CHARACTERIZATION OF THE ROLE OF THE RPF MOTIF IN MYCOBACTERIOPHAGE TAPE MEASURE PROTEINS

by

Laura Jane Marinelli

Bachelor of Science, Youngstown State University, 2001

Submitted to the Graduate Faculty of

Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2008

UNIVERSITY OF PITTSBURGH

FACULTY OF ARTS AND SCIENCES

This dissertation was presented

by

Laura J. Marinelli

It was defended on

May 15th, 2008

and approved by

Karen Arndt, Ph.D. Associate Professor, Department of Biological Sciences

Roger Hendrix, Ph.D. Professor, Department of Biological Sciences

Saleem Khan, Ph.D. Professor, Department of Microbiology and Molecular Genetics

Valerie Oke, Ph.D. Assistant Professor, Department of Biological Sciences

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

CHARACTERIZATION OF THE ROLE OF THE RPF MOTIF IN MYCOBACTERIOPHAGE TAPE MEASURE PROTEINS

Laura J. Marinelli, Ph.D.

University of Pittsburgh, 2008

In order to inject their DNA into the bacterial cytoplasm and establish infection, bacteriophages must ensure their genetic material successfully traverses both the bacterial membrane(s) and the layer of peptidoglycan surrounding the host cell. Phages accomplish this in a variety of ways, and some have virion-associated murein hydrolase enzymes that facilitate this process, particularly in conditions where the peptidoglycan is highly cross-linked. Phages that infect the mycobacteria must also contend with these barriers to infection, as well an impermeable layer of mycolic acids that decorates the cell surface; however, the mechanisms by which they do this are mostly unknown. In this regard, three small sequence motifs have been identified within mycobacteriophage tape measure proteins (TMPs) – extended molecules that span the tail lumen and determine its length – at least two of which have similarity to host proteins with muralytic activity. This suggests that phages may utilize regions of the TMP, which because of its location within the tail might be uniquely primed for host interaction, to facilitate localized peptidoglycan hydrolysis and DNA injection.

The focus of this study is the motif found in the TMPs of mycobacteriophages Barnyard and Giles that has identity to a group of bacterial proteins known as resuscitation promoting factors (Rpfs). These Rpf proteins stimulate growth of non-growing bacteria and seem to exert their activity by cleaving inert peptidoglycan in the cell wall. Notably, the Barnyard Rpf Motif is contained within a 70 kDa C-terminal cleavage product of TMP, which appears to be cell wall-

and/or membrane-associated during infection. Further, mycobacteria expressing TMP fragments containing this motif show aberrant behavior in culture and on solid media, and hybrid proteins in which the Rpf domain of the *Micrococcus luteus* Rpf protein is replaced with either of the phage motifs have muralytic activity *in vivo.* A recombineering-based method for generating mutations on lytically replicating mycobacteriophages has been developed and utilized to make multiple mutations in the Giles TMP motifs. Mutant phages infect host cells in late-stationary phase with a reduced efficiency, an observation that further supports a role for these motifs in cell wall hydrolysis during infection.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

First and foremost, I want to thank Dr. Graham Hatfull, my thesis advisor, for his guidance, patience and overall enthusiasm for science. His positive attitude and support certainly helped to get me through all of the many difficulties and frustrations I encountered early during my graduate work, and for this I am very grateful. I am also appreciative of the way in which he created a work environment that is conducive to both independent thought and to collaboration, and in doing so, has set an excellent example to all of us who have been able to grow as scientists in his laboratory.

I also want to thank all of the other members of the Hatfull lab, whose presence and support contributed greatly to the completion of my dissertation. In particular, I need to acknowledge Dr. Lori Bibb and Dr. Julia van Kessel, who are both friends and valued colleagues. Lori became like a big sister to me in the early years of my graduate work, providing me with guidance and wisdom and, most importantly, a lasting friendship. Julia, my little Hatfull Lab sister also became one of my closest friends and provided me with support and encouragement, as well as valuable critical discussions throughout the completion of my dissertation. She has developed into an amazing scientist, and I am incredibly proud to see the success she has achieved. I am also indebted to her because it was largely through the work she has done in developing the mycobacterial recombineering system that I was able to create the Giles mutants, which facilitated the completion of this work.

Of course, I must also thank my friends: the people who have made my time in graduate school so enriching and enjoyable and are too numerous to mention. In particular, I extend my heartfelt thanks to my classmates, Dr. Heather Hendrickson, Dr. Maggie Braun, Dr. Lisa Sproul and of course Dr. Stephen Hancock, who became my closest friend and who lent his support and encouragement throughout the completion of my graduate work. Each of you means more to mean than I can possibly say here. Heather, Maggie and Lisa, you are inspirations both as women and as scientists. Stephen, you are truly the kindest, most generous person I have ever met, and I am so lucky to share my life with you. Megan Dietz, another fellow graduate student, was friend who impacted my life and will remain forever in my heart. I will always remember her love of life and her science, and I will try to take that inspiration with me, wherever I go and whatever I may do. I miss you Megan.

I want thank to my family as well, especially my parents and my sister Kristine, who provided unwavering love and support throughout my graduate career, and for which I will be forever grateful. I love you all so much, and I am so very proud to be a part of this family. I must also thank the Hancock family, for their kindness, love and generosity, as well as for always treating me like a member of the family.

I must acknowledge my committee for their guidance and patience, and for always helping me see the positive side of things. I am grateful also to Dr. Michael Young, from the University of Aberystwyth, who welcomed me into his laboratory, and to all of the people there who made my stay enjoyable. Thank you to Dr. Martin Pavelka as well, for providing the *M. smegmatis* cell wall mutants and for useful discussion. Lastly, I want to thank all of the people in the Biological Sciences Department at the University of Pittsburgh, who have touched my life in so many ways and who have made my time in Pittsburgh so memorable.

1.0 INTRODUCTION

It has become increasingly evident that tailed bacteriophages are by far the most abundant biological entities on the planet, and it is estimated that there are ~10^{24} new phage infections every second (Wommack & Colwell, 2000, Hendrix, 2002, Wilhelm *et al.*, 2002, Hendrix, 2003). All of these infections are absolutely dependent on the successful injection of viral DNA into the host cytoplasm, and accomplishing this is arguably one of the most challenging obstacles faced by bacteriophages. In order to establish a productive infection, phages must effectively transport their DNA across a formidable set of barriers that includes the bacterial membrane(s) and the layer of cell wall peptidoglycan surrounding the host bacterium (Dijkstra & Keck, 1996, Molineux, 2001, Molineux, 2006). For Gram-negative organisms such as *Escherichia coli*, phages must contend with two membranes and a thin layer of peptidoglycan (Vollmer & Holtje, 2001, Vollmer & Bertsche, 2007). Phages of Gram-positive bacteria must transport their DNA across only a single membrane; however, the peptidoglycan network is typically much thicker than that found in Gram-negative organisms (Vollmer *et al.*, 2008a).

A common solution to this problem that phages have employed is to incorporate proteins that hydrolyze peptidoglycan into various locations within their mature virion structures. Among the *E. coli* phages for example, T4 has a tail-associated protein with lytic activity (Kao & McClain, 1980, Nakagawa *et al.*, 1985, Kanamaru *et al.*, 2002), and others, such as T7, PRD1 and ϕ6, carry similar enzymes either within, or associated with, the phage capsid (Struthers-Schlinke *et al.*,

2000, Moak & Molineux, 2000, Rydman & Bamford, 2000, Mindich & Lehman, 1979). There is also evidence that phages infecting Gram-positive bacteria, such as *Lactococcus lactis*, *Staphylococcus aureus* and *Bacillus subtilis,* have virion-associated muralytic activities, and in the *L. lactis* phage Tuc2009, this protein has been shown to be a component of the phage tail (Moak & Molineux, 2004, Kenny *et al.*, 2004).

Phages that infect the mycobacteria – the mycobacteriophages – must also overcome these same barriers to infection, as well as contend with an impermeable layer of long carbon chain mycolic acids that decorates the cell surface of their hosts (Brennan & Nikaido, 1995, Brennan, 2003). The ways in which they accomplish this are mostly unknown; however, recent work from the Hatfull Laboratory has uncovered a possible solution to this dilemma. Embedded within a number of mycobacteriophage tape measure proteins (TMPs) – molecules that span the tail lumen and determine its length – are three small sequence motifs, which resemble various families of secreted bacterial proteins (Katsura & Hendrix, 1984, Pedulla *et al.*, 2003). It has been shown that proteins containing one of these motifs, Motif 3, have peptidoglycan hydrolase activity, and in mycobacteriophage TM4, this region of the TMP is important for the infection of stationary phase host cells (Piuri & Hatfull, 2006).

The subject of this investigation is another of these motifs referred to as Motif 1 or the Rpf Motif, a name that comes from the fact that this motif has sequence similarity to a family of proteins known as resuscitation promoting factors or Rpfs. These were initially characterized as proteins that have the ability to 'resuscitate' dormant bacterial cells, but it has subsequently been determined that Rpfs have both functional and structural similarity to cell wall hydrolases (Mukamolova *et al.*, 1998, Cohen-Gonsaud *et al.*, 2005, Mukamolova *et al.*, 2006). Thus, in this investigation I present data supporting the hypothesis that the Rpf Motif functions in a manner

similar to the TM4 Motif 3 and likely contributes to localized peptidoglycan hydrolysis during infection.

1.1 BACTERIOPHAGE TAPE MEASURE PROTEINS

The genomes of virtually all phages with flexible, non-contractile tails contain a gene that encodes for a protein known as the tape measure protein (TMP). These molecules are so named because there exists a significant body of evidence demonstrating that TMPs are phage structural proteins that function to precisely determine phage tail length. Although these proteins are often quite diverse in sequence, they are easily identified in that they are predicted to adopt a largely α-helical secondary structure and are frequently encoded by the longest gene in the phage genome (Katsura & Hendrix, 1984, Pedulla *et al.*, 2003). The tape measure gene is also often found directly downstream of two genes that are expressed via a translational frameshift (Levin *et al.*, 1993, Xu *et al.*, 2004), the products of which are believed, at least in some cases, to act as chaperones for the large TMP during synthesis and incorporation into the phage tail (Xu, 2001, Ph.D. Thesis). Further, as would be expected for a protein that determines tail size, there is a strong correlation between the predicted length of the tape measure and tail length. This relationship was first noted in bacteriophage λ and corresponds to ~1.5 Å of tail length per amino acid – the dimensions of an extended alpha helix (Katsura & Hendrix, 1984, Pedulla *et al.*, 2003).

1.1.1 The Tape Measure of Bacteriophage λ

The role of the tape measure has been investigated most extensively in the *E. coli* phage λ, where it has been demonstrated that this protein, gpH, which is essential for tail assembly (Weigle, 1966), is the determinant of phage tail length. Specifically, it was determined that small deletions or duplications in gpH give rise to tails that are proportionately shorter or longer by the expected amount (Katsura & Hendrix, 1984, Katsura, 1987). That is, tail measurements of both wild-type λ and the *tmp* mutants were observed to be consistent with a model in which the entire TMP is utilized to measure the length of the tail as an extended α-helix, displaying ~1.5 Å (0.15 nm) of length per amino acid residue. Thus, although a 100 amino acid C-terminal portion of the full-length protein (853 residues, ~92 kDa) is cleaved prior to head-tail attachment to form gpH* (~80 kDa), it appears that the entire protein is used as a molecular ruler to determine tail length (135 nm) (Hendrix & Casjens, 1974, Tsui & Hendrix, 1983). This protein also has an essential role in tail assembly, as it has been demonstrated that six molecules of gpH interact with a number of other tail proteins to form the initiator complex for tail formation (Casjens & Hendrix, 1974, Katsura & Kuhl, 1975). These are then thought to define tail length by acting as a scaffold upon which the major tail subunit, gpV, assembles (Katsura, 1983). When synthesis of the tail tube is complete and the correct length is achieved, the terminator protein gpU can bind to the top of the tail, triggering gpH cleavage followed by gpZ mediated tail activation and head attachment (Katsura, 1976, Tsui & Hendrix, 1983).

There is also evidence that the λ TMP may play a role in DNA injection. In an experiment where gpH mutants were created by illegitimate recombination between two defective prophages *in vivo* and selected for the ability to form functional phages, two internal deletion mutants – Δ8

(246 bp) and Δ13 (219 bp) – were isolated and analyzed. In addition to having shortened tail lengths, it was determined that the mutant phages with internal deletions in their tape measures displayed a reduced plaque size and a decreased efficiency of plating as compared to wild-type (Katsura & Hendrix, 1984). The mutants were not found to be defective for adsorption or phage release, consistent with the plating deficiency resulting from an impairment in DNA injection. Additional tape measure mutants were also constructed *in vitro* with no selection for phage viability. These included multiple deletions encompassing a wide range of sizes and one duplication, and as expected, the resulting phages displayed corresponding shorter and longer tails, respectively, when examined by electron microscopy (Katsura, 1987). Visual inspection, however, further revealed defects in DNA injection and loss of the tail tip in some mutants. The altered DNA injection properties of the gpH mutants imply multiple distinct functions for this protein. Thus, although there are middle portions of gpH that are dispensable for tail assembly and length determination, as up to 65% of the protein could be deleted and mutants still form tails of a defined length, these regions are clearly important for phage DNA injection and potentially the stability of tail tip. A possible role for gpH in DNA injection is also supported by the observation that mutations in this gene can be isolated that suppress certain mutations in the *E. coli ptsM* locus encoding the ManY membrane protein *(pel* mutants), which bind λ but do not allow phage DNA injection (Scandella & Arber, 1976, Elliott & Arber, 1978). The suppressor mutations are located in the middle of the gpH gene, and unlike wild-type λ, mutants are able to normally infect *pel*- strains.

The tape measure is predicted to span the lumen of the phage tail as an extended α-helix, and as such, would be in the direct path that the phage DNA must traverse during host infection. Localization of the TMP within the tail is supported by data demonstrating that this protein is

resistant to proteolytic digestion within the virion (Roessner & Ihler, 1984), as well as by the observation that loss of gpH* allows dye to penetrate the tail during EM staining (Katsura, 1976). Therefore, due to the diameter of the tail lumen (3 nm), it is likely that this protein must be released prior to the ejection of the viral genome. However, this model, including the destination of the TMP following release from the tail, remain speculative for most phages, and λ is the only phage for which data exist that suggests a possible fate for this protein upon host infection. Specifically, λ has the ability to irreversibly bind and inject its DNA into liposomes that contain the *Shigella* λ receptor (LamB), and when this occurs, the TMP protein is protected from proteinase digestion (Roessner & Ihler, 1984). It was therefore suggested that this protein, which has two predicted transmembrane segments of 35 uncharged amino acid residues, becomes imbedded in the liposomal membrane and thus, may contribute to pore formation in the bacterial membrane during infection.

1.1.2 Tape Measures in Other Bacteriophages

Another phage for which the function of the TMP has been demonstrated is TP901-1, a temperate virus that infects the Gram-positive lactococcal bacteria and which, like λ, is a member of the siphoviridae, having a long flexible tail of 135 nm (Jarvis *et al.*, 1991, Johnsen *et al.*, 1995, Pedersen *et al.*, 2000). The tape measure gene, *orf tmp (orf45,* 2814 bp) is predicted to encode a 100 kDa, 937-residue protein, the size of which correlates closely with the length of the tail, as has been observed in for λ gpH (Pedersen *et al.*, 2000). Further, like gpH, modeling suggests a largely α-helical secondary structure for this protein (Pedersen *et al*., 2000). The TP901-1 *tmp* gene is also located in a similar genomic context; with a translational frameshifting site encoded between the

two upstream genes – *orf46* and *orf47* – that gives rise to both the product of *orf46* and the translational fusion with *orf47 in vivo* (Pedersen *et al*., 2000, Mc Grath *et al.*, 2006).

The function of the TP901-1 tape measure was examined by Pedersen and colleagues, and they demonstrated that this protein is essential for tail assembly as well as for tail length determination (Pedersen *et al*., 2000). The authors utilized a *Lactococcus cremoris* TP901-1 lysogen to introduce specific mutations in the *tmp* gene. These included an amber mutation at codon 16, as well as an in-frame, N-terminal duplication that increased the length of the protein by 29%, and two in-frame deletions in both the 5′-(29% removed) and 3′- (41% removed) ends of the gene. Examination of the mutant phages by electron microscopy revealed a large number of head structures, which lacked tails and were unable to adsorb or efficiently infect lactococcal cells. However, preparations of mutant phage containing either the N-terminal duplication or the N-terminal deletion also contained a number of phages whose tails were lengthened by ~30% (170-180 nm) or shortened by ~30% (85-95 nm), respectively (Pedersen *et al.*, 2000). These and other data have led to the model that the TMPs of TP901-1 and related lactococcal phages, such as Tuc2009, act analogously to λ gpH and interact with other essential tail proteins to form the initiator complex for tail assembly (Pedersen *et al*., 2000, Vegge *et al.*, 2005, Mc Grath *et al*., 2006).

The role of the putative TMP of bacteriophage PSA, which infects the Gram-positive organism *Listeria monocytogenes,* has not been established. However, this protein shares a number of features with known TMPs. PSA has a long, flexible, non-contractile tail of ~180 nm, and the putative tape measure, the 112.1 kDa gp12, has similarity to TMPs from the *Lactococcus* phages bIL285 and TP901-1 (Zimmer *et al.*, 2003). This protein, at 1026 amino acids, is approximately long enough to span the length of the PSA tail as an extended α-helix, and accordingly, gp12 is

predicted to be highly α-helical and has a high coil-coiled potential. Furthermore, the products of the two small ORFs (*orf11* and *orf11a*) located directly upstream of the *tmp* gene are predicted to be produced via a translational frameshift. The second of these, *orf11a* overlaps *orf11* in the -1 frame, and there is a potential slippery sequence immediately upstream of the *orf11* stop codon, similar to what has been observed in λ and many other members of the siphoviridae (Levin *et al*., 1993, Xu *et al*., 2004).

Interestingly, there is evidence that phages with contractile tails – also known as the myoviridae – might also use a molecular ruler to specify tail length, despite the fact that their tails are morphologically quite different from those of the siphoviridae. Specifically, in T4, it has been demonstrated that the putative tape measure, gp29, is present within the tail tube and is long enough to span the length of the tail (Duda *et al.*, 1986). Further, similar to what has been shown for λ, deletions or duplications in this gene give rise to phages with correspondingly shorter or longer tails, respectively (Abuladze *et al.*, 1994). As the sequenced genomes of more bacteriophages become available, it is increasingly evident that although the sequences of the putative TMPs can differ significantly, most of these molecules share a number of the key conserved features described above.

The correlation between tape measure gene length and phage tail length has also been well documented for the mycobacteriophages, the group of phages that infect the mycobacteria, suggesting that these molecules also specify tail length (Pedulla *et al*., 2003). Again, the precise function of this gene product has not been investigated for most of these phages. However, in the case of TM4, it was confirmed that small deletions in the tape measure gene yield phages with correspondingly shorter tails of the expected length (Piuri & Hatfull, 2006). Furthermore, as was observed with λ, mutant phages were seen to be specifically defective in DNA injection (discussed

further in Section 1.4). These data collectively suggest that bacteriophage tape measures represent a family of multifunctional proteins, which are not only necessary for both tail assembly and tail length determination, but may also play a pivotal role in the process of bacteriophage DNA injection. Before discussing this possibility further, it is necessary to consider one of the primary obstacles faced by bacteriophages when attempting to inject their DNA into a host and establish infection: the bacterial cell wall.

1.2 THE CELL WALL

The bacterial cell wall is an extracellular macromolecular structure composed of an interlocked network of a heteropolymer known as peptidoglycan. This peptidoglycan or murein is made up of long glycan strands composed of alternating *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) aminosugars joined to each other by β-1,4-glycosidic bonds and covalently cross-linked by short peptides (Fig. 1A and B) (Vollmer & Bertsche, 2007). The 'peptide stems' are attached to the glycan backbone through an amide bond between the lactyl moiety of MurNAc and the amine group of L-alanine (L-Ala), typically the first amino acid residue in the stem (Vollmer *et al*., 2008a). This mesh-like network provides the bacterial cell with structural support and also functions as a protective barrier against the environment by preventing osmotic lysis and the free diffusion of large macromolecules (Dijkstra & Keck, 1996, Demchick & Koch, 1996). Despite these important protective and structural roles, the peptidoglycan network is a flexible, dynamic structure, and in *E. coli,* it has been estimated that about half of the cell wall peptidoglycan is turned over and recycled during every generation (Goodell, 1985, Koch & Doyle, 1985).

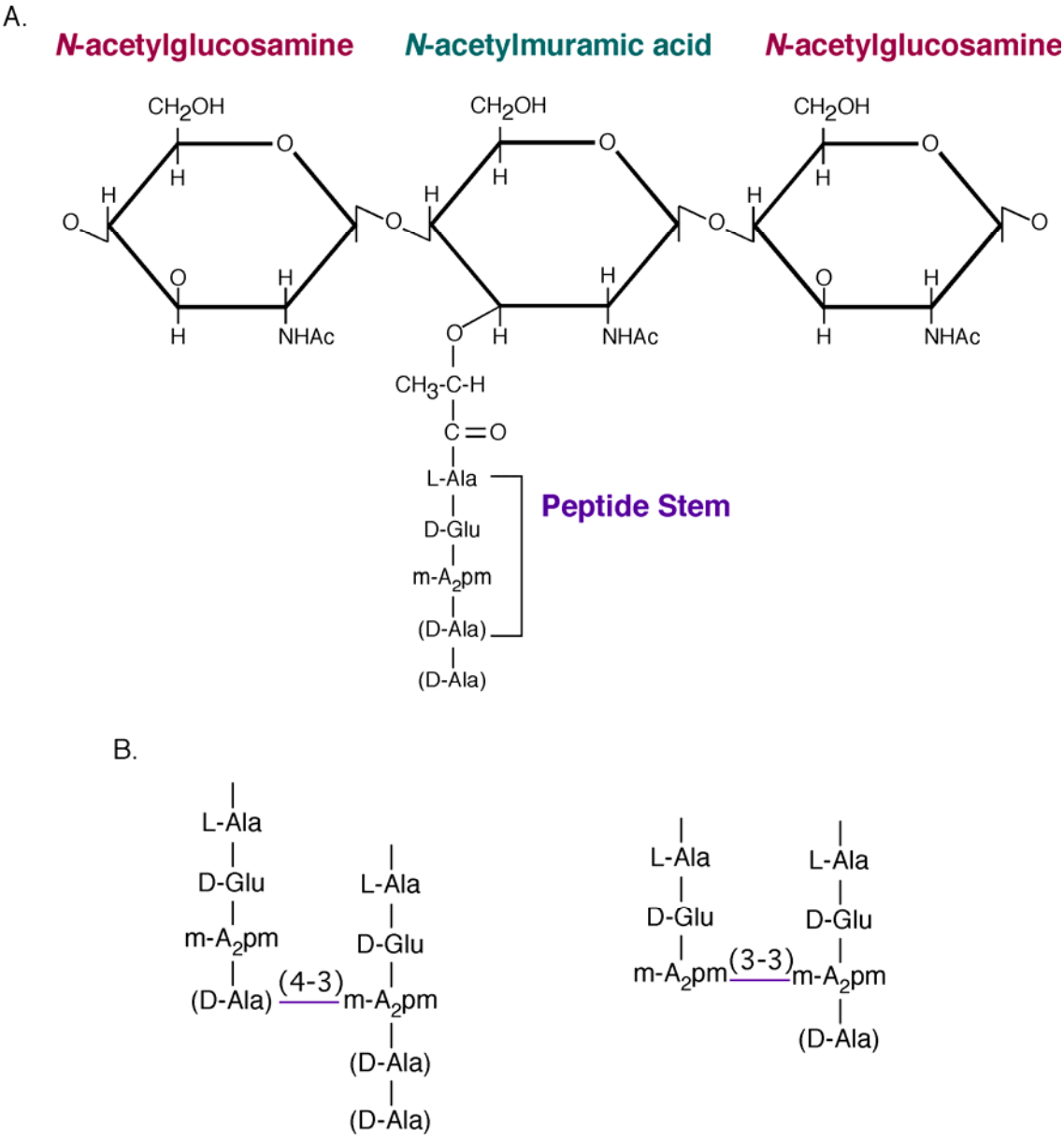


Figure 1. Basic peptidoglycan structure.

Figure 1: The structure of bacterial peptidoglycan. **A.** Bacterial peptidolgycan is composed of alternating units of *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) linked together by β-1,4-glycosidic bonds. Peptides stems are attached to these glycan strands via an amide bond with the lactyl group of MurNAc. **B.** Peptide stems vary in length and are commonly cross-linked through (4–3) or (3–3) linkages, which may be direct or indirect. Adapted from Vollmer & Bertsche, 2007.

Although the basic composition of the cell wall appears to be conserved in most bacteria, there are some striking differences. The glycan strands themselves can vary considerably in length and often contain modifications such as *N*-deacetylation, *O*-acetylyation or *N*-glycolation (Holtje, 1998, Vollmer *et al*., 2008a). The muropeptide (disaccharide peptide units) composition is another feature of the cell wall that can be quite heterogeneous. In general, the newly made peptide stems are composed of five amino acids with the sequence: L-Ala – D-isoglutamate (D-iGlu) – *meso*-diaminopimelic acid (*m*-A_2pm, an intermediate in lysine biosynthesis) – D-Ala – D-Ala (Fig. 1A) (Holtje, 1998, Vollmer & Bertsche, 2007). However, in the Gram-positive bacteria, as well as in the mycobacteria, the free carboxyl group of D-iGlu at position two is amidated to form D-isoglutamine (D-iGln) *(*Vollmer *et al*., 2008a). The *m*-A_2pm residue at position three is also amidated in some bacteria, and in most Gram-positive organisms (other than bacilli and the mycobacteria), this position is occupied by L-lysine (L-Lys) or a diamino acid other than *m*-A_2pm. The D-Ala residues at positions four and five are also subject to variability, and it is common for these to be replaced with glycine (Gly) in some organisms. Notably, in many cases, the pentapeptides are cleaved to form tetra- or tripeptides, such that the cell wall is actually composed of a mixture of protein chains of multiple different lengths (Vollmer & Bertsche, 2007).

There is considerable variability in the ways and extent to which the peptide chains are cross-linked. In *E. coli*, the majority of these are direct DD- or (4–3) peptide bonds occurring between the carboxyl group of D-Ala at position four and the ε-amino group at the D-center of *m*-A_2pm at position three on a neighboring chain (Fig. 1B) (Vollmer & Bertsche, 2007, Vollmer *et al*., 2008a). Another type, LD- or (3–3) links between the L-center of *m*-A_2pm at position three of one strand and the D-center of this residue on another strand (Fig. 1B) occur at a low frequency but increase during stationary phase and in strains resistant to β-lactam antibiotics (Vollmer *et al*.,

2008a). This linking was first discovered in the mycobacteria, and it makes up one-third of all cross-links in *Mycobacterium smegmatis*. It has further been observed that three or four peptide chains can be cross-linked together, and despite the fact that these trimers and tetramers occur at a lower frequency than standard dimers, these arrangements can join multiple glycan strands and therefore, may be important for the overall structure of the cell wall (Vollmer & Bertsche, 2007, Vollmer *et al*., 2008a).

The structure of the peptidoglycan allows it to be quite variable with respect to chain composition, length, modifications and degree of cross-linking, which is generally low in some organisms and higher in others (Vollmer *et al*., 2008a). This suggests that variations between different organisms may reflect optimization of cell walls for different lifestyles and environments. Even within one organism, such as *E. coli*, considerable variability can be observed depending on the strain, growth phase and growth conditions (Vollmer & Bertsche, 2007). This makes it difficult to make broad generalizations about groups of distantly related organisms. However, some general features that distinguish Gram-negative and Gram-positive cell walls, as well as the cell walls of the mycobacteria, are discussed below.

1.2.1 The Gram-Negative Cell Wall

The composition and fine structure of the cell wall has been studied most extensively in the model organism *E. coli.* This bacterium and other Gram-negatives typically have a thin cell wall (~6 nm in *E. coli* and ~2.5 nm in *Pseudomonas aeruginosa*) composed of one to three layers of peptidoglycan that is contained within the double membrane structure unique to these organisms (Vollmer & Bertsche, 2007). The cell walls of these bacteria are also characterized by the fact that they contain large amounts of murein lipoprotein, much of which is covalently attached through its

C-terminal amino acid residues via peptide bonds with *m*-A_2pm residues in the peptide stem (Braun & Wolff, 1970, Holtje, 1998). As the lipoprotein is also attached to the outer membrane through its terminal lipid residues, this serves to link the peptidoglycan and the cell envelope. As stated previously, the cell wall of *E. coli* has been shown to vary in a number of ways depending on variables such as growth phase and growth conditions. Specifically, overall levels of cross-linking are normally low but increase – particularly (3–3) cross-links – during stationary phase (Pisabarro *et al.*, 1985). Stationary phase cells also display an increase in the amount of lipoprotein and in the occurrence of tri-peptides, despite the fact that they also have a shorter average glycan strand length (Pisabarro *et al*., 1985, Vollmer & Bertsche, 2007). This may result partly from down-regulation of cell wall modifying enzymes and could potentially function to fortify the cell and give increased protection against environmental assaults.

1.2.2 The Gram-Positive Cell Wall

The cell wall of the Gram-positive bacteria has the same basic structure and in general serves the same functions as the Gram-negative cell wall. However, a major feature that distinguishes Gram-positive organisms is the thickness of the peptidoglycan layer in their cell walls, which encloses a single cytoplasmic membrane (Vollmer *et al*., 2008a). An examination of these structures by improved high resolution techniques has revealed that Gram-positive cell walls appear to consist of two distinct regions: a low density zone adjacent to the cell membrane and a high density zone located more towards the surface (Matias & Beveridge, 2005, Matias & Beveridge, 2006). The lower density 'inner wall zone' (IWZ) has been likened to the Gram-positive equivalent of the periplasmic space and ranges from 16-22 nm in those species examined. The 'outer wall zone' (OWZ) was observed to range in thickness from 15-30 nm, although this is likely to vary with the

species of bacteria, as well as with bacterial growth phase and conditions. Similar to what has been observed in *E. coli*, the peptidoglycan of some *Bacillus* species has been shown to contain increased levels of cross-linking when the cells are in stationary phase (Fordham & Gilvarg, 1974, Atrih *et al.*, 1999)

In most Gram-positive bacteria, the (4–3) cross links are indirect and occur through a short interpeptide bridge (Vollmer *et al*., 2008a). In general, bridges can be made up of one to seven amino acid residues, and their composition, as well as the extent to which individual residues within the bridge are modified (i.e. amidated, acetylated, etc.), are highly variable (Vollmer *et al*., 2008a). Another main type of cross-linking that has only been seen in the corynebacteria is the (2–4) link that forms between the α-carboxyl group of D-iGlu at position two and the carboxyl group of D-Ala at position four, and which requires a diamino acid containing interpeptide bridge for its formation (Vollmer *et al*., 2008a). As is the case with the Gram-negative organisms, Gram-positive peptidoglycan can also display considerable variability in peptide chain composition and degree of cross-linking (Navarre & Schneewind, 1999).

Another important and unique function of the Gram-positive cell wall is that it acts as a scaffold for the attachment of surface polymers, including teichoic acids, teichuronic acids, lipoteichoic acids, lipoglycans, capsular polysaccharides and cell surface proteins, such as protein A, fibronectin binding proteins, collagen adhesin and surface-layer glycoproteins (Navarre & Schneewind, 1999, Schaffer & Messner, 2005). Teichoic and teichuronic acids are structurally diverse anionic polysaccharide polymers that constitute 10-60% of the total Gram-positive cell wall mass, although like other cell wall features, this varies in different organisms and under different growth conditions (Schaffer & Messner, 2005). Teichoic acids are composed of various combinations of the negatively charged moieties polyglycerol phosphate, poly-ribitol phosphate

and poly-glucosyl phosphate, which are linked together and covalently joined to MurNAc in the peptidoglycan via a disaccharide tether termed the linkage unit (Navarre & Schneewind, 1999). Teichuronic acids are similar but contain glucuronic acid and carboxyl groups rather than phosphate (Schaffer & Messner, 2005). These polymers can contain a variety of sugar and amino acid modifications, and the individual linkage units can vary in different bacteria (Navarre & Schneewind, 1999). Proposed functions of these abundant molecules include the binding of divalent cations and cell surface proteins, including cell wall modifying enzymes, as well as providing increased protection against the environment and preventing the diffusion of valuable nutrients (Navarre & Schneewind, 1999, Schaffer & Messner, 2005). Anionic polymers that are tethered to the cell surface via a lipid group that is inserted into the cell membrane are referred to as lipoteichoic acids, and when the individual subunits within the chains lack phosphate, they are called lipoglycans (Navarre & Schneewind, 1999). These macromolecules can also contain a number of species-specific modifications; however, in most cases their functions are not known.

Cell surface proteins are often covalently linked directly to the peptidoglycan peptide stem by the membrane protein sortase A, although they may also be noncovalently associated with the peptidoglycan or anionic polymers in the cell wall (Navarre & Schneewind, 1999, Vollmer *et al*., 2008a). Some of these proteins are important for pathogenesis and survival within the host, and in general, many are believed to assist bacterial survival in specific environments or to control cell wall biogenesis during growth and cell division (Navarre & Schneewind, 1999, Schaffer & Messner, 2005).

1.2.3 Mycobacteria and the Mycobacterial Cell Wall

The genus *Mycobacterium* includes a number of pathogenic and saprophytic species, many of which can act as opportunistic pathogens, particularly in immunocompromised hosts. The most well known member of this group, *Mycobacterium tuberculosis*, is an extraordinarily successful pathogen that causes tuberculosis (TB) and is responsible for more deaths annually than any other infectious agent (World Health Organization (WHO)). It is estimated that one-third the Earth's population (~2 billion people) is infected with *M. tuberculosis*; however, most of these infections are latent. That is, infected individuals harbor quiescent bacteria in their lungs but are not contagious and asymptomatic, and thus, they act as a reservoir for the disease (Parrish *et al.*, 1998). In most people who become infected, after the initial period of bacterial replication and colonization, the organisms will transition into this latent state, and it may be many years before they reactivate to cause active infection (Manabe & Bishai, 2000). During latency, the bacteria are quite resistant to antibiotics, and a six-month combination therapy is recommended to clear the infection. Despite intensive study, little is known about latent TB, and the factors that control this are not well understood. The immune system is known to play a role, as reactivation is quite common in immunocompromised individuals, such as those infected with HIV (Flynn & Chan, 2001). It is also clear, however, that there are many features of *M. tuberculosis* that contribute to its pathogenicity, one of these being the unique structure of the mycobacterial cell wall (Daffe & Draper, 1998).

The cell wall of the mycobacteria is fundamentally different from the cell walls of both Gram-negative and Gram-positive bacteria in a number of ways; yet, this distinctive structure also has features in common with both. As mentioned previously, the peptide component of the peptidoglycan contains amidated *m*-A_2pm at position three, which is more commonly observed in

the Gram-negatives; however, the D-iGlu residues at position two are amidated to D-iGln, a feature associated with Gram-positive cell walls (Brennan & Nikaido, 1995, Vollmer *et al*., 2008a). A large proportion of (3–3) cross-links between two *m*-A_2pm residues is also found within this network (Wietzerbin *et al.*, 1974), and there is increasing evidence that, similar to the Gram-negative bacteria, these organisms possess an outer membrane lipid bilayer, thus creating a periplasmic space between this structure and the inner membrane (discussed below) (Minnikin, 1982, Brennan & Nikaido, 1995, Hoffmann *et al.*, 2008).

The mycobacteria are grouped in the G+C rich *Actinomycetales* under the suborder *Corynebacterineae*, and as such, they produce branched, long-chain α-alkyl, β-hydroxy fatty acids, the mycolic acids, which are one of the hallmarks of their 'chemotype IV' cell walls (Barry *et al.*, 1998). In the mycobacteria, these average 70-90 carbons in length, making them significantly longer than the 40-60 carbon chains found in *Nocardia* and *Corynebacterium* (Brennan & Nikaido, 1995). They contain a saturated α-branch, which has an average chain length between 20-25 carbons, and a β-hydroxy branch (main chain meromycolate) that averages 60 carbons (Fig. 2C). Mycobacterial mycolic acids are also unique in that their main meromycolate chains contain double bonds and/or cycloprane rings, and some have oxygenated functional groups, all of which can cause the chains to 'kink' and would interfere with regular packing of the strands (Brennan & Nikaido, 1995, Chatterjee, 1997). These kink-introducing groups are typically located toward the cell distal end of the molecules. Therefore, although these regions are likely to be fluid and disordered, the packing is predicted to be more regularly structured and compact closer to the surface of the cell (Brennan & Nikaido, 1995). This is believed to form a gradient of fluidity and permeability that diminishes near the surface of the cell, thus contributing the extreme impermeability of these organisms (Brennan, 2003).

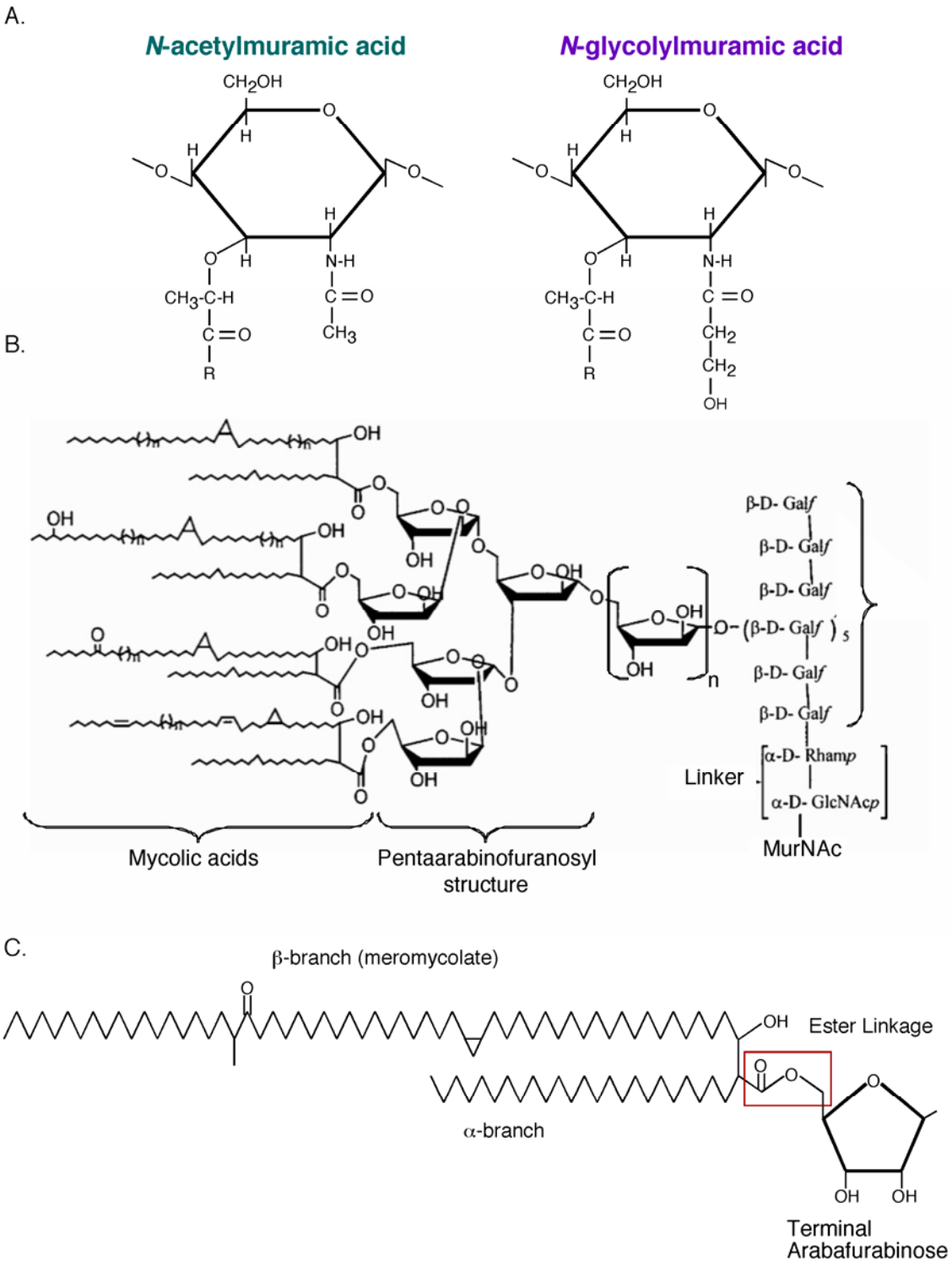


Figure 2. The mycobacterial cell wall.

Figure 2: Components of the mycobacterial cell wall. **A.** Mycobacterial peptidoglycan contains *N*-glycolated rather than *N*-acetylated muramic acid. **B.** Basic structure of the mycolated-arabinogalactan (mAG) (adapted from Tripathi *et al*., 2005 and K. Clemens, 2007). **C.** Mycolic acid esterified to arabinofuranose. Groups predicted to promote 'kink' formation in the long-carbon chains and to interfere with packing are located toward the cell distal end of the molecule.

The mycolic acids are linked to a layer of arabinogalactan, a polysaccharide composed primarily of two rare sugars: a D-galactofuranose core with branches of D-arabinofuranose that terminate in pentaarabinofuraosyl structures, approximately two-thirds of which are esterified by four mycolic acids (Fig. 2B) (Brennan & Nikaido, 1995, Chatterjee, 1997). This structure is therefore referred to as the mycolated arabinogalactan (mAG). The D-galactan core of the mAG is linked to a subset of the muramic acid residues in the glycan portion of the cell wall through a rhamnose-*N*-acetylglucosamine linker that resembles the linkage units connecting muramic acid to the teichoic acids in other Gram-positive bacteria (Chatterjee, 1997). An interesting feature that distinguishes the chemotype IV cell walls is that the muramic acids are *N*-glycolated rather than *N*-acetylated (Fig. 2A) (Azuma *et al.*, 1970). In *M. smegmatis*, this modification is catalyzed by the NamH hydroxylase, a protein that is conserved in all mycobacteria except *Mycobacterium leprae*, and although deletion mutants are viable, they display an increased sensitivity to lysozyme and β–lactam antibiotics (Raymond *et al.*, 2005). The reasons that these specific phenotypes are observed is not known; however it has been proposed that the β–lactam sensitivity may be due to a decreased level of cross-linking that results from the dual function transpeptidase/transglycosylases (see Section 1.2.4) having difficulty in recognizing abnormal muramic acid precursors. In contrast, it was proposed that lysozyme sensitivity increases because

the *N*-glycol group on muramic acid normally functions to block the β-1,4 bond recognized by this enzyme (Raymond *et al.*, 2005).

On the very outside of mycobacterial cells, there are a number of different, and often unique, extractable lipid and glycolipid moieties that are non-covalently associated with the mycolic acids (Brennan & Nikaido, 1995). A large number of these moieties have been identified, and although the specific lipid composition varies in different mycobacterial species, some, such as lipoarabinomannan (LAM), are more broadly distributed (Brennan & Nikaido, 1995). These molecules, together with surface-associated polysaccharides, form a fluid outer leaflet on the surface of mycobacterial cells, and it was proposed that their hydrophobic groups would intercalate with the inner leaflet of mycolic acids to form an asymmetric lipid bilayer (Minnikin, 1982, Brennan & Nikaido, 1995). Recent ultrastructural investigations of the cell wall structure of various mycobacteria and corynebacteria confirm the presence of an outer membrane but suggest that the two leaflets of this structure contain similar phospholipid moieties and thus, are more symmetric than previously thought (Hoffmann *et al.*, 2008). It is not known how the mycolic acids are incorporated into this structure, but they are an essential component, as *Corynebacterium glutamicum* mutants without mycolic acids also lack the outer membrane.

A detailed analysis of changes that might occur to the mycobacterial cell wall during the transition from exponential to stationary phase has not been conducted. However, it has been reported that mycobacterial cells subjected to gradual oxygen depletion in stationary phase appear to have thickened cell walls (Cunningham & Spreadbury, 1998) and similar to what has been observed in *E. coli*, may show an increased incidence of (3–3) cross-links joining the peptide stems (M. Pavelka, personal communication). Further, although recent studies have also shown that at 35 nm, the mycobacterial cell wall is thinner than once believed (Hoffmann *et al*., 2008), it

cannot be disputed that this incredibly complex, multi-layered structure makes the mycobacteria extremely impermeable to assault from many dangers in the environment including antibiotics, chemotherapeutic agents and disinfectants, as well as – one might imagine – bacteriophages.

1.2.4 Cell Wall Modifying Enzymes

The task of maintaining and remodeling the cell wall during growth and cell division by inserting new glycan strands and recycling older ones is performed by cell wall modifying enzymes, which include both murein synthases and murein hydrolases (Holtje, 1998).

The murein biosynthetic enzymes include transglycosylases, which add disaccharides to the growing glycan strand, transpeptidases, which catalyze the peptide cross-linking reactions, as well as proteins having both of these enzymatic activities. Transpeptidases have the ability to covalently bind penicillin and other β-lactam antibiotics, and are therefore often referred to as penicillin binding proteins (PBPs) (Vollmer & Bertsche, 2007). Proteins with these activities, together with ATP-dependent ligases that successively add residues to the peptide stem, are responsible for the expansion and maintenance of the cell wall during cell growth and division (Vollmer & Bertsche, 2007).

The murein hydrolases cleave covalent bonds within the cell wall and are responsible for cell wall degradation and recycling. These enzymes also play an important role in septation, as well as in the incorporation of newly synthesized cell wall material (Holtje, 1995). Similar to the peptidoglycan synthases, this group is made up of a number of proteins with different enzymatic activities (Fig. 3A): endopeptidases hydrolyze the peptide bridge cross-links, *N*-acetylmuramoyl-L-alanine amidases (amidases) cleave the amide bond that links the D-lactyl group of MurNAc to the amino group of the primary L-Ala in the peptide stem, and β-*N*-acetylglucosaminidases

(glucosaminidases) target and hydrolyze the glycosidic bond between GlcNAc and MurNAc (Navarre & Schneewind, 1999). The hydrolases also include the enzymes known as β-*N*-acetylmuramidases (muramidases), which target the bond between MurNAc and GlcNAc and can further be classified as lysozymes or lytic transglycosylases.

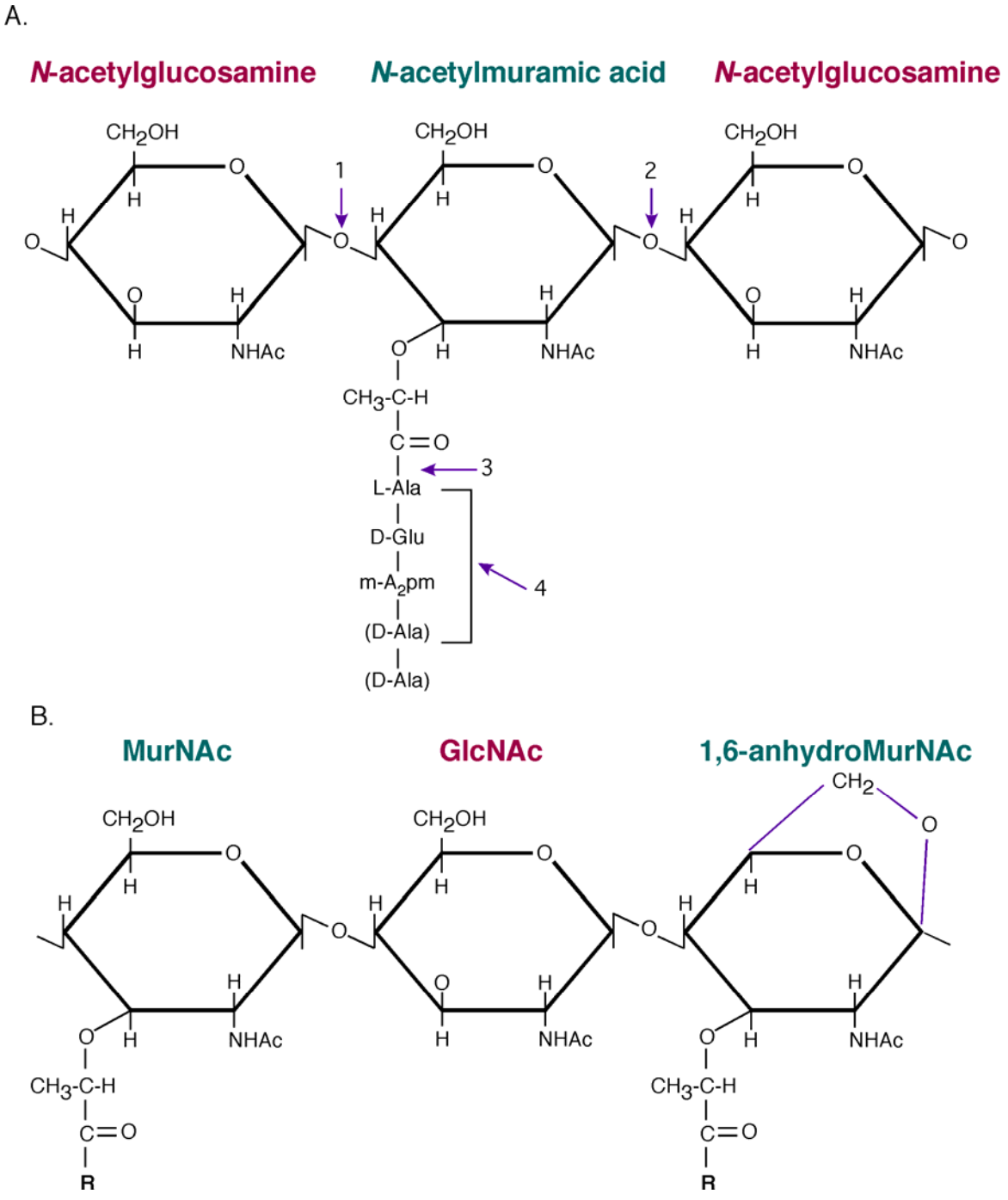


Figure 3. Targets of the murein hydrolases within the cell wall.

Figure 3: A. Sites of action of the various cell wall hydrolases are indicated with arrows: 1) glucosaminidases 2) muramidases 3) amidases 4) endopeptidases. **B.** The product of the lytic transglycosylases reaction is 1,6-anhydroMurNAc acid, a non-reducing intramolecular ring formed between C1 and C6 (adapted from Vollmer & Bertsche, 2007).

Lysozymes are 1,4-β-*N*-acetylmuramidases that cleave the β-1,4-glycosidic bond between MurNAc and GlcNAc residues in the cell wall. This reaction releases the energy of the cleaved bond and produces a MurNAc residue with a reducing end. Lytic transglycosylases are also muramidases that catalyze cleavage of the β-1,4-glycosidic bond between MurNAc and GlcNAc, but these enzymes catalyze an intramolecular transglycosylation reaction that produces a nonreducing, 1,6-anhydro-bond in the MurNAc residue and preserves the energy of the hydrolyzed bond in a ring structure (Fig. 3B) (Holtje *et al.*, 1975). The conservation of the bond energy is what distinguishes these enzymes from lysozymes. Interestingly, 1,6-anhydromuramic acid (the product of this reaction) is found at the free ends of glycan strands in the cell wall, suggesting that they are produced by the action of lytic transglycosylases (Holtje, 1998).

Lytic transglycosylases are found in both Gram-negative and Gram-positive bacteria, and at least six have been described in *E. coli* (Vollmer & Bertsche, 2007). Many of these are membrane-bound; however, one well-characterized member of this group is SltY (70 kDa), a soluble periplasmic enzyme that is strongly associated with the cell wall (Engel *et al.*, 1991, Walderich & Holtje, 1991). These proteins contain an absolutely conserved active site glutamate (E478 in SltY), and mutations at this site completely abolish activity (Thunnissen *et al.*, 1994, van Asselt *et al.*, 1999). Three highly conserved motifs have been identified, one of which surrounds the active site residue, and these are believed to be important for the architecture of active site and substrate binding (Koonin & Rudd, 1994, Dijkstra & Thunnissen, 1994, Mushegian *et al.*, 1996). Notably, although they have little overall sequence similarity with lysozymes, the critical active site residues are conserved. Further, at a structural level, lytic transglycosylases adopt the same 'Lysozyme-like fold' as lysozyme and share many of the same structural features (Thunnissen *et al.*, 1994, Thunnissen *et al.*, 1995, van Asselt *et al.*, 1999).

The cell wall is an interconnected, mesh-like structure, and therefore, its enlargement necessitates both biosynthetic and hydrolytic activities in order to safely incorporate new material (Vollmer & Bertsche, 2007). These activities also need to be tightly coordinated, both spatially and temporally, particularly during cell division, in order to maintain the integrity of the structure (Holtje, 1998). This has been substantiated as a number of studies have detected both genetic and biochemical interactions between many of these cell wall modifying enzymes, both in *E. coli* and in Gram-positive organisms such as *S. aureus* and *B. subtilis* (Holtje, 1998, Vollmer & Bertsche, 2007). These data have led to the proposal that multi-enzyme complexes, containing both synthetic and hydrolytic activities, modulate the structure of the cell wall; an idea supported by the observation that a number of lytic transglycosylases, including Slt70, do not cause cell lysis or are unstable when overexpressed (Holtje, 1998). This suggests that their activity might somehow be modulated through interactions with other proteins, and Slt70 in particular shows increased stability when the endopeptidase BPB7 is added to protein preparations (Romeis & Holtje, 1994). Many of these enzymes also interact closely with cell cycle proteins and the cell division machinery, such that they are directed to the sites of murein synthesis during septum formation.

1.2.5 The Cell Wall as a Barrier

The cell wall generally acts as a barrier against large molecules, such as proteins larger than 50 kDa, a feature that is believed to be strongly influenced by the degree of cross-linking and thus the 'mesh-size' of the peptidoglycan (Demchick & Koch, 1996, Vollmer *et al*., 2008a). Therefore, some processes that require the passage of large compounds across this barrier are believed to utilize lytic transglycosylases to mediate localized peptidoglycan hydrolysis (Dijkstra & Keck, 1996). Flagellar assembly is one example where this occurs, as peptidoglycan hydrolyzing activity

has been shown to be required for this process (Nambu *et al.*, 1999). Localized peptidoglycan hydrolysis has also been predicted to occur in the assembly of the type IV secretion apparatus and during the formation of conjugative pili for plasmid transfer. The VirB1 protein from the *Agrobacterium tumefaciens* plant tumor-inducing T-DNA transfer system has both sequence and predicted structural similarity to lytic transglycosylases, and this protein – specifically the putative active site residue – has been shown to be important for plant pathogenesis and efficient DNA transfer from bacterial cells to plant cells (Mushegian *et al*., 1996). Further, the P19 protein from the plasmid R1 is known to cause localized degradation of the peptidoglycan during conjugal transfer and greatly enhances the efficiency of this process (Bayer *et al.*, 1995, Bayer *et al.*, 2001). Both VirB1 and P19 have sequence similarity to proteins encoded by other protein and DNA transfer systems, including type IV secretion systems found in pathogenic *Bordetella* and *Bartonella spp.,* suggesting that murein hydrolases are broadly utilized to assist in the transport of large molecules across the cell wall. As discussed in the following section, another example of hydrolase proteins being used in this capacity is during the ejection of bacteriophage DNA into the host cytoplasm.

1.3 BACTERIOPHAGE-ASSOCIATED HYDROLASE PROTEINS

In general, the early stages of bacteriophage infection have been well described (Molineux, 2006). Random collision of a phage with its host is thought to result in a transient, non-specific interaction whereby the phage can search for receptors on the cells surface. Reversible adsorption occurs when the phage adsorption protein (often a tail fiber) forms a weak, specific interaction with a structure, such as a protein or carbohydrate, on the cell surface. Irreversible adsorption for many

phages then occurs when all tail fibers are bound and correctly positioned, allowing another tail protein to precisely contact and form a strong, specific interaction with the phage receptor.

Once receptor binding occurs, phages must transport their genetic material across the bacterial cell wall and membrane(s) in order for it to reach the host cytoplasm. How this is accomplished is not fully known for most phages; however, a strategy that many dsDNA phages have employed is to carry a virion-associated protein with murein hydrolase activity to assist in this process by forming or enlarging a hole in the bacterial cell wall peptidoglycan (Moak & Molineux, 2000, Moak & Molineux, 2004).

1.3.1 Phages of Gram-Negative Hosts

Gram-negative organisms possess a relatively thin layer of peptidoglycan that is surrounded by two lipid bilayer membranes (Vollmer & Holtje, 2001). Thus, the viral DNA must traverse all of these barriers as well as a periplasmic space, which is known to contain nuclease enzymes.

For the short-tailed *E. coli* phage T7, the mechanism whereby the phage accomplishes this has been investigated in some detail. Following irreversible adsorption, five virion proteins are ejected into the host. The first to enter are two small proteins: a head protein, gp6.7, and a tail protein, gp7.3, which are thought to function in viral morphogenesis and adsorption and are subsequently degraded by the host (Kemp *et al.*, 2005). The internal head proteins gp14, gp15 and gp16, which in the phage form an internal core composed of four copies of gp16 (144 kDa), eight copies of gp15 (84 kDa) and ten copies of gp14 (21 kDa), are then ejected from the capsid to form a structure that has been likened to an extensible tail (Molineux, 2001, Kemp *et al.*, 2005). These proteins together form a protective channel for the DNA as it crosses the periplasmic space and cell envelope and enters the host cytoplasm (Moak & Molineux, 2000, Molineux, 2001).

Interestingly, because gp14, gp15 and gp16 are located in the phage capsid, they must pass through the extremely narrow head-tail connector (~2.2 nm at the narrowest point) (Donate *et al.*, 1988). Thus, they must be in a completely unfolded conformation during injection and re-fold once they reach their destination, although the mechanism by which this occurs has not been elucidated (Molineux, 2001, Kemp *et al.*, 2005).

One member of this three-protein complex, gp16, has a lytic transglycosylase motif associated with its N-terminus that is important for efficient infection under conditions where the peptidoglycan is predicted to be more highly cross-linked. That is, although the conserved active-site residue, glutamate 37, is not essential, phages with mutations in this residue display a reduced efficiency of infection when *E. coli* cells are grown to higher cell densities in late exponential phase and at low temperatures (Moak & Molineux, 2000). These phenotypes are exacerbated in *E. coli* deleted for the periplasmic lytic transglycosylase SltY and can be partially complemented by overexpression of SltY. Further, gp16 from T7 virions has murein hydrolase activity *in vitro* that is lost when the active site glutamate is mutated (Moak & Molineux, 2004). This also appears to be a multifunctional protein, as mutational analyses of a central region of gp16 suggest that it may act as a clamp to prevent transcription independent internalization of the T7 genome (Struthers-Schlinke *et al*., 2000). These data combine to strongly support a pivotal role for this protein in the early stages of phage infection and DNA injection. Murein hydrolase activity has further been demonstrated *in vitro* for gp16-like proteins in the closely related phages, T3 and the yersiniophage øYeO3-12, as well as for virion-associated proteins in the more distantly related *Salmonella* phage SP6 and coliphage K1-5, suggesting that these might function in a manner similar to T7 gp16 (Moak & Molineux, 2004).

Another well-characterized *E. coli* phage, T4, possesses a tail-associated protein with lysozyme-like (*N*-acetylmuramidase) activity (gp5) that is a constituent of, and essential for the formation of, the central hub of the baseplate (Kao & McClain, 1980, Nakagawa *et al.*, 1985, Mosig *et al.*, 1989, Kanamaru *et al*., 2002). The lysozyme activity is activated by proteolytic cleavage, and mutations that abolish this cleavage show reduced activity and infectivity (Kanamaru *et al.*, 2005). Further, both portions of the protein remain in the mature virion structure: the N-terminal portion that is responsible for catalysis, as well as protein-protein interactions and possibly substrate binding, and a C-terminal piece that forms a needle-like structure at the tip of the baseplate (Kanamaru *et al.*, 1999, Kanamaru *et al*., 2002). The needle is thought to be responsible for puncturing the outer membrane when, following irreversible adsorption, the tail sheath contracts and pushes the tail tube through the cell envelope (Leiman *et al.*, 2004, Rossmann *et al.*, 2004). This would presumably allow the catalytic portion of gp5 to access and digest the cell wall peptidoglycan, creating a hole for the tail tube to contact the inner membrane. Because gp5 is essential, it has been difficult to define a role for the lysozyme function during infection; however, there is recent evidence that cleavage-defective mutants having 10-fold reduced lysozyme activity are cold sensitive, similar to mutants of the T7 transglycosylase gp16 (S. Kanamaru, personal communication with I. J. Molineux, Moak & Molineux, 2004). Genomic analyses have identified gp5 homologues in all of the T-even phages sequenced to date; however their activity has not been demonstrated (Rossmann *et al.*, 2004).

The phage PRD1 – a broad host-range, dsDNA virus that infects cells containing conjugative P, N or W-type plasmids – utilizes a different strategy to initiate infection. Its genome is encased in a membrane vesicle, which contains two proteins with muralytic activity: P15, a 1,4-β-*N*-acetylmuramidase thought to act primarily at the end of infection and P7, a non-essential lytic

transglycosylase believed to facilitate genome entry (Rydman & Bamford, 2000, Rydman & Bamford, 2002). This phage does not have an external tail to mediate host association, but rather, PRD1 has receptor-binding spike-penton complexes located at each vertex of the capsid (Rydman *et al.*, 1999). Following adsorption, an opening in the capsid forms through which a tubular extension of the viral membrane is extruded into the host cell envelope (Rydman *et al*., 1999). The transglycosylase protein P7 is anchored to the viral membrane, and in the current model, it performs localized peptidoglycan hydrolysis to assist the viral 'tube-tail' in reaching the cytoplasmic membrane (Rydman & Bamford, 2000, Grahn *et al.*, 2002). P7 has been hypothesized to form membrane-anchored, multimeric complexes near the phage receptor, and viruses lacking this protein are delayed for DNA injection (Rydman & Bamford, 2000, Grahn *et al*., 2002).

A fourth example is the lytic enzyme, P5, of the dsRNA, *Pseudomonas* phage ϕ6, a protein with apparent endopeptidase activity that is located between the nucleocapsid and surrounding membrane and functions both early and late in infection (Mindich & Lehman, 1979, Caldentey & Bamford, 1992). During the initial stages of ϕ6 infection, a fusion between the phage membrane and the bacterial outer membrane allows P5 to access and locally digest the peptidoglycan (Bamford *et al.*, 1987). This activity is important for subsequent contact and penetration of the host inner membrane by the nucleocapsid (Mindich & Lehman, 1979, Romantschuk *et al.*, 1988). Interestingly, recent extensive bioinformatic analyses have determined that P5 contains conserved sequence motifs found in lysozymes and lytic transglycosylases, and this protein likely adopts a lysozyme-like fold (Pei & Grishin, 2005). Previous data suggested that P5 cleaves peptide bonds between the ε-amino group of *m*-A_2pm and the carboxyl group of D-Ala; however, the assay used to detect glycosidase activity would not have been sensitive to the non-reducing sugars, which result from cleavage by transglycosylases (Caldentey & Bamford, 1992, Pei & Grishin, 2005).

Thus, the question remains as to whether P5 may have two enzymatic functions or whether it represents an example of a transglycosylase-like protein that evolved a peptidase function.

The coliphage T5 provides another interesting case of a virion-associated hydrolase in a phage that infects Gram-negative bacteria. This protein, pb2, was initially identified as the 50 nm-long, straight tail fiber (analogous to λ gpJ) that is inserted into the cell membrane during DNA injection (Guihard *et al.*, 1992). It was further shown that pb2 has pore forming activity and is inserted into liposomal membranes containing the phage receptor *in vitro,* undergoing large conformational changes in the process (Feucht *et al.*, 1990, Lambert *et al.*, 1998, Bohm *et al.*, 2001). Notably, a short metallopeptidase motif has been detected near the C-terminal end of this protein, and a truncated fragment of pb2 containing this region was shown to have muralytic activity *in vitro* and bacteriocidal activity against *E. coli* when added exogenously in culture (Boulanger *et al.*, 2008). These observations are particularly intriguing in light of the fact that in the same study, it was demonstrated that pb2 is predicted to be structurally quite similar to phage tape measures. That is, this protein contains multiple regions predicted to form coiled coils, as well as alanine/glycine rich regions with putative transmembrane domains (Boulanger *et al*., 2008). Similar to λ gpH, it is also proteolytically processed near the C-terminus and is present in approximately six copies within the virion (Zweig & Cummings, 1973a, Zweig & Cummings, 1973b). Thus, it is clear that pb2 shares a number of interesting structural features with bacteriophage tape measures that may be indicative of a common function, and this would therefore represent a case where a virion murein hydrolase is contained within the phage TMP. However, it remains to be determined whether this protein specifically functions in tail length determination and how it might be involved in tail assembly.

In addition to these more well-characterized examples, it has been demonstrated that the *Salmonella* phage P22, the coliphage N4, and the *Xanthomonas campetstris* phage Xp10 all have virion-associated muralytic activity that can be demonstrated *in vitro* (Moak & Molineux, 2004). In the case of P22, there is evidence that the muralytic protein, gp4, is associated with the DNA and/or the head-tail connector along with four other proteins that are known to be injected into the host cell during infection (Steinbacher *et al.*, 1997). Bioinformatic analyses further suggest that other phages, such as P1, may contain enzymes with lytic activity that have yet to be characterized (Lehnherr *et al.*, 1998). Thus, the presence of murein hydrolase enzymes within the virions of the phages of Gram-negative bacteria appears to be a common phenomenon.

1.3.2 Phages of Gram-Positive Hosts

Phages of the Gram-positive bacteria have only a single cytoplasmic membrane to contend with; however, this is protected by a thick layer of peptidoglycan (Vollmer *et al*., 2008a). Thus, it would seem likely that phages infecting these organisms would also employ virion-associated murein hydrolase enzymes to contend with this obstacle. While this has not been well established for most of these viruses, there is a growing body of evidence that this strategy may indeed be commonly utilized.

The best-studied example of this is the *L. lactis* phage Tuc2009, which possesses a tail protein having muralytic activity (Kenny *et al*., 2004). Unlike many of the other viruses where this phenomenon has been observed, Tuc2009 has a λ-like morphology with a long, non-contractile tail. It was shown that an ~100 kDa phage tail structural protein, designated Tal_{2009} for tail-associated lysin, displayed weak murein hydrolase activity against lactococcal cell walls *in vitro*. The catalytically active portion of this protein was localized to a C-terminal region with amino

acid sequence identity to members of the M23/M37 family of cell wall endopeptidases. It was further demonstrated that Tal_{2009} appears to undergo an autocatalytic cleavage when expressed endogenously that liberates the C-terminal catalytically active portion from the larger N-terminus. However, specifically how and why this happens is uncertain, as despite the fact that processing occurs *in vivo* during phage infection, only full-length protein and the larger N-terminal fragment are detected in mature virions (Kenny *et al.*, 2004, Mc Grath *et al.*, 2006). Antibodies raised against the muralytic C-terminus of Tal_{2009} inhibit Tuc2009 infection and associate with the tail tip *in vitro*, further suggesting this protein functions during the early stages of infection (Kenny *et al*., 2004). Tal_{2009} has sequence identity with structural proteins from other lactococcal phages, including TP901-1 and Ul36 and putative structural proteins from the prophage bIL285, suggesting that these viruses may also utilize similar infection strategies.

The terminal protein gp3 of the short-tailed *B. subtilis* phage, ϕ29, was also found to have murein hydrolase activity (Moak & Molineux, 2004). This protein functions in DNA replication and packaging and is covalently bound to the ends of the linear viral genome (Salas *et al.*, 1978, Mellado *et al.*, 1980). However, the DNA priming, binding and packaging motifs are in a separate region of the protein than the putative transglycosylase motif (Moak & Molineux, 2004). This protein is only active when it is not bound to the phage DNA, suggesting that free protein is packaged in the phage capsid and functions to degrade peptidoglycan during DNA injection. Two other short-tailed *Bacillus* phages that are similar to ϕ29, phages M2 and GA-1, are each predicted to encode a terminal protein, and these appear to possess muralytic activity *in vitro* as well (Moak & Molineux, 2004).

Notably, murein hydrolase activity has also been detected *in vitro* for a number of other phages of the Gram-positive bacteria, including the *L. lactis* phages ϕ949, ϕ31, ϕc2, ϕr1t and ϕsk1;

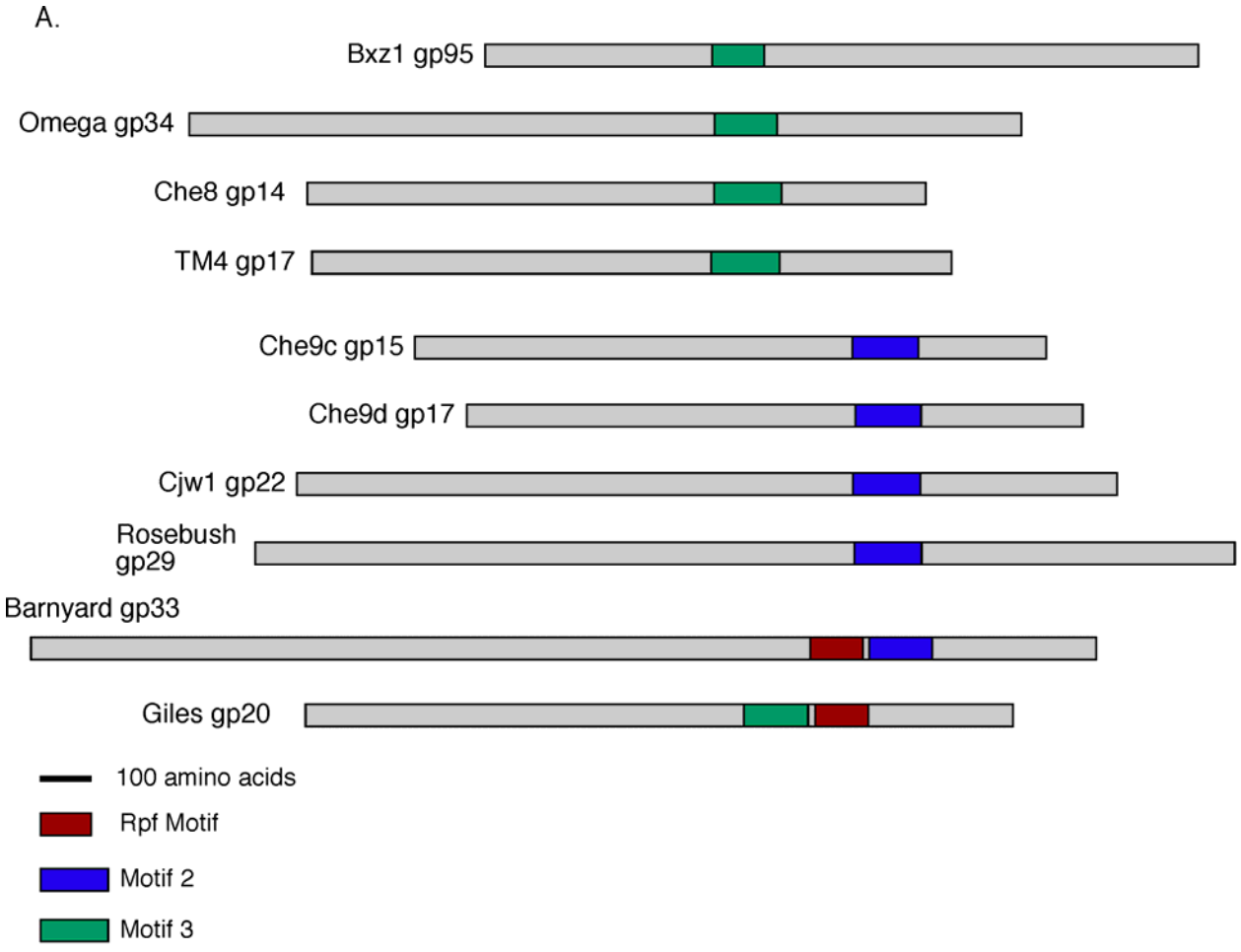
the *S. aureus* phages ϕ11 and ϕ85; and the *B. subtilis* phages SPO1, SPP1 and SPβ (Moak & Molineux, 2004). In all cases, the activity was detected in zymogram gels (polyacrylamide gels embedded with peptidoglycan); however, it is not known whether the muralytic activity of any of these proteins is important for infectivity of the respective viruses that encode them. Further, putative phage structural proteins containing predicted transglycosylase motifs have been identified by bioinformatic analysis in the *Streptococcus* phages Sfi21 and DT-1, as well as in the *Lactocbacillus* phages LL-H, ϕadh and ϕg1e, although enzymatic activity has not yet been demonstrated (Crutz-Le Coq *et al.*, 2002). Thus, while the importance of virion-associated hydrolase proteins is only in the earliest stages of investigation, a pattern is emerging which suggests that this strategy may be widely employed by phages to assist in host cell penetration.

1.4 MYCOBACTERIOPHAGE TAPE MEASURE MOTIFS

Until recently, it was unclear if mycobacteriophages utilized strategies similar to those observed in other phages to assist in the transport of their genetic material across the complex and impermeable cell walls of their hosts, the mycobacteria. The first clue as to a potential solution to this problem came in 2003 with the discovery that many of these phages encode tape measures having regions of similarity to putatively secreted bacterial proteins (Pedulla *et al*., 2003). These regions are small (70-130 amino acids) in relation to the total size of the tape measures (1000-2000 amino acids), and most are situated in approximately the same location, lying in the C-terminal one-third of the protein (Fig. 4A). Three of these tape measure motifs were identified, based on sequence identity, and at least one appeared to be present in eight of the 14 mycobacteriophages completely sequenced at that time (Pedulla *et al*., 2003). Further, these regions were often observed to be

shared by TMPs that otherwise possessed very little sequence identity outside of the motifs. These observations therefore suggested that there was likely a positive selection for the acquisition and maintenance of these sequence motifs within mycobacteriophage tape measures.

The specific function of the tape measure motifs, however, was not initially obvious. At the time of their discovery, the only one that had similarity to a host protein with a known function was Motif 1. This encompasses an ~70 amino acid region and has sequence identity (typically between 55-70%) with a group of proteins known as resuscitation promoting factors (Rpfs; Fig. 4B); thus, this region is often referred to as the Rpf Motif. The Rpf proteins were initially characterized for their apparent ability to resuscitate bacterial cells that had become dormant and, as such, had lost the ability to form colonies (discussed in Section 1.5) (Mukamolova *et al*., 1998). Therefore, it was hypothesized that the phage motifs might be performing the same function as the bacterial Rpfs in order to 'wake-up' dormant hosts prior to infection (Pedulla *et al*., 2003). Of the first 14 mycobacteriophages sequenced, this motif was found only in the Barnyard TMP, but has subsequently also been identified in mycobacteriophage Giles; in both cases it is also found in combination with one of the other motifs (Pedulla *et al*., 2003, Morris *et al*., 2008) (Fig. 4A).



B.

```
M.luteus      ............AAAVGFSAPAQAATVD..............TWDRLAECESNGTWDINTGNGFYGGVQFTLSSWQAVG: 82
Rv2389c_RpfD_ ..........MFVALLGLSTISSKADDI..............DWDAIAQCESGGNWAANTGNGLYGGLQISQATWDSNG: 89
Rv0867c_RpfA_ ............GGGIAMAAQATAATDG..............EWDQVARCESGGNWSINTGNGYLGGLQFTQSTWAAHG: 79
RV1884C_RPFC_ ............ASMSLSTAVAHAGPSP..............NWDAVAQCESGGNWAANTGNGKYGGLQFKPATWAAFG:108
Rv1009_RpfB_  AEVNGVETGRLPVANVVVTPAHEAVVRVGTKPGTEVPPVIDGSIWDAIAGCEAGGNWAINTGNGYYGGVQFDQGTWEANG:320
Rv2450c_RpfE_ ..........PPLAPDFLSPPAEEAPPV.........PVAYSVNWDAIAQCESGGNWSINTGNGYYGGLRFTAGTWRANG:136
Barnyard      ............IEGAVARGTEAGMKAG..............ADWDAIAQKESGGNWAINTGNGYYGGLQFAQSSWEAAG:101
Giles         ............LGGPYGAGLNTGAATQT.............VPDWDAIAQAESGGNWAINTGNGYYGGLQFDQPTWDAYK:101

M.luteus      G...EGYPHQASKAEQIKRAEILQDLQG...WGAWPLCSQKL................GLTQADADAGDVDATEAAPVAV:140
Rv2389c_RpfD_ G...VGSPAAASPQQQIEVADNIMKTQG...PGAWPKCSS...................CSQGDAPLGSLTHILTFLAAE:144
Rv0867c_RpfA_ GGEFAPSAQLASREQQIAVGERVLATQG...RGAWPVCGRGLSNATPREVLPASAAMDAPLDAAAVNGEPAPLAPPPADP:156
RV1884C_RPFC_ G...VGNPAAASREQQIAVANRVLAEQG...LDAWPTCG.......................AASGLPIALWSKPAQG:157
Rv1009_RpfB_  GLRYAPRADLATREEQIAVAEVTRLRQG...WGAWPVCA.......................ARAGAR..........:362
Rv2450c_RpfE_ GSGSAAN...ASREEQIRVAENVLRSQG...IRAWPVCG.......................RRG.............:172
Barnyard      GLAYASRADLASKEQQIAAAEELLKQQG...PGAWPNTFVAARPG...............ATTPGTGGRTRGEISDEDL:162
Giles         PAGAPARADMAPRETQIAAAENLVRDRG.ANAPKAWPNTWKTKDVP..............ATGLSTGTRTAG.IDPETG:164
```

Figure 4. Sequence motifs found in mycobacteriophage TMPs.

Figure 4: Mycobacteriophage tape measure motifs. **A.** Schematic illustrating the three sequence motifs that have been identified in mycobacteriophage tape measures. Motif 1 (the Rpf Motif), found in mycobacteriophages Barnyard and Giles, has sequence identity to a family of bacterial proteins that have growth stimulatory activity, as well as the ability to degrade cell wall peptidolgycan, and this is highlighted in red. Motif 2 has sequence similarity to the *M. tuberculosis* protein Rv1115, a putative secreted protein of unknown function, and is colored in blue. Motif 3, highlighted in green, has similarity to the *M. tuberculosis* proteins Rv0320 and Rv1728c, as well as to a putative protease from *Rhodococcus equi*. Proteins containing the Motif 3 domain also have muralytic activity, and TM4 phages lacking this motif are defective in infecting stationary phase cells of *M. smegmatis*. **B.** ClustalX sequence alignment of the Rpf Motifs from Barnyard and Giles TMPs with the *Micrococcus luteus* Rpf and the five homologues in *Mycobacterium tuberculosis* illustrates a high degree of identity across the conserved region. 100% conserved residues are highlighted in red, identical residues are highlighted in green, and similar residues are highlighted in blue.

It was initially unclear if Motifs 2 and 3 might have similar host-signaling functions due to the fact that the host proteins that they resembled were uncharacterized. Motif 2 was found in mycobacteriophages Che9c, Che9d, Cjw1, Rosebush and Barnyard (Fig. 4A) and has similarity to a small *M. tuberculosis* protein of unknown function, Rv1115, which is predicted to be secreted (Pedulla *et al*., 2003). A deletion of the Rv1115 homologue in *M. smegmatis* or overexpression of a stable version of this protein (minus a C-terminal region that promotes degradation) causes no obvious phenotypic effects, and the precise function of this protein remains unknown (A. Ojha, unpublished observations). Motif 3 was detected in mycobacteriophages TM4, Che8 and Omega and also appeared to be present within the tape measure protein of the contractile-tailed phage Bxz1, although in this latter case, the motif is located towards the N-terminus of the protein (Fig. 4B). The Motif 3 region was found to resemble two small, likely secreted proteins of unknown function in *M. tuberculosis*, Rv0320 and Rv1728c, as well as a putative protease from *Rhodococcus equi* (Pedulla *et al*., 2003).

