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
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 (gp33136) 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 (gp33103), 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

PREFACE........................................................................................................................................ XIV

1.0 INTRODUCTION..................................................................................................................... 1

1.1 WHY STUDY MYCOBACTERIOPHAGES? ................................................................. 2

1.2 GENETIC SWITCHES CONTROLLING TEMPERATE PHAGE LIFECYCLES...................... 3

1.2.1 Bacteriophage Lambda ....................................................................................... 4

1.2.2 Mycobacteriophage L5 .................................................................................... 8

1.2.3 Mycobacteriophage BPs and Integration-Dependent Immunity ...................... 9

1.2.3.1 Genetic Switch .......................................................................................... 9

1.2.3.2 Gene 33-34 Intergenic Regulatory Region ............................................. 12

1.3 PROKARYOTIC TRANSCRIPTION .............................................................................. 13

1.3.1 Bacterial RNA polymerase ........................................................................... 13

1.3.2 Sigma factors .................................................................................................... 14

1.3.3 Initiation, elongation, and termination ............................................................ 15

1.4 PROTEIN-DNA INTERACTIONS .............................................................................. 16

1.4.1 What structures define a DNA-binding protein? ............................................ 16

1.4.2 DNA structure affects protein binding ............................................................ 17

1.4.3 How proteins recognize DNA substrates ....................................................... 17
2.0 CHARACTERIZING THE DNA-BINDING CAPABILITIES OF GP33$^{103}$ AND GP33$^{136}$ 18

2.1 INTRODUCTION ............................................................................................. 18

2.2 MATERIALS AND METHODS ...................................................................... 19

2.2.1 Expression and Purification of Mycobacteriophage BPs gp33$^{103}$ and gp33$^{136}$ 19

2.2.2 Electrophoretic Mobility Shift Assays ...................................................... 20

2.2.3 Creation and radiolabeling of DNA substrates ......................................... 20

2.2.4 Size Exclusion Chromatography ............................................................... 22

2.2.5 Substrates used in this Chapter 2 .............................................................. 22

2.3 RESULTS ........................................................................................................... 24

2.3.1 Mycobacteriophage BPs repressor binds DNA......................................... 24

2.3.1.1 gp33$^{103}$ and gp33$^{136}$ binding to the gene 33-34 intergenic region .... 24

2.3.1.2 Binding in the rest of the BPs genome............................................... 27

2.3.2 Confirmation of the gp33 operator. ............................................................ 29

2.3.2.1 Unradiolabeled operator DNA competes out gp33$^{103}$ binding ....... 29

2.3.2.2 Two single point mutations in the 12bp OR lead to repressor insensitive phage phenotypes ................................................................. 31

2.3.3 Multimeric state of the Mycobacteriophage BPs repressor in solution. 32

2.3.4 Gp33$^{103}$ binds to other DNA substrates ................................................. 34

2.3.4.1 Gp33$^{103}$ binds to ssDNA ................................................................... 34

2.3.4.2 Gp33$^{103}$ binds to operator half-sites ................................................. 36
2.3.4.3 Native sequence surrounding $O_R$ is important for major complex formation ........................................................................................................... 37

2.4 DISCUSSION ..................................................................................................... 39

3.0 MAPPING THE SEQUENCE REQUIREMENTS FOR GP33$^{103}$ BINDING..... 40

3.1 INTRODUCTION ............................................................................................. 40

3.2 MATERIALS AND METHODS ........................................................................ 41

3.2.1 DNase I Footprinting .................................................................................. 41

3.2.2 DNA-Binding Assays ................................................................................. 42

3.3 RESULTS ........................................................................................................... 45

3.3.1 DNase I Footprinting .................................................................................. 45

3.3.2 TIR and End-Deletion Series of Substrates.............................................. 51

3.3.2.1 TIR series of substrates ...................................................................... 51

3.3.2.2 End-Deletion Series of Substrates ..................................................... 53

3.3.2.3 Evidence leads to new theory on operator sequences ...................... 56

3.3.3 Binding to partial sites ............................................................................ 58

3.3.4 Clear plaque mutant analysis .................................................................... 66

3.4 DISCUSSION ..................................................................................................... 70

4.0 HOW GP33$^{103}$ FUNCTIONS TO PREVENT TRANSCRIPTION OF LYtic GENES 72

4.1 INTRODUCTION ............................................................................................. 72

4.2 MATERIALS AND METHODS ........................................................................ 73

4.2.1 DNA binding assays ................................................................................. 73

4.2.2 Preparation of pBend2 plasmid for use in bending assays .................... 74
4.2.2.1 Creation of plasmids ................................................................. 74
4.2.2.2 Preparation of substrates from pVMV21 and pVMV22 .......... 75
4.2.3 RNA Polymerase Purification and Binding Assay Methods........... 75
  4.2.3.1 Purification of RNA polymerase ............................................. 75
  4.2.3.2 Purification of SigA and reconstitution of RNA polymerase holoenzyme ................................................................. 77
  4.2.3.3 Western Blot analyses ............................................................ 78
  4.2.3.4 DNA Binding Assay using agarose gel electrophoresis.......... 79
4.3 DNA BENDING ............................................................................. 79
  4.3.1 Gp33\textsuperscript{103} does not bend DNA at the $O_{R-R}$ site alone............. 80
  4.3.2 Gp33\textsuperscript{103} bends DNA within the 33-34 intergenic region .......... 81
4.4 DNA LOOPING ............................................................................. 83
4.5 SPACER EFFECT ON OPERATOR FUNCTION ................................ 85
  4.5.1 Changes in $P_{R}$ spacer length affect gp33\textsuperscript{103} action \textit{in vivo}, not \textit{in vitro}... 86
  4.5.2 Gp33\textsuperscript{103} binds similarly when half-sites are spaced very close or very far apart 89
4.6 INTERACTION WITH RNA POLYMERASE ................................. 91
  4.6.1 Purification of RNA polymerase ............................................... 91
    4.6.1.1 Purification of \textit{M. smegmatis} RNA polymerase core enzyme......... 92
    4.6.1.2 Purification of \textit{M. smegmatis} SigA ........................................... 94
  4.6.2 Binding Assays with \textit{M. smegmatis} RNA polymerase and gp3\textsuperscript{103} ....... 95
4.7 DISCUSSION ................................................................................. 99
5.0 CONCLUSIONS AND FUTURE PROSPECTIVES .............................. 100
LIST OF TABLES

Table 1 - Substrates used in Chapter 2 .......................................................................................... 22
Table 2 - Binding affinities of selected substrates in Chapter 2 ................................................... 23
Table 3 - Substrates used in Chapter 3 .......................................................................................... 42
Table 4 - Binding affinities of gp33103 to selected substrates in Chapter 3 ......................... 44
Table 5 - Substrates used in Chapter 4 .......................................................................................... 73
LIST OF FIGURES

