Characterization of an unusual phage repressor encoded by mycobacteriophage BPs

by

Valerie Marie Villanueva

B.S. Biology, College of William and Mary, 2009

Submitted to the Graduate Faculty of

School of Medicine in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

This dissertation was presented

by

Valerie Marie Villanueva

It was defended on

July 30, 2015

and approved by

Elodie Ghedin, Professor, Department of Biology, New York University

Roger W. Hendrix, Distinguished Professor, Department of Biological Sciences

Saleem A. Khan, Professor, Department of Microbiology and Molecular Genetics

Andrew P. VanDemark, Associate Professor, Department of Biological Sciences

Dissertation Advisor: Graham F. Hatfull, Eberly Family Professor of Biotechnology, HHMI Professor, Department of Biological Sciences

CHARACTERIZATION OF AN UNUSUAL REPRESSOR ENCODED BY MYCOBACTERIOPHAGE BPS

Valerie Marie Villanueva

University of Pittsburgh, 2015

Temperate phages are capable of both lytic and lysogenic growth. In lytic growth, phages utilize the host machinery to produce new virions and kill their host; during lysogeny, the phage genome integrates into the host chromosome and produces a repressor that prevents expression of lytic genes, thus maintaining the lysogenic state. A few phage genetic switches governing the lytic/lysogenic decision have been well characterized, most of which contain analogous regulatory proteins with similar functions. A new type of genetic switch, which uses phage integration directly in synthesizing the phage repressor, is described. This dependence on integration arises because of the unusual feature that the *attP* core, the site where strand exchange occurs during integration, is located within the repressor gene; in this study we are using the mycobacteriophage BPs as a model system. The BPs repressor gene, 33, encodes a 136-residue protein ($gp33^{136}$) that we have shown is not active for immunity unless overexpressed. However, when the phage genome integrates into the bacterial chromosome, a stop codon is introduced at the *attL* junction site, creating a truncated, but active form of the repressor ($gp33^{103}$), which is expressed within the lysogen. Electrophoretic mobility shift assays (EMSA) have shown that each form of the repressor binds specifically to two operators present within the *33-34* intergenic regulatory region, as well as to four other operators located across the BPs genome. Here we report studies done to map the sequence requirements for repressor

function in repressing lytic growth and show that a previously reported 12 bp operator sequence represents one half of a full binding site. We also show that the BPs repressor does not seem to form inter-DNA bridges for transcriptional regulation, but evidence suggests that bending DNA plays a role. Characterization of the binding sites and behavior of the BPs repressor will deepen our understanding of this new class of integrations-dependent switch. The overarching goal, of which this project represents only a portion, is to dissect BPs so that we have a full understanding of its biology, what its genes do, how they are expressed, how expression is regulated and manipulated to confer immunity, and how it recognizes its mycobacterial hosts.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

The completion of this study would not have been possible without the contribution and support of numerous individuals. I would like to thank my amazing mentor, Dr. Graham Hatfull, for his guidance, support, and patience throughout my development as an independent scientist. I would also like to acknowledge the entire Hatfull lab for helpful discussions, and especially the "BPs team", Gregory Broussard, Lauren Oldfield, Bryce Lunt, and Emilee Shine, who were instrumental in contributing to our initial understanding of the new class of integration-dependent genetic switches and provided useful advice throughout my journey. Some of the work presented in this dissertation (Figures 5-7) was published in our Broussard et. al, 2013 Molecular Cell article. Most of the rest of the work described was recently published in my Villanueva, et al, 2015 PLoS One article. Although I didn't cite this paper throughout the dissertation, at least a portion of all figures shown (with the exception of the RNA polymerase assays and the work done in the appendices, which are not published) were included in the publication. I'd like to acknowledge the contribution of Lauren Oldfield in the Hatfull lab, who performed the experiments in panels A & B of Figure 37. I would also like to acknowledge the invaluable contributions of Dr. Andrew VanDemark (University of Pittsburgh), especially for his expert technical advice during my early work in purifying the BPs repressor, as well as for the contribution of equipment, reagents, and the PBEND2 vector. I would like to thank Dr. Richard Burgess (University of Wisconsin – Madison) for the kind gift of antibodies recognizing various

RNA polymerase subunits, as well as Nancy Thompson in his laboratory who provided helpful discussions and technical advice. I would like to acknowledge the Andrea Berman lab (University of Pittsburgh), especially Roni Lahr for the use of their FPLC and invaluable technical advice. I would also like to express my extreme gratitude to all of my thesis committee members who took the time to share their insight and expert advice.

Finally, I would like to acknowledge the most crucial players in my journey. This would not have been possible without the guidance and support from my wonderful parents Amilcar and Carmen Villanueva, and my brother, Alex Villanueva. I would like to express gratitude to my good friends Sarah Bidula, Benedict Hilldorfer, Lauren Oldfield, and Zaritza Petrova who helped me keep my sanity by providing necessary distractions, and especially to my husband, Scott Whipkey, for his unconditional love and support.

1.0 INTRODUCTION

Bacteriophages can be manipulated and used in a variety of ways related to molecular biology, for example as tools for genetic study. Mycobacteriophages, the phages of mycobacteria, have been instrumental in the development of tools for the manipulation of the *Mycobacterium tuberculosis* genome (1-18). Mycobacteriophages have already been used to develop a mycobacterial site-specific recombination system (15-18), transposon and reporter gene delivery systems (1, 2, 19), selectable markers (3, 20), stable integration-proficient vectors (6-8) and for the creation of recombinant bacterial vaccines (4). With over 850 sequenced mycobacteriophage genomes (www.phagesdb.org) showing a remarkable variety of novel genes, there is significant potential for discovering more genes that will have an impact on the study of *M. tuberculosis* pathogenesis (5, 21, 22).

In order to harness the power of mycobacteriophage genetics, it is important to understand how these viruses control gene expression. A classic example is when a temperate phage makes the decision to enter either lytic or lysogenic growth, a process governed by a genetic switch, which activates or represses relevant genes (12, 23). By studying and understanding these pathways we will get, not only a deeper understanding of how phages function and interact with their hosts, but also insights into how biological switches are regulated by different genetic components.

1.1 WHY STUDY MYCOBACTERIOPHAGES?

Mycobacterium tuberculosis is a major human pathogen affecting one in every three people across the world (24-26). It causes the disease tuberculosis, which, if left untreated, constitutes a potentially deadly respiratory infection. Treatment for tuberculosis is costly and time-consuming (27). When a patient is diagnosed with tuberculosis, they are treated using a combination of drugs (28). Since most of the infection is cleared away and symptoms of the disease disappear quickly after drug treatment begins, patients occasionally cease to continue taking their drugs, allowing the remaining *M. tuberculosis* to proliferate, providing an opportunity for multi (MDR) and extremely (XDR) drug-resistant strains to emerge (29, 30). Drug-resistant strains are a major cause of death in citizens of developing countries across the world (29, 31). They are especially disastrous in patients with Acquired Immunodeficiency Syndrome (AIDS), because immunocompromised patients are unable to fight off infection or keep the infection at bay, thus making the problem of drug-resistance larger (28, 32).

With the emergence of drug-resistant and multi-drug resistant strains of bacteria, the future of antibiotics has come into question and the therapeutic potential of bacteriophages, viruses that infect bacteria, is being explored (33). Phages can be manipulated and used in a variety of ways, including as tools for genetic study and as new technology for treatment of bacterial infections. Already, the development of bacteriophages as cloning vectors and transporters of recombinant DNA has revolutionized the field of molecular microbiology. The clinical prospects of phages are also numerous and these viruses are already being used in some parts of the world as part of primary treatment methods for chronic and drug resistant bacterial infections, as well as in disease prevention. (33-37) Phages also encode proteins with antimicrobial properties, such as lysins, lysterases, and toxin-encoding genes that are only just

beginning to be understood and used as therapeutic agents (38-40). Additionally, phage capsids are being used to display epitopes that help immunize against certain cancers (41-44).

Since mycobacteriophages infect mycobacteria such as *M. tuberculosis*, the expanding database of fully sequenced phage genomes provide a great body of data to explore for therapeutic potential against the etiological agent of tuberculosis.

1.2 GENETIC SWITCHES CONTROLLING TEMPERATE PHAGE LIFECYCLES

A bacterium infected by a temperate bacteriophage will have one of two fates: death due to lytic growth or latent infection during lysogeny (12, 45). If environmental conditions are optimal for bacterial growth, the infecting bacteriophage will replicate its genome and assemble new virions, then lyse the bacterium to release infectious progeny phage. If environmental conditions are not optimal, or if there is a high concentration of infecting phages, an infecting bacteriophage will integrate its genome into the bacterial chromosome, becoming a prophage. The prophage expresses a repressor protein to shut off expression of lytic genes and keep its genome integrated. The presence of repressor also prevents lytic infection from other phages that use similar mechanisms of genetic regulation, thereby conferring immunity to the prophage-containing bacterium, or lysogen. (46, 47) In certain environmental conditions, such as those that cause DNA damage, the SOS pathway leads to induction of lytic growth, prophage excision, and release of infectious virions into the extracellular environment. This process is known as lysogenic induction (45, 48).

The regulatory elements governing the initiation and stability of these phage life cycles make up a bistable genetic switch (23). The most important attribute of a genetic switch is that

once activated, the subsequent series of events will commit the organism to that specific pathway. In the context of this project, the phage is dedicated to either lytic or lysogenic growth.

Phages are the most abundant genetic entities on the planet (49) and due to their genetic diversity and prevalence of the temperate phage life cycle being high (50), phages likely have the largest reservoir of genetic switches. The central components that allow commitment to each state (51, 52) are likely common in all these switches, but differences arise in how the phages flip the switch and maintain it in each state.

A few phage genetic switches have been characterized thus far, the most well-studied being that of bacteriophage lambda (45, 47). Most of these systems contain analogous regulatory proteins with similar functions. Only a few genetic switches have been characterized in mycobacteriophages.

1.2.1 Bacteriophage Lambda

The lambda genetic switch is the most well studied phage switch. The genetic switch is complex and involves many components located throughout the lambda genome but has one goal, the decision and commitment to lytic or lysogenic growth (47).

To begin infection, bacteriophage lambda absorbs to an *Escherichia coli* cell and injects its genome into the cell. Host RNA polymerase initiates expression from two promoters, $P_{R,}$ which transcribes rightwards genes, and P_L for leftwards expression. Cro, produced by the P_R transcript, binds to the P_{RM} promoter and prevents the expression of the repressor, *cI* (53-55). The *N* gene is transcribed from P_L and the N protein is an anti-termination factor that blocks transcription termination at the terminator (t_R1 and t_R2) sites and extends the operons expressed from both P_L and P_R (45). The anti-terminated P_L transcript expresses a number of genes

including two that participate in the genetic switch, *cIII* and *int* (56). From promoter P_R, the genes *cII* and *Q* are transcribed when termination at the t_R1 site in *cro* is prevented. Q is also an anti-termination factor and the $P_{R'}$ promoter region contains a *qut* (Q utilization) site (57). When Q is present, $P_{R'}$ transcribes the genes for lysis of the host cell and the genes for structural proteins that are used to assemble progeny phage. If these conditions persist, lytic growth of the bacteriophage will result. Continued synthesis of Cro, an essential lytic regulator, represses the P_{RM} promoter by its particular binding pattern at its operators and prevents the transcription of *cI* (47).

In order for lysogeny to occur, the CII protein must accumulate and activate the P_{RE} promoter, which will transcribe the repressor gene, *cI*. Under conditions that favor lytic growth, the *cII* gene is transcribed from the N-anti-terminated P_R transcript but the CII protein is susceptible to host protease, FtsH, and is rapidly degraded (58-61). Without CII, the repressor, *cI*, is not expressed as the P_{RE} promoter is not activated and the P_{RM} promoter is repressed by Cro. However, during conditions that favor lysogeny, concentrations of FtsH and other proteases are low and CII is able to escape degradation. This is also due to the actions of CIII, which binds and sequesters FtsH, and therefore reduces the amount of active FtsH able to degrade CII (62, 63). CII is able to act at several promoters and flip the switch towards lysogeny. CII binds to the P_I promoter and increases expression of the lambda Integrase, which integrates the phage genome into the host chromosome at the *attP* site, creating a prophage (47). Also, it regulates the P_{antiq} promoter. This leftwards promoter transcribes an antisense transcript of the *Q* anti-terminator gene, which is necessary for transcription of lytic genes, and switches off production of Q (56). Finally, CII activates the P_{RE} promoter, which transcribes the *cI* gene without the interference of Cro, which is blocking transcription from P_{RM} (45). As CI protein begins to

accumulate, it competes with Cro for operator sites at P_{RM} (64). When CI displaces Cro, it activates its own expression from the leftwards P_{RM} promoter and represses transcription from the P_R promoter that transcribes *cro* and other genes required for lytic growth. This commits the phage to exist as a prophage in a lysogen. If Cro remains bound to its operators, then the production of lytic genes continues and *cI* is not produced, committing the bacteriophage to lytic growth.

In lambda, there are two tripartite operator sites to which CI binds with increasing affinity during lysogeny (65-68). The repressor has two domains and forms dimers via interactions in its C-terminal domain (68). These dimers bind with high affinity to the operator O_R1 through their N-terminal domains, which contain helix-turn- helix DNA binding motifs (64, 67, 69-72). Almost immediately, a second repressor dimer complex interacts with the C-terminal domains of the bound dimer, allowing it to cooperatively bind to O_R2 and form a tetramer (12, 69). By binding to O_R1, the repressor physically blocks RNA polymerase from binding the Cro promoter, which prevents transcription of a cascade of early lytic genes. As a result of binding to O_R2, the repressor physically interacts with domain 4 of the alpha subunit of RNA polymerase (73, 74) and stabilizes the enzyme, which promotes binding to the c*I* promoter with greater affinity, activating transcription of more repressor and creating a positive feedback loop (75).

When repressor concentrations increase enough for a repressor dimer to bind to O_R3, which is located within P_{RM}, it blocks RNA polymerase transcribing *cI* and repressor levels decrease. Binding to O_R3 occurs at a much lower affinity due to lack of cooperative interaction with other dimers (76, 77). As repressors levels fall, dimers no longer bind to O_R3, allowing renewed transcription of *cI*. This difference in affinity allows the repressor to autoregulate its expression during lysogeny and respond to fluctuations in the cellular environment without being

hypersensitive. A threshold concentration of repressor must be achieved before the switch can be flipped (45).

Cooperativity also occurs between repressor proteins bound to three operator sites approximately 2kb away on the opposite end of the *cI* gene, O_L1, O_L2, and O_L3 (69, 78-80). The repressor binds to these three sites with similar affinities as it does to O_R1, O_R2, and O_R3, respectively. Repressor tetramers bound to O_R1 and O_R2 bind cooperatively through its C-terminus with tetramers bound at O_L1 and O_L2 to form an octomer that loops the DNA containing the *cI* gene, increasing its ability to autoregulate its expression by isolating it from the rest of the genome (78). When dimers bind O_R3 and O_L3, they tetramerize and block *cI* expression until the repressor concentration decreases enough for the tetramer to dissociate and transcription to resume.

During lysogen induction, when host DNA is damaged, the SOS response is activated and expression of RecA protease is upregulated (12, 64). This protease cleaves multiple host repressor proteins and also stimulates self-cleavage of the lambda repressor at the linker region between its N-terminal DNA-binding domain and its C-terminal domain (81). When the lambda repressor is cleaved, it can no longer bind cooperatively to the operator sites and when enough proteins are degraded, transcription of Cro begins, allowing the phage to undergo lytic growth.

When Cro proteins are expressed, they form dimers and bind with a greater affinity to O_R3, which creates a physical block against *cI* transcription (51, 52, 82). Unlike the lambda repressors, Cro cannot bind to O_R2 cooperatively and it cannot activate RNA polymerase. Therefore, when Cro levels rise and dimers bind O_R2, *cro* transcription stops until the Cro protein levels fall enough for O_R2 to become unoccupied again. *Cro* is the first gene in an operon of genes needed for prophage excision and DNA replication.

1.2.2 Mycobacteriophage L5

Mycobacteriophage L5 is a temperate phage that uses a single repressor to regulate transcription during the establishment and maintenance of lysogeny (3, 83-85). This repressor, gp71, has two domains separated by a short linker region, similar to the lambda repressor. The N-terminal domain contains a strongly predicted helix-turn-helix DNA-binding motif and the repressor binds as a monomer to a highly conserved, 13bp asymmetric binding motif on double-stranded DNA (not ssDNA or RNA as determined by studies done with the repressor of mycobacteriophage L1, which is homoimmune to L5), even though the protein is capable of dimerization (86-88). Gp71 can bind to operator sites within promoter regions to block transcription initiation (83), such as within the P_{left} promoter responsible for transcribing right-arm genes that are involved in replication and initiating the lytic lifecycle (89-91), as seen in other phage regulatory circuits.

Interestingly, L5 contains approximately 30 copies of the 13bp binding site across the length of its genome and studies have shown that when gp71 binds to sites downstream of the promoter, it is capable of blocking transcription elongation (83). This activity is dependent on the sequence of the binding site being oriented in the direction of transcription, and when multiple sites are clustered together, the blockage of transcription elongation is cumulative. (92-94)

This “stoperator” method of ensuring immunity is shared among all the Cluster A phages (over 200 phages grouped according to sequence similarity (22)). While more work needs to be done to understand the L5 genetic switch, including identifying other players involved, how the switch is flipped, and how lysogeny is maintained, the identification of this repressor, which is capable of exerting an effect on two stages of transcription (initiation and elongation) shows an

example of the types of regulatory elements that are waiting to be discovered in mycobacteriophages.

1.2.3 Mycobacteriophage BPs and Integration-Dependent Immunity

Mycobacteriophage BPs is a temperate phage that has the ability to expand its host range to infect *M. tuberculosis* (95, 96). It has one of the smallest phage genomes sequenced to date (96). This phage is unusual because the core of the phage attachment site (*attP*), where site-specific recombination between the BPs genome and the bacterial chromosome occurs during integration, is located within the open reading frame (ORF) of the repressor gene (*gene 33*) (97). As a consequence, the prophage form of the repressor ($gp33^{103}$) is truncated by 33 amino acids compared to the virally encoded form ($gp33^{136}$).

This unusual organization is seen in over 10 mycobacteriophages (six from Cluster G (96, 98, 99), three from Subcluster I1 (97), two from Cluster N (98), and one from Cluster P (98)) as well as the prophages of other species (97).

1.2.3.1 Genetic Switch

In mycobacteriophage BPs, upon infection of a host cell, two divergent promoters located within the gene *33-34* intergenic region of the genome, P_R and P_{rep}, are active, initiating synthesis of *gene 34*, a putative cro-like protein and the first gene of an early lytic operon, as well as *gene 33*, the repressor gene and *gene 32*, the integrase gene, respectively (**Figure 1**)(97). P_{rep} is twice as active as P_R, indicating that there is a fragile balance in expression between these promoters that is important in determining the outcome of infection. In this genetic switch, the integrase (gp32) determines the outcome of infection. When cellular proteases are low, or when phages infect at a

high multiplicity of infection (moi), then gp32 is active, integration occurs, and lysogeny is initiated (97). Since P_{rep} is twice as active as P_R, gp32 is initially expressed at a higher rate than the early lytic genes, therefore P_{rep} is responsible for lysogenic establishment.

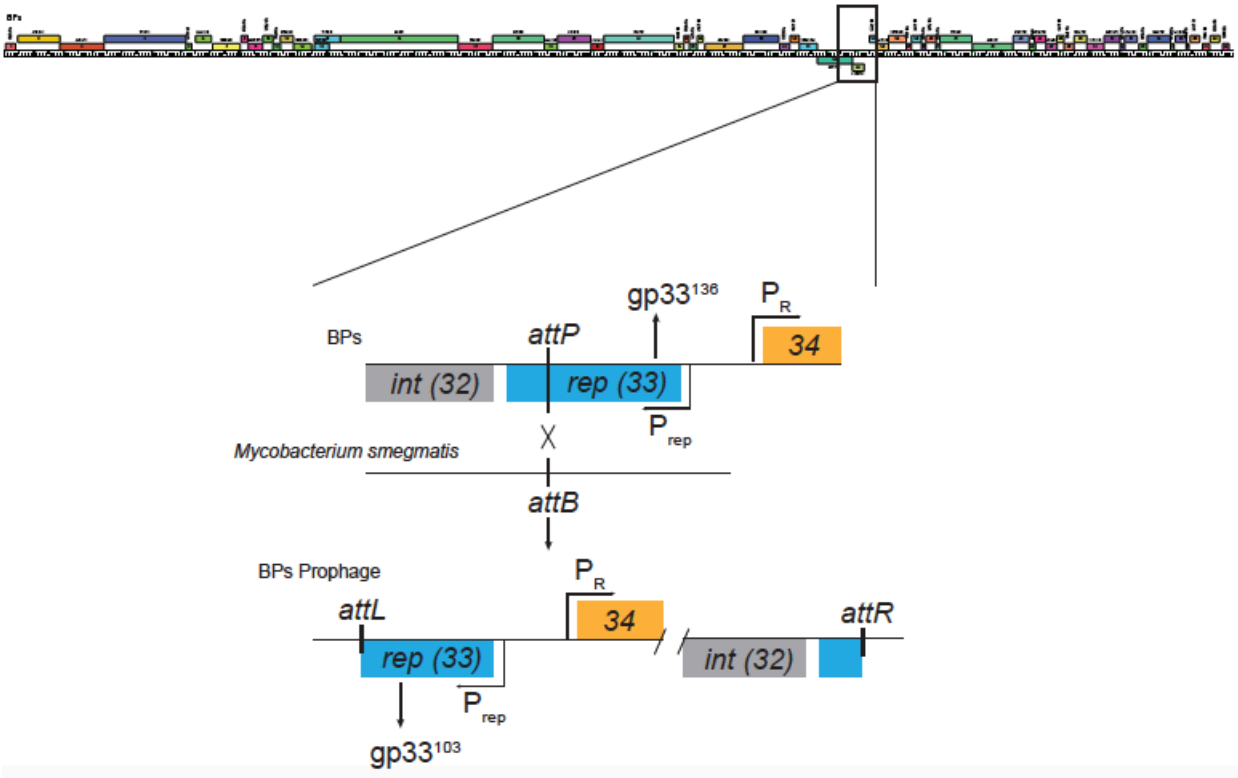

Figure 1 - BPs genome map with immunity cassette highlighted

The mycobacteriophage BPs genome is shown with a ruler representing the length of the genome and colored boxes indicating open reading frames. Boxes above the ruler are transcribed rightward, boxes below the ruler are transcribed leftward. The immunity cassette is zoomed in. There are three genes involved in conferring immunity: gene *32* (integrase), gene *33* (repressor), and gene *34* (putative cro-like protein). The *attP* core is located within the repressor open reading frame. Upon integration the 3' end of the gene is removed, which leads to the expression of a shorter, stable form of the repressor ($gp33^{103}$).

Integration leads to the initiation of lysogeny because the site at which site-specific recombination occurs between the phage genome and host chromosome during genome integration (*attP* core) is located within the repressor open reading frame, such that upon

integration the 3' end is removed from the rest of *gene 33* (**Figure 1**) (96, 97). This leads to a truncated form of the repressor being expressed in the prophage ($gp33^{103}$). This shortened repressor gene is still attached to the constitutive Prep promoter and is expressed throughout lysogeny, therefore P_{rep} is also involved in lysogenic maintenance. The 33 amino acids that are removed from the C-terminus of the protein by this integration process encode an ssrA-like tag for proteolytic degradation, possibly by the ClpXP protease (97, 100, 101). So although the viral form of *gene 33* ($gp33^{136}$) is expressed upon infection, it is rapidly degraded and cannot accumulate to exert an effect. Therefore, integration is the process responsible for flipping the switch. Removing the tag and allowing the active form of the repressor to be expressed and exert its effect on transcription assures lysogeny.

Once enough gp32 is made to allow integration, recombination occurs and *gene 32* is separated from the P_{rep} promoter and expression of integrase ceases (97). The remaining gp32 is degraded, and there is insufficient integrase left to catalyze the excision reaction. If gp32 is not present in sufficient quantities to allow integration (due to low moi or high host protease conditions), the full-length repressor continues to be degraded and lytic proteins accumulate to block transcription of genes *33* and *32*. It is unknown how gp32 accumulates to catalyze the excision reaction during induction of lytic growth from the lysogenic state, since it has no requirement for a recombination directionality factor (RDF) and it is separated from its native promoter by integration (97). We theorize that it is expressed either from a cryptic promoter within the *gene 32-33* intergenic region, which is upstream of *gene 32* after integration, or that it is expressed from a nearby bacterial host promoter whenever conditions are right for phage induction.

This is the first instance of integration-dependent immunity described and it provides yet another example of the rich diversity of genetic regulatory mechanisms found in mycobacteriophages.

1.2.3.2 Gene *33-34* Intergenic Regulatory Region

The *33-34* intergenic region contains two divergent promoters, one for rightward expression of an early lytic operon (P_R) (97, 102) and one for leftward expression of *genes 33* and *32*, encoding repressor and integrase proteins respectively (P_{rep})(97) (**Figure 2**). Both of these promoters have been experimentally determined using RNA seq, promoter trap, and fluorescence reporter assays (102, 103). Fluorescence reporter assays with each promoter fused to a *mCherry* reporter gene have also shown that P_{rep} has twice as much activity as P_R, and that in a lysogen the P_R promoter is tightly downregulated, whereas the P_{rep} promoter is not. In the presence of the active form of repressor, P_{rep} is actually modestly upregulated (97, 104).

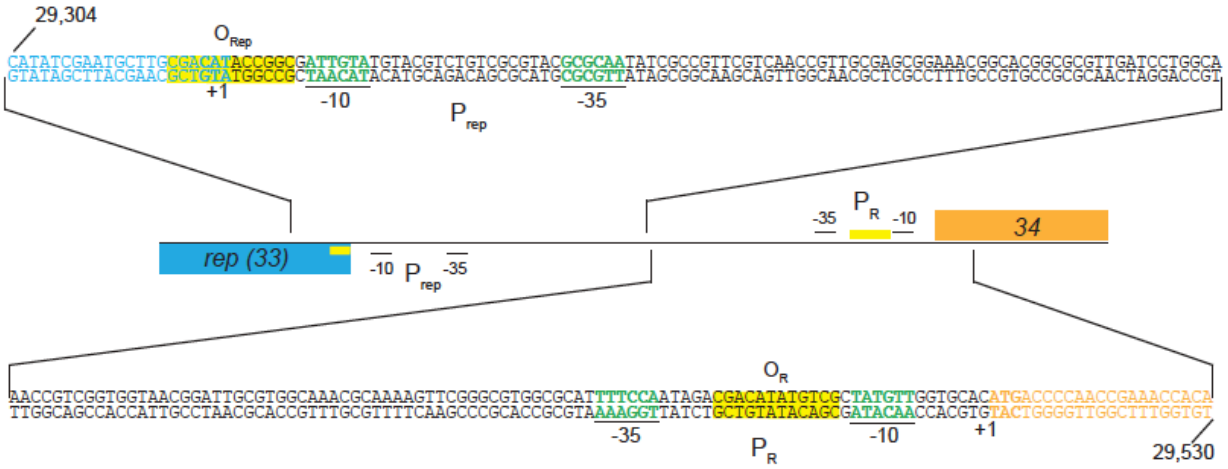

Figure 2 - The BPs *33-34* intergenic region

The sequence of the BPs *33-34* intergenic regulatory region is shown. The -10 and -35 sequences of the P_R and P_{rep} promoters are shown in green text with confirmed transcription start sites indicated with a "+1". The original locations of O_R and O_{Rep} are highlighted in yellow. Some sequence into the gene *33* (blue) and gene *34* (orange) open reading frames is also included.

The transcription start sites of these promoters have been mapped (97, 105) and the promoters are confirmed to be leaderless (which seems to be common with mycobacterial promoters – (105)). **Figure 2** shows the sequence of the *33-34* intergenic region. The P_R and P_{rep} promoters are shown along with their transcription start sites and the first few bases of the open reading frames of *genes 33* and *34*. The divergent nature of these promoters is reminiscent of the lambda phage system where P_R and P_{RM} are located, and suggests that the BPs repressor may exert a function in this area.

1.3 PROKARYOTIC TRANSCRIPTION

Bacteriophage transcription works similarly to bacterial transcription because in most cases, the infecting phage is utilizing host transcription machinery to express its own genes. This section offers a brief overview of prokaryotic transcription, to create a foundation for examining how repressor proteins affect transcription in later chapters.

1.3.1 Bacterial RNA polymerase

The crystal structures of many bacterial DNA-dependent RNA polymerases have been resolved, however the structure of mycobacterial RNA polymerases have only been modeled from structures of *Escherichia coli* and *Thermus aquaticus* RNA polymerase (106-110). These structures show that the core enzyme has four subunits, which are evolutionarily conserved in all bacterial species (111): two alpha subunits, a beta subunit, and a beta prime subunit. This core

enzyme has the ability to bind and travel along DNA (112), but requires an additional subunit, the sigma factor, in order to specifically bind to a promoter sequence and initiate transcription (113, 114).

1.3.2 Sigma factors

Sigma factors are the transcription initiation factors of RNA polymerase that specifically recognize promoter sequences. While some sigma factors are conserved between species (Sigma 70 family - (115)) most are unique to different bacterial species (116-118). Different sigma factors are used in different environmental conditions, for example *E. coli* sigma 32 is expressed whenever the bacteria is in a high heat environment and this sigma factor allows expression of the *E. coli* heat shock proteins (119). Depending on their gene expression needs, some bacteria have more or fewer sigma factors than others. *E. coli* has seven recognized sigma factors (120) whereas *M. smegmatis* and *M. tuberculosis* have 28 and 13 respectively. (118, 121-125).

When sigma subunits bind to an RNA polymerase core enzyme, they allow the polymerases to bind with high affinity to particular DNA sequences, called promoters. Some promoters are easy to recognize, such as the sigma 70-family of promoters (126). These promoters have consensus -10 and -35 motifs, which are appropriately named for their locations 10bp and 35bp upstream of the transcription start sites (127). The promoters of other sigma factors in other species do not always have such readily identifiable sequences in their promoters, making their bioinformatic prediction more difficult (128).

1.3.3 Initiation, elongation, and termination

There are three stages of transcription in prokaryotes: initiation, elongation, and termination. Although elongation and termination are important determinants of gene expression (129), transcription initiation, starting with binding of the RNA polymerase to a specific promoter sequence, is generally considered the most influential for gene expression.

Transcription initiates when an RNA polymerase holoenzyme interacts with a specific promoter sequence (130). The most well understood promoters are the ones recognized by the housekeeping sigma 70-like sigma factors (126). These promoters have a well-characterized -10 and -35 element separated by a ~17bp spacer region. The sigma factor recognizes the -35 element and binds to it. Then the RNA polymerase is able to switch from its unbound closed complex to the open complex where it melts the AT-rich DNA at the -10 site and unwinds ~13bp of DNA to create the transcription bubble (131). At this point the RNA polymerase begins incorporating ribonucleotides and synthesizing an 8-9bp RNA strand. It undergoes a few rounds of abortive initiation until the polymerase synthesizes a long enough RNA strand to release its interactions with the promoter, and then elongation begins (132).

Once the RNA polymerase complex escapes the promoter it begins to move along the DNA strand creating an mRNA transcript in the 5' to 3' direction. This process is known as elongation of the transcript (132). Many bacterial promoters control gene operons, meaning that multiple genes are transcribed in the same transcript, which are then differentially translated by ribosomes (133). RNA polymerase continues elongating the transcript until it encounters a terminator.

When the RNA polymerase elongation complex reaches a terminator, elongation ceases and the nascent RNA strand is released. Prokaryotic terminators come in different forms. A rho-

dependent terminator involves the rho protein actively dislodging RNA polymerase (134). A rho-independent, or intrinsic terminator is typically a DNA structure composed of a T-tract that pauses the RNA polymerase when it becomes a U-tract on the RNA strand, and a hairpin loop, that dislodges the RNA polymerase from the DNA (135, 136).

1.4 PROTEIN-DNA INTERACTIONS

The bacteriophage repressor discussed in the following chapters is a DNA-binding protein. To understand the nature of the interactions this protein has with its DNA-binding sites it will be helpful to have an understanding of how DNA-binding proteins interact with their substrates. This section provides a brief introduction into the structural relationships between DNA binding proteins and their substrates with a focus on the specific nature of their interactions.

1.4.1 What structures define a DNA-binding protein?

DNA binding proteins contain a large range of different structural motifs that lead to their ability to bind DNA. Some DNA binding proteins, such as lambda cI, use alpha helices to interact with the major and minor grooves of DNA (137-142). These include, but are not limited to, Helix-turn-helix motifs (143, 144), helix loop helix motifs, and leucine zipper motifs (145-148). Some proteins, for example TATA box binding proteins, use beta sheets to interact with the minor groove of DNA (149, 150). And some proteins use a combination of both alpha helices and beta sheets, such as those with zinc finger domains (151).

1.4.2 DNA structure affects protein binding

DNA is sometimes a double helix and its structure varies slightly depending on its sequence (152, 153). DNA binding proteins use these variations in structure to recognize specific DNA sequences (154). AT-rich and GC-rich sequences typically have a helix structure of B-DNA, however the electrostatic potential in the minor grooves of DNA are different (155-159). A-tracts (a series of A-T bases) lead to a bend in the DNA due to a slight roll, which leads to expanded grooves on one face of the helix and compacted grooves on the other (160-163). The different structural environments formed by these types of sequences provide optimal binding for different types of DNA-binding proteins.

1.4.3 How proteins recognize DNA substrates

One of the most obvious ways that proteins recognize DNA substrates is through specific interactions between residues in their binding domains and the specific chemistries of base pairs in the major or minor groove of DNA (137, 164, 165). These interactions can be very specific (formed by hydrogen bonds) or less specific (formed by hydrophobic contacts) (164-167). Hydrophobic interactions can also be used by proteins to recognize single-stranded DNA. Another way proteins can specifically recognize DNA is by the structure of their major and minor grooves (152, 154, 168). When major and minor grooves become narrower or wider based on changes in DNA structure, they become amenable to being bound by different proteins. Whether by sequence-specific, or structure-specific interactions, there are a lot of different factors that govern specificity of DNA binding proteins to their substrates and most use a combination of different recognition factors.

2.0 CHARACTERIZING THE DNA-BINDING CAPABILITIES OF GP33^{103} AND GP33^{136}

Work done in this section was published in the following journal articles:

Broussard, G.W., Oldfield, L.M., **Villanueva, V.M.**, Lunt, B.L., Shine, E.E., Hatfull, G.F., January 2013. Integration-Dependent Bacteriophage Immunity Provides Insights into the Evolution of Genetic Switches. *Molecular Cell*, Volume 49, Issue 2, Pages 237-248.

Villanueva, V.M., Hatfull, G.F., September 2015. An Unusual Phage Repressor Encoded by Mycobacteriophage BPs. *PLoS One*, Volume 10:e0137187.

2.1 INTRODUCTION

It has been shown that wild type BPs cannot infect a *M. smegmatis* BPs lysogen because a BPs lysogen confers homoimmunity to superinfection. *M. smegmatis* strains have been engineered to stably express both gp33^{103} and gp33^{136} at similar levels and wild type BPs can lytically infect a strain expressing gp33^{136}, but not a strain expressing gp33^{103}, indicating that gp33^{103} is the active form of the repressor (97). Therefore BPs confers repressor-mediated immunity to superinfection. The first step in characterizing the repressor's function is to identify its operator sites and characterize the repressor's DNA-binding capabilities. In this chapter we will discuss

the initial experiments done to characterize the gp33[103] and gp33[136] binding profiles to the primary operator within the *33-34* intergenic region and the evidence for other operators across the BPs genome.

2.2 MATERIALS AND METHODS

2.2.1 Expression and Purification of Mycobacteriophage BPs gp33[103] and gp33[136]

The gp33[103] and gp33[136] genes were PCR amplified from a BPs lysate using primers 5'-CAA TCG CCC ATA TGT CGC AAG CAT TCG -3' / 5'- GAC TAC AAG CTT TCA GAA GGT TGG GGG TTC GA 3'and 5'-CAA TCG CCC ATA TGT CGC AAG CAT TCG -3'/ 5'- TGC CGG AAG AAG CTT TCA CGA CGC TTT ATC C -3' respectively, which amplified the genes with NdeI recognition sites at the 5' end of the gene and HindIII recognition sites at the 3' end. Each gene was cloned into a maltose-binding fusion vector (pLC3) that was linearized with NdeI and HindIII sites for directional cloning, creating two plasmids pVMV20 and pVMV27 for gp33[103] and gp33[136] respectively. pVMV20 and pVMV27 were transformed into BL21(DE3)star chemically competent cells (Invitrogen) and grown until cultures reached an OD_{600} of 0.4-0.6. Protein expression was induced with 1 mM IPTG at 17°C overnight. Cells were pelleted and frozen at -80°C. Thawed cell pellets were resuspended in 5mL per gram of Lysis Buffer (50 mM Tris pH 8.0, 500 mM NaCl, 8% glycerol, 1 mM EDTA and 1 mM β-mercaptoethanol) and lysed in 200 mL fractions by sonicating 10 times for 10 sec at 30% output with 30 sec of cooling on ice in between bursts. Pooled cell lysates were cleared by centrifugation at 30,000 x g for 40 min at 4°C. Fusion proteins were extracted from soluble cell lysates using amylose resin affinity

chromatography (Invitrogen) and the MBP tag was cleaved (169) from the proteins of interest with TEV protease during overnight dialysis (1,000 MWCO dialysis tubing – Spectra/Por) at 4°C. MBP and TEV protease contain C-terminal His tags and were removed from the gp33 proteins using nickel affinity chromatography. The flow through containing pure gp33 proteins was dialyzed into a storage buffer (50 mM Tris pH 8.0, 500 mM NaCl, 50% glycerol, 1 mM EDTA, 1 mM BME) and stored at -20°C.

2.2.2 Electrophoretic Mobility Shift Assays

DNA binding assays were carried out according to standard protocols (170, 171). Briefly, DNA substrates (either PCR substrates or annealed complimentary synthetic oligonucleotides; Table S1) were 5' radiolabeled using ATP, [γ-32P] with T4 polynucleotide kinase (Roche). Binding reactions contained 5-20 cps radiolabeled DNA probe, 1 μg non-specific calf thymus DNA, and varying concentrations of protein in a binding buffer containing 20 mM Tris pH 7.5, 10 mM EDTA, 25 mM NaCl, 10 mM spermidine, and 1 mM DTT for a total volume of 10 μl. Reactions were incubated at room temperature for 30 min and the resulting protein DNA complexes were resolved on a 5% native gel and detected using autoradiography and a phosphorimaging plate. Substrates used in this Chapter are listed in Table 1 below.

2.2.3 Creation and radiolabeling of DNA substrates

All substrates created by PCR amplification in this chapter were amplified from a wild type BPs lysate using high fidelity Cloned PFU DNA polymerase with the primers listed in Table 1. After PCR amplification, the resulting substrates were run on a 0.8% agarose gel, excised, and gel

extracted using a GeneJet gel extraction kit from Fermentas. Substrate concentrations were determined via NanoDrop and diluted to a concentration of 10-20ng/μl depending on the initial concentration. 25μl of the diluted substrate was 5'-end radiolabeled using ATP, [γ-32P] and T4 polynucleotide kinase (Roche) according to manufacturer recommendations and the reaction volume was brought to 50μl and cleaned up on an illustra ProbeQuant G-50 micro Column (GE Healthcare) according to manufacturer recommendations. The counts per minute (cpm) were measured using a Geiger counter and between 5,000-20,000 cpm were used in each EMSA reaction.

Synthetic complimentary single-stranded DNA oligonucleotides were designed and ordered from Integrated DNA Technologies with standard desalting specifications. These oligonucleotides were resuspended in pure water to a concentration of 100μM. 0.5μl of one strand (typically the forward strand) was 5'-end radiolabeled using ATP, [γ-32P] and T4 polynucleotide kinase (Roche) according to manufacturer recommendations and the reaction volume was brought to 50μl and cleaned up on an illustra ProbeQuant G-50 micro Column (GE Healthcare) according to manufacturer recommendations. To the cleaned up reaction, 5.05μl of 10x annealing buffer (100mM Tris pH 7.5, 500mM NaCl) was added along with 0.5μl of the unradiolabeled (typically the reverse) strand. This mixture was vortexed briefly and spun down, then placed in a 95°C waterbath for 5 minutes. The waterbath was then turned off and allowed to cool to room temperature overnight allowing the complimentary radiolabeled and unradiolabeled strands to anneal. The cpm of the resulting dsDNA oligonucleotides were measured using a Geiger Counter and diluted appropriately with pure water so that between 5,000-20,000 cpm were used in each EMSA reaction.