Although the identity of Motif 1 with the Rpf family of proteins suggested these might be involved in host growth stimulation, a detailed investigation of the mycobacteriophage TM4 Motif 3 and a characterization of the host proteins to which it is similar have provided evidence that suggests another role for these tape measure motifs. This was prompted by bioinformatic analyses, which detected Motif 3 in a number of bacterial proteins from a variety of different species, many of which were predicted to have murein hydrolase activity or to function as metalloendopeptidases (Piuri & Hatfull, 2006). Piuri and Hatfull further demonstrated that the *M. smegmatis* homologues of Rv0320 and Rv1728c, MSMEG0642 and MSMEG3721, respectively, have murein hydrolase activity *in vitro*, and hybrid proteins in which the Motif 3-like region of MSMEG3721 is replaced with the phage TM4 Motif 3 are also active. Additionally, phage lacking Motif 3 or containing a

mutation in the putative active site tryptophan show a decreased plating efficiency on stationary phase cells as compared to exponentially growing *M. smegmatis*. This appears to be due specifically to a defect in DNA injection, not adsorption. It was hypothesized that the tape measure motif, although not essential for phage viability, has a murein hydrolase activity that promotes efficient infection of stationary phase cells, which may have altered cell walls.

These data strongly suggest a new role for bacteriophage TMPs, which have long been known to be essential for tail morphogenesis and tail length determination. The apparent murein hydrolase activity of Motif 3 in TM4, as well as its importance for infection of stationary phase cells, further suggests that this function might be shared by the other mycobacteriophage TMP motifs as well. Little is known about the Motif 2 homologues in bacteria; however at least one, the Rv1115 homologue in *M. tuberculosis* strain C (MtubC_01000897) has been annotated as a cell wall hydrolase, although the evidence for this is unclear. Importantly, the current knowledge of the function of the bacterial Rpf proteins also supports this hypothesis, and this is discussed in the following section.

1.5 BACTERIAL RPF PROTEINS

As stated above, resuscitation promoting factors or Rpfs were initially named for their ability to 'resuscitate' dormant bacteria: non-growing cells which have entered into a reversible state of low metabolic activity as a response to starvation or other conditions unfavorable for growth (Kaprelyants *et al.*, 1993). This dormancy response has been likened to the process of endospore formation, whereby some organisms cease to grow and form highly resistant spores when subjected to nutrient deprivation and stress and can only resume growth when they detect specific

environmental signals (Keep *et al.*, 2006). In the same way, it has been proposed that dormancy increases chances for stationary phase survival under harsh environmental conditions, and although dormant cells retain viability, they require an exogenous stimulus for the resumption of growth (Mukamolova *et al.*, 2003).

1.5.1 The *Micrococcus luteus* Rpf Protein

Members of the Rpf family of proteins are widespread in the high G+C% Gram-positive actinobacteria, although only a few of these have been investigated in any detail (Kell & Young, 2000, Ravagnani *et al.*, 2005, Schroeckh & Martin, 2006). Organisms in which they are found commonly encode multiple Rpf-like proteins; however, the founding member of this family is the single Rpf from *Micrococcus luteus*, a small (~17 kDa) protein that was isolated from the supernatant of exponentially growing cells (Mukamolova *et al*., 1998). This was discovered based on its apparent ability to stimulate the growth of cells that had become dormant, a state that, for *M. luteus*, could be achieved by incubating this organism for prolonged periods of time in spent medium in stationary phase (Kaprelyants & Kell, 1993, Mukamolova *et al*., 1998). In a dormant population, approximately one out of 10^4 observable cells has the ability to form colonies on solid media, but this can be increased several orders of magnitude by adding back Rpf-containing supernatant from growing cells (Kaprelyants *et al.*, 1994, Votyakova *et al.*, 1994, Kaprelyants *et al.*, 1996). Picomolar concentrations of recombinant *M. luteus* Rpf were further shown to decrease the lag phase of viable cells inoculated into fresh media and to be required for growth of extensively washed cells in minimal media, a condition thought to mimic cells in stationary phase that may have become depleted for essential 'growth factors'(Mukamolova *et al*., 1998). The *rpf* gene is essential in *M. luteus,* and expression declines during stationary phase (Mukamolova *et al*.,

1998, Mukamolova *et al.*, 2002a). These characteristics, combined with its requirement for growth under specific conditions and activity at low concentrations, lead to the description of this protein as the first example of a bacterial growth factor, or cytokine. As such, it was proposed that Rpf might interact with extracellular receptors in order to modulate bacterial growth, a model consistent with observed localization of this protein at the cell surface and the fact that the C-terminus of Rpf contains a domain predicted to be involved in cell wall binding (LysM domain) (Bateman & Bycroft, 2000).

1.5.2 Rpf Proteins in the Mycobacteria

Homologues of the Rpf protein are found in all sequenced mycobacterial species, and of these, the five family members from *M. tuberculosis* are the best characterized. Each contains an ~70 amino acid Rpf Motif, and all are predicted to be either membrane anchored or secreted (Fig. 5); in the case of RpfE (Rv2450c) secretion has actually been verified, since a RpfE-PhoA fusion is secreted *in vivo* (Gomez *et al.*, 2000). Unlike the *M. luteus* Rpf, none of the TB Rpf proteins have a LysM-like cell wall binding domain. However, most do have additional domains, such the proline/alanine-rich segment found in RpfA and RpfE or the putative membrane lipoprotein lipid anchor in RpfB (Fig. 5) (Mukamolova *et al.*, 2002b). It has been reported that purified recombinant forms of these proteins have growth-stimulating activity against both the slow- and fast-growing mycobacteria, as well as towards *M. luteus* (Mukamolova *et al*., 2002b). This is consistent with the observation that the *M. luteus* Rpf has cross-species growth-promoting activity towards the mycobacteria and other related organisms, although the conditions required to achieve dormancy vary considerably from species to species (Mukamolova *et al.*, 1998, Shleeva *et al.*, 2002, Mukamolova *et al.*, 2002b, Shleeva *et al.*, 2004).

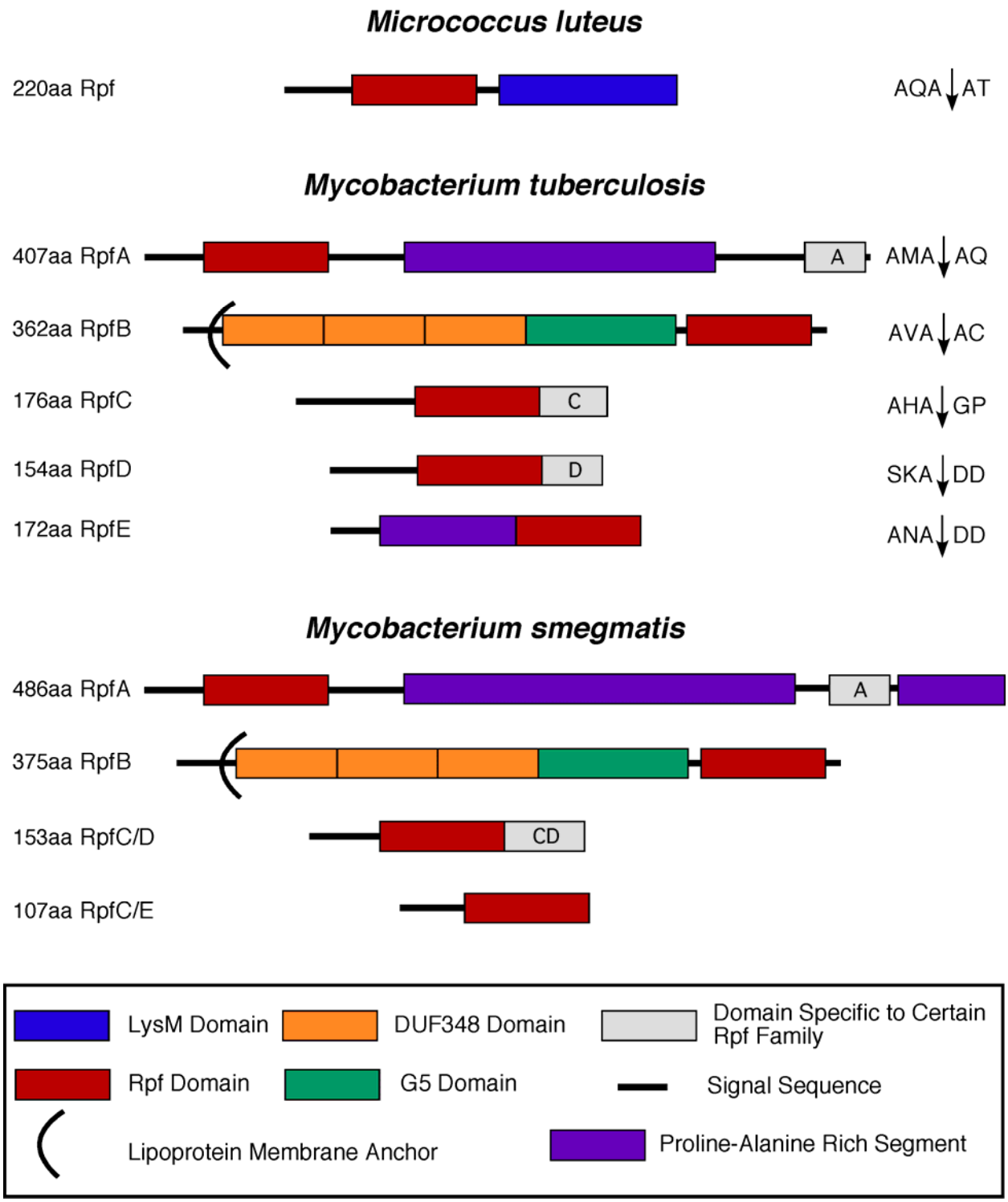


Figure 5. The Rpf proteins of *M. tuberculosis* and *M. smegmatis.*

Figure 5: The Rpf proteins in *M. tuberculosis* and *M. smegmatis*. Unlike the *M. luteus* Rpf, the mycobacterial Rpfs do not contain the LysM cell wall binding domain. However, these contain a variety of other predicted domains, including proline/alanine rich segments, a putative lipoprotein membrane attachment site in the RpfB proteins, and a number of domains whose function is not known, such as the G5 domain that is commonly found in cell wall modifying enzymes. These proteins are all predicted to be either secreted or membrane anchored, and the signal peptide cleavage sites are indicated where known (adapted from Ravagnani *et al*., 2005 and Mukamolova *et al*., 2002b).

Expression of the mycobacterial Rpf proteins has been verified in actively growing cells of both *M. tuberculosis* and *Mycobacterium bovis* BCG. In *M. tuberculosis,* they have slightly different, overlapping expression profiles, but mRNA levels of all five generally diminish as the cells transition from exponential to stationary phase (Mukamolova *et al*., 2002b, Tufariello *et al.*, 2004). Interestingly, however, expression of all five was detected to varying levels in a four-month old culture of *M. tuberculosis,* suggesting regulation that is much more complex than that which is observed with the single Rpf in *M. luteus* (Tufariello *et al*., 2004). This contrasts earlier reports that the five Rpf transcripts were undetectable in late-stationary phase or starved cells of *M. bovis* BCG (Mukamolova *et al.*, 2002b), a discrepancy which could reflect differences between the pathogenic and non-pathogenic mycobacteria or conversely, may be due to experimental variation. Antibodies raised against the Rpf domain weakly detect protein bands – although their identity was not confirmed – in concentrated supernatant from *M. bovis* BCG and *M. smegmatis*, and similar to *M. luteus* Rpf, these were localized at the cell surface by immunofluorescence microscopy (Mukamolova *et al*., 2002b)

The expression of the *M. tuberculosis* Rpf proteins has also been investigated in various animal models of infection, and the patterns observed *in vivo* are similar to what has been observed in culture. Transcripts from all five proteins are present during the initial stages of murine infection, and transcripts from four (*rpfA, B, D* and *E)* are detectable later during chronic, persistent infection (Tufariello *et al*., 2004, Tufariello *et al.*, 2006). Notably, expression levels remain relatively consistent early, following reactivation of disease after a period of latency but decrease at later time points when bacterial levels in the tissues are at a maximum (Tufariello *et al*., 2006). Further evidence of Rpf secretion *in vivo* is derived from the observation that vaccination of mice with RpfE results in decreased bacterial loads and increased survival when

these animals are subsequently infected with *M. tuberculosis* (Yeremeev *et al.*, 2003). Interestingly, there is also evidence that the Rpf proteins are expressed in human lungs infected with pathogenic *M. tuberculosis* (Fenhalls *et al.*, 2002, Rachman *et al.*, 2006).

The ability of the Rpf proteins to stimulate the growth of non-culturable bacteria suggested the intriguing possibility that these proteins might be important for the pathogenesis of *M. tuberculosis*, an organism that can persist in the lungs of infected individuals for many years in a latent, non-transmissible state (Manabe & Bishai, 2000, Flynn & Chan, 2001). Elucidating the function of the mycobacterial Rpf proteins, however, has proved to be a challenging and often confusing endeavor. Although single *rpf* deletion mutants have slightly altered colony morphology under some conditions, in general, they are phenotypically similar to wild-type in culture, as well as *in vivo,* with similar levels of growth and persistence in mice (Downing *et al.*, 2004, Tufariello *et al.*, 2004). Uniquely, a *M. tuberculosis rpfB* (Rv1009) deletion mutant is significantly defective in a mouse model for disease reactivation, resulting in prolonged host survival, which suggests a particular role for this protein at this stage of the disease (Tufariello *et al.*, 2006). Triple deletions – Δ*rpfA* Δ*rpfC* Δ*rpfB* or Δ*rpfA* Δ*rpfC* Δ*rpfD* – are attenuated in an intravenous mouse infection model, as evidenced by decreased bacterial loads in the organs of infected animals, and these mutants are also defective in regrowth after prolonged stationary phase and oxygen depletion in culture (Downing *et al.*, 2005). Quadruple deletions have the same resuscitation defect and are similarly attenuated in mice, as measured by bacterial pathogenesis and host mortality after aerosol infection (Kana *et al.*, 2008). The deficiency in regrowth after prolonged stationary phase can be alleviated in both triple and quadruple mutant backgrounds by genetic complementation or by the addition of supernatant from actively growing cells (Kana *et al.*, 2008). Despite anecdotal reports to the contrary, it has been possible to create *M. tuberculosis* mutants that are deleted for all five

rpf genes (Kana *et al.*, 2008). Further, the *rpf* genes could be added back individually to this strain in order to assess a "hierarchy of function" *in vitro*. Extensive phenotypic analysis of multiple combinations of mutants and various complemented strains suggests that RpfB and RpfE are most important for the phenotypes observed in the *rpf* mutants both *in vitro* (see Section 1.5.4) and *in vivo,* although for unknown reasons, the complementation of quadruple mutants was not possible in mice (Downing *et al.*, 2005, Kana *et al.*, 2008).

These data imply that the Rpf proteins, although non-essential, play somewhat functionally redundant roles in optimizing growth and proliferation under specific conditions and environmental stresses. The observation that the loss of any one of the five results in increased levels of expression of the other *rpf* genes (except *rpfA*) would seem to support this (Downing *et al*., 2004). However, it was further found that this effect on expression is diminished as additional *rpf* mutations are introduced, and some of these genes, such as *rpfB* and *rpfE*, appear to be upregulated in double and triple mutants but downregulated in quadruple deletions of the other four (Kana *et al.*, 2008), an observation suggestive of differential regulation.

The five Rpf proteins are distributed throughout the *M. tuberculosis* (and *M. smegmatis*) chromosome, and the available data suggest that they are differentially regulated, at least to some degree. It has been determined that *rpfA* expression is controlled by the transcription factor Rv3676, a cAMP-receptor-like protein involved in the starvation response (Rickman *et al.*, 2005). There is further evidence that *rpfC* is upregulated by the non-essential alternative sigma factor, SigD, which is expressed during exponential growth but is also induced by starvation conditions (Raman *et al.*, 2004). Conversely, *rpfC* also appears to be negatively regulated by SigE, which is induced by heat-shock, sodium dodecyl sulfate (SDS) treatment and inside human macrophages (Manganelli *et al.*, 2001). The situation is further complicated by the fact that *rpfE* expression

increases following acid-shock (Fisher *et al.*, 2002), and both *rpfC* and *rpfE* are induced during the transition to microaerophilic conditions (Sherman *et al.*, 2001). These data suggest that the *M. tuberculosis* Rpfs may have overlapping but distinctive functions and are perhaps indicative of a general role in stress response; however, this possibility requires further investigation.

1.5.3 Rpf Mechanism of Action

The characterization of Rpf proteins as bacterial growth factors led to the proposal that as such, they were likely to function by binding to receptors on the bacterial cell surface and transmitting their growth signal through the activation of second messenger cascades (Kell & Young, 2000). Rpf proteins were indeed detected on the surface of growing cells of both *M. luteus* and *M. bovis* BCG by immunofluorescence microscopy, and anti-Rpf antibodies could inhibit the growth of both of these organisms, a result that might be expected if these were blocking an essential Rpf-receptor interaction (Mukamolova *et al*., 2002a, Mukamolova *et al*., 2002b). However, despite substantial searching, no Rpf receptors were identified. Subsequent bioinformatic analyses and structural predictions revealed that the conserved Rpf domain has weak similarity to, and may adopt structural conformations resembling, C-type lysozymes (CAZy family:GH22) (Cohen-Gonsaud *et al.*, 2004). The nuclear magnetic resonance (NMR) solution structure of the Rpf domain (residues 280-362) from *M. tuberculosis* RpfB substantiated these observations by demonstrating that the Rpf domain does indeed structurally resemble the lysozyme-like fold of muramidase enzymes. The domain is most similar to C-type lysozymes in regard to conformation of the active site and placement of critical residues, such as the catalytic glutamate. A portion of this domain, however, more closely resembles the active site of the soluble lytic transglycosylases (CAZy family:GH23), particularly Slt70, and therefore, the Rpf domain appears to be a hybrid, although more compact,

of its two closest structural matches. Importantly, the predicted oligosaccharide-binding cleft could bind to N,N′,N′′-triacetylchitotriose (tri-NAG), a polymer of GlcNAc, although the motif itself possessed no detectable lysozyme-like activity (Cohen-Gonsaud *et al*., 2005).

It has further been confirmed that the full-length *M. luteus* Rpf protein also has functional similarity to lysozyme and lytic transglycosylase enzymes both *in vivo* and *in vitro.* This was demonstrated by the ability of Rpf to degrade cell wall material in zymograms, release fluorescein from labeled cell walls, cleave an artificial lysozyme substrate and lyse *E. coli* when secreted into the periplasm of these bacteria (Mukamolova *et al.*, 2006). Mutations in the predicted active site glutamate decreased, but did not abolish activity (see Chapter 5), and the murein hydrolase activity was shown to be important for the resuscitation activity of this protein (Mukamolova *et al*., 2006). This suggests that the growth enhancing ability of Rpf is due, in part, to its muralytic activity.

Although the ability of Rpf to degrade cell wall peptidoglycan is correlated with its physiological growth stimulatory activity, it is unclear how the two are connected. It has been speculated that *M. luteus* cells in stationary phase may have altered cell walls with increased levels of cross-linking, as has been observed for other bacteria (Pisabarro *et al.*, 1985, Atrih *et al.*, 1999). In this case, the murein hydrolase activity of Rpf may be needed to modify the peptidoglycan in a such a way as to allow the cells to resume growth and cell wall synthesis after a dormant state. Dormant *M. luteus* do appear to have thicker cell walls, but these have not been extensively characterized (Mukamolova *et al.*, 1999). Rpf-mediated cleavage might also be required for an essential step of cell division such as septation, as proteins with similar activities function during cell division in spore germination (Atrih & Foster, 1999, Vollmer *et al.*, 2008b).

Another possibility is that the action of the Rpf proteins may release muropeptide degradation products that act as signaling molecules in a manner similar to the way in which the

products of lytic transglycosylases activate expression of β-lactamases in *E. coli* (Jacobs *et al.*, 1997). In support of this, it was previously reported that growth of non-culturable *M. tuberculosis* cells could be stimulated by the addition of phospholipids or peptides, species that potentially could be released by the action of a cell wall hydrolase (Zhang *et al.*, 2001). Conversely, Rpf activity may allow other compounds in the media to penetrate the cell or might mediate the release of signaling compounds or nutrients from lysed cells in a population that promote the growth of neighboring dormant cells (Mukamolova *et al*., 2006). This may be unlikely, however, as it was also reported that trace amounts of lysozyme do not mimic Rpf activity *in vivo*.

1.5.4 Evidence for Rpf-Interacting Proteins

There is an increasing body of evidence that suggests the activity of Rpf-like lytic transglycosylases might be closely linked to that of peptidoglycan peptidases. Several Rpf proteins have been identified in *Streptomyces coelicolor* and *Streptomyces avermitilis*, and both organisms contain one family member that is predicted to be expressed as a fusion with a putative M23 Gly-Gly peptidase domain (Keep *et al*., 2006). Further, the IsaA protein of *S. aureus,* which appears to be distantly related to the Rpfs, is co-regulated with the glycyl-glycine peptidase LytM, and RpfC from *M. tuberculosis* is co-regulated with a putative secreted protease Rv1815 (Dubrac & Msadek, 2004, Raman *et al.*, 2004).

Significantly, it has recently been shown that RpfB from *M. tuberculosis* physically interacts with a putative mycobacterial endopeptidase that has been named Rpf-interacting protein A (RipA) (Hett *et al.*, 2007). Interactions occurred *in vitro* and *in vivo*, and these proteins were observed to co-localize at the septum of dividing cells. Similar interactions were noted between RipA and RpfE, but not with any of the other Rpf proteins. Interestingly, *M. tuberculosis*

quadruple mutants lacking functional copies of *rpfB* or *rpfE* (and derivatives lacking all five), display delayed colony formation on mycobacterial media containing increased amounts of malachite green and show increased sensitivity to SDS (Kana *et al*., 2008). These phenotypes are indicative of a cell wall defect and lend support to a role for these proteins – possibly in conjunction with RipA – in cell wall modification. The data also strongly suggest that at least some of the Rpf proteins may form complexes with endopeptidases in order to hydrolyze peptidoglycan during cell division. As these proteins are collectively non-essential in culture, the likelihood remains that other proteins can compensate for their loss, at least under favorable conditions. Culture conditions such as nutrient deprivation may stress the cell wall to such an extent that the activity of the Rpf proteins becomes essential.

1.5.5 Distribution of Rpf-Like Proteins

Rpf-like proteins have been identified in the sequenced genomes of many other species of actinobacteria, including species of *Mycobacterium*, as well as *Corynebacterium, Nocardia* and *Streptomyces spp*., all of which contain the conserved ~70 amino acid Rpf domain (Kell & Young, 2000, Ravagnani *et al.*, 2005). Many of these proteins also contain putative cell wall binding domains, secretory signal sequences or regions with putative transmembrane helices (Table 1) (Ravagnani *et al*., 2005). The Rpf proteins have been grouped into ten subfamilies based on which of these additional domains are present and how they are arranged. Each of the five Rpf proteins in *M. tuberculosis* represents one of these families, and the LysM containing *M. luteus* Rpf represents another; some of these are illustrated in Figure 6 (Ravagnani *et al*., 2005). Interestingly, only actinomycetes whose cell walls are not decorated with mycolic acids encode LysM-containing Rpfs. Two of these families are comprised of proteins from firmicute bacteria – the low G+C%

Gram-positives – that contain more divergent Rpf domains (not shown). There are also likely to be Rpf-like proteins present in many more related organisms, and this is supported by the observation that degenerate primers to the Rpf domain were also able to detect *rpf* genes in numerous unsequenced species of high G+C% actinobacteria, representing species of *Arthrobacter*, *Brevibacterium* and *Rhodococcus* (Schroeckh & Martin, 2006). Protein expression was detected in many cases, and Rpf-containing supernatant from some was active in resuscitating *M. luteus.*

Table 1. Rpf-domain containing proteins in the *Corynebacterineae.*

Organism	Designation	Size-aa	Additional Domains	Activity Reported
M. luteus	Rpf	223	LysM - cell wall binding domain	Yes
M. tuberculosis	RpfA (Rv0867c)	407	Pro/Ala Domain , Rp fA specific domain	Yes
M. tuberculosis	RpfB (Rv1009)	362	Membrane lipoprotein lipid anchor, DUF348 & G5 domains	Yes
M. tuberculosis	RpfC (Rv1884c)	176	RpfC specific domain	Yes
M. tuberculosis	RpfD (Rv2389c)	154	RpfD specific domain	Yes
M. tuberculosis	RpfE (Rv2450c)	172	Pro/Ala domain	Yes
M. smegmatis	RpfA (MSMEG_5700)	486	RpfA specific domain	ND
M. smegmatis	RpfB (MSMEG_5439)	375	DUF348 & G5 domains	ND
M. smegmatis	RpfC/D (MSMEG_4640)	153	RpfC/D specific domain	ND
M. smegmatis	RpfC/E (MSMEG_4643)	107		ND
C. glutamicum	Rpf1 (NCGL0785)	193	*Corynebacterium* specific domain	Yes
C. glutamicum	Rpf2 (NCGL0872)	374	DUF348 & G5 domains, protein is glycosylated	Yes
S. coelicolor	SCO3097	244	LysM	ND
S. coelicolor	SCO3098	341	LysM, Ser/Gly rich domain	ND
S. coelicolor	SCO0974	439	LysM, M23 peptidase domain	ND

Not Determined.
Consensus sequence: APADLAPP.
Double deletion ha s weak phenotype.

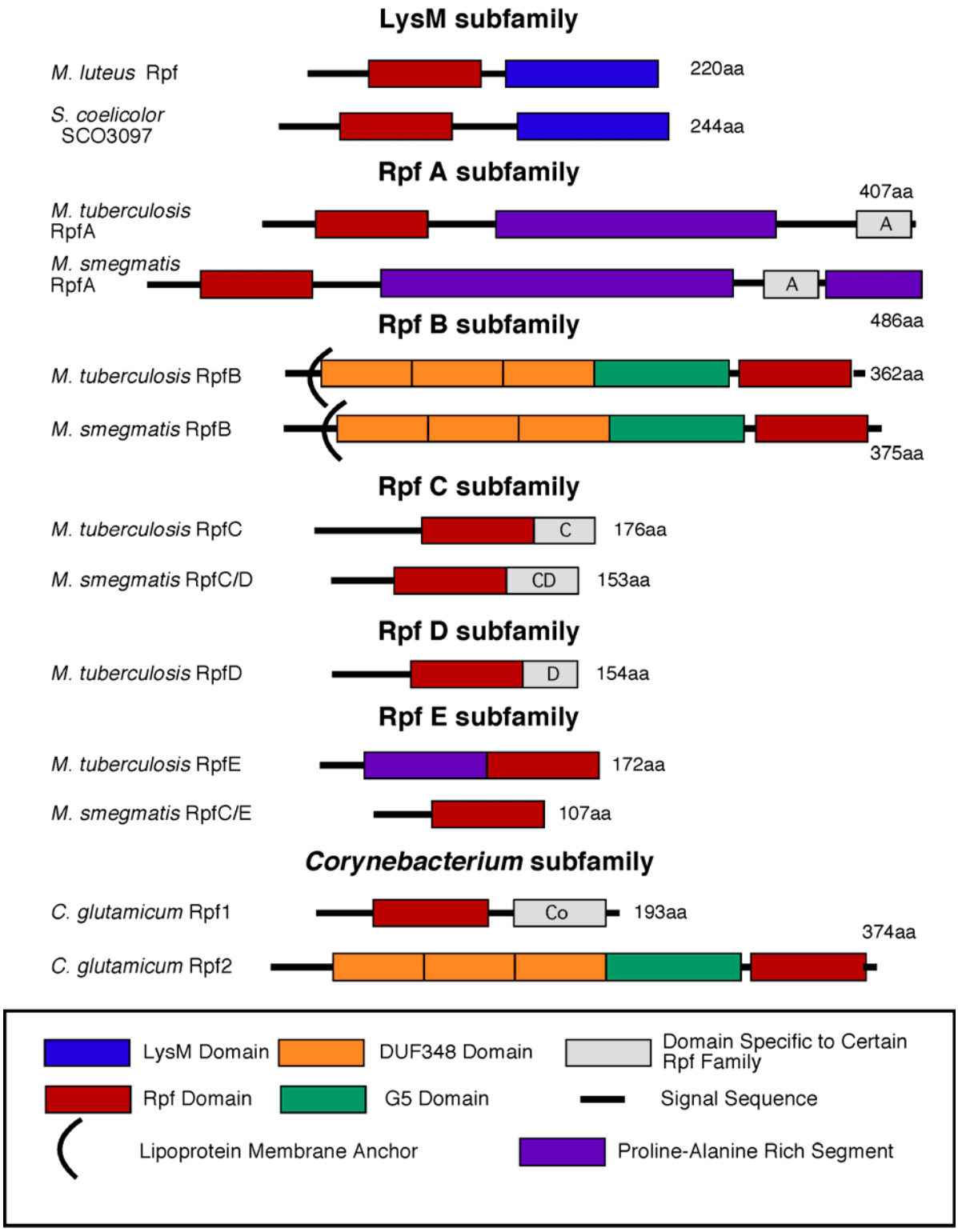


Figure 6. Subfamilies of the Rpf family of proteins.

Figure 6: Subfamilies of the Rpf family of Proteins, adapted from Ravagnani *et al*., 2005.

Most Rpf family members have not been investigated in detail; however, the two Rpf proteins in *C. glutamicum* have been deleted, and although double mutants grow normally, they have an increased lag phase and doubling time after prolonged incubation in stationary phase (Hartmann *et al.*, 2004). One of these proteins, designated Rpf2, has a number of similarities with the RpfB proteins from the mycobacteria (Fig. 6). Further, it has been detected at the cell surface and appears to be glycosylated *in vivo*. A protein more distantly related to the Rpf family members is SceD from *S. aureus* (Ravagnani *et al*., 2005). This and a related protein, IsaA, have been characterized as lytic transglycosylases, which affect cell clumping and salt sensitivity in culture and are important for virulence of *S. aureus in vivo* (Stapleton *et al.*, 2007).

Firmicute bacteria such as *Bacillus* and *Clostriduim spp.* also contain an analogous family of proteins that have been termed stationary phase survival (SPS) proteins, and like the Rpf family members, these often contain secretory signals and additional domains predicted to be involved in membrane anchoring and cell wall binding (Ravagnani *et al*., 2005). The conserved SPS domain has similarity to the *E. coli* lytic transglycosylase MltA and is not related to Rpf domain. However, these proteins were identified because members of one SPS subfamily contain regions with weak sequence identity to conserved domains found in the RpfB proteins. Further, most of these SpsB proteins and the RpfB proteins are located in similar genomic contexts. When mutants of *B. subtilis* deleted for one or more of its *sps* genes are kept in prolonged stationary phase, they display decreased levels of survival as compared to wild-type (Ravagnani *et al*., 2005). Thus, these proteins appear to be involved in bacterial growth regulation and to function somewhat similarly to the Rpf family members.

1.6 SUMMARY

There is good evidence that bacterial proteins related to both the Rpf Motif and Motif 3 have murein hydrolase activity, and at least for Motif 3, the phage motif can substitute for the corresponding domain from its bacterial homologue *in vitro* (Piuri & Hatfull, 2006). Further, the prevalence of muralytic proteins associated with bacteriophage virions has been well documented. These data strongly suggest that mycobacteriophages may have co-opted proteins involved in cell wall metabolism and incorporated them into the TMP to perform localized peptidoglycan hydrolysis during host infection. While the structural role of the TMP has been well characterized in a number of phages, the presence of sequence motifs within these proteins that are involved in cell wall hydrolysis implies a novel function for the tape measure.

Because the TMP is located inside the phage tail, it is in the direct path that the DNA must follow during injection and therefore, must be ejected prior to, or concurrent with, host penetration and viral genome expulsion. Direct evidence for this with most phages has not been obtained; however, there are data suggesting for the coliphage λ that this is indeed the case (Scandella & Arber, 1976, Roessner & Ihler, 1984). Further, the T5 tail spike protein pb2 shares a number of structural features with TMPs, and this protein is known to insert into the bacterial membrane during infection (Guihard *et al*., 1992, Boulanger *et al*., 2008). Therefore, it may be generally true that the bacteriophage TMP contacts and/or enters the host cell during infection, making it a protein ideally suited for host interaction.

Although the evidence strongly suggests a role for the TMP Rpf Motif in murein hydrolysis, a definitive function has not been demonstrated. Therefore, in order to better understand the role of this motif, I have conducted a number of experiments aimed at determining the physical state of the TMP embedded Rpf Motif, the fate of this domain and the entire TMP

during host infection, and whether the Rpf domain possesses an enzymatic activity that is essential or beneficial to the phage. I have also undertaken a global characterization of the mycobacteriophage tape measures in order to better understand these unique proteins, as well as to re-examine the prevalence of the various sequence motifs.

In this study, I present evidence that the mycobacteriophage tape measures share a number of features with one another, as well as with previously characterized TMPs. Further, the tape measure sequence motifs appear to be widespread, not only in this group of phages, but also in phages that infect other Gram-positive bacteria. I have demonstrated that the Barnyard TMP is proteolytically processed in a unique way during or prior to phage assembly, which results in the Rpf Motif (as well as Motif 2) being present on a 70 kDa cleavage product in the mature phage that may be cell wall or membrane associated during infection. I have also shown that both phage Rpf Motifs can substitute for the Rpf domain from the *Micrococcus* Rpf, and these have muralytic activity *in vivo*. Finally, although this motif is not essential, as it can be deleted from mycobacteriophage Giles, phages lacking this motif, as well as those lacking Motif 3, are somewhat defective in infecting stationary phase cells. These data support a role for the mycobacteriophage Rpf Motif in cell wall hydrolysis during host infection. Additionally, in the process of creating the tape measure mutants, I have adapted the mycobacterial recombineering system developed by van Kessel and Hatfull (2007) for use on mycobacteriophages. This system will, for the first time, allow the directed manipulation of mycobacteriophage genomes in an attempt to better understand the immense diversity that is found within this population.

2.0 MATERIALS AND METHODS

2.1 BACTERIAL STRAINS AND MEDIA

Mycobacterium smegmatis mc^2155, a high efficiency transformation strain, has been previously described (Snapper *et al.*, 1990). This was cultured on 7H10 agar (Difco) supplemented with 0.5% glycerol, 10% Albumin Dextrose Complex (ADC) and 1 mM $CaCl_2$. Small (3-5 ml) cultures of *M. smegmatis* and large (50-100 ml) cultures for competent cell preparation were grown in 7H9 liquid media (Difco) that was supplemented with 0.2% glycerol, 0.05% Tween 80, and 10% Albumin Dextrose Complex (ADC). For larger cultures, particularly for use in phage infections, the Tween 80 was omitted and $CaCl_2$ was added at a final concentration of 1 mM. Unless otherwise specified, the antibiotics carbenicillin (CB) (50 μg/ml) and cyclohexamide (CHX) (10 μg/ml) were routinely added to mycobacterial media, and when necessary, kanamycin (KAN) was added to a final concentration of 20 μg/ml and hygromycin (HYG) was added to 150 μg/ml.

Escherichia coli strains were grown in Luria Broth (LB) and on LB agar (Difco) supplemented when necessary with CB (50 μg/ml), chloramphenicol (CM) (15 μg/ml), KAN (20 μg/ml), or HYG (150 μg/ml). High transformation efficiency GC5 competent cells (Gene Choice, Inc.) were used to transform ligation reactions. DH5α (Sambrook *et al.*, 1989) and XL1-Blue (Stratagene) strains were used to propagate plasmids, and HB101 (Promega) was used to propagate cosmids. The overexpression strains BL21(DE3) (Novagen), BL21(DE3)pLysS (Novagen), BL21

Star™(DE3) (Invitrogen), and BL21-CodonPlus(DE3)-RP (Stratagene) were used for protein expression from pET21 derivative plasmids. Top10 (Invitrogen) was used for protein expression from pBAD/gIII derivative plasmids.

2.2 PLASMIDS USED IN THIS STUDY

2.2.1 Commercially Available Plasmids

pET21a (Novagen) is a protein expression vector that contains an ampicillin resistance gene, the *lacI* gene, and the T7 ϕ10 promoter adjacent to a *lac* operator and a multiple cloning site (MCS). Vector includes an optional C-terminal polyhistadine (6×-His) tag for recombinant protein purification.

pET28c (Novagen) is similar to pET21a, but it encodes a kanamycin resistance gene and contains an optional N-terminal 6×-His tag for protein purification.

pBAD/gIII A (Invitrogen) is a pBR322-derived protein expression vector that contains an ampicillin resistance gene, the *araC* gene, and the *araBAD* promoter upstream of an MCS. It also encodes the leader peptide from the bacteriophage fd gene III protein (*gIII*) for secreted expression of recombinant protein into the *E. coli* periplasm. Vector includes an optional C-terminal 6×-His tag for recombinant protein purification and an optional C-terminal c-myc epitope for detection and analysis with an Anti-myc antibody.

pBAD/gIII C (Invitrogen) is identical to pBAD/gIII A, except the 6×-His tag is in a different reading frame relative to the MCS.

2.2.2 Plasmids Constructed by Others

pJL37 is a mycobacterial-*E. coli* shuttle vector that replicates extrachromosomally in both mycobacteria and *E. coli*. It is a derivative of pMV261 (Stover *et al.*, 1991) constructed by J. A. Lewis (Lewis & Hatfull, 2000, Bibb & Hatfull, 2002) that confers kanamycin resistance (*aph* gene) and contains an MCS as well as an NdeI site at the translation start site for the *M. bovis* BCG *hsp60* promoter (P_{hsp60}), which allows for constitutive expression of genes in mycobacteria.

pYUB854 (Bardarov *et al.*, 2002) was a gift from the laboratory of W. R. Jacobs. Jr. This is a cosmid vector that replicates extrachromosomally in *E. coli* (ColEI *oriE*) and contains a *cos* site for packaging in λ and a hygromycin resistance gene flanked by two MCSs. pYUB854 is a derivative of pYUB572, in which the *bla* gene was removed by digestion with BspHI and replaced with a *res–hyg–res gene* cassette flanked by MCSs.

pLC3 was a gift from gift from Craig Cassidy and James C. Sacchettini. This vector replicates extrachromosomally in *E. coli*, has a kanamycin resistance marker, and encodes an N-terminal 6×-His-maltose binding protein (MBP) tag linked to a TEV protease cleavage site. Expression of cloned genes is controlled by the T7 ϕ10 promoter, which is adjacent to a *lac* operator and the MCS. It also contains the *lacI* gene to control expression from this promoter.

pSD26 was a gift from the laboratory of W. R. Jacobs. Jr. This is a mycobacterial-*E. coli* shuttle vector that confers hygromycin resistance and contains an MCS upstream of the inducible *M. smegmatis acetamidase* promoter ($P_{acetamidase}$) (Parish *et al.*, 1997).

pTH1-1 (T. Huang, unpublished data) is a derivative of pSD26 that was modified to contain an NdeI site at the translational start site of $P_{acetamidase}$.

pJV53 (van Kessel & Hatfull, 2007) is a derivative of pLAM12 (see below) constructed by J. C. van Kessel that contains the mycobacteriophage Che9c *genes 60* and *61* downstream of $P_{acetamidase}$.

2.2.3 pLAM Plasmid Constructs

pLAM4 is a derivative of pET21a that encodes the mycobacteriophage Barnyard tape measure protein (TMP, gp33) with a C-terminal 6×-His tag. The *tmp* gene was amplified from Barnyard genomic DNA as a 6164 bp product with LJM15 and LJM16 and cloned into the NdeI and XhoI sites of pET21a.

pLAM5 is a derivative of pET21a that encodes the Rpf Motif from the Barnyard TMP (Rpf-like fragment, RLF) (amino acid residues 1501-1610) with a C-terminal 6×-His tag. The 356 bp product was amplified from Barnyard genomic DNA with LJM17 and LJM18 and cloned into the NdeI and XhoI sites of pET21a.

pLAM6 is a derivative of pJL37 that encodes the Barnyard RLF. The 336 bp insert was excised from pLAM5 with NdeI and XhoI and ligated to pJL37 that was cut with NdeI and SalI.

pLAM7 is a derivative of pLAM6 with the predicted signal sequence from the *M. tuberculosis* H37Rv RpfA protein (Rv0867c) fused to the N-terminus of the RLF to make a secreted form of the protein (sRLF). The 99 base oligonucleotides LJM119 and LJM120a were annealed and kinased, generating ends with 5′ overhangs compatible with NdeI cut DNA. This was ligated to NdeI digested pLAM6, and clones were screened for having the signal sequence ligated in the correct orientation. Upon ligation, only one NdeI site is reconstituted at the 5′ end of the inserted signal sequence.

pLAM8 is a derivative of pTH1-1 that encodes the RLF. The 344 bp insert was excised from pLAM6 with NdeI and HincII and ligated to pTH1-1 that was cut with BamHI, Klenow treated (to generate a blunt end) and then cut with NdeI.

pLAM9 is a derivative of pTH1-1 that encodes the sRLF. The 443 bp insert was excised from pLAM7 and cloned into pTH1-1 with the same strategy used to generate pLAM8.

pLAM10 is a derivative of pLAM8 with a leucine codon at amino acid 112 (CTC) changed to an amber stop codon (TAG) by site-directed mutagenesis (L112→Stop).

pLAM11 is a derivative of pLAM9 with a leucine codon at amino acid 145 (CTC) changed to an amber stop codon (TAG) by site-directed mutagenesis (L145→Stop).

pLAM12 is a derivative of pJL37 in which P_{hsp60} was excised with BamHI and XbaI and replaced with the $P_{acetamidase}$ contained on a 2660 bp BamHI-XbaI fragment of pTH1-1.

pLAM13 is a derivative of pLAM12 that encodes the RLF (357 bp) sub-cloned from pLAM10 with NdeI and ClaI.

pLAM14 is a derivative of pLAM12 that encodes the sRLF (456 bp) sub-cloned from pLAM11 with NdeI and ClaI.

pLAM15 is a derivative of pJL37 that encodes the Barnyard TMP. The 6141 bp fragment was excised from pLAM4 by digestion with XhoI, followed by Klenow treatment (to blunt) and digestion with NdeI, and this fragment was ligated to pJL37 that was cut with NheI, treated with Klenow to blunt, then cut with NdeI.

pLAM16 is a derivative of pLAM12 that encodes the Barnyard TMP, which was constructed with the same strategy as pLAM15.

pLAM19 is a derivative of pET21a that encodes Barnyard gp33 Segment 1 (residues 1-102) with a C-terminal 6×-His tag (11.8 kDa). The 333 bp insert was amplified from EcoRV-linearized pLAM4 with LJM25 and LJM26 and cloned into the NdeI and XhoI sites of pET21a.

pLAM20 is a derivative of pET21a that encodes Barnyard gp33 Segment 2 (residues 103-242) with a C-terminal 6×-His tag (14.9 kDa). The 446 bp insert was amplified from EcoRV-linearized pLAM4 with LJM27 and LJM28 and cloned into the NdeI and XhoI sites of pET21a.

pLAM21 is a derivative of pET21a that encodes Barnyard gp33 Segment 3 (residues 243-445) with a C-terminal 6×-His tag (21.9 kDa). The 635 bp insert was amplified from EcoRV-linearized pLAM4 with LJM29 and LJM30 and cloned into the NdeI and XhoI sites of pET21a.

pLAM22 is a derivative of pET21a that encodes Barnyard gp33 Segment 4 (residues 446-721) with a C-terminal 6×-His tag (30 kDa). The 854 bp insert was amplified from EcoRV-linearized pLAM4 with LJM31 and LJM32 and cloned into the NdeI and XhoI sites of pET21a.

pLAM23 is a derivative of pET21a that encodes Barnyard gp33 Segment 6 (residues 951-1109) with a C-terminal 6×-His tag (18 kDa). The 505 bp insert was amplified from EcoRV-linearized pLAM4 with LJM35 and LJM36 and cloned into the NdeI and XhoI sites of pET21a.

pLAM24 is a derivative of pET21a that encodes Barnyard gp33 Segment 7 (residues 1112-1261) with a C-terminal 6×-His tag (16 kDa). The 476 bp insert was amplified from EcoRV-linearized pLAM4 with LJM37 and LJM38 and cloned into the NdeI and XhoI sites of pET21a.

pLAM25 is a derivative of pET21a that encodes Barnyard gp33 Segment 9 (residues 1611-1821) with a C-terminal 6×-His tag (22 kDa). The 659 bp insert was amplified from EcoRV-linearized pLAM4 with LJM41 and LJM42 and cloned into the NdeI and XhoI sites of pET21a.

pLAM26 is a derivative of pET21a that encodes the 70 kDa C-terminal fragment of the Barnyard TMP (SKF) with a C-terminal 6×-His tag. The 2015 bp fragment was amplified from EcoRV-linearized pLAM4 with LJM15 and LJM16 and cloned into the NdeI and XhoI sites of pET21a.

pLAM27 is a derivative of pLJ37 that encodes the SKF. The 1992 bp fragment was excised from pLAM26 by digestion with XhoI, followed by Klenow treatment (to blunt) and digestion with NdeI, and this was ligated to pJL37 cut with NheI, treated with Klenow to blunt and cut with NdeI.

pLAM28 is a derivative of pLAM12 that encodes the SKF, which was constructed with the same strategy as pLAM27.

pLAM29 is a derivative of pLAM27 with the RpfA signal sequence fused to the N-terminus of the SKF to generate a secreted form of the protein (sSKF). The 99 base oligonucleotides LJM119 and LJM120a were annealed and kinased, generating ends with 5′ overhangs compatible with NdeI cut DNA. This was then ligated to NdeI digested pLAM27, and clones were screened for having the signal sequence ligated in the correct orientation. Upon ligation, only one NdeI site is reconstituted at the 5′ end of the inserted signal sequence.

pLAM30 is a derivative of pLAM28 that encodes the sSKF, which was constructed with the same strategy as pLAM29

pLAM32 is a derivative of pYUB854 containing 500 bp homology upstream of the Barnyard Rpf Motif (22,111 bp-22,610 bp) amplified with LJM58 and LJM59 (526 bp) and inserted at the BglII and HindIII sites of pYUB854, and 500 bp of homology downstream of the Rpf Motif (22,941 bp-23,441 bp) amplified with LJM60 and LJM61 (529 bp) and inserted at the HindIII and NcoI sites of pYUB854 using a 3-way ligation strategy. Cloning inserts the sequence GCA directly upstream of the HindIII site, and regions of homology were amplified from EcoRV-linearized pLAM4.

pLAM33 is a derivative of pLAM26 that encodes Barnyard gp31 and gp32 immediately upstream of the BY-SKF. The 905 bp insert was amplified from Barnyard genomic DNA with LJM63 and LJM64 and cloned into the NdeI site of pLAM26.

pLAM34 is a derivative of pET28c that encodes the Barnyard TMP with an N-terminal 6×-His tag. The 1992 bp insert was excised from pLAM26 by digestion with XhoI, followed by Klenow treatment (to blunt) and digestion with NdeI, and this was ligated to pET28c that was cut with NheI, treated with Klenow to blunt, then cut with NdeI.

pLAM35 is a derivative of pLC3 that encodes the Barnyard TMP with an N-terminal 6×-His-MBP tag, which was constructed with the same strategy as pLAM34.

pLAM37 is a derivative of pLAM7 with the ATG codon at amino acid 34 (1370 bp) removed by site-directed mutagenesis with LJM65 and LJM66 in order to remove a methionine codon and make the signal peptide cleavage site more closely match that of RpfA (Rv0867c).

pLAM38 is a derivative of pLAM14 with the ATG codon at amino acid 34 (3646 bp) removed by site-directed mutagenesis with LJM65 and LJM66 in order to remove a methionine codon and make the signal peptide cleavage site more closely match that of RpfA.

pLAM39 is a derivative of pLAM29 with the ATG codon at amino acid 34 (1370 bp) removed by site-directed mutagenesis with LJM67 and LJM68 in order to remove a methionine codon and make the signal peptide cleavage site more closely match that of RpfA.

pLAM40 is a derivative of pLAM30 with the ATG codon at amino acid 34 (3646 bp) removed by site-directed mutagenesis with LJM67 and LJM68 in order to remove a methionine codon and make the signal peptide cleavage site more closely match that of RpfA.

pLAM41 is a derivative of pBAD/gIII A that encodes the Barnyard RLF with a C-terminal 6×-His tag. The 355 bp insert was amplified from EcoRV-linearized pLAM4 with LJM69 and LJM71 and inserted into the NcoI and SalI sites of pBAD/gIII A.

pLAM42 is a derivative of pBAD/gIII A that encodes the Barnyard RLF. The 358 bp insert was amplified from EcoRV-linearized pLAM4 with LJM69 and LJM72 and inserted into the NcoI and XbaI sites of pBAD/gIII A.

pLAM43 is a derivative of pBAD/gIII C that encodes the Barnyard SKF with a C-terminal 6×-His tag. The 2012 bp insert was amplified from EcoRV-linearized pLAM4 with LJM70 and LJM73 and inserted into the NotI and SalI sites of pBAD/gIII C.

pLAM44 is a derivative of pBAD/gIII C that encodes the Barnyard SKF. The 2012 bp insert was amplified from EcoRV-linearized pLAM4 with LJM70 and LJM74 and inserted into the NotI and XbaI sites of pBAD/gIII C.

pLAM45 is a derivative of pBAD/gIII A that encodes the Rpf protein from *M. luteus* with a C-terminal 6×-His tag. The 586 bp insert was amplified from *M. luteus* genomic DNA with LJM77 and LJM78 and cloned into the NcoII and SalI sites of pBAD/gIII A.

pLAM46 is a derivative of pBAD/gIII A that encodes a fusion of the Barnyard RLF and the cell wall binding domain of the *M. luteus* Rpf (RLF-LysM). The RLF was amplified as a 280 bp fragment from pLAM41 with LJM69 and LJM79, which engineer an NcoI site at the 5′ end and a BglII site at the 3′ end. The Rpf LysM was amplified as a 347 bp fragment from pLAM45 with LJM80 and LJM78, which engineer a BglII site at the 5′ end and a SalI site at the 3′ end. These were cloned into the NcoI and SalI sites of pBAD/gIII A in a 3-way ligation.

pLAM47 is a derivative of pLAM46 with lysine 53 (numbers refer to the residue number in *M. luteus* Rpf) mutated to a cysteine (K53C) by site-directed mutagenesis with LJM84 and LJM85.

pLAM48 is a derivative of pLAM46 with arginine 114 mutated to a cysteine residue (R114C) by site-directed mutagenesis with LJM86 and LJM87.

pLAM49 is a derivative of pLAM48 (contains R114C) with the K53C mutation inserted by site-directed mutagenesis with LJM84 and LJM85.

*In the following set of plasmids, 42 bp encoding for 14 amino acids at the N-termini of the RLF-LysM fusion proteins were removed and replaced by 12 bp encoding four amino acids (Figure 7), such that their N-termini (directly downstream of the signal peptide in the pBAD constructs and upstream of the putative active site) match that of the *M. luteus* Rpf.

pLAM50 is a derivative of pBAD/gIII A that was created by PCR amplifying the fusion protein containing K53C and R114C (577 bp) from pLAM49 with LJM91 and LJM78 and inserting this into the NcoI and SalI sites of pBAD/gIII A.

pLAM51 is a derivative of pBAD/gIII A that was created by PCR amplifying the 'wild-type' fusion protein (577 bp) from pLAM46 with LJM98 and LJM78 and inserting this into the NcoI and SalI sites of pBAD/gIII A.

pLAM52 is a derivative of pBAD/gIII A that was created by PCR amplifying the fusion protein containing K53C (577 bp) from pLAM47 with LJM91 and LJM78 and inserting this into the NcoI and SalI sites of pBAD/gIII A.

pLAM53 is a derivative of pBAD/gIII A that was created by PCR amplifying the fusion protein containing R114C (577 bp) from pLAM48 with LJM98 and LJM78 and inserting this into the NcoI and SalI sites of pBAD/gIII A.

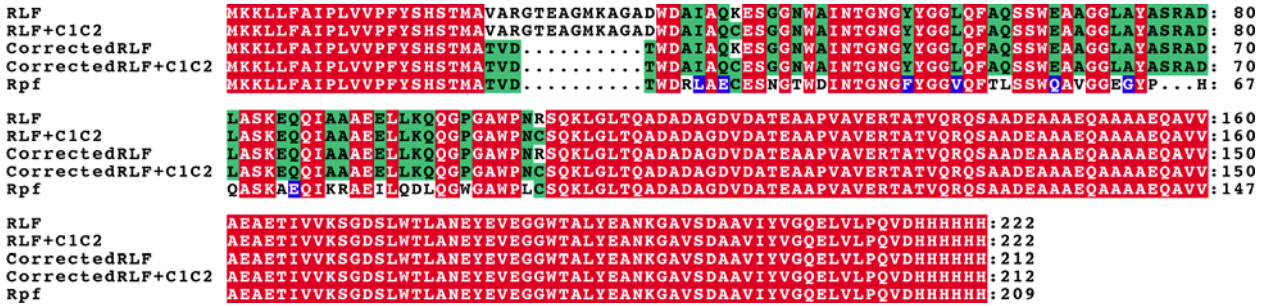


Figure 7. Barnyard RLF-Rpf fusion proteins with corrected N-termini.

Figure 7: Sequence of Barnyard RLF-Rpf fusion proteins with corrected N-termini. The inserts were amplified from plasmids containing the original fusions using a 5′ primer that replaced 14 codons at the N-terminus of the protein with the four corresponding codons from the *M. luteus* Rpf. The primers also included the NcoI and SalI sites from the orginal constructs for cloning the corrected fusions back into the expression vector pBAD/gIII A.

pLAM54 is a derivative of pBAD/gIII A that encodes the *M. luteus* Rpf LysM with a C-terminal 6×-His tag. The 337 bp insert was amplified from pLAM45 with LJM99 and LJM78 and cloned into the NcoI and SalI sites of pBAD/gIII A.

pLAM55 is a derivative of pJL37 that encodes the *M. luteus* Rpf with its native signal sequence. The 690 bp insert was amplified from *M. luteus* genomic DNA with LJM81 and LJM82 and cloned into the NdeI and NheI sites of pJL37.

pLAM60 is a derivative of pJL37 that encodes the RLF-LysM fusion with K53C and R114C fused to the RpfA signal sequence. The 110 bp RpfA signal sequence was amplified from pLAM38 with LJM83 and LJM92, which insert an NdeI site at the 5′ end and create a blunt 3′ end. The 567 bp fusion was amplified from pLAM50 with LJM94 and LJM82, which create a blunt 5′ end and insert an NheI site at the 3′ end. The fragments were cloned into the NdeI and NheI site of pJL37 in a 3-way ligation.

pLAM61 is a derivative of pJL37 that encodes the RLF-LysM fusion joined to the RpfA signal sequence. The 567 bp product was amplified from pLAM51 with LJM94 and LJM82 and cloned into pJL37 with the same strategy used to construct pLAM60.

pLAM62 is a derivative of pJL37 that encodes the RLF-LysM fusion with K53C fused to the RpfA signal sequence. The 567 bp product was amplified from pLAM52 with LJM94 and LJM82 and cloned into pJL37 with the same strategy used to construct of pLAM60.

pLAM63 is a derivative of pJL37 that encodes the RLF-LysM fusion with R114C fused to the RpfA signal sequence. The 567 bp product was amplified from pLAM53 with LJM94 and LJM82 and cloned into pJL37 with the same strategy used to construct pLAM60.

pLAM64 is a derivative of pJL37 that encodes *M. luteus* Rpf LysM fused to the RpfA signal sequence. The 230 bp product was amplified from pLAM45 with LJM100 and LJM82 and cloned into pJL37 with the same strategy used to construct pLAM60.

pLAM65 is a derivative of pLAM12 that encodes the *M. luteus* Rpf with its native signal sequence. The 690 bp insert was amplified from *M. luteus* genomic DNA with LJM81 and LJM82 and cloned into the NdeI and NheI sites of pLAM12.

pLAM70 is a derivative of pLAM12 that encodes the RLF-LysM fusion with K53C and R114C fused to the RpfA signal sequence. The 110 bp RpfA signal sequence was amplified from pLAM38 with LJM83 and LJM92, which insert an NdeI site at the 5′ end and create a blunt 3′ end. The 567 bp fusion was amplified from pLAM50 with LJM94 and LJM82, which create a blunt 5′ end and insert an NheI site at the 3′ end. The fragments were cloned into the NdeI and NheI site of pLAM12 in a 3-way ligation.

pLAM71 is a derivative of pLAM12 that encodes the RLF-LysM fusion joined to the RpfA signal sequence. The 567 bp product was amplified from pLAM51 with LJM94 and LJM82 and cloned into pLAM12 with the same strategy used to construct pLAM70.

pLAM72 is a derivative of pLAM12 that encodes the RLF-LysM fusion with K53C fused to the RpfA signal sequence. The 567 bp product was amplified from pLAM52 with LJM94 and LJM82 and cloned into pLAM12 with the same strategy used to construct of pLAM70.

pLAM73 is a derivative of pLAM12 that encodes the RLF-LysM fusion with R114C fused to the RpfA signal sequence. The 567 bp product was amplified from pLAM53 with LJM94 and LJM82 and cloned into pLAM12 with the same strategy used to construct pLAM70.

pLAM74 is a derivative of pLAM12 that encodes *M. luteus* Rpf LysM fused to the RpfA signal sequence. The 230 bp product was amplified from pLAM45 with LJM100 and LJM82 and cloned into pLAM12 with the same strategy used to construct pLAM70.

pLAM75 is a derivative of pBAD/gIII A that encodes the Rpf Motif from mycobacteriophage Giles TMP fused to the Rpf LysM. The 259 bp fragment of the tape measure gene (2968 bp-3187 bp) was amplified from Giles genomic DNA with LJM101 and LJM103, which insert an NcoI site at the 5′ end and a BglII at the 3′ end. The Rpf LysM was amplified as a 347 bp fragment from pLAM45 with LJM80 and LJM78, which engineer a BglII site at the 5′ end and a SalI site at the 3′ end. These were cloned into the NcoI and SalI sites of pBAD/gIII A in a 3-way ligation.

pLAM79 is a derivative of pLAM51 with the glutamate codon at position 54 (GAA) mutated to an alanine (GCA) (E54A) by site-directed mutagenesis with LJM110 and LJM111.

pLAM80 is a derivative of pLAM50 with the glutamate codon at position 54 (GAA) mutated to an alanine (GCA) (E54A) by site-directed mutagenesis with LJM112 and LJM113.

2.3 GENERAL CLONING AND DNA MANIPULATION

PCR reactions were performed with *Pfu* Polymerase (Stratagene) according to manufacturer's instructions with deoxyribonucleotides (dNTPs) added at a final concentration of 1 mM (0.25 mM each), primers added at a final concentration of 0.25 μM each and 5% DMSO added to most reactions. Reactions were cleaned and concentrated using QIAquick PCR Purification Kit or Gel Extraction Kit (QIAGEN). Restriction enzymes were obtained from NEB, and parent plasmids were treated with calf intestinal phosphatase (CIP) (NEB) according to manufacturer's instructions. When necessary, Klenow (NEB) treatment or T4 polynucleotide kinase (Roche) treatment was performed according to manufacturer's instructions. Vectors and inserts were quantified on 0.8% agarose gels using Gene Choice DNA Ladder I, and all ligations were performed with Fast-Link DNA Ligation Kit (Epicentre) according to manufacturer's instructions.

2.3.1 Oligonucleotides and Primers Used in this Study

The oligonucleotides and primers used for various applications, including cloning, sequencing, mutagenesis and recombineering, are listed in Table 2.

Table 2. Oligonucleotides and primers used in this study.

Name	*Sequence*	*Use*
LJM15	TTTTTTCATATGCCCAATTACGATTTGGGAAC	pLAM4
LJM16	TTTTTTCTCGAGTTGCCCTGTAGCTCCTGC	pLAM4/pLAM16
LJM17	TTTTTTCATATGGCAGTGGCTCGCGG	pLAM5
LJM18	TTTTTTCTCGAGCTCGTCACTGATTTCGC	pLAM5
LJM19	TATGAGTGGACGCCACCGTAAGCCCACCACATCCAACGTCAGCGTCGCCA AGATCGCCTTTACCGGCGCAGTACTCGGTGGCGGCGGCATCGCCATGGC	Rv0867c Signal sequence
LJM20a	TAGCCATGGCGATGCCGCCGCCACCGAGTACTGCGCCGGTAAAGGCGATC TTGGCGA C GCTGACGTTGGATGTGGTGGGCTTACGGTGGCGTCCACTCA	Rv0867c Signal sequence
LJM20b	CCGCAGTTGTTCTCGCATACCCCATC	Sequence pLAM12
LJM21	CCATGATGGCCGGACAAACAACAG	Sequencing
LJM22	GTTTGCAAGCAGCAGATTACGCGCAG	Sequence pYUB854
LJM23	CGGACGGTTGCTAGCACGCGCACCAT	Sequence pYUB854
LJM24a	GTGGCGAAATCAGTGACGAGTAGGACGTAGTTGATC CAAGCTT	Site-directed mutagenesis
LJM24b	AAGCTTGGATCAACTACGTCCTACTCGTCACTGATTTCGCCAC	Site-directed mutagenesis
LJM25	TTTTTTCATATGCCCAATTACGATTTGGGAAC	pLAM19
LJM26	TTTTTTCTCGAGCAGTCCCTTCAAACGG	pLAM19
LJM27	TTTTTTCATATGCACGGAGCAACAGAGAAG	pLAM20
LJM28	TTTTTTCTCGAGCAGTGAGAGCTTGTCCAT	pLAM20
LJM29	TTTTTTCATATGGGCACTGTGCAGTTGATCG	pLAM21
LJM30	TTTTTTCTCGAGACCCTGCCCGATGAATTCAA	pLAM21
LJM31	TTTTTTCATATGCAGACGATCACCGACG	pLAM22
LJM32	TTTTTTCTCGAGGAGCTTGTCGCCAAC	pLAM22
LJM33	TTTTTTCATATGAAGGCAGGTATCGCC	BY gp33 Seg.5
LJM34	TTTTTTCTCGAGGGAGTCTCCGGCTG	BY gp33 Seg.5
LJM35	TTTTTTCATATGCTTGAAGGTCTGGGCTTG	pLAM23
LJM36	TTTTTTCTCGAGGTCCTGCACCAGCA	pLAM23
LJM37	TTTTTTCATATGACATTGGAAGCCGAAATCG	pLAM24
LJM38	TTTTTTCTCGAGCTCGTAACCCTTTGCAAC	pLAM24
LJM39	TTTTTTCATATGGAAACGACTTTCGTCGTC	BY gp33 Seg.8
LJM40	TTTTTTCTCGAGGCCTTCGATGTCACC	BY gp33 Seg.8
LJM41	TTTTTTCATATGGACTTGCTGAGCCGTG	pLAM25
LJM42	TTTTTTCTCGAGTTCAACCATGCCACGTTC	pLAM25
LJM43	TTTTTTCATATGAACCAGAATCCCATGGAC	BY gp33 Seg.10
LJM44	TTTTTTCTCGAGTCATTGCCCTGTAGCTC	BY gp33 Seg.10
LJM51	TTTTTTCATATGGCCGCAGATGGATTGGCCG	pLAM26
LJM52	GCACGGAGCAACAGAGAAGCTG	RT-PCR
LJM53	CCAGTGAGAGCTTGTCCATGTTGTTG	RT-PCR
LJM54	GTCAATCAGGTCAAGGACGCGATCAC	RT-PCR
LJM55	CTTGCGCTTGCGAGCTTCTTCAC	RT-PCR
LJM56	GAAATCAGTGACGAGGACTTGCTGAG	RT-PCR
LJM57	GCCTGCCGGAGTGTCATTCAC	RT-PCR
LJM58	TTTTTTAGATCTACCGACTGCTCGCAGACTTC	pLAM32
LJM59	TTTTTTAAGCTTGCCTTCGATGTCACCGGGA	pLAM32
LJM60	TTTTTTAAGCTTGCAGACTTGCTGAGCCGTGTGC	pLAM32
LJM61	TTTTTTCCATGGACCTGCTGATCGCTTGAACTTGG	pLAM32
LJM62	TGCTCCCGGTGACATCGAAGGCAAGCTTGCAGACTTGCTGAGCCGTGTGC	RLF deletion probe

*Restriction sites for cloning are underlined

(Table 2 Continued)

Name	*Sequence*	*Use*
LJM63	TTTTTTCATATGGGTAATCACACCGGTTCG	pLAM33
LJM64	TTTTTTCATATGCTATCTCACCCCGTTTCTC	pLAM33
LJM65	CATCGCCATGGCTGCAGTGGCTCGCG	Site-directed mutagenesis
LJM66	CGCGAGCCACTGCAGCCATGGCGATG	Site-directed mutagenesis
LJM67	GCATCGCCATGGCTGCCGCAGATGGATTG	Site-directed mutagenesis
LJM68	CAATCCATCTGCGGCAGCCATGGCGATGC	Site-directed mutagenesis
LJM69	TTTTTTCCATGGCAGTGGCTCGCGGCA	pLAM41/42/46
LJM70	TTTTTTGCGGCCGCAGATGGATTGGC	pLAM43/44
LJM71	TTTTTTGTCGACCTCGTCACTGATTTCGCCAC	pLAM41
LJM72	TTTTTTCTAGAGGTCCTCGTCACTGATTTCGCCA	pLAM42
LJM73	TTTTTTGTCGACTTGCCCTGTAGCTCCTG	pLAM43
LJM74	TTTTTTCTAGAGTTGCCCTGTAGCTCCTGC	pLAM44
LJM75	CCATAAGATTAGCGGATCCTACCTGACGC	Sequence pBAD
LJM76	GCTACTGCCGCCAGGCAAATTCTG	Sequence pBAD
LJM77	TTTTTTCCATGGCCACCGTGGACACCTGGGACC	pLAM45
LJM78	TTTTTTGTCGACCTGCGGCAGGACGAGCTCCTGG	pLAM45/46/50-54/75
LJM79	TTTTTTAGATCTGTTCGGCCATGCGCCCG	pLAM46
LJM80	TTTTTTAGATCTCAGAAGCTGGGCCTGACCCAGGCT	pLAM46/75
LJM81	TTTTTTCATATGGACACCATGACTCTCTTCACCACTTCCG	pLAM55/65
LJM82	TTTTTTGCTAGCCCTGCGGCAGGACGAGCTCC	pLAM55/60-64/65/70-74
LJM83	TTTTTTCATATGAGTGGACGCCACCGTAAGCCCACC	pLAM60-64/70-74
LJM84	CTGGGATGCCATAGCGCAA**TGC**GAATCTGGTGGAAACTGGG	Site-Directed Mutagenesis
LJM85	CCCAGTTTCCACCAGATTC**GCA**TTGCGCTATGGCATCCCAG	Site-Directed Mutagenesis
LJM86	GGCGCATGGCCGAAC**TGC**TCTCAGAAGCTGGGC	Site-Directed Mutagenesis
LJM87	GCCCAGCTTCTGAGA**GCA**GTTCGGCCATGCGCC	Site-Directed Mutagenesis
LJM88	AGGCGCGCGGCGGCGGGGTGCCCAACCTCGGCGGGCCGTACGGTGCCGGGATCGACCCGGAGACCGGCGAGCGCGGGTTCTACACCCCGGACCCGAAGAA	Giles Rpf Δ
LJM89	CGCGAACCACATGGACCATGTGCACGTCTTCGCAAACGACATCGCGGCGCAGGCGCGCGGCGGCGGGGTGCCCAA	Giles Rpf Δ Extender 1
LJM90	CCCGGATCCGCTGTTCGGCGTCGGCCACCTTTTGGTCGGCCTCCCGGATTTTCTTCGGGTCCGGGGTGTAGAACC	Giles Rpf Δ Extender 2
LJM91	TTTTTTCCATGGCAACCGTGGACACCTGGGATGCCATAGCGCAATGCGA	pLAM50/52
LJM92	AGCCATGGCGATGCCGCCGCCACCGAGTA	pLAM60-64/70-74
LJM94	GCAACCGTGGACACCTGGGATGCCATAG	pLAM60/70-73
LJM95	GGCCGATGTGATGATCCCGAACTACAACACCCCGCA	Screen Giles Mutants
LJM96	ATTCGGCCTGCGCGATGGCATCCCAATCGGGCA	Screen Giles Mutants
LJM97	CGTCGGCCTTCGCCGCGTCCAACTGCCGT	Screen Giles Mutants
LJM98	TTTTTTCCATGGCAACCGTGGACACCTGGGATGCCATAGCGCAAAAGGA	pLAM51/53

*Restriction sites for cloning are underlined

(Table 2 Continued)

Name	*Sequence*	*Use*
LJM99	TTTTTTCCATGGCCTGACCCAGGCTGAC	pLAM54
LJM100	GCGGGCCTGACCCAGGCTGAC	pLAM74
LJM101	TTTTTTCCATGGCAACCGTGGACACCTGGGATGCCATCGCGCAGGCCGAA	pLAM75
LJM103	TTTTTTAGATCTGTTCGGCCACGCCTTCGGCGCGTT	pLAM75
LJM114	CACCGAACGCGTGATGTGGCAACACC	Sequencing-Giles
LJM115	CGATCCCTCACGAGGTTCTCCGC	Sequencing-Giles
LJM116	GACGGTGCCCGATTGGGATGCCATCGCGCAGGCAGCTTCCGGCGGCAACT GGGCGATCAACACCGGCAAC	Giles E→A Point Mutant + AluI Site
LJM117	GTTGCCGGTGTTGATCGCCCAGTTGCCGCCGGAAGCTGCCTGCGCGATGG CATCCCAATCGGGCACCGTC	Giles E→A Point Mutant + AluI Site
LJM119	TTGCCACCGGCCGGGCGTTGGGGATCCTCCAATCGTTCGCTAACGGCGGG TTGAACACCGGCGCGGCGACGCAGACGGTGCCCGATTGGGATGCCATCGC	Giles Mt3Δ
LJM120	GCGATGGCATCCCAATCGGGCACCGTCTGCGTCGCCGCGCCGGTGTTCAA CCCGCCGTTAGCGAACGATTGGAGGATCCCCAACGCCCGGCCGGTGGCAA	Giles Mt3Δ
LJM121	CGAAGCGTTTATCCCGATCAACGGGTCGCAACGGTCGAAGGACATCTGGG TTGCCACCGGCCGGGCGTTGGGGAT	Mt3Δ Extender 1
LJM122	AGTACCCGTTGCCGGTGTTGATCGCCCAGTTGCCGCCGGATTCGGCCTGC GCGATGGCATCCCAATCGGGCACCG	Mt3Δ Extender 2
LJM123	GACATGCAGGCACTGTTGCAGCAGTACCCGATG	Screen Giles Mutants
LJM124	TTGCCACCGGCCGGGCGTTGGGGATCCTCCAATCGTTCGCTAACGGCGGG ATCGACCCGGAGACCGGCGAGCGCGGGTTCTACACCCCGGACCCGAAGAA	Giles Rpf and Motif3 Double Δ (DΔ)
LJM125	TTCTTCGGGTCCGGGGTGTAGAACCCGCGCTCGCCGGTCTCCGGGTCGAT CCCGCCGTTAGCGAACGATTGGAGGATCCCCAACGCCCGGCCGGTGGCAA	Giles Rpf and Motif3 Double Δ (DΔ)
LJM126	CCTCCAATCGTTCGCTAACGGCGGGATC	MAMA-PCR DΔ
LJM127	CCTCCAATCGTTCGCTAACGGCGGGAAC	MAMA-PCR DΔ
LJM128	CGGCGTAACCCGTCTGCCACGGGAAC	Screen Giles Mutants
LJM129	AGTACCCGTTGCCGGTGTTGATCGCCCAGTTGCCGCCGGAAGCTGCCTGC GCGATGGCATCCCAATCGGGCACCG	Giles MtΔ/ RpfE→A Extender 2
LJM130	GTGGATCCGCCGGTATGCCGGTGTACGCCATCCAATCCGGCACCGTCATC CCGCTCACGGCCGCGAAGATTGGGTTGGTGCTCTGATGGCGTGGAAACCT	Giles LysAΔ
LJM131	CACGACCGGCCGGTCCGCCCCGCACGGCGGCGTTGACTTCGGCCGGGCCG GTGGATCCGCCGGTATGCCGGTGTA	Giles LysAΔ Extender 1
LJM132	TCCTTCGCGTCGGCGATGTCGGGGTGCCGATCTCCGATTTGGTATTCGGT AGGTTTCCACGCCATCAGAGCACCA	Giles LysAΔ Extender 2
LJM133	CCAGCTTCTGTACCGCGACTAGCGCGTTC	Screen LysAΔ
LJM134	CACGCCTGCAACACCCGCGACACGTAGTC	Screen LysAΔ
LJM135	TACGCCATCCAATCCGGCACCGTCATCCCG	MAMA-PCR LysAΔ
LJM136	GATCCTCCAATCGTTCGCTAACGGCGGGTTGAA	MAMA-PCR Mt3Δ
LJM137	GTACCGGTAGACCGTTCGGCAGCGTTTGG	Screen Mt3Δ & DΔ
LJM144	AGAGACTGGGCCGGGCGCACCCCGGCCGTCACCGCACACCACCACCACCA CCACTAACACACCGAAGGGATCGCAACAGTGAGCACGAAT	Giles LysB-His
LJM145	TACGAATTCCGCGAAGTGTGGCCCGGTCAAACGTATCTGGGCCTGGCCAT CCAACACGTCAGAGACTGGGCCGGGCGCACCCCGGCCGTC	Giles LysB-His Extender 1
LJM146	GAGCGTGCGTAGCGGATTGACACCCGGCGCGGGGACCGTGGCCGGGTTCG GATTGGGCGGATTCGTGCTCACTGTTGCGATCCCTTCGGT	Giles LysB-His Extender 2
LJM147	CCGTCACCGCACACCACCACCACCAC	Screen LysB-His
LJM148	CATCCCCTGCGGTGCCGGTTCGTCG	Screen LysB-His

*Restriction sites for cloning are underlined

2.3.2 Transformation and Electroporation

Plasmids were transformed into the *E. coli* strain DH5α by heat shock transformation methods using $CaCl_2$ competent cells (Sambrook *et al.*, 1989) or into XL1-Blue cells by electroporation using a Bio-Rad Gene Pulser II set to 200 Ω, 2.5 kV and 25 μF. All other strains were transformed according to the manufacturer's instructions.

M. smegmatis competent cells were generated as previously described (Bibb & Hatfull, 2002); in some cases, cells were washed four times with ice-cold 10% glycerol prior to resuspension in ice-cold 10% glycerol. Competent cells were stored in single aliquots of 100 μl or larger aliquots of 200 μl or 400 μl. Recombineering strains of *M. smegmatis* (containing the plasmid pJV53) were made competent in the same way, except that ADC was omitted from overnight cultures and KAN and succinate (at a final concentration of 0.2%) were added to the media. When cultures reached an OD_{600} of 0.4-0.45, recombineering functions were induced by adding 0.2% acetamide, and the cells were grown an additional 3 hours. Plasmid DNA (50-100 ng), bacteriophage DNA and/or recombineering substrates (50-500 ng) were introduced into 50-100 μl electrocompetent *M. smegmatis* cells using a Bio-Rad Gene Pulser II set to 1000 Ω, 2.5 kV and 25 μF. Cells were recovered in 900 μl 7H9 media with 10% ADC and 0.05% Tween 80 at 37°C for 2 hours. For transformations with phage DNA, 1 mM $CaCl_2$ was added and Tween 80 was omitted.

2.3.3 Sequencing

Cloned DNA was sequenced according to protocols for PE Applied Biosystems dRhodamine terminator chemistry using an Applied Biosystems GeneAmp 2700 thermal cycler. Sequencing reactions were analyzed on an ABI 310 capillary sequencer. In some cases, cloned DNA was sequenced by the University of Pittsburgh Genome Center.

2.3.4 Site-Directed Mutagenesis

The QuikChange Site-Directed Mutagenesis (Stratagene) kit was employed, and protocols were followed according to manufacturer's instructions. DpnI digests were recovered by transforming into GC5 cells.

2.3.5 Cosmid Packaging Reactions

Packaging reactions were performed using Gigapack III Gold Packaging Extract (Stratagene) according to manufacturer's instructions.

2.4 PROTEIN EXPRESSION, PURIFICATION AND ANAYLSIS

2.4.1 Expression and Purification of the Barnyard TMP Rpf-Like Fragment

Expression of the Barnyard RLF from pLAM5 was induced in BL21(DE3)pLysS and BL21(DE3) *E. coli* strains in generally the same way. Colonies were inoculated into ~2 ml of LB CB (CM was

added for pLysS cells) and grown overnight with vigorous agitation at 37ºC. Saturated cultures were subcultured 1:200 – 1:50 into fresh media and grown to an OD_{600} of 0.4 – 0.6 (usually 3-3½ hours); small (2-5 ml) cultures were used for test inductions, and large cultures (1-2 L) were used for purifications. Expression was induced with 0.8-1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG, Sigma) at 37ºC with vigorous shaking for 2 hours. Longer expression times resulted in degradation and reduced levels of expression of the recombinant protein. Cell aliquots from inductions were pelleted, resuspended in 2X SDS-sample buffer (typically 50 µl), boiled at 95ºC for 5 minutes and usually separated on 3-layer (4%, 10% and 16.5%) tris-tricine SDS-PAGE gels for accurate separation of small proteins. At 2 hours post-induction, cultures for purifications were chilled on ice for 10 minutes and then pelleted at 6000 × g for 10 minutes. Pellets were resuspended in 4 ml chilled lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8) per 200 ml of culture, frozen on dry ice, and stored at -80ºC.

Protein purifications were performed using TALON™ Metal Affinity Resin (Clontech), according to manufacturer's instructions. Briefly, pellets were thawed on ice and subjected to an additional round of freezing on dry ice and thawing on ice; lysozyme (1 mg/ml) was often added to purifications from BL21(DE3) cells after the first thaw. These were sonicated 3-6 times, using 10-45 second bursts with 30 second – 1 minute rests in between, until (fairly) smooth, and the insoluble material was pelleted at 12,000 × g for 20 minutes. The clarified lysate was applied to TALON™ cobalt resin (~1 ml per L of induced culture) that had been equilibrated with lysis buffer and incubated at room temperature for 20 minutes. Protein bound resin was washed twice with 20 bed volumes lysis buffer and resuspended in one bed volume of lysis buffer before adding to the column. This was washed a fourth time with five bed volumes of lysis buffer and a fifth time with the same amount of lysis buffer at pH 7 prior to elution with TALON™ 1 X elution buffer pH

7. Fractions (0.5 ml) were collected, analyzed by SDS-PAGE and the ones containing the largest amount of protein were dialyzed against appropriate buffers, which varied in different experiments.