Figure 1 - BPs genome map with immunity cassette highlighted ................................................ 10
Figure 2 - The BPs 33-34 intergenic region ................................................................................. 12
Figure 3 - Binding Profile of gp33^{103} to the 33-34 intergenic region ....................................... 25
Figure 4 - Comparison of the binding profiles of gp33^{103} and gp33^{136} to the 33-34 intergenic region ............................................................................................................................................ 26
Figure 5 - Binding Profile of gp33^{103} to other areas of the BPs genome ................................. 28
Figure 6 - Competition assay with unradiolabeled O_R DNA ..................................................... 30
Figure 7 - gp33^{103} binding to two repressor-insensitive mutants ............................................... 32
Figure 8 - Size-exclusion chromatogram ...................................................................................... 33
Figure 9 - gp33^{103} exists as a tetramer in solution ..................................................................... 34
Figure 10 - gp33^{103} binds faintly to single-stranded DNA ....................................................... 35
Figure 11 - gp33^{103} does not bind to a 6bp O_R half site ......................................................... 36
Figure 12 - gp33^{103} binding to O^R in a native and non-native sequence context .................... 38
Figure 13 - DNase I footprinting assay of gp33^{103} binding to the 33-34 intergenic region ....... 46
Figure 14 - Regions of protection from DNase I footprinting ....................................................... 48
Figure 15 - gp33^{103} binding to Regions of DNase I protection ................................................. 49
Figure 16 - gp33^{103} binding to grouped regions of DNase I protection .................................... 50
Figure 17 - TIR series of substrates in the context of the full 33-34 intergenic region

Figure 18 - gp33\textsuperscript{103} binding to the TIR series substrates

Figure 19 - gp33\textsuperscript{103} binding to substrates within TIR-5

Figure 20 - End-deletion series of substrates

Figure 21 - gp33\textsuperscript{103} binding to end-deletion series of substrates

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

Figure 23 - Map of BPs genome with operators marked

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

Figure 25 - Substrates testing O\textsubscript{Rep} and O\textsubscript{R}

Figure 26 - gp33\textsuperscript{103} binding to the O\textsubscript{6} operator

Figure 27 - gp33\textsuperscript{103} binding to O\textsubscript{27}

Figure 28 - gp33\textsuperscript{103} binding to O\textsubscript{55}

Figure 29 - gp33\textsuperscript{103} binding to O\textsubscript{61}

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

Figure 31 - gp33\textsuperscript{103} binding to substrate variants of the 33-34 intergenic region derived from repressor-insensitive mutants

Figure 32 - Locations of O\textsubscript{Rep} and O\textsubscript{R} relative to P\textsubscript{Rep} and P\textsubscript{R}

Figure 33 - gp33\textsuperscript{103} does not bend DNA at O\textsubscript{R-R}

Figure 34 - gp33\textsuperscript{103} bends DNA within the 33-34 intergenic region

Figure 35 - gp33\textsuperscript{103} does not create inter-DNA bridges

Figure 36 - gp33\textsuperscript{103} binds to two adjacent 12bp palindromes

Figure 37 - gp33\textsuperscript{103} binding to various promoter mutants
Figure 38 - gp33\textsuperscript{103} binding to two half sites with varying spacer regions ........................................ 90

Figure 39 - Western blots confirming 8RB13 and 4RA2 antibodies .................................................... 92

Figure 40 - Purification of \textit{M. smegmatis} RNA polymerase ................................................................ 93

Figure 41 - Purification of \textit{M. smegmatis} SigA protein ...................................................................... 94

Figure 42 - RNA polymerase and gp33\textsuperscript{103} binding to the 33-34 intergenic region ............... 96

Figure 43 - RNA polymerase and gp33\textsuperscript{103} binding to other intergenic regions across the BPs genome .......................................................................................................................................... 97

Figure 44 - RNA polymerase and gp33\textsuperscript{103} binding within the P\textsubscript{R} and P\textsubscript{hsp60} promoters ............. 98

Figure 45 - Models of gp33\textsuperscript{103} binding ....................................................................................... 102

Figure 46 - Full site alignment for BPs operators ......................................................................................... 104

Figure 47 - BPs operator half site alignment ................................................................................................. 104

Figure 48 - Brujita and Charlie repressors bind to their respective regulatory intergenic regions.................................................................................................................................................. 112

Figure 49 - gp33\textsuperscript{103} binds to the Brujita 34-35 intergenic region .............................................. 114

Figure 50 – gp34 purification and binding to BPs 33-34 intergenic region .............................................. 119

Figure 51 - gp33\textsuperscript{97} does not bind DNA .............................................................................................. 124

Figure 52 - Repressor insensitive BPs mutants with mutations in the N-terminus of gp33 ...... 125

Figure 53 - Confirmation of the BPs lysogen ................................................................................................. 130

Figure 54 - using an anti-gp33\textsuperscript{103} antibody to determine the concentration of repressor in a lysogen ........................................................................................................................................ 131
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 \(PR\), the genes \(cII\) and \(Q\) are transcribed when termination at the \(trI\) site in \(cro\) is prevented. \(Q\) is also an anti-termination factor and the \(PR\) promoter region contains a \(qut\) (\(Q\) utilization) site \(57\). When \(Q\) is present, \(PR\) 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 \(PRM\) 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 \(PRE\) promoter, which will transcribe the repressor gene, \(cI\). Under conditions that favor lytic growth, the \(cII\) gene is transcribed from the \(N\)-anti-terminated \(PR\) 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 \(PRE\) promoter is not activated and the \(PRM\) 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 \(P_1\) 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 \(PRE\) promoter, which transcribes the \(cI\) gene without the interference of \(Cro\), which is blocking transcription from \(PRM\) \(45\). As \(CI\) protein begins to
accumulate, it competes with Cro for operator sites at \( P_{\text{RM}} \) (64). When CI displaces Cro, it activates its own expression from the leftwards \( P_{\text{RM}} \) promoter and represses transcription from the \( P_R \) promoter that transcribes \( \text{cro} \) 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 \( O_R \) 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 \( cI \) 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 \( P_{\text{RM}} \), 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 cl gene, O₁, O₂, and O₃ (69, 78-80). The repressor binds to these three sites with similar affinities as it does to O₁₁, O₂₂, and O₃₃, respectively. Repressor tetramers bound to O₁₁ and O₂₂ bind cooperatively through its C-terminus with tetramers bound at O₁ and O₂ to form an octomer that loops the DNA containing the cl gene, increasing its ability to autoregulate its expression by isolating it from the rest of the genome (78). When dimers bind O₃₃ and O₃, they tetramerize and block cl 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 O₃₃, which creates a physical block against cl transcription (51, 52, 82). Unlike the lambda repressors, Cro cannot bind to O₂₂ cooperatively and it cannot activate RNA polymerase. Therefore, when Cro levels rise and dimers bind O₂₂, cro transcription stops until the Cro protein levels fall enough for O₂₂ 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 \(P_{\text{left}}\) 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 \( P_{\text{rep}} \) is twice as active as \( P_R \), gp32 is initially expressed at a higher rate than the early lytic genes, therefore \( P_{\text{rep}} \) is responsible for lysogenic establishment.

![Figure 1 - BPs genome map with immunity cassette highlighted](image)

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 \( \text{attP} \) 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 (\( \text{attP} \) 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\textsuperscript{103}). This shortened repressor gene is still attached to the constitutive Prep promoter and is expressed throughout lysogeny, therefore P\textsubscript{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\textsuperscript{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\textsubscript{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 (PR) (97, 102) and one for leftward expression of genes 33 and 32, encoding repressor and integrase proteins respectively (Prep)(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 Prep has twice as much activity as PR, and that in a lysogen the PR promoter is tightly downregulated, whereas the Prep promoter is not. In the presence of the active form of repressor, Prep is actually modestly upregulated (97, 104).

