antisense	1781199 to 1781093	*spyM18_2133*	putative transposase
IVI38	89	intergenic	626749 to 626661	*spyM18_0789*	conserved hypothetical protein
IVI41	203	cryptic	396484 to 396686	*spyM18_0486*	putative UDP-N-acetylglucosamine pyrophosphorylase
IVI43	49	antisense	1871198 to 1871246	*spyM18_2232*	hypothetical protein
IVI46	210	antisense	1835539 to 1835748	*spyM18_2185*	conserved hypothetical protein
IVI48	310	antisense	913289 to 913598	*spyM18_1098*	putative thiamine biosynthesis lipoprotein precursor
IVI49	1029	promoter	1893418 to 1894446	*spyM18_2257*	putative chromosome segregation protein
IVI52	249	cryptic	1448278 to 1448030	*spyM18_1741*	putative transcription regulator
IVI53	481	promoter	1027534 to 1027054	*spyM18_1223*	putative amino acid ABC transporter, periplasmic amino acid-binding protein
IVI55	54	antisense	971714 to 971767	*spyM18_1163*	hypothetical protein
IVI57a	185	antisense	146737 to 146921	*spyM18_0155*	hypothetical protein
IVI57b	83	antisense	841031 to 841113	*spyM18_1014*	extracellular hyaluronate lyase
IVI59	296	cryptic	1482683 to 1482388	*spyM18_1793*	hypothetical phage protein
IVI60	959	promoter	182746 to 183704	*spyM18_0197*	hypothetical protein
IVI62	163	cryptic	1146888 to 1146726	*spyM18_1373*	putative internalin A precursor
IVI65a	257	antisense	21654 to 21398	*spyM18_r02*	rRNA-23S ribosomal RNA
IVI65a	257	antisense	27655 to 27399	*spyM18_r04*	rRNA-23S ribosomal RNA
IVI65a	257	antisense	84557 to 84301	*spyM18_r06*	rRNA-23S ribosomal RNA
IVI65a	257	antisense	262614 to 262358	*spyM18_r08*	rRNA-23S ribosomal RNA

IVI65a	257	antisense	1344577 to 1344833	*spyM18_r09*	rRNA-23S ribosomal RNA
IVI65a	257	cryptic	1634201 to 1634457	*spyM18_r11*	rRNA-23S ribosomal RNA
IVI65b	54	cryptic	1603230 to 1603177	*spyM18_1929*	conserved hypothetical protein
IVI69a	193	antisense	186190 to 185998	*spyM18_0204*	glucose-6-phosphate isomerase
IVI69b	51	cryptic	1611161 to 1611111	*spyM18_1939*	glycoprotein endopeptidase
IVI70	203	cryptic	566923 to 567125	*spyM18_0707*	cell-division protein
IVI72	600	promoter	622769 to 623367	*spyM18_0784*	transcriptional regulator
IVI78	92	antisense	641030 to 640939	*spyM18_0802*	hypothetical protein
IVI81	145	cryptic	1248054 to 1247910	*spyM18_1515*	hemolysin
IVI82a	682	antisense	466442 to 467123	*spyM18_0583*	DhaKLM operon coactivator DhaQ
IVI82b	348	cryptic	850043 to 850390	*spyM18_1022*	coproporphyrinogen III oxidase
IVI84	842	promoter	1669042 to 1668201	*spyM18_2004*	hypothetical protein
IVI87	1001	promoter	1587104 to 1588104	*spyM18_1912*	pyruvate formate-lyase
IVI91	210	cryptic	988364 to 988155	*spyM18_1179*	sugar ABC transporter ATP-binding protein
IVI92	151	antisense	1500024 to 1500174	*spyM18_1815*	acetyl-CoA carboxylase subunit alpha
IVI93a	325	cryptic	752667 to 752343	*spyM18_0921*	hypothetical protein
IVI93b	131	antisense	863452 to 863322	*spyM18_1041*	succinic semialdehyde dehydrogenase
IVI94	237	antisense	396966 to 396730	*spyM18_0486*	bifunctional N-acetylglucosamine-1-phosphate uridyltransferase/glucosamine-1-phosphate acetyltransferase
IVI95a	132	antisense	1139356 to 1139487	*spyM18_1367*	hypothetical protein
IVI95b	78	intergenic	1391342 to 1391419	*spyM18_1669* / *spyM18_1670*	7 bp at 5' side: amino acid ABC transporter permease / 20 bp at 3' side: ATP-dependent RNA helicase
IVI97a	191	cryptic	571500 to 571689	*spyM18_0710*	bifunctional ATP-dependent DNA helicase/DNA polymerase III subunit epsilon
IVI97b	115	antisense	252617 to 252503	*spyM18_0282*	oligopeptide permease
IVI100	97	promoter	612261 to 612357	*spyM18_0771*	hypothetical protein
IVI101	111	antisense	1377561 to 1377671	*spyM18_1656*	hypothetical protein
IVI106	248	antisense	401018 to 400771	*spyM18_0494*	metal binding protein of ABC transporter (lipoprotein)
IVI112	107	antisense	591092 to	*spyM18_0742*	hypothetical protein