2.4.2 Expression and Purification of the Barnyard TMP 70 kDa C-Terminal Fragment

Expression and purification were performed essentially as described for the RLF. Various expression strains of *E. coli* were tested; however, protein was eventually obtained from BL21-CodonPlus(DE3)-RP cells containing the plasmid pLAM35 that were induced for 4 hours with 0.8 mM IPTG prior to harvesting. Imidazole was omitted from the lysis buffer (the protein did not bind to the column in the presence of 10 mM imidazole), and lysozyme (0.5 mg/ml) was added to thawed pellets prior to the second round of freezing and thawing. Additionally, the fifth column wash used in the RLF purifications was omitted, and all samples from inductions and purifications were analyzed on 10% SDS-PAGE gels. Substantial amounts of protein were obtained when purified in this way; however, there were a great deal of contaminants, and/or the recombinant protein was subjected to considerable proteolytic degradation. Therefore, future purifications would need to be performed in the presence of protease inhibitors and possibly with slightly more stringent wash conditions.

2.4.3 Production of Polyclonal Antibodies to the Barnyard Rpf Motif

Expression of the Barnyard RLF was induced from pLAM5 in BL21(DE3)pLysS or BL21(DE3) and purified with TALON™ Metal Affinity Resin as described above (lysis buffer utilized in initial purifications was at pH 7). Fractions containing the largest amount of recombinant protein

were dialyzed against 500 ml of EB (50 mM NaH_2PO_4, 500 mM NaCl, 1 mM EDTA and 1 mM DTT) + 50% glycerol at 4ºC overnight, and the concentration of protein was determined using a Bradford Assay (Bio-Rad). Partially purified protein (~400 µg total, estimated 80% purity) was run on a large tris-tricine gel to separate the recombinant protein from co-purifying contaminants, and the RLF bands were excised. These were soaked in dH_20 + 20% ethanol and sent to Pocono Rabbit Farm and Laboratory, Inc. (PRFL) for polyclonal antibody production in rabbits. Pre-bleeds from three rabbits were screened for cross-reactivity prior to injection. Additional protein (~260 µg and ~150 µg) was prepared in the same way and sent for boost injections. Dialyzed protein, 0.8-1 mg and later ~2.1 mg in phosphate-buffered saline (PBS) was also sent for affinity purification.

2.4.4 Immunoblot Analysis

Protein samples were separated by Laemmli SDS-PAGE (Laemmli, 1970), using the percentage of acrylamide giving the best resolution of target proteins; 7% gels were used for large (>70 kDa) proteins, 10%-12% gels were used for mid-sized proteins, and 15% gels were used for small proteins (<20 kDa). In some cases, high-resolution (4%, 10% and 16.5%) tris-tricine gels were utilized for accurate separation of small proteins. Proteins were transferred to Sequi-Blot PVDF membrane (Bio-Rad) using a Bio-Rad Trans-Blot SD Semi-Dry Transfer Cell. Both the SDS-gel and the membrane were equilibrated in transfer buffer for 20 minutes prior to transfer. Proteins were transferred for 90 minutes at (area of gel(s) (cm^2) × 0.8) mA. Following transfer, PVDF membranes were air-dried and when necessary, stored in saran wrap overnight.

Membranes were re-wetted with methanol, rinsed with dH_2O, and blocked with 5% non-fat dry milk in TBS-T for 45 minutes to 1 hour. Primary antibody was then added in 5% or 1% milk in

TBS-T for ~1 hour. Antibody dilutions are noted for individual experiments; α-RLF pre-bleeds were used at an optimum dilution of 1:5000, and affinity purified α-RLF was typically used at a dilution of 1:125. Blots were then washed three times with TBS-T for 5 minutes each time, and secondary antibody (horseradish peroxidase-conjugated goat anti-rabbit IgG, Sigma) was added in 0.5% milk in TBS-T at a concentration of 1:10,000 for 45 minutes – 1 hour. Membranes were washed again with TBS-T, three times for 5 minutes each time. The chemiluminescent substrate (Western Lightning™ Chemiluminescence Reagent Plus, PerkinElmer) was then prepared, and individual membranes were incubated with substrate in the dark for 1-2 minutes. Membranes were wrapped in saran wrap and exposed to film (Kodak BioMax Light) for various lengths of time.

2.4.5 *M. smegmatis* Cell Fractionation

These experiments were performed essentially as described by Piuri & Hatfull, 2006. *M. smegmatis* cells (in 200 ml 7H9 CB supplemented with 1 mM $CaCl_2$ and 4% glucose) were subjected to high M.O.I. (15 and 30 were both attempted with similar results) infections with mycobacteriophage Barnyard for various lengths of time prior to sonication and differential centrifugation. Tween 80 (0.05%) was added after a 30-minute adsorption to minimize clumping, and at each time-point, 80 ml of cells were collected and washed with PBS (in later experiments, washes were with dH_20 containing 10% glycerol and 0.05% Tween 80 to further minimize clumping). These samples were resuspended in 3 ml of PBS with 1 mM PMSF and subjected to about six rounds of sonication (45 second bursts with 1 minute rests on ice). The insoluble material and unbroken cells were removed by centrifugation at 11,000 × g for 10 minutes at 4°C; cell wall material was collected by centrifugation at 27,000 × g for 1 hour at 4°C; cell membranes were

collected by ultracentrifugation in an 80Ti type rotor at 100,000 × g for 3-4 hours at 4°C; and the remaining supernatant was considered to be cytoplasmic material. All pellets were resuspended in 100-300 µl PBS with 1 mM PMSF, and samples were quantified using the BioRad Bradford Assay. Equivalent amounts of each sample were separated by SDS-PAGE on 7% polyacrylamide gels and assayed to immunoblot analysis as described above.

2.5 BACTERIOPHAGE PREPARATION AND ANALYSIS

2.5.1 Preparation of Bacteriophage Stocks

High-titer bacteriophage stocks and plate lysates were essentially prepared as described by Sarkis and Hatfull (Sarkis & Hatfull, 1998). Barnyard infections were performed with exponentially growing *M. smegmatis* cells at an OD_{600} of 0.9; although phenotypic characterization later revealed that cells from a fresh, saturated culture diluted to the same optical density would have likely worked as well. Bacteria were diluted to an OD_{600} of 0.1-0.2 in 7H9 with 1 mM $CaCl_2$ and infected with approximately 10^6-10^7 plaque forming units (pfu) of phage per large plate. This is higher than used for most phages, as Barnyard makes extremely small, often turbid plaques. Phage were adsorbed for 30 – 45 minutes at room temperature and plated as top agar lawns in 0.35% mycobacterial top agar (MBTA) with 1 mM $CaCl_2$. Plates were incubated at 37°C for two to three days to allow the low-density lawns to fully grow. Plates were then flooded with phage buffer, and the lysate was cleared by centrifugation at 3500 × g.

For initial phage preparations, the phage was precipitated with 10% polyethylene glycol (PEG) 8000 (Sigma) and 1 M NaCl, overnight with continual stirring at 4°C. Precipitated phage

were pelleted at 3500 × g for 10 minutes at 4°C and resuspended in ~10.5 ml phage buffer for 1 – 4 hours at 4°C. This was brought to a density of 1.5 mg/ml with CsCl (Fisher Scientific) and pelleted again at 3500 × g to clear, if necessary. The solution was transferred to a 14 ml polyallomer quick-seal centrifuge tube (Beckman) and topped off with phage buffer that had been adjusted to a density of 1.5 g/ml. The sealed tube was then subjected to ultracentrifugation at 38,000 × g in a type Ti65 rotor for 16 hours at 10°C. After puncturing the top of sealed tubes, opaque bands containing the concentrated phage were removed with a 1-3 ml syringe and an 18G or larger needled and dialyzed into phage buffer before use.

Alternatively, cleared lysate from initial plates was subjected to ultracentrifugation in a type Ti45 rotor tubes (elephant foot rotor), ~60 ml per tube at 20,000 × g for 1.5 hours. Phage pellets were then resuspended overnight at 4°C in 10 ml phage buffer and purified using a continuous CsCl gradient as described above. In some cases, a CsCl step-gradient was utilized to increase the final yield. For this, phage pellets were resuspended in 8 ml phage buffer and transferred to an appropriately sized centrifuge tube. Using a Pasteur pipette, 1 ml of 10% glycerol was then added to the bottom of the tube, below the phage. Next, 1.5 ml phage buffer brought to a density of 1.4 g/ml was added in the same way, under the glycerol. Finally, 1.5 ml of phage buffer brought to a density of 1.6 g/ml was added in the same way to the very bottom of the gradient. This was subjected to ultracentrifugation in a swinging-bucket type SW41 rotor at 30,000 × g for 1.5 hours. The band containing the phage was extracted with a syringe.

2.5.2 Bacteriophage DNA Isolation

Bacteriophage DNA was isolated essentially according to Sarkis and Hatfull (Sarkis & Hatfull, 1998). Briefly, CsCl purified bacteriophage were dialyzed into phage buffer 4°C (two times against 500 ml). This was extracted with an equal volume of buffer-equilibrated phenol (Invitrogen) and pelleted for 5 minutes. The aqueous phase was removed, and this was repeated until the white interface disappeared. The phenolic phases were simultaneously back-extracted with TE to increase the yield. The sample was then extracted once with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) and a final time with an equal volume of chloroform. The DNA was precipitated by adding 1/10 volume of 3 M sodium acetate (NaOAc) and 2.5 volumes 95% ethanol (EtOH); this was then frozen on dry ice and pelleted at 14,000 × g at 4°C for 30 minutes. Pellet was washed with 1 ml 70% EtOH, air dried and resuspended in TE.

2.5.3 Phage Protein Analysis

Barnyard proteins were initially prepared as described previously by Sarkis and Hatfull (Sarkis & Hatfull, 1998). Briefly, proteins were precipitated from CsCl-banded virions dialyzed into phage buffer with 20% TCA. Pellets were washed with acetone and resuspended in 2X SDS-sample buffer. These samples were then boiled for 5 minutes, separated by SDS-PAGE on 10% or 12% denaturing gels and visualized by staining with Coomassie Blue for 15 minutes. This method resulted in considerable protein degradation, so the protocol described below was adopted.

CsCl-banded virions were dialyzed into phage buffer and subjected to three rounds of freezing on dry ice and thawing on ice. This sample was mixed with 4X SDS-sample buffer and

heated to 95°C for 1-3 minutes. Proteins were separated by SDS-PAGE on 10% or 12% denaturing gels and visualized by staining with Coomassie Blue for 15 minutes.

2.5.4 N-Terminal Sequencing

Phage proteins were prepared by the freeze/thaw method and separated by SDS-PAGE as described above. Proteins were transferred to Sequi-Blot PVDF membrane (Bio-Rad) as described for immunoblot analysis, except that following transfer, membranes were stained with Coomassie Blue for 5 minutes and briefly destained with methanol. Protein bands were excised from air-dried membranes and sequenced using automated Edman degradation chemistry by the University of Pittsburgh Protein Microanalytical Laboratory.

2.5.5 Mass Spectrometry

Phage proteins were prepared by the freeze/thaw method and separated by SDS-PAGE as described above, except that SDS-gels were stained with Colloidal Coomassie (Invitrogen) according to the manufacturer's instructions. Bands were excised and sequenced using NanoLC-MS/MS peptide sequencing technology (ProTech, Inc.). Briefly, protein bands were in-gel digested with sequencing grade modified Trypsin (Promega). Peptide mixtures were then analyzed by a LC-MS/MS system (Thermo), consisting of a high-pressure liquid chromatography (HPLC) with a 75 μm inner diameter reverse phage C18 column on-line coupled to an ion trap mass spectrometer. The spectrometric data for Barnyard proteins were analyzed against the GenBank protein database, and the data for Giles were analyzed against a database consisting of all predicted Giles ORFs. In

both cases the output was manually verified. When not indicated, all other chemicals used in proteolytic digestion and HPLC were obtained from Sigma.

2.5.6 Mycobacteriophage Barnyard Liquid Infections

An overnight culture of *M. smegmatis* was grown overnight at 37°C, shaking at ~250 rpm; typically 25 ml cultures grown in 125 ml baffled flasks were used. When the OD_{600} reached approximately 1.00-1.25, the overnight was diluted to an OD_{600} of ~0.1 in 7H9 supplemented with 4% glucose, CB, and 1 mM $CaCl_2$. In some experiments, cells were washed with dH_2O or phage buffer plus 10% glycerol. For RT-PCR experiments, overnight cultures at an OD_{600} of 1.5-2.0 were used, and cells were not washed cells prior to dilution. Barnyard was added to infection cultures at the desired multiplicity of infection (M.O.I.), which varied depending on the experiment. A culture at an OD_{600} of 0.1 was estimated to contain 3.5×10^7 cfu/ml. Phage were adsorbed at room temperature for 30 minutes, and 20% Tween was then added to final concentration of 0.025% to prevent cells from clumping. Cultures were incubated at 37°C and agitated at 250-300 rpm for desired length of infection. The OD_{600} was monitored throughout the course of the infection, and time points were taken, as necessary. For RT-PCR analyses, 10 ml samples were removed at each time point; cells were pelleted and immediately frozen at -80°C.

2.5.7 RNA Isolation

RNA was isolated from infected cells using the RiboPure-Bacteria Kit (Ambion), according to manufacturer's instructions. Cells were pelleted, resuspended in the RNAwiz solution and disrupted with ice-cold Zirconia beads using a vortex adaptor for a total of 8 minutes at maximum

speed, in 30 second intervals. Bacterial lysate was extracted with chloroform, diluted with ethanol and passed over the silica filter. RNA bound to the filter was washed a number of times to remove any contaminants and eluted with elution solution pre-warmed to 95°C. RNAseOUT Ribonuclease Inhibitor (Invitrogen) (1 μl) was added to 100 μl purified RNA samples. DNAseI treatment (supplied with the kit) was performed according to supplied instructions. Purified RNA samples were generally checked on an agarose gel and quantified with a Nano-drop apparatus.

2.5.8 RT-PCR

RT-PCR was performed using the QIAGEN OneStep RT-PCR Kit, according to manufacturer's instructions. Primer pairs LJM52/53, LJM54/55 and LJM56/67 were used to amplify three regions of the Barnyard tape measure transcript. Primers were used at a concentration of 6 μM, and 50 ng of RNA was added to each reaction. In some cases, the number of cycles was optimized in order to give non-saturating amounts of product and allow quantification.

2.6 MICROBIOLOGY

2.6.1 Induction of pBAD/gIII Derivative Constructs in *E. coli*

Plasmids were verified by sequencing and restriction digest and transformed into *E. coli* Top 10 competent cells. Strains were grown in LB CB overnight and subcultured 1:100 into fresh LB CB in the morning. These were grown to an OD_{600} of 0.4-0.6, and expression was induced by adding

2% arabinose to a final concentration of 0.02%. The optical density was monitored over time and plotted on a semi-log curve.

2.6.2 Induction of pLAM12 Derivative Constructs in *M. smegmatis*

To induce expression from $P_{acetamidase}$ on solid media, succinate and acetamide were added to 7H10 agar at a final concentration of 0.2%, and ADC was omitted. Control plates were prepared identically, except that inducer (acetamide) was not added.

For inductions in liquid media, overnight cultures of the strains to be assayed were inoculated into fresh 7H9 CB CHX KAN supplemented with 10% ADC at dilutions of 1:100 to 1:200 and agitated at ~250 rpm overnight at 37°C. Typically, 25 ml cultures grown in 125 ml baffled flasks were used. When cultures reached an OD_{600} of 1.0-1.5, cells were pelleted at 5000 rpm for 10 minutes and washed with 1/2 volume 7H9 containing 0.05% Tween 80 or (base induction medium). Cells were spun a second time and resuspended to a final OD_{600} of ~0.1 in 50 ml base induction medium (7H9 CB [sometimes omitted] KAN supplemented with 0.2% succinate and 0.05% Tween). One-half of each resuspended culture was transferred to a fresh sterile flask, and sterile acetamide was added to a final concentration of 0.2% (an equal volume of sterile water was added to the other to keep the volume consistent). Induced and non-induced cultures were shaken at 200-250 rpm at 37°C for 24 hours, and the OD_{600} was measured every 2 hours for 8 hours. Periodically, viability was measured by serially diluting cells in 7H9 with 10% ADC and plating on 7H10 CB CHX KAN agar with 10% ADC. In some experiments, cells were washed in 7H9 containing 10% ADC prior to dilution and plating. Also, in some experiments, larger induction cultures were used; 10 ml samples were removed periodically, washed with 10% glycerol, and subjected to immunoblot analysis.

In other experiments, the protocol was modified as described below. Strains to be assayed were inoculated into 25 ml 7H9 induction medium (in 125 ml baffled flasks) at an OD_{600} of ~0.02 and agitated at ~250 rpm overnight (~17 hours) at 37°C. When cultures reached an OD_{600} of ~0.5, samples were removed for viability determination, and the cultures were induced by adding acetamide to a final concentration of 0.2%. Induced cultures were shaken at 200-250 rpm at 37°C for 24 hours, and the OD_{600} was measured every 2 hours for 8 hours. Periodically, viability was measured by serially diluting cells in 7H9 containing 10% ADC and plating on 7H10 CB CHX KAN agar with 10% ADC.

2.7 RECOMBINEERING IN *M. SMEGMATIS* TO GENERATE BACTERIOPHAGE GILES TMP MUTANTS

2.7.1 Mycobacteriophage Mutant Construction and PCR Screening

In order to generate the 200 bp double-stranded DNA (dsDNA) deletion substrate, a 100 nucleotide (nt) PAGE-purified 'deletion oligonucleotide (oligo)' was synthesized (IDT), containing 50 nt of homology upstream and downstream of the region to be deleted. This was extended by PCR (100 µl reactions) with two 75 nt PAGE-purified 'extender primers', which add 50 bp of homology to either end of the deletion oligo. The 200 bp dsDNA product was checked on a 1% agarose gel, and the reaction was cleaned using the QIAquick PCR Purification Kit (QIAGEN) and eluted off the column with dH_20. In order to obtain concentrated deletion substrate, multiple PCR reactions were often purified over one column.

Electrocompetent cells of the recombineering strain (mc^2155:pJV53) were transformed with 50-100 ng phage DNA and 50-500 ng of the deletion substrate. Reactions were recovered for ~2 hours in 7H9 with 10% ADC and 1 mM $CaCl_2$ and plated as top agar lawns with ~300 µl fresh *M. smegmatis* cells in 0.35% molten MBTA with 1 mM $CaCl_2$. Resulting plaques were picked into 100 µl phage buffer, and 1 µl of these were typically screened by PCR (as described previously, using 20 µl reactions) with 'diagnostic primers' (25-35 bp) that anneal upstream and downstream of the deletion. In some cases, the highly sensitive mismatch amplification mutation assay (MAMA)-PCR (Swaminathan *et al.*, 2001) using Platinum Taq High Fidelity DNA Polymerase (Invitrogen) was employed to screen for the mutation (25 µl reactions supplemented with 2 mM $MgSO_4$ to increase fidelity of the polymerase). In this case, the upstream primer was designed such that its 3′ end anneals over the new junction created by the deletion. This creates a mismatch with the wild-type sequence and thus, favors the formation of the mutant product. Plaques containing a mixture of deletion and wild-type DNA were serially diluted, and ~300 µl fresh *M. smegmatis* was infected with 10 µl of a 10^{-3}, 10^{-4} and 10^{-5} dilution. Individual plaques from the 10^{-4} and 10^{-5} plates and a lysate from the 10^{-3} or 10^{-4} plate were then screened for the presence of the deletion with diagnostic primers or by MAMA-PCR as described above.

2.7.2 Infectivity Assay with Mutant Phages

A culture of *M. smegmatis* (60 ml in 250 ml baffled flasks) was grown in 7H9 CB CHX supplemented with 10% ADC and 1 mM $CaCl_2$, and this was sampled at various time-points as it transitioned from exponential growth into early- and then late-stationary phase. At each time-point, 300 µl of cells were infected with multiple dilutions (in phage buffer) of wild-type Giles and the

mutant phages. Phage were adsorbed at room temperature for 30 minutes, and the cells were plated as top agar overlays in molten 0.35% MBTA with 1 mM $CaCl_2$. The same dilutions were also utilized to infect cells from an exponentially growing *M. smegmatis* culture, and the titers obtained on the aged culture were averaged and normalized against these values.

2.8 BUFFERS AND REAGENTS

Tris-EDTA (TE):

10 mM Tris-HCl, pH 7.5; 1.25 mM EDTA

Phage Buffer:

10 mM Tris-HCl, pH 7.5; 10 mM MgSO4; 68.5 mM NaCl; 1 mM $CaCl_2$

4X SDS Sample Buffer:

62.5 mM Tris-HCl, pH 6.8; 20% glycerol; 2% (w/v) SDS; 5% 2-mercaptoethanol; 0.1% (w/v) bromophenol blue

Laemmli 4X Separating Buffer:

1.5 M Tris-HCl, pH 8.8; 0.4% (w/v) SDS

Laemmli 4X Stacking Buffer:

0.5 M Tris-HCl, pH 6.8; 0.4% (w/v) SDS

Protein Gel Running Buffer:

25 mM Tris base; 192 mM glycine; 0.035% (w/v) SDS; pH 8.3

Standard SDS-PAGE, run according to Laemmli (Laemmli, 1970):

Separating gel: 1X Separating buffer; 7-15% acrylamide:bis-acrylamide (29:1); 50 µl 10% ammonium persulfate (APS) and 10 µl tetramethylethylenediamine (TEMED) (10 ml)

Stacking gel: 1X Stacking buffer; 4% acrylamide:bis-acrylamide (29:1); 50 µl 10% APS and 10 µl TEMED (10 ml)

High Resolution Tris-Tricine Gels: for separation of low molecular weight polypeptides (Schagger & von Jagow, 1987).

Separating Layer: 1 M Tris-HCl, pH 8.45; 0.1% SDS (w/v); 13.3% glycerol; 16.5% acrylamide:bisacrylamide (29:1); 225 µl 10% APS and 22.5 µl TEMED (45 ml)

Spacer Layer: 1 M Tris-HCl, pH 8.45; 0.1% SDS (w/v); 10% acrylamide:bisacrylamide (29:1); 60 µl 10% APS and 6 µl TEMED (6 ml)

Stacking Layer: 1 M Tris-HCl, pH 8.45; 0.1% SDS (w/v); 4% acrylamide:bisacrylamide (29:1); 100 µl 10% APS and 20 µl TEMED (12.5 ml)

Anode Buffer for Tris-Tricine Gels:

0.2 M Tris-HCl, pH 8.9

Cathode Buffer for Tris-Tricine Gels:

0.1 M Tris base; 0.1 M Tricine; 0.1% w/v SDS; pH 8.25

2X Tris-Tricine Sample Buffer:

0.1 M Tris-HCl, pH 6.8; 24% glycerol; 8% (w/v) SDS; 0.2 M dithiothreitol (DTT); 0.02% (w/v) Coomassie Blue R-250

1X TBS-T:

25 mM Tris-HCl, pH 8; 125 mM NaCl; 0.1% Tween 20

Transfer Buffer:

48 mM Tris base; 39 mM glycine; 0.037% (w/v) SDS; 20% methanol

1X Lysis Buffer:

50 mM NaH_2PO_4; 300 mM NaCl; pH 8

1X Elution Buffer:

50 mM NaH_2PO_4; 300 mM NaCl; 150 mM imidazole; pH 7

3.0 GLOBAL ANALYSIS OF MYCOBACTERIOPHAGE TAPE MEASURES AND TAPE MEASURE MOTIFS

3.1 INTRODUCTION

A comparison of the bacteriophage tape measures that have been described reveals a number of characteristics shared by this group of proteins. Notably, despite the fact that TMPs often display only limited sequence similarity with one another, structurally these are predicted to be quite similar, with a largely α-helical secondary structure containing multiple regions with the potential to form coiled-coil interactions (Katsura & Hendrix, 1984, Xu, 2001, Ph.D. Thesis). In the case of bacteriophage λ, it has been noted that while the N-terminus of gpH is rich in α-helical content, the C-terminus is not and rather, contains a large proportion of glycine residues. The reason for this is not known; however, it may be significant that the C-terminal region of the protein is required for tail initiator formation and is thought to interact with other tail tip proteins. Interestingly, the same structural features have been noted for the TMP-like tail fiber protein pb2 in phage T5 (Boulanger *et al*., 2008), and like gpH, this protein is proteolytically cleaved at the C-terminus during the process of tail assembly (Zweig & Cummings, 1973a).

Tape measure proteins are also often predicted to contain membrane-spanning domains, although the significance of this remains to be determined as well. An attractive hypothesis, based on the location of these proteins and on the DNA injection phenotypes of various gpH mutants, is

that tape measures might insert into the bacterial membrane during infection to form the pore for DNA injection. Currently, there is little evidence to support this for most of these proteins; however, for phage λ, there are data suggesting that gpH* may be involved in DNA injection and can at least insert into liposomes containing the λ receptor (Scandella & Arber, 1976, Roessner & Ihler, 1984). Strikingly, T5 pb2 is also predicted to contain multiple membrane-spanning domains, and in this case, it has been shown that this is a fusogenic protein that inserts into the bacterial membrane during infection and *in vitro* (Guihard *et al*., 1992, Lambert *et al*., 1998, Boulanger *et al*., 2008). The large number of sequenced mycobacteriophages provides us with the opportunity to determine if and how these structural characteristics are shared by members of a diverse population of phages that infect a common host. In order to answer this question, I have, therefore, subjected these phages to various structural prediction algorithms, and the data are presented in the first part of this chapter.

An additional benefit of comparative genomic studies of large populations of related entities – such as the mycobacteriophages – is that they can sometimes reveal features that are not necessarily apparent from the study of one or two individuals. One example of this was the discovery of the three sequence motifs with identity to host proteins embedded within a number of mycobacteriophage TMPs (Pedulla *et al*., 2003). The observation that the same motif was often found in otherwise dissimilar tape measures, combined with the fact that these were present in eight of the 14 sequenced phages, provided evidence that, rather than just representing the remnants of an ancient illegitimate recombination event, these might actually be serving an important function for the phage. The nature of this function was, however, not immediately obvious.

One of these motifs, found in mycobacteriophage Barnyard, was observed to have similarity with bacterial proteins that appeared to function as microbial growth factors, the Rpfs. This suggested that phages had perhaps co-opted a host signaling function, in this case growth stimulation, for its own purposes, which might be advantageous when attempting to infect dormant, or non-growing bacteria (Pedulla *et al*., 2003). The other motifs, however, had similarity with proteins of unknown function, making it difficult to determine if they might serve the same purpose. It has subsequently been shown that the Rpf proteins structurally resemble lytic transglycosylase enzymes and display muralytic activity both *in vitro* and *in vivo*, and this activity may underlie the resuscitation promoting properties of these proteins (Cohen-Gonsaud *et al*., 2005, Mukamolova *et al*., 2006). Further, the *M. smegmatis* proteins containing regions of identity with the tape measure Motif 3 have murein hydrolase activity *in vitro*, and when the phage motif is inserted in its place, the hybrid protein retains function (Piuri & Hatfull, 2006). Thus, although the Motif 2 homologues have yet to be characterized, the data suggest that the tape measure motifs may in fact, function to perform localized peptidolgycan hydrolysis during host infection. This activity has been found to be associated with a number of phage structural proteins; however, the mycobacteriophages appear to represent the first demonstration of such an activity associated with the TMP.

In mycobacteriophage TM4, Motif 3 is non-essential; however, phage lacking this motif display reduced ability to infect stationary phase cells (Piuri & Hatfull, 2006). Although this corresponds to only a 50% reduction in plaquing efficiency, such a defect might place a bacteriophage and its progeny at a severe disadvantage in nature, where most, if not all of the host cells are expected to be in a non-growing, stationary phase-like state. For example, starting with a population of 200 phages, half of which contain a Motif 3-like activity and half of which do not,

and assuming a burst size of 100 for both phages, it would take only 10 rounds of infection for the 'mutant' phages to be out-numbered ~10,000 to one.

Calculations such as these suggest that all phages are likely to have similar virion-associated hydrolytic activities. In accordance with this, such an activity has been found to be associated with the virion structures of a number of diverse phages, and even more are predicted to contain such an activity by bioinformatics (Moak & Molineux, 2004). There are now over 50 fully sequenced mycobacteriophage genomes, and further, the number of protein sequences in the database has increased substantially since the tape measure motifs were first detected. Therefore, I have re-examined the expanded, sequenced mycobacteriophage population and the protein database as a whole for the presence of these motifs, in an attempt to better characterize their distribution, and these data will be discussed in the second part of this chapter.

3.2 STRUCTURAL CHARACTERIZATION OF MYCOBACTERIOPHAGE TAPE MEASURES

3.2.1 Predicted Secondary Structure

There are various lines of evidence suggesting that the λ tape measure extends and measures the length of the tail as an extended α-helical protein (Katsura, 1990). This configuration is one of the two main secondary structural elements that make-up all native proteins, and these tightly packed structures can range in size from four to 40 amino acids. Residues within an α-helix face outward and are arranged in a right-handed 5.4 Å (0.54 nm) wide helical coil, which is stabilized by hydrogen bonding between each residue's amino groups and the carboxyl group of the amino acid

four residues in front of it, a bonding pattern that defines the α-helix. Each amino acid comprises a 100° turn in the helix, resulting in 3.6 residues per turn, and occupies a space of 1.5 Å (0.15 nm) along the helical axis, a property that is evident in the relationship between tape measure length and tail length (Katsura & Hendrix, 1984).

Multiple α-helices can also wind around each other to form a supercoil known as a coiled coil. These are stable, structural motifs involved in the oligomerization of a number of proteins, including motor proteins and cytoskeletal elements (Burkhard *et al.*, 2001). Regular coiled coil structures contain two to five amphipathic α-helices that wind around each other to form a left-handed supercoil. This is stabilized by the 'knobs-into-holes' packing of apolar side chain residues that face inward and insert into the hydrophobic core of the coiled coil (Crick, 1953). The amino acid sequences of these structures display a characteristic heptad repeat: the repeating pattern of a seven-amino acid motif, usually noted (abcdefg)n, with the (a) and (d) positions occupied by mostly hydrophobic residues. The structure can be further stabilized by electrostatic interactions between opposing charges on the fifth (e) and seventh (g) residues of the repeat. Tape measures are often predicted to contain multiple regions with the propensity to form coiled coils, and therefore, it is thought that the six molecules of this protein present within the phage tail are likely stabilized by such interactions.

Importantly, although detailed genetic and structural studies such as those performed on phage λ have not been undertaken for other tape measures, the levels of structural similarity predicted between gpH and the TMPs from other phages, such as the coliphage HK022, are suggestive of a common function, even when only limited sequence identity can be detected (Xu, 2001, Ph.D. Thesis). Therefore, in order to extend these analyses to the flexible-tailed mycobacteriophages, the amino acid sequences of their TMPs were subjected to the secondary

structure prediction algorithms (PSIPRED) (Jones, 1999, McGuffin *et al.*, 2000), (PredictProtein, PROF) (Rost & Sander, 1993, Rost *et al.*, 1996, Rost *et al.*, 2004) and (PredictProtein, COILS v2.2) (Lupas *et al.*, 1991, Rost *et al.*, 2004). The data are summarized in Table 3, and as expected, all of these proteins are predicted to be highly α-helical with multiple regions that are strongly predicted to form coiled-coil interactions. Notably, the percent α-helical content was observed to vary considerably in different groups of phages, which are clustered based on overall levels of relatedness (Hatfull *et al.*, 2006). For example, while the tape measures from Cluster A phages are >85% α-helical, those from Clusters E and F are significantly less so (30-50%). However, regardless of the percentage of predicted α-helical structure, the percentage of β–strands is always low, and therefore, the majority of residues not within an α-helix are predicted to form loops or random coil structure. Further, as has been noted previously for λ, in most cases, the N-termini of the mycobacteriophage TMPs are richer in α-helical content than the C-termini (this is quantified for Barnyard and Giles TMPs in Chapter 4).

Table 3. Secondary structure predictions of the mycobacteriophage TMPs.

Phage	*Tape Measure*	*Cluster*	*Size*	*α-helical Content*	*Predicted Coiled Coil Regions*	*Predicted Trans-membrane Domains*
Bethlehem	gp23	A	823	85.91%	3 (I & C)	6/7 or 7/6
Bxb1	gp22	A	823	91.37%	3 (I & C)	6/7 or 6/5
Bxz2	gp26	A	1008	95.63%	2 (C)	7/6
Che12	gp28	A	841	91.32%	2-3 (C)	7/6
D29	gp26	A	837	91.88%	2-3 (N & C)	7/6
DD5¤	gp25	A	826	93.1%	3 (N & C)	10/9 or 6/5
Jasper¤	gp25	A	826	95.52%	3 (N & C)	10/9 or 6/5
KBG¤	gp25	A	823	92.83%	3 (I & C)	6/7 or 8/7
L5	gp26	A	837	94.15%	2-3 (N & C)	8/7 or 7/6
Madison¤	gp23	A	823	96.11%	3 (I & C)	8/7
Pukovnik¤	gp28	A	798	90.98%	3-5 (I & C)	8/7 or 6/5
U2	gp23	A	822	96.23%	3 (I & C)	7 or 7/6
Cooper	gp25	B	2023	55.56%	4-5 (N & I)	12/11 or 5/4
Nigel¤	gp23	B	2026	65.3%	4-5 (N & I)	12/11 or 7/6
Orion	gp28	B	1991	64.79%	4 (N & I)	15/14 or 7/6
PG1	gp28	B	1932	57.66%	2-3 (I)	15/14 or 7/6
Phaedrus¤	gp26	B	1914	50.31%	2 (N & I)	17/16 or 7/6
Pipefish	gp28	B	1910	54.21%	1-2 (N & I)	17/16 or 7/6
Qyrzula	gp24	B	1882	58.61%	1-2 (N & I)	14/13 or 7/6
Rosebush	gp29	B	1882	59.25%	1-2 (N & I)	14/13 or 7/6
Adjutor¤	gp28	D	1590	68.49%	3 (N & I)	9/8 or 5/4
PBI1	gp27	D	1590	68.49%	3 (N & I)	9/8 or 5/4
Plot	gp28	D	1590	68.49%	3 (N & I)	9/8 or 5/4
244	gp24	E	1578	35.23%	2-3 (N & I)	8/7 or 5/4
Cjw1	gp22	E	1576	35.53%	3 (N & I)	8/7 or 3/2
Kostya¤	gp22	E	1577	38.24%	3-4 (N, I)	8/7 or 2/1
Porky¤	gp19	E	1578	35.61%	3 (N & I)	8/7 or 3/2
Boomer¤	gp15	F	1168	42.47%	2 (I)	8/7 or 1/0
Che8	gp14	F	1188	32.83%	2-4 (N & I)	12/11 or 2/1
Che9d	gp17	F	1183	46.91%	1 (N)	5/4 or 2/1
Llij	gp14	F	1170	37.52%	2-3 (I)	8/7 or 2/1
PMC	gp14	F	1170	37.52%	2-3 (I)	8/7 or 2/1
Tweety	gp14	F	1176	35.03%	2-3 (I)	8/7 or 1/0
BPs¤	gp16	G	1337	71.43%	2-3 (C)	14/13 or 3/2
Halo	gp16	G	1337	71.43%	2-3 (C)	14/13 or 3/2
Barnyard	gp33	H	2047	72.45%	3-5 (Dispersed)	10/9 or 7/6
Predator¤	gp25	H	2121	71.05%	4-6 (N & I)	12/11 or 6/5
Che9c	gp15		1213	40.07%	3-4 (N & I)	4/3 or **?**
Corndog	gp57		1461	37.3%	3-4 (N & I)	8/7 or 2/1
Giles	gp20		1359	62.1%	5-6 (N, I, C)	7/6 or 2/1
Omega	gp34		1599	39.4%	5-6 (N & I)	10/9 or 4/3
TM4	gp17		1229	67.37%	1-2 (I)	8/7 or 5/4
Wildcat	gp38		1983	45.39%	4-6 (N & I)	11/10 or 4/3

Ge neral location of coiled coil region: N = N-terminus, I = internal and C = C-terminus.

Num bers in first column are TMpred prediction and PHD prediction is listed second; in for both methods, the preferred prediction is given first.

¤Unpublished phage.

3.2.2 Putative Transmembrane Domains

It has also been observed that tape measures are often predicted to contain multiple transmembrane domains, although the significance and function of these have not been demonstrated for most phages (Roessner & Ihler, 1984, Pedersen *et al*., 2000, Zimmer *et al*., 2003). In order to determine if this feature is also shared by the mycobacteriophage tape measures, the sequences of these proteins were analyzed for possible membrane spanning regions using TMpred (Hofmann & W., 1993) and PredictProtein, PHD (Rost, 1996, Rost *et al.*, 1996, Rost *et al*., 2004). Strikingly, all of the TMPs from the mycobacteriophages with flexible, non-contractile tails are predicted to contain multiple transmembrane domains, although the number varies depending on the prediction method used. Importantly, for most proteins, multiple high-scoring regions, well above the value considered to be statistically significant, were found. Therefore, although membrane insertion has yet to be demonstrated for most phage tape measures, these data provide the argument that such an activity might be possible.

3.2.3 Amino Acid Composition

Different amino acid residues have varying propensities for forming α-helical structures. For example, alanine, glutamate, leucine, lysine, and methionine all have especially high helix-forming propensities; whereas, proline, glycine, tyrosine and serine are unlikely to be found in these structures. Proline tends to break or kink helices because it lacks an amide hydrogen and therefore, cannot form a hydrogen bond. Further, its sidechain sterically interferes with the backbone conformation. However, proline is often seen as the first residue of a helix, possibly due to its structural rigidity. Glycine, which also disrupts α-helices, has a high conformational flexibility,

and thus, its incorporation into the constrained α-helical structure is entropically unfavorable. These observations suggest that proteins adopting the same structural conformations are likely to contain similar amino acid content, even if they are unrelated at the sequence level. This property was therefore examined for the mycobacteriophage tape measures, and the data are presented in Figure 8. Importantly, only one protein from each cluster of related phages was included to eliminate any bias in the analysis due to the presence of multiple, highly similar proteins, so none of the proteins analyzed were more than 30% identical to one another at the sequence level. The TMPs from λ, TP901-1 and T5, as well as the T7 capsid/extensible tail protein that contains a transglycosylase motif, gp16, were also included for comparison.

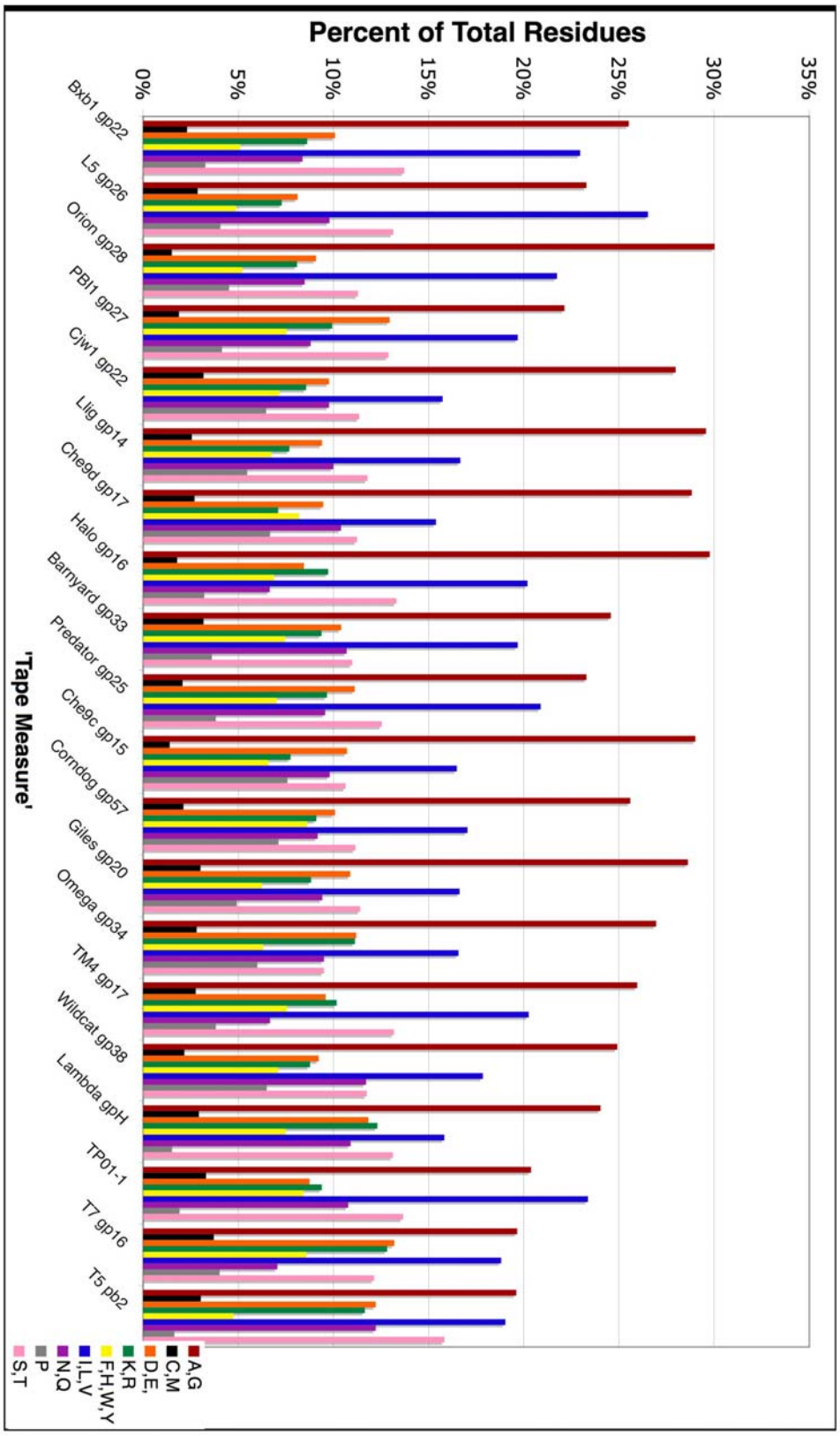


Figure 8. Amino acid composition of various tape measures and tape measure like proteins.

Figure 8: Amino acid composition of various tape measures and tape measure-like proteins. The frequency of occurrence for similar amino acids were calculated together: alanine (A) and glycine (G) are very small and polar; cysteine (C) and methionine (M) are hydrophobic and contain sulfur; glutamate (E) and aspartate (D) are negatively charged; lysine (K) and arginine (R) are positively charged; phenylalanine (F), histadine (H), tryptophan (W) and tyrosine (Y) are aromatic; isoleucine (I), leucine (L) and valine (V) are hydrophobic; asparagine (N) and glutamine (Q) are polar; serine (S) and threonine (T) are small and polar; and proline (P), due to its unique structure was considered separately.

Intriguingly, the pattern of amino acid usage is quite consistent for all of these proteins, and, for the most part, it is in accordance with what might be predicted for highly α-helical proteins. However, significant deviations are observed, and one example of this is the relatively high proportion of glycine residues, which are not favored in α-helices. The relatively high proportion of both glycine and alanine residues is a feature that has been noted previously for TMPs, and therefore, this may be important for some aspect of their function. Notably, these appear to cluster in the C-terminus of the protein, which, as has been noted for gpH, is generally predicted to be lower in α-helical content. For example, glycine residues make-up 8.8% of the residues in the N-terminal half of the Cjw1 TMP and 17.5% of the residues in the C-terminal portion. It is also noteworthy, that none of the TMPs from any of the flexible-tailed mycobacteriophages examined was found to contain more than one cysteine residue, including the coliphage T5 tail fiber pb2, which has zero. This may be due to the fact that aberrant intramolecular disulfide bonds within these proteins might be expected to interfere with the extended conformation that they must adopt within the phage tail. Analogously, T7 gp16, which is a capsid protein, must exit through the narrow head-tail connector (2.2 nm) prior to membrane insertion during infection, and therefore, it too must be able to freely adopt an extended conformation.

3.3 PREVALENCE OF THE TAPE MEASURE MOTIFS

Three distinct sequence motifs with similarity to host secreted proteins were identified in eight of the first 14 mycobacteriophages sequenced, suggesting these might be under positive selection within the population (Pedulla *et al.*, 2003). However, there are presently 51 fully sequenced mycobacteriophage genomes, and thus, we now have a much larger population from which to sample in order to determine if the same pattern is observed (Hatfull *et al*., 2006, Pham *et al.*, 2007, Morris *et al*., 2008, Hatfull Lab, unpublished data). The sequences of the TMPs from these phages were, therefore, subjected to BLAST analyses against the complete protein database. When possible motifs were detected, these were individually examined and compared against the database as well as with other members of the motif family. Strikingly, it was observed that 38 out of the 51 completed mycobacteriophage genomes encode TMPs containing at least one of these motifs. Further, these analyses have revealed the presence of two additional sequence motifs not detected in the tape measures of the original 14 phages. The results are summarized below and in Figure 9.

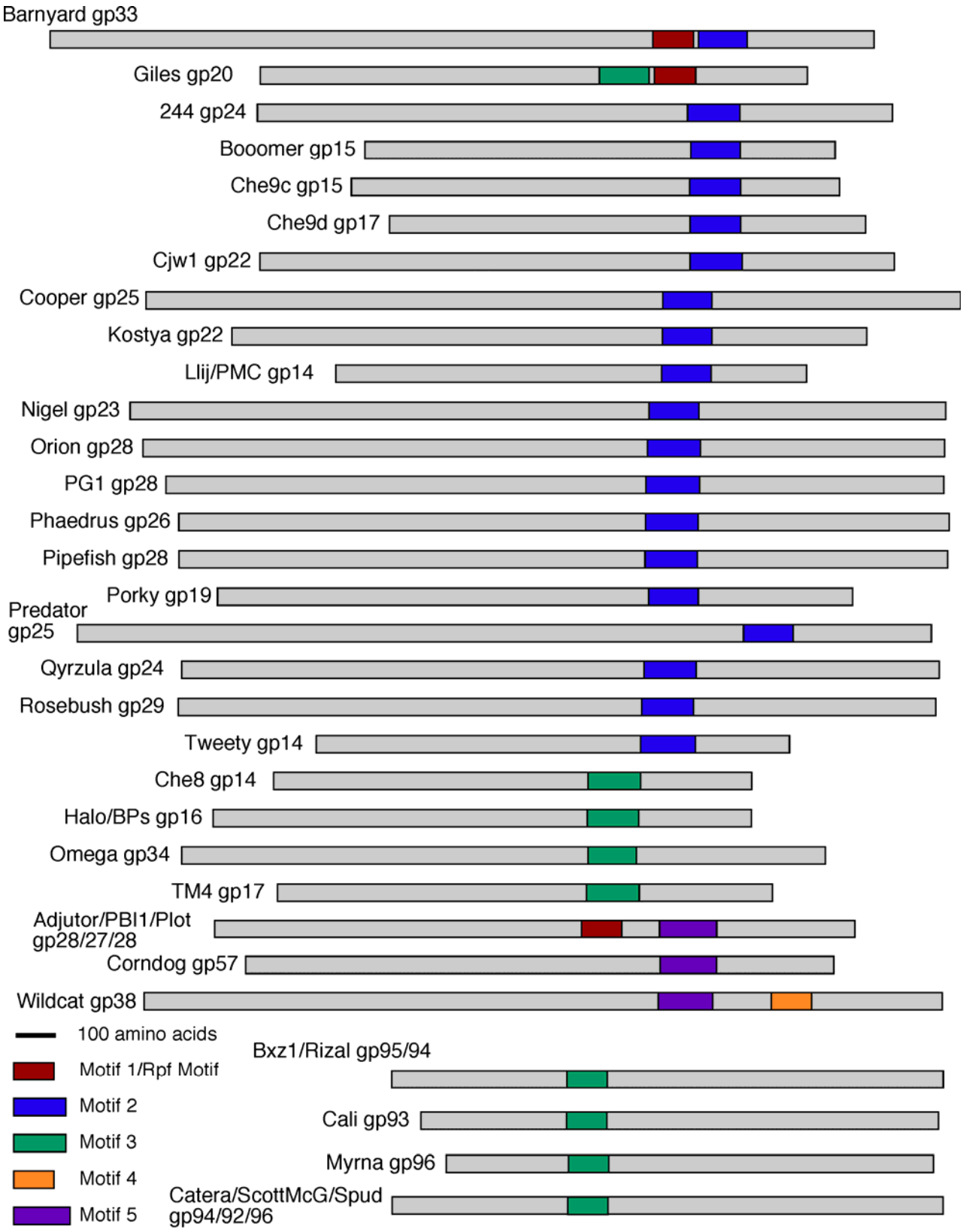


Figure 9. Prevalence of the mycobacteriophage tape measure motifs.

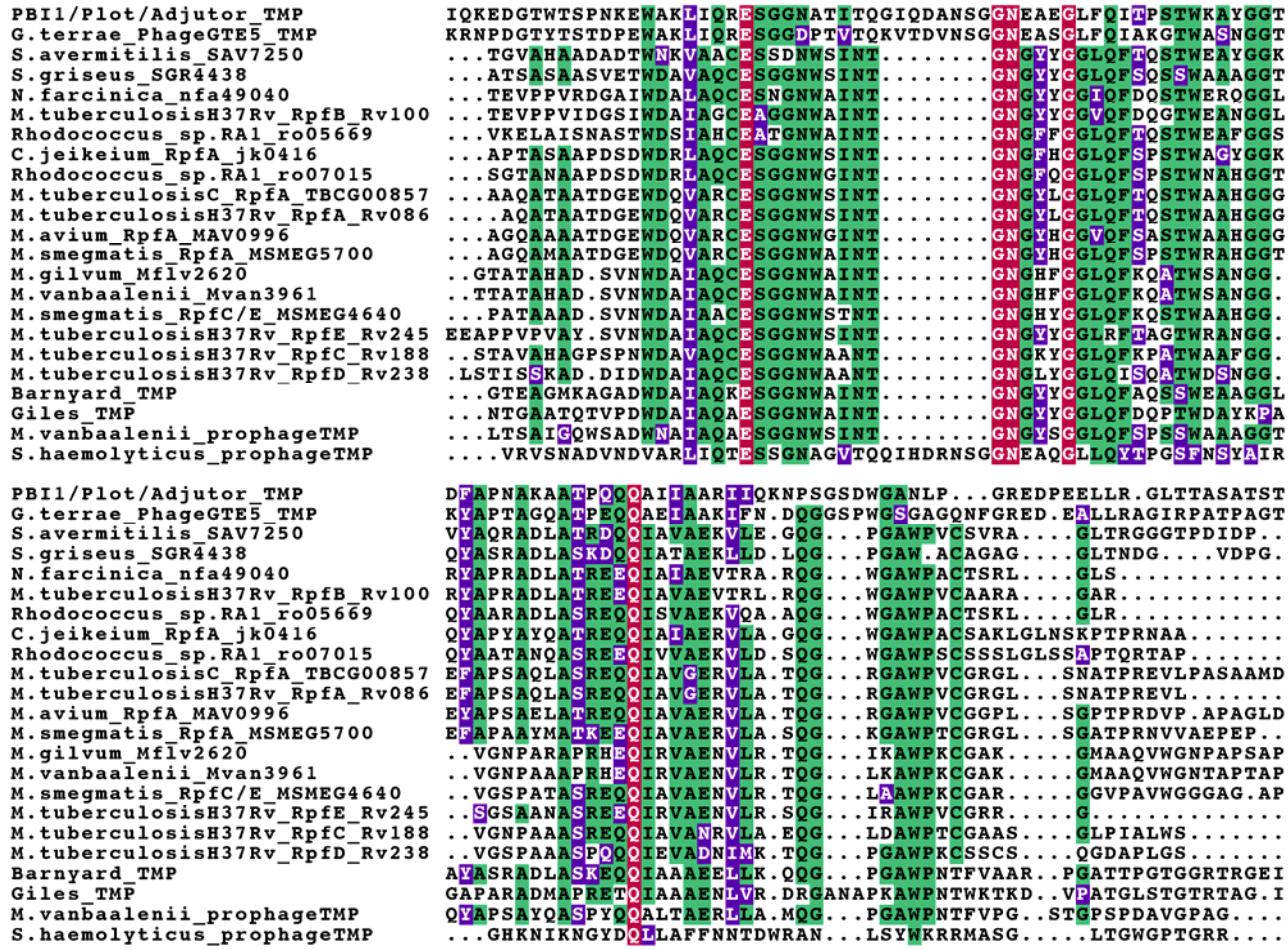
Figure 9: Prevalence of the mycobacteriophage tape measure motifs. At least one of the five tape measure motifs is present in 38 of the 51 fully sequenced mycobacteriophage genomes. Only the Cluster A phages encode TMPs lacking one of these motifs. Proteins >99.9% identical were drawn once as indicated. Contractile-tailed phages contain Motif 3 and are grouped together at the bottom.

3.3.1 The Rpf Motif

The examination of additional mycobacteriophage tape measures did not reveal the 'canonical' Rpf Motif in any of the newly characterized and sequenced phages; however, a more divergent version of this motif was uncovered in one group of related phages, and these are described below. The Barnyard and Giles TMPs contain an easily identifiable Rpf Motif, and these regions are closely related, with 56% identity over an ~100 residue span. BLAST analysis performed with the Barnyard Rpf Motif yields over 100 significant hits (E $<10^{-5}$) to proteins from both the mycobacteria and other high G+C% Gram-positives, such as *Streptomyces* and *Rhodococcus*. The best match is to the TMP of a *Mycobacterium vanbaalenii* prophage (E = 2×10^{-23}, 69% identity). Other high scoring matches include a hypothetical protein in *Nocardia farcinica* (E = 8×10^{-23}, 66% identity), a transglycosylase domain in a protein annotated as a peptidase from *S. avermitilis* (E = 1×10^{-22}, 62% identity), the RpfB proteins from multiple strains of *M. tuberculosis* (for strain H37Rv, E = 7×10^{-22} and identity is 61%) and the Rpf domain from mycobacteriophage Giles (E = 1×10^{-21}). A BLAST search with the Giles Rpf domain yields similar results: the Rpf domain from Barnyard TMP is the highest scoring hit (E = 8×10^{-22}). Other high scoring matches include putative secreted transglycosylases from *Mycobacterium gilvum* (E = 4×10^{-17}, 61% identity) and *M. vanbaalenii* (E = 7×10^{-17}, 55% identity), and the RpfC/E protein (MSMEG_4640) from *M. smegmatis* (E = 3×10^{-16}, 54% identity). Interestingly, the TMP from mycobacteriophage Predator (grouped with Barnyard in Cluster H) is almost 40% identical to the Barnyard TMP. However, although both of these contain Motif 2 (discussed below), the Predator TMP does not have the Rpf Motif.

PBI1, Plot and Adjutor (Cluster D) tape measures are virtually identical to one another and contain a region that has weak similarity to the Rpf Motif. This is a 70-80 residue segment that most closely matches a region from what is likely the tape measure protein from the *Gordonia terrae* phage GTE5 ($E = 7\times10^{-27}$, 54% identity). This is particularly interesting because this *Gordonia* protein also has a separate region with identity to Motif 3 (discussed below). The next best hits are to proteins from various species of *Corynebacterium,* some of which are annotated as Rpfs, including a likely TMP in a *Staphylococcus haemolyticus* prophage (SH1756) ($E = 5\times10^{-7}$, 43% identity) that also contains an M23 peptidase domain and the *M. smegmatis* RpfA protein (MSMEG_5700) ($E = 9\times10^{-7}$, 42% identity). Significant hits ($E < 10^{-5}$) are also obtained to proteins from *Kineococcus radiotolerans*, additional *Corynebacterium spp*, *Rhodococcus spp.,* and *N. farcinica,* as well as to the *M. vanbaalenii* prophage Rpf and the RpfA proteins from *Mycobacterium avium, M. avium* subsp. *paratuberculosis* and *M. tuberculosis* strain C. The relationship between the phage Rpf Motifs and some of their closest bacterial homologues is shown in Figure 10.

A.


PBI1/Plot/Adjutor_TMP
G.terrae_PhageGTE5_TMP
S.avermitilis_SAV7250
S.griseus_SGR4438
N.farcinica_nfa49040
M.tuberculosisH37Rv_RpfB_Rv100
Rhodococcus_sp.RA1_ro05669
C.jeikeium_RpfA_jk0416
Rhodococcus_sp.RA1_ro07015
M.tuberculosisC_RpfA_TBCG00857
M.tuberculosisH37Rv_RpfA_Rv086
M.avium_RpfA_MAV0996
M.smegmatis_RpfA_MSMEG5700
M.gilvum_Mflv2620
M.vanbaalenii_Mvan3961
M.smegmatis_RpfC/E_MSMEG4640
M.tuberculosisH37Rv_RpfE_Rv245
M.tuberculosisH37Rv_RpfC_Rv188
M.tuberculosisH37Rv_RpfD_Rv238
Barnyard_TMP
Giles_TMP
M.vanbaalenii_prophageTMP
S.haemolyticus_prophageTMP

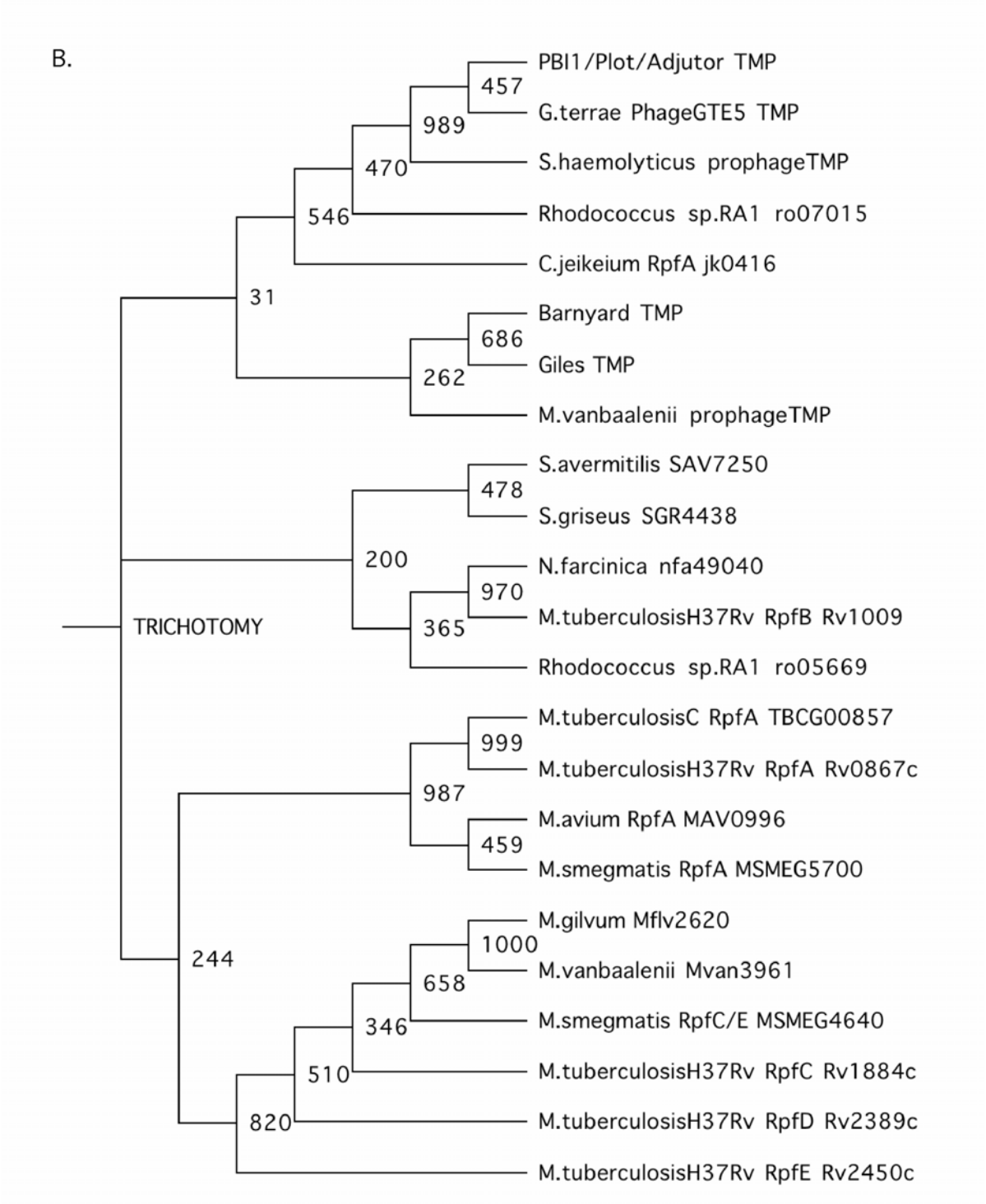


Figure 10. The Mycobacteriophage Rpf Motifs.

Figure10: The mycobacteriophage Rpf Motif. **A.** ClustalX sequence alignment of the mycobacteriophage Rpf Motifs and related bacterial proteins. 100% conserved residues are highlighted in red, identical residues are highlighted in green, and similar residues are highlighted in purple. **B.** Phylogenetic tree displaying the relationships among the various TMP Rpf Motifs and Rpf-like proteins. Neighbor-joining, un-rooted tree was drawn correcting for multiple substitutions, and confidence values displayed are from 1000 bootstrap trials.

3.3.2 Motif 2

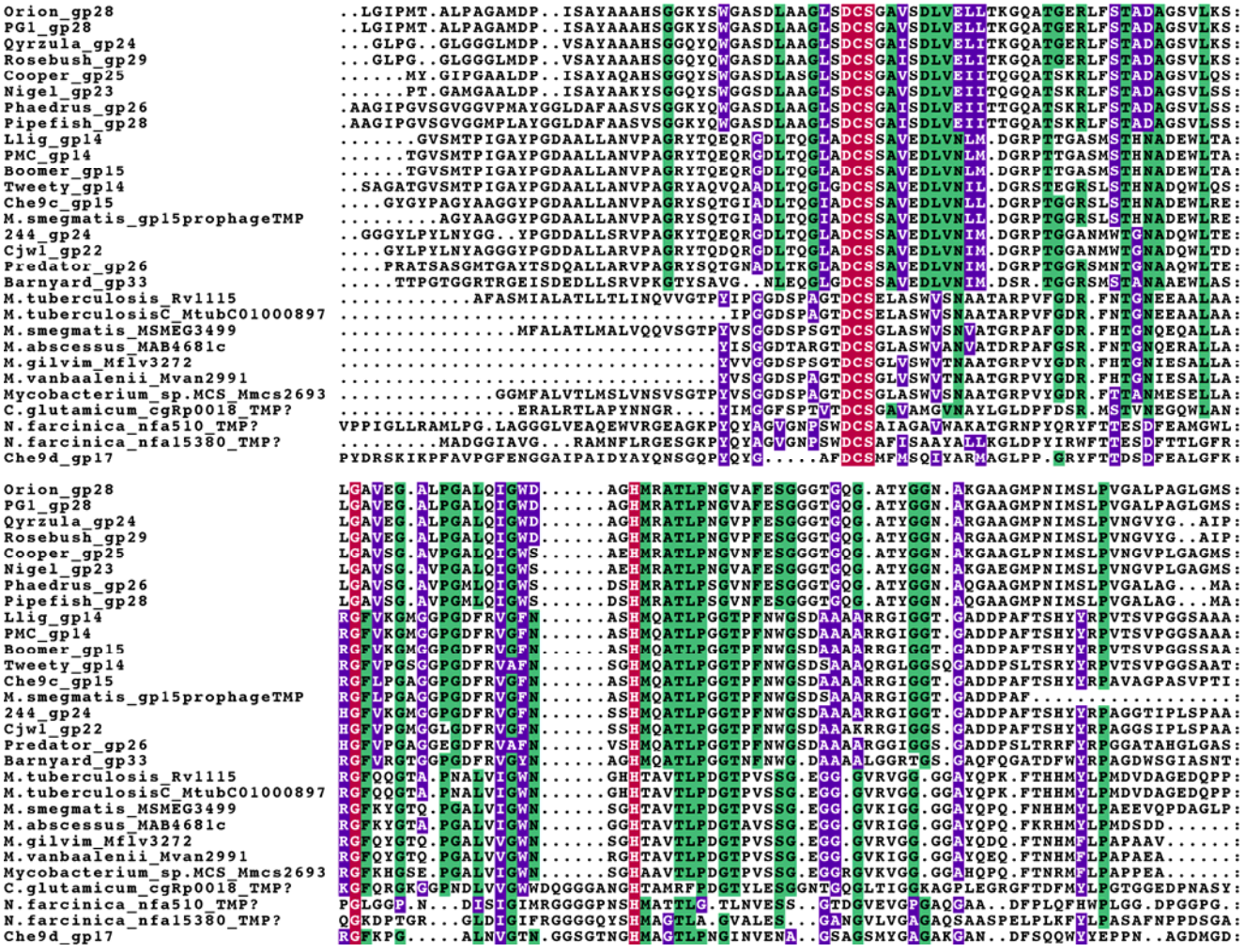
Motif 2 is the most prevalent motif and was detected in 20 of the 51 phages examined. These were grouped and searched by cluster, and in general, the number of significant hits was much lower than observed for either the Rpf Motif or Motif 3. When a Motif 2 from any of the Cluster B phages (Orion/PG1, Qyrula/Rosebush, Phaedrus/Pipefish, Cooper/Nigel) is used as the BLAST query, not surprisingly, the best matches are to the phages from the same cluster. These are followed – in variable order – by other mycobacteriophage Motif 2s (Che9c, Tweety, Barnyard, Llij/PMC, 244, Cjw1 and sometimes Che9d) and gp15 (MSMEG_2146), a likely TMP from an *M. smegmatis* prophage ($E = 4\times10^{-10}$, 47% identity from BLAST with Orion Motif 2 but similar regardless of query). The next hits, which are significant in some cases but generally have E-values in the range of 10^{-5}-10^{-3}, are to a number of mycobacterial proteins of unknown function; many of these are predicted to be secreted, contain domains found in lipoproteins, and/or contain a proline-rich region at the C-terminus.

The phages in Cluster E (244, Cjw1, Kostya, Porky) also contain Motif 2, and all of their tape measure genes are >90% identical. A BLAST search with any of these motifs detects, after the other Cluster E phage in the database: Llij, Che9c, Tweety, Barnyard, the cluster B's and basically the same mycobacterial matches – including the prophage TMP – observed with the cluster B motifs. The scores and E-values for these are generally in the range of 10^{-6}-10^{-4}, although they are somewhat lower when Cjw1 is the query. One of these, in *M. tuberculosis* strain C is annotated as an 'invasion-associated cell wall hydrolase', although it is not clear if there are data to support this, as there was no corresponding publication referenced.

An interesting case is encountered when it comes to the Cluster F phages: Llij, PMC, Boomer (which are nearly identical), Che8, Tweety and Che9d. Strikingly, although the Che8 TMP contains a large region of identity to Llij in the C-terminal half of the protein, in Che8, Motif 2 is replaced by Motif 3; therefore, Che8 will be considered with the Motif 3 containing phages. A BLAST search with the Llij or Tweety Motif 2 detects the Motif 2 of the other phages and essentially the same hits as the Cluster B and E phages considered previously ($E = 10^{-6}$-10^{-4} for most non-phage hits). Che9d is interesting because it encodes a more divergent Motif 2; the two best matches are to what appear to be TMPs from prophages in *N. farcinica* nfa510 ($E = 4\times10^{-16}$, 40% identity) and nfa15380 ($E = 4\times10^{-13}$, 40% identity). The next best matches are to the Cluster B phages, although some of the other mycobacterial proteins and phage motifs discussed above were detected with weaker scores.

Cluster H is composed of Barnyard and Predator, and overall their tape measures are 38.5% identical. The best matches to the Barnyard Motif 2 are to the Motif 2 domains in 244, Che9c, the *M. smegmatis* prophage gp15, Llij, Cjw1, Tweety, followed by the Cluster B TMPs ($E < 10^{-9}$). A hypothetical protein annotated as an efflux protein that resembles a TMP in *C. glutamicum* ($E = 4\times10^{-6}$, 40% identity) is detected, as well as the same mycobacterial proteins (weakly) discussed above. For Predator, the same phage motifs are found (including the Barnyard Motif 2), and no bacterial proteins are detected with any significance. Che9c doesn't cluster with any of the other phages, and its TMP Motif 2 is nearly identical to the motif from the *M. smegmatis* prophage gp15 (8×10^{-81}, 98% identity), although it also has significant similarity to Llij, Cjw1, 244, Tweety, Barnyard and the Cluster B phage TMPs, as well as to some of the mycobacterial proteins that were found to match the Cluster B Motif 2 domains. The Motif 2 domains and the proteins to which they are most similar were aligned, and the relationships are illustrated in Figure 11.

A.


Orion_gp28
PG1_gp28
Qyrzula_gp24
Rosebush_gp29
Cooper_gp25
Nigel_gp23
Phaedrus_gp26
Pipefish_gp28
Llig_gp14
PMC_gp14
Boomer_gp15
Tweety_gp14
Che9c_gp15
M.smegmatis_gp15prophageTMP
244_gp24
Cjwl_gp22
Predator_gp26
Barnyard_gp33
M.tuberculosis_Rv1115
M.tuberculosisC_MtubC01000897
M.smegmatis_MSMEG3499
M.abscessus_MAB4681c
M.gilvim_Mflv3272
M.vanbaalenii_Mvan2991
Mycobacterium_sp.MCS_Mmcs2693
C.glutamicum_cgRp0018_TMP?
N.farcinica_nfa510_TMP?
N.farcinica_nfa15380_TMP?
Che9d_gp17

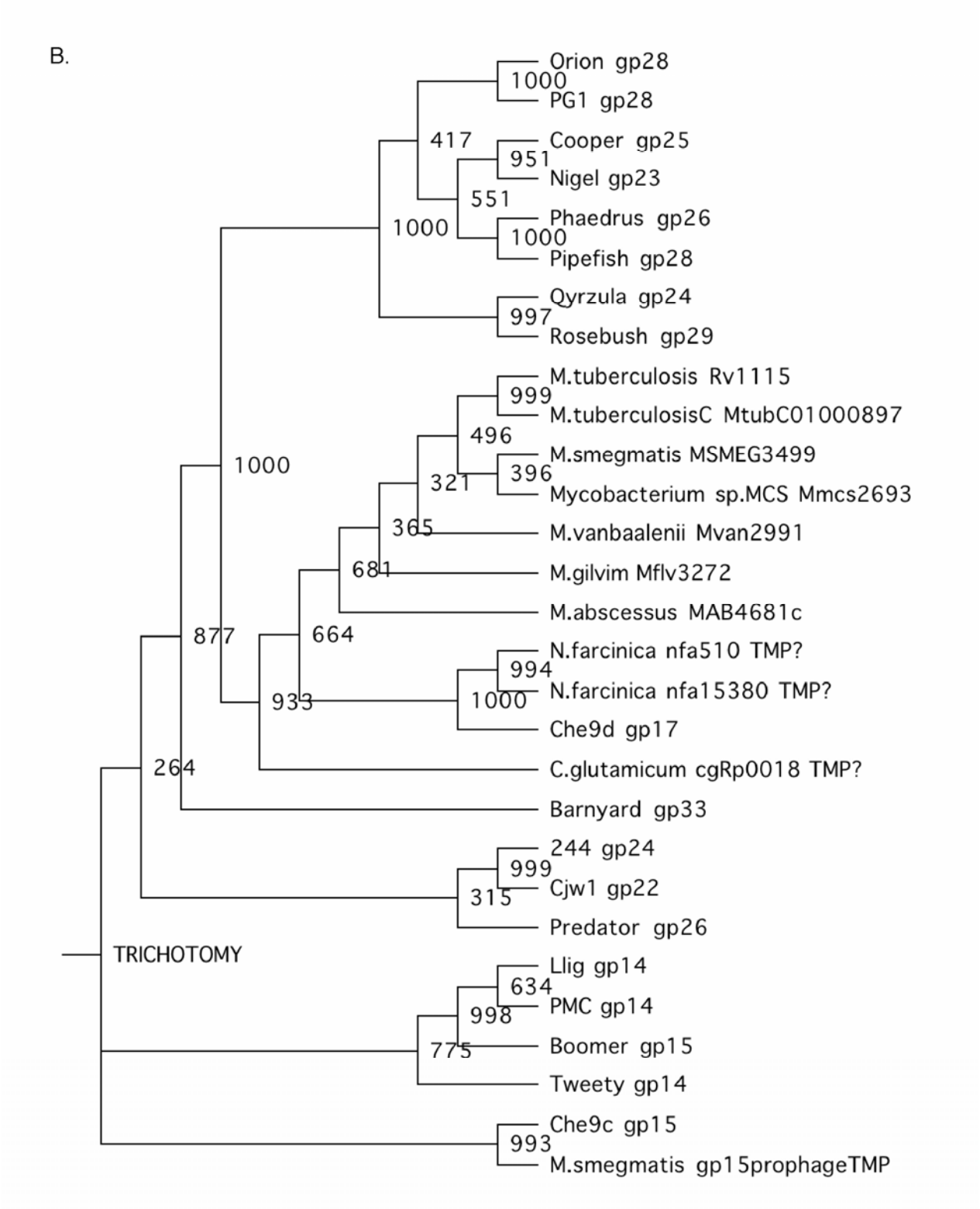


Figure 11. Mycobacteriophage TMPs containing Motif 2.

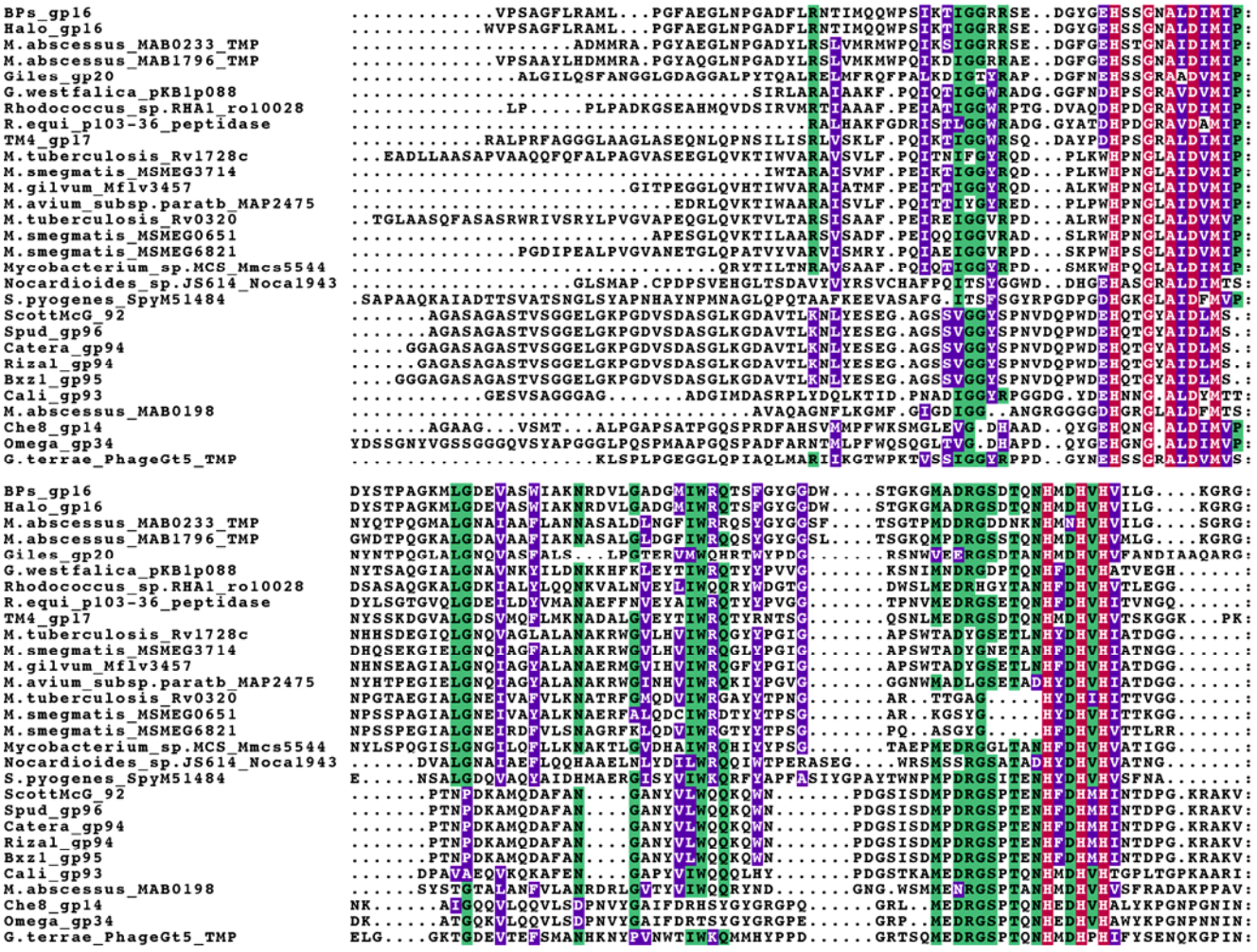
Figure 11: The mycobacteriophage tape measure Motif 2. **A.** ClustalX sequence alignment of the mycobacteriophage tape measure Motif 2s and related bacterial proteins. 100% conserved residues are highlighted in red, identical residues are highlighted in green, and similar residues are highlighted in purple. **B.** Phylogenetic tree displaying the relationships among the various TMP Motif 2s and Motif 2-like proteins. Neighbor-joining, un-rooted tree was drawn correcting for multiple substitutions, and confidence values displayed are from 1000 bootstrap trials.

3.3.3 Motif 3

Motif 3 was detected in 13 out of the 51 phages analyzed. For Che8 – the unique Cluster F phage with Motif 3 – the best match is to Omega Motif 3, which is 81% identical over this region. Other significant hits ($E = 3\times10^{-13}$-5×10^{-4}) include the tape measures from Halo, two *M. abscessus* prophages (MAB_0233 and MAB_1796) and TM4, as well as a putative secreted protein in *Nocardioide*s sp. JS614 and the TMPs from the *G. terrae* phage GTE5 and Giles. The best mycobacterial matches are to hypothetical secreted proteins in *M. abscessus* ($E = 6\times10^{-4}$) and *Mycobacterium* sp. MCS and KMS ($E = 7\times10^{-4}$), and these are followed by the peptidase in *R. equi* ($E = 0.002$), which was detected previously as a Motif 3 homologue. Not surprisingly due to their similarity, a BLAST with the Omega Motif 3 as a query detects the same proteins with similar scores, although E-value for the *R. equi* peptidase is slightly better ($E = 3\times10^{-4}$).

The phages Halo and BPs comprise Cluster G, and their TMPs are identical. Unlike the Che8 and Omega Motif 2s, which have only six non-mycobacteriophage hits with E-values lower than 10^{-4}, this motif has over 100 significant matches. It most closely matches the two tape measures in the *M. abscessus* prophages ($E = \sim10^{-49}$, 60-70% identity) mentioned above. Other matches include the TM4 and Giles TMPs, peptidases in *Rhodococcus* sp. RHA1 and *R. equi,* and the GTE5 TMP. The list also includes a number of hypothetical mycobacterial proteins, with the best match being a putative secreted protein in *Mycobacterium spp.* MCS and KMS ($E = 8\times10^{-15}$). Notably, many of these proteins are found in various species of high G+C% Gram-positive bacteria and are annotated as peptidases and/or secreted proteins.

A.