2.2.4 Size Exclusion Chromatography

Purified gp33[103] was dialyzed into a buffer containing 10 mM Tris, pH 8.0, 1 mM BME, 500 mM NaCl, and 4.5% glycerol using 1,000 MWCO dialysis tubing (Specra/Por). 1.2 mL of protein was run over a G-120 column using FPLC, 0.5 ml elution fractions were collected and peaks was identified with UV 280 absorbance measurements. Gel filtration molecular weight markers (Sigma-Aldrich) were resuspended in the same buffer at manufacturer recommended concentrations and run over the same column. The molecular mass for the gel filtration standards were plotted against elution volume (Ve) over void volume (Vo) to determine the molecular mass of gp33[103] based on its elution volume as previously described (169, 172).

2.2.5 Susbtrates used in this Chapter 2

Table 1 - Substrates used in Chapter 2

DNA Substrate Name	Sequence/Primers used
33-34 intergenic region (226bp) – Fig. 7	Fwd: GTTCTTCTTGCTTCAAACCAGCTTCAAGG Rev: GTGTAGTGATCGCTTCGGAATGTCATGC
33-34 intergenic region (366bp) – Fig. 7	Fwd: CACGGGGTAGCGGCTGTTCGTCG Rev: TGCGGGTTCTGGCTGGTGGGGTC
5-6 intergenic region – Fig. 7	Fwd: AACGTAGCCACACACGACGCTTGC Rev: ATTCGCCATGCTGCTTGCTCCCGGT
26-27 intergenic region – Fig. 7	Fwd: TGCCGTGGCGCTGATTCTGAGGTAA Rev: GCATCGGGAAGAAACGATCGGCCAT
54-55 intergenic region – Fig. 7	Fwd: CACGGGGTAGCGGCTGTTCGTCG Rev: TGCGGGTTCTGGCTGGTGGGGTC
60-61 intergenic region – Fig. 7	Fwd: GGCGAGTTCGACGCGCCCGT Rev: TGGCCTTGTTGGTGTCGCGTGGAG
Mt-3 – Fig. 6	Fwd: TTCGGGCGTGGCGCATTTTCCAATAGA Rev: GTGTAGTGATCGCTTCGGAATGTCATGC

O_R competition assay	oVMV-6: GCATTTTCCAATAGACGACATATGTCGCTATGTTGGTGCACA oVMV-6.2: TGTGCACCAACATAGCGACATATGTCGTCTATTGGAAAATGC
O_{R4} competition assay	oVMV-7: GCATTTTCCAATAGACcAgATgTGTtGCTATGTTGGTGCACA oVMV-8: TGTGCACCAACATAGCaACAcATcTgGTCTATTGGAAAATGC
ssDNA	Top Strand: GCATTTTCCAATAGACGACATATGTCGCTATGTTGGTGCACA Bottom Strand: TGTGCACCAACATAGCGACATATGTCGTCTATTGGAAAATGC
1 12bp O_R	oVMV-37: agctccgagctcgcg*CGACATATGTCG*gcgagctcgcgagct oVMV-38: agctcgcgagctcgc*CGACATATGTCG*cgcgagctcggagct
½ 12bp O_R	oVMV-55: GCTCTCCCAGCACTCCGTCCACCGACG*CGACAT*CAGCCTGGCAAGGTGCCTTGGCAGCGG oVMV-56: CCGCTGCCAAGGCACCTTGCCAGGCTG*ATGTCG*CGTCGGTGGACGGAGTGCTGGGAGAGC
1 12bp O_R in native sequence	oVMV-79: GGGCGTGGCGCATTTTCCAATAGACGACATATGTCGCTATGTTGGTGCACATGACCCCAA oVMV-80: TTGGGGTCATGTGCACCAACATAGCGACATATGTCGTCTATTGGAAAATGCGCCACGCCC
1 12bp O_R in non-native sequence	oVMV-39: CTCCCAGCACTCCGTCCACCGACGCGACATATGTCGCAGCCTGGCAAGGTGCCTTGGCAG oVMV-40: CTGCCAAGGCACCTTGCCAGGCTGCGACATATGTCGCGTCGGTGGACGGAGTGCTGGGAG

Table 2 - Binding affinities of selected substrates in Chapter 2

Substrate Name	**Average Calculated K_d (Molar)**
Wild type *33-34* intergenic region	2.61×10^{-6}
5-6 intergenic region	2.75×10^{-6}
26-27 intergenic region	9.00×10^{-7}
54-55 intergenic region	2.60×10^{-6}
60-61 intergenic region	3.40×10^{-6}

2.3 RESULTS

2.3.1 Mycobacteriophage BPs repressor binds DNA

The active mycobacteriophage BPs repressor, gp33^{103}, was purified as described in 2.2.1 and used in DNA binding assays as described in 2.2.2. The substrates tested were prepared according to section 2.2.3 and the substrates used are listed in Table 1.

2.3.1.1 gp33^{103} and gp33^{136} binding to the gene *33-34* intergenic region

Electrophoretic mobility shift assays (EMSA) show that gp33^{103} binds to DNA substrates containing the *33-34* intergenic control region to form several distinct complexes (C1, C2, C3, C4) as protein concentration increases (**Figure 3A**) (97). The combined affinity for gp33^{103} binding is relatively weak (2.6 μM) and the prominent complex (C4) forms at protein concentrations of 16 μM and above. Three faster migrating complexes are observed at lower protein concentrations and are likely binding intermediates. Although DNA binding is relatively weak compared to other repressors, it is specific as little or no binding is observed with a control substrate (**Figure 3B**), and the binding reactions all contain 1 μg calf thymus DNA.

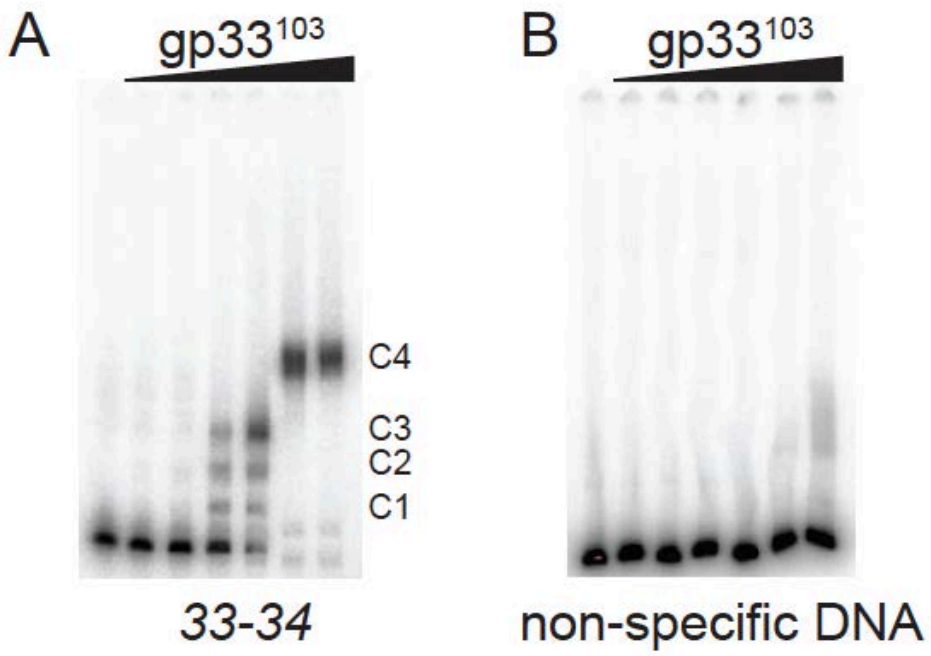

Figure 3 - Binding Profile of gp33[103] to the *33-34* intergenic region

A) EMSA analysis shows that gp33[103] binds to the *33-34* intergenic region with a complicated pattern of binding with four distinct complexes, labeled C1, C2, C3, and C4. B) No binding was seen between gp33[103] and a non-speciic DNA substrate. Protein concentrations are as follows: 1) none 2) 0.16 μM 3) 0.54 μM, 4) 1.6 μM, 5) 5.4 μM, 6) 16 μM, 7) 54 μM. Substrates are shown in Table 1 and Table 2.

BPs gp33[136] is longer than gp33[103] because of an additional 33 residues at its C-terminus. The C-terminal extension includes the ssrA tag that targets the protein for proteolysis, and stabilization of the protein by an A135E substitution gives higher levels of lysogeny. However, it was noted that the gp33[136] A135E mutant, which acts as an active repressor due to inactivation of the ssrA-like tag, appears to give a modest increase in activity of the P_{rep} promoter in a reporter fusion assay (97), perhaps providing transcriptional activation that is dependent on the C-terminal 33 residues. We therefore compared the binding profiles of gp33[103] and gp33[136] for any differences in binding to the *33-34* intergenic region that contains both P_{rep} and P_R (**Figure 4**). Both proteins give similar profiles and the additional 33 C-terminal residues in gp33[136] do not

appear to substantially influence DNA binding. It is likely that any functional activation of P_{rep} results from protein-protein interactions, perhaps reflecting direct contacts between BPs $gp33^{136}$ and RNA Polymerase.

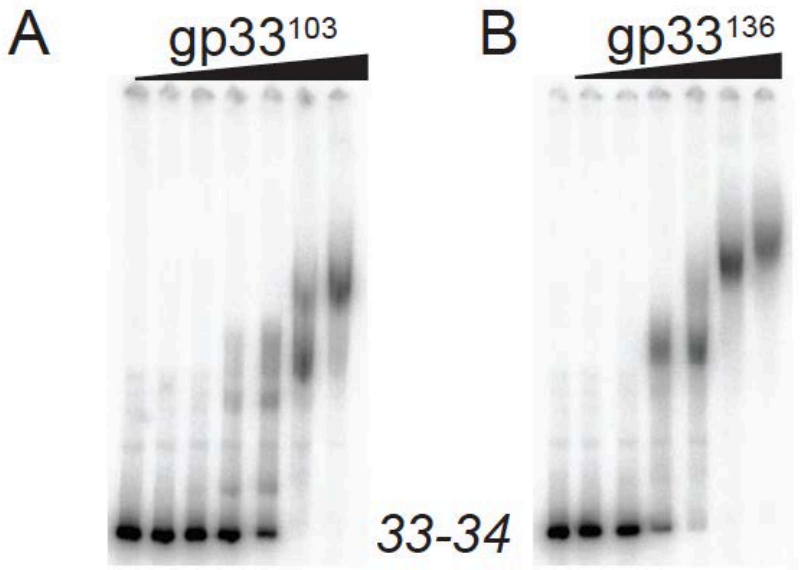

Figure 4 - Comparison of the binding profiles of $gp33^{103}$ and $gp33^{136}$ to the *33-34* intergenic region

DNA binding assays show that both (A) $gp33^{103}$ and (B) $gp33^{136}$ bind similarly to a DNA substrate containing PCR amplified DNA from the BPs gene *33-34* intergenic region (Table 1). Protein concentrations are as follows: 1) none 2) 0.16 μM 3) 0.54 μM, 4) 1.6 μM, 5) 5.4 μM, 6) 16 μM, 7) 54 μM.

The 185 bp DNA segment contains only a single copy of the 12 bp sequence 5'-CGACATATGTCG that was previously predicted to be recognized by $gp33^{103}$, but the complexes formed are more varied than would be expected for a single protein-DNA interaction. Presumably, the protein binds to additional sites within this region, binds with varying protein-DNA stoichiometries, or imposes significant DNA distortions. It is important to note that there is a related sequence (5'-CGACATACCGGC) at the left end (*33*-proximal) of the intergenic region that overlaps the P_{rep} transcription start site (**Figure 2**), although the similarity is restricted to the

left-most seven base pairs of the sequence motif. We therefore sought to dissect the various determinants of $gp33^{103}$ binding to this region, and to compare this to the binding of $gp33^{103}$ to the other sites located elsewhere in the BPs genome. Most assays were done with $gp33^{103}$ because it is more stable than $gp33^{136}$.

2.3.1.2 Binding in the rest of the BPs genome

Both versions of the repressor are capable of binding DNA within the gene *33-34* intergenic region, where the bioinformatically predicted operator site (O_R) is located (**Figure 3**). O_R is composed of a 12bp palindrome located between the -10 and -35 regions of the P_R promoter. Since we predicted that O_R is the probable repressor binding site within the *33-34* intergenic region, we searched the rest of the BPs genome for similar sequences. Three other areas were found that contain sites identical to the 12bp O_R operator within the *5-6*, *26-27* and *54-55* intergenic regions. The intergenic regions containing these other potential operators were PCR amplified from a wild type BPs lysate and used as substrates to test for $gp33^{103}$ binding in EMSA assays. They all showed binding, although each binding profile is different and the protein binds with slightly different affinities to each substrate (**Figure 5**).

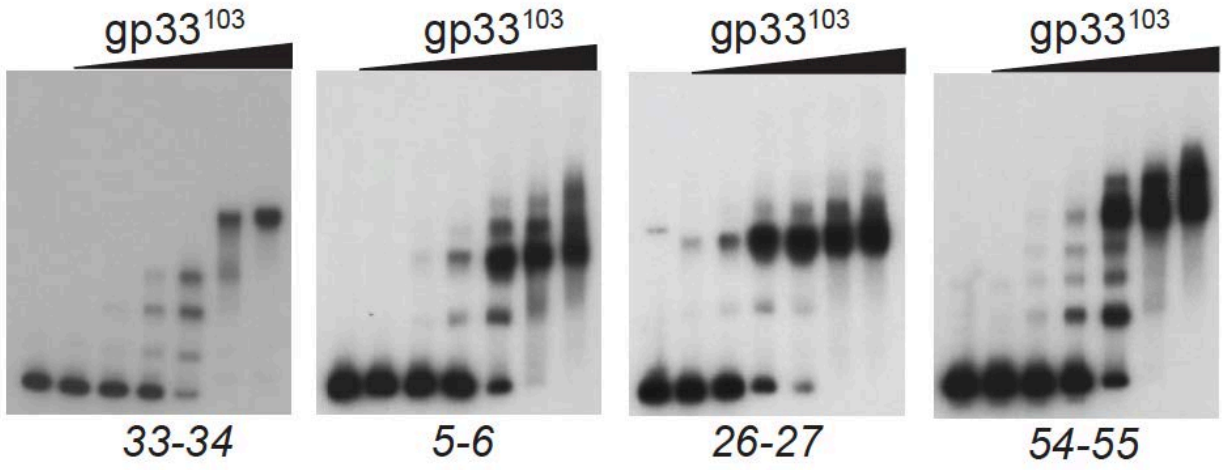

Figure 5 - Binding Profile of gp33[103] to other areas of the BPs genome

The BPs genome contains three occurrences of the 12bp palindrome 5' – CGACATATGTCG outside of the 33-34 intergenic region. EMSAs were performed on substrates containing the intergenic regions where these sequences are located to show that gp33[103] binds to all of them, albeit with simpler binding profiles. Substrates are shown in Table 1 and Table 2. Protein concentrations are as follows: 1) none 2) 0.16 μM 3) 0.54 μM, 4) 1.6 μM, 5) 5.4 μM, 6) 16 μM, 7) 54 μM.

These findings suggest that the predicted 12bp palindromic operator is a true binding site, and that the repressor could exert an effect in multiple areas of the phage genome. The *5-6* and *54-55* intergenic regions have been found to contain promoter activity, so presumably the repressor could be shutting down multiple operons during lysogeny (97, 102). The *26-27* intergenic region does not contain strong promoter activity (103), however it lies upstream of two late-lytic lysin genes, so gp33[103] could perhaps play a role in blocking any leaky expression that could lead to expression of these lytic genes (103).

2.3.2 Confirmation of the gp33 operator.

2.3.2.1 Unradiolabeled operator DNA competes out gp33^{103} binding

To prove that the repressor binds to the 12bp O_R, we performed a competition assay by adding specific un-radiolabeled probes containing the 12bp O_R sequence to a reaction where gp33^{103} was interacting with a radiolabeled probe derived from the right half of the *33-34* intergenic region (amplified from Mt3 primers; Table 1). 42 bp complimentary single-stranded DNA oligonucleotides were engineered to contain the 12bp O_R operator site plus 15 base pairs on each side of surrounding native sequence with either wild type 12bp O_R sequence (denoted as O_R) or with the 12bp O_R containing four point mutations (denoted as O_{R4}) as shown in Table 1. These single-stranded oligonucleotides were annealed as described in section 2.2.3 but without radiolabeling. Binding reactions were set up containing 16μM gp33^{103} and ~15ng of radiolabeled Mt3 DNA (containing O_R and some surrounding sequence - conditions shown to cause a single shift on a gel - Table 1, **Figure 6**). Increasing concentrations of unradiolabeled O_R and O_{R4} were added to the reactions to compete for gp33^{103} binding.

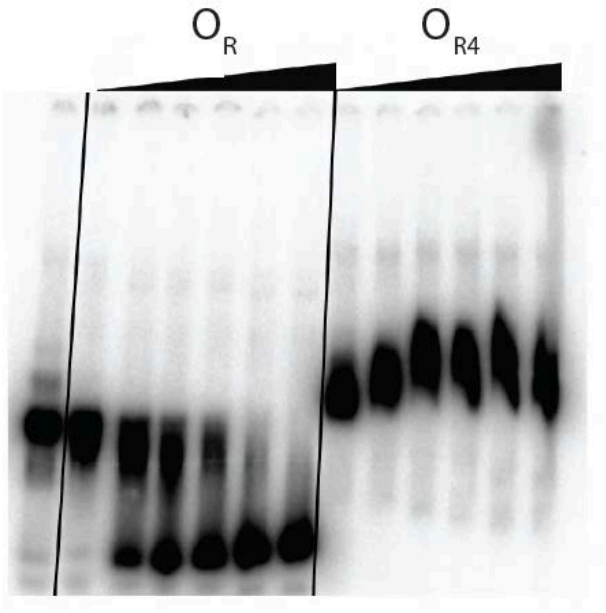

Figure 6 - Competition assay with unradiolabeled O_R DNA

A DNA substrate containing the operator site (Mt3 – Table 3 and Table 4) was radiolabeled and incubated with gp33[103] at a concentration of 16μM. Increasing concentrations of unradiolabeled probe containing the operator site (wild type, O_R, and with four point mutations, O_{R4}) and 15bp of native sequence on either side were added and the binding was outcompeted. Concentrations of unradiolabeled probe in lanes are as follows: 1) 10μM, 2) 30μM, 3) 50μM, 4) 100μM, 5) 300μM, 6) 500μM.

Our results show that the presence of cold O_R outcompetes the radiolabeled probe for gp33[103] binding, whereas O_{R4} does not (**Figure 6**). These results suggest that the 12bp O_R is an operator site and that its sequence is important for binding specificity, but binding assays done with the full *33-34* intergenic region, which show four complexes being formed, indicates that the left side of the *33-34* intergenic region (not containing O_R) might also be important for binding.

2.3.2.2 Two single point mutations in the 12bp O_R lead to repressor insensitive phage phenotypes

Repressor-insensitive mutants are temperate phages that have accumulated mutations allowing them to be able to grow lytically on a lawn of host bacteria expressing its active repressor protein. Despite having copious amounts of repressor present, the proteins have no effect on the infecting phage genome and the lytic pathway dominates the lifecycle. Multiple repressor-insensitive mutants have been isolated from BPs. The sequences of these phages revealed that two of them had single point mutations mapping to the *33-34* intergenic region. Upon closer examination, we noted that these point mutations were both located within the 12bp O_R sequence (Table 3 and Table 4).

We amplified the entire *33-34* intergenic region from the two repressor insensitive mutants containing single point mutations and performed binding assays on them. Compared to $gp33^{103}$ binding to a wild type BPs *33-34* intergenic region, the two point mutants only had a 3- to 4-fold decrease in binding affinity (**Figure 7A**). However when we amplified the right-most regions of the *33-34* intergenic region from each phage (using the same Mt3 primers as were used in the competition assay in section 2.3.2.1) and did an EMSA with $gp33_{103}$ on them, they showed a 10-fold reduction in binding affinity compared to wild type (**Figure 7B**).

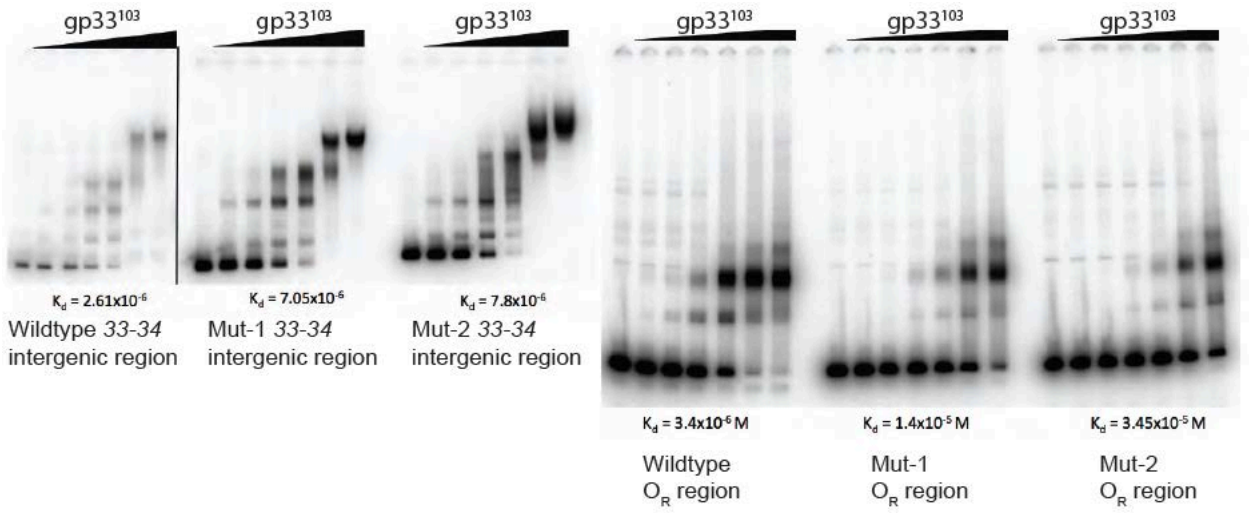

Figure 7 - gp33[103] binding to two repressor-insensitive mutants

BPs gp33[103] binding to the *33-34* intergenic region (left 3 panels – Table 1) and the right-most portion of the 33-34 intergenic region (using Mt3 primers, Table 3) derived from two repressor-insensitive BPs mutants with single base pair substitutions, 102a (Mut-1) and 102e (Mut-2). Binding profiles of gp33[103] to each mutated substrate show a decrease in binding affinity compared to wild type. The concentrations of protein are as follows: 1) none, 2) 0.16μM, 3) 0.54μM, 4) 1.6μM, 5) 5.4μM, 6) 16μM, 7) 54μM.

This evidence indicated that not only is the sequence of the palindrome important for the repressor's biological function, but that it is so important that a single point mutation, which decreases the repressor's binding affinity only modestly, is able to prevent host immunity from superinfection. Despite these promising results, the requirements for DNA binding at O_R and at related sequences elsewhere in the phage genome and the nature of the protein-DNA interactions are not well understood.

2.3.3 Multimeric state of the Mycobacteriophage BPs repressor in solution

We estimated the oligomeric state of gp33[103] in solution. A single protein peak was observed using size-exclusion chromatography (**Figure 8**), and when compared with protein

markers of globular shape and low molecular weights, gp33^{103} has an apparent molecular mass of 40.4 kDa (**Figure 9**). The monomeric mass of gp33^{103} is 11.2 kDa, and the simple interpretation is that the major peak corresponds to a gp33^{103} tetramer, although we cannot rule out the possibility that it is an alternative multimer shaped to give altered elution in the chromatography. The 12 bp operator described in the previous sections has dyad symmetry consistent with recognition by either a dimer or tetramer of gp33^{103}. The gp33^{103} elution profile did not show any other prevalent forms and the oligomeric state is quite homogenous (**Figure 8**).

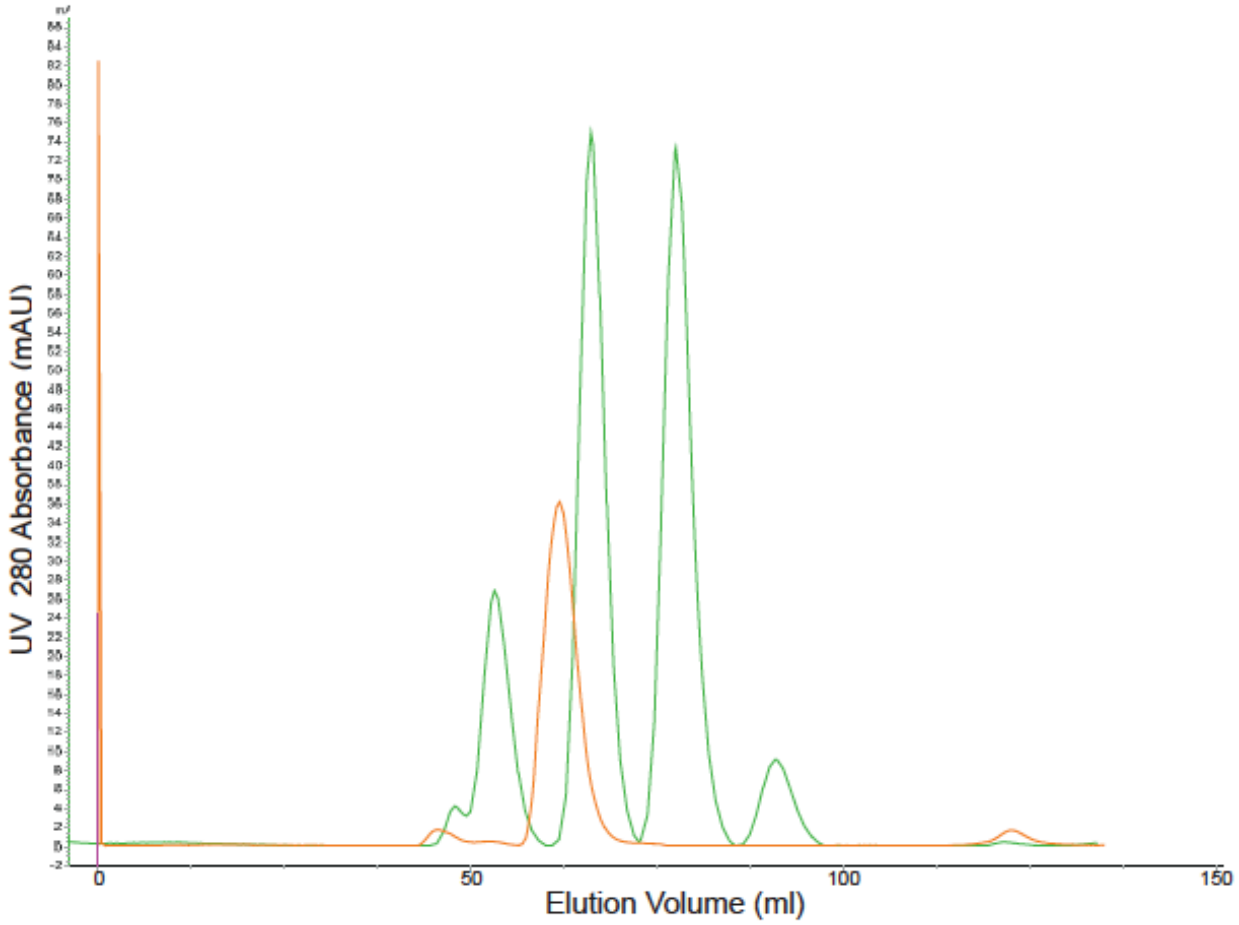

Figure 8 - Size-exclusion chromatogram

Size-exclusion chromatograms for gp33^{103} (orange) and the molecular weight markers (green; 66kDa, 29kDa, 12.4kDa, 6.5kDa) are overlayed. Gp33^{103} elutes in a single peak.

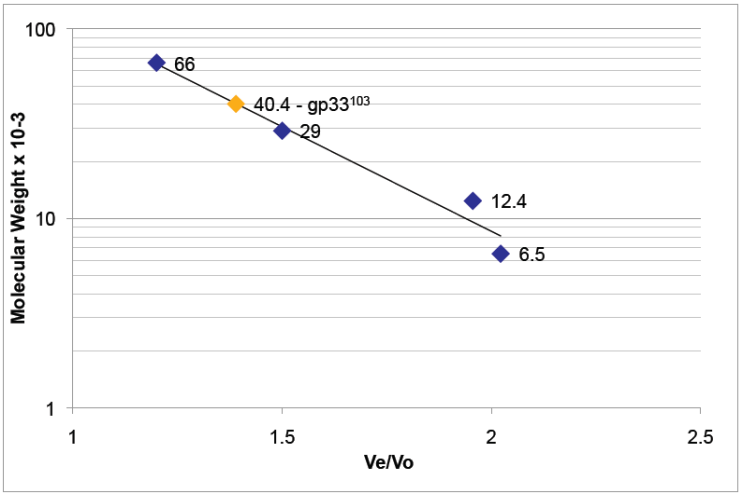

Figure 9 - gp33[103] exists as a tetramer in solution

The graph shows a standard plot for the molecular mass for the protein standards against the ratio of their elution volume (Ve) and void volume (Vo) (blue diamonds); the molecular mass of gp33[103] is predicted to be 40.4kDa (orange diamond). Gp33[103] is an 11.2 kDa protein, so it is likely a tetramer in solution.

2.3.4 Gp33[103] binds to other DNA substrates

2.3.4.1 Gp33[103] binds to ssDNA

To determine that the operator site for gp33[103] is a double-stranded DNA binding site, we tested whether gp33[103] was able to bind single-stranded DNA as well as double-stranded DNA. In this assay, we radiolabeled two complimentary strands of synthetic oligonucleotides engineered to contain a central 12bp operator site flanked by 15bp of native sequence on each side. We tested gp33[103]'s ability to bind to each single-stranded DNA oligonucleotide separately, and then annealed them to create a double-stranded DNA substrate of identical sequence for reference and did an EMSA on that as well.

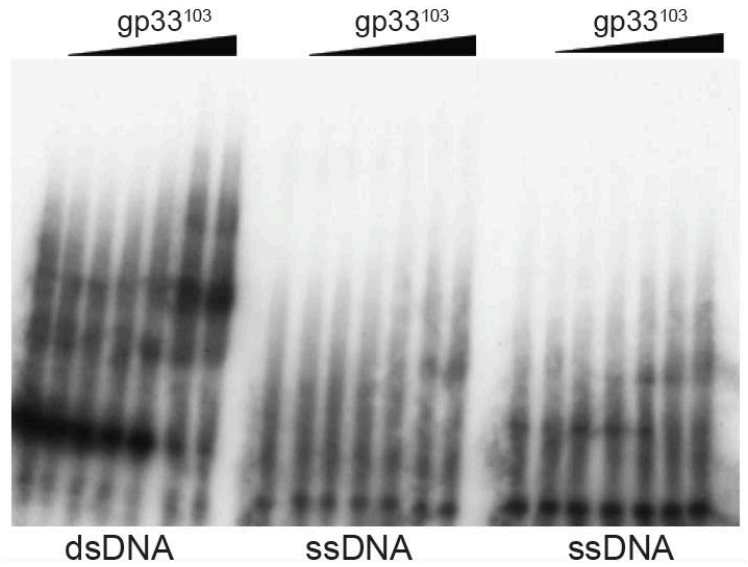

Figure 10 - gp33^{103} binds faintly to single-stranded DNA

EMSAS were done on a 42bp double stranded DNA substrate containing O_R and 15bp of native sequence on either side (Table 1). Each strand of the double-stranded DNA substrate was radiolabeled separately prior to annealing. EMSAs done on these substrates show that gp33^{103} binds very weakly to single-stranded DNA. The left panel is the double stranded DNA substrate, the center panel is top strand ssDNA, and the right panel is bottom strand ssDNA. Protein concentrations were as follows: 1) none, 2) 0.16μM, 3) 0.54μM, 4) 1.6μM, 5) 5.4μM, 6) 16μM, 7) 54μM.

The results show that gp33^{103} is able to bind each single-stranded DNA substrate only very modestly, but it binds strongly to the double-stranded DNA substrate with identical sequence (**Figure 10**). Interestingly, the mobility of the complexes formed with each of the single-stranded DNA substrates runs similarly to the faint fast-migrating complex seen in the double-stranded DNA complex. It is plausible that the faint, faster-moving complexes formed when gp33^{103} binds to synthetically created, annealed oligos could be due to the protein binding to un-annealed single-stranded DNA contaminants in the substrate preparation. Not many phage repressors have been shown to bind to single-stranded DNA (88) and the probability of the BPs repressor to bind single-stranded DNA and, if it does bind, that it performs a specific biological function via that interaction *in vivo* is unknown.

2.3.4.2 Gp33[103] binds to operator half-sites

The leftmost portion of the *33-34* intergenic region, just inside the *gene 33* open reading frame, contains a 6bp operator half-site. The *33-34* intergenic region substrates in this chapter all contain this half-site due to the location of the primers used to amplify the region (**Table 1**). One of the potential reasons for multiple complexes being formed when EMSAs are performed on these substrates is that gp33[103] is able to bind the O_R half-site, leading to different protein-DNA stoichiometries in each reaction. To test if gp33[103] is able to bind this half-site, we engineered synthetic oligonucleotides where we placed the 6bp sequence in a non-native sequence context and compared the binding profile of the protein to this substrate to that of a substrate containing a full 12bp O_R site in a non-native sequence context (**Table 1; Figure 11**).

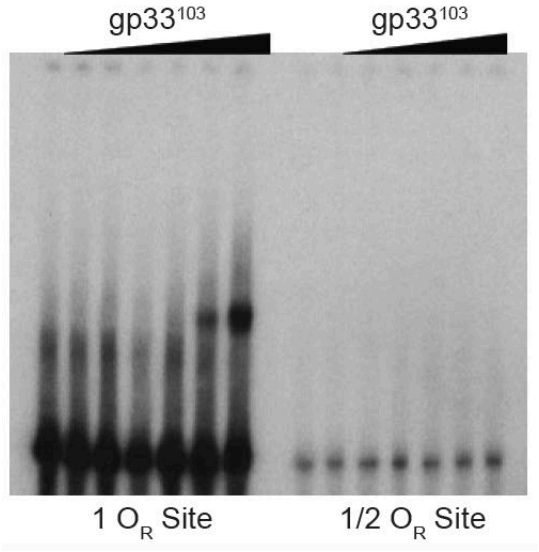

Figure 11 - gp33[103] does not bind to a 6bp O_R half site

Gp33[103] is able to bind to a 60 bp substrate containing the 12 bp O_R centrally but not to a 60 bp substrate containing a 6 bp O_R half site located centrally (Table 1).

The results indicate that while gp33[103] can bind to a full 12bp operator site out of native sequence context, it cannot bind to a 6bp half-site (**Figure 11**). However, these results do not necessarily rule out involvement of the half-site in formation of one or more of the complexes

seen when gp33^{103} binds to the entire *33-34* intergenic region. In fact, gp33^{103} may only bind to the 6bp half-site cooperatively, and there could be a requirement for a full 12bp site nearby for the half-site to be occupied. This theory suggests that perhaps cooperative binding and DNA bending or looping may play a role in formation of the complexes seen when gp33^{103} binds to the *33-34* intergenic region.

2.3.4.3 Native sequence surrounding O_R is important for major complex formation

In its natural sequence, gp33^{103} binds O_R and forms two complexes: a slower-migrating major complex, and a faster-migrating complex (**Figure 12**). If gp33^{103} is in fact a tetramer in solution, the major band could be a tetramer bound at the O_R site and the minor complex could be either a dimer or monomer binding to the O_R site. When O_R is placed outside of its native sequence context, two complexes are again formed, but interestingly, the major complex shifts from being the slower-migrating complexes to being the faster moving complex (**Figure 12**).

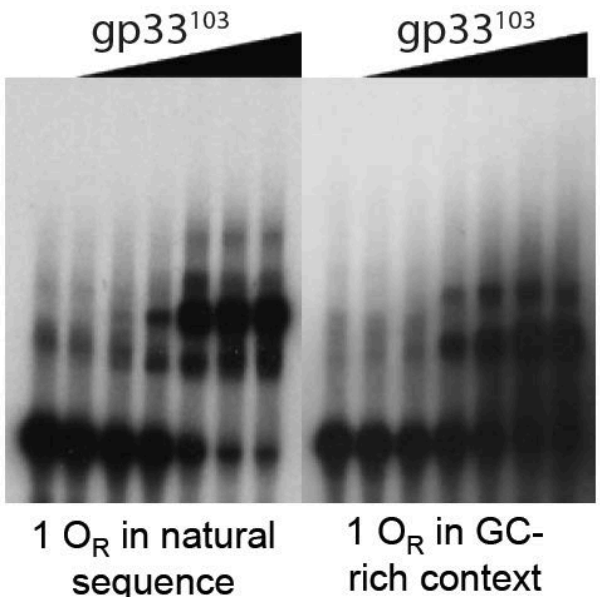

Figure 12 - gp33^{103} binding to O^{R} in a native and non-native sequence context

EMSAs were done with gp33^{103} binding to two 60 bp substrates containing a single 12 bp O_R site located centrally. One substrate contained 24 bp of native sequence surrounding O_R, the other contained non-native sequence (Table 1). Gp33^{103} has a slower migrating major complex when O_R is in its native sequence compared to when it is in a non-native sequence.

In section 2.3.4.2 we suggest a role for DNA bending within the *33-34* intergenic region, and such a theory could explain the changes seen in major complex formation between the substrates shown in Figure 12. When O_R is centered on a 60bp substrate that is not its natural sequence, gp33^{103} binds to O_R and causes a shift (faster-moving complex), but the lack of an AT-righ region surrounding the operator may not allow the protein to bend DNA very efficiently, leading to weaker formation of the slower-moving complex. However in its native sequence, gp33^{103} could not only bind DNA, but also efficiently bend DNA, which would lead to high incidence of a slower-moving complex corresponding to gp33^{103} bending the DNA substrate it is attached to, and also possibly a faster-moving complex where gp33^{103} binds, but does not bend the DNA.

2.4 DISCUSSION

In this chapter we have shown that gp33^{103} binds with a complicated pattern of binding, but that it binds specifically, and that it binds similarly to the gp33^{136} long form of the protein. We have also identified a 12bp operator site that is present within four intergenic regions across the BPs genome. The primary regulatory region, the *33-34* intergenic region, where two divergent promoters exist is where we first identified this binding site and we called it O_R. gp33^{103} also binds within the other three intergenic regions where identical O_R-like sites are located. To confirm this as an operator, we performed a competition assay in which an unradiolabeled probe competed out the gp33^{103} binding, and we also examined some repressor-insensitive mutation data that showed that a single base pair change in this operator sequence led to dramatic consequences *in vivo*. Gel filtration chromatography reveals that this protein seems to exist as a tetramer in solution, which could help explain some of the many complexes seen when gp33^{103} binds to the *33-34* intergenic region.

We also showed that gp33^{103} is able to bind to single-stranded DNA weakly. The significance of this experiment in understanding the integration-dependent genetic switch as a whole is not great, however it has been shown that proteins that bind to single stranded DNA often make hydrophobic interactions with DNA (173). Bacteriophages P22 and 434 encode a cro and repressor protein, respectively, that recognizes DNA via hydrophobic interactions (166, 167). So perhaps it is not far-fetched to suggest that this is how gp33^{103} recognizes its operator sequence. Also, there seems to be some dependence on the surrounding sequence of the operator for proper binding. This will be explored in subsequent chapters.

3.0 MAPPING THE SEQUENCE REQUIREMENTS FOR GP33^{103} BINDING

Work done in this section was published in the following publications:

Broussard, G.W., Oldfield, L.M., **Villanueva, V.M.**, Lunt, B.L., Shine, E.E., Hatfull, G.F., January 2013. Integration-Dependent Bacteriophage Immunity Provides Insights into the Evolution of Genetic Switches. *Molecular Cell*, Volume 49, Issue 2, Pages 237-248.

