GENE EXPRESSION IN MYCOBACTERIOPHAGE BPs

by

Lauren Marie Oldfield

Bachelor of Science, University of Akron, 2008

Submitted to the Graduate Faculty of the

Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Lauren Marie Oldfield

It was defended on

May 1, 2014

and approved by

Karen Arndt, Professor, Department of Biological Sciences

Jon P. Boyle, Assistant Professor, Department of Biological Sciences

Roger Hendrix, Professor, Department of Biological Sciences

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

Dissertation Advisor: Graham F. Hatfull, Professor, Department of Biological Sciences

GENE EXPRESSION IN MYCOBACTERIOPHAGE BPs

Lauren M. Oldfield, PhD

University of Pittsburgh, 2014

Temperate bacteriophages can grow lytically, producing progeny phage and lysing the host, or lysogenically, integrating into the chromosome and shutting off the expression of lytic genes. The decision is controlled by a genetic switch, which senses host intracellular conditions and chooses a program of gene expression to employ. The genetic switch of mycobacteriophage BPs requires integration to produce stable repressor protein and maintain lysogeny. The regulation of this switch is controlled at the protein level, not at transcription initiation. Both integrase and repressor are tagged for proteolytic degradation and their concentrations are sensitive to host protease levels. The concentration of integrase determines the fate of BPs. The overall result of the genetic switch is determining the program of transcription. In lytic growth, genes are temporally expressed to replicate the phage genome, assemble progeny phage and lyse the host. In lysogenic growth, transcription is silenced in the phage and the repressor protein maintains repression of lytic phage genes and prevents superinfecting phage from growing lytically as well. We sought to understand the genetic switch of mycobacteriophage BPs and the transcriptional profile of gene expression. Transcription in mycobacteriophage BPs is temporally regulated and the mapping of promoter and terminator signals demonstrate additional methods of regulation in the expression and timing of transcription in mycobacteriophage BPs.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

Mycobacterium tuberculosis, the causative agent of human tuberculosis, kills two million people annually [1]. The study of mycobacteriophages, viruses that infect mycobacterial hosts, has led to the development of tools for the genetic manipulation and detection of mycobacteria. The isolation, sequencing and characterization of mycobacteriophages have also led to deeper understanding of phage diversity and evolution. More than six hundred mycobacteriophages have been sequenced (phagesdb.org) and a comparative analysis of 471 phages shows that there are around 30 distinct types, organized into clusters [2]. Several individual phages and mycobacteriophage clusters have been characterized [3-10].

Though it has been more than twenty years since the first mycobacteriophage genome was published [4], relatively little is known about mycobacteriophage gene expression. The detailed studies of the gene expression patterns and signals of mycobacteriophages conducted here will increase understanding of how bacteriophages function. At approximately 42 kilobases, the genomes of mycobacteriophages in cluster G, which includes mycobacteriophage BPs, are the shortest of any of the sequenced mycobacteriophages [11]. BPs is an ideal model for this investigation of gene expression due to its small size, as well as other interesting features including a novel genetic switch and ultra small genetic mobile element. This study aims to provide a comprehensive picture of mycobacteriophage BPs gene expression by examining the transcriptome and the transcription initiation and termination signals of BPs during lysogenic

growth and lytic infection of *Mycobacterium smegmatis* mc^2155, which is related to *Mycobacterium tuberculosis*.

1.1 MYCOBACTERIOPHAGES

The study of mycobacteriophages pursues two primary goals. First, these phages can, and have, been used to advance the study of their mycobacterial hosts, notably *M. tuberculosis*. Second, the detailed investigation of mycobacteriophages has increased our understanding of phage diversity and evolution. A detailed characterization of mycobacteriophage BPs's gene expression will broaden understanding of phage gene expression patterns overall and has the potential to develop BPs as a model of mycobacteriophage gene expression. The detailed study of the transcription initiation and termination signals could elucidate these mechanisms in the host and be useful in the generation of further genetic tools for mycobacteria.

The genus *Mycobacterium* contains an array of environmental bacteria and pathogens, including *M. leprae* and *M. tuberculosis*. *Mycobacterium tuberculosis*, the causative agent of human tuberculosis, is a slow-growing (>20 hour doubling time) intracellular pathogen. A saprophytic member of the genus, *Mycobacterium smegmatis*, is often used as a model for *M. tuberculosis* in the laboratory. *M. smegmatis* has a shorter doubling time, approximately 3 hours, and is nonpathogenic. In 1993, the World Health Organization (WHO) declared tuberculosis (TB) a global health emergency [12]. It is estimated that one-third of the world's population is infected with *M. tuberculosis* and nearly two million people die annually of TB [1,13]. The recent emergence and spread of multidrug-resistant and extensively drug-resistant strains of *M. tuberculosis* has led to a need for new and faster tools to test antibiotic susceptibility [14], and

one approach is utilizing mycobacteriophages that infect *M. tuberculosis* to detect drug resistant specimens [15]. Mycobacteriophages have been engineered to express luciferase or green fluorescent protein (GFP) for visual detection [15-18] and examined as tools to detect drug resistance in *M. tuberculosis* clinical samples but have mostly demonstrated only variable sensitivity [19]. A recent endeavor by the Hatfull laboratory, which employs fluoromycobacteriophages to detect live *M. tuberculosis* cells after the addition of rifampicin or streptomycin, has been shown reporter phages to be as sensitive as traditional methods of drug-susceptibility testing and able to provide results much faster than and at lower cost than expensive molecular techniques [15].

Mycobacteriophages have been used to develop a number of tools used for the genetic manipulation of mycobacteria, including facilitating the delivery of transposons and of substrates for allelic exchange through the use of specialized transducing shuttle phasmids [20]. Although these techniques enable the creation and recovery of *M. tuberculosis* mutants, they also possess considerable drawbacks, requiring multiple rounds of screening or the construction of complicated genetic substrates [21]. Therefore, a mycobacterial recombineering system was developed, using the *Escherichia coli* Rac prophage recombineering system as a model [22]. The mycobacterial recombineering system, which utilizes mycobacteriophage homologues of RecE and RecT from Che9c, is simple, efficient and has been successful in making targeted mutations in *M. smegmatis*, *M. tuberculosis* and *M. abscessus* [23,24]. This recombineering approach for mycobacterial host manipulation also led to the development of a new approach to manipulate the phages themselves, called bacteriophage recombineering of electroporated DNA (BRED), which has been shown to efficiently make deletions in essential and nonessential genes, point mutations, small insertions and gene replacements in several diverse mycobacteriophages [25].

It is estimated that there are 10^{31} bacteriophages on earth, making them the most abundant organisms in the biosphere [26]. As the number of sequenced bacteriophages grows, more becomes known about how phages have evolved and some trends of overall phage genomic diversity and architecture have been observed [reviewed in [27]]. There is a surprisingly large amount of diversity within mycobacteriophages, and it is not uniform, which allows similar mycobacteriophages to be clustered [28]. The genomes are highly mosaic, which is attributed to acquiring genetic information through horizontal gene transfer. The annotated genes have also been clustered into "phamilies," groups of related phage genes, based on their amino acid sequence. Of the 1,536 "phamilies" of mycobacteriophage genes, only 15% have significant sequence similarity to non-mycobacteriophage database entries and many of the phamilies contain only a single member [29]. This demonstrates the wealth of novel genes that are found in phage, specifically mycobacteriophage, genomes. The ability to create mutations using BRED has allowed for a closer assessment of the functional genomics of mycobacteriophages. The function of *lysB*, a gene predicted to be involved in lysis, was determined, in part, through a deletion created by BRED. Phages lacking *lysB* were shown to have a small plaque phenotype and display delayed and incomplete lysis [30]. This demonstrates that BRED is a particularly valuable tool and will aid in determining the functions of novel mycobacteriophage genes.

BPs is a temperate mycobacteriophage, capable of undergoing growth as a lysogen, where the prophage is integrated into its host's chromosome, or lytically, which results in the production of phage progeny and lysis of the host cell. BPs was chosen for detailed gene expression studies due to its small genomic size and interesting features, for example the location of the *attP* within the repressor gene and the discovery of an ultra small genetic mobile element

that warranted further exploration [10]. Also, BPs is known to be amenable to creation of unmarked mutations using BRED [25].

1.1.1 Cluster G

Cluster G contains six members that have been published in GenBank (phagesDB.org): Angel, Avrafan, BPs, Halo, Hope and Liefie (Fig. 1-1). They range in size from 41,441 base pairs (bp) to 42,289 bp with a guanine and cytosine (GC) content that ranges from 66.6% to 66.8%. Cluster G phages Angel, BPs and Halo have been comparatively investigated in greater depth [10]. The genomes of these phages are remarkably similar to each other with 99% nucleotide identity, with the exception of some small insertions in the far right end of the genome [10]. The general organization of the genomes of these phages is simple. The left arm consists of structure and assembly genes, which demonstrate canonical synteny [11] and a recognizable *lysA* and *lysB* (genes *27* and *28* in BPs) [10]. The genome arms are split by an immunity cassette (genes *32* and *33* in BPs), the only genes annotated that are transcribed from the opposite strand, in the leftwards direction. BPs gene *32* and its homologs in the other cluster G phages can be bioinformatically identified as tyrosine integrases. The right arm of BPs contains many small open reading frames (ORFs) and only four can be assigned functions by database searches. Putatively involved in replication or recombination functions of BPs, *42* and *43* encode genes with RecE and RecT homology, respectively, gene *51* is a putative Holliday junction resolvase and finally, gene *62* has similarity to HNH endonucleases [10]. The simple organization of the genome, with only one small cassette of the immunity genes transcribed from the opposite strand, and the small genomic size of cluster G phages below led to the utilization of BPs in the gene expression studies discussed here.

One notable difference between these three cluster G phages is the absence or presence and location of the insertion of a putative mobile element in the right arm. The BPs and Halo genomes demonstrate the insertion of a small gene belonging to Pham139, which are annotated as gene *58* and *56*, respectively (Fig. 1-1). Angel, on the other hand, lacks both of these insertions. These phages are closely related such that the pre-integration sites can be determined, demonstrating the insertion of the putative mobile element into a previously intact ORF. Pham139 members are present in 53 mycobacteriophage genomes from diverse clusters, including F, G, I, N, O, and T (phagesdb.org). This evidence supports the hypothesis that these genes are mobile elements and have transposase activity, though *in vitro* activity has not been demonstrated. Additionally, if verified, these would be the smallest known prokaryotic mobile elements [10].

Another compelling feature observed from examination of the sequences of Cluster G mycobacteriophages was the arrangement of the *rep* gene (gene *33* in BPs) and the *attP* core (See Chapter 5, Fig. 5-1). In these phages, *attP* occurs within the C-terminal portion of the coding region of the *rep* gene, which produces the immunity repressor (Rep) responsible for conferring immunity to superinfection [10]. BPs and Halo are able to lysogenize *M. smegmatis* mc^2155 at a frequency of ~5% [10]. Though Angel is nearly identical to BPs and Halo, it is a virulent phage, unable to form lysogens [10,31]. The coding region of *rep* contains the *attP* and the consequence of integration into the host chromosome is the loss of the C-terminal end of Rep. Halo is homoimmune to BPs, however expression of full-length Halo or BPs repressor protein does not confer immunity [10,31]. The placement of the *attP* site within the *rep* gene and functional consequences for repressor activity leads to the hypothesis that this is a novel genetic switch mechanism that controls the decision between lytic and lysogenic growth in temperate

phage BPs. Chapter 5 of these studies will further elucidate the role of the *attP* location in the *rep* gene in the establishment of lysogeny.

BPs is amenable to genetic manipulation by bacteriophage recombineering of electroporated DNA (BRED), a technique derived from mycobacterial recombineering to create unmarked mutations in mycobacteriophages [25]. Specifically, deletions of genes *44*, *50*, *52*, *54* and *58* were engineered in BPs and a replacement of gene *54* with *gfp* was created [[25], Appendix B]. The ability to easily create unmarked mutations in the BPs genome is essential for these and future studies of the determinants of mycobacteriophage gene expression and the further use of that knowledge to help in ascertaining gene functions.

Further, although BPs does not infect *M. tuberculosis*, host range expansion mutants can be isolated at a frequency of approximately 10^{-5} and are able to subsequently infect *M. tuberculosis* and *M. smegmatis* with equal efficiencies [10]. This host range expansion makes it possible to utilize BPs as a diagnostic tool for *M. tuberculosis* and also, demonstrates that the gene products of BPs allow it to grow lytically in *M. tuberculosis*.

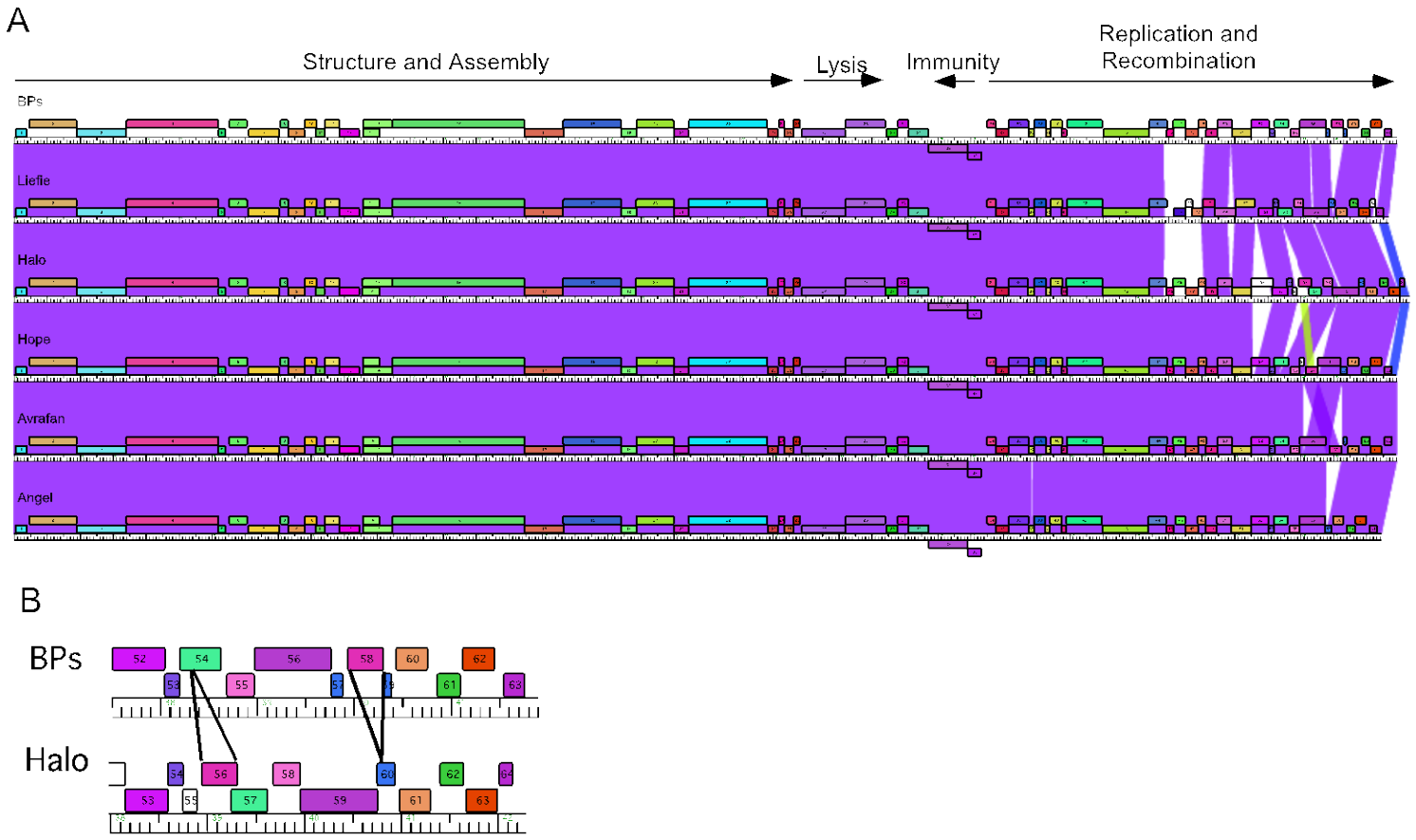

Figure 1-1. Cluster G mycobacteriophages are genetically similar.

Fig. 1-1. (A) Alignment in Phamerator of the Cluster G genomes of BPs, Liefie, Halo, Hope, Avrafan and Angel shows the genetic similarity among them based on the color spectrum with violet being the most similar [32]. (B) Location of mobile element insertions, BPs gene *58* and Halo gene *56*, using the genome of the other phage as a reference for the unmodified parent genome.

1.2 GENE EXPRESSION IN MYCOBACTERIOPHAGES

Gene expression in bacteriophages is temporally controlled and in temperate phages the decision of whether to enter lysogeny or lytic growth is controlled by a genetic switch. The organization of the structural and assembly genes in bacteriophage genomes generally follows a conserved synteny [33]. Due to conserved genome architecture of bacteriophages, it is possible to use gene location to hypothesize gene functions without direct evidence. Additional evidence for the functions of genes, and therefore the overall biology and the lifecycle of the bacteriophage, can be provided by the patterns of gene expression can provide additional insight into the gene functions.

1.2.1 Mycobacteriophage gene expression through analysis of L5

Gene expression studies of mycobacteriophage L5, the first complete mycobacteriophage genome sequenced [4], and the action of its repressor gene have provided an insight into mycobacteriophage gene expression. An examination of gene expression during lytic growth by ^{35}S-methionine labeling revealed L5 has two phases of protein expression: early, which occurs from 10 minutes to 45 minutes post-infection, and late, which happens from 25 minutes post-infection until lysis [4]. The repressor protein was identified as gp71 through its ability to confer immunity to superinfection [34]. Further study of the regulation of gene *71* showed that mRNA from this locus was significantly reduced by 10 minutes after the induction of lytic growth from

a lysogenic state [35]. This corresponds to the beginning of early phage protein expression. A promoter upstream of gene *71*, P_{left}, was identified and shown to be active in early lytic growth [35]. P_{left} is negatively regulated by the binding of gp71 and the identification of this binding site revealed 26 additional sites in the L5 genome bound by gp71 [36]. Some of these sites are found in structural genes in the left arm, which are not expressed until late lytic growth and not all are coupled with promoters. Thus, these sites were suspected to have an effect on transcription elongation, not initiation. It was further demonstrated that the placement of a binding site in the same orientation as transcription and in the presence of gp71 strongly decreases the activity of a reporter [36]. These binding sites, termed stoperators, have been used to determine a model for transcriptional silencing of the prophage and also for the lysogenic-lytic switch of L5. P_{left} may be responsible for most gene expression in early lytic growth; however, another a promoter was identified in gene *83* immediately upstream of three cytotoxic genes, whose activity is controlled by a nearby stoperator [37].

Little is known about the expression of other genes in the right arm of the genome, though the L5 repressor expression and regulation is well characterized, except that they are predicted to be expressed early in lytic growth. In addition, almost nothing is known about late gene expression. No late promoters have been identified, nor any genes confirmed to be expressed late in infection, though it is expected that the structural genes located in the left arm of the genome are expressed in late lytic infection.

Protein expression profiling of three mycobacteriophages, including L5, demonstrated the phages undergo two phases of protein expression, one soon after lytic growth begins and one later in lytic growth [3-5]. Newly synthesized proteins were labeled with ^{35}S-methionine at several time points throughout lytic growth and two distinct patterns were detected throughout

lytic growth, indicating that protein expression is temporally controlled. Additional shifts in protein expression were not apparent though 35S-methionine labeling might miss subtle shifts in gene expression. Also, in L5 and Bxb1, host protein synthesis is shut down during lytic growth through an unknown mechanism [4,5], while in TM4 host protein synthesis does not appear to be affected [3].

Recently, an RNA-Seq analysis of temperate mycobacteriophage Giles was conducted [see Chapter 3;[8]]. Giles was found to have different patterns of expression in early lytic (30 minutes post adsorption), late lytic (2.5 hours post adsorption) and lysogenic samples. During lysogeny, the *rep* gene (*47*) gives the strongest signal in RNA-Seq and very few other regions are expressed, though a 100 bp intergenic region between genes *74* and *75* is also highly expressed [8]. In early lytic growth, an operon from genes *49* to *59* is expressed at a modest level, along with three other small regions, however, the genes involved in virion structure and assembly are not expressed [8]. In late lytic growth, the expression during early lytic growth is retained, while the structure and assembly genes are very highly expressed [8]. Also expressed differentially in late lytic growth an operon containing the lysis genes (*36* and *37*) is moderately expressed and the same intergenic found to be strongly expressed in the lysogen is also strongly expressed [8]. Giles displays distinct patterns of expression at these three different points in its lifecycle, lysogenic growth and early and late lytic growth. As with the protein profiles of mycobacteriophages L5, Bxb1 and TM4, Giles has at least two phases of gene expression during lytic growth, which are temporally regulated. However, the expression pattern may be more finely tuned than this analysis demonstrates because only two time points post-adsorption were examined.

1.2.2 Mycobacteriophage promoters

Mycobacteriophage genomes are organized into long operons allowing for the possibility of expression of a large number of genes from a small number of promoters. A relatively small number of mycobacteriophage promoters have been identified.

Though mycobacterial promoters generally function poorly in *Escherichia coli* [38], some mycobacteriophage promoters have been detected through screens of libraries of clones containing fragments of genomic phage DNA in *E. coli*. The sequences of four promoters from mycobacteriophage I3 were identified through a screen in *E. coli*, though the transcription start sites were not determined and the promoters never tested for activity in mycobacteria [39]. The locations of the promoter sequences in I3 are unknown, as a complete genome sequence has not been published. In L1, a mycobacteriophage that is homoimmune with L5, three promoters were identified in a similar screen in *E. coli*, however only one of these promoters showed marked reporter activity in *M. smegmatis* [40].

Five promoters from mycobacteriophage L5 have been characterized. A collection of three promoters (P1, P2 and P3) are employed to express the repressor, gp71, of L5 [35]. The transcription start sites of all three have been mapped [35]. The sequences of two of the promoters, P1 and P2, resemble *E. coli* Sig70 promoters, and though activity is low, these promoters are recognized by *E. coli* RNA polymerase *in vitro* [35]. The transcripts from these promoters are detected in an L5 lysogen, but not in lytically growing phage [35]. The expression changes in these three promoters could be due to the shut down of transcription from promoters recognized by host RNA polymerase shortly after L5 infection during lytic growth [4]. A strong early promoter, P_{left}, was also identified in L5 [35]. This promoter is regulated by the repressor gp71, and is not active in an L5 lysogen [35,36]. P_{left} is characterized as an early promoter but

transcripts are also present in late lytic growth, even though P_{left}, like P1, P2 and P3, is recognized by the host RNA polymerase [35]. The 5' end of transcripts from P_{left} are stable and persist for at least 20 minutes, which marks the beginning of late protein expression [35]. Recently, an additional L5 promoter was described that expresses a set of cytotoxic genes, thought to be responsible for the shut down of host gene expression functions [37].

In Bxb1, divergent promoters P_{left} and P_{right}, which are both regulated by its repressor, gp69, were identified [41]. P_{left} of Bxb1 is analogous to P_{left} in L5, and though Bxb1 and L5 are genetically similar, they are heteroimmune, due to variation in the binding sites for the repressors [41]. Finally, two Sig70-like promoters, P1 and P2, which express the lysis gene is Ms6 were identified [42].

A comparison of the sequence conservation at the -35 and -10 motifs of all mycobacteriophage promoters described shows that the best conserved nucleotides are 5'-TTGACN and 5'-TANNNT for the -35 and -10 hexamers, respectively. Though the mycobacteriophage consensus resembles the *E. coli* Sig70 consensus sequence, most of these promoters do not function well when tested in *E. coli*. Promoters active in lysogenic and lytic growth have been described. No late lytic promoters, which are expected to have the highest activity in mycobacteriophages, have been identified. Late lytic promoters may require phage-encoded activators and therefore being more difficult to identify in screens using reporter genes.

1.2.3 Gene expression in other bacteriophages

As the details of mycobacteriophage gene expression are not well characterized, a brief examination of the gene expression patterns and features of other bacteriophages is helpful to understand the commonalities and the diversity of mechanisms employed. Gene expression from

a number of different host bacteria, including *Escherichia coli*, *Yersinia* spp., *Bacillus subtilis*, *Lactococcus lactis* and *Lactobacillus*, have been studied in varying levels of detail [43-45].

A common feature of gene expression patterns in bacteriophages is the categorization of genes based on their temporal regulation (i.e. early, middle, late). Early genes are transcribed soon after DNA injection into the host cell and sometimes before injection is complete. In the T-even phages, early genes are classified as those that do not require other phage proteins for expression [44]. The functions of many of the early genes in T-even phages is to disrupt host gene expression and increase the preference for T-even promoters [46]. For example, early protein Alc selectively terminates transcription elongation of the host cytosine-containing DNA, while allowing continued transcription of the phage's DNA, which instead contains hydroxymethylcytosine [47,48]. Alt, a protein that is injected from the viral particle, is responsible for ADP-ribosylation at Arg^{265} of one of the α subunits of RNAP, which is involved in the recognition of UP elements in promoters [49]. This modification results in a 2-fold activation of T4 promoters [50].

1.2.4 Genetic switches control decision between lysogenic and lytic growth

The genetic switch of *E. coli* phage lambda is the most well studied switch. It is a complex pathway with many components, but a single goal, to switch between lytic and lysogenic growth [51].

During lytic growth, when lambda injects its genome into an *E. coli* cell, RNA polymerase binds two promoters, one for rightwards expression (P_R) and one for leftward expression (P_L). The first gene expressed from P_L is *N*. The product of the *N* gene acts as anti-termination factor and blocks transcript termination, allowing more genes to be expressed from

both P_R and P_L [52]. Downstream of P_L are genes required for recombination as well as two proteins required for lysogeny, *cIII* and further down, *int* [53]. Downstream of P_R are genes *cro*, *cII*, *O*, *P*, and *Q*. The Q protein acts as another antitermination factor, which allows increased translation of the products of the $P_{R'}$ promoter, which are lytic genes [54]. When these are transcribed, new phage particles will be assembled and lyse the host cell, releasing progeny phage. To ensure the lytic pathway continues, *cro*, the first gene transcribed from P_R, accumulates and binds operators within P_{RM}, preventing the repressor gene, *cI*, from being transcribed [55]. The repressor, CI, represses lytic genes, therefore Cro is an essential pro-lytic regulator [51].

During lysogeny, the N protein functions to antiterminate the P_R transcript that contains the *cII* gene. This gene is also transcribed during lytic growth, but it is susceptible to host proteases such as FtsH [53]. Therefore, it is rapidly degraded during lytic growth and not involved further in lytic growth. During host cell conditions that promote lysogeny, however, host protease levels are typically low and the CIII protein, transcribed from P_L after N antitermination, is able to sequester FtsH and allow CII to accumulate [56]. When CII accumulates, it performs a few important functions. The first is that it binds to a promoter, pINT, and permits expression of the integrase, which is responsible for integrating the phage genome into the bacterial host chromosome, creating a prophage [51]. This stabilizes the phage genome. Second, CII binds to the beginning of *Q* to create an antisense RNA to the *Q* gene, which prevents further expression of Q, effectively hindering further lytic gene activation [54]. The third function of CII is to bind to a leftward facing promoter, P_{RE}, which transcribes the repressor gene, *cI*. Once CI begins to accumulate, it binds upstream of its own promoter, P_{RM}, and activates its own transcription, and also binds within the adjacent P_R to prevent further

downstream lytic gene expression [53]. CI competes with the lytic protein Cro for operator sites [57]. When CI is bound to its main operators, P_R is shut down and *cro* cannot be transcribed. When Cro is bound to its main operator, P_{RM} is shut down and *cI* is not transcribed.

1.3 MYCOBACTERIAL TRANSCRIPTION INITIATION

Gene expression in prokaryotes is generally controlled and regulated at the level of transcription, specifically transcription initiation. Transcription initiation is a process that includes the binding of the RNA polymerase (RNAP) holoenyzme to a promoter element of the DNA and the melting of several base pairs around the transcription start site. Though the later steps of transcription initiation are important determinants of gene expression [58], the binding of the RNAP to the promoter is generally regarded as the most influential for transcription of the gene. The affinity of the RNAP holoenzyme for the promoter site is mostly determined by three factors: the promoter sequence, the identity of the sigma factor bound to the holoenzyme and the effect of any transcription factors at the specific promoter.

1.3.1 Promoter sequence determinants

Unlike the promoter sequence determinants for gene expression for *E. coli* σ^{70} promoters, mycobacterial promoter sequences have been difficult to identify. Many of the first promoters studied did not follow the consensus sequence predicted for prokaryotes based on the *Escherichia coli* σ^{70} promoter consensus and no other consensus sequence could be generated

from the few known promoters. After many years of study and the characterization of many more mycobacterial promoters a pattern has begun to emerge.

The -10 hexamer of mycobacterial promoters was characterized from randomly isolated mycobacterial promoters. The -10 region of *M. smegmatis* and *M. tuberculosis* promoters was conserved, with the consensus sequence $T_{100}A_{93}T_{50}A_{57}A_{43}T_{71}$ for *M. smegmatis* and $T_{80}A_{90}Y_{60}G_{40}A_{60}T_{100}$ for *M. tuberculosis*, where the capital letters indicate the base and the subscript indicates its frequency [59]. These -10 sequences resemble the consensus for *E. coli* σ^{70} promoters, $T_{82}A_{90}T_{52}A_{59}a_{49}T_{89}$ [60]. No mycobacterial consensus for the -35 element, either similar or dissimilar to that of the *E. coli* promoter, was found [59]. However, the -35 region is an important contributor to the strength of the promoter. When Bashyam and colleagues deleted the region upstream of the -10 element, including the -35 element, promoter activity decreased 90% [59]. However, the -10 region can provide sufficient promoter activity alone to drive expression of a chloramphenicol resistance cassette and provide resistance [59]. Single point mutations in the -10 hexamer of mycobacterial and mycobacteriophage promoters decrease activity by between 50% and 95% [42,61]. Providing different sequences, whether by swapping -35 regions intentionally or inadvertently through insertions in the spacer region or cloning mishaps, can change promoter activity between 2- and 10-fold [16,62,63]. Taken together, these data show that the presence and sequence of the -35 and -10 hexamers are vital for promoter activity in mycobacteria.

To further elucidate the optimal -35 sequence, Agarwal and Tyagi completed a mutational analysis of the -35 element of a synthetic strong mycobacterial promoter [64]. They found that single base pair substitutions showed the optimal -35 sequence for this promoter was 5'-TTGCGA [64]. When the -35 sequences of native mycobacterial promoters were mutated to

this sequence, the RNAP bound to the promoter DNA with higher affinity [64]. This optimal mycobacterial -35 sequence is very similar to the *E. coli* -35 consensus, $T_{82}T_{83}G_{78}A_{64}C_{53}A_{44}$ [60], indicating *E. coli* σ^{70}-like -35 elements would function well in this promoter and mycobacterial promoters in general but had not been identified from native promoters isolated.

Another determinant of promoter activity is the presence of an extended -10 motif, which consists of a 5'-TGN sequence immediately upstream of the -10 hexamer. Extended -10 regions have been indentified in many bacterial species, including *E. coli* [65], *Bacillus subtilis* [66] and *Streptococcus pneumoniae* [67], as well as the *rpsL* promoter of *M. smegmatis* [61]. In *E. coli*, region 2.5 of the σ factor contacts the extended -10 [68] and transcription can initiate at promoters with an extended -10 motif but lacking a -35 hexamer [69,70]. This led to the hypothesis that since mycobacterial promoters seemed to lack a functional -35 region, extended -10 motifs might play an important role in promoter activity. However, even in the presence of an extended -10 motif, providing various -35 sequences modulated promoter activity [63].

Extended -10 motifs are reported to occur in approximately 25% of mycobacterial promoters [63], compared to 16% of *E. coli* promoters [70] and 60% in gram-positive bacterial genera with low GC contents [63]. Mutational analysis of a naturally occurring 5'-TGN motif in an *M. smegmatis* promoter decreased promoter strength between 2- and 13-fold [63,71]. Even more striking, adding a 5'-TGN motif to a promoter that does not naturally contain an extended -10 element increased promoter strength by between 2- and 75-fold [63].

Gomez and Smith compiled 102 mycobacterial promoter sequences in which the transcription start site had been experimentally determined [72]. The promoters were divided into four groups based on the sequence of the -35 and -10 motifs. Group A contained promoters that had a conserved -35 and -10 sequences. Comparing the consensus for Group A promoters

from *M. smegmatis* to *E. coli* σ^{70}-like promoters, the consensus sequences are the same for both the -35 and -10 elements. The consensus sequences for the organisms only differ in their relative conservation—the frequencies for the consensus nucleotide; the *E. coli* promoters are more well conserved than the mycobacterial promoters. The *M. tuberculosis* Group A promoter elements were also very similar to the *E. coli* consensus [60] but generally less well conserved than the *M. smegmatis* promoters [72]. Group B mycobacterial promoters contain a more divergent -10 hexamer, but no conserved -35 sequence, and had a higher GC content, which may constrain the sequence [72].

Gomez and Smith found that there are also promoters that either do not conform to any known consensus sequence or do contain any conserved consensus. Group C promoters do not resemble the *E. coli* σ^{70} consensus and are likely upstream of genes regulated by alternative sigma factors or activators [72]. Group D promoters are a distinct class of promoters with a common consensus found in *M. paratuberculosis* and represent a different consensus [72].

Additional mycobacterial promoter consensus sequences have been published [59,73-76]. Most use different methods of organization, making it difficult to compare and contrast the findings directly. The other prevailing method of organization, besides the Gomez and Smith Groups A to D, is to divide the promoters into two groups. One encompasses σ^{70}-like promoters, which are SigA mycobacterial promoters, the housekeeping sigma factor in mycobacteria. Although the recognition of these promoters by σ^{A}. The second class is promoters that have a higher GC content, referred to as SigGC, and are likely transcribed by an alternative sigma factor [73,74]. The consensus sequences generated for the SigA group correlates to the Group A promoters of Gomez and Smith while the SigGC promoters correlate to the Group D promoters,

which is unexpected because the Group D promoters are a specific group of *M. paratuberculosis* promoters while the SigGC are general mycobacteria or *M. tuberculosis* promoters [72-74,77].

In summary, many mycobacterial promoters have identifiable σ^{70}-like consensus sequences and are presumably recognized by SigA, the housekeeping sigma factor in mycobacteria. Conversely, promoters from the Group C or SigGC classes [72-74] do not contain promoter sequences that are similar to the SigA class of promoters and are likely recognized by unspecified alternative sigma factors. The high number of sigma factors in mycobacteria could contribute to the inability to determine any similarity upstream of the transcription start site. These promoters with unidentified consensus sequences actually represent many different groups of regulated promoters.

1.3.2 Mycobacterial sigma factors

The RNA polymerase holoenzyme consists of a five-subunit core polymerase and a sigma factor (σ). The core enzyme is capable of catalytic activity, but the sigma factor is required for specific binding of the polymerase to a particular DNA sequence. Most bacteria encode several sigma factors, which are used to modulate gene expression under differing conditions. The sigma factor provides most, if not all, the determinants for promoter recognition and therefore is responsible for conferring promoter specificity to the RNAP. The association and dissociation of different sigma factors with the core polymerase, along with a number of transcription factors, are responsible for changes in gene expression. Typically, one sigma factor, which is related to the *Escherichia coli* σ^{70}, is responsible for the transcription of housekeeping genes. Additional alternative sigma factors vary in number depending on the bacterial species and are responsible for the expression of regulons induced during particular growth conditions or stresses.

The genomes of *M. smegmatis* and *M. tuberculosis* encode 28 and 13 sigma factors, respectively [78-80]. Both encode a sigma factor A gene (*sigA*), which is the housekeeping sigma factor and is essential [81]. The structural characteristics and phylogenetic relationships of sigma factors related to σ^{70} are used to categorize them into groups 1 to 4 [82]. Group 1 sigma factors contain all 4 conserved regions and are the housekeeping factors [82,83]; SigA is the sole group 1 sigma factor in mycobacteria [84]. Transcripts of *sigA* are maintained at constant levels, though there are reports of regulation in infection of human macrophages and in stationary phase [85-87]. Though SigA is the housekeeping sigma factor, it has been proposed that, in *M. tuberculosis*, it is also involved in the transcription of virulence factors and is utilized for the transcription of genes required for intracellular survival [88-90].

Generally, organisms that live in more varied environments encode for great numbers of sigma factors, in order to deal with changing conditions and respond to stresses [83]. Both of these mycobacterial organisms live in distinct and complicated environments. *M. smegmatis* is found in soil, which contains many, diverse and possibly competitive microorganisms, as well as variable chemical and physical properties. *M. tuberculosis* is a pathogen adapted to grow in harsh, dynamic intracellular environments and sigma factors control many of the host-pathogen interactions [91,92]. Of all obligate pathogens, *M. tuberculosis* has the highest sigma factor to genome size ratio, indicating a complicated regulatory mechanism and numerous stresses [93]. The genus *Mycobacterium* displays more variation in the number of sigma factors between members of the genus than any other bacterial genus, with 28 sigma factor genes in *M. smegmatis* and only 4 functional sigma factors in *M. leprae* [93].

Though the remaining sigma factors are not essential for laboratory growth, their functions and regulons of the alternative sigma factors of *M. tuberculosis* have been difficult to

elucidate [reviewed in [94]]. SigB is a principle-like sigma factor that is very similar to SigA [93,95], however it is non-essential [81,96]. With the exception of a few genes, SigB has a distinct regulon from SigA that does not overlap [89,97]. Though the *sigB* gene is upregulated with exposure to a number of stressors, inactivation of *sigB* did not impact survival of *M. tuberculosis* in models of infection [86,96,98]. SigB is regulated by four other alternative sigma factors in mycobacteria and seems to be a central player in a sigma network [94].

Many of the additional sigma factors in mycobacterial species are influenced by either external stressors or, for the pathogenic species, in models of infection. SigI, SigJ and SigM are expressed at high levels in stationary phase [94,99] and SigH is involved in the regulation of genes involved in response to oxidative and heat stress [100,101]. Some other alternative sigma factors have increased expression when infecting human macrophages or in other animal models of infection. Only *sigG* has been reported to be required for survival in macrophages [97]. However, a phenotype where growth of bacteria to high titers is possible but mortality is decreased, called an immunopathology (*imp*) defect phenotype, is observed with the inactivation of a number of sigma factors, including SigC, SigE, SigF and SigH [102-106]. Presumably, the deletion of these sigma factors permits growth of the pathogen but decreases virulence.

While sigma factors control overall gene expression in bacteria at the level of transcription initiation, much of the control over the pool of sigma factors available for recruitment to core RNAP is exerted post-translationally through the sequestration of sigma factors by anti-sigma factors. Anti-sigma factor proteins bind and sequester sigma factors, thereby preventing their interaction with RNAP. The anti-sigma factors are able to sense external signals through phosphorylation and redox conditions and release their cognate sigma factor

under the appropriate conditions [107-109]. Adding another layer of complexity, the anti-sigma factors can be regulated by anti-anti-sigma factors as well [110].

The identification of the promoter sequences recognized by distinct mycobacterial sigma factors has been undertaken by comparative analysis of gene expression profiles in strains with an inactivated sigma factor [94]. Because of the complicated regulatory network amongst sigma factors and their ability to affect transcription indirectly through the control of other sigma factors and transcription factors, the ability to pinpoint the promoter sequences recognized by each sigma factor has been limited. Another method for determining promoter sequences is comparing the sigma factor-specific sequences in other related organisms, which though potentially helpful, may bias the promoter search [111]. Promoter consensus sequences for 12 out of the 13 sigma factors of *M. tuberculosis* have been proposed [reviewed in [94]]. Notably, the -35 motifs of the consensus sequences are quite similar for SigD, SigE, SigH, SigL and SigM and are GC-rich[101,112-115]. These sigma factors, which are structurally categorized as group 4 and lack the region 3.0 responsible for binding to extended -10 motifs, likely depend on the -35 for binding while the -10 confers specificity [114].

1.4 TRANSCRIPTION TERMINATION IN MYCOBACTERIA

Transcription termination occurs when the RNA polymerase dissociates from the RNA molecule being elongated, the RNA:DNA hybrid that is contained in the transcription bubble and the template DNA. Termination events are categorized broadly in one of two groups, Rho-dependent termination, which requires the presence of Rho protein to cause termination, and intrinsic termination, which does not require any additional factors. An intrinsic, also called Rho-

independent or simple, terminator is defined as a DNA sequence that has the ability to cause termination in an *in vitro* reaction with no additional termination factors present [116,117]. Classically, they are composed of a GC-rich palindrome, which can form a stem-loop structure, also called a hairpin, followed by a short stretch of T nucleotides on the coding strand of the DNA, which become a stretch of U nucleotides in the RNA [118]. The U-stretch is responsible for pausing the elongating RNAP [118] and decreasing the rate of elongation has been shown to promote transcript release [119]. The hairpin is responsible for the destabilization of the elongation complex by eliminating contacts between the RNA and the complex and the RNA:DNA hybrid and the complex [118].

The mechanism of intrinsic transcription termination has mostly been studied in *E. coli* and is not well understood for mycobacteria. It was suggested, in the past, that *Mycobacterium*, along with many other types of bacteria, employed a different mechanism of termination because classic intrinsic terminators with both hairpin structures and T-stretches had not been identified [120,121]. However, intrinsic terminators from *E. coli* function in other bacteria, including mycobacteria, which demonstrates that the overall mechanism is conserved [122].

New prediction algorithms that do not rely on identifying stretches of T nucleotides have found that many *Mycobacterium sp.* do contain sequences that are likely to form stable hairpins but most lack the canonical T-stretch [122-124]. These terminators are described as I-shaped, consisting of a hairpin structure with no U-tract [124]. The necessity of the U-tract for termination depends on the system examined and has been shown to be necessary [125], not necessary [126] and dependent on other downstream elements [127]. As a group, actinobacteria have a high preponderance of I-shaped terminators [123]. While this might be affected by the high GC content of these organisms, this is not the only factor. *M. leprae*, which has a lower GC

content than many other mycobacterial species, maintains a strong preference for I-shaped terminators

[123]. One possible explanation for the lack of T-stretches in mycobacterial I-shaped terminators is that the pausing function, which is the function of U nucleotides, is not necessary for mycobacteria because the rate of elongation is 10-fold slower than *E. coli* [124,128].

1.4.1 Antitermination is vital to timing of gene expression in bacteriophage λ

Gene expression can be controlled by the efficiency of termination at specified locations in prokaryotes and eukaryotes [129-131]. Antitermination mechanisms are employed to regulate the efficiency of some transcriptional terminators. As mentioned previously (see Chapter 1.2.4), *E. coli* bacteriophage λ utilizes two antitermination factors, N and Q, to regulate the timing of gene expression. Both N and Q proteins modify the RNA polymerase so that the complex transcribes through multiple transcriptional terminators located downstream.

After infection, N is made from the P_L promoter of λ and prevents termination at both early operons through recognition of N-utilization (*nut*) sites [131]. The N-modified elongation complex also consists of several host proteins, NusA, NusB, S10 and NusG [131-133]. All of these factors are necessary to form a stable complex and for processive antitermination [134-136]. The *nut* site contains two elements, *boxA* and *boxB*, and is in the nascent RNA [136-140]. It has been shown by electrophoretic mobility shift assay that N binds to *boxB* [141]. The *E. coli* ribosomal RNA operons have *boxA* elements that are closely related to the λ *nut* site *boxA* and function as antiterminators [142]. However, the λ *boxA* is not sufficient for antitermination and the λ *nut* site requires N and *boxB* [136]. *In vivo*, N-modified RNA polymerase is able to prevent

termination far downstream of the *nut* site, which likely occurs through RNA looping [143,144], although the mechanism of action is not understood [143].

The transcript of the other antiterminator in bacteriophage λ, *Q*, is produced from the N-antiterminated transcript of P_R and prevents termination caused by pausing of RNA Polymerase 16-17 nucleotides downstream of the $P_{R'}$ promoter [131,143,145]. The Q protein recognizes the paused transcriptional complex and accelerates through the pause site, allowing transcription of late bacteriophage genes, involved in phage morphogenesis and cell lysis [131]. Unlike N, which requires several host factors for its action, there is not evidence of host factor involvement with Q-mediated antitermination [143]. The Q-utilization (*qut*) site overlaps the $P_{R'}$ promoter and is located on the DNA, in contrast to N which recognizes *nut* sites on the nascent RNA [131]. The mechanism of Q antitermination may be analogous to that proposed for N and work through DNA looping or through modifications to the conformation of the RNA polymerase [143].

Although antitermination has not been explored for the regulation of timing of gene expression in mycobacteriophages, the elongation factors such as NusA, NusB, S10 are present in *M. tuberculosis* and the expression of ribosomal RNA from the *rrn* operon relies on antitermination mechanisms involving NusA [79,146,147]. At the only *rrn* operon in *M. tuberculosis*, elongation factors including NusA function similarly, but not identically, to the antitermination mechanism at the seven *rrn* operons of *E. coli* [146,148].

2.0 MATERIALS AND METHODS

2.1 BACTERIAL GROWTH AND MANIPULATIONS

2.1.1 Mycobacterial strains and growth conditions

Mycobacterium smegmatis mc^2155 strains were grown on 7H10 agar (Difco) supplemented with $CaCl_2$ (1 mM), CB (50 μg/ml) and CHX (10 μg/ml) and either 10% dextrose or 10% albumin dextrose catalase growth supplement (ADC). Appropriate antibiotics were added at the following concentrations: kanamycin (20 μg/ml), hygromycin (50 μg/ml). Colonies were grown at 37°C for between 3 and 5 days. Broth cultures of *M. smegmatis* mc^2155 strains were grown in 7H9 media supplemented with 10% ADC, CB and CHX. Additional appropriate were added when necessary. Tween80 (0.05%) was added when the cultures would not be used for phage infections, as it can interfere with the ability of some mycobacteriophages to infect mycobacteria. Liquid cultures were grown at 37°C for 2 to 3 days with shaking.

Mycobacterium tuberculosis mc^27000 [149] strains were grown on 7H11 agar (Difco) supplemented with $CaCl_2$ (1mM), CB (50 μg/ml), CHX (10 μg/ml), pantothenate (100 μg/ml) and 10% oleic albumin dextrose catalase growth supplement (OADC). Appropriate antibiotics were added at the following concentrations: kanamycin (20 μg/ml), hygromycin (50 μg/ml). Colonies were grown at 37°C for approximately 4 weeks. Broth cultures of *M. tuberculosis*

$mc^2$7000 strains were grown in 7H9 media supplemented with 10% OADC, Pan, CB, CHX and 0.05% tween and grown standing at 37°C for 2 to 3 weeks.

Mycobacterial strains were stored at -80°C in 20% glycerol and were typically streaked on solid media (either 7H10 or 7H11 agar), though direct culture into 7H9 broth was used at times for *M. smegmatis* strains.

2.1.2 *Escherichia coli* growth conditions

E. coli DH5α, *E. coli* 5α and *E. coli* DY331 strains were grown on LB agar (Difco) and in LB broth (Difco) supplemented with the following antibiotics when needed: kanamycin (20 μg/ml), hygromycin (150 μg/ml) and tetracycline (6.5 μg/ml). Colonies were grown overnight at 37°C standing and liquid cultures were grown overnight at 37°C shaking. Strains were stored at -80°C in 20% glycerol.

2.2 BACTERIOPHAGE GROWTH AND MANIPULATIONS

2.2.1 Bacteriophage infections of *M. smegmatis* $mc^2$155 on solid media

Dilutions of phage lysate, often 10-fold serial dilutions, are made in phage buffer supplemented with 1 mM $CaCl_2$. Infections were carried out by mixing approximately 300 μl of *M. smegmatis* liquid culture grown to saturation with 10 μl of diluted BPs or other mycobacteriophage and incubating the infection at 37°C standing for 30 minutes to allow for adsorption and the beginning stages of phage growth. The phage-infected *M. smegmatis* is then mixed with 4 ml of

phage top agar mixture (2.5 ml MBTA, 1.5 ml 7H9, 50 µl 1M $CaCl_2$) and immediately layered on top of a 7H10 plate containing ADC, CB, CHX, and $CaCl_2$. The top agar layer is allowed to solidify at room temperature and then the plate is incubated overnight at 37°C. The presence of plaques within the bacterial lawn or of complete clearing of bacterial growth indicates successful infection and lysis of host *M. smegmatis* cells.

2.2.2 Preparation of high titer phage lysate of BPs

Infections of BPs phage were prepared from a BPs lysate stock. To obtain a plate with the maximum number of phage, 10-fold serial dilutions of the lysate stock were used to infect an excess of *M. smegmatis* mc^2155 cells grown to stationary phase. The optimal number of infecting phage will leave a webbed pattern on the *M. smegmatis* lawn, which results in the most progeny phage produced. The infection was replicated so that at least 5 plates with a webbed pattern of clearing were produced. These plates were then flooded with 5ml of phage buffer with 1 mM $CaCl_2$ and incubated for several hours at room temperature. The phage buffer is collected from the plates and filter sterilized with a 0.22 µm filter to remove *M. smegmatis* cells.

2.2.3 Concentration of phage by ammonium sulfate precipitation

To further concentrate the lysate, the phage were precipitated by treatment with saturated ammonium sulfate solution. Briefly, an equal volume of saturated ammonium sulfate was added to the BPs lysate, gently mixed by inverting the tube and incubated on ice for 2 hours. The phage was pelleted by two rounds of centrifugation at 3500 x *g* at 4°C for 10 and 30 minutes, respectively. The supernatant was not disposed of between rounds. The phage pellet was located

and the supernatant was carefully discarded immediately. The pellet was resuspended in 500µl of phage buffer with 1mM $CaCl_2$. The resuspended phage was dialyzed to remove the ammonium sulfate in 15,000 MWCO dialysis tubing in phage buffer at 4°C twice (overnight and for 2 hours). Any unprotected nucleic acids, such as *M. smegmatis* DNA and ribosomes, were removed from the phage sample by addition of 2U of DNaseI and a final concentration of 100µg/ml of RNase A.

2.2.4 Mycobacteriophage infections in liquid culture

A liquid culture of *M. smegmatis* mc^2155 was grown to mid-logarithmic phase (between OD_{600nm} of 0.8 and 1.5) in media without Tween80 detergent. The optical density of the culture was measured and the number of cells was calculated (assume OD_{600nm}= 0.1 has 3.5 x 10^7 cfu/ml). The cells were infected with phage by adding the appropriate amount of lysate, calculated from the titer of the lysate and the number of bacterial cells, for the desired multiplicity of infection (MOI). By definition, an MOI equal to 1.0 is an equal number of plaque forming units (pfu) to the number of colony forming units (cfu). The infected cultures were allowed to stand for adsorption of the phage at room temperature for a short amount of time. Then the infected cultures were transferred to 37°C, shaking for the remainder of the infection time.

2.2.5 Bacteriophage recombineering of electroporated DNA (BRED)

Bacteriophage recombineering of electroporated DNA was carried out as described previously [25]). Substrates for recombineering were designed as previously described [25,150]). To insert the *gfp* gene, two 75-base primers were used to amplify *gfp* from plasmid pMN437 (a

generous gift from Michael Niederweis), with 25 bases complementary to each end of *gfp* and 50 bp of homology upstream and downstream of the inserted sequence. The PCR product was further extended by a second round of PCR to add an additional 50 bp of homology to each end (to generate a substrate with 100 bp homology on each end). All oligonucleotides were purchased from IDT Inc. and were gel purified.

Recombineering *M. smegmatis* mc^2155 cells were induced as described previously [22]. Briefly, cells were grown to OD_{600nm} = ~0.4 in 7H9 with 0.2% glycerol, 0.05% Tween 80, and 0.2% succinate. Cells were then induced with 0.2% acetamide, grown for 3 hours, washed three times with ice-cold 10% glycerol, resuspended in ~3 mL of 10% gylcerol and stored at −80°C. Aliquots of 100 µl were co-electroporated with phage genomic DNA and recombineering substrate, recovered at 37°C in 7H9 containing 10% ADC and 1 mM CaCl2 for 2 hours, and plated on 7H10 agar as top agar lawns with approximately 300 µl of *M. smegmatis* mc^2155.

Plaques were picked into 100 µl of phage buffer (10 mM Tris-HCl, pH 7.5; 10 mM $MgSO_4$; 68.5 mM NaCl; 1 mM $CaCl_2$). 1 µl was PCR amplified with detection primers (25–35 bp) where one primer anneals within the replacement allele and another in the flanking phage genomic DNA. Plaques containing mixtures of mutant and wild-type DNA were picked into 100 µl buffer, and 10 µl of 10^{-3}, 10^{-4} and 10^{-5} dilutions were plated with 300 µl *M. smegmatis* cells. Individual plaques were screened for the presence of the mutation by PCR as described above.

2.3 GENERAL ASSAYS

2.3.1 Genomic phage DNA Isolation

Genomic phage DNA was isolated by phenol-chloroform extraction. An equal volume of buffer equilibrated phenol was added to the processed phage sample. The phenol was mixed with the phage by gentle inversion of the tube. The phenol was separated from the aqueous portion by centrifugation at room temperature for 5 minutes. The aqueous phase is removed and the phenol extraction is repeated between 3 and 5 times until the white interface between layers is gone. Back extraction of additional DNA from the processed phenol was performed with 600 µl of TE. To both extractions, an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added to the DNA in aqueous solution, mixed gently and separated by centrifugation again. For the final step, an equal volume of chloroform was added to the aqueous phase containing the DNA, which was mixed and centrifuged. The aqueous phase is removed and processed further.

To concentrate the DNA, the DNA was ethanol precipitated from the cleaned aqueous solutions and resuspended in 50-100µl of TE. The concentration of DNA was determined via spectrophotroscopy using a NanoDrop (Thermo Scientific).

2.3.2 Ethanol precipitation

One tenth volume of 3M sodium acetate and 3 volumes of 95% ethanol were added to the DNA. The tube was mixed gently by inversion until the DNA formed a visible precipitate. The sample was frozen on dry ice or in the -80°C freezer for between 30 minutes and overnight. The DNA was centrifuged at 13000 rpm at 4°C for 30 minutes to pellet the DNA. The supernatant was

removed and the DNA pellet was washed with 1 ml 70% ethanol. After gentle mixing, the sample was centrifuged at 13000 rpm at room temperature for 1 min. The supernatant was discarded and the pellet was air dried to allow the evaporation of any remaining ethanol. The pellet was resuspended in 50 to 100 µl TE at 42°C. DNA was stored at 4°C.

2.3.3 Miniprep of *E. coli* cultures with Qiagen Biorobot

Cultures of *E. coli* were grown in 1ml LB with the appropriate antibiotic shaking at 37°C for at least 48 hours in 96-well plates with 2ml per well capacity. The 96-well plate was centrifuged at 1000 x *g* for 10 minutes. Then the 96-well plate was placed on the Biorobot (Qiagen) and minipreps were performed according to manufacturer's instruction. The plasmids were eluted with between 50 µl and 120 µl of elution buffer.

2.3.4 Sequencing of plasmid inserts by Sanger sequencing

Some plasmids were sequenced in the University of Pittsburgh Genome Center using an ABI 3730 Sanger sequencing machine. First, a sequencing reaction composed of 3µl of plasmid, 0.5µl of Big Dye, 1x sequencing buffer 200nM LMO51 primer and the necessary volume of HPLC water in a total volume of 10µl was mixed and PCR was conducted under the following cycling conditions: 95°C for 5 minutes one time; 50 cycles of 95°C for 30 seconds, 50°C for 10 seconds, and 60°C for 4 minutes; hold at 4°C. Then the reactions were treated with 1µl of 2% SDS and incubated the thermocycler for 5 minutes at 98°C, 10 minutes at 25°C and hold at 4°C. The reactions were then sequenced with ABI 3730 according to manufacturer's instructions. The sequencing reads generated were analyzed with Consed software [151].

2.3.5 ATP Release Assay

M. smegmatis mc^2155 was grown in 7H9 broth supplemented with ADC, CB, CHX and $CaCl_2$ to an OD_{600nm} of approximately 1.0. The culture was then diluted to OD_{600} of 0.03. Uninfected cultures were left untreated. BPs-infected cultures were treated with BPs phage lysate at an MOI of 10. Following adsorption for 30 min at room temperature, Tween80 was added to a final concentration of 0.05% to the cultures to inhibit the adsorption of additional phage and cells were incubated shaking at 37°C. Aliquots of 100µl were removed from each culture at the recorded time points. Free ATP in the media was measured by adding 100µl of ENLITEN rLuciferase/Luciferin reagent (Promega) to the 100µl of culture, as the only source of ATP; luminescence was measured for a 10 s interval in a Monolight 2010 luminometer.