![Figure 2 - The BPs 33-34 intergenic region](image)

The sequence of the BPs 33-34 intergenic regulatory region is shown. The -10 and -35 sequences of the PR and Prep promoters are shown in green text with confirmed transcription start sites indicated with a “+1”. The original locations of OR and ORep 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 GP33103 AND GP33136

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


2.1 INTRODUCTION

It has been shown that wild type BPs cannot infect a *M. smegmatis* BPs lysogen because a BPs lysogen confers homoinmunity to superinfection. *M. smegmatis* strains have been engineered to stably express both gp33103 and gp33136 at similar levels and wild type BPs can lytically infect a strain expressing gp33136, but not a strain expressing gp33103, indicating that gp33103 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\textsuperscript{103} and gp33\textsuperscript{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\textsuperscript{103} and gp33\textsuperscript{136}

The gp33\textsuperscript{103} and gp33\textsuperscript{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\textsuperscript{103} and gp33\textsuperscript{136} respectively. pVMV20 and pVMV27 were transformed into BL21(DE3)star chemically competent cells (Invitrogen) and grown until cultures reached an OD\textsubscript{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 un radiolabeled (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 un radiolabeled 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\textsuperscript{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\textsuperscript{103} based on its elution volume as previously described (169, 172).

2.2.5 Substrates used in this Chapter 2

<table>
<thead>
<tr>
<th>DNA Substrate Name</th>
<th>Sequence/Primers used</th>
</tr>
</thead>
</table>
| 33-34 intergenic region (226bp) – Fig. 7 | Fwd: GTTCTTCTCTGTCTCAAACCAGCTGTTCAAGG  
Rev: GTGTA GTGATCGCTTCCGGAAATGTCATGC |
| 33-34 intergenic region (366bp) – Fig. 7 | Fwd: CACGGGTAGCCGGCTGTTGCTCAGG  
Rev: TGGGTTCTGGCTGGTGGGGGTC |
| 5-6 intergenic region – Fig. 7 | Fwd: AACGTTAGCACACGCGCTTGC  
Rev: ATTGGGCAATGCTGCTGCTGGGCA |
| 26-27 intergenic region – Fig. 7 | Fwd: TGGGTTGGGCTGATTTGAGGATTA  
Rev: GCATCGGAAAGACGATCGGTCAT |
| 54-55 intergenic region – Fig. 7 | Fwd: CACTCGGCTGGCCAGGCTTCGCGCC  
Rev: TGCTGCGGCTGCTGCTGCTGGTGGGTC |
| 60-61 intergenic region – Fig. 7 | Fwd: GGCGAGTTCGACGCGCGTGCG  
Rev: TGGGTTGGGCTGATTTGAGGATTA |
| Mt-3 – Fig. 6 | Fwd: TTTGCGGCTGGGCGGCTTTCTGCAATAGA  
Rev: GTGTA GTGATCGCTTCCGGAAATGTCATGC |
Table 2 - Binding affinities of selected substrates in Chapter 2

<table>
<thead>
<tr>
<th>Substrate Name</th>
<th>Average Calculated Kd (Molar)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type 33-34 intergenic region</td>
<td>2.61 x 10^{-6}</td>
</tr>
<tr>
<td>5-6 intergenic region</td>
<td>2.75 x 10^{-6}</td>
</tr>
<tr>
<td>26-27 intergenic region</td>
<td>9.00 x 10^{-7}</td>
</tr>
<tr>
<td>54-55 intergenic region</td>
<td>2.60 x 10^{-6}</td>
</tr>
<tr>
<td>60-61 intergenic region</td>
<td>3.40 x 10^{-6}</td>
</tr>
</tbody>
</table>
2.3 RESULTS

2.3.1 Mycobacteriophage BPs repressor binds DNA

The active mycobacteriophage BPs repressor, gp33\textsuperscript{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\textsuperscript{103} and gp33\textsuperscript{136} binding to the gene 33-34 intergenic region

Electrophoretic mobility shift assays (EMSA) show that gp33\textsuperscript{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\textsuperscript{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\textsuperscript{103} to the 33-34 intergenic region

A) EMSA analysis shows that gp33\textsuperscript{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\textsuperscript{103} and a non-specific 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\textsuperscript{136} is longer than gp33\textsuperscript{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\textsuperscript{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\textsubscript{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\textsuperscript{103} and gp33\textsuperscript{136} for any differences in binding to the 33-34 intergenic region that contains both P\textsubscript{rep} and P\textsubscript{R} (Figure 4). Both proteins give similar profiles and the additional 33 C-terminal residues in gp33\textsuperscript{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](image)

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).
The BPs genome contains three occurrences of the 12bp palindromic 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 gp33103 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 gp33103 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\textsuperscript{103} binding

To prove that the repressor binds to the 12bp O\textsubscript{R}, we performed a competition assay by adding specific un-radiolabeled probes containing the 12bp O\textsubscript{R} sequence to a reaction where gp33\textsuperscript{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\textsubscript{R} operator site plus 15 base pairs on each side of surrounding native sequence with either wild type 12bp O\textsubscript{R} sequence (denoted as O\textsubscript{R}) or with the 12bp O\textsubscript{R} containing four point mutations (denoted as O\textsubscript{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 16\textmu M gp33\textsuperscript{103} and \sim 15ng of radiolabeled Mt3 DNA (containing O\textsubscript{R} and some surrounding sequence - conditions shown to cause a single shift on a gel - Table 1, Figure 6). Increasing concentrations of unradiolabeled O\textsubscript{R} and O\textsubscript{R4} were added to the reactions to compete for gp33\textsuperscript{103} binding.
A DNA substrate containing the operator site (Mt3 – Table 3 and Table 4) was radiolabeled and incubated with gp33\textsuperscript{103} at a concentration of 16\textmu M. Increasing concentrations of unradiolabeled probe containing the operator site (wild type, OR, and with four point mutations, OR\textsubscript{4}) 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) 10\textmu M, 2) 30\textmu M, 3) 50\textmu M, 4) 100\textmu M, 5) 300\textmu M, 6) 500\textmu M.

Our results show that the presence of cold OR outcompetes the radiolabeled probe for gp33\textsuperscript{103} binding, whereas OR\textsubscript{4} does not (Figure 6). These results suggest that the 12bp OR 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 OR) 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\textsuperscript{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\textsuperscript{103} on them, they showed a 10-fold reduction in binding affinity compared to wild type (Figure 7B).
Figure 7 - gp33\textsuperscript{103} binding to two repressor-insensitive mutants

BPs gp33\textsuperscript{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\textsuperscript{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) 0.16\mu M, 3) 0.54\mu M, 4) 1.6\mu M, 5) 5.4\mu M, 6) 16\mu M, 7) 54\mu M.

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 OR 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\textsuperscript{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<sup>103</sup> 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<sup>103</sup> is predicted to be 40.4kDa (orange diamond). Gp33<sup>103</sup> is an 11.2 kDa protein, so it is likely a tetramer in solution.