Villanueva, V.M., Hatfull, G.F., September 2015. An Unusual Phage Repressor Encoded by Mycobacteriophage BPs. *PLoS One*, Volume 10:e0137187.

3.1 INTRODUCTION

We have established that gp33^{103} is the active form of the BPs repressor and that it is capable of binding DNA at a putative 12bp palindromic operator, O_R, as well as to three identical sites across the BPs genome. When gp33^{103} binds to the other identical 12bp sites, a much simpler binding pattern is seen than when gp33^{103} binds within the *33-34* intergenic region. We sought to explain the nature of the complexes seen when gp33^{103} binds to the *gene 33-34* intergenic regulatory region and explain why these are formed in this particular area, but not in other regions of the genome. In this chapter we will discuss experiments done to elucidate the nature of the C1, C2, C3, and C4 complexes formed when gp33^{103} binds the *33-34* intergenic region, as well as work done to elucidate the true sequence of the operator site, which is two 12bp half sites

separated by a spacer region, and the sequence requirements for gp33[103] binding and relate that to its function *in vivo*.

3.2 MATERIALS AND METHODS

3.2.1 DNase I Footprinting

Footprinting assays were carried out as previously described (170, 174, 175). Briefly, binding reactions were carried out in a final volume of 50 µl containing various concentrations of protein, 20cps radio labeled probe, 25 mM Tris-HCL pH 8.0, 50 mM KCl, 6.25 mM $MgCl_2$, 0.5 mM EDTA, 10% Glycerol, 0.5 mM DTT. Binding reactions were incubated at room temperatures for 30min. After incubation, 50 µl of a solution containing 5 mM $CaCl_2$ and 10 mM $MgCl_2$ was added to the samples and incubated for 1 minute. Samples were treated with 1.5U DNaseI (Sigma) for exactly one minute, then the digestion reaction was stopped by addition of 90µl of a pre-warmed (37°C) Stop solution (200 mM NaCl, 30 mM EDTA, 1% SDS, 100 µg/mL yeast RNA). Samples were PCI (Invitrogen) extracted and ethanol precipitated. Samples were resuspended in 2-3 µl formamide loading buffer and heated to 95°C for 2 minutes before loading onto a 6% polyacrylamide 7M urea denaturing gel for resolution. Bands were detected using autoradiography or exposed on a phosphorimageing plate and detected on a FLA-5100; FujiFilm imaging system. Sequencing ladders were made as previously described (175, 176).

3.2.2 DNA-Binding Assays

EMSAs were done according to protocol described in 2.2.2. Substrates used in this chapter were prepared as described in 2.2.3 are listed in Table 2 below.

Table 3 - Substrates used in Chapter 3

DNA Substrate Name	Sequence/Primers used
DNase I Footprinting substrate – Fig. 13	Fwd: TCTAGAACCGGCGATTGTATGTAC Rev: AAGCTTGTGCACCAACATAGCGAC
Region 1 – Fig. 15	AACCGTGCATATCGCGCAATACACTTGCGACATACCGGCAACGTTGAAACAATCACGTTG
Region 2 – Fig. 15	AACCGTGCATATCGCGCAATACAAACATTGTATGGTTGAAACAATCACGTTGCACGTTGC
Region 3 – Fig. 15	AACCGTGCATATCGCGCAATACAAACTTCGGGCGTTGAAACAATCACGTTGCACGTTGCA
Region 4 – Fig. 15	AACCGTGCATATCGCGCAATACAAACGGCGCATGTTGAAACAATCACGTTGCACGTTGCA
Region 5 – Fig. 15	AACCGTGCATATCGCGCAATAAATAGACGACATATGTCGCAAACGTTGAAACAATCACGT
TIR-1 – Fig. 18	ACCGGCGATTGTATGTACGTCTGTCGCGTACGCGCAATAT
TIR-2 – Fig. 18	CGCGCAATATCGCCGTTCGTCAACCGTTGCGAGCGGAAAC
TIR-3 – Fig. 18	GAGCGGAAACGGCACGGCGCGTTGATCCTGGCAAACCGTC
TIR-4 – Fig. 18	GCAAACCGTCGGTGGTAACGGATTGCGTGGCAAACGCAAA
TIR-5 – Fig. 18	CAAACGCAAAAGTTCGGGCGTGGCGCATTTTCCAATAGAC
TIR-6 – Fig. 18	TCCAATAGACGACATATGTCGCT
TIR-5.1 – Fig. 19	CAAACGCAAAAGTTCGGGCG
TIR-5.2 – Fig. 19	TGGCGCATTTTCCAATAGAC
TIR-5.3 – Fig. 19	GATTGCGTGGCAAACGCAAA
TIR-5.4 – Fig. 19	AGTTCGGGCGTGGCGCATTT
TIR-5.5 – Fig. 19	TCCAATAGACGACATATGTCGCT
TIR-5.6 – Fig. 19	GCAAAAGTTCGGGCGTGGCG
TIR-5.7 – Fig. 19	GGGCGTGGCGCATTTTCCAA
Mt-2 – Fig. 21	Fwd: TTGGTGCACATGACCCCAACCGAAAC Rev: GTGTAGTGATCGCTTCGGAATGTCATGC
Mt-3 – Fig. 21	Fwd: TTCGGGCGTGGCGCATTTTCCAATAGA Rev: GTGTAGTGATCGCTTCGGAATGTCATGC
Mt-4 – Fig. 21	Fwd: TGGCAAACCGTCGGTGGTAACGGAT Rev: GTGTAGTGATCGCTTCGGAATGTCATGC
Mt-5 – Fig. 21	Fwd: GTTCGTCAACCGTTGCGAGCGGAAA Rev: GTGTAGTGATCGCTTCGGAATGTCATGC
Mt-6 – Fig. 21	Fwd: ACATACCGGCGATTGTATGTACGTCTGTC Rev: GTGTAGTGATCGCTTCGGAATGTCATGC
Mt-7 – Fig. 21	Fwd: ACGAAAGCCTGCTCGGCGGGATC Rev: GTGTAGTGATCGCTTCGGAATGTCATGC
Mt-8 – Fig. 21	Fwd: GTTCTTCTTGCTTCAAACCAGCTTCAAGG Rev: GTTTCGGTTGGGGTCATGTGCACCAA
Mt-9 – Fig. 21	Fwd: GTTCTTCTTGCTTCAAACCAGCTTCAAGG Rev: TCTATTGGAAAATGCGCCACGCCCGAA

Mt-10 – Fig. 21	Fwd: GTTCTTCTTGCTTCAAACCAGCTTCAAGG Rev: ATCCGTTACCACCGACGGTTTGCCA
Mt-11 – Fig. 21	Fwd: GTTCTTCTTGCTTCAAACCAGCTTCAAGG Rev: tttccGCTCGCAACGGTTGACGAAC
Mt-12 – Fig. 21	Fwd: GTTCTTCTTGCTTCAAACCAGCTTCAAGG Rev: GACAGACGTACATACAATCGCCGGTATGT
Mt-13 – Fig. 21	Fwd: GTTCTTCTTGCTTCAAACCAGCTTCAAGG Rev: GATCCCGCCGAGCAGGCTTTCGT
$O_{6\text{--}L}$ – Fig. 26	CACACACGACGCTTGCACCGCGACATATGTCGGTTTTGCAGACGTGCAGGTT
$O_{6\text{-}R}$ – Fig. 26	GCATTTTCCAATAGAGGACATATGCGGCTATGTTGGTGCACA
$O_{27\text{-}R}$ – Fig. 27	GCATCACGGACGCTGCTATCCGACATATGTCGTGGATGATCGGGATATGGCC
$O_{27\text{-}L}$ – Fig. 27	GCATTTTCCAATAGACCGCATATGTCGCTATGTTGGTGCACA
$O_{R\text{-}R}$ – Fig. 25	GCATTTTCCAATAGACGACATATGTCGCTATGTTGGTGCACA
$O_{Rep\text{-}R}$ – Fig. 25	TTCGGGCGTGGCGCATTTTCCAATAGACGACATCTATGTTGGTGCACATGACCCCAACCG
$O_{55\text{-}L}$ – Fig. 28	TCGGACATGGTGAACGCTTGCGACATATGTCGCCAATGTGTGCATCTTTGCA
$O_{55\text{-}R}$ – Fig. 28	GCATTTTCCAATAGATGACACTTGTCGCTATGTTGGTGCACA
$O_{61\text{-}R}$ – Fig. 29	GCATTTTCCAATAGACGACATATGTGCCTATGTTGGTGCACA
$O_{61\text{-}L}$ – Fig. 29	GCATTTTCCAATAGACGACATTCGCGCCTATGTTGGTGCACA
5-6 intergenic region – Fig. 26	Fwd: AACGTAGCCACACACGACGCTTGC Rev: ATTCGCCATGCTGCTTGCTCCCGGT
26-27 intergenic region – Fig. 27	Fwd: TGCCGTGGCGCTGATTCTGAGGTAA Rev: GCATCGGGAAGAAACGATCGGCCAT
33-34 intergenic region (366bp) – Fig. 25	Fwd: GTTCTTCTTGCTTCAAACCAGCTTCAAGG Rev: GTGTAGTGATCGCTTCGGAATGTCATGC
54-55 intergenic region – Fig. 28	Fwd: CACGGGGTAGCGGCTGTTCGTCG Rev: TGCGGGTTCTGGCTGGTGGGGTC
60-61 intergenic region – Fig. 29	Fwd: GGCGAGTTCGACGCGCCCGT Rev: TGGCCTTGTTGGTGTCGCGTGGAG
Repressor-insensitive BPs mutants - *33-34* intergenic region – Fig. 31	Fwd: GTTCTTCTTGCTTCAAACCAGCTTCAAGG Rev: GTGTAGTGATCGCTTCGGAATGTCATGC

Table 4 - Binding affinities of gp33[103] to selected substrates in Chapter 3

Mutant Name	**Specific Mutation within *33-34* intergenic region**	**Average Calculated K_d (Molar)**
Wild type *33-34* intergenic region	None	2.61×10^{-6}
102a	T29489C	7.05×10^{-6}
102e	A29486C	7.80×10^{-6}
102j	151bp Inversion (29282-29432)	4.23×10^{-6}
102k	24bp Duplication (29472-29495)	6.67×10^{-6}
127b	151bp Deletion (29323-29473)	9.50×10^{-6}
127c	33bp Duplication (29495-29527)	3.50×10^{-6}
127d	12bp Duplication (29474-29485)	2.99×10^{-6}
127e	6bp Duplication (29483-29488)	3.81×10^{-6}
Δ32 Clr1	T29336C	1.74×10^{-6}
Δ32 Clr4	G29500A	1.45×10^{-6}
33A129E Clr5	C29372T	2.90×10^{-6}
33A129E Clr6	T29475C	2.65×10^{-6}
Δ32 Clr8	T29501C	2.80×10^{-6}
5-6 intergenic region	None	2.75×10^{-6}
26-27 intergenic region	None	9.00×10^{-7}
54-55 intergenic region	None	2.60×10^{-6}
60-61 intergenic region	None	3.40×10^{-6}
$O_{6\text{-}L}$	$O_{6\text{-}R}$ ablated	1.05×10^{-5}
$O_{27\text{-}R}$	$O_{27\text{-}L}$ ablated	1.10×10^{-5}
$O_{27\text{-}L}$	None	3.85×10^{-5}
$O_{55\text{-}L}$	$O_{55\text{-}R}$ ablated	8.50×10^{-6}
$O_{61\text{-}R}$	$O_{61\text{-}L}$ ablated	1.95×10^{-5}
Region 5	None	1.20×10^{-5}
TIR-5	None	1.25×10^{-5}
TIR-6	None	9.50×10^{-6}

3.3 RESULTS

3.3.1 DNase I Footprinting

Both the viral and prophage forms of $gp33^{103}$ have been purified. Previous studies have shown that the prophage form of $gp33^{103}$ binds within the gene *33-34* intergenic region, which contains two divergent promoters: one for rightwards expression of an early lytic operon, P_R, and one for leftward expression of $gp33^{103}$ and Int (102). Studies done by Broussard *et al*. showed that $gp33^{103}$ binds specifically to a 12 base pair palindrome, O_R. DNA binding assays from that study showed that the binding affinity of the BPs $gp33^{103}$ to its operator sites is low, with a dissociation constant of approximately 1 μM, which is 10-fold higher than that expected for a phage repressor (97). To confirm that $gp33^{103}$ binds to O_R, we performed DNase I footprinting assays (**Figure 13**).

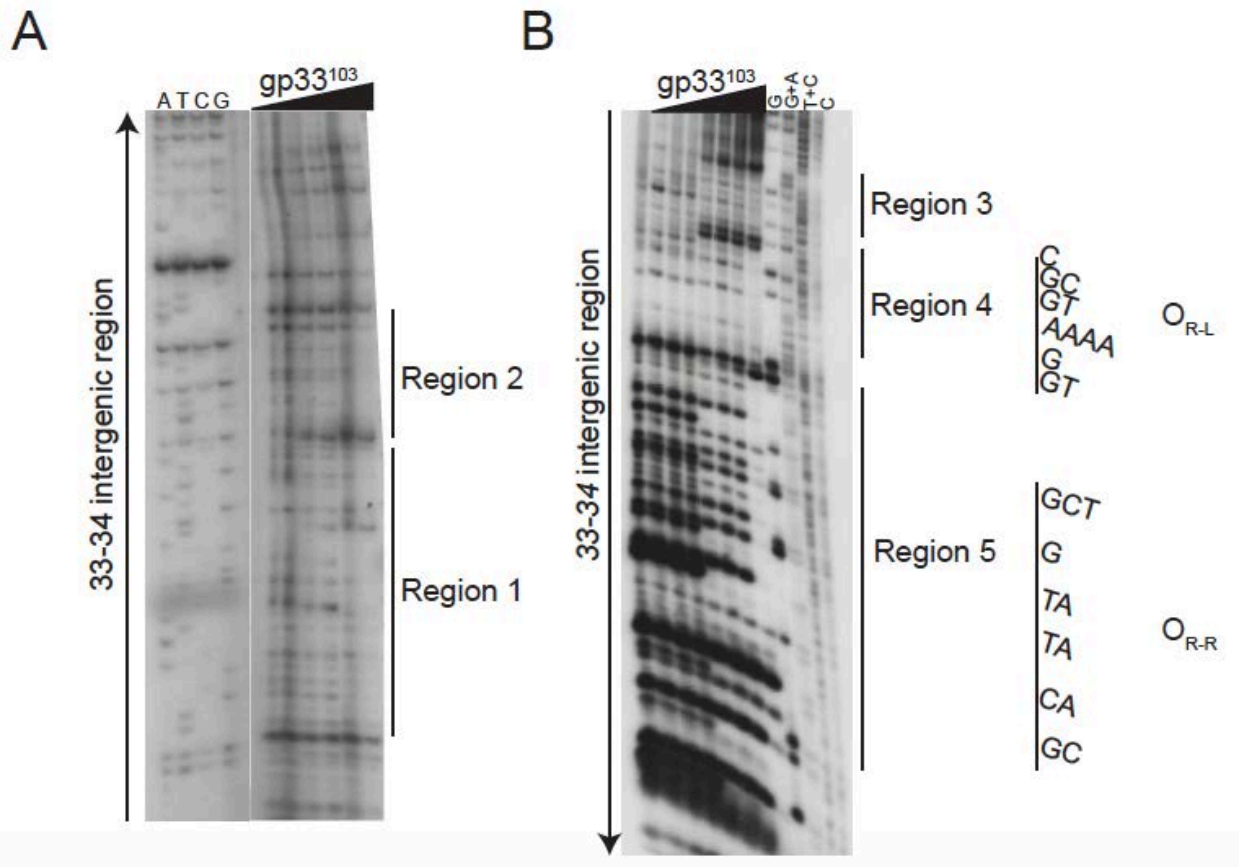

Figure 13 - DNase I footprinting assay of gp33[103] binding to the *33-34* intergenic region

DNase I footprinting assays (panels A and B) were done using 5' end-labeled probes of the *33-34* intergenic region, which were incubated with increasing concentrations of gp33[103], subjected to DNase I cleavage, and run on a denaturing gel alongside a sequencing ladder. Black arrows show the direction of the sequence from 5' to 3' for each DNA strand. Regions of protection from DNase I cleavage are indicated by black bars and labeled. Two regions of protection observed near P_{rep}, one of which contains O_{Rep} (Regions 1 and 2). Three regions of protection are observed near P_R (Regions 3 and 4, Region 5), one of which includes O_R (Region 5). The sequence of O_R is shown (O_{R-R}) as well as a predicted second binding site (O_{R-L}) (section 3.3.2.3). Protein concentrations are as follows: 1) none 2) 0.16 μM 3) 0.54 μM, 4) 1.6 μM, 5) 5.4 μM, 6) 16 μM, 7) 54 μM, 8) 160 μM.

Figure 13 shows representative DNase I Foot printing assays – one for each strand of the probe. It is evident upon first glance that there seem to be two patterns of altered DNase I protection: one when intermediate protein concentrations are added and one when the most concentrated protein is used. When the forward strand of the *33-34* intergenic region was

probed, we found two areas of protection from DNase I cleavage separated by an area of enhanced cleavage (**Figure 13A**). The first area of protection, Region 1, is 16bp long and is located just inside of the *gene 33* open reading frame. Upon further examination, we found that this area contains an O_R half-site that was originally dismissed as a potential binding site due to the lack of binding seen when it was isolated in a non-native sequence; however there was no evidence to show that the site was unoccupied when $gp33^{103}$ binds in the presence of the rest of the *33-34* intergenic region. The second area, Region 2, is an 8bp site that could be a continuation of Region 1. The binding of $gp33^{103}$ to these sites could be dependent upon both of them being present on the same substrate, though this second region of protection contains no obvious sequence similarity to the 12bp O_R. One curious aspect of this binding profile is that there are two areas of DNase I cleavage enhancement surrounding Region 2. This enhancement could mean multiple things, including increased cleavage due to favorable protein-protein interactions between $gp33^{103}$ and DNase I, or perhaps $gp33^{103}$ caused a change of DNA secondary structure, such as DNA bending, leading to greater exposure of those base pairs to the enzyme. From this assay we cannot confirm the cause of the enhancements.

When the antiparallel (reverse) strand of DNA was probed, we found three areas of protection from DNase I cleavage (**Figure 13B**). Region 3 and Region 4 are short binding sites upstream of O_R with no remarkable sequence identity to the 12bp operator site, although it has been noted previously (section 2.3.4.3) that the AT-rich nature of the sequence surrounding the 12bp O_R site could have a function in allowing $gp33^{103}$ to bend DNA. Surrounding Site 3 we again see areas of enhanced DNase I cleavage, but we cannot conclude a reason for the enhancements from this assay alone. Region 5 contains the entire 12bp O_R site. Overall we can see an area of clearing spanning at least 19 base pairs (**Figure 13B**). The sequence gets cut off at

the right edge of O_R at the bottom of the gel so it is unclear whether there is a larger area of protection in that direction.

Figure 14 - Regions of protection from DNase I footprinting

The sequences of the areas of protection from DNase I cleavage are shown in the context of the entire *33-34* intergenic region.

The assays showed that gp33[103] does indeed bind to the 12bp operator O_R, and also to four other sites across the region (**Figure 14**). To confirm these as true binding sites, we performed an EMSA on each site separately in a non-native sequence context (**Figure 15**) as well as each set of sites together based on the area of the *33-34* intergenic region in which they are located (**Figure 16**). Each probe used was standardized to the same length for easy comparison.

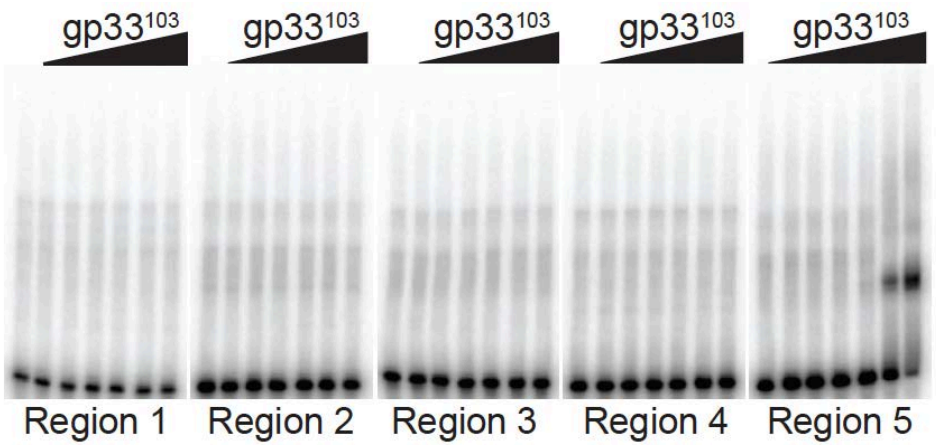

Figure 15 - gp33[103] binding to Regions of DNase I protection

DNA binding assays with gp33[103] using synthetic oligonucleotide substrates that contain the sequence of each region identified in Figure 13, each in a non-native 60 bp sequence context. BPs gp33[103] only binds to Region 5, which contains O_R. Protein concentrations are as follows: 1) none 2) 0.16 μM 3) 0.54 μM, 4) 1.6 μM, 5) 5.4 μM, 6) 16 μM, 7) 54 μM.

When the five sites were isolated from each other in a non-native sequence context, only the site containing the full O_R showed binding (**Figure 15**). But grouped together, Sites 1 & 2 showed modest binding and Sites 3, 4, and 5 show a more complex binding pattern (multiple shifts) than that seen with Site 5 alone (single shift) (**Figure 16**).

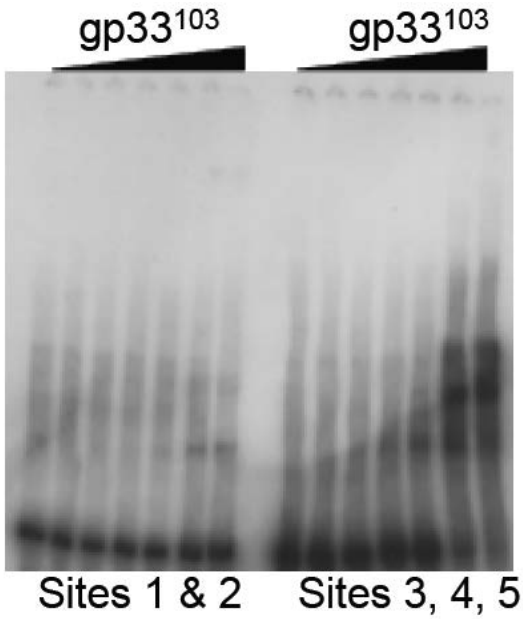

Figure 16 - gp33[103] binding to grouped regions of DNase I protection

Regions of protection from DNase I cleavage were grouped on a 60 bp DNA substrate according to which side of the genome they were on (33-proximal or 34-proximal). Substrates have non-native sequence surrounding the Regions, but native sequence between them. Gp33[103] is able to bind to Regions 1 and 2, as well as 3, 4, and 5. Protein concentrations are as follows: 1) none 2) 0.16 μM 3) 0.54 μM, 4) 1.6 μM, 5) 5.4 μM, 6) 16 μM, 7) 54 μM.

These results, along with previous genetic and biochemical data collected (97), show that the BPs gp33[103] binds specifically to the 12bp O_R site as well as the region immediately upstream of O_R. These assays also show that gp33[103] binds to the opposite end of the *33-34* intergenic region where an operator half-site exists, which could be the reason we see such a complicated binding pattern in mobility shift assays on this substrate. The regions of protection at P_{rep} will be referred to as O_{Rep} from here on out. It is likely that the slowest moving of the protein-DNA complexes seen when gp33[103] binds to the *33-34* intergenic region (C4) corresponds to the more extensive DNase I protections seen at the highest protein concentration, and that the faster migrating complexes correspond to those seen at intermediate concentrations.

3.3.2 TIR and End-Deletion Series of Substrates

3.3.2.1 TIR series of substrates

We created a series of synthetic double-stranded DNA oligonucleotide substrates spanning the *33-34* intergenic region and overlapping with each other by only 10bp so as to further explore where gp33[103] may be binding (**Figure 17A**). The results from performing EMSAs on these substrates did not show binding in any of the substrates except TIR-5 and TIR-6 (**Figure 18**). We did not see any binding to the TIR-1 substrate containing Region 2 from the DNase I footprinting results (section 3.3.1), which is not surprising since we already showed in **Figure 15** that gp33[103] did not bind to Region 1 or Region 2 when isolated from each other.

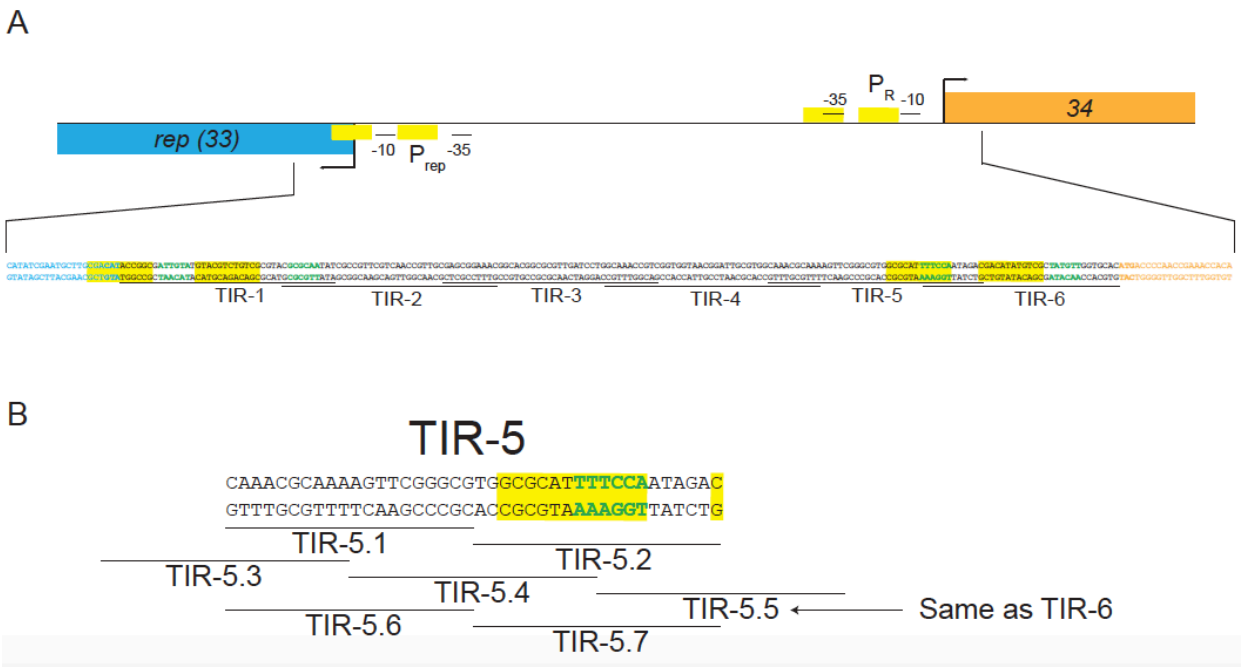

Figure 17 - TIR series of substrates in the context of the full *33-34* intergenic region

(A) The sequences and relative positions of the TIR series of substrates within the *33-34* intergenic region are shown. (B) The sequences of the derivatives of the TIR-5 substrate are shown.

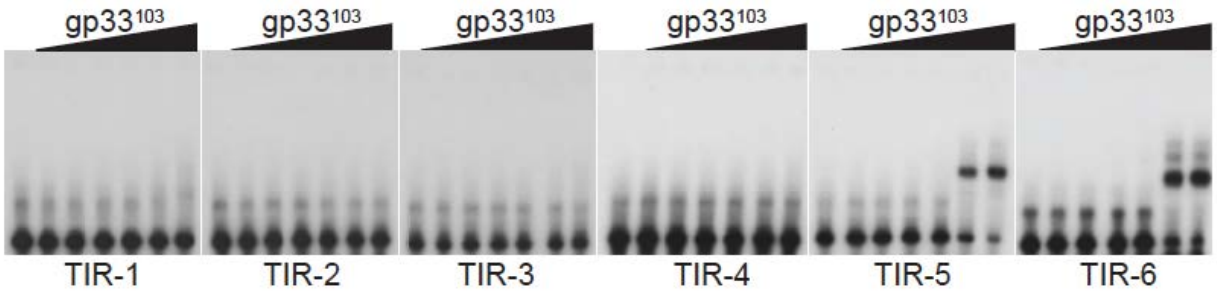

Figure 18 - gp33[103] binding to the TIR series substrates

EMSAs showing gp33[103] binding to 40 bp substrates spanning the *33-34* intergenic region (see Figure 17). Binding occurs to the TIR-6 fragment containing the 12bp O_R ($O_{R\text{-}R}$) and to substrate TIR-5, which contains Region 3 and Region 4 ($O_{R\text{-}L}$) and only one base pair of $O_{R\text{-}R}$. Protein concentrations are as follows: 1) none 2) 0.16 μM 3) 0.54 μM, 4) 1.6 μM, 5) 5.4 μM, 6) 16 μM, 7) 54 μM. See Table 3 for substrate sequences and Table 4 for binding affinities.

Interestingly, the area we do see binding to (apart from TIR-6, which contains the 12bp O_R) is the TIR-5 substrate that is adjacent to, but does not include, the 12bp O_R (**Figure 18**). This substrate contains one base pair from the O_R site and 39 base pairs of sequence to the left of O_R, including Region 3 and Region 4 from DNase I footprinting. The binding seen is not very strong, however it is present. To narrow down where within this TIR-5 substrate gp33[103] was binding to, we created seven new double-stranded DNA substrates (20bp) spanning the TIR-5 sequence (**Figure 17B**). The results of doing EMSAs on these substrates showed that gp33[103] binds to TIR-5.2, which contains the 19bp directly adjacent to the 12bp O_R (**Figure 19**).

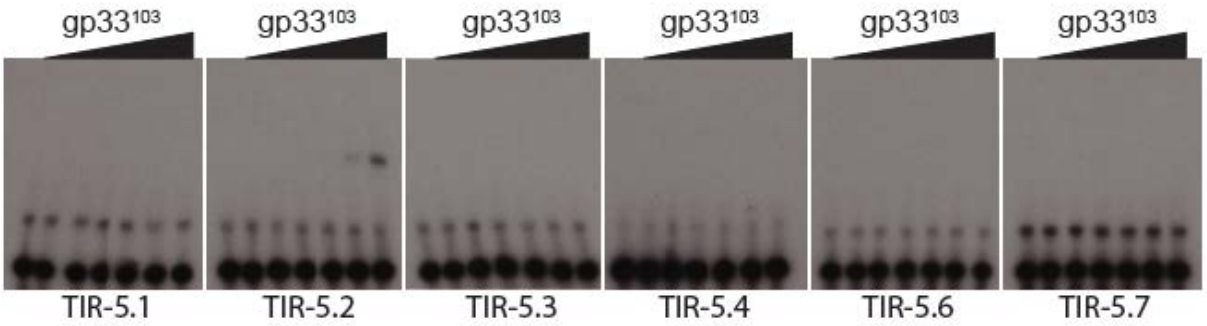

Figure 19 - gp33[103] binding to substrates within TIR-5

EMSAs showing gp33[103] binding to 20 bp substrates spanning the TIR-5 substrate (see **Figure 17B**). Binding occurs to the TIR-5.2 fragment containing regions 3 and 4 from DNase I footprinting and what we predict to be $O_{R\text{-}L}$. Protein concentrations are as follows: 1) none 2) 0.16 μM 3) 0.54 μM, 4) 1.6 μM, 5) 5.4 μM, 6) 16 μM, 7) 54 μM. See Table 3 for substrate sequences and Table 4 for binding affinities.

3.3.2.2 End-Deletion Series of Substrates

We wished to determine whether the O_{Rep} half-site present near the P_{rep} promoter was indeed an area where gp33[103] bound on the left half of the *33-34* intergenic region as the footprinting results suggested so we created a series of substrates spanning the *33-34* intergenic region that gradually and sequentially removed ~40bp increments of sequence from each end of the substrate to be used in EMSAs (**Figure 20**). We created these substrates two ways, first with sequence being removed from the left (or alternatively viewed as added in from the right side of the intergenic region - substrates Mt2-Mt7 – **Table 2**) and second with sequence being removed from the right (or added from the left - substrates Mt8-Mt13 – **Table 2**). Our hope with the first set of substrates was to determine at which sequence addition gp33[103] was able to form more complicated binding patterns to draw conclusions as to the nature of these complexes. The second set of substrates were also designed to allow us to draw conclusions on the nature of gp33[103] complexes, but also to help pinpoint more exactly the location of the second gp33[103] binding site, O_{Rep}.

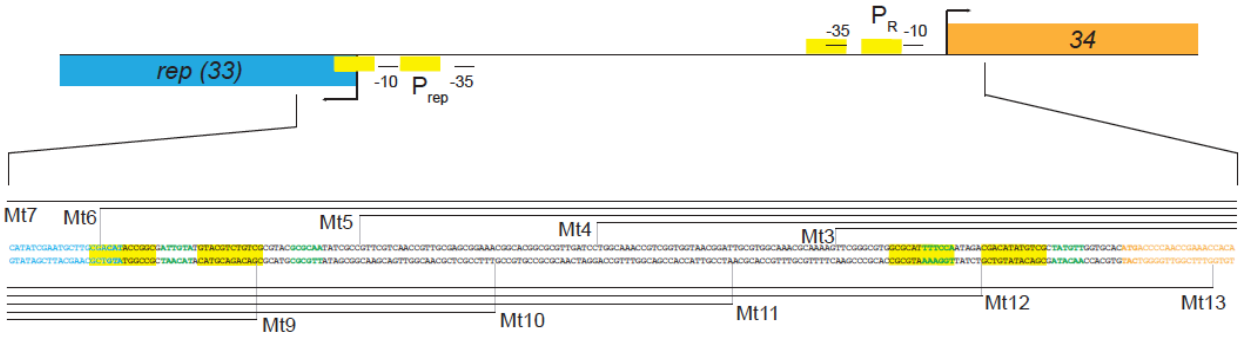

Figure 20 - End-deletion series of substrates

The sequences and relative positions of the end-deletion series of substrates within the *33-34* intergenic region are shown.

We used each of these substrates in EMSAs (**Figure 21**). Our results when sequence is sequentially added from the right side of the *33-34* intergenic region shows that binding begins when sequence containing the 12bp O_R site is added (Mt3 substrate). The binding profile is that of protein binding a single site and causing a single mobility shift. Substrate Mt4 also has a relatively simple binding pattern and both Mt3 and Mt4 contain all of O_R. However, the relative mobilities (R_f) of the major complexes differ (0.69 and 0.55 for Mt6 and Mt5 respectively). Although many factors can influence complex mobility, we propose that the stoichiometries of both are the same, but that there is a protein-induced DNA distortion influencing complex mobility. As more sequence is gradually introduced onto the substrate, we see an increase in the number of complexes formed until finally once the entire sequence of the *33-34* intergenic region is included, we see the four complexes that are typically formed in the binding profile within this region (as measured by R_f).

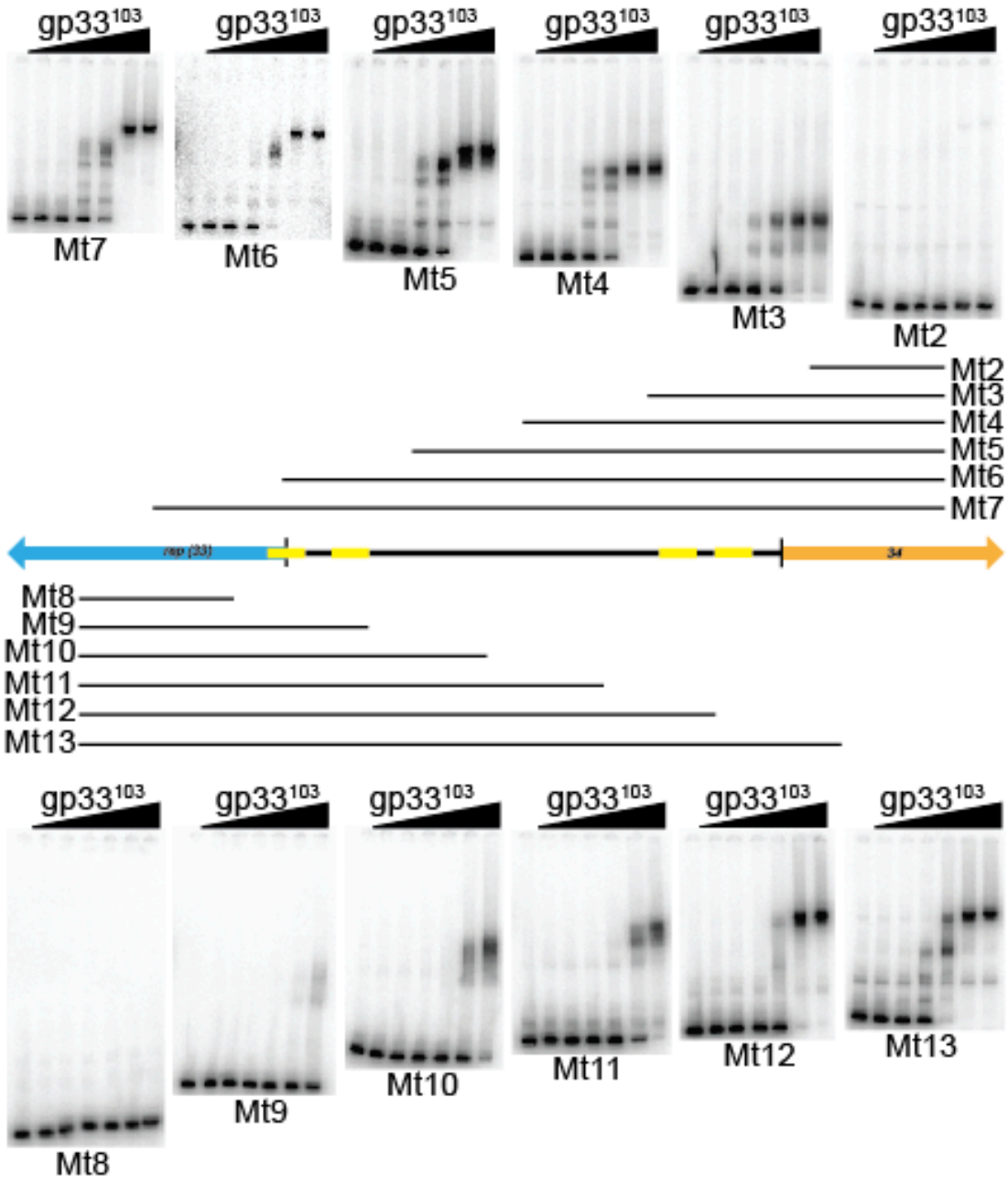

Figure 21 - gp33[103] binding to end-deletion series of substrates

BPs gp33[103] binds with variable affinities to a series of end-deletion substrates. Substrates were PCR amplified from the BPs *33-34* intergenic region using primers located at regular intervals from each end to sequentially shorten the *33-34* intergenic region from each end (Table 3). The sequences included in each substrate are shown as horizontal black lines either above or below a schematic representation of the *33-34* intergenic region. Genes *33* and *34* are represented by blue and orange arrows respectively, and O_R and O_{Rep} are shown in yellow. Protein concentrations are as follows: 1) none 2) 0.16 μM 3) 0.54 μM, 4) 1.6 μM, 5) 5.4 μM, 6) 16 μM, 7) 54 μM. Protein affinities are shown in Table 4.

Our results when sequence is added in from the left (**Figure 21**) show that once the O_{Rep} 6bp half-site is included in the sequence (Mt9), $gp33^{103}$ is still not binding. It is not until the subsequent 40bp of sequence are added to the substrate (Mt10), which include the adjacent Region 2 of DNase I protection (section 3.3.1) that we see $gp33^{103}$ beginning to bind. Regardless of the absence of O_R, binding to substrates Mt10 and Mt11 is relatively weak, and binding to this *33*-promixal part of the site presumably also contributes to complexes formed with the full-length *33-34* intergenic region. It is not until substrate Mt12 that we see $gp33^{103}$ binding at a similar affinity to wild type with the typical complexes being formed (as determined by R_f). Interestingly, the binding patterns seen with the Mt12 and Mt13 substrates are similar, even though the 12bp O_R is present in Mt13 but absent from Mt12. The affinity for Mt12 is slightly reduced compared to that of Mt13 (**Table 4**), but we assume that the slowest migrating complexes (C4) have similar stoichiometries, with each containing $gp33^{103}$ bound at O_R or the areas encompassing Region 3 and Region 4 from DNase I footprinting, in addition to binding elsewhere in the substrate. Also of note is that one of the Mt13 complexes (C3) appears to be absent from the Mt12 substrate, and probably requires the 12bp O_R.