			590986		
IVI115	431	cryptic	1690774 to 1690344	*spyM18_2030*	prolyl-tRNA synthetase
IVI116	155	cryptic	1601101 to 1601255	*spyM18_1926*	hypothetical protein
IVI118a	724	antisense	6604 to 5881	*spyM18_0007*	transcription-repair coupling factor
IVI118b	252	antisense	1407068 to 1407319	*spyM18_1692*	trans-acting positive regulator
IVI119a	198	antisense	1391644 to 1391841	*spyM18_1670*	ATP-dependent RNA helicase
IVI119b	183	cryptic	1725606 to 1725788	*spyM18_2064*	mitogenic exotoxin Z
IVI125	103	antisense	1017508 to 1017610	*spyM18_1213*	hypothetical protein
IVI129	170	cryptic	1220056 to 1219887	*spyM18_1463*	hypothetical protein
IVI132	251	antisense	416432 to 416182	*spyM18_0513*	phosphate starvation-induced protein
IVI135	288	cryptic	1073881 to 1073594	*spyM18_1286*	hypothetical protein
IVI138a	468	cryptic	1067947 to 1067480	*spyM18_1276*	phage protein
IVI138b	295	antisense	1027614 to 1027908	*spyM18_1224*	amino acid ABC transporter ATP-binding protein
IVI138c	103	antisense	441923 to 441821	*spyM18_0556*	dephospho-CoA kinase
IVI138d	84	cryptic	1721939 to 1722022	*spyM18_2057*	anthranilate synthase component II
IVI140	327	antisense	968624 to 968950	*spyM18_1160*	pyridoxamine kinase
IVI142	184	intergenic	562668 to 562851	*spyM18_0703*	2 bp at 5' side: conserved hypothetical protein
				spyM18_0704	510 bp at 3' side: conserved hypothetical protein
IVI143	168	cryptic	233754 to 233921	*spyM18_0261*	glyceraldehyde-3-phosphate dehydrogenase, plasmin receptor
IVI146	77	antisense	1500543 to 1500619	*spyM18_1815*	putative acetyl-CoA carboxylase alpha subunit
IVI148	1016	antisense	1760729 to 1761744	*spyM18_2110*	putative pyruvate formate-lyase 2
IVI152	1029	antisense	446733 to 445706	*spyM18_0562*	putative exoribonuclease R
IVI154	185	cryptic	1196279 to 1196095	*spyM18_1429*	putative D-Ala-D-Ala adding enzyme
IVI156	985	promoter	433733 to 434717	*spyM18_0544*	putative BlpM homologue
IVI157a	148	antisense	107943 to 107796	*spyM18_0111*	acetate kinase
IVI157b	105	cryptic	1571714 to 1571818	*spyM18_1893*	putative divalent cation transport
IVI157c	65	antisense	1893419 to 1893355	*spyM18_2256*	putative serine protease
IVI157d	35	cryptic	14043 to 14077	*spyM18_0014*	putative cell division protein