2.3.4 Gp33<sup>103</sup> binds to other DNA substrates

2.3.4.1 Gp33<sup>103</sup> binds to ssDNA

To determine that the operator site for gp33<sup>103</sup> is a double-stranded DNA binding site, we tested whether gp33<sup>103</sup> 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<sup>103</sup>’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.
EMSAS were done on a 42bp double stranded DNA substrate containing OR 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\textsuperscript{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\textsuperscript{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\textsuperscript{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 \textit{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](image)

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 OR is important for major complex formation

In its natural sequence, gp33$^{103}$ binds OR 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 OR site and the minor complex could be either a dimer or monomer binding to the OR site. When OR 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).
EMSAs were done with gp33\(^{103}\) binding to two 60 bp substrates containing a single 12 bp OR site located centrally. One substrate contained 24 bp of native sequence surrounding OR, the other contained non-native sequence (Table 1). Gp33\(^{103}\) has a slower migrating major complex when OR 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 OR is centered on a 60bp substrate that is not its natural sequence, gp33\(^{103}\) binds to OR and causes a shift (faster-moving complex), but the lack of an AT-rich 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 \textit{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\textsuperscript{103} BINDING

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


3.1 INTRODUCTION

We have established that gp33\textsuperscript{103} is the active form of the BPs repressor and that it is capable of binding DNA at a putative 12bp palindromic operator, O\textsubscript{R}, as well as to three identical sites across the BPs genome. When gp33\textsuperscript{103} binds to the other identical 12bp sites, a much simpler binding pattern is seen than when gp33\textsuperscript{103} binds within the 33-34 intergenic region. We sought to explain the nature of the complexes seen when gp33\textsuperscript{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\textsuperscript{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\textsuperscript{103} binding and relate that to its function \textit{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\textsubscript{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\textsubscript{2} and 10 mM MgCl\textsubscript{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>
<thead>
<tr>
<th>DNA Substrate Name</th>
<th>Sequence/Primers used</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNase I Footprinting substrate – Fig. 13</td>
<td>Fwd: TCTAGAACCAGCGATTGATATGTAC&lt;br&gt;Rev: AAGCTTGTGCACCAACATAGCGAC</td>
</tr>
<tr>
<td>Region 1 – Fig. 15</td>
<td>AACCGTGCATATCGCGCAATACCTTGGCAGATACCGGCAACGGGTTGAACAACATACGTTG</td>
</tr>
<tr>
<td>Region 2 – Fig. 15</td>
<td>AACCGTGCATATCGCGCAAATACACATTTGGTGAACAACATACGTTG</td>
</tr>
<tr>
<td>Region 3 – Fig. 15</td>
<td>AACCGTGCATATCGCGCAATACACATTTGGTGAACAACATACGTTG</td>
</tr>
<tr>
<td>Region 4 – Fig. 15</td>
<td>AACCGTGCATATCGCGCAATACACATTTGGTGAACAACATACGTTG</td>
</tr>
<tr>
<td>Region 5 – Fig. 15</td>
<td>AACCGTGCATATCGCGCAATACACATTTGGTGAACAACATACGTTG</td>
</tr>
<tr>
<td>TIR-1 – Fig. 18</td>
<td>ACCGGGATTGTATTGAACGCTGCTGCTACGCTACGGAATAT</td>
</tr>
<tr>
<td>TIR-2 – Fig. 18</td>
<td>CGCGCAATATCGCGCAATACCGGCGATTGTATGTACGTCTGTCGCGTACGCGCAATAT</td>
</tr>
<tr>
<td>TIR-3 – Fig. 18</td>
<td>GAGCCGAAACGGCCGACGGCGCTTGGATCGGTCAACCGTGGCAGCAGAAC</td>
</tr>
<tr>
<td>TIR-4 – Fig. 18</td>
<td>GCAAACCGTGGTGGTAAACGGGATTGTATTGAACGCTGCTGCTACGCTACGGAATAT</td>
</tr>
<tr>
<td>TIR-5 – Fig. 18</td>
<td>CAAAGCCTACCGGCGATTGTATGTACGTCTGTCGCGTACGCGCAATAT</td>
</tr>
<tr>
<td>TIR-6 – Fig. 18</td>
<td>TCCAATAGACGACATATGTCGCT</td>
</tr>
<tr>
<td>TIR-5.1 – Fig. 19</td>
<td>CAAAGCCTACCGGCGATTGTATGTACGTCTGTCGCGTACGCGCAATAT</td>
</tr>
<tr>
<td>TIR-5.2 – Fig. 19</td>
<td>TGGCGCAATATCGCGCAATACCGGCGATTGTATGTACGTCTGTCGCGTACGCGCAATAT</td>
</tr>
<tr>
<td>TIR-5.3 – Fig. 19</td>
<td>GAAATCGCGGCAATACCGGCGATTGTATGTACGTCTGTCGCGTACGCGCAATAT</td>
</tr>
<tr>
<td>TIR-5.4 – Fig. 19</td>
<td>AGTGCGGCGATTGTATGTACGTCTGTCGCGTACGCGCAATAT</td>
</tr>
<tr>
<td>TIR-5.5 – Fig. 19</td>
<td>TCCAATAGACGACATATGTCGCT</td>
</tr>
<tr>
<td>TIR-5.6 – Fig. 19</td>
<td>CAAAGCCTACCGGCGATTGTATGTACGTCTGTCGCGTACGCGCAATAT</td>
</tr>
<tr>
<td>TIR-5.7 – Fig. 19</td>
<td>GGGCGTTCGCGCCCTATCTTCCAA</td>
</tr>
<tr>
<td>Mt-2 – Fig. 21</td>
<td>Fwd: TGGGTGACATGACCCCAACCGAAAC&lt;br&gt;Rev: GTGTAGTGATCGCTTCGGAATGTCATGC</td>
</tr>
<tr>
<td>Mt-3 – Fig. 21</td>
<td>Fwd: TGGGTGACATGACCCCAACCGAAAC&lt;br&gt;Rev: GTGTAGTGATCGCTTCGGAATGTCATGC</td>
</tr>
<tr>
<td>Mt-4 – Fig. 21</td>
<td>Fwd: TGGGTGACATGACCCCAACCGAAAC&lt;br&gt;Rev: GTGTAGTGATCGCTTCGGAATGTCATGC</td>
</tr>
<tr>
<td>Mt-5 – Fig. 21</td>
<td>Fwd: TGGGTGACATGACCCCAACCGAAAC&lt;br&gt;Rev: GTGTAGTGATCGCTTCGGAATGTCATGC</td>
</tr>
<tr>
<td>Mt-6 – Fig. 21</td>
<td>Fwd: TGGGTGACATGACCCCAACCGAAAC&lt;br&gt;Rev: GTGTAGTGATCGCTTCGGAATGTCATGC</td>
</tr>
<tr>
<td>Mt-7 – Fig. 21</td>
<td>Fwd: TGGGTGACATGACCCCAACCGAAAC&lt;br&gt;Rev: GTGTAGTGATCGCTTCGGAATGTCATGC</td>
</tr>
<tr>
<td>Mt-8 – Fig. 21</td>
<td>Fwd: TGGGTGACATGACCCCAACCGAAAC&lt;br&gt;Rev: GTGTAGTGATCGCTTCGGAATGTCATGC</td>
</tr>
<tr>
<td>Mt-9 – Fig. 21</td>
<td>Fwd: TGGGTGACATGACCCCAACCGAAAC&lt;br&gt;Rev: GTGTAGTGATCGCTTCGGAATGTCATGC</td>
</tr>
</tbody>
</table>
| Mt-10 – Fig. 21 | Fwd: GTTCTTTCTGCTTCAAAAACCAGCCTCAAGG  
Rev: ATCCGTTACCACGCGGTTGCTCA |
| Mt-11 – Fig. 21 | Fwd: GTTCTTTCTGCTTCAAAAACCAGCCTCAAGG  
Rev: tttcGCTGCAACCGTTGCTGAGA |
| Mt-12 – Fig. 21 | Fwd: GTTCTTTCTGCTTCAAAAACCAGCCTCAAGG  
Rev: GACAGACATCATATCGCCGCTATGT |
| Mt-13 – Fig. 21 | Fwd: GTTCTTTCTGCTTCAAAAACCAGCCTCAAGG  
Rev: GATCCCGCCGAGCAGGCTTTCGT |
| O6—L – Fig. 26 | Fwd: CACACAGACGCTTGCACCGAGCTATATGCAGTGCTTTCATTGACGAGTCAGGT |
| O6—R – Fig. 26 | GCATTTTTCAATAAGAGACCATATGGGCTATGTTGGTCACA |
| O27—R – Fig. 27 | GCATCAGGACGCTGCTATCGGACATATGTCGATGGATGATGAGGATGCA |
| O27—L – Fig. 27 | GCATTTTTCAATAAGAGACATATGTCGATGGATGTCGACCA |
| O5—L – Fig. 25 | GCATTTTTCAATAAGAGACATATGTCGATGGATGTCGACCA |
| O5—R – Fig. 25 | GCATTTTTCAATAAGAGACATATGTCGATGGATGTCGACCA |
| O33—L – Fig. 25 | Fwd: AACGTAGCCACACAGACGCTTGC  
Rev: ATTCGCCATGCTGCTTGTCGACG |
| O33—R – Fig. 25 | Fwd: TGCCGTGGCGCTGATTCTGAGGTAA  
Rev: GCATCGGGAAGAAACGATCGGCCAT |
| O35—L – Fig. 26 | GCATTTTTCAATAAGAGACATATGTCGATGGATGTCGACCA |
| O35—R – Fig. 26 | GCATTTTTCAATAAGAGACATATGTCGATGGATGTCGACCA |
| O51—L – Fig. 26 | GCATTTTTCAATAAGAGACATATGTCGATGGATGTCGACCA |
| O51—R – Fig. 26 | GCATTTTTCAATAAGAGACATATGTCGATGGATGTCGACCA |
| 5—6 intergenic region – Fig. 26 | Fwd: AACGTAGCCACACAGACGCTTGC  
Rev: ATTCGCCATGCTGCTTGTCGACG |
| 26—27 intergenic region – Fig. 27 | Fwd: TGCCGTGGCGCTGATTCTGAGGTAA  
Rev: GCATCGGGAAGAAACGATCGGCCAT |
| 33—34 intergenic region (366bp) – Fig. 25 | Fwd: GTTCTTTCTGCTTCAAAAACCAGCCTCAAGG  
Rev: GTGTAGTGATCGCTTCGGAATGTCATGC |
| 54—55 intergenic region – Fig. 28 | Fwd: CACGGGGTACGGCGCTTGGTCG  
Rev: TGGGTTCTGTGGCGGTTGCGTC |
| 60—61 intergenic region – Fig. 29 | Fwd: GCCAGTTCCAGCGCCGCGGTG  
Rev: TGGCCTTGTTGGTGCTGCGTGGAG |
| Repressor-insensitive BPs mutants - 33—34 intergenic region – Fig. 31 | Fwd: GTTCTTTCTGCTTCAAAAACCAGCCTCAAGG  
Rev: GTGTAGTGATCGCTTCGGAATGTCATGC |
Table 4 - Binding affinities of gp33\textsuperscript{103} to selected substrates in Chapter 3