These results showed that $gp33^{103}$ binds to both the *33*- and *34*- proximal regions of the *33-34* intergenic region, but the evidence also strongly suggests that the operator O_{Rep} stretches beyond the 6bp half-site into a full-sized operator, even though there is not much sequence similarity to the full palindrome, and that the 12bp O_R operator also stretches further to the left.

3.3.2.3 Evidence leads to new theory on operator sequences

Because of the DNase I protection patterns shown in **Figure 13**, and the results of the assays done with the end deletion and TIR series of substrates (**Figures 18, 19, 21**), we propose that the $gp33^{103}$ binding site at P_R spans a larger region than just the 12 bp palindrome, and that

this palindrome is equivalent to a half site, of which Region 4 (and perhaps part of Region 3) constitute the other half site, not withstanding the sequence dissimilarities (see **Figure 22** - all substrates together and operator sites marked clearly). We propose the sequence 5'-GCGCATTTTCCA for $O_{R\text{-}L}$ which retains five bases of the 12 bp palindrome (**Figure 22**). To simplify the presentation and discussion of data from here on out, we will refer to the 12 bp palindrome as $O_{R\text{-}R}$ and the region to the left that includes Region 4 as $O_{R\text{-}L}$ (**Figure 22**) reflecting the right and left half sites of O_R respectively. The TIR-5 and TIR-5.2 substrates contain $O_{R\text{-}L}$ but do not contain $O_{R\text{-}R}$ (**Figure 22**). A plausible explanation is that gp33[103] binds as a tetramer to $O_{R\text{-}L}$ and $O_{R\text{-}R}$ independently and at reduced affinity, despite the sequence differences. In this model, the tetramer would also have two unoccupied helix-turn-helix DNA binding domains, which would be available for binding either to another site within the same substrate (perhaps O_{Rep}) with introduction of a DNA loop, or by forming intramolecular bridges between two different DNA molecules.

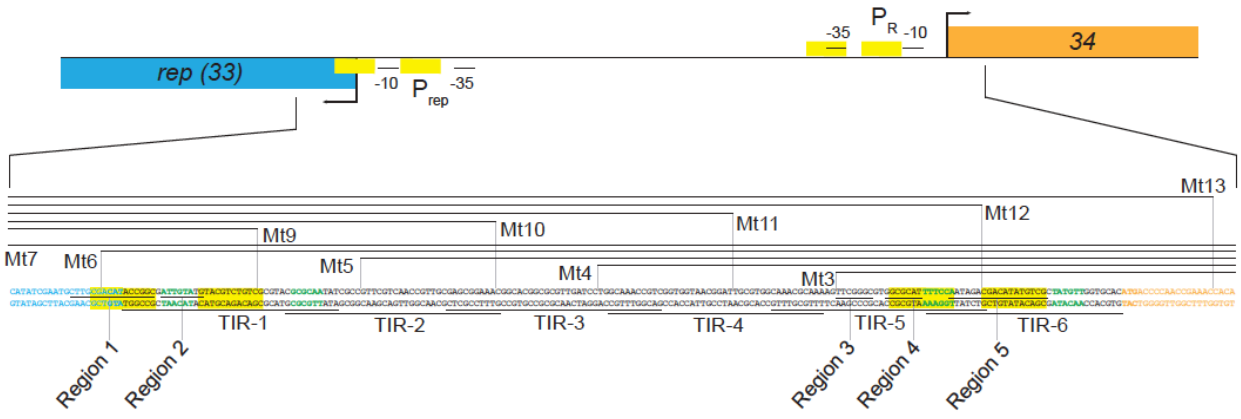

Figure 22 - Locations of predicted complete operators in the *33-34* intergenic region

The locations of the left and right half sites of O_R and O_{Rep} are shown, along with the sequences of all substrates used that led to this hypothesis.

3.3.3 Binding to partial sites

As previously mentioned in section 2.3.1.2, BPs gp33[103] is known to bind to several additional sites within the BPs genomes (97). We decided to investigate whether there are similar complexities to the binding of gp33[103] at these sites. The three other instances of sequences identical to $O_{R\text{-}R}$ are located in small intergenic regions, between genes *5* and *6* (O_6), between genes *26* and *27* (O_{27}), and between genes *54* and *55* (O_{55}) (**Figure 23**). We also found a site with two base pair differences that is located within the *60-61* intergenic region (O_{61}) (**Figure 23**).

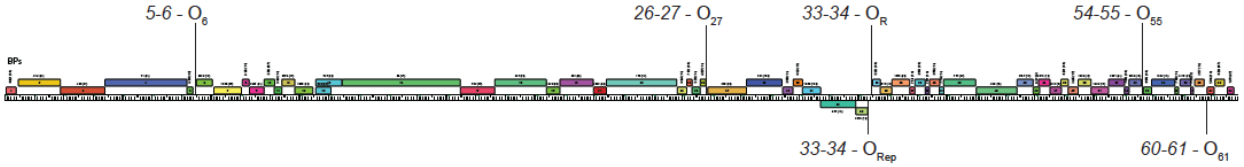

Figure 23 - Map of BPs genome with operators marked

The BPs genome is shown with black lines indicating positions of operators.

To test whether these regions play a role in the phage transcriptional program and are subject to repressor control, each was inserted upstream of an *mcherry* reporter gene and transformed into *M. smegmatis* mc[2]155. The *5-6*, *26-27*, *54-55*, and *60-61* regions all have promoter activity (P_6, P_{27}, P_{55}, and P_{61}), although the strengths vary, with P_{55} being the most active, and P_{27} the weakest (103). Putative promoter elements containing -10 and -35 hexameric motifs are predicted for P_6, P_{55}, and P_{61}, but not confidently for P_{27} (**Figure 24**). The promoter-reporter plasmids were also transformed into a BPs lysogen and the promoter activities

determined. The P_6, P_{27}, and P_{55} promoters are clearly down regulated in a lysogen, presumably by gp33[103], as is P_R (97). No regulation for either P_{rep} or P_{61} was observed.

In view of the hypothesis that the 12 bp palindrome at $O_{R\text{-}R}$ corresponds to a half-site for gp33[103] binding, we examined the other intergenic region sequences for additional sites related to $O_{R\text{-}R}$. Interestingly, in all four regions we identified different, but related, sequences that are positioned either 5 bp (O_6, O_{27}, O_{61}), or 8 bp (O_{55}) away from the other half site, and that differed from the palindrome by 2-5 base pairs. This finding was consistent with the 12 bp palindrome constituting just one of two half sites. From here on out we will refer to the left and right halves of each of these as $O_{6\text{-}L}$ and $O_{6\text{-}R}$, $O_{27\text{-}L}$ and $O_{27\text{-}R}$, $O_{55\text{-}L}$ and $O_{55\text{-}R}$, and $O_{61\text{-}L}$ and $O_{61\text{-}R}$ for O_6, O_{27}, O_{55} and O_{61} respectively (**Figure 24**).

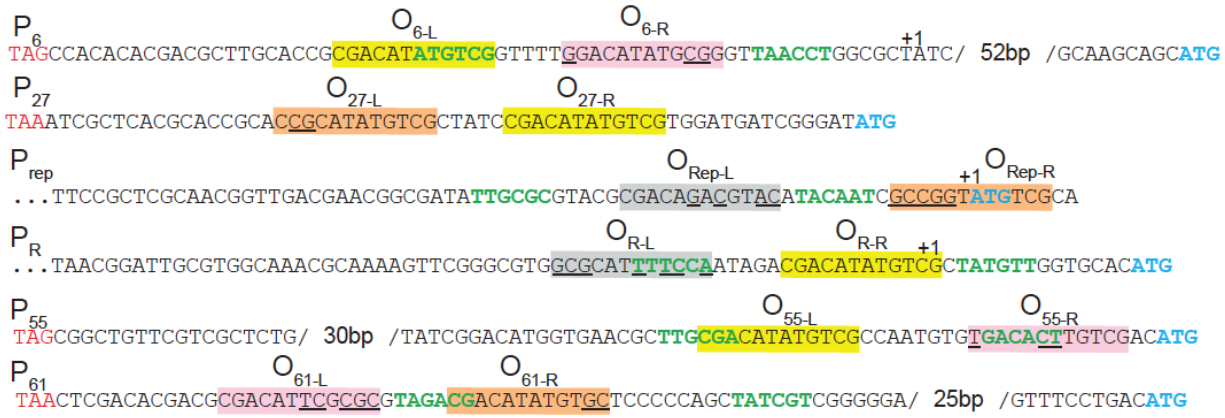

Figure 24 - Sequences of all BPs operators within their respective promoter regions

Sequences of the intergenic regions containing promoters with operators are shown. Operator half sites are labeled. Half sites highlighted in yellow indicate full sequence identity to the 12bp palindrome. Sites highlighted in orange are half sites that are not full palindromes, but still show an interaction with gp33[103]. Sites highlighted in pink do not bind gp33[103] alone. Sites highlighted in gray have not been tested.

To explore each of these sequences in more detail, we created two types of substrates for each operator. The first had the half site containing the 12bp palindrome intact with the second

half site mutationally ablated to have no sequence identity with the palindrome, which theoretically would prevent gp33^{103} from binding to that half site as it normally would. The second had the non-palindromic half site separated from its native sequence to see if gp33^{103} binds it without the full palindrome present on the same substrate. For reference, binding assays done with O_R and O_{Rep} are shown in **Figure 25**, however these particular substrates were created before the theory of operators consisting of two half sites and are therefore not as informative as the other operator substrates.

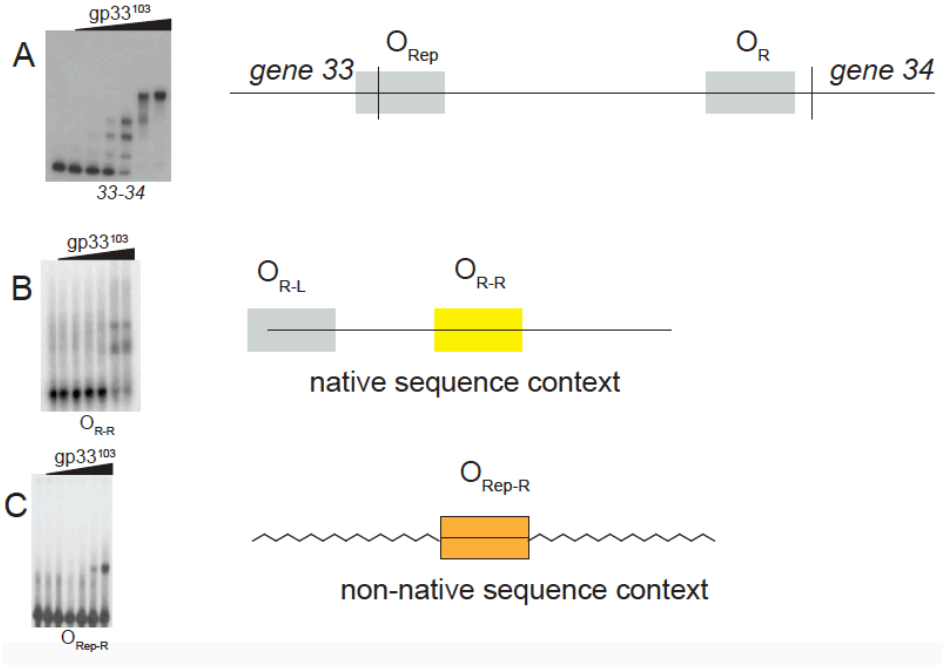

Figure 25 - Substrates testing O_{Rep} and O_R

EMSAs show binding profiles of gp33^{103} to the substrates shown in the diagrams. Protein concentrations were as follows: 1) none 2) 0.16 µM 3) 0.54 µM, 4) 1.6 µM, 5) 5.4 µM, 6) 16 µM, 7) 54 µM.

Gp33^{103} binds within the *5-6* intergenic region where studies by Oldfield *et al.* (97, 103) have established the presence of a late lytic promoter. This promoter is downregulated in a lysogen, so we know that gp33^{103} plays a role in this region during lysogeny. Along with a full 12bp palindromic half site ($O_{6\text{-}L}$), we found a partial binding site in this region that differs from

the palindrome by 3 base pairs ($O_{6\text{-}R}$) (**Figure 24**). When $gp33^{103}$ binds to a 45bp substrate where $O_{6\text{-}R}$ is mutationally ablated, a single protein-DNA complex is formed (**Figure 26B**). A binding assay done on this $O_{6\text{-}R}$ site while isolated from its native sequence by placing it in a non-native sequence context shows that there is no binding (**Figure 26C**, **Table 3**). This shows us that some or all of the three base pairs that differ from the palindrome are necessary for binding.

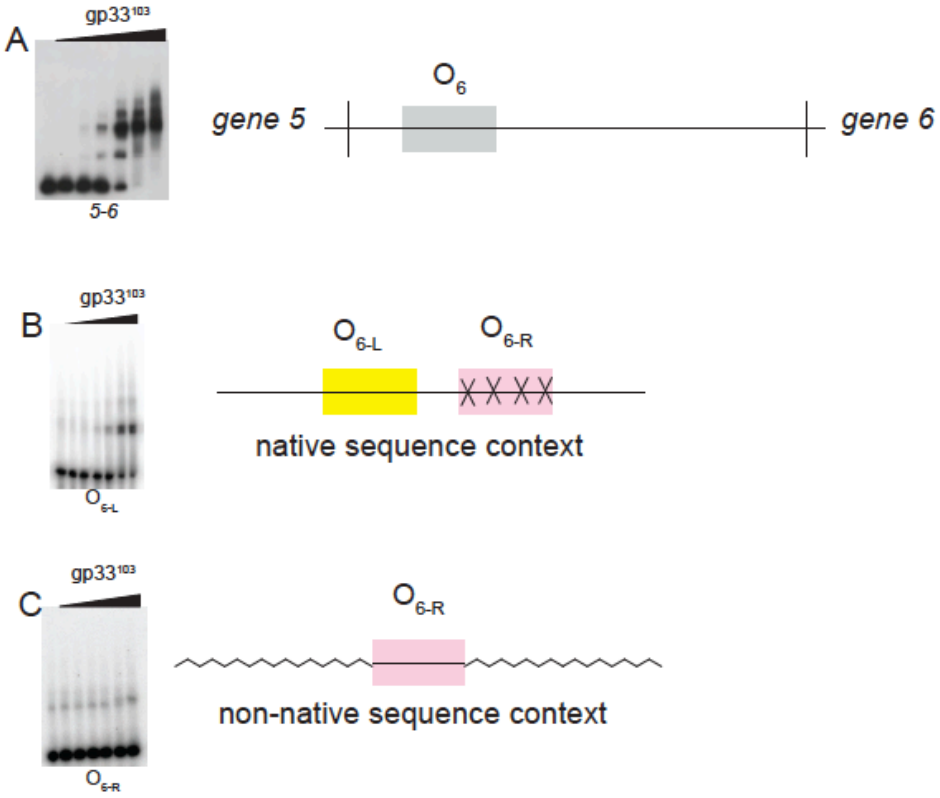

Figure 26 - $gp33^{103}$ binding to the O_6 operator

EMSAs show binding profiles of $gp33^{103}$ to the substrates shown in the diagrams. $Gp33^{103}$ binds to $O_{6\text{-}L}$, but not $O_{6\text{-}R}$ alone. Protein concentrations were as follows: 1) none 2) 0.16 µM 3) 0.54 µM, 4) 1.6 µM, 5) 5.4 µM, 6) 16 µM, 7) 54 µM.

Gp33[103] binds within the *26-27* intergenic region (**Figure 27A**) where studies by Broussard *et al.* predicted that no promoter was present (**Figure 24**), but since this intergenic region lies upstream of the lysis cassette, we theorized that gp33[103] may still play a role in preventing read-through transcription. In her doctoral dissertation, Lauren Oldfield showed that the *26-27* intergenic region does have very slight promoter activity, and that this weak promoter is downregulated in a lysogen, however no promoter sequences have been identified. This evidence suggests that gp33[103] indeed plays a role in this region during lysogeny.

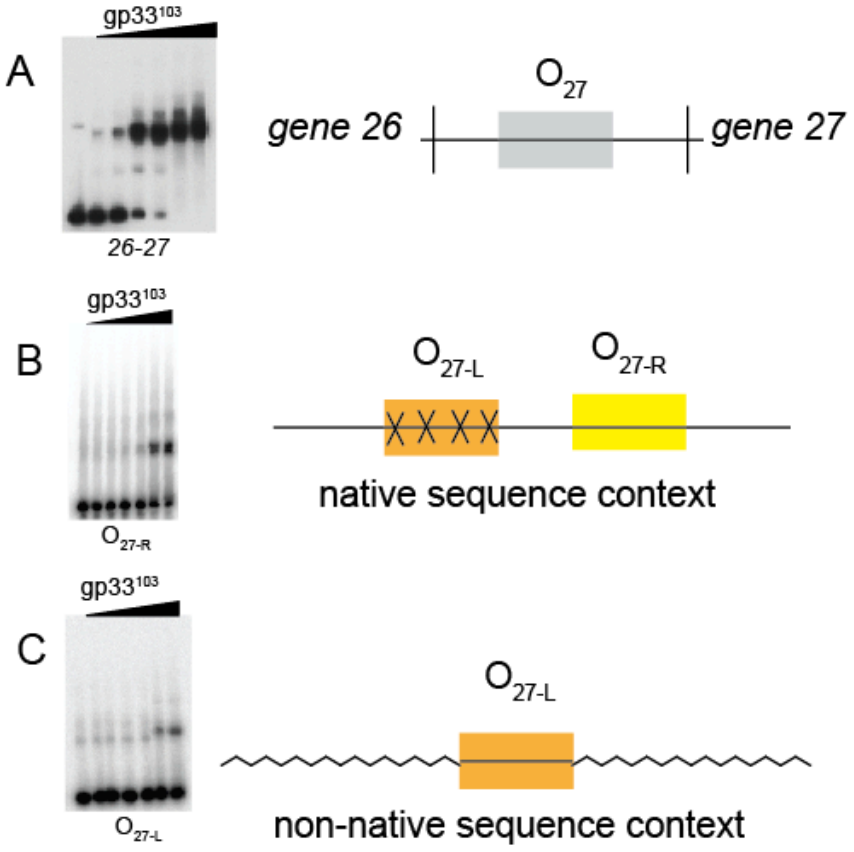

Figure 27 - gp33[103] binding to O_{27}

EMSAs show binding profiles of gp33[103] to the substrates shown in the diagrams. Gp33[103] binds to both $O_{27\text{-}L}$, and $O_{27\text{-}R}$ alone. Protein concentrations were as follows: 1) none 2) 0.16 µM 3) 0.54 µM, 4) 1.6 µM, 5) 5.4 µM, 6) 16 µM, 7) 54 µM.

Along with a full 12bp palindromic half site ($O_{27\text{-}R}$), we found a partial binding site in this region that differs from the palindrome by 2 base pairs ($O_{27\text{-}L}$) (**Figure 24**). When $gp33^{103}$ binds to a 45bp substrate of native 26-27 intergenic region sequence where $O_{27\text{-}L}$ is mutationally ablated, a single protein-DNA complex is formed (**Figure 27B**). A binding assay done on this $O_{27\text{-}L}$ site while isolated from its native sequence by placing it in a non-native sequence context shows that there is slight binding (**Figure 27C; Table 3**), which could explain why the binding affinity of $gp33^{103}$ to the *26-27* intergenic region is stronger than to any other intergenic region tested.

If $gp33^{103}$ binds strongly to both half sites rather than just one as the binding assays suggest, perhaps the role of $gp33^{103}$ in this region is more important than in other operators. This shows us that the two base pairs that differ from the palindrome are not as necessary for binding as other bases.

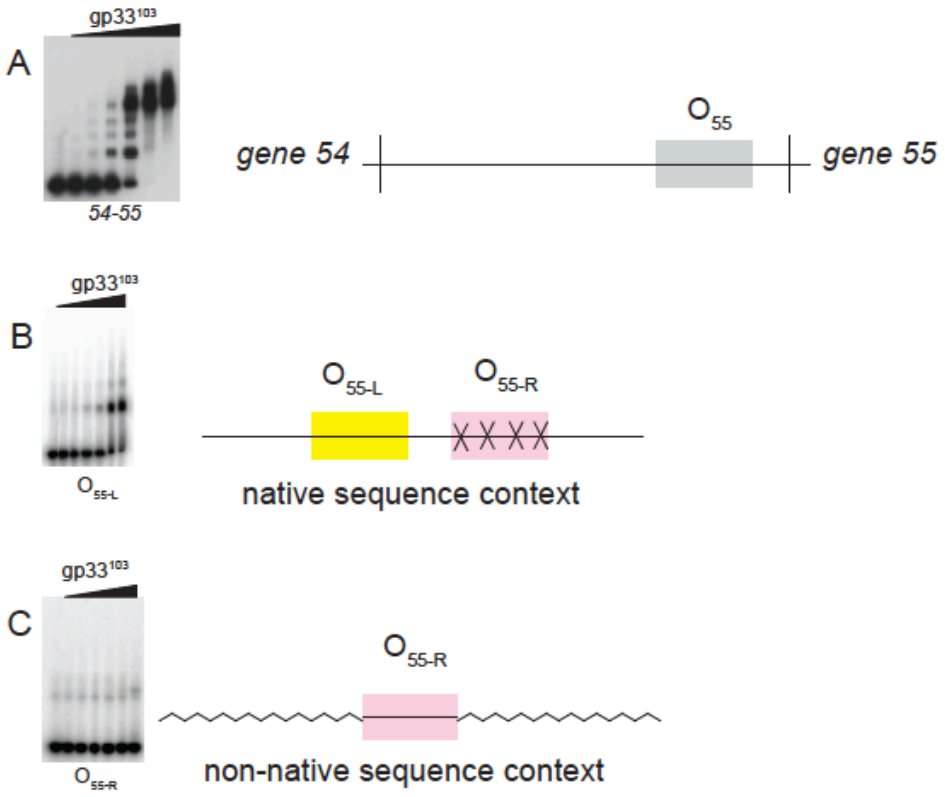

Figure 28 - gp33^{103} binding to O_{55}

EMSAs show binding profiles of gp33^{103} to the substrates shown in the diagrams. Gp33^{103} binds to $O_{55\text{-}L}$, but not $O_{55\text{-}R}$ alone. Protein concentrations were as follows: 1) none 2) 0.16 µM 3) 0.54 µM, 4) 1.6 µM, 5) 5.4 µM, 6) 16 µM, 7) 54 µM.

When gp33^{103} binds within the *54-55* intergenic region where studies by Oldfield *et al.* (97, 103) have established the presence of a lytic promoter (**Figure 24**), a larger number of complexes are formed than with any of the other non-P_R intergenic regions (**Figure 28A**). This promoter is downregulated in a lysogen, so we know that gp33^{103} plays a role in this region during lysogeny. Along with a full 12bp palindromic half site ($O_{55\text{-}L}$), we found a partial binding site in this region that differs from the palindrome by 3 base pairs ($O_{55\text{-}R}$). When gp33^{103} binds to a 45bp substrate where $O_{55\text{-}R}$ is mutationally ablated, a single protein-DNA complex is formed (**Figure 28B**). A binding assay done on this $O_{55\text{-}R}$ site while isolated from its native sequence by

placing it in a non-native sequence context shows that there is no binding (**Figure 28C**, **Table 3**). This shows us that some or all of the three base pairs that differ from the palindrome are necessary for binding; however these results don't help us understand why there are more complexes in the binding profile of gp33[103] within the *54-55* intergenic region than there are in other intergenic regions.

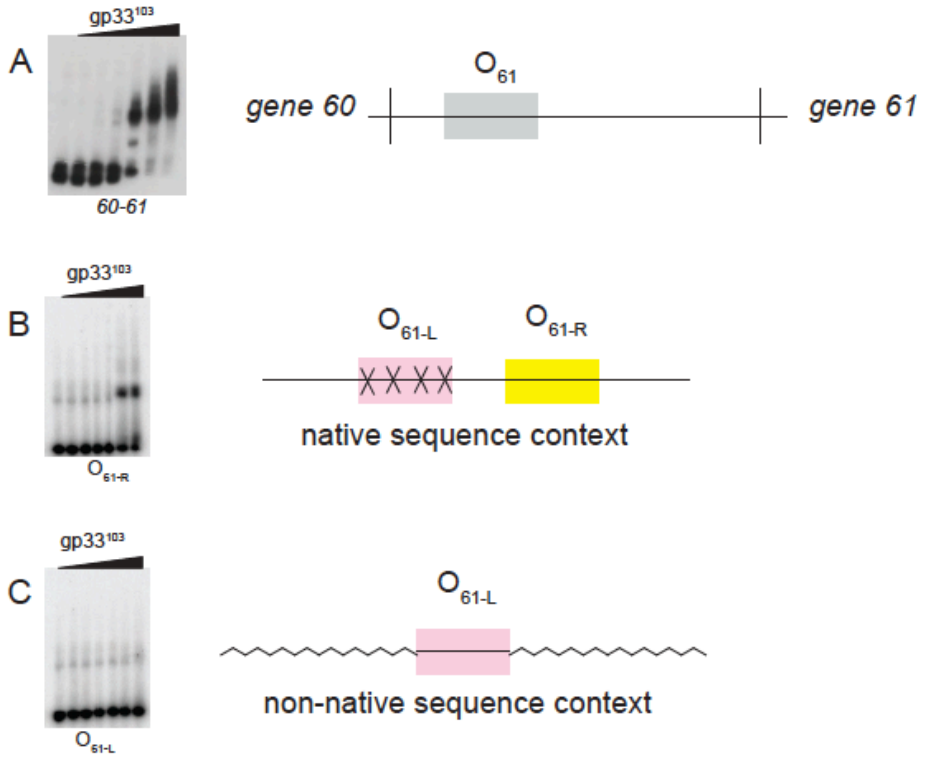

Figure 29 - gp33[103] binding to O_{61}

EMSAs show binding profiles of gp33[103] to the substrates shown in the diagrams. Gp33[103] binds to $O_{61\text{-}R}$, but not $O_{61\text{-}L}$ alone. Protein concentrations were as follows: 1) none 2) 0.16 µM 3) 0.54 µM, 4) 1.6 µM, 5) 5.4 µM, 6) 16 µM, 7) 54 µM.

The *60-61* intergenic region contains a promoter for a three gene operon (103). Within the predicted promoter region we found two operator half sites, one with 2bp differences from

the palindrome ($O_{61\text{-}R}$) and one with 5bp differences ($O_{61\text{-}L}$) (**Figure 24**). When we tested $gp33^{103}$ binding to the two half sites only $O_{61\text{-}R}$ showed binding (**Figure 29B**). Both partial sites contain the same final two base pair substitutions and the left half site also contains three other base pair substitutions. This accumulation of differences from the 12bp palindromic sequence is probably why there is no binding to $O_{61\text{-}L}$ (**Figure 29C**). The right half site, however, has a binding profile similar to that of other fully palindromic operator half sites and $gp33^{103}$ binds to this site with a comparable affinity to those sites (**Table 4**), suggesting that the two base pair substitutions are not very important for binding. Despite the promising binding results and the presence of a promoter near the O_{61} operator, fluorescence reporter assays have shown that the P_{61} promoter is not downregulated in a lysogen, so although $gp33^{103}$ can bind to O_{61}, it may not be exerting an effect at that location *in vivo*. It is entirely possible, however, that if $gp33^{103}$ is capable of preventing read-through transcription, that it could do so at this operator.

3.3.4 Clear plaque mutant analysis

In the studies done by Broussard *et al.*, multiple repressor-insensitive and clear-plaque phage mutants were discovered and sequenced. Sequencing results showed that all of these mutants had either mutations within the repressor gene, *gene 34*, or mutations within the *33-34* intergenic region (**Figure 30 – Table 3**). This evidence led to the hypothesis that the *gene 32, 33, 34* cassette were the only genes involved in regulating lysogeny (97).

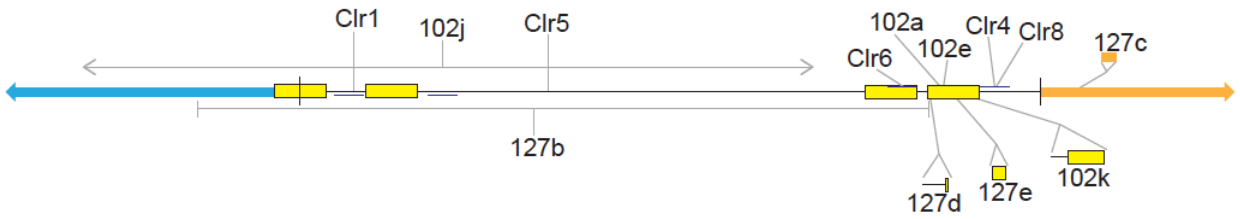

Figure 30 - Locations of BPs repressor insensitive mutations that map to the *33-34* intergenic region

Locations of the mutations in each substrate (see Table 3).

Most of the mutations mapped within the repressor gene itself (97), which is not surprising since the repressor is necessary for maintenance of lysogeny. For the mutations that mapped within the *33-34* intergenic region, Clr1 has a T29336C base pair substitution in the -10 element of P_{rep} which strongly reduces the promoter's activity, and without proper expression of the repressor gene, lysogeny cannot occur. A binding assay done on this mutant shows a similar binding profile as wild type (**Figure 31**). Clr4, Clr6, and Clr8 have base pair substitutions in the -10 or -35 promoter elements of P_R. The Clr4 and Clr8 base pair substitutions have been shown to lead to an upregulation in promoter activity, however the Clr6 mutation does not (97, 102). With P_R upregulated in the Clr4 and Clr8 mutants, the genetic switch is pushed toward lytic growth. The Clr6 point mutation is, however, located within $O_{R\text{-}L}$ and could influence $gp33^{103}$ activity within O_R. Binding assays done with these three mutants show that Clr 8 has a binding profile similar to wild type, Clr6 has an early appearance of the C1 complex compared to wild type but all other complexes appear at similar protein concentrations, and Clr 4 has an earlier appearance of the C4 major complex (**Figure 31**).

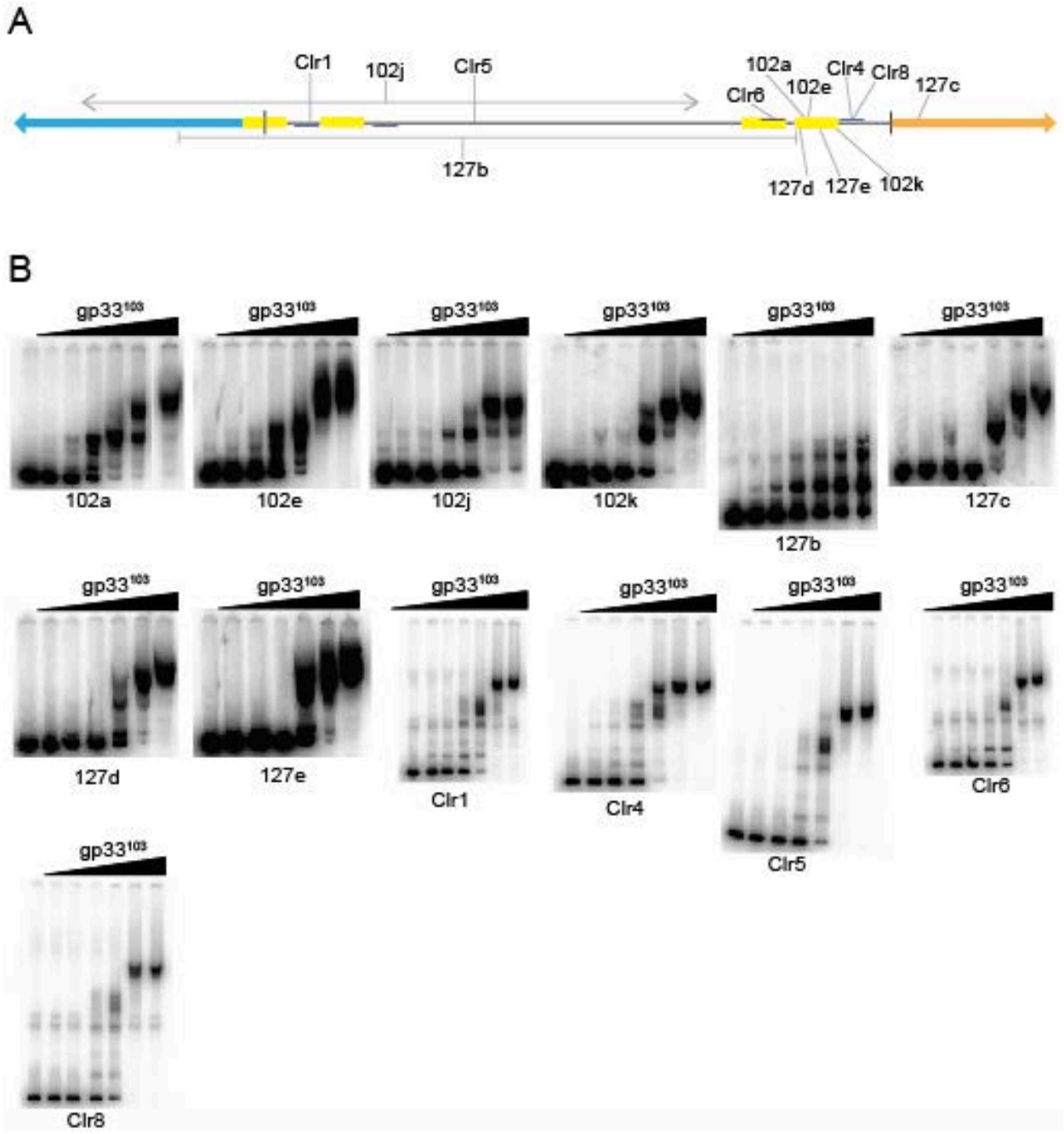

Figure 31 - gp33[103] binding to substrate variants of the *33-34* intergenic region derived from repressor-insensitive mutants

(A) Locations of the mutations in each substrate (see Table 3). (B) Binding profiles of gp33[103] to each *33-34* substrate. The concentrations of protein are as follows: 1) none, 2) 0.16μM, 3) 0.54μM, 4) 1.6μM, 5) 5.4μM, 6) 16μM, 7) 54μM. Protein affinities are shown in Table 4.

Two mutants described in section 2.3.2.2 (102a and 102e) were shown to have point mutations in $O_{R\text{-}R}$ that reduce the affinity of $gp33^{103}$ by about 3-10-fold (depending on specific DNA substrate used), which allows expression of the early lytic genes that outcompete the repressor and lead to lytic growth (**Figures 7, 31**). Six other mutants have more substantial sequence rearrangements. Mutant 102j has a 151bp inversion that flips the direction of P_{rep} and also removes the 5' end of the repressor gene from the 3' end. With the repressor gene interrupted and P_{rep} facing the opposite direction, lysogeny cannot occur because gp32 is not expressed, and no functional repressor can be made. The binding profile of $gp33^{103}$ to 102j shows similar binding affinity as to a wild type substrate, but only complexes C3 and C4 are formed (**Figure 31**). Mutant 127b has a 151bp deletion that removes the 5' end of the repressor gene, P_{rep}, and $O_{R\text{-}L}$. Without the P_{rep} or the 5' end of the repressor gene, gp32 cannot be made and no functional repressor can be synthesized. The binding profile of $gp33^{103}$ to this mutant reflects the removal of O_{Rep} and $O_{R\text{-}L}$, as C1 becomes the major complex and a moderate amount of C2 complexes are made (**Figure 31**). 127d, 127e, and 102k have duplications within $O_{R\text{-}R}$, which presumably have an effect on $gp33^{103}$ binding and affect P_R repression in a lysogen. These mutants have similar binding profiles as wild type, however some of the faster-moving complexes (C1 and C2) are missing (**Figure 31**). Mutant 127c has a duplication within the *gene 34* open reading frame that probably affects the proteins' function and could allow it to more strongly counteract the repressor's actions. The binding profile of $gp33^{103}$ to this mutant shows a wild type binding profile (**Figure 31**). Clr 5 is the hardest mutant to explain. It is a single base pair substitution in the central part of the *33-34* intergenic region that does not interfere with any promoter or operator sequences. Binding assays done on this mutant shows a binding profile similar to wild type (**Figure 31**). The effect this mutation has on the genetic switch is unclear.

3.4 DISCUSSION

In this chapter we have discussed the experiments that led to identification of the true operator sequence being composed of two 12bp half sites separated by either 5bp or 8bp, with a 5bp spacer being the most common. DNase I footprinting assays showed areas of protection near the P_{rep} promoter, and the P_R promoter. Interestingly, the footprinting showed a large area of protection extending toward the center of the *33-34* intergenic region where we would not have predicted the protein bound. Binding assays with the TIR series of substrates confirmed that $gp33^{103}$ does indeed bind to DNA to the left of O_R, and the end-deletion series of substrates showed that inclusion of the area to the left of O_R led to a similar binding phenotype as if O_R itself were present (Mt12). In light of these data we proposed that O_R is not just a 12bp site, but that it extended further into the *33-34* intergenic region.

We did not hypothesize the presence of a second half site until we tested the data of the other intergenic regions (*5-6, 26-27, 54-55*) and their operators (O_6, O_{27}, O_{55}). In each of these regions we found a second partial site that contained 2-3bp departures from the palindrome. We also found two partial sites in the *60-61* intergenic region and found the operator O_{61}. Although O_R and O_{Rep} did not have sites with such similar sequence identity (>75% identical to palindrome), we did find some similar sequences. The new O_R and O_{Rep} are annotated correctly in the diagram in Figure 32. How this knowledge allows us to better understand how $gp33^{103}$ functions is the subject of the next chapter.

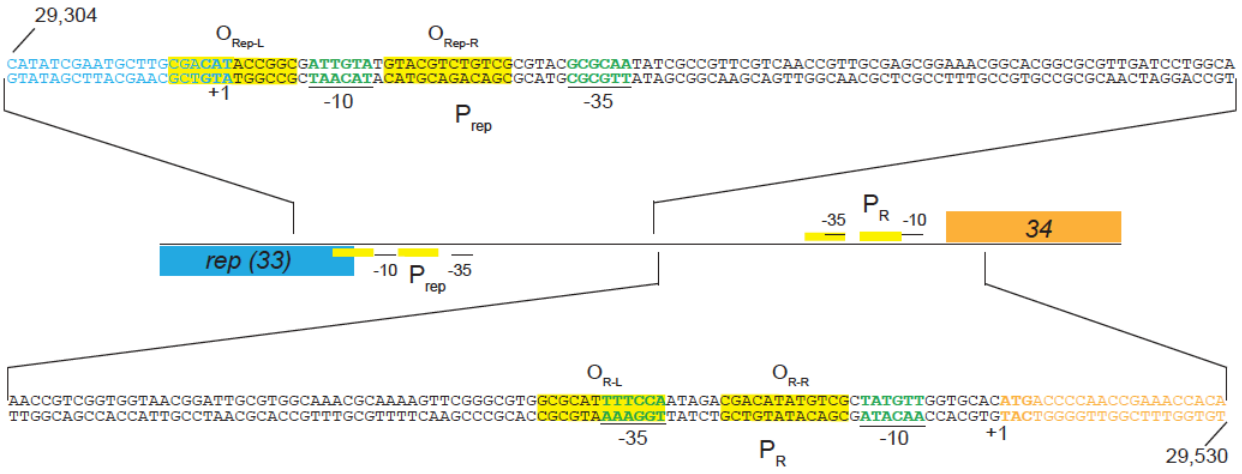

Figure 32 - Locations of O_{Rep} and O_R relative to P_{rep} and P_R

Diagram showing the sequence of the *33-34* intergenic region. Operators are labeled and half sites are highlighted in yellow. Promoter sequences are in green text with transcription start sites indicated with +1.

4.0 HOW GP33^{103} FUNCTIONS TO PREVENT TRANSCRIPTION OF LYTIC GENES

Work done in this section was published in the following publication:

Villanueva, V.M., Hatfull, G.F., September 2015. An Unusual Phage Repressor Encoded by Mycobacteriophage BPs. *PLoS One*, Volume 10:e0137187.