2.3.6 RNA Isolation from *M. smegmatis*

M. smegmatis bacterial cultures were grown up and any phage infections or treatments were performed. Up to 6mL of culture were collected in 500µl aliquots and each aliquot was treated immediately with 1mL of RNAprotect Bacteria Reagent (Qiagen) to stabilize the RNA. The mixture of culture and RNAprotect was vortexed for 5 sec and incubated at room temperature for 5 min. The mixture was centrifuge at 5000 x *g* for 1 min and the supernatant was removed and the pellet was dried upside down on a KimWipe. The pellet was stored at -20°C or used immediately for RNA isolation.

RNA isolation from *M. smegmatis* cells was performed using a modified protocol from the RNeasy Mini Kit (Qiagen). For each RNA sample, the cell pellets previously collected were resuspended in 700 µl total of RLT buffer with 2-mercaptoethanol and the resuspension of each

sample was added to one tube of Lysing Matrix B (MP Biomedicals) 0.1mm silica beads. The cells were disrupted by shaking for 45 sec at maximum speed on Bead Beater 2 times with 1 min incubation on ice between rounds. 1 volume of 80% ethanol was added to the bacterial lysate and well mixed by shaking the tube and up to 700 µl at a time was added to the RNeasy spin column. Washes were performed with RW1 and RPE buffer, according to the manufacturer's instruction. RNA was eluted with 2 subsequent rounds of elution with 50 µl RNase-free water.

RNA was treated with RNaseOut Recombinant Ribonuclease Inhibitor (Invitrogen) according to manufacturer's instructions and digested with DNase I with the DNA-*free* kit (Ambion). The concentration of the RNA was determined with a NanoDrop.

2.3.7 Fluorescence quantification

Bacterial strains containing plasmids carrying *gfp* or *mCherry* reporter genes were grown in liquid culture. C-terminally tagged GFP vectors were constructed in nitrile inducible vectors [152], electroporated into *M. smegmatis* mc^2155, grown in liquid culture and 1 ml aliquots were induced with a final concentration of 0.5% ε-caprolactam at 37°C for approximately 4 hours. Genomic regions to be tested for promoter activity were PCR amplified, cloned upstream of *mCherry*, and cultures of *M. smegmatis* were grown for 48 hrs at 37°C. Fluorescence of 50 µl aliquots was measured at 473 nm and 532 nm for GFP and mCherry, respectively, on a FLA-5100 (FujiFilm). Reported fluorescence units are (LAU/mm^2)/OD of the culture.

2.3.8 Site-directed mutagenesis

Site-directed mutagenesis (SDM) was carried out according to the instructions of the QuikChange Site-Directed Mutagenesis Kit (Aglient) with some modifications. SDM primers were designed be 25 to 45 bases in length with the desired mutation in the middle of the sequence. Two primers were ordered (IDT DNA); one as designed and the other its complement. The protocol from Aglient states that primers must be PAGE-purified but due to the high cost, primers were prepared only with standard desalting purification. Comparison of SDM using 3 sets of primers ordered with PAGE or standard desalting protocol showed no difference in the ability to isolate the desired mutations (unpublished data, L. Oldfield). The template plasmid was amplified with the primers via a 50 µl PCR reaction using cloned pfu polymerase (Invitrogen). A control with the template plasmid but no primers was also run side-by-side, and checks for effective digestion of the template by DpnI. PCR was carried out under the conditions stated in the protocol. Any mutations that proved difficult to isolate were subjected to PCR for 18 cycles, the cycle number designated for larger amino acid deletions or insertions, no matter the type of mutations being created. The PCR products were digested by addition of 1 µl of DpnI (Promega) directly to the PCR reaction, which was incubated at 37°C for 1 or more hours. One microliter of the digested PCR was immediately transformed into NEB5-alpha chemically competent *E. coli* (New England BioLabs), according to manufacturer's instructions. The transformation was plated on LB agar supplemented with the appropriate antibiotics. The colony numbers between the control and SDM transformations were compared and if favorable, transformants from the SDM were picked, grown in liquid media, and the plasmids are isolated by miniprep (GeneJET Miniprep Kit from Fermentas) and sequenced (GeneWiz).

2.3.9 *E. coli* recombineering of plasmids

Some plasmids were modified by *E. coli* recombineering in strain DY331 using the method outlined in *Current Protocols in Molecular Biology* Support Protocol 5 [153]. Briefly, *E. coli* DY331 cells were grown overnight at 30°C shaking in LB supplemented with tetracycline. A 35 ml culture was inoculated and grown to an OD_{600nm} of between 0.4 and 0.6 at 30°C. These cells were induced for 15 minutes shaking in a 42°C water bath. The culture was incubated on ice for ~5 minutes and then the cells were pelleted and washed 3 times with cold sterile water and resuspended in approximately 500 µl of cold water. To aliquots (50 to 100 µl) of these electrocompetent, induced recombineering cells, ~100 ng of the parental plasmid and ~100 ng of the recombineering substrate were added and co-electroporated at 200 Ω, 25 µF, 2.5 kV. The transformations were recovered in 1 ml of LB at 30°C for 1 to 2 hours. Then 9 ml of LB with the appropriate antibiotic for plasmid selection was added and the transformations were grown overnight at 30°C shaking. The following day the plasmids from the transformation were isolated (GeneJET Miniprep Kit, Fermentas) and transformed at low concentration into NEB5-alpha chemically competent *E. coli* (New England BioLabs). Plasmids from these transformants were isolated by miniprep and sequenced (GeneWiz).

2.3.10 Western blot

To perform isolate whole cell lysates from mycobacterial, the cultures were grown as desired and 2 ml of culture was pelleted and resuspended in 500 µl of lysis buffer. Cells were lysed by 3 rounds of sonication for 20-30 seconds with incubation on ice between rounds. The lysates were boiled in sample buffer and 14 µl of lysate and buffer was loaded into a 10% PAGE gel with a

4.5% stacking top. 10 µl of molecular weight marker ladder (Bio-Rad Precision Plus All Blue) was also loaded. The proteins were separated at 150 V for approximately 1 hour in 1X TGS buffer (Bio-Rad). The proteins from the PAGE gel were transferred onto PVDF in transfer buffer overnight at 4°C at 40 mAmp. The PVDF was washed with 1X TBS, blocked with TBS+5% milk. The PVDF was blotted with primary antibody in TBS+5% BSA. Following this, the blot was washed 3 times with 1X TBS and then blotted with the secondary antibody in TBS+5% milk. The PVDF was washed 3 times with 1X TBS again. The blot was developed using BCIP/NBT Substrate Kit (Invitrogen) according to manufacturer's instructions. Color development took around 10 minutes. Photographs of the blots were in case further color development distorted the results.

2.4 RNA-SEQ OF BPS DURING LYTIC AND LYSOGENIC GROWTH

2.4.1 Removing ribosomal RNA from total isolated RNA

Total RNA was isolated from *M. smegmatis* mc^2155 cells infected with BPs at an MOI of 3 at 30 minutes for early infection and 2 hours for late infection and also from a BPs lysogen of *M. smegmatis* mc^2155 grown to mid-logarithmic phase. Ribosomal RNA (rRNA) was depleted from between 1 and 5 µg total RNA using the RiboZero Magnetic Kit (Epicentre) according to manufacturer's instructions. The rRNA-depleted RNA pellet was resuspended in RNase-free water and stored at -80°C. Samples were qualitatively analyzed using the Agilent 2100 Bioanalyzer.

2.4.2 RNA-Seq with Illumina platform and data analysis with Galaxy

Libraries were prepared according to the True Seq RNA Sample Prep v2 (Illumina) according to manufacturer's instructions. Sequencing was carried out at Tufts Genomic Core Facility (Boston, MA), where the samples were subjected to 50 bp Hi-Seq Illumina sequencing. Analysis of raw FASTQ files was performed on the Galaxy platform (Penn State University). Bowtie was used to map the reads to the mycobacteriophage BPs reference genome. This file was then filtered of unmapped reads (Filter SAM), and converted into a BAM file format (SAM-to-BAM). The BAM file was then processed into "chunks" of 10 million randomly assigned reads because Galaxy was unable to process all of the reads at one time. Pileups with each of the "chunks" of reads were generated against the BPs reference genome. Pileups align the mapped reads with each individual nucleotide of reference genome, which allows the user to graph the number of reads that hit each nucleotide. The "chunk" files of filtered BAM reads were also aligned to the genes and intergenic regions of the *M. smegmatis* mc^2155 genome using the feature coverage tool of Galaxy and, in particular, to the rRNA genes, to examine the level of rRNA depletion.

2.5 REVERSE-TRANSCRIPTASE PCR

2.5.1 BPs infection and RNA isolation

M, smegmatis cells were grown as previously described to an OD_{600nm} of approximately 1.0 for mid-logarithmic growth phase cells. A standard liquid phage infection was carried out with BPs at an MOI of 10. Six milliliters of infected cells were collected in 500µl aliquots at early (32

minutes post-infection) and late (2 hours post-infection) time points. To the cells, 1000 µl of RNAprotect (Qiagen) was added and cells were stored at -20°C according to manufacturer's instructions.

RNA was isolated from infected *M. smegmatis* cells as described for the RNA-Seq analysis using RNeasy (Qiagen). RNA samples were treated with DNase I three times using the DNA-free kit (Ambion). The RNA concentration was determined via Nanodrop to be between 500 ng/µl and 800 ng/µl.

2.5.2 Reverse-transcriptase PCR

Maxima Reverse Transcriptase (Thermo Scientific) was used to converted 800ng of RNA to cDNA according to manufacturer's instructions. The cDNA was used as a template for PCR using Taq DNA polymerase. These reactions were carried out in 50µl with 10ng of cDNA as template. Thermocycling was performed using the standard parameters for Taq and for 25 or 30 cycles with an annealing temperature of 58°C and an extension time of 1 minute. The forward primer is located in the upstream gene and the reverse primer is located in the downstream gene of BPs. To detect the PCR products, 5 µl of the PCR reaction was loaded with Ficoll dye into a 1% agarose gel. Electrophoresis was carried out for 1 hour at 100V and the gel was stained with ethidium bromaide to examine the products produced. All of the primers were tested using genomic BPs DNA and were able to form a product of the correct size (data not shown).

Table 2-1. Primers utilized to determine presence of transcripts at gene junctions.

BPs gene junction	Foward primer	Reverse primer	Product size (bp)
1-2	LMO204	LMO205	357
2-3	LMO206	LMO207	461
3-4	LMO208	LMO209	417
4-5	LMO210	LMO211	394
5-6	LMO212	LMO213	431
6-7	LMO214	LMO215	411
7-8	LMO216	LMO217	386
8-9	LMO218	LMO219	413
9-10	LMO220	LMO221	420
10-11	LMO222	LMO223	362
11-12	LMO224	LMO225	395
12-13	LMO226	LMO227	386
13-14	LMO228	LMO229	429
15-16	LMO230	LMO231	430
16-17	LMO232	LMO233	422
17-18	LMO234	LMO235	427
18-19	LMO236	LMO237	448
19-20	LMO238	LMO239	441
20-21	LMO240	LMO241	446
21-22	LMO242	LMO243	417
22-23	LMO244	LMO245	426
23-24	LMO246	LMO247	360
24-25	LMO248	LMO249	356
25-26	LMO250	LMO251	397
26-27	LMO252	LMO253	426
27-28	LMO254	LMO255	482
28-29	LMO256	LMO257	457
29-30	LMO258	LMO259	396
30-31	LMO260	LMO261	401
31-32	LMO262	LMO263	428
32-33	LMO264	LMO265	419
33-34	LMO266	LMO267	428
34-35	LMO333	LMO334	389
35-36	LMO270	LMO271	410
36-37	LMO272	LMO273	350
37-38	LMO274	LMO275	363
38-39	LMO276	LMO277	307
39-40	LMO335	LMO336	353
40-41	LMO280	LMO281	413
41-42	LMO282	LMO283	324
42-43	LMO284	LMO285	430
43-44	LMO286	LMO287	426
44-45	LMO288	LMO289	398
45-46	LMO290	LMO291	373
46-47	LMO292	LMO293	400
47-48	LMO294	LMO295	404
48-49	LMO296	LMO297	415
49-50	LMO298	LMO299	369
50-51	LMO300	LMO301	468
51-52	LMO302	LMO303	379
52-53	LMO304	LMO305	333
53-54	LMO306	LMO307	396
54-55	LMO308	LMO309	458
55-56	LMO337	LMO338	357
56-57	LMO312	LMO313	354
57-58	LMO314	LMO315	293
58-59	LMO316	LMO317	339
59-60	LMO318	LMO319	297
60-61	LMO320	LMO321	447
61-62	LMO322	LMO323	426
62-63	LMO324	LMO325	412
63-1	LMO326	LMO327	252

2.6 PROMOTER IDENTIFICATION

2.6.1 Fluorescence assays

The vectors constructed for these experiments were electroporated into *M. smegmatis* mc^2155, *M. smegmatis* mc^2155(BPs) or were transformed into NEB 5-alpha chemically competent *E. coli* (New England BioLabs), according to manufacturer's instructions. Transformants were grown under selection on solid media and then in liquid culture with selection, as previously described. *M. tuberculosis* strains were fixed with paraformaldehyde. Fluorescence assays for cells in liquid culture were conducted to detect fluorescence at 532 nm, as previously described. The optical density of the cultures was measured at 595 nm. Fluorescence units were reported as $(LAU)/mm^2)/OD_{595nm}$.

2.6.2 Transcription start site mapping

Total RNA was isolated from *M. smegmatis* mc^2155 strains carrying plasmids with promoter-mCherry transcriptional fusions derived from pLO86 as described previously. Up to 5 μg of total RNA was processed with tobacco acid pyrophosphatase, then 200 ng was circularized with T4 RNA ligase and the maximum amount allowed by volume was reverse transcribed with Maxima Reverse Transcriptase (Thermo Scientific) according to manufacturer's instructions. The circularized cDNA was amplified by PCR over the ligated junction using primers LMO363 and LMO364 in a 50 μl reaction with cloned pfu (Invitrogen). All of the PCR product produced was separated by electrophoresis and slices of the gel were removed based. DNA was extracted using GeneJET Gel Extraction kit (Fermentas) and another round of PCR was performed with primers

LMO363 and LMO364. These PCR products were cleaned with GeneJET PCR Purification kit (Fermentas) and sequenced (Genewiz, Germantown, MD).

2.7 TERMINATOR IDENTIFICATION

2.7.1 Creation of terminator trap vector

pLO32 was manipulated by SDM to insert an ScaI blunt restriction site between the P_6 promoter and the *mCherry* gene. pLO32 was amplified with primer LMO382 and its reverse complement and an SDM procedure was carried out, as described previously. The vector with the ScaI site, now called pLO106, was prepared for ligation by digestion with ScaI, treatment with CIP and was gel purified and extracted using the GeneJET Gel Extraction kit (Fermentas).

2.7.2 Amplification of putative terminators

Putative terminator locations were chosen by screening the intergenic regions of the BPs genome for intrinsic terminators using the program TransTermHP [154]. All genomic locations singled out by the search were cloned into the ScaI site of pLO106. The putative terminators were amplified by standard 50µl PCR reaction with cloned pfu (Invitrogen), as previously described, from BPs genomic DNA or lysate. For primers and plasmid names see Appendix D.1. The amplified products were phosphorylated with T4 PNK (New England Biolabs). Ligations were performed utilizing Fast-Link DNA Ligation Kit (Epicentre) according to manufacturer's instructions. Plasmids were sequenced (GeneWiz) and transformed into *M. smegmatis* mc^2155

electrocompetent cells, as previously described. Also, the *E. coli rrnB* terminator from pLO86 was cloned in pLO106 as a positive control for a known terminator. Fluorescence quantification experiments were carried out, as previously described.

Table 2-1. Plasmids used in BPs terminator identification.

Plasmid	BPs Gene*	BPs coordinate start	BPs coordinate end	Forward primer	Reverse primer
pLO124	3-4	3188	3604	LMO208	LMO209
pLO125	5-6	6311	6703	LMO472	LMO473
pLO126	13-14	10351	10779	LMO228	LMO229
pLO127	16-17	15268	15689	LMO232	LMO233
pLO128	25-26	23422	23818	LMO250	LMO251
pLO129	26-27	23662	24087	LMO252	LMO253
pLO130	30-31	26918	27318	LMO260	LMO261
pLO131	31R	27537	27162	LMO320	LMO435
pLO132	32-33R	29099	28681	LMO314	LMO315
pLO133	32	27946	28274	LMO474	LMO475
pLO134	43-44	34185	34610	LMO423	LMO424
pLO135	45	34749	35127	LMO286	LMO287
pLO136	47	35342	35850	LMO425	LMO426
pLO137	47R	35715	35350	LMO314	LMO315
pLO138	54-55	38435	38855	LMO427	LMO428
pLO139	55-56	38873	39199	LMO429	LMO430
pLO140	56-57	39499	39852	LMO431	LMO432
pLO141	56R	39519	39180	LMO433	LMO434
pLO142	58	39832	40124	LMO310	LMO311
pLO143	60-61	40623	40939	LMO312	LMO313
pLO175	*Eco* rrnB from pLO86	NA	NA	LMO476	LMO477

*Except pLO175 insert from pLO86

2.7.3 Elimination of putative promoters from pLO132 and pLO134

SDM was conducted as previously described to delete 5bp at the 3' cloning junction from pLO132 and pLO134 using primers LMO486 and LMO487 and their reverse complements, respectively. As a control, 5bp at the same location was also deleted from pLO141 though no promoter element was detected and no effect on termination was found (data not shown).

2.8 DISSECTION OF PROMOTER P_R

2.8.1 Creation of extrachromosomal vectors containing P_R-mCherry fusion

The promoterless fluorescence construct, pLO86, was constructed by replacing *gfp* in pJL37 with *mCherry* and placing the TM4 gene *9* ribosome binding site (5973-5998) immediately upstream. Fluorescence expression vectors pLO07 and pLO08 were created by PCR amplifying BPs coordinates 29,224-29,598 and cloning into the Dra I site of pLO86 in the forward and reverse orientations, respectively. Similarly, pLO32 was created by PCR amplifying BPs coordinates 6,285-6,494 and cloning the fragment in the Dra I site of pLO86.

2.8.2 Fluorescence assays in liquid culture for *M. smegmatis, M. tuberculosis* and *E. coli*

The vectors constructed for these experiments were electroporated into *M. smegmatis* mc^2155, *M. smegmatis* mc^2155(BPs) or *M. tuberculosis* mc^27000 as previously described or were

transformed into NEB 5-alpha chemically competent *E. coli* (New England BioLabs), according to manufacturer's instructions. Transformants were grown under selection on solid media and then in liquid culture with selection, as previously described. *M. tuberculosis* strains were fixed with paraformaldehyde. Fluorescence assays for cells in liquid culture were conducted to detect fluorescence at 532 nm, as previously described. The optical density of the cultures was measured at 595 nm. Fluorescence units were reported as (LAU)/mm^2)/OD_{595nm}.

2.8.3 Detection of mCherry fluorescence by colony color

The strains expressing mCherry from a control or P_R derivative promoter were patched onto solid media. *M. smegmatis* $mc^2$155 strains were patched onto 7H10 supplemented with dextrose and $CaCl_2$ and containing CB, CHX and Kan antibiotics and grown at 37°C for approximately 3 days. The *M. tuberculosis* $mc^2$7000 strains were patched on 7H11 supplemented with OADC, Pan and $CaCl_2$ and containing CB, CHX and Kan antibiotics and grown at 37°C for between 1 and 2 weeks. Photographs were taken of the plates and the colony colors were compared between samples on the same plate.

2.8.4 Detection of mCherry fluorescence by fluorescence microscopy

The strains were grown and treated as described for the fluorescence assay in liquid culture. The cells were concentrated and 5 µl was placed on a glass slide, a coverslip placed over top and sealed into place using coverslip sealant. The fluorescence was examined with a fluorescence microscope (Axiostar Plus; Carl Zeiss). Bright field and multiple fluorescence images at varying controlled exposure times (20 ms, 100 ms and 1000 ms) were taken for each field of view using

an AxioCam MRc5 camera (Carl Zeiss) and Carl Zeiss AxioVision Rel. 4.6 software. For the detection of fluorescent cells a HQ:R NX (41002c- HQ545/30X, HQ620/60m, Q570LP) filter from Chroma Technology Corporation was used. The brightness/contrast of the images was not altered.

2.9 CHARACTERIZATION OF BPS GENETIC SWITCH

2.9.1 Plasmid construction

Point mutants pLO07:A-24C, pLO07:T-21C and pLO26 (pLO08 derivative with P_{rep} A-12G mutation) were created by site-directed mutagenesis to create mutations in P_R synonymous with BPs A29486C, T29489C and T29336C, respectively. BPs gp33 expressing plasmids pLO09, pLO10, pLO15 and pLO21 are derivatives of pGWB43, pGWB48, pGWB43GoF1 and pGWB43GoF2, where the Kan resistance cassette was replaced with a Hyg resistance cassette through lambda Red recombineering [153]. Inducible tagged GFP construct pLO16 was created by generating a fusion of the 13 C-terminal residues of BPs gp33 (coordinates 28,916- 28,955) to *gfp* of pNIT:eGFP synthetically through PCR and pLO24 was created by generating a substrate of GFP with the 64 C-terminal residues of BPs gp32 by PCR and cloning the GFP+64 aa gp32 tag into pNIT:GFP at the NdeI and HindIII. Site-directed mutatgenesis of pLO24 generated plasmids pLO27, pLO28, pLO29, pLO56, pLO57, pLO70, pLO71 and pLO72, which have deletions or amino acid substitutions in the 64 aa gp32 tag.pLO64 was created adding the 5 C-terminal residues of Brujita Int (coordinates 28,055- 28,072) to *gfp* of pNIT:eGFP through site-directed mutagenesis. A substitution of the penultimate Ala to Glu of the tags was made by site-

directed mutagenesis in both pLO16 and pLO64 to create pLO20 and pLO65, respectively. BPs gp34 expressing strains were created by blunt cloning BPs coordinates 29,299- 29,755—either the wild-type sequence (pLO59) or with T29489C point mutation (pLO61)—into a Giles integrative vector, pGH1000b, at the Nru I site.

A table of all plasmids used in Chapter 5 can be found in Appendix D.

2.9.2 Induction of nitrile inducible vectors for fluorescence assay and Western blot

The plasmids constructed in nitrile inducible vectors were electroporated into *M. smegmatis* mc^2155 and a BPs lysogen, *M. smegmatis* mc^2155(BPs). Liquid cultures were grown to mid-logarithmic phases and 1 ml aliquots were induced with a final concentration of 0.5% ε-caprolactam at 37°C for approximately 4 hours. Fluorescence of 50 µl aliquots was quanitified as described at 437 nm. The OD measurements were taken at 600 nm or 595 nm depending on the experiment. Figure legends note which wavelength and method was used. Fluorescence was reported in fluorescence units (LAU/mm^2)/OD.

Western blots were conducted as previously described with anti-GFP rabbit polyclonal IgG fraction (Invitrogen A-11122) primary antibody at a 1 to 1000 dilution and goat anti-rabbit secondary antibody at a 1 to 5000 dilution.

2.9.3 Fluorescence of promoter-reporter fusions

Fluorescence assays of mCherry activity were conducted as previously described at 533 nm (see Chapter 2.8.2). Fluorescence was reported in fluorescence units (LAU/mm^2)/OD_{595nm}.

3.0 TRANSCRIPTOME OF MYCOBACTERIOPHAGE BPS

3.1 INTRODUCTION

The patterns of gene expression and their temporal regulation have been studied in mycobacteriophages by examining the changes in the overall profile of protein expression over time [3-5]. RNA sequencing (RNA-Seq) is a considerably more detailed approach, which can determine the identity of the genes being expressed at a given moment, the exact boundaries of the transcript they are being expressed from and the quantity of the transcript present at the resolution of the nucleotide. Though an RNA-Seq profile of expression can generate information about possible transcription start sites, terminators and overall operon structure, further experiments are necessary to confirm details of the transcriptional organization and regulation.

The expression of proteins in mycobacteriophages is temporally regulated. L5, Bxb1 and TM4 have an early phase of protein expression, which begins shortly after lytic growth is induced or phage are adsorbed, and a late phase of expression, which occurs after early expression, though the two phases may overlap [4,5]. Though the identity of the proteins expressed in these phases is not known, it is hypothesized that early proteins are involved in DNA replication and recombination of the phage and host-phage interactions and that late proteins are involved in virion structure and assembly and lysis of the host cell.

The gene expression of mycobacteriophage Giles has been examined by RNA-Seq and two phases of gene expression were detected [8]. The early expression includes genes of unknown function in the right arm of phage Giles and late expression retains the expression of these early genes and includes very high expression of the structure and assembly genes in the left arm of Giles [8]. The RNA-Seq examination of Giles revealed likely transcription start sites but no SigA-like mycobacterial promoters could be identified immediately upstream [8].

To study gene expression in mycobacteriophages in further depth, BPs was used as a model organism. The cluster G mycobacteriophages, which BPs is a member of, have the shortest genome length and lowest number of genes of any mycobacteriophages that have been sequenced to date (phagesdb.org). The organization of the genome of BPs is very simple (see Chapter 1, Fig. 1-1). The structure and assembly genes, followed by the lysin genes constitute the left arm of the phage from genes *1* to *31* and are transcribed rightwards. The only genes in BPs transcribed leftwards are *int* (*32*) and *rep* (*33*), which—with gene *34*—constitute the immunity cassette. The right arm of BPs, genes *34* to *63*, is transcribed rightwards and is made up of many small ORFs with unknown functions. The right arm does contain three genes with predicted recombination functions, and therefore it is hypothesized that the right arm of BPs is expressed in early lytic growth and is involved in replication and recombination in the bacteriophage. The structure and assembly genes in the left arm of the phage are predicted to be expressed during late lytic growth.

3.2 TIMING OF BPS INFECTION CYCLE

The timing of cell lysis varies depending on the specific mycobacteriophage. This latent period of a bacteriophage is a single infection cycle, including adsorption and injection of nucleic acids, replication and assembly of progeny phage, and host cell lysis. The latent period of BPs is hypothesized to be approximately the same as the doubling time of *M. smegmatis*, approximately three hours. Determining the amount of time BPs requires to lyse *M. smegmatis* cells is important for the examination of more detailed patterns of gene expression and proper assignment of temporal designations, such as early and late, for BPs lytic growth.

One indicator of lysed bacterial cells is the amount of free ATP present in the culture [155], which is released when the cells lyse or the membrane becomes permeabilized [30]. Free ATP can be measured indirectly through a luciferase assay, where the limiting reagent in the assay is ATP and is provided only by the bacterial culture and therefore a function of the number of lysed cells. *M. smegmatis* mc^2155 cells were infected with BPs phage and an aliquot of the culture was added to a reaction that contained luciferase and substrate, but no ATP. The output of light, measured with a luminometer, is therefore determined by the amount of free ATP in the culture.

Using a BPs-infected culture, luciferase activity (relative light units; RLU) increases dramatically between 150 and 200 minutes post-adsorption and then the activity plateaus after 200 minutes (Fig. 3-1). In an uninfected control, very little change in the luciferase activity is observed. The midpoint of the lysis of *M. smegmatis* infected with BPs is three hours post-infection, confirming the hypothesis that time required for BPs phage replication and lysis is similar to the doubling time of *M. smegmatis*. The latent period of mycobacteriophage Giles was previously examined by tracking the optical density (OD) of cultures infected with Giles and

also, by the ATP release assay; Giles lyses *M. smegmatis* beginning at three hours post-infection and proceeds until five hours post-infection [30]. The agreement between the OD and ATP release assay helps to demonstrate that the ATP release assay is reliable. The latent period for BPs is notably shorter than that of Giles in *M. smegmatis*. Further work is necessary to determine the range of the latent periods found in mycobacteriophages.

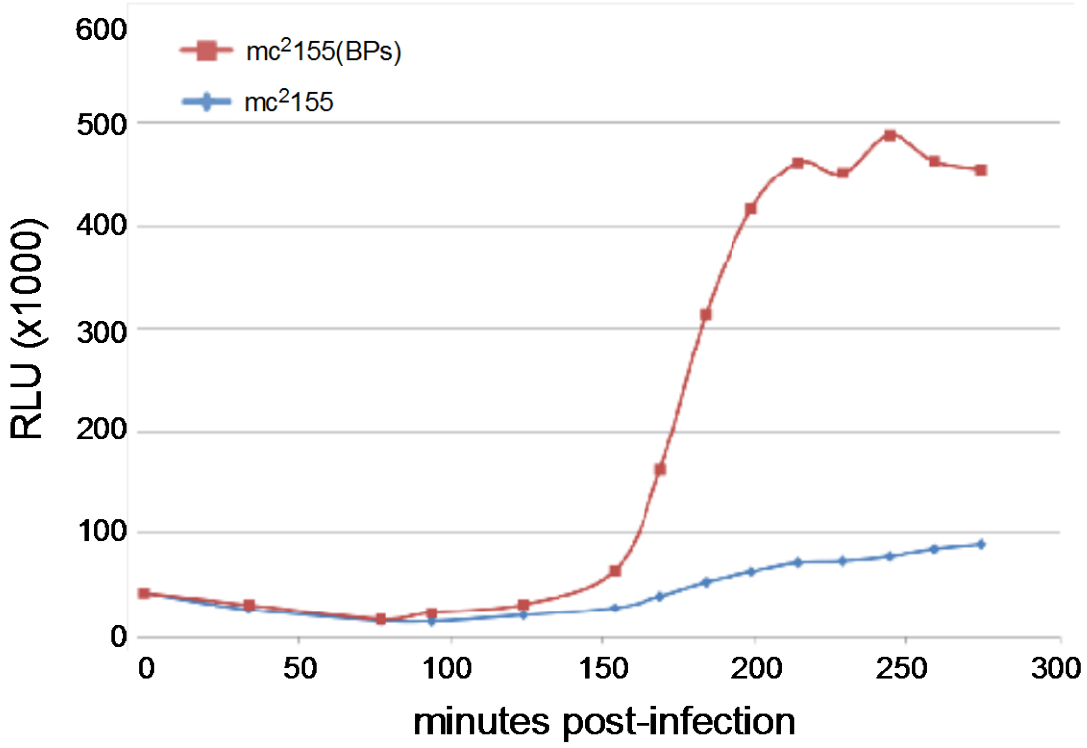

Figure 3-1. Lysis of *M. smegmatis* by mycobacteriophage BPs.

Fig. 3-1. The time required to lyse host cells with BPs (red) was measured in a luciferase assay. Luciferase was measure in relative light units (RLU) over intervals spanning of 0 to 270 minutes post-infection. Uninfected cells were used a control (blue).

3.3 MAPPING THE BPS TRANSCRIPTOME DURING LYTIC AND LYSOGENIC GROWTH

To map the transcriptome, a BPs lysogen of *M. smegmatis* mc^{2}155 and BPs-infected cells were subjected to RNA-Seq. RNA-Seq is a method in which RNA is isolated from growing cells, processed to enrich for messenger RNA (mRNA) by depleting ribosomal RNA (rRNA), converted to cDNA and sequenced by high-throughput sequencing, such as Illumina. The sequencing reads obtained are then aligned to the reference genome of the organism being investigated. The data reveal which genes are transcribed, generates hypotheses of the locations of the 3' and 5' ends of the transcripts and establishes the relative abundances of the transcripts at a single nucleotide resolution. Overall, RNA-Seq provides a detailed snapshot at the time of sample collection for the transcriptome of the organism.

3.3.1 Selection of time points to be investigated and hypotheses

A complete picture of BPs gene expression includes an examination of transcript levels during both lytic and lysogenic growth. As a temperate phage, BPs can grow in either state. Similar to many other well-studied bacteriophages, BPs is hypothesized to temporally regulate gene expression during lytic growth, expressing genes required for phage replication first and later expressing the genes for the assembly of progeny phage and for cell lysis. As previously stated, lysis of *M. smegmatis* mc^{2}155 after infection with BPs requires approximately 3 hours, which indicates that a single infection cycle is completed (Fig. 3-1). ^{35}S-methionine labeling of newly synthesized proteins during lytic growth indicated a putative change in the profile of proteins at 45 minutes post-adsorption, though the results were not conclusive (see Appendix A). This may

be the shift between early and late protein expression and the time points chosen for examination with RNA-Seq during lytic growth are before and after this putative shift in expression.

To capture different stages of temporal regulation during lytic growth, an early (30 min) and a late (2 hours) post-adsorption time point were selected for RNA-Seq. *M. smegmatis* mc^2155 cells were infected with BPs at a multiplicity of infection (MOI) of 3 and total RNA was isolated at 30 minutes and 2 hours post-adsorption. Many additional times post-infection could also be examined to determine more precisely the succession of gene expression.

Throughout the entire cycle of lytic growth, most BPs genes are expected to be expressed by the phage during lytic infection. The early genes are expected to be those located in the right arm of the BPs genome (genes *34* to *63*) and transcripts from this region will be present in lower abundances than the late structure and assembly genes. Though most of these small open reading frames (ORFs) have no predicted function, RecE (*42*), RecT (*43*) and RuvC (*51*) homologues have been identified [10]; the location of these recombination genes leads to the prediction that other genes required for phage DNA replication and recombination, which would likely be expressed early in lytic growth, are located similarly. The transcripts generated late in infection are hypothesized to be those in the left arm of the genome (genes *1* to *31*), where the predicted structural genes are located. These genes are likely transcribed at particularly high levels, since the structural components are required in high copy number to produce progeny phage. Phage growing lytically are not expected to produce significant amounts of repressor (gene *33*) or integrase (gene *32*) mRNA, but transcripts may be detected at low levels due to the creation of lysogens at a low frequency.

Gene expression patterns during lysogenic growth of BPs are expected to be very different from those observed during lytic growth. During lysogenic growth, we hypothesize that

the only genes expressed are hypothesized to be the integrase (*32*), responsible for carrying out site-specific recombination between the BPs *attP* and *M. smegmatis attB* sequences, resulting in the integration of the phage genome into the host chromosome, and the repressor (*33*), which is responsible for the transcriptional repression of lytic genes in the phage and preventing the superinfection of homoimmune phages. Both functions are required for stable integration and maintenance of the prophage and the identities of genes *int* and *rep*, *32* and *33*, respectively, has been confirmed [10,31]. Total RNA was isolated from a culture of a stably replicating BPs lysogen of *M. smegmatis* mc^2155 and the profile of transcription was examined.

3.3.2 Sequencing for RNA-Seq

The sequencing of cDNA libraries from uninfected *M. smegmatis*, an *M. smegmatis* lysogen of BPs, and *M. smegmatis* cells infected with BPs at early and late time points produced between 175 and 201 million reads per sample. The raw sequencing reads were processed and mapped to the BPs and *M. smegmatis* mc^2155 reference genomes in Galaxy (Pennsylvania State University; http://galaxyproject.org/). Approximately 95% of the reads mapped to either the host or phage genome (Table 3-1). RNA-Seq sample preparation and sequencing reported here were carried out on two separate occasions. The early and late lytic samples were prepared from the same BPs-infected culture of *M. smegmatis* mc^2155 and were sequenced in 2012. The wild-type *M. smegmatis* mc^2155 control (no BPs phage) and the BPs lysogen of *M. smegmatis* were grown side by side and the samples were prepared and sequenced in 2013. This is important for comparisons between the quality of the samples and of the sequencing between the 2012 and 2013 runs.

Sequencing performed in 2012 produced slightly more total reads than the 2013 sequencing but both sequencing runs were within a comparable range. The 2013 experiments produced better quality sequencing reads as a lower percentage of the reads were unmapped. These differences illustrate variation in the technical quality of the experiment and demonstrate that, if possible, samples to be directly compared should be prepared and sequenced concurrently.

An important step in the sample preparation for RNA-Seq is the removal of ribosomal RNA (rRNA) from the total isolated RNA. If rRNA is not efficiently depleted, up to 97% of the RNA isolated and sequenced is rRNA and tRNA [156]. The removal of rRNA relies upon high quality, intact total RNA and an effective method of separating rRNA from other cellular RNAs. To assess the quality of the total RNA isolated, a small amount was evaluated with a 2100 Bioanalyzer (Agilent). The RNA integrity number (RIN), which is based on the ratio of 16S and 28S rRNA, is an indicator of the quality of the RNA and is scored between 1 and 10 [157]. A very high quality sample of total RNA will score between 8 and 10. The 2012 early and late lytic samples had RIN scores of 5.9 and 6.5, respectively. The 2013 wild-type *M. smegmatis* and BPs lysogen samples had RIN scores of 8.3 and 9.1, respectively. The rRNA was depleted from the samples using the RiboZero rRNA Removal Kit (Epicentre) and resulted in very effective depletion of intact rRNA, determined by a lack of rRNA peaks detected by a second round on the Bioanalyzer. For the four RNA-Seq samples, an average of only 5% of the total number of reads mapped to rRNA sequences in the *M. smegmatis* mc^2155 genome (Table 3-1). Using other methods of rRNA depletion, sequencing data with rRNA proportions around 40% of total mapped reads were found (B. Dedrick, unpublished results), which decreases the sensitivity to detect transcripts in other parts of the host and in the bacteriophage.

Table 3-1. Overall count of RNA-Seq reads.

Run	Total reads	Mapped to BPs	Mapped to *Msm*	Mapped to *Msm* rRNA	Unmapped
2012 Early	201,648,367	360,000	190,000,000	5,647,475	11,288,367
2012 Late	199,837,982	6,500,000	180,000,000	6,667,774	13,337,982
2013 Lysogen	191,634,887	123,169	184,115,736	8,314,042	7,395,982
2013 *Msm*	176,556,511	408	170,255,494	7,851,575	6,300,609
average	192,419,437	1,745,894	181,092,808	7,120,217	9,580,735

Run	Percent mapped (%)	Percent mapped to BPs (%)	Percent mapped to *Msm* (%)	Percent mapped to *Msm* rRNA (%)	Percent unmapped (%)
2012 Early	94.40	0.18	94.22	2.80	5.60
2012 Late	93.33	3.25	90.07	3.34	6.67
2013 Lysogen	96.14	0.06	96.08	4.34	3.86
2013 *Msm*	96.43	0.00	96.43	4.45	3.57
average	95.07	0.87	94.20	3.73	4.93

In the BPs lytic and lysogenic samples, a low proportion of the total reads (between 0.06% and 3.25%) of the total reads align to the BPs genome, while over 90% of the reads map to the host *M. smegmatis* genome (Table 3-1). However, the host bacterial genome is nearly 7 million base pairs in length, while the BPs genome is only 41,901 bp, which is around 0.5% the length of the host chromosome. Comparing the ratio of the number of sequencing reads mapped to the genomic length of the phage and host during late lytic growth, the amount of BPs transcription in late lytic growth is above the average amount of transcription in *M. smegmatis*, which indicates that there is a greater amount of transcription occurring at the phage genome than in the host chromosome, on average. In early lytic and lysogenic growth, the transcription from BPs is on average below that of *M. smegmatis*.

3.3.3 Genome-wide patterns of BPs gene expression in lytic and lysogenic growth

A detailed analysis of the genome-wide transcription patterns present in BPs-infected cells and in a BPs prophage was conducted. On the Galaxy platform, a pileup was generated from the reads mapped to BPs. The number of times all of the sequencing reads overlay a particular nucleotide of the phage genome is calculated and then these "hits" can be plotted along the BPs genome to create a visual representation of the RNA-Seq results. The pileups were plotted as hits per nucleotide (hits/nt) for the early lytic, late lytic and lysogenic BPs RNA-Seq experiments (Fig. 3-2).

As the statistics for mapped reads indicated, the late lytic sample was more transcriptionally active than the others (green line; Fig. 3-2). The maximally expressed genes are the structural genes beginning with the scaffold (gene *6*) and have, at most, over 60,000 hits/nt. This was well above the cut off for sequencing coverage of 8,000 reads for a particular location arbitrarily set by the Galaxy and mpileup software, which had to be adjusted to observe peak expression in the late lytic RNA-Seq data. Also, almost all of the genes of BPs are transcribed. Low levels of transcription in the immunity cassette (*32*-*33*) and around gene *56* in the right arm.

In the early BPs-infected RNA-Seq, a lesser number of genes were transcribed and at lower levels of expression (Fig. 3-2). The maximum number of hits per nucleotide in this sample was approximately 4000 hits/nt. The genes transcribed were concentrated mostly in the right arm of the genome, especially from *34* to *44*.

In the lysogen, very few genes are expressed. Notably, there is increased expression of the immunity cassette (genes *32* and *33*) in the lysogen, above the amount found in either of the lytic samples (Fig. 3-2). Another area of high expression is the mobile element, gene *58*, an ultra small mobile element [10].

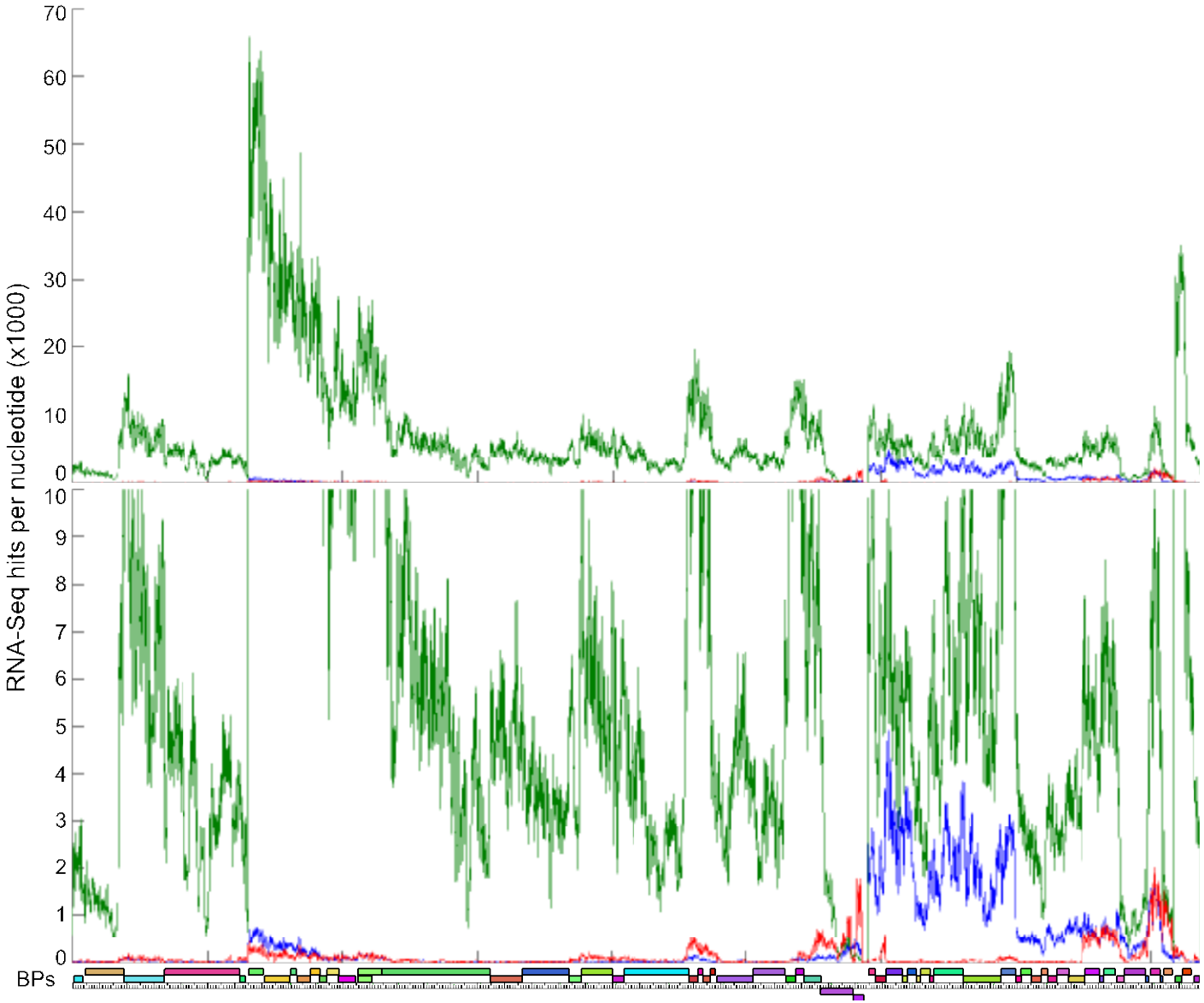

Figure 3-2. RNA-Seq profile of lytic and lysogenic BPs.

Fig. 3-2: The RNA-Seq reads were aligned against the BPs reference genome and recorded as hits per nucleotide. The top panel shows the maximal level of expression with over 60,000 hits/nt. The bottom panel zooms in to 10,000 hits/nt to show lower levels of expression. The BPs genome is aligned along the bottom of the graphs with boxes noting annotated genes. Green is late lytic infection. Blue is early lytic infection. Red is a BPs lysogen.

The genes for phage head and tail structure and assembly are located in the left arm of BPs. Gene *6* is the scaffold protein, which is involved in the assembly of capsid proteins, is the location of the strongest region of gene expression, which occurs during late lytic growth (Fig. 3-3A). Very low amounts of expression can be distinctly observed in this region during early lytic growth and in the lysogen (Fig. 3-3A). The prediction for the cause of expression of these genes, which were predicted to occur only late in infection, is that the promoter located upstream of gene *6* has very high activity and is not completely repressible. Expression from the immunity cassette, genes *32* and *33* is lowest during late lytic growth and highest during lysogenic growth (Fig. 3-3B). In the middle of gene *33*, a sharp decrease in the level of transcripts is observed, which corresponds to the phage *attP* core (Fig. 3-3B). Integration into the host chromosome occurs at this location in the prophage, thus the arrangement of the genome is different in a BPs lysogen and the 5' end of gene *33* is not contiguous with the 3' end and also explains markedly different levels of expression between genes *33* and *32* in the lysogenic sample. Outside of the immunity cassette, the most notable region with low levels of gene expression is within gene *56*. The function of gene *56* is not known, and we hypothesize that a strong terminator is present near the beginning of gene *56*, where transcript levels fall.

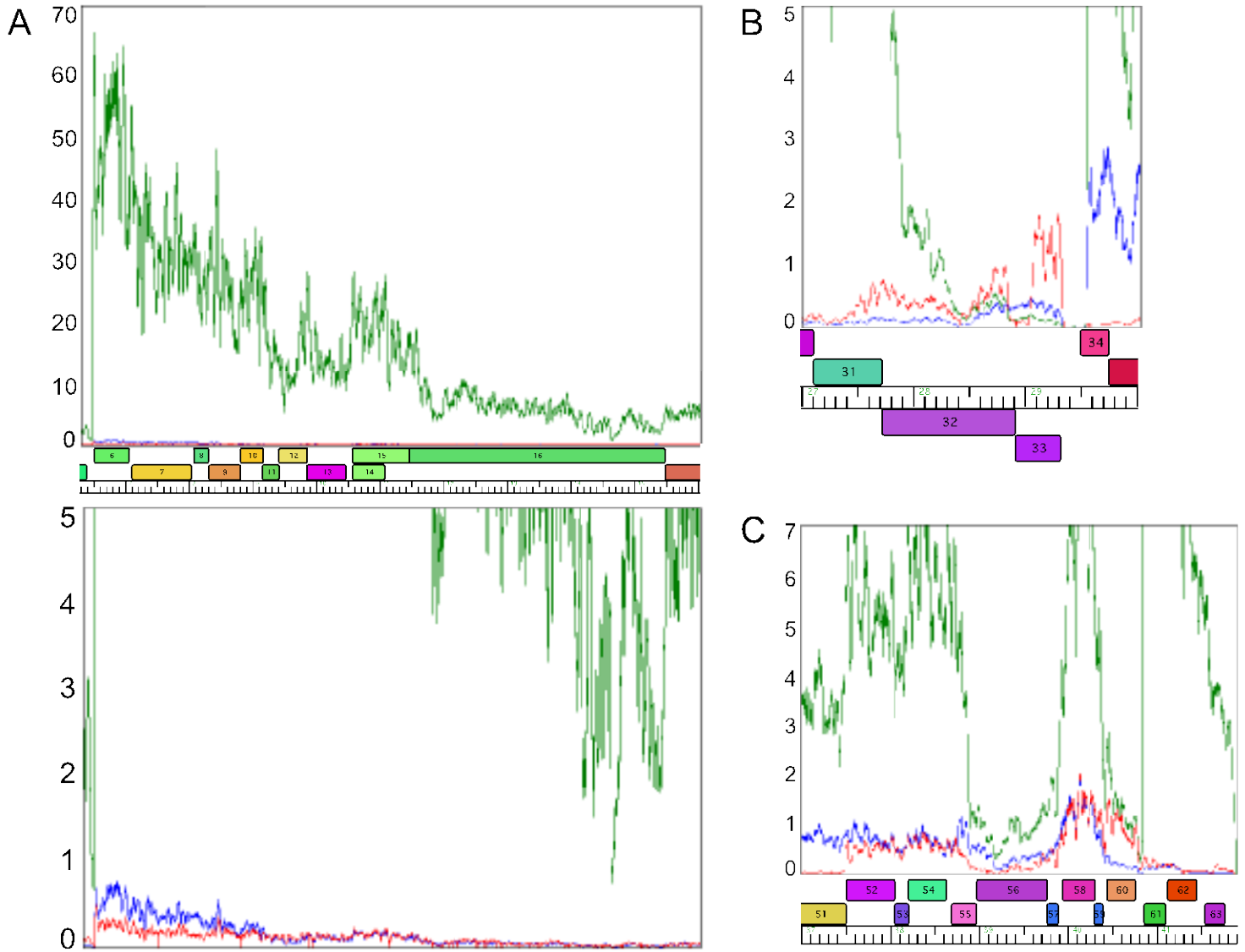

Figure 3-3. RNA-Seq profiles at specific loci.

Fig. 3-3. A. The RNA-Seq profile for genes *6-16* with a zoomed out view showing expression up to 70,000 hits/nt (top) and a zoomed in view showing 5,000 hits/nt (bottom). (A) BPs genome map is provided in the middle to show locations of each gene. Green is late lytic infection. Blue is early lytic infection. Red is a BPs lysogen. (B) The RNA-Seq profile for genes *33-32*, the BPs immunity cassette. The graph displays values up to 5,000 hits/nt with the BPs genome map of the region located below for context. Green is late lytic infection. Blue is early lytic infection. Red is a BPs lysogen. (C) The RNA-Seq profile of genes *51-63*. The graph displays values up to 7,000 hits/nt with the BPs genome map of the region located below for context. Green is late lytic infection. Blue is early lytic infection. Red is a BPs lysogen.

3.3.4 Conclusions from RNA-Seq profiles

When the RNA-Seq reads were aligned to the BPs genome and graphed along the coordinates of BPs (hits per nucleotide), the transcriptome of BPs during lytic infection and lysogeny was represented (Fig. 3-2). The late lytic sample gave the highest levels of BPs expression. Nearly all genes are expressed in the late lytic sample. A much smaller proportion of the genes are expressed early in infection.

A small number of genes are expressed at very low levels, including the immunity cassette, genes *32* and *33*, which encode the integrase and repressor, respectively (Fig. 3-3B). These genes are expressed at higher levels in the lysogenic sample compared to the lytic samples. The small amount of expression of *32* and *33* during lytic growth is likely due to a small percentage of infecting phage that undergo lysogenic growth instead of lytic growth. In BPs, the frequency of lysogeny is about 5% [10,31]. The lysogen, however, expresses *rep* (*33*) and Rep protein shuts down expression of lytic genes [31].

3.3.5 Operon predictions from RNA-Seq

The RNA-Seq data give clues to the locations of BPs operons. Expression of the immunity cassette as a single operon is expected because genes *33* and *32* are the only two genes facing left

and no promoter can be predicted between the two for separate expression of gene *32*. However, these genes cannot be expressed as an operon in the lysogen because integration at the *attP* site within *33*, which physically separates these genes in the prophage. The clearest operon prediction from the RNA-Seq profiles is that genes *34* to *44* are expressed as an operon (Fig. 3-2). These genes are expressed both in early and late infection and the transcription has apparent defined start and end points. Also, at the far right arm of the phage genome, genes *61* to *63* are clearly expressed at a higher level that the preceding genes during late infection, which indicates that these genes may be an operon. The late RNA-Seq data show a sharp increase in expression near the beginning of gene *3* and expression continues through gene *5*. There is clear evidence for a transcription start site in the intergenic region between genes *5* and *6*, which precedes the most highly expressed region of BPs. Operons consisting of genes *3-5* and *6-16* are consistent with these data. The rest of the BPs genome is expressed at moderate levels late in transcription and sharp changes in transcription are not present. In summary, with evidence from the RNA-Seq, operons are predicted for genes *3-5*, *6-16*, *32-33*, *34-44* and *61-63*.

3.4 DETERMINING OPERON STRUCTURE OF BPS

Elucidating the operon structure of a bacterium or bacteriophage can provide insight into the gene functions and location of transcriptional signals present. The BPs genome has very few intergenic regions and the locations of transcription initiation and termination are not straightforwardly discernible. The genomic structure requires the presence of a minimum of only two operons to express all genes annotated. One operon is needed for the leftwards-transcribed

immunity cassette and another for all of the rightwards genes, which can be expressed together in circularly replicating lytic BPs.

To determine the transcriptional units of BPs, the transcriptional read through between genes was examined by reverse transcriptase PCR (RT-PCR) over gene junctions (Fig. 3-4). RNA was isolated from *M. smegmatis* infected with BPs, reverse transcribed and PCR was performed with unique primers for each junction of two genes in BPs. To achieve a semi-quantifiable result, the number of PCR cycles was varied. When a product is produced over the junction, it indicates that transcription continues from one gene to the next and the genes are transcribed in the same transcriptional unit. The intensity of the signal observed from the RT-PCR product was also used to analyze the possible quantity of transcription at the location and likelihood that an operon was present.

Junctions that appear at similar intensities and under the same experimental conditions are genes *1*-5, *6-8, 9-16, 34-38, 38-41, 42-45*, *50-60,* and *61-63* (Fig. 3-4). The RT-PCR results display a similar pattern of expression for the early and the late samples. For example, a product is produced at similar intensities for the junctions of genes *6-7* and *7-8* and no product is at *8-9*, indicating that genes *6-8* are transcribed as an operon. This pattern is observed in both the early and late samples.

This is true for the right arm of the phage (gene *34-63*), although one noticeable difference is that junctions *61-62*, *62-63*, and *63-1* appear to be more highly expressed at late infection, which is supported by RNA-Seq gene expression (Fig. 3-2). The assembly genes at the extreme left end of the genome (genes *1-2*) are expressed in an operon with genes *61* to *63*.

Differences in the intensity of the bands in the left arm of the genome (genes *1-31*) are apparent, though the patterns are very similar between early and late. Only faint bands at genes

6-8 and *9-16* can be seen in the early-infected sample when a lesser number of cycles was used for amplification. RT-PCR for early lytic sample with a higher number of cycles reveals that there are lower levels of expression detectable in all of the other genes from *1-23* (with the exception of the gene *8-9* and *19-20* junctions). Additional bands at some of the junctions from genes *23-33* are also present. In the late sample, there is an increase in the intensity of the bands but a similar pattern remains. The most intense bands are located between the junctions from *6* to *16*, as with the early sample. Lower levels of transcription are found at other locations between genes *1* to *24*. A few bands (*5-6* and *16-17*) are absent from the late infected RT-PCR reactions with the lower cycle number, which does not fit the pattern established in early infection. The products at these locations do appear to give weak signals, which could indicate lesser expression or differences in primer efficiencies. In the region from genes *23* to *33*, similar bands are present as with the early sample. There is no obvious operon structure in this region, which may be the result of a lower amount of expression or due to expression of only one or two genes from weak promoters.

A number of gene junctions did not yield a PCR product in our studies, despite being able to amplify genomic BPs DNA. Products were not detected for junctions *8-9*, *19-20*, *24-25* and *45-48* (all three junctions). Further explanation is provided below.

The overall RT-PCR results indicate that genes *1-5*, *6-8*, *9-16*, *34-38*, *39-41*, *42-45*, *50-60*, and *61-63* are operons (see Fig. 3-12).