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

3.3.1 DNase I Footprinting

Both the viral and prophage forms of gp33\textsuperscript{103} have been purified. Previous studies have shown that the prophage form of gp33\textsuperscript{103} binds within the gene 33-34 intergenic region, which contains two divergent promoters: one for rightwards expression of an early lytic operon, P\textsubscript{R}, and one for leftward expression of gp33\textsuperscript{103} and Int (102). Studies done by Broussard et al. showed that gp33\textsuperscript{103} binds specifically to a 12 base pair palindrome, O\textsubscript{R}. DNA binding assays from that study showed that the binding affinity of the BPs gp33\textsuperscript{103} to its operator sites is low, with a dissociation constant of approximately 1 $\mu$M, which is 10-fold higher than that expected for a phage repressor (97). To confirm that gp33\textsuperscript{103} binds to O\textsubscript{R}, we performed DNase I footprinting assays (Figure 13).
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\textsuperscript{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 Prep, one of which contains O\textsubscript{Rep} (Regions 1 and 2). Three regions of protection are observed near P\textsubscript{R} (Regions 3 and 4, Region 5), one of which includes O\textsubscript{R} (Region 5). The sequence of O\textsubscript{R} is shown (O\textsubscript{R-R}) as well as a predicted second binding site (O\textsubscript{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.

![Diagram of regions of protection from DNase I footprinting](image)

**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.
DNA binding assays with gp33\textsuperscript{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\textsuperscript{103} only binds to Region 5, which contains OR. 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 OR 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).
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\textsuperscript{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\textsuperscript{103} binds specifically to the 12bp OR site as well as the region immediately upstream of OR. These assays also show that gp33\textsuperscript{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\textsubscript{rep} will be referred to as O\textsubscript{Rep} from here on out. It is likely that the slowest moving of the protein-DNA complexes seen when gp33\textsuperscript{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 gp33103 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 gp33103 did not bind to Region 1 or Region 2 when isolated from each other.

![Diagram of TIR series substrates](image)

**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.
EMSAs showing gp33<sup>103</sup> 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<sub>R</sub> (O<sub>R-R</sub>) and to substrate TIR-5, which contains Region 3 and Region 4 (O<sub>R-L</sub>) and only one base pair of O<sub>R-R</sub>. 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<sub>R</sub>) is the TIR-5 substrate that is adjacent to, but does not include, the 12bp O<sub>R</sub> (Figure 18). This substrate contains one base pair from the O<sub>R</sub> site and 39 base pairs of sequence to the left of O<sub>R</sub>, 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<sup>103</sup> 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<sup>103</sup> binds to TIR-5.2, which contains the 19bp directly adjacent to the 12bp O<sub>R</sub> (Figure 19).
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,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}.
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_\text{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_\text{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).
BPs gp33\textsuperscript{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 \text{OR} and \text{OR}_{\text{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 ORep
6bp half-site is included in the sequence (Mt9), gp33103 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 gp33103 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 gp33103 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 gp33103 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 gp33103 binds to both the 33- and 34- proximal regions of the
33-34 intergenic region, but the evidence also strongly suggests that the operator ORep 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 gp33103 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-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-R} and the region to the left that includes Region 4 as O_{R-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-L} but do not contain O_{R-R} (Figure 22). A plausible explanation is that gp33^{103} binds as a tetramer to O_{R-L} and O_{R-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\textsuperscript{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\textsuperscript{103} at these sites. The three other instances of sequences identical to O\textsubscript{R-R} are located in small intergenic regions, between genes 5 and 6 (O\textsubscript{6}), between genes 26 and 27 (O\textsubscript{27}), and between genes 54 and 55 (O\textsubscript{55}) (Figure 23). We also found a site with two base pair differences that is located within the 60-61 intergenic region (O\textsubscript{61}) (Figure 23).