4.1 INTRODUCTION

Repressor proteins can function in a variety of ways. The first and most obvious way a repressor can block lytic gene expression is by promoter occlusion. The presence of repressor binding to its operator(s) within the promoter elements sterically prevents the bacterial RNA polymerase from binding (177, 178). Another way is by binding within or near a promoter and changing the secondary structure of the DNA such that the RNA polymerase cannot bind, or if it does, the secondary structure does not let the polymerase proceed to elongate the transcript (134, 136). A third way is with direct interaction with RNA polymerase to either prevent promoter recognition or prevent transcript elongation (179). In this section we describe the experiments done to identify how the repressor exerts its function. We show that gp33^{103} likely exerts its effect via

DNA bending and that the spacer region between operator half-sites is important for proper bending.

4.2 MATERIALS AND METHODS

4.2.1 DNA binding assays

Electrophoretic mobility shift assays were performed as described previously in section 2.3.2 with minor changes. Substrates resulting from PCR or synthetic oligonucleotides were prepared as previously described. Substrates resulting from digested DNAs were first treated with calf intestinal phosphatase (NEB) to remove 5' phosphates and then run on a 0.8% agarose gel and extracted using a GeneJet gel extraction kit from Fermentas according to manufacturer recommendations. The concentration of these substrates and subsequent radiolabeling and use in EMSAs were performed as previously described. The substrates used in this chapter are listed below in Table 5.

Table 5 - Substrates used in Chapter 4

DNA Substrate Name	Sequence/Primers used
33-34 intergenic region (226bp)	Fwd: TCCGGCCCATATCGAATGCTTGCGACAT Rev: CGGTTGGGGTCATGTGCACCAACAT
33-34 intergenic region (366bp)	Fwd: GTTCTTCTTGCTTCAAACCAGCTTCAAGG Rev: GTGTAGTGATCGCTTCGGAATGTCATGC
2 half sites 5bp apart	CACCGACGCGACATATGTCGACTGACGACATATGTCGCAGCCTGG
Left half site ablated (5bp apart)	CACCGACGAAGTGCCACGATACTGACGACATATGTCGCAGCCTGG
Right half site ablated (5bp	CACCGACGCGACATATGTCGACTGAAAGTGCCACGATCAGCCTGG

apart)	
2 half sites 8bp apart	ACCGACGCGACATATGTCGACGCTTGACGACATATGTCGCAGCCT
Left half site ablated (8bp apart)	CCGACGAAGTGCCACGATACGCTTGACGACATATGTCGCAGCCTG
Right half site ablated (8bp apart)	CCGACGCGACATATGTCGACGCTTGAAAGTGCCACGATCAGCCTG

4.2.2 Preparation of pBend2 plasmid for use in bending assays

4.2.2.1 Creation of plasmids

The *33-34* intergenic region was PCR amplified from a BPs lysate using primers 5'-ATA TCG AAT GCT TGC GAC ATT CTA GAG ATT GTA TGT AC -3' and 5'- GTC ATG TGC ACC AAC ATG TCG ACT ATG TCG T 3', which amplified the region with a SalI recognition site at the 5' end of the gene and a XbaI recognition site at the 3' end. This substrate was run on a 0.8% agarose gel to check its purity, digested with SalI and XbaI (Fermentas fast digest enzymes), PCR purified using a GeneJet PCR purification kit from Fermentas, and was ligated using an Epicentre Fast-Link DNA ligase kit into the pBend2 vector (pLC3) that was linearized with SalI and XbaI, phosphatase treated (calf intestinal phosphatase – NEB), and run on a 0.8% agarose gel and gel extracted using a GeneJet gel extraction kit from Fermentas. The resulting plasmid, pVMV21, was transformed into chemically competent NEB5-α cells (New England BioLabs) and colonies were picked into Luria-Bertani Broth (Difco) with Kanomycin selection. Plasmids were extracted from overnight cultures using a GeneJet plasmid miniprep kit from Fermentas according to manufacturer recommendations. The plasmids were sent to GeneWiz for

sequencing according to company instructions and the plasmid with the correct sequence was stored at -20°C.

pVMV22 was created in a similar manner to pVMV21, except that rather than PCR amplification, the insert was a synthetic oligonucleotide engineered with $O_{R\text{-}R}$ in the center of the substrate and SalI and XbaI restriction endonuclease sites flanking the 12bp sequence. Complimentary synthetic oligonucleotides were ordered from IDT and annealed according to the previously described protocol. The double stranded DNA substrate was digested with SalI and XbaI Fermentas fast digest enzymes and gel extracted with a GeneJet gel extraction kit from Fermentas according to manufacturer protocol. This was inserted into a linearized pBend2 vector and the plasmid with correct sequence was stored.

4.2.2.2 Preparation of substrates from pVMV21 and pVMV22

pVMV21 and pVMV22 were subjected to various enzyme cleavage reactions. For each digestion reaction, ~500ng of plasmid DNA was digested with 2 units of the corresponding enzyme. The resulting digested pieces were then treated with phosphatase and gel extracted. These substrates were radiolabeled and used in EMSAs. The restriction enzymes used to create each substrate are: MluI, BglII, NheI, StyI, SpeI, XhoI, EcoRV, SmaI, NRuI, NcoI, BamHI.

4.2.3 RNA Polymerase Purification and Binding Assay Methods

4.2.3.1 Purification of RNA polymerase

The *M. smegmatis* RNA polymerase core enzyme was purified as previously described (180, 181). Briefly, a 1L *M. smegmatis* mc^2155 culture was grown to saturation in 7H9 broth (difco) supplemented with CB/CHX/ADC/Tween80. The pellet was spun down and frozen. The

cell pellet was thawed and 5 g of cells were resuspended in 15 mL Buffer A (50mM Tris, pH 7.9, 0.5 mM EDTA, 50 mM NaCl, 5% glycerol, 0.1 mM DTT) with 260 μg lysozyme per mL and incubated on ice for 20 min. 200 μl of 10% sodium deoxycholate was added and the mixture was incubated on ice for another 10 minutes. The cells were sonicated with a VirSonic sonicater at 30% power output 10 times for 30 sec with 2 minutes of ice in between bursts. 15 mL Buffer A was added to the cells, mixed, and then the cells were centrifuged for 10 minutes at 10,000 rpm. The supernatant was PEI precipitated (182) by adding 900 μl 10% PEI, mixing for ten minutes at 4°C, and centrifuging 10 min at 6,000 rpm. Pellet was washed by adding 20 mL Buffer A + 0.3 M NaCl and resuspended using a handheld Dounce homogenizer, then spun for ten minutes at 6,000 rpm. The polymerse was eluted from the pellet by resuspending 20 mL Buffer A + 1 M NaCl and centrifuging for ten minutes at 6,000rpm. To 20mL supernatant, added 7.6 g ammonium sulfate for an ammonium sulfate precipitation (182). The mixture was incubated at 4°C with rocking for 10 min and then spun down at 6,000 rpm for 10 min. The pellet was dissolved in 20 mL TE (50mM Tris-HCl pH 7.9, 0.5mM EDTA), centrifuged at 6,000 rpm for ten minutes, and the supernatant was collected and kept on ice.

Meanwhile, 2 mL of MAb 8RB13-Sepharose (A kind gift from the Richard Burgess laboratory (183-185)) was washed on a fritted glass funnel with 20 mL TE + 0.15 M NaCl. This was then incubated with the supernatant fluid containing the RNA polymerase and mixed end-over-end for 1hr at 4°C. At room temperature, the mAB 8RB13-Sepharose was collected on the fritted glass funnel, washed with 20 mL TE + 0.15 M NaCl, poured into a disposable BioRad column and washed with 5mL TE + 0.5 M NaCl. The RNA polymerase was eluted with 10 mL TE + 0.75 mM Ammonium sulfate and 40% propylene glycol (this step must be done at room temperature) and fractions were collected. DTT was added to each fraction (0.1 mM final

concentration) and fractions were pooled and dialyzed into storage buffer (Buffer A with 50% glycerol). Purified RNA polymerase core enzyme was stored at -20°C.

4.2.3.2 Purification of SigA and reconstitution of RNA polymerase holoenzyme

Creation of pVMV36

To create pVMV36, The *Mycobacterium smegmatis* s*igA* gene, formerly *mysA*, (125, 186, 187) was PCR amplified using primers 5' – TTT TTT CAT ATG CTG ATC TCT CAG CGT CCC AC – 3' and 5' – TTT TTT AAG CTT AAG CTG CTC GGT CTC GGC GTA GT – 3' from purified wild type *M. smegmatis* mc^2155 DNA, taken with permission from a DNA extraction performed by Zaritza Petrova in the Hatfull lab. The gene was amplified with NdeI (5' end) and HindIII (3' end) restriction enzyme recognition sequences. The resulting PCR product was digested with NdeI and HindIII Fermentas Fast Digest Enzymes and gel extracted using a GeneJet gel extraction kit from Fermentas according to manufacturer protocol. This substrate was ligated into a NdeI/HindIII linearized, phosphatase-treated (NEB CIP), gel extracted pET21a vector (Invitrogen), which has an IPTG-inducible promoter and introduces a C-terminal histidine tag to the expressed protein. The ligation was transformed into chemically competent NEB5-alpha cells (NEB) and colonies were picked into LB broth (Difco). The resulting cultures were miniprepped using a GeneJet Plasmid miniprep kit from Fermentas according to manufacturer protocol to extract the plasmids. Plasmid samples were sent to GeneWiz for sequencing according to company guidelines and the correct plasmid was stored at -20°C.

Expression and Purification of SigA

pVMV36 was transformed into RIPL chemically competent cells (gift from the VanDemark laboratory) and colonies were picked into 5mL LB broth (Difco) with selection and grown overnight at 37°C. 1L of media was inoculated at an OD_{600} of 0.01-0.05 and grown until cultures reached an OD_{600} of 0.4-0.6. His-tagged SigA expression was induced with 1 mM IPTG at 17°C overnight. Cells were pelleted and frozen at -80°C. Thawed cell pellets were resuspended in 5mL per gram of Buffer A (50mM Tris, pH 7.9, 0.5 mM EDTA, 50 mM NaCl, 5% glycerol, 0.1 mM DTT) and lysed in 200 mL fractions by sonicating 10 times for 10 sec at 30% output with 30 sec of cooling on ice in between bursts. Pooled cell lysates were cleared by centrifugation at 30,000 x g for 40 min at 4°C. His-tagged SigA was extracted from the soluble cell lysate using nickel affinity chromatography (Ni-NTA resin, Invitrogen) and fractions were evaluated for protein purity and fractions with pure protein were pooled and dialyzed into a storage buffer (50mM Tris, pH 7.9, 0.5 mM EDTA, 50 mM NaCl, 50% glycerol, 0.1 mM DTT) and stored at -20°C.

4.2.3.3 Western Blot analyses

Western blots were performed as previously described (102). Briefly, protein samples were run on an SDS polyacrylamide gel and transferred to a PVDF membrane (BioRad) using a BioRad wet transfer apparatus according to manufacturer recommendations. The membrane was blocked with TBS + 5% milk for 1-3 hours, washed with TBST three times for 5min, probed with primary antibody in TBS + 5% BSA (8RB13 or 4RA2 – both gifted from the Richard Burgess laboratory) at a 1:5000 dilution for 30-60min, washed with TBST three times for 5min, and probed with anti-mouse AP-conjugated secondary antibody in TBS + 5% milk for 30min. After 2 quick rinses with TBS, a BCIP/NBT Substrate Kit (Invitrogen) was used to induce a

colorimetric reaction from the secondary antibodies and presence or absence of probed protein was detected.

4.2.3.4 DNA Binding Assay using agarose gel electrophoresis

A binding assay using a large protein complex can be too large to fit through the matrix of a 4-5% native gel, so in these instances a variation of a DNA binding assay can be done on an agarose gel. In this assay, no non-specific carrier DNA is added and the binding substrates are not radiolabeled. Increasing concentrations of protein are added to a constant concentration of substrate DNA in a binding buffer containing 20 mM Tris pH 7.5, 10 mM EDTA, 25 mM NaCl, 10 mM spermidine, and 1 mM DTT for a total volume of 10 μl. Reactions were incubated at room temperature for 30 min and the resulting protein DNA complexes were resolved on a 1% agarose gel. The agrose gel was stained with 500mL 1x TBE and 5μl Ethidium Bromide (sigma) for 15 minutes, and destained for 20 minutes in 1x TBE. A BioRad geldoc XR+ geldoc imaging system was used to detect bands.

4.3 DNA BENDING

After theorizing in various sections of previous chapters, we decided to conclusively test if one of the complexes present in the binding profile of $gp33^{103}$ to the *33-34* intergenic region was due to $gp33^{103}$ affecting the structure of the DNA and bending it. To look at this we utilized the pBend2 plasmid ((188-190) a kind gift from the laboratory of Andrew VanDemark), which contains two identical sets of restriction endonuclease recognition sites surrounding a unique cloning site. Cutting with each enzyme in succession results in multiple substrates of equal

length with the unique site in the middle shifting from one end of the substrate, toward the middle and on to the other side of the substrate (191). In this assay we cloned both $O_{R\text{-}R}$ (pVMV22) and the *33-34* intergenic region (pVMV21) into the unique cloning site and cleaved the plasmid with each enzyme in parallel reactions. Since bent DNA has a lower mobility through a gel than unbent DNA, we ran each substrate on a gel with $gp33^{103}$ at a concentration at which it binds ~50% of the DNA (5.4μM) and looked for an upward shift when the unique site was located in the middle of the substrate.

4.3.1 $Gp33^{103}$ does not bend DNA at the $O_{R\text{-}R}$ site alone

We decided to check if $gp33^{103}$ is able to bend DNA if only $O_{R\text{-}R}$ is present and not the rest of the *33-34* intergenic region so we cloned the 12bp $O_{R\text{-}R}$ into the unique cloning region of pBend2, cut the various substrates from the plasmid, and performed DNA binding assays with each substrate (**Figure 33**). The results on the gel show that $gp33^{103}$ is unable to bend the DNA when $O_{R\text{-}R}$ is present by itself. What this suggests is that there are other sequences present in the *33-34* intergenic region that are necessary for bending, the most important of which is probably $O_{R\text{-}L}$.

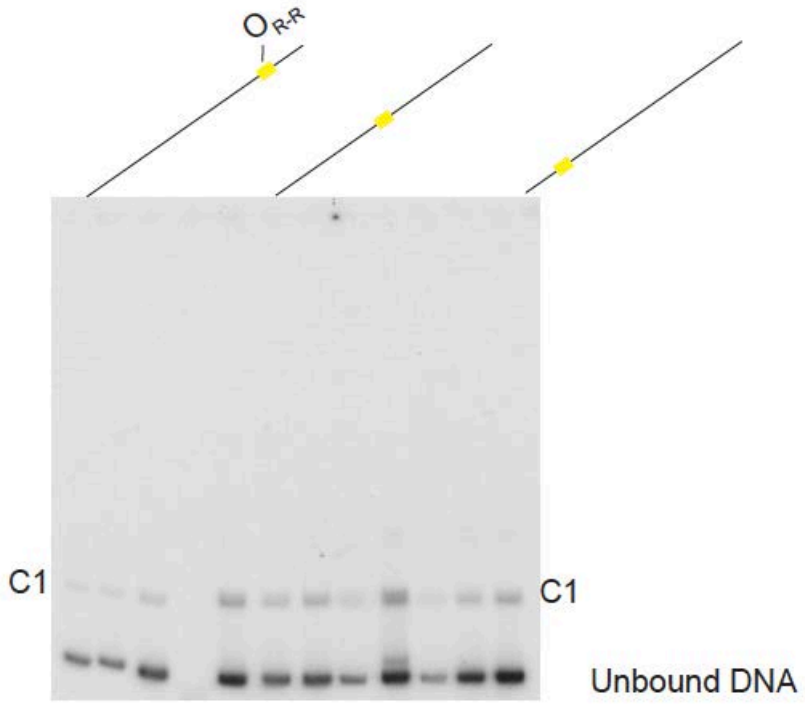

Figure 33 - gp33[103] does not bend DNA at O_{R-R}

The 12bp of O_R was cloned into plasmid pBend2 (191) using unique restriction enzyme sites SalI and XbaI to create pVMV22. Equal-sized fragments of DNA were excised from pVMV22 with O_R located at different positions relative to the edge of the probe by cleaving with appropriate enzymes. BPs gp33[103] (0.54μM) was incubated with each radiolabeled probe and run on a native polyacrylamide gel. The only gp33[103]-DNA complex (C1) does not show any differences in relative mobility with different substrates.

4.3.2 Gp33[103] bends DNA within the *33-34* intergenic region

When we tested the full *33-34* intergenic region for binding using pVMV21, what we saw on the gel was not only a change in mobility as the *33-34* intergenic region reached the center of the substrate, but a continuing upward shift that did not lose mobility when the *33-34* intergenic region reached the other end of the substrate (**Figure 34**). Only complexes C2 and C3 show a change in mobility; complex C1 does not. The simplest explanation for this is that bending is

occurring at O_R, not O_{Rep}, and due to the location of O_R at the right-most end of the *33-34* intergenic region, the operator never makes it past the center of the substrate, and therefore we don't see the decreasing mobility corresponding to the bend migrating to the other end of the substrate (192). The C1 complex could be the interaction between $gp33^{103}$ and O_{Rep}. No intrinsic bending of *33-34* intergenic region was observed (data not shown). The magnitude of the bend (191) is estimated to be approximately 48°, which is consistent with $gp33^{103}$ being bound and introducing a DNA distortion as a tetramer to $O_{R\text{-}L}$ and $O_{R\text{-}R}$.

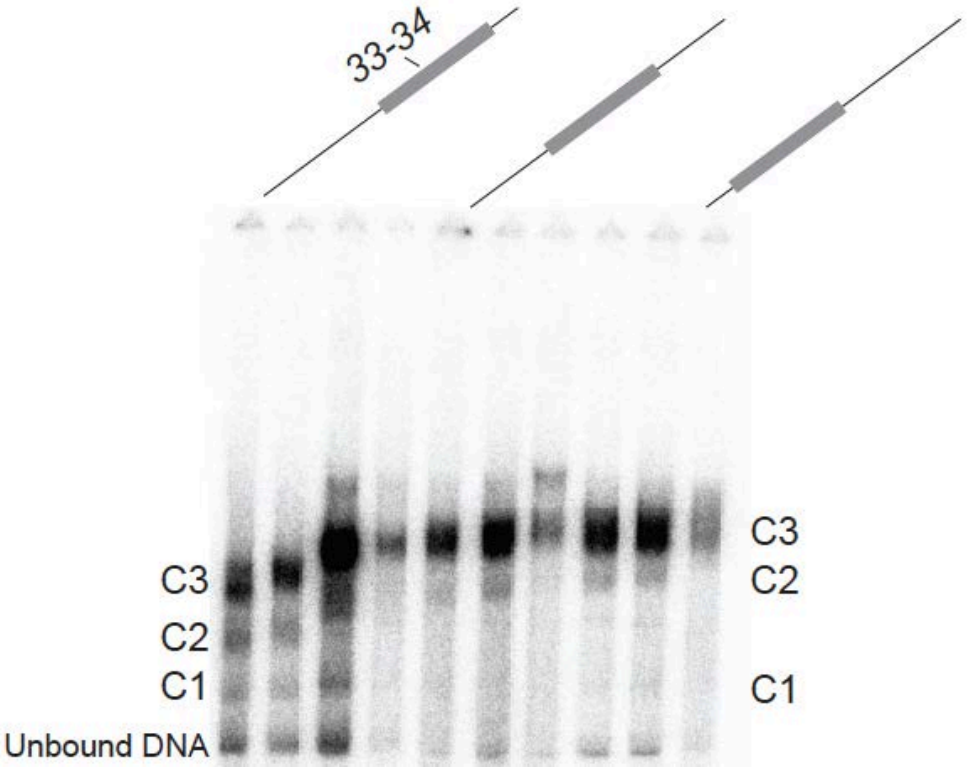

Figure 34 - $gp33^{103}$ bends DNA within the *33-34* intergenic region

The 33-34 intergenic region was cloned into plasmid pBend2 (191) using unique restriction enzyme sites SalI and XbaI to create pVMV21. Equal-sized fragments of DNA were excised from pVMV21 with the insert located at different positions relative to the edge of the probe by cleaving with appropriate enzymes. BPs $gp33^{103}$ (5.4μM) was incubated with each radiolabeled probe and run on a native polyacrylamide gel. The results suggest that a bend of about 48° is introduced into the major complex (C4) by $gp33^{103}$ binding.

4.4 DNA LOOPING

We wondered if one of the complexes present in the binding profile of gp33^{103} to the *33-34* intergenic region was due to gp33^{103} binding to multiple areas and interacting, thus creating a and Inter-DNA bridge, or DNA loop, as has been demonstrated with lambda cI (192, 193). To test this we created two different sized substrates (366bp and 226bp; **Table 5**), both containing the *33-34* intergenic region centered. We compared the binding profiles of gp33^{103} to each substrate separately, and then of gp33^{103} binding to both substrates present in the same reaction (**Figure 35**). If one of the complexes is due to looping, we would expect to see a complex present when both sized substrates are added together, that is not present when gp33^{103} binds to each substrate separately.

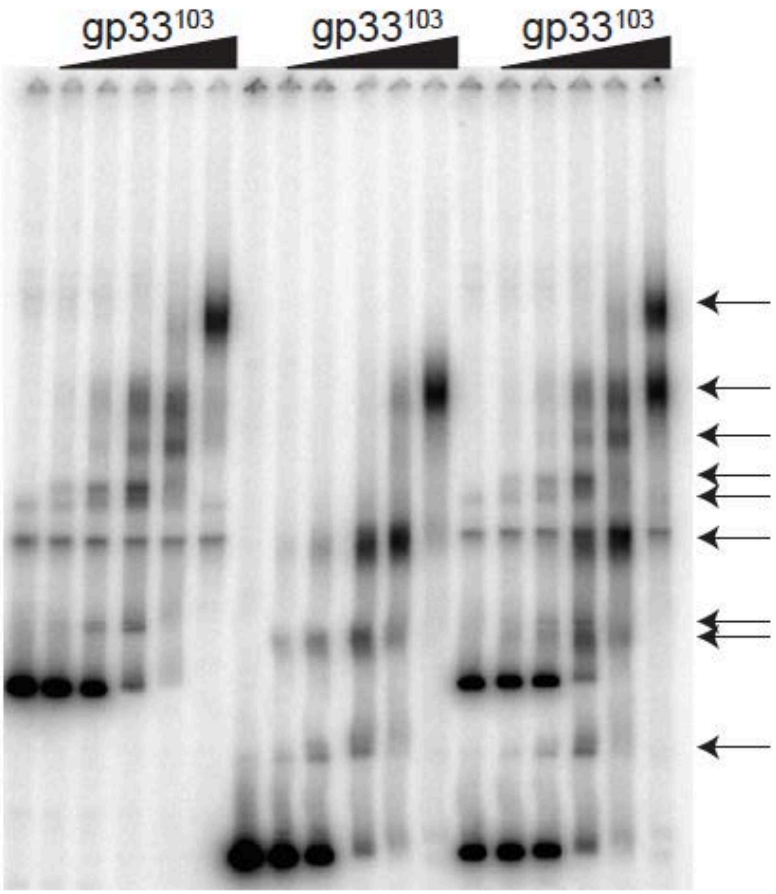

Figure 35 - gp33[103] does not create inter-DNA bridges

To determine if intermolecular DNA bridges are formed by gp33[103], two different sized fragments of DNA containing the *33-34* intergenic region (366 bp and 256 bp) were PCR amplified and radiolabeled (Table 5). The leftmost six lanes contain only the 366 bp substrate, the middle six lanes contain only the 256 bp substrate, and the rightmost six lanes contain both substrates. Complexes formed with each substrate are indicated by arrows. The concentration of protein in each series of substrates is 1) none, 2) 0.54μM, 3) 1.6μM, 4) 5.4μM, 5) 16μM, 6) 54 μM.

We did not see any such bands and therefore concluded that gp33[103] is unable to facilitate loop formation within the *33-34* intergenic region (**Figure 35**). This assay has not necessarily ruled out binding between regions of the genome that contain operators, even though both substrates contained a complete O_R sequence and a bound gp33[103] tetramer was unable to interact with another bound gp33[103] tetramer. This assay also does not rule out the formation of hairpin-like

structures within the same DNA fragments that do not interact with neighboring strands, even though bending assays do not support the hypothesis of $gp33^{103}$ bending DNA at a magnitude necessary for protein bound at O_{Rep} and O_R to interact.

4.5 SPACER EFFECT ON OPERATOR FUNCTION

O_6, O_R, O_{27}, and O_{61} each contain half-sites that are 5bp apart. O_{55} and O_{Rep} on the other hand, contain half-sites that are 8bp apart. We hypothesized that spacer length has an effect on $gp33^{103}$'s ability to bend DNA at these operator sites. We created a series of 42 bp synthetic oligonucleotide substrates, each containing two 12 bp palindromic sequences spaced either 5 bp or 8 bp apart (**Figure 36**). We also made derivatives of these in which either the left or right half site was mutationally ablated, and examined $gp33^{103}$ binding to these substrates.

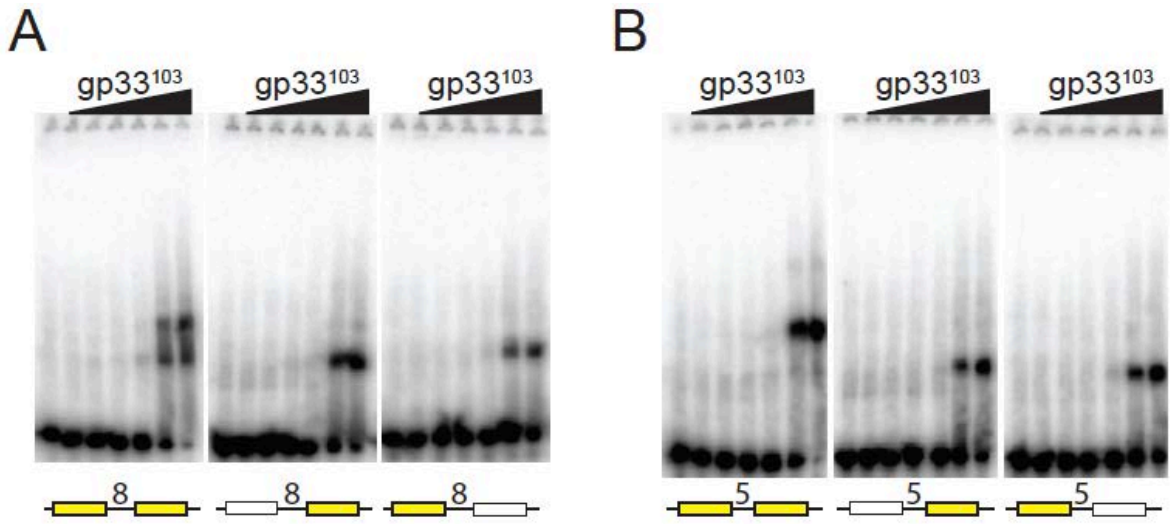

Figure 36 – $gp33^{103}$ binds to two adjacent 12bp palindromes

(A) and (B) EMSAs showing $gp33^{103}$ binding to substrates containing two 12bp palindromic half sites (yellow boxes) spaced either 8 bp (A) or 5 bp apart (B). Substrates in which one half site is mutationally ablated is shown as a white box.

With a 5 bp spacer, BPs gp33^{103} binds to form a single complex that has a slower mobility than those formed when each of the half sites is ablated (**Figure 36**). Their affinities are not greatly different suggesting a lack of cooperativity, and we propose that each complex contains a gp33^{103} tetramer bound to DNA, but that the presence of two 12 bp palindromes enables the protein to correctly dock at each half site and induce a DNA bend, whereas the single half-site substrates do not.

When substrates with an 8 bp inter-site spacing are tested, two complexes are formed, one with similar mobility to the single half-site substrates, and a slower one that migrates similarly to the 5 bp spaced substrate (**Figure 36**). We propose that the larger spacing disfavors proper docking of gp33^{103} to its DNA sites, resulting in a mixture of two complexes, one that is properly docked and the DNA is bent, and one in which only one half site is engaged and the DNA is not bent.

4.5.1 Changes in P_R spacer length affect gp33^{103} action *in vivo*, not *in vitro*

Taking a closer look at where the operators we've confirmed are present with respect to the nearby promoters, we drew some conclusions about possible mechanisms of repression. Except for P_{27}, which does not have confidently predicted -10 and -35 motifs, all BPs operators have at least one half site that covers part or all of a predicted -10 or -35 motif.

Usually operators are located over one of the promoter elements (194), so this could be indicative of a promoter occlusion mechanism of action, or also, due to the DNA bending we see in section 4.3.2, a change in DNA structure could be the cause of RNA polymerase not binding.

Some data that Lauren Oldfield collected has shed some light on another possible mechanism of action for gp33^{103} repression. In her studies of the P_R promoter, she fused P_R to a *mCherry* reporter gene and expressed it both in wild type *M. smegmatis* mc^{2}155 and in a BPs lysogen of *M. smegmatis*. Wild type P_R expressed mCherry in wild type *M. smegmatis*, but expression was dramatically reduced in a BPs lysogen, as expected. She created various mutations looking for an effect on promoter strength and measured the bacteria for mCherry expression by quantifying the fluorescence.

In wild type P_R, the spacer length is 18 base pairs and she saw expression of mCherry in wild type *M. smegmatis*, but decreased expression in a lysogen (**Figure 37A/B**). When she added one basepair upstream of $O_{R\text{-}R}$, but before the -35 motif, she saw decreased expression from P_R in wild type *M. smegnatis*, but no change in the lysogen's ability to repress the promoter (**Figure 37A/B**). When she removed one or two base pairs between $O_{R\text{-}R}$ and the -35 element, however, she saw that while there were some changes in WT *M. smegmatis* expression, there were markedly less decreases in expression in the lysogen. This means that the repressor in the lysogens is unable to block this lytic P_R promoter as effectively when $O_{R\text{-}R}$ is closer to the -35 element and $O_{R\text{-}L}$.

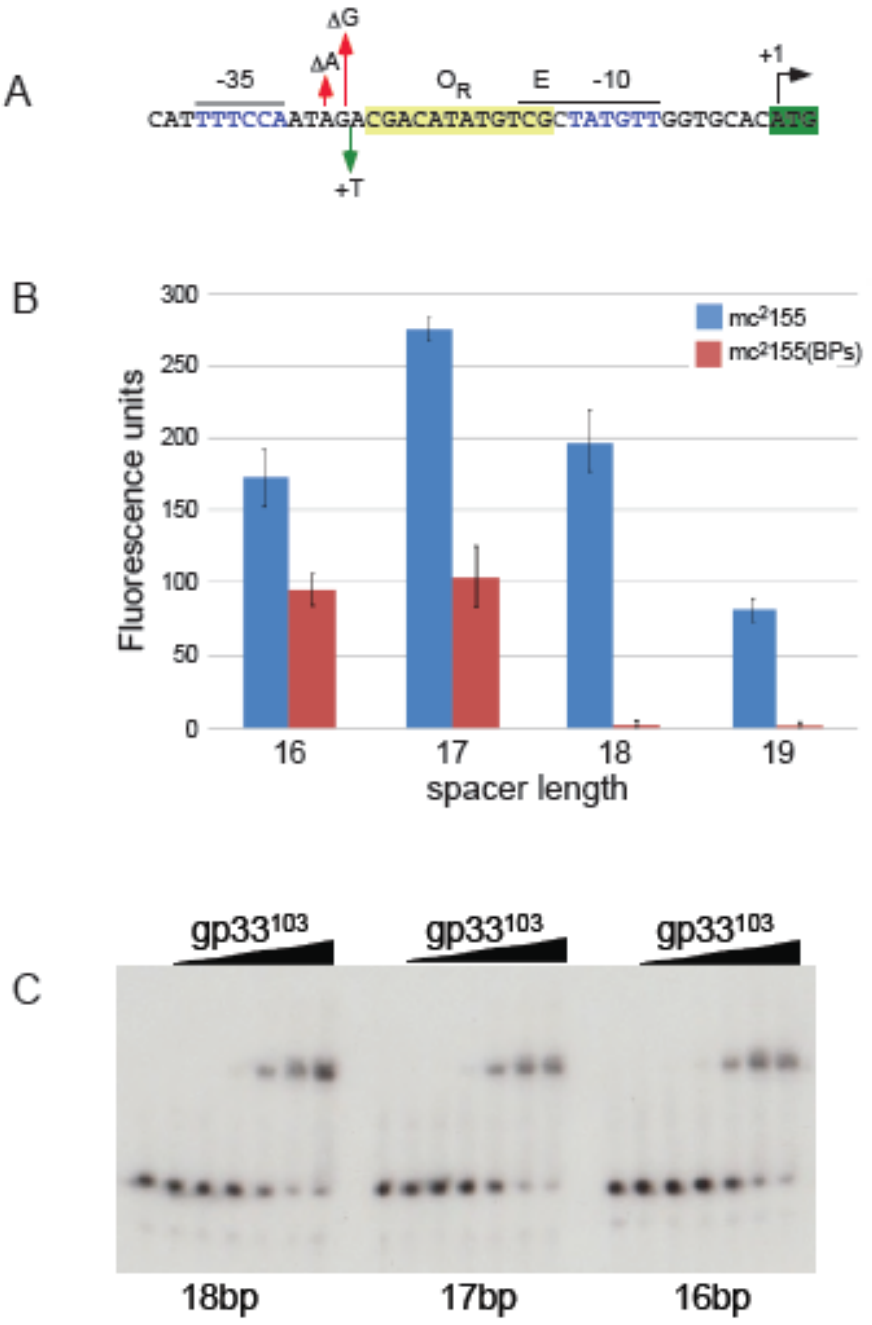

Figure 37 - gp33[103] binding to various promoter mutants

(A) Shows the locations of the base pairs that Lauren Oldfield added and removed. (B) Shows the florescent reporter assays she did in which plasmids with P_R (with or without mutations) driving mCherry expression were transformed into wild type (blue) and BPs lysogens (red) of *Mycobacterium smegmatis*. Her result showed dramatic de-repression effects in vivo as a result of removing base pairs from the P_R spacer. In (C) EMSAs were done on substrates containing just the promoter sequence and showed that the base pair deletions had no effect on binding.

4.5.2 Gp33^{103} binds similarly when half-sites are spaced very close or very far apart

To test whether this effect was due to an inability of gp33^{103} to bind to O_R, we created a series of substrates containing just the promoter sequence from the left end of the -35 motif to the right end of the -10 motif. These substrates were created before the hypothesis of O_R consisting of two 12bp half-sites, so only 6bp of $O_{R\text{-}L}$ are included. The synthetically engineered substrates had a wild type spacer length, and either one or two base pairs missing from the spacer, reflecting the mutations made in the data from Oldfield and Hatfull (**Figure 37C**). Gp33^{103} binds just as well to each of the promoter mutants.

Only one complex is formed when gp33^{103} binds to these substrates, leading us to believe that this complex consists of a gp33^{103} tetramer binding to $O_{R\text{-}R}$, but not bending DNA, as $O_{R\text{-}L}$ is not fully present. We wondered how close (or far) the operator half-sites could be and still bend DNA.

To look at this, we created two 60bp oligonucleotide substrates that had two 12bp operaters either 2bp apart, or 23bp apart (**Figure 38**). When gp33^{103} binds to each of these, it binds with similar binding profiles, which is unexpected. Based on results from **Figure 36**, Gp33^{103} binding to the operator with the 2bp spacer would presumably have a binding profile indicating either a tetramer bound to the DNA substrate, or a tetramer bound and bending the DNA substrate. We do see two major complexes being formed, as we do with an 8bp spacer, which indicates that some DNA is being bound, but not all of it.

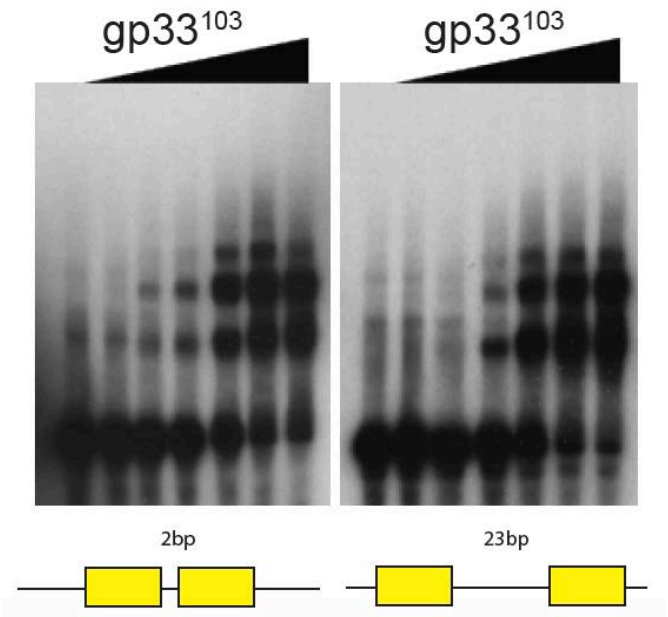

Figure 38 - gp33[103] binding to two half sites with varying spacer regions

EMSAs were performed on 60 bp synthetic oligonucleotide substrates (Table 5) containing two 12bp palindrome half sites separated by 2 or 23bp.

Curiously, when the half-sites are 23bp apart, we see a very similar binding profile with two major complexes being formed, although the faster-moving complex looks slightly more dominant (**Figure 38**). The best explanation for what is happening is that the faster-moving complex is a single tetramer bound to one of the two half sites and the slower moving substrate corresponds to both half-sites being occupied. Due to the relative mobilities of the slower moving complex between the 2bp and 23bp spacer substrates being the same, it is unlikely that the complexes of bent DNA and DNA with two tetramers bound run at the same mobility. Therefore, we propose a second theory for what could be happening at the operator sites. Gp33[103] could be binding as a tetramer to each half-site and with the correct spacing, interacting and bending DNA, but without the correct spacing, just binding without bending.

4.6 INTERACTION WITH RNA POLYMERASE

Some repressors, such as lambda cI, are capable of directly interacting with RNA polymerase to exert an effect (195). In the example of lambda, CI interacts with the alpha subunit of the RNA polymerase to enhance transcription of its own gene during lysogeny (195). In other systems, repressors interact with RNA polymerase in solution to prevent it from binding to a particular promoter, changing its specificity toward a different promoter, or even interacting with RNA polymerase while it is bound at a promoter sequence and preventing it from elongating a transcript.

In the case of BPs, we wanted to discover whether $gp33^{103}$ could interact with RNA polymerase at O_R within the P_R promoter. We started by purifying the mycobacterial RNA polymerase and then performed binding assays using $gp33^{103}$ and RNA polymerase on the *33-34* intergenic region.

4.6.1 Purification of RNA polymerase

The Burgess lab kindly gifted us some of their monoclonal antibodies recognizing bacterial RNA polymerase. Polyol responsiveness means that the monoclonal antibodies bind tightly, but allow easy elution so that you can pull proteins down in complex during immunoaffinity chromatography. This polyol responsive antibody binds to the beta flap tip helix of the beta subunit of bacterial RNAP's and this is one of the major binding sites of SigA to the core enzyme (125, 180, 183-187, 196). Therefore, this style of purification will only bring down the core RNA polymerase, not the holoenzyme. So we had to purify the mycobacterial SigA sigma factor and reconstitute the holoenzyme.

4.6.1.1 Purification of *M. smegmatis* RNA polymerase core enzyme

The Burgess lab sent us a small sample of their monoclonal antibodies that broadly recognize bacterial RNA polymerase beta (8RB13) and alpha (4RA2) subunits (197, 198). The first thing we had to do was make sure that their antibodies recognized the *M. smegmatis* RNA polymerase subunits. To do this, we performed a Western blot on a *M. smegmatis* mc^2155 whole cell lysate next to an *E. coli* positive control whole cell lysate (**Figure 39**). Both antibodies recognized the *M. smegmatis* RNA polymerase subunits, so we proceeded with the purification as described in section 4.2.4.1.