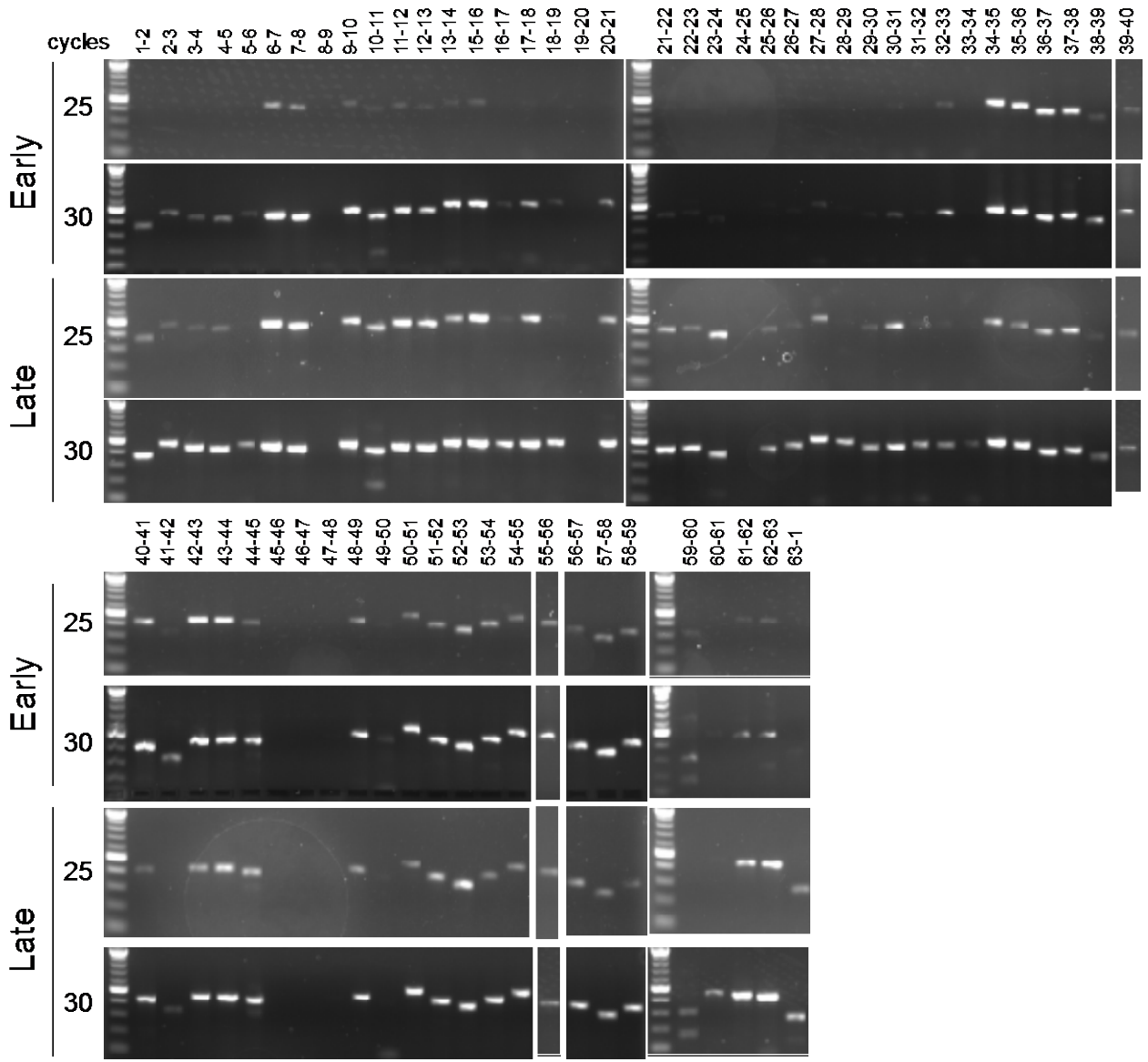

Figure 3-4. Reverse-transcriptase PCR of gene junctions.

Fig. 3-4. RT-PCR to define transcriptional units in BPS was performed with primers that overlapped the junction between genes as indicated above the photos for 25 or 30 cycles to give semi-quantitative results. The PCRs were performed on RNA isolated from early (~30 minutes post-adsorption) and late (~2 hours post-adsorption) lytic BPs samples.

3.4.1 Comparison of reverse-transcriptase PCR and RNA-Seq results

The results of the RT-PCR generally confirm the results of the RNA-Seq of early and late infected samples. The discrepancies between the early and late RT-PCR samples can be reasoned with the help of the RNA-Seq profiles.

The RNA-Seq data do not suggest that there are strong transcriptional terminators at the locations where RT-PCR products are never produced (*8-9, 19-20, 24-25* and *45-48*) as there is expression at all of these locations in the late infection sample. However, the lack of products in the RT-PCR suggests that the target is not present. This may be due to RNA modifications that prevent primer recognition. It is unlikely that the reverse transcription from RNA to cDNA could be the reason that product is not available because this would also affect the RNA-Seq samples and profiles.

In the late lytic infection RNA-Seq data, there is evidence of an operon from gene *6* to *16*, which agrees with the RT-PCR results, except for the lack of a product between genes *8* and *9*. The expression analysis indicates that this gap is not a real consequence of changes in expression. Therefore, combining the evidence from both experiments together, the operon is most likely composed of genes *6* to *16* (see Fig. 3-3A).

Combining the data from both the RNA-Seq and RT-PCR experiments, evidence for operons from genes *61-2, 3-4, 5-16, 33-32,* and *34-44* (possibly actually split into two operons between genes *38* and *39*) is supported. There is evidence to support additional operons but it is not as conclusive, possibly due to the lower levels of expression in some of these regions.

The expression of these genes together as an operon may indicate similarity in function or coordinated function. Many of the structural genes, necessary to make and assemble phage particles, are expressed together supporting this conclusion. Genes with a predicted function in operon *61-2* are gene *2*, a terminase, and gene *62*, a predicted nuclease [10], which would be predicted to be involved at separate times during infection, calling into question if genes *1-2* are expressed with genes *61-63*. Gene *3* is the portal protein for BPs [10] indicating that gene *4* may also be involved in portal formation. Genes *5* to *16* contain a number of structure and assembly genes including the scaffold (*6*), capsid (*7*), major tail subunit (*13*) and tape measure protein (*16*) [10]. The functions of genes *33*, the repressor, and *32*, the integrase, are well established [31]. The operon of genes *33* to *44*, which is present in both early and late infection, is likely involved in the beginning stages of lytic growth such as phage replication and DNA recombination. Gene *34* has a putative Cro-like function [31] and genes *43* and *44* are RecE and RecT homologs, respectively.

Gene functions will need to be more carefully examined in further studies to substantiate these associations, but evidence of the operon structure of BPs may help to provide a starting point for future functional studies.

3.5 MAPPING PROMOTER SEQUENCES IN BPS

Promoter sequences in a small number of mycobacteriophages have been determined [35,37,40-42,158]. The sequences of these promoters resemble SigA promoters in mycobacteria, with a consensus sequence of 5'-TTGACN and 5'-TANNNT for the -10 and -35 hexamers (see Chapter 1). This remarkable similarity to SigA mycobacterial and Sig70 *E. coli* promoter consensus

sequences may be partly due to screens for promoter sequences that function in *E. coli* [40,158]. A bias for sequences that are easily detected by bioinformatics is also a concern, however, a number of the promoters were determined by promoter mapping of transcripts without narrowing down putative promoter regions beforehand [35,37,41].

We wished to determine the promoters responsible for gene expression in mycobacteriophage BPs. Determining the signals responsible for transcription initiation and the regulation of this process is key to understanding how and why BPs transcriptional patterns determined by RNA-Seq come about and will give further clues about the operon structure of BPs.

3.5.1 Determining regions of the BPs genome to test for promoter activity

We took a number of approaches to determining regions of the BPs genome that showed evidence of being locations for promoters. The overall BPs genomic arrangement, promoter prediction software, a promoter trap experiment to identify fragments of BPs genomic DNA with promoter activity, and the putative locations of transcription start sites from the RNA-Seq transcriptomic profiles were all used to determine regions of interest. In all, we constructed and examined 33 locations for promoter activity by creating promoter-reporter transcriptional fusions (Table 3-2 and Fig. 3-5).

The regions of DNA were cloned upstream of a codon-optimized mCherry fluorescence gene, which is translated from a ribosome binding site from the capsid protein of TM4. When expressed under the strong mycobacterial hsp60 promoter [159], colonies in *M. smegmatis* and *M. tuberculosis* have a vibrant purple color. For weaker promoters, a pink color is present and can be discerned visually from a strain with a promoterless mCherry construct, which is white or

yellowish for *M. smegmatis*. The BPs promoter-reporter constructs were transformed into *M. smegmatis* mc^2155 and examined for mCherry fluorescence visually, looking for pink or purple color in the colonies. The fluorescence of most of the constructs were also grown in liquid culture and assayed quantitatively for fluorescence. Promoters active in wild-type *M. smegmatis* were also transformed into a strain of *M. smegmatis* mc^2155 that is lysogenic for BPs [denoted *M. smegmatis* mc^2155(BPs)], to determine if promoter activity is regulated by the BPs repressor or other prophage proteins.

Table 3-2. Promoter-reporter vectors constructed.

Plasmid	BPs location	Insert BPs coordinates (bp) start	Insert BPs coordinates (bp) end	Orientation of insert
pLO07	33-34	29224	29598	F
pLO08	33-34	29224	29598	R
pLO11	32-33	28448	29108	F
pLO12	32-33	28448	29108	R
pLO30	63-1	41801	250	R
pLO31	middle 2	1541	1740	F
pLO32	5-6	6285	6494	F
pLO33	end 57	39720	39919	F
pLO34	60-61	40653	40848	F
pLO46	54-55	38400	39000	F
pLO55	26-27	23663	24084	F
pLO108	63-1	41777	296	F
pLO109	middle 2 middle	1445	1818	F
pLO110	16	11471	12260	F
pLO111	end 17	16098	16575	F
pLO112	18-19	18231	18581	F
pLO113	19-20	18625	19108	F
pLO114	22-23 middle	22215	23005	F
pLO115	27	24231	24968	F
pLO116	end 40 middle	31516	31827	F
pLO117	42	31997	32469	F
pLO118	43-44	33959	34440	F
pLO119	47-49	35832	36303	F
pLO120	51-52	37192	37705	F
pLO121	57-58	39721	40099	F
pLO122	28-29 middle	25837	26625	F
pLO193	54	38254	38625	F
pLO202	28-27	24895	25312	R
pLO206	56-58	39343	39973	F
pLO209	29-31	26767	27225	F
pLO216	63-end	41588	41901	F
pLO219	49-50	36364	36792	F
pLO221	51-52	37315	37835	F
pLO223	36-38	30532	31032	F

3.5.1.1 The structure of the BPs genome can be used to predict promoter locations

The arrangement of the open reading frames in BPs, with a structural operon of genes in the left arm of the genome, the immunity cassette with two leftwards-facing genes, and an operon of early lytic genes in the right arm of the phage, indicates that minimally two promoters are required for expression. Between genes *33* and *34*, there is a short, approximately 200 bp intergenic region and two divergent promoters can be identified by bioinformatics to predict *E. coli* Sig70 promoters (using the following software: Softberry and DNAmaster). Transcription from the leftwards-facing promoter, designated P_{rep} (pLO08; Table 3-2 and Fig. 3-5), would be able to transcribe genes *32* (*int*) and *33* (*rep*), the BPs immunity cassette. The rightwards-facing promoter, designated P_R (pLO07; Table 3-2 and Fig. 3-5), could conceivably transcribe all of the rightwards-facing genes in the genome, as BPs circularizes when growing lytically. The promoter-reporter fusions of P_R and P_{rep} produced pink colonies and were quantitatively assayed (see below).

In a BPs lysogen, the location of the *attP* core truncates gene *33*, the repressor, and separates gene *33* from the integrase gene *32*. The integrase is required for excision of the prophage and therefore must be expressed when prophage induction occurs. Expression of *32* in the prophage could occur by a BPs promoter located upstream of the *32* coding region, though no obvious promoter sequence can be identified, or by a bacterial promoter. A promoter-reporter fusion was constructed spanning the *32-33* region in the forward (pLO11) and reverse (pLO12) orientation (Table 3-2 and Fig. 3-5). Promoter activity was not observed in either orientation for the *32-33* intergenic region visually or by fluorescence microscopy (data not shown).

In BPs, genes *27* and *28* encode LysA and LysB, respectively, which are lysin proteins utilized by the bacteriophage to lyse the host cell and release progeny phage [29,30]. Between genes *26* and *27*, a short intergenic region is located and a promoter in this location could exist to direct the expression of the lysin genes, as has been shown in mycobacteriophage Ms6 [42] (pLO55, Table 3-2). No promoter activity was observed from the region between genes *27* and *28*.

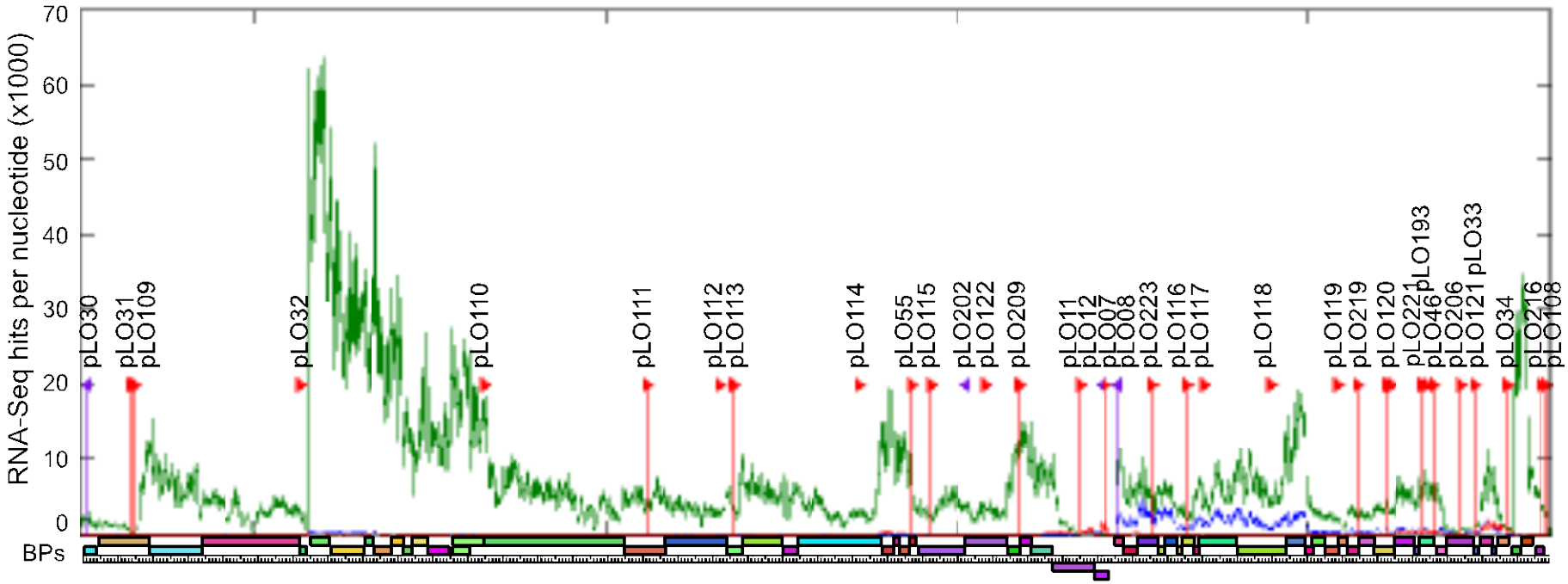

Figure 3-5. Locations of promoters examined with RNA-Seq profile

Fig. 3-5. Genomic locations of the promoters tested are noted with red (rightwards putative promoters) or purple (leftwards putative promoters) arrows and with the name of the promoter-reporter construct created to assay for promoter activity above. The BPs genome along the bottom of the figure indicates the genomic position and the RNA-Seq gene expression data are graphed in hits/nt x1000. Graph created in MatLab (MathWorks).

3.5.1.2 Screening for BPs promoters by bioinformatics and promoter trap analysis

Bioinformatic screens of the BPs genome using Softberry [160], predict a large number of promoters, but most have very weak scores. Five promoters with high scores were chosen (pLO30 through pLO34) and the promoter activities were examined (Table 3-1). Two of these regions produced mCherry fluorescence and were further characterized [(see below) (pLO32 and pLO34; Table 3-2)].

We also performed a promoter trap experiment, which allowed for an unbiased approach to finding promoter regions in BPs (see Appendix C). Fragments of BPs genomic DNA were randomly cloned upstream of mCherry and a library of these clones was screened for promoter activity visually, by identifying pink or purple colonies in *M. smegmatis*. The approach allows the identification of promoters with unexpected -35 and -10 sequences or promoters that may be located in open reading frames and not further examined. Though a large number of clones with promoter activity were identified, including promoters from BPs identified by other methods, technical issues made determining the actual genomic regions with promoter activity difficult. Namely, most of the clones contain multiple inserts of genomic DNA. From this experiment, we identified a region with promoter activity that spans the intergenic region between genes *54* and *55*, which was amplified by PCR from the BPs genome and cloned into a new vector (pLO46). Seven additional regions identified were cloned upstream of mCherry, but showed no promoter activity (pLO202, pLO206, pLO209, pLO216, pLO219, pLO221 and pLO223; Table 3-2 and Fig. 3-5).

3.5.1.3 Putative promoter locations determined from RNA-Seq profile

Lastly, we utilized a preliminary RNA-Seq profile to identify potential transcription start sites in BPs during early and late lytic growth. We identified locations throughout the genome where sharp increases in the number of transcripts were observed and constructed promoter-reporter fusions (pLO108-122; Table 3-2 and Fig. 3-5). The poor quality of the initial RNA-Seq (performed using a different method than with the data presented in this chapter) led us to choose a number of locations that do not appear to be transcription start sites with the final RNA-Seq profile, and thus only one of these promoter-reporter fusions showed promoter activity (pLO121; Table 3-2 and Fig. 3-5).

3.5.2 Activity of BPs promoters

Using the above methods, we tested 33 distinct genomic regions of the BPs genome and found six promoters in constructs pLO07, pLO08, pLO32, pLO34, pLO46 and pLO121 (Table 3-2), designated P_R, P_{rep}, P_6, P_{61}, P_{55} and P_{58}, respectively. The activities of these promoters range from modest, around 10-fold lower than the strong hsp60 promoter, to very high, approximately equal to the hsp60 promoter (Fig. 3-6A). Of the six promoters, five were also examined in a BPs lysogen [*M. smegmatis* mc^2155(BPs)], and three are strongly repressed (P_R, P_6, and P_{55}; Fig. 3-6A).

Surprisingly, the strongest promoter we identified, P_{55}, was located between genes *54* and *55* (Fig. 3-6B), and the RNA-Seq profile does not show high expression of these genes or a putative transcription start site. The locations of the other promoters do correlate with regions of gene expression seen in the RNA-Seq transcriptomic profile (Fig. 3-6B).

O_R, the binding site of the BPs immunity repressor gp33, at P_R has been characterized in genetic, biochemical and promoter activity studies [see Chapters 4 and 5, [31]]. Related operator sites can be found at five intergenic locations in the BPs genome, *5-6, 26-27, 33-34*, *54-55* and *60-61*. We found promoter activity at four of these locations, between genes *5* and *6* (P_6), between *33* and *34* (P_R and P_{rep}), between *54* and *55* (P_{55}), and between *60* and *61* (P_{61}) (Fig. 3-6C). However, in a lysogen, only three of the five promoters are repressed by the lysogen (Fig. 3-6A). The two promoters P_{rep} and P_{61}, which are not shut down in the BPs lysogen, do not contain operators that are a perfect match to O_R. These regions have imperfect operator sites and half operator sites. These data demonstrates that a perfect match to the O_R operator site found at P_R is required for repression of BPs promoters.

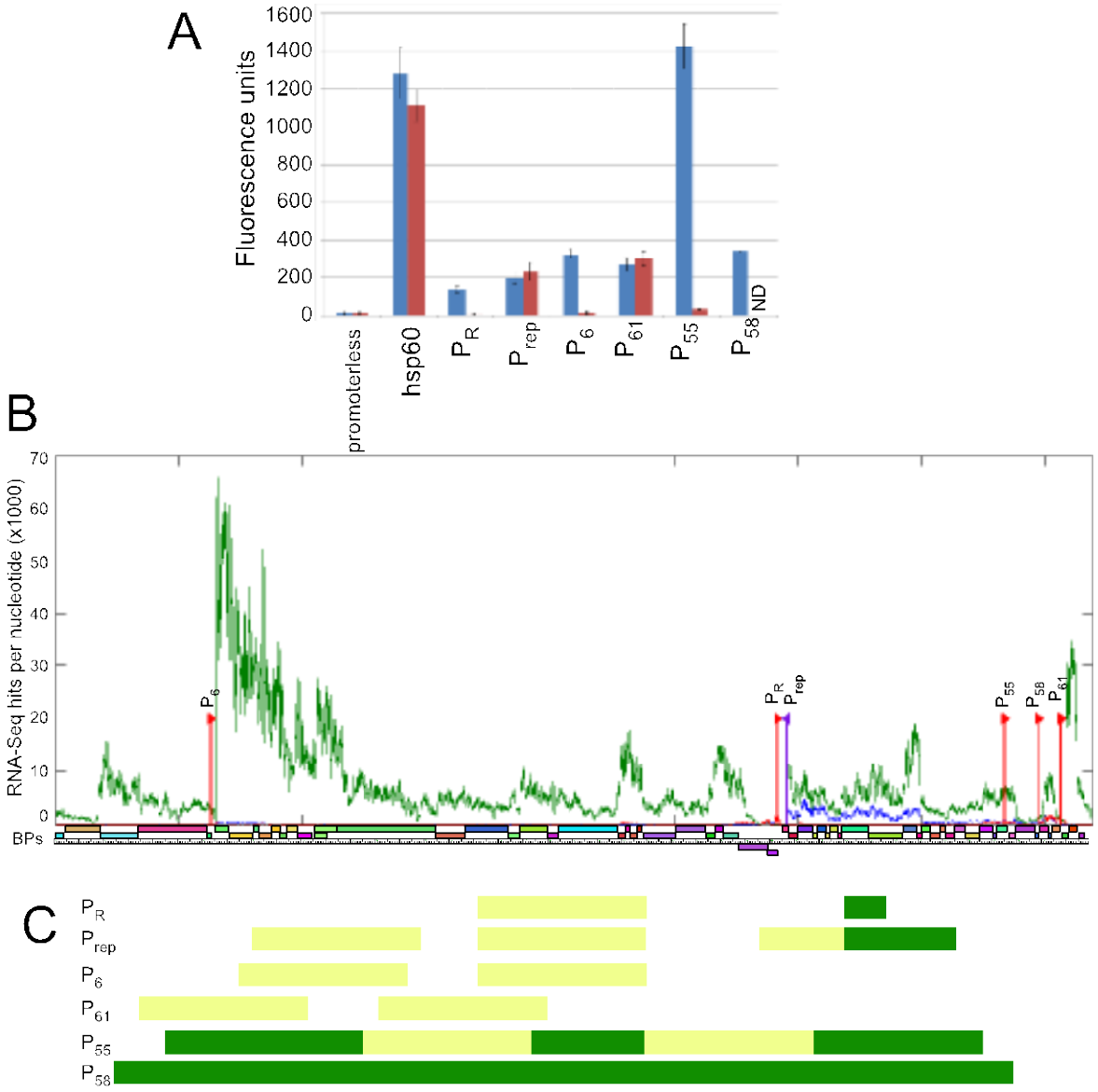

Figure 3-6. Locations of active promoters identified by promoter-reporter fusions.

Fig 3-6. (A) Promoter activity of BPs promoters identified. Fluorescence units measured in (LAU/mm^2)/OD_{600nm}. (B) The locations of the promoters identified from BPs are indicated with the RNA-Seq profiles (see Fig. 3-2) and BPs genome map provided for context. Red arrows indicate regions examined in the forward orientation and purple indicates those in the reverse. (C) The sequences of the identified BPs promoters. The -35 and -10 regions are noted with blue letter, potential operator sites are in yellow boxes and open reading frames are in green.

3.5.3 Transcription start site mapping of BPs promoters

Using a novel method of transcription start site (TSS) identification recently developed [161], we mapped the TSS of three of the six promoters. The TSSs were mapped from the promoter-reporter fusion constructs. Briefly, total RNA isolated from strains of *M. smegmatis* carrying the constructs was treated with tobacco acid pyrophosphatase to convert the 5' triphosphates to 5' monophosphates, the RNA circularized with T4 RNA ligase and converted to cDNA. Primers were designed to amplify over the 5' junction of the promoter and mCherry by placing the forward primer at the 3' end of the mCherry gene and a reverse primer in the 5' end of the reporter gene. Sequencing of products from this PCR reaction and analysis of the sequencing trace allows the determination of the 5' end of the transcript because the 3' end is variable and poor sequencing quality is achieved.

Both the P_R and P_{rep} promoters identified are transcribed as leaderless transcripts, as the TSSs are the first base in the translation start codon of genes *34* and *33*, respectively (Fig. 3-7A and B). The -10 motif of the predicted promoters is located 7 bp upstream of the TSS, confirming the location of the promoters. The TSS for P_6 is located in the intergenic region upstream of *6* and the predicted -10 is also 7 bp upstream (Fig. 3-7C).

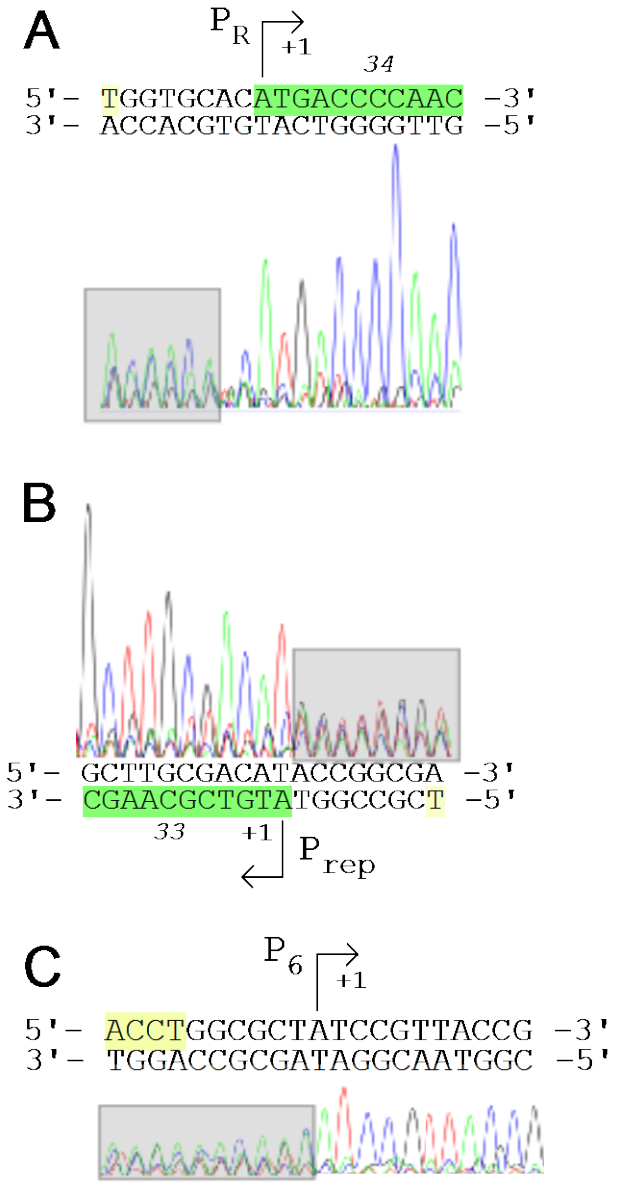

Figure 3-7. The TSS of three BPs promoters.

Fig. 3-7. The promoter transcription start sites were mapped with a novel method involving the circularization of transcripts from the promoter-reporter fusions. The sequences are provided and green boxes indicate the coding regions of downstream genes. The sequence trace files are shown with gray boxes located where the called sequences did not match the promoter region. Bent arrows indicate the TSSs of P_R, P_{rep} and P_6.

3.5.4 Summary of BPs promoters

The sequences of the BPs promoters are similar to mycobacterial SigA promoter sequences with spacers ranging from 18 to 21 bp in length (Fig. 3-8A). The most well conserved bases are 5'TTGNNN in the -35 motif and 5'-TANNNT in the -10 motif (Fig. 3-8B), and each of the promoters with mapped TSSs have a 7 bp spacer between the -10 and TSS and initiate at an A base (Fig. 3-8A). None of the BPs promoters we identified contains an extended -10 5'-TGN motif.

A

Promoter	-35 motif	Spacer	Extended -10	-10 motif	Spacer	TSS
P_R	TTTCCA	18	CGC	TATGTT	7	A
P_{rep}	TTGCGC	18	ACA	TACAAT	7	A
P_6	ATGTCG	19	GGT	TAACCT	7	A
P_{61}	TAGACG	19	AGC	TATCGT		
P_{55}	TTGCGA	18	TGT	GACACT		
P_{48}	TCGGTC	21	ATG	CATGAT		

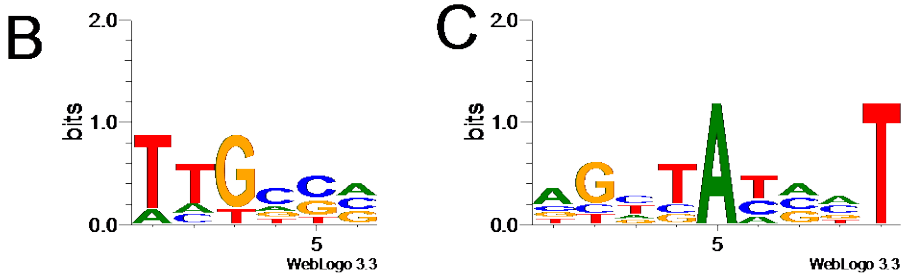

Figure 3-8. The sequences of the six BPs promoters identified and their conservation.

Fig. 3-8. (A) The sequences of the -35, extended -10 and -10 motifs for each BPs promoter is provided. The identity of the nucleotide of the transcription start site (TSS) is included in the three promoters that it has been mapped. Also the number of nucleotides between the features (spacer) has been provided. Note that the spacer between the -35 and -10 motifs is the number of nucleotides between the hexamers and includes the 3 bases of the extended -10 in its count. (B) The conservation of the -35 motif of the BPs promoters is shown with WebLogo (Berkeley, Version 2.8.2). (C) The conservation of the extended -10 and -10 motifs of the BPs promoters is shown with WebLogo (Berkeley, Version 2.8.2).

3.6 MAPPING TRANSCRIPTION TEMRINATORS IN BPS

Transcription termination is another important part of transcriptional regulation. While some transcriptional start sites can be predicted from the sharp increases in RNA-Seq hits, the location of transcriptional terminators in BPs is more difficult to predict from the RNA-Seq data because of the gradual reduction in transcript levels through out the operons. The location of Rho-independent transcriptional terminators can be bioinformatically predicted but this process can be computationally slow and unreliable because terminators are not a conserved sequence but rather RNA stem-loop secondary structure. Terminators in BPs were predicted using TransTermHP [154] and then were experimentally checked with a "terminator trap" experiment, similar to the promoter trap described earlier this chapter.

3.6.1 Prediction of intrinsic terminators with TransTermHP

Sequences that might possibly function as terminators were identified in the BPs genome by the TransTermHP prediction program [154]. The algorithm searches the intergenic regions of the genome (which were determined by Glimmer3, not the annotation of the genome) for short stretches of sequence enriched for thymines and then examines the upstream region for potential hairpin structures, which are scored based on the energetics of their conformations [154]. TransTermHP identified 20 sequences as putative intrinsic terminators (Table 3-3).

Of the 20 predicted terminators in BPs, 16 are predicted for rightwards-transcribed regions and 4 are for leftwards-facing genes. Most of the predicted terminators have weak ΔG values (17 out of 20 have ΔG less negative than -10). The authors of a different terminator prediction algorithm used a $\Delta G_{cut\ off}$ based on the GC content of the genome to reduce false

positives and had a false negative rate of less than 10% [123]. The $\Delta G_{cut\ off}$ for BPs with a GC content of 66.6% (phagesdb.org) would be -17.34. Only one of the TransTermHP predicted terminators (at gene *13-14*) would be considered a likely terminator candidate using these parameters. Though a $\Delta G_{cut\ off}$ was not used in these studies to narrow the candidate list, this demonstrates the weak ΔG values of these predicted terminators.

A potential weakness in using TransTermHP with BPs is that it is not able to predict I-shaped terminators, defined as terminators with a stem-loop structure but less than 3 U bases in the 10bp downstream [123,154], due to its search method that first identifies T residues. Two of the reverse terminators predicted are I-shaped terminators. All of the rest of the terminators predicted are L-shaped with a stem-loop structure followed by a 10bp stretch with at least 3 Ts (from 3 to 7). Finding 18 putative L-shaped terminators within the BPs genome is surprising because most mycobacterial terminators (>80%) are I-shaped [122], which prevented their detection for many years [121,123].

Table 3-3. Terminators predicted in BPs by TransTermHP.

BPs Gene	BPs coordinate start	BPs coordinate stop	Strand	*! G*	5' tail	Stem-loop	3'tail	Plasmid
3-4	3389	3415	+	-10.1	CGCTGGGCCGGCCGA	CGCTGGTCGGC-TAAAT-GTCGATCGGCG	TTCCTGAGTTCCGCG	pLO124
5-6	6416	6429	+	-2.3	CACGACGCTTGCACC	GCGAC-ATAT-GTCGG	TTTTGGACATATGCG	pLO125
13-14	10474	10513	+	-24.2	CGCCGCGTAAGCCGG	CGGGCCCGCGCTCCGGTC-TGGT-GGGCGGCGCGCGGGCCCG	TCTCTACGCAATGCA	pLO126
16-17	15474	15488	+	-4.2	GACGACGAAGGTTCA	CTGATG-ACG-CAACAG	TTTCTCGGCGACGAC	pLO127
25-26	23612	23630	+	-2.3	GGCGCTGGCCGCGGC	GCAGGGC-AACCA-GTCCTGA	TGATTCGTCGATTGC	pLO128
26-27	23875	23895	+	-4.7	GACATATGTCGTGGA	TGATCGG-GATATGG-CCGATCGT	TTCTTCCCGATGCG	pLO129
30-31	27106	27125	+	-2.8	ACGAGGGGGTGGCAC	GGT-GACGA-TCG-TCGACGACC	TATGGTCCTGGCCGT	pLO130
32	28007	28063	+	-5.1	CCGAGCATCGTTTGT	ACTGCG-CGCAGATTGCGCGATCCGCG-GTAGG-CGCGCGTGGCG-AATCGGTGGCGCAGC	TTGTGCATGGTCCAG	pLO133
43-44	34397	34407	+	-2.4	GTGAAGGGAACGAGT	GACC-ATG-GGAC	TGATTCAGGACCAGC	pLO134, pLO178
45	34969	34999	+	-6.7	GGCAGACGCCGGCCA	CTCGGCGCACTGG-ATCAG-CGAGTGCGCGGCG	TGCTTCGCGATGTTC	pLO135
47	35612	35654	+	-4.1	CACCTTGCACGGAAA	CCGTGCAACAT-CAGCGACT-TTCG-AGACGCACAATCTTGCACGG	TTTTTTGTGCAACAT	pLO136
54-55	38607	38635	+	-8	TCCGTACACGGTGGC	GCCGATGCCCAG-ATACA-CGGGGTAGCGGC	TGTTCGTCGCTCTGC	pLO138
55-56	38964	38987	+	-2.2	ACCTGCAACAACGGG	CGGGCGTGAC-CATG-GCCGCGATCG	TTCTCCGCTCCGAAC	pLO139
56-57	39754	39781	+	-2	CGTCGCCGCTCGACT	GGTGGCCGCAT-AATGCAC-AT-CGGCGACC	TGTTCATGCTGCCCG	pLO140
58	40032	40046	+	-7.8	CCTGAACACCGCGA	TCGTGC-TTG-GCGCGA	TAGGTTCAACGCCCG	pLO142
60-61	40797	40828	+	-14.4	CGCGTAGACGACATA	TGTGCTCCCCCAGC-TATC-GTCGGGGGAGCACG	TTTGTTTCTACGACC	pLO143
56R	39382	39410	-	-4.5	GTGCGGCGCCCAGAA	GCTCGAC-ACCG-CGGCGC-CGGCCGGCGAGC	TCTACACGTCCGCTC	pLO141
47R	35576	35608	-	-2.6	CGTCATGCACGAAAA	CCGTGCAACATCAG-CGGTC-CACACCTTGCACGG	AAACCGTGCAACATC	pLO137
32-33R	28845	28855	-	-2.4	CCGCGGCATCTGAAA	CGGG-TTT-CACG	GACCTTATTGTCCGG	pLO132, pLO177
31R	27352	27367	-	-6.8	GAACGCCAACGAGAA	CGC-GCG-GAT-CGCCGCG	GAAGCGAAGTCGTTC	pLO131

!

Table 3-3. Putative terminators predicted by TransTermHP in BPs intergenic regions, which were predicted by Glimmer3 (pepper.molgenrug.nl). The BPs sequence coordinates indicate the predicted terminator sequence location and the pLO plasmids include additional surrounding sequence. The + indicates terminators predicted for rightwards transcribed regions and – indicates terminators for transcription leftwards.

Two of the predicted terminators are located in regions known to have promoter activity. The first is within the intergenic region between genes *5* and *6* and the other is with the intergenic region between genes *60* and *61* (Fig. 3-9). The sequence of the putative terminator overlaps the promoters.

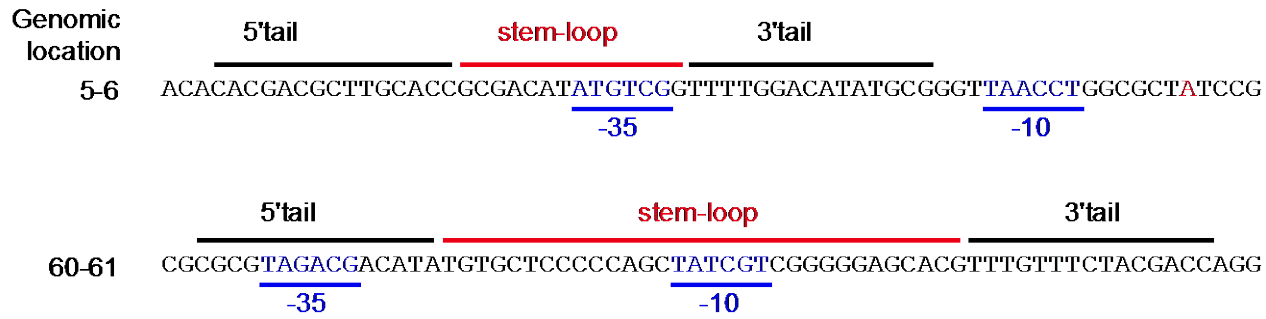

Figure 3-9. Genomic regions with both promoters and predicted terminators.

Fig. 3-9. The locations of the terminators predicted for the *5-6* and *60-61* intergenic regions in BPs are shown with the locations of the P_6 and P_{55} promoter sequences (blue text). The experimentally verified TSS for P_6 is shown in red.

3.6.2 Terminator trap to confirm TransTermHP predictions

To test the ability of these sequences to decrease the expression of a downstream gene, a vector was constructed to allow the insertion of sequences between a mycobacterial promoter and the mCherry reporter gene. Sequences that are able to function as efficient terminators will have decreased fluorescence activity, due to decreased expression of mCherry.

A moderately active promoter, the BPs P_6 promoter previously described, was chosen for expression of mCherry to reduce the number of false positives for terminator sequences. The hsp60 promoter can be more easily detected but the instability of the promoter would lead to false positive results [[162]; M. Olm and L. Oldfield, unpublished results).

The putative terminators were cloned between the P_6 promoter and the ribosome binding site. Of the 20 sequences predicted by TransTermHP, 15 of the plasmids had fluorescence activity at 50% or lower of the vector alone (Fig. 3-10A). The levels of fluorescence from these 15 plasmids is similar to that of a positive control with the *E. coli rrnB* terminator.

Five of the plasmids did not show termination. The fluorescence activity from the genomic locations 58 and 60-61 was significantly less than the vector control but above the 50% cut off. Three others clones, locations 5-6, 32-33R and 43-44, had fluorescence readings significantly higher than the vector only control. The lack of termination in 4 out of 5 of these clones can be attributed to the presence of promoters within the sequence.

The increased fluorescence from 32-33R and 43-44 was unexpected and further analysis led to the discovery of promoters within the plasmid sequences. The -10 region of these promoters, TACACT, was formed by the 3' cloning junction of the insert to the vector, with 3bp from the insert, TAC, and the 3bp of the half cut site of ScaI, ACT. SDM was performed to remove 5 bases (TACAC) within the -10 of the predicted promoter from these. The activity of 32-33RΔ5bp and 43-44Δ5bp indicated that these DNA sequences are able to reduce fluorescence to less than 20% of the vector alone (Fig. 3-10B). Thus, the evidence gathered so far supports that 17 of the 20 regions predicted by TransTermHP and tested for an ability to reduce fluorescence in the vector are terminators.

The increased fluorescence from clones with the 5-6 and 60-61 genomic locations is hypothesized to be a result of the native promoters found in these regions of the BPs genome (Fig. 3-9).

The only genomic region predicted by TransTermHP that was not determined to have the ability to decrease fluorescence below 50% of vector is region 58. The sequence does not have a

promoter sequence that is detectable with standard promoter prediction software (Softberry). The translation start codon is present however which may lead to translation of the mRNA produced from P_{58} and the coupling of translation with transcription can prevent termination from otherwise active terminators [124].

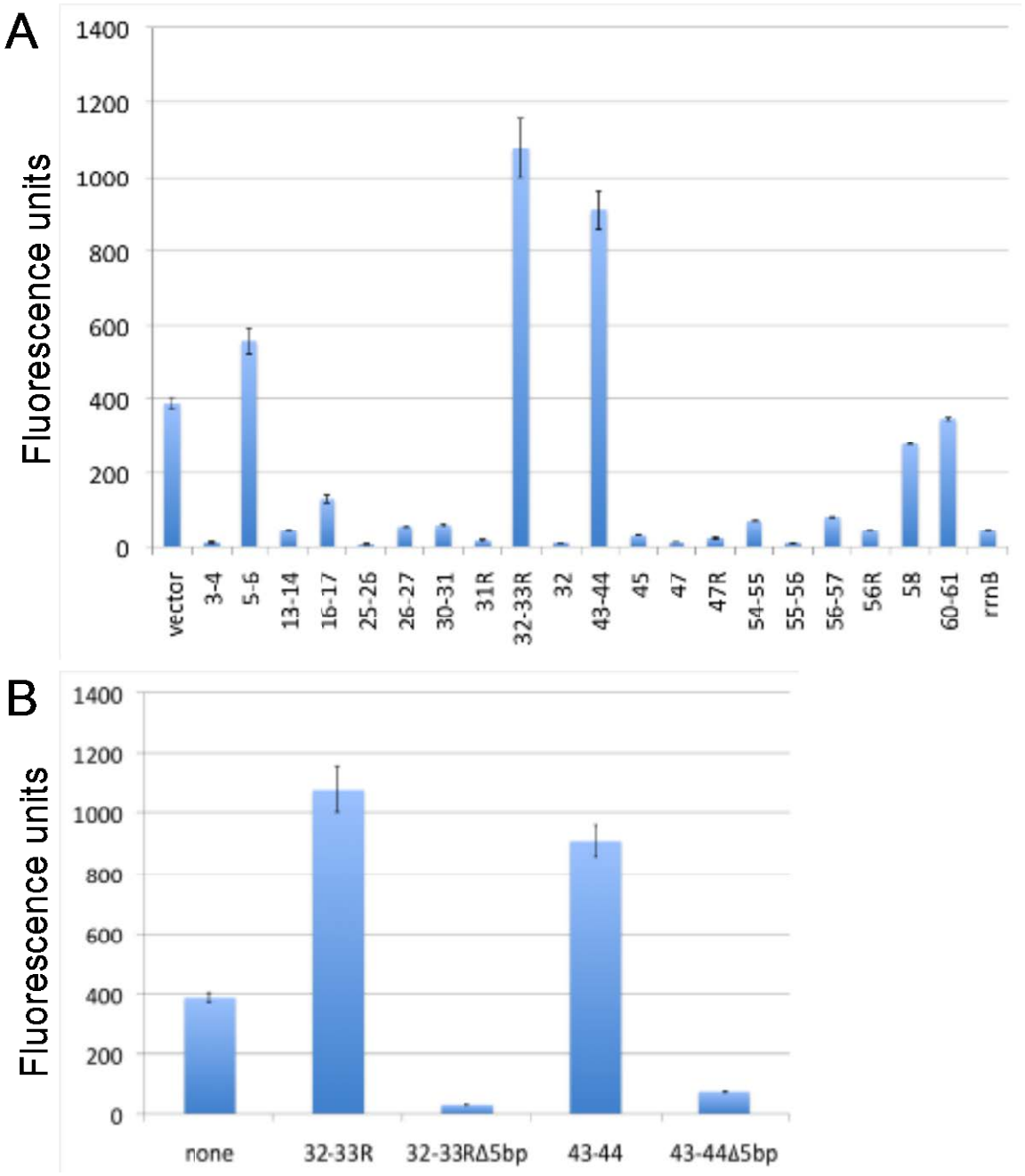

Figure 3-10. The quantification of BPs terminator activities.

Fig. 3-10. (A) Fluorescence of vectors with putative terminators from BPs inserted between the P_6 BPs promoter and an mCherry fluorescence gene, to identify regions that can eliminate fluorescence. (B) In two of the terminator assay vectors, promoters were created at the cloning junction. The Δ5bp vectors show that these regions do eliminate mCherry fluorescence when these promoters are removed. Fluorescence units measured in $(LAU/mm^2)/OD_{600nm}$.

3.6.3 Further studies required to confirm transcriptional termination

These regions are bioinformatically predicted to produce stem-loop structures indicative of Rho-independent transcriptional terminators and have been shown to reduce fluorescence output in when placed upstream of a reporter gene. These data suggest that the genomic regions identified function as terminators in BPs. However, the levels of mRNA and the 3' ends of the RNA itself were not examined in these experiments and need to be to directly show that transcription cannot continue or is dramatically reduced at these locations.

3.6.4 Transcriptomic profile does not agree with terminator locations

When the location of the functional terminators is examined with the RNA-Seq transcriptomic profile, many of these terminators are found within regions of the genome that have high expression during late lytic growth and do not seem to correlate with decreasing amounts of transcript abundance (Fig. 3-11). Four of the terminators identified (26-27, 32, 47, 56-57) do appear in genomic locations where decreasing transcription is observed and might represent terminator sequences used by BPs. Other bacteriophages employ anti-termination as a mechanism of gene regulation, and BPs may harbor an anti-termination mechanism that is responsible for overcoming the large number of functional terminators we identified in BPs.

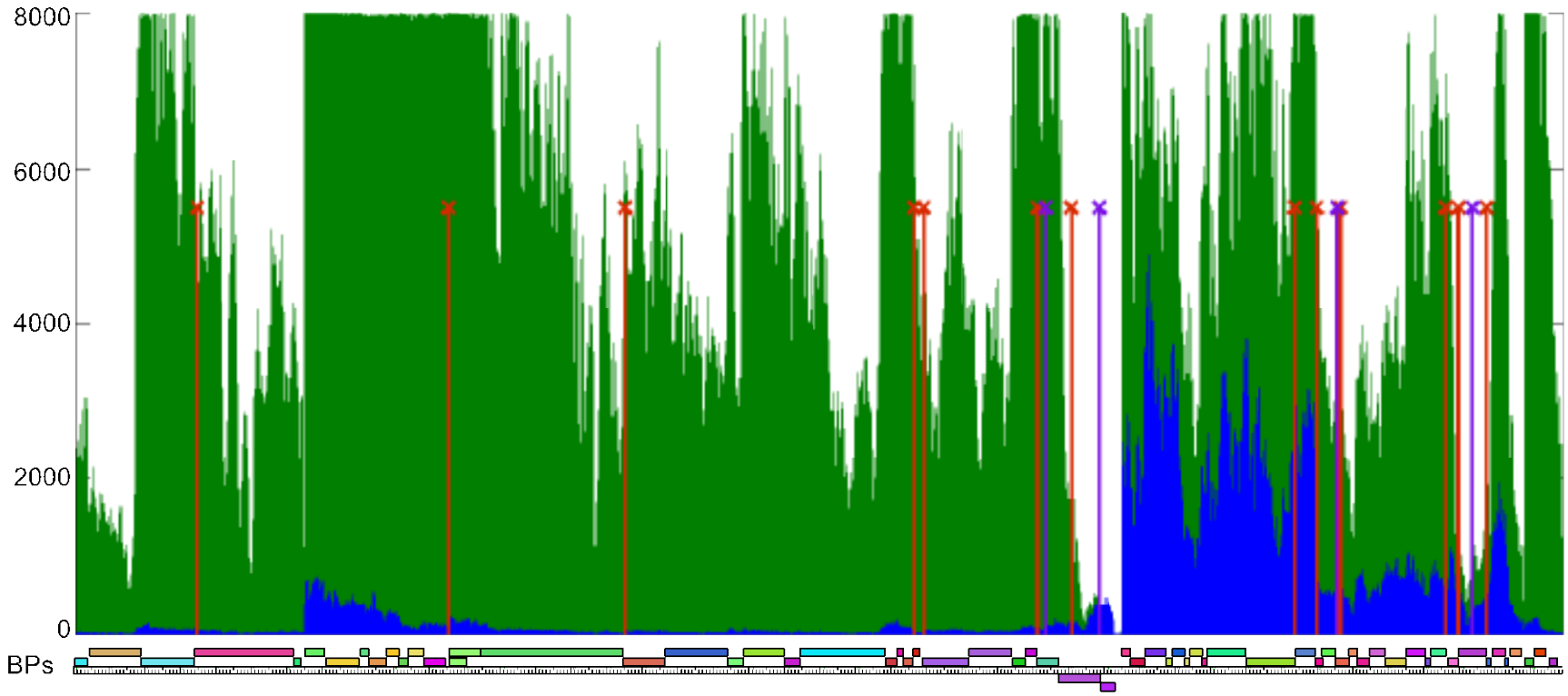

Figure 3-11. Location of functional terminators in BPs and with lytic transcriptome.

Fig. 3-11. The genomic locations, indicated by BPs genomic map along x-axis, of functional terminators are marked in the forward orientation (red X) and in the reverse orientation (purple X). The RNA-Seq profiles (see Fig. 3-2) and BPs genome map provided for context. Gene expression (hits/nt) in late lytic growth (green) and early lytic growth (blue) from the RNA-Seq transcriptomic profile is provided for comparison. The maximum gene expression (hits/nt) is artificially capped at 8000 hits/nt. Graph created in MatLab (Mathworks).

3.7 DISCUSSION

3.7.1 Giles and BPs

BPs shows a distinct pattern of gene expression in lytic growth relative to mycobacteriophage Giles [8], the only other mycobacteriophage for which an RNA-Seq analysis has been performed. Much of this may be explained by a more complex arrangement of the Giles genome, with three separate instances of switches from rightwards-facing to leftwards-facing genes, whereas, in BPs, only the repressor and integrase are expressed leftwards and neatly divide the early and late genes. Gene expression in Giles is divided into three portions, the structural, lysis and early genes. The structural and lysis regions are both expressed late in infection [8]. The early genes are expressed to the same levels in both late and early infection.

In both phages, the early transcribed genes consist mostly of a subset of genes in the right arm of the phage. The function of the majority of these genes are unknown for both BPs and Giles, though in Giles, deletion analysis has demonstrated that some of these genes are essential for growth of the phage [8]. The level of expression in the end quartile of the Giles genome is very low for all samples examined. This is not the case with BPs and transcripts are recorded all the way to the end of the genome. Very little expression of the structural genes is observed in early infection for Giles or BPs.

Like BPs, the structural genes in Giles are expressed only in late infection and account for some of the highest expression in the RNA-Seq samples. A putative operon that begins at Giles gene *6* appears to be analogous to the structural gene expression in BPs, which begins

sharply also at gene *6*. The early genes are expressed at nearly identical levels between early and late infection in Giles. In BPs, the early genes are expressed more highly in the late infection sample.

3.7.2 Analysis of transcription in BPs

Using several different approaches, we have analyzed the transcriptomic profile of mycobacteriophage BPs during lytic and lysogenic growth. Using the RNA-Seq profile and reverse transcriptase PCR (RT-PCR), we have identified five regions which are transcribed as operons, genes *62-1*, *3-4*, *5-16*, *33-32*, and *34-44*. In four of these operons, a promoter has been identified which drives their expression. The location of terminators for these operons is harder to define.

The operon *61-2* is driven by the P_{61} promoter but no terminator at the end of this operon was identified by the TransTermHP program. The P_{61} promoter is moderately constitutively active, though the RNA-Seq profile indicates that this operon is temporally regulated, as there is pronounced expression in late lytic growth and none during early infection.

The RNA-Seq profile indicates a putative TSS near the beginning of gene *3* and an operon containing genes *3-4*. However, no promoter was found upstream of this operon though two slightly different regions were tested near the end of gene *2* (pLO31 and pLO109). A terminator was identified in the middle of this operon (between genes *3* and *4*), and no terminator was predicted at the end of the operon between genes *4* and *5*.

A number of the structure and assembly genes are located in the operon that extends from gene *6* to *16*. The operon is driven by promoter P_6, and its TSS has been mapped and corresponds exactly with the sharp increase in expression in the RNA-Seq profile. Also, P_6 is

regulated by the lysogen and very little expression is seen in early lytic growth or in a BPs lysogen. A terminator located between genes *16* and *17* is found at the end of the operon. An additional terminator is located between genes *13* and *14*.

The immunity cassette of BPs, *33-32*, forms another operon, though these genes can only be transcribed together prior to integration of the prophage into the host chromosome, and relocates one gene to the *attL* and the other to the *attR*. P_{rep} drives expression of this operon and is constitutively active, though only very low levels of genes *33* and *32* are observed during lytic growth. The TSS of P_{rep} has been mapped. A terminator in gene *31* has been identified and can eliminate transcription in the leftwards direction.

Finally, genes *34* to *44* are expressed in an operon that can be clearly observed in the RNA-Seq profile as the predominant early operon and is also present during late lytic infection. A terminator located in the beginning of gene *45* may be responsible for the decrease in expression levels after gene *44*. The expression of this operon is driven by P_R, which is tightly regulated in the lysogen.

A promoter, P_{58}, located upstream of the mobile element, gene *58*, may be responsible for its expression [10]. Transcription of this region is seen in both lytic and lysogenic growth, which suggests that P_{58} is not regulated by the repressor but P_{58} activity has not yet been assayed in *M. smegmatis* mc^2155. The location of P_{55} is puzzling as the promoter-reporter experiment shows that this is very strong promoter, equal to the level of the hsp60 promoter. However, only moderate levels of transcript abundance are seen at this location, which decreases to very low expression in gene *56*, and no increase indicative of a TSS is present.

The promoter and terminator results presents several new questions about the regulation of gene expression in mycobacteriophage BPs. The promoters are active in the absence of phage-

encoded activators. However gene expression from P_6 and P_{61} are only observed during late lytic infection and not early in infection. A mechanism that controls the temporal regulation of these promoters has not been described but it appears that some mechanism is present. Also P_{61} is not repressed by the lysogen in the promoter-reporter fusion assay but transcription of the downstream region is not observed in lysogenic growth. A putative operator site is located near the promoter and was included in the construct tested, but there is a possibility that an additional regulatory sequence is missing.

A high number of terminators have been identified from the BPs genome, but do not correspond to the RNA-Seq profiles of late and early infection. This suggests that a phage-encoded protein is required for anti-termination and may present a method of temporal regulation of gene expression. In the early operon transcribed from P_R, no terminators were predicted with the TransTermHP program between genes *34* and *43*, indicating that if an anti-termination protein is present it may be encoded by one of these ten genes.

The RNA-Seq profile demonstrates that BPs genes are temporally regulated. An early operon, where no terminators have been identified, may contain an anti-termination factor that allows for the expression of additional genes and helps to control the temporal program of lytic gene expression. Phage lambda gene expression is well studied and it encodes for two anti-termination factors, N and Q, which are important for appropriate timing of gene expression [53]. The regulation of some promoters is also not well understood, and BPs does not encode a sigma-like protein or its own RNA polymerase, though other mechanisms of modification of host RNA polymerases have been described in T4 phage of *E. coli* [50].