IVI159	63	antisense	1184631 to 1184693	*spyM18_1416*	putative competence protein
IVI163	88	antisense	332105 to 332192	*spyM18_0393*	exotoxin type A precursor
IVI164	245	cryptic	367319 to 367563	*spyM18_0446*	putative ATP-dependent protease proteolytic subunit
IVI165	84	cryptic	13964 to 14047	*spyM18_0014*	putative cell division protein
IVI167a	327	cryptic	1220838 to 1220512	*spyM18_1464*	conserved hypothetical phage protein
IVI167b	133	antisense	1455118 to 1455250	*spyM18_1750*	hypothetical phage protein
IVI167c	126	antisense	616307 to 616182	*spyM18_0777*	hypothetical phage protein
IVI173	91	cryptic	484019 to 484109	*spyM18_0600*	putative shikimate 5-dehydrogenase
IVI176	42	promoter	344746 to 344787	*spyM18_0414*	hypothetical protein
IVI177	113	cryptic	1530798 to 1530686	*spyM18_1856*	ABC transporter ATP-binding protein
IVI178	50	antisense	1207032 to 1206983	*spyM18_1446*	MF3
IVI181a	773	antisense	983703 to 982931	*spyM18_1174*	phosphopantothenoylcysteine decarboxylase
IVI181b	186	antisense	677648 to 677463	*spyM18_0836*	ATP-dependent exonuclease subunit A
IVI185	309	antisense	1546690 to 1546998	*spyM18_1868*	hypothetical protein

CURRICULUM VITAE

Brent Armstrong, B.Sc. (Hons)

EDUCATION

2009- present The University of Western Ontario, London, On, Canada

             Ph.D. candidate in Microbiology and Immunology

2007-2009 The University of Western Ontario, London, On, Canada

             M.Sc. candidate in Microbiology and Immunology (Transferred to Ph.D. program)

2003-2007 University of Guelph, Guelph, On, Canada

             B.Sc. (Hons.) in Molecular Biology and Genetics

1998-2003 Resurrection Catholic Secondary School, Waterloo, On, Canada

HONOURS AND AWARDS

2008-2013 Western Graduate Research Scholarship ($2235 per term)

2007-2008 Western Graduate Research Scholarship ($1367 per term)

2006-2009 Dean's Honour List

2003-2004 University of Guelph Entrance Scholarship ($1000)

1998-2003 Honour Roll

RESEARCH EXPERIENCE

2007-present Doctorate Thesis Project, The University of Western Ontario.

             Thesis title: "The Identification of *in vivo* Induced Genes in *Streptococcus pyogenes*".

             Research Advisor: Dr. John K. McCormick.

2006-2007   Honours Thesis Project, University of Guelph.

Thesis title: "Molecular Analysis of the *apxIBD* Genes of *Actinobacillus suis*".

Research Advisor: Dr. Janet I. MacInnes

TEACHING EXPERIENCE

2009   Training and direct supervision of an Honour's Project student

2009   Training and direct supervision of an NSERC summer student

2008-2013   Training and direct supervision of all work study students.

2008-2011   Teaching Assistant for Biology of Prokaryotes (2100a) at the University of Western Ontario.

PRESENTATIONS

Armstrong, B.D., Heinrichs, D.H. and McCormick, J.K. "Identification of *in vivo* induced genes in *Streptococcus pyogenes*". Infection and Immunity Research Forum. London, Ontario, Canada (2012).

Armstrong, B.D. and McCormick, J.K. "*Counter-selection in Streptococcus pyogenes*". Banff Conference on Infectious Diseases. Banff, Alberta, Canada (2010).

Armstrong, B.D. and McCormick, J.K. "*Counter-selection in Streptococcus pyogenes*". Canadian Society of Microbiologist's annual meeting poster presentation. Calgary, Alberta, Canada (2008).

Armstrong, B.D. and MacInnes, J.I. "*Molecular Analysis of the apxIBD Genes of Actinobacillus suis*". Honours student poster presentation. University of Guelph, Guelph, Ontario, Canada (2007).