![Figure 23 - Map of BPs genome with operators marked](image)

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 \textit{mcherry} reporter gene and transformed into \textit{M. smegmatis} mc\textsuperscript{2}155. The 5-6, 26-27, 54-55, and 60-61 regions all have promoter activity (P\textsubscript{6}, P\textsubscript{27}, P\textsubscript{55}, and P\textsubscript{61}), although the strengths vary, with P\textsubscript{55} being the most active, and P\textsubscript{27} the weakest (103). Putative promoter elements containing -10 and -35 hexameric motifs are predicted for P\textsubscript{6}, P\textsubscript{55}, and P\textsubscript{61}, but not confidently for P\textsubscript{27} (Figure 24). The promoter-reporter plasmids were also transformed into a BPs lysogen and the promoter activities...
determined. The P₆, P₂₇, and P₅₅ promoters are clearly down regulated in a lysogen, presumably by gp33<sup>103</sup>, as is P₇ (97). No regulation for either P<sub>rep</sub> or P₆₁ was observed.

In view of the hypothesis that the 12 bp palindrome at O₉⁻R corresponds to a half-site for gp33<sup>103</sup> binding, we examined the other intergenic region sequences for additional sites related to O₉⁻R. Interestingly, in all four regions we identified different, but related, sequences that are positioned either 5 bp (O₆, O₂₇, O₆₁), or 8 bp (O₅₅) 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 O₆⁻L and O₆⁻R, O₂₇⁻L and O₂₇⁻R, O₅₅⁻L and O₅₅⁻R, and O₆₁⁻L and O₆₁⁻R for O₆, O₂₇, O₅₅ and O₆₁ respectively (Figure 24).

![Figure 24 - Sequences of all BPs operators within their respective promoter regions](image)

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<sup>103</sup>. Sites highlighted in pink do not bind gp33<sup>103</sup> 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\textsuperscript{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\textsuperscript{103} binds it without the full palindrome present on the same substrate. For reference, binding assays done with O\textsubscript{R} and O\textsubscript{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\textsubscript{Rep} and O\textsubscript{R}**

EMSAs show binding profiles of gp33\textsuperscript{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\textsuperscript{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\textsuperscript{103} plays a role in this region during lysogeny. Along with a full 12bp palindromic half site (O\textsubscript{6-1}), we found a partial binding site in this region that differs from
the palindrome by 3 base pairs (O₆-R) (Figure 24). When gp33¹⁰³ binds to a 45bp substrate where O₆-R is mutationally ablated, a single protein-DNA complex is formed (Figure 26B). A binding assay done on this O₆-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 - gp33¹⁰³ binding to the O₆ operator](image)

EMSAs show binding profiles of gp33¹⁰³ to the substrates shown in the diagrams. Gp33¹⁰³ binds to O₆-L, but not O₆-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\textsuperscript{103} binds within the 26-27 intergenic region (Figure 27A) where studies by Broussard \textit{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\textsuperscript{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\textsuperscript{103} indeed plays a role in this region during lysogeny.

EMSAs show binding profiles of gp33\textsuperscript{103} to the substrates shown in the diagrams. Gp33\textsuperscript{103} binds to both O\textsubscript{27-L}, and O\textsubscript{27-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-R}), we found a partial binding site in this region that differs from the palindrome by 2 base pairs (O_{27-L}) (Figure 24). When gp33^{103} binds to a 45bp substrate of native 26-27 intergenic region sequence where O_{27-L} is mutationally ablated, a single protein-DNA complex is formed (Figure 27B). A binding assay done on this O_{27-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.
EMSAs show binding profiles of gp33\textsuperscript{103} to the substrates shown in the diagrams. Gp33\textsuperscript{103} binds to O\textsubscript{55-L}, but not O\textsubscript{55-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\textsuperscript{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\textsubscript{R} intergenic regions (Figure 28A). This promoter is downregulated in a lysogen, so we know that gp33\textsuperscript{103} plays a role in this region during lysogeny. Along with a full 12bp palindromic half site (O\textsubscript{55-L}), we found a partial binding site in this region that differs from the palindrome by 3 base pairs (O\textsubscript{55-R}). When gp33\textsuperscript{103} binds to a 45bp substrate where O\textsubscript{55-R} is mutationally ablated, a single protein-DNA complex is formed (Figure 28B). A binding assay done on this O\textsubscript{55-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 gp33103 within the 54-55 intergenic region than there are in other intergenic regions.

Figure 29 - gp33103 binding to O61

EMSAs show binding profiles of gp33103 to the substrates shown in the diagrams. Gp33103 binds to O61-R, but not O61-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-R}) and one with 5bp differences (O_{61-L}) (Figure 24). When we tested gp33^{103} binding to the two half sites only O_{61-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-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 \textit{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 \textit{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, \textit{gene 34}, or mutations within the 33-34 intergenic region (Figure 30 – Table 3). This evidence led to the hypothesis that the \textit{gene 32, 33, 34} cassette were the only genes involved in regulating lysogeny (97).
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 Prep 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 PR. 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 PR upregulated in the Clr4 and Clr8 mutants, the genetic switch is pushed toward lytic growth. The Clr6 point mutation is, however, located within OR-L and could influence gp33 activity within OR. 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<sup>103</sup> 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<sup>103</sup> 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\textsubscript{R-R} that reduce the affinity of gp33\textsuperscript{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\textsubscript{rep} and also removes the 5’ end of the repressor gene from the 3’ end. With the repressor gene interrupted and P\textsubscript{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\textsuperscript{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\textsubscript{rep}, and O\textsubscript{R-L}. Without the P\textsubscript{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\textsuperscript{103} to this mutant reflects the removal of O\textsubscript{Rep} and O\textsubscript{R-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\textsubscript{R-R}, which presumably have an effect on gp33\textsuperscript{103} binding and affect P\textsubscript{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\textsuperscript{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\textsuperscript{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\textsuperscript{103} functions is the subject of the next chapter.
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\textsuperscript{103} FUNCTIONS TO PREVENT TRANSCRIPTION OF LYTIC GENES

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


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\textsuperscript{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

<table>
<thead>
<tr>
<th>DNA Substrate Name</th>
<th>Sequence/Primers used</th>
</tr>
</thead>
<tbody>
<tr>
<td>33-34 intergenic region (226bp)</td>
<td>Fwd: TCCGGGCCCATACTGAAATGGTTGGCGACAT</td>
</tr>
<tr>
<td></td>
<td>Rev: CGTTFGGGTCATGGCACCACAT</td>
</tr>
<tr>
<td>33-34 intergenic region (366bp)</td>
<td>Fwd: GTTCTTCTTGCTTCAAAACCAGCTTCAAGG</td>
</tr>
<tr>
<td></td>
<td>Rev: GTGTAGTGATCGCTCGGAATGTCATGC</td>
</tr>
<tr>
<td>2 half sites 5bp apart</td>
<td>CACCGACGCACCATATGCTGCACGACATATGCTGCAGCCTGG</td>
</tr>
<tr>
<td>Left half site ablated (5bp apart)</td>
<td>CACCGACGAAGTGCCAGCATATGCTGACATATGCTGCAAGCTTG</td>
</tr>
<tr>
<td>Right half site ablated (5bp)</td>
<td>CACCGACGCACCATATGCTGACATGAAAGTGCACGATCAGCCTTG</td>
</tr>
</tbody>
</table>
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 $O_{R-R}$ 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, Styl, 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 $M. \text{sme}g\text{mat}is$ RNA polymerase core enzyme was purified as previously described (180, 181). Briefly, a 1L $M. \text{sme}g\text{mat}is$ mc$^2$155 culture was grown to saturation in 7H9 broth (difco) supplemented with CB/CHX/ADC/Tween80. The pellet was spun down and frozen. The
A 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 sonicator 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 polymerase was eluted from the pellet by resuspending 20 mL Buffer A + 1 M NaCl and centrifuging for ten minutes at 6,000 rpm. 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* sigA 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* mc² 155 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 agarose 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 gelpoc XR+ gelpoc 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 gp33103 to the 33-34 intergenic region was due to gp33103 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-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 \( \sim 50\% \) of the DNA (5.4\( \mu \)M) 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-R} site alone

We decided to check if gp33^{103} is able to bend DNA if only O_{R-R} is present and not the rest of the 33-34 intergenic region so we cloned the 12bp O_{R-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-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-L}. 
The 12bp of OR 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 OR located at different positions relative to the edge of the probe by cleaving with appropriate enzymes. BPs gp33\textsuperscript{103} (0.54μM) was incubated with each radiolabeled probe and run on a native polyacrylamide gel. The only gp33\textsuperscript{103}-DNA complex (C1) does not show any differences in relative mobility with different substrates.