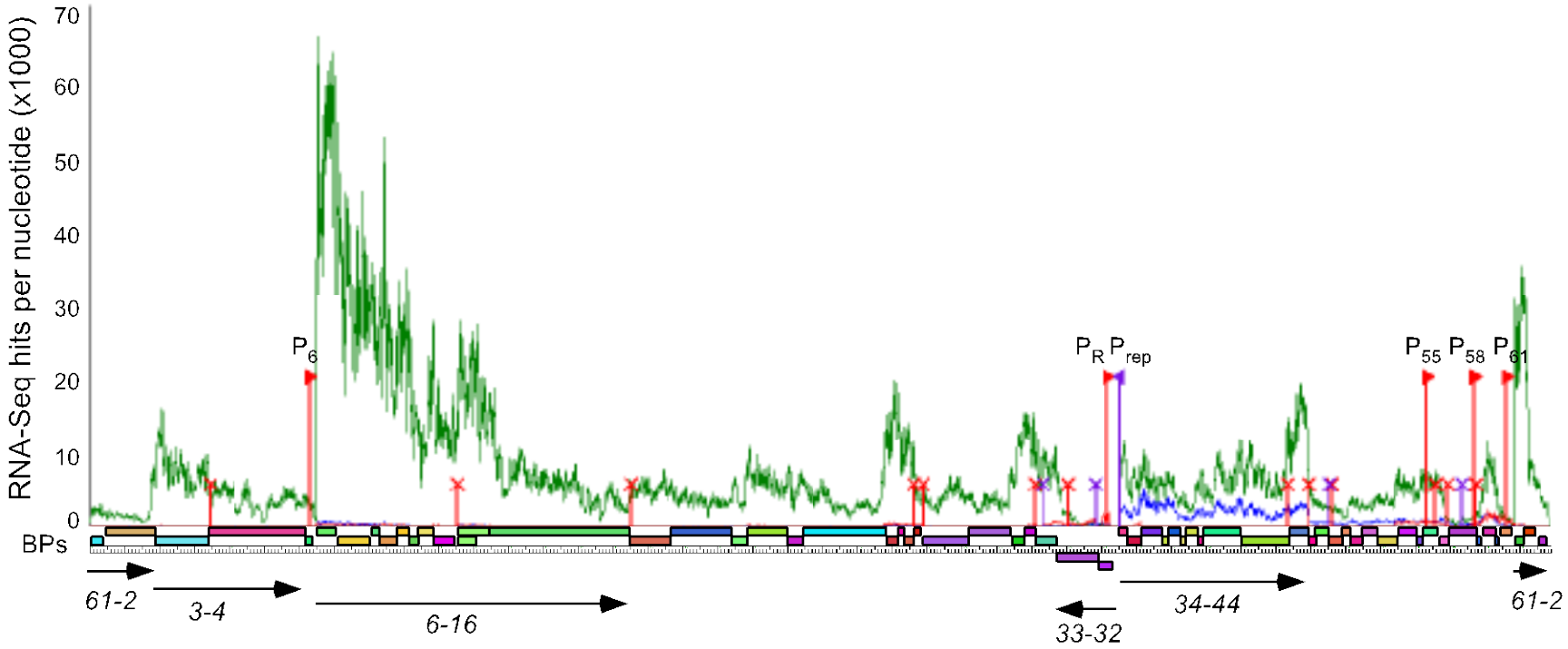

Figure 3-12. Full transcriptomic analysis of BPs in lytic and lysogenic growth.

Fig. 3-12. The RNA-Seq transcriptome is displayed in a line graph with early (blue) and late (green) lytic infection and lysogenic (red) growth. RNA-Seq data is represented as hits per nucleotide (x1000). The five rightwards (red) and one leftwards (purple) promoters are indicated by stem-arrows. The functional terminators in the forward (red X) and reverse (purple X) direction. The arrows below the BPs genome map indicate operons based on the RNA-Seq and RT-PCR experiments.

4.0 DISSECTION OF SEQUENCE REQUIREMENTS OF BPS PROMOTER P_R

4.1 INTRODUCTION

Gene expression in prokaryotes is usually controlled at the transcriptional level, and transcription regulation in mycobacteria is complex, involving many factors. Binding of RNAP to DNA is directed by the DNA binding of the sigma factor of the holoenzyme and the genomes of *M. smegmatis* and *M. tuberculosis* contain genes for 28 and 13 sigma factors, respectively [79,84,93]. Each of these sigma factors recognizes a different promoter sequence [reviewed in [75,94]]. However, SigA has been designated at the housekeeping sigma factor and is necessary for growth in both *M. smegmatis* [81] and *M. tuberculosis*. The consensus sequence of SigA promoters in mycobacteria has been reported by multiple groups [59,72-76], and the sequences derived differ slightly from one another. It is clear from these reports that the sequences -10 and -35 of SigA promoters are similar to the consensus sequence for Sig70 promoters in *E. coli* [60].

One study compiled 102 promoters, which had experimentally verified transcription start sites, with conserved -35 and -10 sequences and found a consensus sequence of 5'-TTGACA for the -35 hexamer and 5'-TATAAT for the -10 hexamer [72]. The best conserved nucleotides are the 5'-TTG of the -35 and the 5'-TA and 3'-T (TANNNT) of the -10 [72,73]. Mycobacterial promoters may also contain an extended -10 motif, which plays a major role in promoter

strength, and may be especially important for promoters that do not have a well conserved -35 motif [71].

Several strong promoters identified from mycobacteria have been used for gene manipulation in mycobacteria [36,159,163]. More tools for controlling gene expression in mycobacteria are needed to advance a number of experimental approaches, including the development of recombinant mycobacterial vaccines [159,164]. While synthetic promoter libraries have been developed for other bacteria [165-168], no collection of mycobacterial promoters with variable strengths has been created and characterized.

We completed a detailed dissection of the BPs P_R promoter and determined the contributions of each nucleotide of the -35 and -10 motifs to promoter activity. P_R resembles a SigA promoter, closely following the consensus sequence in its -10 region, though mutations in the -35 were more divergent. We also examined the role of the length and sequence of the spacer region between the -35 and -10 hexamer and found that changes in spacer length and sequence affect promoter activity. Combining the mutations these mutations, we created a collection of promoters with calibrated activities in *M. smegmatis* and *M. tuberculosis*, and which can be used to fine-tune the expression of heterologous genes in mycobacteria.

4.2 EXPRESSION OF P_R IN INTEGRATED AND EXTRACHROMOSOMAL CONTEXTS

The BPs P_R promoter is located immediately upstream of the early lytic operon that begins at gene *34* and is regulated by the BPs repressor gp33 (Fig. 4-1A). The start site for transcription has been mapped and corresponds to the first base of the initiation codon of gene *34* (Fig. 4-1B).

Putative -10 and -35 motifs have been predicted with reasonable correspondence to the mycobacterial consensus sequences (Fig. 4-1B), and are separated by 18 bp. P_R lacks the extended -10 5'-TGN reported to strongly influence promoter activity in some sigma factor (SigA) promoters [71]. In the spacer region between the -35 and -10 motifs, a 12bp operator, O_R, is located and is the binding site for gp33 [31].

When fused to an mCherry reporter gene in an extrachromosomal vector, P_R has moderate activity, about 7-fold lower than the strong hsp60 promoter (Fig. 4-1C), although when integrated into the chromosome using an integration vector derived from phage Tweety [169], P_R activity is barely detectable above the background of a promoterless vector (Fig. 4-1C). The extrachromosomal:integrated activity ratio for P_R of 43-fold is considerably greater than the ratio of 7-fold for the hsp60 promoter (Fig. 4-1). The copy numbers of pAL5000-derived extrachromosomal vector has been experimentally determined by qPCR to be 23, but may vary when carrying strong promoters [170]. Activation of P_R in the extrachromosomal context is consistent with a model in which its reduced chromosomal activity contributes to its repression in the prophage state, but is fully active when extrachromosomal during lytic growth. DNA supercoiling could play a role in this activation as described previously for other mycobacterial promoters [171].

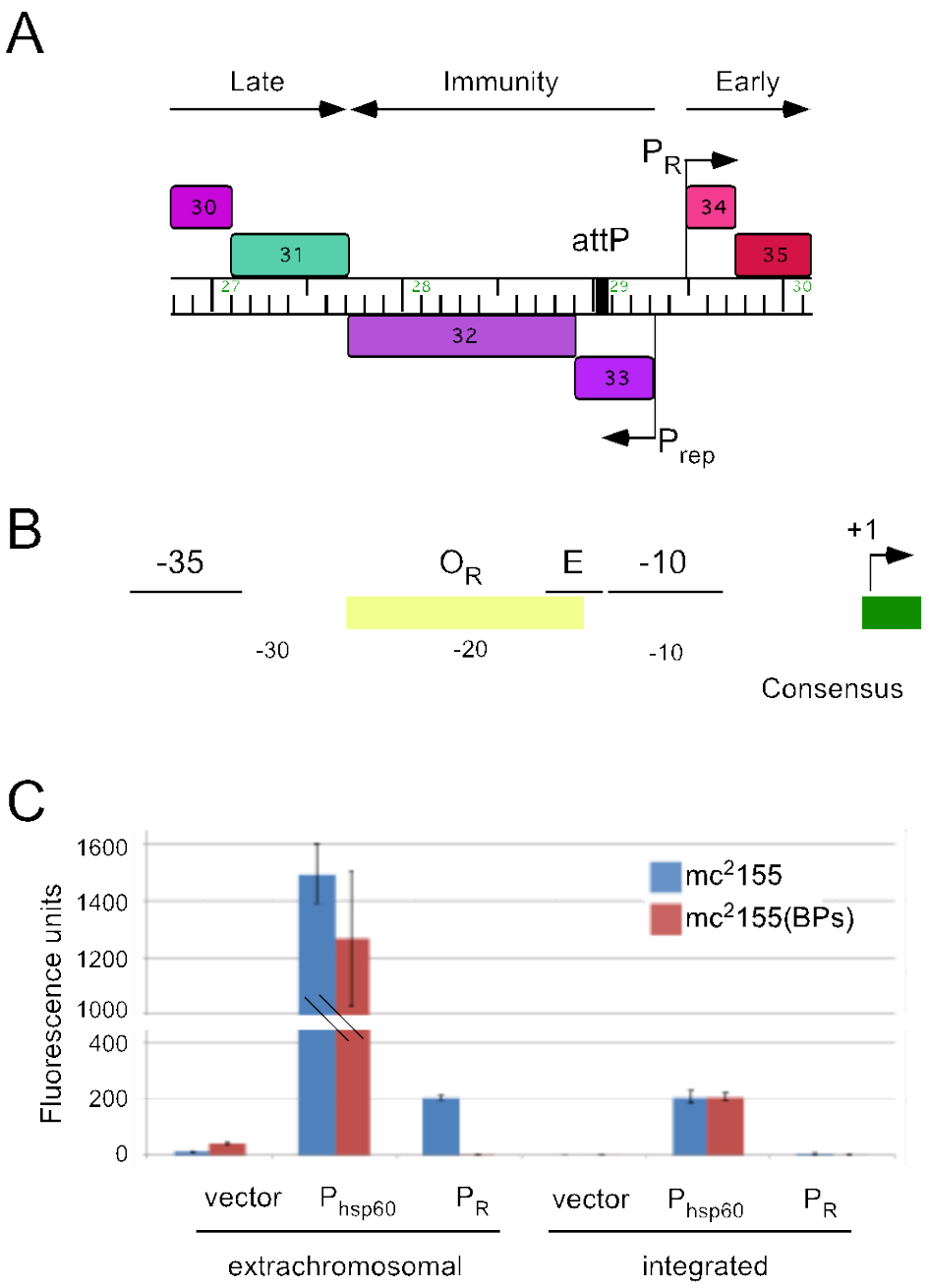

Figure 4-1. P_R is located upstream of a putative early operon.

Fig. 4-1. (A) A map of the BPs genome from approximately coordinates 26790 to 30150 generated in Phamerator ([32]), where the ruler indicates the genomic position, the boxes represent open reading frames, which are transcribed rightwards when located above the ruler and leftwards when located below. The P_R and P_{rep} promoters and the BPs attachment site (*attP*) are noted along with some putative gene expression information. (B) The sequence (BPs coordinates 29470-29512) of the -10 and -35 regions (blue letters) of P_R contains a recognizable mycobacterial SigA promoter. The transcription start site was previously mapped to the A of the

translation start codon of gene *34* [31]. The 12 bp operator (O_R), the binding site for Rep, is located between the -35 and -10 regions and overlaps the location that an extended -10 (E) 5'-TGN would be located. The consensus sequence for mycobacterial SigA promoters is included below the hexamers [72]. (C) The activity of the P_R promoter driving mCherry expression was examined in an extrachromosomal vector and integrated into the *M. smegmatis* mc^2155 chromosome using a Tweety integration vector [169] with a promoterless and hsp60 promoter control. The activity of the constructs was measured in wild-type *M. smegmatis* mc^2155 and *M. smegmatis* mc^2155(BPs), a lysogen carrying the BPs prophage.

4.3 INFLUENCES OF SINGLE BASE SUBSTITUTIONS IN THE -10 MOTIF ON P_R ACTIVITY

The similarity of the -10 hexamer of P_R (positions -8 – -13; Fig. 4-1B) to the previously reported 5'-TATAAT consensus strongly suggests that P_R is recognized by SigA. To examine the sequence contributions at the -10, we constructed a series of base substitution mutants and measured their promoter activity using an mCherry gene reporter (Fig. 2A). Positions -8 and -13 showed a strong preference on the wild-type T at that location, and the wild-type A base is preferred at -12 . Most of the substitutions at these three positions are severely deleterious to P_R activity. At the -9 position, substitution of the wild type T base with either A, C or G substantially increased activity, and both G-10A and G-10C have elevated activities (Fig. 4-2A). Bases at the -10 contributing to maximal activity can be expressed as 5'-TATAMT (Fig. 4-2A). The activities of the mutant promoters were also tested in a BPs lysogen (Fig. 4-2B), and moderate negative effects on repression were observed with some mutations, primarily T-36A and C-34G (see below).

Some SigA promoters have an extended -10 sequence 5'-TGN (Fig. 4-2B) that substantially elevates promoter activity and renders such promoters relatively insensitive to

changes in the -35 motif [64,71]. The corresponding region of P_R has the sequence 5'-CGC (positions -16, -15, -14) and thus lacks an extended -10 motif. We examined the roles of bases in these positions and found that several substitutions, including C-14G and G-15A, increased promoter activity more than two-fold. However, we were unable to transform the C-16T mutant P_R-mCherry fusion into *M. smegmatis*, even after multiple attempts. This is the only mutant out of more than seventy we have constructed that displays this phenotype. Because strong promoter activity is known to interfere with plasmid replication, we

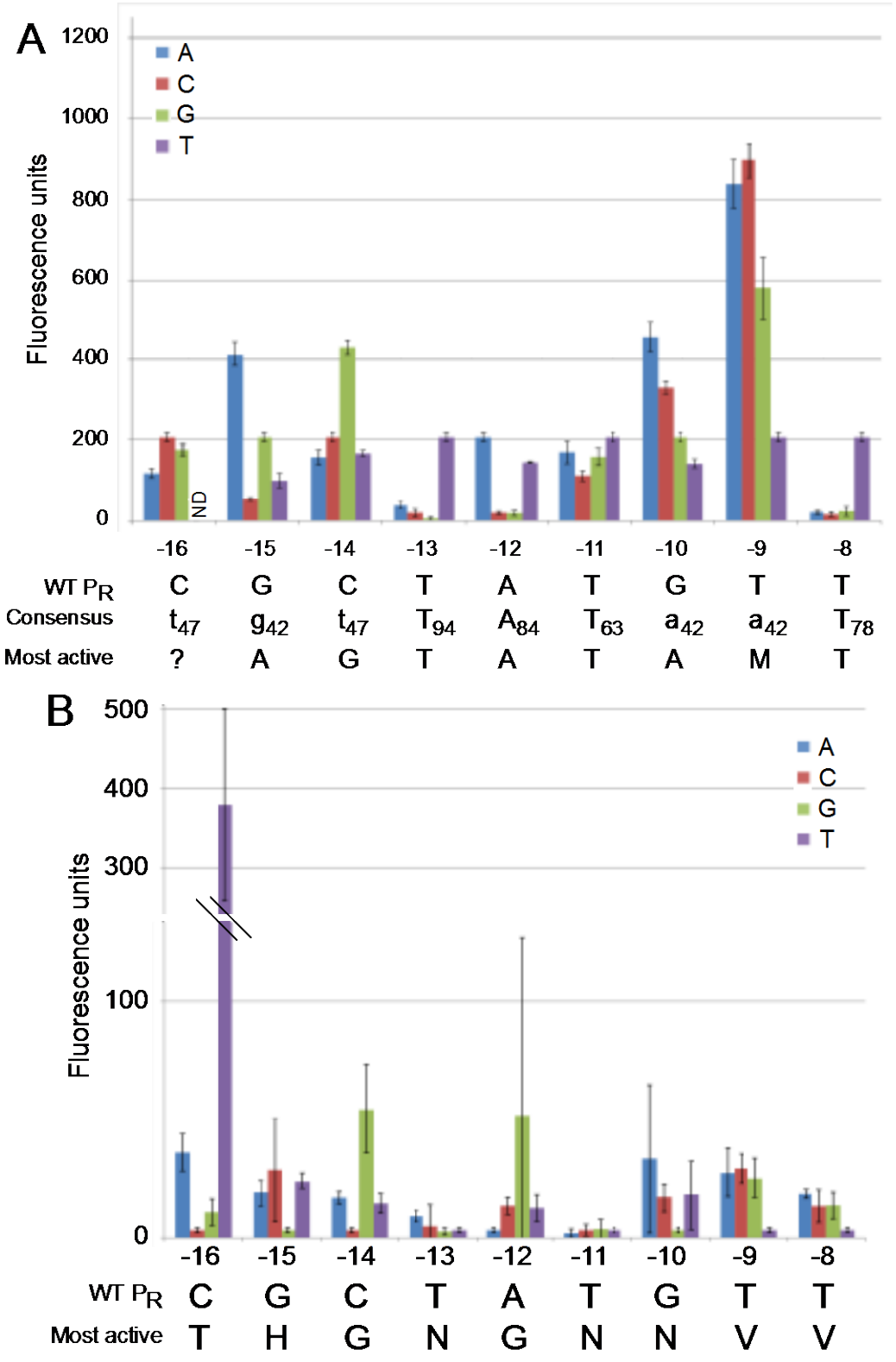

Figure 4-2. Complete mutational analysis extended -10 and -10 motifs of P_R.

Fig. 4-2. (A) The fluorescence of P_R-mCherry transcriptional fusions in *M. smegmatis* mc^2155 containing single point mutations in the extended -10 (positions -16 to -14) and the -10 hexamer (positions -13 to -8) of P_R were quantified. The wild-type sequence of P_R, the consensus bases with their respective frequencies (Gomez and Smith, 2000) and the substitution we determined to be most active in P_R are included below. ND: not determined. (B) The fluorescence of same P_R-mCherry transcriptional fusions as (A) in a BPs lysogen, *M. smegmatis* mc^2155(BPs). The wild-type P_R sequence and most active nucleotide are included below. Fluorescence units are $(LAU/mm^2)/OD_{595nm}$ and the data are represented as mean ± 95% confidence interval. N= any base; M= A or C; H= A, C, or T; V= A,C or G.

hypothesize that the C-16T mutation strongly enhances promoter activity by generating an extended -10 sequence. We were able to transform the C-16T mutant into a BPs lysogen and it has substantial activity (Fig. 4-2B), consistent with this interpretation (the C-16T substitution also alters O_R, but this does not itself substantially impair repression; see below).

4.4 INFLUENCE OF SINGLE BASE SUBSTITUTIONS IN THE -35 MOTIF ON P_R ACTIVITY

To determine the role of individual base pairs in the -35 region of P_R (positions -32 – -37; Fig. 1B), we constructed a similar set of mutant derivatives in which each base in the -35 hexamer was changed and the promoter activities determined (Fig. 4-3A). At the -33 position none of the substitutions influenced activity, and at -32 replacement of the A with C had no effect, and A-32G and A-32T mutations modestly reduced activity (Fig. 4-3A). In contrast, at positions -34, -35, -36 and -37, several of the substitutions increased activity, some substantially so, especially substitutions of G at -34 and -35 (Fig. 4-3A). Activity is also increased in the C-34A mutant, which corresponds to the consensus sequence (Fig. 4-3A). Although T is well conserved at PR -

36 and -37, it is somewhat surprising that either C or G substitutions increase promoter activity at these positions. A prior mutagenesis study on promoter activity in mycobacteria demonstrated that any substitution at these positions resulted in around a 2-fold decrease in transcription [64]. The bases within the -35 hexamer of P_R that contribute to optimal promoter strength (5'SSGGCA) thus differ from both the mycobacterial SigA consensus - sequence (5'-TTGACA) and the most active sequence 5'-TTGCGA derived by mutagenesis [64]. It seems unlikely that P_R uses a sigma factor other than SigA, with which mutations in the P_R -10 conform, since promoters described for other sigma factors differ greatly from the SigA consensus and none have an AT-rich -10 [[94], see Chapter 1].

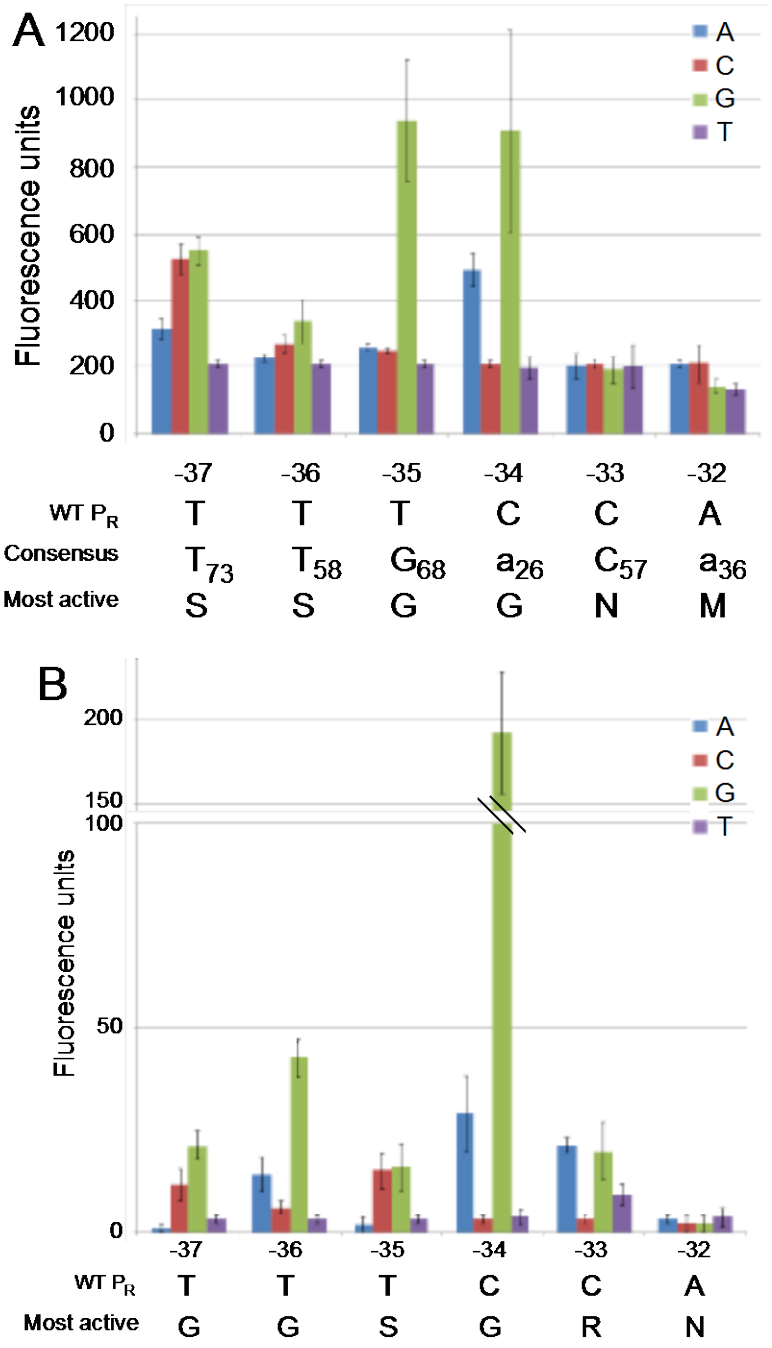

Figure 4-3. Complete mutational analysis -35 hexamer of P_R.

Fig. 4-3. (A) The fluorescence of P_R-mCherry transcriptional fusions in *M. smegmatis* mc^2155 with single base substitutions in the -35 hexamer (-37 to -32) were quantified. The wild-type sequence of P_R, the consensus bases with their respective frequencies [72] and the most active substitution are included below. (B) The fluorescence of same P_R-mCherry transcriptional fusions as (A) in a BPs lysogen, *M. smegmatis* mc^2155(BPs). The wild-type P_R sequence and most active nucleotide are included below. Fluorescence units are $(LAU/mm^2)/OD_{595nm}$ and the data are represented as mean ± 95% confidence interval. N= any base; S= C or G; R= A or G.

4.5 ROLE OF SPACER LENGTH ON P_R ACTIVITY

Promoters are strongly influenced by interhexamer spacing, and we therefore examined the effect of this spacing on P_R activity. A study of mycobacterial SigA promoters found that 70% had spacer regions between 16 and 18bp in length [72]. A previous mutational analysis found that optimal activity was achieved with an 18bp spacer [64]. The interhexamer spacing in P_R is 18 bp (Fig. 1B), and we altered this by inserting a base (T between G-28 and A-27) to increase it to 19 bp, or deleting either one (ΔA-29) or two bases (ΔA-29/ΔG-28) to decrease it to 17 bp and 16 bp, respectively (Fig. 4-4A). These mutations were located in a region of the spacer near the -35 hexamer and outside of the O_R (Fig. 4-4A).

We find that a spacing of 17 bp is optimal for P_R activity, and is somewhat more active than the 18 bp wild-type spacer (Fig. 4-4B). We also tested P_R activity in a strain lysogenic for BPs, in which wild-type PR is tightly down-regulated by the repressor. Although the 17 bp spacer mutant has elevated activity in the absence of repressor, there is substantial de-repression in the lysogen. Because the spacer mutation lies outside of the O_R operator (Fig. 1B), it is not expected to influence binding of repressor directly [31], and poor repression presumably is the consequence of the altered configuration of O_R and P_R. We note that poor repression also is seen in the 16 bp spacer mutant, and normal repression is observed in the 19 bp mutant (Fig. 3), indicating that removing bases from the spacer had a deleterious effect on repression but the insertion of a base did not. Thus the architecture of the P_R promoter presumably represents a compromise between promoter strength in the absence of repressor, and the ability to tightly repress P_R in the prophage.

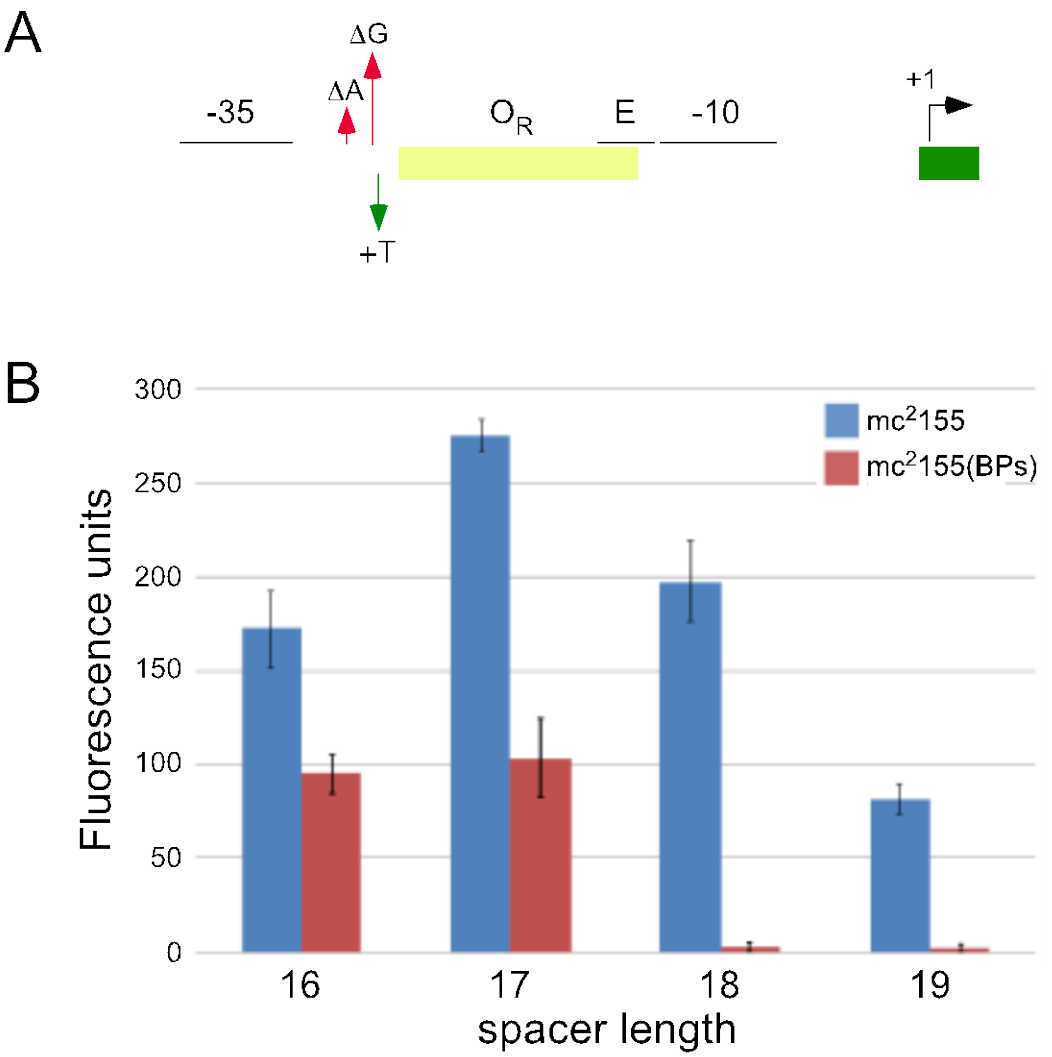

Figure 4-4. Effect of spacer length on promoter activity.

Fig. 4-4 (A) The sequence of P_R (BPs coordinates 29470-29512) with locations of the spacer deletion and insertion mutations. ΔA-29, ΔG-28 and the T insertion between G-28 and A-27 were utilized to create spacers ranging from 16 to 19 bp in length. Noted are the -10 and -35 regions (blue letters), the extended -10 (E), O_R (yellow box), and the translation start codon (green box) and transcription start site (+1 arrow) of BPs *34*. (B) The activities of the spacer length mutants were quantified in wild-type *M. smegmatis* mc^2155 and a BPs lysogen of *M. smegmatis* mc^2155(BPs). Fluorescence units are (LAU/mm^2)/OD_{595nm} and the data are represented as mean ± 95% confidence interval.

4.6 INFLUENCE OF SINGLE BASE SUBSTITUTIONS IN THE SPACER REGION ON P_R ACTIVITY

The spacer sequence is known to have effects on the activity of a promoter in other prokaryotes [172-174]. However, a previous study in mycobacteria demonstrated that replacing the GC-rich spacer around the -15 area with an AT-rich sequence, which increased activity in *E. coli* ~15-fold, did not change the activity in mycobacteria [64]. We constructed a series of single base mutations in the 12 bp O_R region, which is located between the -35 and -10 motifs and overlaps the extended -10 (-26 — -15; Fig 1B) and examined their effects on P_R promoter activity and repression in a BPs lysogen. The effects on promoter activity of substitutions from -26 to -17 were modest, less than two-fold increase or decrease (Fig. 4-5). Mutations in the two positions that are also part of the extended -10 motif (-16 to -14) and the O_R, C-16 and G-15, give the most extreme changes in promoter activity with G-15A having the highest activity of any of these mutations and G-15C having the lowest (Fig. 4-5). The modest effect of the spacer mutants shows that while they do affect P_R activity, no single base change in the spacer is able to dramatically increase promoter activity.

As previously stated, we were not able to transform the C-16T mutant into *M. smegmatis* but we were able to transform the mutant into a BPs lysogen of *M. smegmatis*. The activity of C-16T in a lysogen is approximately 2-fold higher than wild-type P_R in a non-lysogen and 95-fold higher than P_R in the BPs lysogen. The remarkably high activity of C-16T in a BPs lysogen we hypothesize is not due to complete de-repression but is due to very high activity of a non-repressed C-16T promoter and is also supported by our inability to transform C-16T into a non-lysogen.

None of the single mutations we examined in O_R were able to completely alleviate repression of P_R in a lysogen (Fig. 4-5A), but most did de-repress P_R to some extent (Fig. 4-5B). With the exception of C-26G, C-26T and T-21A, all of the mutations relieved repression by at least 2-fold (Fig. 4-5B). This supports biochemical data that provide evidence that the 12 bp O_R palindrome is the binding site for the BPs repressor, gp33 [31]. The ability of O_R to retain repression of P_R in the presence of single base substitutions demonstrates that there is some flexibility in the sequence of O_R. Also, the lack of substitutions that repress O_R more tightly than the wild-type sequence indicates selective pressure to maintain tight control of this promoter in the BPs lysogen.

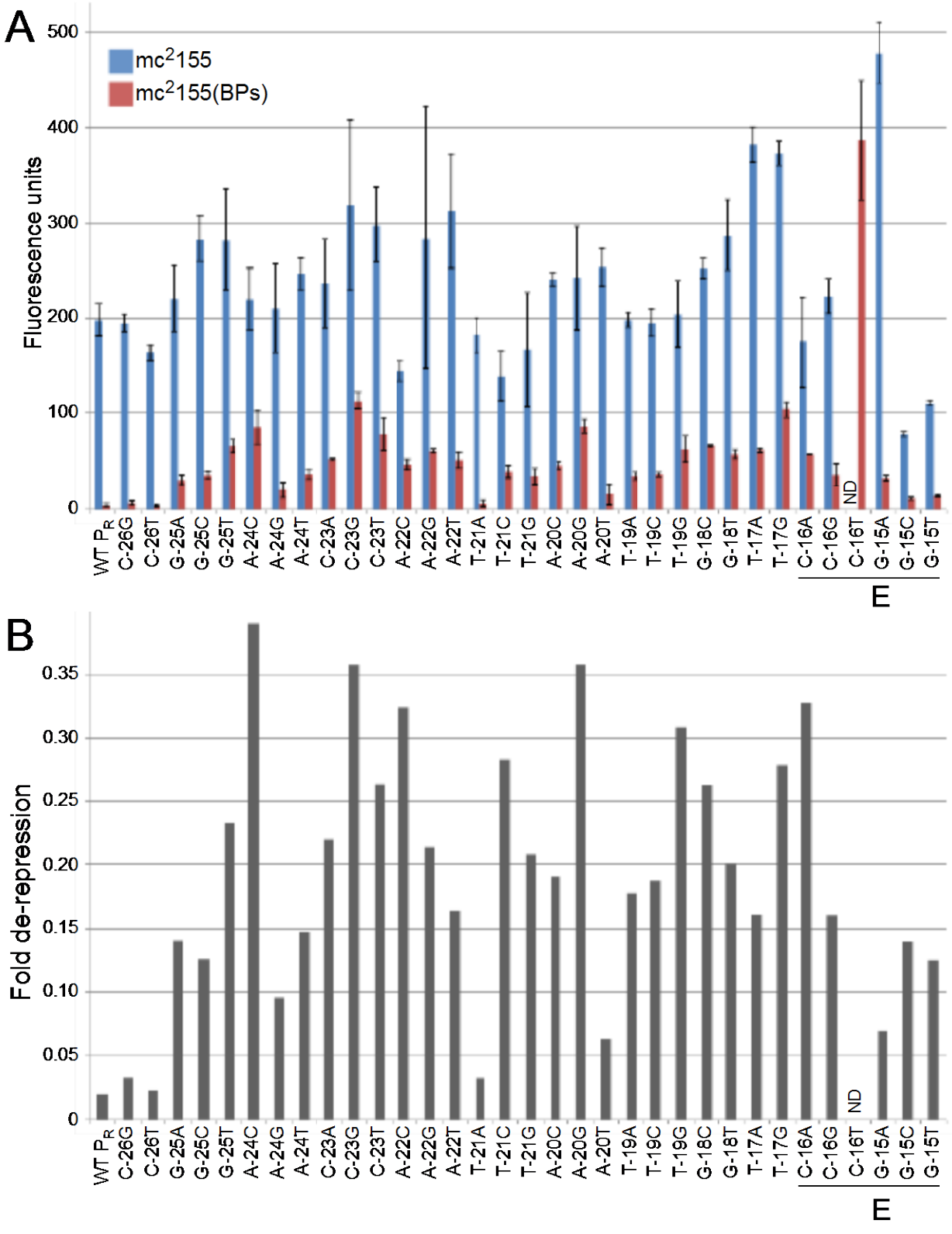

Figure 4-5. Effect of single base pair substitutions in the spacer region on promoter activity and de-repression of P_R in a BPs lysogen.

Fig. 4-5. (A) The fluorescence of P_R-mCherry transcriptional fusions in *M. smegmatis* mc^2155 and *M. smegmatis* mc^2155(BPs) containing single base substitutions in the spacer region, namely O_R (BPs coordinates 29484-29495). The bases from O_R that overlap with the extended -10 motif are noted (E). Fluorescence units are $(LAU/mm^2)/OD_{595nm}$ and the data are represented as mean ± 95% confidence interval. (B) The amount of de-repression caused by the substitutions from (A) was calculated by taking the reciprocal of ratio of activity in mc^2155 to mc^2155(BPs). The bases from O_R that overlap with the extended -10 motif are noted (E). ND= not determined.

4.7 COMBINATORIAL EFFECTS OF P_R MUTATIONS

The behaviors of single base substitution mutants are consistent with the P_R promoter being recognized by SigA, but with base changes that reflect the particular context of P_R. Because base substitutions in both the -35 and -10 hexamers and changes in spacing can result in increases in promoter activity, we constructed several series of mutants with different combinations of mutations to achieve the strongest promoter possible in the P_R context. However, because of the concern that the T substitution as the -16 position generated a non-transformable phenotype, hypothetically due to strong promoter activity, and because combinations of mutations that increase promoter strength could yield a similar phenotype, these mutant combinations were constructed in a vector that integrates into the *M. smegmatis* chromosome at the *attB* site used by phage Tweety and provides a single copy of the vector [169].

First, we tested whether the C-16T substitution in an integrated vector is able to be transformed into *M. smegmatis* mc^2155. Integrating plasmid pLO76 (C-16T) efficiently transforms, has activity that is nearly 15-fold greater than wild-type P_R, and is fully repressed in a BPs lysogen indicating that in the extrachromosomal context PR activity would be at least 2-fold higher than the hsp60 promoter without additional activation due to supercoiling. Thus even though C-16 is within O_R, the mutant tolerates repressor binding (Fig. 4-6). The C-16T

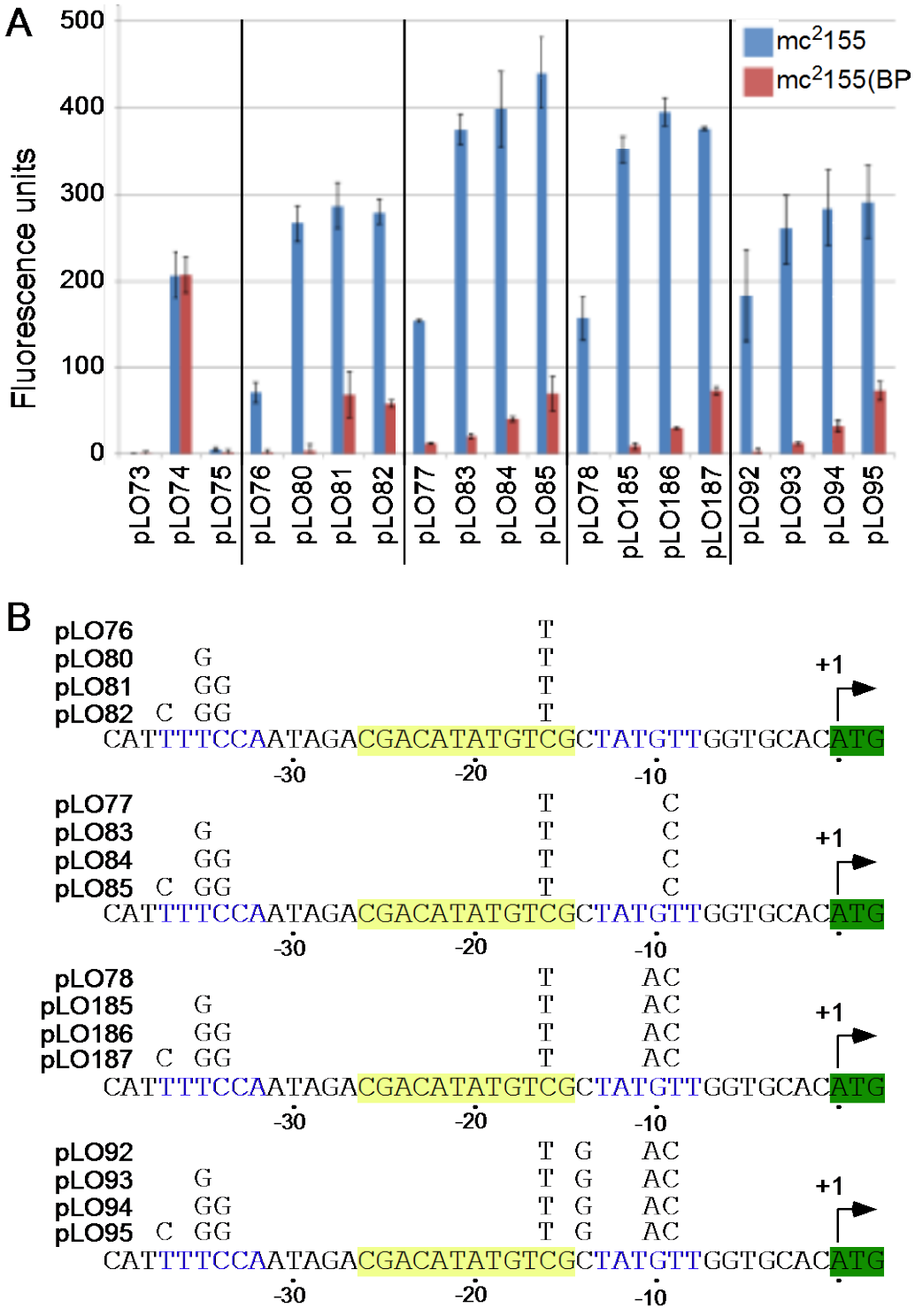

Figure 4-6. Activity of combinatorial mutations in P_R in a Tweety integrative vector.

Fig 4-6. (A) The activity of combinations of P_R mutations driving mCherry expression in *M. smegmatis* mc^2155 and a BPs lysogen of *M. smegmatis* mc^2155(BPs) compared to the activities of a promoterless vector (pLO73), the hsp60 promoter (pLO74), and wild-type P_R (pLO75). The mutants are arranged in four groups. Fluorescence units are $(LAU/mm^2)/OD_{595nm}$ and the data are represented as mean ± 95% confidence interval. (B) The mutations contained in each plasmid and their positions are indicated. The mutants are arranged in four groups, where the -10 mutations are common between all members and consecutive -35 mutations are added subsequently (no -35 mutation; T-35G; T-35G and C-34G; T-35G, C-34G and T-37C). The -10 mutations are C-16T (pLO76), C-16T and T-9C (pLO77), C-16T, T-9C and G-10A (pLO78) and C-16T, T-9C, G-10A and C-14G (pLO92). The sequence of P_R (BPs coordinates 29470-29512), and locations of the -35 and -10 motifs (blue letters), O_R (yellow box) and transcription (+1 arrow) and translation (gene box) start sites are included.

substitution generates an extended -10 type promoter (i.e. 5'-TGC). The C-16T mutant is still about 3–fold lower in activity than the hsp60 promoter (Fig. 4-6).

We then sequentially introduced additional -10 hexamer mutations that enhanced activity as single base substitutions (Fig. 4-2A). Addition of the T-9C substitution (pLO77) gives a further two-fold increase in activity (Fig. 4-6) in agreement with the single substitution observations (Fig. 4-2), and small additional increases in activity resulted from sequential addition of the G-10A and C-14G mutations (pLO78 and pLO92 respectively, Fig. 4-6), achieving levels of activity close to 90% of the hsp60 promoter.

Although it has been reported previously that -10 extended promoters are relatively insensitive to substitutions in the -35 region [64,71], we tested the influence of -35 substitutions on the C-16T mutant (Fig. 4-6). Addition of a C-35G substitution – which gives a substantial increase in P_R activity as a single base change (Fig. 4-2) – increases activity of the C-16T mutant over three-fold, and above that of hsp60 (Fig. 4-6). Addition of the C-34G and then T-37C gave only small further increases in function (Fig. 4-6). We then tested the impacts of these -35 substitutions in the context of more complex substitutions in the -10 hexamer, with a purpose to further elevating P_R activity. All of these combinations of -10 and -35 substitutions gave

substantial increases in activities that are greater than hsp60 (see Fig. 4-6), with the optimal being a combination of C-16T, T-9C, T-35G, C-34G, T-37C (i.e. pLO85) (Fig. 4-6). These data show that a version of the P_R promoter with an extended -10 (i.e. 5-'TGC) is sensitive to -35 substitutions, and that although combining mutations can have additive effectives on the single substitutions alone, this is not always observed. For example, pLO95 that differs from pLO85 by also having additional -10 mutations, C-14G and G-10A, each of which individually gives about a two-fold increase in P_R activity (Fig. 4-2A), has lower activity than pLO85 (Fig. 4-6). Removing a base to decrease the interhexameric spacer to 17 bp did not have a substantial impact on any of the multiple -10/-35 mutants that we tested (Fig. 4-7). Finally, we note that repression in a BPs lysogen is compromised in many of the mutants, with the T-37C substitution, located furthest from the operator, unexpectedly having the greatest impact (Fig. 4-6). It is possible that the effect on repression may be due to the combination of mutations and other combinations of -35 mutations were not tested.

The difficulty in predicting the best overall combination of mutations for P_R provides further evidence for context dependence for promoter activity at P_R. Even with detailed knowledge of the effects of base changes at all locations within the -35 and -10 motifs, simply adding all of the mutations that enhance strength at each position is not sufficient to create the strongest P_R promoter sequence. The changes made at other positions change the effects of the tested substitutions made in this new promoter context.

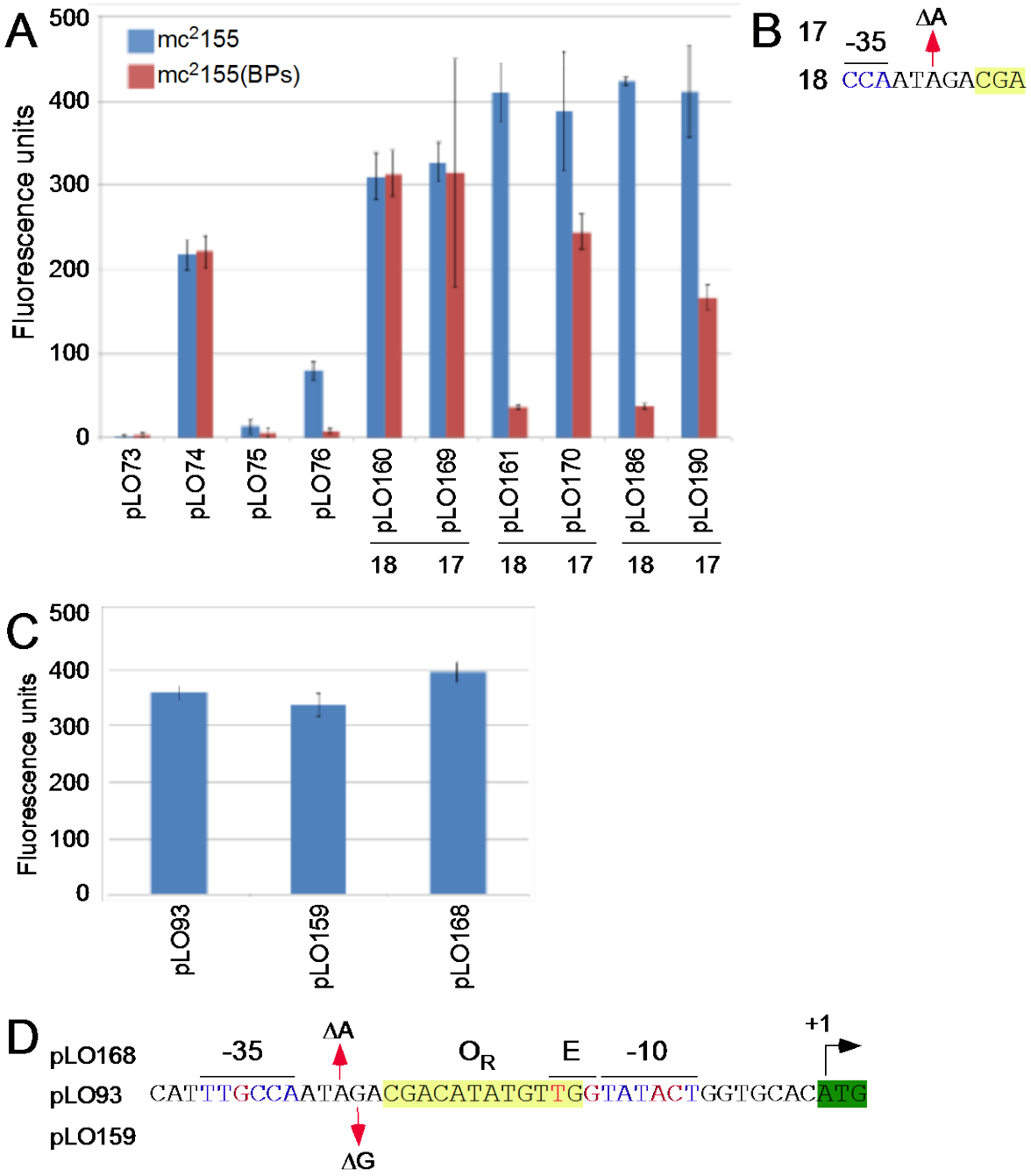

Figure 4-7. The effect of a single base deletion in the spacer region on combinatorial P_R mutants.

Fig. 4-7. (A) A comparison of the fluorescence of P_R-mCherry transcriptional fusions in *M. smegmatis* mc^{2}155 and *M. smegmatis* mc^{2}155(BPs) containing equivalent series promoter-up mutations and spacer regions of 18 or 17 bp (indicated below the graph). The activities of a promoterless vector (pLO73), the hsp60 promoter (pLO74), wild-type P_R (pLO75) and the C-16T (pLO76) mutant are included. Fluorescence units are (LAU/mm^2)/OD_{595nm} and the data are

represented as mean ± 95% confidence interval. (B) The sequence surrounding the spacer deletion (BPs coordinates 29476-29486) showing the deletion of A-29 which creates a 17 bp spacer in pLO169, pLO170 and pLO190. (C) The fluorescence of a strong P_R combination mutant with a wild-type 18 bp spacer (pLO93) and equivalent mutants also containing 17 bp spacers by either deletion of G-28 (pLO159) or A-29 (pLO168) in *M. smegmatis* mc^2155. Fluorescence units are $(LAU/mm^2)/OD_{595nm}$ and the data are represented as mean ± 95% confidence interval. (D) The sequence of P_R (BPs coordinates 29470-29512) with locations of the spacer deletions ΔA-29 and ΔG-28, found in pLO168 and pLO159, respectively and the mutations found in the promoter are shown in red. Noted are the -10 and -35 regions (blue letters), the extended -10 (E), O_R (yellow box), and the translation start codon (green box) and transcription start site (+1 arrow) of BPs *34*.

4.7.1 Additional evidence for context dependence of P_R

While the -10 single base substitutions generally fit well with the predicted SigA consensus, mutations in the -35 motif was more complicated (Fig 4-2 and 4-3). One minor difference was the preference for a G at position -34 but the consensus predicts an A this location, albeit at a low frequency. We decided to test the function of C-34A and made this mutation in promoters with additional mutations. We note that the strongest mutation C-34G did not provide significantly increase activity when added to a promoter already containing T-35G (Fig 4-8). However, addition of C-34A (pLO161) to pLO93, which contains T-35G, we did get a small (~20%) increase in promoter strength (Fig).

Also, the single base substitutions in the -35 motif of P_R surprisingly showed that C and G gave the highest activity substitutions at the -37 and -36 positions, as previously noted (Fig4-3). To test if this held true in a different promoter context, we compared added the T-37C and T-36G mutations to a promoter that already had mutations in the -10 and -35. The new mutant has a -35 sequence of 5'CGGACA. The promoter containing these two mutations at -37 and -36 (pLO162) was significantly less active than the analogous promoter without these mutations (pLO161), with the sequence of 5'TTGACA (Fig 4-8).

The activities of these mutants lend additional evidence that the activity of mutations in P_R are context dependent.

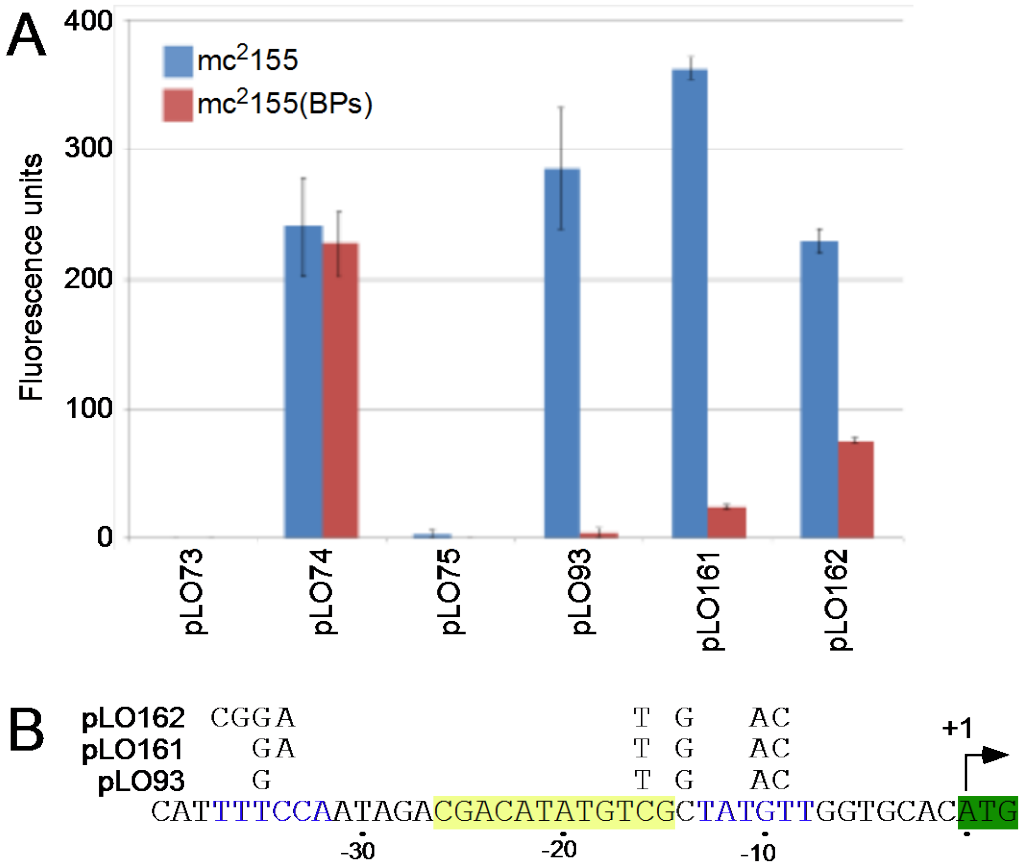

Figure 4-8. The effect of additional -35 mutations that mimic the consensus or most active single base substitutions.

Fig. 4-8. (A) The fluorescence of a strong P_R combination mutant (pLO93) compared a mutant with the addition of C-34A (pLO161), which mimics the consensus sequence, and the addition of C-34A, T-37C and T-36G (pLO162). (B) The sequence of P_R (BPs coordinates 29470-29512) with locations and identity of these substitutions for pLO93, pLO161 and pLO162. The activities of a promoterless vector (pLO73), the hsp60 promoter (pLO74), and wild-type P_R (pLO75) are included. Fluorescence units are $(LAU/mm^2)/OD_{595nm}$ and the data are represented as mean ± 95% confidence interval.