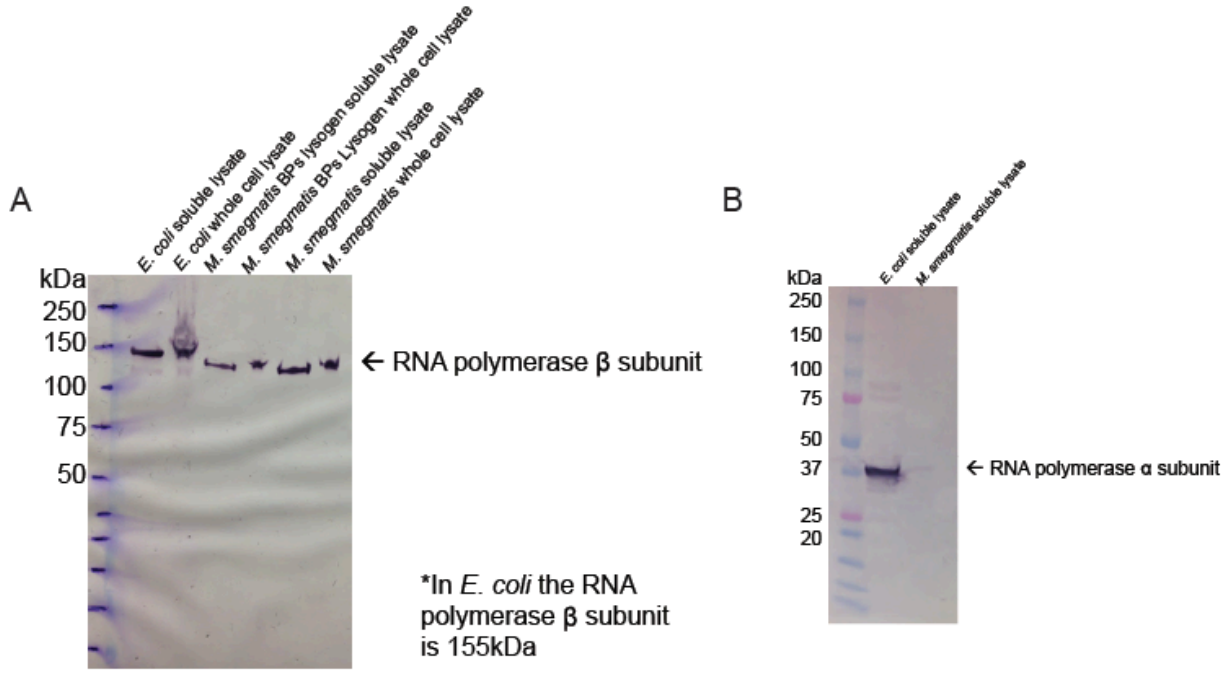

Figure 39 - Western blots confirming 8RB13 and 4RA2 antibodies

Western blots showing the (A) beta (8RB13) and (B) alpha (4RA2) RNA polymerase subunits of *E. coli* and *M. smegmatis*.

The purification was successful, although the antibodies pulled down an unidentified contaminant (**Figure 40A**). Western blots of the purified *M. smegmatis* RNA polymerase showed that all of the correct subunits are present (**Figure 40B**), so we proceeded with purification of SigA.

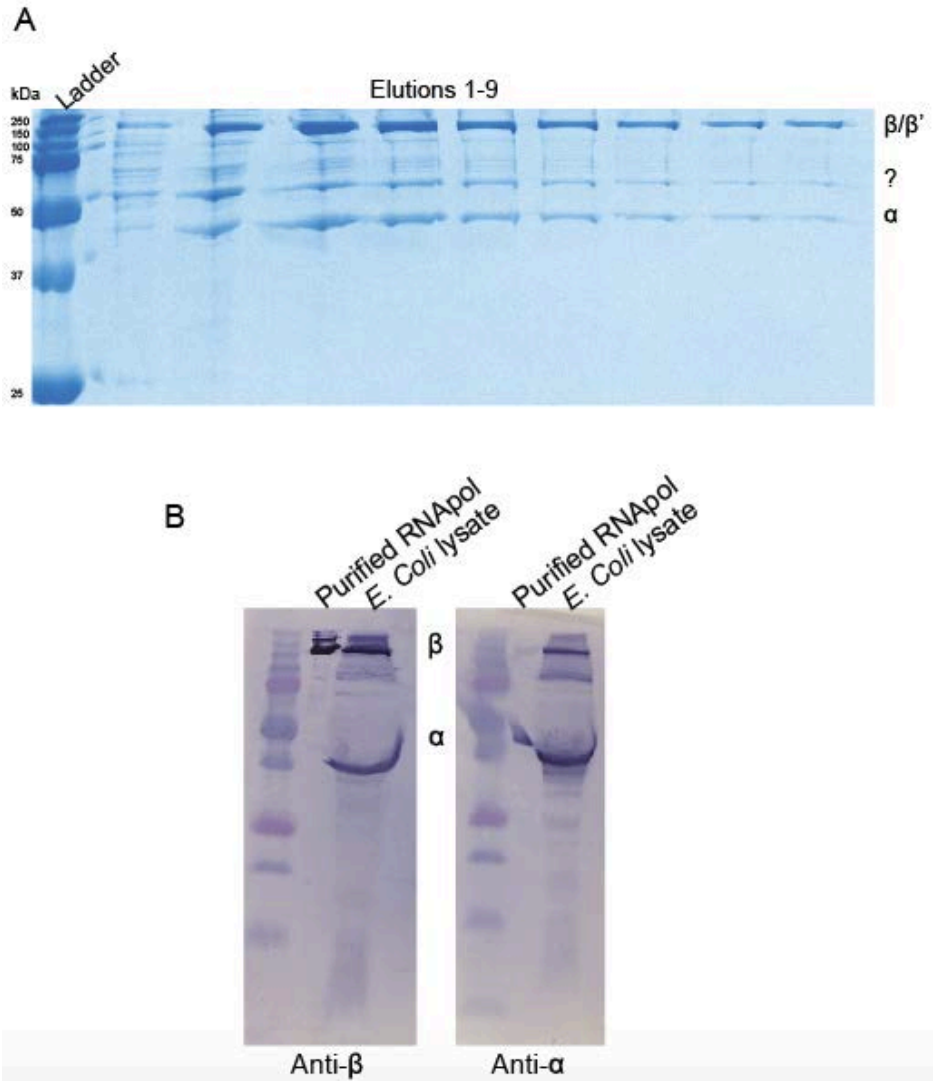

Figure 40 - Purification of *M. smegmatis* RNA polymerase

(A) Shows an SDS PAGE gel of the *M. smegmatis* RNA polymerase purification. (B) Shows Western blot confirmation of beta and alpha subunits in the purified pooled fractions.

4.6.1.2 Purification of *M. smegmatis* SigA

His-tagged SigA was insoluble in the first two strains used to overexpress the protein, so we screened six new strains to try to find one that expresses soluble SigA (**Figure 41A**). Since the band we presumed to be SigA ran higher than expected on an SDS PAGE gel (70kDa vs. 55kDa MW), we performed Western blots on the soluble lysates from each strain with an anti-Sigma factor antibody from the Burgess lab (2G10) (**Figure 41B**), and when those results were inconclusive, we performed a Western blots using an anti-His antibody (Invitrogen) (**Figure 41B**). We chose to continue with the RIL strain to perform a full-scale purification.

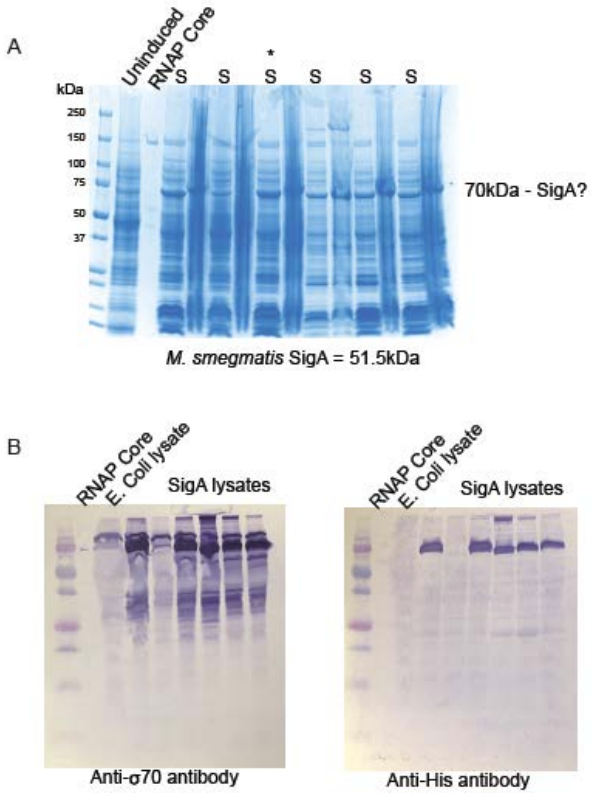

Figure 41 - Purification of *M. smegmatis* SigA protein

(A) The *M. smegmatis* SigA protein was overexpressed in multiple overexpression strains but it was unclear whether SigA was present. (B) Western blots confirm the presence of SigA with a 2G10 anti-sigma 70 antibody as well as an anti-His antibody to recognize the His-tag these proteins were expressed with.

To reconstitute the *M. smegmatis* RNA polymerase holoenzyme we mixed purified SigA and RNA polymerase core enzyme in a 10:1 ratio.

4.6.2 Binding Assays with *M. smegmatis* RNA polymerase and gp3^{103}

We performed EMSAs with RNA polymerase and gp33^{103} on the 33-34 intergenic region substrate according to the protocol described in section 2.2.2 (**Figure 42**). We started by adding RNA polymerase and gp33^{103} alone to determine their binding profiles, and then mixed RNA polymerase and gp33^{103} in different molar ratios. RNA polymerase by itself did not cause an upward shift on the gel indicating that it does not bind the promoter DNA (**Figure 42**). This could either be because the RNA polymerase is not active, or because the RNA polymerase complex is too big to enter the gel matrix. It is evident that some radiolabelled DNA is retained in the well, indicating that something is holding it there and presumably that is where the RNA polymerase complex is.

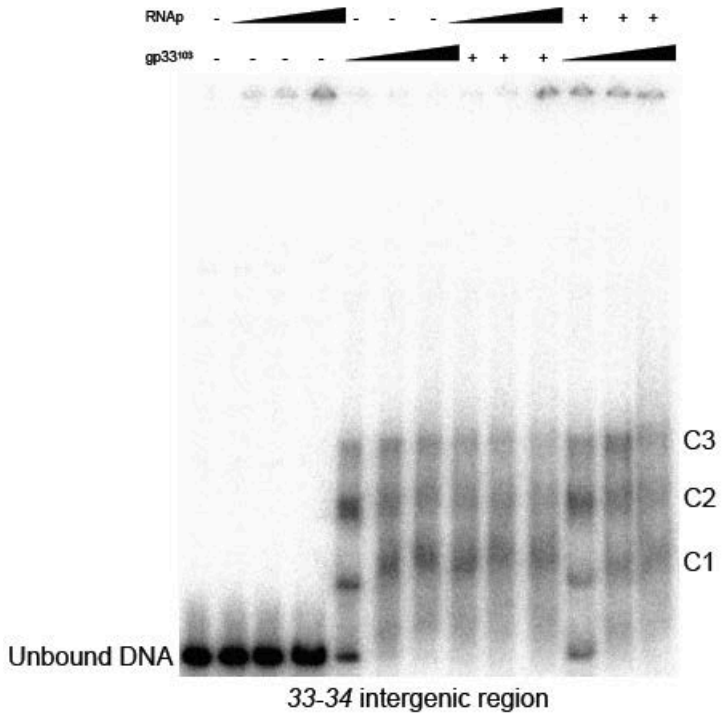

Figure 42 - RNA polymerase and gp33[103] binding to the *33-34* intergenic region

EMSA showing the binding profiles of RNA polymerase, gp33[103], and both together. RNA polymerase was unable to bind DNA in this assay.

Gp33[103] does bind DNA, but with the addition of RNAP there is no difference in gp33[103]'s binding profile (**Figure 42**). Again the DNA trapped in the well indicates that perhaps the RNA polymerase complex is too large to fit through the gel matrix.

We also performed the binding assays on the other intergenic regions with the same results (**Figure 43**).

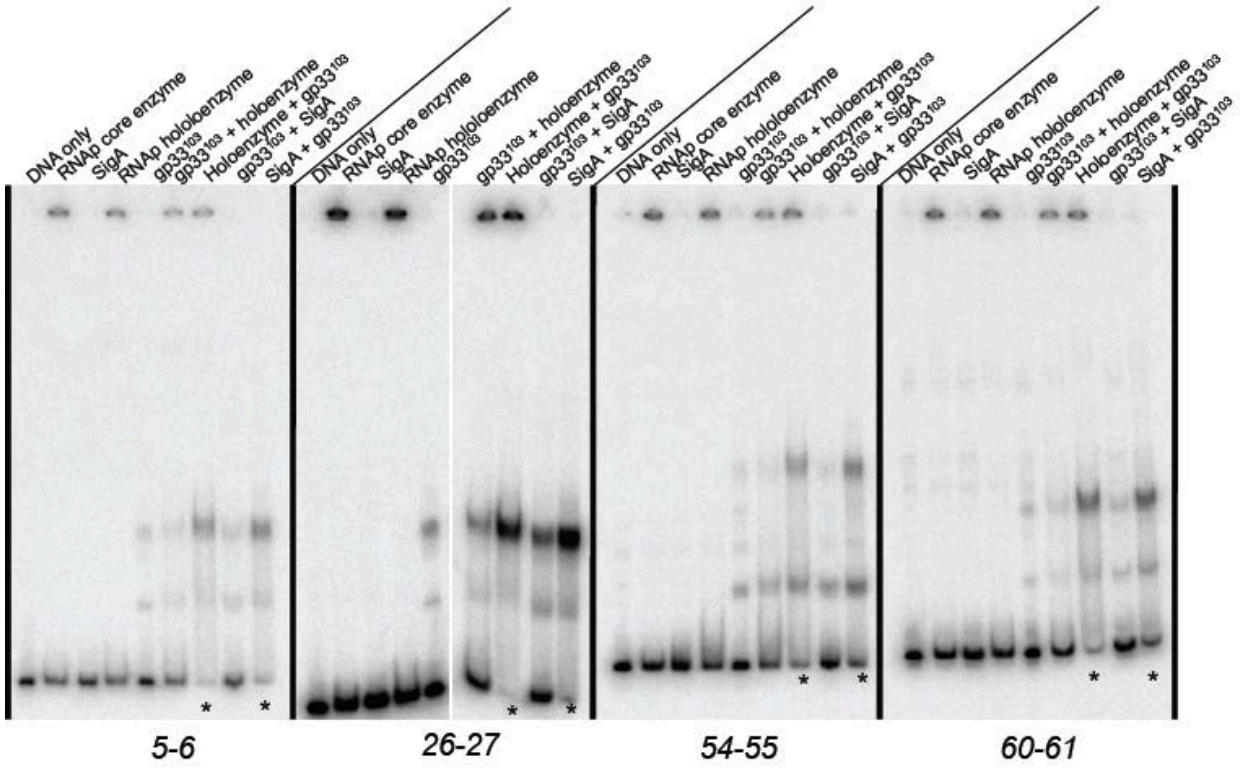

Figure 43 - RNA polymerase and gp33[103] binding to othe rintergenic regions across the BPs genome

EMSAs showing the binding profiles of RNA polymerase, gp33[103], and both together to various intergenic regions across the BPs genome. RNA polymerase was unable to bind DNA in this assay. Gp33[103] protein concentrations are 5.4μM or 16μM. Asterisks mark lanes where a higher concentration of gp33[103] was used.

We decided to perform a binding assay on an agarose gel rather than a 5% native polyacrylamide gel to create a larger gel matrix through which RNA polymerase could migrate (**Figure 44**).

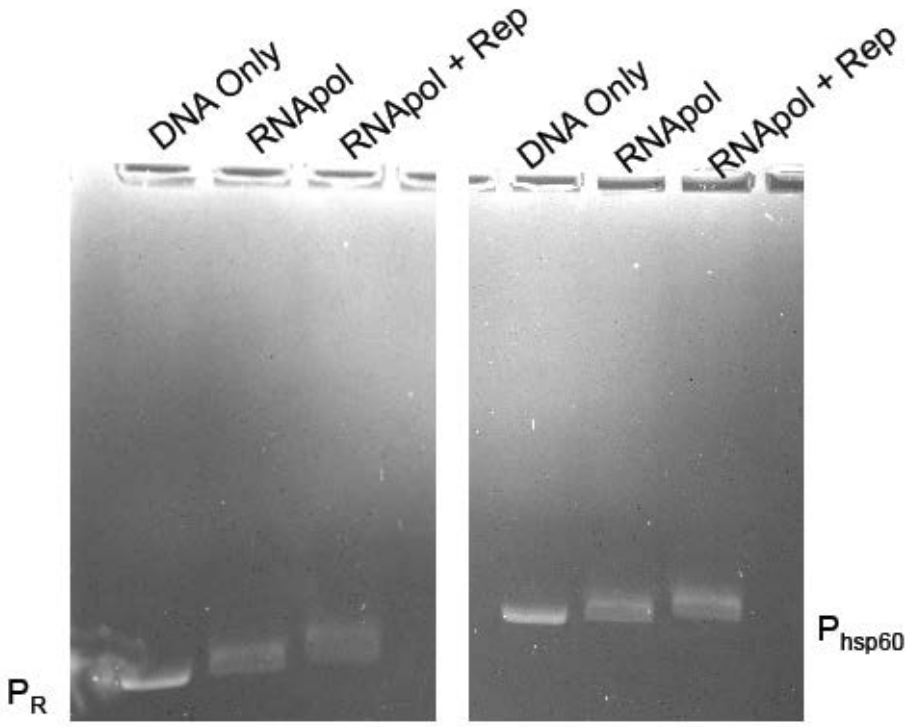

Figure 44 - RNA polymerase and gp33[103] binding within the P_R and P_{hsp60} promoters

RNA polymerase and gp33[103] were incubated with plasmid DNA containing the P_R and P_{hsp60} promoters as indicated. RNA polymerase faintly binds DNA in this assay, and gp33[103] seems to increase gp33[103] binding, suggesting an interaction between the proteins.

So we repeated the binding assay using a 1% agarose gel. This gel is ethidium bromide stained and therefore no competitor DNA was added to prevent non-specific binding, because it would show up on the gel. The substrates used in this assay are plasmids containing either P_R or P_{hsp60} upstream of an *mCherry* reporter gene, because the *M. smegmatis* RNA polymerase has a dependence on supercoiling to work, so perhaps that is why the assay did not work in previous gels with linear substrates (199). The results show that in both cases, RNA polymerase is able to bind DNA, at least slightly, and there is a small decrease in mobility of the P_R substrate when gp33[103] is added (**Figure 44**).

The results indicate that perhaps RNA polymerase and gp33^{103} do interact, because when O_R is present (in the P_R substrate) and gp33^{103} is added to the RNA polymerase complex, a higher shift occurs than with RNA polymerase alone. However it is a smaller shift that we would anticipate for a huge holoenzyme binding to a DNA substrate, so these results are not definitive and need to be repeated.

4.7 DISCUSSION

In this section we show that gp33^{103} does not form inter-DNA bridges, but instead has the ability to bend DNA within the *33-34* intergenic region, likely within O_R. It cannot form a DNA bend when O_{R-R} is by itself, however it does seem to bend DNA when two half sites are present on the same DNA substrate. It would be informative to put two 12bp sites with varying spacer regions into the pBend2 plasmid and compare their abilities to bend DNA. It would be interesting to see if there is a difference in, not just ability to bend, but also in the angles of DNA-bending. It would also be informative to repeat the bending assay with the other BPs operators to determine if gp33^{103} bends at those sites and if it does, if it bends with a similar angle as in the *33-34* intergenic region.

Gp33^{103} also may have an ability to interact with the *M. smegmatis* RNA polymerase as indicated by the small shift in a binding assay. However more experiments need to be done to determine if this is the case, and if there is an interaction, the nature of the interaction should be resolved.

5.0 CONCLUSIONS AND FUTURE PROSPECTIVES

We have described here the unusual binding properties of the repressor encoded by mycobacteriophage BPs. The repressor is non-canonical in that it is encoded in two forms, a virally-encoded product 136-residues long ($gp33^{136}$), and prophage-encoded version ($gp33^{103}$) that is 33 residues shorter because the site where site-specific recombination works during integration (*attP*) is located within the gene *33* open reading frame. The two proteins bind similarly to a DNA substrate containing the regulatory intergenic between genes *33* and *34*, but the binding pattern is complex with multiple protein-DNA complexes observed (**Figures 3, 4**).

BPs $gp33^{103}$ binds to a total of six operators (O_6, O_{27}, O_{Rep}, O_R, O_{55}, O_{61}) within the BPs genome, all of which are situated over promoters (P_6, P_{rep}, P_R, P_{55}, P_{61}), except P_{27}, for which the promoter sequences are not confidently predicted (**Figures 5, 24-29**). The simplest explanation for DNA recognition is that each site contains two half sites, one of which corresponds closely to the palindrome 5'-CGACATATGTCG, and the other which is related to it but has some or many sequence departures. $Gp33^{103}$ probably occupies each of the full sites as a tetramer because this is the predominant and homogenous form observed in solution (**Figure 8, 9**), although it is also possible that each full site is occupied by two tetramers. Occupancy as a tetramer is an unusual feature, although other phage repressors can form DNA loops involving tetrameric interactions (79). However, at least with the substrates used here, there is no evidence of simultaneous

occupancy of all four protomers within a tetramer, and the significance of the stoichiometry remains unclear.

Another unusual feature of $gp33^{103}$ is the apparent ability of $gp33^{103}$ to bind to sites in which different distances separate the two putative half sites (**Figure 36, 38**), although this is not unprecedented and is also seen in transposon resolvases (200). There are clearly some cooperative interactions that enhance binding affinity for $gp33^{103}$, even though for many substrates these are of slight magnitudes.

Our first model of how $gp33^{103}$ could be binding involves a tetramer binding to each half site and interacting to create a DNA bend. The depiction in **Figure 45A** shows a hypothetical explanation for what each of the complexes in the $gp33^{103}$ binding profile mean under the premise of this model. This model could also be applied to dimers binding to each half site (**Figure 45A**). In both cases, two protomers bind to each half site and interact to bend DNA when the inter-site spacing is correct.

Our second model postulates one protmer binding to each half site. In this model, a dimer would bind to the full site and the spacing between half sites determines whether or not both protomers can interact with and bend DNA. This also applies for a tetramer binding to a full site as illustrated in **Figure 45B**.

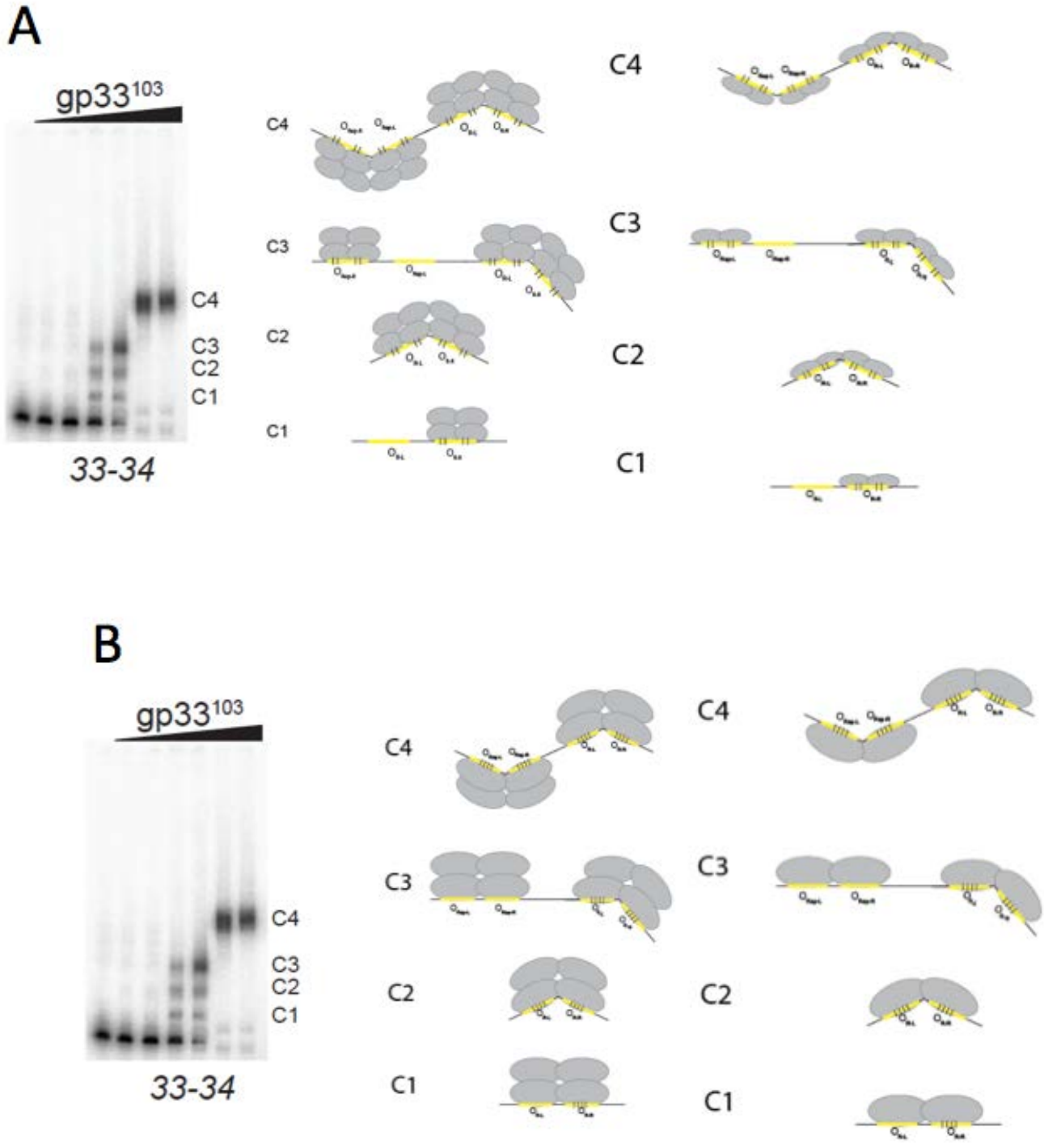

Figure 45 - Models of gp33[103] binidng

(A) Our first model for gp33[103] binding with two protomers binding each half site. (B) Our second model with one protomer binding at each half site.

To conclusively determine the actual stoichiometries of protein binding to each operator, there are a variety of experiments that can be performed. One approach is to create Ferguson plots with relative mobility of substrates vs. the concentration of acrylamide to deduce the absolute molecular weight of the complexes being formed. Another method is to perform an experiment whereby $gp33^{103}$ would be synthesized with ^{35}S-methionine incorporated and the counts per protomer could be calculated. Using known quantities and counts of ^{32}P-labeled probe would allow the ratio of ^{35}S to ^{32}P counts to be calculated and the stoichiometries of protein per DNA molecule could be surmised. This would require very careful calibration and may be time consuming and difficult to accomplish. An easier and faster experiment would be to crosslink the proteins to DNA during the binding reaction and then run the samples on a denaturing SDS PAGE gel. If the protein binds as a monomer, a ladder-like pattern will be seen, however if the protein does indeed bind as a tetramer, a step-like pattern will occur. An alternative approach to the size-exclusion chromatography used is to perform sedimentation equilibrium to isolate protein complexes. If multiple complexes are formed, this can help determine what intermediate binding profiles are composed of. Also, mass spectrometry, surface plasmon resonance (201) or fluorescence anisotropy could be used (202).

Alignment of the six putative DNA binding sites (O_6, O_{27}, O_{Rep}, O_R, O_{55}, and O_{61}) shows the juxtaposition of the half sites within each operator (**Figure 46**). In four of the sites (O_6, O_{27}, O_R, and O_{55}), one half site contains the 12 bp palindrome ($O_{6\text{-}L}$, $O_{27\text{-}R}$, $O_{R\text{-}R}$, and $O_{55\text{-}L}$) and is easily recognizable, and the other half sites ($O_{6\text{-}R}$, $O_{27\text{-}L}$, $O_{R\text{-}L}$, and $O_{55\text{-}R}$) contain sequence departures at two or more positions (**Figure 46**). At O_{61}, both half sites have sequence departures from the consensus (two in $O_{61\text{-}R}$ and five in $O_{61\text{-}L}$) although $gp33^{103}$ binds with

similar affinity as it does to O_6 and O_{55} (**Figure 46**). A half-site alignment shows that the 12bp palindrome is indeed the consensus sequence for each half-site (**Figure 47**).

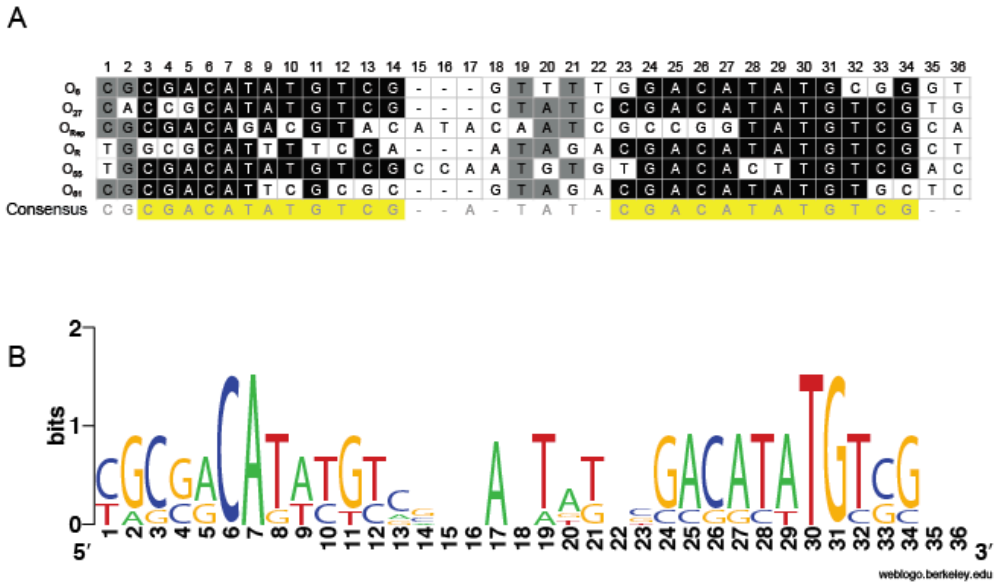

Figure 46 - Full site alignment for BPs operators

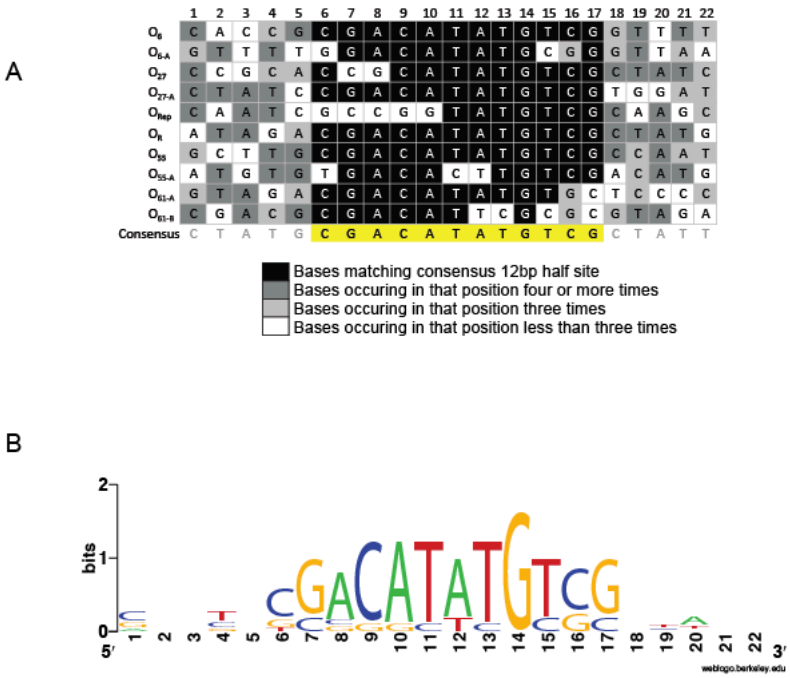

Figure 47 - BPs operator half site alignment

With most half-sites, more than two base pair departures from the consensus do not support gp33^{103} binding, but O_R is peculiar in that the $O_{R\text{-}L}$ half site is only distantly related, although gp33^{103} clearly binds to it, even in the absence of $O_{R\text{-}R}$ (**Figures 18, 19, 21**). $O_{R\text{-}L}$ is spaced five bases away from $O_{R\text{-}R}$, which is similar to the geometries of O_6, O_{27}, and O_{61} (**Figure 46**). We also note that position eight of this proposed $O_{R\text{-}L}$ is part of the P_R -35 motif, and substitution with a G base both increases promoter activity, but also reduces the efficiency of repression (102), consistent with a role for $O_{R\text{-}L}$ in binding and regulation.

The sequences of the O_{Rep} half sites are the most divergent, and the best alignment suggests that $O_{Rep\text{-}L}$ and $O_{Rep\text{-}R}$ have eight and seven positions conserved respectively (**Figure 46**), spaced eight bases apart (**Figure 47**), similarly to O_{55}. DNase I footprinting is consistent with binding to these sites, and complexes are observed with the small substrate tested that contains the complete sequence (**Mt10, Figure 21**). At least at higher protein concentrations, binding of gp33^{103} to the *33-34* intergenic regions is expected to involve occupancy of both O_R and O_{Rep} although this likely involves multiple gp33^{103} tetramers, as there is not evidence of protein-mediated bridging, or loop formation. If loop formation did occur, there would probably need to be another DNA-bending protein involved, since O_{Rep} and O_R are only ~150bp apart and our evidence suggests that gp33^{103} can only bend DNA up to 48°. Due to the low binding affinity of gp33^{103} to O_{Rep}, it is possible that this operator is only bound when protein levels get too high in the lysogen and that this operator functions to regulate repressor levels during lysogeny, similarly to O_R3 in lambda (69, 77).

Despite the complexity of the binding profile of gp33^{103} to the *33-34* intergenic region, formation of properly configured complexes is necessary for normal repression in a lysogen. BPs mutants with repressor-insensitive phenotypes that have mutations mapping to the *33-34*

intergenic region demonstrate the importance of this sequence and the ability of $gp33^{103}$ to bind to it (**Figures 30, 31**). Loss of normal repression does not closely correlate with large changes in binding affinity, but the relatively subtle sequence changes give rise to altered configurations that interfere with repression. This notwithstanding, some DNA substrates of the repressor-insensitive mutants have binding patterns that are more consistent with binding of monomers or dimers (e.g. 127b, **Figure 31**), and it is unclear what determines this behavior.

While inter-site spacing seems to have an effect on the $gp33^{103}$ binding profile *in vitro*, a logical question is to ask whether the spacing between half sites is important *in vivo*. We know that narrowing the space between O_R half sites leads to a decrease in repression in a lysogen based on Lauren Oldfield's fluorescent reporter experiments with mutated versions of P_R (**Figure 37**). In order to test whether this spacing is important for all of the phages *in vivo*, BRED can be used to insert and delete base pairs between operator half sites in all BPs operators. For example, we can ask if changing the spacing between O_{27} or O_6 half sites from 5 to 8 changes the phage's ability to enter lysogeny or maintain the lysogenic state. We can similarly examine whether changing the spacing of O_{55} and O_{Rep} from 8 to 5 base pairs changes the gene expression in a lysogen or increases lysogenization efficiency. Alternatively, fluorescent reporter assays can be done with the promoters containing the above-mentioned operators and spacing can be tested in the same manner as P_R and O_R.

It is unknown whether $gp33^{103}$ has the ability to interfere with transcript elongation, however an elegant design to test this would be to create fluorescent reporter assays with a promoter unregulated by $gp33^{103}$ (such as P_{hsp60}) and insert two full 12bp palindromes spaced 5bp apart downstream of the promoter and upstream of the reporter gene, and then express this construct in wild type *M. smegmatis* as a control, and into a BPs lysogen to see if the reporter

gene is repressed by gp33[103] due to the presence of an operator (83). If it did exert an effect over transcription elongation, this "stoperator" function may be used at any of the operators found in the BPs genome, but would probably be most likely used in O_{61}, where the P_{61} promoter is not downregulated in a lysogen in fluorescent reporter assays. Binding to O_{61} may prevent the RNA polymerase from continuing to elongate the P_{55} or P_{58} operon transcripts (103). It could also be used at O_{27} to prevent any possible read-through transcription from expressing the lysis gene cassette.

Gp33[103] may also interact with the mycobacterial RNA polymerase, however the assays performed thus far have not shown any conclusive evidence for this interaction (**Figures 42, 43, 44**). The next step to check if gp33[103] and RNA polymerase truly interact, is to perform a pull-down assay. We have constructed a vector with an HA-tagged *M. smegmatis rpoA* gene placed downstream of an Hsp60 constitutive promoter (pVMV36). When transformed into wild type *M. smegmatis* and a BPs lysogen, the cell lysates containing RNA polymerase with tagged alpha subunits can be used in immunoaffinity chromatography with an anti-HA antibody to pull down the RNA polymerase complex, along with anything it interacts with (203). Mass spectrometry can reveal if the pull-down results from the lysogen include gp33[103]. This technique could also be used to identify any other phage proteins that interact with RNA polymerase, or used to determine the unidentified repressors of other phages by performing co-immunoprecipitation with this system in other phage lysogens.

The system of integration-dependent immunity seen in BPs and other phages offers a different perspective on phage life style decision making than seen in phage lambda and its relatives, and may represent an ancestral state for temperate phages (204). In light of this, it is not surprising that the repressor has non-canonical binding properties including tetramerization

and binding to dispersed sites in the phage genome. Because the repressor can be expressed in two forms that differ in their C-termini, this raises the possibility that the virally-encoded product $gp33^{136}$ forms mixed tetramers with the shorter protein ($gp33^{103}$) and exerts a dominant effect. Thus the observed tetramerization and DNA binding profiles may play roles in modulating the overall genetic switch in these phages.

The work done in this study has shed light on the mechanism action of a new phage repressor. This information has helped uncover how a novel type of integration-dependent genetic switch works, and has provided insights into how mycobacteriophages can regulate gene expression.

APPENDIX A

OTHER PHAGE REPRESSORS THAT CONFER INTEGRATION-DEPENDENT IMMUNITY

A.1 INTRODUCTION

Integration-dependent immunity systems have been found on phages from different clusters as described in Broussard *et al*, 2013. Mycobacteriophage Brujita (Cluster I1) and Charlie (Cluster N) immunity cassettes have a similar structure to the one in BPs (Cluster G). They include an Integrase gene and a Repressor gene amplified leftward from a P_{rep}-like promoter, and a rightward facing P_R-like promoter, which transcribes an early lytic operon. Because of this, we initially started this project by working with the repressors of all three representative phages, not just BPs. In this appendix we describe the purification of the long and short forms of the Brujita and Charlie repressors and the assays done to show that all of these repressors bind to their respective primary regulatory intergenic regions (similar to the BPs *33-34* intergenic region).

A.2 MATERIALS AND METHODS

A.2.1 Plasmid Construction and Protein Purification

The short and long versions of the mycobacteriophage Brujita and Charlie repressor genes (Gene *34* $gp34^{101}$ & $gp34^{129}$ and gene *35* $gp35^{103}$ & $gp35^{124}$, respectively) were cloned into a pLC3 vector that expresses the genes with an N-terminal maltose binding protein fusion. The resulting plasmids, pVMV18 (Brujita $gp34^{101}$), pVMV19 (Brujita $gp34^{129}$), pVMV16 (Charlie $gp35^{103}$), and pVMV17 (Charlie $gp35^{124}$), were transformed into BL21(DE3)pLysS chemically competent *E. coli* cells (Invitrogen) and confirmed colonies were inoculated into LB/Kan/CM media to saturation. The saturated cultures were used to inoculate new cultures at 0.01 OD_{600} and once cultures reached an optical density of OD_{600}=0.4-0.6, protein expression was induced with addition of 1mM IPTG and incubation at 37°C with shaking for 3 hours. Cells were harvested using centrifugation and pellets were resuspended in 5mL/g with Lysis Buffer (300mM NaCl, 5% Glycerol, 50mM Tris, pH8.0). Resuspended cell pellets were thawed, incubated with protease inhibitors, and lysed using a VirSonic Sonicator (ten 10s pulses at 20-30% output with 30s of ice in between pulses). Lysates were cleared using centrifugation and the resulting soluble lysate was incubated with equilibrated Amylose resin for 2hrs at 4°C. Resin was poured over a disposable BioRad column and washed with Lysis Buffer. Bound protein was eluted using Lysis Buffer + 0.25% Maltose. Eluted fractions containing the purified fusion protein were pooled and dialyzed overnight with TEV Protease (from the VanDemark Lab) to

cleave the MBP tag off of the proteins. The resulting partial purification was used in Binding assays.