BPs_gp16
Halo_gp16
M.abscessus_MAB0233_TMP
M.abscessus_MAB1796_TMP
Giles_gp20
G.westfalica_pKB1p088
Rhodococcus_sp.RHA1_ro10028
R.equi_p103-36_peptidase
TM4_gp17
M.tuberculosis_Rv1728c
M.smegmatis_MSMEG3714
M.gilvum_Mflv3457
M.avium_subsp.paratb_MAP2475
M.tuberculosis_Rv0320
M.smegmatis_MSMEG0651
M.smegmatis_MSMEG6821
Mycobacterium_sp.MCS_Mmcs5544
Nocardioides_sp.JS614_Noca1943
S.pyogenes_SpyM51484
ScottMcG_92
Spud_gp96
Catera_gp94
Rizal_gp94
Bxz1_gp95
Cali_gp93
M.abscessus_MAB0198
Che8_gp14
Omega_gp34
G.terrae_PhageGt5_TMP

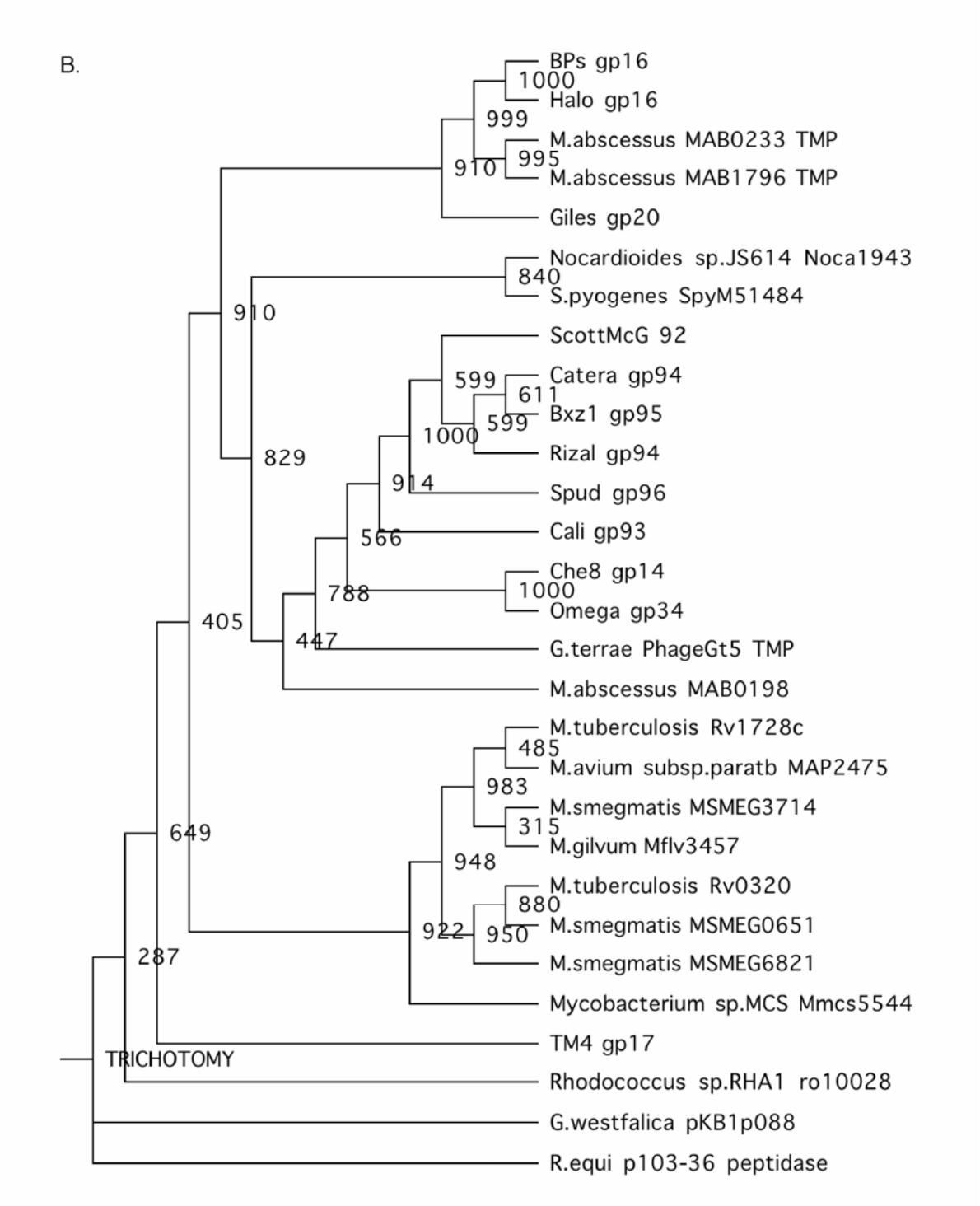


Figure 12. Mycobacteriophage TMPs containing Motif 3.

Figure 12: The mycobacteriophage tape measure Motif 3. **A.** ClustalX sequence alignment of the mycobacteriophage tape measure Motif 3s and related bacterial proteins. 100% conserved residues are highlighted in red, identical residues are highlighted in green, and similar residues are highlighted in purple. **B.** Phylogenetic tree displaying the relationships among the various TMP Motif 3s and Motif 3-like proteins. Neighbor-joining, un-rooted tree was drawn correcting for multiple substitutions, and confidence values displayed are from 1000 bootstrap trials.

The TM4 Motif 3 is similar to Halo and BPs in that it has over 100 significant matches when used as a query in a BLAST search. The best match is to a M23/M37 peptidase in *Gordonia westfalica* (E = 1×10^{-33}, 53% identity), a protein that is also predicted to contain a separate transglycosylases domain. Other significant hits are to some of the proteins described for Halo and BPs, and the highly conserved *M. tuberculosis* proteins Rv1728c and Rv0320 are detected with E-values of 7×10^{-25} and 4×10^{-20}, respectively. Giles Motif 3 is most similar to the TM4 and Halo Motif 3s and the *M. abscessus* prophage TMPs, and it also has over 100 significant matches in the database. Interestingly, the contractile-tailed mycobacteriophages, Bxz1, Cali, Catera, Myrna, Rizal, ScottMcG, and Spud also contain Motif 3-like regions within their predicted TMPs. These are located more toward the N-terminal regions of these proteins (Fig. 9), and it is tempting to speculate that this notable difference might related to the distinct mechanism of tail assembly likely employed by these phages. The relationships between these motifs and their various bacterial homologues are shown in Figure 12.

3.3.4 Evidence for New Motifs in Mycobacteriophage TMPs

When the TMP of mycobacteriophage Wildcat was used as the query in a BLAST search against the nr database, a conserved lytic transglycosylase/lysozyme domain (cd00254: LT_GEWL) was detected in the C-terminal third of the protein. This is a 70-80 residue core region that significantly matches over 50 proteins. Many of the best matches are likely TMPs from both phages and prophages, including the *Streptomyces* phage VWB gp47 (E = 7×10^{-16}, 42% identity), a *C. glutamicum* hypothetical protein cgR_1882 (E = 9×10^{-14}, 42% identity), the *Corynebacterium* phage BFK20 gp15 (E = 1×10^{-13}, 48% identity); a prophage TMP in *Bacillus clausii* (E = 2×10^{-12},

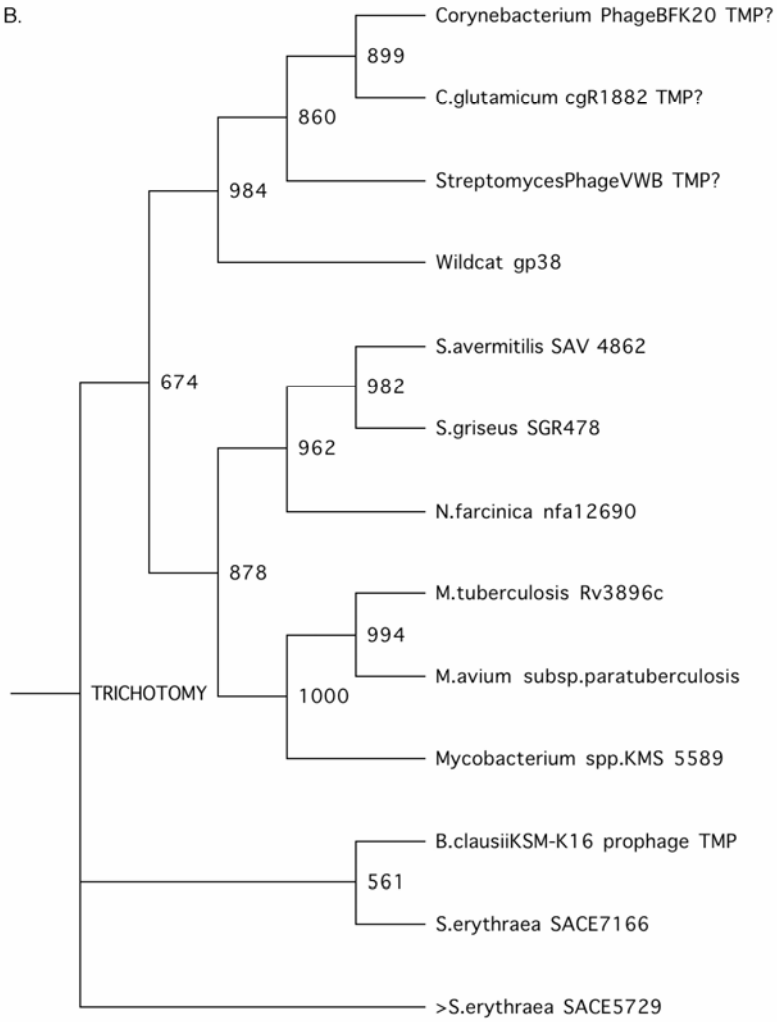
46% identity) that also contains a separate M23 peptidase domain, and a likely TMP from a prophage in *B. subtilis* (E = 2×10^{-12}, 44% identity) that is annotated as a possible membrane protein. The list of significant bacterial hits contains representatives from many species of Gram-positive bacteria, such as *Streptomyces, Nocardia, Lactobacillus* and *Bacillus spp*. Some of the best matches include a hypothetical protein in *Streptomyces griseus* SGR_478 (E = 7×10^{-15}, 48% identity) that also contains a LysM domain, a hypothetical protein *S. avermitilis* SAV_4862 (E = 6×10^{-14}, 47% identity), and two proteins in *Saccharopolyspora erythraea* (E = ~10^{-12}, 52% identity). Importantly, none of the Rpfs and no bacterial proteins related to the other motifs were detected, even after multiple rounds of PSI-BLAST, indicating that this represents a new tape measure motif, designated Motif 4. The best mycobacterial hits include a hypothetical protein in *Mycobacterium spp*. KMS Mkms_5589 (E = 7×10^{-11}, 45% identity) and a hypothetical protein in *M. tuberculosis* strain H37Rv, Rv3896c (E = 3×10^{-8}, 40% identity) that is also conserved in other sequenced *M. tuberculosis* strains. Currently Wildcat is the only mycobacteriophage to contain this motif, and therefore, the relationships between this phage motif and the various proteins to which it has similarity are illustrated in Figure 13.

A.

```
Corynebacterium_PhageBFK20_TMP ........TLWNWVKSKLPFGGGETGVQGGVAAGVEQWRDLVNRVLAAKGFD........TSMADTVLRRMNQESGGNFA:  64
C.glutamicum_cgR1882_TMP?      ............EESGAYFG........DVGAGVEQWRPLVISILKAKGFP........ESLADTTLRRMDQESGGNSR:  51
StreptomycesPhageVWB_TMP?      ....LGGLKDKVLDAAKSLFGGS.SGAADIGGSGVQRWSPVVLQALQMVGQS........ASLLPVVLRRMNQESGGNFA:  67
Wildcat_gp38                   ........NGALGVDIASLFGGQ.KPATKSQGEGVDQWDDTIFGVLMQLAQMGLINPSEIPDWMDAVKRQMKIESGGNPG:  71
B.clausiiKSM-K16_prophage_TMP  ............DPLGAASGGG.......AFKGSGASGMVKTWIMQAINATG......VPKSWLEPLATIAQKESGGNPR:  55
S.erythraea_SACE7166           ......................GG.......PPPGNVEEWIREAIKLIQANGIP......VTEENIDEIWTIIEKESGGNPH:  47
S.erythraea_SACE5729           ......................AD.......PPKDQLDGWIKEAVAVIEANGTP......KSKADVEDIRTIIEHESGGDPK:  47
S.avermitilis_SAV_4862         AVLALPLLGATGASAATPATTAAAPSSAVTYTNNLDGWIRESLDIMAKHNIP.........GTYNGIYRNVIRESSGNPL:  71
S.griseus_SGR478               ....................TAPAAEKAVTYTNDLDGWIKESLAIMAEQGIP.........GTYEGIHRNIMRESSGNPA:  51
N.farcinica_nfa12690           AAPAAEPAPAPAPEPVAEAVAQQAPPPPPVYENNLDGWIRQALDIMAANGIP.........GSYEGIYRNIMRESTGNPQ:  71
M.tuberculosis_Rv3896c         ..ARLLGLRLRRLRYLRTAAARRPQVTTPGGRAQVLAAIQKALDIQGVHDPAAR......ARWTRGMDLVARRESNYNAN:  72
M.avium_subsp.paratuberculosis .........LRRLHYRRRRPAAGP.TNQPGGRPAVLAAIRKALDIKGIHDPAAR......ARWERGMDLVARRESNYNAG:  64
Mycobacterium_spp.KMS_5589     ....PLADRQAGTRLRRSDIASSP...VAGG...IDAAISRALDIKGITDPRAR......RNWATGMKVVTTRESGNDIN:  64

Corynebacterium_PhageBFK20_TMP AVNNWDINAVNGTPSKGLMQVIDPTFAAHKDAGYD..DIWDPESNIRASMNYAINRYGSLSAAYDRAGGYHDGGLAPAGQ:142
C.glutamicum_cgR1882_TMP?      AINNWDSNAAAGIPSKGLMQVIDPTFAAHKDPGFD..DIWDPEANIRASMNYAVSRYGSLPAAYNRAGGYHDGGLAGYGR:129
StreptomycesPhageVWB_TMP?      AINNWDINAKNGVPSKGLMQVIDPTFAAYAGALRGR.GVWDPLANIYASMRVALSRYGSLASAYNRPGGYANGGRPRPGE:146
Wildcat_gp38                   AVNNWDSNAAKGTPSKGLMQMVEGTFNQHKVPGIGSGQWLDPVSQIAASVMYAKSKGG..PSYMTGKQGYATGGMTPSDL:149
B.clausiiKSM-K16_prophage_TMP  AINLWDSNYRRGTPSKGLMQTIDPTFNAYKKKGMD..NIWNPVHNAVAAIRVIVKRYGSVFETPGIKSMMRGGGYKGYAD:133
S.erythraea_SACE7166           AINNWDSNAAKGTPSKGLMQCIDPTFQSYKLPGHD..DIYNPVDNIIAGVRYTFSRYGGFEGHPGLKSMSQGGGYRGY..:123
>S.erythraea_SACE5729          IVNNWDSNAAMGTPSKGLMQTIEPTFESYALPGHG..DILDPVDNIIAGVRYSIDRYGSVSDVPGVVGVDSGGGSYRGY..:123
S.avermitilis_SAV_4862         AINNWDSNAAKGTPSKGLLQVIDPTFAAYHVPGTA.FDPYNPVANITAACNYAAARYGSIDNVFGAY..............:137
S.griseus_SGR478               AINNWDSNAVKGTPSKGLLQVIDPTFQAYHVPGTS.TDSYDPVANITAACNYAADRYGSIDNVFGAY..............:117
N.farcinica_nfa12690           AINLWDSNAAAGIPSKGLLQVIDPTFAAYHVPGTP.FDVWDPVANIVAACNYAAHRYGSMDNVNSAY..............:137
M.tuberculosis_Rv3896c         AINHWDSNAARGTPSRGVWQFIAPTFAAYHEPGTS.TNIHDLVAQACAFINYARGHYGVAADASNLADLIQQADPRRSPR:151
M.avium_subsp.paratuberculosis AVNDWDANAARGTPSKGAWQFIEPTFAAYHQRGTS.RDIHNLVAQACAFINYAMGRYGVAADASNLADRIQQADPRRSPK:143
Mycobacterium_spp.KMS_5589     AVNNWDSNAANGTPSQGAFQFIEPTFRAYHEPGTS.PFLRDPVAQGAAFVNYARGRYNVSADGSDLAQKIQQADPTRPPR:143
```

B.



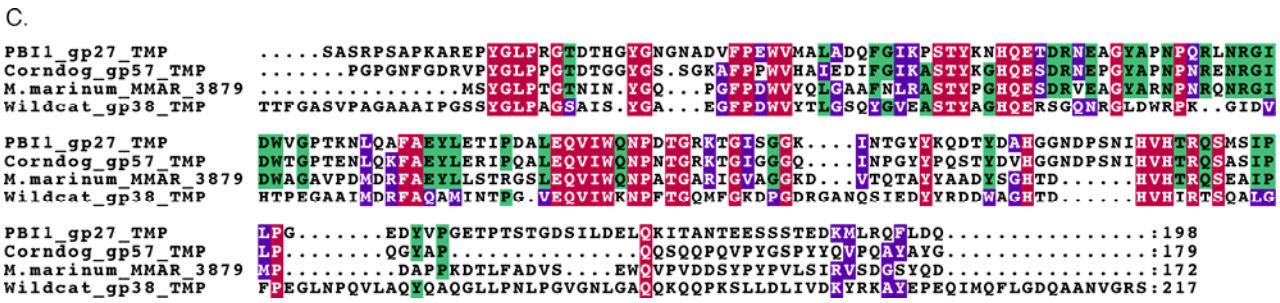


Figure 13. Tape measure Motifs 4 and 5.

Figure 13: Mycobacteriophage tape measure Motifs 4 and 5. **A.** ClustalX sequence alignment of the mycobacteriophage tape measure Motif 4s and related bacterial proteins. 100% conserved residues are highlighted in red, identical residues are highlighted in green, and similar residues are highlighted in purple. **B.** Phylogenetic tree displaying the relationships between the Wildcat Motif 4 and various Motif 4-like proteins. Neighbor-joining, un-rooted tree was drawn correcting for multiple substitutions, and confidence values displayed are from 1000 bootstrap trials. **C.** ClustalX sequence alignment of the mycobacteriophage tape measure Motif 5s and the putative lysin in *M. marinum.* 100% conserved residues are highlighted in red, identical residues are highlighted in green, and similar residues are highlighted in purple.

Examination of the Wildcat TMP also revealed what may be a fifth tape measure motif, Motif 5, N-terminal to the Motif 4 region. This segment is 130-140 residues, and it matches a conserved hypothetical bacteriophage protein in *Mycobacterium marinum* ($E = 6\times10^{-24}$, 45% identity), as well as another phage protein in *M. vanbaalenii* ($E = 2\times10^{-20}$, 48% identity). The second of these is likely a TMP; however, the *M. marinum* protein (MMAR_3879) at 407 residues is likely too small to be a tape measure. Interestingly, this motif also matches a region near the C-terminus in the tape measures of PBI1, Plot, Adjutor, and Corndog, none of which were previously observed to contain motifs (although PBI1, Plot and Adjutor may contain a more divergent Rpf Motif as noted above). Although these results alone are somewhat inconclusive and do not suggest a possible function for this new motif, a BLAST search with the *M. marinum* protein reveals that this also has a conserved glycoside hydrolase domain (pfam:01183, GH25), which is separate from the region of the protein that matches 'Motif 5'. This GH25 domain is commonly found in mycobacteriophage endolysins: proteins that hydrolyze peptidolgycan and mediate cell lysis at the end of the lytic cycle. Further, it has been shown that mycobacteriophage lysins have a modular nature and often contain multiple domains, some of which are uncharacterized (K. Clemens, unpublished data). This suggests that *M. marinum* MMAR_3879 might be a multi-domain prophage lysin and may further imply that the region of this protein that matches the mycobacteriophage tape measures also has muralytic activity. However, further work will be required to determine if this indeed represents an additional hydrolase motif.

3.4 CONCLUSIONS

The data presented in this chapter support the hypothesis that the mycobacteriophage TMPs are structurally similar to one another, as well as to the tape measures from other phages. This is based on the following observations: they are predicted to be highly α-helical, contain regions likely to form coiled coils and possess multiple putative transmembrane domains, features that have been noted for λ gpH, as well as for other TMPs that have been described. Further, an analysis of the amino acid composition of representatives from the various clusters of mycobacteriophage TMPs reveals that the residue profiles are remarkably similar to one another, even when the proteins themselves are not otherwise related. All of these features may be due to the fact that TMPs from different phages are likely to perform similar functions and must adopt very similar and constrained conformations within the phage tail.

There are a number of other interesting features, however, that emerge from these analyses. Notably, the predicted α-helical content varied significantly for the different tape measures examined; this was highest among the Cluster A phages (as high as 96.23% for U2) and lowest for the Cluster E and F phages (32.83% for Che8). Further, all of the tape measures examined, except those from Cluster A, contain regions near their C-termini that are predicted to be lower in α-helical content. This has also been noted for λ gpH, and in this case, it is known that the C-terminal region of the protein is required for tail assembly and is cleaved prior to head-tail attachment (Weigle, 1966, Hendrix & Casjens, 1974). It has further been noted that the C-terminus of gpH is more glycine-rich than the rest of the protein (Xu, 2001, Ph.D. Thesis), and this same feature has generally been observed for the mycobacteriophage tape measures as well. Therefore, the data suggest that these proteins are likely to function similarly to gpH during tail assembly.

Siphoviridae such as λ, TP901-1 and the mycobacteriophages considered here are morphologically very much alike; therefore, it is perhaps not surprising that they all encode a virion protein with similar properties. However, it is quite striking that many of these properties – including a high α-helical content, multiple putative transmembrane domains, and a similar amino acid composition – are also exhibited by T7 gp16. This large structural protein is clearly not what can classically be thought of as a tape measure, as T7, a member of the podoviridae, only has a very short tail stump. Thus, characteristics of TMPs that are also shared with gp16 are unlikely to be related to their role in tail length determination. Notably, gp16 is noted for the fact that it, along with two other proteins, forms an extensible tail that exits from the phage capsid during infection, inserts into the bacterial membrane and facilitates DNA injection (Garcia & Molineux, 1996, Struthers-Schlinke *et al.*, 2000, Molineux, 2001, Kemp *et al.*, 2004). Further, this protein also has an active lytic transglycosylase region that is important for the infection of stationary phase *E. coli* (Moak & Molineux, 2000, Moak & Molineux, 2004), a function that is proposed for the mycobacteriophage tape measures motifs. This suggests that the TMP might exit from the tail during infection and function in a manner that is analogous to gp16. In support of this, it has been observed that the tape measure-like protein pb2 from phage T5 also has many of these same structural features, and this protein is known to insert into lipid bilayer membranes during infection and *in vitro* (Guihard *et al*., 1992, Lambert *et al*., 1998, Boulanger *et al*., 2008). Intriguingly, this activity is thought to be mediated by glycine/alanine rich portions of pb2, which resemble the regions of eukaryotic viral envelope proteins that insert into the membrane and mediate fusion with the host membrane (Del Angel *et al.*, 2002).

This bioinformatic investigation of the mycobacteriophage TMPs has also confirmed that Giles and Barnyard remain the only two mycobacteriophages with a canonical Rpf Motif. These

regions appear to be closely related to one another and to numerous Rpf family members from high G+C% corynebacteria (>55% identity across this region), and both are highly similar to what appears to be an Rpf Motif within the TMP of a *M. vanbaalenii* prophage (Fig. 10A and B). However, a region that is more distantly related to the Giles and Barnyard motifs is found in the TMPs of PBI1, Plot, and Adjutor, which are nearly identical (>99%) to one another. This putative Rpf Motif is highly similar to a region found in the TMP of the *G. terrae* phage GTE5 and to a segment of what is likely a TMP from a prophage in *S. haemolyticus*, neither of which is otherwise related to the mycobacteriophage tape measures, displaying less than 15% identity across the length of the proteins. Although these Rpf-like motifs are more divergent and have an insertion near the protein active site, they have identity (~40%) to Rpf proteins from various species of *Corynebacterium* and also contain residues that are highly conserved in the Rpf family of proteins, such as the putative active site glutamate (Fig. 10A), This might suggest that PBI1, Plot and Adjutor should be included within the list of mycobacteriophages whose TMPs contain Rpf Motifs. Intriguingly, the tape measures from these phages also contain one of the newly identified motifs, discussed below. Further, the GTE5 tape measure contains a separate Motif 3-like region (~110 residues) within a larger segment of the protein (~360 residues) that has identity to the mycobacteriophage Halo (and BPs) TMP, and the *S. haemolyticus* protein contains a putative M23 peptidase domain distinct from its Rpf-like region. This is consistent with the pattern observed in the mycobacteriophages, where the Rpf Motif is always found in conjunction with one of the other motifs.

Motif 2 was found to be the most prevalent of the tape measure motifs, as it was identified in 20 of the 51 phages examined. Unfortunately, the bacterial proteins most similar to Motif 2 have not been characterized, and therefore, it is difficult to predict a function for this motif. However,

many of the mycobacterial proteins with Motif 2-like regions are predicted to be secreted and/or contain domains found in lipoproteins, and one of these – in *M. tuberculosis* strain C – is annotated as an 'invasion-associated cell wall hydrolase', although it is unclear if there are experimental data to support this. These observations, combined with the functions of the bacterial homologues of the other tape measure motifs, suggest that Motif 2 may also have some type of muralytic activity.

Motif 3 is also quite common and was found in 13 of the 51 mycobacteriophage tape measures investigated. Proteins containing this motif have muralytic activity *in vitro*, and TM4 phages deleted for this motif infect late-stationary phase *M. smegmatis* with a reduced efficiency as compared to wild-type. These data strongly suggest that Motif 3 functions analogously to the lytic transglycosylase in T7, gp16. However, many of the bacterial proteins containing this motif are annotated as peptidoglycan peptidases, suggesting that Motif 3 may hydrolyze the peptide bonds that form cross-links within the peptidoglycan network. This motif is also interesting because it is found within the 'tape-measure proteins' of the contractile-tailed mycobacteriophages. These phages have tails that are morphologically distinct from those of the siphoviridae, and their tape measures are different from the tape measures considered in this study in a number of different ways. For example, they commonly contain 10-20 cysteine residues, whereas multiple cysteines were not observed in other tape measures. However, there is evidence that the TMP from the coliphage T4 is found within the phage tail and functions to specify the length of this structure (Duda *et al*., 1986, Abuladze *et al.*, 1994). Therefore, it may also contact the host during infection, which would make it ideally suited to hydrolyze cell wall peptidoglycan in the process.

These analyses have also revealed the presence of potentially two additional distinct motifs within the mycobacteriophage tape measures. Motif 4, which is found only in mycobacteriophage

Wildcat, corresponds to a region of the tape measures that is annotated as having a conserved lytic transglycosylases (LT)/goose egg white lysozyme (GEWL) domain and has a high degree of amino acid similarity to a number of proteins with muralytic or predicted-muralytic activity. One of the bacterial proteins to which it is similar is the *M. tuberculosis* protein Rv3896c. This is distinct from the five Rpf proteins and the Motif 2 and Motif 3 containing proteins encoded by this organism, further suggesting that Motif 4 represents a separate motif that is unrelated to the tape measure Rpf Motifs. The other new motif, Motif 5, is also present in the Wildcat TMP and is additionally found within the tape measures of Corndog and the PBI/Plot/Adjutor Cluster. The region does not resemble any known murein hydrolase domains; however, it does have identity to the N-terminus of a protein from *M. marinum* MMAR3879 that is annotated as a bacteriophage protein. This is not a tape measure, but rather, appears to be a phage endolysin, which has a glycoside hydrolase (GH25) domain (residues 144-314) that is distinct from the portion of the protein that matches the tape measures (residues 1-133). Mycobacteriophage endolysins are modular proteins that have been observed to contain a number of different catalytic domains (K. Clemens, unpublished data). Uniquely, some of these proteins have two separate domains with distinct catalytic activities (e.g. peptidase and amidase); however, many contain one characterized catalytic domain plus a region at the N-terminus of the protein with unknown function. It has been hypothesized that these uncharacterized N-terminal domains may possess a unique muralytic activity that is specific for a particular bond within the mycobacterial cell wall, and the observation that one of these regions is also present in mycobacteriophage TMPs in the same location as motifs thought to be involved in murein hydrolysis lends support to this hypothesis.

Notably, the identification of Motif 4 and Motif 5 has revealed that at least one of the five motifs is present in every sequenced mycobacteriophage, except for those in Cluster A.

Intriguingly, these are phages whose TMPs are generally shorter (~800 residues) and have a higher α-helical content, as discussed above. This might suggest that the C-terminal regions of tape measures that contain these motifs perform a specialized function (e.g. murein hydrolysis) during infection. The Cluster A phages on the other hand, may encode an additional structural protein that performs the role of the C-terminal portion of the TMPs from other phages. Importantly, however, these structurally distinct C-termini must also participate in tail length determination, as phages with longer TMPs have correspondingly longer tails. Future studies aimed at addressing the roles of the various regions of the mycobacteriophage tape measures from individual phages will be necessary to elucidate the basis for their observed structural features.

4.0 DETERMINATION OF THE STATE OF THE RPF MOTIF IN MATURE VIRIONS AND DURING THE COURSE OF INFECTION

4.1 INTRODUCTION

The majority of TMPs that have been identified and annoted in bacteriophage genomes have not been studied in detail. Despite this limitation, the current available data suggest that it is quite common for these proteins to be proteolytically processed in some way during phage assembly. For the *E. coli* bacteriophage λ, it has been shown that 100 amino acid residues are cleaved from the C-terminus of the tape measure, gpH (92 kDa) to form gpH* (~80 kDa) (Hendrix & Casjens, 1974). Processing occurs after TMP incorporation and tail tube polymerization – gpH is a member of the tail-initiator complex and is therefore one of the first components added to the tail – and prior to head-tail attachment; thus, the entire length of the protein determines the λ tail length (Tsui & Hendrix, 1983). In this case, the cleaved fragment does not remain associated with the phage, as mature particles contain only the ~80 kDa gpH*, and the ~10 kDa fragment presumably diffuses out of the tail shaft after cleavage. The role of this processing as well as how it is catalyzed is not known; however, the C-terminus of the protein is required for initiator (and therefore tail) formation and likely interacts with other members of the initiator complex (Murialdo & Siminovitch, 1972, Katsura, 1990)

There is also evidence for processing among the phages of the Gram-positive bacteria. The *tmp* gene (*orf tmp)* of the *L. lactis* phage TP901-1 at 2814 bp is predicted to encode an ~100 kDa (937 residue) protein with a largely α-helical secondary structure, which at this size, would span the length of the 135 nm phage tail (Pedersen *et al*., 2000). In a 2000 study, Pedersen and colleagues made a number of mutations in this gene in order to determine its function (discussed in Section 1.1.2). Examination of mutant virions by electron microscopy clearly demonstrated that this protein was essential for tail formation as well as tail length determination, as insertion and deletion mutants had tails that were longer and shorter, respectively, by the expected amounts (Pedersen *et al*., 2000). Interestingly, however, an examination of the protein profiles of *tmp* mutant phages suggested that the product of *orf tmp* actually corresponds to a structural protein of ~70 kDa. A protein of this size and an ~35 kDa protein were missing in mutants with a C-terminal deletion in *tmp* and in phage containing an amber mutation at codon 16. Phages with both an N-terminal duplication and a deletion in this region were also missing the ~70 kDa protein and had reduced amounts of the 35 kDa protein. However, mutants with the duplication contained an extra 92 kDa protein. The authors suggest that since a 100 kDa protein was not detected, the TP901-1 TMP is likely proteolytically processed, and the observed ~70 and ~35 kDa proteins may represent the two pieces of TMP. In a later study, however, it was determined that the ~35 kDa protein is actually Orf48 (BppU), a component of the upper baseplate (Vegge *et al.*, 2005). Therefore, the fate of the cleavage product of the TMP remains unknown.

It is noteworthy that processing of the TMP protein has also been found for the related phage, Tuc2009, as polyclonal antibodies generated to the C-terminal 638 residues of this protein (1024 residues total) recognize an ~60 kDa structural protein (Mc Grath *et al*., 2006). Intriguingly, they also recognize the full-length 110 kDa protein, suggesting that both are actually present in the

phage. Antibodies recognizing the N-terminal region of the protein were not tested, and therefore in this case as well, the fate of the cleavage product is not known, although it was not detected by N-terminal sequencing of virion proteins.

A particularly unusual form of proteolytic processing has been observed for the TMP of the temperate *Listeria monocytogenes* phage PSA (Zimmer *et al.*, 2003). This phage has a long flexible non-contractile tail of ~180 nm, and the putative tape measure gene is the 3078 bp *orf12*, which is predicted to encode a 112.1 kDa protein. In contrast to this, an ~84 kDa structural protein comprising ~2% of the total protein was identified as gp12, which is ~30 kDa (25%) smaller than the expected size of the protein. Peptides from both ends of the protein were sequenced by mass spectrometry, but none were identified from the 270-residue region between amino acid positions 524 and 812. The mass of these missing residues is predicted at 29.7 kDa, which corresponds well to the size reduction observed by SDS-PAGE and suggests that this protein, although it does not contain any recognizable inteins, undergoes a novel, internal processing event during tail assembly. Structure prediction software predicts four putative transmembrane helices in this missing region, but it is unclear what their role is. It is also unclear from this analysis if the cleaved internal portion of gp12 is also present in virion particles. It was not identified by mass spectrometry or by N-terminal sequencing; however, it is possible that the band corresponding to this fragment was masked by the highly abundant ~30 kDa major capsid protein.

Notably, there are also data suggesting that mycobacteriophage tape measures may be processed as well. It has previously been shown that the mycobacteriophage TM4 tape measure motif, Motif 3, has similarity to proteins with muralytic activity and likely performs the same function during infection, as mutant phage missing this motif are impaired in infecting late stationary phase *M. smegmatis* cells (Piuri & Hatfull, 2006). Interestingly, when the fates of

labeled TM4 proteins are charted throughout the course of infection, it appears as though the TMP is proteolytically processed (Ford *et al.*, 1998), although this has not been further investigated.

Furthermore, a number of phage structural proteins that are known to be involved in peptidoglycan hydrolysis are also subjected to proteolytic cleavage. The T4 tail lysozyme gp5 is proteolytically cleaved during assembly; the N-terminal product contains the catalytically active residues, and the C-terminus forms a needle-apparatus believed to puncture the bacterial outer membrane (Kanamaru *et al*., 2002). Importantly, this processing is necessary for full enzymatic activation of the lysozyme (Kanamaru *et al*., 1999, Kanamaru *et al.*, 2005). It has also been shown that the Tuc2009 tail lysin Tal_{2009}, which is located at the tail tip, undergoes autolytic processing, and both the full-length form and the N-terminal cleavage product are part of the mature virion (Kenny *et al*., 2004, Mc Grath *et al.*, 2006). The tail fiber protein (ORF47) from the related phage, TP901-1, which is also found at the tail tip, has a high degree of identity to Tal_{2009} – 93% across the entire protein – and is also processed in a similar way, suggesting that both proteins perform a similar function (Vegge *et al.*, 2005).

4.2 BARNYARD TAPE MEASURE IS PROTEOLYTICALLY PROCESSED

Little is known about the mycobacteriophage TMPs. Therefore, in order to better characterize these proteins and begin to understand the function of the mycobacteriophage TMP embedded Rpf Motifs, I have conducted a number of experiments aimed at determining the physical state of the Barnyard motif both within mature phage particles as well as during the course of host infection.

4.2.1 The State of the TMP in Mature Barnyard Virions

Tape measures are typically large proteins, which constitute only a minor fraction of the total bacteriophage protein profile. Further, as noted above, there is an increasing body of evidence that these are often subjected to various forms of proteolytic processing. I, therefore, generated polyclonal antibodies to the Barnyard Rpf Motif in order better visualize this protein as well as to confirm the presence of the Rpf Motif within the mature polypeptide.

To this end, I PCR-amplified the 330 bp region of the Barnyard *tmp* gene that encodes the Rpf Motif and cloned this fragment in-frame with a C-terminal 6×-His tag in an *E. coli* expression vector under control of the inducible T7 promoter. Expression from this construct, designated pLAM5, was induced in *E. coli* cells containing the T7 RNA polymerase by addition of IPTG, and samples were analyzed by SDS-PAGE. Interestingly, the 115 amino acid recombinant protein was expressed at only low levels and appeared to be unstable, as maximum expression levels were reached at approximately 2 hours post-induction, after which the protein was no longer detectable by standard SDS-PAGE (Fig. 14A). Induced cultures were therefore harvested during the window of maximum expression, and the polyhistadine-tagged protein was partially purified with metal affinity chromatography (Fig. 14B). Sufficient quantities of this Rpf-like fragment (RLF) were generated to permit the production of polyclonal antibodies in rabbits, which were screened and affinity purified.

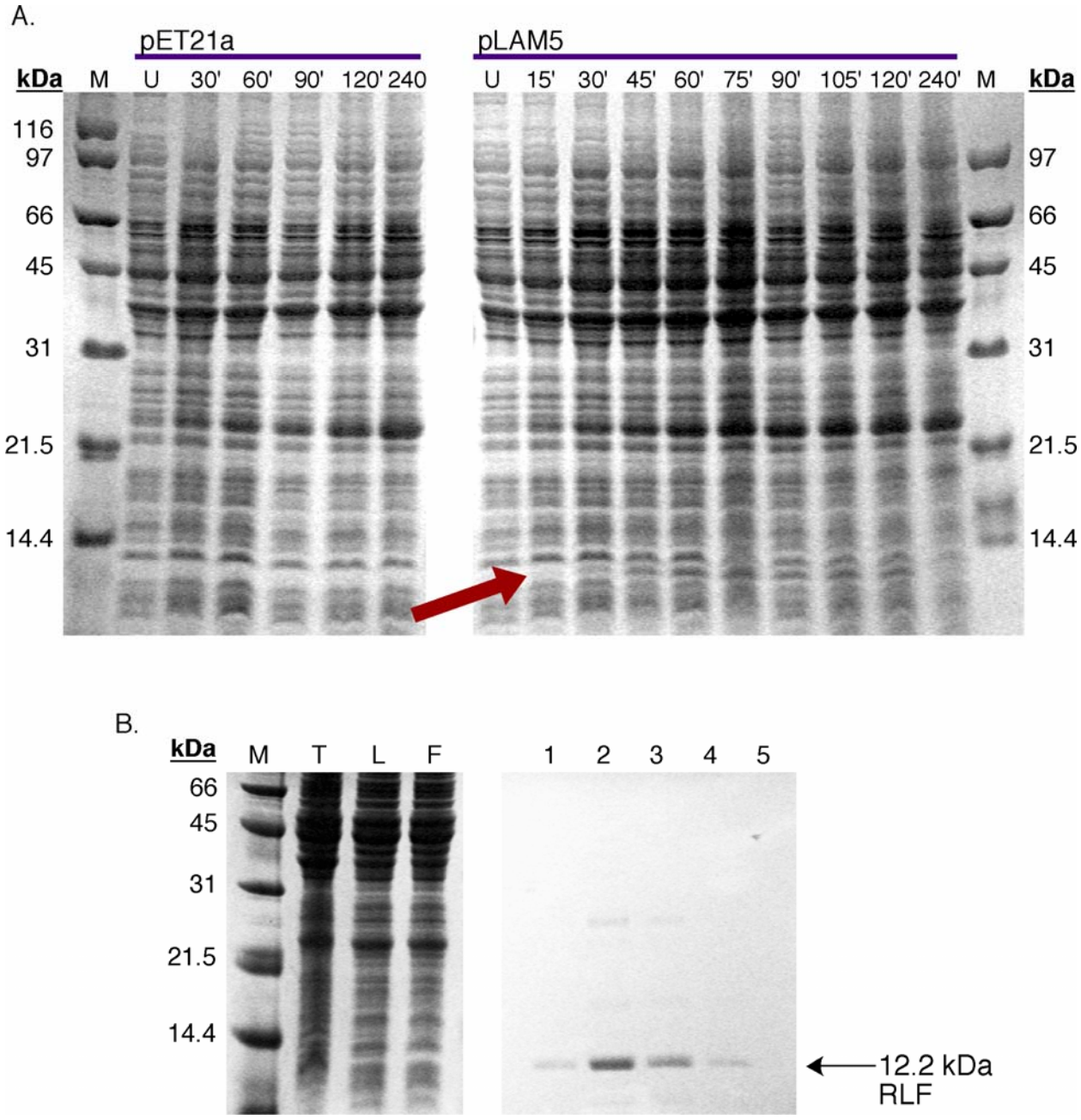


Figure 14. Expression and purification of the Barnyard Rpf-like fragment (RLF).

Figure 14: Expression and purification of the Barnyard Rpf-like fragment (RLF). **A.** Induced *E. coli* cells containing pLAM5 and the empty vector control were harvested at various time points and proteins were separated by high resolution Tris-tricine SDS-PAGE. Numbers at the top indicate minutes post induction, and U = uninduced. A protein band in the expected size range (12.2 kDa) can be seen in the pLAM5 samples, 30 minutes – 2 hours post induction. **B.** The His-tagged recombinant protein was purified from a 2 hour induction in *E. coli* using TALON metal affinity resin. Protein samples from the RLF purification were analyzed by high-resolution Tris-tricine SDS-PAGE: T = total protein, L = load, F = flow through, and numbers indicate elution fractions.

Immunoblot analyses of total phage protein separated by SDS-PAGE initially suggested that the Rpf Motif was present in two phage structural proteins of ~65 and ~70 kDa. This was initially quite puzzling, but subsequent analyses employing reduced boiling times during SDS-PAGE sample preparation revealed that the smaller of the two bands was a degradation product and could be eliminated by utilizing gentler methods of protein preparation. It remained striking however, that the α-RLF antibodies detected a 70 kDa protein due to the fact that the *tmp* gene is predicted to encode an ~217 kDa protein, and this was the first piece of evidence that this protein may be processed.

In order to identify this ~70 kDa band detected by the α-RLF antibodies and determine the state of mature TMP, LC MS/MS mass spectrometry was performed on a number of high molecular weight Barnyard structural proteins. These analyses confirmed that the 70 kDa protein that is immunoreactive with the α-RLF antibodies represents the C-terminal third of the TMP, corresponding to residues 1385-2047 (Fig. 15). The remaining portion of TMP is also present in the mature virion protein profile, as peptides from the N-terminal region of TMP (between residues 87-1384) were sequenced from the 138 kDa band highlighted in Figure 15. These data indicate that mature Barnyard TMP is indeed processed into two fragments, both of which are present in mature phage, and predict a putative cleavage site between K1384 and A1385. These results also suggest that up to 86 amino acid residues of the Barnyard TMP are proteolytically cleaved from the N-terminus of the protein, as these residues were not detected in the 138 kDa fragment of TMP.

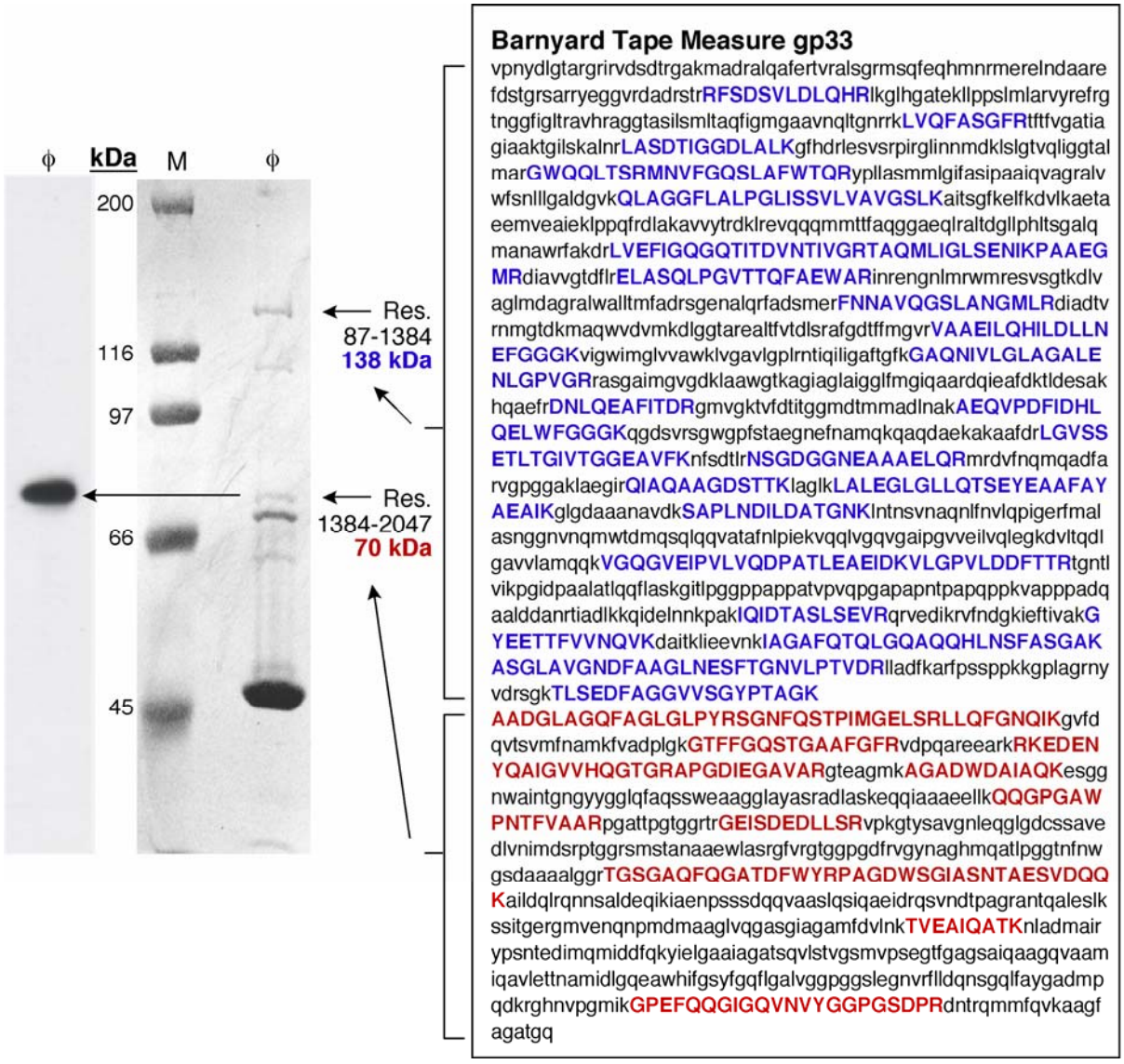


Figure 15. Identification of the Barnyard tape measure fragments by mass spectrometry.

Figure 15: Identification of the Barnyard tape measure fragments by mass spectrometry. Polyclonal antibodies to the Barnyard RLF detect an ~70 kDa phage structural protein; this is significantly smaller than the predicted size of Barnyard TMP (217 kDa). Total phage protein was separated on a 7% SDS polyacrylamide gel and analyzed by western blot with affinity purified α-RLF antibodies at a 1:125 dilution. The right panel illustrates total phage protein stained with Coomassie Blue, and the western blot analysis is shown on the left. In order to identify the band detected by the α-RLF antibodies, Barnyard structural proteins were subjected to mass spectrometric analysis (ProtTech, Inc.). LC-MS/MS spectra were analyzed against the GenBank protein database in order to provide identification. These experiments confirm that TMP is present in two pieces within mature phage; a 138 kDa N-terminal portion and a 70 kDa C-terminal fragment containing the RLF. Peptides sequenced from the N-terminus are bolded in blue and those from the C-terminal fragment are in red.

4.2.2 Tape Measure Expression and Cleavage During Infection

Tape measure processing was further examined *in vivo* by monitoring the state of this protein during time-course infections of *M. smegmatis.* Immunoblot analyses with α-RLF antibodies were performed on total cell protein harvested at 30-minute intervals from bacteria infected with Barnyard at a high multiplicity of infection (M.O.I.). The data, shown in Figure 16, indicate that full-length TMP (~200 kDa) is expressed during infection and is first detectable by this method at ~2 hours post-infection (P.I.). However, as the infection progresses, this protein appears to be cleaved, which is evident by an increase in the 70 kDa Rpf Motif-containing TMP fragment that is concurrent with the appearance of newly synthesized full-length protein. The initial appearance of the 200 kDa band (likely corresponding to full length TMP) and the subsequent re-appearance of the cleaved 70 kDa fragment are both blocked by treating cells with either the protein synthesis inhibitor chloramphenicol or the inhibitor of host RNA polymerase, rifampicin (Fig. 16). These results further suggest that Barnyard TMP is subjected to proteolytic processing at some point during virion assembly.

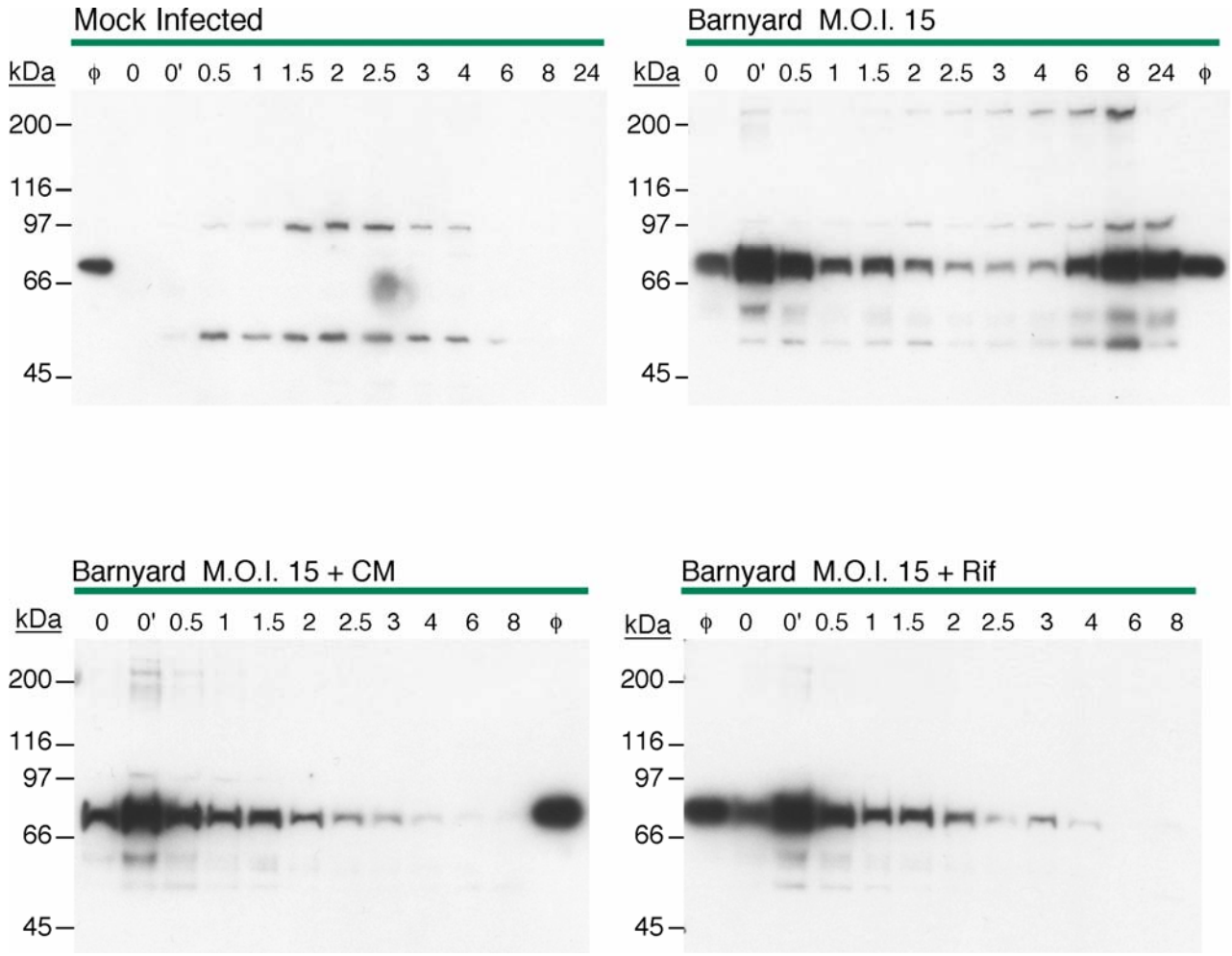


Figure 16. Expression and cleavage of the Barnyard TMP during infection of *M. smegmatis.*

Figure 16: Expression and cleavage of the Barnyard TMP during infection of *M. smegmatis.* An overnight culture of *M. smegmatis* was diluted to an OD_{600} of ~0.1 in 7H9 medium supplemented with ADC and $CaCl_2$ and infected with Barnyard at an M.O.I. of 15. Three otherwise identical cultures were prepared; a no phage control, one to which rifampicin (Rif) was added at a final concentration of 200 μM, and one to which chloramphenicol (CM) was added at a final concentration of 40 μg/ml (both at 1 hour post-infection). All cultures were held at room temperature for 20 minutes to allow phage adsorption and then transferred to 37°C. Samples were removed at the time-points indicated, with the numbers above indicating minutes after transfer to 37°C (0 is pre-adsorption and 0' is post-adsorption). These were washed with phage buffer with 10% glycerol, and total protein was separated on a 7% SDS-PAGE gel and probed with a 1:125 dilution of affinity purified α-RLF antibodies. Expression of an ~200 kDa protein that likely corresponds to full-length TMP can be detected at ~120 minutes post-infection. This appears to be processed into the 70 kDa RLF-containing species present in mature phage. Addition of chloramphenicol and rifampicin inhibits the appearance of the ~200 kDa product, indicating that this represents newly synthesized protein.

4.2.3 The Barnyard Tape Measure mRNA Transcript is Not Cleaved

Although unlikely, it remained a possibility that the tape measure mRNA transcript might be processed in some way. Therefore, to rule out this possibility, reverse transcription (RT)-PCR was performed on mRNA extracted from *M. smegmatis* cells from similar time-course infection experiments. Three primer sets – one to the N-terminus of the protein, one to the C-terminus, and one which spans the putative cleavage site – were utilized and all generated product at all time points examined, providing no evidence for mRNA processing, as expected (Fig. 17A). Semi-quantitative RT-PCR utilizing the primers spanning the cleavage site, indicates that, although the protein is not visible by western blot until later in the infection cycle (~120 minutes P. I.), TMP mRNA is visible at very early time points (30 minutes P.I.) and appears to increase up to the time when full-length TMP is first visible by western blot (Fig. 17B). The reason for this discrepancy is unclear; however, these data support the hypothesis that Barnyard TMP is post-translationally processed prior to or during virion assembly.

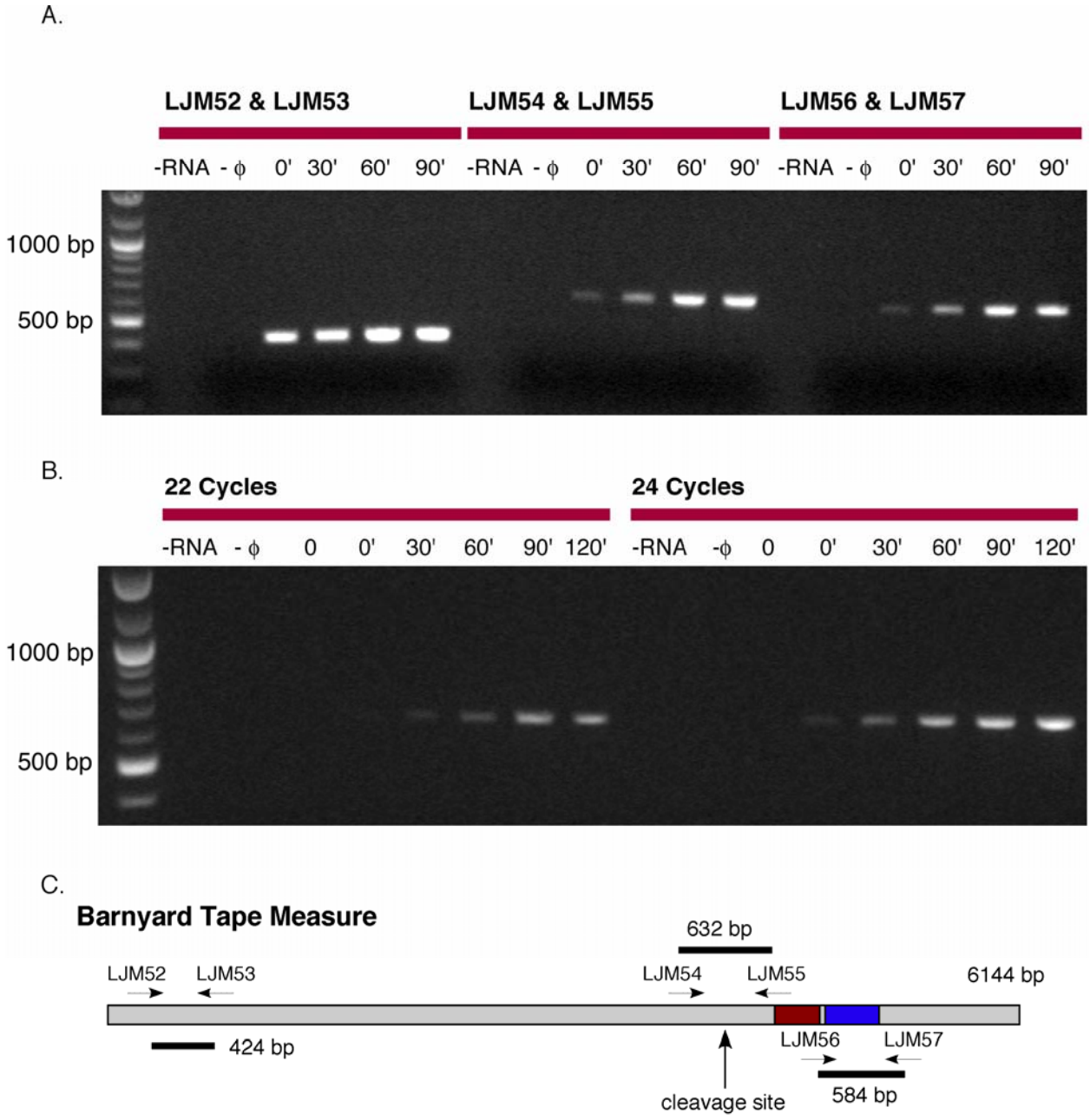


Figure 17. RT-PCR analysis of TMP expression during host infection.

Figure 17: RT-PCR analysis of TMP expression during infection of *M. smegmatis*. Total mRNA was prepared (Ambion) from cells infected with Barnyard at an M.O.I. of 12, and 50 ng of total mRNA from various time-points was subjected to **A.** qualitative RT-PCR (QIAGEN) with three primer pairs that amplify three separate regions of the *tmp* and **B.** semi-quantitative RT-PCR analysis with the pair (LJM54 and LJM55) and amplifies over the cleavage site. **C.** Sites of annealing for the primer pairs relative to the *tmp* are indicated. Transcription of tape measure is evident at even the earliest stages of infection and increases in amount at up to 120 minutes post-infection, which is approximately when lysis is first observed.

4.2.4 Infection-Independent Expression of the Barnyard TMP

In order to further dissect the regulation of this TMP cleavage, a construct was created (pLAM16) in which full-length TMP is expressed under the control of the inducible *acetamidase* promoter ($P_{acetamidase}$), and this was transformed into *M. smegmatis*. Time-course inductions were then carried out in media containing the inducer acetamide, and α-RLF antibodies were utilized to probe total protein samples from these inductions for the presence of full-length or cleaved TMP (Fig. 18A). These data show that full-length TMP can be expressed at low levels under inducible control in *M. smegmatis,* and importantly, this protein is not cleaved to generate the 70 kDa Rpf Motif containing fragment observed in mature phage. This suggests that the processing observed during infection is phage-mediated, either directly by a Barnyard encoded protein or through stimulation of a host protease.

Notably, prolonged expression of TMP is detrimental to the growth of *M. smegmatis* (Fig. 18B and C); TMP expressing cells show a 3- to 4-fold decrease in colony forming units (cfu) relative to initial viability when incubated for 24 hours in the presence of acetamide, whereas cells grown in identical media lacking inducer show a 10- to 20-fold increase in cfu over the same time period. Empty vector control cells increase 3- to 9-fold when grown with inducer and 2- to 7-fold when grown without it (data are from two independent experiments). Further, *M. smegmatis* containing pLAM16 will not form colonies on solid media containing inducer, and these cells appear to give rise to suppressor mutants at a fairly high frequency. The mechanism by which the growth inhibition of TMP expressing cells occurs, however, is unknown.

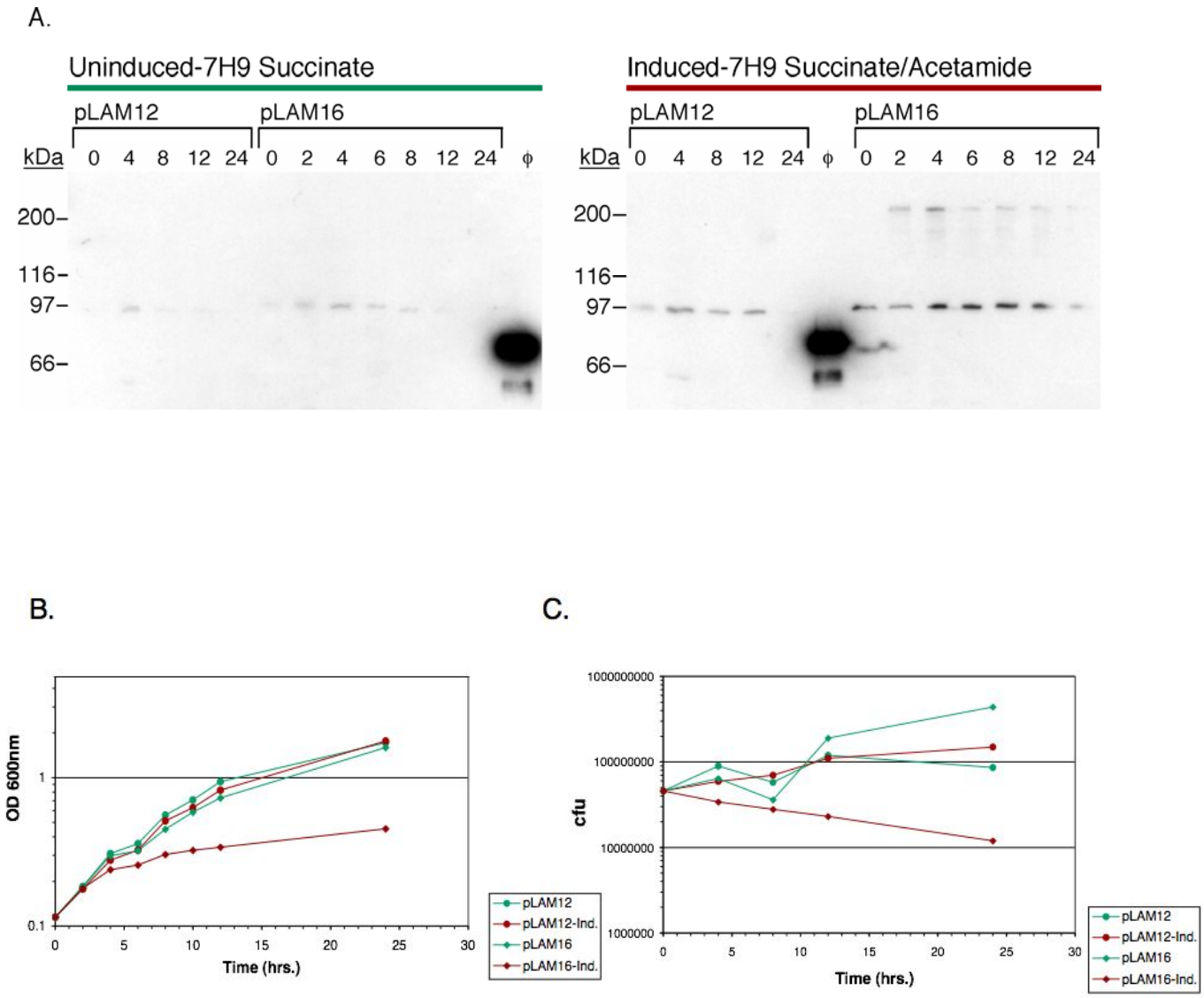


Figure 18. Expression of the Barnyard TMP in *M. smegmatis.*

Figure 18: Expression of the full-length Barnyard TMP in *M. smegmatis*. Plasmid pLAM16 contains the Barnyard TMP gene expressed under the control of the inducible *M. smegmatis acetamidase* promoter ($P_{acetamidase}$). **A.** *M. smegmatis* strains containing pLAM16 and the vector control (pLAM12) were grown in 7H9 with ADC overnight to early stationary phase and subcultured at an OD_{600} of 0.1 in 7H9 induction media (containing 0.2% succinate, 0.05% Tween 80 and KAN) with or without inducer (0.2% acetamide). These were grown with agitation at 37°C, and at the time points shown above, samples were removed for immunoblot analysis. Total protein was separated on a 7% SDS-PAGE gel and probed with a 1:125 dilution of purified α-RLF antibodies. Expression of an ~200 kDa protein band that likely represents full length TMP can be detected in the presence of inducer. This protein is not produced at high levels and is not processed to the 70 kDa form present during infection and in phage particles, indicating that this processing event is likely to be specific and dependent upon one or more phage proteins. **B.** *M. smegmatis* strains containing pLAM16 and the vector control (pLAM12) were grown as described above, and samples were removed for optical density determination at the time points indicated. **C.** Total colony forming units were also measured at various time points throughout the course of the experiment. Growth of the pLAM16 containing strain is comparable to vector control under non-inducing conditions. However, these cells show reduced growth rate and decreased viability in the presence of acetamide.

4.2.5 Localization of Barnyard TMP During Infection

A final important question concerns the localization and the fate of the Rpf Motif-containing portion of TMP during phage infection. I hypothesized that if this protein is acting to promote peptidoglycan cleavage, it might be associated with the host cell wall or membrane during the early stages of infection. Therefore, I performed experiments in which *M. smegmatis* cells were adsorbed with Barnyard at a high M.O.I. (15 or 30) and subsequently fractionated by differential centrifugation to separate host cell wall, membrane and cytoplasm. The α–RLF antibodies were then used to perform immunoblot analyses with equivalent amounts of protein from each fraction. Results of these experiments indeed suggest that the RLF-containing species is cell wall and/or membrane-associated (Fig. 19). However, although these data have been consistent and repeatable, the signal from the 70 kDa TMP fragment is typically quite low and is often obscured by the high background in these blots. The reason for this is unclear, although it is possible that a portion of the host-associated protein is lost during the fractionation process.

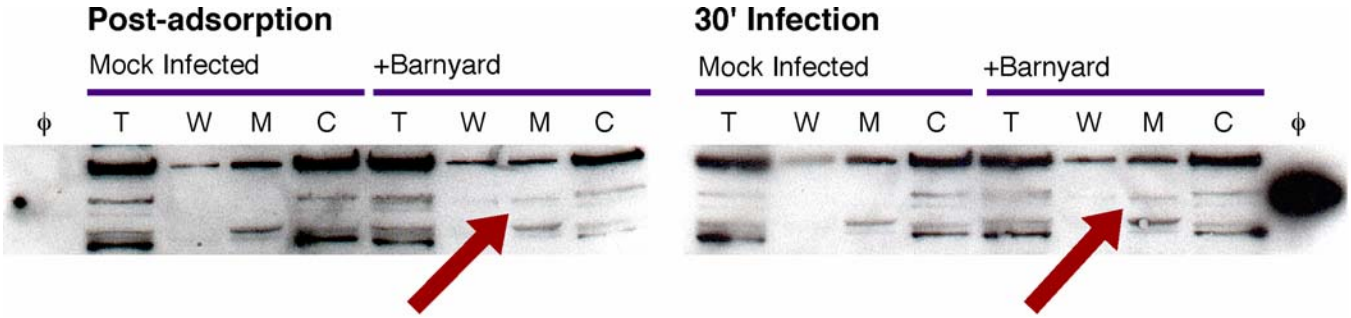


Figure 19. Subcellular localization of the C-terminal fragment of Barnyard TMP during infection of M. smegmatis.

Figure 19: Subscellular localization of the 70 kDa C-terminal fragment of Barnyard tape measure during infection of *M. smegmatis*. Cells from an overnight culture were diluted to an OD_{600} of 0.1 and infected with Barnyard at an M.O.I. of 15. Samples from this culture and an uninfected control were removed after 30 minutes of stationary incubation (post-adsorption) and after an additional 30 minutes of agitation at 37°C, and these were subjected to differential centrifugation to separate cell walls and cell membranes from the cytosolic components. Equivalent amounts of each fraction were separated on 7% SDS polyacrylamide gels and probed with a 1:125 dilution of affinity purified α-RLF antibodies. T = total protein, W = cell wall fraction, M = cell membrane fraction, and C = cytoplasmic fraction. These data indicate that the 70 kDa RLF-containing portion of TMP may localize to the cell wall and/or membrane during infection of *M. smegmatis*.

4.3 EVIDENCE FOR PROCESSING OF GILES TMP

Mycobacteriophage Giles is only the second phage found to contain a canonical Rpf Motif in its TMP. Therefore, it was hypothesized that this protein might also be subjected to similar processing. The α–RLF antibodies do not cross-react with the Giles TMP and thus, could not be utilized to address this question. However, LC MS/MS mass spectrometry performed on Giles virion proteins does suggest this TMP might be processed as well. That is, although this protein has a predicted size of 141 kDa, a number of peptides spanning the length of the protein were sequenced from an ~70 kDa band visible by SDS-PAGE and Coomassie staining (Fig. 20). This suggests that the Giles tape measure is processed near the middle of the protein; and therefore, both pieces migrate to the same location during electrophoresis. However, the potential for a more complex internal processing such as that observed for the TMP of the *L. monocytogenes* phage PSA (Zimmer *et al*., 2003) or a non-specific proteolytic degradation cannot be ruled out in this case. Interestingly, although the RLF antibodies do not recognize Giles TMP, polyclonal antibodies generated against the mycobacteriophage TM4 Motif 3 (gift from M. Piuri), which is the other motif found in this tape measure, detect a protein of ~40-50 kDa, which is further suggestive of proteolytic processing. However, a mechanism by which a fragment of this size could be generated that is consistent with the mass spectrometry data is unclear.

Figure 20. Identification of the Giles tape measure by mass spectrometry.

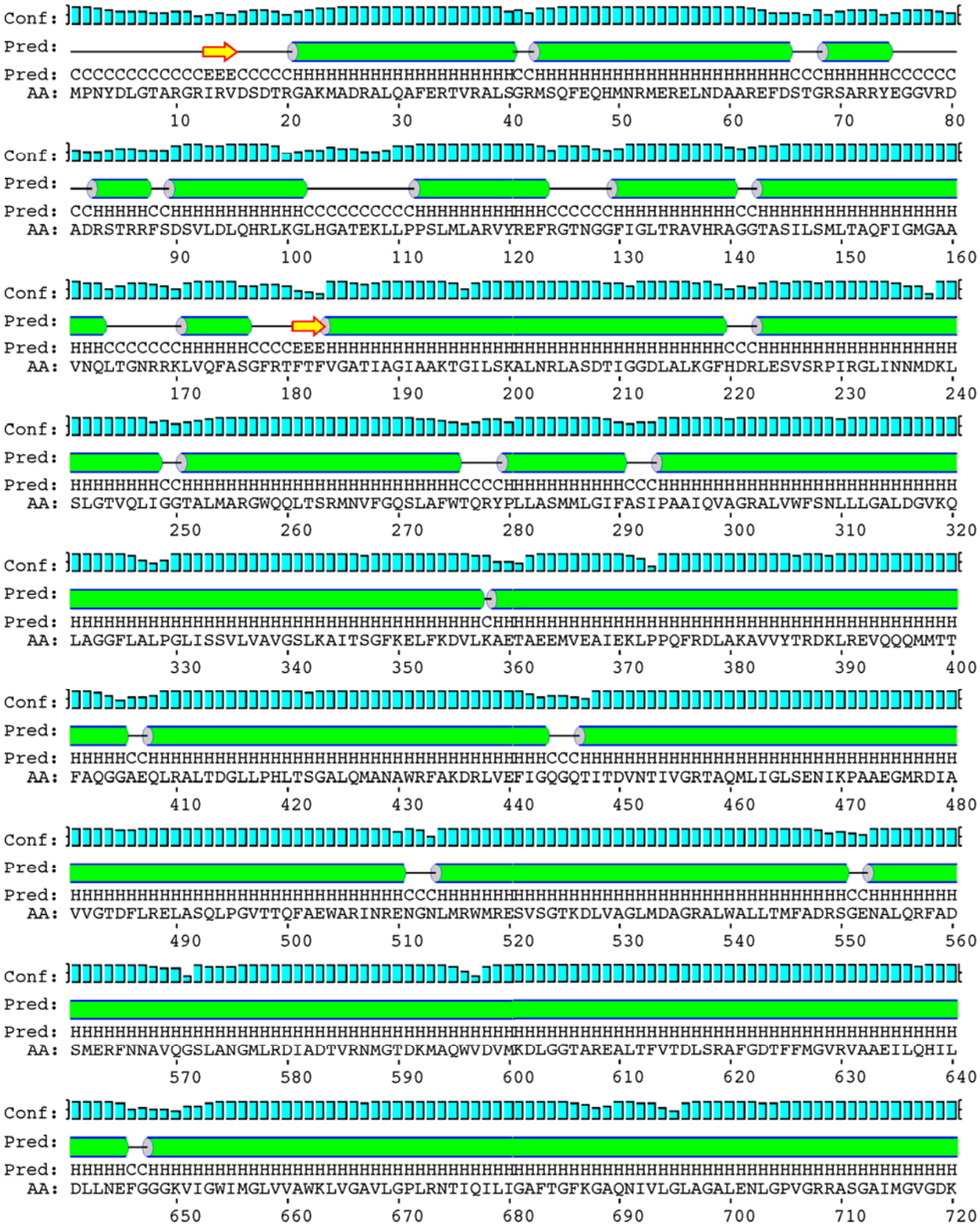
Figure 20: Indentification of the Giles tape measure by mass spectrometry. Phage structural proteins were separated on a 10% SDS polyacrylamide gel, and bands were excised and sequenced by mass spectrometry (ProtTech, Inc.). LC-MS/MS spectra were analyzed against a unique protein database consisting of all predicted Giles ORFs. Most TMP peptides sequenced (highlighted in red) were present in a band ~70 kDa, suggesting Giles TMP is processed near the middle of the protein.

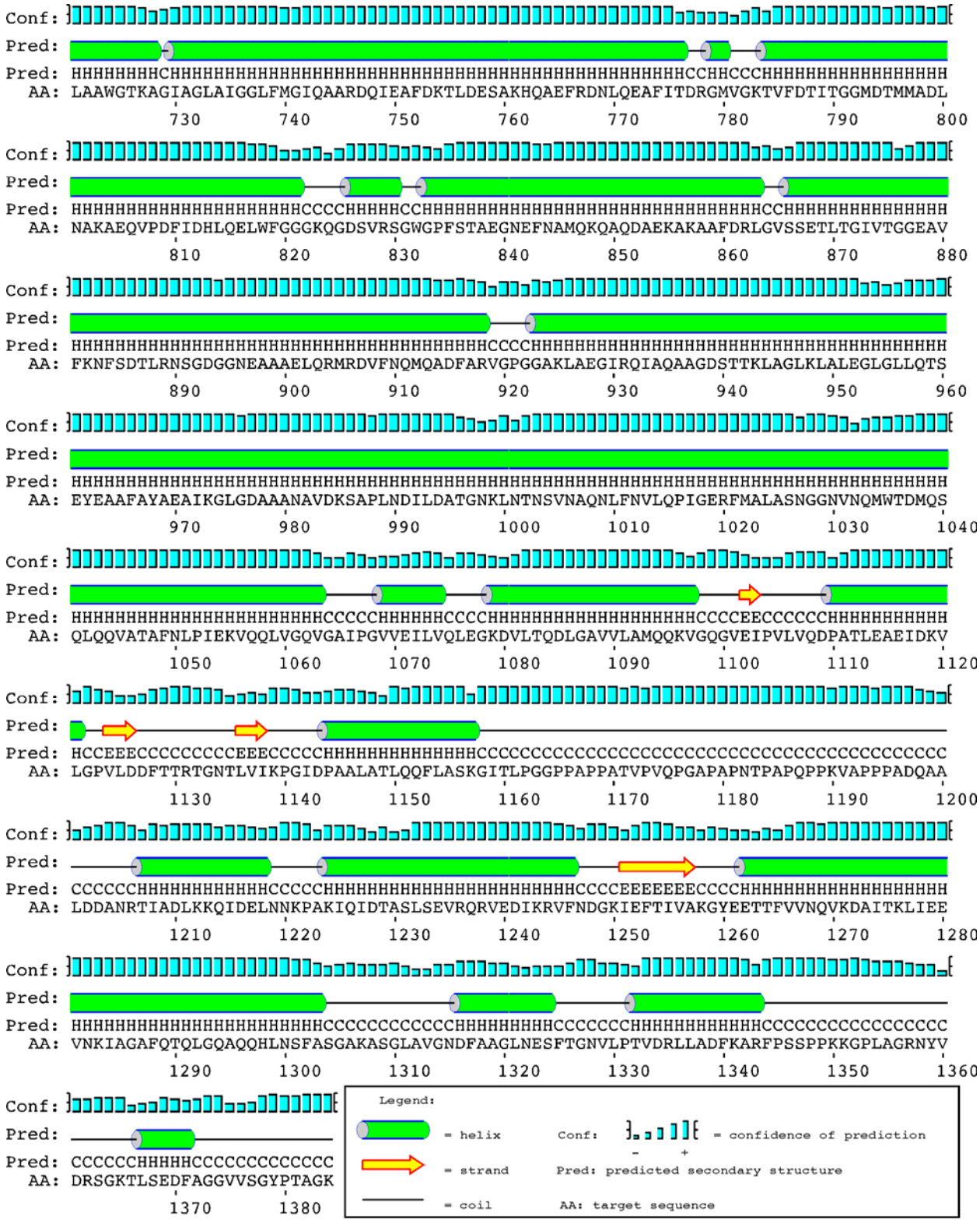
4.4 SECONDARY STRUCTURE PREDICTION OF BARNYARD AND GILES TMP FRAGMENTS

I have demonstrated that many of the mycobacteriophage tape measures are similar to TMPs from other phages in that they are predicted to be largely α-helical and to contain multiple regions with the potential to form coiled-coil structures (Chapter 3). However, I have further shown that the Barnyard, and possibly the Giles, TMPs are proteolytically processed in such a way as to generate two large protein fragments, both of which are present in mature phages. In addition, for many of the mycobacteriophage tape measures, the α-helical content did not appear to be evenly distributed throughout the protein, but rather, was concentrated more towards the N-termini of the proteins. Therefore, in order to better quantify this secondary structure distribution for the Barnyard and Giles TMPs, the predicted secondary structures of the individual tape measure fragments were examined.