4.8 COMPARATIVE EXPRESSION IN *E. COLI*

A subset of the promoter mutants we have described were tested for activity in *E. coli*. Though the mycobacterial SigA consensus sequence is similar to the *E. coli* Sig70 consensus, the sequence requirements in mycobacteria appear to be relaxed in comparison and non-conformity better tolerated. Many previous studies have noted that mycobacterial promoter have limited or no function in *E. coli* [38,175]. For this experiment, we used the promoters with combinatorial mutations in the Tweety integrative vector, which contains an origin of replication for *E. coli*. We found that the native mycobacterial promoters P_{hsp60}, our control throughout these experiments for strong mycobacterial activity, and P_R gave low activity in *E. coli* (Fig 4-9). The mutations in the -10 and extended -10 (pLO76, pLO77, pLO78 and pLO92), which gave increasing activity in *M. smegmatis* (Fig 4-2), are largely ineffective in raising the activity of P_R in *E. coli* (Fig 4-9). Both mutations in the extended -10 motif are detrimental to promoter activity, though *E. coli* promoters also have extended -10 sequences [65]. The -10 mutation that has the largest effect is G-10A, which is the only of the mutations in the -10 which substitutes a nonconserved base for a base Sig70 consensus sequence [60]. As with mycobacteria, the addition of T-35G alone in the -35 (pLO93) augments promoter activity, as does C-34A alone (pLO160) and additional mutations of T-34G and T-37C (pLO94 and pLO95) do not enhance promoter strength (Fig 4-9). In mycobacteria, we found that adjusting the spacer length to 17 bp in combination with many of these other mutations did not increase promoter strength (Fig 4-7). In contrast, the deletion of a nucleotide in the spacer of pLO93 results in a 2-fold increase in promoter activity (pLO168 and pLO159, Fig 4-9). The effect of T-35G and C-34A together is another departure from the activities in mycobacteria. The addition of C-34A (pLO161) results in a 3-fold increase in promoter strength in *E. coli* (Fig 4-9), while the same mutation in *M.*

smegmatis gives only a slight increase in activity (data not shown). This promoter with the sequence of 5'TTGACA at the -35 and 5'TATACT at the -10 is the strongest in *E. coli* and though it is also strong in *M. smegmatis* many other promoters are equally active. The activity of pLO161 is not improved by shortening the spacer to 17 bp (pLO170; Fig 4-9).

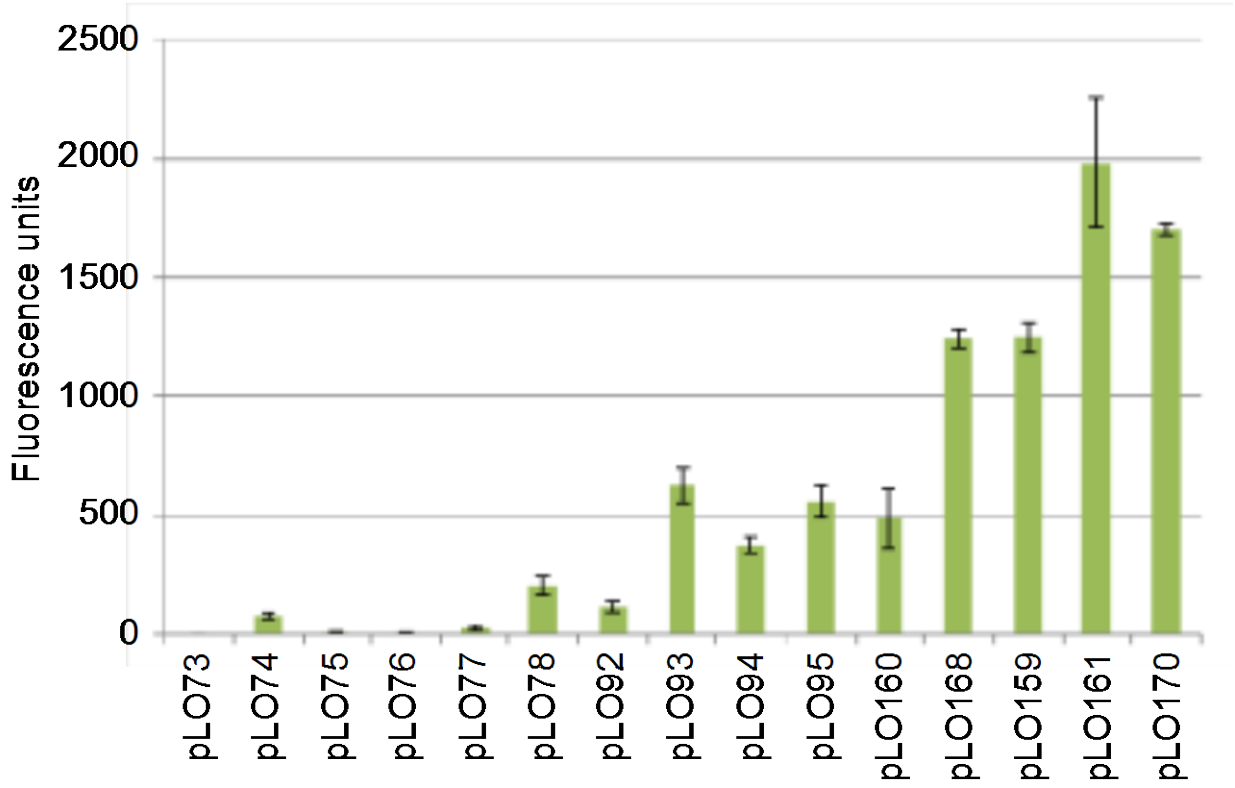

Figure 4-9. The activities of a subset of the combinatorial mutants in *E. coli*.

Fig. 4-9. The fluorescence of wild-type P_R (pLO75) and a number of the combinatorial mutants generated and tested in *E. coli*. Fluorescence units are $(LAU/mm^2)/OD_{595nm}$ and the data are represented as mean ± 95% confidence interval.

4.9 A CALIBRATED SERIES OF *M. SMEGMATIS* PROMOTERS

The large number of promoter substitutions in P_R generated represents a pool of promoters with differing strengths in *M. smegmatis* mc^2155 and varying degrees of repression in a BPs lysogen. From among these we selected six promoters, along with wild-type P_R, that span a range of activity barely greater than the promoterless vector (pLO75; wild-type P_R) to about 2-fold stronger than hsp60 (Fig. 4-10), and which are tightly repressed in the BPs lysogen. When tested together, they show a gradual increase in activities with each contributing approximately one-fifth higher activity to the promoter previous to it in the series (Fig. 4-6). The behaviors of these promoters are illustrated in the colors of the *M. smegmatis* colonies reflecting expression of the mCherry reporter gene (Fig. 4-10).

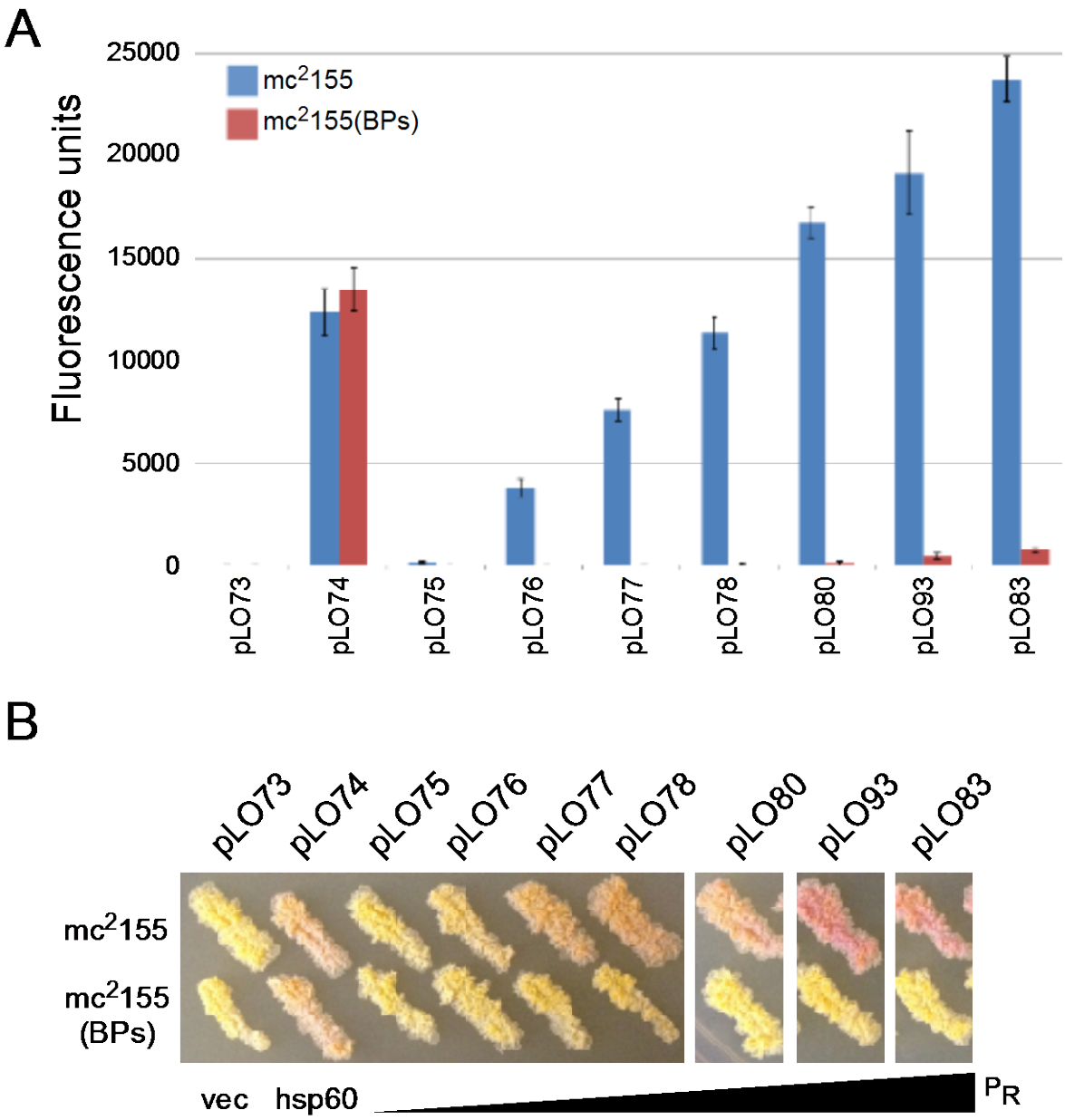

Figure 4-10. Collection of calibrated promoters for *M. smegmatis.*

Fig. 4-10. (A) The activity of a subset of the P_R derivative promoters tested in *M. smegmatis* mc^{2}155 and a BPs lysogen, *M. smegmatis* mc^{2}155(BPs). Fluorescence units are (LAU/mm^2)/OD_{595nm} and the data are represented as mean ± 95% confidence interval. (B) Visual comparison of the integrated mCherry transcriptional fusions shows differences in promoter strength by variation in the colony color in *M. smegmatis* mc^{2}155 and *M. smegmatis* mc^{2}155(BPs) compared to a promoterless vector (pLO73), the hsp60 promoter (pLO74) and wild-type P_R (pLO75).

4.10 A CALIBRATED SERIES OF *M. TUBERCULOSIS* PROMOTERS

We also tested a subset of the promoter mutants in M. tuberculosis $mc^2$7000, including all those shown in Figure 5 and some of those in Figure 4-11. We used three methods to observe the fluorescence in these strains, a quantitative measure of fluorescence from liquid cultures (Fig. 4-11), fluorescence microscopy (Fig. 4-11) and the visual colors of colonies (Fig. 4-11). The general trend of the mutants tested in *M. smegmatis* (Fig. 4-10) is the same in *M. tuberculosis* (Fig. 4-11), although none is greater than the hsp60 promoter using the quantitative assay on liquid cultures (Fig. 4-11). This may be due to clumping of the bacterial cultures in growth and preparation, which would affect the cell density measurement that is a component of the fluorescence units reported. The promoters with spacing changes as well as additional hexamer mutations (pLO159, pLO168 and pLO170) tested do not have substantially better activity than pLO83 and pLO93 in the quantitative assay (Fig. 4-10), as noted in *M. smegmatis* (Fig 4-10). It was observed that several of the P_R mutants appear as bright as hsp60 by fluorescence microscopy, notably all of the strains above and including pLO83 in the series (Fig. 4-11). Also, the colony colors appear as dark or even darker than hsp60 for some mutants (Fig. 4-11). The quantitative assay of mCherry in *M. tuberculosis* may under represent promoter strength. However, although the activities of this series of promoters may differ according to the assay used, the strength relative to each other generally reflects activities in *M. smegmatis* (Fig. 4-11).

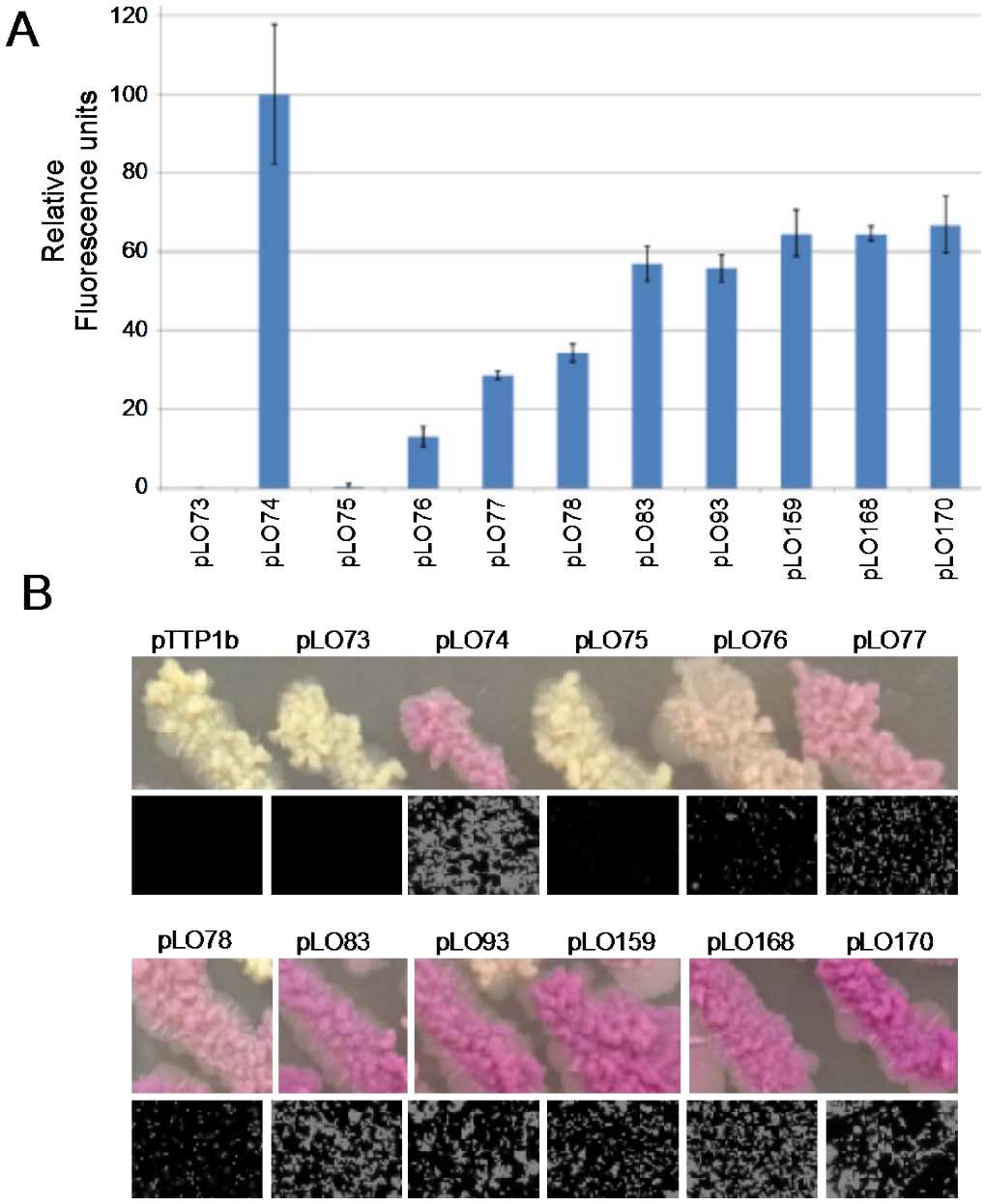

Figure 4-11. Collection of calibrated promoters for *M. tuberculosis.*

Fig. 4-11. (A) Fluorescence assay with fixed *M. tuberculosis* mc^{2}7000 strains carrying the integrative Tweety vector. The fluorescence is reported in relative fluorescence units, where the expression from the hsp60 promoter has been set to 100. The P_R mutants used include those from Fig. 5 and two with 17 bp spacers that contain the equivalent mutations as pLO93 (pLO159 and pLO168) and pLO170, which is equivalent to pLO168, plus C-34A. Fluorescence units are (LAU/mm^2)/OD_{595nm} and the data are represented as mean ± 95% confidence interval. (B) Visual comparison of *M tuberculosis* mc^{2}7000 colony colors and fluorescence microscopy of fixed *M. tuberculosis* strains carrying integrative plasmids with promoter-mCherry transcriptional fusions. Fluorescence microscopy images exposed for 100 ms. pTTP1b is the Tweety integrative vector with no promoter and no mCherry gene.

4.11 DISCUSSION

Although mycobacteria encode many sigma factors, the mutational analysis we performed strongly suggests that SigA, the housekeeping sigma factor in *M. smegmatis* and *M. tuberculosis*, recognizes BPs P_R. Substitutions in the -10 hexamer are highly detrimental at positions -13, -12 and -8, which correspond to the well conserved bases 5'-TANNNT. Only SigA promoters follow this consensus, and SigA promoters contain the most AT-rich -10 regions [75,94]. To our knowledge, this is the first complete mutagenesis of the -10 hexamer of a mycobacterial promoter, and it is of interest that the bases at the -9 and-10 positions have considerable variations in activity. Other reported mutagenesis studies of equivalent positions in mycobacterial promoters indicate that these positions are important for promoter activity but the effects of particular bases is not consistent with the results for P_R, demonstrating the effect of sequence context [61,176]. These positions may thus play a prominent role is modulating the activity of individual promoters in M. smegmatis and M. tuberculosis. The greatest sequence variation allowed is at the -11 position, where although T has strongest activity, function is reduced by less than two-fold by any of the substitutions.

The wild-type P_R promoter does not contain an extended -10 5'-TGN motif, but it can be constructed by addition of the C-16T substitution. However, an extrachromosomal vector carrying this mutation is not transformable into wild-type *M. smegmatis* mc^2155 but is able to be transformed into *M. smegmatis* mc^2155(BPs). The extremely elevated activity of the extrachromosomal vector carrying C-16T in a BPs lysogen suggests that the mutation results in very high promoter activity or complete de-repression of P_R. When the C-16T P_R-mCherry fusion is tested in an integrative context, the C-16T promoter shows elevated activity compared to wild-type P_R but is approximately 3-fold less active than the hsp60 promoter. C-16T is fully

repressed in a BPs lysogen when integrated, and therefore it does not appear that high promoter activity or de-repression can explain the inability to transform into *M. smegmatis.* However, if the C-16T P_R mutant is also activated preferentially in the extrachromosomal context, like the wild-type promoter which is 43-fold more active when expressed extrachromosomally (Fig. 4-1) though the copy number difference is only 23:1 [170], this would account for the lower than expected activity of C-16T when integrated and the inability to transform when extrachromosomal. Activation of C-16T by 43-fold would produce a promoter slightly more than 2-fold stronger than the hsp60 promoter, which may be sufficient to interfere with plasmid replication. Unfortunately, an inducible system for P_R expression to test this is not yet available.

The P_R promoter is sensitive to base changes in the -35 hexamer in both the wild-type context and in combination with the extended -10 C-16T substitution, contrary to the assertions made in some prior studies that the -35 does not play a role in promoter strength when an extended -10 motif is present [61,71]. Prior mutational analysis of the -35 hexamer of a synthetic promoter showed that the sequence 5'-TTGCGA represents the bases at each position giving optimal expression [64]. Dissection of the P_R promoter shows a distinctly different set of preferences, with strong activity represented by individual mutations towards the sequence 5'-SSGGNM (where S= G or C; M= A or C). The preference for C or G at -37 and -36 where T is well conserved was surprising, however, the addition of these mutations (T-37C and T-36G) to a strong P_R mutant promoter, decreased its activity (Fig. 4-8). This demonstrates context dependence for the effect of substitutions made in the -35 hexamer. Additionally, a previous study found that mutations to single substitutions to C at positions equivalent to -37 and -36 were not detrimental to promoter activity, lending support to our observation that in an otherwise weak SigA promoter [61].

The length and sequence of the spacer region between the -35 and -10 hexamers does affect the promoter activity of P_R, although with more modest changes than mutations in the -35 and -10 regions. The length of the spacer sequence is important for the proper configuration to allow contact of the sigma factor with the -35 and -10 simultaneously [177]. Interhexameric spacing in mycobacterial promoters is 16 to 19 bp, with optimal activity at 18 bp [64]. We found that an interhexameric spacer of 17 bp gave optimal P_R activity, however when added to strong P_R mutants, no activity was gained with the 17 bp spacer. The dependence of promoter strength on the sequence of spacer region has been demonstrated in other prokaryotes [172-174] but not tested in mycobacteria. This may be a function of SigA, as it was shown that Sig70 mediates the spacer sequence effects in *E. coli* [173].

Utilizing the mutational data we collected from the single base substitutions, we generated promoters with combinations of substitutions, and were able to create several promoters with activity greater than the hsp60 promoter in an integrated context. The strongest of these are approximately 2-fold more active than hsp60, which may be approaching the maximal promoter strength achievable, as additional mutations do not confer additional promoter activity. The strongest P_R mutants, when integrated and in single copy, have about one third the activity of the extrachromosomal hsp60 promoter, which is present in 23 copies [170]. For the construction of strains where very high levels of expression of heterologous genes is necessary, for example in recombinant vaccine strains, multiple integrative vectors can be used to increase gene expression levels [170]. Many phage-derived integrative vectors, which integrate at different chromosomal locations, have been constructed for use in mycobacteria [10,31,169,170,178,179].

The activities of the combinatorial mutants did not strictly adhere to our predictions. The activities of strong P_R mutants were not enhanced by additional mutations that were shown to increase promoter strength, such as the C-34G substitution (Fig. 4-6) in the -35 and altering the spacer length (Fig. 4-7). This may indicate that we have achieved the maximal promoter activity that P_R is capable of. The lack of enhanced activity with these mutants may also be a function of the dependence on the surrounding sequence. We found some substitutions that increased activity when they were the sole mutation (Fig. 4-2 and Fig. 4-3), but decreased activity when placed in a strong P_R mutant (Fig. 4-8), namely T-73C and T-36G.

As has been noted in other studies, the native mycobacterial promoters hsp60 and P_R were not active in *E. coli*. However, the P_R mutants that followed the *E. coli* Sig70 consensus sequence had dramatically increased activity. In contrast to the results from *M. smegmatis*, addition of mutations in the spacer length (to 17 bp) and of mutation C-34A were each able to double promoter activity. However, the strongest P_R mutant has a spacer of 18 bp (pLO161; Fig. 4-9) and its equivalent P_R mutant with 17 bp is weaker, demonstrating that *E. coli* promoter strength may also be context dependent.

From the many P_R mutants we have generated, a collection of promoters with varying strengths can be assembled. This library of promoters with calibrated activities can be used to fine-tune gene expression for experimental approaches where careful control of gene expression is necessary. Activity of the strongest promoters in *M. tuberculosis* do not reach the level of the hsp60 promoter in the quantitative assay, but the relative strengths of the P_R mutants and the hsp60 promoter is assay dependent. The visual and microscopy assays indicate that the activity of the P_R mutants is equal or greater than that of hsp60. The P_R mutants form a library of

synthetic promoters for use in mycobacteria, and add to the tools available for gene expression in bacterial systems, as similar libraries have been generated for other prokaryotes [165-168].

BPs P_R, in contrast to many other early lytic promoters, is not highly active in the absence of repressor. In an integrated context, which mimics the prophage, activity is barely detectable. The lower activity level likely leads to more complete repression than other analogous phage promoters, and this is supported by evidence that P_{left} in L5 is repressed to a similar fold change but activity of the repressed P_{left} is still 10-fold higher than a promoterless control [36]. Excision of the BPs prophage leads to synthesis of gp^{136} is proteolytically degraded and de-represses $P_{R,}$ and additionally, P_R is activated in the extrachromosomal state, further increasing its activity.

5.0 REGULATION OF THE BPS GENETIC SWITCH

5.1 AUTHOR CONTRIBUTIONS

This chapter is the result of collaboration with Gregory Broussard, PhD, Valerie Villanueva, Bryce Lunt and Emilee Shine within the Graham Hatfull laboratory, which has been published in *Molecular Cell* under the title "Integration-dependent bacteriophage immunity provides insights into the evolution of genetic switches." The s in this chapter are adapted from that publication and were prepared by all of the authors. To maintain transparency, a list of contributions of the authors are listed here. GB performed genetic experiments which provided the hypotheses for many of the experiments presented in this chapter, including the gain of function (GoF) mutants in Figure 5-2, the isolation and sequencing of clear plaque and repressor insensitive mutants in Figure 5-3, and the alignments of integrase protein sequences in Figure 5-4. The results of these experiments are provided as they directly explain the rationale for later experiments. The data and images in Figure 5-1, 5-2A, 5-2B, 5-3B, 5-3D, and 5-4A were prepared by GB with assistance from GH. The images and data for Figure 5-2C, 5-2D, 5-3A, 5-3C, 5-3E, 5-4B, 5-5, and 5-6 were prepared by LO with assistance from GH. No data from VV, BL or ES is presented in the Figures in this chapter and any references are properly cited.

5.2 INTRODUCTION

Regulating gene expression and, ultimately, the phenotype of many diverse organisms is a vital role performed by varied genetic switches. Investigations of bacteriophages, prokaryotes and eukaryotes have demonstrated the importance of genetic switches as a method of gene regulation [53,180,181]. The study of genetic switches also has implications for bioengineers attempting to design genetic circuits using basic biological principles [182,183]. The characteristic feature of a genetic switch is a stable shift in gene expression in response to a particular stimulus, leading to a change in phenotype. Due to their overwhelming number, temperate bacteriophages display the greatest potential diversity of mechanisms used to accomplish this gene expression shift.

The best characterized genetic switch is the lytic-lysogenic switch of bacteriophage λ, a temperate *E. coli* phage [53]. Briefly, upon entry into the bacterial cell, protease levels determine the concentration of protein CII, which, if present at high enough concentration, activates the establishment of lysogeny through initial synthesis of the *cI* repressor gene. The repressor, CI, shuts down the expression of lytic genes and maintains the lysogenic state. There are many other players involved in the λ genetic switch—including antiterminators Q and N; a competitive inhibitor of the protease, CIII; integrase and excise proteins, which are responsible for the incorporation and excision of the prophage from the bacterial chromosome; and Cro, which is a protein involved in properly timing the expression of lytic genes. The locations of these genes are spread throughout much of the λ genome.

In contrast, BPs encodes a novel, self-contained genetic switch, which accomplishes the same phenotypic switch as λ's lytic-lysogenic switch, using only three genes and one small intergenic region, contained in a contiguous 2 kbp segment of the BPs genome. We believe all the necessary machinery for the entire switch is contained in this cassette. An interesting feature

of the switch is that the *attP* core is found within the BPs repressor gene (*33*), resulting in a shortened gene product in the prophage as compared to the viral repressor (gp33). The shortened $gp33^{103}$ is the active repressor, while $gp33^{136}$ is unable to confer immunity to superinfection when expressed from a single-copy integrative vector [10].

In BPs, this novel switch determines whether to enter lytic or lysogenic growth and appropriately controls the gene expression of either phenotype. Understanding the mechanism of decision making of this genetic switch is crucial to understanding the gene expression patterns of BPs and an important part of uncovering the life cycle of BPs.

5.3 GENETIC ORGANIZATION OF THE IMMUNITY REGION

BPs and the other mycobacteriophages in cluster G encode a predicted repressor and a tyrosine integrase in a small operon and are the only genes transcribed leftwards [10,11]. The phage attachment site, *attP*, which is the location of site-specific recombination between the host and phage genomes, is typically found near the integrase gene (*int)*. Here, the *attP* core is located within the open reading frame of the predicted repressor gene, gene *33* in BPs (Fig5-1A). Integration at this location results in the separation of the 5' and 3' portions of the gene in the prophage. This is curious because the main function required by a prophage to maintain lysogenic stability is the ability to repress lytic gene expression and prevent the expression of lytic genes of superinfecting phages; roles performed primarily by the repressor. Most cluster G mycobacteriophages are temperate and are able to lysogenize *M. smegmatis* mc^2155 [10] so how the repressor protein is able to function is called into question.

A similar arrangement of an operon with *int-rep* and the *attP* located within the repressor gene can be found in mycobacteriophages from other clusters, including I1, N, and P. Though phages from the same cluster have similarity at the nucleotide level [29], the sequences of the immunity cassettes between clusters do not share significant nucleotide similarity [31]. The regions from three of these mycobacteriophages, BPs from cluster G, Brujita from cluster I1 and Charlie from cluster N were examined in depth [31]. These studies focus mainly on the genetic switch of BPs, though other mycobacteriophages will be used to draw comparisons between the systems.

The prophage of BPs integrates into tRNA-Arg (MSMEG_6349) and the result is a truncated form of the repressor, as a stop codon is formed at the *attL* junction [Fig. 5-1A; [31]]. The viral form of the protein is 136 residues ($gp33^{136}$) and the prophage form is 103 residues ($gp33^{103}$). BPs is able to form a stable lysogen that is immune to superinfection (Fig. 5-1B), indicating that the prophage version of the protein is able to function as an active repressor. To investigate the role of the viral and prophage forms of gp33, we tested the ability of strains expressing either the viral 136 amino acid protein or prophage 103 amino acid protein to confer immunity. Results show that $gp33^{103}$ was able to confer immunity to BPs, while $gp33^{136}$ was not (Fig. 5-1B). However, a strain that overexpresses $gp33^{136}$ was able to confer near-complete immunity to superinfecting BPs (Fig. 5-1B).

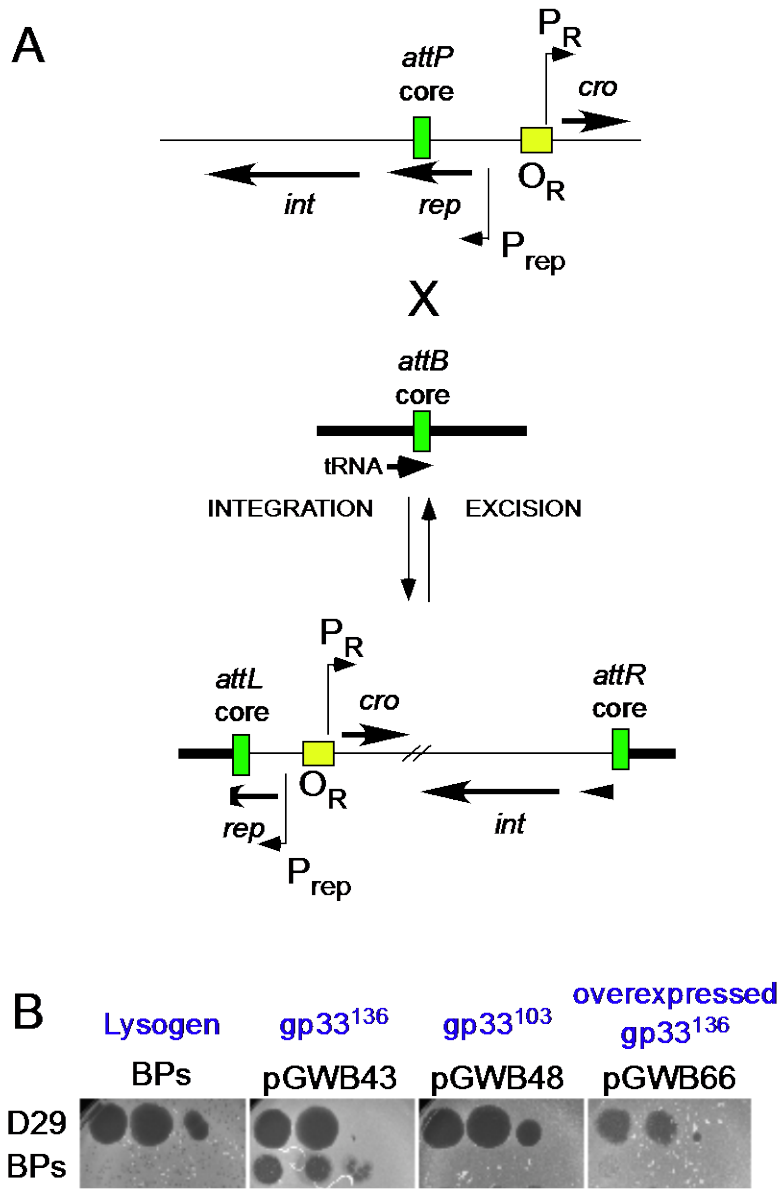

Figure 5-1. The truncated prophage form of the repressor, gp33[103], confers immunity to superinfection.

Fig. 5-1. (A) Organization of the immunity cassette in mycobacteriophages with a novel arrangement of the *attP* and repressor gene. The *attP* common core (green box) is located within the open reading frame of the repressor. The rearrangement of the genes after site-specific recombination with the *attB* results in a truncated repressor gene adjacent to the *attL* site and the smaller 3' end of the gene (arrowhead) located at the *attR*. (B) The ability of the repressor to confer immunity. Serial dilutions of mycobacteriophages D29, a control for lytic infection, and BPs were spotted with decreasing numbers of phage from left to right on a BPs lysogen and strains expressing different forms of the repressor. pGWB43 and pGWB48 express gp33[136] and gp33[103], respectively, from an integrated plasmid at the L5 integration site. pGWB66 expresses gp33[136] from the hsp60 promoter on an extrachromosomal vector. This figure is adapted from Broussard et al., 2013.

5.4 THE C-TERMINAL PORTION OF GP33 IS A DEGRADATION TAG

We wished to determine the means by which expression of the long viral form of the repressor, $gp33^{136}$, and the short, prophage form of the repressor, $gp33^{103}$ resulted in differing phenotypes. Because the viral form of the protein is inactive, we can isolate phages with gain of function (GoF) mutations by plating a strain of *M. smegmatis* expressing $gp33^{136}$ on a plate seeded with BPs phage that are repressor defective. Surviving bacteria will have gained a functional version of gp33, and are able to prevent superinfection from the repressor defective BPs phage. Seven independent mutants were isolated and found to have mutations in the C-terminal region of gp33 (Fig. 5-2A). GoF1 was a single amino acid substitution in the penultimate amino acid of gp33, corresponding to A135E. The other mutants contained frameshifts between codons 94 and 119, resulting in truncations and the substitutions of out-of-frame amino acids (Fig. 5-2A). Three of the GoF mutants, GoF2, GoF3 and GoF4, are nearly as long as the viral form of gp33 (133 amino acid versus 136 amino acid), indicating that the sequence of the C-terminal region is important, not just its length. The ability of the substitution at residue A135 in GoF1 to produce a functional repressor protein indicates that this residue is vitally important.

Alignment of the C-terminal regions of BPs gp33, a number of other mycobacteriophage repressors, which share a similar organization of the immunity cassette, and with the *M. smegmatis* ssrA tag demonstrates that there is limited similarity at the 3' end (Fig. 5-2B). The ssrA tag is added to incompletely translated proteins and tags them for degradation by the protease ClpXP [184]. The characteristics of the GoF mutations and the limited similarity with ssrA led us to investigate the possibility that the C-terminal region of $gp33^{136}$ destabilized the

protein and functioned as a degradation tag. The ability of a strain overexpressing $gp33^{136}$ to confer near-complete immunity also suggests that the amount of gp33 protein is important for immunity (Fig. 5-1B).

To examine if the extreme C-terminal end of gp33 functions as a degradation tag, we placed the C-terminal 13 amino acids of gp33 at the 3' end of GFP in a nitrile inducible vector [152] and assayed for fluorescence as an indicator of the amount of GFP protein. The uninduced strains produce a low level of fluorescence, which is unaffected by the presence or absence of GFP, and is due to autofluoresence of *M. smegmatis* (Fig. 5-3C). When inducer is added, the fluorescence of the untagged GFP increased approximately 7-fold compared to uninduced (Fig5-3C). The GFP tagged with the terminal 13 amino acids of gp33 was only induced 2-fold and the mutation A135E restored GFP fluorescence to more than 70% that of the untagged GFP and is 8-fold induced above its own uninduced levels (Fig. 5-3C). Western blot analysis with an antibody against GFP showed that varying levels of fluorescence are due to differences in protein levels. GFP protein was evident in the untagged and A135E mutant when induced, but no band is present in the empty vector or when GFP was tagged with the wild-type gp33 sequence (Fig. 5-2D).

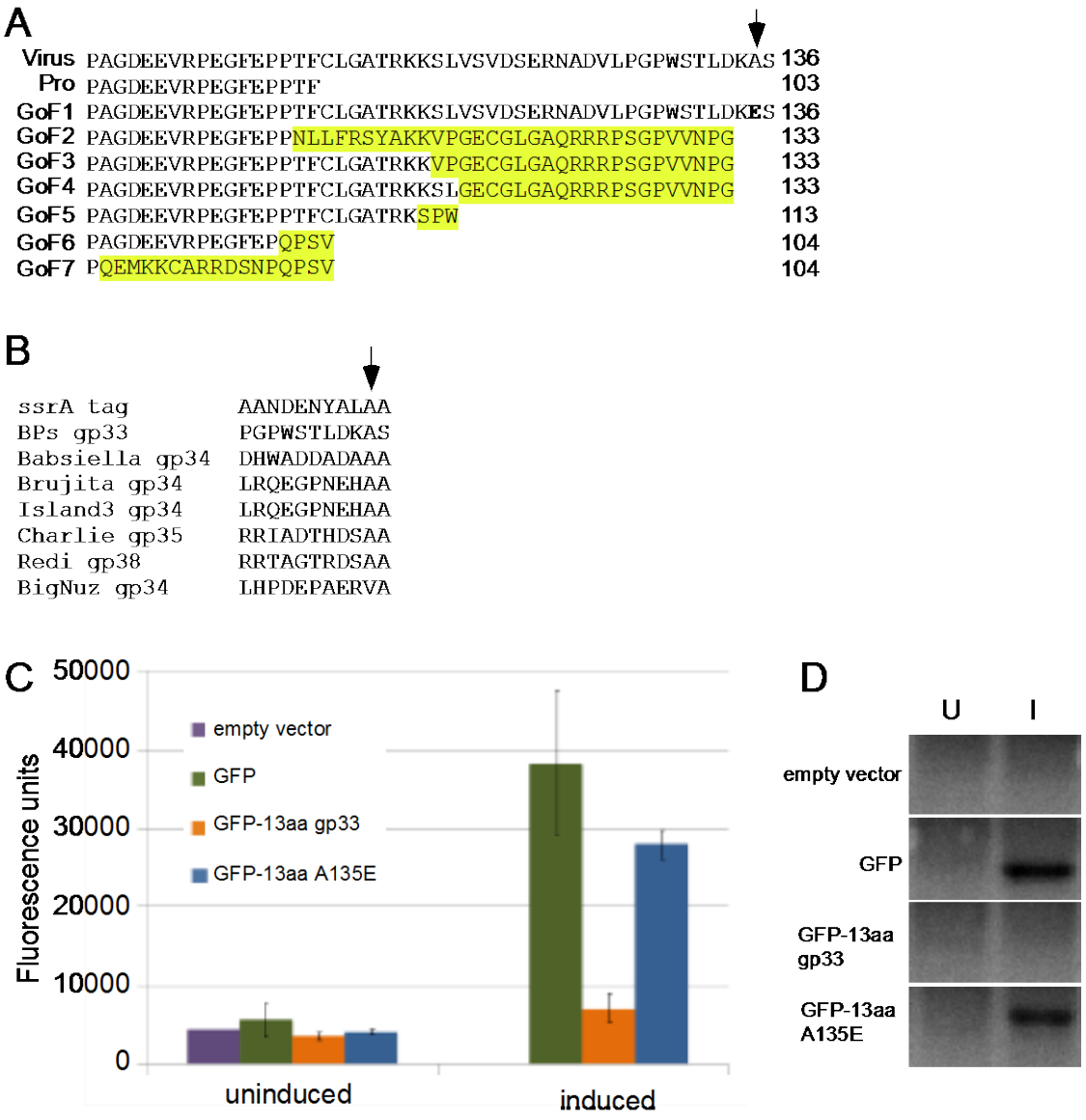

Figure 5-2. The C-terminus of gp33[136] is a tag for proteolytic degradation.

Fig. 5-2. (A) BPs gp33 amino acid sequences of gain of function (GoF) mutants. The virus and prophage (pro) amino acid sequences are provided. GoF1 contains an amino acid substitution (bold) at the penultimate residue (arrow). GoF2-7 contain frameshift mutations (yellow boxes). The number of amino acids in each of the proteins is noted. (B) Alignment of the ssrA proteolytic tag and the last 11 amino acids (aa) of the repressors from phages in cluster G (BPs), I1 (Babsiella, Brujita, Island3), N (Charlie, Redi) and P (BigNuz). (C) Effect of the addition of the last 13 aa of gp33[136] to GFP. A nitrile inducible promoter drives expression in an empty vector with no GFP, an untagged GFP, GFP plus the last 13 aa of BPs gp33[136] and GFP plus 13 aa with the A135E mutation, equivalent to the mutation isolated in GoF1. Fluorescence units are $(LAU/mm^2)/OD_{600nm}$ and the data are represented as mean ± SD. (D) Western blot of uninduced (U) and induced (I) cultures carrying the same vectors described in C with anti-GFP antibody. Portions of this figure are adapted from Broussard et al., 2013.

5.5 EXPRESSION OF REP AND CRO BY DIVERGENT PROMOTERS

Two divergent promoters are contained within the intergenic region between genes *33* and *34*, P_R and P_{rep} (see Chapter 3). P_R is responsible for expression of the early lytic operon beginning with gene *34* and P_{rep} expresses the repressor, *33*. P_R, as previously discussed in Chapters 3 and 4, is tightly down-regulated when expression is measured in a BPs lysogen of *M. smegmatis*. To determine if the repression of P_R is due to the action of gp33, we assayed for transcription from the P_R promoter, measured by mCherry fluorescence of a promoter-reporter transcriptional fusion. The P_R promoter is shut down in a BPs lysogen (as previously noted in Chapter 3), and when the $gp33^{103}$ active repressor is provided on a plasmid, it is also sufficient to eliminate P_R activity (Fig. 5-3A). The inactive form of the repressor $gp33^{136}$ is unable to repress P_R activity under the same conditions (Fig. 5-3A), unless it is stabilized by a gain of function mutation, such as A135E (gp33A135E) or a frameshift (gp33shift) in the C-terminal tag, which correspond to the GoF1 and GoF2 mutants, respectively (Fig. 5-2A). P_R, in the presence of both of these gp33 mutants, has a partial restoration of promoter activity although activity is considerably weaker than with wild-type $gp33^{136}$, which contains the full-length degradation tag. However, this small amount of expression of the early lytic operon from P_R must be sufficient for lytic activity as BPs phage containing these GoF repressor mutations are able to form plaques on a BPs lysogen [31].

The P_{rep} promoter is more active than P_R and is not regulated in a BPs lysogen or by $gp33^{103}$ or $gp33^{136}$ (Fig. 5-3A). P_{rep} activity is up-regulated in strains expressing the stabilized versions of the repressor, A135E from GoF1 and the frameshift from GoF2 (Fig. 5-3A), but the mechanism for this activation is unknown. TSS mapping confirmed the location of the P_{rep}

promoter (see Chapter 3) and a single base substitution in the -10 hexamer for P_{rep}, A-12G (Fig. 5-3B), eliminates activity of P_{rep} (Fig. 5-3C). This confirms the promoter location identified by TSS mapping and provides evidence that P_{rep} is the only leftwards promoter in this region. This is contrary to the regulation in the phage lambda genetic switch, where two promoters are present, one for the establishment and another for maintenance of lysogeny. The P_{rep} A-12G mutation was discovered in a genetic screen for clear plaque mutants, which are unable to form lysogens, therefore P_{rep} promoter activity is required for lysogeny [31].

In order to determine the site of action of the BPs $gp33^{103}$ repressor, repressor-insensitive BPs mutants were isolated. These phage mutants were able to grow lytically on a strain expressing $gp33^{103}$ that confers immunity to superinfecting phage. Thirteen independent mutants were sequenced and all had mutations in the *33-34* intergenic region [31]. Two of these were single nucleotide substitutions in O_R, A-24C and T-21C (Fig. 5-3D). These mutations helped to define the operator for gp33 and biochemical assays show that binding of $gp33^{103}$ to the *33-34* intergenic regions is specific and binding affinity in reduced ~10-fold in the O_R mutants [31]. Both A-24C and T-21C cause de-repression of the P_R promoter in a BPs lysogen (Fig. 5-3E). The promoter activity of both of these mutants is modestly elevated compared to wild-type. The inability of the repressor to fully repress P_R in these mutants allows for lytic infection of a strain of *M. smegmatis* expressing the repressor. Interestingly, the cluster G phage Angel, which is highly similar in sequence identity to BPs [10], contains the equivalent substitution as BPs O_R T-21C, and is repressor insensitive [31].

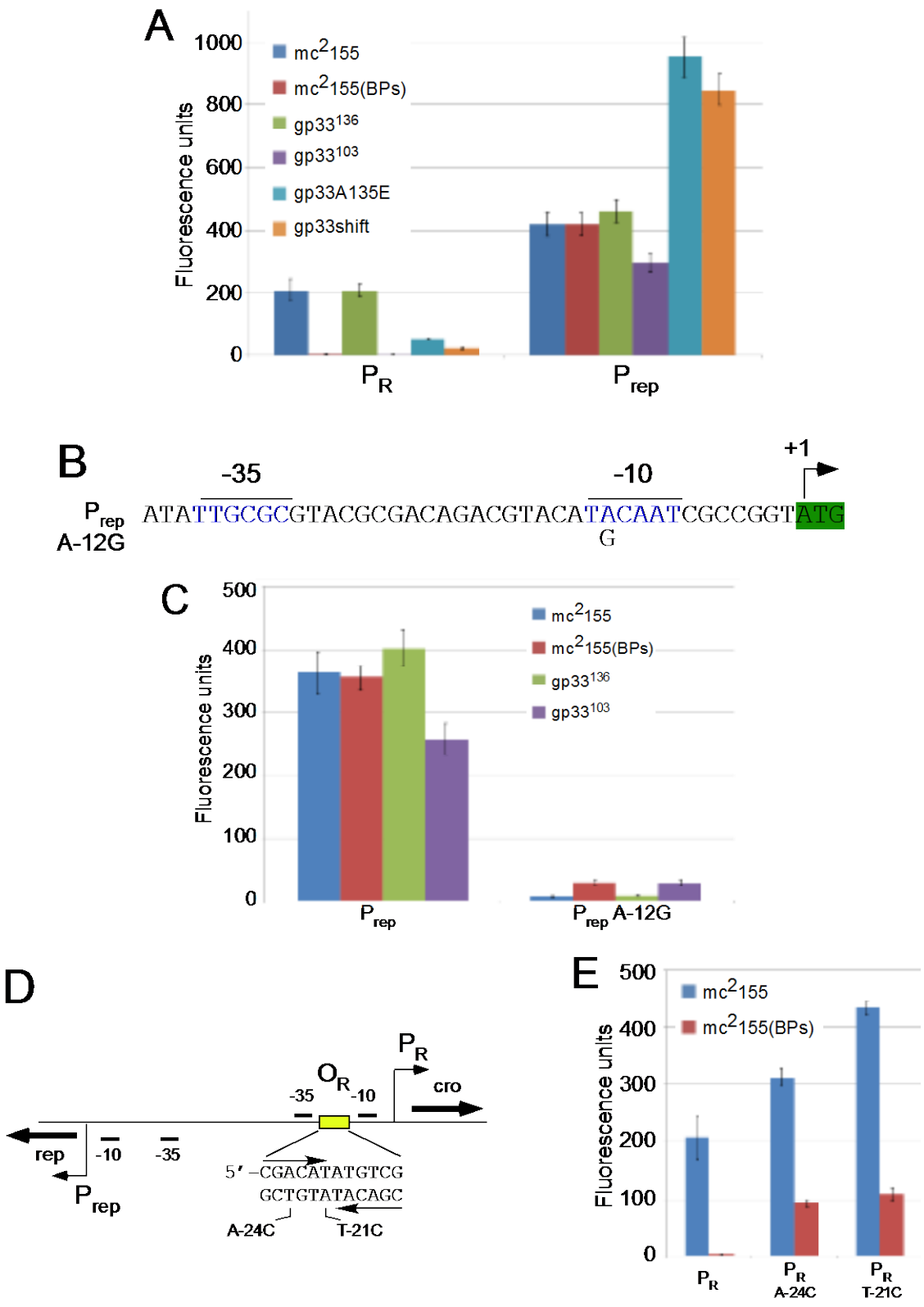

Figure 5-3. Divergent transcription in BPs from P_R and P_{rep}.

Fig 5-3. (A) Expression from promoter-mCherry transcriptional fusions and the effect of gp33. The promoters P_R and P_{rep} were fused to mCherry and examined in *M. smegmatis* mc^2155, a BPs lysogen [mc^2155(BPs)] and strains expressing $gp33^{136}$, $gp33^{103}$, gp33A135E and gp33shift. Fluorescence units are (LAU/mm^2)/OD_{595nm} and the data are represented as mean ± SD. (B) The sequence of P_{rep} and the location of A-12G mutation. The -35 and -10 hexamers of P_{rep}, the TSS (+1) and translation start codon of *33* (green box) are denoted. (C) The A-12G mutation eliminates promoter activity. The activity of wild-type P_{rep} and P_{rep}A-12G mutant are shown in *M. smegmatis* mc^2155, a BPs lysogen [mc^2155(BPs)] and strains expressing $gp33^{136}$ and $gp33^{103}$. Fluorescence units are (LAU/mm^2)/OD_{595nm} and the data are represented as mean ± SD. (D) The location of the repressor-insensitive mutations within O_R, A-24C and T-21C. (E) The promoter activities of P_R and two of the repressor-insensitive mutants A-24C and T-12C in *M. smegmatis* mc^2155 and a BPs lysogen [mc^2155(BPs)]. Fluorescence units are (LAU/mm^2)/OD_{595nm} and the data are represented as mean ± SD. Portions of this figure are adapted from Broussard et al., 2013.

5.6 REGULATION OF INTEGRASE LEVELS

As integration at the BPs *attP* site is required to produce active $gp33^{103}$, site-specific integration and, therefore, the expression and stability of integrase are vital to the ability to establish and maintain lysogeny. We believe that gene *32*, a noncanonical tyrosine integrase [31], is expressed from P_{rep} and therefore its expression and the expression of *33* appear to be constitutive and do not require activation (Fig. 5-3A). This indicates a different mode of regulation than in the phage lambda genetic switch, where CII is required for activation of P_{RE} and the production of CI, to establish lysogeny.

A BPs phage with a mutation in the catalytic tyrosine of Int (Y256A) does not produce lysogens, demonstrating that functional Int is required for lysogeny [31]. If Int concentration is not controlled at the transcriptional level by regulation of P_{rep}, what is the mechanism of Int regulation? The integrases of other mycobacteriophages that have *attP-rep* organizations similar to BPs, have C-terminal sequences that resemble mycobacterial ssrA tags (Fig. 5-4A), like the repressors (Fig. 5-2B). The BPs Int does not have a recognizable ssrA-like sequence, but has a

longer C-terminal sequence than the other integrases. Addition of the C-terminal 5 amino acids of the mycobacteriophage Brujita Int, Brujita gene *33,* to GFP results in a decrease in fluorescence and fluorescence is restored with a mutation in the penultimate residue, A296E (Fig. 5-4B), which is equivalent to the stabilized mutant in the BPs repressor, gp33A135E (Fig. 5-2C). The restoration of GFP fluorescence indicates that the Brujita Int and others with penultimate alanine residues may be targeted for degradation by ClpXP, like the repressors of these phages.

Integration proficient vectors derived from BPs, Brujita and Charlie have reduced transformation efficiencies, which is in line with the evidence that the integrases from these mycobacteriophages are unstable [31]. Addition of stabilizing mutations to the integrase genes in these vectors can improve transformation efficiencies by between one and three orders of magnitude [31]. The C-terminal sequence of the Charlie Int contains a penultimate valine (Fig. 5-4A) and also has the highest transformation efficiency, although efficiency can also be improved by additional stabilizing mutations changing the last three residues from AVS to ADD [31]. This provides additional evidence that these integrases are unstable and proteolytically degraded.

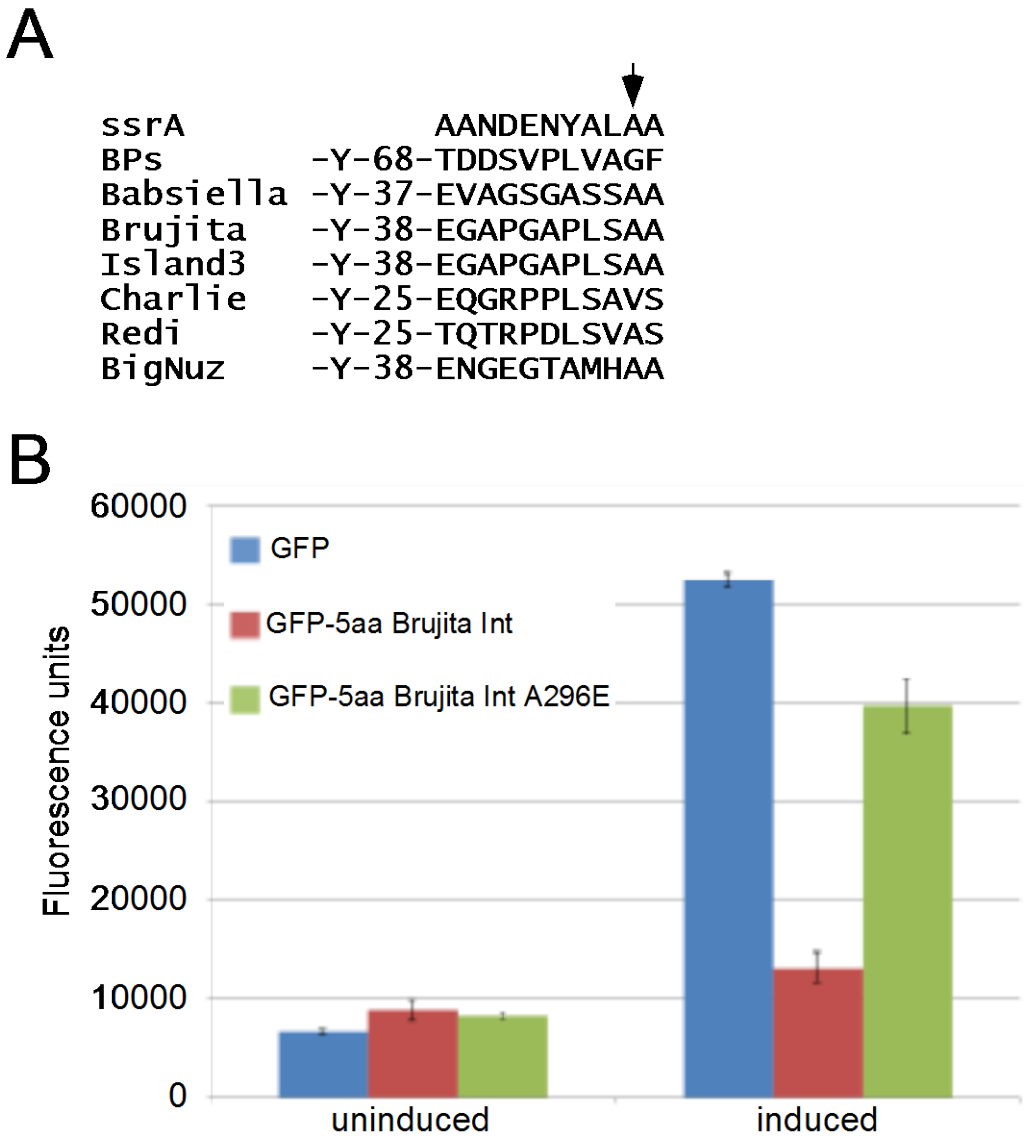

Figure 5-4. The C-terminal 5 amino acids of Brujita Int are a tag for proteolytic degradation.