4.3.2 \textbf{Gp33\textsuperscript{103} bends DNA within the \textit{33-34} intergenic region}

When we tested the full \textit{33-34} intergenic region for binding using pVMV21, what we saw on the gel was not only a change in mobility as the \textit{33-34} intergenic region reached the center of the substrate, but a continuing upward shift that did not lose mobility when the \textit{33-34} intergenic region reached the other end of the substrate (\textbf{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 Oₐ, not Oₑₗ, and due to the location of Oₐ 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 gp33nano³ and Oₑₗ. 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 gp33nano³ being bound and introducing a DNA distortion as a tetramer to Oₑₗ and Oₑₗ₊₁.

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 gp33nano³ (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 gp33nano³ binding.

Figure 34 - gp33nano³ bends DNA within the 33-34 intergenic region
4.4 DNA LOOPING

We wondered if one of the complexes present in the binding profile of gp33\textsuperscript{103} to the 33-34 intergenic region was due to gp33\textsuperscript{103} binding to multiple areas and interacting, thus creating an 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\textsuperscript{103} to each substrate separately, and then of gp33\textsuperscript{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\textsuperscript{103} binds to each substrate separately.
Figure 35 - gp33\textsuperscript{103} does not create inter-DNA bridges

To determine if intermolecular DNA bridges are formed by gp33\textsuperscript{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\textsuperscript{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\textsubscript{R} sequence and a bound gp33\textsuperscript{103} tetramer was unable to interact with another bound gp33\textsuperscript{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\textsuperscript{103} bending DNA at a magnitude necessary for protein bound at O\textsubscript{Rep} and O\textsubscript{R} to interact.

### 4.5 Spacer Effect on Operator Function

O\textsubscript{6}, O\textsubscript{R}, O\textsubscript{27}, and O\textsubscript{61} each contain half-sites that are 5bp apart. O\textsubscript{55} and O\textsubscript{Rep} on the other hand, contain half-sites that are 8bp apart. We hypothesized that spacer length has an effect on gp33\textsuperscript{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\textsuperscript{103} binding to these substrates.

![Figure 36](image)

**Figure 36 – gp33\textsuperscript{103} binds to two adjacent 12bp palindromes**

(A) and (B) EMSAs showing gp33\textsuperscript{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\textsuperscript{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\textsuperscript{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\textsuperscript{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\textsubscript{R} spacer length affect gp33\textsuperscript{103} action \textit{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\textsubscript{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\textsuperscript{103} repression. In her studies of the P\textsubscript{R} promoter, she fused P\textsubscript{R} to a \textit{mCherry} reporter gene and expressed it both in wild type \textit{M. smegmatis} mc\textsuperscript{2}155 and in a BPs lysogen of \textit{M. smegmatis}. Wild type P\textsubscript{R} expressed mCherry in wild type \textit{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\textsubscript{R}, the spacer length is 18 base pairs and she saw expression of mCherry in wild type \textit{M. smegmatis}, but decreased expression in a lysogen (\textbf{Figure 37A/B}). When she added one basepair upstream of O\textsubscript{R-R}, but before the -35 motif, she saw decreased expression from P\textsubscript{R} in wild type \textit{M. smegmatis}, but no change in the lysogen’s ability to repress the promoter (\textbf{Figure 37A/B}). When she removed one or two base pairs between O\textsubscript{R-R} and the -35 element, however, she saw that while there were some changes in WT \textit{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\textsubscript{R} promoter as effectively when O\textsubscript{R-R} is closer to the -35 element and O\textsubscript{R-L}. 
Figure 37 - gp33\textsuperscript{103} binding to various promoter mutants

(A) Shows the locations of the base pairs that Lauren Oldfield added and removed. (B) Shows the fluorescent reporter assays she did in which plasmids with P\textsubscript{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\textsubscript{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\textsuperscript{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\textsuperscript{103} to bind to O\textsubscript{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\textsubscript{R} consisting of two 12bp half-sites, so only 6bp of O\textsubscript{R,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\textsuperscript{103} binds just as well to each of the promoter mutants.

Only one complex is formed when gp33\textsuperscript{103} binds to these substrates, leading us to believe that this complex consists of a gp33\textsuperscript{103} tetramer binding to O\textsubscript{R,R}, but not bending DNA, as O\textsubscript{R,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 operators either 2bp apart, or 23bp apart (Figure 38). When gp33\textsuperscript{103} binds to each of these, it binds with similar binding profiles, which is unexpected. Based on results from Figure 36, Gp33\textsuperscript{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.
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\textsuperscript{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\textsuperscript{103} could interact with RNA polymerase at OR within the PR promoter. We started by purifying the mycobacterial RNA polymerase and then performed binding assays using gp33\textsuperscript{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* mc²155 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.

![Western blots confirming 8RB13 and 4RA2 antibodies](image)

*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](image)

(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 gp3103

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\textsuperscript{103} binding to the 33-34 intergenic region

EMSA showing the binding profiles of RNA polymerase, gp33\textsuperscript{103}, and both together. RNA polymerase was unable to bind DNA in this assay.

Gp33\textsuperscript{103} does bind DNA, but with the addition of RNAP there is no difference in gp33\textsuperscript{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\textsuperscript{103} binding to other intergenic regions across the BPs genome

EMSAs showing the binding profiles of RNA polymerase, gp33\textsuperscript{103}, and both together to various intergenic regions across the BPs genome. RNA polymerase was unable to bind DNA in this assay. Gp33\textsuperscript{103} protein concentrations are 5.4\,\mu M or 16\,\mu M. Asterisks mark lanes where a higher concentration of gp33\textsuperscript{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).
RNA polymerase and gp33\textsuperscript{103} were incubated with plasmid DNA containing the P\textsubscript{R} and P\textsubscript{hsp60} promoters as indicated. RNA polymerase faintly binds DNA in this assay, and gp33\textsuperscript{103} seems to increase gp33\textsuperscript{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\textsubscript{R} or P\textsubscript{hsp60} upstream of an \textit{mCherry} reporter gene, because the \textit{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\textsubscript{R} substrate when gp33\textsuperscript{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 protomer 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 gp33103 binding