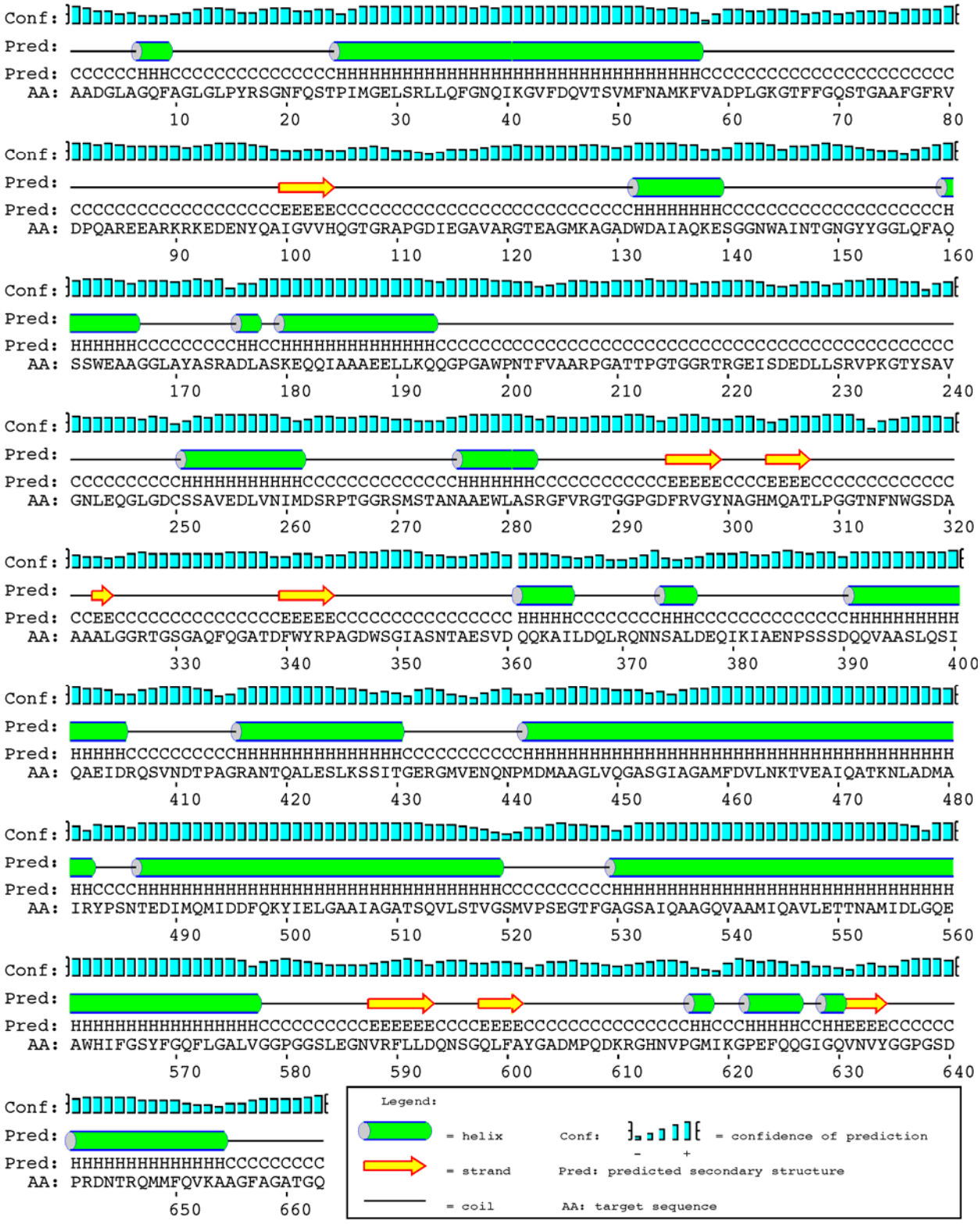
When the amino acid sequences of the Barnyard tape measure fragments were subjected to the secondary structure prediction algorithms PSIPRED (Jones, 1999, McGuffin *et al.*, 2000) and PredictProtein, PROF (Rost & Sander, 1993, Rost *et al.*, 1996, Rost *et al.*, 2004) it was observed that 87.4% of the residues in the N-terminal fragment (Fig. 21A) but only 37.7% in the 70 kDa C-terminus (Fig. 21B) were predicted to form α-helices (overall the estimate was 72.5%). Despite the differing α-helical content, both fragments contained regions with the potential to form coiled-coil interactions; two were strongly predicted in the N-terminus, and one was predicted in the C-terminal fragment (PredictProtein, COILS v2.2) (Lupas *et al*., 1991, Rost *et al*., 2004).

A.

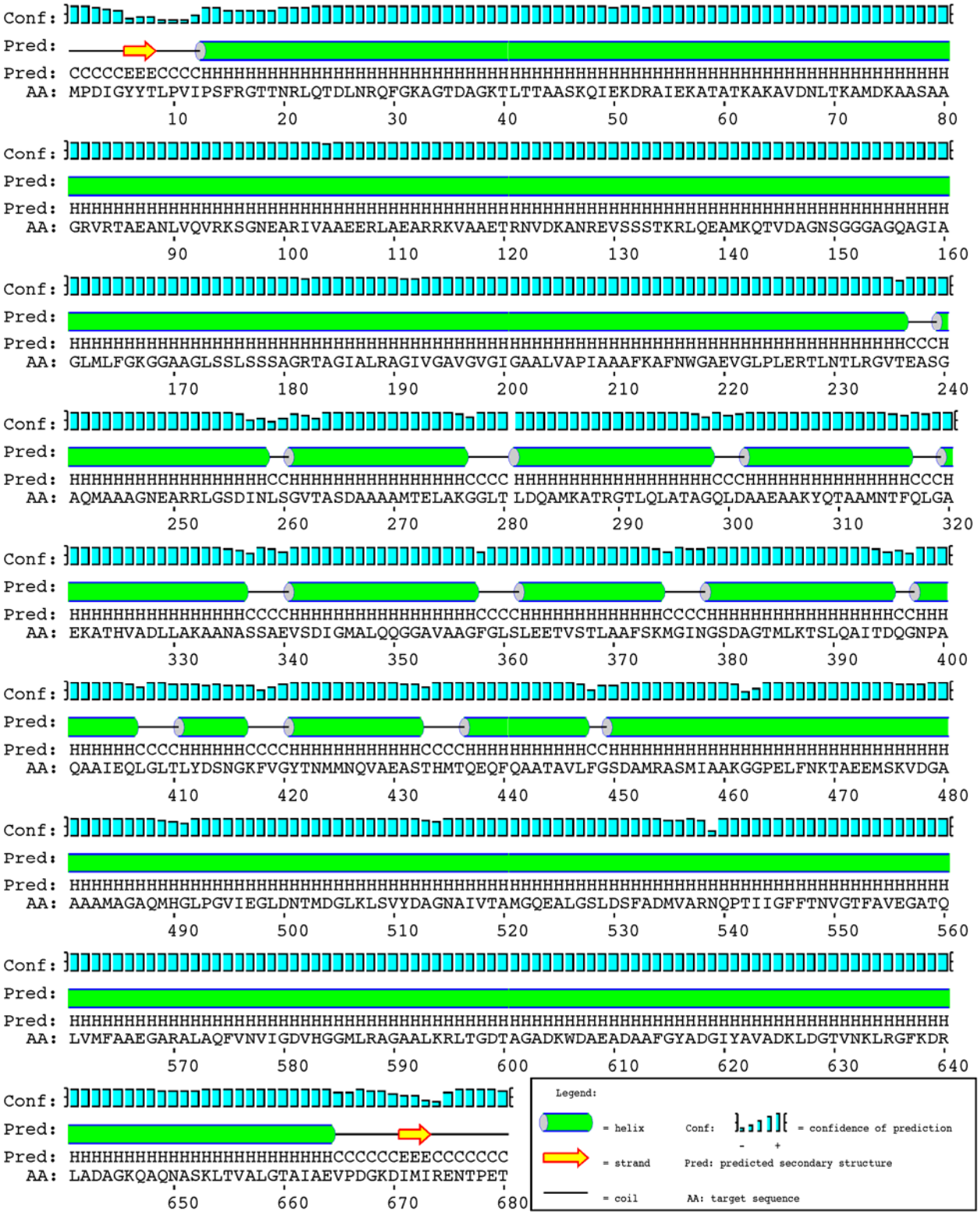

Conf:
Pred:
Pred: CCCCCCCCCCCCCEEECCCCCHHHHHHHHHHHHHHHHHHHHCCHHHHHHHHHHHHHHHHHHHHHHCCCHHHHHHCCCCCC
AA: MPNYDLGTARGRIRVDSDTRGAKMADRALQAFERTVRALSGRMSQFEQHMNRMERELNDAAREFDSTGRSARRYEGGVRD
10 20 30 40 50 60 70 80
Conf:
Pred:
Pred: CCHHHHHCCHHHHHHHHHHHHCCCCCCCCCCCHHHHHHHHHHHHCCCCCCHHHHHHHHHHCCHHHHHHHHHHHHHHHH
AA: ADRSTRRFSDSVLDLQHRLKGLHGATEKLLPPSLMLARVYREFRGTNGGFIGLTRAVHRAGGTASILSMLTAQFIGMGAA
90 100 110 120 130 140 150 160
Conf:
Pred:
Pred: HHHCCCCCCCHHHHHHCCCCEEEHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCHHHHHHHHHHHHHHHH
AA: VNQLTGNRRKLVQFASGFRTFTFVGATIAGIAAKTGILSKALNRLASDTIGGDLALKGFHDRLESVSRPIRGLINNMDKL
170 180 190 200 210 220 230 240
Conf:
Pred:
Pred: HHHHHHHHCCHHHHHHHHHHHHHHHHHHHHHHHHCCCCHHHHHHHHHHHCCCHHHHHHHHHHHHHHHHHHHHHHHHH
AA: SLGTVQLIGGTALMARGWQQLTSRMNVFGQSLAFWTQRYPLLASMMLGIFASIPAAIQVAGRALVWFSNLLLGALDGVKQ
250 260 270 280 290 300 310 320
Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
AA: LAGGFLALPGLISSVLVAVGSLKAITSGFKELFKDVLKAETAEEMVEAIEKLPPQFRDLAKAVVYTRDKLREVQQQMMTT
330 340 350 360 370 380 390 400
Conf:
Pred:
Pred: HHHHHCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
AA: FAQGGAEQLRALTDGLLPHLTSGALQMANAWRFAKDRLVEFIGQGQTITDVNTIVGRTAQMLIGLSENIKPAAEGMRDIA
410 420 430 440 450 460 470 480
Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCHHHHHHHH
AA: VVGTDFLRELASQLPGVTTQFAEWARINRENGNLMRWMRESVSGTKDLVAGLMDAGRALWALLTMFADRSGENALQRFAD
490 500 510 520 530 540 550 560
Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
AA: SMERFNNAVQGSLANGMLRDIADTVRNMGTDKMAQWVDVMKDLGGTAREALTFVTDLSRAFGDTFFMGVRVAAEILQHIL
570 580 590 600 610 620 630 640
Conf:
Pred:
Pred: HHHHHCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
AA: DLLNEFGGGKVIGWIMGLVVAWKLVGAVLGPLRNTIQILIGAFTGFKGAQNIVLGLAGALENLGPVGRRASGAIMGVGDK
650 660 670 680 690 700 710 720


Conf:
Pred:
Pred: HHHHHHHHCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCHHCCCHHHHHHHHHHHHHHHH
AA: LAAWGTKAGIAGLAIGGLFMGIQAARDQIEAFDKTLDESAKHQAEFRDNLQEAFITDRGMVGKTVFDTITGGMDTMMADL
730 740 750 760 770 780 790 800
Pred: HHHHHHHHHHHHHHHHHHHHCCCCHHHHHCCHHHHHHHHHHHHHHHHHHHHHHHHHHHCCHHHHHHHHHHHHHH
AA: NAKAEQVPDFIDHLQELWFGGGKQGDSVRSGWGPFSTAEGNEFNAMQKQAQDAEKAKAAFDRLGVSSETLTGIVTGGEAV
810 820 830 840 850 860 870 880
Pred: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
AA: FKNFSDTLRNSGDGGNEAAAELQRMRDVFNQMQADFARVGPGGAKLAEGIRQIAQAAGDSTTKLAGLKLALEGLGLLQTS
890 900 910 920 930 940 950 960
Pred: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
AA: EYEAAFAYAEAIKGLGDAAANAVDKSAPLNDILDATGNKLNTNSVNAQNLFNVLQPIGERFMALASNGGNVNQMWTDMQS
970 980 990 1000 1010 1020 1030 1040
Pred: HHHHHHHHHHHHHHHHHHHHHHHHCCCCCHHHHHHCCCCHHHHHHHHHHHHHHHHHHHCCCCEECCCCCCHHHHHHHHHHH
AA: QLQQVATAFNLPIEKVQQLVGQVGAIPGVVEILVQLEGKDVLTQDLGAVVLAMQQKVGQGVEIPVLVQDPATLEAEIDKV
1050 1060 1070 1080 1090 1100 1110 1120
Pred: HCCEEECCCCCCCCCEEECCCCCCHHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: LGPVLDDFTTRTGNTLVIKPGIDPAALATLQQFLASKGITLPGGPPAPPATVPVQPGAPAPNTPAPQPPKVAPPPADQAA
1130 1140 1150 1160 1170 1180 1190 1200
Pred: CCCCCCHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHHHHHCCCCEEEEEEECCCCHHHHHHHHHHHHHHHHH
AA: LDDANRTIADLKKQIDELNNKPAKIQIDTASLSEVRQRVEDIKRVFNDGKIEFTIVAKGYEETTFVVNQVKDAITKLIEE
1210 1220 1230 1240 1250 1260 1270 1280
Pred: HHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCCHHHHHHHHCCCCCCCHHHHHHHHHHCCCCCCCCCCCCCCCCC
AA: VNKIAGAFQTQLGQAQQHLNSFASGAKASGLAVGNDFAAGLNESFTGNVLPTVDRLLADFKARFPSSPPKKGPLAGRNYV
1290 1300 1310 1320 1330 1340 1350 1360
Pred: CCCCCCHHHHHCCCCCCCCCCCCC
AA: DRSGKTLSEDFAGGVVSGYPTAGK
1370 1380
Legend:
= helix
= strand
= coil
Conf: = confidence of prediction
Pred: predicted secondary structure
AA: target sequence

B.


Conf:
Pred:
Pred: CCCCCCHHHCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCC
AA: AADGLAGQFAGLGLPYRSGNFQSTPIMGELSRLLQFGNQIKGVFDQVTSVMFNAMKFVADPLGKGTFFGQSTGAAFGFRV
10 20 30 40 50 60 70 80
Pred: CCCCCCCCCCCCCCCCCCEEEEECCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHCCCCCCCCCCCCCCCCCCCH
AA: DPQAREEARKRKEDENYQAIGVVHQGTGRAPGDIEGAVARGTEAGMKAGADWDAIAQKESGGNWAINTGNGYYGGLQFAQ
90 100 110 120 130 140 150 160
Pred: HHHHHHCCCCCCCCCHHCCHHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: SSWEAAGGLAYASRADLASKEQQIAAAEELLKQQGPGAWPNTFVAARPGATTPGTGGRTRGEISDEDLLSRVPKGTYSAV
170 180 190 200 210 220 230 240
Pred: CCCCCCCCCCHHHHHHHHHHHCCCCCCCCCCCCCCCHHHHHHHCCCCCCCCCCCCCEEEEECCCCEEEECCCCCCCCCCCCC
AA: GNLEQGLGDCSSAVEDLVNIMDSRPTGGRSMSTANAAEWLASRGFVRGTGGPGDFRVGYNAGHMQATLPGGTNFNWGSDA
250 260 270 280 290 300 310 320
Pred: CCEECCCCCCCCCCCCCCCEEEEECCCCCCCCCCCCCCCCC HHHHHCCCCCCCCHHHCCCCCCCCCCCCCCHHHHHHHHHH
AA: AAALGGRTGSGAQFQGATDFWYRPAGDWSGIASNTAESVD QQKAILDQLRQNNSALDEQIKIAENPSSSDQQVAASLQSI
330 340 350 360 370 380 390 400
Pred: HHHHHCCCCCCCCCCHHHHHHHHHHHHHHHHCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
AA: QAEIDRQSVNDTPAGRANTQALESLKSSITGERGMVENQNPMDMAAGLVQGASGIAGAMFDVLNKTVEAIQATKNLADMA
410 420 430 440 450 460 470 480
Pred: HHCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
AA: IRYPSNTEDIMQMIDDFQKYIELGAAIAGATSQVLSTVGSMVPSEGTFGAGSAIQAAGQVAAMIQAVLETTNAMIDLGQE
490 500 510 520 530 540 550 560
Pred: HHHHHHHHHHHHHHHHHCCCCCCCCCCEEEEEECCCCEEEECCCCCCCCCCCCCCCCHHCCCHHHHHCCHHEEEECCCCCC
AA: AWHIFGSYFGQFLGALVGGPGGSLEGNVRFLLDQNSGQLFAYGADMPQDKRGHNVPGMIKGPEFQQGIGQVNVYGGPGSD
570 580 590 600 610 620 630 640
Pred: HHHHHHHHHHHHHHHCCCCCCCCC
AA: PRDNTRQMMFQVKAAGFAGATGQ
650 660
Legend:
= helix
= strand
= coil
Conf: - +
= confidence of prediction
Pred: predicted secondary structure
AA: target sequence

C.


Conf:
Pred:
Pred: CCCCCEEECCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
AA: MPDIGYYTLPVIPSFRGTTNRLQTDLNRQFGKAGTDAGKTLTTAASKQIEKDRAIEKATATKAKAVDNLTKAMDKAASAA
10 20 30 40 50 60 70 80
Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
AA: GRVRTAEANLVQVRKSGNEARIVAAEERLAEARRKVAAETRNVDKANREVSSSTKRLQEAMKQTVDAGNSGGGAGQAGIA
90 100 110 120 130 140 150 160
Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCH
AA: GLMLFGKGGAAGLSSLSSSAGRTAGIALRAGIVGAVGVGIGAALVAPIAAAFKAFNWGAEVGLPLERTLNTLRGVTEASG
170 180 190 200 210 220 230 240
Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHHCCHHHHHHHHHHHHHHCCCC HHHHHHHHHHHHHHHHHHCCCHHHHHHHHHHHHHHCCCH
AA: AQMAAAGNEARRLGSDINLSGVTASDAAAAMTELAKGGLT LDQAMKATRGTLQLATAGQLDAAEAAKYQTAAMNTFQLGA
250 260 270 280 290 300 310 320
Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHCCCCHHHHHHHHHHHHHHHHCCCCHHHHHHHHHHHHCCCCHHHHHHHHHHHHHHHHCCHHH
AA: EKATHVADLLAKAANASSAEVSDIGMALQQGGAVAAGFGLSLEETVSTLAAFSKMGINGSDAGTMLKTSLQAITDQGNPA
330 340 350 360 370 380 390 400
Conf:
Pred:
Pred: HHHHHHCCCCHHHHHHCCCCHHHHHHHHHHHHCCCCHHHHHHHHHHCCHHHHHHHHHHHHHHHHHHHHHHHHHHHH
AA: QAAIEQLGLTLYDSNGKFVGYTNMMNQVAEASTHMTQEQFQAATAVLFGSDAMRASMIAAKGGPELFNKTAEEMSKVDGA
410 420 430 440 450 460 470 480
Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
AA: AAAMAGAQMHGLPGVIEGLDNTMDGLKLSVYDAGNAIVTAMGQEALGSLDSFADMVARNQPTIIGFFTNVGTFAVEGATQ
490 500 510 520 530 540 550 560
Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
AA: LVMFAAEGARALAQFVNVIGDVHGGMLRAGAALKRLTGDTAGADKWDAEADAAFGYADGIYAVADKLDGTVNKLRGFKDR
570 580 590 600 610 620 630 640
Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHHHHHHHHCCCCCCEEECCCCCCC
AA: LADAGKQAQNASKLTVALGTAIAEVPDGKDIMIRENTPET
650 660 670 680
Legend:
= helix
= strand
= coil
Conf: = confidence of prediction
Pred: predicted secondary structure
AA: target sequence

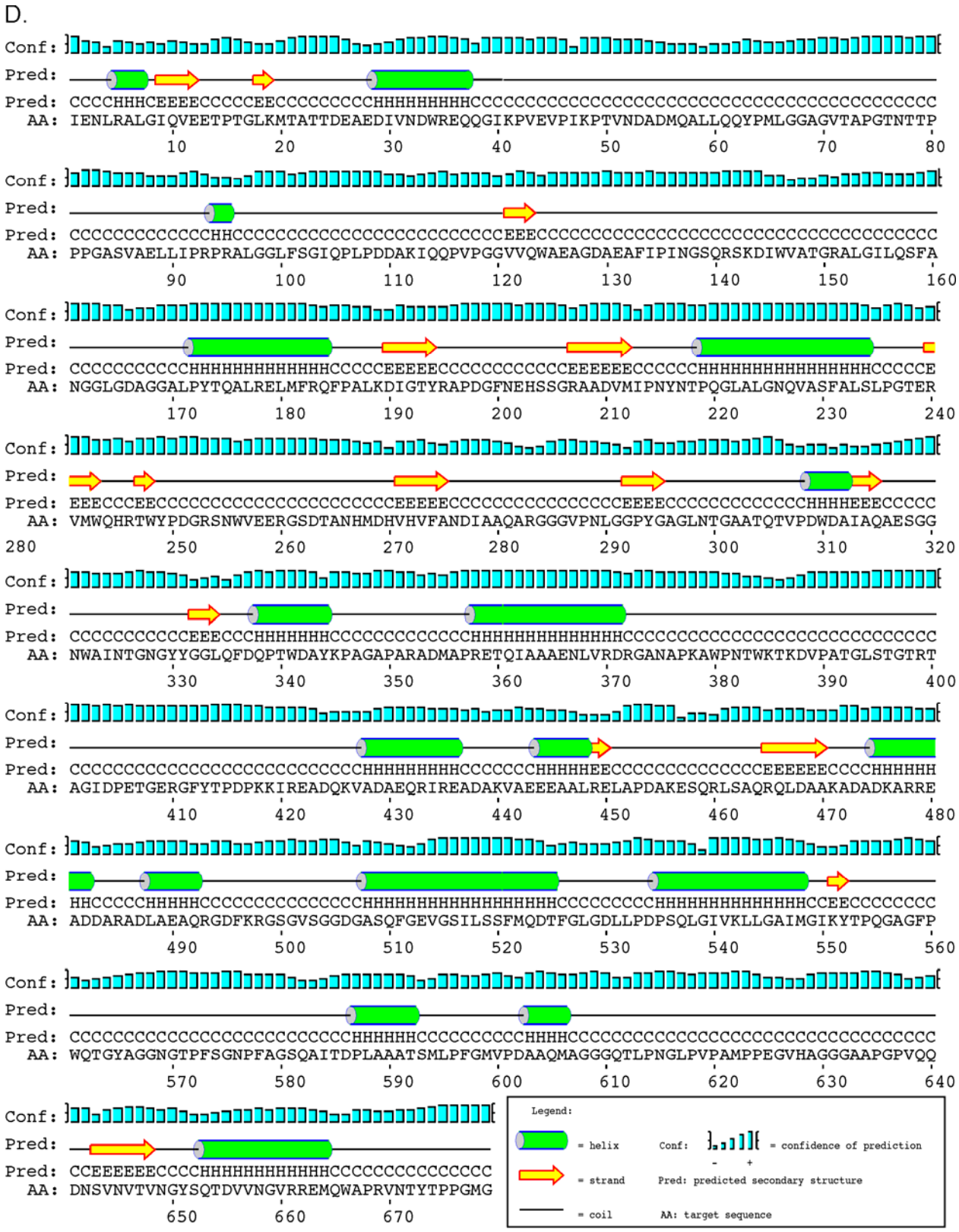


Figure 21. Secondary structure predictions of the Barnyard and Giles TMPs.

Figure 21: Secondary structure predictions of the Barnyard and Giles TMPs. The Barnyard TMP fragments, **A.** the 138 kDa N-terminus and **B.** the 70 kDa C-terminus, were subjected to secondary structure prediction analysis (PSIPRED) and the graphical output is presented below. The Giles tape measure was arbitrarily split into two pieces of ~70 kDa, **C.** a 680 residue N-terminal piece and **D.** a 679 residue C-terminal piece, and the same analysis was performed. These data illustrate that similar to what has been observed for other phages, the Barnyard and Giles TMPs are more highly α-helical near their N-termini with their C-termini predicted to be mostly random coil.

Similar observations were made for the Giles TMP gp20. The entire protein was predicted to be 62.1% α-helical; however, the α-helical content of the C-terminus of the protein was strikingly lower than that of the N-terminus (Fig. 21C and D). When this protein is arbitrarily split in the middle and each portion is examined (PredictProtein, PROF), it is found that 91.8% of the N-terminus, but only 19.2% of the C-terminus, is predicted to be in an α-helical conformation. Each half, however, contains regions with a high coiled-coil probability (PredictProtein, COILS v2.2). The reasons for this disparity in α-helical content between the N- and C-terminal regions of the tape measure is not known; however, this could be reflective of differing roles for these protein fragments during virion assembly or host infection.

Interestingly, both fragments of Barnyard tape measure were also strongly predicted to contain a number of putative transmembrane domains (TMpred) (Hofmann & W., 1993); depending on topology, six or seven were predicted in the 138 kDa N-terminal portion, and two or three were found in the 70 kDa C-terminal fragment of the protein. Even when a more stringent method of prediction is employed, four are predicted in the N-terminus, and one is found in the C-terminal fragment (PredictProtein, PHD) (Rost *et al.*, 1996, Rost, 1996). The residues encompassing the Rpf Motif and Motif 2 are not predicted to be membrane-associated. These data suggest that both portions of the Barnyard TMP may insert into the membrane during infection. Again, the same was also true for the Giles TMP, which was predicted to contain as many as six or seven transmembrane helices (TMpred); four in the N-terminal half and two or three in the C-terminus. Although it has not been definitely determined if the Giles tape measure is similarly processed, the individual halves of this protein also contain a number of regions with the potential to insert into the membrane. Further analyses will be required to determine if this is indeed the case.

4.5 CONCLUSIONS

In the course of investigating the state of the Rpf Motif within the Barnyard TMP, I have shown that this protein is actually present in two pieces within mature virion particles: an ~138 kDa N-terminal portion and an ~70 kDa C-terminal fragment that contains the Rpf Motif (Fig. 15). Mass spectrometric analysis of the tape measure fragments have revealed a putative cleavage site between K1384 and A1385, although the mechanism by which the processing occurs is unknown. When TMP expression is induced in *M. smegmatis,* α-RLF antibodies detect only low levels of the full-length protein, and cleavage is not observed (Fig. 18A). This suggests that processing is phage-mediated, either directly by a Barnyard encoded protein or through stimulation of a host protease. It is also notable that both fragments of the cleaved protein are present in mature virions, unlike what has been observed in other phages such as λ and the lactococcal phages Tuc2009 and TP901-1, although in this latter case, the fate of the cleaved TMP fragments have not been definitively determined (Hendrix & Casjens, 1974, Pedersen *et al.*, 2000, Kenny *et al.*, 2004, Mc Grath *et al.*, 2006). Further investigation is required to dissect the regulation of the Barnyard TMP processing event and to determine if it is assembly dependent, as has been observed with λ TMP cleavage (Tsui & Hendrix, 1983). As there is evidence that the Giles tape measure may also be processed in a similar way, future work should focus on this cleavage event in Giles and potentially in other mycobacteriophage tape measures to determine if proteolytic cleavage is a general feature of these proteins.

The individual fragments of the Barnyard and Giles TMPs – assuming cleavage near the middle of the Giles protein – are predicted to possess very different secondary structures. While the N-terminal fragments are largely α-helical, the C-terminal fragments are predicted to be mostly

random coil. The significance of this remains to be determined; however, it is noteworthy that the Cluster A tape measures, which at ~800 residues are smaller than the other mycobacteriophage TMPs, do not display the same heterogeneity. Rather, these are largely α-helical (>85%) across their entire length and do not contain any of the identified tape measure motifs (discussed in Chapter 3). It is tempting to speculate that longer, motif-containing tape measures that display a structural disparity between their N- and C-termini perhaps arose from a fusion of two separate structural proteins; and therefore, although the whole protein now functions in tail length determination, the C-terminal fragment must be cleaved to assume a separate role, either in assembly or infection (e.g. murein hydrolysis). In λ, it is the C-terminal (cleaved) part of the protein that interacts with the baseplate to form the initiator, but little is known about mycobacteriophage assembly, and it is difficult to draw clear parallels. Further work would be required to decipher the importance of the proteolytic cleavage of these tape measures.

Interestingly, the Barnyard tape measure is somewhat longer than would be needed to span the length of the phage tail as an extended α-helix; the 2047 residue protein would extend ~307 nm, and the Barnyard tail is 250-260 nm. However, neither the N-terminal cleavage product (an estimated 208 nm) nor the C-terminal fragment (an estimated 99 nm) would be long enough to span a tail of this size. An additional layer of complexity is added by the fact that these analyses have also suggested that an additional processing event may occur, and up to 86 amino acid residues may be cleaved from the N-terminus of the Barnyard TMP. The uncleaved TMP without this fragment would generate a tail of ~294 nm, and the processed form of the TMP missing the 86 residue N-terminus would produce a 195 nm tail, neither of which matches the size of the mature tail. Intriguingly, if both Motif 2 and the Rpf Motif (~230 residues) are subtracted from the full-length protein, a tail of 273 nm would be generated, although the mechanism by which these

regions might be excluded is hard to fathom. Therefore, it is remains unclear if there are specific regions of the Barnyard TMP that do not contribute to tail-length determination, and if so, the identity of such regions cannot be inferred from these analyses. In contrast to what is observed with Barnyard, the Giles tape measure at 1360 residues would predict a tail of 204 nm, and this closely agrees with the observed size of the tail (Morris *et al*., 2008). Therefore, this peculiar feature of the relationship between the Barnyard tape measure and tail length may be unrelated to the presence of the motifs embedded within this protein.

Tape measure cleavage can be visualized during time-course infections of *M. smegmatis*, as the full-length protein, which is initially expressed, is subsequently cleaved into the processed form of the protein present in mature phage (Fig. 16). The characterization of this as proteolytic cleavage is supported by the fact that no evidence for mRNA processing is observed in RT-PCR analysis of the tape measure transcript in infected cells (Fig. 17A and B). The cleaved form of the protein that is observed at the earliest time points of time-course infections – prior to the appearance of full-length protein – likely corresponds to the tape measures from infecting phages associating with the host cells. This is supported by the observation that the early-appearing, cleaved form of the protein is also present in cells treated with inhibitors of transcription or translation, while the subsequent appearance of full-length protein and later appearance of the cleaved form are no longer observed (Fig. 16). Interestingly, expression of full-length TMP and its subsequent cleavage are visible in infected cells up to eight hours P.I., despite the fact that these cultures were infected at a high M.O.I. and as such were expected to undergo only one round of infection. This suggests that multiple rounds of infection are in fact occurring, and this is discussed in further detail in Appendix.

Intriguingly, the amount of the cleaved, Rpf-Motif containing fragment of TMP that is associated with *M. smegmatis* appears to increase at the earliest time points following phage adsorption (Fig. 16). As the cell samples analyzed in these time-course infection experiments were washed prior to separation by SDS-PAGE, this observation strongly suggests that the TMP protein becomes stably associated with the host bacterium during infection. In λ, there is evidence that the TMP is also injected into the host cell and may be associated with the membrane, although the subcellular destination of this protein was not determined (Roessner & Ihler, 1984). Further, the tape measure-like protein of T5 has been shown to insert into the cellular membrane during infection and into liposomes *in vitro* (Guihard *et al*., 1992, Lambert *et al*., 1998). In my experiments with mycobacteriophage Barnyard, infected cells were fractionated, and it was determined that the 70 kDa Rpf Motif–containing fragment of Barnyard TMP appears to be associated with the host cell wall and/or membrane during infection (Fig. 19). Although the signal in these experiments was low and the background was high, possibly due to the α–RLF antibodies reacting with *M. smegmatis* Rpf proteins, the results were consistent and reproducible. Secondary structure predictions of the Barnyard TMP indeed indicate that both cleaved fragments contain multiple putative transmembrane domains; as many as six or seven in the N-terminal fragment and two to three in the 70 kDa C-terminus. This suggests the intriguing possibility that both fragments of the tape measure may insert into the host membrane and contribute to pore formation. Probing infected cells with antibodies generated against other Barnyard structural proteins as well as the N-terminal fragment of tape measure would be useful in testing the hypothesis.

5.0 ASSESSMENT OF THE ACTIVITY OF THE MYCOBACTERIOPHAGE RPF MOTIFS

5.1 INTRODUCTION

There are several lines of evidence suggesting that the Rpf proteins found in a number of bacteria function as resuscitation factors and possess muralytic activity that stimulates bacterial growth under certain conditions. Importantly, although Rpfs often contain accessory domains, this activity appears to be mediated by the ~70 amino acid conserved 'Rpf Motif' found in these proteins. In several cases it has been reported that this motif alone – residues A42-L118 of *M. luteus* Rpf and residues N255-R362 of *M. tuberculosis* RpfB – is sufficient for resuscitation of *M. luteus* (Mukamolova *et al*., 2002a, Cohen-Gonsaud *et al*., 2005). However, the data are far from definitive, as the same fragment of RpfB could bind to an artificial murein–like polymer but did not show activity in an *in vitro* test for lysozyme activity (Cohen-Gonsaud *et al*., 2005). One possible reason for this is that the artificial lysozyme substrate provided in these experiments may not have contained the particular bond recognized and cleaved by the Rpf proteins. It has further been observed that Rpf proteins are generally quite unstable and lose activity readily during storage, even for short periods of time. Therefore, it is also possible that the purified Rpf fragments are unstable or catalytically dead, and the resuscitation activity observed in culture results from the interaction of this motif with other cell wall modifying enzymes, such as the putative

mycobacterial endopeptidase RipA, a protein that associates with RpfB and RpfE in *M. tuberculosis* (see Section 1.5.4) (Hett *et al.*, 2007).

The instability of Rpf proteins and the highly specific culture conditions under which they are apparently required has made Rpf activity assays particularly challenging, especially in hardy, fast-growing organisms such as *M. smegmatis*. One way in which this problem has been overcome is to introduce a plasmid into these bacteria that expresses the *rpf* gene under control of the inducible *acetamidase* promoter (Mukamolova *et al*., 2002b, Parish *et al*., 1997). Growth stimulation is apparent in *M. smegmatis* cells containing such a plasmid, as they have been reported to display a reduced lag phase when inoculated at low densities in minimal media and an increased colony size relative to controls on minimal media, even in the absence of inducer (Mukamolova *et al*., 2002b). Interestingly, however, these same strains display a reduced growth rate on rich media (M. Young and G. Mukamolova, personal communication), a phenotype that may be attributed to the ability of the Rpf proteins to hydrolyze bonds within the peptidoglycan of the cell wall.

The muralytic activity of the *M. luteus* Rpf protein has also been demonstrated *in vivo* in the Gram-negative model organism *E. coli*. Strikingly, expression from a vector that directs secretion of this protein into the *E. coli* periplasm results in significant cell lysis (Mukamolova *et al*., 2006). No activity is observed in cells expressing only the Rpf Motif, as this construct is not expressed well. It is notable that the Barnyard Rpf Motif is also not expressed to high levels in *E. coli* (Chapter 4), and this might reflect an inherent instability of the motif when overexpressed in this manner. Importantly, lytic activity is dependent on the Rpf domain, as lysis is not observed in control cells expressing only the Rpf LysM domain or in those containing the empty plasmid (Mukamolova *et al*., 2006).

The Rpf domain is structurally similar to lysozymes and lytic transglycosylase enzymes, and there are numerous lines of evidence that Rpfs have murein hydrolase activity, both *in vivo* and *in vitro*. However, there are also data suggesting that these proteins might interact with peptidoglycan peptidases, and the specific biochemical activity of the Rpf proteins has not been determined (Hett *et al*., 2007). Lysozymes and lytic transglycosylases have limited sequence identity with one another and with the Rpf family of proteins; however, these enzymes do share a set of conserved residues (Koonin & Rudd, 1994, Dijkstra & Thunnissen, 1994, Mushegian *et al.*, 1996). One of these is the predicted active site glutamate – E54 in *M. luteus* Rpf – which corresponds to E35 in lysozyme, E162 in Slt35 and E478 in Slt70, all of which have been shown to be absolutely required for enzymatic activity (Malcolm *et al.*, 1989, Thunnissen *et al.*, 1994, van Asselt *et al.*, 1999).

The importance of this residue and a number of other conserved residues in the *M. luteus* Rpf was investigated for by Mukamolova and colleagues (2006). Mutant versions of this protein were assayed both *in vitro,* utilizing zymogram analyses and by measuring the release of fluorescein from labeled cell walls; and *in vivo*, as determined by measuring lysis of *E. coli* when mutant Rpfs were expressed into the periplasm. Biological activity was also monitored by expressing some of the mutants in *M. luteus* and *M. smegmatis* and measuring the growth of the strains under various conditions where Rpf activity can be detected; relevant results are summarized in Table 4 *(*Mukamolova *et al*., 2006). These data clearly indicate that while the conserved active-site glutamate is important for Rpf function, this is not absolutely required for activity, as proteins with mutations in this residue showed variable levels of impairment in the different assays utilized. Further, the loss of muralytic activity was somewhat correlated with a loss of physiological, growth stimulatory activity. The variability that is observed *in vivo* may be

due, in part, to the fact that all assays were performed in a wild-type Rpf background; or alternatively, the mutants may have retained the ability to interact with various protein partners. It is notable, however, that even the enzymatic activity of this protein was retained in some mutants, in contrast to what has been observed with other similar enzymes.

Table 4. Activity of various Rpf point mutants.

Mutant	*Zymograms*	*Hydrolyze Fluorescamine*	*Lyse E. coli*	*M. luteus Liquid media*	*M. luteus Solid media*	*M. smegmatis*
WT	Active	+++++	++++++	+++++	+++++	1
E54Q	RA	++	+++	++/+++	++	2×10^{-2}
E54A	N/A	+/++	+++	+++	++/+	6×10^{-4}
E54K	N/A	+	++	++/+++	+++	NDA
C53K	N/A	+/++	+++	+++	+	1×10^{-2}
C114T	N/A	++	+++/++++	++	++++	2×10^{-2}
C53K/ C114T	N/A	++	+++	++	NDA	NDA

RA = Reduced activity (non-quantitative); NDA = no detectable activity; N/A = not measured.
Percent activity: +++++ = Full; ++++ = 100-75%; +++ = 75-50%; ++ = 50-25%; + = 25-1%.
Resuscitation Index: r efers to calculated number of viable cells; value for cells expressing the wild-type *M. luteus* Rpf is set to 1.

A number of other resides conserved in the Rpf proteins were mutated and did not appear to be important for activity, although a few did show slight impairment in some assays, including Q72, D48, E94, D45, D60, E84, E100 and D104. However, members of the Rpf family also contain a pair of cysteine residues that are absolutely conserved in all bacterial representatives but are absent in both the Barnyard and Giles tape measure motifs. Strikingly, replacement of either or both of theses cysteines in the *M. luteus* Rpf with the corresponding residues from the Barnyard Rpf Motif resulted in a loss of activity, although the extent of the effect varied considerably depending on which assay was utilized (Table 4) (Mukamolova *et al*., 2006). It is unclear why this

is the case, although it was postulated that these cysteines may form a disulfide bond, and therefore, Rpf derivatives lacking these might be somewhat structurally unstable, a defect that could manifest to different degrees depending on the conditions. This is supported by the observation that recombinant wild-type Rpf only displayed muralytic activity *in vitro* when it was purified from the *E. coli* periplasm (Mukamolova *et al*., 2006). It was also hypothesized that these cysteines might be involved in the regulation of enzymatic activity or in the chelation of a metal ion, such as zinc, in the active site, as addition of DTT or mercaptoethanol did not abolish Rpf activity in zymograms (Mukamolova *et al*., 2006).

Although the situation is by no means resolved, these data do suggest that the resuscitation and growth enhancing activities of the Rpf proteins are due, at least in part, to their murein hydrolase activity. The question I have addressed relates to the function of the Rpf Motifs that are found embedded with mycobacteriophage tape measure proteins. The observation that bacteriophages often contain structural proteins with the ability to hydrolyze peptidoglycan, which are thought to assist in localized host cell wall degradation and viral DNA injection, suggests that these tape measure motifs are likely to perform the same function for the phages in which they are found. However, this would also not rule out the possibility that phages might also utilize these motifs to manipulate the growth of their hosts. Therefore, a second major focus of this project has involved determining if the phage Rpf domain has murein hydrolase and/or growth-stimulating activity.

5.2 IN VITRO ACTIVITY ASSAYS

In setting out to answer this question, a number of possible avenues were considered and, in some cases, explored. Prior to the discovery that the Barnyard tape measure is cleaved, attempts were made to express and purify a full-length recombinant His-tagged version of the protein for use in activity assays. This was unsuccessful, however, due to the fact that expression from this construct (pLAM4) was not detectable in *E. coli*. This is not entirely surprising, as it has been previously reported that TMPs are often difficult to express and can have toxic effects when they are overexpressed in bacteria (Abuladze *et al.*, 1994, Xu, 2001). Further, the Barnyard protein, at ~217 kDa, is exceptionally large. I, therefore, decided to express and purify a recombinant, His-tagged form of the 109 amino acid Rpf Motif (see Chapter 3). Although this did not express to high levels in *E. coli*, sufficient amounts were produced to generate polyclonal antibodies and for use in a variety of activity assays. However this protein did not display activity in zymograms – which involve separating proteins on SDS-PAGE gels containing *Micrococcus* cell wall material, renaturing in buffer, and staining with methylene blue to visualize clear areas that result from peptidoglycan degradation by proteins within the gel – or under any conditions in which resuscitation was tested, although there were a number of confounding variables, such as the lack of adequate positive controls, making these experiments difficult to interpret.

Whole phage particles were also separated in zymogram gels in order to determine if any structural proteins had the ability to hydrolyze lyophilized *Micrococcus lysodeikticus* cells (Sigma) or *M. luteus* cells. However, although preliminary results indicated that Barnyard might possess one or two proteins that have the ability to degrade cell wall material, no activity was found to be associated with proteins that would likely correspond to either the cleaved or full-length TMP, and

these results were difficult to confirm. Therefore, an alternate method known as a zymoblot was utilized to determine if Barnyard particles have hydrolase activity. This assay involves spotting whole phage particles (or the protein of interest) onto wells of agarose embedded with whole cell material (lyophilized *M. lysodeikticus* or killed *M. smegmatis*). These are then stained with methylene blue, and clearing is indicative of cell wall hydrolysis. This assay eliminates the denaturation of the proteins and thus the need for renaturation inherent to zymogram analyses. While these data suggest that the phage particles have an inherent hydrolase activity (Fig. 22), this is only a crude estimation of function, and it is difficult to draw any clear conclusions from these data. It was therefore necessary to probe for activity of the Rpf Motifs in alternate ways.

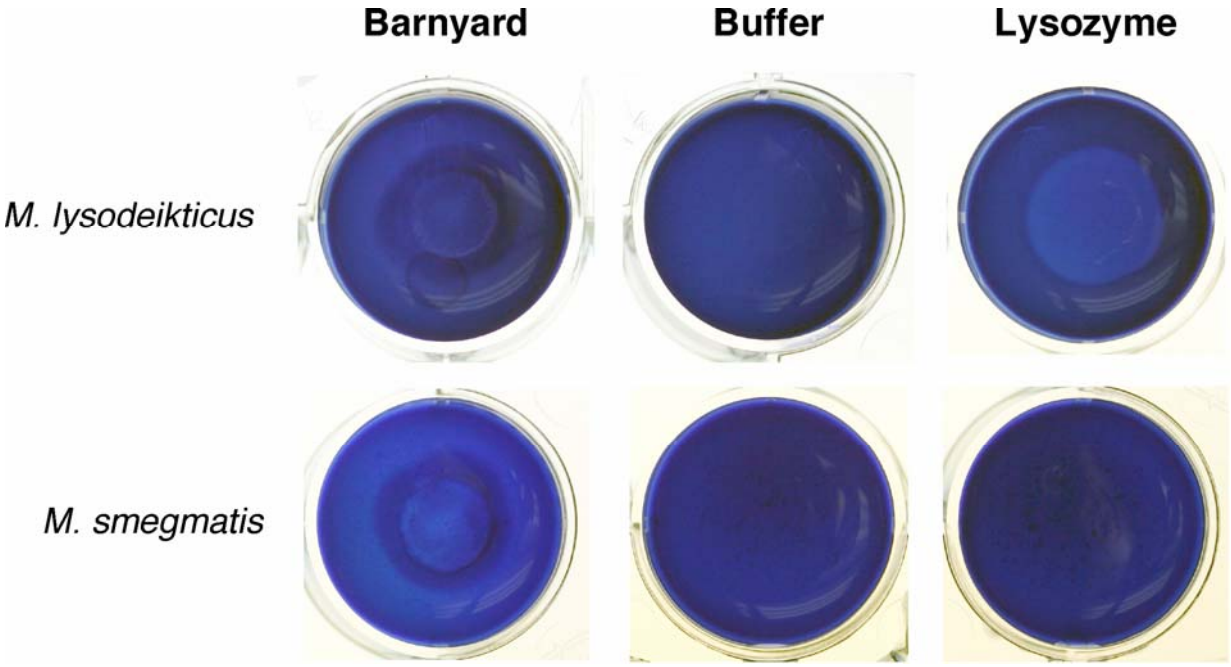


Figure 22. Zymoblot analysis of Barnyard virions.

Figure 22: Zymoblot analysis of Barnyard virions. Buffer (25 mM $NaPO_4$, pH 7, 10 mM $MgCl_2$, 0.1% Triton X-100) containing 0.1% lyophilized *M. lysodeikticus* or heat-killed *M. smegmatis* was solidified with 1% agarose in a 12-well flat-bottomed tissue culture plate (750 µl-1 ml/well). Barnyard (~10^{12} pfu), lysozyme (5 µg) or phage buffer control (20 µl) was spotted onto sterile paper discs in individual wells and incubated at room temperature overnight. Wells were stained with methylene blue for one hour and destained with dH_2O. Peptidoglycan hydrolysis can be visualized in the well containing *M. lysodeikticus* spotted with lysozyme and possibly in both wells spotted with Barnyard.

5.3 EXPRESSION OF TAPE MEASURE FRAGMENTS IN *M. SMEGMATIS*

It has previously been reported that the Rpf protein from *M. luteus* has the ability to modulate the growth of *M. smegmatis* when expressed in these cells. Therefore, vectors were designed and constructed to allow constitutive expression of both a cytoplasmic (pLAM6) and secreted (pLAM7) form of the 109-residue Rpf-like fragment (RLF) from the Barnyard TMP in the mycobacteria. The secreted form of the protein was generated by utilizing complementary 99 base oligonucleotides containing compatible restriction sites for cloning to insert the signal sequence from the *M. tuberculosis* RpfA protein (Rv0867c) upstream of, and in-frame with, the RLF coding sequence. While the construct expressing the RLF transformed *M. smegmatis* with approximately the same efficiency as the parent-plasmid (pJL37) (Lewis & Hatfull, 2000, Bibb & Hatfull, 2002), transformation with the construct expressing the secreted RLF (sRLF) yielded no detectable transformants (Table 5). I hypothesized that this could have resulted from the fact that these vectors utilize the strong, constitutive *M. bovis* BCG *hsp60* promoter (P_{hsp60}), and high levels of the secreted RLF might be toxic to the cell. This would be in agreement with anecdotal evidence that high concentrations of the Rpf proteins can have growth-inhibitory effects. *M. smegmatis* containing pLAM6 appeared identical to vector controls, and further, no RLF expression from this strain was detected by western blot. However, as cell extracts were utilized in these experiments, it is possible that the protein was in the insoluble fraction.

Table 5. Constructs expressing the RLF in *M. smegmatis.*

Plasmid	*Description*	*Transformation Efficiency (cfu/μg)*	*Phenotype*
pJL37	Parent-P_{hsp60}	2.1×10^5	Wild-type
pLAM6	Constitutive RLF	1×10^5	Wild-type
pLAM7	Constitutive sRLF	0	Does not transform
pLAM12	Parent-$P_{acetamidase}$	1.1×10^6	Wild-type
pLAM13	Inducible RLF	1.2×10^6	Wild-type
pLAM14	Inducible sRLF	1.1×10^6	Wild-type

Sec reted form of the RLF utilizes the *M. tuberculosis* RpfA signal sequence.

Due to the problems encountered with the plasmids containing the constitutive promoter, a derivative of pJL37, designated pLAM12, was constructed in which the P_{hsp60} was replaced with the inducible *acetamidase* promoter ($P_{acetamidase}$) (Parish *et al*., 1997), and this was utilized to create plasmids in which the RLF and sRLF are expressed under inducible control. These transformed *M. smegmatis* with approximately the same efficiency as parent vector (Table 5), and RLF expression was verified by western blot with α–RLF antibodies. A number of growth analyses were then performed in the presence and absence of inducer; however, these strains did not behave measurably different than vector control strains on solid or in liquid media. Specifically, no differences in colony morphology were observed on standard 7H9 mycobacterial media – with or without inducer – and on Sauton's minimal media (Fig 23A). It is notable, however, that a strain expressing the *M. tuberculosis rpfE* gene with its native signal sequence from the same parent plasmid (pLAM12-RpfE, M. Piuri) also appeared no different than controls on this media (Fig. 23A). RLF expressing strains did not display an altered lag phase when inoculated at low density in minimal media or any media tested (Fig. 23B), and no differences were observed in growth when measured in standard 7H9 or induction media, with or without inducer (Fig. 23C).

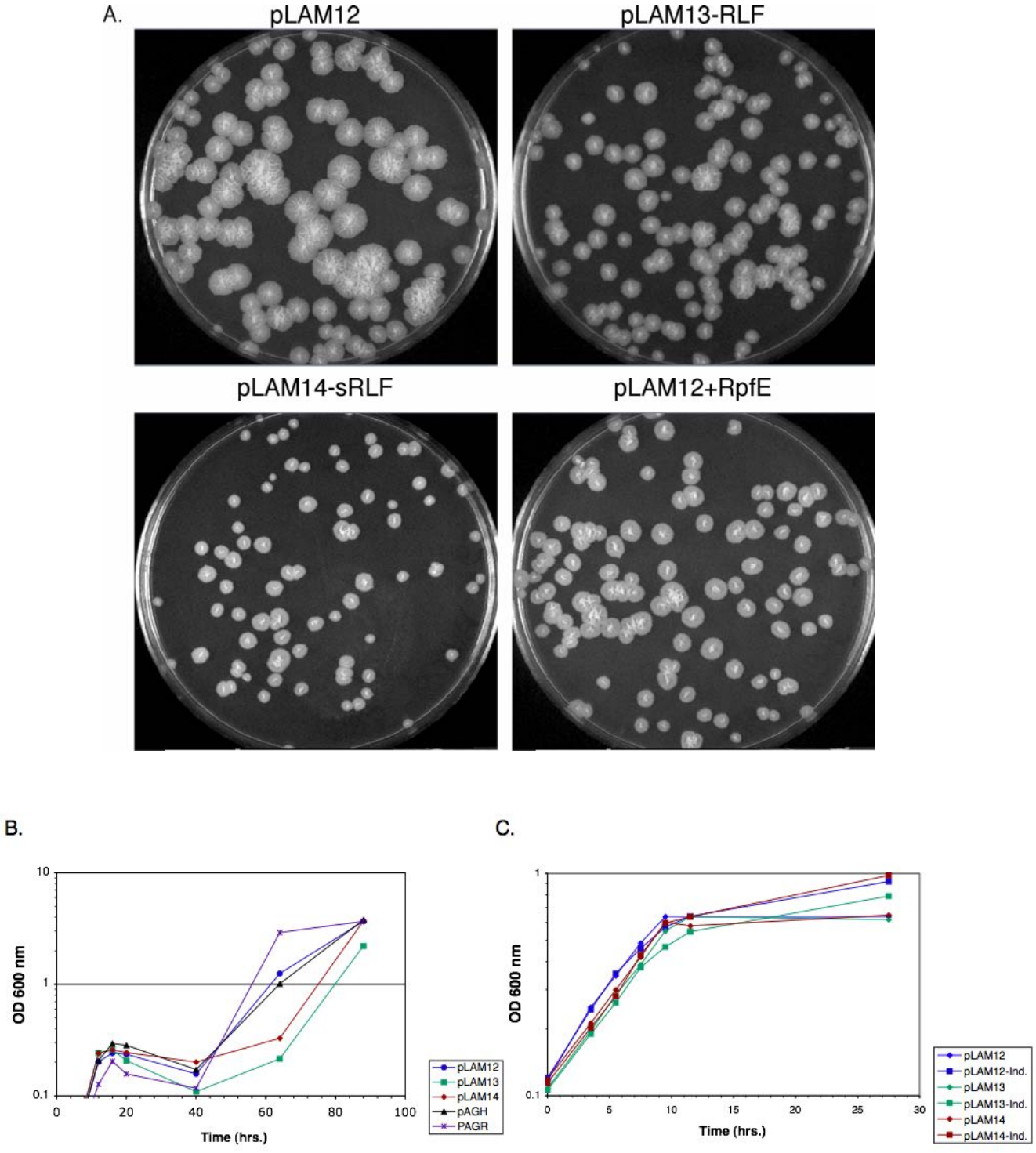


Figure 23. Growth analyses of RLF expressing strains of *M. smegmatis.*

Figure 23: Growth analyses of RLF expressing strains of *M. smegmatis*. A series of plasmids was constructed to express cytoplasmic and secreted forms of the Barnyard Rpf-like fragment (RLF and sRLF, respectively) under the control of the inducible *M. smegmatis acetamidase* promoter ($P_{acetamidase}$). The predicted signal peptide from *M. tuberculosis* RpfA (Rv0867c) was engineered at the N-terminus of the RLF to generate the secreted form of the protein. **A.** *M. smegmatis* strains expressing the RLF, the sRLF, or the *M. tuberculosis* RpfE protein (Rv2450c) and the vector control strain were serially diluted into Sauton's media and spread onto agar-solidified Sauton's media supplemented with 10% ADC; no differences in colony morphology were noted. **B.** Strains expressing the RLF or the sRLF do not display an altered lag phase, as measured by time to detectable turbidity, when inoculated to low density (3×10^2 cfu/ml) in 7H9 with 10% ADC. Strains containing pAGR and pAGH were a gift from M. Young, University of Aberystwyth, Wales; pAGR expresses the *M. luteus* Rpf protein from $P_{acetamidase}$ but also contains its native promoter sequence and pAGH is the parent plasmid (Shleeva *et al*., 2004). **C.** *M. smegmatis* strains inducibly expressing the RLF or sRLF and the empty vector control (pLAM12) were grown in 7H9 with 10% ADC overnight and subcultured in duplicate at an OD_{600} of 0.1 in 7H9 base induction media (contains 0.2% succinate, 0.05% Tween 80 and KAN). These were grown for 3.5 hours with agitation at 37°C, and inducer (0.2% acetamide) was added to one culture from each strain. Growth was continued, and samples were removed at various time-points for optical density determination. No differences in growth were noted in the presence or absence of inducer.

Vectors were also constructed to express the entire 70 kDa C-terminal fragment (SKF) containing the Rpf Motif and Motif 2 in *M. smegmatis* under both constitutive and inducible promoters. As in the previous case, all transformed with efficiencies similar to the parent vector, except for the construct pLAM29, which constitutively expresses a secreted SKF (sSKF) and was not tolerated (Table 6). Interestingly, cells expressing both the SKF and the sSKF under inducible control grow poorly on plates containing acetamide (Fig. 24A) and display altered growth patterns in liquid induction media (Fig. 24B), with the secreted form of the protein producing a slightly more severe phenotype in both cases. Specifically, although SKF expressing strains appear to grow slowly in liquid media, they actually show a drastically increased incidence of clumping, which could potentially be due to an altered cell surface. There was no change in viability, as no significant differences in colony counts were noted between *M. smegmatis* stains expressing the SKF and vector controls after growth in the presence of inducer (Fig. 24C). In separate experiments it was observed that the viability of *M. smegmatis* containing pLAM12-RpfE was also indistinguishable from controls upon growth in acetamide. No alterations in colony morphology were observed on Sauton's minimal media, and as with the RLF expressing strains, expression from pLAM28 and pLAM30 was verified by western blot with α-RLF antibodies performed on whole cell material (Fig. 24D).

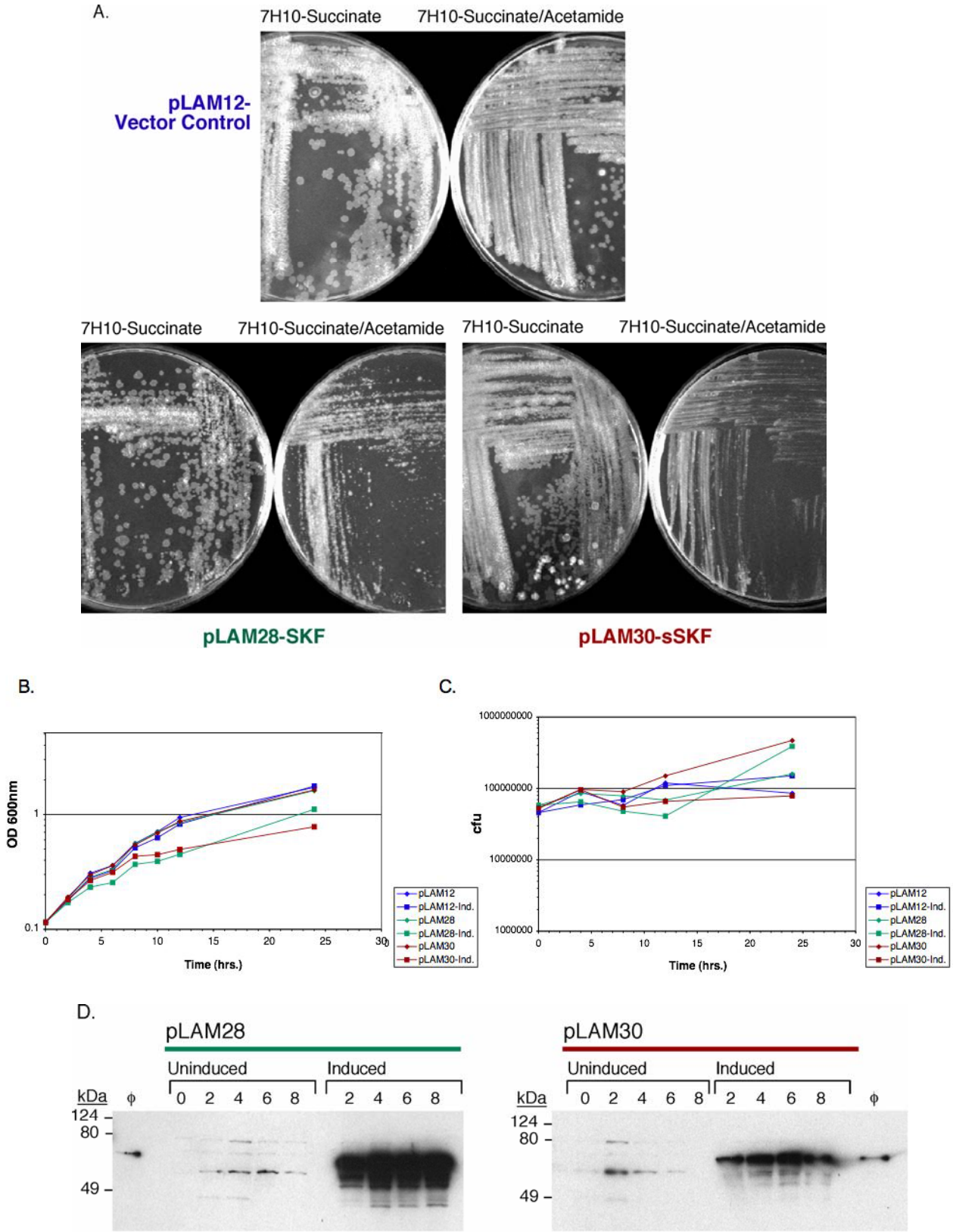


Figure 24. Growth analyses of SKF expressing strains of *M. smegmatis.*

Figure 24: Growth analyses of SKF expressing strains of *M. smegmatis*. A series of plasmids was constructed to express cytoplasmic and secreted forms of the Barnyard 70 kDa C-terminal fragment (SKF and sSKF, respectively) under the control of $P_{acetamidase}$. The predicted signal peptide from *M. tuberculosis* RpfA (Rv0867c) was engineered at the N-terminus of the SKF to generate a secreted form of the protein. **A.** *M. smegmatis* strains expressing either the SKF or the sSKF under inducible control grow slowly on 7H10 plates supplemented with 0.2% acetamide. **B.** These strains and the empty vector control (pLAM12) were grown in 7H9 with 10% ADC overnight and subcultured at an OD_{600} of 0.1 in 7H9 media base induction media (0.2% succinate, 0.05% Tween 80 and KAN) with or without inducer (0.2% acetamide). Cultures were grown with agitation at 37°C, and samples were removed at various time-points for optical density determination and **C.** viability determination. Under these conditions, strains expressing either the secreted or the non-secreted form of the protein appear to show slightly reduced growth rates in the presence of inducer. However, despite this apparent difference in growth rate, there does not appear to be a significant difference in viability between strains expressing the 70 kDa C-terminal portion of tape measure and the vector control. **D.** Strains were induced as described above, and aliquots were removed at the indicated-time points for immunoblot analysis with α–RLF antibodies (1:125).

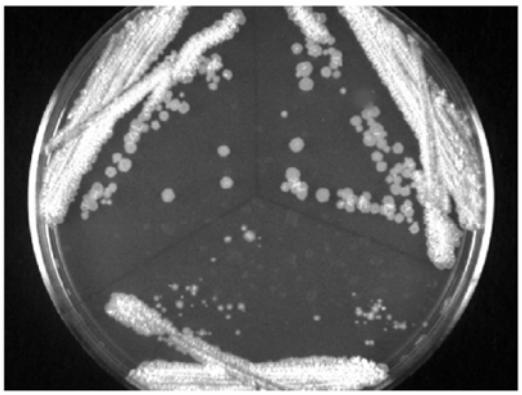
Table 6. Constructs expressing the SKF in *M. smegmatis.*

Plasmid	*Description*	*Transformation Efficiency (cfu/μg)*	*Phenotype*
pJL37	Parent-P_{hsp60}	7.6×10^4	Wild-type
pLAM27	Constitutive SKF	1.7×10^5	Wild-type
pLAM29	Constitutive sSKF	6.7×10^1	Does not transform
pLAM12	Parent-$P_{acetamidase}$	6.6×10^4	Wild-type
pLAM28	Inducible SKF	1.6×10^5	Slow growth on plates, clumps in presence of inducer
pLAM30	Inducible sSKF	1.4×10^5	Slow growth on plates, clumps in presence of inducer

Sec reted form of the SKF utilizes the *M. tuberculosis* RpfA signal sequence.
Des cribed as wild-type because no obvious phenotype was noted; however, SKF expression was not verified.

A potential problem with the previously described constructs is that in the original plasmids the cloning of the signal sequence resulted in a methionine residue immediately flanking the putative cleavage site (AMA^**M**A) that is not present in RpfA. Therefore, site-directed mutagenesis was utilized to remove this ATG codon in all plasmids containing the RpfA signal sequence at the N-terminus of the expressed protein so that the cleavage sites would match that of RpfA (AMA^**A**). Most 'corrected' plasmids transformed with efficiencies similar to the non-mutated parent and behaved similarly in liquid media and on plates containing inducer (Table 7; Fig. 25B and C). However, pLAM37 – the mutated derivative of pLAM7 that constitutively expresses a secreted form of the RLF – was able to transform *M. smegmatis* with an efficiency similar to parent (2.2×10^5 cfu/μg); whereas the non-mutated plasmid cannot (2.7×10^3 cfu/μg and the few colonies obtained likely represent mutants). Cells containing pLAM37 grow more slowly on solid media than the parent strain (Fig. 25A), although the reason for this is not known.

A.



pJL37-
Vector Control

pLAM6-RLF

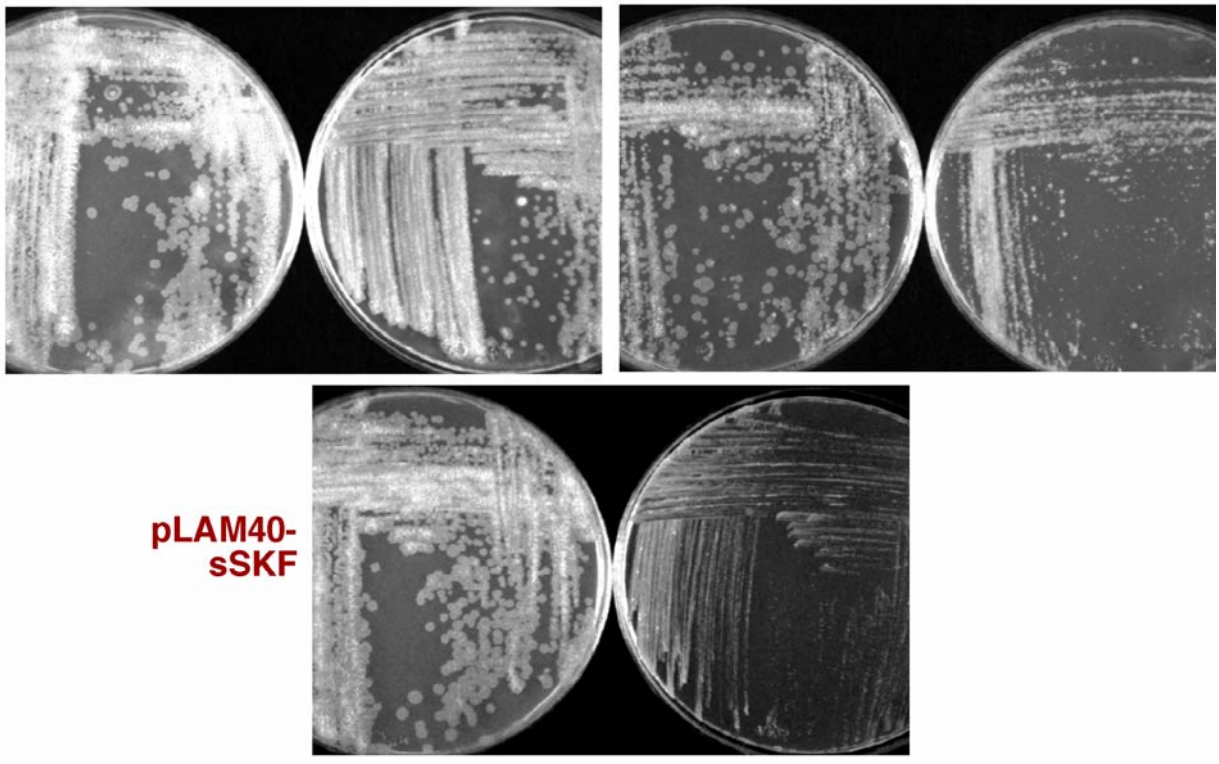
pLAM37-sRLF

B.

pLAM12-Vector Control

pLAM28-SKF

7H10-Succinate 7H10-Succinate/Acetamide 7H10-Succinate 7H10-Succinate/Acetamide



pLAM40-
sSKF

7H10-Succinate 7H10-Succinate/Acetamide

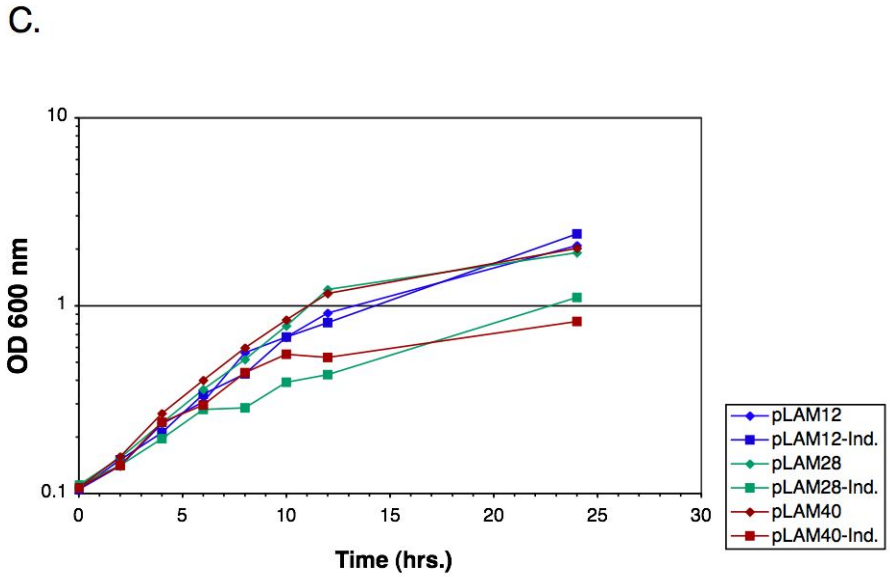


Figure 25. RLF and SKF expressing strains encoding the corrected RpfA signal sequence.

Figure 25: RLF and SKF expressing strains with the corrected RpfA signal sequence. Site-directed mutagenesis was utilized to remove a methionine codon in the signal sequence cleavage site of all plasmids expressing secreted forms of the RLF and SKF (sRLF and sSKF, respectively) so that it would be identical to the corresponding site in the *M. tuberculosis* RpfA (Rv0867c) protein. **A.** *M. smegmatis* constitutively expressing the sRLF grows slowly on solid 7H10 media with 10% ADC and has an altered colony morphology. **B.** A strain expressing the sSKF with the corrected signal sequence from $P_{acetamidase}$ grows slowly on 7H10 plates containing 0.2% acetamide similar to cells containing the original (uncorrected) plasmid (pLAM30; Fig. 24A). **C.** *M. smegmatis* strains inducibly expressing the SKF (pLAM28) or the sSKF (pLAM40) and the empty vector control (pLAM12) were grown in 7H9 with 10% ADC overnight and subcultured at an OD_{600} of 0.1 in base induction media (0.2% succinate, 0.05% Tween 80 and KAN) with or without inducer. These were grown with agitation at 37°C, and samples were removed at various time-points for optical density determination. *M. smegmatis* expressing the sSKF with the corrected signal sequence grew similarly to cells containing the original (uncorrected) plasmid (pLAM30; Fig. 24B).

Table 7. Constructs expressing secreted proteins in *M. smegmatis* with corrected signal sequence cleavage sites.

Plasmid	*Description*	*Transformation Efficiency (cfu/μg)*	*Phenotype*
pJL37	Parent-P_{hsp60}	5.3×10^5	Wild-type
pLAM7	Constitutive sRLF	2.7×10^3	Does not transform
pLAM37	Constitutive sRLF	2.2×10^5	Slow growth on plates
pLAM29	Constitutive sSKF	3.5×10^3	Does not transform
pLAM39	Constitutive sSKF	4.8×10^2	Does not transform
pLAM12	Parent-$P_{acetamidase}$	2.7×10^5	Wild-type
pLAM14	Inducible sRLF	6.1×10^5	Wild-type
pLAM38	Inducible sRLF	1.5×10^5	Wild-type
pLAM30	Inducible sSKF	4.4×10^5	Slow growth on plates, clumps in presence of inducer
pLAM40	Inducible sSKF	4.4×10^5	Slow growth on plates, clumps in presence of inducer

In dicates construct in which the methionine codon at the putative signal sequence cleavage site was removed by site-directed mutagenesis.

5.4 EXPRESSION OF TAPE MEASURE FRAGMENTS IN THE *E. COLI* PERIPLASM

Since the data obtained from the previous expression experiments were difficult to interpret, an alternative strategy to assay for function of the Rpf Motif *in vivo* was devised. It has been shown that secretion of the *M. luteus* Rpf into the periplasm causes lysis of *E. coli* cells (Mukamolova *et al*., 2006). Therefore, the same expression system was employed in this study – an arabinose inducible pBAD (Invitrogen) vector that expresses a protein of interest fused to an *E. coli* secretion signal – to express both the Barnyard RLF as well as the SKF as secreted proteins in *E. coli* (Fig. 26A). While strains expressing the RLF did not behave differently than controls, strains expressing the SKF grew poorly in the presence of arabinose (Fig. 26B). This slow growth did not appear to be actual lysis, and the cells were able to recover by 24 hours of growth. Thus, the reason for this growth inhibition in the SKF expressing strains is not clear. It has also been reported that the *M. luteus* Rpf domain alone did not express well or demonstrate activity in this assay either, making the negative results obtained with the RLF difficult to interpret.

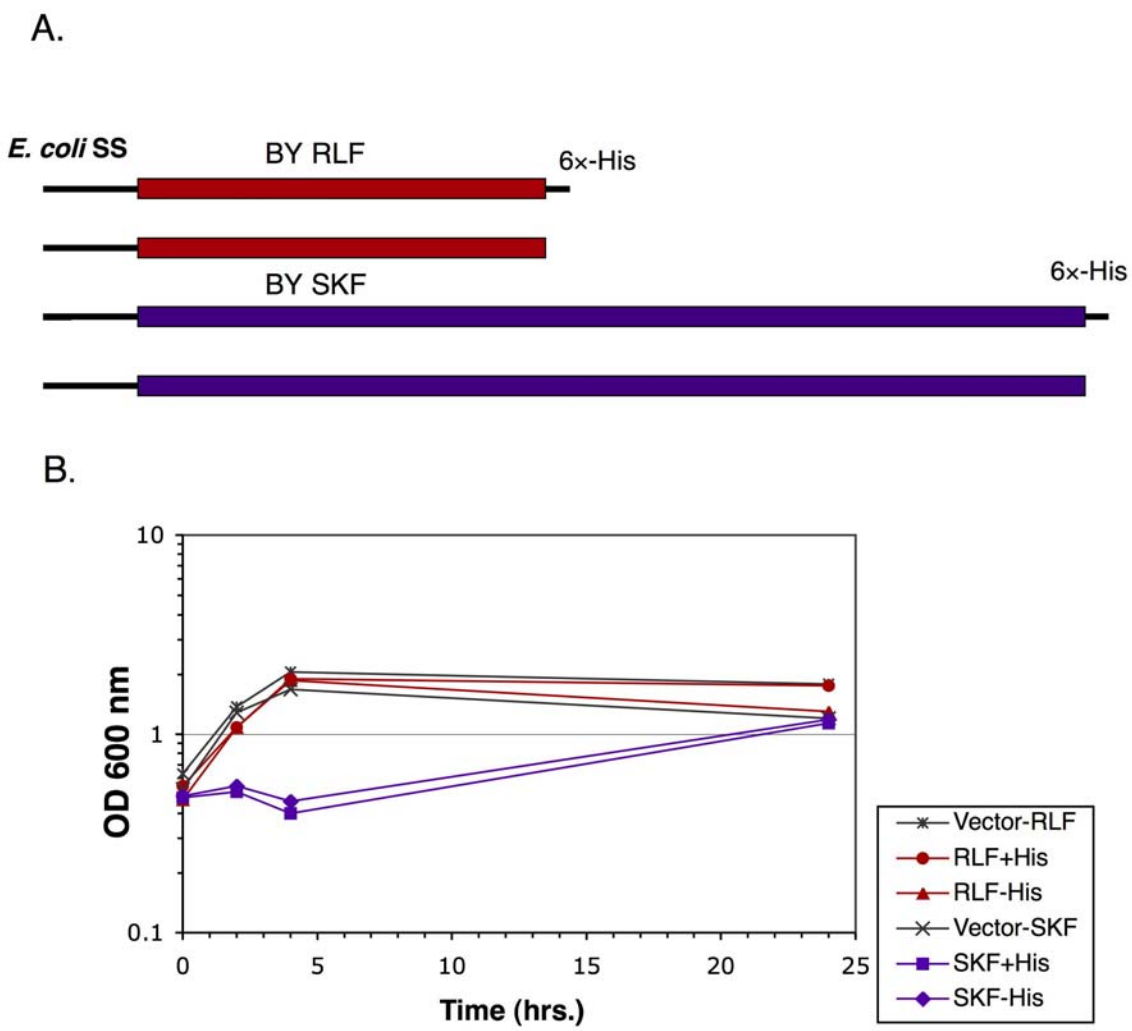


Figure 26. Expression of the RLF and the SKF as secreted proteins in *E. coli.*

Figure 26: Expression of the RLF and the SKF as secreted proteins in *E. coli*. **A.** Plasmids were constructed to express the Barnyard RLF and the SKF (with or without a C-terminal 6×-His tag) into the periplasm of *E. coli* using the pBAD/gIII vector system (Invitrogen). **B.** Expression was induced from Top 10 cells with 0.02% arabinose, and cultures were sampled at various time-points for optical density determination. Growth of cells expressing either the tagged or untagged RLF was indistinguishable from vector controls. Cells expressing either form of the SKF grew poorly; however, lysis was not observed, and cultures were fully recovered by 24 hours of growth.

5.4.1 Expression of RLF-Rpf Fusion Proteins

The lytic activity of the *M. luteus* Rpf protein in *E. coli* was observed to be dependent on the presence of the C-terminal peptidoglycan-binding (LysM) domain in the full-length protein. Therefore, a pBAD derivative plasmid (pLAM46) was constructed that expresses a fusion of the Barnyard RLF to LysM domain of the *M. luteus* Rpf protein (Fig. 27A). The phage Rpf domain is missing two critical cysteine residues located near the active site of the *M. tuberculosis* RpfB motif (Fig. 27C), which were previously found to be important for the activity of the *Micrococcus* protein. Thus, site-directed mutagenesis was utilized to create plasmid derivatives of pLAM46 that express fusion proteins containing either or both of these cysteine residues in the respective locations in which they are found in the bacterial Rpfs, illustrated in Figure 27A. When induced with arabinose, a fusion protein with both cysteines (pLAM49) or with only the first one (pLAM47; C53 in *M. luteus* Rpf) killed *E. coli* similar to Rpf controls; however, those constructs with either C114 alone (pLAM48) or neither cysteine behaved similarly to the vector control (Fig. 27B). This initially suggested that the cysteine flanking the active site glutamate was indeed important for enzymatic activity, at least when the Rpf domain is contained within a soluble secreted protein. However, the phage Rpf Motifs are not soluble, but rather, are imbedded within the larger context of the TMP, a structural component of the phage particle. Thus, these data might also suggest that the function of the cysteines is compensated for by other residues within the tape measure or perhaps by its specific structural conformation within the phage tail.