Fig. 5-4. (A) The C-terminal 11 residues of several Int proteins aligned with the ssrA proteolytic tag. For Int, spacing from the catalytic tyrosine (Y) is noted. The conserved penultimate residue is indicated (arrow) and is not present in BPs and Charlie. (B) Effect of the addition of the last 5 aa of Brujita Int to GFP. A nitrile inducible promoter drives expression of untagged GFP, GFP plus the last 5 aa of Brujita Int and GFP plus 5 aa with the A296E mutation, in the penultimate residue of the tag. Fluorescence units are $(LAU/mm^2)/OD_{595nm}$ and the data are represented as mean ± SD. Portions of this figure are adapted from Broussard et al., 2013.

5.6.1 Stability of BPs Int

The behavior of the BPs integration vector suggests that the BPs Int is unstable although no obvious ssrA-like tag is observed at its C-terminus (Fig. 5-4A). BPs Int has an extended C-terminus that is 30 aa longer than any of the other integrases examined (Fig. 5-4A). We fused the C-terminal 64 residues of Int to GFP (pLO24; Fig. 5-5A) and found that fluorescence is decreased to levels equivalent to uninduced controls (pLO24; Fig. 5-5B) indicating that this 64 aa sequence destabilizes GFP. Ten amino acid segments were deleted from the C-terminal end by creating stop codons in the 64 aa tag by site directed mutagenesis (pLO27-57; Fig. 5-5A). GFP fluorescence is still largely absent with a tag of 14 aa (pLO57; Fig. 5-5B). The removal of next three amino acids AGA, which are residues 346 to 348 in BPs Int, restores fluorescence to the same level as untagged GFP (pLO71; Fig. 5-5B). However, mutating the two residues that would be important for function if this was a similar ssrA-like tag, G347D and A348D, did not result in stabilization of GFP (pLO70). We believe that BPs Int is proteolytically degraded and likely has multiple tags in its C-terminus, the most 5’ of which includes the residues A346 G347 and A348.

A

```
pLO24 GDGDGRPIWTRAGASIIAAAMILLGLGISDAIGGVLDRTADVAHVLVLDNAVDTDDSVPLVAGF*
pLO27 GDGDGRPIWTRAGASIIAAAMILLGLGISDAIGGVLDRTADVAHVLVLDNAVDT*
pLO28 GDGDGRPIWTRAGASIIAAAMILLGLGISDAIGGVLDRTADVAH*
pLO29 GDGDGRPIWTRAGASIIAAAMILLGLGISDAIGG*
pLO56 GDGDGRPIWTRAGASIIAAAMILL*
pLO57 GDGDGRPIWTRAGA*
pLO70 GDGDGRPIWTRADDSIIAAAMILLGLGISDAIGGVLDRTADVAHVLVLDNAVDTDDSVPLVAGF*
pLO71 GDGDGRPIWTR*
pLO72 GDGD*
```

B

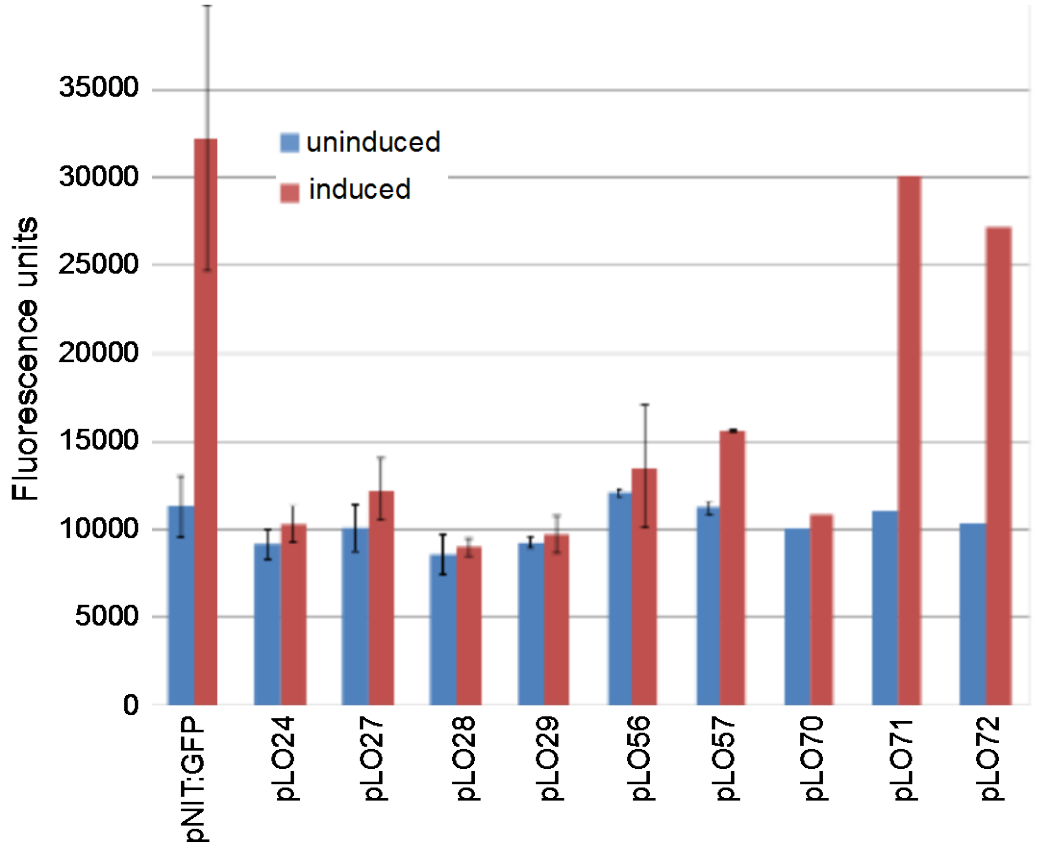

Figure 5-5. The C-terminal tag of BPs destabilizes GFP.

Fig. 5-5. (A) The sequence of tags added to GFP from BPs Int. pLO24 includes GFP plus the last 64 aa of BPs Int. Subsequent tags are subtractions of 10 aa by addition of stop codons (asterisk), with the exception of pLO70 which includes all 64 aa with substitution of 2 residues AGA→ADD (red letters) and pLO71, in which three residues were removed (AGA) compared to pLO57. (B) The fluorescence of uninduced (blue bars) and induced (red bars) GFP with tags from A and an untagged control (pNIT:GFP) all under the control of a nitrile inducible promoter. All experiments include at least 2 biological replicates except pLO70-72 (no SD reported). Fluorescence units are $(LAU/mm^2)/OD_{595nm}$ and the data are represented as mean ± SD.

5.7 P_R AND P_{REP} ARE NOT REGULATED BY GP34

The integrase plays an important role in the decision between lytic and lysogenic growth in our switch, but in other genetic switches there is a counterbalancing function that promotes lytic growth. In lambda, the Cro protein competes for binding to the operators and shuts down the production of *cI*, the lambda repressor. In the BPs switch, gene *34*, the first gene in the early lytic operon expressed from P_R, has a predicted helix-turn-helix domain and is a likely candidate for a Cro-like function because of its location [31]. When BPs is plated on bacterial strains expressing gp34 from its native P_R or a repressor-insensitive P_R T-21C, the rate of lysogeny is reduced, especially when gp34 is expressed from the repressor-insensitive promoter [31]. This phenotype is also observed with BPsΔint phage and is therefore not due to increased excision mediated by gp34 [31].

We also examined if gp34 could directly regulate promoter activity from P_R or P_{rep} using similar strains that express gp34 from wild-type P_R or P_R T-21C. Expression of gp34 did not alter activity of the P_R or P_{rep} promoters (Fig. 5-6) and the mechanism of action by which gp34 promotes lytic growth is still unknown.

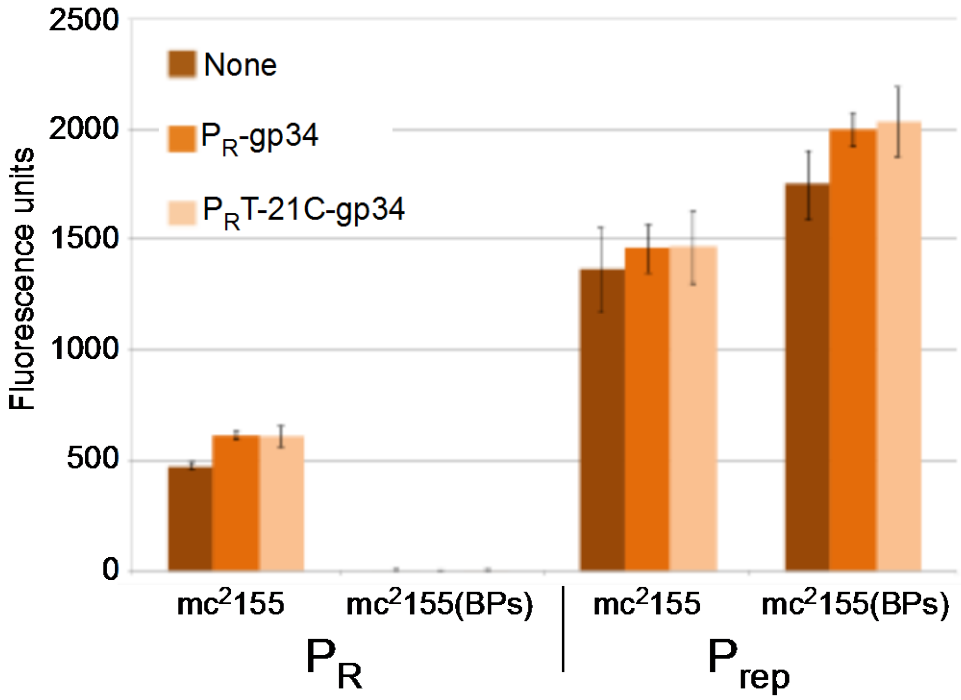

Figure 5-6. The activities of P_R and P_{rep} are not affected by gp34.

Fig. 5-6. The activity of P_R and P_{rep} promoters in *M. smegmatis* mc^2155 and a BPs lysogen [mc^2155(BPs)] with either no gp34, gp34 expressed from wild-type P_R, or gp34 expressed from repressor-insensitive P_RT -12C. Fluorescence units are (LAU/mm^2)/OD_{595nm} and the data are represented as mean ± SD.

5.8 DISCUSSION

We believe that the entirety of the BPs genetic switch is housed in the ~ 2 kb region from gene *32* to *34*. This is supported by our analysis of the functions of genes. Gene *32* encodes the integrase, the concentration of which control the lytic-lysogenic decision, as integration is required for the establishment of lysogeny and the continued synthesis of stable repressor. Gene *33* encodes the repressor, and the active truncated $gp33^{103}$, which is generated when site-specific recombination occurs and a stop codon is formed at *attL*, is capable of conferring immunity to superinfection. The protein encoded by gene *34* stimulates lytic growth, similar to the function of

Cro in phage lambda, but the mechanism for this activity is unresolved. The intergenic region between genes *33* and *34* contains divergent promoters, P_R and P_{rep}. P_R is tightly regulated by $gp33^{103}$ and small increases in its promoter activity in a BPs lysogen increase the frequency of lytic growth. P_{rep}, which is constitutively active, is the sole promoter for the establishment and maintenance of lysogeny and is not autoregulated by the repressor. A search for clear plaque mutants defective in lysogeny, which could indicate other genes involved in the lytic-lysogenic switch, found that all mutations isolated were located in the *int-rep-cro* immunity cassette [31].

These small, self-contained genetic switches are found in diverse mycobacteriophages genomes and could be transferred through horizontal gene transfer, as they are modular in structure. Integration-dependent immunity systems may represent an ancestral form of genetic switch [31,185]. These switches lack many of the constituents of the phage lambda genetic switch, including Xis, CI and CII but these could be added as further refinements to simpler systems. Moreover, bioinformatic searches indicate that similar switches are found outside of mycobacteriophages; potential integration-dependent immunity systems can be found in prophages in *Rhodococcus*, *Nocardia* and *Corynebacterium* among others [31].

The process of excision and induction of the prophage is less well understood in these bacteriophages. Of the mycobacteriophages with this immunity system, none are inducible with DNA damaging agents. There is also no clear role for an excise protein, as the integrase of BPs does not appear to have an ability to control the directionality of site-specific recombination [31]. It appears that Int is synthesized from P_{rep}, and once integration has occurred, the promoter is uncoupled from *int*. There is no evidence for a promoter between genes *32* and *33* that would be able to synthesize *int* in the prophage (L. Oldfield, unpublished results). It is possible that

synthesis of Int in the prophage occurs from a promoter upstream of the chromosomal gene adjacent to *int* in the prophage, MSMEG_6348 [31].

This novel self-contained switch is the first described where integration, through site-specific recombination, is responsible for ability to switch between bistable states, here lytic versus lysogenic growth. Potentially, simple genetic switches like this may be utilized in the construction of complex genetic circuits for synthetic biology, such as in the recently described biological counting and data storage systems being developed [186-188].

6.0 CONCLUSIONS

We have completed a detailed study of the gene expression, promoter sequence requirements and the genetic switch that controls gene expression in mycobacteriophage BPs. This gives insight on the basic biology of mycobacteriophages and generates hypotheses for further studies of newly uncovered putative regulation in mycobacteriophages and for future studies of additional mycobacteriophages.

The transcriptome of BPs is temporally controlled, as is the gene expression of other bacteriophages, including T4 [44], lambda [189], phages of *Yersinia* spp. [43], and phages of *Streptomyces* [190]. Distinct patterns of gene expression in mycobacteriophage BPs were observed at two independent time points during lytic growth, early and late. Though the mechanism of the temporal control of gene expression in BPs is unknown, other bacteriophage use phage encoded activators or sigma factors, antiterminators or modifications to the host RNAP. No mechanisms have been proposed for control of temporal regulation of gene expression in mycobacteriophages, though other interesting regulatory mechanisms have been identified. In mycobacteriophage L5, stoperators sequences which prevents transcription elongation [36]. We propose that BPs gene expression is regulated through an antitermination system, which could, like in the bacteriophage lambda system, control the timing of gene expression [53]. In our model, a phage-encoded antitermination factor, putatively located within the early lytic operon (*34* to *44*), is able to promote read through of the terminators that are

clustered near the end of this operon and around genes *55* to *59* in the right arm. There are nine additional terminators throughout the left arm of BPs, which may also be affected by an antitermination system.

Our analysis only examined two time points during lytic infection, at 30 minutes and 2 hours post-adsorption. Potentially, there are additional temporal changes in BPs gene expression yet to be uncovered. Other mycobacteriophages examined. L5, TM4, Bxb1 and Giles, also show only two phases of expression [3-5,8]. L5, TM4 and Bxb1 expression was examined by ^{35}S-methionine protein labeling, which is not a sensitive measure of changes in protein expression as all newly synthesized proteins in phage-infected bacterial cells are displayed. Changes in the synthesis of a few proteins made at relatively low concentrations would not be observable. Therefore, additional phases of protein expression in these mycobacteriophages may have been missed. The transcriptome of Giles was examined by RNA-Seq, which avoids many of the pitfalls of ^{35}S-methionine labeling. Two phases of gene expression were observed for Giles as well, however, like with BPs in these studies, only two time points were examined [8]. Additional, fine-tuned regulation of the gene expression of BPs almost certainly occurs. Though RNA for the late lytic sample was collected near the time that BPs begins to lyse host cells, the lysis cassette, which includes the essential *lysA* gene [25], is not being transcribed. The *lysA* and *lysB* lysin genes are presumably expressed near the time of lysis. Potentially, isolating RNA from a sample closer to the time of lysis would show a different pattern of expression that included the lysin genes. The gene expression of some other bacteriophages also contains middle time points between early and late gene expression, which suggests that additional middle time points in BPs should be investigated.

In late lytic growth, the structure and assembly genes, beginning with gene *6*, the scaffold protein, are highly expressed, but the method of temporal control over the P_6 promoter is not obvious. No phage activator is required for expression from P_6, as P_6 is active in *M. smegmatis* with no phage proteins present. However, a method of temporal control must be in place that restricts expression from P_6. A similar situation is seen for P_{61}, which is active without phage-encoded proteins and is not repressed in a lysogen. However, expression downstream of the P_{61} location is only found in late lytic growth, and not early in infection or in a lysogen.

While the method of regulation of some of the BPs promoters is unknown, the P_R and P_{rep} promoters have been examined in detail because of their role in the genetic switch of BPs that controls the decision between lytic and lysogenic growth. The immunity cassette, genes *32* and *33*, are expressed most strongly in a BPs lysogen. The repressor (*33*) is expressed from the constitutively active P_{rep} promoter and the two bidirectional promoters are located in the intergenic region between genes *33* and *34*. The P_{rep} promoter is unusual compared to the two promoters that express the lambda repressor, *cI*, where CI autoregulates its own expression [53]. Constitutive expression of the repressor should lead to very high rates of lysogeny, however the frequency of lysogeny for BPs is only 5% [10]. In depth examination reveals that the level of BPs repressor present in cells is controlled by proteolytic degradation and, therefore, regulation at the transcriptional level is not required. Adding an additional layer of complexity, the production of stable BPs repressor requires site-specific recombination of the phage into the host chromosome. The ability to perform site-specific recombination is determined by the concentration of integrase. Integrase is also controlled by proteolytic degradation. The integration of BPs allows the phage to generate continuous levels of stable repressor from the P_{rep} promoter. Understanding the novel mechanism of integration-dependent immunity of the

genetic switch of BPs is vital to understanding how the P_R and P_{rep} promoters function. The overall function of the switch is to decide which program of gene expression is undertaken—lytic growth and the expression of genes necessary to replicate, assemble progeny phage and lyse the host, or lysogenic growth and the production of stable repressor and efficient shutdown of expression of lytic genes.

In addition to examining the role of P_R in the genetic switch, the role of the sequence of the -35 hexamer, the extended -10 region, the -10 motif, the spacer region and the length of the spacer in P_R were all determined. This is the most complete analysis of any mycobacterial or mycobacteriophage promoter ever conducted. This promoter appears to be transcribed using host RNA polymerase and SigA, the housekeeping sigma factor. The regulation of P_R by repressor binding to the operator between the -10 and -35 motifs was indirectly examined by substitutions in O_R. Through genetic screens, P_R mutants were discovered that do not allow lysogenic growth [31]. Even a moderate amount of de-repression of P_R and upregulated expression of lytic genes is able to unbalance the switch and prevent BPs from growing as a lysogen.

Examination of the P_R promoter also identified that P_R is activated by supercoiling through comparison of P_R activity in an extrachromosomal versus integrated context. The activation by supercoiling provides an additional increase in early lytic gene expression during lytic growth when the BPs genome is replicating extrachrosomally, while P_R is still being fully repressible by the repressor in the prophage, integrated form.

The examination of gene expression in mycobacteriophage BPs sheds light on the biology and mechanisms of regulation in mycobacteriophages and generates interesting hypotheses for future experiments on transcription initiation and termination in mycobacteriophages. The development of a library of synthetic promoters also demonstrates how

deeper understanding and examination of mycobacteriophages can lead to the development of genetic tools for mycobacteria.

APPENDIX A

PROFILE OF PROTEIN EXPRESSION IN BPS LYTIC GROWTH

A.1 INTRODUCTION

^{35}S-Methionine labeling of proteins has been used to demonstrate the timing of gene expression and the state of host protein synthesis in mycobacteriophages L5, TM4 and Bxb1 [3-5]. Each of the phages was shown to have two phases of protein expression, early and late, albeit with slightly different timing. L5 early protein synthesis begins around 10 minutes post-infection and the pattern persists until around 45 minutes and late protein expression overlaps beginning around 20-25 minutes until cell lysis around 2.5 hours post-infection [4]. TM4 early expression was seen from 20 to 60 minutes post-infection and late from 60 minutes onwards with lysis occurring by around 4 hours [3]. Bxb1 early proteins were observed from 15 to around 40 minutes, which overlaps with late expression from 20 minutes onwards [5]. In L5 and Bxb1, host protein synthesis is shut off soon after infection [4,5], but not in TM4 [3].

We hypothesized that BPs would also demonstrate temporal regulation of protein synthesis, most likely in two phases similar to L5, TM4 and Bxb1. Characterizing the pattern and timing of protein expression was a useful first step for further gene expression experiments and would allow the selection of defined early and late time points in these experiments. *M.*

smegmatis cells infected with BPs were pulse labeled with ^{35}S-Methionine, which labels newly synthesized proteins, at 15 to 30 minute intervals until 180 minutes post-adsorption, which corresponds to the time of cell lysis for BPs (see Chapter 3).

A.2 MATERIALS AND METHODS

M. smegmatis mc^{2}155 cultures were grown in liquid media to mid-logarithmic phase (OD_{600nm} = 1.0-1.5) and then diluted to an OD_{600nm} of 0.1 in 7H9 supplemented with glucose and $CaCl_2$. BPs phage were added to the culture at an MOI of 10 and allowed to adsorb to bacterial cells at room temperature, standing for 30 minutes. 1 ml aliquots of control or infected culture were incubated with 1 µl ^{35}S-Methionine for 3 minutes at 37°C. Labeled samples were flash frozen on dry ice. The cells were then pelleted for 15 minutes at 4°C and resuspended in 100 µl of SDS gel-loading buffer. The samples were denatured at 95°C for 5 minutes and 10 µl per sample was separated on a 10% polyacrylamide gel at 150V for 5 hours. The gels were stained with 0.5% coomassie brillant blue and destained with a methanol-based solution overnight. The gels were dried and exposed to an imaging plate (IP).

A.3 RESULTS

The ^{35}S-Methionine labeling demonstrated that host protein synthesis is not shut down completely by BPs infection, which is similar to TM4. Many of the bands observed in the uninfected control present throughout the infection.

One caveat with this experiment is that the first sample was take after the BPs phage were allowed to adsorb to the bacterial cells at room temperature for 30 minutes. This was not counted as part of the infection time. This is different than the other gene expression studies conducted on BPs, which had very short or no time allotted for phage adsorption. The post-adsorption lane displays many bands that appear to be regulated throughout the course of infection (black arrows). It is unclear if these are phage or host proteins. One of these bands (indicated by the red arrow) shows a protein, which appears to change expression from up to down regulated several times in the course of BPs infection.

One band (green arrow) appears to be a higher molecular weight species that appears to be a candidate for a late BPs protein, which gradually increases in levels starting at around 45 minutes post-infection. There are no other candidates for late or early proteins because it is hard to discern what is present in the uninfected control. Though this experiment was not conclusive, but time points for further gene expression studies were chosen to before or after 45 minutes due to the changes in host and possibly phage protein expression between 30 and 45 minutes post-infection.

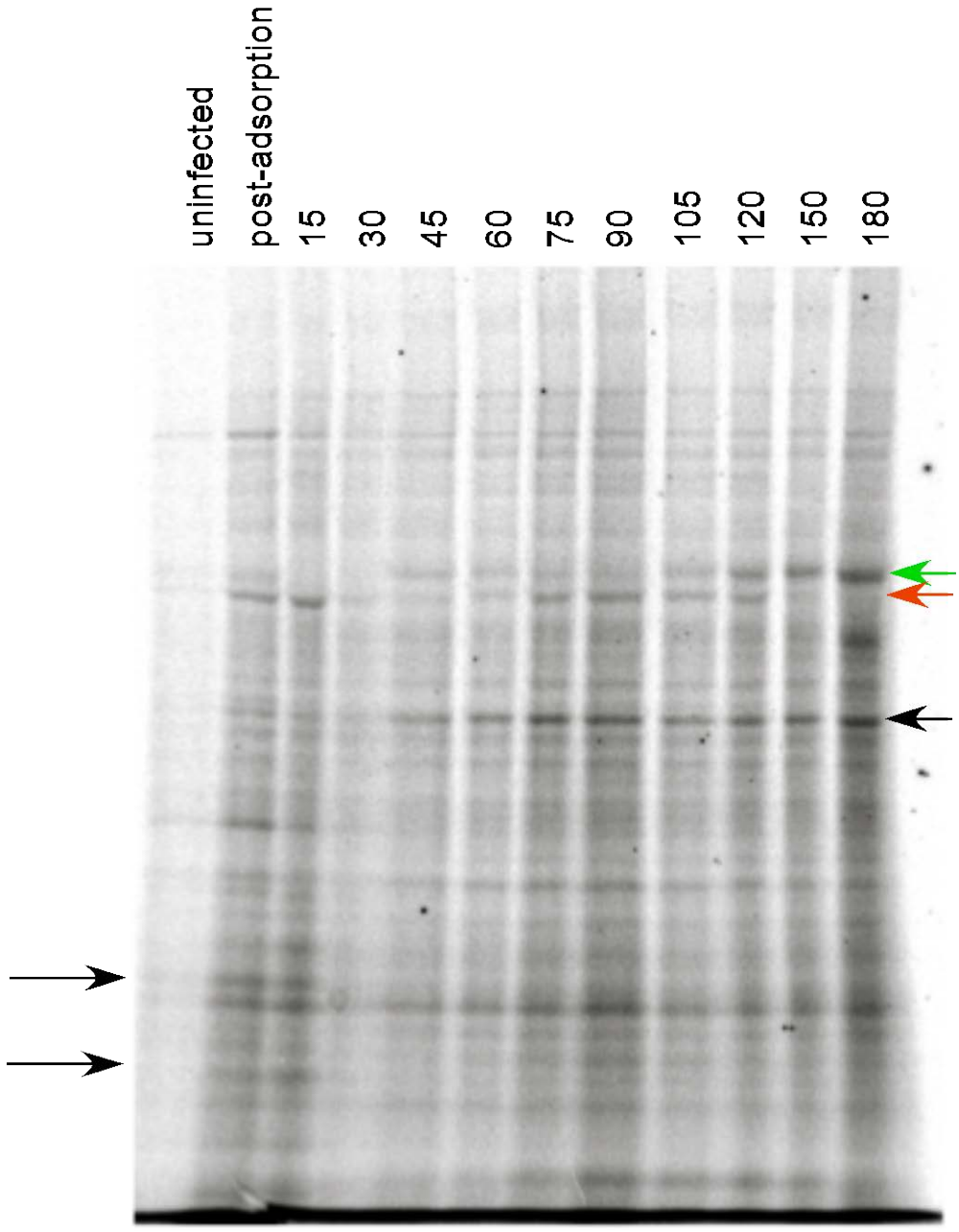

Figure 6-1. Profile of protein expression in BPs-infected *M. smegmatis* cells.

Fig. 6-1. ^{35}S-methionine labeling of newly synthesized proteins in *M. smegmatis* mc^{2}155 cells infected with BPs. Samples were taken at adsorption and every 15 to 30 minutes post-adsorption. An uninfected sample of *M. smegmatis* cells serves as a control. The arrows indicate proteins regulated over the course of infection.

A.4 CONCLUSIONS

We were not able to conclusively determine the timing of BPs protein synthesis due to the design of this experiment, but we decided to move forward with other gene expression experiments using early and late time points of 30 minutes and 120 minutes post-adsorption, respectively. Using this design for the RNA-Seq experiment (see Chapter 3), we were able to observe two phases of gene expression. More detailed analysis will need to be performed to determine the timing of the shift from the early phase into the late phase of gene expression. Most of the proteins that were differentially regulated throughout infection were present in the sample taken immediately after the 30 min room temperature adsorption (Fig. 6-1, lane “post-adsorption”) and therefore we are unable to determine if these are host or phage proteins. The difficulty observing obvious phage protein synthesis, especially for putative highly expressed proteins like the capsid, is likely due to the presence of a high concentration of newly synthesized host proteins. The shifts in expression that were observed were seen around 30 to 45 minutes post-adsorption. This generally agrees with the shifts in phage protein synthesis observed for L5, Bxb1 and TM4. In these mycobacteriophages late protein synthesis begins at between 20 minutes and 60 minutes post-adsorption [3-5].

APPENDIX B

REPLACEMENT OF BPS GENE *54* WITH *GFP* USING BRED

B.1 INTRODUCTION

The construction of mycobacteriophage mutants had previously been possible through the creation of shuttle phasmids, which are chimeras that replicate as plasmids in *E. coli* and as phage in *Mycobacterium* [17]. The development of bacteriophage recombineering of electroporated DNA (BRED) allowed for the construction of unmarked deletions in lytically growing mycobacteriophages [25]. BRED was derived from the mycobacterial recombineering system that promotes high levels of homologous recombination in mycobacteria through the expression of RecE and RecT-like proteins found in Che9c [23]. This technique has been successfully utilized to create unmarked mutations, including point mutations, deletions, and insertions, in numerous different mycobacteriophages [8,25]. Cluster G phages are amenable to BRED mutagenesis and deletions BPs genes *44*, *50*, *52*, *54* and *58* and Halo genes *49* and *52* have been reported [25].

B.2 MATERIALS AND METHODS

Bacteriophage recombineering of electroporated DNA (BRED) was carried out as described in Chapter 2.2.5. Specifically, to generate the recombineering substrate, the *gfp* gene from pMN437 (gift of the Michael Neiderweis) was PCR amplified using cloned pfu polymerase (Invitrogen) and the primers LMO01 and LMO02. All primers used to generate the recombineering substrate were PAGE purified. LMO01 and LMO02 amplified the *gfp* gene with 25 bp of homology to it and extended the substrate sequence with 50 bp of homology to the location of desired insertion in the BPs genome, flanking gene *54*. Another amplification with primers LMO03 and LMO04 was performed to add an additional 50 bp of homology to BPs, resulting in a substrate containing the *gfp* gene and 100 bp of flanking BPs genomic sequence on both ends for targeted insertion. The PCR product was cleaned up using the MinElute PCR Purification Kit (Qiagen) and eluting in sterile H_2O.

The recombineering substrate and BPs genomic DNA were co-transformed into a 100 μl aliquot of frozen recombineering *M. smegmatis* mc^2155: pJV53 cells. Plaques were picked into 100 μl of phage buffer using P200 pipet tips and vortexed. 1 μl was added to a diagnostic PCR reaction with primers LMO06 and LMO08 to detect plaques containing the replacement mutant. One primer annealed within the *gfp* gene and one in the flanking genomic sequence. Plaques containing the mutant were serially diluted and plated again on lawns of *M. smegmatis* mc^2155 to isolate pure mutant phage populations. Diagnostic PCR was again performed to identify mutant plaques.

To test for the functionality of GFP, liquid infections of *M. smegmatis* mc^2155 cells, which were grown to OD_{600nm} = ~1.0, were carried out using BPs *gfpΔ54* or mycobacteriophage

TM4 that expresses GFP [15] at an MOI of 100. An uninfected control, BPs *gfpΔ54*-infected and TM4:gfp-infected cells were grown for 4 hours at 37°C and fixed with 2% paraformaldehyde for 30 minutes at room temperature to preserve fluorescence. Fixed cells were washed with 10% glycerol two times and resuspended in 100 µl of PBS. 5 µl of fixed cells were spotted onto a glass microscope slide, a coverslip placed over top and sealed into place using coverslip sealant (1:1:1 vaseline: lanoline: paraffin).

The fluorescence was examined with a fluorescence microscope (Axiostar Plus; Carl Zeiss). Bright field and multiple fluorescence images at varying controlled exposure times (20 ms, 100 ms and 1000 ms) were taken for each field of view using an AxioCam MRc5 camera (Carl Zeiss) and Carl Zeiss AxioVision Rel. 4.6 software. For the detection of fluorescent cells a HQ:R NX (41002c- HQ545/30X, HQ620/60m, Q570LP) filter from Chroma Technology Corporation was used. The brightness/contrast of the images was not altered.

B.3 *GFPΔ54* BPS PRODUCES LOW LEVELS OF FLUORESCENCE

At the time, an insertion of a foreign gene into a mycobacteriophage had not been accomplished using BRED. It was uncertain if a longer genomic DNA sequences could be tolerated by BPs or other mycobacteriophages, gene *54* was replaced by *gfp* gene. Gene *54* of BPs was known to be non-essential for growth of the phage under laboratory conditions and was approximately the same size as the *gfp* gene. Another consideration was that BPs might be able to be developed into a reporter phage for *M. tuberculosis* diagnostics, as host range expansion mutants can be readily isolated.

Through BRED, a *gfp* replacement of *54*, named *gfpΔ54*, was constructed and examined by fluorescence microscopy. The mutant *gfpΔ54* was determined to display relatively weak expression of GFP (Fig. 6-2), possibly due to low levels of expression of gene *54* and other neighboring genes *in vivo*. This hypothesis is supported by our RNA-Seq data (see Chapter 3), which indicates that gene *54* is not highly expressed in BPs early or late lytic growth.

In order to increase fluorescence from *gfpΔ54*, constructs were designed to insert the *M. bovis* BCG promoter for *hsp60* and the L5 P_{left} promoter upstream of the *gfp* insertion using BRED. However, the isolation of pure mutants for both of these promoter insertions failed (data not shown). Though the mutants were readily detected as mixed phage populations during the BRED isolation procedure, pure mutants were not. The inability to purify these mutations could be due these strong promoters not being tolerated within the early BPs region or the phage not tolerating the increased length of its genome. The combination of a mutant strongly expressing *gfp* with the host range expansion mutations makes BPs an ideal candidate for diagnostic and reporter phages for *M. tuberculosis*.

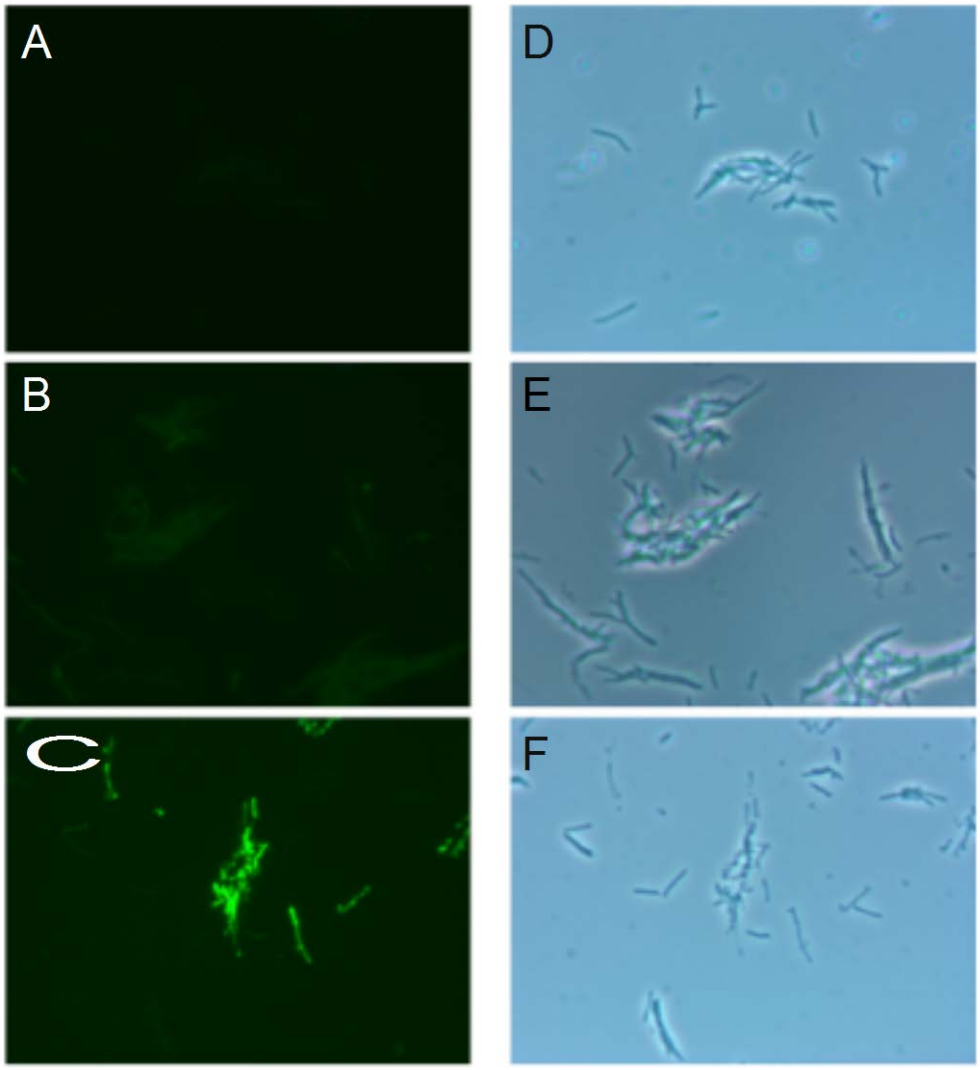

Figure 6-2. Expression of gfp from BPs gfpΔ54.

Fig. 6-2. Fluorescence microscopy of (A) mock-infected *M. smegmatis* mc^2155 (B) BPs gfpΔ54 and (C) a TM4 mycobacteriophage expressing GFP as a control.

APPENDIX C

IDENTIFYING REGIONS OF THE BPS GENOME WITH PROMOTER ACTIVITY THROUGH PROMOTER TRAP

There are many types of mycobacterial promoters that do not conform to a predictable SigA-like sequence [8,72,94], and though these promoters do have specific sequence requirements, they are not well characterized enough to inform facile predictions in genomic sequences. To examine mycobacteriophage BPs for all kinds of promoters, we screened for promoter activity from random fragments of BPs genomic DNA in a promoter trap. This does not bias our search to only find sequences that look similar to SigA-like promoters.

C.1 MATERIALS AND METHODS

C.1.1 Prepare BPs genomic DNA

A high titer lysate of BPs was prepared, precipitated with ammonium sulfate to further concentrate the sample, and genomic DNA was isolated, as stated in Chapter 2.

C.1.2 Sonicate DNA to fragment

The phage genomic DNA was sonicated to fragment the DNA. Approximately 35µg of DNA was resuspended in a total of 500µl of TE. The solution was sonicated for 10 seconds and chilled on ice for 1 minute. This was repeated 10 times. The DNA was concentrated by ethanol precipitation and resuspended in 30-50µl of TE at 42°C for 10 minutes. The fragmented DNA was separated on a 1% agarose gel for 45 min at 100V. The gel was briefly visualized with a handle-held UV lamp and a gel slice corresponding to all of the fragmented DNA between 150 and 500bp was removed. Extraction of the DNA from the agarose gel fragment was achieved using the MinElute kit (Qiagen) via the manufacturer's instructions.

C.1.3 End repair and ligation of BPs genomic DNA fragments

To prepare the DNA for ligation, the fragmented DNA was end repaired. In 100µl total volume, a mixture of dNTPs to a final concentration 4mM for each nucleotide and polynucleotide kinase buffer to a final concentration of 1x were added and mixed. Then 1µl of Klenow fragment, 5µl of T4 DNA polymerase and 5µl of T4 polynucleotide were added and gently mixed. The end repair reaction was incubated at 37°C for 30 minutes. The reaction was cleaned up with the MinElute kit (Qiagen) according to manufacturer's instructions. The concentration of DNA was determined via Nanodrop.

C.1.4 Preparing pool of clones

The reporter vector, pLO86, which contains no promoter upstream of the *mCherry* reporter gene, was digested with DraI, treated with 1µl of calf inostiol phosphotase and gel purified. The ligation was performed with the FastLink DNA Ligation kit (Epicentre) according to the manufacturer's protocol for ligations with blunt ends. The amount of fragmented DNA insert included varied between 2ng and 100ng of DNA. A control ligation reaction containing no insert was also prepared. Ligations were carried out at room temperature for between 1 hour and overnight. Ligations were transformed into One Shot TOP10 chemically competent *E. coli* cells (Life Technologies) or chemically competent NEB5α (NewEngland Biolabs) cells according to manufacturer's instructions. Two different amounts, 100µl and the rest of the transformation volume, were plated on LBA Kan and were incubated at 37°C overnight.

The transformations of control ligations with no insert produced very few colonies. The experimental ligations with insert produced a total of 1637 individual colonies, which also indicates that the maximal number of unique clones that could be present. All *E. coli* colonies were removed from the plate. 2ml of LB broth was added to the plates and colonies were harvested using a sterile cell spreader. The LB broth containing the transformants was centrifuged at 5000 x *g* for 10 minutes to collect the cells. The plasmids were isolated from the concentrated cells using the GeneJet Plasmid MiniPrep kit (Fermentas) according to the manufacturer's instructions. Three independent pools of ligated clones were created.

C.1.5 Visual screen to identify active promoters

The pools of promoter trap clones were transformed into electrocompetent wild-type and BPs lysogen *M. smegmatis* mc^2155 cells. 20 transformations were performed via electroporation with between 2ng and 85ng of the 3 pools of clones. The transformations were split into 100µl and 900µl volumes, plated on 7H10 CB CHX KAN $CaCl_2$ and incubated at 37°C for between 3 and 5 days. 3985 *M. smegmatis* transformants (wild-type 2802, BPs 1183) were produced and assessed visually for the presence a pink or purple color. The transformation plates were also scanned by fluorimetry for the presence of mCherry fluorescence on the FLA5000 fluorimeter (FujiFilm). 176 colonies were identified as positive for fluorescence and followed up on. The transformants were picked and struck out to ensure purity on 7H10 CB CHX KAN $CaCl_2$ plates.

C.1.6 Electroduction of plasmids into *E. coli*

Isolated, pure *M. smegmatis* clones were electroduced from *M. smegmatis* into *E. coli* XL1Blue cells.. *M. smegmatis* colonies were picked into 20µl of 10% glycerol, which was mixed by vortexing and frozen at -80°C and thawed to break apart cells. The glycerol containing *M. smegmatis* cells was added to 50µl of electrocompetent XL1Blue cells on ice. The cells were electroporated using the ECM 630 Electroporation System (BTX Harvard Appartatus) at 2.5kV, 25µF and 200Ω. Recovery media, 250µl of SOC, was added to the electroporated samples and they were recovered shaking at 37°C for 1 hour. Approximately 200µl of the recovered transformations were plated on LBA with Kan and grown overnight at 37°C. Liquid cultures inoculated from isolated colonies of the electroduction were immediately grown up. Plasmids from these cultures were isolated by GeneJet Miniprep Kit (Fermentas) according to

manufacturer's instructions or by miniprep using the Qiagen Biorobot. The DNA concentration of the plasmids were determined by Nanodrop. The plasmids or PCR products using primers LMO51 and LMO52 were sequenced through an outside service (GeneWiz) or in the University of Pittsburgh Genome Center (ABI 3730).

C.2 IDENTIFICATION OF BPS GENOMIC REGIONS WITH PROMOTER ACTIVITY

Genomic DNA (gDNA) from mycobacteriophage BPs was isolated and fragmented into 150 to 500 bp segments and ligated upstream of an mCherry reporter gene. Pools of clones that potentially had fragments of BPs gDNA were collected and transformed into electrocompetent *M. smegmatis* mc^2155 cells. Several transformations of each of the multiple separate ligations were screened for promoter activity visually, looking for pink color in the colonies with the naked eye. A total of approximately 1650 colonies were generated from ligations into the promoter trap vector in *E. coli*, and thus was the maximum number of possible different clones screened for promoter activity. A total of 20 independent transformations were performed into *M. smegmatis* mc^2155 or a BPs lysogen of *M. smegmatis* and 3985 colonies were screened for pink or purple color. Of these nearly 4000 colonies, 235 pink or purple colonies were identified and 171 isolated for further examination.

The plasmids from these 171 transformants were electroduced into *E. coli* and the plasmids were isolated and sequenced. 51 unique clones with promoter activity were identified (see Table 6-1). Many of these clones contained more than one fragment from separate regions of the BPs genome, comprising a total of 88 inserts at 55 distinct locations in the BPs genome.

The screen detected all of the previously identified BPs promoters: the gene *5-6* intergenic region (P_6), the gene *60-61* intergenic region (P_{61}) and in the forward and reverse direction in the gene *33-34* intergenic region (P_R and P_{rep}, respectively), providing proof that this method can identify promoters from the BPs genome.

Table 6-1. Plasmid clones with promoter activity isolated in promoter trap screen.

Promoter trap clone	BPs coordinates of insert	Genomic location	Times isolated
1A	R 25040-25252 R 32900-33358	27R, 42-43R	8
1B	38587-38828 19630-19749 40372-40790	IR54-55, gene20, gene60	7
1C	6260-6690	IR5-6	6
1D	8843-9110 R29366-29426	10, IR33-34	5
1E	40654-40849	60+IR	4
1F	38431-38922	IR54-55	2
1G	R 38354-38525	54	2
1H	35677-35815 33193-33566 35672-35869	47, 43, 47	1
1I	20588-20730 1575-1768	22, 2	1
1J	R 20536-21044 39983-40214	22R, 58	1
1K	R 16532-16715	17-18R	1
1L	R 37537-37677	52R	1
1M	37187-37290	51	1
1N	38453-38832?	IR54-55	1
1O	16919-17228	18	1
1P	15187-15403 39311-39403	16, 56	1
2A	R29373-29494 30662-30932 R604-889	IR33-34R, 36-38, 2R	2
2B	1065-1236	2	3
2C	R1987-2369	2-3R	1
2D	26867-27125 R2272-2645 R38704-38765	29-31, 3R, IR54-55R	1
2E	~15281-15483 R2264-2517 R11111-11253	16-17, 3R, 14-15R	3
2F	3662-4001	4	6
2G	5647-6189	4-5	1
2H	~R5574-5869 R15407-15624 38709-38874	4R, 16-17R, 55	2
2J	6163-6463	4-6	1
2K	R10432-10568 6402-6675	IR13-14R, IR5-6	2

Promoter trap clone	BPs coordinates of insert	Genomic location	Times isolated
2L	6527-6921 R975-1311	6, 2R	1
2M	12887-13105 7920-8316 R29239-25301	16, 7-9, 28-33R	2
2N	7840-8080	7-8	3
2O	9889-10341	12-13	3
2P	4404-4491 14190-14464 R10701-11000	4, 16, 14-15R	1
2Q	R14410-14656 3083-3121	16R, 3	2
2R	R14395-14975	16R	1
2S	R16101-16345 41688-41890	17R, 63-end	2
2T	37751-38006 18090-18358 6288-6429	52, 18, IR5-6	2
2U	R21198-21471 16952-17128	22R, 18	4
2V	R21429-21597 R21789-22018	22R, 22R	2
2W	R21830-22234	22R	1
2X	R24995-25212	27-28R	2
2Y	26625-26870	29-30	2
2Z	R28006-28229	32-33R	2
2AA	R28501-28766	32-33R	1
2AB	28481-28735	32	1
2AC	R29178-29400	IR33-34R	2
2AD	29288-29551	IR33-34	1
2AE	21943-22088 29391-29597	22, IR33-34	2
2AF	35401-35713	46-47	1
2AG	R1437-1609 36264-36692	2R, 49-50	2
2AH	15607-15761 37415-37735	17, 51-52	1
2AI	38707-38986 28729-28787	55-56, 32	1
2AJ	39443-39873	56-58	1
2AK	39980-40117	58	4
2AL	R40069-40333	58-59R	2
2AM	R41541-41763	63R	1
2AN	R41689-41890	63R	3

Only nine of the clones had particularly high activity (1F, 1N, 2C, 2K, 2R, 2X, 2Y, 2Z, 2AC; Table 6-1). Two of these clones with high activity (clones 1F and 1N) contained a similar fragment of gDNA that encompassed the region surrounding the gene *54-55* intergenic region. To confirm promoter activity in the *54-55* intergenic region, a promoter-reporter fusion was constructed (pLO46) and this segment of gDNA gave strong promoter activity (Fig. 6-3). In this independent clone of the *54-55* intergenic region, a slightly larger fragment of gDNA was designed into the construct to vary junctions in the reporter vector. The *54-55* promoter (P_{55}) is active in wild-type *M. smegmatis* but is shut off in a BPs lysogen (Fig. 6-3). However, when 31 additional regions were screened with independently constructed clones, none of the 31 clones gave promoter activity that could be identified on a plate with a visual screen. The promoter activity detected in the promoter trap clones from these regions could be due to promoters formed at the cloning junctions between the inserts and the vector.

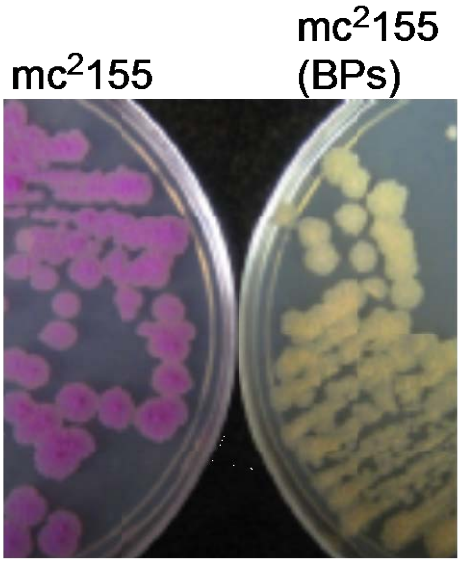

Figure 6-3. Fluorescent reporter activity of gene *54-55* intergenic region.

Fig. 6-3. The visible color to the naked eye of strains containing the promoter-reporter fusion of the *54-55* intergenic region driving mCherry expression (pLO46) in wild-type *M. smegmatis* and a BPs lysogen.

C.3 CONCLUSIONS

Using promoter trap to identify promoters throughout the BPs genome, the P_{55} promoter of BPs was identified. Transcription start site mapping has not yet been successful for P_{55} so the exact location of the promoter has not been identified, but the fragment of genomic DNA included spans from the end of gene *54*, covering the entire intergenic region, to the end of gene *55* (BPs coordinates 38587-38828). The only putative promoter sequence within this region falls within the gene *55* ORF and has the sequence 5'-TTGCGA and 5'-TGTGACACT for the -35 hexamer and the -10 nonamer, respectively (see Chapter 3, Fig. 3-6). Though the -10 hexamer (5'-GACACT) is not very similar to the mycobacterial -10 consensus sequence (5'-TATATT, see Chapter 4, Fig. 4-1), the presence of an extended -10 motif (5'-TGN) upstream most likely contributes to the strength of this promoter. P_{55} is the strongest promoter identified in mycobacteriophage BPs (see Chapter 3, Fig. 3-6).

The promoter trap method was able to identify P_{55}, a previously unidentified promoter, and was able to identify the promoter already identified in BPs, which served as a test that the experimental design and execution was sound. However, many of the potential promoters that were identified could not be verified with independent clones. These false positives were likely the result of the creation of active promoter sequences at the cloning junctions between the BPs gDNA fragment and the vector backbone.

APPENDIX D

TABLES OF PLASMIDS

D.1 CHAPTER 3 PLASMIDS

Table 6-2. Plasmids used in Chapter 3 of this study.

Number	BPs coordinates*	Description
pLO07	29224-29598	pJ promless gp9RBS mCherry bomb deriv, extended IR33-34 in forward orientation
pLO08	29224-29598	pJ promless gp9RBS mCherry bomb deriv, extended IR33-34 in reverse orientation
pLO30	41801-250	pJ promless mCherry, dig NotI+KpnI, ins PCR LMO154+155
pLO31	1541-1740	pJ promless mCherry, dig NotI+KpnI, ins PCR LMO156+157
pLO32	6285-6494	pJ promless mCherry, dig NotI+KpnI, ins PCR LMO158+159
pLO33	39720-39919	pJ promless mCherry, dig NotI+KpnI, ins PCR LMO160+161
pLO34	40653-40848	pJ promless mCherry, dig NotI+KpnI, ins PCR LMO162+163
pLO46	38400-39000	pJ promless deriv, ins IR54-55
pLO55	23663-24084	pJ promless deriv, ins IR26-27 (not located at DraI site)
pLO106	NA	pLO32 deriv, use SDM to add ScaI cut site immediately upstream of RBS
pLO108	41777-296	pJ promless deriv, dig NotI+KpnI, ins PCR LMO413+414redo
pLO109	1445-1818	pJ promless deriv, dig NotI+KpnI, ins PCR LMO385+386
pLO110	11471-12260	pJ promless deriv, dig NotI+KpnI, ins PCR LMO415+416redo
pLO111	16098-16575	pJ promless deriv, dig NotI+KpnI, ins PCR LMO389+390
pLO112	18231-18581	pJ promless deriv, dig NotI+KpnI, ins PCR LMO391+392
pLO113	18625-19108	pJ promless deriv, dig NotI+KpnI, ins PCR LMO393+394
pLO114	22215-23005	pJ promless deriv, dig NotI+KpnI, ins PCR LMO417+418redo
pLO115	24231-24968	pJ promless deriv, dig NotI+KpnI, ins PCR LMO397+398
pLO116	31516-31827	pJ promless deriv, dig NotI+KpnI, ins PCR LMO399+400
pLO117	31997-32469	pJ promless deriv, dig NotI+KpnI, ins PCR LMO401+402
pLO118	33959-34440	pJ promless deriv, dig NotI+KpnI, ins PCR LMO403+404
pLO119	35832-36303	pJ promless deriv, dig NotI+KpnI, ins PCR LMO405+406
pLO120	37192-37705	pJ promless deriv, dig NotI+KpnI, ins PCR LMO407+408
pLO121	39721-40099	pJ promless deriv, dig NotI+KpnI, ins PCR LMO419+420redo
pLO122	25837-26625	pJ promless deriv, dig NotI+KpnI, ins PCR LMO411+412
pLO124	3188-3604	pLO106 deriv, dig ScaI, ins PCR LMO208+209 term3-4
pLO125	6311-6703	pLO106 deriv, dig ScaI, ins PCR LMO 472+473 term5-6

pLO126	10351-10779	pLO106 deriv, dig ScaI, ins PCR LMO228+229 term13-14
pLO127	15268-15689	pLO106 deriv, dig ScaI, ins PCR LMO232+233 term16-17
pLO128	23422-23818	pLO106 deriv, dig ScaI, ins PCR LMO250+251 term25-26
pLO129	23662-24087	pLO106 deriv, dig ScaI, ins PCR LMO252+253 term26-27
pLO130	26918-27318	pLO106 deriv, dig ScaI, ins PCR LMO260+261 term30-31
pLO131	27537-27162	pLO106 deriv, dig ScaI, ins PCR LMO320+435 term31R
pLO132	29099-28681	pLO106 deriv, dig ScaI, ins PCR LMO314+315 term32-33R
pLO133	27946-28274	pLO106 deriv, dig ScaI, ins PCR LMO474+475 term32
pLO134	34185-34610	pLO106 deriv, dig ScaI, ins PCR LMO423+424 term43-44
pLO135	34749-35127	pLO106 deriv, dig ScaI, ins PCR LMO286+287 term45
pLO136	35342-35850	pLO106 deriv, dig ScaI, ins PCR LMO425+426 term47
pLO137	35715-35350	pLO106 deriv, dig ScaI, ins PCR LMO314+315 term47R
pLO138	38435-38855	pLO106 deriv, dig ScaI, ins PCR LMO427+428 term54-55
pLO139	38873-39199	pLO106 deriv, dig ScaI, ins PCR LMO429+430 term55-56
pLO140	39499-39852	pLO106 deriv, dig ScaI, ins PCR LMO431+432 term56-57
pLO141	39519-39180	pLO106 deriv, dig ScaI, ins PCR LMO433+434 term56R
pLO142	39832-40124	pLO106 deriv, dig ScaI, ins PCR LMO310+311 term58
pLO143	40623-40939	pLO106 deriv, dig ScaI, ins PCR LMO312+313 term60-61
pLO175	NA	deriv pLO106: ins PCR LMO476+477 of pJ promless (rrnB term), positive control terminator
pLO177	29099-28679	deriv pLO132: SDM with LMO486 to delete 5 bp from 3' cloning junction and remove a putative predicted promoter
pLO178	34185-34608	deriv pLO134: SDM with LMO487 to delete 5 bp from 3' cloning junction and remove a putative predicted promoter
pLO193	38254-38625	pJ promless deriv, dig NotI+KpnI, ins PCR LMO492+493
pLO202	24895-25312	pJ promless deriv, dig NotI+KpnI, ins PCR LMO510+511
pLO206	39343-39973	pJ promless deriv, dig NotI+KpnI, ins PCR LMO518+519
pLO209	26767-27225	pJ promless deriv, dig NotI+KpnI, ins PCR LMO524+525
pLO216	41588-41901	pJ promless deriv, dig NotI+KpnI, ins PCR LMO538+539
pLO219	36364-36792	pJ promless deriv, dig NotI+KpnI, ins PCR LMO544+545
pLO221	37315-37835	pJ promless deriv, dig NotI+KpnI, ins PCR LMO548+549
pLO223	30532-31032	pJ promless deriv, dig NotI+KpnI, ins PCR LMO552+553

* Red BPs coordinates indicate that BPs regions are found in the reverse orientation within the constructs.

D.2 CHAPTER 4 PLASMIDS

Table 6-3. Plasmids used in Chapter 4 of this study.