(A) Our first model for $\text{gp33}^{103}$ 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<sup>103</sup> would be synthesized with <sup>35</sup>S-methionine incorporated and the counts per protomer could be calculated. Using known quantities and counts of <sup>32</sup>P-labeled probe would allow the ratio of <sup>35</sup>S to <sup>32</sup>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<sub>6</sub>, O<sub>27</sub>, O<sub>Rep</sub>, O<sub>R</sub>, O<sub>55</sub>, and O<sub>61</sub>) shows the juxtaposition of the half sites within each operator (**Figure 46**). In four of the sites (O<sub>6</sub>, O<sub>27</sub>, O<sub>R</sub>, and O<sub>55</sub>), one half site contains the 12 bp palindrome (O<sub>6-L</sub>, O<sub>27-R</sub>, O<sub>R-R</sub>, and O<sub>55-L</sub>) and is easily recognizable, and the other half sites (O<sub>6-R</sub>, O<sub>27-L</sub>, O<sub>R-L</sub>, and O<sub>55-R</sub>) contain sequence departures at two or more positions (**Figure 46**). At O<sub>61</sub>, both half sites have sequence departures from the consensus (two in O<sub>61-R</sub> and five in O<sub>61-L</sub>) although gp33<sup>103</sup> binds with
similar affinity as it does to O₆ and O₅₅ (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$_{R-L}$ half site is only distantly related, although gp33$^{103}$ clearly binds to it, even in the absence of O$_{R-R}$ (Figures 18, 19, 21). O$_{R-L}$ is spaced five bases away from O$_{R-R}$, 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$_{R-L}$ 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$_{R-L}$ in binding and regulation.

The sequences of the O$_{Rep}$ half sites are the most divergent, and the best alignment suggests that O$_{Rep-L}$ and O$_{Rep-R}$ 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\textsuperscript{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\textsuperscript{103} binding profile \textit{in vitro}, a logical question is to ask whether the spacing between half sites is important \textit{in vivo}. We know that narrowing the space between O\textsubscript{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\textsubscript{R} (Figure 37). In order to test whether this spacing is important for all of the phages \textit{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\textsubscript{27} or O\textsubscript{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\textsubscript{55} and O\textsubscript{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\textsubscript{R} and O\textsubscript{R}.

It is unknown whether gp33\textsuperscript{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\textsuperscript{103} (such as P\textsubscript{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 \textit{M. smegmatis} as a control, and into a BPs lysogen to see if the reporter

106
gene is repressed by gp33\textsuperscript{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\textsubscript{61}, where the P\textsubscript{61} promoter is not downregulated in a lysogen in fluorescent reporter assays. Binding to O\textsubscript{61} may prevent the RNA polymerase from continuing to elongate the P\textsubscript{55} or P\textsubscript{58} operon transcripts (103). It could also be used at O\textsubscript{27} to prevent any possible read-through transcription from expressing the lysis gene cassette.

Gp33\textsuperscript{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\textsuperscript{103} and RNA polymerase truly interact, is to perform a pull-down assay. We have constructed a vector with an HA-tagged \textit{M. smegmatis} \textit{rpoA} gene placed downstream of an Hsp60 constitutive promoter (pVMV36). When transformed into wild type \textit{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\textsuperscript{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\textsuperscript{136} forms mixed tetramers with the shorter protein (gp33\textsuperscript{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.
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\textsuperscript{101} & gp34\textsuperscript{129} and gene 35 gp35\textsuperscript{103} & gp35\textsuperscript{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\textsuperscript{101}), pVMV19 (Brujita gp34\textsuperscript{129}), pVMV16 (Charlie gp35\textsuperscript{103}), and pVMV17 (Charlie gp35\textsuperscript{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\textsubscript{600} and once cultures reached an optical density of OD\textsubscript{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\textsuperscript{101} and gp34\textsuperscript{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.
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\textsuperscript{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\textsubscript{R-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\textsuperscript{103} was able to bind to the Brujita 34-35 intergenic region, but not to the Charlie 35-36 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 substrates 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.
Interestingly, 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 $O_{R-R}$ 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\textsuperscript{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 substrates (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\textsubscript{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.
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 quantities 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\textsuperscript{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\textsuperscript{103} binds with a $K_d$ of $\sim$2\textmu M, 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\textsuperscript{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\textsuperscript{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\textsuperscript{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\textsuperscript{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\textsubscript{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\textsuperscript{97} proteins using nickel affinity chromatography. The flow through containing pure gp33\textsuperscript{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\textsuperscript{97} does not bind DNA, but gp33\textsuperscript{103} does

Before we knew that gp33\textsuperscript{103} was expressed from a leaderless transcript, we purified a shorter form, gp33\textsuperscript{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, 5\textmu M. 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\textsuperscript{103} is very low (~2\textmu M) and 5\textmu M may not be enough protein to see a shift with gp33\textsuperscript{97}.
EMSAs done with gp33\textsuperscript{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\textsuperscript{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 6\textsuperscript{th} residue (D6G). Another was an arginine to glycine substitution at the 14\textsuperscript{th} 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 \textit{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 binding affinity. Since lambad cI’s binding is reduced 8,000-fold when it’s first 7 residues are missing (206, 207), perhaps gp33\textsuperscript{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\textsuperscript{97} binds when present at higher concentrations. It would be interesting to see, if gp33\textsuperscript{97} does bind at higher concentrations, if the specificity of binding is affected by removal of the 6 N-terminal residues, if gp33\textsuperscript{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\textsuperscript{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\textsuperscript{103} is in a BPs lysogen of \textit{Mycobacterium 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\textsuperscript{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\textsuperscript{103} in a BPs lysogen.

D.2 MATERIALS AND METHODS

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

Polyclonal antibodies to gp33\textsuperscript{103} were created through the Pocono Rabbit Farm & lab. We sent 5mg of pure gp33\textsuperscript{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* mc²155**

7H10 agar (Difco) media plates supplemented with carbenicilin and cyclohexamide were streaked with the lab strain of *M. smegmatis* mc²155 (212) as well as a BPs lysogen of *M. smegmatis* mc²155 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).
A BPs lysogen strain was streaked onto a lawn of wild type *M. smegmatis* mc²155 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\textsuperscript{103} was sent to Pocono Rabbit Farm & Laboratory to create polyclonal rabbit antibodies that recognize the active BPs repressor, gp33\textsuperscript{103}. Western blots using the antibodies have shown specific recognition to gp33\textsuperscript{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\textsuperscript{103} to maintain the prophage) with the anti-gp33\textsuperscript{103} antibody to check native levels of repressor shows no bands in the lysogen corresponding to gp33\textsuperscript{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\textsuperscript{103} (~50ng). We used 1x10\textsuperscript{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\textsuperscript{103} antibody to determine the concentration of repressor in a lysogen

(A) Testing the 28951 anti-gp33\textsuperscript{103} antibody on a series of gp33\textsuperscript{103} dilutions. (B) Testing the anti-gp33\textsuperscript{103} antibody on a BPs lysogen. No gp33\textsuperscript{103} was detected.

Gp33\textsuperscript{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 \textit{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.
Our approach was to use custom-made polyclonal antibodies that recognize the BPs repressor in Western blots to detect \( \text{gp33}^{103} \) in a lysogen and compare that to known quantities of protein to roughly quantify the protein concentrations. We were unsuccessful in detecting \( \text{gp33}^{103} \) in a lysogen. This could be for a variety of reasons; either \( \text{gp33}^{103} \) is not present in high enough levels to meet the detection threshold of the antibodies, \( \text{gp33}^{103} \) could be insoluble and/or be unrecognizable by my antibody in the context of a whole cell lysate, or perhaps \( \text{gp33}^{103} \) 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 \( \text{gp33}^{103} \) in a BPs lysogen, these results don’t necessarily mean that the \textit{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.


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