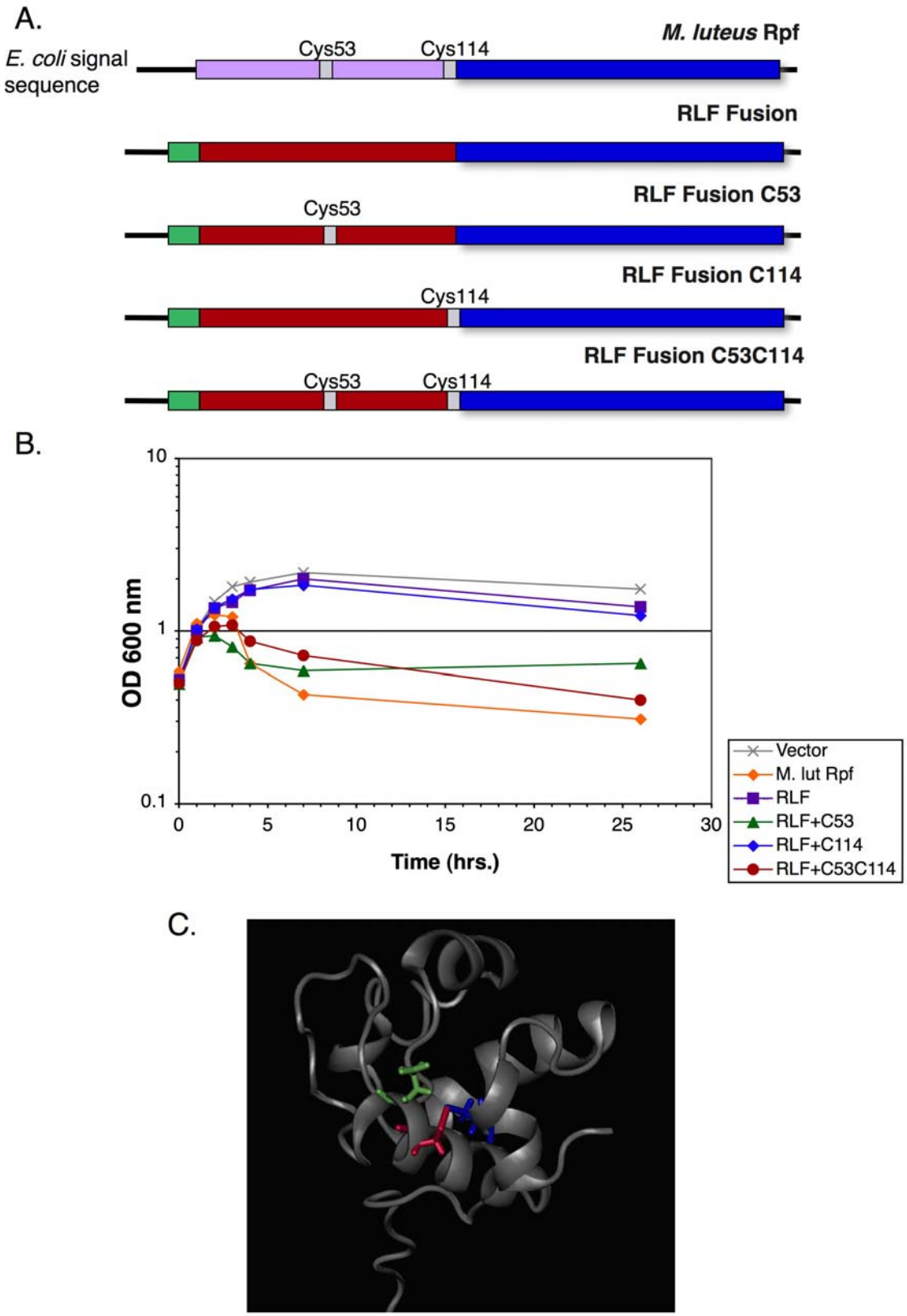


Figure 27. Expression of the RLF as a fusion with the *M. luteus* Rpf LysM domain in *E. coli.*

Figure 27: Expression of the RLF as a fusion with the *M. luteus* Rpf LysM domain in *E. coli.* **A.** The Barnyard Rpf Motif (RLF) was fused to the peptidoglycan binding domain (LysM) of the *M. luteus* Rpf and expressed as a secreted protein in *E. coli* using the pBAD/gIII vector system (Invitrogen). The two cysteine residues that are absolutely conserved in the bacterial Rpfs but missing in the phage motifs were engineered back into the Barnyard fusion protein, either singly or in combination, using site-directed mutagenesis (Stratagene). **B.** Expression was induced from Top 10 *E. coli* cells (Invitrogen) with 0.02% arabinose, and growth was monitored by optical density determination. Growth of strains expressing fusions containing the wild-type Barnyard Rpf Motif or this motif with C114 alone was comparable to the empty vector control, and strains expressing the fusion with C53 alone or both cysteine mutations were lysed to a similar degree as the wild-type Rpf control cells. **C.** Conserved cysteine residues (C53 and C114 in the *M. luteus* Rpf) were highlighted in the solution structure of the Rpf domain (residues 280-362) from the *M. tuberculosis* RpfB (Rv1009) (Image created using VMD 1.8.5 molecular graphics program). These are predicted to form a disulfide bond near the active site and may contribute to the stability of these proteins.

5.4.2 Corrected RLF-Rpf Fusion Proteins

This first set of fusions was not ideal, however, because the phage Rpf Motif encoded by these constructs contained 14 extra residues at the N-terminus of the protein that are not conserved in Rpf family members (Chapter 2, Fig. 7). These are located upstream of the Rpf Motif in the Barnyard TMP and were included in the initial cloning of this region. Therefore, another set of fusions was made that are identical to the first except that these residues were removed, and the N-termini of these proteins – directly downstream of the signal sequence cleavage site and upstream of the active site – now matched that of the Rpf protein in size and sequence (Fig. 28A and Chapter 2, Fig. 7).

Strikingly, when induced in *E. coli*, these proteins were now all lethal. Fusions with either the wild-type RLF (pLAM51) or the second (C114) cysteine residue (pLAM53) were now slightly more toxic than *M. luteus* Rpf (Fig. 28B). The fusion containing the first cysteine (C53 in pLAM52) displayed an even greater degree of killing, and interestingly, a fusion with both cysteines (pLAM50) was significantly more lethal than any of the other RLF constructs and the wild-type Rpf control, both in the timing of and degree of *E. coli* lysis (Fig. 28B).

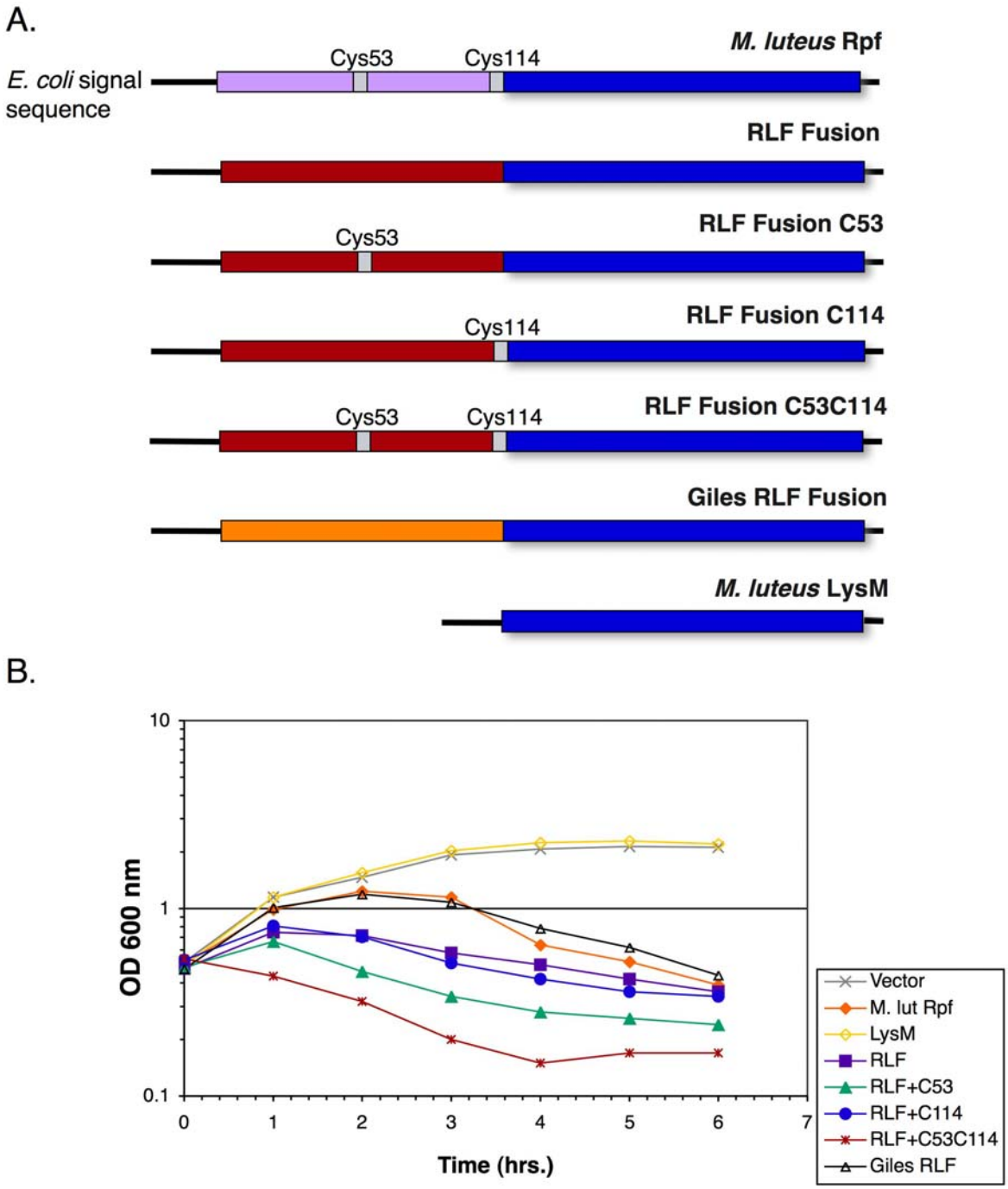


Figure 28. Corrected RLF-Rpf fusion proteins.

Figure 28: Corrected RLF-Rpf LysM fusion proteins and expression in *E. coli*. **A.** Plasmids were constructed in which 14 amino acid residues at the N-terminus (directly downstream of the signal sequence) of the RLF-Rpf LysM fusion proteins were removed and replaced by the four corresponding residues in the *M. luteus* Rpf. A similar construct containing the Giles Rpf Motif, as well as a control expressing only the *M. luteus* LysM domain, were also made. **B.** Expression was induced from Top10 *E. coli* cells (Invitrogen) with 0.02% arabinose, and growth was monitored by optical density determination. Fusions containing either phage Rpf Motif or the Barnyard motif with the C114 mutation lysed *E. coli* to a similar degree as the Rpf control. The fusion containing C53 was slightly more lethal than these, and the one with both cysteine mutations was significantly more lethal than any other protein tested, including the wild-type Rpf control. As expected, growth of cells expressing the LysM domain was comparable to vector controls in all experiments.

Since the TMP of mycobacteriophage Giles also contains an Rpf Motif, it was important to determine if this domain is also catalytically active. In order to assay this, a plasmid construct (pLAM75) was created that expresses the Giles Rpf Motif fused to the *M. luteus* Rpf LysM domain (Fig. 28A). This was also tested and shown to be active in lysing *E. coli*, although perhaps to a slightly lesser extent than the Barnyard Rpf Motif (Fig. 28B). Importantly, as has been previously reported, controls expressing only the LysM domain were not lethal and showed no lytic activity in this assay (Fig. 28B), confirming that the activity observed was due to the presence of the Rpf Motif in these fusions. These data therefore suggest that the bacteriophage Rpf Motifs have a catalytic potential that is at least equal to, and possibly greater than the *M. luteus* Rpf.

5.4.3 RLF-Rpf LysM Fusion Proteins with E54A Mutation

It was previously reported that a mutation in the putative active site residue (E54) of the *M. luteus* Rpf results in diminished enzymatic activity (Mukamolova *et al*., 2006). Therefore, site-directed mutagenesis was utilized to replace the corresponding glutamate residue with an alanine in the RLF-Rpf LysM fusion proteins containing either the wild-type Barnyard RLF or the RLF with both cysteine mutations (C53 and C114). Strikingly, when these constructs were introduced into *E. coli* and induced with arabinose, the fusion with E54A alone lysed *E. coli* to a similar degree as the wild-type RLF-Rpf LysM fusion protein. However, the protein with E54A and both cysteine mutations was markedly less toxic than the RLF-Rpf LysM fusion containing the cysteine mutations alone, and its activity was reduced to the level observed for the *M. luteus* Rpf and the wild-type RLF-Rpf LysM protein (Fig. 29).

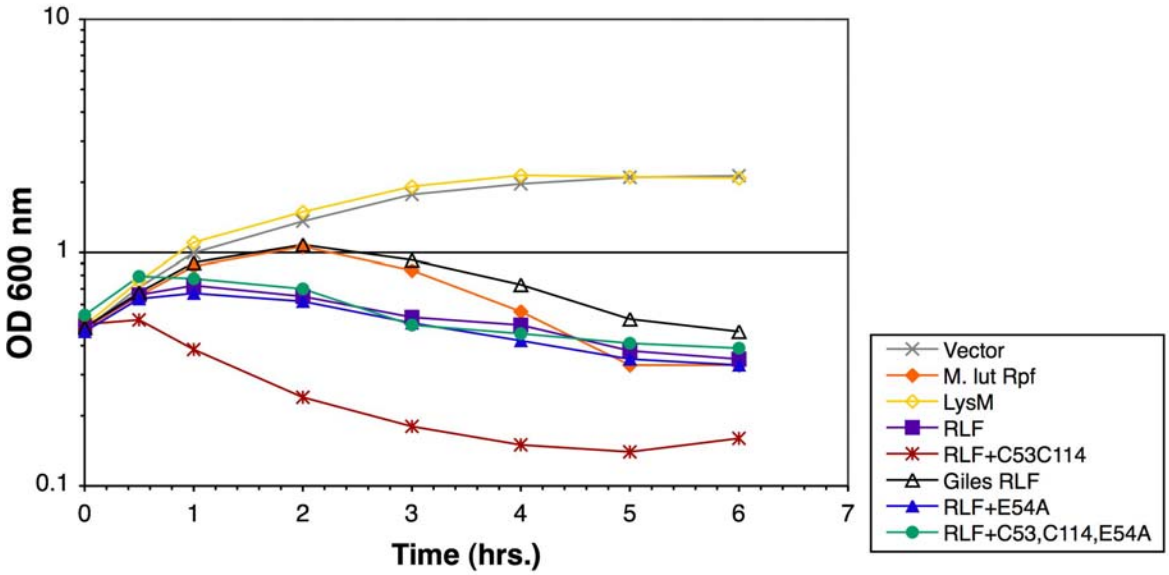


Figure 29. Expression of Fusions Containing E54A in *E. coli*.

Figure 29: Expression of RLF-Rpf LysM fusion proteins with E54A in *E. coli*. Plasmids were constructed in which the putative active-site residue (E54) in the wild-type RLF-Rpf LysM fusion, or in the fusion containing both cysteine mutations (C53 and C114), was replaced with an alanine. Expression was induced from Top10 *E. coli* cells (Invitrogen) with 0.02% arabinose, and growth was monitored by optical density determination. The fusion with E54A alone lysed *E. coli* to a similar degree as the wild-type RLF-Rpf LysM fusion protein. The fusion with E54A and both cysteines was less toxic that than the protein containing only the cysteine mutations, and its activity was reduced to approximately the level observed for the *M. luteus* Rpf and the wild-type RLF-Rpf LysM protein.

5.5 EXPRESSION OF RLF-LYSM FUSION PROTEINS IN *M. SMEGMATIS*

The *M. luteus* Rpf protein is not known to be toxic or to cause cell lysis when expressed in *M. smegmatis* and in fact, was reported to have growth stimulatory activities under some conditions (Mukamolova *et al*., 2002b). However, the same researchers also observed growth inhibition when this protein was overexpressed or when strains that express this protein were grown on rich media (M. Young and G. Mukamolova, personal communication). Therefore, although the specific activity of this protein remains unclear, the toxicity of the RLF-Rpf LysM fusion constructs in *E. coli* – particularly those containing the cysteine residues – led to the speculation that these might have similar phenotypes in *M. smegmatis*.

In order to test this possibility, I designed and created constructs to express the fusion proteins, as well as the *M. luteus* Rpf and the LysM domain alone, in mycobacteria utilizing both P_{hsp60} and $P_{acetamidase}$. Since the *E. coli* signal sequence from the pBAD constructs was unlikely to be functional in the mycobacteria, the RpfA signal sequence utilized in the *M. smegmatis* constructs described previously in Section 5.3 was inserted at the N-terminus of all proteins except the *M. luteus* Rpf control, which was cloned with its native signal sequence. Importantly, these were cloned such that the signal sequence cleavage sites matched that of RpfA. Notably, many of these plasmids were not well tolerated during cloning steps in *E. coli*, and multiple clones had to be sequenced before ones lacking mutations were found. Further, most of the constructs in which the fusions were constitutively expressed were unable to transform *M. smegmatis*. Even controls expressing only the *M. luteus* LysM domain were unable to transform. This might suggest that high levels of expression of full-length Rpf-like proteins are lethal in *M. smegmatis*, although the

toxicity of even the LysM controls suggests that it is merely the interaction with the cell wall is that is detrimental. This is supported by the observation that a previously constructed plasmid that expresses only the Rpf domain fused to the 'corrected' RpfA signal sequence is tolerated in *M. smegmatis*, although cells containing this plasmid display a slow growth phenotype (Fig. 25A). Further, the *M. tuberculosis* Rpf proteins have been cloned and expressed under control of P_{hsp60} in *M. smegmatis* with no detrimental effects (J. Tufariello, unpublished data).

In contrast to plasmids with the constitutive promoter, all constructs in which the fusion proteins were expressed under control of the inducible $P_{acetamidase}$ transformed *M. smegmatis* with efficiencies similar to the vector control. Although an extensive phenotypic analysis of these strains under a complete array of growth conditions has not been conducted, it has been observed that all strains display normal growth on standard mycobacterial media or on the minimal induction media lacking inducer. Interestingly, *M. smegmatis* cells expressing RLF-LysM fusion proteins containing either C53 or both cysteines were unable to grow on media containing the inducer (Fig. 30). Importantly, under these conditions, strains expressing the *M. luteus* Rpf, the wild-type RLF-Rpf LysM fusion, or the RLF-Rpf LysM fusion with C114, as well as the LysM control strain were viable and grew similar to vector controls. This suggests that addition of C53 to the RLF-Rpf LysM fusion increases the activity of the protein such that even the low levels of expression mediated by $P_{acetamidase}$ are now toxic. This is in contrast to what is observed for the wild-type Rpf protein, which is not toxic under the same conditions but contains C53. Unlike what is observed on solid media, however, when induced in liquid media, all strains displayed similar growth patterns, and no significant alterations in viability were noted.

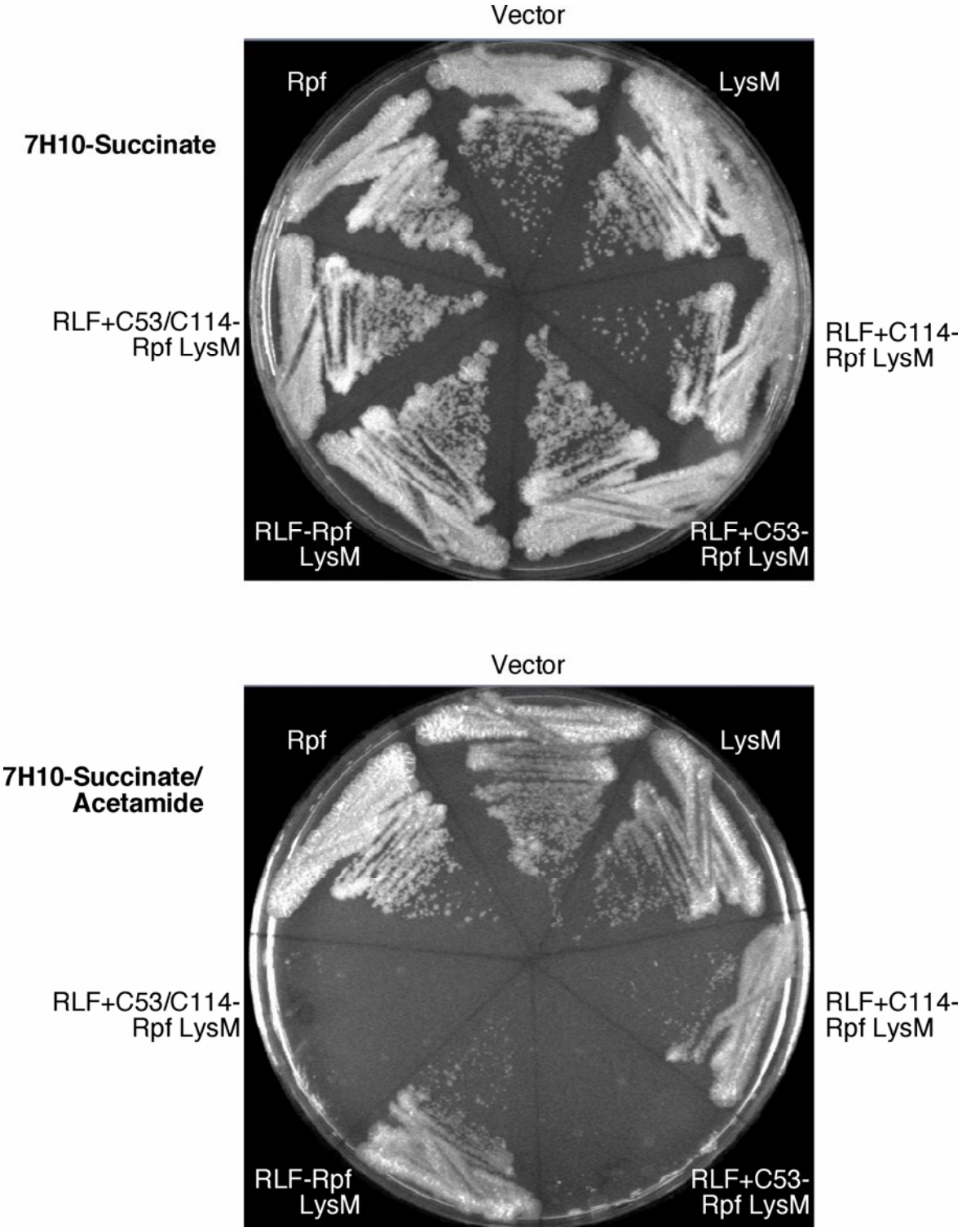


Figure 30. Expression of RLF-Rpf fusion proteins under inducible control in *M. smegmatis*.

Figure 30: Expression of RLF-Rpf LysM fusion proteins under inducible control in *M. smegmatis.* *M. smegmatis* strains expressing secreted forms of the various RLF-Rpf LysM fusion proteins, the *M. luteus* Rpf or the Rpf LysM domain alone under control of $P_{acetamidase}$, as well as the empty vector control, were grown on 7H10 media containing 0.2% succinate, with or without inducer (0.2% acetamide). Strains expressing fusions containing the C53 mutation were unable to grown in the presence of inducer.

5.6 CONCLUSIONS

There are a number of factors that have contributed to the difficulty of assessing the function of the bacteriophage Rpf Motifs *in vitro*. A His-tagged, recombinant RLF was partially purified from *E. coli*; however, I was unable to demonstrate activity for this protein in zymograms and in other assays (EnzChek Lysozyme Assay Kit, Molecular Probes). This is perhaps not surprising as the full-length *M. luteus* Rpf was reported to have only weak activity in zymogram assays, and the Rpf domain from the *M. tuberculosis* RpfB was also inactive in the EnzChek assay (Cohen-Gonsaud *et al.*, 2005, Mukamolova *et al.*, 2006), although the reason for this is not clear. The *M. luteus* Rpf protein is reportedly only active *in vitro* when isolated from *M. luteus* supernatant or when purified from the *E. coli* periplasm (Mukamolova *et al*., 2006). This suggests that disulfide bond formation is important for activity, which could be problematic as the phage motifs lack the two cysteine residues thought to form a disulfide bridge near the active site. However, it was also reported that addition of DTT or mercaptoethanol did not abolish Rpf activity in zymograms. Although these data are difficult to interpret, it has been well established that recombinant Rpf is quite unstable, so the absence of activity *in vitro* assays is not necessarily indicative of a lack of activity.

Another possible reason for lack of activity of recombinant RLF in zymograms is that the phage motif itself may be not be functional unless it is in the context of the complete tape measure protein, or possibly the 70 kDa C-terminal cleavage fragment. Although the phage Rpfs are typically quite similar to the bacterial proteins (55-70%), this does not rule out the possibility that they may differ fundamentally in how they work. That is, the bacterial Rpf family members must function in solution or perhaps tethered to the cell membrane, but presumably the phage motifs

must exert their activity while constrained within the context of the TMP and the bacteriophage tail. In this case, it would be difficult to envision an adequate *in vitro* assay for enzymatic activity. Zymogram analyses involve separating proteins of interest on denaturing SDS-polyacrylamide gels, and the detection of activity then depends on the proteins renaturing in buffer and having the ability to digest a visible amount of the cell wall material embedded within the gel. It is therefore possible that the 70 kDa protein containing the Rpf Motif is too large to effectively renature in the gel, or perhaps was unable to renature under the range of conditions tested. Alternatively, it may not be able to renature to an active conformation once it is separated from the phage particle. Accordingly, zymoblot analyses, which do not involve denaturation, detect a possible hydrolytic activity associated with the Barnyard particle, although the origin of this is unknown. Furthermore, based on what has been shown for other phages, there are likely six copies of the tape measure per tail (Casjens & Hendrix, 1974, Katsura, 1983, Abuladze *et al.*, 1994); and thus, it is also possible that there was not a sufficient amount of this protein to make a visible clearing within the zymogram gel. This might be particularly true if the phage motifs had limited or low levels of enzymatic activity. It has been observed that this is the case for the Rpf proteins, as well as for a number of other bacteriophage structure proteins with muralytic activity, such as T7 gp16, and it has been postulated that it would be disadvantageous for phages to carry cell wall hydrolases that are highly active or unconstrained in some way, as this might kill the host before the phage had the chance to establish productive infection (Moak & Molineux, 2004).

If this were true and the 70 kDa C-terminal fragment (SKF) of Barnyard TMP has only weak or constrained hydrolytic activity, one might expect that large amounts of this protein might be required to detect activity in zymograms. This is a testable hypothesis as I have successfully expressed (from the plasmid pLAM35) and partially purified the entire SKF as an N-terminal

fusion to the 6×-His tagged maltose binding protein (MBP) in *E. coli*. This is known to stabilize difficult proteins and to increase their solubility and was utilized because attempts to express the protein with either an N-terminal (pLAM34) or C-terminal (pLAM26) 6×-His tag or to express it simultaneously with the two upstream proteins (gp31 and gp32) predicted to undergo a translational frame-shift (pLAM33) – analogous to the proteins required for λ TMP expression in *E. coli* (Xu, 2001, Ph.D. Thesis) – were unsuccessful. Interestingly, a large portion of the MBP tag appears to be cleaved from the partially purified SKF, and much of this is further proteolyzed into what appears to be the ~65 kDa degradation product observed in initial phage protein preps (Chapter 4). There is a predicted acid cleavage site (Asp-Pro) between residues 1444 and 1445 of Barnyard TMP (residues 60 and 61 of the SKF), and cleavage of the SKF at this site would yield a product of 63 kDa, which closely corresponds to what is observed. Further, there are a large number of contaminating proteins and/or smaller degradation products. Therefore, these purification protocols will need to be optimized in order to improve upon these results and obtain protein for use *in vitro*; although, since individual proteins are separated during zymogram analysis, it may not be necessary to utilize highly purified protein for this assay.

It has also proved challenging to measure the activity of the phage Rpf Motifs *in vivo*. Although it has been reported that expression of the *M. luteus* Rpf can modulate the growth of *M. smegmatis,* expression of either the RLF or the SKF in this organism did not allow a clear elucidation of function for the Rpf Motif. The various Rpf proteins are predicted to be either secreted or cell wall-associated, and accordingly, at least some of these have been detected in the supernatant or associated with the cell surface (Mukamolova *et al*., 2002b, Mukamolova *et al*., 2002a). Therefore, it was necessary to generate secreted versions of the phage proteins, and this was accomplished by fusing the signal sequence from the *M. tuberculosis* RpfA (Rv0867c) protein

to the N-terminus of the various constructs. Initial plasmids created in this way encoded a Met residue directly downstream of the signal sequence cleavage site, and although there is no direct evidence that this would inhibit cleavage, constructs were also created in which this codon was removed, resulting in a cleavage site identical to that of RpfA.

When the various RLF expressing strains were assayed under a variety of conditions, even those inducibly expressing the secreted RLF with the 'corrected' signal sequence were phenotypically wild-type in all cases. This might be indicative of a lack of activity, although similar constructs inducibly expressing the *M. luteus* Rpf or the *M. tuberculosis* RpfE protein with their native signal sequences were also indistinguishable from vector controls in some of the same assays, making these data exceedingly difficult to interpret. Interestingly, initial constructs constitutively expressing the sRLF did not transform *M. smegmatis*; however, the plasmid derivative with the corrected signal sequence produced viable transformants. This indirectly suggests that the altered cleavage site may have inhibited cleavage to some degree, although this hypothesis would require further testing. *M. smegmatis* constitutively expressing the sRLF did display slow growth on solid and in liquid media; however, the reason for this is also not known. Interpretation of these results is further complicated by the observation that plasmids encoding either the *M. luteus* Rpf or the Rpf LysM domain under control of P_{hsp60} are unable to transform *M. smegmatis*, although strains constitutively expressing the individual *M. tuberculosis* Rpf proteins under the same promoter in a similar parent vector (J. Tufariello, unpublished data) appear normal.

In contrast to what was observed for the RLF, constructs encoding a constitutively expressed SKF with an N-terminal secretion signal – even those containing the corrected signal sequence cleavage site – were unable to transform *M. smegmatis*. Strains containing plasmid constructs from which expression of the SKF is controlled by the inducible *acetamidase* promoter,

however, can be constructed and grow normally under non-inducing conditions or on Sauton's minimal media. Interestingly, these display altered growth in media containing the acetamide inducer. This phenotype manifests as slow growth on plates and a seemingly reduced growth rate in liquid induction media. However, this actually appears to represent an increased incidence of clumping, as overall viability of induced strains is not severely affected. Intriguingly, this clumping phenomenon has also been observed for *M. smegmatis* cells in stationary phase that are treated with lysozyme (M. Pavelka, personal communication). Therefore, this might be indicative of alterations in the bacterial cell wall, although this cannot be definitely asserted. Interestingly, when expression of the SKF lacking the RpfA signal sequence is induced in *M. smegmatis*, these cells also behave similarly to strains expressing the secreted form of the protein. This may be due to the fact that the SKF is predicted to contain two to three transmembrane domains (Chapter 4), and it is therefore possible that even the version of the protein without the signal sequence is being targeted to and disrupting the cell membrane. As these data proved inconclusive, it was necessary to assay the *in vivo* activity of the Rpf Motifs in another way.

The *M. luteus* Rpf protein has the ability to lyse *E. coli* when expressed into the periplasm of these bacteria. Strikingly, I have shown that replacing the Rpf domain of this protein with either the Barnyard or the Giles Rpf Motif results in fusion proteins having the same activity. This is therefore the first demonstration that the phage Rpf Motifs have murein hydrolase activity *in vivo*. Both of these motifs lack two critical conserved cysteine residues that are found in all other Rpf-like proteins and which are reportedly required for activity in the *M. luteus* Rpf. These were engineered back into the Barnyard RLF-Rpf LysM fusions, and in the initial plasmids constructed, at least the first of these (C53) was indeed necessary for activity. *E. coli* expressing fusions lacking this cysteine residue behaved similarly to vector control cells and were not lysed (Fig. 27B).

However, a second set of fusions was created from which 14 N-terminal amino acid residues were removed and replaced by the four corresponding residues found at the N-terminus (directly downstream of the signal sequence) of the *M. luteus* Rpf. When expressed in *E. coli*, all of these proteins were now able to mediate cell lysis, regardless of whether or not they contained the cysteine residues (Fig. 28B). Fusions with either the wild-type RLF or the second (C114) cysteine residue were slightly more lethal than *M. luteus* Rpf, and the fusion containing C53 displayed an even greater degree of killing. Most striking, the fusion with both cysteines was significantly more lethal than any of the others and the wild-type Rpf control, both in the timing of and degree of *E. coli* lysis. The Rpf Motif from Giles tape measure showed activity similar to wild-type Rpf, and importantly, control strains expressing only the LysM cell wall binding domain were phenotypically normal.

At the moment, it is not understood why addition of the cysteine residues results in proteins with seemingly increased levels of enzymatic activity. However, cysteine residues are quite rare in phage TMPs, and even though these molecules are quite large, those of the siphoviridae were never observed to contain more than one cysteine (discussed in Chapter 3), which is a much lower frequency than would be expected in proteins of these size. Further, despite the observation that these residues appear to be important for the activity of the *M. luteus* Rpf, their specific function is unknown, although their proximity to one another suggests that they can form a disulfide bride that enhances protein stability (Fig. 27C). Conversely, the observation of increased activity when only the first (C53) is present suggests that this residue might directly stimulate hydrolysis, and its proximity to the putative active site residue (E54) supports this. Regardless of their specific function, it seems likely that phage tape measure Rpf Motifs have evolved to compensate for this to some degree. Thus, regions of this motif that differ from other Rpf family members may confer

increased stability and increased levels of catalytic activity in the absence of the cysteine residues, and further stabilization or stimulation of the phage domain by the "re-addition" of these residues results in proteins with enhanced levels of activity.

The observed phenotypes of the RLF-Rpf LysM fusion proteins containing the putative active-site mutation E54A are also puzzling. Introduction of this mutation clearly decreases the lethality of the fusion containing the cysteine mutations; however, introduction of E54A seems to have no effect on the activity of the wild-type RLF-Rpf LysM fusion. When this same mutation is engineered into the *M. luteu*s Rpf protein, which contains the conserved cysteine residues, it has been reported to result in decreased lysis of *E. coli* in the same assay (Mukamolova *et al*., 2006). This suggests an interaction between at least one of these cysteines – likely the adjacent C53 – and the active-site glutamate; an interpretation that is consistent with the data presented in this study. In order to address this possibility, it may prove useful to investigate the phenotypes of fusions containing the individual cysteine residues combined with the E54A mutation. Importantly, both fusions containing E54A retain some level of activity, which further suggests that there are other residues within the Barnyard RLF that are important for enzymatic activity. Expression in *E. coli* provides a rapid and reliable readout for enzymatic activity; therefore, random or directed mutagenesis of these fusions could prove useful in uncovering the identity of the 'compensating' residues in the phage Rpf Motifs. These and other fusion proteins can also be purified from *E. coli,* which will to allow the activity levels of various mutants to be monitored and assessed *in vitro.*

The complexity of the situation also somewhat mirrors what has been observed with murein hydrolase enzymes found in other DNA transfer systems. In one example, mutations in the active site glutamate of the VirB1 protein from the *A. tumefaciens* the type IV secretion system were found to impair but not abolish biological activity, although enzymatic activity was not

assayed (Hoppner *et al.*, 2004). Conversely, a truncation of the P19 protein from the plasmid R1 transfer (*tra*) operon, which displays a decreased conjugation efficiency and decreased susceptibility to 'male-specific' bacteriophages, cannot be complemented by versions of this protein with mutations in the active-site residue (Bayer *et al*., 1995, Bayer *et al*., 2001). The mutant versions of P19 also lose the ability to mediate lysis of *E. coli* cells in which they are overexpressed (Bayer *et al.*, 2001). It has also been observed that mutations in the active site residue of the T7 gp16 hydrolase that abolish enzymatic activity do not completely abolish function (infectivity); however, this can be explained by the multifunctional nature of this phage structural protein (Moak & Molineux, 2000, Moak & Molineux, 2004). It seems, therefore, that the situation is far from clear, and different systems that utilize muralytic proteins may display fundamental differences in the requirement for this activity.

It is also currently not clear whether the RLF-Rpf LysM fusion proteins have similar activity in the mycobacteria. None are tolerated when constitutively expressed in *M. smegmatis*; however, a secreted version of only the LysM control protein is also not tolerated. This suggests that overexpression of proteins with the ability to interact with the cell wall, perhaps not surprisingly, is detrimental for growth. This phenomenon also appears to be specific to proteins containing the *M. luteus* LysM domain, as strains expressing each of the five *M. tuberculosis* Rpf proteins from the same promoter in a similar parent vector are phenotypically normal. *M. smegmatis* expressing any of these proteins under the inducible $P_{acetamidase}$ appear to grow normally under both non-inducing and inducing conditions in liquid media. However, strains expressing fusions with C53, either alone or in combination with C114, are unable to grow on solid media containing acetamide. As these were the proteins that showed greater levels of lethality in *E. coli*, this strongly indicates that this phenotype is due to the muralytic activity of these proteins. It will

be interesting to assess the growth of these strains under a broad array of growth conditions, specifically those under which the bacterial cell wall is subjected to increased stress.

These analyses have shown that the Rpf Motifs found in bacteriophage tape measures are indeed catalytically active. Although it has proved difficult to assess their function while in the context of the complete tape measure and in the phage particle, the phage Rpf Motifs can effectively substitute for the bacterial Rpf Motif *in vivo*. Critical residues required for activity of the bacterial proteins are dispensable for activity of the phage derivatives, suggesting they have evolved features that compensate for their activity. In a similar manner, the mycobacteriophage TM4 tape measure Motif 3 was shown to be able to substitute for the homologous domain in host proteins and to have murein hydrolase activity *in vitro* (Piuri & Hatfull, 2006). These data therefore lend increasing evidence to the hypothesis that phages utilize the tape measure motifs to mediate localized peptidoglycan hydrolysis during infection.

6.0 DEVELOPMENT OF A STRATEGY TO DELETE THE RPF MOTIF FROM MYCOBACTERIOPHAGE TAPE MEASURES AND CHARACTERIZATION OF MUTANT PHAGES

6.1 INTRODUCTION

An important consideration concerning the mycobacteriophage tape measure Rpf Motifs is the question of whether these are essential for the phage. Based on what is known about other phage virion-associated murein hydrolases, one might predict that the answer is 'no'. For example, an active site mutation in the T7 gp16 hydrolase abolishes enzymatic activity *in vitro*; however, phage containing the mutant protein show normal infectivity in most circumstances (Moak & Molineux, 2000, Moak & Molineux, 2004). Importantly, the mutant viruses are only impaired when infecting host bacteria grown under conditions in which the peptidoglycan is expected to be more highly cross-linked, suggesting that these proteins are only important to confer an evolutionary advantage in specific environmental conditions (Moak & Molineux, 2000, Molineux, 2001). Notably, these are the conditions in which one might expect to find most bacteria in nature (i.e. in stationary phase), as the logarithmic growth observed in culture is likely to be rare in the nutrient-limiting conditions of most natural environments.

A similar phenotype has been observed for the mutants of mycobacteriophage TM4, which lack the region of the TMP known as Motif 3. As noted previously, this motif has similarity to

proteins that have murein hydrolase activity *in vitro*, and importantly, the phage motif can effectively substitute for the corresponding region in its bacterial homologues (Piuri & Hatfull, 2006). Further, TM4 derivatives encoding a TMP with an in-frame deletion of Motif 3 are indistinguishable from wild-type under most conditions. These mutants, however, are impaired in infecting *M. smegmatis* cells that are in late-stationary phase, an effect that appears to result from a specific defect in DNA injection (Piuri & Hatfull, 2006). These data suggest that mycobacteriophages with deletions of the Rpf Motif might have similar phenotypes. It is also noteworthy that in both cases where the Rpf Motif is found, it is present in combination with one of the other tape measure motifs, Motif 2 in Barnyard and Motif 3 in Giles. There are a number of lines of evidence suggesting that the bacterial Rpf proteins might function in conjunction with peptidoglycan peptidases (Hett *et al*., 2007), and this raises the intriguing possibility that these motifs may somehow work together to facilitate localized murein hydrolysis during infection.

Therefore, in light of these observations, another major focus of my work has been to generate a mutant phage from which the Rpf Motif in the TMP has been deleted. This has proved quite difficult due to the fact that a reliable system for mycobacteriophage genetics simply does not exist. In most cases, bacteriophage mutant construction is dependent on the ability of the phage in question to form stable lysogens. That is, once the viral genome is integrated into the host chromosome, and provided adequate systems are available, it can be manipulated in the same way as the host DNA. Mutant phages are then produced by inducing the lysogen and collecting the lysate. This type of strategy has proved immensely useful in lysogenic phages such as λ, in which countless mutants have been constructed and analyzed (Katsura, 1983, Gottesman, 1999, Murray & Gann, 2007).

Unfortunately, most mycobacteriophages, including Barnyard, do not form stable lysogens, and even for those that do, such as Giles (see Appendix), the cues that trigger induction are not known. In some select cases, however, this problem has been overcome by the use of shuttle phasmids. This term is used to refer to bacteriophage derivatives from which a non-essential portion of the genome has been deleted and replaced by an *E. coli* plasmid (Jacobs *et al.*, 1987, Snapper *et al.*, 1988). Thus, shuttle phasmids will replicate as a phage in the mycobacteria but as a plasmid in *E. coli*, which allows for genetic manipulation and mutant construction by the exploiting the plethora of genetic techniques available in this organism. This technology was, in fact, utilized to generate the Motif 3 deletion (Mt3Δ) in TM4 (Piuri & Hatfull, 2006), as a number of phasmid derivatives of this phage have been constructed and are available (Jacobs *et al.*, 1987, Bardarov *et al.*, 2002). The ability to use this technique, however, is absolutely dependent on the ability to make phasmid derivatives of the phage of interest, and because phasmid construction requires packaging in λ particles, the packaging constraints are prohibitive for phages larger than ~50 kb. This precludes the use of this technology for most mycobacteriophages, which contain genomes that are on average, quite large (Pedulla *et al.*, 2003, Hatfull *et al.*, 2006). Thus, phasmid derivatives are available for only a very limited number of mycobacteriophages, meaning that this strategy, while useful, is not widely applicable.

Functional genomic studies of the host mycobacteria have been greatly facilitated by the development of mycobacterial recombineering, a term that generally refers to genetic engineering mediated by recombination proteins (Court *et al.*, 2002). Recombineering was first developed in the model organism *E. coli* by utilizing the Bacteriophage λ Red recombination proteins (Exo and Beta) or the Rac Prophage RecE and RecT proteins, which efficiently promote homologous recombination between double-stranded DNA (dsDNA) or single-stranded DNA (ssDNA)

substrates and homologous targets in the bacterial chromosome (Murphy, 1998, Zhang *et al.*, 1998, Yu *et al.*, 2000, Ellis *et al.*, 2001). Exo/RecE is an exonuclease that binds to and degrades dsDNA in a 5′→3′ manner (Little, 1967, Joseph & Kolodner, 1983). Beta/RecT is a single-stranded binding protein that binds to ssDNA, such as the 3′ tails generated by Exo/RecE, and promotes homologous recombination by catalyzing strand annealing, strand exchange or strand invasion (Hall & Kolodner, 1994, Noirot & Kolodner, 1998, Li *et al.*, 1998, Iyer *et al.*, 2002, Rybalchenko *et al.*, 2004). This system utilizes substrates with short regions of homology and allows the high frequency generation of targeted gene replacements, deletions and point mutations in the chromosome and on exogenously replicating pieces of DNA such as plasmids and bacterial artificial chromosomes (BACs) (Zhang *et al.*, 1998, Muyrers *et al.*, 1999, Yu *et al.*, 2000, Murphy *et al.*, 2000, Ellis *et al*., 2001, Lee *et al.*, 2001) (Reviewed in Muyrers *et al.*, 2001, Court *et al*., 2002). This has made recombineering an extremely useful tool in the study of *E. coli* and other Gram-negative organisms.

Recently, homologues of the Rac prophage recombineering proteins RecE and RecT have been identified in the mycobacteriophage Che9c (van Kessel & Hatfull, 2007). When inducibly expressed in *M. smegmatis* and *M. tuberculosis*, these function in a manner very similar to the *E. coli* λ Red and RecET recombineering systems and have been utilized to generate gene replacements and point mutations at a high frequency on both on the mycobacterial chromosomes and on replicating plasmids (van Kessel & Hatfull, 2007, van Kessel & Hatfull, 2008). This system has greatly facilitated and expedited the construction of mutants in the mycobacteria, particularly in the slow-growing pathogen *M. tuberculosis*.

It has further been shown the λ Red recombineering can be used to make similar mutations on bacteriophage λ DNA (Oppenheim *et al.*, 2004). I therefore hypothesized that mycobacterial

recombineering system could be utilized in a similar fashion to delete the region encoding the Rpf Motif from phage *tmp* genes. More broadly, since genetic manipulation of most of the mycobacteriophages isolated and sequenced to date is difficult, if not impossible, due to the limitations discussed above, it would be tremendously useful to have a system whereby mutations could be made on mycobacteriophage genomic DNA during the lytic cycle. In this chapter, I describe the extension of the mycobacterial recombineering system for use on mycobacteriophages. This was utilized to generate multiple mutations of the conserved motifs within the Giles tape measure, and the phenotypes of mutant phages are described.

6.2 DIFFICULTIES ENCOUNTERED IN THE ATTEMPTS TO PERFORM MYCOBACTERIOPHAGE GENETIC MANIPULATION

Previous attempts to generate mutant bacteriophages from which the Rpf Motif was deleted from the tape measure were met with little success, and some of the various strategies employed with both Barnyard and Giles are briefly discussed below.

6.2.1 Mycobacteriophage Barnyard

Early work focused on mycobacteriophage Barnyard, which is in itself challenging to work with due to its extremely small, turbid plaque morphology and high sensitivity to host cell density (discussed in Appendix). The Barnyard genome is 70,797 kb (Pedulla *et al*., 2003), which means that it far exceeds the packaging constraints of λ and thus is not amenable to shuttle phasmid construction. However, I hypothesized that it should be possible to create a Barnyard cosmid

library, with inserts in the range of ~45-50 kb. Individual cosmids would be missing large regions of the Barnyard genome and would, therefore, be non-infectious when introduced into *M. smegmatis* by electroporation. However, two cosmids that have overlapping sequence and together constitute a complete Barnyard genome (referred to as complementary) potentially could recombine in *M. smegmatis* to yield viable phage. The phenomenon has been noted for overlapping TM4 shuttle phasmids, and it has been hypothesized that the phage might encode recombination proteins that catalyze this recombination event (Jacobs *et al.*, 1987). None, however, can be detected by sequence homology to known recombination proteins, suggesting that the phage proteins are highly divergent, and further implying that other phages, such as Barnyard, might also encode similar proteins. If this were possible, recombineering in *E. coli* could be utilized to generate the Rpf Motif deletion on the cosmid containing the *tmp* gene. This could then be recombined with its complementary partner to yield mutant phage.

A Barnyard cosmid library was therefore constructed by partially digesting Barnyard DNA to generate fragments of ~45-50 kb and ligating these fragments to a linearized vector (pYUB854) backbone that contains *cos* sites for packaging in λ. Concatamers were packaged, and these were used to infect *E. coli*. Individual cosmids from the resulting colonies were isolated, and these were mapped by restriction digest and sequencing. This library, however, proved to be of limited utility for a number of reasons. When these cosmids were mapped, it was discovered that the insert sizes were generally smaller than the 50 kb target, and therefore, it was difficult to find two that were complementary. This was further complicated by the observation that many appeared to contain multiple inserts or to have undergone large-scale inversions. Two were isolated that appeared to be complementary by sequencing (BYcos20 and BYcos27) and these were transformed into electrocompetent *M. smegmatis*. However, while Barnyard DNA transformed in a similar manner

produced plaques (with an efficiency of 8×10^2 pfu/μg), the cosmids were unable to recombine to form viable phage, even when as much as 4 μg of DNA was transformed. Additional complementary pairs were not found and thus could not be tested for recombination. It is therefore not known whether this strategy was unsuccessful due to the inability of complementary Barnyard cosmids to recombine in *M. smegmatis,* or if there were mutations in BYcos20 and/or BYcos27, which made recombinant products unable to produce viable phage.

The difficulties encountered with the Barnyard cosmid library led me to explore other possible strategies to construct the tape measure motif deletion. One of these that would eliminate the necessity of finding complementary cosmid pairs, as well as the dependence on the ability of these to recombine, would be if the entire Barnyard genome could be cloned on one circular molecule. In this regard, bacterial artificial chromosomes (BACs) have proved to be quite useful for the cloning and propagation of large pieces (up to 350 kb) of DNA. Therefore, I attempted to clone the entire Barnyard genome utilizing the pEZ BAC system (Lucigen). A range of different amounts of Barnyard DNA – that was; untreated, phosphorylated or treated with Klenow to ensure blunt ends – was ligated to the blunted and phosphorylated vector supplied by the manufacturer. This system utilizes LacZ-based blue/white screening; therefore, white colonies were picked and screened by restriction digest. Despite numerous attempts to generate these constructs, however, I was unable to clone the complete Barnyard DNA into the vector backbone under any of the conditions tested. The reason for this is unclear; however, it is possible that toxic phage gene products on the constructs were expressed in *E. coli*, or alternatively, the genome ends may be modified in way that inhibits ligation.

By this point in time, J. van Kessel had succeeded in developing the mycobacterial recombineering system and had shown that dsDNA substrates containing 500 bp of homology

upstream and downstream of a gene of interest flanking a resistance cassette could be utilized to generate targeted knock-outs in *M. smegmatis* and *M. tuberculosis* (van Kessel & Hatfull, 2007). I hypothesized that such a system might prove useful for generating mutations in the mycobacteriophages as well. Therefore, I designed and generated a plasmid (pLAM32) that contained 500 bp of homology upstream and downstream of the Rpf Motif. One caveat was that since the mutation was within a phage structural gene that was likely to be essential, I could not disrupt this by adding a bulky insert; the mutation would need to be in-frame. Furthermore, antibiotic selection is not useful in lytically replicating phages. Therefore, I designed the construct such that the regions of homology flanked a unique 9 bp 'linker sequence'. I could then screen potential recombinant plaques by performing plaque lift assays and probing the membranes with a 50 nt radiolabeled probe containing homology to the linker sequence, as well as to sequences upstream and downstream of the new junction created by the deletion. The deletion substrate was liberated from the plasmid pLAM32 by restriction digest, and this linear dsDNA substrate was transformed into electrocompetent recombineering cells containing the plasmid pJV24, which expresses Che9c gp60 and gp61, as well as the two small flanking ORFs, under control of $P_{acetamidase}$ (van Kessel & Hatfull, 2007). These cells were then infected with Barnyard, plated as top agar lawns and the resulting plaques were screened using the plaque-lift strategy described above. From multiple such experiments, a small number of plaques that weakly hybridized with the probe were detected and isolated. However, these were always found to contain only the wild-type *tmp* sequence by PCR analysis. These data suggested that the recombineering might not be amenable for use on phages; however, subsequent analyses have revealed likely reasons why these experiments were unsuccessful, and these are discussed in the conclusions of this chapter.

6.2.2 Mycobacteriophage Giles

A potential solution to the problems encountered with Barnyard was presented with the isolation and sequencing of mycobacteriophage Giles, the only other phage to contain a canonical Rpf Motif within its TMP (Morris *et al*., 2008). The Giles genome is ~54 kb, and therefore, I speculated that this might be just small enough to allow the construction of a Giles shuttle phasmid. These would be generated in much the same way as the Barnyard cosmid library. Briefly, Giles DNA was ligated, subjected to partial digestion to generate fragments of ~45-50 kb, and ligated to linearized pYUB854 as before. The initial ligation step was required due to the fact that Giles contains cohesive rather than terminally redundant ends. Concatamers were packaged into λ, and these were used to infect *E. coli*. The resulting colonies were pooled, and the DNA isolated from these cells was transformed into *M. smegmatis*. Most of these molecules were expected to be non-infectious; however, the prediction was that a small proportion would represent true shuttle phasmids and form plaques. I subsequently screened multiple putative Giles shuttle phasmids; however, in every case the isolated DNA was wild-type and had lost the *E. coli* plasmid. Thus, these plaques had likely resulted from a low level of recombination occurring between complementary cosmids in the pooled DNA. These data therefore suggested that the Giles genome might be slightly too large to be used for shuttle phasmid construction.

These results, however, led me to revisit an earlier strategy that was employed with the Barnyard cosmids. Notably, the Giles library was significantly larger, had a larger average insert size (43.6 kb) and contained a number of complementary cosmid pairs (Fig. 31A). Further, I had already observed that recombination between such pairs (present in the pooled DNA) could likely occur at low levels in wild-type *M. smegmatis*, and the mycobacterial recombineering strains now in existence could potentially increase the efficiency of this homologous recombination between

selected cosmids. I tested this hypothesis and determined that indeed, two complementary Giles cosmids, which did not form plaques individually, could recombine with variable frequencies when electroporated into recombineering *M. smegmatis* and form wild-type plaques (Table 8). A number of these pairs were tested in wild-type *M. smegmatis*, and in some case recombination was observed, although the frequency was reduced. Importantly, it was now conceivably possible to generate mutant phage by making the Giles Rpf Motif deletion on a cosmid containing the TMP gene and recombining this construct in *M. smegmatis* with a complementary partner.

In order to do this, I attempted to utilize ssDNA recombineering in *E. coli* to make the deletion, as described in Piuri and Hatfull, 2006. Briefly, recombineering functions were induced in recombineering strains of *E. coli* (Court *et al*., 2002), and these cells were transformed with the Giles cosmid containing the *tmp* gene and a 200 bp deletion substrate (this was generated by PCR as described by Piuri & Hatfull, 2006, with a strategy similar to that illustrated in Figure 32B). DNA from co-transformations with candidate cosmids and the recombineering substrate was pooled and screened for the presence of the deletion by PCR; these data indicated that the deletion was generated and present at low levels in the pooled cosmid DNA (Fig. 31B). Large-scale PCR screening of subsets of these populations as well as individual clones was therefore undertaken in order to locate and isolate cosmids with the RLF deletion. However, a number of difficulties were encountered, and the available data suggest that many of the constructs were either lost or were mutated in all of the *E. coli* strains utilized for their propagation. Therefore, I attempted yet another recombineering based strategy, and decided to return, once again, to the mycobacterial system.

Table 8. Giles cosmid recombination frequencies in recombineering *M. smegmatis*.

Cosmids	*Plaques*	*pfu/μg*	*Plaquing Efficiency*
40 × 62	56; 40	70; 50	N/A; 0.36%
54 × 33	3	3.8	N/A
54 × 45	7	8.8	N/A
54 × 62	3	3.8	N/A
54 × 113	22	27.5	N/A
75 × 45	3	3.8	N/A
97 × 82	1	1.3	N/A
99 × 62	1	1.3	N/A
102 × 82	2	2.5	N/A
105 × 45	8	10	N/A
120 ×113	83, 256	104.8; 320	N/A; 2.29%
124 × 62	19	23.8	N/A
10 × 40	55; 21	68.8; 26.3	1.15%; 0.19%
12 × 17	>400	~500	~8.33%
32 × 17	109	136.3	2.27%
41 × 54	5	6.3	N/A

Plaquing efficiency is equal to pfu/*μg* obtained with total cosmid DNA divided by pfu/*μg* obtained with WT Giles DNA.

Pair is compatible, but both cosmids contain *tmp* gene.

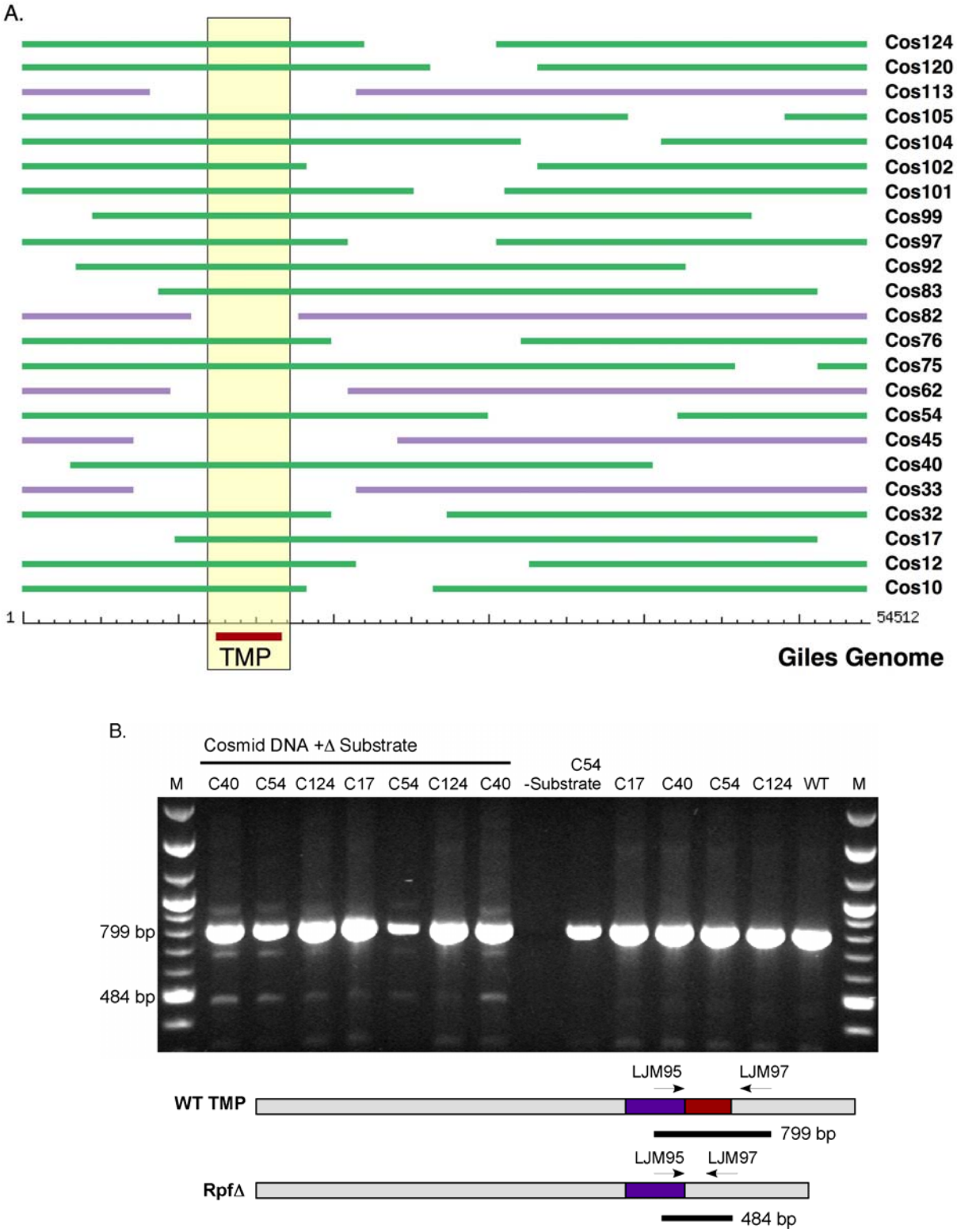


Figure 31. Construction of the Giles cosmid library and recombineering in *E. coli.*

Figure 31: Construction of the Giles cosmid library and recombineering in *E. coli*. **A.** A mycobacteriophage Giles cosmid library was constructed (ave. insert size 43.6 kb) and mapped. Complementary pairs were chosen such that one contained the entire *tmp* – this would be the molecule on which the deletion was generated – and one was deleted for the entire *tmp* gene. **B.** Candidate cosmids and the deletion substrate were transformed into the *E. coli* recombineering strain DY331 (kindly provided by Donald Court), which contains a defective λ prophage that expresses the Red recombineering functions under the control of a temperature sensitive cI repressor. Cells were induced prior to electroporation by incubating a culture – which was grown at 30°C to an OD_{600} of ~0.6 – at 42°C for 15 minutes. These cells were then chilled on ice for 15 minutes, washed three times with ice-cold water and used immediately. Total DNA from recombineering experiments was pooled (C = Cosmid) and used as a template in PCR reactions with primers that anneal upstream and downstream of the motif. Sites of annealing of the diagnostic primers relative to the wild-type *tmp* and *rpfΔ tmp* genes and sizes of the expected products are illustrated. Control PCRs were also performed on cosmid DNA alone or on DNA isolated from a recombineering experiment from which the deletion substrate was omitted. If cosmids with the deletion are present in the pools, a 484 bp band should be visible, in addition to the 799 bp wild-type band. These data indicate that the Rpf Motif was deleted from a small subset of the cosmid DNA molecules in transformations to which the deletion substrate was added.

6.3 UTILIZING RECOMBINEERING IN *M. SMEGMATIS* TO GENERATE MUTATIONS ON MYCOBACTIOPHAGE DNA

In addition to gene replacements, which require dsDNA substrates and both the Exo and Beta proteins, the *E. coli* Red recombineering system can be utilized to generate point mutations, deletions and small insertions in the bacterial chromosome and on plasmids. These can be made with ssDNA substrates, and this requires only the single-stranded binding protein (Beta/RecT). It was subsequently determined that mycobacterial recombineering, in much the same way, could utilize ssDNA substrates to engineer non-selectable point mutations on both the chromosome and on extrachromosomal plasmids (van Kessel & Hatfull, 2008). Importantly, due to the fact that the transformation efficiency of mycobacteria is quite low as compared to *E. coli*, the construction and detection of these mutants requires an essential co-selection step. This involves transforming recombineering cells with both the ssDNA substrate to make the desired mutation, as well as with a selectable substrate (either a plasmid or an additional recombineering substrate which makes a point mutation that can be selected for). Transformed cells are plated in the presence of antibiotic selection, and this effectively selects against the non-transformable cells that make up the majority (99.9%) of the population. These can then be screened for the presence of the point mutation that was not selected for, which can typically be found at a frequency of ~10% (van Kessel & Hatfull, 2008).

These observations suggested that it might be possible to use a similar strategy to generate non-selectable mutations on bacteriophages. Rather than transforming the recombineering substrate and infecting these cells with phage (as I had previously done with Barnyard), co-

transformations could be utilized in which the phage DNA performs the same function as the selectable marker. That is, only competent cells will be able to take-up the phage DNA and subsequently, if transformations are plated prior to lysis, only these cells will form plaques in top agar overlays. These cells will have also been competent to take up the DNA substrate, thus greatly increasing the likelihood of recovering the mutation. I therefore decided to utilize this strategy to construct the Rpf mutation in mycobacteriophage Giles, which was selected over Barnyard due to the fact that it is, in general, much less difficult to work with.

6.3.1 Generation of the Rpf Motif Deletion in Mycobacteriophage Giles

In mycobacterial recombineering, the minimum length requirement for efficiently generating point mutations with ssDNA oligonucleotides (oligos) is 48 nt (van Kessel & Hatfull, 2008). However, it was not known if the same would be true for phage recombineering or for the construction of unmarked deletions, as these had not been yet been tested. It was also reported that heat-denatured dsDNA substrates containing 100 nt of homology upstream and downstream of the deletion were utilized to construct the Mt3Δ in TM4 (Piuri & Hatfull, 2006), and attempts with smaller regions of homology (50 nt) were unsuccessful (M. Piuri, unpublished data). Further, there is evidence in *E. coli* (with both the λ Red and RecET recombineering systems) that protein-protein interactions between cognate protein partners (Exo/Beta and RecE/T) may promote increased levels of recombination, and DNA substrates generated by the exonuclease might be a preferentially bound by the recombinase (Muyrers *et al.*, 2000, Zhang *et al.*, 2003, Yu *et al.*, 2003). Therefore, despite the fact that single-stranded oligos are typically utilized to make deletions with Red recombineering system, I decided to attempt to construct the 315 bp Rpf Motif deletion (RpfΔ) using both a 100 nt oligo (LJM88) as well as a 200 bp dsDNA substrate (ds200ΔRpf) (Fig. 32A

and B). This was generated by extending LJM88 (which has 50 nt of homology upstream and downstream of the motif) by PCR with two 'extender primers' that add an additional 50 bp of homology to each end of the substrate, illustrated in Figure 32B (Piuri & Hatfull, 2006).

Each of these substrates was co-transformed with Giles genomic DNA into electrocompetent *M. smegmatis* cells expressing mycobacteriophage Che9c gp60 and gp61 proteins (Exo and Beta functional analogs) under control of the *acetamidase* promoter in the plasmid pJV53 (van Kessel & Hatfull, 2007, Parish *et al.*, 1997). Transformations were recovered for two hours and plated as top agar overlays with additional 'plating cells' of wild-type *M. smegmatis*. Since the frequency with which the mutations might be recovered was not known, individual plaques were combined and screened in groups of three by performing PCR with primers that anneal outside of the Rpf Motif (Fig. 32C). Thus, if the deletion is present, a PCR product that is smaller than the wild-type band by the size of the deletion will be amplified, in this case ~300 bp smaller. When screened in this way, it was observed that the deletion was present in five out of eight pools screened from transformations with ds200ΔRpf and possibly in two of eight from transformations with LJM88 (Fig. 32D), which indicates that the recombineering was successful in a subset of the phage DNA amplified in the initial infectious center.

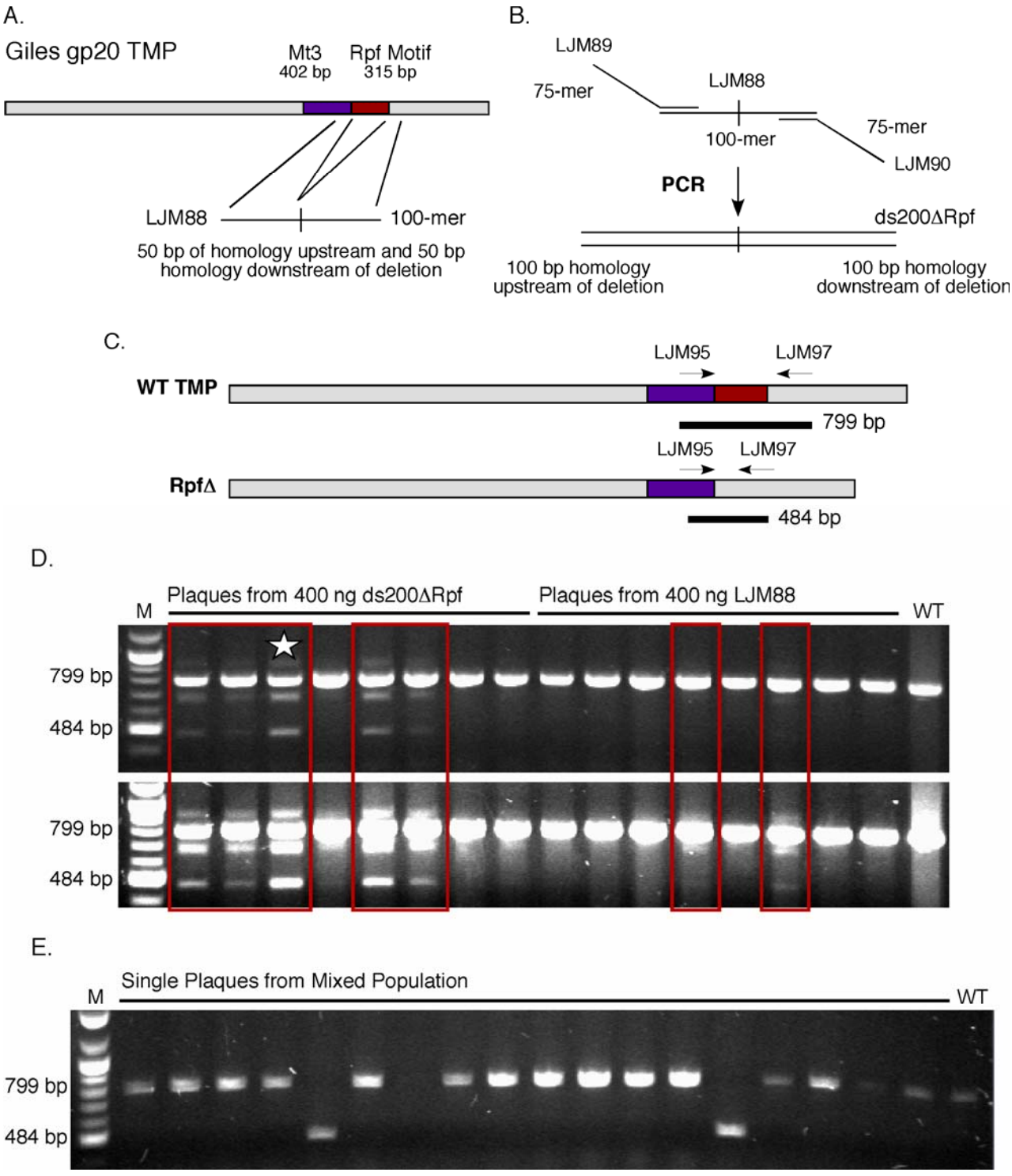


Figure 32. Utilizing recombineering in *M. smegmatis* to delete the Giles Rpf Motif.

Figure 32: Utilizing recombineering in *M. smegmatis* to delete the Giles Rpf Motif. **A.** Schematic illustrating the Giles TMP and the region containing the Rpf Motif that is deleted by recombineering. **B.** Strategy for generating the 200 bp deletion substrate. An oligo (LJM88) was purchased (IDT) containing 50 nt of homology upstream and downstream of the Rpf Motif. This was extended by PCR, so that the final product is double-stranded and has 100 bp homology upstream and downstream of the Rpf Motif. **C.** Sites of annealing of the diagnostic primers relative to the wild-type *tmp* and *rpfΔ tmp* genes and sizes of the expected products. **D.** Electrocompetent *M. smegmatis* cells containing the recombineering plasmid pJV53 from which the recombination functions had been induced were transformed with Giles DNA (50 ng) and the deletion substrates (400 ng), recovered for 2 hours and plated as top agar lawns in 0.35% MBTA with 1 mM $CaCl_2$ and additional wild-type plating cells of *M. smegmatis*. Diagnostic PCRs were performed on pooled DNA (3 plaques per reaction) from recombineering experiments with ds200ΔRpf or LJM88. Pools containing the deletion are boxed in red. **E.** Phages from one of the mixed populations detected in D. (starred lane) were serially diluted, and individual plaques were screened by PCR with the diagnostic primers. These data illustrate that it is possible to use recombineering in *M. smegmatis* to generate deletions on lytically replicating mycobacteriophages, and further, the Rpf Motif is not essential in mycobacteriophage Giles.

Strikingly, the groups of plaques from the transformations with the dsDNA substrate appeared to contain a much greater proportion of mutant DNA than those from transformations with LJM88. Two of these pools that seemed to have the greatest amount of the mutant were serially diluted and plated in top agar lawns, and individual plaques were screened for the presence of the deletion by PCR with the same diagnostic primers. Pure populations of mutant Giles were present in two of 18 plaques screened from one pool and in zero out of 18 screened from the other (Fig. 32E). These were plaque purified one additional time, and the presence of the deletion was confirmed by DNA sequencing. The mutant phage also appeared otherwise normal by restriction digest. These data show for the first time that it is possible to utilize mycobacterial recombineering to make deletions on bacteriophage genomes, and suggest that 200 bp substrates with 100 bp of homology flanking the deletion are more efficient for this process than ssDNA substrates.

6.3.2 Generation of the Motif 3 Deletion in Mycobacteriophage Giles

Due to the ease with which the Rpf mutant phage was constructed, I decided to utilize recombineering to generate a Giles mutant with a deletion of the tape measure Motif 3 (Mt3Δ), as well as to test some additional parameters and optimize the system. In both *E. coli* and the mycobacteria, oligos that are complementary to the lagging strand are more efficient substrates for recombineering than those that anneal to the strand undergoing leading strand synthesis (van Kessel & Hatfull, 2008). However, in Giles the direction and mode of replication are unknown. Therefore, although the results of the previous experiment suggested that dsDNA substrates are more efficient for this recombineering than ssDNA substrates with 50 nt of homology, it is also

possible that the ssDNA oligo (LJM88) was inefficient in generating the Rpf deletion because it is complementary to the leading strand.

Therefore, in this case I synthesized two complementary 100 nt oligos – LJM119 and LJM120 – and tested both separately for recombineering (Fig. 33A). In order to determine if a dsDNA substrate of the same length was more efficient than the ssDNA oligos, these were annealed to generate a 100 bp product (ds100ΔMt3). Finally, a 200 bp dsDNA product (ds200ΔMt3) was created by extending the 100 nt oligo LJM119 by PCR with two 75 nt 'extender primers', as described above (Fig. 33B). Each of the substrates – LJM119, LJM120, ds100ΔMt3, and ds200ΔMt3 – was co-transformed into *M. smegmatis* mc^2155:pJV53 with mycobacteriophage Giles DNA and plated prior to lysis as before. This time, individual plaques were screened for the presence of the deletion by PCR (Fig. 33C), in order to better calculate frequency. In this way, I was able to identify plaques containing the deletion product with the highest frequency in transformations with ds200ΔMt3 (Table 9). Further, the mutant product appeared to represent a greater 'proportion' of the mixed population in plaques positive for the deletion obtained from transformations with ds200ΔMt3 (Fig. 33C), although this is difficult to quantify. The deletion was not detected when ds200ΔMt3 was transformed into control cells that do not express the Che9c recombineering proteins. This indicates that although Giles encodes a putative RecT homologue, the recombineering is dependent upon the presence of the Che9c proteins in the recombineering strain of *M. smegmatis*.

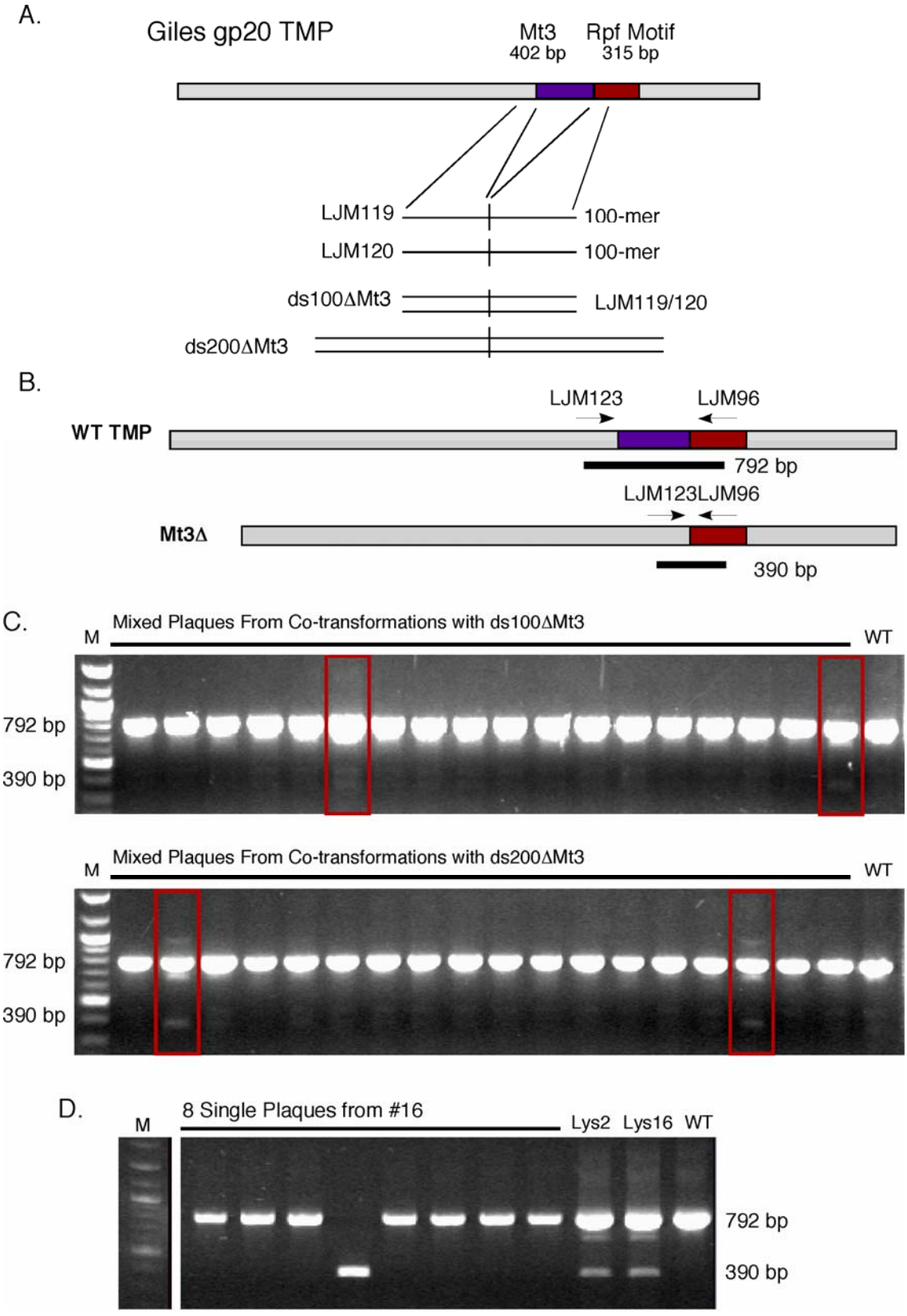


Figure 33. Construction of the Giles Mt3Δ by recombineering in *M. smegmatis.*

Figure 33: Construction of the Giles Mt3**Δ** by recombineering in *M. smegmatis*. **A.** Schematic illustrating the region of the Giles TMP containing Motif 3 that is deleted by recombineering, as well as the various recombineering substrates tested in these experiments. **B.** Sites of annealing of the diagnostic primers relative to the wild-type *tmp* and *mt3Δ tmp* gene and sizes of the expected products. **C.** *M. smegmatis* cells containing the recombineering plasmid pJV53 were transformed with 50 ng Giles DNA and 200 ng of each of the substrates illustrated in part A, as described in Figure 32D. Diagnostic PCRs performed on individual plaques from recombineering experiments with ds200ΔMt3 or ds100ΔMt3 illustrate that the mixed populations obtained from transformations with the longer substrate appear to contain a higher proportion of mutant product. Populations containing the deletion are boxed in red. **D.** Phages from the mixed populations detected in the first round of screening (from ds200ΔMt3) were serially diluted, and both individual plaques and lysates (Lys) collected from plates with ~10,000 plaques were screened by PCR with the diagnostic primers. PCRs performed on the individual plaques from one set of dilutions are shown, and well as the PCRs from both lysates (Lys2 and Lys16). These data demonstrate that the Giles Motif 3 is not essential.

Table 9. Mycobacteriophage Giles recombineering frequencies in *M. smegmatis*.

Mutation	*Size*	*Substrate*	*Amount*	*PCR*	*Frequency*	*MAMA -PCR*	*Frequency*
Mt3Δ	402 bp	LJM119-100 nt	200 ng	1/29	3.4%	N/A	N/A
Mt3Δ	402 bp	LJM120-100 nt	200 ng	0/29	0%	N/A	N/A
Mt3Δ	402 bp	dsDNA-100 bp	200 ng	1/19	5.3%	N/A	N/A
Mt3Δ	402 bp	dsDNA-200 bp	200 ng	3/29	10.3%	N/A	N/A
RpfΔ/Mt3Δ	717 bp	LJM123-100 nt	150 ng	0/18	0%	N/A	N/A
RpfΔ/Mt3Δ	717 bp	LJM124-100 nt	150 ng	0/18	0%	N/A	N/A
RpfΔ/Mt3Δ	717 bp	dsDNA-100 bp	150 ng	1/18	5.6%	N/A	N/A
RpfΔ/Mt3Δ	717 bp	dsDNA-200 bp	150 ng	4/18	22.2%	N/A	N/A
RpfE-A/Mt3Δ	402 bp	dsDNA-200 bp	100 ng	6/36	16.7%	12/36	33.3%
RpfE-A/Mt3Δ	402 bp	dsDNA-200 bp	200 ng	6/36	16.7%	12/36	33.3%
RpfE-A/Mt3Δ	402 bp	dsDNA-200 bp	300 ng	6/36	16.7%	14/36	38.9%
RpfE-A/Mt3Δ	402 bp	dsDNA-200 bp	400 ng	4/36	11.1%	15/36	41.7%

Ge nerated by dividing the number of positive mixed-plaques by total number screened.

Two positive plaques that appeared to have the largest deletion to wild-type ratio were serially diluted and plated, and individual plaques were screened by PCR with the diagnostic primers as before. From one mixed population, a seemingly pure population of mutant phage was recovered in one out of eight plaques screened (Fig. 32D); zero of eight were recovered from the other mixed population. Additionally, single plate lysates were generated from each set of

dilutions, from plates containing ~10,000 plaques. These were screened by PCR, and both showed wild-type and mutant product (Fig. 33D). Thus, although the mutant was only recovered from one of the two mixed plaques examined, it is likely that further screening would have revealed more mutants. The presence of the deletion was confirmed by sequencing, and again the mutant phage appeared otherwise normal by restriction digest. These results confirm that mycobacterial recombineering can be used to generate unmarked deletions in mycobacteriophage genomes, and strongly suggest that this is more efficient with 200 bp dsDNA substrates than with shorter double-stranded or single-stranded substrates.