Plasmid	Promoter mutations	BPs genomic reference	Description	Reference
pLO86	NA	NA	Replicative parental vector with no promoter driving mCherry expression	Broussard et al 2013
pLO87	NA	NA	Replicative parental vector with P_{hsp60} driving mCherry expression	Broussard et al 2013

pLO07	none	none	Replicative vector with P_R (genes *33-34* intergenic region, coordinates 29224-29598) driving mCherry expression	Broussard et al 2013
pLO224	T-37A	T29473A	pLO07 site directed mutant with point mutation T>A at P_R -37 (BPs coordinate 29473)	this study
pLO225	T-37C	T29473C	pLO07 site directed mutant with point mutation T>C at P_R -37 (BPs coordinate 29473)	this study
pLO226	T-37G	T29473G	pLO07 site directed mutant with point mutation T>G at P_R -37 (BPs coordinate 29473)	this study
pLO227	T-36A	T29474A	pLO07 site directed mutant with point mutation T>A at P_R -36 (BPs coordinate 29474)	this study
pLO228	T-36C	T29474C	pLO07 site directed mutant with point mutation T>C at P_R -36 (BPs coordinate 29474)	this study
pLO229	T-36G	T29474G	pLO07 site directed mutant with point mutation T>G at P_R -36 (BPs coordinate 29474)	this study
pLO230	T-35A	T29475A	pLO07 site directed mutant with point mutation T>A at P_R -35 (BPs coordinate 29475)	this study
pLO231	T-35C	T29475C	pLO07 site directed mutant with point mutation T>C at P_R -35 (BPs coordinate 29475)	this study
pLO232	T-35G	T29475G	pLO07 site directed mutant with point mutation T>G at P_R -35 (BPs coordinate 29475)	this study
pLO233	C-34A	C29476A	pLO07 site directed mutant with point mutation C>A at P_R -34 (BPs coordinate 29476)	this study
pLO234	C-34G	C29476G	pLO07 site directed mutant with point mutation C>G at P_R -34 (BPs coordinate 29476)	this study
pLO235	C-34T	C29476T	pLO07 site directed mutant with point mutation C>T at P_R -34 (BPs coordinate 29476)	this study
pLO236	C-33A	C29477A	pLO07 site directed mutant with point mutation C>A at P_R -33 (BPs coordinate 29477)	this study
pLO237	C-33G	C29477G	pLO07 site directed mutant with point mutation C>G at P_R -33 (BPs coordinate 29477)	this study
pLO238	C-33T	C29477T	pLO07 site directed mutant with point mutation C>T at P_R -33 (BPs coordinate 29477)	this study
pLO239	A-32C	A29478C	pLO07 site directed mutant with point mutation A>C at P_R -32 (BPs coordinate 29478)	this study
pLO240	A-32G	A29478G	pLO07 site directed mutant with point mutation A>G at P_R -32 (BPs coordinate 29478)	this study
pLO241	A-32T	A29478T	pLO07 site directed mutant with point mutation A>T at P_R -32 (BPs coordinate 29478)	this study
pLO242	C-16A	C29494A	pLO07 site directed mutant with point mutation C>A at P_R -16 (BPs coordinate 29494)	this study
pLO243	C-16G	C29494G	pLO07 site directed mutant with point mutation C>G at P_R -16 (BPs coordinate 29494)	this study
pLO244	C-16T	C29494T	pLO07 site directed mutant with point mutation C>T at P_R -16 (BPs coordinate 29494)	this study

pLO245	G-15A	G29495A	pLO07 site directed mutant with point mutation G>A at P_R -15 (BPs coordinate 29495)	this study
pLO246	G-15C	G29495C	pLO07 site directed mutant with point mutation G>C at P_R -15 (BPs coordinate 29495)	this study
pLO247	G-15T	G29495T	pLO07 site directed mutant with point mutation G>T at P_R -15 (BPs coordinate 29495)	this study
pLO248	C-14A	C29496A	pLO07 site directed mutant with point mutation C>A at P_R -14 (BPs coordinate 29496)	this study
pLO249	C-14G	C29496G	pLO07 site directed mutant with point mutation C>G at P_R -14 (BPs coordinate 29496)	this study
pLO250	C-14T	C29496T	pLO07 site directed mutant with point mutation C>T at P_R -14 (BPs coordinate 29496)	this study
pLO251	T-13A	T29497A	pLO07 site directed mutant with point mutation T>A at P_R -13 (BPs coordinate 29497)	this study
pLO252	T-13C	T29497C	pLO07 site directed mutant with point mutation T>C at P_R -13 (BPs coordinate 29497)	this study
pLO253	T-13G	T29497G	pLO07 site directed mutant with point mutation T>G at P_R -13 (BPs coordinate 29497)	this study
pLO254	A-12C	A29498C	pLO07 site directed mutant with point mutation A>C at P_R -12 (BPs coordinate 29498)	this study
pLO255	A-12G	A29498G	pLO07 site directed mutant with point mutation A>G at P_R -12 (BPs coordinate 29498)	this study
pLO256	A-12T	A29498T	pLO07 site directed mutant with point mutation A>T at P_R -12 (BPs coordinate 29498)	this study
pLO257	T-11A	T29499A	pLO07 site directed mutant with point mutation T>A at P_R -11 (BPs coordinate 29499)	this study
pLO258	T-11C	T29499C	pLO07 site directed mutant with point mutation T>C at P_R -11 (BPs coordinate 29499)	this study
pLO259	T-11G	T29499G	pLO07 site directed mutant with point mutation T>G at P_R -11 (BPs coordinate 29499)	this study
pLO260	G-10A	G29500A	pLO07 site directed mutant with point mutation G>A at P_R -10 (BPs coordinate 29500)	this study
pLO261	G-10C	G29500C	pLO07 site directed mutant with point mutation G>C at P_R -10 (BPs coordinate 29500)	this study
pLO262	G-10T	G29500T	pLO07 site directed mutant with point mutation G>T at P_R -10 (BPs coordinate 29500)	this study
pLO263	T-9A	T29501A	pLO07 site directed mutant with point mutation T>A at P_R -9 (BPs coordinate 29501)	this study
pLO264	T-9C	T29501C	pLO07 site directed mutant with point mutation T>C at P_R -9 (BPs coordinate 29501)	this study
pLO265	T-9G	T29501G	pLO07 site directed mutant with point mutation T>G at P_R -9 (BPs coordinate 29501)	this study
pLO266	T-8A	T29502A	pLO07 site directed mutant with point mutation T>A at P_R -8 (BPs coordinate 29502)	this study
pLO267	T-8C	T29502C	pLO07 site directed mutant with point mutation T>C at P_R -8 (BPs coordinate 29502)	this study

pLO268	T-8G	T29502G	pLO07 site directed mutant with point mutation T>G at P_R -8 (BPs coordinate 29502)	this study
pLO156	Δ1 bp	Δ29481	pLO07 site directed mutant with deletion of 1 bp from the P_R spacer region (BPs coordinate 29481)	this study
pLO157	Δ2 bp	Δ29481 Δ29482	pLO07 site directed mutant with deletion of 2 bp from the P_R spacer region (BPs coordinate 29481-29482)	this study
pLO158	+1 bp	ins29482T	pLO07 site directed mutant with insertion of 1 bp from the P_R spacer region (BPs coordinate 29482)	this study
pLO269	C-26G	C29484G	pLO07 site directed mutant with point mutation C>G at O_R -26 (BPs coordinate 29484)	this study
pLO270	C-26T	C29484T	pLO07 site directed mutant with point mutation C>T at O_R -26 (BPs coordinate 29484)	this study
pLO271	G-25A	G29485A	pLO07 site directed mutant with point mutation G>A at O_R -25 (BPs coordinate 29485)	this study
pLO272	G-25C	G29485C	pLO07 site directed mutant with point mutation G>C at O_R -25 (BPs coordinate 29485)	this study
pLO273	G-25T	G29485T	pLO07 site directed mutant with point mutation G>T at O_R -25 (BPs coordinate 29485)	this study
pLO274	A-24C	A29486C	pLO07 site directed mutant with point mutation A>C at O_R -24 (BPs coordinate 29486)	this study
pLO275	A-24G	A29486G	pLO07 site directed mutant with point mutation A>G at O_R -24 (BPs coordinate 29486)	this study
pLO276	A-24T	A29486T	pLO07 site directed mutant with point mutation A>T at O_R -24 (BPs coordinate 29486)	this study
pLO277	C-23A	C29487A	pLO07 site directed mutant with point mutation C>A at O_R -23 (BPs coordinate 29487)	this study
pLO278	C-23G	C29487G	pLO07 site directed mutant with point mutation C>G at O_R -23 (BPs coordinate 29487)	this study
pLO279	C-23T	C29487T	pLO07 site directed mutant with point mutation C>T at O_R -23 (BPs coordinate 29487)	this study
pLO280	A-22C	A29488C	pLO07 site directed mutant with point mutation A>C at O_R -22 (BPs coordinate 29488)	this study

pLO281	A-22G	A29488G	pLO07 site directed mutant with point mutation A>G at O_R -22 (BPs coordinate 29488)	this study
pLO282	A-22T	A29488T	pLO07 site directed mutant with point mutation A>T at O_R -22 (BPs coordinate 29488)	this study
pLO283	T-21A	T29489A	pLO07 site directed mutant with point mutation T>A at O_R -21 (BPs coordinate 29489)	this study
pLO284	T-21C	T29489C	pLO07 site directed mutant with point mutation T>C at O_R -21 (BPs coordinate 29489)	this study
pLO285	T-21G	T29489G	pLO07 site directed mutant with point mutation T>G at O_R -21 (BPs coordinate 29489)	this study
pLO286	A-20C	A29490C	pLO07 site directed mutant with point mutation A>C at O_R -20 (BPs coordinate 29490)	this study
pLO287	A-20G	A29490G	pLO07 site directed mutant with point mutation A>G at O_R -20 (BPs coordinate 29490)	this study
pLO289	A-20T	A29490T	pLO07 site directed mutant with point mutation A>T at O_R -20 (BPs coordinate 29490)	this study
pLO290	T-19A	T29491A	pLO07 site directed mutant with point mutation T>A at O_R -19 (BPs coordinate 29491)	this study
pLO291	T-19C	T29491C	pLO07 site directed mutant with point mutation T>C at O_R -19 (BPs coordinate 29491)	this study
pLO291	T-19G	T29491G	pLO07 site directed mutant with point mutation T>G at O_R -19 (BPs coordinate 29491)	this study
pLO292	G-18C	G29492C	pLO07 site directed mutant with point mutation G>C at O_R -18 (BPs coordinate 29492)	this study
pLO293	G-18T	G29492T	pLO07 site directed mutant with point mutation G>T at O_R -18 (BPs coordinate 29492)	this study
pLO294	T-17A	T29493A	pLO07 site directed mutant with point mutation T>A at O_R -18 (BPs coordinate 29493)	this study
pLO295	T-17G	T29493A	pLO07 site directed mutant with point mutation T>G at O_R -18 (BPs coordinate 29493)	this study
pTTB1b	NA	NA	Tweety integrative parental vector	Pham et al 2007

pLO73	NA	NA	pTTP1b with promoterless mCherry (pLO86)	this study
pLO74	NA	NA	pTTP1b with P_{hsp60} mCherry (pLO87)	this study
pLO75	none	none	pTTP1b with P_R mCherry (pLO07)	this study
pLO76	C-16T	C29494T	pTTP1b with P_R C29494T-mCherry (pLO244)	this study
pLO77	C-16T T-9C	C29494T T29501C	pLO76 site directed mutant with 2 point mutations: C-16T in extended -10 and T-9C in -10	this study
pLO78	C-16T T-9C G-10A	C29494T T29501C G29500A	pLO77 site directed mutant with 2 point mutations: C-16T in extended -10 and T-9C, G-10A in -10	this study
pLO80	T-35G C-16T	T29475G C29494T	pLO76 site directed mutant with 2 point mutations: C-16T in extended -10 and T-35G in -35	this study
pLO81	T-35G T-34G C-16T	T29475G T29476G C29494T	pLO76 site directed mutant with 3 point mutations: C-16T in extended -10 and T-35G, T-34G in -35	this study
pLO82	T-37C T-35G T-34G C-16T	T29473C T29475G T29476G C29494T	pLO76 site directed mutant with 4 point mutations: C-16T in extended -10 and T-35G, T-34G, T-37C in -35	this study
pLO83	T-35G C-16T T-9C	T29475G C29494T T29501C	pLO77 site directed mutant with 3 point mutations: C-16T in extended -10, T-9C in -10, and T-35G in -35	this study
pLO84	T-35G T-34G C-16T T-9C	T29475G T29476G C29494T T29501C	pLO77 site directed mutant with 4 point mutations: C-16T in extended -10, T-9C in -10, and T-35G, T-34G in -35	this study
pLO85	T-37C T-35G T-34G C-16T T-9C	T29473C T29475G T29476G C29494T T29501C	pLO77 site directed mutant with 4 point mutations: C-16T in extended -10, T-9C in -10, and T-35G, T-34G, T-37C in -35	this study
pLO92	C-16T C-14G T-9C G-10A	C29494T C29496G T29501C G29500A	pLO78 site directed mutant with 3 point mutations: C-16T, C-14G in extended -10 and T-9C, G-10A in -10	this study
pLO93	T-35G C-16T C-14G T-9C G-10A	T29475G C29494T C29496G T29501C G29500A	pLO83 site directed mutant with 4 point mutations: C-16T, C-14G in extended -10 and T-9C, G-10A in -10 and T-35G in -35	this study
pLO94	T-35G T-34G C-16T C-14G T-9C G-10A	T29475G T29476G C29494T C29496G T29501C G29500A	pLO84 site directed mutant with 4 point mutations: C-16T, C-14G in extended -10 and T-9C, G-10A in -10 and T-35G, T-34G in -35	this study
pLO95	T-37C T-35G T-34G C-16T C-14G T-9C G-10A	T29473C T29475G T29476G C29494T C29496G T29501C G29500A	pLO85 site directed mutant with 4 point mutations: C-16T, C-14G in extended -10 and T-9C, G-10A in -10 and T-35G, T-34G, T-37C in -35	this study

pLO159	T-35G C-16T C-14G T-9C G-10A Δ1	T29475G C29494T C29496G T29501C G29500A Δ29482	pLO93 site directed mutant with 4 point mutations: C-16T, C-14G in extended -10 and T-9C, G-10A in -10 and T-35G in -35, and single base deletion in spacer	this study
pLO160	C-34A C-16T C-14G T-9C G-10A	C29476A C29494T C29496G T29501C G29500A	pLO83 site directed mutant with 4 point mutations: C-16T, C-14G in extended -10 and T-9C, G-10A in -10 and C-34A in -35	this study
pLO161	T-35G C-34A C-16T C-14G T-9C G-10A	T29475G C29476A C29494T C29496G T29501C G29500A	pLO83 site directed mutant with 4 point mutations: C-16T, C-14G in extended -10 and T-9C, G-10A in -10 and T-35G, C-34A in -35	this study
pLO162	T-37C T-36G T-35G C-34A C-16T C-14G T-9C G-10A	T29473C T29474G T29475G C29476A C29494T C29496G T29501C G29500A	pLO83 site directed mutant with 4 point mutations: C-16T, C-14G in extended -10 and T-9C, G-10A in -10 and T-37C, T-36G, T-35G, C-34A in -35	this study
pLO168	T-35G C-16T C-14G T-9C G-10A Δ1	T29475G C29494T C29496G T29501C G29500A Δ29481	pLO93 site directed mutant with 4 point mutations: C-16T, C-14G in extended -10 and T-9C, G-10A in -10 and T-35G in -35, and single base deletion in spacer	this study
pLO169	C-34A C-16T C-14G T-9C G-10A Δ1	C29476A C29494T C29496G T29501C G29500A Δ29481	pLO83 site directed mutant with 4 point mutations: C-16T, C-14G in extended -10 and T-9C, G-10A in -10 and C-34A in -35 and single base deletion in spacer	this study
pLO170	T-35G T-34A C-16T C-14G T-9C G-10A Δ1	T29475G T29476G C29494T C29496G T29501C G29500A Δ29481	pLO161 site directed mutant with 7 point mutations: C-16T, C-14G in extended -10 and T-9C, G-10A in -10 and T-35G, T-34G in -35, and single base deletion in spacer	this study
pLO185	T-35G C-16T T-9C G-10A	T29475G C29494T T29501C G29500A	pLO78 site directed mutant with 4 point mutations: C-16T in extended -10 and T-9C, G-10A in -10 and T-35G in -35	this study
pLO186	T-35G T-34G C-16T T-9C G-10A	T29475G T29476G C29494T T29501C G29500A	pLO78 site directed mutant with 5 point mutations: C-16T in extended -10 and T-9C, G-10A in -10 and T-35G, T-34G in -35	this study

pLO187	T-35G T-34G T-37C C-16T T-9C G-10A	T29473C T29475G T29476G C29494T T29501C G29500A	pLO78 site directed mutant with 6 point mutations: C-16T in extended -10 and T-9C, G-10A in -10 and T-35G, T-34G, T-37G in -35	this study
pLO190	T-35G T-34G C-16T T-9C G-10A Δ1	T29475G T29476G C29494T T29501C G29500A Δ29481	pLO78 site directed mutant with 5 point mutations: C-16T in extended -10 and T-9C, G-10A in -10 and T-35G, T-34G in -35 and single base deletion in spacer	this study

D.3 CHAPTER 5 PLASMIDS

Table 6-4. Plasmids used in Chapter 5 of this study.

Plasmid	Features	Description	Antibiotic Resistance	Reference
pGWB43	$IR^{rep,cro}$-BPs $gp33^{136}$	L5 integrative vector expressing BPs viral repressor	Kan	Broussard et al., 2013
pGWB48	$IR^{rep,cro}$-BPs $gp33^{103}$	L5 integrative vector expressing BPs prophage repressor	Kan	Broussard et al., 2013
pGWB66	P_{hsp60}:BPs $gp33^{136}$	Replicative vector, expresses BPs $gp33^{136}$ from *hsp60* promoter	Kan	Broussard et al., 2013
pGWB76	P_R-BPs gp34	L5 integrative vector expressing BPs *cro* from P_R	Kan	Broussard et al., 2013
pGWB78	P_RT-21C-BPs gp34	L5 integrative vector expressing BPs *cro* from P_R with the point mutation found in the repressor insensitive mutant BPs 102a	Kan	Broussard et al., 2013
pGWB81	Brujita gp33-gp34-$IR^{rep,cro}$	Brujita integrative vector containing *int*, *rep*, *rep-cro* intergenic region	Kan	Broussard et al., 2013
pGWB87	Brujita gp33A296E	pGWB81 site-directed mutant expresses stable form of Brujita integrase with penultimate alanine to glutamic acid substitution	Kan	Broussard et al., 2013
pGWB43GoF1	BPs gp33A135E	BPs repressor gain of function mutant, penultimate alanine to glutamic acid substitution	Kan	Broussard et al., 2013.
pGWB43GoF2	BPs $gp33^{102Shift}$	BPs repressor gain of function mutant, frame shift at amino acid 102, C insertion at base 298	Kan	Broussard et al., 2013
pGWB43GoF3	BPs $gp33^{112Shift}$	BPs repressor gain of function mutant, frame shift at amino acid 112, A insertion at base 327	Kan	Broussard et al., 2013
pGWB43GoF4	BPs $gp33^{114Shift}$	BPs repressor gain of function mutant, frame shift at amino acid 114, T insertion at base 338	Kan	Broussard et al., 2013
pGWB43GoF5	BPs $gp33^{111Shift}$	BPs repressor gain of function mutant, frame shift at amino acid 111, ΔA327	Kan	Broussard et al., 2013
pGWB43GoF6	BPs $gp33^{101Shift}$	BPs repressor gain of function mutant, frame shift at amino acid 101, ΔC298	Kan	Broussard et al., 2013
pGWB43GoF7	BPs $gp33^{88Shift}$	BPs repressor gain of function mutant, frame shift at amino acid 88, ΔG261	Kan	Broussard et al., 2013
pNIT	P_{NIT}	Replicative vector with nitrile inducible promoter	Kan	Pandey et al., 2009

pNIT:GFP	eGFP	eGFP expressed from nitrile inducible promoter	Kan	Pandey et al., 2009
pLO16	GFP-VLPGPWSTLDKAS	pNIT:GFP with C-term 13 aa from BPs viral repressor fused to GFP	Kan	Broussard et al., 2013
pLO20	GFP-VLPGPWSTLDKES	pLO16 site directed mutant to make A135E substitution	Kan	Broussard et al., 2013
pLO07	P_R-mCherry	Replicative vector with P_R (forward orientation of BPs *rep, cro* intergenic region, coordinates 29224-29598) driving mCherry expression	Kan	Broussard et al., 2013
pLO08	P_{rep}-mCherry	Replicative vector with P_{rep} (reverse orientation of BPs *rep, cro* intergenic region, coordinates 29224-29598) driving mCherry expression	Kan	Broussard et al., 2013
pLO09	BPs gp33^{136}	pGWB43 Kan^R cassette replaced with Hyg^R cassette	Hyg	Broussard et al., 2013
pLO10	BPs gp33^{103}	pGWB48 Kan^R cassette replaced with Hyg^R cassette	Hyg	Broussard et al., 2013
pLO15	BPs gp33A135E	pGWB43GoF1 Kan^R cassette replaced with Hyg^R cassette	Hyg	Broussard et al., 2013
pLO21	BPs gp33^{102Shift}	pGWB43GoF2 Kan^R cassette replaced with Hyg^R cassette	Hyg	this study
pLO26	P_{rep}A-12G-mCherry	pLO08 with point mutation found in clear-plaque mutant BPs Δ32 Clr2 (BPs coordinate T29336C)	Kan	Broussard et al., 2013
pLO07:A-24C	P_RA-24C-mCherry	pLO07 site directed mutant with point mutation found in repressor insensitive mutant BPs 102e (BPs coordinate A29486C)	Kan	Broussard et al., 2013
pLO07:T-21C	P_RT-21C-mCherry	pLO07 site directed mutant with point mutation found in repressor insensitive mutant BPs 102a (BPs coordinate T29489C)	Kan	Broussard et al., 2013
pLO64	GFP-Brujita gp33 C-term 5aa	pNIT:GFP with C-term 5aa from Brujita integrase fused to GFP	Kan	Broussard et al., 2013
pLO65	GFP-Brujita gp33A296E C-term 5aa	pLO64 site directed mutant to make A296E substitution	Kan	Broussard et al., 2013
pLO24	GFP-64 aa of BPs gp32	pNIT:GFP with C-term 64 aa from BPs integrase fused to GFP	Kan	this study
pLO27	GFP-54 aa of BPs gp32	pLO24 site directed mutant with stop codon to form 54 aa tag	Kan	this study
pLO28	GFP-44 aa of BPs gp32	pLO24 site directed mutant with stop codon to form 44 aa tag	Kan	this study
pLO29	GFP-34 aa of BPs gp32	pLO24 site directed mutant with stop codon to form 34 aa tag	Kan	this study
pLO56	GFP-24 aa of BPs gp32	pLO24 site directed mutant with stop codon to form 24 aa tag	Kan	this study
pLO57	GFP-14 aa of BPs gp32	pLO24 site directed mutant with stop codon to form 14 aa tag	Kan	this study
pLO70	GFP-64 aa of BPs gp32	pLO24 site directed mutant with BPs gp32 G347D A348D amino acid subsititutions in context of full length 64 aa gp32 tag on GFP	Kan	this study
pLO71	GFP-11 aa of BPs gp32	pLO24 site directed mutant with stop codon to form 11 aa tag	Kan	this study
pLO72	GFP-4 aa of BPs gp32	pLO24 site directed mutant with stop codon to form 4 aa tag	Kan	this study
pLO59	P_R-BPs gp34	Giles integrative vector expressing BPs *cro* from P_R	Hyg	Broussard et al., 2013
pLO61	P_RT29489C-BPs gp34	Giles integrative vector expressing BPs *cro* from P_R with the point mutation found in the repressor insensitive mutant BPs 102a	Hyg	Broussard et al., 2013
pMH94		L5 integrative vector, used for plasmid construction and cloning	Kan	Lee et al., 1991
pJL37	P_{hsp60}	Replicative vector with *hsp60* promoter	Kan	Lewis and Hatfull, 2000
pGH1000b		Giles integrative vector, used for plasmid construction	Hyg	Morris et al.,

		and cloning		2008

D.4 PRIMERS

Table 6-5. Primers used in this study.

Name	Sequence	Description
LMO01	ccgcggacctacgggcgatcggccgagcacaacggcgaaaggcgaagtgattaattaacagaaaggaggttaata	BPs forward primer 50bp homology BPs to gene54 25bp homology GFP
LMO02	taccggaatattcccgcgccctcagcggcccgcagagcgacgaacagccgatctacttgtacagctcgtccatgc	BPs reverse primer 50bp homolgy BPs to gene54 25bp homology GFP
LMO03	cgtgttccgggccgtgatgcagtccgaatggatcttcggccggatcctggccgcggacctacgggcgatcggccg	BPs forward primer 50bp homology BPs to gene54 25bp homology product of LMO01/ LMO02
LMO04	acaagtgtcacacattggcgacatatgtcgcaagcgttcaccatgtccgataccggaatattcccgcgccctcag	BPs reverse primer 50bp homology BPs to gene54 25bp homology product of LMO01/LMO02
LMO05	gcatgagctgggaaccgattctggc	Upstream Diagnostic primer for BPs gene54 v GFP
LMO06	cattggcgacatatgtcgcaagcgttcac	Downstream Diagnostic primer for BPs gene54 v GFP
LMO07	gcacaacggcgaaaggcgaagtgattaa	MAMA PCR screen for mutant BPs with GFP
LMO08	cgtcaacggccacaagttctccgtc	Upstream diagnostic primer that sits within GFP
LMO09	cacgcaccgcaccgcatatgtcgctatccgacatatgtcgtggatgatcgcgcatgcttaattaacagaaaggag	BPs Lysin A forward primer with GFP homology
LMO10	gatcagcgggtgcgtgtcgccgacgtgcgcggtgcttggtaccgccatcacttgtacagctcgtccatgccgtgg	BPs Lysin A reverse primer with GFP homology, eliminates GFP stop codon and replaces with LysA stop codon
LMO11	gggcgcaccggcgcgcggctgccgtggcgctgattctgaggtaaatcgctcacgcaccgcaccgcatatgtcgct	Extender for LMO09
LMO12	ccgatcgccttgccgtatgagttgccggccagcttccgtttcgccgccgggatcagcgggtgcgtgtcgccgacg	Extender for LMO10
LMO13	cgcaccgcatatgtcgctatccgac	Forward diagnostic primer in front of LysA
LMO14	gccgacgttgcgcggtgcttggtac	Reverse diagnostic primer after LysA
LMO15	ggtacgtcttcaccgaccgcttcgtggtcatcgaaaacgccgtcgcgtaagcatgcttaattaacagaaaggagg	BPs insertion after capsid gene forward primer with GFP homology
LMO16	ttcatgagtcgcatgtgtcgccccttccgagcggtgagttcggatcgggctacttgtacagctcgtccatg	BPs insertion after capsid gene reverse primer with GFP homology
LMO17	cctgaagcgccacaaccagatcgcgcttcggctcgaaatcgtctatggctggtacgtcttcaccgaccgcttcgt	Extender for LMO15
LMO18	ccagctggtcgccgagctcgtcgggcacgttgaccacgacgccgtgttcgttcatgagtcgcatgtgtcgcccct	Extender for LMO16
LMO19	gtcgagctcatcaagtacggcgatc	Forward diagnostic primer within capsid
LMO20	gttgaccacgacgccgtgttcgttc	Reverse diagnostic primer within gene 8
LMO21	catcgaaaacgccgtcgcgtaagca	MAMA PCR screen for GFP insertion after capsid, forward primer
LMO22	gttcggttagggcctcgactacgcg	reverse flanking outside substrate to replace gene54
LMO23	cttcgagccgccgagcacgtcccag	reverse flanking outside substrate of capsid GFP insertion
LMO25	ccgcggacctacgggcgatcggccgagcacaacggcgaaaggcgaagtgatctagaggtgaccacaacgacgcgc	Forward to make hsp60promoter before GFPgene54 insertion

LMO26	tcgaccaggatcgggacgacgccggtgaacagctcctcgcccttcgacatatggaagtgattcctccggatcggg	Reverse to make hsp60promoter before GFPgene54 insertion
LMO27	cgtgttccgggccgtgatgcagtccgaatggatcttcggccggatcctggccgcggacctacgggcgatc	Extender for LMO25
LMO28	caccctcgccggagacggagaacttgtggccgttgacgtcaccgtccagctcgaccaggatcgggacgacgc	Extender for LMO26
LMO29	cgctcggcggatatctgttc	Forward diagnostic for hsp60promoter insertion (outside substrate)
LMO30	gcttctcgttcgggtccttg	Reverse diagnostic for hsp60promoter insertion (outside substrate)
LMO31	ggtcttgttgtcgttggcgg	Forward diagnostic within hsp60promoter
LMO33	accggcgattgtatgtacgtctgtcgc	Forward to amplify intergenic region between rep and xis of BPs
LMO34	agcgacatatgtcgtctattggaaaatgcgc	Reverse to amplify intergenic region between rep and xis of BPs
LMO35	caccatcaccattgagcactggc	Reverse to diagnose "bad" hsp60pro mutant (bad = 174bp)
LMO36	cagacgtccccgatcaaagcg	Reverse to diagnose "good" hsp60pro mutant (good = 181bp)
LMO37	gcgagttcttcttgcttcaaaccagcttcaag	Forward to amplify extended (+100bp) intergenic region between genes 33 and 34 of BPs
LMO38	gcgcggtgtagtgatcgcttcggaa	Reverse to amplify extended intergenic region between genes 33 and 34 of BPs
LMO39	tcaccctcgccggagacggag	Forward for sequencing derivatives of pMN437 (sits ~100bp inside gfp)
LMO40	gagcgcccaatacgcaaaccgcc	Reverse for sequencing derivatives of pMN437 (sits between insert and colE ori)
LMO41	cggacctacgggcgatcggccgagcacaacggcgaaaggcgaagtgattactgcagctagggcaccaatttgcgattag	Forward to amplify L5 Pleft and make recombineering substrate to place in front of gfp in BPs GFPg54
LMO42	cggtgaacagctcctcgcccttcgacattattaacctcctttctgttaatccggtccaagttggtgatgcttgcag	Reverse to amplify L5 Pleft and make recombineering substrate to place in front of gfp in BPs GFPg54
LMO43	tacaccgcgctcgtaggggacttcgtgcaatgacattccgagaagaagccgatccggccggcgaaacgtagccacacacgacgcttgcaccgcgacatat	Deletion oligo to delete BPs gene5. Leaves 21bp of coding seq at both the 5' and 3' ends of gene5.
LMO44	ggccgatcatgagacagatgattttggacggctcgacgaacttcagttcgtacaccgcgctcgtaggggacttcg	Forward extender for LMO43. Adds another 50bp of homology to BPs.
LMO45	cgttcggtaacggatagccgccaggttaacccgcatatgtccaaaaccgacatatgtcgcggtgcaagcgtcgtgtgtg	Reverse extender for LMO43. Adds another 50bp of homology to BPs.
LMO46	gcagctagggcaccaatttgcgattag	Foward diagnostic, sits inside Pleft promoter
LMO47	gtccacaaccacattctcgagcggatc	Forward diagnostic flanking for BPs gene5 deletion
LMO48	ctggtcctggttgtcgtgctggtc	Reverse diagnostic flanking for BPs gene5 deletion
LMO49	gttccgggccgtgatgcagtccgaatggatcttcggccggatcctggccgcggacctacgggcgatcggcc	F extender for LMO41
LMO50	cttgtggccgttgacgtcaccgtccagctcgaccaggatcgggacgacgccggtgaacagctcctcgcccttcga	Reverse extender for LMO42
LMO51	gccatgcctgtctcgttgcgtg	F for sequencing of pJ promless gp9RBS mCherry bomb and derivatives
LMO52	ggaagctcagcttcaggtagtcg	R for sequencing of pJ promless gp9RBS mCherry bomb and derivatives
LMO53	ggcgctcactcgacgagcac	F new flanking diagnostic for Pleft insertion at GFPg54 (binds upstream of LMO29)
LMO54	ccgtcctccttgaagtcgatgc	R new flanking diagnostic for Pleft insertion at GFPg54 (binds upstream of LMO30)

LMO55	gacattccgagaagaagccgatcc	DADA primer for BPs gene 5 deletion
LMO56	gaacaataaaactgtctgcttacataaacagtaatacaaggcatgcctgcaggtcgactctagagg	F E. coli recombineering substrate-- to replace Kan with Hyg in pGWB43&48 lower -- 40bp homology to pGWB43&48, UPPER -- to amp Hyg of pJV39 (is Greg's primer Hyg_ampl_fwd)
LMO57	gataacaatttcacacaggaaacagctatgaccatgattacgcggccagtgaattcgagctcggtacc	R E. coli recombineering substrate-- to replace Kan with Hyg in pGWB43&48 lower -- 43bp homology to pGWB43&48, UPPER -- to amp Hyg of pJV39 (is Greg's primer Hyg_ampl_rvs)
LMO58	ctggaacacggcgacggtggttg	R to amp BPs putative Preg without putative term/RNaseIII site
LMO59	ggcgaatcggtggcgcagcttg	F to amp BPs putative Preg (can be used with LMO58 or 60 to make with or without term/RNaseIII)
LMO60	ctactgatcgcgcgccttgaagc	R to amp BPs putative Preg WITH putative term/RNaseIII site (stops before start codon of gene 33)
LMO61	cctgccaaacacaagcaacggaggtacgc	F to amp GFP from pNITepiEGFP, keeps NdeI cloning site
LMO62	ggttcgaaaagcttttacgacgctttatccagggttgaccacgggcccggaaggaccttgtacagctcgtccatgc	R to amp and tag GFP of pNITepiEGFP. tag = WT BPs gp33 tag (last 13aa), keeps HindIII cloning site
LMO63	ggttcgaaaagcttttacgactctttatccagggttgaccacgggcccggaaggaccttgtacagctcgtccatgc	R to amp and tag GFP of pNITepiEGFP. tag = point mut (KES) BPs gp33 tag (last 13aa), keeps HindIII cloning site
LMO64	ggttcgaaaagcttttatgccgcgtgctcattcggaccctcctggcggaggggatcttgtacagctcgtccatgc	R to amp and tag GFP of pNITepiEGFP. tag = Brujita gp34 tag (last 13aa), keeps HindIII cloning site
LMO65	ggttcgaaaagcttttaggcagcgagggcgtagtcgcgctgattggaatcggcgctcttgtacagctcgtccatgc	R to amp and tag GFP of pNITepiEGFP. tag = smeg ssrA tag (SADSNQRDYALAA), keeps HindIII cloning site
LMO66	gcccggaaggaccttgtacagctcgtccatgc	1 of 3, R to add WT BPs gp33 13aa C-terminal tag to GFP of pNIT:GFP (notebook LMO4, p.63)
LMO67	cagggttgaccacgggcccggaaggacctt	2 of 3 (see LMO66)
LMO68	ggttcgaaaagcttttacgacgctttatccagggttgaccacgg	3 of 3 (see LMO66)
LMO69	caagggtcctggcattctgcgcc	R to seq pNIT:GFP plus tags, sits ~100bp downstream of GFP
LMO70	aaatctagattaaattaaaccgatacaattaaag	F to amp fd terminator of pNIT, add XbaI site
LMO71	tttggtaccggagttcgcct	R to amp fd terminator of pNIT, add KpnI site
LMO72	caaccctggataaagagtcgtaaaagcttttcgaac	F to make point mut in pLO16 (to make tag KES/A129E mutant) "QuikChange SDM" PAGE purified
LMO73	gttcgaaaagcttttacgactctttatccagggttg	antiparallel of LMO72, needed for QuikChange, PAGE purified
LMO74	cgaaactcgtcgcggatatc	F qRT-PCR primer for expression of BPs gene32 (integrase)
LMO75	aatccgaccctcacgacatg	R qRT-PCR primer for expression of BPs gene32 (integrase)
LMO76	cgtccgcaaaatcgttctg	F qRT-PCR primer for expression of BPs gene33 (repressor)
LMO77	gccgtgttccagccaactaa	R qRT-PCR primer for expression of BPs gene33 (repressor)
LMO78	caccggcgaattcgactt	F qRT-PCR primer to look at expression levels of BPs gene7 (capsid)

LMO79	tcggcccactgaacctctt	R qRT-PCR primer to look at expression levels of BPs gene7 (capsid)
LMO80	gcaatctggcatggcatcta	F qRT-PCR primer to look at expression levels of BPs gene51 (ruvC)
LMO81	tgcccgtcgagttgttgac	R qRT-PCR primer to look at expression levels of BPs gene51 (ruvC)
LMO82	ttttttttaagcttctagaagccggcgaccagcg	R to amp BPs gp32 WT C-term tag (64aa) + HindIII site and poly-a for cleavage
LMO83	ggcgacggcgacggccgg	F to amp gp32 WT C-term tag (64aa, from ~lambda alignment end) from BPs
LMO84	cactctcggcatggacgagctgtacaagggcgacggcgacggccgg	F to extend gp32 WT tag by adding homology to GFP (lower, 28bp)
LMO85	tacgcatatggtgagcaagggcgag	F to amp GFP of pNIT:GFP (include NdeI site)
LMO86	cttgtacagctcgtccatgccgaga	R to amp GFP (without stop codon) of pNIT:GFP
LMO87	agatcggccggccgtcgccgtcgcccttgtacagctcgtccatgccgaga	R to extend GFP by adding homology to BPs gp32 WT tag (upper, 25bp)
LMO88	tgcgacataccggcgattgcatgtacgtctgtcgcgtacg	F (relative to genomic seq) for SDM to make pt mut in P-rep that gives phenotype of "clearer" mutant (T29336C)
LMO89	cgtacgcgacagacgtacatgcaatcgccggtatgtcgca	R (relative to genomic seq) for SDM to make pt mut in P-rep that gives phenotype of "clearer" mutant (antiparellel of LMO88)
LMO90	ctccgggcgcacttc	R primer for 5'RACE for P-rep promoter
LMO91	cagccggcggaacgt	R primer for 5'RACE for P-right promoter
LMO92	gcattttccaatagacgacacatgtcgctatgttggtgcac	F SDM to make GWB's BPs 102a -- T to C in putative P-R operator
LMO93	gtgcaccaacatagcgacatgtgtcgtctattggaaaatgc	R of LMO92
LMO94	gcattttccaatagacgacavatgtcgctatgttggtgcac	same as LMO92, F SDM to make GWB's BPs 102a (T to C in putative P-rep) with degenerative base (V=all but T)
LMO95	gtgcaccaacatagcgacatbtgtcgtctattggaaaatgc	R of LMO95 (with degen. B=all but A)
LMO96	ggcgcattttccaatagacgbcatatgtcgctatgttggtg	for SDM to make BPs 102e mutation NOTE--c is the actual mutation but primer is B (all but A, which is the WT seq)
LMO97	caccaacatagcgacatatgvcgtctattggaaaatgcgcc	R of LMO96 SDM to make BPS 102e mutant (V=all but T, which is WT)
LMO98	gaaagccacgttgtgtctcaaaatctctgatgttacattggaggatccagacatgataag	F recombineering primer to change pNIT:GFP (and derivatives) from KanR to ZeoR (UPPER = 40bp homology to pNIT, lower = to amp zeoR from pER10)
LMO99	gtaagttgagtgcatcaggcgtgatctgatccttcaactcagcaaaagttcgatttattc	R recombineering primer to change pNIT:GFP (and derivatives) from KanR to ZeoR (UPPER = 40bp homology to pNIT, lower = to amp zeoR from pER10)
LMO100	gtggcgcattttccaatagadgacatatgtcgctatgttg	F SDM for putative BPs rep operator, C29484AGorT C--is WT, primer is D at this position
LMO101	Caacatagcgacatatgtchtctattggaaaatgcgccac	R for LMO100 SDM for putative BPs rep operator, C29484AGorT G--is WT, primer is H at this position
LMO102	ggcgcattttccaatagachacatatgtcgctatgttggt	F SDM for putative BPs rep operator, G29485ACorT G--is WT, primer is H at this position
LMO103	accaacatagcgacatatgtdgtctattggaaaatgcgcc	R for LMO102 SDM for putative BPs rep operator, G29485ACorT

		C--is WT, primer is D at this position
LMO104	cgcattttccaatagacgadatatgtcgctatgttggtgc	F SDM for putative BPs rep operator, C29487AGorT C--is WT, primer is D at this position
LMO105	gcaccaacatagcgacatathtcgtctattggaaaatgcg	R for LMO104 SDM for putative BPs rep operator, C29487AGorT G--is WT, primer is H at this position
LMO106	cgcattttccaatagacgacbtatgtcgctatgttggtgc	F SDM for putative BPs rep operator, A29488AGorT A--is WT, primer is B at this position
LMO107	gcaccaacatagcgacatavgtcgtctattggaaaatgcg	R for LMO106 SDM for putative BPs rep operator, A29488AGorT T--is WT, primer is V at this position
LMO108	cattttccaatagacgacatbtgtcgctatgttggtgcac	F SDM for putative BPs rep operator, A29490AGorT A--is WT, primer is B at this position
LMO109	gtgcaccaacatagcgacavatgtcgtctattggaaaatg	R for LMO108 SDM for putative BPs rep operator, A29490AGorT T--is WT, primer is V at this position
LMO110	attttccaatagacgacatavgtcgctatgttggtgcaca	F SDM for putative BPs rep operator, T29491V T--is WT, primer is V at this position
LMO111	tgtgcaccaacatagcgacbtatgtcgtctattggaaaat	R for LMO110 SDM for putative BPs rep operator, T29491V A--is WT, primer is B at this position
LMO112	tttccaatagacgacatathtcgctatgttggtgcacatg	F SDM for putative BPs rep operator, G29491H G--is WT, primer is H at this position
LMO113	catgtgcaccaacatagcgadatatgtcgtctattggaaa	R for LMO112 SDM for putative BPs rep operator, G29491H C--is WT, primer is D at this position
LMO114	tttccaatagacgacatatgvcgctatgttggtgcacatg	F SDM for putative BPs rep operator, T29493V T--is WT, primer is V at this position
LMO115	catgtgcaccaacatagcgbcatatgtcgtctattggaaa	R for LMO114 SDM for putative BPs rep operator, T29493V A--is WT, primer is B at this position
LMO116	tccaatagacgacatatgtdgctatgttggtgcacatgac	F SDM for putative BPs rep operator, C29494D C--is WT, primer is D at this position
LMO117	gtcatgtgcaccaacatagchacatatgtcgtctattgga	R of LMO116 SDM for putative BPs rep operator, C29494D G--is WT, primer is H at this position
LMO118	ccaatagacgacatatgtchctatgttggtgcacatgacc	F SDM for putative BPs rep operator, G29495H G--is WT, primer is H at this position
LMO119	ggtcatgtgcaccaacatagdgacatatgtcgtctattgg	R for LMO118 SDM for putative BPs rep operator, G29495H C--is WT, primer is D at this position
LMO120	cgccgtcgacacgtaggattccgttccgctgg	F SDM for to make GFP+gp32tag of 54aa adds earlier (-10aa) stop to pLO24's GFP+gp32 tag (64aa)
LMO121	ccagcggaacggaatcctacgtgtcgacggcg	R for LMO120 SDM for to make GFP+gp32tag of 54aa adds earlier (-10aa) stop to pLO24's GFP+gp32 tag (64aa)
LMO122	ggccgacgtcgcacactagctggtcctcgacaac	F SDM for to make GFP+gp32tag of 44aa adds earlier (-20aa) stop to pLO24's GFP+gp32 tag (64aa)
LMO123	gttgtcgaggaccagctagtgtgcgacgtcggcc	R for LMO122 SDM for to make GFP+gp32tag of

		44aa adds earlier (-20aa) stop to pLO24's GFP+gp32 tag (64aa)
LMO124	gacgcgatcggcggttagctcgacaggacggcc	F SDM for to make GFP+gp32tag 34aa adds earlier (-30aa) stop to pLO24's GFP+gp32 tag (64aa)
LMO125	ggccgtcctgtcgagctaaccgccgatcgcgtc	R for LMO124 SDM for to make GFP+gp32tag 34aa adds earlier (-30aa) stop to pLO24's GFP+gp32 tag (64aa)
LMO126	caatagacgacatatgtcgdtatgttggtgcacatgaccc	F SDM for putative BPs PR -10 promoter, C29496D C--is WT, primer is D at this position
LMO127	gggtcatgtgcaccaacatahcgacatatgtcgtctattg	R SDM for putative BPs PR -10 promoter, C29496D G--is WT, primer is H at this position
LMO128	ccaatagacgacatatgtcgcvatgttggtgcacatgacc	F SDM for putative BPs PR -10 promoter, T29497V T--is WT, primer is V at this position
LMO129	ggtcatgtgcaccaacatbgcgacatatgtcgtctattgg	R SDM for putative BPs PR -10 promoter, T29497V A--is WT, primer is B at this position
LMO130	caatagacgacatatgtcgctbtgttggtgcacatgacc	F SDM for putative BPs PR -10 promoter, A29498B A--is WT, primer is B at this position
LMO131	ggtcatgtgcaccaacavagcgacatatgtcgtctattg	R SDM for putative BPs PR -10 promoter, A29498B T--is WT, primer is V at this position
LMO132	ccaatagacgacatatgtcgctavgttggtgcacatgacc	F SDM for putative BPs PR -10 promoter, T29499V T--is WT, primer is V at this position
LMO133	ggtcatgtgcaccaacbtagcgacatatgtcgtctattgg	R SDM for putative BPs PR -10 promoter, T29499V A--is WT, primer is B at this position
LMO134	gacgacatatgtcgctathttggtgcacatgaccccaac	F SDM for putative BPs PR -10 promoter, G29500H G--is WT, primer is H at this position
LMO135	gttggggtcatgtgcaccaadatagcgacatatgtcgtc	R SDM for putative BPs PR -10 promoter, G29500H C--is WT, primer is D at this position
LMO136	gacgacatatgtcgctatgvtggtgcacatgaccccaac	F SDM for putative BPs PR -10 promoter, T29501V T--is WT, primer is V at this position
LMO137	gttggggtcatgtgcaccabcatagcgacatatgtcgtc	R SDM for putative BPs PR -10 promoter, T29501V A--is WT, primer is B at this position
LMO138	acgacatatgtcgctatgtvggtgcacatgaccccaac	F SDM for putative BPs PR -10 promoter, T29502V T--is WT, primer is V at this position
LMO139	gttggggtcatgtgcaccbacatagcgacatatgtcgt	R SDM for putative BPs PR -10 promoter, T29502V A--is WT, primer is B at this position
LMO140	gaggatccagacatgataag	just pER10 homology of LMO98+99 to make substrate easier
LMO141	gtaagttgagtgcatcagg	just pER10 homology of LMO98+99 to make

		substrate easier
LMO142	caatagacgacatatgtcggtatgttggtgcacatgaccc	F SDM for putative BPs PR -10 promoter, C29496G
LMO143	gggtcatgtgcaccaacataccgacatatgtcgtctattg	R SDM for putative BPs PR -10 promoter, C29496G
LMO144	ccaatagacgacatatgtcgccatgttggtgcacatgacc	F SDM for putative BPs PR -10 promoter, T29497C
LMO145	ggtcatgtgcaccaacatggcgacatatgtcgtctattgg	R SDM for putative BPs PR -10 promoter, T29497C
LMO146	caatagacgacatatgtcgctgtgttggtgcacatgacc	F SDM for putative BPs PR -10 promoter, A29498G
LMO147	ggtcatgtgcaccaacacagcgacatatgtcgtctattg	R SDM for putative BPs PR -10 promoter, A29498G
LMO148	gacgacatatgtcgctatattggtgcacatgaccccaac	F SDM for putative BPs PR -10 promoter, G29500A
LMO149	gttggggtcatgtgcaccaatatagcgacatatgtcgtc	R SDM for putative BPs PR -10 promoter, G29500A
LMO150	gacgacatatgtcgctatgctggtgcacatgaccccaac	F SDM for putative BPs PR -10 promoter, T29501C
LMO151	gttggggtcatgtgcaccagcatagcgacatatgtcgtc	R SDM for putative BPs PR -10 promoter, T29501C
LMO152	acgacatatgtcgctatgtcggtgcacatgaccccaac	F SDM for putative BPs PR -10 promoter, T29502C
LMO153	gttggggtcatgtgcaccgacatagcgacatatgtcgt	R SDM for putative BPs PR -10 promoter, T29502C
LMO154	aaagcggccgccgttggcgatctt	F for cloning prom1 into pJ promless mCherry
LMO155	aaaggtacccccctggtccctcacaa	R for cloning prom1
LMO156	aaagcggccgcgactccgaatcgc	F for cloning prom2 into pJ promless mCherry
LMO157	aaaggtaccccgcggatcggccga	R for cloning prom2 into pJ promless mCherry
LMO158	aaagcggccgcgaagtgctggcaac	F for cloning prom3 into pJ promless mCherry
LMO159	aaaggtaccggtgtctggtgggtc	R for cloning prom3 in pJ promless mCherry
LMO160	aaagcggccgcgtatcggataccag	F for cloning prom4 into pJ promless mCherry, mobile element
LMO161	aaaggtaccggctgcctacctggag	R for cloning prom4 into pJ promless mCherry
LMO162	aaagcggccgcgtcttccaagccat	F for cloning prom5 into pJ promless mCherry
LMO163	aaaggtaccctcctggtcgtagaaac	R for cloning prom5 into pJ promless mCherry
LMO164	aaagcggccgccatgctcatgccag	F to clone putative promoter at P6, prom6
LMO165	aaaggtaccccggtgtcgtgctgc	R to clone putative promoter at P6 (without Softberry predicted promoter)
LMO166	aaaggtaccgttcagatcctcttcgg	R to clone putative promoter at P6 (WITH Softberry predicted promoter)
LMO167	aaagcggccgccagtcggtcgg	F to clone prom7 into pJ promless mCherry, R prom in P7
LMO168	aaaggtacccgacgagatcgacaact	R to clone prom7 into mCherry
LMO172	cgtttcatccccgatgccggtttaaacggcgggagaaaggagaaa	F SDM to add DraI site between hsp60 and mCherry in pJ hsp60 mCherry
LMO173	tttctccttctcccgccgtttaaaccggcatcggggatgaaacg	R SDM to add DraI site between hsp60 and mCherry in pJ hsp60 mCherry
LMO174	cgggcgtggcgcatvttccaatagacgac	F SDM to make T29473V, PR -35
GC-LMO174	gtcgtctattggaabatgcgccacgcccg	R SDM to make T29473V, PR -35
LMO175	cgggcgtggcgcattvtccaatagacgac	F SDM to make T29474V, PR -35
GC-	gtcgtctattggabaatgcgccacgcccg	R SDM to make T29474V, PR -35

LMO175		
LMO176	cgggcgtggcgcatttvccaatagacgac	F SDM to make T29475V, PR -35
GC-LMO176	gtcgtctattggbaaatgcgccacgcccg	R SDM to make T29475V, PR -35
LMO177	cgggcgtggcgcattttdcaatagacgacatatg	F SDM to make C29476D, PR -35
GC-LMO177	catatgtcgtctattghaaaatgcgccacgcccg	R SDM to make C29476D, PR -35
LMO178	cgggcgtggcgcattttcdaatagacgacatatg	F SDM to make C29477D, PR -35
GC-LMO178	catatgtcgtctatthgaaaatgcgccacgcccg	R SDM to make C29477D, PR -35
LMO179	cgggcgtggcgcattttccbatagacgacatatg	F SDM for A29478B, PR -35
GC-LMO179	catatgtcgtctatvggaaaatgcgccacgcccg	R SDM for A29478B, PR -35
LMO180	ttttttgcatgcggtggcatccgtggcg	F to amp prom+mCherry and clone into pML2357 (L5 int vector) with SphI site
LMO181	catcgataagcttcgaattctgcagctggatcc	R to amp prom+mCherry and clone into pML2357
LMO182	aaagcggccgccgacgagcaactgattctc	F to clone region around BPs genes54 and 55 into pJ promless mCherry vector, NotI site added
LMO183	aaaggtaccgtccgataccggaatattccc	R to clone region around BPs genes 54 and 55, through IR54-55 (up to gene55 start codon)
LMO184	aaaggtacctcggagcggagaacgatcg	R to clone region around BPs genes 54 and 55, whole region (to end of gene 55 + ~30bp)
LMO185	ggcgcattttccaatagacgccatatgtcgctatgttggtg	SDM to make A29486C (Greg's 102e vir mut)
LMO186	ggcgcattttccaatagacggcatatgtcgctatgttggtg	SDM to make A29486G
LMO187	cgggcgtggcgcatattccaatagacgac	F SDM to make T29473A
LMO188	gaaattgcaggtcgtagaagcgcg	F seq primer for pML2357 derivatives
LMO189	gatctctccggcttcaccgatc	R seq primer for pML2357 derivatives
LMO190	cgatgatcttgctctagctcggtatcagcgacgcgatc	F SDM to make GFP+gp32 tag 24aa (-40aa truncation) to find location of degradation tag on BPs gp32 (int)
LMO191	gcgcgccggcgcctagatcattgccgcgg	F SDM to make GFP+gp32 tag 14aa (-50aa truncation of WT 64aa tag)
LMO192	cactctcggcatggacgagctgtacaaggagggcgccccgg gtgcgccgttgtcggcagcctgaaagcttaaaaaa	pNIT GFP homology, 11aa Brujita gp33 (int) Cterm tag, HindIII site
LMO193	cgccgttgtcggcagcctgaaagcttaaaaaa	R to amp GFP+Brujita gp33 tag and HindIII site
LMO194	cactctcggcatggacgagctgtacaaggagcaaggccgac cgccgttgtccgcagtcagctagaagcttaaaaaa	pNIT GFP homology, 11aa Charlie gp34 (int) Cterm tag, HindIII site
LMO195	gccgttgtccgcagtcagctagaagcttaaaaaa	R to amp GFP+Charlie gp34 tag and HindIII site
LMO196	cactctcggcatggacgagctgtacaagacgcagacgcgac cagatctgtcggtcgcgagctgaaagcttaaaaaa	pNIT GFP homology, 11aa Redi gp37 (int) Cterm tag, HindIII site
LMO197	Cagatctgtcggtcgcgagctgaaagcttaaaaaa	R to amp GFP+Redi gp37 tag and HindIII site
LMO198	gaatgggtgtctgccgaccaca	new F to seq derivatives of pML2357
LMO199	gatctggactcccgacgattact	F qRT-PCR beginning of BPs gene 36
LMO200	tgtaatgcggatcttcgttga	R qRT-PCR beginning of BPs gene 36
LMO201	cggtcaaggcggattcg	F qRT-PCR middle of BPs gene 36
LMO202	tggaacggaatcaaccagaac	R qRT-PCR middle of BPs gene 36
LMO203	ctggacgcgcgccggcgccagcatc	F SDM to return cloned (pLO41 and pLO44) plasmid to the original sequence (silent mut--c->t)
LMO328	aaagcggccgcgtacggcaccgattggcc	R to make prom clone of BPs gene 26-27 IR, NotI
LMO329	aaaaggtaccgccgggtggtcaatcacg	R to make prom clone of BPs gene 26-27 IR, KpnI
LMO330	gtgcgccgttgtcggaggcctgaaagcttttc	F SDM to make penultimate A->E mut in Brujita int tag (pLO49) and create pLO52
LMO331	ccgccgttgtccgcagagagctagaagcttttcg	F SDM to make penultimate V->E mut in Charlie

		int tag (pLO50) and create pLO53
LMO332	gaccagatctgtcggtcgagagctgaaagcttttcg	F SDM to make penultimate A->E mut in Redi int tag (pLO51) and create pLO54
LMO333	gatcgaacgcggtcagattcccg	F redo RT primer (redo of well C9, LMO268, 34-35 junction) 389bp
LMO334	gcgatgatggtcgcggcgatg	R redo RT primer (redo of well C9, LMO269, 34-35 junction)
LMO335	gcctggtgatcggcgtcctg	F redo RT primer (redo of well D2, LMO278, 39-40 junction) 353bp
LMO336	cggtgtcggtcgtcgagaacg	R redo RT primer (redo of well D2, LMO279, 39-40 junction)
LMO337	ctaaccgaactcggcgccacca	F redo RT primer (redo of well E6, LMO310, 55-56 junction) 357bp
LMO338	ggtgctggctgtggcagcg	F redo RT primer (redo of well E6, LMO311, 55-56 junction)
LMO339	gcaatatcgccgtttgtcaaccgttgcg	SDM to make C29372T pt mut (clr plaque phenotype)
LMO340	cgggcgtggcgcatttcccaatagacgac	SDM to make T75C specifically
LMO341	gctgtacaaggagggcgccccgggtgcgccgttgtcggcagcctgaaagcttttcgaacc	SDM to make Brujita gp33 (int) tag to GFP in pNIT backbone --pLO49
LMO342	gctgtacaagccgttgtcggcagcctgaaagcttttcgaacc	SDM to make insertion of Brujita Int (gp33) 5aa tag--pLO64
LMO343	caagccgttgtcggaggcctgaaagcttttc	SDM to make penultimate A->E in pLO64--create pLO65
LMO344	gctgtacaaggattccaatcagcgcgactacgccctcgctgcctaaaagcttttcgaacc	SDM to make insertion of Msmeg ssrA 11aa tag--pLO66
LMO345	ctgtacaagtacgccctcgctgcctaaaagcttttcgaacc	SDM to make insertion of Msmeg ssrA 5aa tag--pLO67
LMO346	cgcgactacgccctcgaggcctaaaagcttttcg	SDM to make penultimate A->E mut in 11aa ssrA tag --pLO68
LMO347	gctgtacaagtacgccctcgaggcctaaaagcttttcgaacc	SDM to make penultimate A->E mut in 5aa ssrA tag --pLO69
LMO348	gttccgcaggctcgcgtag	F seq for pGH1000 plasmids (in HygR gene)
LMO349	cattcgccattcaggctgcgc	R seq for pGH1000 plasmids (near M13 ori)
LMO350	ttttttctagagcttgtggtggcatccgtgg	F to amp promoter + mCherry and add XbaI
LMO351	ttttttctagacgtcaggtggctagctgatcac	R to amp promoter + mCherry and add XbaI
LMO352	gatctggacgcgctaggccggcgccagc	SDM to add stop codon (ins) in pLO24 right before AGA (is 11aa internal tag)
LMO353	gggcgacggcgactagggccggccgatc	SDM to add stop codon (ins) in pLO24 -60aa (only 4aa tag left)
LMO354	gctattacgccagctggcgaaag	F seq pTTP1b derivatives
LMO355	gcgagtgtggatgcgcagcg	R seq pTTP1b derivatives
LMO356	gctgtacaagctcgctgcctaaaagcttttcgaacc	SDM to add C-terminal 3aa of ssrA tag to GFP, in pNIT:GFP
LMO357	gcgactacgctctcgaggcctaggacccag	SDM to create penultimate A->E mutation in pmv_GFPssra(zeo) plasmid (Rubin lab construct)
LMO358	gacgacatatgttgctatgctggtgcacatgaccccaac	SDM to make T01C in pLO76 (C94T) (TGCTATGCT) --to make pLO77
LMO359	gacgacatatgttgctatactggtgcacatgaccccaac	SDM to make G00A in pLO77 (TGCTATACT) --to make pLO78
LMO360	cgggcgtggcgcatttgccaatagacgac	F SDM to make T29475G, PR -35
LMO361	cgggcgtggcgcatttggcaatagacgacatatg	F SDM to make C29476G with T75G already, PR -35
LMO362	cgggcgtggcgcatctggcaatagacgac	F SDM to make T29473C with T75G, C76G