A.2.2 DNA binding assays

EMSAs were performed as previously described in section 2.2.3. Substrates were PCR amplified from the lysates of the respective mycobacteriophages (Lysates acquired from Gregory Broussard).

A.3 RESULTS

A.3.1 Brujita gp34 binds to the *34-35* intergenic region

The Brujita $gp34^{101}$ and $gp34^{129}$ were used in EMSAs with a *34-35* intergenic region probe. As expected, both the long and short forms of the Brujita repressor are able to bind the *34-35* intergenic region substrate (**Figure 48A**). The binding is very modest, and only occurs at the highest protein concentrations, and no binding assay has been done on a non-specific DNA substrate for reference, however the concentrations of partially purified Brujita repressor used are unknown and estimated to be quite low based on SDS PAGE analysis.

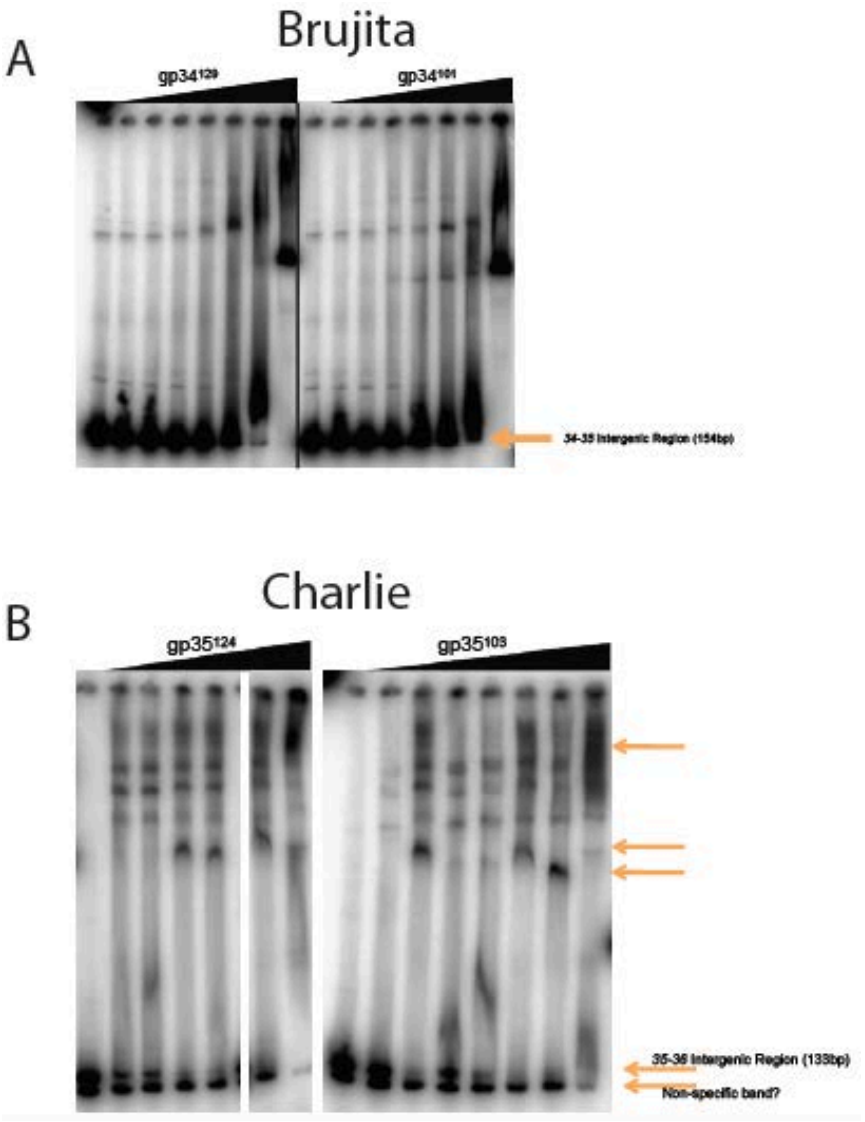

Figure 48 - Brujita and Charlie repressors bind to their respective regulatory intergenic regions

A.3.2 Charlie's gp35 binds to the *35-36* intergenic region and not a non-specific contaminant

The Charlie repressors $gp35^{103}$ and $gp35^{124}$ were also used in EMSAs to the regulatory *35-36* intergenic region. In this assay, a Charlie phage lysate was used as a template for PCR

purification. This lysate was later confirmed to have a contaminating phage present (Gregory Broussard) and therefore there the contaminating band present in the unbound DNA is likely a non-specific substrate that amplified along with the *35-36* intergenic region (**Figure 48B**). With this in mind, Charlie's repressors each bind to one substrate (presumably the *35-36* intergenic region) and not to the other (**Figure 48B**).

A.3.3 BPs gp33[103] is able to bind the Brujita *34-35* intergenic region, but not Charlie's *35-36* intergenic region

In the Brujita *33-34* intergenic region we discovered a 12bp sequence that diverges from the BPs $O_{R\text{-}R}$ by 4 base pairs. We wanted to test if the BPs repressor was able to bind to this region, even though immunity assays have shown that Brujita can grow lytically on a BPs lysogen as well as in strains expressing the active form of the BPs repressor (97). The BPs repressor was used in binding assays with the Brujita and Charlie primary regulatory intergenic regions (**Figure 49**). $Gp33^{103}$ was able to bind to the Brujita *34-35* intergenic region, but not to the Charlie *35-36* intergenic region.

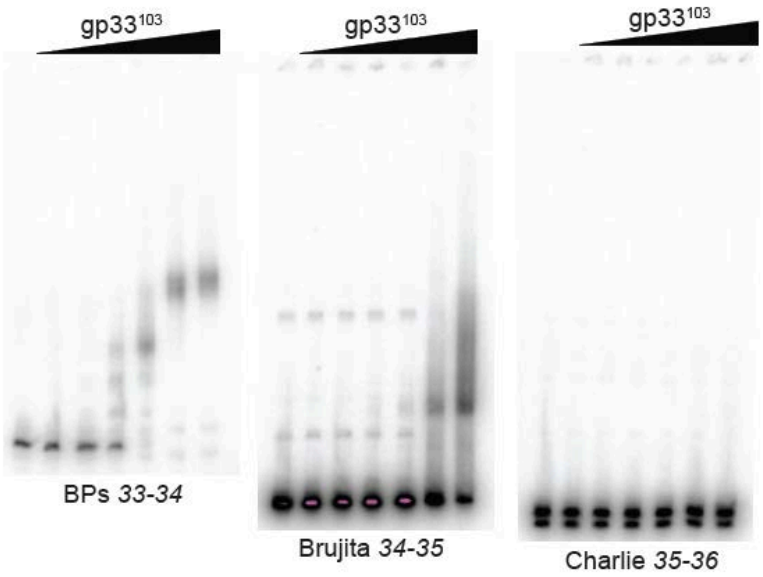

Figure 49 - gp33[103] binds to the Brujita *34-35* intergenic region

A.4 DISCUSSION

Further analysis needs to be done to prove that the Charlie and Brujita repressors bind specifically to their respective primary regulatory regions (**Figure 48**). Some experiments that can be done initially include fully purifying the repressors to get accurate protein concentrations and convincingly show that the binding profiles seen belong to the respective proteins and not to a contaminant in the protein preparations. Also, performing binding assays with non-specific DNA susbtrates to show that these proteins are binding specifically would be useful.

The most confident results we saw came from mycobacteriophage Charlie, because the presence of an unanticipated non-specific band (resulting from an impure phage lysate being used in PCR) shows that the binding of the Charlie repressor is specific (**Figure 48B**). The binding profile of these proteins contain multiple complexes, just as the BPs repressors create multiple complexes with its *33-34* intergenic region.

Interstingly, although BPs and Brujita's repressors do not confer immunity to the other phage, the BPs repressor is able to bind to Brujita's regulatory *34-35* intergenic region (**Figure 49**). The Brujita *34-35* intergenic region contains a 12bp sequence with four base departures from the $O_{R\text{-}R}$ palindrome, and presumable this is where $gp33^{103}$ binds. Two reasons for $gp33^{103}$ being unable to exert an effect include no second half site being located nearby, and the half-site not being situated over the P_R promoter. The Charlie and Brujita repressor proteins could be used to supplement the data gathered on the BPs $gp33^{103}$ to create a more convincing argument for the function of this new class of repressor in all of the clusters this integration-dependent immunity system is found.

APPENDIX B

OTHER PROTEINS INVOLVED IN CONFERRING IMMUNITY IN BPS LYSOGENS

B.1 INTRODUCTION

The first gene of an early lytic BPs operon, gp34, is predicted to have a Cro-like function that antagonizes the function of gp33. The evidence for this was presented by Broussard et. al. in 2013. They created strains expressing gp34 and tested the ability of BPs to infect these strains. A marked decrease of lysogenization frequency occurred, leading to the hypothesis that gp34 is a cro-like protein. Gp34 is small (81aa) and has a predicted helix-turn-helix DNA binding motif. If this protein has a cro-like function, it should compete with $gp33^{103}$ to bind the appropriate sites on the *33-34* intergenic region to reduce P_{rep} activity. In this appendix we describe the attempted purification of BPs gp34 and the binding assays done with a partially purified protein preparation.

B.2 MATERIALS AND METHODS

B.2.1 Plasmid construction and protein purification

The BPs gene *34* was cloned into a pLC3 vector that expresses the gene with a maltose binding protein fusion using directional cloning with NdeI and HindIII restriction enzyme sites. The resulting plasmid, pVMV6, was transformed into BL21(DE3)pLysS and confirmed colonies were inoculated into 5mL LB/Kan/CM media to saturation. The saturated cultures were used to inoculate new cultures at 10% and once cultures reached an optical density of OD_{600}=0.4-0.6, protein expression was induced with addition of 1mM IPTG and incubation at 37°C with shaking for 3 hours. Cells were harvested using centrifugation and pellets were resuspended in 5mL/g with Lysis Buffer (300mM NaCl, 5% Glycerol, 50mM Tris, pH8.0). Resuspended cell pellets were thawed, incubated with protease inhibitors, and lysed using a VirSonic Sonicator (ten 10s pulses at 20-30% output with 30s of ice in between pulses). Lysates were cleared using centrifugation and the resulting soluble lysate was incubated with equilibrated Amylose resin for 2hrs at 4°C. Resin was poured over a disposable BioRad column and washed with Lysis Buffer. Bound protein was eluted using Lysis Buffer + 0.25% Maltose. Eluted fractions containing the purified fusion protein were pooled and dialyzed overnight with TEV Protease to cleave the MBP tag off of the gp34 proteins. The cleaved protein mixture was run over a Nickel column to remove the His-tagged TEV protease and His-tagged MBP tag, however the tag and protease eluted with gp34 and we were unable to separate gp34 from the other proteins. This partially purified protein was used in DNA binding assays as described below.

B.2.2 DNA Binding assays

EMSAs were performed as previously described in section 2.2.3. Substrates were PCR amplified from a lysates mycobacteriophage BPs (Lysates acquired from Gregory Broussard).

B.3 RESULTS

We used a partially purified gp34 protein preparation to do binding assays on various susbtrates (**Figure 50A**). Cro did not bind to a non-specific substrate or to a substrate containing the intergenic region between genes *32-33* which was initially hypothesized to contain a cryptic P_{Int} promoter (although that theory was disproven by Broussard *et al*., 2013) (**Figure 50B**). Cro does bind DNA within the *33-34* intergenic region, however it looks like it binds with an extremely low affinity (**Figure 50B**). Without knowing the concentration of the protein in the preparation, we cannot calculate the exact affinity.

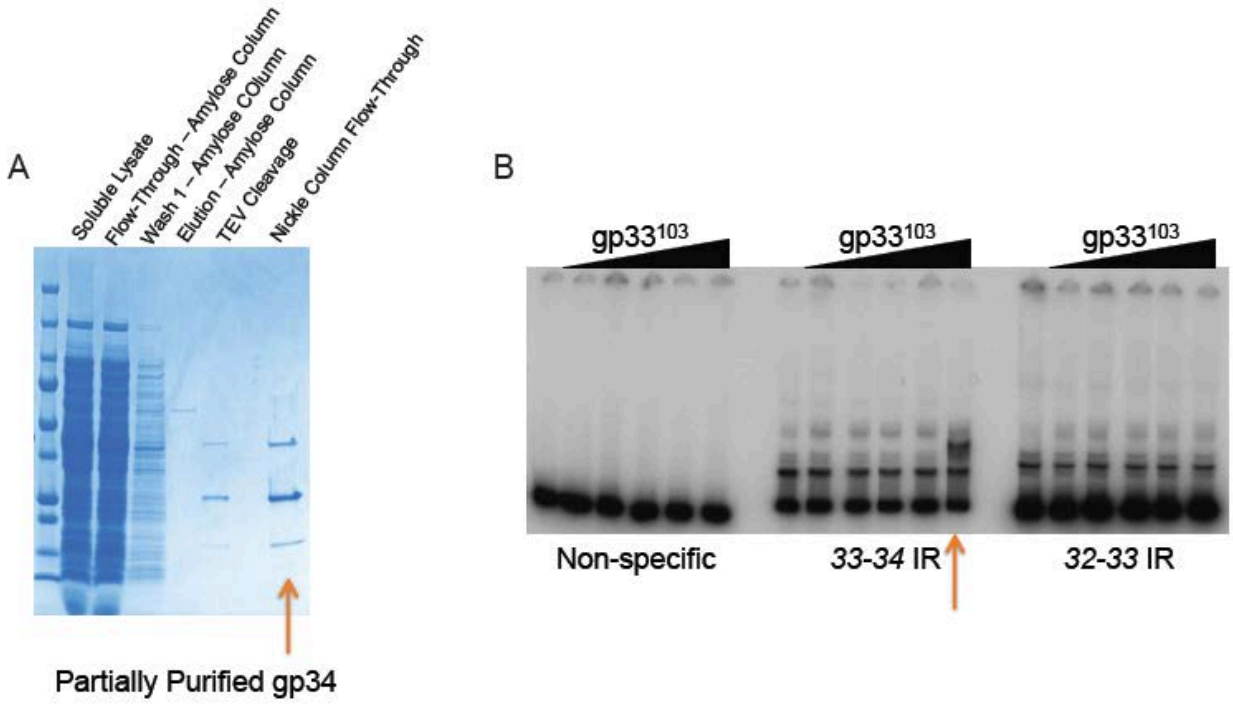

Figure 50 – gp34 purification and pinding to BPs *33-34* intergenic region

Reasons for why we believe that the small shift seen indicates that gp34 truly does bind to the *33-34* intergenic region (**Figure 50B**) include the low proportion of gp34 in the partially purified preparation leading to a low concentration of protein that is not present in enough quanitities to see its full binding profile. Also, the fact that gp34 does not create any shifts when non-specific DNA or DNA without operator sites are used as probes indicates that the shift seen when gp34 is mixed with the *33-34* intergenic region reflects a real protein-DNA interaction.

B.4 DISCUSSION

The BPs gp34 may have a cro-like function in the integration-dependent immunity genetic switches, however there is not currently enough evidence to convincingly report this. Showing that gp34 binds to the intergenic region where $gp33^{103}$ is known to bind is a step in the right direction, but these results need to be solidified. To do this, gp34 must be fully purified and concentrated to a useable concentration. BPs $gp33^{103}$ binds with a K_d of ~2μM, so it is logical to assume that gp34 binds to DNA with a similar binding affinity. If it bound with a higher affinity than $gp33^{103}$, lysogeny would never occur.

It would also be useful to show where the gp34 protein binds within the *33-34* intergenic region to see if it competes with $gp33^{103}$ for binding sites. To do this, DNase I footprinting could be used. Once the gp34 binding sites are identified, conjectures can be made as to how the proteins interact and compete for binding. The characterization of gp34's binding sites and elucidation of the nature of its interaction with $gp33^{103}$ would provide valuable knowledge about how the lytic pathway of the intergration-dependent genetic switches are established.

APPENDIX C

BPS GP33^{103} N-TERMINAL SIX RESIDUES ARE IMPORTANT FOR DNA BINDING

C.1 INTRODUCTION

Before the transcription start site of P_{rep} was confirmed and the promoter was known to be leaderless, gene 33 had a different predicted start site. This gene was 130 aa long instead of 136 and the expressed protein (gp33^{130}) was missing its N-terminal 6 amino acids ([M]SQAFD). We purified this protein using the incorrect start site located 18 base pairs into the now-confirmed correct open reading frame and the short, "active" form of the protein, gp33^{97}, was unable to bind DNA.

It is known that without its first seven residues ([M]STKKKP), the binding affinity of lambda cI decreases 8000-fold because the N-terminus comprises a flexible arm that wraps around the DNA helix and assists in binding specificity (45, 205). This appendix examines the evidence collected that shows that the N-terminal residues of gp33^{103} are important for binding.

C.2 MATERIALS AND METHODS

C.2.1 Plasmid construction and protein purification

The gp33^{97} gene was PCR amplified from a BPs lysate using primers 5'-CAA TCG CCC ATA TGT CGC AAG CAT TCG -3' and 5'-CAA TCG CCC ATA TGT CGC AAG CAT TCG -3', which amplified the gene with a NdeI recognition site at the 5' end of the gene and a HindIII recognition site at the 3' end. The PCR product was digested with NdeI and HindIII fermentas fast digest enzymes and ligated into a maltose-binding fusion vector (pLC3) that was linearized with NdeI and HindIII sites for directional cloning, creating the plasmid pVMV1. Confirmed pVMV1 was transformed into BL21(DE3)star chemically competent cells (Invitrogen) and grown until cultures reached an OD_{600} of 0.4-0.6. Protein expression was induced with 1 mM IPTG at 17°C overnight. Cells were pelleted and frozen at -80°C. Thawed cell pellets were resuspended in 5mL per gram of Lysis Buffer (50 mM Tris pH 8.0, 500 mM NaCl, 8% glycerol, 1 mM EDTA and 1 mM β-mercaptoethanol) and lysed in 200 mL fractions by sonicating 10 times for 10 sec at 30% output with 30 sec of cooling on ice in between bursts. Pooled cell lysates were cleared by centrifugation at 30,000 x g for 40 min at 4°C. Fusion proteins were extracted from soluble cell lysates using amylose resin affinity chromatography (Invitrogen) and the MBP tag was cleaved from the proteins of interest with TEV protease during overnight dialysis at 4°C. MBP and TEV protease contain C-terminal His tags and were removed from the gp33^{97} proteins using nickel affinity chromatography. The flow through containing pure gp33^{97} proteins was dialyzed into a storage buffer (50 mM Tris pH 8.0, 500 mM NaCl, 50% glycerol, 1 mM EDTA, 1 mM BME) and stored at -20°C.

C.2.2 DNA Binding Assays

EMSAs were performed as previously described in section 2.2.3. Substrates were PCR amplified from a BPs lysate. Lysates were acquired with permission from Gregory Broussard.

C.3 RESULTS

C.3.1 $Gp33^{97}$ does not bind DNA, but $gp33^{103}$ does

Before we knew that $gp33^{103}$ was expressed from a leaderless transcript, we purified a shorter form, $gp33^{97}$, using an incorrect start site located 18 base pairs into the correct open reading frame. This form of the protein was unable to bind DNA (**Figure 51**) even at its highest concentration, 5μM. The most obvious explanation for these results is that the missing N-terminal amino acids are important for binding activity, although an alternative explanation is that the low amount of protein, even at the highest concentrations, may not be enough to see binding. We know that the binding affinity of $gp33^{103}$ is very low (~2μM) and 5μM may not be enough protein to see a shift with $gp33^{97}$.

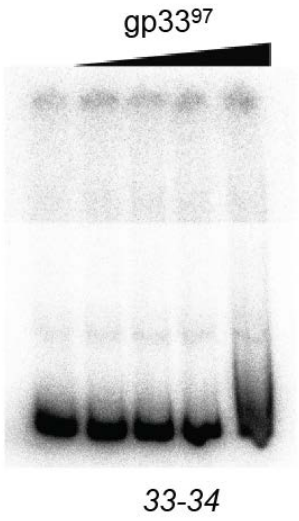

Figure 51 - gp33^{97} does not bind DNA

EMSAs done with gp33^{97} show that the protein only binds DNA very faintly at the highest concentration of protein.

C.3.2 Clear-plaque and repressor-insensitive mutations map to N-terminal gp33^{103} residues

Some repressor-insensitive mutants also suggest that the region between the N-terminus of gp33 and its helix-turn-helix DNA binding motif is important for binding. Broussard et. Al isolated 13 repressor-insensitive and clear plaque mutants that had mutations mapping within the repressor gene. Five of these were located between the N-terminus and the predicted helix-turn-helix DNA binding motif (**Figure 52**). One of these was a substitution of the 6th residue (D6G). Another was an arginine to glycine substitution at the 14th amino acid. Interestingly, three mutants were isolated that had amino acid substitutions in the 11th residue (P11L, P11S, P11S), suggesting an important role for this particular residue in repressor function.

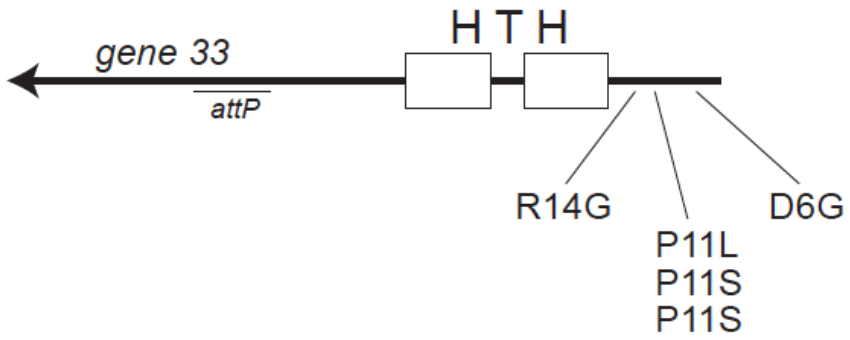

Figure 52 - Repressor insensitive BPs mutants with mutations in the N-terminus of gp33

Diagram shows an arrow as a representation of the repressor gene, white boxes show the helix turn helix motif and the *attP* core is underlined. Mutations are indicated with black lined.

C.4 DISCUSSION

The N-terminus of the BPs repressor protein is important for DNA binding. Removal of the protein's six N-terminal amino acids led to a binding-deficient phenotype at the concentrations used. To explore this result in more detail, the next step in this experiment would be to determine if the protein is truly unable to bind, or if it just binds at a very reduced binging affinity. Since lambad cI's binding is reduced 8,000-fold when it's first 7 residues are missing (206, 207), perhaps gp33^{97} also has a very reduced binding affinity. To test this, a higher concentration of pure protein will be needed and binding assays will need to be done with this more concentrated protein preparation to see if gp33^{97} binds when present at higher concentrations. It would be interesting to see, if gp33^{97} does bind at higher concentrations, if the specificity of binding is affected by removal of the 6 N-terminal residues, if gp33^{97} can still bend DNA, and if the protein still retains its ability to tetramerize. DNase I footprinting, DNA

bending assays, and size exclusion chromatography could be utilized to look at these properties in greater detail.

Results from clear plaque and repressor-insensitive mutants showed that the residues between the N-terminus and helix-turn-helix binding motif are important, especially the 11th residue. Whether this proline makes up part of a flexible arm that helps in DNA recognition like in the lambda cI protein (206, 208-210) or a different aspect of protein-DNA recognition, or if it's somehow involved in a direct interaction with another protein, such as a subunit of RNA polymerase (178, 179), or even if it's involved in something like facilitating tetramerization of the protein (211), it seems to have an important role because without it, gp33 does not function normally and cannot repress lytic genes.

APPENDIX D

CONCENTRATION OF GP33^{103} IN A BPS LYSOGEN

D.1 INTRODUCTION

An interesting question from the perspective of the BPs genetic switch as a whole is what the concentration of gp33^{103} is in a BPs lysogen of *Mycobcaterium smegmatis*. We would predict that the levels of protein needed to maintain lysogeny would be higher than most other phages considering the low binding affinity that gp33^{103} has for its operator sites. We attempted to determine the concentration of the BPs repressor in a lysogen through Western blot assays. This appendix details the experiments done to try to quantify the levels of go33^{103} in a BPs lysogen.

D.2 MATERIALS AND METHODS

D.2.1 Creation of polyclonal antibodies against gp33^{103}

Polyclonal antibodies to gp33^{103} were created through the Pocono Rabbit Farm & lab. We sent 5mg of pure gp33^{103} to the facility and through the 28-day Mighty Quick Protocol with one final

boost from a gel slice on week 29, two rabbits (28951 & 28952) were immunized with the protein three times and bled every two weeks. Titermax was used as an adjuvant because the standard Freud's Complete adjuvant contains *M. tuberculosis* proteins that can be cross-reactive with mycobacteriophage proteins. Serum from the bleeds was collected and sent back to us for analysis. Western Blots showed that antibodies from all the bleeds worked and the final exsanguination was collected and sent to us. Antibody aliquots were stored at -80°C.

D.2.2 Growing *M. smegmatis* mc^{2}155

7H10 agar (Difco) media plates supplemented with carbenicilin and cyclohexamide were streaked with the lab strain of *M. smegmatis* mc^{2}155 (212) as well as a BPs lysogen of *M. smegmatis* mc^{2}155 acquired from Gregory W. Broussard, PhD. Single colonies were picked into 7H9 broth (Difco) supplemented with CB/CHX/ADC and Tween 80 according to standard practice and grown at 37°C overnight with shaking.

D.2.3 Western blot analysis

Western blots were performed as previously described (213). Briefly, protein samples were run on an SDS polyacrylamide gel and transferred to a PVDF membrane (BioRad) using a BioRad wet transfer apparatus according to manufacturer recommendations. The membrane was blocked with TBS + 5% milk for 1-3 hours, washed with TBST three times for 5min, probed with primary antibody in TBS + 5% BSA (28951 or 28952) at a 1:500 dilution for 30-60min, washed with TBST three times for 5min, and probed with anti-mouse AP-conjugated secondary antibody in TBS + 5% milk for 30min. After two quick rinses with TBS, a BCIP/NBT Substrate Kit

(Invitrogen) was used to induce a colorimetric reaction from the secondary antibodies and presence or absence of probed protein was detected. Blots were dried between sheets of filter paper and stored.

Tris-Buffered Saline (TBS): 20mM Tris pH 7.5, 150mM NaCl

TBS with Tween 20 (TBST): 20mM Tris pH 7.5, 150mM NaCl, 0.1% Tween20

Transfer Buffer: 25mM Tris, 190mM Glycine, 20% methanol and pH to 8.3

D.3 RESULTS

D.3.1 Confirmation of BPs lysogen

To confirm that the BPs lysogen was in fact a lysogen, and lysogeny assay was performed. In this assay, a lawn of *M. smegmatis* mc2155 was plated on a 7H10 agar plate supplemented with CB/CHX. The BPs lysogen was streaked on this lawn and the plate was incubated at 37°C for 48 hours (**Figure 53**). This showed cleared areas of the *M. smegmatis* lawn, indicating release of lytic BPs, as well as lysogenic colonies growing in the middle of the zones of clearing, which indicated that the culture is in fact a lysogen. To prove that this was a BPs lysogen, a spot test was also performed as previously described (214).

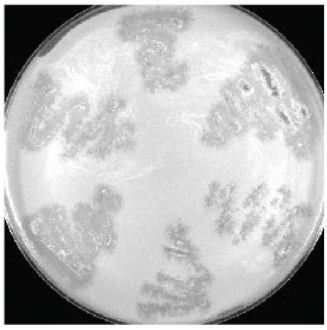

Figure 53 - Confirmation of the BPs lysogen

A BPs lysogen strain was streaked onto a lawn of wild type *M. smegmatis* mc^2155 to look for areas of clearing due to phage release. The strain tested is a confirmed lysogen.

D.3.2 Detection of repressor in a lysogen lysate

Purified $gp33^{103}$ was sent to Pocono Rabbit Farm & Laboratory to create polyclonal rabbit antibodies that recognize the active BPs repressor, $gp33^{103}$. Western blots using the antibodies have shown specific recognition to $gp33^{103}$ and not to negative control phage Bxb2 C-terminal domain (**Figure 54B&C**). Probing the proteins in whole cell lysate of a *Mycobacterium smegmatis* BPs lysogen (which expresses $gp33^{103}$ to maintain the prophage) with the anti-$gp33^{103}$ antibody to check native levels of repressor shows no bands in the lysogen corresponding to $gp33^{103}$ (**Figure 54A**). We looked up the concentration of CI repressor in a lambda lysogen (12) and used that as a template to determine the number of cells needed to get a detectable quantity of $gp33^{103}$ (~50ng). We used $1x10^9$ cells in the lysogen preparation but we still did not detect any bands that were not also present in a wild type *M. smegmatis* lysate.

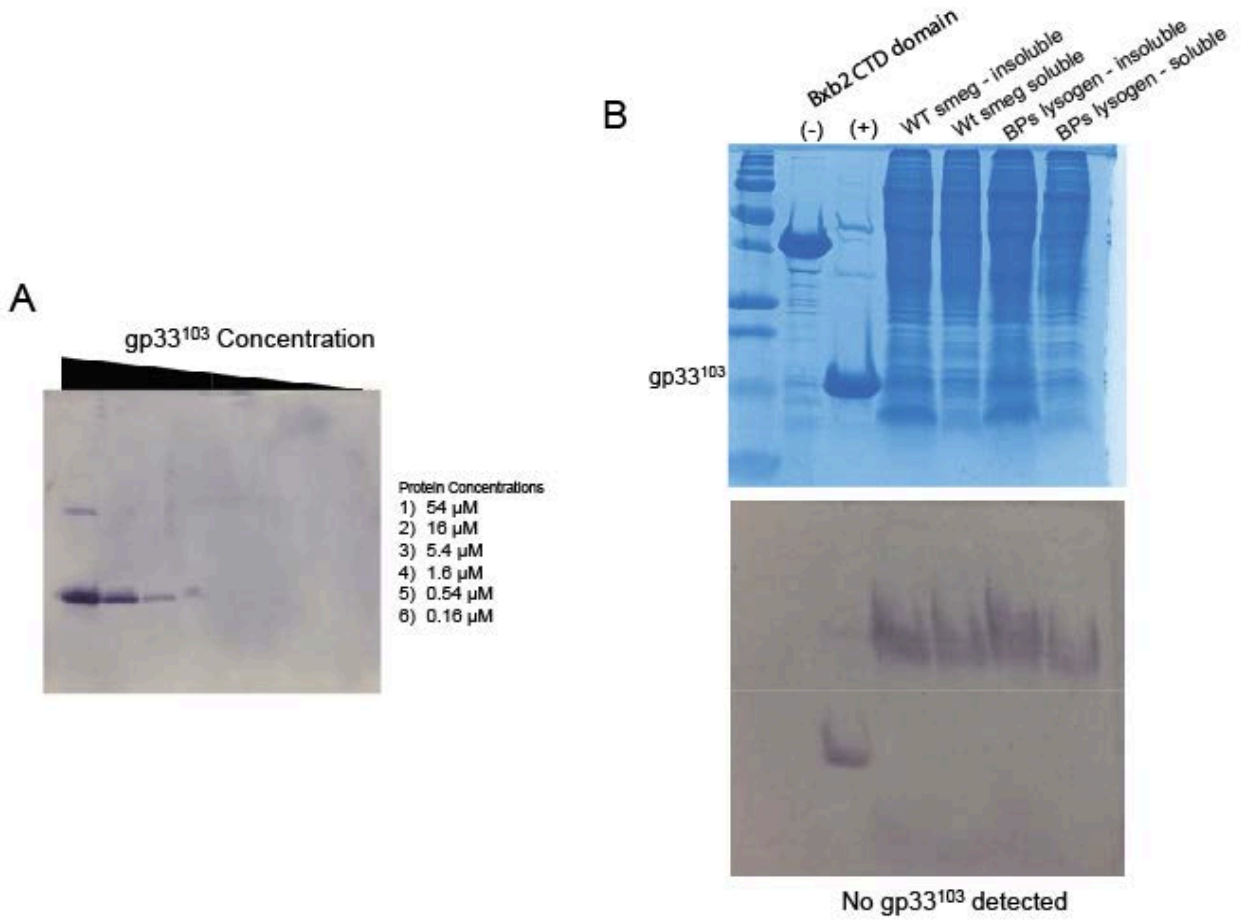

Figure 54 - using an anti-gp33[103] antibody to determine the concentration of repressor in a lysogen

(A) Testing the 28951 anti-gp33[103] antibody on a series of gp33[103] dilutions. (B) Testing the anti-gp33[103] antibody on a BPs lysogen. No gp33[103] was detected.

Gp33[103] is an 11.2 kDa protein that typically runs at 14kDa on an SDS PAGE gel and the band in the lanes corresponding to the BPs lysogen and wild type *M. smegmatis* is approximately 30 kDa. This band is probably recognized because the protein sample sent to Pocono Rabbit Farm & Laboratory was not 100% pure and the co-purified contaminants could have affected antibody production.

D.4 DISCUSSION

Our approach was to use custom-made polyclonal antibodies that recognize the BPs repressor in Western blots to detect $gp33^{103}$ in a lysogen and compare that to known quantities of protein to roughly quantify the protein concentrations. We were unsuccessful in detecting $gp33^{103}$ in a lysogen. This could be for a variety of reasons; either $gp33^{103}$ is not present in high enough levels to meet the detection threshold of the antibodies, $gp33^{103}$ could be insoluble and/or be unrecognizable by my antibody in the context of a whole cell lysate, or perhaps $gp33^{103}$ has a very short half life and is rapidly degraded in the preparation stages before the blot is prepared.

Even though the antibodies we created were not sensitive enough to detect $gp33^{103}$ in a BPs lysogen, these results don't necessarily mean that the *in vitro* work we have done does not pertain to biologically relevant actions of BPs. It is known that bacteriophages localize to poles when they adsorb to a host cell during infection (215) and perhaps when lysogeny occurs, the repressor is localized to the area of the cell where the chromosome is situated, leading to a greater concentration in that area than in the cell as a whole.

These antibodies can be useful in a variety of genetic experiments and can be used in the future by others in the lab studying different aspects of this Integration-dependent genetic switch.

BIBLIOGRAPHY

1. **Bardarov S, Jr., Dou H, Eisenach K, Banaiee N, Ya S, Chan J, Jacobs WR, Jr., Riska PF.** 2003. Detection and drug-susceptibility testing of M. tuberculosis from sputum samples using luciferase reporter phage: comparison with the Mycobacteria Growth Indicator Tube (MGIT) system. Diagn Microbiol Infect Dis **45:**53-61.

2. **Bardarov S, Kriakov J, Carriere C, Yu S, Vaamonde C, McAdam RA, Bloom BR, Hatfull GF, Jacobs WR, Jr.** 1997. Conditionally replicating mycobacteriophages: a system for transposon delivery to Mycobacterium tuberculosis. Proc Natl Acad Sci U S A **94:**10961-10966.

3. **Donnelly-Wu MK, Jacobs WR, Jr., Hatfull GF.** 1993. Superinfection immunity of mycobacteriophage L5: applications for genetic transformation of mycobacteria. Mol Microbiol **7:**407-417.

4. **Hatfull GF, Barsom L, Chang L, Donnelly-Wu M, Lee MH, Levin M, Nesbit C, Sarkis GJ.** 1994. Bacteriophages as tools for vaccine development. Dev Biol Stand **82:**43-47.

5. **Hatfull GF, Jacobs-Sera D, Lawrence JG, Pope WH, Russell DA, Ko CC, Weber RJ, Patel MC, Germane KL, Edgar RH, Hoyte NN, Bowman CA, Tantoco AT, Paladin EC, Myers MS, Smith AL, Grace MS, Pham TT, O'Brien MB, Vogelsberger AM, Hryckowian AJ, Wynalek JL, Donis-Keller H, Bogel MW, Peebles CL, Cresawn SG, Hendrix RW.** 2010. Comparative Genomic Analysis of 60 Mycobacteriophage Genomes: Genome Clustering, Gene Acquisition, and Gene Size. J Mol Biol **397:**119-143.

6. **Huff J, Czyz A, Landick R, Niederweis M.** 2010. Taking phage integration to the next level as a genetic tool for mycobacteria. Gene **468:**8-19.

7. **Lee MH, Pascopella L, Jacobs WR, Jr., Hatfull GF.** 1991. Site-specific integration of mycobacteriophage L5: integration-proficient vectors for Mycobacterium smegmatis, Mycobacterium tuberculosis, and bacille Calmette-Guerin. Proc Natl Acad Sci U S A **88:**3111-3115.

8. **Morris P, Marinelli LJ, Jacobs-Sera D, Hendrix RW, Hatfull GF.** 2008. Genomic characterization of mycobacteriophage Giles: evidence for phage acquisition of host DNA by illegitimate recombination. J Bacteriol **190:**2172-2182.

9. **Murry JP, Rubin EJ.** 2005. New genetic approaches shed light on TB virulence. Trends Microbiol **13:**366-372.

10. **Onozaki I, Raviglione M.** 2010. Stopping tuberculosis in the 21st century: goals and strategies. Respirology **15:**32-43.

11. **Pham TT, Jacobs-Sera D, Pedulla ML, Hendrix RW, Hatfull GF.** 2007. Comparative genomic analysis of mycobacteriophage Tweety: evolutionary insights and construction of compatible site-specific integration vectors for mycobacteria. Microbiology **153:**2711-2723.

12. **Ptashne M, Johnson AD, Pabo CO.** 1982. A genetic switch in a bacterial virus. Sci Am **247:**128-130, 132, 134-140.

13. **Raviglione MC.** 2008. Facing extensively drug-resistant tuberculosis--a hope and a challenge. N Engl J Med **359:**636-638.

14. **Raviglione MC, Smith IM.** 2007. XDR tuberculosis--implications for global public health. N Engl J Med **356:**656-659.

15. **van Kessel JC, Hatfull GF.** 2007. Recombineering in *Mycobacterium tuberculosis*. Nature Methods **4:**147-152.

16. **van Kessel JC, Hatfull GF.** 2008. Efficient point mutagenesis in mycobacteria using single-stranded DNA recombineering: characterization of antimycobacterial drug targets. Mol Microbiol **67:**1094-1107.

17. **van Kessel JC, Hatfull GF.** 2008. Mycobacterial recombineering. Methods Mol Biol **435:**203-215.

18. **van Kessel JC, Marinelli LJ, Hatfull GF.** 2008. Recombineering mycobacteria and their phages. Nat Rev Microbiol **6:**851-857.

19. **Bardarov S, Bardarov Jr S, Jr., Pavelka Jr MS, Jr., Sambandamurthy V, Larsen M, Tufariello J, Chan J, Hatfull G, Jacobs Jr WR, Jr.** 2002. Specialized transduction: an efficient method for generating marked and unmarked targeted gene disruptions in Mycobacterium tuberculosis, M. bovis BCG and M. smegmatis. Microbiology **148:**3007-3017.

20. **Petrova ZO, Broussard GW, Hatfull GF.** 2015. Mycobacteriophage-repressor mediated immunity as selectable genetic markers: Adephagia and BPs repressor-selection. Microbiology doi:10.1099/mic.0.000120.

21. **Pope WH, Bowman CA, Russell DA, Jacobs-Sera D, Asai DJ, Cresawn SG, Jacobs WR, Hendrix RW, Lawrence JG, Hatfull GF, Science Education Alliance Phage Hunters Advancing G, Evolutionary S, Phage Hunters Integrating R, Education, Mycobacterial Genetics C.** 2015. Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. Elife **4:**e06416.

22. **Pope WH, Jacobs-Sera D, Russell DA, Peebles CL, Al-Atrache Z, Alcoser TA, Alexander LM, Alfano MB, Alford ST, Amy NE, Anderson MD, Anderson AG, Ang AA, Ares M, Jr., Barber AJ, Barker LP, Barrett JM, Barshop WD, Bauerle CM, Bayles IM, Belfield KL, Best AA, Borjon A, Jr., Bowman CA, Boyer CA, Bradley KW, Bradley VA, Broadway LN, Budwal K, Busby KN, Campbell IW, Campbell AM, Carey A, Caruso SM, Chew RD, Cockburn CL, Cohen LB, Corajod JM, Cresawn SG, Davis KR, Deng L, Denver DR, Dixon BR, Ekram S, Elgin SC, Engelsen AE, English BE, Erb ML, Estrada C, Filliger LZ, et al.** 2011. Expanding the diversity of mycobacteriophages: insights into genome architecture and evolution. PLoS One **6:**e16329.