6.3.3 A Double Deletion of the Rpf Motif and Motif 3 in the Giles Tape Measure is Not Viable

Once the two single tape measure motif deletions were constructed, I wanted to determine if I could create the double-deletion mutant (ΔΔ). The deletion substrates were therefore designed such that they would remove 717 bp of the Giles *tmp* gene encompassing the 402 bp of Motif 3 plus the 315 bp encoding the Rpf Motif downstream of this (Fig. 34A). Again, to confirm the previous results, I tested 100 nt ssDNA substrates (LJM123 and LJM124), a 100 bp substrate resulting from annealing of the 100 nt complementary ssDNA substrates (ds100ΔΔ) and a 200 bp PCR-generated substrate (ds200ΔΔ), as illustrated in Figure 34A and B. When these were transformed into the recombineering strain, the deletion product was detected in 22.2% of plaques screened from transformations with ds200ΔΔ and was found in only one other plaque from the transformation with the ds100ΔΔ substrate (Table 9; Fig. 32C). This lends further support to the assertion that phage recombineering is more efficient with 200 bp dsDNA substrates.

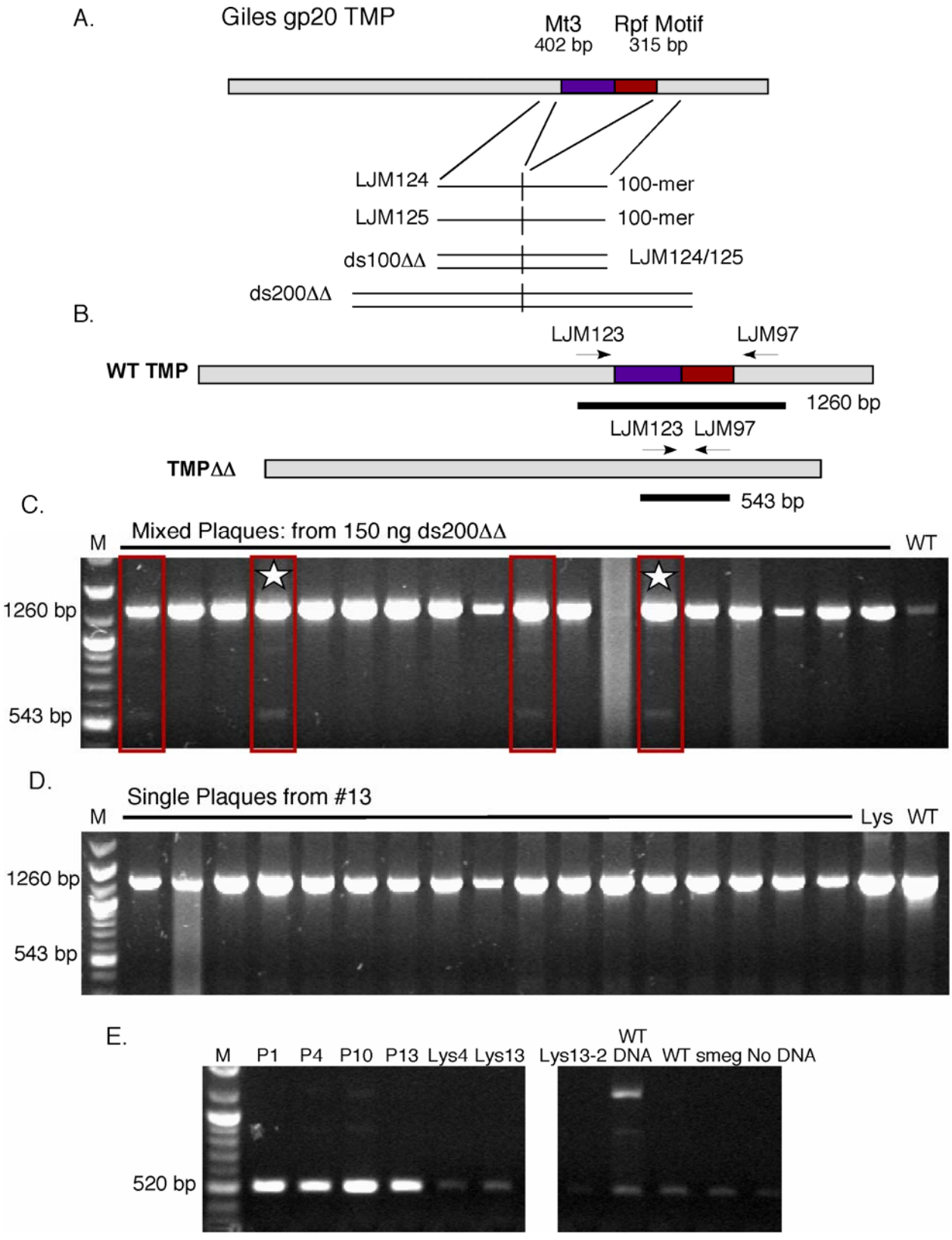


Figure 34. The Motif 3/Rpf double deletion is not viable.

Figure 34: The Motif 3/Rpf double deletion is not viable. **A.** Schematic illustrating the region of the Giles TMP containing Motif 3 and the Rpf Motif that is deleted by recombineering, as well as the various recombineering substrates tested in these experimints. **B.** Sites of annealing of the diagnostic primers relative to the wild-type *tmp* and *ΔΔtmp* gene and sizes of the expected products. **C.** *M. smegmatis* cells containing the recombineering plasmid pJV53 were transformed with 50 ng Giles DNA and 150 ng of each of the substrates illustrated in part A., as described in Figure 32D. Diagnostic PCRs performed on individual plaques from recombineering experiments with ds200ΔΔ illustrate that the double-deletion is generated and can be detected in the initial round of screening. Populations containing the deletion are boxed in red. **D.** Phages from the mixed populations detected in the first round of screening were serially diluted, and both individual plaques and lysates (Lys) collected from plates with ~10,000 plaques were screened by PCR with the diagnostic primers. These data suggest that phage containing the double-deletion are inviable. **E.** Mismatch amplification mutation assay (MAMA)-PCR was performed on the individual plaques that were positive for the deletion by diagnostic PCR, as well as on lysates generated from those populations, wild-type phage and both *M. smegmatis* and no DNA controls. These results confirm (despite the low levels of background visible even in the control lanes) that the double-deletion cannot be propagated.

As with the other mutants, two positive plaques that appeared to have the largest deletion to wild-type ratio were serially diluted and plated, and individual plaques were screened by PCR with the diagnostic primers. In this case, however, all of these were wild-type (Fig. 34D). Importantly, when single plate lysates from each set of dilutions – again from plates containing ~10,000 plaques – were screened by PCR, these only showed wild-type product as well (Fig. 34D). This strongly suggests that the mutant phages present in the initial mixed populations were non-infectious and could therefore not be propagated.

In order to confirm this, a primer was designed such that the 3′ end (3 nt) anneals only to the mutant template across the new junction created by the deletion; this creates a mismatch between the primer and the wild-type *tmp* sequence, which ensures that the mutant product is synthesized more efficiently even if it is found at very low levels in the mixed population. This primer was then used, along with a downstream primer complementary to both wild-type and mutant templates to perform a variation of mismatch amplification mutation assay (MAMA)-PCR on the plaques that were positive by diagnostic PCR, as well as on the lysates generated from the re-plating of these plaques (Swaminathan *et al.*, 2001) (Fig. 34E). These results confirm that while the mutant product is present in the initial mixed populations, it is unable to be propagated in the second round of infection.

The reason for this is not known, although it is possible that the TMP motifs work together to perform an essential function, and therefore, phage lacking both of these are not viable. Perhaps a more plausible explanation is that this deletion gives rise to phage whose tails are simply too short to function properly; a defect that may actually arise from the fact that the TMP itself is too small to perform some essential function in DNA injection. It has been observed that although mutant derivatives of λ with large deletions in the *tmp* gene form tails of defined length, these are

non-infectious, and even viable phages with smaller deletions are somewhat defective for DNA injection (Katsura & Hendrix, 1984, Katsura, 1987). Due to the way in which the Giles mutants were constructed, it was not possible to examine the ΔΔ phages microscopically, and therefore, the possibility that these are completely defective for tail formation can also not be ruled out.

6.3.4 Simultaneous Construction of Giles Mutants with a Motif 3 Deletion and a Point Mutation in the Putative Active Site Residue of the Rpf Motif

While I was in the initial phases of testing the utility of recombineering in mycobacteriophages, I determined that it was also possible to generate point mutations by mutating the predicted active site glutamate (GAA) codon in the Rpf Motif to an alanine (GCT). In this case, I utilized 100 nt complementary substrates (200 ng of each transformed together but not annealed) that create the non-synonymous mutation and simultaneously incorporate synonymous substitutions, which create an AluI restriction site for screening (Fig. 35A). These mutants were detected by PCR amplifying the region of the *tmp* containing the Rpf Motif and digesting the products with AluI (Fig. 35A). Pure populations of the mutant were then isolated essentially as described above, identified by AluI digest of diagnostic PCRs (Fig. 35B), and confirmed by DNA sequencing. As with the Rpf and Motif 3 deletions, DNA from mutant phages appeared otherwise normal by restriction digest.

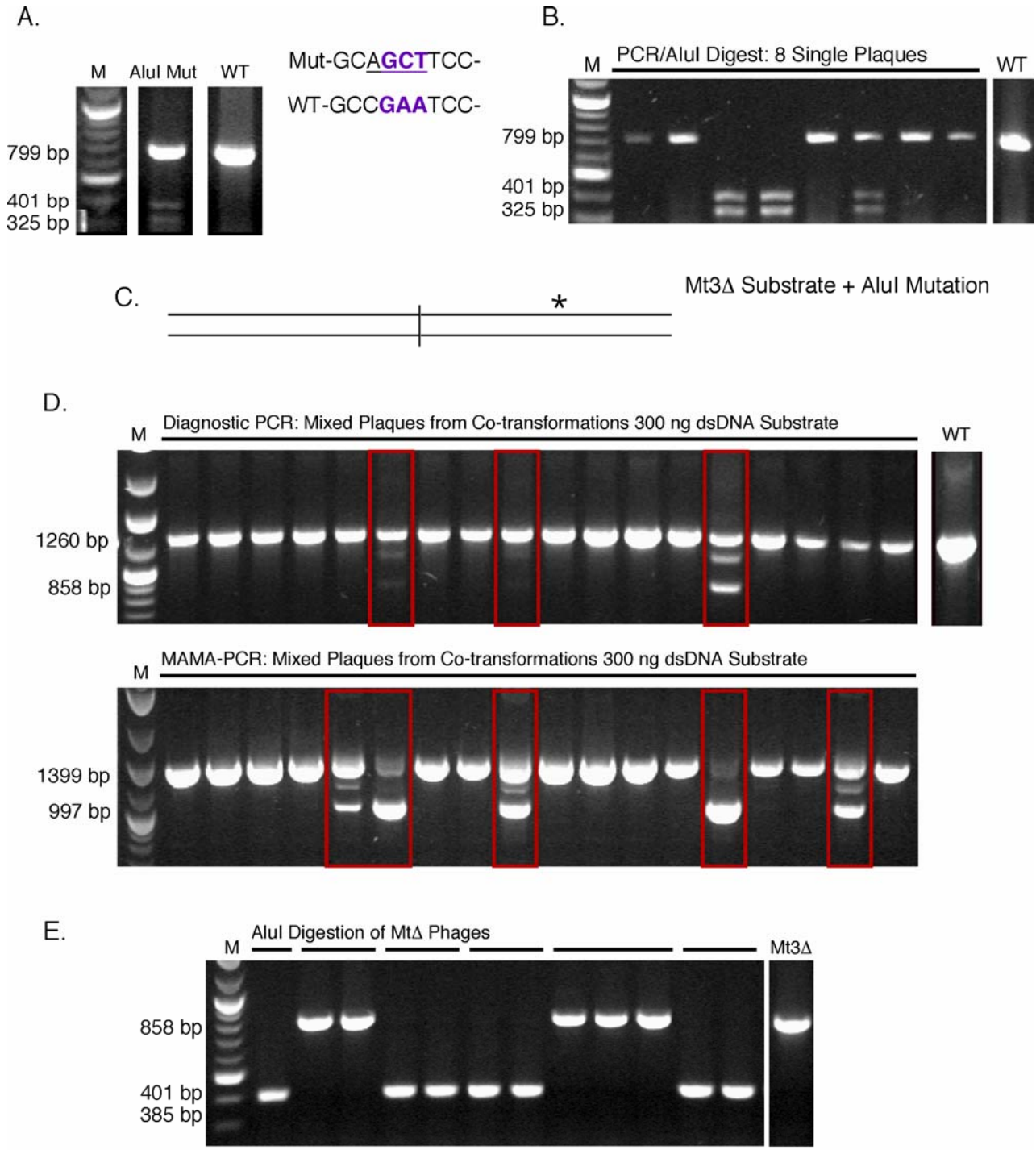


Figure 35. Simultaneous incorporation of an Rpf point mutation and a Mt3Δ deletion in mycobacteriophage Giles.

Figure 35: Simultaneous incorporation of an Rpf point mutation and a Mt3**Δ** in mycobacteriophage Giles. **A.** Generation of a point mutation in the mycobacteriophage Giles *tmp* that changes the putative active-site glutamate (GAA) in the Rpf Motif to an alanine (GCT) and incorporates an AluI restriction site for screening. *M. smegmatis* cells containing the recombineering plasmid pJV53 were transformed with 50 ng Giles and two complementary 100 nt oligos that incorporate the mutation (200 ng each), as described in Figure 32D. Plaques containing mutant phage – and thus the AluI restriction site – can be detected in the first round of screening by PCR amplification and restriction digest of the products with AluI. **B** Mutant phages can be isolated from mixed populations by serial dilutation and re-plating, and these are detected by PCR from isolated plaques and restriction digest of the products with AluI. **C.** A 200 bp Motif 3 deletion substate was generated that incorporates the Rpf point mutation 57 bp downstream of the Motif 3 deletion site. **D.** *M. smegmatis* containing pJV53 were transformed with 50 ng Giles DNA and various amounts of the 200 bp dsDNA deletion substrate, as described in Figure 32D. Shown are diagnostic PCRs and MAMA-PCRs performed on plaques from transformations with 300 ng dsDNA deletion substrate. Phages containing the Mt3Δ can be detected by both methods; however, the frequency with which they are found is higher with the more sensitive MAMA-PCR. Pools containing the Mt3Δ were serially diluted and screened for pure populations of Mt3Δ phage by diagnostic PCR. **E.** When pure populations of Mt3Δ phage were found, these were screened again by PCR and AluI restriction digest to determine how many of these contained the AluI/E-A mutation. The Mt3Δ phage previously constructed was amplified as a control, and plaques arising from the same population are indicated by the discontinuous line. The point mutation was found in four out of seven mixed-plaques screened, and mutants isolated from the same mixed population were identical with regards to the presence of the Rpf point mutation.

The region altered by the incorporation of the point mutation and the AluI site is located approximately 50 bp downstream of Motif 3, so when I was unsuccessful in creating the double deletion, I decided to test whether I could simultaneously incorporate the Rpf point mutation (RpfE-A) and the Mt3Δ in wild-type Giles. At the same time, I titrated the amount of deletion substrate transformed in order to see if it was possible to determine an optimal concentration for recombineering. In this case, the 200 bp deletion substrate was generated as described above, except that the downstream extender primer incorporated the AluI mutation and the glutamate to alanine substitution 57 bp – 60 bp downstream of the deletion junction (Fig. 34C). Various amounts of this substrate were co-transformed into *M. smegmatis* mc^2155:pJV53 with Giles DNA, and the resulting mixed plaques were initially screened for the deletion by diagnostic PCR as described above (Fig. 35D). Using this method of identification, I was unable to detect a difference in the frequency with the different amounts of 200 bp substrate (Table 9). Interestingly, however, positive plaques obtained from transformations with more substrate DNA in general, appeared to contain a greater proportion of mutant DNA.

This suggested that perhaps there were plaques containing small amounts of mutant DNA that were below the level of detection. I therefore designed a MAMA-PCR primer similar to the one described above, whose 3′ end anneals to the mutant template across the new junction created by the deletion. Again this was used along with a downstream primer complementary to both wild-type and mutant templates to perform MAMA-PCR (Swaminathan *et al.*, 2001) on the same plaques screened by diagnostic PCR (Fig. 35D). Unlike standard MAMA-PCR, in our assays the wild-type product – which is larger than the mutant product – is sometimes amplified; however, the mutant product is still favored. Using this more sensitive method of detection, it was confirmed that more positive plaques containing the Mt3Δ could be detected with the MAMA-PCR primers

than with the standard diagnostic PCR primers (Fig. 35D and Table 9). Mutants were detected at the greatest frequency from transformations with the most deletion substrate (400 ng); however, the differences were minimal (Table 9).

In order to determine the proportion of Mt3Δ plaques that also contained the AluI site mutation, I isolated 22 individual Mt3Δ mutants arising from seven different mixed plaques as described above. These were screened by PCR followed by restriction digest of the resulting products with AluI (Fig. 35E). I found that Mt3Δ phages from four out of seven mixed plaques also contained the AluI site mutation; phages from three of the seven mixed populations contained only the Mt3Δ. Further, I was unable to isolate any deletion mutants from three mixed plaques (from which I screened 25, 26 and 27 isolated plaques) that appeared to contain a small proportion of mutant DNA, and in general, as one might expect, it was easier to isolate mutants from the plaques that appeared to contain more mutant DNA by standard diagnostic PCR. In the cases where mutants were found, the frequency with which these were able to be isolated from the mixed plaques ranged from 6.7% to 66.7% (~14% on average). In one striking example, ten Mt3Δ phages were isolated from one mixed population that appeared to contain more mutant DNA than wild-type DNA, and in this case 10/15 phages that were screened (66.7%) were mutant. These results demonstrate that it is possible to insert multiple different kinds of mutations with one substrate, although both must be confirmed, as not all progeny phage will contain both mutations. Interestingly, it was generally observed that all mutant phages arising from the same mixed plaque were identical. As it was only possible to isolate mutants from seven mixed plaques (and from one population only one was found), I cannot say for certain that this will always hold true. However, that fact that in one case, ten identical mutants were isolated from one mixed population suggests that it may be.

6.4 PHENOTYPIC CHARACTERIZATION OF MUTANT PHAGES

6.4.1 Mutant Phages Have a Normal Plaque Morphology

All mutant phages, as a consequence of their isolation and purification, were shown to be viable, and furthermore, no obvious defects were observed for these phages during this process. In order to examine this more closely, mutant lysates were serially diluted and spotted along with wild-type Giles onto freshly poured top agar lawns of *M. smegmatis*. In all cases, the plaques generated by the mutant phages were indistinguishable from those of wild-type phage (Fig. 36). Differences were also not observed when older cultures of *M. smegmatis* (two weeks to one month old) were utilized to make the lawn or when the bacteria were infected in liquid medium prior to plating. Although when larger amounts of bacteria were used to seed the lawn, it was generally observed that the plaque sizes of all phages were decreased. Further, it was noted that the titers of plate lysates from the double mutant were often lower than that of the other phages, but the differences were slight and therefore, difficult to quantify.

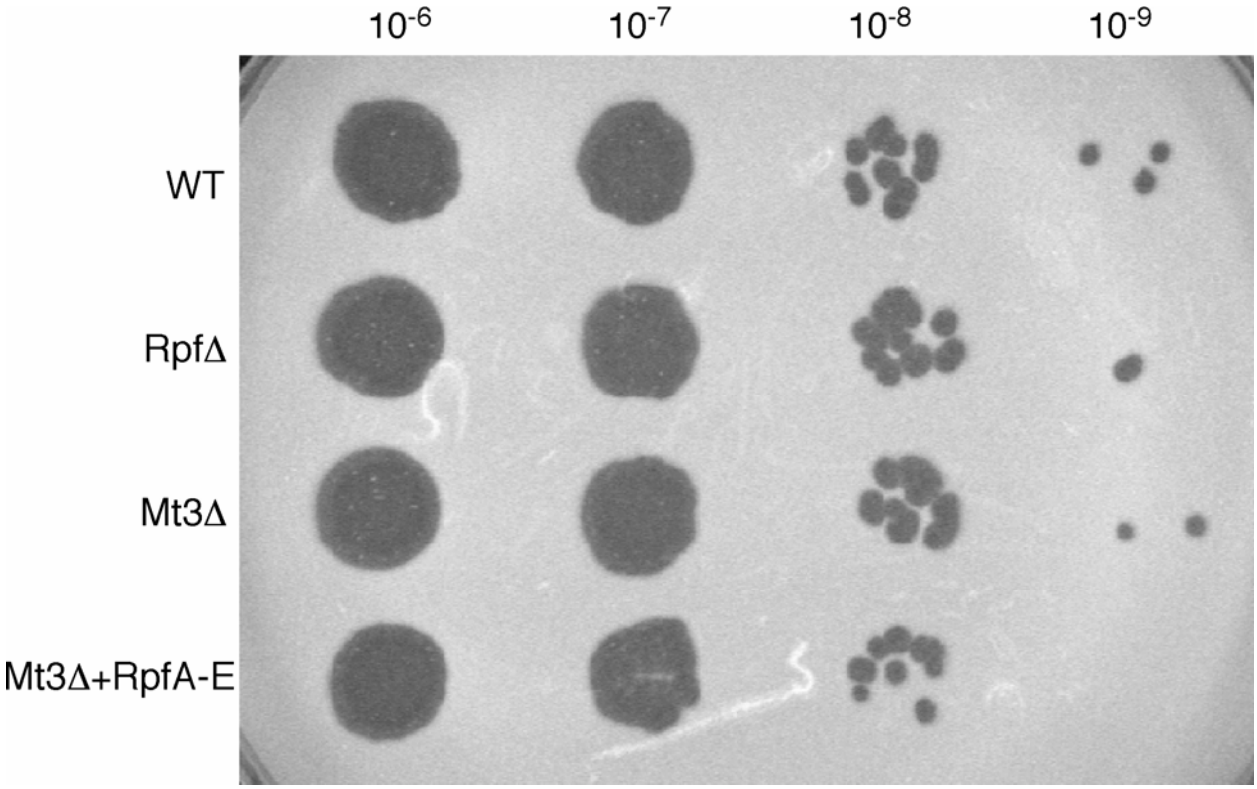


Figure 36. Plaque morphology of mutant phages.

Figure 36: Mutant phages have a normal plaque morphology. A wild-type Giles lysate and lysates from the various tape measure mutants were serially diluted and spottted (5 μl) onto a fresh top agar lawn (0.35% MBTA with 1 mM $CaCl_2$) seeded with an early stationary-phase culture of *M. smegmatis*. The plaques formed by the Giles *tmp* mutants were indistinguishable from those formed by wild-type Giles.

6.4.2 Mutant Phages Are Defective in Infecting Late-Stationary Phase Cells

It has previously been shown that a shuttle phasmid derivative of mycobacteriophage TM4, which is deleted for the tape measure Motif 3, infects late-stationary phase *M. smegmatis* cells with a reduced (~50%) efficiency as compared to wild-type phasmids (Piuri & Hatfull, 2006). Therefore, similar experiments were performed with the various Giles TMP mutants in order to determine if they would display the same phenotype. For each experimental replicate, a culture of *M. smegmatis* was grown and sampled at various time-points as it transitioned from exponential growth into early- and then late-stationary phase. At each of these time-points, the cells were infected with multiple dilutions of wild-type Giles and the mutant phages. The same dilutions were also utilized to infect cells from an exponentially growing *M. smegmatis* culture, and the titers obtained on the aged culture were normalized to these values. Multiple replicates were performed and the data are summarized in Figure 37.

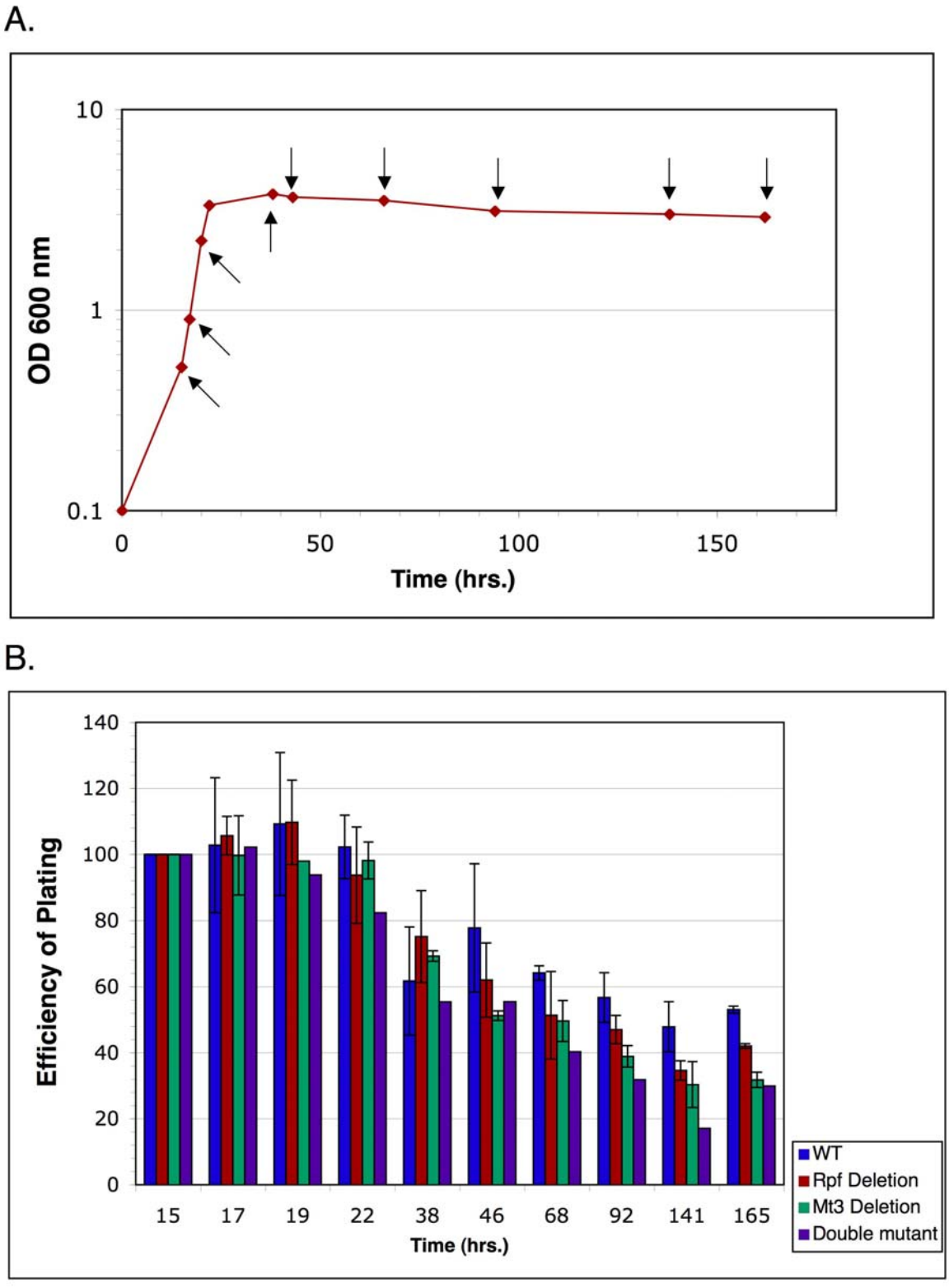


Figure 37. Giles tape measure mutants infect late stationary phase *M. smegmatis* with a reduced efficiency.

Figure 37: Giles tape measure mutants infect late-stationary phase *M. smegmatis* with a reduced efficiency as compared to wild-type. Mycobacteriophage Giles and each of the tape measure mutants (except the one containing the Rpf point mutation alone) were repeatedly titered in triplicate on a *M. smegmatis* culture as it transtioned from exponential into stationary phase. **A.** Growth curve of a culture utilized in one such experiment is shown, and the time-points sampled are indicated. **B.** Titers obtained from the experimental culture (OD_{600} ~0.6) were normalized against the same dilutions plated on exponentially growing *M. smegmatis*, and expressed as a percent of this value. All of the plotted values for wild-type and the RpfΔ represent the average of three independent experiments, except the 19 hour time-point for which there were only two values averaged. All of the plotted values for the Mt3Δ phage represent the average of two independent experiments, except the 19 hour time-point, which is from one experiment. All of the values for the Mt3Δ/RpfE-A phage are from one experiment.

In most cases, the data presented for wild-type Giles and the RpfΔ are from three separate experiments, the values for the Mt3Δ are from two independent experiments, and the plating efficiency for the double mutant is from one trial. Deviations from this are noted in the legend for Figure 37. These results would seem to indicate that similar to TM4 Motif 3, the Giles tape measure motifs are involved in the infection of stationary phase cells. Interestingly, even the wild-type phage was seen to infect with a reduced efficiency when the host cells are in stationary phase, and this is unlike what was observed in similar experiments with TM4. Importantly, however, the efficiencies calculated for the mutant phages were subtly but reproducibly lower than for wild-type Giles. When the plating culture had been growing for longer than ~40 hours, the efficiencies for the RpfΔ phage were on average ~13% lower than wild-type, the values for Mt3Δ phage were ~20% lower than wild-type, and the double mutant was ~25% reduced as compared to wild-type.

6.4.3 Mutant Phages Can Infect *M. smegmatis* with Various Cell Wall Mutations

Mutant phages were also assayed for their ability to infect strains of *M. smegmatis* with various mutations in the cell wall peptidoglycan. One of these, PM1081 (parent is PM759), contains a transposon insertion in the *dapB* gene, the product of which (dihydropicolinate reductase) is involved in the synthesis of *meso*-diaminopimelic acid (*m*-A_2pm or DAP). This residue is incorporated into the third position of the peptide stems of the mycobacterial cell wall and is involved in the formation of both (4–3) and (3–3) cross-links (discussed in Chapter 1). The *dapB* gene essential, and mutants cannot grow in the absence of DAP (Pavelka & Jacobs, 1996); however, PM1081 has an uncharacterized suppressor mutation that allows it to grow in the absence of DAP, and has thus been termed a Sud (suppressor of DAP-less death) mutant (Flores *et al.*,

2005a). The parent strains for all mutants are deleted for the major β–lactamase gene in *M. smegmatis* (*blaS*), making them somewhat susceptible to β–lactam antibiotics (Flores *et al.*, 2005b); however, the PM1081 mutant derivative is hyper-susceptible to β–lactam antibiotics in the absence of DAP and shows increased sensitivity to lysozyme (Flores *et al*., 2005a). PM979 (parent is PM975) is a *namH* mutant that does not N-glycolate the muramic acid residues in the glycan strands of the cell wall and is also hyper-susceptible to lysozyme and β–lactam antibiotics (Raymond *et al*., 2005). PM1536 (parent is PM1482) contains a deletion in the *ponA2* gene, which is believed to encode a high-molecular weight penicillin binding protein (PBP) with transpeptidase activity that may be involved in the formation of (3–3) cross-links in the cell wall (Goffin & Ghuysen, 2002). This mutant shows increased resistance to lysozyme and β–lactam antibiotics as compared to the parent and has a defect in stationary phase survival (M. Pavelka, unpublished data).

Wild-type Giles and mutant lysates were serially diluted and spotted onto freshly poured top agar lawns seeded with each of the *M. smegmatis* cell wall mutant strains and the respective parent. As was observed for wild-type *M. smegmatis*, in all cases, the plaques generated by the mutant phages were indistinguishable from those of wild-type phage (Fig. 36). These data suggest that the tape measure motifs are not required for infection of host bacteria containing certain modifications in the cell wall. It will be important, however, to assay these strains under a wider range of conditions, such as in liquid media or in stationary phase. Furthermore, there are additional strains of *M. smegmatis* with cell wall mutations that have yet to be assayed.

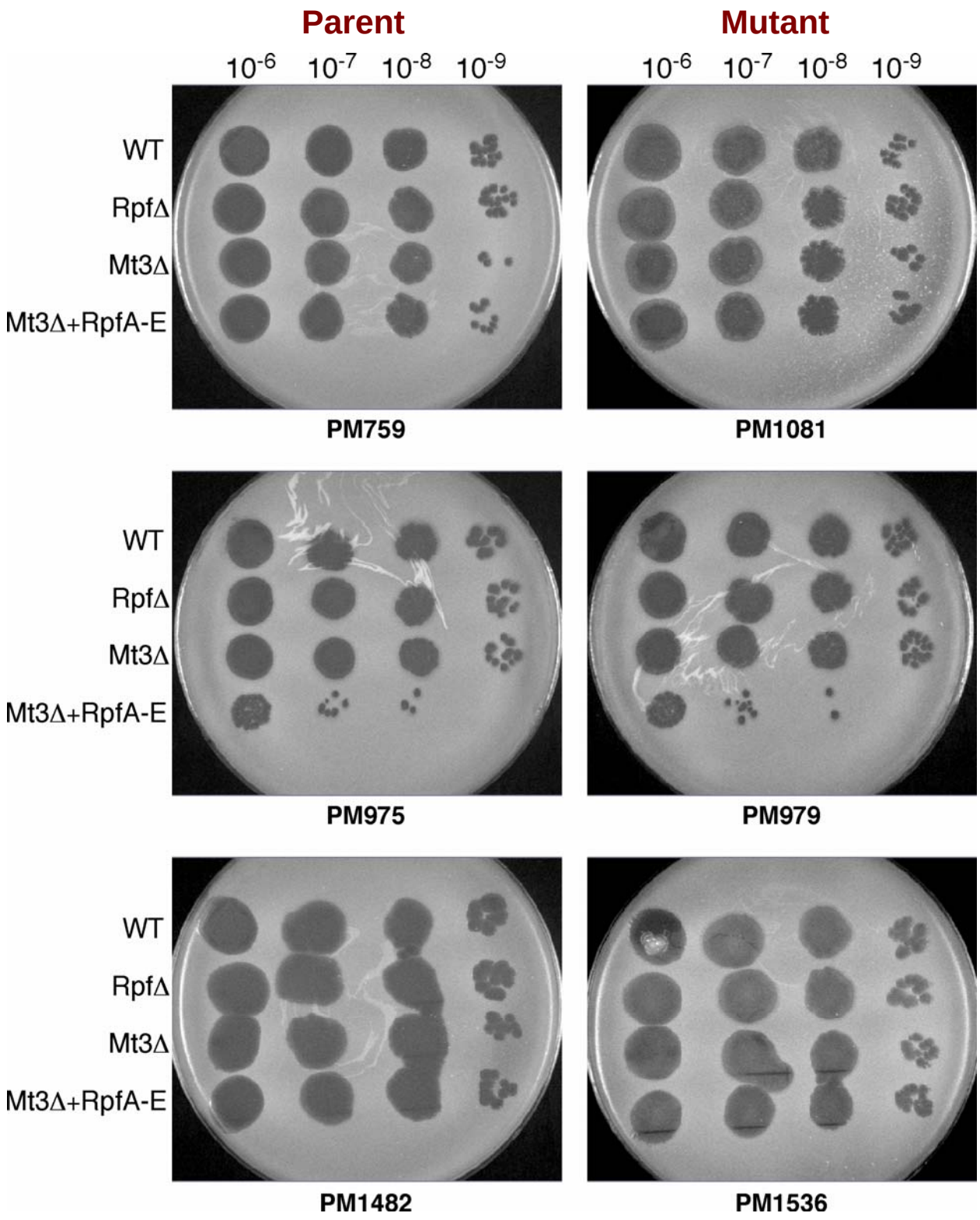


Figure 38. Giles tape measure mutants can infect M. smegmatis strains with various cell wall defects.

Figure 38: Mutant phages can infect *M. smegmatis* strains with various defects in the cell wall. A wild-type Giles lysate and lysates from the tape measure mutants were serially diluted and spottted (5 μl) onto a freshly prepared top agar lawns (0.35% MBTA with 1 mM $CaCl_2$) seeded with a panel of *M. smegmatis* stains containing mutations in various cell wall biosynthetic genes (and their respective parents). Strains PM759, PM1081, PM1482 and PM1536 are lysine auxotrophs, and therefore, all media was supplemented with L-lysine (40 μg/ml). The titer of the double-mutant lysate spotted on PM975 and PM979 was lower than the lysates from the other phages, and lower than that of the double-mutant lysate used with the other strains.

6.5 CONCLUSIONS

The phenotypes of the Giles tape measure mutants, the construction and characterization of which are described in this chapter, suggest a role for the tape measure motifs in cell wall hydrolysis during infection. This is in accordance with what has been observed previously for a mycobacteriophage TM4 phasmid-derivative that is deleted for Motif 3. Notably, as with other virion-associated murein hydrolases, the Giles motifs are not essential and show no obvious plating deficiencies under most conditions. However, similar to TM4 with a Mt3Δ, the Giles mutants showed specific defects when infecting late-stationary phase *M. smegmatis* cells from a culture >38 hours old. The observed impairment was more subtle than that which was measured for the TM4 TMP mutant, and this might be due to the fact that, unlike TM4, wild-type Giles also infects stationary cells with a reduced efficiency.

It is also likely that the two motifs work together to facilitate murein hydrolysis, and therefore, deletion of one is partially compensated for by the other. As I was unable to delete both tape measure motifs, this possibility cannot be directly addressed. However, it was observed that a double-mutant phage containing a mutation in the predicted active-site residue of the Rpf Motif (E997A; E54 in *M. luteus* Rpf) and deleted for Motif 3 was slightly more defective than phage containing a deletion of either motif alone. These data, however, must be considered in light of the observation that incorporation of this active-site mutation reduces, but does not abolish, activity of a Barnyard RLF-Rpf LysM fusion protein (containing C53 and C114) *in vivo* (Chapter 5), and the E54A mutation has an even lesser effect on a fusion containing the wild-type Rpf Motif. Importantly, this suggests that the Giles Rpf Motif with E997A likely retains some level of

activity, and therefore, the double-mutant phage cannot be likened to a true double-deletion. In this regard, it is not known why it was impossible to construct the double-motif deletion phage. It is possible that the motifs work together to perform some essential function; however, it seems equally likely that the deletion encompassing both motifs is so large (707 bp) that it renders the TMP itself non-functional. This can be addressed by using recombineering to construct identically sized deletions in other regions of the *tmp*.

Although an extensive phenotypic characterization of the Giles TMP mutants has not been carried out, it was observed that these phages are not defective in infecting *M. smegmatis* strains containing mutations in a number of cell wall biosynthetic genes. Notably, however, two of these strains (PM1081 and PM979) are predicted to have decreased levels of cross-linking and are more susceptible β–lactam antibiotics and degradation by lysozyme. This feature may, in actuality, obviate the need for the tape measure motifs, and therefore, one might predict that the mutant phages would be able to infect the mutant strains *more* effectively than the parent strains, particularly in stationary phase, and this possibility should be addressed in future studies. It will also be important to assay these and additional mutant phages on the complete panel of available mycobacterial cell wall mutants, displaying a wide range of phenotypic traits. Further, the infectivity of the tape measure mutants (towards both mutant and wild-type *M. smegmatis*) can be examined under an extensive range of environmental conditions in order to better understand the biological role of the mycobacteriophages tape measure motifs.

Importantly, the construction of the Giles mutants was made possible by the mycobacterial recombineering system, the development of which represents an important advance in mycobacterial genetics (van Kessel & Hatfull, 2007, van Kessel & Hatfull, 2008). In just a few years, this system, which has both extended and expedited the process of mutant construction in

pathogenic mycobacteria, is already coming into widespread use in the field (J. van Kessel, personal communication). In the processing of creating the Giles mutants, I have also developed a method to extend the use of mycobacterial recombineering to lytically replicating bacteriophages. As an extension of this, I have shown that this system can be utilized to create inframe unmarked deletions with a high efficiency in mycobacteriophages. The ability to create mutations of this kind is particularly useful for the manipulation of phage genomes, where strict packaging constraints make it difficult to insert large cassettes, and most selectable markers are not useful in non-lysogenic phages. Constructing mutations in this way will also minimize the likelihood that they will have polar effects on the expression of downstream genes, often a concern when making mutations in phages that tend to express their genes in large polycistronic operons. Furthermore, the ability to construct unmarked deletions is a powerful application of the λ Red recombineering system in bacteria (Ellis *et al.*, 2001, Court *et al.*, 2002), and there is evidence that it may also prove to be useful in the mycobacteria (J. van Kessel, personal communication).

The successful development of this technique was absolutely dependent on the critical observation that the transformation efficiency in both slow- and fast-growing mycobacteria is much lower than in other organisms, such as *E. coli* (van Kessel & Hatfull, 2008). A possible reason for this is the impermeability of the mycobacterial cell wall, and this feature necessitates the use of a co-selection strategy when constructing non-selectable mutants in the mycobacteria. The use of a co-selectable marker, essentially allows one to select against all cells in the population that are not competent to take up DNA and drastically reduces the amount of screening required to recover a non-selectable mutation (van Kessel & Hatfull, 2008). The poor transformation efficiency of the mycobacteria likely explains why initial attempts to utilize mycobacterial recombineering to delete the Rpf Motif from the Barnyard tape measure were unsuccessful. In

these experiments, competent cells were infected with Barnyard after the electroporation introducing the dsDNA deletion substrate. However, it is now known that only a small fraction (1/1000) of the cells infected with phage were actually competent to take up the substrate. This means that if the mutation was generated in 10-20% of these cells (as was observed for mycobacteriophage Giles), in reality 1/5000-1/10,000 plaques would be predicted to contain the mutation. This is in agreement with the results of the plaque lift assays, which detected only a few weak hybridization signals from plates containing ~10,00-20,000 plaques. Further, as the plaques that result from these recombineering experiments always represent a mixture of wild-type and mutant DNA – with the vast majority of this being wild-type – this also explains the weak signal and the difficulty in recovering the mutation by standard PCR. In light of what is known now, it should be possible to repeat those experiments with identical substrates and recover the mutation at a frequency similar to what was observed in Giles.

In the course of constructing these tape measure mutants, I have also been able to optimize the mycobacterial recombineering system for use on phages. Importantly, I have shown that 200 bp substrates, having 100 bp of homology upstream and downstream of the deletion, work better than 100 bp dsDNA substrates or 100 nt ssDNA oligos, regardless of which DNA strand they anneal to. The larger dsDNA substrates increase both the frequency with which the mutants are recovered and seemingly, the amount of mutant product present in the initial mixed-populations. There is evidence that mycobacterial recombineering mediated through dsDNA works more efficiently when longer homology lengths are employed (van Kessel & Hatfull, 2007), therefore, this may explain the difference observed between the double-stranded substrates of varying length.

Critically, both the length and the double-stranded nature of these substrates appear to be important, as in most cases examined, 100 bp dsDNA substrates work better than individual

ssDNA oligos of the same length. This seems to be in stark contrast to the *E. coli* λ Red recombineering system, where deletion mutations are routinely made using short, ssDNA oligos (Court *et al*., 2002). However, to my knowledge, the potential for using dsDNA substrates for making deletions has not been systematically investigated for this system. Interestingly, the Motif 3 deletion on the mycobacteriophage TM4 shuttle phasmid was constructed using 200 bp substrates dsDNA substrates, which were heat-denatured prior to electroporation (Piuri & Hatfull, 2006). These were utilized because attempts with 100 nt substrates were unsuccessful (M. Piuri, personal communication). It was hypothesized that the low efficiency of recombineering encountered when making mutations on large phasmids could be increased by utilizing ssDNA substrates with increased homology. However, in light of my observations, it is possible that part of the reason for the success with the larger substrate was due to the fact that some of it was double-stranded, even after the heat-denaturing. Accordingly, it was observed that the heat-denaturation step was not required for the recovery of the TM4 mutation (M. Piuri, personal communication). Further, my own attempts at recombineering in *E. coli* are in agreement with this. Although I was never able to recover the Giles cosmid containing the Rpf Motif deletion, I was able to detect the mutation in pooled DNA from the recombineering experiments, even in cases where the initial heat-denaturation was omitted. These data suggest that the efficiency of this type of recombineering, even in *E. coli,* can potentially be increased by using dsDNA substrates.

A possible reason for this can be derived from the observation made by Muyrers *et al*. that the dsDNA recombineering activities of the RecE/RecT or Exo/Beta systems are dependent on the presence of the cognate protein partners (RecE with RecT and Exo with Beta) (Muyrers *et al*., 2000). This suggests that specific protein-protein interactions may facilitate the strand annealing and strand invasion activities of the SSBs, and indeed it was observed in the same study that RecE

(but not λ Exo) can be co-purified with a tagged version of RecT. Further, in a later study by the same group, it was demonstrated that a 46 bp dsDNA substrate was ~5-fold more efficient for recombineering than a 46 nt ssDNA oligo complementary to the lagging strand of DNA replication, as assayed by the ability to either insert or delete 4 bp from a kanamycin resistance gene on a replicating plasmid in *E. coli* strains expressing either Exo/Beta or RecE/RecT (Zhang *et al*., 2003). Although a direct comparison is not possible, these data are similar to my own observations for deletion construction with mycobacterial recombineering on mycobacteriophages. Since it is known that recombineering can proceed through both ssDNA- and dsDNA-dependent mechanisms, it is possible that these contribute additive effects and thus increase the overall efficiency observed with dsDNA substrates. Intriguingly, in the same study it was noted that even when only Beta or RecT is present, recombineering is ~1.5-fold more efficient with the dsDNA substrate by the same assay (Zhang *et al*., 2003). The reason for this is not known, but it was hypothesized that in this case, the increased efficiency results from denaturation of the dsDNA substrate, providing ssDNA substrates that can anneal to either DNA strand and/or the dsDNA substrates are more stable *in vivo*. As all of my assays were performed in *M. smegmatis* strains expressing both gp60 and gp61, I cannot ascertain whether the same factors may have contributed to the increased efficiency observed with dsDNA substrates. It is of further note that Zhang *et al*. also reported that recombineering with ssDNA substrates was ~1.5-fold more efficient in the presence of the cognate RecE or Exo protein, and since the exonuclease activity is not required under these conditions, this is further suggestive of the importance of protein-protein interaction between the partners of these systems (Zhang *et al*., 2003).

Conversely, Ellis *et al.* have reported that λ Exo is completely dispensable for point mutagenesis with oligonucleotides, and its absence has no effect on the efficiency of ssDNA

recombineering, as measured by the ability to engineer a point mutation in the *galK* gene in the *E. coli* chromosome; construction of a 3.3 kb deletion in the same locus was similarly efficient with a 70 nt oligo, but deletion efficiency was not measured with dsDNA substrates or in strains expressing only Beta (Ellis *et al*., 2001). In a separate study from the same group it was observed that the efficiency of recombineering a point mutation at this locus is similar with a 70 bp dsDNA or 70 nt ssDNA substrates (Yu *et al.*, 2003). However, this study also presented data supporting the notion that the Exo protein may function to load Beta onto DNA substrates; as dsDNA recombineering with substrates having 3′ overhangs generated independently of Exo was less efficient than experiments with standard dsDNA substrates that presumably can be acted on by Exo (Yu *et al.*, 2003). This situation is further complicated by the observation that recombineering in the mycobacteria with ssDNA oligos also occurs independently of the Exo/RecE-like protein (gp60) and is highly efficient (van Kessel & Hatfull, 2008). Clearly the issue remains unresolved, and it seems likely that factors such as the type of mutation being generated, the context into which it is being constructed (i.e. the chromosome, a replicating plasmid, or a bacteriophage genome) and the recombination proteins utilized may contribute to the differences observed. Specifically, since these proteins are thought to act at the replication fork, the mode of replication may be important, and therefore, further studies should be directed at testing this possibility.

It is also intriguing, from a mechanistic perspective, why it is observed that when using substrates that insert both a deletion and a point mutation, only some phages containing the deletion mutation also contain the point mutation. Notably, a similar assay was performed by Zhang *et al*. in which they utilized 92 nt ssDNA oligos that incorporated two mutations in a mutant kanamycin resistance gene on a replicating plasmid in *E. coli*: a small (4 bp) deletion or insertion to restore gene function and a synonymous point mutation (marked by a unique restriction site).

The region of the oligo incorporating the deletion or insertion was not centrally located, and the region incorporating the point mutation was located on the longer, 59 nt 5′ arm, either 14 nt or 29 nt from the deletion/insertion. The authors were able to calculate the efficiency of recombineering by selecting for kanamycin resistance – thus the incorporation of the deletion/insertion – and surprisingly, found that the presence of the point mutation in the oligo decreased the overall efficiency of recombineering. Further, when recombinants were screened for the point mutation, only two out of 168 were positive, and these were only isolated from transformations with the oligo in which the point mutation was closer (14 nt) to the other mutation.

In my experiments with substrates inserting multiple mutations, I found that a much higher proportion of the isolated mutant plaques containing the deletion also contained the point mutation; Mt3Δ mutants from four out of seven mixed plaques also contained the AluI point mutation. The higher efficiency of incorporation might be due to the fact that dsDNA substrates were used to construct this mutation. Alternatively, an intriguing possibility is that incorporation of a second mutation is dependent on its location within the mutagenic substrate. In this case, this was located 3′ of the deletion, whereas Zhang *et al.* utilized oligos in which the point mutation was 5′ of the substrate, and this may be significant. In support of this, in the same study it was observed that ssDNA oligos incorporated a point mutation more efficiently when the substrate contained more homology 5′ of the site being mutagenized, suggesting that the 5′ end of the substrate is important for stability of the hybrid (Zhang *et al.*, 2003). The two mutations might also be expected to occur together more frequently if they are closer together on the substrate, and mutagenic substrates that differ with respect to the position of the point mutation relative to the deletion can be tested in order to address these possibilities.

It is also interesting that all Mt3Δ phages isolated from a single mixed population were identical for the presence (or absence) of the point mutation, and the reason for this is not known. Specifically, the large amount of replication that might be expected to occur in 'infected cells' might predict that the substrates could act multiple times, some of which incorporate the point mutation and some of which do not. Conversely, these results suggest that the mutagenic substrates act early during the infection; either because they are unstable and are degraded or perhaps because they can only act during early theta replication, and this possibility will require further testing.

It is also not completely clear why some of the other methods attempted in the construction of tape measure mutants were unsuccessful. In the case of the Barnyard cosmids, it is likely that recombination between complementary cosmids was not observed due to the fact that this was assayed in wild-type mycobacteria, which are known to undergo only low levels of homologous recombination. Recombination between Giles cosmids was greatly enhanced in recombineering strains of *M. smegmatis*, therefore, it is likely that the same would be true for Barnyard cosmid derivatives. Further, some complementary Giles pairs were observed to recombine at a higher frequency than others, suggesting that had other Barnyard pairs been available, it may have been possible to find ones that could recombine even in wild-type *M. smegmatis*. The reason(s) I was unable to clone the entire Barnyard genome on a BAC is harder to ascertain. It is possible that expression of toxic gene products may have contributed to this; however, it is unclear why this would be problematic for BAC construction but not cosmid construction. It is also possible that the ends of the Barnyard DNA are modified in some way that inhibits ligation, as attempts to ligate the linear genomic DNA to itself to generate concatamers were unsuccessful, and treatment with Klenow or polynucleotide kinase did not improve upon the results.

These issues, along with all of the other difficulties encountered in the search for a way of generating the tape measure motif mutants, serve to highlight the importance of having a reliable way of performing functional genomic studies in the mycobacteriophages. The recombineering system has already been utilized to make mutations on multiple phages, and more are being investigated. I have also shown that this system can be utilized as a test for gene essentiality in the mycobacteriophages by demonstrating that, similar to what was observed for the Giles double motif deletion, a deletion of the Giles *lysA* gene can be detected in primary mixed-plaques; however, these mutant phages cannot be propagated in the secondary round of infection and are undetectable even by MAMA-PCR. Further, I have shown that it is possible to utilize mycobacterial recombineering to introduce small insertions in mycobacteriophage genes by inserting a region encoding a 6x-His tag onto the C-terminus of the Giles LysB protein. This technique will permit the purification of phage proteins directly from infected mycobacterial cells, thus obviating the need for cloning and expression of a gene of interest in *E. coli* and will be of particular utility for phage proteins that are toxic or insoluble in this organism. Thus, it appears that the mycobacterial recombineering system will be widely applicable to the rapidly growing population of sequenced mycobacteriophages and should make struggles, such as those described here, a thing of the past. This mycobacteriophage population is also characterized by the fact that it is incredibly diverse, and the majority of open reading frames identified in these viruses are unique, having no recognizable amino acid similarity with other proteins in the database (Pedulla *et al*., 2003, Hatfull *et al*., 2006, Morris *et al*., 2008). Therefore, the advent of bacteriophage recombineering can also hopefully help us begin to explore this vast reservoir of genetic information in the hopes of uncovering the function and importance of the myriad of unknown genes in these organisms.

7.0 DISCUSSION

The study of bacteriophages has laid the very groundwork for our current understanding of biology and has been indispensable for the development of modern molecular biology and genetic techniques. The insights that arose directly or indirectly from these studies are too numerous to recount, but important examples include fundamental advances in the study of gene regulation, replication, homologous recombination and molecular chaperones (Murray & Gann, 2007, Weigel & Seitz, 2006). Much of this research was performed with the classic representative of this group, the *E. coli* phage λ, and the exquisitely detailed characterization of almost every aspect of this phage, from its genetic circuitry to capsid structure, have been invaluable in this regard (Gottesman, 1999). However, although it cannot be disputed that early characterizations of model phages such as λ and T4 have established the paradigm for our understanding of bacteriophage biology – and by extension, all biology – it has also become evident with the advent of the genomic and proteomic eras that we still have much to learn. Specifically, the enormous diversity present in the phage population suggests that, although studying models can teach us a lot about basic mechanisms underlying phage structure and function, comparative studies aimed at addressing how these fundamental themes are elaborated upon and varied within the phage population can also enhance and extend our understanding in valuable ways.

One area in which this is particularly true is in the study of how organisms utilize different strategies to accomplish the same general goal. All phages share a number of basic features, such

as a capsid to enclose and protect the viral genome, and these are clearly important for the survival and propagation of these entities. However, another indication of importance is when it can be observed that a function has been evolved multiple times and in multiple different ways within a diverse population. An examination of the various virion-associated cell wall hydrolases found within the phage population seems to be reflective of this. These proteins share a number of features, including the apparent ability to hydrolyze the bonds that join together the components of the cell walls of their hosts. However, they also differ in a number of striking ways, which seem to indicate a considerable amount of plasticity in the manner in which this function can be effectively performed. In most cases where these enzymes have been found, the way in which they are incorporated into the virion is somewhat different. The studies on TM4 Motif 3 and its bacterial homologues as well as the data presented in this investigation suggest that the mycobacteriophage tape measure motifs, including the Rpf Motif, represent yet another variation on this theme.

7.1 EVIDENCE TO SUPPORT A ROLE FOR THE TAPE MEASURE IN PHAGE DNA INJECTION

Before discussing the various ways in which phages have incorporated muralytic proteins into their virion structures, it is necessary to consider tape measure proteins and why these proteins might be ideally suited to such a purpose. TMPs are found in virtually all tailed phages, and their essential roles in tail assembly and length determination have been most clearly demonstrated through a number of elegant studies in bacteriophage λ (Katsura, 1990). This work has shaped our understanding of these proteins; however, another function for the tape measure suggested from many of these studies is often overlooked. Specifically, it was noted that tape measure mutants that

retained the ability to initiate tail formation and to specify a defined (albeit altered) tail length had specific defects in DNA injection; some of these retained viability but others, with larger tape measure deletions were non-infectious (Katsura & Hendrix, 1984, Katsura, 1987). Further, evidence directly supporting a role for the λ TMP in DNA injection can be gleaned from the observation that certain mutations in an *E. coli* inner membrane protein (*pel* mutants) that are no longer susceptible to λ can be suppressed by mutations that lie within the *tmp* gene (Scandella & Arber, 1976). These data suggest that rather than acting merely as an inert scaffold for tail assembly that must be removed before the exciting process of host infection can really begin, tape measures may actually play an active role in this process, perhaps by inserting into the bacterial membrane and facilitating pore formation. Indeed, λ gpH is predicted to have multiple transmembrane domains, and there is evidence that this protein can insert into the lipid bilayer following irreversible adsorption (Roessner & Ihler, 1984).

For the coliphage T5, it has been definitively shown that a similar function is performed by the tail fiber protein pb2, which inserts into the bacterial membrane during infection (Guihard *et al*., 1992) and displays the same activity *in vitro* (Feucht *et al*., 1990, Lambert *et al.*, 1998). Significantly, this protein shares a number of features with tape measure proteins, including a high predicted α-helical content, multiple putative coiled coil and transmembrane domains, similar amino acid content and a peptidase domain with muralytic activity. These features have led to the proposal that this protein acts as both the tail fiber and the tape measure for T5, although this remains to be shown. It was further demonstrated that a muralytic fragment of pb2 containing the peptidase motif as well as both a putative coiled coil and a transmembrane domain was able to oligomerize and interact with the membrane *in vivo,* and further, this portion of the protein can insert into liposomes and displays fusogenic activity *in vitro* (Boulanger *et al*., 2008). Strikingly,

this ability to insert into and mediate the fusion of two lipid bilayers is a feature that has been noted for glycine/alanine-rich peptides within the envelope glycoproteins synthesized by eukaryotic viruses (Del Angel *et al*., 2002). This is also an activity that one can imagine might be useful to phages of Gram-negative bacteria that must contend with the double-membrane and periplasm of their hosts. Thus, it has been hypothesized that these phages must have a way to bridge the two membranes during pore formation in order to shield the viral DNA from destructive nucleases in the periplasmic space (Molineux, 2006), and in T5, this function is performed by the tape measure-like pb2 protein (Boulanger *et al*., 2008). These observations combined with the many similarities between pb2 and other phage tape measures support the hypothesis that all of these proteins might perform similar functions during host infection.

Intriguingly, the mycobacteriophage tape measures also share many predicted structural features with gpH and pb2, as well as with the T7 internal capsid protein that contributes to channel formation, gp16 (discussed below). Further, although the mycobacteria are considered to be Gram-positive, their cell walls have features in common with both Gram-positive and Gram-negative bacteria, and importantly, these organisms have long been noted for their impermeability, hydrophobicity (i.e. they are more permeable to lipophilic compounds) and the pseudo-lipid bilayer-like character of their of cell envelope (Minnikin, 1982, Brennan & Nikaido, 1995). A recent ultra-structural investigation of the mycobacterial cell wall from both intact and thin-sectioned cells suggests that the outer layer of this structure is indeed a bilayer, 8 nm thick, which responds to detergent treatment in a way that is consistent with this having a predominantly lipidic-composition (Hoffmann *et al*., 2008). Notably, upon exposure to detergent, regions can be observed where the outer layer is disrupted, and below these sections, the inner membrane is also dissolved. The thickness and appearance of the individual leaflets of this structure are very similar,

which suggests an unanticipated symmetry; however, the specific composition of the outer bilayer and how the mycolic acids might be incorporated are unknown. Despite this, the lipophilic nature and bilayer organization of the outermost layer of mycobacterial cells suggest that similar to T5 pb2, the hydrophobic regions of mycobacteriophage tape measures may be directed into this outer membrane–like structure by glycine/alanine-rich regions of the protein during the initial stages of infection. This would serve to fuse this outer layer with the inner membrane and facilitate the formation of the pore for DNA translocation into the bacterial cell. Indeed, if a critical role in DNA injection was common to all TMPs, than it is perhaps not surprising that in performing this important function, the tape measures of a large portion of the mycobacteriophages have also evolved the ability to catalyze peptidoglycan hydrolysis. During this process of channel formation, it is likely that localized hydrolysis of covalent bonds within the cell wall might sometimes be necessary or advantageous, and this would explain the acquisition by the phage tape measures of various sequence motifs with hydrolytic activity.

Although it has not been definitely demonstrated, my experimental observations support the hypothesis that TMPs can insert into the bacterial membrane. Specifically, I have observed that although the Barnyard TMP can be expressed to low levels in the mycobacteria, it also displays the toxicity in bacteria common to these proteins. Further, expression of the 70 kDa C-terminal fragment of this protein appears to alter the bacterial cell surface. Thus, it is tempting to speculate that these features are due to the propensity of the tape measure to insert into the bacterial membrane. Accordingly, most of these TMPs examined, including the proteins from λ and the lactococcal phages discussed previously, are predicted to contain hydrophobic membrane-spanning domains, and the same is true for the Giles and Barnyard tape measures. Extending this analysis (TMpred) to all sequenced mycobacteriophages, the same pattern is observed, and we see that all

are predicted to have multiple transmembrane domains (Table 3). While many of these numbers are likely to be overestimates, the ubiquitous nature of this phenomenon suggests that it is indicative of a function that is common to all tape measures. Cell localization data indeed suggest that the 70 kDa C-terminal fragment of the Barnyard tape measure that contains the Rpf Motif and Motif 2 is associated with the cell membrane and/or cell wall during infection of *M. smegmatis*. These experiments were complicated by a high background and lack of availability of adequate controls, and therefore, further proof will be required to affirm this for Barnyard as well as to determine if this is indeed a feature common to these proteins.

In addition to the structural features described above, a number of additional parallels can be drawn between the tape measures investigated in this study and the TMPs that have been characterized from other phages. In this regard, I have shown that similar to λ gpH, the lactococcal phage TMPs and the TMP from the *Listeria* phage PSA, the Barnyard TMP gp33 is proteolytically processed, and this is at least dependent on the presence of one or more phage proteins. Cleavage does not occur when the TMP is expressed independently in *M. smegmatis,* implying that proteolytic processing of the tape measure is likely to occur prior to or concurrent with virion assembly. Based on these data, an autocatalytic mechanism for this cleavage cannot be ruled out; however, this activity is at least dependent on additional phage and/or host proteins, or possibly by stimulation resulting from a precise conformation adopted by the protein within the tail. Unlike gpH however, both portions of the cleaved protein can be detected in mature phages. Conversely, although the cleavage of the lactococcal phage tape measures has not been investigated in detail, the available data *suggest* that in these phages, only the larger cleavage product remains in the virion. Mass spectrometric analysis further suggests that the Giles TMP may also be cleaved; however, the nature of this processing has not been determined. Notably, proteolytic processing is

a feature that has been found for all phage tape measures that have been investigated to date, which suggests that this might represent an essential step in phage tail assembly. Another possibility is that cleavage – which for λ is known to occur late during tail assembly – primes or activates the tape measure such that it becomes competent to fulfill its role in DNA injection.

The ability to generate mutations with a high efficiency on the bacteriophage genome will permit future studies aimed at elucidating the importance of this processing event (e.g. by mutating specific residues within the cleavage site). It is also not known what roles the TMP fragments generated by this cleavage play in infection. An attractive possibility is that they are directed to the bacterial cell membrane via their hydrophobic domains and assemble to form the pore complex for DNA entry. Perhaps the most striking feature, however, that distinguishes the mycobacteriophage tape measures is the presence of the sequence motifs embedded within them that appear to have muralytic activity. Further, the presence of these motifs on the smaller cleaved fragment of the tape measure, which is structurally distinct from the N-terminal portion, is perhaps further indicative of a specialized role for this portion of the protein during infection.

7.2 THE IMPORTANCE OF VIRION-ASSOCIATED HYDROLASES

Another critical difference that is evident from a comparison of various tape measures is that some of these, including the λ protein, lack the murein hydrolase sequence motifs found in a subset of mycobacteriophage TMPs. This suggests a number of possibilities; on one hand, it is possible that the activity performed by the motifs is mediated by a different structural protein. This certainly appears to be the case for the lactococcal phages, and this is discussed in further detail below. It is also possible that there are, in fact, regions of these tape measures with hydrolytic activity;

however, they do not have similarity to known cell wall hydrolases and therefore, are not detectable by sequence homology. In this regard, it may be noteworthy that the Cluster A mycobacteriophage TMPs that lack any of the five sequence motifs do possess regions with amino acid similarity to the tape measure of the lysogenic *Streptococcus mitis* phage SM1, a protein also known as PblA. The SM1 TMP is unique in that it was characterized (along with the tail fiber protein, PblB) as a bacterial *cell-wall associated* virulence factor, which promotes binding of *S. mitis* to human platelets (Bensing *et al.*, 2001, Siboo *et al.*, 2003). In these studies it was also observed to exist in a proteolytically processed form, both in the phage tail and in the cell wall. While the mycobacteriophage tape measures are unlikely to act as virulence factors, this at least suggests that they may contain regions that direct localization to the bacterial cell wall. Finally, a third possibility is that phages lacking the sequence motifs simply do not encode for proteins that performs their function. As the non-essentiality of these motifs has been demonstrated both in this study and for TM4, this cannot be excluded.

As I alluded to above, the TMPs of the sequenced lactococcal phages do not appear to contain hydrolase motifs such as those found in the mycobacteriophages. However, Tuc2009 represents an example of a phage in which this function seems to have been acquired by another protein, the muralytic tail lysin Tal_{2009} (Kenny *et al*., 2004). Similar to the TMP, this is a large phage structural protein that is found in the phage tail. The catalytic activity has been localized to a specific region of the C-terminus, suggesting that other parts of the protein may perform additional function(s). Accordingly, it has been observed that antibodies against the C-terminal region of Tal_{2009} bind to the tail tip and inhibit phage infection. The homologous protein in the related phage TP901-1 has been shown to be required for tail formation and is predicted to form the phage tail fiber (Vegge *et al*., 2005), suggesting that Tal_{2009} may have the same function.

Interestingly, there is evidence that both the Tuc2009 tape measure (discussed in Section 4.1) and the Tal_{2009} protein are proteolytically processed, although for the tail lysin, this phenomenon is somewhat puzzling. Cleavage occurs by an autocatalytic mechanism at a Gly rich region of the protein and is abolished by mutating residues in this cleavage site (Kenny *et al*., 2004). This processing can be observed both when the protein is expressed in *L. lactis* and *E. coli,* as well as during infection; however, only *full-length* protein and possibly the *non-catalytic* N-terminal portion are detected in mature virions. Thus, the cleavage observed during infection may function to release catalytically active species to assist in host cell lysis, and/or the N-terminal fragment may play an important structural role in the in the virion These data also suggest that that upon infection, the full-length Tal_{2009} may be released from a constrained position, permitting it to undergo self-cleavage and releasing the C-terminal portion to locally degrade the cell wall. This is supported by the observation that the cleavage product displays a higher activity in zymograms than the full-length protein, which has only weak activity.

In general, the observation of cell wall hydrolytic activities associated with two distinct tail proteins in two different groups of phages highlights the likely importance of such activities. It also raises the question of how such configurations may have arisen. While the precise mechanism is open to debate, it may be noteworthy that for the lactococcal phages, TP901-1 and Tuc2009, there are data suggesting that analogously to λ, both the TMP and the tail fiber protein (the tail lysin) are components of the initiator complex for tail formation. Located in between theses genes is one additional gene that encodes for a cone-like structure at the tail tip. This arrangement is somewhat colinear with the λ tail-initiator genes (Mc Grath *et al*., 2006); however, in λ, there are multiple genes (encoding for multiple components of the initiator, M-I) in between the tape measure and the tail fiber gene. One of these encodes gpK, a small tail protein, which weakly

resembles proteins with the ability to hydrolyze peptidoglycan (Moak & Molineux, 2004). This suggests that the lambda-like coliphages may have evolved yet another strategy to mediate host cell wall degradation during infection. It is also tempting to speculate, however, that the hydrolytic activity of the lactococcal tail fibers proteins was acquired by an ancient recombination event between the tail fiber gene and a gpK like-gene in a lactococcal phage ancestor. Taking this one step further, perhaps similar recombination events within this region could have given rise to a lineage of phages where the gpK-like protein became incorporated into the phage tape measure, thus giving rise to the hydrolase motifs we see today. In this regard, it may be significant that proteins with predicted peptidase activity are found in the tail gene regions of a number of mycobacteriophages, including but not excluded to Bxb1, Che8, Corndog, and Omega; some of which have a recognizable tape measure motif and some of which do not. Although, what this means, if anything, for the evolution of the TMP motifs remains to be discerned.

The phages considered so far in this chapter have been members of the siphoviridae; however, the coliphage T4 represents an example of how a contractile-tailed phage has utilized yet another strategy to incorporate a muralytic protein into its virion structure. Once again, despite the differences, we can see some of the same general themes emerging. The T4 hydrolase gp5 is both essential for baseplate – and therefore, tail – formation and is also a lysozyme; thus, this protein also plays dual roles (Kao & McClain, 1980, Nakagawa *et al*., 1985). Additionally, gp5 also requires proteolytic cleavage for full-activation (Kanamaru *et al*., 2005). Structurally, the active site of the lysozyme is blocked by a loop on the C-terminus of the protein, and mutations that abolish cleavage result in virions with decreased levels of infectivity (Kanamaru *et al*., 2005). In this case, both fragments of gp5 remain in the virion, and have different, important functions (Kanamaru *et al.*, 1999). The N-terminal portion encompasses the catalytically active residues and

may also be important for protein-protein interactions with other members of the baseplate and tail (Leiman *et al*., 2004). The C-terminal fragment forms a hollow, needle-like structure that appears to release the lysozyme to its site of action by piercing the outer membrane during DNA injection, and this may also form the channel for DNA translocation (Kanamaru *et al*., 2002). Thus, again we see that the virion hydrolase is incorporated into a tail structure protein, which requires cleavage for activation and may contribute to the formation of a DNA conduit into the bacterial cell. A search of the database reveals gp5-like proteins in many of the related T-even and T-even-like phages, which might suggest that this mechanism represents common strategy employed by the myoviridae. However, the mycobacteriophages with contractile tails, such as Bxz1 and Catera, encode for TMPs containing Motif 3-like regions, and these motifs have identity to tape measures from phages that infect other organisms. This implies therefore, that as was observed for the siphoviridae, there are multiple ways by which morphologically similar phages might mediate the same function.

In the previous examples, the virion hydrolases were seen to be associated with some part of the phage tail. However, PRD1 and ϕ6 represent two examples of how tail-less phages have incorporated muralytic proteins into their virion structures. Although PRD1 doesn't have what is classically thought of as a tail, it uses the viral membrane within the capsid to form a membranous, extensible tail. In the current model, a viral protein (P11) permeabilizes the OM, which then allows a suite of viral membrane proteins to form the 'tail-tube' believed to be required for DNA injection (Grahn *et al*., 2002). Following the paradigm that has been observed in other phages, the viral transglycosylase, P7, is important, but not required for this process, and P7 mutants show delayed DNA injection (Rydman & Bamford, 2000). Similarly, the lytic enzyme P5 of dsRNA *Pseudomonas* phage ϕ6 is located between the nucleocapsid and surrounding membrane (Mindich

& Lehman, 1979, Caldentey & Bamford, 1992). Following an initial fusion of the viral and host OM, P5 is believed to locally digest the cell wall in order to facilitate a second fusion event between the encapsulated phage and the cytoplasmic membrane (Bamford *et al.*, 1987, Romantschuk *et al.*, 1988). Further, a similar phenomenon has been noted for the related dsRNA phage ϕ13 (Daugelavicius *et al.*, 2005). These provide examples of how, in the absence of tails, phages have incorporated proteins required for DNA injection – including those with muralytic activity – into their capsid structures.

Continuing on the same theme, an interesting observation is that in many ways the virion-associated hydrolase that is most comparable to the mycobacteriophage TMPs is the T7 protein, gp16, a similarity that has been noted previously (Struthers-Schlinke *et al*., 2000). This might not be expected at first due to the fact that most of the mycobacteriophages have long, flexible tails, and T7, as a member of the podoviridae, barely has a tail. However, it has been shown that T7 does possess what has been described as an 'extensible tail'. That is, during the initiation of infection, three proteins that are contained within the capsid, exit through the head-tail connector and insert into the bacterial membrane. It is believed that these proteins form a channel that bridges the inner and outer membranes and forms a conduit for the viral DNA entering the host cell (Molineux, 2001). The muralytic activity is contained with a discrete region of the large, multifunctional protein gp16, which has also been shown to be important for phage structure and assembly, as well as DNA injection (Moak & Molineux, 2000, Struthers-Schlinke *et al.*, 2000, Kemp *et al.*, 2004). Similar to what has been observed for TMPs, gp16 is largely α-helical (composed of 50% short α-helices and 30% random coils), is predicted to contain one or two transmembrane helices, and is somewhat toxic when expressed in *E. coli* (Molineux, 2001). Further, because it has to exit the narrow (2.2 nm) head-tail connector, at some point this protein must unfold and exit the capsid as a

long linear molecule, which is reminiscent of the conformation of the tape measure within and as it exits the tail (Moak & Molineux, 2000, Molineux, 2006).

Despite these striking similarities, gp16 clearly has features that distinguish it from TMPs. For gp16, muralytic activity could be demonstrated *in vitro*, and mutations that abolish this were observed to render the phage defective in infecting bacteria under conditions where the peptidoglycan is predicted to be more highly cross-linked (Moak & Molineux, 2000, Moak & Molineux, 2004). Similar *in vitro* activity could not be demonstrated for the mycobacteriophage tape measures, although this may have been due to the lack of the correct substrate in the zymogram gels. Additionally, mutations in the predicted active site residues did not appear to completely abolish function of either the TM4 Motif 3 (Piuri & Hatfull, 2006) or the Rpf Motif. Finally, the conditions where the tape measure motifs were shown to be important were somewhat different than those observed for the gp16 transglycosylase. Specifically mycobacteriophage TM4 tape measure mutants did not display the same cold-sensitive phenotype observed for the T7 gp16 mutants (M. Piuri, personal communication). Finally, gp16 has been proposed to form part a motor complex that directs the transcription-independent T7 DNA injection (Kemp *et al.*, 2004, Kemp *et al.*, 2005). The mechanism of mycobacteriophage DNA injection is not known; however, a similar role for the tape measure seems unlikely.

7.3 FUNCTION OF THE MYCOBACTERIOPHAGE TAPE MEASURE MOTIFS

This leads us back to critical question regarding the activity and importance of the phage tape measure motifs, specifically the Rpf Motif, which is the subject of this investigation. A survey of the hydrolytic proteins found in many different phages and in a variety of distinct but similar

structural proteins, supports a role for these motifs in host cell wall hydrolysis. This is a function that was not initially apparent, as the host Rpf proteins were characterized as bacterial growth factors. However, the structural similarity of these proteins with lytic transglycosylase enzymes, as well as their enzymatic activity observed both *in vivo* and *in vitro* suggest that the Rpf proteins are indeed involved in limited murein hydrolysis during growth and cell wall remodeling and further implies a similar function for the phage Rpfs. Accordingly, I have shown that recombinant proteins in which the Rpf domain from the *M. luteus* Rpf is replaced by either of the mycobacteriophage Rpf Motifs have the ability to lyse *E. coli* when expressed into the periplasm. This demonstrates that the phage motifs have the potential for catalytic activity within the TMP, and further, the conserved cysteine residues previously shown to be required for Rpf activity are dispensable for the phage Rpfs, even when they are expressed as soluble fusions with the Rpf LysM domain. Interestingly, engineering these residues back into the fusions resulted in proteins with increased levels of activity, and some of these appear to be toxic when inducibly expressed as secreted proteins in *M. smegmatis*, at least on solid media. This suggests that although the Rpf Motifs have catalytic activity (as demonstrated in *E. coli*), the addition of one or both cysteine residues increases this activity through some yet undeciphered mechanism (discussed in Chapter 4).

These observations imply multiple distinct possibilities regarding the level of hydrolytic activity of the Rpf Motifs within the tape measure. On the one hand, it appears that residues within the phage motifs can compensate completely for the missing cysteines, resulting in fusion proteins that are phenotypically comparable to wild-type Rpf (Fig. 28). This would suggest a low inherent enzymatic activity for the phage motifs, as the *M. luteus* Rpf was observed to have limited or weak activity in most assays utilized (Mukamolova *et al*., 2006). This apparent 'weakness', however, might actually be important due to the fact that lytic enzymes functioning during the infection

process need to do minimal damage to the host cell. That is, the bacteria must remain intact and metabolically active in order to produce progeny phage, and therefore, it may be advantageous for these hydrolases to have limited activity or to be constrained in some way, such as by being contained within a large structural protein like gp16, the Tuc2009 tail fiber, or the tape measure (Moak & Molineux, 2004, Kenny *et al.*, 2004, Molineux, 2006). However, this also does not preclude that possibility that additional residues outside of this motif may stimulate the activity of the phage Rpfs. The observation that the addition of one or more cysteine residues enhances the lytic activity of the Rpf Motif *in vivo* implies that it has the *potential* to display greater levels of muralytic activity in the tail than that which have been observed for the bacterial Rpfs. It is possible, even likely, that this activity is in some way mediated by the specific conformation of the protein within the phage tail, thus explaining the lack of activity of the tape measure fragments observed in most assays.

The lack of structural data regarding the conformation of the tape measure within the tail and importantly during host infection makes it difficult to predict specifically how this might occur. However, it seems likely that the muralytic activity of the Rpf Motifs is stimulated by, or mediated in some way through, an interaction with the additional motifs present within the tape measures of both Barnyard and Giles. In this regard, it should be noted that a number of lytic transglycosylases, including Slt70, do not cause cell lysis or are unstable when overexpressed in bacteria (Holtje, 1998). These data, combined with the genetic and biochemical interactions that have been detected between multiple cell wall modifying enzymes, both in *E. coli* and in Gram-positive organisms such as *S. aureus* and *B. subtilis,* have led to the proposal that these proteins are likely to function in multi-enzyme complexes (Holtje, 1998, Vollmer & Bertsche, 2007). This further implies that their activity is likely to be modulated through interactions with other proteins;

Slt70 in particular shows increased stability in the presence of the endopeptidase BPB7 (Romeis & Holtje, 1994). Accordingly, there are now data suggesting that the mycobacterial Rpfs are likely to work in conjunction with additional cell wall hydrolases, specifically the putative endopeptidase RipA (Hett *et al*., 2007). Furthermore, Rpf proteins have been identified in *Streptomyces,* which contain putative endopeptidase domains.

The hypothesis that Rpf proteins must interact with other cell wall remodelers in order to exert their full activity is supported by the fact that the Rpf Motif has only ever been found in together with one of the other tape measure motifs (Motif 2 or Motif 3). In addition, the TM4 Motif 3 was shown to have muralytic activity *in vitro*, and a number of bacterial proteins with similarity to Motif 3 appear to function as peptidoglycan peptidases (Piuri & Hatfull, 2006). This also seems to be substantiated by the observation that the Giles tape measure mutant containing a mutation in the putative active site residue of the Rpf Motif combined with the Mt3Δ, is slightly more defective than mutant derivatives containing only the deletion or the point mutation separately. I was unable to construct the double motif deletion; therefore, it is tempting to speculate that these motifs work together to perform some essential function. However, the non-essentiality of the virion-hydrolase activity in all other phages examined argues against this. It is more likely that deleting both motifs – potentially due to the size of the region removed – renders the tape measure defective for some other essential function such as tail assembly or possibly, DNA injection, as has been observed for large deletion of λ gpH.

Importantly, the phenotypes of mutant phages deleted for either of the tape measure motifs supports a role for these in host cell wall hydrolysis during infection. However, the enzymatic nature of this activity remains to be determined. As mentioned above, the bioinformatic data suggest that Motif 3 is likely to act as a peptidoglycan peptidase, and such, mediates cleavage of

the bonds linking the peptides stems within the cell wall. The Rpf proteins, however, have characteristics of lytic transglycosylases, and therefore, it seems that this motif functions as a muramidase and cleaves the β-1,4-linkages between MurGlc (the glycolated variant of MurNAc found in the mycobacteria) and GlcNAc residues in the glycan backbone. It is more difficult to assign a function for Motif 2, although there is weak bioinformatic data suggesting that this too may have some sort of muralytic activity. Additional work will be required do determine how these motifs might work together to mediate peptidoglycan hydrolysis in the host cell wall.