LMO363	cggcgcgtacaacgtgaacatc	F to ID TSS in mCherry plasmids
LMO364	gacaggatgtcccaggcgaac	R to ID TSS in mCherry plasmids
LMO365	gccacaacatggaagatggctcc	F seq pmv_GFPssra
LMO366	Gatcaccgcggccatgatgg	R seq pmv_GFPssra
LMO367	cacttccatatggtgagcaagggc	F cloning to move GFP (+tag) from pLO42-44 to pLAM12 (~785bp)
LMO368	cgtcgacatcgataagcttttacgac	R cloning to move GFP (+tag) from pLO42-44 to pLAM12
LMO369	gacgacatatgttggtatactggtgcacatgaccccaac	SDM for -10 PR mutants-- C94T, C96G, G00A, T01C
LMO370	cgggcgtggcgcatcttccaatagacgac	SDM PR -35 only T73C
LMO371	gccagcatcattgccgcggc	F 150bp from end of BPs gp32 (to ID int TSS in prophage)
LMO372	gacgattgctgccggaagcgc	R at very beginning of BPs gp32 (to ID int TSS in prophage)
LMO373	ggccgaccttcgccaggatg	R at very beginning of MSEMG_6348 (to ID int TSS in prophage)
LMO374	gaccttggctgtcgccatggc	R at very beginning of MSEMG_6346 (to ID int TSS in prophage)
LMO375	gcgacgacagagattgcaacggccgtcgaggac	SDM to substitute Tyr->Ala in BPs int gp32
LMO376	cgtagctcctaaatctcctgccggcg	SDM to delete BPs attP
LMO377	cggaggtacgcatatggtgag	F to amp GFP from pNIT:GFP and deriv, cut with NdeI
LMO378	aaccatggttcgaaatcgataagctt	R to amp GFP from pNIT:GFP and deriv, inserted and cut with ClaI
LMO379	cggaggtacgcatatggtgagcaag	F to amp GFP from pNIT vectors (for pNIT:GFP, pLO67/91)
LMO380	ggttcgaaatcgataagcttttacttgtacagctc	R to amp GFP from pNIT vectors, +ClaI cut site
LMO381	ccgatgccggaagtactcggcgggagaaag	SDM to add ScaI cut site to pJ hsp60
LMO382	cccaccagacaccagtactaaatctagccggac	SDM to add ScaI cut site to pLO32
LMO383	aaagcggccgcatgatggccggtc	check for additional promoters
LMO384	aaaggtaccggcgatcttgtcgcg	check for additional promoters
LMO385	aaagcggccgcaccgaggaagtctg	check for additional promoters
LMO386	aaaggtacctagagcgccagcgtc	check for additional promoters
LMO387	aaagcggccgctcggcaagatgctc	check for additional promoters
LMO388	aaaggtacccagagttgaagatcgcg	check for additional promoters
LMO389	aaagcggccgccacatgccctgg	check for additional promoters
LMO390	aaaggtacctgatgtgtgcgaccgtg	check for additional promoters
LMO391	aaagcggccgcaagaccatgtgtcgg	check for additional promoters
LMO392	aaaggtaccaggtggtacgccagaac	check for additional promoters
LMO393	aaagcggccgcaagcgacgggagta	check for additional promoters
LMO394	aaaggtaccatcatcggcgcgttcttg	check for additional promoters
LMO395	aaagcggccgcagttcgcgtttctcga	check for additional promoters
LMO396	aaaggtaccatcaccgaaccgacatac	check for additional promoters
LMO397	aaagcggccgcaagttcatccgaccgt	check for additional promoters
LMO398	aaaggtaccgattcgtgccaggcag	check for additional promoters
LMO399	aaagcggccgcaacccaccgtgaac	check for additional promoters
LMO400	aaaggtacctcagccggtcgagg	check for additional promoters
LMO401	aaagcggccgcttctgcacggttcac	check for additional promoters
LMO402	aaaggtacctgcgccttctccatgtc	check for additional promoters
LMO403	aaagcggccgcagcaatccgaaacggta	check for additional promoters
LMO404	aaaggtacctggtgggcaggtcttc	check for additional promoters
LMO405	aaagcggccgccttgaccaaaccagc	check for additional promoters
LMO406	aaaggtaccagccaggacggcag	check for additional promoters

LMO407	aaagcggccgccaatctcgacgcg	check for additional promoters
LMO408	aaaggtaccagcgcggcgtggat	check for additional promoters
LMO409	aaagcggccgctggccgcataatgca	check for additional promoters
LMO410	aaaggtaccatccacccgccgtagt	check for additional promoters
LMO411	aaagcggccgctcgttcggcgatc	check for additional promoters
LMO412	aaaggtaccaggaacgtcgcggtg	check for additional promoters
LMO413	aaagcggccgccatgtgccagcac	check for additional promoters
LMO414	gcgtcgatctggttctgaacggtaccttt	check for additional promoters
LMO414 redo	aaaggtaccgcgtcgatctggttctgaac	check for additional promoters
LMO415	aaagcggccgccgatacgaagaccg	check for additional promoters
LMO416	catcgccttaccggatgtcggtaccttt	check for additional promoters
LMO416 redo	aaaggtacccatcgccttaccggatgtc	check for additional promoters
LMO417	aaagcggccgcaagaccggcacac	check for additional promoters
LMO418	ctcttggaatgcctcgatgggtaccttt	check for additional promoters
LMO418 redo	aaaggtaccctcttggaatgcctcgatg	check for additional promoters
LMO419	aaagcggccgctatcggataccagc	check for additional promoters
LMO420	gcctggtacgaatccacggtaccttt	check for additional promoters
LMO420 redo	aaaggtaccgcctggtacgaatccac	check for additional promoters
LMO421	gattcctgctctcgcgtgcctggt	check for terminators
LMO422	gccaccagcgatccgagccattc	check for terminators
LMO423	gagcaagatcatcgccgcggcaatg	check for terminators
LMO424	cgaatcgttcgacggttacgtcctg	check for terminators
LMO425	gaggaagccgcacgcgaggaag	check for terminators
LMO426	cggctcgccgatgacgtggatag	check for terminators
LMO427	cagccaaagacaaacgtctgctgatcc	check for terminators
LMO428	ggtggctggtttggtcaaggtgattg	check for terminators
LMO429	gacaaacgtctgctgatccgtaactacc	check for terminators
LMO430	cgtgcatgacatcttgcacggctg	check for terminators
LMO431	gcgtcgcgcacagctacgagtc	check for terminators
LMO432	cttggcgccgatcacggcggt	check for terminators
LMO433	gcgccttcgccaacgccgtc	check for terminators
LMO434	gggtcgacgagtccgtggagc	check for terminators
LMO435	ggctgtggcgccttcggtgg	check for terminators
LMO436	ctagaggatctactagtcatatggatcga	F seq for pGH1000B derivs (esp pLO123)
LMO437	gatgcctggcagtcgatcgtac	R seq for pGH1000B derivs (esp pLO123)
LMO438	gcttgtggtggcatccgtggcgcggccgcggtaccagatcttt aaatctagcgaactcccttatgcgggt	F recombineering to place pNIT in pLO73 (Tweety int vector)
LMO439	ctagttaactacgtcgacatcgataagcttcgaattctgcagct ggatccgttcacggctctagcgccga	R recombineering to place pNIT in pLO73 (Tweety int vector)
LMO440	ccttttcacctcgtactgtcctgccaaacacaagcaacggagg tacgcatcggcgggagaaaggagaaag	F recombineering to replace EGFP with RBS+mCherry in pNIT
LMO441	taattgtatcggtttaatttaaatttaattaaccatggttcgaaaag cttccttacttgtacagctcgtc	R recombineering to replace EGFP with RBS+mCherry in pNIT
LMO442	ggtggcatccgtggcgcggccgcggtaccagatctttaaatct agccggataaagaagtgacgcggtctc	F recombineering to add acetamide promoter to pLO73 (Tweety int vector)
LMO443	gtcctcctcgcccttcgagaccatgattggctttctcctttctccc gccgcgttacatatgtggactccc	R recombineering to add acetamide promoter to pLO73 (Tweety int vector)
LMO444	gcttgtggtggcatccgtggcgcggccgcggtaccagatcttt aaatctacatggtcattaggagccgct	F recombineering to add insert TetON system to pLO73 (Tweety int vector)
LMO445	ctcgcccttcgagaccatgattggctttctcctttctcccgccgt	R recombineering to add insert TetON system to

	ccggcgcggatcgtgctcatttcgg	pLO73 (Tweety int vector)
LMO446	cgggcgtggcgcattttcaaatagacgacatatg	SDM to make single mutation C77A (aka C29477A, from LMO178)
LMO447	cgggcgtggcgcatttgcaaatagacgacatatg	SDM to make single mutation C77A and T75G
LMO448	gacgacatatgttggtatgctggtgcacatgaccccaac	SDM to put 3 mut in -10 (C94T, T01C and C96G)
LMO449	tggcgcattttccaatgacgacatatgtcgcta	SDM to delete one bp from between -35 and -10
LMO450	tggcgcattttccaatacgacatatgtcgcta	SDM to delete two bp from between -35 and -10
LMO451	tggcgcattttccaatagtacgacatatgtcgcta	SDM to ins one bp from between -35 and -10
LMO452	tggcgcatttgccaataacgacatatgttggta	SDM to delete one bp from between -35 and -10 from pLO93 (diff base than LMO449)
LMO453	cgggcgtggcgcattttacaatagacgacatatg	SDM to make C76A
LMO454	cgggcgtggcgcatttgacaatagacgacatatg	SDM to make T75G and C76A
LMO455	cgggcgtggcgcatgggacaatagacgacatatg	SDM to make "best" -35 GGGACA (T73G, T74G, T75G and C76A)
LMO456	gacgacatatgttagtatactggtgcacatgaccccaac	SDM to make "best" -10 (C94T, C96G, G00A, T01C, G95A)
LMO457	gacgacatatgtcgctatactggtgcacatgaccccaaca	SDM to make best -10 without extended cgctATACt
LMO458	agaatgccaggacccttgtcattccacgtcaattcatgcgccttttcacctcgtactgtc	F to extend homology another 40bp of LMO440+1 recombineering substrate (RBS + mCherry gene into pNIT)
LMO459	cgccttaagctccaaaaaaaaaggctccaaaaggagccttaattgtatcggtttaatttaaatt	R to extend homology another 40bp of LMO440+1 recombineering substrate (RBS + mCherry gene into pNIT)
LMO460	tggcgcattttccaatagaccagatgtgttgctatgttggtgcacatgac	SDM to make Val's 4 point mutations in O-R and find out if the mutant is fully derepressed (Val has found that it does NOT bind Rep at all?)
LMO461	aaaaaaattaatcggcgggagaaaggagaaag	F to amp mCherry with AseI site
LMO462	ttttaagcttttacttgtacagctcgtccatg	R to amp mCherry with HindIII site
LMO463	ttttgaattcttacttgtacagctcgtccatg	R to amp mCherry with EcoRI site
LMO464	tggcgcatttgccaatgacgacatatgttggta	SDM -1bp from spacer of pLO93 (AT-GA)
LMO465	cgtggcgcattttacaatgacgacatatgttggta	SDM -1bp from spacer of pLO160 (AT-GA)
LMO466	tggcgcatttgacaatgacgacatatgttggta	SDM -1bp from spacer of pLO161 (AT-GA)
LMO467	tccaatagacgacatatgttagtatactggtgcacatgacc	SDM "best" -10 --> pLO92 + G95A
LMO468	gcttgtggtggcatccgtggc	F to amp NIT prom from pLO166
LMO469	tctagcgccgatggtagtgtgg	R to amp NIT prom from pLO166 (NIT+mCherry)
LMO470	ctttaaatctagataaagaagtgacgcggtc	F to amp acet prom from pLO167
LMO471	tcgacatcgataagcttcgaattcttacttg	R to amp acet prom +mCherry from pLO167
LMO472	gacgatgctcacgatccgcac	F to amp term 5-6
LMO473	gaaccttgttctcggccttgagc	R to amp term 5-6
LMO474	gtcctcgacggccgtgtaaatc	F to amp term32
LMO475	tcctggtcgtgcacggcaag	R to amp term32
LMO476	cagctgcagaattcgaagcttatcgatg	F to amp rrnB terminator from pJ promless mCh (+ control for terminator search)
LMO477	cgtttcccgttgaatatggctcataacac	R to amp rrnB terminator from pJ promless mCh
LMO478	ggtttcgtagtctagatatgacgacagg	F to amp terminator pJEB from pNIT:GFP
LMO479	gctctagcgccgatggtagtgtg	R to amp terminator pJEB from pNIT:GFP
LMO480	aaagcggccgcgaacctactggctca	F prom inside 16, redo of 387+8
LMO481	gcatcctgcaaccacgtgaacggtaccttt	R prom inside 16, redo of 387+8
LMO482	ggttgtgatgcagcgtgatt	F for qRT MSMEG_0873
LMO483	catcgcgcacgacaacttc	R for qRT MSMEG_0873
LMO484	agcggcggtgtggagaa	F for qRT MSMEG_4892

LMO485	gcgccctcttacaaatgca	R for qRT MSMEG_4892
LMO486	ctcgcagctgccggataaatctagccggac	SDM to remove promoter from cloning junction in pLO132
LMO487	ggaaggccacaagtaaatctagccggac	SDM to remove promoter from cloning junction in pLO134
LMO488	gcgttggcgaaggtaaatctagccggac	SDM to copy mut made with LMO486 and 487 on pLO141
LMO489	tggcgcatttgacaatgacgacatatgttgcta	SDM to -1bp from spacer in T75G, C76A, no C96G
LMO490	tggcgcatttggcaatgacgacatatgttgcta	SDM to -1bp from spacer in T75G, C76G, no C96G
LMO491	tggcgcatctggcaatgacgacatatgttgcta	SDM to -1bp from spacer in T75G, C76G, T73C, no C96G
LMO492	aaagcggccgccgtgtatctgggcat	check for additional promoters
LMO493	aaaggtaccaacgcgccggcggtc	check for additional promoters
LMO494	aaagcggccgcccacaagtcattcgac	check for additional promoters
LMO495	aaaggtaccaactgcatctggatcgggtg	check for additional promoters
LMO496	aaagcggccgcaccgacaaggccat	check for additional promoters
LMO497	aaaggtaccttccagccggcgccac	check for additional promoters
LMO498	aaagcggccgccaggtccgcggtc	check for additional promoters
LMO499	aaaggtacctgtgacggaggaaacgacgt	check for additional promoters
LMO500	aaagcggccgctgaccgcgacgtggta	check for additional promoters
LMO501	aaaggtacccagcgcggcgatcttgc	check for additional promoters
LMO502	aaagcggccgcgctcgttcgagggaat	check for additional promoters
LMO503	aaaggtaccccggcttcgagccgc	check for additional promoters
LMO504	aaagcggccgcctgttcggccaagcaattt	check for additional promoters
LMO505	aaaggtaccgctgtactggacgacgtgat	check for additional promoters
LMO506	aaagcggccgccccaagatcgccgtag	check for additional promoters
LMO507	aaaggtaccgcggaacacgatcatgcagc	check for additional promoters
LMO508	aaagcggccgccgcggtgtcggtgtaa	check for additional promoters
LMO509	aaaggtaccacggcacgctcaccgc	check for additional promoters
LMO510	aaagcggccgccgctgcggtcgtc	check for additional promoters
LMO511	aaaggtaccacgcgcgctgctcgg	check for additional promoters
LMO512	aaagcggccgcttcgtcgccgacgt	check for additional promoters
LMO513	aaaggtacccgccggacactgcgcc	check for additional promoters
LMO514	aaagcggccgccgaaaaaagtccctggtg	check for additional promoters
LMO515	aaaggtacccgtttgtactgcgcgcagat	check for additional promoters
LMO516	aaagcggccgcgccgtaaccgtgagc	check for additional promoters
LMO517	aaaggtaccgcggtaggggagaacggc	check for additional promoters
LMO518	aaagcggccgcggcccgatcgtcgt	check for additional promoters
LMO519	aaaggtacctcgtcatctccctacgagtg	check for additional promoters
LMO520	aaagcggccgcgggtttcatccctccatcat	check for additional promoters
LMO521	aaaggtaccgacgactatgccgaccacag	check for additional promoters
LMO522	aaagcggccgctgccatggggtggc	check for additional promoters
LMO523	aaaggtacccgcgaagctgcactgggt	check for additional promoters

References

1. World Health Organization Report 2009 G (2009) Global tuberculosis control 2009: Epidemiology, strategy, financing. Geneva: World Health Organization.
2. Hatfull GF (2014) Molecular Genetics of Mycobacteriophages. Microbiology Spectrum 2: 1-36.
3. Ford ME, Stenstrom C, Hendrix RW, Hatfull GF (1998) Mycobacteriophage TM4: genome structure and gene expression. Tuber Lung Dis 79: 63-73.
4. Hatfull GF, Sarkis GJ (1993) DNA sequence, structure and gene expression of mycobacteriophage L5: a phage system for mycobacterial genetics. Mol Microbiol 7: 395-405.
5. Mediavilla J, Jain S, Kriakov J, Ford ME, Duda RL, et al. (2000) Genome organization and characterization of mycobacteriophage Bxb1. Mol Microbiol 38: 955-970.
6. Pope WH, Anders KR, Baird M, Bowman CA, Boyle MM, et al. (2014) Cluster M Mycobacteriophages Bongo, PegLeg, and Rey with Unusually Large Repertoires of tRNA Isotypes. J Virol 88: 2461-2480.
7. Pope WH, Ferreira CM, Jacobs-Sera D, Benjamin RC, Davis AJ, et al. (2011) Cluster K Mycobacteriophages: Insights into the Evolutionary Origins of Mycobacteriophage TM4. PLoS ONE 6: e26750.
8. Dedrick RM, Marinelli LJ, Newton GL, Pogliano K, Pogliano J, et al. (2013) Functional requirements for bacteriophage growth: gene essentiality and expression in mycobacteriophage Giles. Mol Microbiol 88: 577-589.
9. Mageeney C, Pope WH, Harrison M, Moran D, Cross T, et al. (2012) Mycobacteriophage Marvin: a new singleton phage with an unusual genome organization. J Virol 86: 4762-4775.
10. Sampson T, Broussard GW, Marinelli LJ, Jacobs-Sera D, Ray M, et al. (2009) Mycobacteriophages BPs, Angel and Halo: comparative genomics reveals a novel class of ultra-small mobile genetic elements. Microbiology 155: 2962-2977.
11. Hatfull GF (2012) The secret lives of mycobacteriophages. Adv Virus Res 82: 179-288.
12. Kumaresan J, Heitkamp P, Smith I, Billo N (2004) Global Partnership to Stop TB: a model of an effective public health partnership. International Journal of Tuberculosis and Lung Disease 8: 120-129.
13. Dye C, Scheele S, Dolin P, Pathania V, Raviglione MC (1999) Consensus statement. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project. Jama 282: 677-686.
14. Gandhi NR, Moll A, Sturm AW, Pawinski R, Govender T, et al. (2006) Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. Lancet 368: 1575-1580.
15. Piuri M, Jacobs WR, Jr., Hatfull GF (2009) Fluoromycobacteriophages for rapid, specific, and sensitive antibiotic susceptibility testing of Mycobacterium tuberculosis. PLoS ONE 4: e4870.

16. Sarkis GJ, Jacobs WR, Jr., Hatfull GF (1995) L5 luciferase reporter mycobacteriophages: a sensitive tool for the detection and assay of live mycobacteria. Mol Microbiol 15: 1055-1067.
17. Jacobs WR, Jr., Barletta RG, Udani R, Chan J, Kalkut G, et al. (1993) Rapid assessment of drug susceptibilities of Mycobacterium tuberculosis by means of luciferase reporter phages. Science 260: 819-822.
18. Rondon L, Piuri M, Jacobs WR, Jr., de Waard J, Hatfull GF, et al. (2011) Evaluation of fluoromycobacteriophages for detecting drug resistance in Mycobacterium tuberculosis. J Clin Microbiol 49: 1838-1842.
19. Kalantri S, Pai M, Pascopella L, Riley L, Reingold A (2005) Bacteriophage- based tests for the detection of Mycobacterium tuberculosis in clinical specimens: a systematic review and meta- analysis. BMC Infect Dis 5: 59.
20. Bardarov S, Bardarov Jr S, Jr., Pavelka Jr MS, Jr., Sambandamurthy V, Larsen M, et al. (2002) Specialized transduction: an efficient method for generating marked and unmarked targeted gene disruptions in Mycobacterium tuberculosis, M. bovis BCG and M. smegmatis. Microbiology 148: 3007-3017.
21. van Kessel JC, Marinelli LJ, Hatfull GF (2008) Recombineering mycobacteria and their phages. Nat Rev Microbiol 6: 851-857.
22. van Kessel JC, Hatfull GF (2007) Recombineering in *Mycobacterium tuberculosis*. Nature Methods 4: 147-152.
23. van Kessel JC, Hatfull GF (2008) Mycobacterial recombineering. Methods Mol Biol 435: 203-215.
24. Medjahed H, Reyrat JM (2009) Construction of *Mycobacterium abscessus* defined glycopeptidolipid mutants: Comparison of genetic tools. Applied and Environmental Microbiology 75: 1331-1338
25. Marinelli LJ, Piuri M, Swigonova Z, Balachandran A, Oldfield LM, et al. (2008) BRED: a simple and powerful tool for constructing mutant and recombinant bacteriophage genomes. PLoS ONE 3: e3957.
26. Hendrix RW (2002) Bacteriophages: evolution of the majority. Theor Popul Biol 61: 471-480.
27. Hatfull GF (2008) Bacteriophage genomics. Curr Opin Microbiol 11: 447-453.
28. Hatfull GF, Cresawn SG, Hendrix RW (2008) Comparative genomics of the mycobacteriophages: insights into bacteriophage evolution. Res Microbiol 159: 332-339.
29. Hatfull GF, Pedulla ML, Jacobs-Sera D, Cichon PM, Foley A, et al. (2006) Exploring the mycobacteriophage metaproteome: phage genomics as an educational platform. PLoS Genet 2: e92.
30. Payne K, Sun Q, Sacchettini J, Hatfull GF (2009) Mycobacteriophage Lysin B is a novel mycolylarabinogalactan esterase. Mol Microbiol 73: 367-381.
31. Broussard GW, Oldfield LM, Villanueva VM, Lunt BL, Shine EE, et al. (2013) Integration-dependent bacteriophage immunity provides insights into the evolution of genetic switches. Mol Cell 49: 237-248.
32. Cresawn SG, Bogel M, Day N, Jacobs-Sera D, Hendrix RW, et al. (2011) Phamerator: a bioinformatic tool for comparative bacteriophage genomics. BMC Bioinformatics 12: 395.
33. Brussow H, Desiere F (2001) Comparative phage genomics and the evolution of Siphoviridae: insights from dairy phages. Mol Microbiol 39: 213-222.

34. Donnelly-Wu MK, Jacobs WR, Jr., Hatfull GF (1993) Superinfection immunity of mycobacteriophage L5: applications for genetic transformation of mycobacteria. Mol Microbiol 7: 407-417.
35. Nesbit CE, Levin ME, Donnelly-Wu MK, Hatfull GF (1995) Transcriptional regulation of repressor synthesis in mycobacteriophage L5. Mol Microbiol 17: 1045-1056.
36. Brown KL, Sarkis GJ, Wadsworth C, Hatfull GF (1997) Transcriptional silencing by the mycobacteriophage L5 repressor. Embo J 16: 5914-5921.
37. Rybniker J, Plum G, Robinson N, Small PL, Hartmann P (2008) Identification of three cytotoxic early proteins of mycobacteriophage L5 leading to growth inhibition in Mycobacterium smegmatis. Microbiology 154: 2304-2314.
38. Das Gupta SK, Bashyam MD, Tyagi AK (1993) Cloning and assessment of mycobacterial promoters by using a plasmid shuttle vector. J Bacteriol 175: 5186-5192.
39. Ramesh G, Gopinathan KP (1995) Cloning and characterization of mycobacteriophage I3 promoters. Indian J Biochem Biophys 32: 361-367.
40. Chattopadhyay C, Sau S, Mandal NC (2003) Cloning and characterization of the promoters of temperate mycobacteriophage L1. J Biochem Mol Biol 36: 586-592.
41. Jain S, Hatfull GF (2000) Transcriptional regulation and immunity in mycobacteriophage Bxb1. Mol Microbiol 38: 971-985.
42. Garcia M, Pimentel M, Moniz-Pereira J (2002) Expression of Mycobacteriophage Ms6 lysis genes is driven by two sigma(70)-like promoters and is dependent on a transcription termination signal present in the leader RNA. J Bacteriol 184: 3034-3043.
43. Hertwig S, Skurnik M, Appel B (2006) Yersinia phages. In: Calendar R, editor. The Bacteriophages. Oxford: Oxford University Pess. pp. 545-557.
44. Mosig G, Eiserling FA (2006) T4 and Related Phages: Structure and Development. In: Calendar R, editor. The Bacteriophages. Oxford: Oxford University Press. pp. 225-267.
45. Brondsted L, Hammer K (2006) Phages of Lactococcus lactis. In: Calendar R, editor. The Bacteriophages. Oxford: Oxford University Press.
46. Kutter E, Guttman B, Carlson K (1994) The transition from host to phage metabolism after T4 infection; Karam JD, editor. 343-346 p.
47. Snyder L, Gold L, Kutter E (1976) A gene of bacteriophage T4 whose product prevents true late transcription on cytosine-containing T4 DNA. Proceedings of the National Academy of Sciences of the United States of America 73: 3098-3102.
48. Drivdahl RH, Kutter EM (1990) Inhibition of transcription of cytosine-containing DNA in vitro by the alc gene product of bacteriophage T4. Journal of Bacteriology 172: 2716-2727.
49. Koch T, Raudonikiene A, Wilkens K, Rueger W (1995) Overexpression, purification, and characterization of the ADP-ribosyltransferase (gpAlt) of bacteriophage T4: ADP-ribosylation of E. coli RNA polymerase modulates T4 "early" transcription. Gene Expression 4: 253-264.
50. Wilkens K, Tiemann B, Bazan F, Ruger W (1997) ADP-ribosylation and early transcription regulation by bacteriophage T4. In: Haag F, KochNolte F, editors. Adp-Ribosylation in Animal Tissues: Structure, Function, and Biology of Mono. pp. 71-82.
51. Oppenheim AB, Kobiler O, Stavans J, Court DL, Adhya S (2005) Switches in bacteriophage lambda development. Annu Rev Genet 39: 409-429.

52. Greenblatt J, Mah TF, Legault P, Mogridge J, Li J, et al. (1998) Structure and mechanism in transcriptional antitermination by the bacteriophage lambda N protein. Cold Spring Harbor Symposia on Quantitative Biology 63: 327-336.
53. Ptashne M (2004) A genetic switch: phage lambda revisited, 3rd ed. Cold Spring Harbor, NY.: Cold Spring Harbor Laboratory Press.
54. Marr MT, Datwyler SA, Meares CF, Roberts JW (2001) Restructuring of an RNA polymerase holoenzyme elongation complex by lambdoid phage Q proteins. Proceedings of the National Academy of Sciences of the United States of America 98: 8972-8978.
55. Johnson A, Meyer BJ, Ptashne M (1978) Mechanism of action of the cro protein of bacteriophage lambda. Proc Natl Acad Sci U S A 75: 1783-1787.
56. Kobiler O, Rokney A, Friedman N, Court DL, Stavans J, et al. (2005) Quantitative kinetic analysis of the bacteriophage lambda genetic network. Proc Natl Acad Sci U S A 102: 4470-4475.
57. Ptashne M, Jeffrey A, Johnson AD, Maurer R, Meyer BJ, et al. (1980) How the lambda repressor and cro work. Cell 19: 1-11.
58. China A, Tare P, Nagaraja V (2010) Comparison of promoter-specific events during transcription initiation in mycobacteria. Microbiology-Sgm 156: 1942-1952.
59. Bashyam MD, Kaushal D, Dasgupta SK, Tyagi AK (1996) A study of mycobacterial transcriptional apparatus: identification of novel features in promoter elements. J Bacteriol 178: 4847-4853.
60. Hawley DK, McClure WR (1983) Compilation and analysis of Escherichia coli promoter DNA sequences. Nucleic Acids Research 11: 2237-2255.
61. Kenney TJ, Churchward G (1996) Genetic analysis of the Mycobacterium smegmatis rpsL promoter. Journal of Bacteriology 178: 3564-3571.
62. Kremer L, Baulard A, Estaquier J, Content J, Capron A, et al. (1995) Analysis of the Mycobacterium tuberculosis 85A antigen promoter region. Journal of Bacteriology 177: 642-653.
63. Bashyam MD, Tyagi AK (1998) Identification and analysis of "extended -10" promoters from mycobacteria. J Bacteriol 180: 2568-2573.
64. Agarwal N, Tyagi AK (2006) Mycobacterial transcriptional signals: requirements for recognition by RNA polymerase and optimal transcriptional activity. Nucleic Acids Res 34: 4245-4257.
65. Belyaeva T, Griffiths L, Minchin S, Cole J, Busby S (1993) The Escherichia coli cysG promoter belongs to the'extended-10'class of bacterial promoters. Biochemical Journal 296: 851-857.
66. Voskuil MI, Voepel K, Chambliss GH (1995) The -16 region, a vital sequence for the utilization of a promoter in Bacillus subtilis and Escherichia coli. Molecular Microbiology 17: 271-279.
67. Sabelnikov AG, Greenberg B, Lacks SA (1995) An" Extended-10" promoter alone directs the transcription of the DpnII operon of Streptococcus pneumoniae. Journal of Molecular Biology 250: 144-155.
68. Barne KA, Bown JA, Busby SJW, Minchin SD (1997) Region 2.5 of the Escherichia coli RNA polymerase sigma(70) subunit is responsible for the recognition of the 'extended -10' motif at promoters. Embo Journal 16: 4034-4040.

69. Ponnambalam S, Webster C, Bingham A, Busby S (1986) Transcription initiation at the Escherichia coli galactose operon promoters in the absence of the normal -35 region sequences. Journal of Biological Chemistry 261: 6043-6048.
70. Kumar A, Malloch RA, Fujita N, Smillie DA, Ishihama A, et al. (1993) The minus 35 recognition region of Escherichia coli sigma 70 is inessential for initiation of transcription at the "extended minus 10" promoter. Journal of Molecular Biology 232: 406-418.
71. Agarwal N, Tyagi AK (2003) Role of 5'-TGN-3' motif in the interaction of mycobacterial RNA polymerase with a promoter of 'extended -10' class. FEMS Microbiol Lett 225: 75-83.
72. Gomez M, Smith I (2000) Determinants of mycobacterial gene expression. In: Hatfull GF, Jacobs Jr. WR, editors. Molecular genetics of the mycobacteria. Washington, DC: ASM Press. pp. In press.
73. Unniraman S, Chatterji M, Nagaraja V (2002) DNA gyrase genes in Mycobacterium tuberculosis: a single operon driven by multiple promoters. J Bacteriol 184: 5449-5456.
74. Smith I, Bishai WR, Nagaraja V (2005) Control of mycobacterial transcription. In: Cole ST, Eisenach K, McMurray DN, Jr. JWR, editors. Tuberculosis and the Tubercle Bacillus. Washington D.C.: ASM Press. pp. 219-231.
75. Newton-Foot M, Gey van Pittius NC (2013) The complex architecture of mycobacterial promoters. Tuberculosis (Edinb) 93: 60-74.
76. Tare P, Nagaraja V (2013) Regulation of transcription initiation in mycobacteria. Current Science 105: 632-642.
77. Bannantine JP, Barletta RG, Thoen CO, Andrews RE (1997) Identification of Mycobacterium paratuberculosis gene expression signals. Microbiology-Uk 143: 921-928.
78. Waagmeester A, Thompson J, Reyrat JM (2005) Identifying sigma factors in *Mycobacterium smegmatis* by comparative genomic analysis. Trends in Microbiology 13: 505-509.
79. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, et al. (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 393: 537-544.
80. Tekaia F, Gordon SV, Garnier T, Brosch R, Barrell BG, et al. (1999) Analysis of the proteomone of *Mycobacterium tuberculosis* in silico. Tuber Lung Dis 79: 329-342.
81. Gomez M, Doukhan L, Nair G, Smith I (1998) sigA is an essential gene in Mycobacterium smegmatis. Mol Microbiol 29: 617-628.
82. Lonetto M, Gribskov M, Gross CA (1992) The sigma 70 family: sequence conservation and evolutionary relationships. J Bacteriol 174: 3843-3849.
83. Gruber TM, Gross CA (2003) Multiple sigma subunits and the partioning of bacterial transcription space. Annual Review Microbiology 57: 441-466.
84. Manganelli R (2014) Sigma Factors: Key Molecules in Mycobacterium tuberculosis Physiology and Virulence. Microbiology Spectrum 2: 1-23.
85. Volpe E, Cappelli G, Grassi M, Martino A, Serafino A, et al. (2006) Gene expression profiling of human macrophages at late time of infection with *Mycobacterium tuberculosis*. Mol Microbiol 188: 449-460.
86. Manganelli R, Dubnau E, Tyagi S, Kramer FR, Smith I (1999) Differential expression of 10 sigma factor genes in Mycobacterium tuberculosis. Mol Microbiol 31: 715-724.

87. Dubnau E, Fontan P, Manganelli R, Soares-Appel S, Smith I (2002) Mycobacterium tuberculosis genes induced during infection of human macrophages. Infect Immun 70: 2787-2795.
88. Wu S, Howard ST, Lakey DL, Kipnis A, Samten B, et al. (2004) The prinicple sigma factor sigA mediates enhanced growth of *Mycobacterium tuberculosis* in vivo. Mol Microbiol 51: 1551-1562.
89. Vallecillo AJ, Espitia C (2009) Expression of *Mycobacterium tuberculosis* pe_pgrs33 is repressed during stationary phase and stress conditions, and its transcription us mediated by sigma factor A. Microb Pathog 46: 119-127.
90. Rengarajan J, Bloom BR, Rubin EJ (2005) Genome-wide requirements for *Mycobacterium tuberculosis* adaption and survival in macrophages. Proc Natl Acad Sci U S A 102: 8327-8332.
91. Vergne I, Chua J, Singh SB, Deretic V (2004) Cell biology of *Mycobacterium tuberculosis* phagosome. Annu Rev Cell Biol 20: 367-394.
92. Smith I (2003) Mycobacterium tuberculosis pathogenesis and molecular determinants of virulence. Clin Microbiol Rev 16: 463-496.
93. Rodrigue S, Provvedi R, Jacques PE, Gaudreau L, Manganelli R (2006) The sigma factors of Mycobacterium tuberculosis. FEMS Microbiol Rev 30: 926-941.
94. Sachdeva P, Misra R, Tyagi AK, Singh Y (2010) The sigma factors of Mycobacterium tuberculosis: regulation of the regulators. FEBS J 277: 605-626.
95. Doukhan L, Predich M, Nair G, Dussurget O, Mandic-Mulec I, et al. (1995) Genomic organization of the mycobacterial sigma gene cluster. Gene 165: 67-70.
96. Fontan PA, Voskuil MI, Gomez M, Tan D, Pardini M, et al. (2009) The *Mycobacterium tuberculosis* sigma factor sigmaB is required for full response to cell envelope stress and hypoxia in vitro, but it is dispensable for in vivo growth. J Bacteriol 191: 5628-5633.
97. Lee J, Karakousis PC, Bishai WR (2008) Roles of SigB and SigF in the *Mycobacterium tuberculosis* sigma factor netweork. J Bacteriol 190: 699-707.
98. Hu Y, Coates ARM (1999) Transcription of two sigma 70 homologue genes, sigA and sigB, in stationary-phase Mycobacterium tuberculosis. J Bacteriol 181: 469-476.
99. Agarwal N, Woolwine SC, Tyagi S, Bishai WR (2007) Characterization of the *Mycobacterium tuberculosis* sigma factor SigM by assessment of virulence and identification of SigM-dependent genes. Infect Immun 75: 452-461.
100. Manganelli R, Voskuil MI, Schoolnik GK, Dubnau E, Gomez M, et al. (2002) Role of the extracytoplasmic-function sigma factor sigmaH in *Mycobacterium tuberculosis* global gene expression. Mol Microbiol 45: 365-374.
101. Raman S, Song T, Puyang X, Bardarov S, Jacobs Jr WR, Jr., et al. (2001) The alternative sigma factor SigH regulates major components of oxidative and heat stress responses in *Mycobacterium tuberculosis*. J Bacteriol 183: 6119-6125.
102. Abdul-Majid KB, Ly LH, Converse PJ, Geiman DE, Mcmurray DN, et al. (2008) Altered cellular infiltration and cytokine levels during early *Mycobacterium tuberculosis* sigC mutant infection are associated with late-stage disease attentuation and milder immunopathology in mice. BMC Microbiol 8.
103. Ando M, Yoshimatsu T, Ko C, Converse PJ, Bishai WR (2003) Deletion of *Mycobacterium tuberculosis* sigma factor E results in delayed time of death with bacterial persistance in the lungs of aerosol-infected mico. Infect Immun 71: 7170-7172.

104. Geiman DE, Kaushal D, Ko C, Tyagi S, Manabe YC, et al. (2004) Attenuation of late-stage disease in mice infected by the *Mycobacterium tuberculosis* mutant lacking the SigF alternative sigma factor and identification of SigF-dependent genes by microarray analysis. Infect Immun 72: 1733-1745.
105. Kaushal D, Schroeder BG, Tyagi S, Yoshimatsu T, Scott C, et al. (2002) Reduced immunopathology and mortality despite tissue persistance in a *Mycobacterium tuberculosis* mutant lacking alternative sigma factor, SigH. Proc Natl Acad Sci U S A 99: 8330-8335.
106. Mehra S, Golden NA, Stuckey K, Didier PJ, Doyle LA, et al. (2012) The *Mycobacterium tuberculosis* stress response factor SigH is required for bacterial burden as well as immunopathology in primate lungs. J Infect Dis 205: 1203-1213.
107. Dutta R, Inouye M (2000) GHKL, an emergent ATPase/kinase superfamily. Trends Biochem Sci 25: 24-28.
108. Kang JG, Paget MS, Seok YJ, Hahn MY, Bae JB, et al. (1999) RsrA, an anti-sigma factor regulated by redox change. Embo J 18: 4292-4298.
109. Paget MS, Bae JB, Hahn MY, Li W, Kleanthous C, et al. (2001) Mutational anaylsis of RsrA, a zinc-binding anti-sigma factor with a thiol-disulphide redox switch. Mol Microbiol 39: 1036-1047.
110. Hughes KT, Mathee K (1998) The anti-sigma factors. Annu Rev Microbiol 52: 231-286.
111. Mulder MA, Zappe H, Steyn LM (1997) Mycobacterial promoters. Tuber Lung Dis 78: 211-223.
112. Raman S, Harza R, Dascher CC, Husson RN (2004) Transcription regulation by the *Mycobacterium tuberculosis* alternative sigma factor SigD and its role in virulence. J Bacteriol 186: 6605-6616.
113. Hahn MY, Raman S, Anaya M, Husson RN (2005) The *Mycobacterium tuberculosis* extracytoplasmic-function sigma factor SigL regulated polyketide synthases and secreted or membrane proteins and is required for virulence. J Bacteriol 187: 7062-7071.
114. Raman S, Puyang X, Cheng T, Young DC, Moody DB, et al. (2006) *Mycobacterium tuberculosis* SigM positively regulates Esx secreted protein and nonribosomal peptide synthetase genes and down regulates virulence-associated surface lipid synthesis. J Bacteriol 188: 8460-8468.
115. Song T, Song SE, Raman S, Anaya M, Husson RN (2008) Critical role of a single position in the -35 element for promoter recognition by *Mycobacterium tuberculosis* SigE and SigH. J Bacteriol 190: 2227-2230.
116. Peters JM, Vangeloff AD, Landick R (2011) Bacterial Transcription Terminators: The RNA 3 '-End Chronicles. Journal of Molecular Biology 412: 793-813.
117. von Hippel PH (1998) Transcription - An integrated model of the transcription complex in elongation, termination, and editing. Science 281: 660-665.
118. Gusarov I, Nudler E (1999) The mechanism of intrinsic transcription termination. Molecular Cell 3: 495-504.
119. McDowell JC, Roberts JW, Jin DJ, Gross C (1994) Determination of intrinsic transcription termination efficiency by RNA polymerase elongation rate. Science 266: 822-825.
120. Ermolaeva MD, Khalak HG, White O, Smith HO, Salzberg SL (2000) Prediction of transcription terminators in bacterial genomes. Journal of Molecular Biology 301: 27-33.

121. Washio T, Sasayama J, Tomita M (1998) Analysis of complete genomes suggests that many prokaryotes do not rely on hairpin formation in transcription termination. Nucleic Acids Research 26: 5456-5463.
122. Mitra A, Angamuthu K, Nagaraja V (2008) Genome-wide analysis of the intrinsic terminators of transcription across the genus Mycobacterium. Tuberculosis 88: 566-575.
123. Mitra A, Angamuthu K, Jayashree HV, Nagaraja V (2009) Occurrence, divergence and evolution of intrinsic terminators across Eubacteria. Genomics 94: 110-116.
124. Unniraman S, Prakash R, Nagaraja V (2001) Alternate paradigm for intrinsic transcription termination in eubacteria. Journal of Biological Chemistry 276: 41850-41855.
125. Wilson KS, Vonhippel PH (1995) Transcription termination at intrinsic terminators: the role of the RNA hairpin. Proceedings of the National Academy of Sciences of the United States of America 92: 8793-8797.
126. Abe H, Aiba H (1996) Differential contributions of two elements of rho-independent terminator to transcription termination and mRNA stabilization. Biochimie 78: 1035-1042.
127. Reynolds R, Bermudezcruz RM, Chamberlin MJ (1992) Parameters affecting transcription termination by< i> Escherichia coli</i> RNA polymerase: I. Analysis of 13 rho-independent terminators. Journal of Molecular Biology 224: 31-51.
128. Harshey RM, Ramakrishnan T (1977) Rate of ribonucleic acid chain growth in Mycobacterium tuberculosis H37Rv. Journal of Bacteriology 129: 616-622.
129. Yager TD, von Hippel PH (1987) Transcript elongation and termination in *Escherichia coli*. In: Neidhardt FC, editor. *Escherichia coli* and *Salmonella typhimurium*: Cellular and molecular biology. Washington DC: American Society for Microbiology. pp. 1241-1275.
130. Spencer CA, Groudine M (1990) Transcription elongation and eukaryotic gene regulation. Oncogene 5: 777-785.
131. Roberts JW (1988) Phage lambda and the regulation of transcription termination. Cell 52: 5-6.
132. Friedman DI, Schauer AT, Baumann MR, Baron LS, Adhya S (1981) Evidence that ribosomal protein S10 participates in control of transcription termination. Proc Natl Acad Sci U S A 78: 1115-1118.
133. Sullivan SL, Ward DF, Gottesman ME (1992) Effect of *Escherichia coli* nusG function on lambda N-mediated transcription antitermination. J Bacteriol 174: 1339-1344.
134. Mason SW, Li J, Greenblatt J (1992) Host factor requirements for processive antitermination of transcription and suppression of pausing by the N protein of bacteriophage lambda. J Biol Chem 267: 19418-19426.
135. Mason SW, Greenblatt J (1991) Assembly of transcription elongation complexes containing the N protein of phage lambda and the *Escherichia coli* elongation factors NusA, NusB, NusG, and S10. Genes Dev 5: 1504-1512.
136. Nodwell JR, Greenblatt J (1991) The nut site of bacteriophage lambda is made of RNA and is bound by transcription antitermination factors on the surface of RNA polymerase. Genes Dev 5: 2141-2151.
137. Horwitz RJ, Li J, Greenblatt J (1987) An elongation control particle containing the N gene transcriptional antitermination protein of bacteriophage lambda. Cell 51: 631-641.
138. Whalen WA, Das A (1990) Action of an RNA site at a distance: role of the nut genetic signal in transcription antitermination by phage-lambda N gene product. New Biol 2: 975-991.

139. Olsen ER, Tomich CS, Friedman DI (1984) The nusA recognition site. Alteration in its sequence or position relative to upstream translation interferes with the action of the N antitermination function of phage lambda. J Mol Biol 180: 1053-1063.
140. Warren F, Das A (1984) Formation of termination-resistant transcription complex at phage lambda nut locus: effects of altered translation and a ribosomal mutation. Proc Natl Acad Sci U S A 81: 3612-3616.
141. Das A (1992) How the phage lambda N gene product supresses transcription termination: communication of RNA polymerase with regulatory proteins mediated by signals in nascent RNA. J Bacteriol 174: 6711-6716.
142. Morgan EA (1986) Antitermination mechanisms in rRNA operons of *Escherichia coli*. J Bacteriol 168: 1-5.
143. Greenblatt J, Nodwell JR, Mason SW (1993) Transcriptional antitermination. Nature 364: 401-406.
144. Greenblatt J (1984) Regulation of transcription termination in *Escherichia coli*. Can J biochem cell Biol 62: 79-88.
145. Kainz M, Roberts J (1992) Structure of transcription elongation complexes in vivo. Science 255: 838-841.
146. Gopal B, Haire LF, Gamblin SJ, Dodson EJ, Lane AN, et al. (2001) Crystal structure of the transcription elongation/anti-termination factor NusA from *Mycobacterium tuberculosis* at 1.7 A resolution. J Mol Biol 314: 1087-1095.
147. Richardson JP, Greenblatt J (1996) Control of RNA chain elongation and termination. In: Neidhardt FC, editor. *Escherichia coli* and *Salmonella*: Cellular and molecular biology. Washington DC: American Society for Microbiology. pp. 822-848.
148. Gopal B, Haire LF, Cox RA, Colston MJ, Major S, et al. (2000) The crystal structure of NusB from *Mycobacterium tuberculosis*. Nat Struct Biol 7: 475-478.
149. Ojha AK, Baughn AD, Sambandan D, Hsu T, Trivelli X, et al. (2008) Growth of *Mycobacterium tuberculosis* biofilms containing free mycolic acids and harbouring drug-tolerant bacteria. Mol Microbiol 69: 164-174.
150. Piuri M, Hatfull GF (2006) A peptidoglycan hydrolase motif within the mycobacteriophage TM4 tape measure protein promotes efficient infection of stationary phase cells. Mol Microbiol 62: 1569-1585.
151. Gordon D, Abajian C, Green P (1998) Consed: a graphical tool for sequence finishing. Genome Res 8: 195-202.
152. Pandey AK, Raman S, Proff R, Joshi S, Kang CM, et al. (2009) Nitrile-inducible gene expression in mycobacteria. Tuberculosis (Edinb) 89: 12-16.
153. Thomason LC, Costantino N, Shaw DV, Court DL (2007) Multicopy plasmid modification with phage lambda Red recombineering. Plasmid 58: 148-158.
154. Kingsford CL, Ayanbule K, Salzberg SL (2007) Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. Genome Biology 8.
155. McElroy WD, DeLuca MA (1983) Firefly and bacterial luminescence: basic science and applications. Journal of applied biochemistry 5: 197-209.
156. Rosenow C, Saxena RM, Durst M, Gingeras TR (2001) Prokaryotic RNA preparation methods useful for high density array analysis: comparison of two approaches. Nucleic Acids Research 29: art. no.-e112.

157. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, et al. (2006) The RIN: an RNA integrity number for assigning integrity values to RNA measurements. Bmc Molecular Biology 7.
158. Ramesh G, Gopinathan KP (1995) Cloning and characterization of mycobacteriophage I3 promoters [published erratum appears in Indian J Biochem Biophys 1996 Feb;33(1):83]. Indian J Biochem Biophys 32: 361-367.
159. Stover CK, de la Cruz VF, Fuerst TR, Burlein JE, Benson LA, et al. (1991) New use of BCG for recombinant vaccines. Nature 351: 456-460.
160. Solovyev V, Salamov A (2011) Automatic Annotation of Microbial Genomes and Metagenomic Sequences. In: Li RW, editor. Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies: Nova Science Publishers. pp. 61-78.
161. Shell SS, Prestwich EG, Baek S-H, Shah RR, Sassetti CM, et al. (2013) DNA Methylation Impacts Gene Expression and Ensures Hypoxic Survival of Mycobacterium tuberculosis. Plos Pathogens 9.
162. Al-Zarouni M, Dale JW (2002) Expression of foreign genes in Mycobacterium bovis BCG strains using different promoters reveals instability of the hsp60 promoter for expression of foreign genes in Mycobacterium bovis BCG strains. Tuberculosis 82: 283-291.
163. Jain P, Hartman TE, Eisenberg N, O'Donnell MR, Kriakov J, et al. (2012) phi(2)GFP10, a high-intensity fluorophage, enables detection and rapid drug susceptibility testing of Mycobacterium tuberculosis directly from sputum samples. J Clin Microbiol 50: 1362-1369.
164. Kaufmann SH, Gengenbacher M (2012) Recombinant live vaccine candidates against tuberculosis. Curr Opin Biotechnol 23: 900-907.
165. Blount BA, Weenink T, Vasylechko S, Ellis T (2012) Rational diversification of a promoter providing fine-tuned expression and orthogonal regulation for synthetic biology. PLoS One 7: e33279.
166. Wang W, Li X, Wang J, Xiang S, Feng X, et al. (2013) An engineered strong promoter for streptomycetes. Appl Environ Microbiol 79: 4484-4492.
167. Seghezzi N, Amar P, Koebmann B, Jensen PR, Virolle MJ (2011) The construction of a library of synthetic promoters revealed some specific features of strong Streptomyces promoters. Appl Microbiol Biotechnol 90: 615-623.
168. Siegl T, Tokovenko B, Myronovskyi M, Luzhetskyy A (2013) Design, construction and characterisation of a synthetic promoter library for fine-tuned gene expression in actinomycetes. Metabolic Engineering 19: 98-106.
169. Pham TT, Jacobs-Sera D, Pedulla ML, Hendrix RW, Hatfull GF (2007) Comparative genomic analysis of mycobacteriophage Tweety: evolutionary insights and construction of compatible site-specific integration vectors for mycobacteria. Microbiology 153: 2711-2723.
170. Huff J, Czyz A, Landick R, Niederweis M (2010) Taking phage integration to the next level as a genetic tool for mycobacteria. Gene 468: 8-19.
171. Levin ME, Hatfull GF (1993) Mycobacterium smegmatis RNA polymerase: DNA supercoiling, action of rifampicin and mechanism of rifampicin resistance. Mol Microbiol 8: 277-285.

172. Jensen PR, Hammer K (1998) The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. Applied and Environmental Microbiology 64: 82-87.
173. Singh SS, Typas A, Hengge R, Grainger DC (2011) Escherichia coli Sigma(70) senses sequence and conformation of the promoter spacer region. Nucleic Acids Research 39: 5109-5118.
174. Liu MF, Tolstorukov M, Zhurkin V, Garges S, Adhya S (2004) A mutant spacer sequence between-35 and-10 elements makes the P-lac promoter hyperactive and cAMP receptor protein-independent. Proceedings of the National Academy of Sciences of the United States of America 101: 6911-6916.
175. Jain V, Sujatha S, Ojha AK, Chatterji D (2005) Identification and characterization of rel promoter element of Mycobacterium tuberculosis. Gene 351: 149-157.
176. Chauhan S, Singh A, Tyagi JS (2010) A single-nucleotide mutation in the-10 promoter region inactivates the narK2X promoter in Mycobacterium bovis and Mycobacterium bovis BCG and has an application in diagnosis. Fems Microbiology Letters 303: 190-196.
177. Shultzaberger RK, Chen Z, Lewis KA, Schneider TD (2007) Anatomy of Escherichia coli sigma(70) promoters. Nucleic Acids Research 35: 771-788.
178. Lewis JA, Hatfull GF (2000) Identification and characterization of mycobacteriophage L5 excisionase. Mol Microbiol 35: 350-360.
179. Morris P, Marinelli LJ, Jacobs-Sera D, Hendrix RW, Hatfull GF (2008) Genomic characterization of mycobacteriophage Giles: evidence for phage acquisition of host DNA by illegitimate recombination. J Bacteriol 190: 2172-2182.
180. Ferrell JE (2002) Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. Current Opinion in Cell Biology 14: 140-148.
181. Dubnau D, Losick R (2006) Bistability in bacteria. Molecular Microbiology 61: 564-572.
182. Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in Escherichia coli. Nature 403: 339-342.
183. Becskei A, Seraphin B, Serrano L (2001) Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. Embo Journal 20: 2528-2535.
184. Raju RM, Unnikrishnan M, Rubin DH, Krishnamoorthy V, Kandror O, et al. (2012) Mycobacterium tuberculosis ClpP1 and ClpP2 function together in protein degradation and are required for viability in vitro and during infection. PLoS Pathog 8: e1002511.
185. Broussard GW, Hatfull GF (2013) Evolution of genetic switch complexity. Bacteriophage 3: e24186.
186. Bonnet J, Subsoontorn P, Endy D (2012) Rewritable digital data storage in live cells via engineered control of recombination directionality. Proc Natl Acad Sci U S A 109: 8884-8889.
187. Friedland AE, Lu TK, Wang X, Shi D, Church G, et al. (2009) Synthetic gene networks that count. Science 324: 1199-1202.
188. Lu TK, Khalil AS, Collins JJ (2009) Next-generation synthetic gene networks. Nat Biotechnol 27: 1139-1150.
189. Hendrix RW, Casjens S (2006) Bacteriophage lambda and its genetic neighborhood. In: Calendar R, editor. The Bacteriophages. Oxford, UK: Oxford Univeristy Press. pp. 409-447.

190. Smith M (2006) Molecular genetics of streptomyces phages. In: Calendar R, editor. The Bacteriophages. Oxford: Oxford University Press. pp. 621-3635.