23. **Dubnau D, Losick R.** 2006. Bistability in bacteria. Mol Microbiol **61:**564-572.

24. **Corbett EL, Watt CJ, Walker N, Maher D, Williams BG, Raviglione MC, Dye C.** 2003. The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. Arch Intern Med **163:**1009-1021.

25. **Dye C, Watt CJ, Bleed DM, Hosseini SM, Raviglione MC.** 2005. Evolution of tuberculosis control and prospects for reducing tuberculosis incidence, prevalence, and deaths globally. JAMA **293:**2767-2775.

26. **Velmurugan K, Chen B, Miller JL, Azogue S, Gurses S, Hsu T, Glickman M, Jacobs WR, Jr., Porcelli SA, Briken V.** 2007. Mycobacterium tuberculosis nuoG is a virulence gene that inhibits apoptosis of infected host cells. PLoS Pathog **3:**e110.

27. **Pooran A, Pieterson E, Davids M, Theron G, Dheda K.** 2013. What is the cost of diagnosis and management of drug resistant tuberculosis in South Africa? PLoS One **8:**e54587.

28. **McIlleron H, Abdel-Rahman S, Dave JA, Blockman M, Owen A.** 2015. Special populations and pharmacogenetic issues in tuberculosis drug development and clinical research. J Infect Dis **211 Suppl 3:**S115-125.

29. **Honer zu Bentrup K, Russell DG.** 2001. Mycobacterial persistence: adaptation to a changing environment. Trends Microbiol **9:**597-605.

30. **Manabe YC, Bishai WR.** 2000. Latent Mycobacterium tuberculosis-persistence, patience, and winning by waiting. Nat Med **6:**1327-1329.

31. **Hingley-Wilson SM, Sambandamurthy VK, Jacobs WR, Jr.** 2003. Survival perspectives from the world's most successful pathogen, Mycobacterium tuberculosis. Nat Immunol **4:**949-955.

32. **Smith CV, Sharma V, Sacchettini JC.** 2004. TB drug discovery: addressing issues of persistence and resistance. Tuberculosis (Edinb) **84:**45-55.

33. **Parisien A, Allain B, Zhang J, Mandeville R, Lan CQ.** 2008. Novel alternatives to antibiotics: bacteriophages, bacterial cell wall hydrolases, and antimicrobial peptides. J Appl Microbiol **104:**1-13.

34. **Carlton RM.** 1999. Phage therapy: past history and future prospects. Arch Immunol Ther Exp (Warsz) **47:**267-274.

35. **Kutateladze M, Adamia R.** 2010. Bacteriophages as potential new therapeutics to replace or supplement antibiotics. Trends Biotechnol **28:**591-595.

36. **Kutateladze M, Adamia R.** 2008. Phage therapy experience at the Eliava Institute. Med Mal Infect **38:**426-430.

37. **Kutateladze M.** 2015. Experience of the Eliava Institute in bacteriophage therapy. Virol Sin **30:**80-81.

38. **Diez-Martinez R, De Paz HD, Garcia-Fernandez E, Bustamante N, Euler CW, Fischetti VA, Menendez M, Garcia P.** 2015. A novel chimeric phage lysin with high in vitro and in vivo bactericidal activity against Streptococcus pneumoniae. J Antimicrob Chemother **70:**1763-1773.

39. **Lood R, Winer BY, Pelzek AJ, Diez-Martinez R, Thandar M, Euler CW, Schuch R, Fischetti VA.** 2015. Novel phage lysin capable of killing the multidrug-resistant gram-negative bacterium Acinetobacter baumannii in a mouse bacteremia model. Antimicrob Agents Chemother **59:**1983-1991.

40. **Payne K, Sun Q, Sacchettini J, Hatfull GF.** 2009. Mycobacteriophage Lysin B is a novel mycolylarabinogalactan esterase. Mol Microbiol **73:**367-381.

41. **Dabrowska K, Opolski A, Wietrzyk J, Nevozhay D, Szczaurska K, Switala-Jelen K, Boratynski J, Gorski A.** 2005. Activity of bacteriophages in murine tumor models depends on the route of phage administration. Oncol Res **15:**183-187.

42. **Budynek P, Dabrowska K, Skaradzinski G, Gorski A.** 2010. Bacteriophages and cancer. Arch Microbiol **192:**315-320.

43. **Eriksson F, Tsagozis P, Lundberg K, Parsa R, Mangsbo SM, Persson MA, Harris RA, Pisa P.** 2009. Tumor-specific bacteriophages induce tumor destruction through activation of tumor-associated macrophages. J Immunol **182:**3105-3111.

44. **Eriksson F, Culp WD, Massey R, Egevad L, Garland D, Persson MA, Pisa P.** 2007. Tumor specific phage particles promote tumor regression in a mouse melanoma model. Cancer Immunol Immunother **56:**677-687.

45. **Ptashne M.** 2004. A genetic switch: phage lambda revisited, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

46. **Little JW.** 2010. Evolution of complex gene regulatory circuits by addition of refinements. Curr Biol **20:**R724-734.

47. **Oppenheim AB, Kobiler O, Stavans J, Court DL, Adhya S.** 2005. Switches in bacteriophage lambda development. Annu Rev Genet **39:**409-429.

48. **Bloch S, Nejman-Falenczyk B, Los JM, Baranska S, Lepek K, Felczykowska A, Los M, Wegrzyn G, Wegrzyn A.** 2013. Genes from the exo-xis region of lambda and Shiga toxin-converting bacteriophages influence lysogenization and prophage induction. Arch Microbiol **195:**693-703.

49. **Suttle CA.** 2007. Marine viruses--major players in the global ecosystem. Nat Rev Microbiol **5:**801-812.

50. **Hatfull GF, Hendrix RW.** 2011. Bacteriophages and their genomes. Curr Opin Virol **1:**298-303.

51. **Hochschild A, Douhan J, 3rd, Ptashne M.** 1986. How lambda repressor and lambda Cro distinguish between OR1 and OR3. Cell **47:**807-816.

52. **Hochschild A, Ptashne M.** 1986. Homologous interactions of lambda repressor and lambda Cro with the lambda operator. Cell **44:**925-933.

53. **Greenblatt J, Mah TF, Legault P, Mogridge J, Li J, Kay LE.** 1998. Structure and mechanism in transcriptional antitermination by the bacteriophage lambda N protein. Cold Spring Harb Symp Quant Biol **63:**327-336.

54. **Scharpf M, Sticht H, Schweimer K, Boehm M, Hoffmann S, Rosch P.** 2000. Antitermination in bacteriophage lambda. The structure of the N36 peptide-boxB RNA complex. Eur J Biochem **267:**2397-2408.

55. **Mogridge J, Mah TF, Greenblatt J.** 1998. Involvement of boxA nucleotides in the formation of a stable ribonucleoprotein complex containing the bacteriophage lambda N protein. J Biol Chem **273:**4143-4148.

56. **Marr MT, Datwyler SA, Meares CF, Roberts JW.** 2001. Restructuring of an RNA polymerase holoenzyme elongation complex by lambdoid phage Q proteins. Proc Natl Acad Sci U S A **98:**8972-8978.

57. **Johnson A, Meyer BJ, Ptashne M.** 1978. Mechanism of action of the cro protein of bacteriophage lambda. Proc Natl Acad Sci U S A **75:**1783-1787.

58. **Shotland Y, Koby S, Teff D, Mansur N, Oren DA, Tatematsu K, Tomoyasu T, Kessel M, Bukau B, Ogura T, Oppenheim AB.** 1997. Proteolysis of the phage lambda CII regulatory protein by FtsH (HflB) of Escherichia coli. Mol Microbiol **24:**1303-1310.

59. **Shotland Y, Shifrin A, Ziv T, Teff D, Koby S, Kobiler O, Oppenheim AB.** 2000. Proteolysis of bacteriophage lambda CII by Escherichia coli FtsH (HflB). J Bacteriol **182:**3111-3116.

60. **Srinivasan R, Anilkumar G, Rajeswari H, Ajitkumar P.** 2006. Functional characterization of AAA family FtsH protease of Mycobacterium tuberculosis. FEMS Microbiol Lett **259:**97-105.

61. **Hoyt MA, Knight DM, Das A, Miller HI, Echols H.** 1982. Control of phage lambda development by stability and synthesis of cII protein: role of the viral cIII and host hflA, himA and himD genes. Cell **31:**565-573.

62. **Kobiler O, Koby S, Teff D, Court D, Oppenheim AB.** 2002. The phage lambda CII transcriptional activator carries a C-terminal domain signaling for rapid proteolysis. Proc Natl Acad Sci U S A **99:**14964-14969.

63. **Kobiler O, Rokney A, Friedman N, Court DL, Stavans J, Oppenheim AB.** 2005. Quantitative kinetic analysis of the bacteriophage lambda genetic network. Proc Natl Acad Sci U S A **102:**4470-4475.

64. **Ptashne M, Jeffrey A, Johnson AD, Maurer R, Meyer BJ, Pabo CO, Roberts TM, Sauer RT.** 1980. How the lambda repressor and cro work. Cell **19:**1-11.

65. **Humayun Z, Jeffrey A, Ptashne M.** 1977. Completed DNA sequences and organization of repressor-binding sites in the operators of phage lambda. J Mol Biol **112:**265-277.

66. **Humayun Z, Kleid D, Ptashne M.** 1977. Sites of contact between lambda operators and lambda repressor. Nucleic Acids Res **4:**1595-1607.

67. **Pabo CO, Lewis M.** 1982. The operator-binding domain of lambda repressor: structure and DNA recognition. Nature **298:**443-447.

68. **Pabo CO, Sauer RT, Sturtevant JM, Ptashne M.** 1979. The lambda repressor contains two domains. Proc Natl Acad Sci U S A **76:**1608-1612.

69. **Meyer BJ, Maurer R, Ptashne M.** 1980. Gene regulation at the right operator (OR) of bacteriophage lambda. II. OR1, OR2, and OR3: their roles in mediating the effects of repressor and cro. J Mol Biol **139:**163-194.

70. **Pabo CO, Krovatin W, Jeffrey A, Sauer RT.** 1982. The N-terminal arms of lambda repressor wrap around the operator DNA. Nature **298:**441-443.

71. **Pabo CO, Sauer RT.** 1984. Protein-DNA recognition. Annu Rev Biochem **53:**293-321.

72. **Sauer RT, Jordan SR, Pabo CO.** 1990. Lambda repressor: a model system for understanding protein-DNA interactions and protein stability. Adv Protein Chem **40:**1-61.

73. **Meyer BJ, Ptashne M.** 1980. Gene regulation at the right operator (OR) of bacteriophage lambda. III. lambda repressor directly activates gene transcription. J Mol Biol **139:**195-205.

74. **Kedzierska B, Szambowska A, Herman-Antosiewicz A, Lee DJ, Busby SJ, Wegrzyn G, Thomas MS.** 2007. The C-terminal domain of the Escherichia coli RNA polymerase alpha subunit plays a role in the CI-dependent activation of the bacteriophage lambda pM promoter. Nucleic Acids Res **35:**2311-2320.

75. **Joung JK, Koepp DM, Hochschild A.** 1994. Synergistic activation of transcription by bacteriophage lambda cI protein and E. coli cAMP receptor protein. Science **265:**1863-1866.

76. **Ackers GK, Johnson AD, Shea MA.** 1982. Quantitative model for gene regulation by lambda phage repressor. Proc Natl Acad Sci U S A **79:**1129-1133.

77. **Maurer R, Meyer B, Ptashne M.** 1980. Gene regulation at the right operator (OR) bacteriophage lambda. I. OR3 and autogenous negative control by repressor. J Mol Biol **139:**147-161.

78. **Revet B, von Wilcken-Bergmann B, Bessert H, Barker A, Muller-Hill B.** 1999. Four dimers of lambda repressor bound to two suitably spaced pairs of lambda operators form octamers and DNA loops over large distances. Curr Biol **9:**151-154.

79. **Lewis D, Le P, Zurla C, Finzi L, Adhya S.** 2011. Multilevel autoregulation of lambda repressor protein CI by DNA looping in vitro. Proc Natl Acad Sci U S A **108:**14807-14812.

80. **Griffith J, Hochschild A, Ptashne M.** 1986. DNA loops induced by cooperative binding of lambda repressor. Nature **322:**750-752.

81. **Ghosh K, Pal A, Chattopadhyaya R.** 2004. pH-dependent autocleavage of lambda repressor occurs in the operator-bound form: characterization of lambda repressor autocleavage. Biochem J **379:**325-330.

82. **Schubert RA, Dodd IB, Egan JB, Shearwin KE.** 2007. Cro's role in the CI Cro bistable switch is critical for {lambda}'s transition from lysogeny to lytic development. Genes Dev **21:**2461-2472.

83. **Brown KL, Sarkis GJ, Wadsworth C, Hatfull GF.** 1997. Transcriptional silencing by the mycobacteriophage L5 repressor. Embo J **16:**5914-5921.

84. **Jain S, Hatfull GF.** 2000. Transcriptional regulation and immunity in mycobacteriophage Bxb1. Mol Microbiol **38:**971-985.

85. **Sau S, Chattoraj P, Ganguly T, Lee CY, Mandal NC.** 2004. Cloning and sequencing analysis of the repressor gene of temperate mycobacteriophage L1. J Biochem Mol Biol **37:**254-259.

86. **Bandhu A, Ganguly T, Jana B, Mondal R, Sau S.** 2010. Regions and residues of an asymmetric operator DNA interacting with the monomeric repressor of temperate mycobacteriophage L1. Biochemistry **49:**4235-4243.

87. **Browning DF, Busby SJ.** 2004. The regulation of bacterial transcription initiation. Nat Rev Microbiol **2:**57-65.

88. **Ganguly T, Bandhu A, Chattoraj P, Chanda PK, Das M, Mandal NC, Sau S.** 2007. Repressor of temperate mycobacteriophage L1 harbors a stable C-terminal domain and binds to different asymmetric operator DNAs with variable affinity. Virol J **4:**64.

89. **Hatfull GF, Sarkis GJ.** 1993. DNA sequence, structure and gene expression of mycobacteriophage L5: a phage system for mycobacterial genetics. Mol Microbiol **7:**395-405.

90. **Mediavilla J, Jain S, Kriakov J, Ford ME, Duda RL, Jacobs WR, Jr., Hendrix RW, Hatfull GF.** 2000. Genome organization and characterization of mycobacteriophage Bxb1. Mol Microbiol **38:**955-970.

91. **Pedulla ML, Ford ME, Houtz JM, Karthikeyan T, Wadsworth C, Lewis JA, Jacobs-Sera D, Falbo J, Gross J, Pannunzio NR, Brucker W, Kumar V, Kandasamy J, Keenan L, Bardarov S, Kriakov J, Lawrence JG, Jacobs WR, Hendrix RW, Hatfull GF.** 2003. Origins of highly mosaic mycobacteriophage genomes. Cell **113:**171-182.

92. **Belitsky BR, Sonenshein AL.** 2011. Roadblock repression of transcription by Bacillus subtilis CodY. J Mol Biol **411:**729-743.

93. **Colin J, Candelli T, Porrua O, Boulay J, Zhu C, Lacroute F, Steinmetz LM, Libri D.** 2014. Roadblock termination by reb1p restricts cryptic and readthrough transcription. Mol Cell **56:**667-680.

94. **Nemeth A, Perez-Fernandez J, Merkl P, Hamperl S, Gerber J, Griesenbeck J, Tschochner H.** 2013. RNA polymerase I termination: Where is the end? Biochim Biophys Acta **1829:**306-317.

95. **Jacobs-Sera D, Marinelli LJ, Bowman C, Broussard GW, Guerrero Bustamante C, Boyle MM, Petrova ZO, Dedrick RM, Pope WH, Science Education Alliance Phage Hunters Advancing G, Evolutionary Science Sea-Phages P, Modlin RL, Hendrix RW, Hatfull GF.** 2012. On the nature of mycobacteriophage diversity and host preference. Virology **434:**187-201.

96. **Sampson T, Broussard GW, Marinelli LJ, Jacobs-Sera D, Ray M, Ko CC, Russell D, Hendrix RW, Hatfull GF.** 2009. Mycobacteriophages BPs, Angel and Halo: comparative genomics reveals a novel class of ultra-small mobile genetic elements. Microbiology **155:**2962-2977.

97. **Broussard GW, Oldfield LM, Villanueva VM, Lunt BL, Shine EE, Hatfull GF.** 2013. Integration-dependent bacteriophage immunity provides insights into the evolution of genetic switches. Mol Cell **49:**237-248.

98. **Hatfull GF.** 2012. The secret lives of mycobacteriophages. Adv Virus Res **82:**179-288.

99. **Hatfull GF, Pedulla ML, Jacobs-Sera D, Cichon PM, Foley A, Ford ME, Gonda RM, Houtz JM, Hryckowian AJ, Kelchner VA, Namburi S, Pajcini KV, Popovich MG, Schleicher DT, Simanek BZ, Smith AL, Zdanowicz GM, Kumar V, Peebles CL, Jacobs WR, Jr., Lawrence JG, Hendrix RW.** 2006. Exploring the mycobacteriophage metaproteome: phage genomics as an educational platform. PLoS Genet **2:**e92.

100. **Flynn JM, Levchenko I, Seidel M, Wickner SH, Sauer RT, Baker TA.** 2001. Overlapping recognition determinants within the ssrA degradation tag allow modulation of proteolysis. Proc Natl Acad Sci U S A **98:**10584-10589.

101. **Gottesman S, Roche E, Zhou Y, Sauer RT.** 1998. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. Genes Dev **12:**1338-1347.

102. **Oldfield LM, Hatfull GF.** 2014. Mutational analysis of the mycobacteriophage BPs promoter PR reveals context-dependent sequences for mycobacterial gene expression. J Bacteriol **196:**3589-3597.

103. **Oldfield LM.** 2015. Gene Expression in Mycobacteriophage BPs. PhD - Doctor of Philosophy. University of Pittsburgh.

104. **Villanueva VM, Oldfield LM, Hatfull GF.** 2015. An Unusual Phage Repressor Encoded by Mycobacteriophage BPs. PLoS One **10:**e0137187.

105. **Cortes T, Schubert OT, Rose G, Arnvig KB, Comas I, Aebersold R, Young DB.** 2013. Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in Mycobacterium tuberculosis. Cell Rep **5:**1121-1131.

106. **Campbell EA, Korzheva N, Mustaev A, Murakami K, Nair S, Goldfarb A, Darst SA.** 2001. Structural mechanism for rifampicin inhibition of bacterial rna polymerase. Cell **104:**901-912.

107. **Gopal B, Haire LF, Gamblin SJ, Dodson EJ, Lane AN, Papavinasasundaram KG, Colston MJ, Dodson G.** 2001. Crystal structure of the transcription elongation/anti-termination factor NusA from Mycobacterium tuberculosis at 1.7 A resolution. J Mol Biol **314:**1087-1095.

108. **Josa D, da Cunha EF, Ramalho TC, Souza TC, Caetano MS.** 2008. Homology modeling of wild-type, D516V, and H526L Mycobacterium tuberculosis RNA polymerase and their molecular docking study with inhibitors. J Biomol Struct Dyn **25:**373-376.

109. **Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K, Darst SA.** 1999. Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. Cell **98:**811-824.

110. **Vassylyev DG, Sekine S, Laptenko O, Lee J, Vassylyeva MN, Borukhov S, Yokoyama S.** 2002. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. Nature **417:**712-719.

111. **Archambault J, Friesen JD.** 1993. Genetics of eukaryotic RNA polymerases I, II, and III. Microbiol Rev **57:**703-724.

112. **deHaseth PL, Lohman TM, Burgess RR, Record MT, Jr.** 1978. Nonspecific interactions of Escherichia coli RNA polymerase with native and denatured DNA: differences in the binding behavior of core and holoenzyme. Biochemistry **17:**1612-1622.

113. **Callaci S, Heyduk T.** 1998. Conformation and DNA binding properties of a single-stranded DNA binding region of sigma 70 subunit from Escherichia coli RNA polymerase are modulated by an interaction with the core enzyme. Biochemistry **37:**3312-3320.

114. **Dombroski AJ, Walter WA, Record MT, Jr., Siegele DA, Gross CA.** 1992. Polypeptides containing highly conserved regions of transcription initiation factor sigma 70 exhibit specificity of binding to promoter DNA. Cell **70:**501-512.

115. **Bashyam MD, Hasnain SE.** 2004. The extracytoplasmic function sigma factors: role in bacterial pathogenesis. Infect Genet Evol **4:**301-308.

116. **Kazmierczak MJ, Wiedmann M, Boor KJ.** 2005. Alternative sigma factors and their roles in bacterial virulence. Microbiol Mol Biol Rev **69:**527-543.

117. **Helmann JD, Chamberlin MJ.** 1988. Structure and function of bacterial sigma factors. Annu Rev Biochem **57:**839-872.

118. **Gomez JE, Chen JM, Bishai WR.** 1997. Sigma factors of Mycobacterium tuberculosis. Tuber Lung Dis **78:**175-183.

119. **Muller A, Hoffmann JH, Meyer HE, Narberhaus F, Jakob U, Leichert LI.** 2013. Nonnative disulfide bond formation activates the sigma32-dependent heat shock response in Escherichia coli. J Bacteriol **195:**2807-2816.

120. **Osterberg S, del Peso-Santos T, Shingler V.** 2011. Regulation of alternative sigma factor use. Annu Rev Microbiol **65:**37-55.

121. **Manganelli R, Voskuil MI, Schoolnik GK, Dubnau E, Gomez M, Smith I.** 2002. Role of the extracytoplasmic-function sigma factor sigma(H) in Mycobacterium tuberculosis global gene expression. Mol Microbiol **45:**365-374.

122. **Waagmeester A, Thompson J, Reyrat JM.** 2005. Identifying sigma factors in Mycobacterium smegmatis by comparative genomic analysis. Trends Microbiol **13:**505-509.

123. **Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE, 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG.** 1998. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature **393:**537-544.

124. **Tekaia F, Gordon SV, Garnier T, Brosch R, Barrell BG, Cole ST.** 1999. Analysis of the proteome of Mycobacterium tuberculosis in silico. Tuber Lung Dis **79:**329-342.

125. **Gomez M, Doukhan L, Nair G, Smith I.** 1998. sigA is an essential gene in Mycobacterium smegmatis. Mol Microbiol **29:**617-628.

126. **Lonetto M, Gribskov M, Gross CA.** 1992. The sigma 70 family: sequence conservation and evolutionary relationships. J Bacteriol **174:**3843-3849.

127. **Kumar A, Malloch RA, Fujita N, Smillie DA, Ishihama A, Hayward RS.** 1993. The minus 35-recognition region of Escherichia coli sigma 70 is inessential for initiation of transcription at an "extended minus 10" promoter. J Mol Biol **232:**406-418.

128. **Bashyam MD, Kaushal D, Dasgupta SK, Tyagi AK.** 1996. A study of mycobacterial transcriptional apparatus: identification of novel features in promoter elements. J Bacteriol **178:**4847-4853.

129. **China A, Tare P, Nagaraja V.** 2010. Comparison of promoter-specific events during transcription initiation in mycobacteria. Microbiology **156:**1942-1952.

130. **McClure WR.** 1985. Mechanism and control of transcription initiation in prokaryotes. Annu Rev Biochem **54:**171-204.

131. **Korzheva N, Mustaev A, Kozlov M, Malhotra A, Nikiforov V, Goldfarb A, Darst SA.** 2000. A structural model of transcription elongation. Science **289:**619-625.

132. **Murakami KS, Darst SA.** 2003. Bacterial RNA polymerases: the wholo story. Curr Opin Struct Biol **13:**31-39.

133. **Oehler S, Eismann ER, Kramer H, Muller-Hill B.** 1990. The three operators of the lac operon cooperate in repression. EMBO J **9:**973-979.

134. **Platt T.** 1986. Transcription termination and the regulation of gene expression. Annu Rev Biochem **55:**339-372.

135. **Gusarov I, Nudler E.** 1999. The mechanism of intrinsic transcription termination. Mol Cell **3:**495-504.

136. **McDowell JC, Roberts JW, Jin DJ, Gross C.** 1994. Determination of intrinsic transcription termination efficiency by RNA polymerase elongation rate. Science **266:**822-825.

137. **Garvie CW, Wolberger C.** 2001. Recognition of specific DNA sequences. Mol Cell **8:**937-946.

138. **Schumacher MA, Choi KY, Zalkin H, Brennan RG.** 1994. Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices. Science **266:**763-770.

139. **Lewis M, Chang G, Horton NC, Kercher MA, Pace HC, Schumacher MA, Brennan RG, Lu P.** 1996. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. Science **271:**1247-1254.

140. **Van Roey P, Waddling CA, Fox KM, Belfort M, Derbyshire V.** 2001. Intertwined structure of the DNA-binding domain of intron endonuclease I-TevI with its substrate. EMBO J **20:**3631-3637.

141. **Edgell DR, Derbyshire V, Van Roey P, LaBonne S, Stanger MJ, Li Z, Boyd TM, Shub DA, Belfort M.** 2004. Intron-encoded homing endonuclease I-TevI also functions as a transcriptional autorepressor. Nat Struct Mol Biol **11:**936-944.

142. **Shen BW, Landthaler M, Shub DA, Stoddard BL.** 2004. DNA binding and cleavage by the HNH homing endonuclease I-HmuI. J Mol Biol **342:**43-56.

143. **Shen A, Higgins DE, Panne D.** 2009. Recognition of AT-rich DNA binding sites by the MogR repressor. Structure **17:**769-777.

144. **Daniels DS, Woo TT, Luu KX, Noll DM, Clarke ND, Pegg AE, Tainer JA.** 2004. DNA binding and nucleotide flipping by the human DNA repair protein AGT. Nat Struct Mol Biol **11:**714-720.

145. **Ferre-D'Amare AR, Pognonec P, Roeder RG, Burley SK.** 1994. Structure and function of the b/HLH/Z domain of USF. EMBO J **13:**180-189.

146. **Ma PC, Rould MA, Weintraub H, Pabo CO.** 1994. Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. Cell **77:**451-459.

147. **Nair SK, Burley SK.** 2003. X-ray structures of Myc-Max and Mad-Max recognizing DNA. Molecular bases of regulation by proto-oncogenic transcription factors. Cell **112:**193-205.

148. **Parraga A, Bellsolell L, Ferre-D'Amare AR, Burley SK.** 1998. Co-crystal structure of sterol regulatory element binding protein 1a at 2.3 A resolution. Structure **6:**661-672.

149. **Kim Y, Geiger JH, Hahn S, Sigler PB.** 1993. Crystal structure of a yeast TBP/TATA-box complex. Nature **365:**512-520.

150. **Kim JL, Nikolov DB, Burley SK.** 1993. Co-crystal structure of TBP recognizing the minor groove of a TATA element. Nature **365:**520-527.

151. **Pavletich NP, Pabo CO.** 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science **252:**809-817.

152. **Rohs R, West SM, Liu P, Honig B.** 2009. Nuance in the double-helix and its role in protein-DNA recognition. Curr Opin Struct Biol **19:**171-177.

153. **Shakked Z, Rabinovich D.** 1986. The effect of the base sequence on the fine structure of the DNA double helix. Prog Biophys Mol Biol **47:**159-195.

154. **Rohs R, West SM, Sosinsky A, Liu P, Mann RS, Honig B.** 2009. The role of DNA shape in protein-DNA recognition. Nature **461:**1248-1253.

155. **Leslie AG, Arnott S, Chandrasekaran R, Ratliff RL.** 1980. Polymorphism of DNA double helices. J Mol Biol **143:**49-72.

156. **Drew HR, Wing RM, Takano T, Broka C, Tanaka S, Itakura K, Dickerson RE.** 1981. Structure of a B-DNA dodecamer: conformation and dynamics. Proc Natl Acad Sci U S A **78:**2179-2183.

157. **Shakked Z, Guerstein-Guzikevich G, Eisenstein M, Frolow F, Rabinovich D.** 1989. The conformation of the DNA double helix in the crystal is dependent on its environment. Nature **342:**456-460.

158. **Ng HL, Kopka ML, Dickerson RE.** 2000. The structure of a stable intermediate in the A <--> B DNA helix transition. Proc Natl Acad Sci U S A **97:**2035-2039.

159. **Eisenstein M, Shakked Z.** 1995. Hydration patterns and intermolecular interactions in A-DNA crystal structures. Implications for DNA recognition. J Mol Biol **248:**662-678.

160. **Nelson HC, Finch JT, Luisi BF, Klug A.** 1987. The structure of an oligo(dA).oligo(dT) tract and its biological implications. Nature **330:**221-226.

161. **Hizver J, Rozenberg H, Frolow F, Rabinovich D, Shakked Z.** 2001. DNA bending by an adenine--thymine tract and its role in gene regulation. Proc Natl Acad Sci U S A **98:**8490-8495.

162. **Haran TE, Mohanty U.** 2009. The unique structure of A-tracts and intrinsic DNA bending. Q Rev Biophys **42:**41-81.

163. **Goodsell DS, Kaczor-Grzeskowiak M, Dickerson RE.** 1994. The crystal structure of C-C-A-T-T-A-A-T-G-G. Implications for bending of B-DNA at T-A steps. J Mol Biol **239:**79-96.

164. **Seeman NC, Rosenberg JM, Rich A.** 1976. Sequence-specific recognition of double helical nucleic acids by proteins. Proc Natl Acad Sci U S A **73:**804-808.

165. **Harrison SC, Aggarwal AK.** 1990. DNA recognition by proteins with the helix-turn-helix motif. Annu Rev Biochem **59:**933-969.

166. **Aggarwal AK, Rodgers DW, Drottar M, Ptashne M, Harrison SC.** 1988. Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. Science **242:**899-907.

167. **Wolberger C, Dong YC, Ptashne M, Harrison SC.** 1988. Structure of a phage 434 Cro/DNA complex. Nature **335:**789-795.

168. **Travers AA.** 1989. DNA conformation and protein binding. Annu Rev Biochem **58:**427-452.

169. **Folta-Stogniew E.** 2006. Oligomeric states of proteins determined by size-exclusion chromatography coupled with light scattering, absorbance, and refractive index detectors. Methods Mol Biol **328:**97-112.

170. **Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K.** 1996. Current Protocols in Molecular Biology. Wiley Intersciences, New York.

171. **Garner MM, Revzin A.** 1981. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. Nucleic Acids Res **9:**3047-3060.

172. **Porath J, Flodin P.** 1959. Gel filtration: a method for desalting and group separation. Nature **183:**1657-1659.

173. **Max KE, Zeeb M, Bienert R, Balbach J, Heinemann U.** 2007. Common mode of DNA binding to cold shock domains. Crystal structure of hexathymidine bound to the domain-swapped form of a major cold shock protein from Bacillus caldolyticus. FEBS J **274:**1265-1279.

174. **Hampshire AJ, Rusling DA, Broughton-Head VJ, Fox KR.** 2007. Footprinting: a method for determining the sequence selectivity, affinity and kinetics of DNA-binding ligands. Methods **42:**128-140.

175. **Eckert RL.** 2001. DNA sequencing by the chemical method. Curr Protoc Mol Biol **Chapter 7:**Unit7 5.

176. **Maxam AM, Gilbert W.** 1977. A new method for sequencing DNA. Proc Natl Acad Sci U S A **74:**560-564.

177. **Shea MA, Ackers GK.** 1985. The OR control system of bacteriophage lambda. A physical-chemical model for gene regulation. J Mol Biol **181:**211-230.

178. **Hwang JJ, Gussin GN.** 1988. Interactions between Escherichia coli RNA polymerase and lambda repressor. Mutations in PRM affect repression of PR. J Mol Biol **200:**735-739.

179. **Liu B, Shadrin A, Sheppard C, Mekler V, Xu Y, Severinov K, Matthews S, Wigneshweraraj S.** 2014. A bacteriophage transcription regulator inhibits bacterial transcription initiation by sigma-factor displacement. Nucleic Acids Res **42:**4294-4305.

180. **Anthony JR, Green HA, Donohue TJ.** 2003. Purification of Rhodobacter sphaeroides RNA polymerase and its sigma factors. Methods Enzymol **370:**54-65.

181. **China A, Nagaraja V.** 2010. Purification of RNA polymerase from mycobacteria for optimized promoter-polymerase interactions. Protein Expr Purif **69:**235-242.

182. **Burgess RR.** 2009. Important but little known (or forgotten) artifacts in protein biochemistry. Methods Enzymol **463:**813-820.

183. **Thompson NE, Foley KM, Stalder ES, Burgess RR.** 2009. Identification, production, and use of polyol-responsive monoclonal antibodies for immunoaffinity chromatography. Methods Enzymol **463:**475-494.

184. **Thompson NE, Hager DA, Burgess RR.** 1992. Isolation and characterization of a polyol-responsive monoclonal antibody useful for gentle purification of Escherichia coli RNA polymerase. Biochemistry **31:**7003-7008.

185. **Bergendahl V, Thompson NE, Foley KM, Olson BM, Burgess RR.** 2003. A cross-reactive polyol-responsive monoclonal antibody useful for isolation of core RNA polymerase from many bacterial species. Protein Expr Purif **31:**155-160.

186. **Wu S, Howard ST, Lakey DL, Kipnis A, Samten B, Safi H, Gruppo V, Wizel B, Shams H, Basaraba RJ, Orme IM, Barnes PF.** 2004. The principal sigma factor sigA mediates enhanced growth of Mycobacterium tuberculosis in vivo. Mol Microbiol **51:**1551-1562.

187. **Kuznedelov K, Minakhin L, Niedziela-Majka A, Dove SL, Rogulja D, Nickels BE, Hochschild A, Heyduk T, Severinov K.** 2002. A role for interaction of the RNA polymerase flap domain with the sigma subunit in promoter recognition. Science **295:**855-857.

188. **Kim J, Zwieb C, Wu C, Adhya S.** 1989. Bending of DNA by gene-regulatory proteins: construction and use of a DNA bending vector. Gene **85:**15-23.

189. **Zwieb C, Kim J, Adhya S.** 1989. DNA bending by negative regulatory proteins: Gal and Lac repressors. Genes Dev **3:**606-611.

190. **Matlova R, Horska K, Benes V, Cvekl A, Sponar J.** 1995. Protein-induced bending or flexing at the 5'-end of the duck beta A-globin promoter. Gene **156:**277-281.

191. **Zwieb C, Adhya S.** 2009. Plasmid vectors for the analysis of protein-induced DNA bending. Methods Mol Biol **543:**547-562.

192. **Chen JM, Ren H, Shaw JE, Wang YJ, Li M, Leung AS, Tran V, Berbenetz NM, Kocincova D, Yip CM, Reyrat JM, Liu J.** 2008. Lsr2 of Mycobacterium tuberculosis is a DNA-bridging protein. Nucleic Acids Res **36:**2123-2135.

193. **Cui L, Murchland I, Shearwin KE, Dodd IB.** 2013. Enhancer-like long-range transcriptional activation by lambda CI-mediated DNA looping. Proc Natl Acad Sci U S A **110:**2922-2927.

194. **Benson N, Youderian P.** 1989. Phage lambda Cro protein and cI repressor use two different patterns of specific protein-DNA interactions to achieve sequence specificity in vivo. Genetics **121:**5-12.

195. **Li M, Moyle H, Susskind MM.** 1994. Target of the transcriptional activation function of phage lambda cI protein. Science **263:**75-77.

196. **Arthur TM, Burgess RR.** 1998. Localization of a sigma70 binding site on the N terminus of the Escherichia coli RNA polymerase beta' subunit. J Biol Chem **273:**31381-31387.

197. **Probasco MD, Thompson NE, Burgess RR.** 2007. Immunoaffinity purification and characterization of RNA polymerase from Shewanella oneidensis. Protein Expr Purif **55:**23-30.

198. **Breyer MJ, Thompson NE, Burgess RR.** 1997. Identification of the epitope for a highly cross-reactive monoclonal antibody on the major sigma factor of bacterial RNA polymerase. J Bacteriol **179:**1404-1408.

199. **Levin ME, Hatfull GF.** 1993. Mycobacterium smegmatis RNA polymerase: DNA supercoiling, action of rifampicin and mechanism of rifampicin resistance. Mol Microbiol **8:**277-285.

200. **Hatfull GF, Noble SM, Grindley ND.** 1987. The gamma delta resolvase induces an unusual DNA structure at the recombinational crossover point. Cell **49:**103-110.

201. **Eisenberg S, Korza G, Carson J, Liachko I, Tye BK.** 2009. Novel DNA binding properties of the Mcm10 protein from Saccharomyces cerevisiae. J Biol Chem **284:**25412-25420.

202. **Updegrove TB, Correia JJ, Chen Y, Terry C, Wartell RM.** 2011. The stoichiometry of the Escherichia coli Hfq protein bound to RNA. RNA **17:**489-500.

203. **Plocinski P, Laubitz D, Cysewski D, Stodus K, Kowalska K, Dziembowski A.** 2014. Identification of protein partners in mycobacteria using a single-step affinity purification method. PLoS One **9:**e91380.

204. **Broussard GW, Hatfull GF.** 2013. Evolution of genetic switch complexity. Bacteriophage **3:**e24186.

205. **Kim YI, Hu JC.** 1995. Operator binding by lambda repressor heterodimers with one or two N-terminal arms. Proc Natl Acad Sci U S A **92:**7510-7514.

206. **Eliason JL, Weiss MA, Ptashne M.** 1985. NH2-terminal arm of phage lambda repressor contributes energy and specificity to repressor binding and determines the effects of operator mutations. Proc Natl Acad Sci U S A **82:**2339-2343.

207. **Weiss MA, Sauer RT, Patel DJ, Karplus M.** 1984. Amino-terminal arm of the lambda repressor: a 1H NMR study. Biochemistry **23:**5090-5095.

208. **Pabo CO, Sauer RT.** 1992. Transcription factors: structural families and principles of DNA recognition. Annu Rev Biochem **61:**1053-1095.

209. **Clarke ND, Beamer LJ, Goldberg HR, Berkower C, Pabo CO.** 1991. The DNA binding arm of lambda repressor: critical contacts from a flexible region. Science **254:**267-270.

210. **Sarai A, Takeda Y.** 1989. Lambda repressor recognizes the approximately 2-fold symmetric half-operator sequences asymmetrically. Proc Natl Acad Sci U S A **86:**6513-6517.

211. **Biswas A, Mandal S, Sau S.** 2014. The N-terminal domain of the repressor of Staphylococcus aureus phage Phi11 possesses an unusual dimerization ability and DNA binding affinity. PLoS One **9:**e95012.

212. **Jacobs WR, Jr., Kalpana GV, Cirillo JD, Pascopella L, Snapper SB, Udani RA, Jones W, Barletta RG, Bloom BR.** 1991. Genetic systems for mycobacteria. Methods Enzymol **204:**537-555.

213. **Taylor SC, Posch A.** 2014. The design of a quantitative western blot experiment. Biomed Res Int **2014:**361590.

214. **Pope WH, Ferreira CM, Jacobs-Sera D, Benjamin RC, Davis AJ, DeJong RJ, Elgin SC, Guilfoile FR, Forsyth MH, Harris AD, Harvey SE, Hughes LE, Hynes PM, Jackson AS, Jalal MD, MacMurray EA, Manley CM, McDonough MJ, Mosier JL, Osterbann LJ, Rabinowitz HS, Rhyan CN, Russell DA, Saha MS, Shaffer CD, Simon SE, Sims EF, Tovar IG, Weisser EG, Wertz JT, Weston-Hafer KA, Williamson KE, Zhang B, Cresawn SG, Jain P, Piuri M, Jacobs WR, Jr., Hendrix RW, Hatfull GF.** 2011. Cluster K mycobacteriophages: insights into the evolutionary origins of mycobacteriophage TM4. PLoS One **6:**e26750.

215. **Edgar R, Rokney A, Feeney M, Semsey S, Kessel M, Goldberg MB, Adhya S, Oppenheim AB.** 2008. Bacteriophage infection is targeted to cellular poles. Mol Microbiol **68:**1107-1116.