It should also be noted here, that in light of the phenotypes observed for the mutant phages and considering the observation that the growth-stimulating activity of Rpf appears to be mediated by its enzymatic murein hydrolase activity, I also cannot rule out a possible growth stimulatory role for these motifs, particularly for the Rpf Motif. This activity might only even be a 'side-effect' of the hydrolytic function of the motif, but the argument can be made that such an activity would be extremely useful for viruses attempting to infect non-growing hosts, which are likely represent the majority of bacterial cells in the environment. This suggests that such an activity might be selected for in the phage population, and may explain why some phage TMPs have the Rpf Motif in addition to one of the other cell wall hydrolase motifs. This hypothesis has proved more difficult to address due to the fact that adequate assays for dormancy of resuscitation of the mycobacteria, particularly for *M. smegmatis*, are not readily available; however, it cannot be overlooked and should be considered in future studies.

Regardless of this possibility, the muralytic activity and non-essential nature of these motifs (and other virion hydrolases) implies that they are important for infection only under specific cell wall conditions. Further, the deficiency observed for the mutant phages specifically towards hosts in late-stationary phase suggests we could gain a better understanding of their

function by characterizing the changes that occur to the mycobacterial peptidoglycan network as these bacteria transition into, and are maintained in stationary phase. This process has been well-studied in *E. coli*, where it has been shown that the overall levels of cross-linking as well as the occurrence of (3–3) cross-links between *m*-A_2pm residues increases as these bacteria transition from exponential to stationary phase (Pisabarro *et al*., 1985). It has been hypothesized, therefore, and in some cases demonstrated, that phage hydrolases such as T7 gp16 become important as active cell wall remodeling declines and levels of cross-linking increase. This is believed to decrease the average pore size in the peptidoglycan network, thus making it more difficult for phages – and other complexes that mediate molecular transport – to associate with and assemble on the cell surface (Dijkstra & Keck, 1996). Accordingly, there are proteins with predicted murein hydrolase activity associated with a number of complexes that function in macromolecular transport or the assembly of extracellular structures that span the cell wall. These include the conjugative transfer apparatus, flagellar assembly, and the Type IV secretion complex (Koonin & Rudd, 1994, Bayer *et al*., 1995, Mushegian *et al*., 1996, Nambu *et al*., 1999). In most cases, the muralytic activity has been shown to increase efficiency of transfer but to be non-essential for the general functioning of the system, similar to what has been observed for the phage virion hydrolases.

Unfortunately, the changes that might occur in the mycobacterial cell wall during the transition to stationary phase have not been extensively analyzed. There are reports, however, that describe an apparent cell wall thickening in these populations (Cunningham & Spreadbury, 1998) and preliminary evidence suggests that similar to *E. coli*, there is an increase in the occurrence (3–3) cross-links (M. Pavelka, personal communication). Therefore, although it is difficult to definitively assert which hydrolytic activities might be most useful for phages attempting to infect

non-growing mycobacterial hosts, enzymes that cleave the peptide cross-bridges or disrupt a (potentially) thickened glycan network are good candidates. Despite the fact that the dynamic nature of the cell wall is likely to render such proteins nonessential, these may confer a selective advantage to the phages that carry them, allowing these viruses to infect hosts whose physiological conditions might render them otherwise inaccessible.

In accordance with this, BLAST analyses with any of the tape measure motifs identify similar putative murein hydrolase activities in a number of other phage tail proteins, including ones from phages that infect *G. terrae, S. aureus, Lactobacillus spp., Bacillus megaterium* and *M. vanbaalenii* (Piuri, 2006 and my own observations). Further, these are often predicted to lie within the phage tape measure, suggesting that the mycobacteria are not the only ones to have exploited this protein for such a purpose. Although for most of these motifs, a definitive function has not been demonstrated, they certainly appear to be ubiquitous in the phage world.

7.4 CONCLUDING REMARKS: WHY DO PHAGE HAVE TAILS?

Returning once again to the variant nature of the proteins with which we see these murein hydrolases associated, we can begin to think about the broader question of phage morphology. Specifically, the seemingly simple question, why do phages have tails? It is known that these structures are not absolutely required, as there are numerous phages with short stubby tails (the podoviridae), such as T7 and P22, as well as others that lack a tail structure altogether, such as PRD1 and ϕ6. Importantly, examination of these phages reveals that in each case, they possess a protein with murein hydrolase activity that is associated with the capsid. Although, by and large, this hydrolytic activity is non-essential, the process with which it is associated, namely viral DNA

injection, is absolutely essential. Further, in every case examined, tail-less and short-tailed phages have a unique, capsid-associated mechanism for DNA injection. A clear example is the trio of internal head proteins in T7 that are ejected upon irreversible binding to host receptor and insert into the membrane to form the DNA injection apparatus.

This leads to the ultimate conclusion that all phages must carry some apparatus that facilitates genome entry into the host. Further, if these proteins are not associated with, or contained within the head, then they must be in, or associated with, the tail. For the siphoviridae, I propose that the protein that performs this function is the tape measure. It has long been known that these proteins specify tail length, but why make a tail in the first place? Tape measures are large and expensive proteins to synthesize if their function is merely to assist in the assembly of a tail that some phages are happy to do without. Further, even assuming a tail is required (as it clearly is for some phages), there are other mechanisms by which its length can be specified that do not require the synthesis of multiple copies an enormous protein and its associated chaperones, some of which were proposed by Kellenberger (Kellenberger, 1972). However, this expenditure of energy is justified if the fundamental function of the tape measure is to facilitate the critical process of DNA injection into the host. Ultimately, rather than viewing the tape measure as something that exists because of the tail, perhaps quite the opposite is true and in fact, the tail exists merely as a vehicle for its ‘tape measure’. By synthesizing a structure that is exactly the right size for this protein to be contained in an extended conformation, the phage, like an armed invader, is now equipped with a means by which to scale the ‘walls’ and mount a successful attack.

APPENDIX

GENERAL CHARACTERIZATION OF MYCOBACTERIOPHAGES BARNYARD AND GILES

A.1 MYCOBACTERIOPHAGE BARNYARD

A separate aspect of this project has involved a general characterization of mycobacteriophage Barnyard, which was observed to possess a number of interesting features that distinguish it from other mycobacteriophages. For one, during the isolation and initial characterization of this phage, it was noticed that Barnyard forms very small, seemingly turbid plaques, and only appeared to be able do so on exponentially growing cultures of *M. smegmatis* (M. Pedulla, unpublished data). It was postulated that perhaps this dependence on the growth state of the host was somehow mediated by the Rpf Motif within the TMP, although a proclivity towards infecting exponentially growing versus stationary cells was somewhat opposite of what might be expected for a phage that contained a motif thought to mediate resuscitation of dormant cells.

Of further interest was the observation made upon the complete sequencing of the Barnyard genome, that this phage – at the time at least – was quite unique; 89.9% of its predicted open reading frames (ORFs) did not match any other mycobacteriophage proteins, and 91.8% did not have detectable similarity to any other proteins in the database (Pedulla *et al*., 2003). Barnyard was

thus an excellent illustration of the enormous extent of sequence diversity that was likely to be present within this population and the phage population as a whole. This also, however, made it exceedingly difficult to assign putative functions to individual Barnyard proteins, and even genes encoding for phage structural proteins such as the major tail protein or capsid, were impossible detect. Therefore, I have conducted a number of experiments aimed at characterizing and better understanding some of the features of this distinct phage.

Cell Density Dependence

In order to investigate the apparent growth-phase dependent phenotype of mycobacteriophage Barnyard, I spotted serial dilutions of this phage on lawns seeded with *M. smegmatis* cultures of varying ages and cell densities. Representative plates from these experiments are shown in Figure 39. Importantly, regardless of the age of the culture, when the lawns were seeded with a greater number of cells (generally more than 500 μl of culture at an OD_{600} of 0.2), Barnyard was unable to form visible plaques. Conversely, when stationary phase and exponential cultures were diluted to an identical optical density and utilized to seed top agar lawns, the plaque morphology of Barnyard spotted on either plate was comparable. If anything, the plaques were observed to be slightly less turbid on lawns generated from stationary phase cells, although this is difficult to quantify. These data strongly suggest that, although Barnyard is indeed extremely sensitive to cell density and will only form visible plaques on very thin bacterial lawns (and infect cells in liquid culture at a low starting cell density), this phage will form plaques on lawns seeded with both exponentially growing and stationary phase *M. smegmatis* to a similar degree.

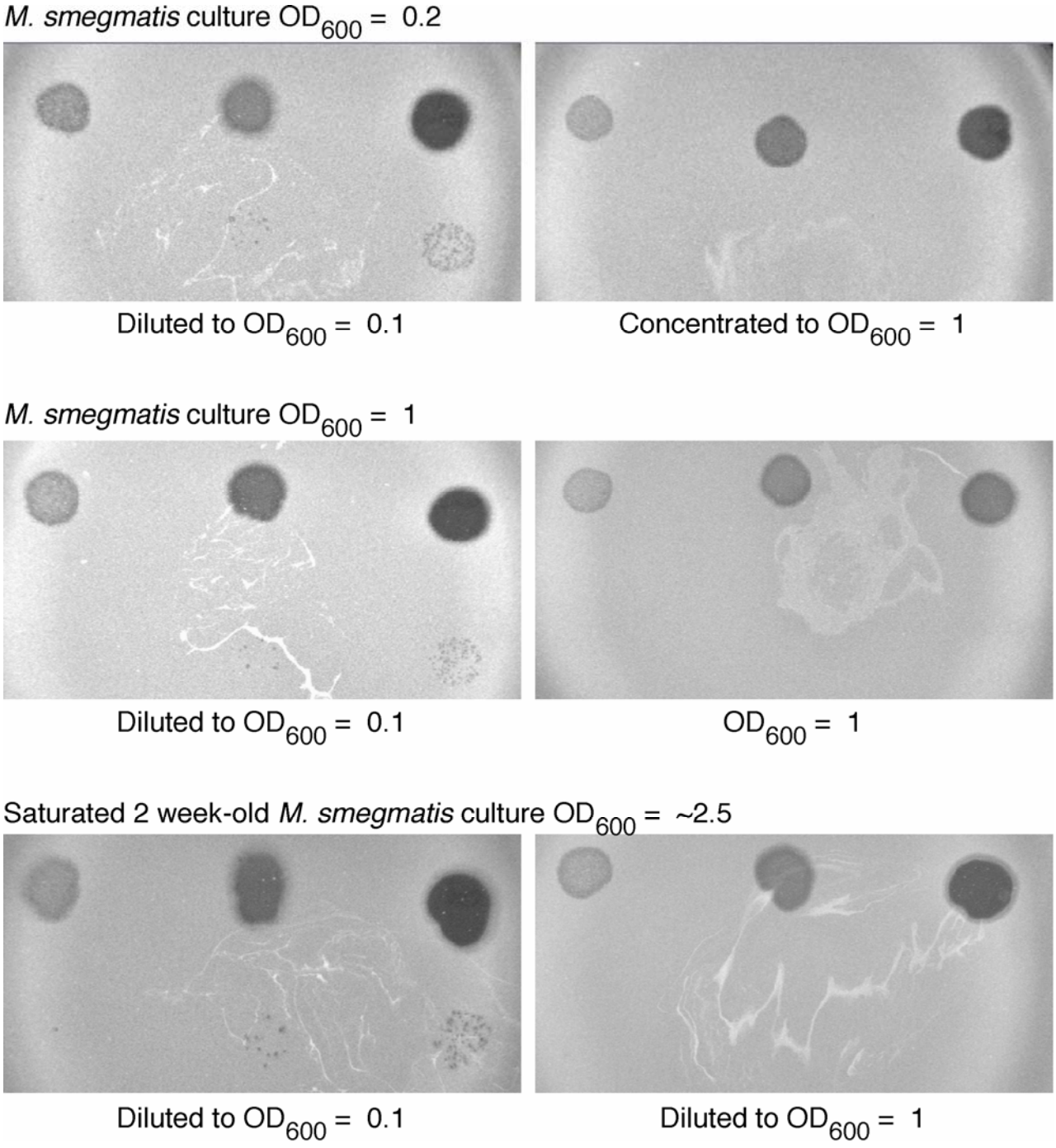


Figure 39: Mycobacteriophage Barnyard is sensitive to host cell plating density.

Figure 39: Barnyard forms small turbid plaques and is exquisitely sensitive to host cell plating density. *M. smegmatis* cells from a pre-exponential (OD_{600} 0.2), a late-exponential (OD_{600} 1.0) and a late-stationary phase (~2 weeks old, OD_{600} ~2.5) were diluted to an OD_{600} of 0.1 in fresh growth medium, and 500 µl of this was used to seed a top agar lawn in 0.35% MBTA with 1 mM $CaCl_2$. Similar lawns were also generated using 500 µl of the pre-exponential culture concentrated to an OD_{600} of 1.0; 500 µl of neat culture at an OD_{600} of 1.0; and 500 µl of the late-stationary culture diluted to an OD_{600} of 1.0. Serial dilutions of Barnyard were spotted on all lawns, and the plates were incubated at 37°C for 2 days. These data illustrate that, regardless of the age of the 'seed'-culture, Barnyard can only form visible plaques when plated on lawns seeded with relatively low numbers of *M. smegmatis* cells.

Unusual Behavior of Barnyard During Liquid Infections of *M. smegmatis*

As noted in Chapter 4, liquid infections of *M. smegmatis* with mycobacteriophage Barnyard have been utilized to investigate the proteolytic cleavage of the tape measure protein (TMP). An additional interesting observation that was gleaned from these time-course experiments, however, was that the levels of TMP protein (both processed and full-length) do not appear to reach a clear peak at a time-point that would correspond to the period immediately prior to host lysis. Rather, the protein levels gradually increase up to eight hours post-infection (P.I.) and only appear to level off after 24 hour of infection (Fig. 16). Monitoring the optical density (OD_{600}) of cells infected at a high M.O.I (15 or 30) also does not give a clear picture of when phage-mediated lysis is occurring. Infected cells appear to grow slower than control, and while in most experiments there is a slight drop in OD_{600} at ~3-4 h P.I., this is not the dramatic lysis often observed with infection by lytic phages, and in most cases cells seem to continue growing to some extent.

Interestingly, when I monitor the titer the supernatant from a high M.O.I. time-course infection, I see a similar phenomenon, despite the fact that ~80% of the phage particles are adsorbed, which suggests that most cells in the population are infected; the titer begins to increase at about two hours P.I. and continues to rise for up to 24 hours (Fig. 40A). Importantly, this pattern is also observed when detergent is added to the cultures after adsorption to prevent multiple rounds of lysis and *re*-infection from occurring and when *M. smegmatis* cells are infected with Barnyard at an M.O.I. as high as 50 (Fig. 40B). These data suggest that all host cells are not being infected synchronously, and – although I have been unable to isolate stable lysogens – perhaps Barnyard can enter into an unstable lysogeny with its host, *M. smegmatis*, during liquid infection. This is consistent with the observation that during the process of lysogen isolation, colonies were often

detected that displayed properties of lysogens (i.e. phage release); however, these were lost upon repeated streaking. This would also be consistent with the patterns of protein expression and phage release observed in liquid culture. A more detailed and quantitative analysis of the patterns of gene expression in infected cultures will be required, however, to definitely ascertain if this sort of phenomenon is occurring.

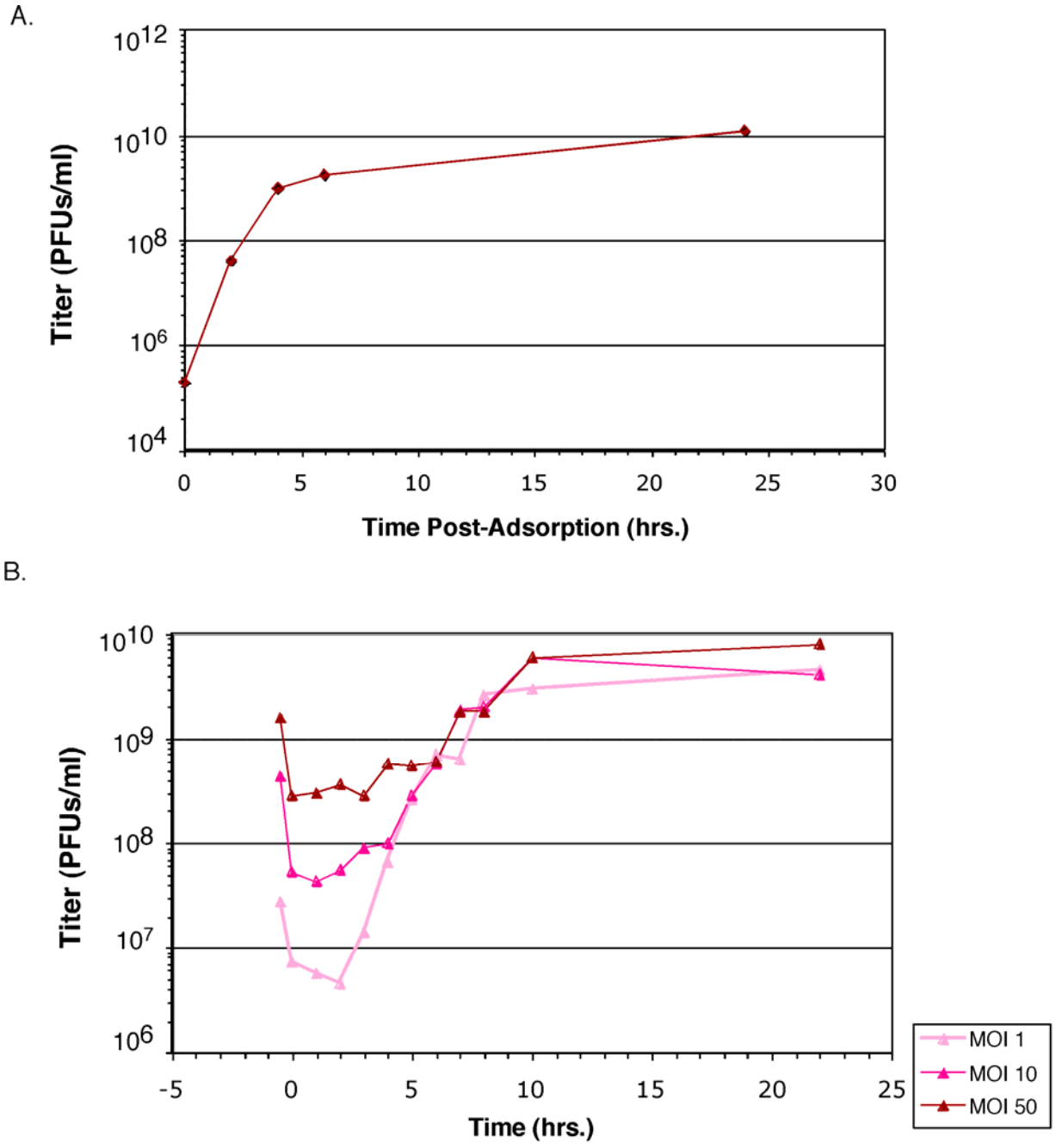


Figure 40: Phage release during mycobacteriophage Barnyard infection of *M. smegmatis* in liquid culture.

Figure 40: Phage release from a liquid infection with mycobacteriophage Barnyard. A. An overnight culture of *M. smegmatis* was grown to an OD_{600} of ~1 in 7H9, CB, CHX supplemented with 10% ADC, as it was previously determined that late exponential cultures were optimal for Barnyard infection. *M. smegmatis* cells were pelleted, resuspended in 7H9-CB supplemented with 4% glucose and 1 mM $CaCl_2$ and diluted to an OD_{600} of 0.1 in the same media. Barnyard particles were added to the culture at an M.O.I. of ~30 and adsorbed at room-temperature for 30 minutes. The entire culture was then washed 3× (dH_2O + 10% glycerol, 0.025% Tween 80) and resuspended in fresh growth media with Tween 80 (0.025%) to prevent cell clumping and re-infection and grown with agitation (~250 rpm) at 37°C. Supernatant was sampled every two hours and titered in duplicate on *M. smegmatis*. The titers of initial (~2×10^9 pfu/ml) and post-adsorption (~4×10^8 pfu/ml) time-points suggest that ~80% of particles are competent to infect. The supernatant titer increases throughout the course of the experiment, and appears to level off by 24 hours of infection. Notably it increases ~10 fold between the six hour and 24 hour time-points. B. An overnight culture of *M. smegmatis* was grown to an OD_{600} of ~1.3 in 7H9, CB, CHX supplemented with 10% ADC. This was diluted to an OD_{600} of 0.1 in 7H9-CB supplemented with 4% glucose and 1 mM $CaCl_2$ and infected with Barnyard at the M.O.I.s noted in the legend. A pre-adsorption time-point was taken immediately following phage addition. Following a 30-minute adsorption (at room temperature), a post-adsorption sample was taken, and Tween 80 was added (0.025%) to all cultures. These were then transferred to 37°C and grown with agitation at ~250 rpm. Culture supernatants from infected cultures were sampled every hour and titered on *M. smegmatis*. One again, the titers obtained for pre- and post-adsorption cultures suggest that in all cases, 74-90% of phage particles are able to infect *M. smegmatis*, and multiple rounds of phage release occur even in cultures infected at an M.O.I. of 10 or 50.

Barnyard Protein Profile

At the time of its isolation, Barnyard was observed to be quite dissimilar from all other previously isolated and sequenced mycobacteriophages at the nucleotide level; and further, the predicted ORFs had very little sequence similarity other proteins in the database (Pedulla *et al.*, 2003). Thus, it was impossible to predict which genes might encode for the basic structural components of the virion, such as the capsid protein or the major tail subunit, or how these might be organized. I therefore, subjected Barnyard structural proteins to both N-terminal sequencing and mass spectrometric analyses, and the results of these experiments are summarized in Figure 41. These data have allowed us to assign functions to a number of these previously unknown proteins and to begin to more clearly define the regions of the genome that encode for the phage structural components. Importantly, these can also assist in the identification of the structural proteins of more recently sequence phages that contain regions with similarity to Barnyard, including PBI1, P-lot (Hatfull *et al*., 2006) and Predator (Hatfull Lab, unpublished data).

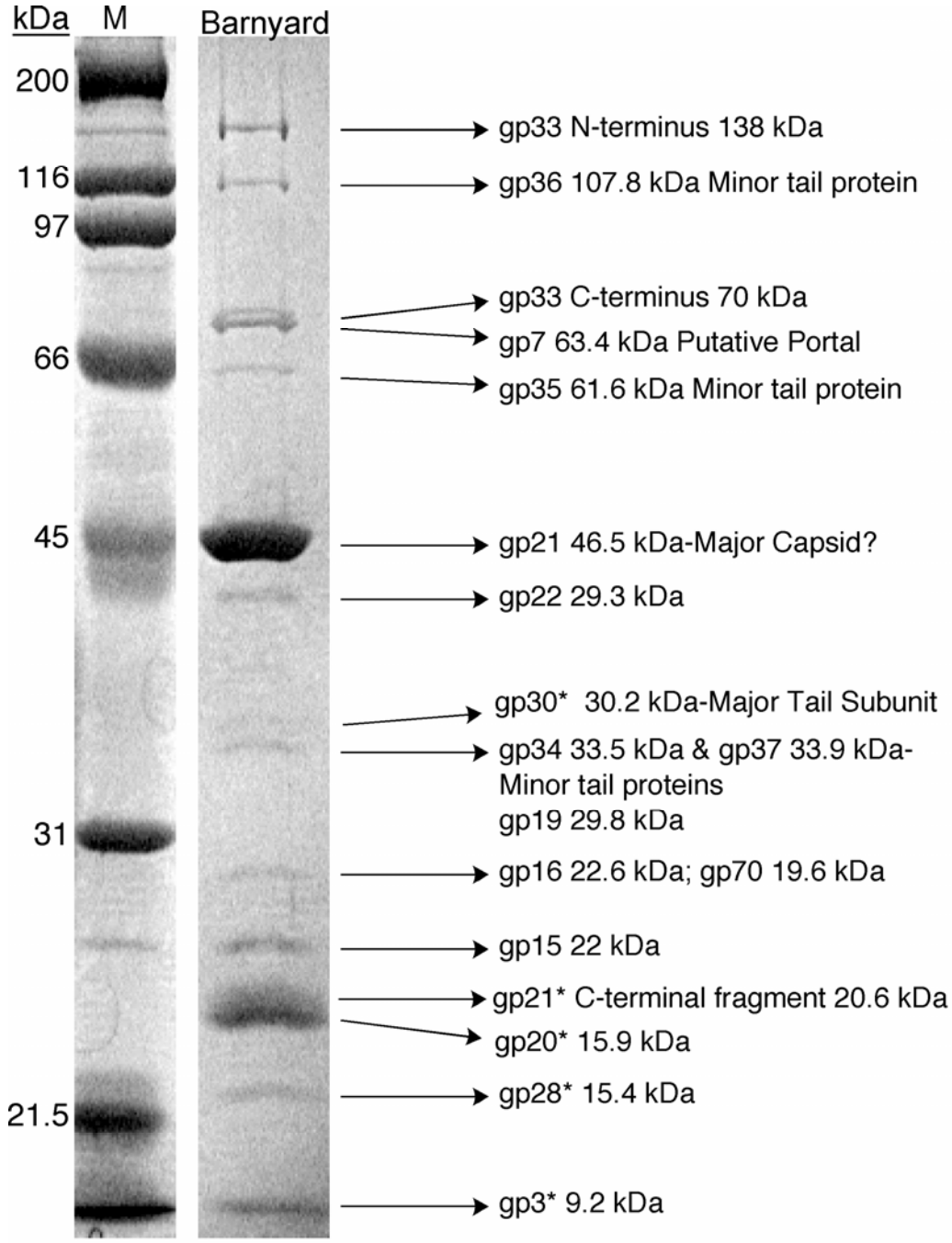


Figure 41: Identification of Barnyard structural proteins by mass spectrometry and N-terminal sequencing.

Figure 41: Identification of Barnyard virion proteins by mass spectrometry. Total phage protein was separated on a 10% SDS polyacrylamide gel and stained with Coomassie Blue. Individual bands were excised and subjected to mass spectrometric analysis (ProtTech, Inc.), and LC-MS/MS spectra were analyzed against the GenBank protein database in order to provide identification. In alternative experiments, proteins were transferred to a PVDF membrane, stained with Coomassie Blue and subjected to N-terminal sequencing.

Putative Identification of the Major Tail Protein

The protein band identified by mass spectrometry as gp30 appears to represent only a minor fraction of the Barnyard protein profile (Fig. 41). However, this protein was initially identified by N-terminal sequencing, and in some preparations of total Barnyard proteins – utilizing an alternative isolation method and longer sample-boiling times – this protein band appeared to be significantly more abundant. Further, BLAST analysis using gp30 as the query sequence detects the presence of a Fibronectin type 3 domain (FN3) in the C-terminus of this protein. This domain, which is similar to the immunoglobulin (Ig) fold, is found in 2% of all animal proteins and if often involved in protein-protein interactions, such as receptor-ligand binding. The FN3 and other Ig-like domains are also found in phage structural proteins, such as the major tail subunit, and may be important for their polymerization (Fraser *et al.*, 2006). Another feature of proteins containing Ig-like domains is that they are highly insoluble, and this might explain the 'disappearance' of this protein in SDS-PAGE gels where shorter boiling times were utilized during sample preparation. Therefore, a 'boiling time-course' was performed in which aliquots of total Barnyard protein, prepared by standard methods, was boiled at 95°C for varying lengths of time prior to SDS-PAGE analysis (Fig. 42). As expected, the protein that likely corresponds to gp30 was barely visible when shorter boiling times were employed; however, this appeared to increase in abundance with increased time of incubation at 95°C. This indicates that gp30 is highly insoluble, a feature that is likely to due its C-terminal FN3 domain, and these data also support the identification of this protein as the major tail subunit. When longer boiling times are employed it is evident that this protein appears to be present in sufficient amounts to correspond to the major tail protein.

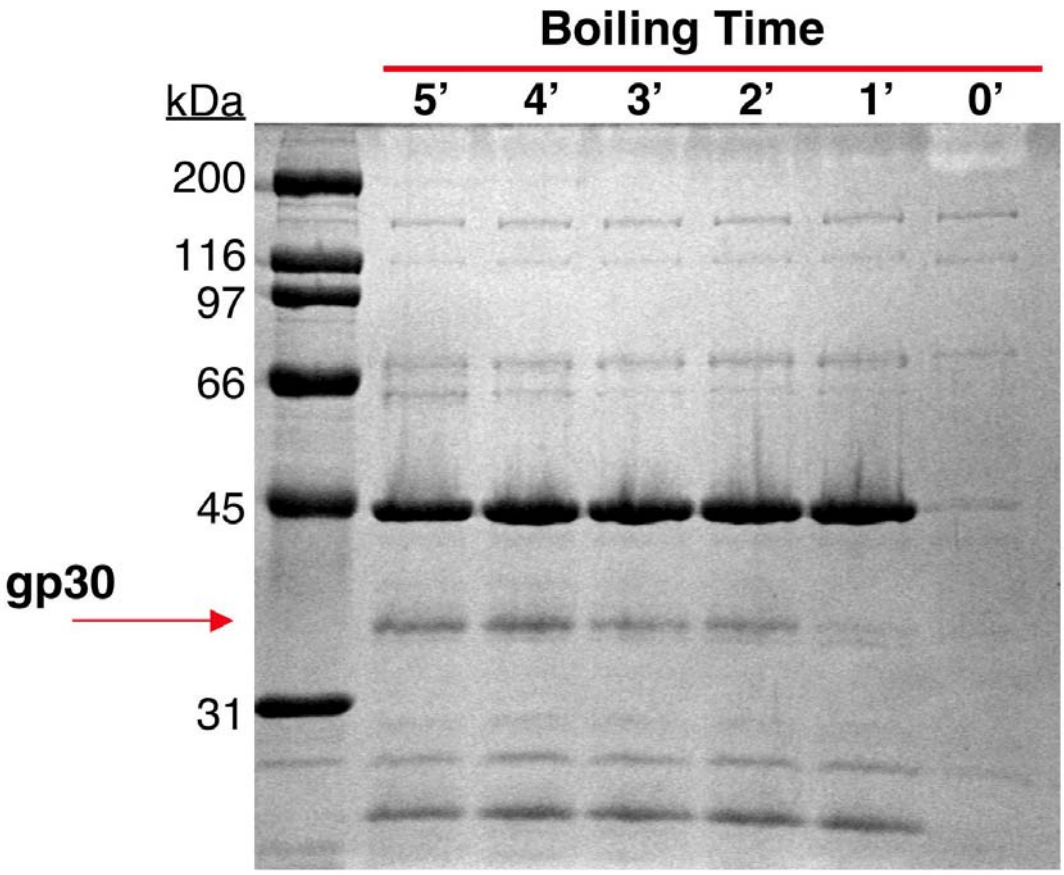


Figure 42: Putative identification of Barnyard gp30 as the major tail subunit.

Figure 42: Putative Identification of gp30 as the major tail protein. Barnyard particles were subjected to three rounds of freezing on dry ice and thawing on ice with vigorous vortexing in between. SDS-Sample buffer (4X) was added was added to the disrupted particles, and samples were boiled at 95°C for the lengths of times indicated above. Virion proteins were then separated on a 10% SDS-PAGE gel, and bands were stained with Coomassie Blue. The protein that likely corresponds to gp30 becomes increasingly insoluble and appears to 'disappear' when shorter boiling times are employed.

Conclusions Pt. 1

Through these studies I have been able to characterize some of the unique features of mycobacteriophage Barnyard, including the dependence on host cell density and the patterns of phage release observed in liquid culture. Notably I have shown that Barnyard only forms visible plaques when bacterial lawns are seeded with a low number of cells (~500 μl of cells at an OD_{600} of 0.1) and appears to undergo multiple rounds of phage release in high M.O.I. liquid culture infections, despite the fact that ~80% of the phage are adsorbed. Further, I have identified components the virion protein profile through a combination of N-terminal sequencing and mass spectrometry. This has allowed putative functions to be assigned for a number of Barnyard ORFs, and may reveal new classes of structural proteins as well as permit the tentative identification of highly similar genes products predicted in the related phages PBI1, Plot and Predator, which have subsequently been isolated.

Although the basis of the Barnyard dependence on host cell density is not known, it is possible that this as well as the extremely small plaque morphology is due to this phages ability to enter into a transient lysogeny with its host. That is, a large proportion of host cells that are infected – either on solid or in liquid media – do not immediately enter into the lytic pathway. Particularly on solid media, this would allow the uninfected hosts to multiply and enter into stationary phase before many of the infected cells undergo lysis and release progeny phage, resulting in a small plaque phenotype that would be more severe the larger the number of host cells are utilized to seed the bacterial lawns. The plaque size of most phages would be expected to depend to some degree on host lawn cell density; however, this might be expected to manifest to a greater degree if lysis was delayed in a large proportion of cells that are infected. It is more

difficult to speculate why such a phenotype would be advantageous to the phage, and future investigation will be required to determine the genetic basis for this unique phenomenon.

A.2 MYCOBACTERIOPHAGE GILES

Mycobacteriophage Giles

Mycobacteriophage Giles is also particularly interesting for a number of reasons. In one respect, as discussed previously, the TMP of this phage contains a canonical Rpf Motif, and it is only the second mycobacteriophage – Barnyard being the first – found to contain this region. Further, similar to Barnyard (at the time of its isolation and characterization), Giles is a unique phage that does not have a high degree of nucleotide sequence similarity with any other sequenced mycobacteriophage genome. Notably, Giles is novel both with respect to its genomic organization and content, as over 50% of its ORFs have no significant matches in the database and only 39% have significant sequence similarity with other mycobacteriophage proteins (Morris *et al.*, 2008). Many of the unique genomic features of Giles, particularly in regards to the insights these might provide about the fundamental processes that govern phage evolution and generate the mosacism that characterizes this population, are addressed in Morris *et al*., 2007, and the results of a number of experiments I performed in order to characterize this phage are presented below.

Giles Protein Profile

The left arm of the Giles genome is predicted to encode for the phage structural proteins; however, only 12 of these have detectable similarity with other mycobacteriophage predicted ORFs. Therefore, Giles proteins were prepared and sequenced by mass spectrometry as described previously for Barnyard, and the results of these analyses are summarized in Figure 43. Notably, similar to Barnyard, the predicted major capsid protein (gp9) does not have significant similarity with any other mycobacteriophage capsid proteins, suggesting that this represents a novel member of this protein family. Based the abundance of gp15 and weak similarity with the Che9c major tail protein (gp14), this function has been putatively assigned to Giles gp15 as well (Morris *et al*., 2008).

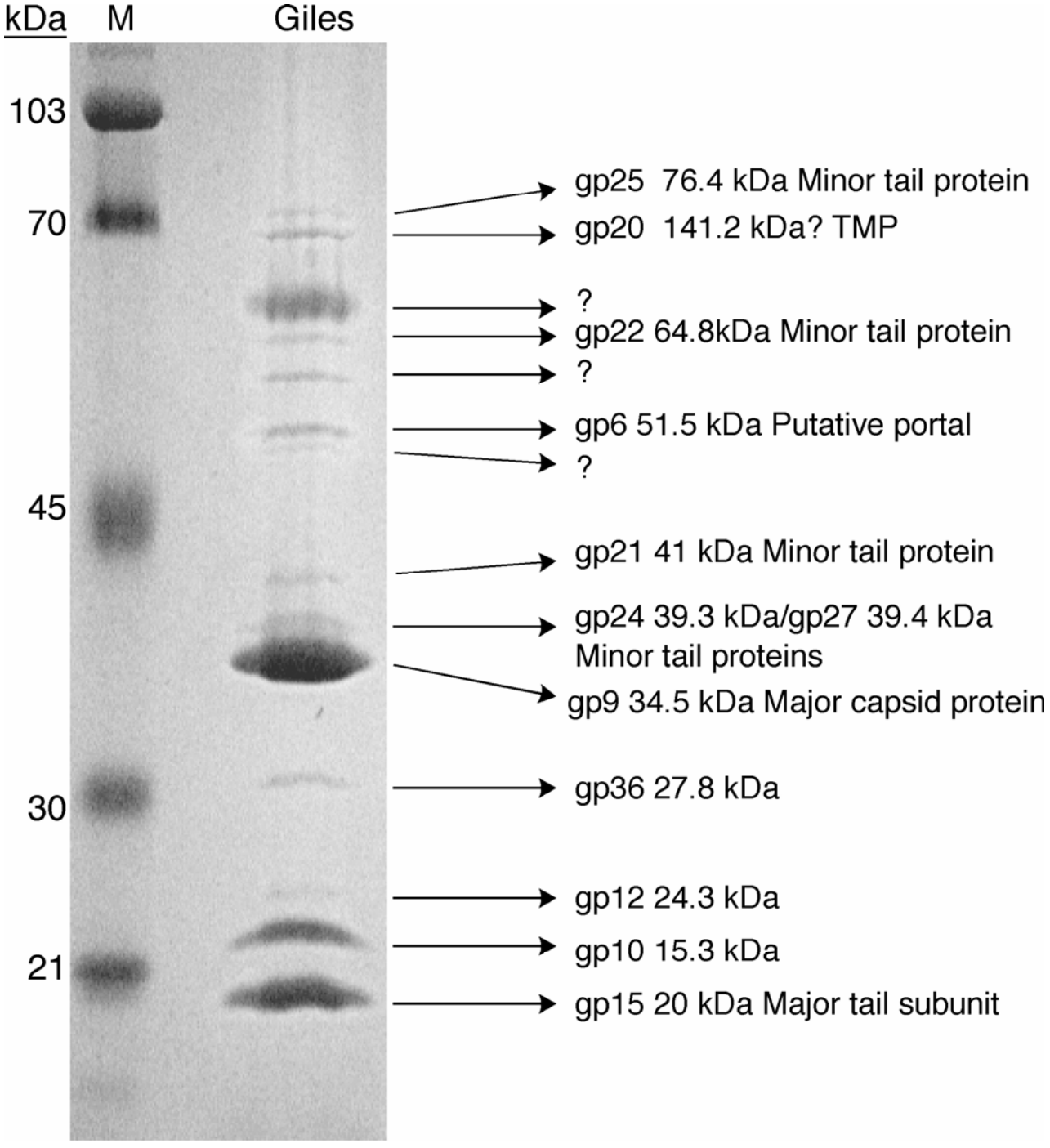


Figure 43: Identification of Giles virion proteins by mass spectrometry.

Figure 43: Identification of Giles virion proteins by mass spectrometry. Total phage protein was separated on a 10% SDS polyacrylamide gel and stained with Coomassie Blue. Individual bands were excised and subjected to mass spectrometric analysis (ProtTech, Inc.), and the LC-MS/MS spectra were analyzed against a unique database containing all predicted Giles ORFs in order to provide identification.

Isolation and Characterization of Giles Lysogens

Sequencing of the Mycobacteriophage Giles genome revealed the presence of genes encoding for a putative tyrosine integrase and an excise protein located near the middle of the genome (Morris *et al*., 2008). It was unknown, however, if this integration cassette was active and if Giles had the ability to form stable lysogens. Further, Giles plaques are neither clear nor turbid, making it difficult to visibly determine if Giles is a lysogenic phage. Therefore, bacterial cells from the clear area of a mycobacteriophage Giles spot titer were isolated and streaked onto solid media. Two independent colonies were then re-streaked multiple times to remove contaminating phage particles, and after each round of isolation, these were checked for phage release by patching onto a lawn of *M. smegmatis,* as illustrated in Figure 44A. Both colonies isolated in this were observed to undergo phage release and were immune to superinfection by Giles (Fig. 44B); however, one appeared to be unstable are large numbers of plaques were observed in lawns generated from this strain. The stable lysogen strain was further shown to be heteroimmune with (i.e. susceptible to infection by) all other phages tested (Fig. 44C), and importantly, the presence of the integrated Giles genome in the host chromosomal *attB* (located within a $tRNA^{Pro}$ gene) was confirmed by PCR with primers that amplify the Giles *attL* (Fig. 44D). Further, the Giles frequency of lysogenation was determined by plating serial dilutions of *M. smegmatis* (containing approximately 2000, 200 and 20 cfus) on agar plates seeded with ~10^9 Giles particles. Under these conditions, the frequency of lysogenation was determined to be ~2%, which is ~10-fold lower than that which was calculated for mycobacteriophage L5 (Donnelly-Wu *et al.*, 1993), although the reason for this is not known. The identity of the Giles repressor protein is also not know; however,

gp68, a WhiB family protein that is located ~5.5 kb from the right end of the genome has been suggested as a possible candidate (Morris *et al*., 2008).

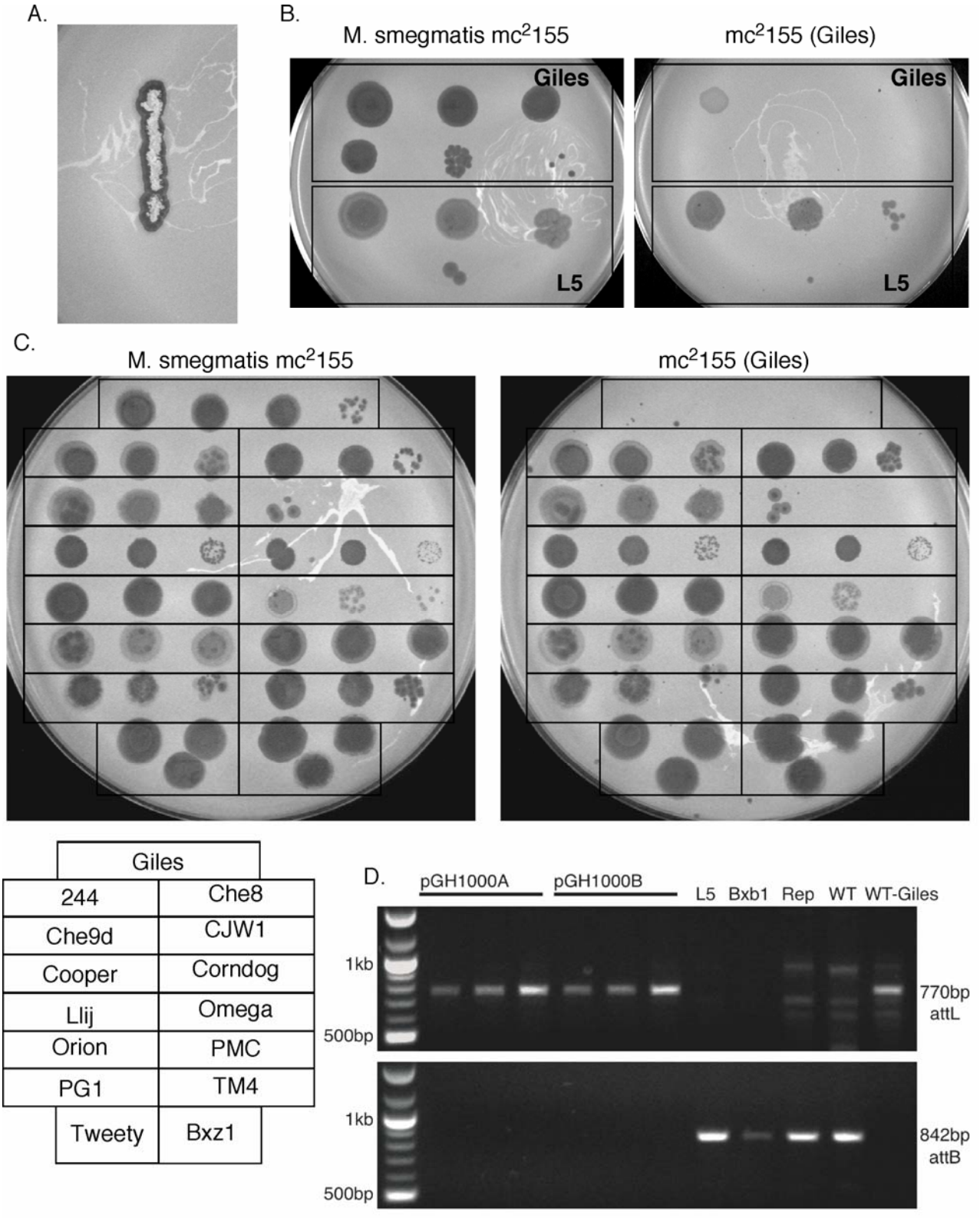


Figure 44: Analysis of mycobacteriophage Giles lysogens.

Figure 44: Analysis of Giles lysogens. A. Phage release from a Giles lysogen. A single colony from a putative Giles lysogen was picked and spotted onto a freshly prepared lawn of *M. smegmatis*. After one day of growth, a zone of clearing can be observed around the spots, indicative of phage release. B. Giles lysogens are immune to Giles infection. Top agar lawns were prepared with both a Giles lysogen and wild-type *M. smegmatis*. Phage L5 is able to infect both strains; however, Giles bacteriophage are unable to infect the lysogen lawn. C. Giles lysogens are sensitive to infection by all mycobacteriophages tested. Top agar lawns were prepared with both a Giles lysogen and wild-type *M. smegmatis*. All phages show identical infectivity on wild-type and Giles lawns, except for Giles, which is unable to infect the Giles lysogen. D. Site-specific integration into the *M. smegmatis attB* site was confirmed by PCR amplification of the attachment junctions in a Giles lysogens and in strains containing integration-proficient plasmids (Morris *et al.*, 2008). Isolated colonies from transformations with Giles integrating vectors and control plasmids, both integrating and extrachromosomal, and a single colony from the Giles lysogen were resuspended in 200 μl sterile dH_2O and boiled at 95°C for 6 minutes; 1 μl was of each was then used as a template in PCRs with primer pairs that amplify either a 770 bp *attL* fragment or an 842 bp *attB* fragment. The *attL* product is amplified from the Giles lysogen and from all Giles transformants, and the *attB* product is obtained with control colonies and wild-type *M. smegmatis.*

Conclusions Pt. 2

The investigation and characterization of mycobacteriophage Giles has provided a number of insights about the biology of this interesting phage, as well as provided evidence for the illegitimate recombination that is believed to be largely responsible for the extensive diversity and mosacism present within the mycobacteriophage population and the phage population as a whole (Hendrix *et al.*, 2003, Morris *et al.*, 2008). Through the experiments presented in this study, I have identified a number of the – in many cases – unique components of the Giles virion protein profile, the characterization of which further illustrates the large amount of diversity present in the phage population and highlights the myriad of ways by which phages can perform the same functions (i.e. assemble their heads and tails). I have also shown that Giles has the ability to form stable lysogens in *M. smegmatis*, albeit at a lower frequency than that which was observed for other lysogenic phages such as L5. This suggests that mycobacteriophages, which are similar to Giles in that they do not form obviously turbid or clear plaques but have predicted integrase functions (Pedulla *et al*., 2003), might display the same activity. Further, because Giles integrates in a unique location in the *M. smegmatis* chromosome relative to the other characterized integrating phages, the Giles integration cassette has been utilized to create integration-proficient vectors, adding to the genetic toolbox that has been provided by these phages. Importantly, these studies collectively serve to highlight the utility and importance of characterizing novel phages and suggest that we have only begun to tap into and explore the enormous amount of information that this population may hold.

BIBLIOGRAPHY

Abuladze, N. K., M. Gingery, J. Tsai & F. A. Eiserling, (1994) Tail length determination in bacteriophage T4. *Virology* **199**: 301-310.

Atrih, A., G. Bacher, G. Allmaier, M. P. Williamson & S. J. Foster, (1999) Analysis of peptidoglycan structure from vegetative cells of Bacillus subtilis 168 and role of PBP 5 in peptidoglycan maturation. *Journal of bacteriology* **181**: 3956-3966.

Atrih, A. & S. J. Foster, (1999) The role of peptidoglycan structure and structural dynamics during endospore dormancy and germination. *Antonie van Leeuwenhoek* **75**: 299-307.

Azuma, I., D. W. Thomas, A. Adam, J. M. Ghuysen, R. Bonaly, J. F. Petit & E. Lederer, (1970) Occurrence of N-glycolylmuramic acid in bacterial cell walls. A preliminary survey. *Biochim Biophys Acta* **208**: 444-451.

Bamford, D. H., M. Romantschuk & P. J. Somerharju, (1987) Membrane fusion in prokaryotes: bacteriophage phi 6 membrane fuses with the Pseudomonas syringae outer membrane. *The EMBO journal* **6**: 1467-1473.

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

Barry, C. E., 3rd, R. E. Lee, K. Mdluli, A. E. Sampson, B. G. Schroeder, R. A. Slayden & Y. Yuan, (1998) Mycolic acids: structure, biosynthesis and physiological functions. *Progress in lipid research* **37**: 143-179.

Bateman, A. & M. Bycroft, (2000) The structure of a LysM domain from E. coli membrane-bound lytic murein transglycosylase D (MltD). *Journal of molecular biology* **299**: 1113-1119.

Bayer, M., R. Eferl, G. Zellnig, K. Teferle, A. Dijkstra, G. Koraimann & G. Hogenauer, (1995) Gene 19 of plasmid R1 is required for both efficient conjugative DNA transfer and bacteriophage R17 infection. *Journal of bacteriology* **177**: 4279-4288.

Bayer, M., R. Iberer, K. Bischof, E. Rassi, E. Stabentheiner, G. Zellnig & G. Koraimann, (2001) Functional and mutational analysis of p19, a DNA transfer protein with muramidase activity. *Journal of bacteriology* **183**: 3176-3183.

Bensing, B. A., I. R. Siboo & P. M. Sullam, (2001) Proteins PblA and PblB of Streptococcus mitis, which promote binding to human platelets, are encoded within a lysogenic bacteriophage. *Infection and immunity* **69**: 6186-6192.

Bibb, L. A. & G. F. Hatfull, (2002) Integration and excision of the Mycobacterium tuberculosis prophage-like element, phiRv1. *Molecular microbiology* **45**: 1515-1526.

Bohm, J., O. Lambert, A. S. Frangakis, L. Letellier, W. Baumeister & J. L. Rigaud, (2001) FhuA-mediated phage genome transfer into liposomes: a cryo-electron tomography study. *Curr Biol* **11**: 1168-1175.

Boulanger, P., P. Jacquot, L. Plancon, M. Chami, A. Engel, C. Parquet, C. Herbeuval & L. Letellier, (2008) Phage T5 straight tail fiber is a multifunctional protein acting as a tape measure and carrying fusogenic and muralytic activities. *The Journal of biological chemistry*.

Braun, V. & H. Wolff, (1970) The murein-lipoprotein linkage in the cell wall of Escherichia coli. *European journal of biochemistry / FEBS* **14**: 387-391.

Brennan, P. J., (2003) Structure, function, and biogenesis of the cell wall of Mycobacterium tuberculosis. *Tuberculosis (Edinburgh, Scotland)* **83**: 91-97.

Brennan, P. J. & H. Nikaido, (1995) The envelope of mycobacteria. *Annual review of biochemistry* **64**: 29-63.

Burkhard, P., J. Stetefeld & S. V. Strelkov, (2001) Coiled coils: a highly versatile protein folding motif. *Trends in cell biology* **11**: 82-88.

Caldentey, J. & D. H. Bamford, (1992) The lytic enzyme of the Pseudomonas phage phi 6. Purification and biochemical characterization. *Biochimica et biophysica acta* **1159**: 44-50.

Casjens, S. R. & R. W. Hendrix, (1974) Locations and amounts of major structural proteins in bacteriophage lambda. *Journal of molecular biology* **88**: 535-545.

Chatterjee, D., (1997) The mycobacterial cell wall: structure, biosynthesis and sites of drug action. *Current opinion in chemical biology* **1**: 579-588.

Cohen-Gonsaud, M., P. Barthe, C. Bagneris, B. Henderson, J. Ward, C. Roumestand & N. H. Keep, (2005) The structure of a resuscitation-promoting factor domain from Mycobacterium tuberculosis shows homology to lysozymes. *Nature structural & molecular biology* **12**: 270-273.

Cohen-Gonsaud, M., N. H. Keep, A. P. Davies, J. Ward, B. Henderson & G. Labesse, (2004) Resuscitation-promoting factors possess a lysozyme-like domain. *Trends in biochemical sciences* **29**: 7-10.

Court, D. L., J. A. Sawitzke & L. C. Thomason, (2002) Genetic engineering using homologous recombination. *Annual review of genetics* **36**: 361-388.

Crick, F. H. C., (1953) The packing of alpha-helices: simple coiled coils. *Acta. Crystallogr.* **6**: 689-697.

Crutz-Le Coq, A. M., B. Cesselin, J. Commissaire & J. Anba, (2002) Sequence analysis of the lactococcal bacteriophage bIL170: insights into structural proteins and HNH endonucleases in dairy phages. *Microbiology (Reading, England)* **148**: 985-1001.

Cunningham, A. F. & C. L. Spreadbury, (1998) Mycobacterial stationary phase induced by low oxygen tension: cell wall thickening and localization of the 16-kilodalton alpha-crystallin homolog. *Journal of bacteriology* **180**: 801-808.

Daffe, M. & P. Draper, (1998) The envelope layers of mycobacteria with reference to their pathogenicity. *Advances in microbial physiology* **39**: 131-203.

Daugelavicius, R., V. Cvirkaite, A. Gaidelyte, E. Bakiene, R. Gabrenaite-Verkhovskaya & D. H. Bamford, (2005) Penetration of enveloped double-stranded RNA bacteriophages phi13 and phi6 into Pseudomonas syringae cells. *Journal of virology* **79**: 5017-5026.

Del Angel, V. D., F. Dupuis, J. P. Mornon & I. Callebaut, (2002) Viral fusion peptides and identification of membrane-interacting segments. *Biochemical and biophysical research communications* **293**: 1153-1160.

Demchick, P. & A. L. Koch, (1996) The permeability of the wall fabric of Escherichia coli and Bacillus subtilis. *Journal of bacteriology* **178**: 768-773.

Dijkstra, A. J. & W. Keck, (1996) Peptidoglycan as a barrier to transenvelope transport. *Journal of bacteriology* **178**: 5555-5562.

Dijkstra, B. W. & A. M. Thunnissen, (1994) 'Holy' proteins. II: The soluble lytic transglycosylase. *Current opinion in structural biology* **4**: 810-813.

Donate, L. E., L. Herranz, J. P. Secilla, J. M. Carazo, H. Fujisawa & J. L. Carrascosa, (1988) Bacteriophage T3 connector: three-dimensional structure and comparison with other viral head-tail connecting regions. *Journal of molecular biology* **201**: 91-100.

Donnelly-Wu, M. K., W. R. Jacobs, Jr. & G. F. Hatfull, (1993) Superinfection immunity of mycobacteriophage L5: applications for genetic transformation of mycobacteria. *Molecular microbiology* **7**: 407-417.

Downing, K. J., J. C. Betts, D. I. Young, R. A. McAdam, F. Kelly, M. Young & V. Mizrahi, (2004) Global expression profiling of strains harbouring null mutations reveals that the five rpf-like genes of Mycobacterium tuberculosis show functional redundancy. *Tuberculosis (Edinburgh, Scotland)* **84**: 167-179.

Downing, K. J., V. V. Mischenko, M. O. Shleeva, D. I. Young, M. Young, A. S. Kaprelyants, A. S. Apt & V. Mizrahi, (2005) Mutants of Mycobacterium tuberculosis lacking three of the five rpf-like genes are defective for growth in vivo and for resuscitation in vitro. *Infection and immunity* **73**: 3038-3043.

Dubrac, S. & T. Msadek, (2004) Identification of genes controlled by the essential YycG/YycF two-component system of Staphylococcus aureus. *Journal of bacteriology* **186**: 1175-1181.

Duda, R. L., M. Gingery & F. A. Eiserling, (1986) Potential length determiner and DNA injection protein is extruded from bacteriophage T4 tail tubes in vitro. *Virology* **151**: 296-314.

Elliott, J. & W. Arber, (1978) E. coli K-12 pel mutants, which block phage lambda DNA injection, coincide with ptsM, which determines a component of a sugar transport system. *Mol Gen Genet* **161**: 1-8.

Ellis, H. M., D. Yu, T. DiTizio & D. L. Court, (2001) High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 6742-6746.

Engel, H., B. Kazemier & W. Keck, (1991) Murein-metabolizing enzymes from Escherichia coli: sequence analysis and controlled overexpression of the slt gene, which encodes the soluble lytic transglycosylase. *Journal of bacteriology* **173**: 6773-6782.

Fenhalls, G., L. Stevens, L. Moses, J. Bezuidenhout, J. C. Betts, P. Helden Pv, P. T. Lukey & K. Duncan, (2002) In situ detection of Mycobacterium tuberculosis transcripts in human lung granulomas reveals differential gene expression in necrotic lesions. *Infection and immunity* **70**: 6330-6338.

Feucht, A., A. Schmid, R. Benz, H. Schwarz & K. J. Heller, (1990) Pore formation associated with the tail-tip protein pb2 of bacteriophage T5. *The Journal of biological chemistry* **265**: 18561-18567.

Fisher, M. A., B. B. Plikaytis & T. M. Shinnick, (2002) Microarray analysis of the Mycobacterium tuberculosis transcriptional response to the acidic conditions found in phagosomes. *Journal of bacteriology* **184**: 4025-4032.

Flores, A. R., L. M. Parsons & M. S. Pavelka, Jr., (2005a) Characterization of novel Mycobacterium tuberculosis and Mycobacterium smegmatis mutants hypersusceptible to beta-lactam antibiotics. *Journal of bacteriology* **187**: 1892-1900.

Flores, A. R., L. M. Parsons & M. S. Pavelka, Jr., (2005b) Genetic analysis of the beta-lactamases of Mycobacterium tuberculosis and Mycobacterium smegmatis and susceptibility to beta-lactam antibiotics. *Microbiology (Reading, England)* **151**: 521-532.

Flynn, J. L. & J. Chan, (2001) Tuberculosis: latency and reactivation. *Infection and immunity* **69**: 4195-4201.

Ford, M. E., C. Stenstrom, R. W. Hendrix & G. F. Hatfull, (1998) Mycobacteriophage TM4: genome structure and gene expression. *Tuber Lung Dis* **79**: 63-73.

Fordham, W. D. & C. Gilvarg, (1974) Kinetics of cross-linking of peptidoglycan in Bacillus megaterium. *The Journal of biological chemistry* **249**: 2478-2482.

Fraser, J. S., Z. Yu, K. L. Maxwell & A. R. Davidson, (2006) Ig-like domains on bacteriophages: a tale of promiscuity and deceit. *Journal of molecular biology* **359**: 496-507.

Garcia, L. R. & I. J. Molineux, (1996) Transcription-independent DNA translocation of bacteriophage T7 DNA into Escherichia coli. *Journal of bacteriology* **178**: 6921-6929.

Goffin, C. & J. M. Ghuysen, (2002) Biochemistry and comparative genomics of SxxK superfamily acyltransferases offer a clue to the mycobacterial paradox: presence of penicillin-susceptible target proteins versus lack of efficiency of penicillin as therapeutic agent. *Microbiol Mol Biol Rev* **66**: 702-738, table of contents.

Gomez, M., S. Johnson & M. L. Gennaro, (2000) Identification of secreted proteins of Mycobacterium tuberculosis by a bioinformatic approach. *Infection and immunity* **68**: 2323-2327.

Goodell, E. W., (1985) Recycling of murein by Escherichia coli. *Journal of bacteriology* **163**: 305-310.

Gottesman, M., (1999) Bacteriophage lambda: the untold story. *Journal of molecular biology* **293**: 177-180.

Grahn, A. M., R. Daugelavicius & D. H. Bamford, (2002) Sequential model of phage PRD1 DNA delivery: active involvement of the viral membrane. *Molecular microbiology* **46**: 1199-1209.

Guihard, G., P. Boulanger & L. Letellier, (1992) Involvement of phage T5 tail proteins and contact sites between the outer and inner membrane of Escherichia coli in phage T5 DNA injection. *The Journal of biological chemistry* **267**: 3173-3178.

Hall, S. D. & R. D. Kolodner, (1994) Homologous pairing and strand exchange promoted by the Escherichia coli RecT protein. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 3205-3209.

Hartmann, M., A. Barsch, K. Niehaus, A. Puhler, A. Tauch & J. Kalinowski, (2004) The glycosylated cell surface protein Rpf2, containing a resuscitation-promoting factor motif, is involved in intercellular communication of Corynebacterium glutamicum. *Archives of microbiology* **182**: 299-312.

Hatfull, G. F., M. L. Pedulla, D. Jacobs-Sera, P. M. Cichon, A. Foley, M. E. Ford, R. M. Gonda, J. M. Houtz, A. J. Hryckowian, V. A. Kelchner, S. Namburi, K. V. Pajcini, M. G. Popovich, D. T. Schleicher, B. Z. Simanek, A. L. Smith, G. M. Zdanowicz, V. Kumar, C. L. Peebles, W. R. Jacobs, Jr., J. G. Lawrence & R. W. Hendrix, (2006) Exploring the mycobacteriophage metaproteome: phage genomics as an educational platform. *PLoS genetics* **2**: e92.

Hatfull Lab, (unpublished data).

Hendrix, R. W., (2002) Bacteriophages: evolution of the majority. *Theoretical population biology* **61**: 471-480.

Hendrix, R. W., (2003) Bacteriophage genomics. *Current opinion in microbiology* **6**: 506-511.
Hendrix, R. W. & S. R. Casjens, (1974) Protein cleavage in bacteriophage lambda tail assembly. *Virology* **61**: 156-159.
Hendrix, R. W., G. F. Hatfull & M. C. Smith, (2003) Bacteriophages with tails: chasing their origins and evolution. *Research in microbiology* **154**: 253-257.
Hett, E. C., M. C. Chao, A. J. Steyn, S. M. Fortune, L. L. Deng & E. J. Rubin, (2007) A partner for the resuscitation-promoting factors of Mycobacterium tuberculosis. *Molecular microbiology* **66**: 658-668.
Hoffmann, C., A. Leis, M. Niederweis, J. M. Plitzko & H. Engelhardt, (2008) Disclosure of the mycobacterial outer membrane: cryo-electron tomography and vitreous sections reveal the lipid bilayer structure. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 3963-3967.
Hofmann, K. & S. W., (1993) TMbase - A database of membrane spanning proteins segments. *Biol Chem* **374**: 166.
Holtje, J. V., (1995) From growth to autolysis: the murein hydrolases in Escherichia coli. *Archives of microbiology* **164**: 243-254.
Holtje, J. V., (1998) Growth of the stress-bearing and shape-maintaining murein sacculus of Escherichia coli. *Microbiol Mol Biol Rev* **62**: 181-203.
Holtje, J. V., D. Mirelman, N. Sharon & U. Schwarz, (1975) Novel type of murein transglycosylase in Escherichia coli. *Journal of bacteriology* **124**: 1067-1076.
Hoppner, C., Z. Liu, N. Domke, A. N. Binns & C. Baron, (2004) VirB1 orthologs from Brucella suis and pKM101 complement defects of the lytic transglycosylase required for efficient type IV secretion from Agrobacterium tumefaciens. *Journal of bacteriology* **186**: 1415-1422.
Iyer, L. M., E. V. Koonin & L. Aravind, (2002) Classification and evolutionary history of the single-strand annealing proteins, RecT, Redbeta, ERF and RAD52. *BMC genomics* **3**: 8.
Jacobs, C., J. M. Frere & S. Normark, (1997) Cytosolic intermediates for cell wall biosynthesis and degradation control inducible beta-lactam resistance in gram-negative bacteria. *Cell* **88**: 823-832.
Jacobs, W. R., Jr., M. Tuckman & B. R. Bloom, (1987) Introduction of foreign DNA into mycobacteria using a shuttle phasmid. *Nature* **327**: 532-535.
Jarvis, A. W., G. F. Fitzgerald, M. Mata, A. Mercenier, H. Neve, I. B. Powell, C. Ronda, M. Saxelin & M. Teuber, (1991) Species and type phages of lactococcal bacteriophages. *Intervirology* **32**: 2-9.
Johnsen, M. G., H. Neve, F. K. Vogensen & K. Hammer, (1995) Virion positions and relationships of lactococcal temperate bacteriophage TP901-1 proteins. *Virology* **212**: 595-606.
Jones, D. T., (1999) Protein secondary structure prediction based on position-specific scoring matrices. *Journal of molecular biology* **292**: 195-202.
Joseph, J. W. & R. Kolodner, (1983) Exonuclease VIII of Escherichia coli. I. Purification and physical properties. *The Journal of biological chemistry* **258**: 10411-10417.
Kana, B. D., B. G. Gordhan, K. J. Downing, N. Sung, G. Vostroktunova, E. E. Machowski, L. Tsenova, M. Young, A. Kaprelyants, G. Kaplan & V. Mizrahi, (2008) The resuscitation-promoting factors of Mycobacterium tuberculosis are required for virulence and resuscitation from dormancy but are collectively dispensable for growth in vitro. *Molecular microbiology* **67**: 672-684.

Kanamaru, S., N. C. Gassner, N. Ye, S. Takeda & F. Arisaka, (1999) The C-terminal fragment of the precursor tail lysozyme of bacteriophage T4 stays as a structural component of the baseplate after cleavage. *Journal of bacteriology* **181**: 2739-2744.

Kanamaru, S., Y. Ishiwata, T. Suzuki, M. G. Rossmann & F. Arisaka, (2005) Control of bacteriophage T4 tail lysozyme activity during the infection process. *Journal of molecular biology* **346**: 1013-1020.

Kanamaru, S., P. G. Leiman, V. A. Kostyuchenko, P. R. Chipman, V. V. Mesyanzhinov, F. Arisaka & M. G. Rossmann, (2002) Structure of the cell-puncturing device of bacteriophage T4. *Nature* **415**: 553-557.

Kao, S. H. & W. H. McClain, (1980) Baseplate protein of bacteriophage T4 with both structural and lytic functions. *Journal of virology* **34**: 95-103.

Kaprelyants, A. S., J. C. Gottschal & D. B. Kell, (1993) Dormancy in non-sporulating bacteria. *FEMS microbiology reviews* **10**: 271-285.

Kaprelyants, A. S. & D. B. Kell, (1993) Dormancy in Stationary-Phase Cultures of Micrococcus luteus: Flow Cytometric Analysis of Starvation and Resuscitation. *Appl Environ Microbiol* **59**: 3187-3196.

Kaprelyants, A. S., G. V. Mukamolova, H. M. Davey & D. B. Kell, (1996) Quantitative Analysis of the Physiological Heterogeneity within Starved Cultures of Micrococcus luteus by Flow Cytometry and Cell Sorting. *Appl Environ Microbiol* **62**: 1311-1316.

Kaprelyants, A. S., G. V. Mukamolova & D. B. Kell, (1994) Estimation of dormant Micrococcus luteus cells by penicillin lysis and by resuscitation in cell-free spent culture medium at high dilution. *FEMS Microbiol Lett* **115**: 347-352.

Katsura, I., (1976) Morphogenesis of bacteriophage lambda tail. Polymorphism in the assembly of the major tail protein. *Journal of molecular biology* **107**: 307-326.

Katsura, I., (1983) Tail Assembly and Injection In: Lambda II. R. W. Hendrix, J. W. Roberts, F. W. Stahl & R. A. Weisberg (eds). Cold Spring Harbor, New York: Cold Spring Harbor Laboratory, pp. 331-346.

Katsura, I., (1987) Determination of bacteriophage lambda tail length by a protein ruler. *Nature* **327**: 73-75.

Katsura, I., (1990) Mechanism of length determination in bacteriophage lambda tails. *Advances in biophysics* **26**: 1-18.

Katsura, I. & R. W. Hendrix, (1984) Length determination in bacteriophage lambda tails. *Cell* **39**: 691-698.

Katsura, I. & P. W. Kuhl, (1975) Morphogenesis of the tail of bacteriophage lambda. III. Morphogenetic pathway. *Journal of molecular biology* **91**: 257-273.

Keep, N. H., J. M. Ward, M. Cohen-Gonsaud & B. Henderson, (2006) Wake up! Peptidoglycan lysis and bacterial non-growth states. *Trends in microbiology* **14**: 271-276.

Kell, D. B. & M. Young, (2000) Bacterial dormancy and culturability: the role of autocrine growth factors. *Current opinion in microbiology* **3**: 238-243.

Kellenberger, E., (1972). In: Polymerization in Biological Systems. G. E. W. Wolstenholme & M. O'Conner (eds). New York: Elsevier, pp. 295-299.

Kemp, P., L. R. Garcia & I. J. Molineux, (2005) Changes in bacteriophage T7 virion structure at the initiation of infection. *Virology* **340**: 307-317.

Kemp, P., M. Gupta & I. J. Molineux, (2004) Bacteriophage T7 DNA ejection into cells is initiated by an enzyme-like mechanism. *Molecular microbiology* **53**: 1251-1265.

Kenny, J. G., S. McGrath, G. F. Fitzgerald & D. van Sinderen, (2004) Bacteriophage Tuc2009 encodes a tail-associated cell wall-degrading activity. *Journal of bacteriology* **186**: 3480-3491.

Koch, A. L. & R. J. Doyle, (1985) Inside-to-outside growth and turnover of the wall of gram-positive rods. *Journal of theoretical biology* **117**: 137-157.

Koonin, E. V. & K. E. Rudd, (1994) A conserved domain in putative bacterial and bacteriophage transglycosylases. *Trends in biochemical sciences* **19**: 106-107.

Laemmli, U. K., (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.

Lambert, O., L. Plancon, J. L. Rigaud & L. Letellier, (1998) Protein-mediated DNA transfer into liposomes. *Molecular microbiology* **30**: 761-765.

Lee, E. C., D. Yu, J. Martinez de Velasco, L. Tessarollo, D. A. Swing, D. L. Court, N. A. Jenkins & N. G. Copeland, (2001) A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* **73**: 56-65.

Lehnherr, H., A. M. Hansen & T. Ilyina, (1998) Penetration of the bacterial cell wall: a family of lytic transglycosylases in bacteriophages and conjugative plasmids. *Molecular microbiology* **30**: 454-457.

Leiman, P. G., P. R. Chipman, V. A. Kostyuchenko, V. V. Mesyanzhinov & M. G. Rossmann, (2004) Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. *Cell* **118**: 419-429.

Levin, M. E., R. W. Hendrix & S. R. Casjens, (1993) A programmed translational frameshift is required for the synthesis of a bacteriophage lambda tail assembly protein. *Journal of molecular biology* **234**: 124-139.

Lewis, J. A. & G. F. Hatfull, (2000) Identification and characterization of mycobacteriophage L5 excisionase. *Molecular microbiology* **35**: 350-360.

Li, Z., G. Karakousis, S. K. Chiu, G. Reddy & C. M. Radding, (1998) The beta protein of phage lambda promotes strand exchange. *Journal of molecular biology* **276**: 733-744.

Little, J. W., (1967) An exonuclease induced by bacteriophage lambda. II. Nature of the enzymatic reaction. *The Journal of biological chemistry* **242**: 679-686.

Lupas, A., M. Van Dyke & J. Stock, (1991) Predicting coiled coils from protein sequences. *Science* **252**: 1162-1164.

Malcolm, B. A., S. Rosenberg, M. J. Corey, J. S. Allen, A. de Baetselier & J. F. Kirsch, (1989) Site-directed mutagenesis of the catalytic residues Asp-52 and Glu-35 of chicken egg white lysozyme. *Proceedings of the National Academy of Sciences of the United States of America* **86**: 133-137.

Manabe, Y. C. & W. R. Bishai, (2000) Latent Mycobacterium tuberculosis-persistence, patience, and winning by waiting. *Nature medicine* **6**: 1327-1329.

Manganelli, R., M. I. Voskuil, G. K. Schoolnik & I. Smith, (2001) The Mycobacterium tuberculosis ECF sigma factor sigmaE: role in global gene expression and survival in macrophages. *Molecular microbiology* **41**: 423-437.

Matias, V. R. & T. J. Beveridge, (2005) Cryo-electron microscopy reveals native polymeric cell wall structure in Bacillus subtilis 168 and the existence of a periplasmic space. *Molecular microbiology* **56**: 240-251.

Matias, V. R. & T. J. Beveridge, (2006) Native cell wall organization shown by cryo-electron microscopy confirms the existence of a periplasmic space in Staphylococcus aureus. *Journal of bacteriology* **188**: 1011-1021.

Mc Grath, S., H. Neve, J. F. Seegers, R. Eijlander, C. S. Vegge, L. Brondsted, K. J. Heller, G. F. Fitzgerald, F. K. Vogensen & D. van Sinderen, (2006) Anatomy of a lactococcal phage tail. *Journal of bacteriology* **188**: 3972-3982.

McGuffin, L. J., K. Bryson & D. T. Jones, (2000) The PSIPRED protein structure prediction server. *Bioinformatics (Oxford, England)* **16**: 404-405.

Mellado, R. P., M. A. Penalva, M. R. Inciarte & M. Salas, (1980) The protein covalently linked to the 5' termini of the DNA of Bacillus subtilis phage phi 29 is involved in the initiation of DNA replication. *Virology* **104**: 84-96.

Mindich, L. & J. Lehman, (1979) Cell wall lysin as a component of the bacteriophage phi 6 virion. *Journal of virology* **30**: 489-496.

Minnikin, D., (1982) Lipids: complex lipids, their chemistry, biosynthesis and roles. In: The Biology of Mycobacteria. C. Tatledge & J. Stanford (eds). London: Academic Press, pp. 94-184.

Moak, M. & I. J. Molineux, (2000) Role of the Gp16 lytic transglycosylase motif in bacteriophage T7 virions at the initiation of infection. *Molecular microbiology* **37**: 345-355.

Moak, M. & I. J. Molineux, (2004) Peptidoglycan hydrolytic activities associated with bacteriophage virions. *Molecular microbiology* **51**: 1169-1183.

Molineux, I. J., (2001) No syringes please, ejection of phage T7 DNA from the virion is enzyme driven. *Molecular microbiology* **40**: 1-8.

Molineux, I. J., (2006) Fifty-three years since Hershey and Chase; much ado about pressure but which pressure is it? *Virology* **344**: 221-229.

Morris, P., L. J. Marinelli, D. Jacobs-Sera, R. W. Hendrix & G. F. Hatfull, (2008) Genomic characterization of mycobacteriophage Giles: Evidence for phage acquisition of host DNA by illegitimate recombination. *Journal of bacteriology*.

Mosig, G., G. W. Lin, J. Franklin & W. H. Fan, (1989) Functional relationships and structural determinants of two bacteriophage T4 lysozymes: a soluble (gene e) and a baseplate-associated (gene 5) protein. *The New biologist* **1**: 171-179.

Mukamolova, G. V., A. S. Kaprelyants, D. B. Kell & M. Young, (2003) Adoption of the transiently non-culturable state--a bacterial survival strategy? *Advances in microbial physiology* **47**: 65-129.

Mukamolova, G. V., A. S. Kaprelyants, D. I. Young, M. Young & D. B. Kell, (1998) A bacterial cytokine. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 8916-8921.

Mukamolova, G. V., S. S. Kormer, D. B. Kell & A. S. Kaprelyants, (1999) Stimulation of the multiplication of Micrococcus luteus by an autocrine growth factor. *Archives of microbiology* **172**: 9-14.

Mukamolova, G. V., A. G. Murzin, E. G. Salina, G. R. Demina, D. B. Kell, A. S. Kaprelyants & M. Young, (2006) Muralytic activity of Micrococcus luteus Rpf and its relationship to physiological activity in promoting bacterial growth and resuscitation. *Molecular microbiology* **59**: 84-98.

Mukamolova, G. V., O. A. Turapov, K. Kazarian, M. Telkov, A. S. Kaprelyants, D. B. Kell & M. Young, (2002a) The rpf gene of Micrococcus luteus encodes an essential secreted growth factor. *Molecular microbiology* **46**: 611-621.

Mukamolova, G. V., O. A. Turapov, D. I. Young, A. S. Kaprelyants, D. B. Kell & M. Young, (2002b) A family of autocrine growth factors in Mycobacterium tuberculosis. *Molecular microbiology* **46**: 623-635.

Murialdo, H. & L. Siminovitch, (1972) The morphogenesis of bacteriophage lambda. IV. Identification of gene products and control of the expression of the morphogenetic information. *Virology* **48**: 785-823.

Murphy, K. C., (1998) Use of bacteriophage lambda recombination functions to promote gene replacement in Escherichia coli. *Journal of bacteriology* **180**: 2063-2071.

Murphy, K. C., K. G. Campellone & A. R. Poteete, (2000) PCR-mediated gene replacement in Escherichia coli. *Gene* **246**: 321-330.

Murray, N. E. & A. Gann, (2007) What has phage lambda ever done for us? *Curr Biol* **17**: R305-312.

Mushegian, A. R., K. J. Fullner, E. V. Koonin & E. W. Nester, (1996) A family of lysozyme-like virulence factors in bacterial pathogens of plants and animals. *Proceedings of the National Academy of Sciences of the United States of America* **93**: 7321-7326.

Muyrers, J. P., Y. Zhang, F. Buchholz & A. F. Stewart, (2000) RecE/RecT and Redalpha/Redbeta initiate double-stranded break repair by specifically interacting with their respective partners. *Genes & development* **14**: 1971-1982.

Muyrers, J. P., Y. Zhang & A. F. Stewart, (2001) Techniques: Recombinogenic engineering--new options for cloning and manipulating DNA. *Trends in biochemical sciences* **26**: 325-331.

Muyrers, J. P., Y. Zhang, G. Testa & A. F. Stewart, (1999) Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic acids research* **27**: 1555-1557.

Nakagawa, H., F. Arisaka & S. Ishii, (1985) Isolation and characterization of the bacteriophage T4 tail-associated lysozyme. *Journal of virology* **54**: 460-466.

Nambu, T., T. Minamino, R. M. Macnab & K. Kutsukake, (1999) Peptidoglycan-hydrolyzing activity of the FlgJ protein, essential for flagellar rod formation in Salmonella typhimurium. *Journal of bacteriology* **181**: 1555-1561.

Navarre, W. W. & O. Schneewind, (1999) Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiol Mol Biol Rev* **63**: 174-229.

Noirot, P. & R. D. Kolodner, (1998) DNA strand invasion promoted by Escherichia coli RecT protein. *The Journal of biological chemistry* **273**: 12274-12280.

Oppenheim, A. B., A. J. Rattray, M. Bubunenko, L. C. Thomason & D. L. Court, (2004) In vivo recombineering of bacteriophage lambda by PCR fragments and single-strand oligonucleotides. *Virology* **319**: 185-189.

Parish, T., E. Mahenthiralingam, P. Draper, E. O. Davis & M. J. Colston, (1997) Regulation of the inducible acetamidase gene of Mycobacterium smegmatis. *Microbiology (Reading, England)* **143 (Pt 7)**: 2267-2276.

Parrish, N. M., J. D. Dick & W. R. Bishai, (1998) Mechanisms of latency in Mycobacterium tuberculosis. *Trends in microbiology* **6**: 107-112.

Pavelka, M. S., Jr. & W. R. Jacobs, Jr., (1996) Biosynthesis of diaminopimelate, the precursor of lysine and a component of peptidoglycan, is an essential function of Mycobacterium smegmatis. *Journal of bacteriology* **178**: 6496-6507.

Pedersen, M., S. Ostergaard, J. Bresciani & F. K. Vogensen, (2000) Mutational analysis of two structural genes of the temperate lactococcal bacteriophage TP901-1 involved in tail length determination and baseplate assembly. *Virology* **276**: 315-328.

Pedulla, M. L., M. E. Ford, J. M. Houtz, T. Karthikeyan, C. Wadsworth, J. A. Lewis, D. Jacobs-Sera, J. Falbo, J. Gross, N. R. Pannunzio, W. Brucker, V. Kumar, J. Kandasamy, L. Keenan, S. Bardarov, J. Kriakov, J. G. Lawrence, W. R. Jacobs, Jr., R. W. Hendrix & G. F. Hatfull, (2003) Origins of highly mosaic mycobacteriophage genomes. *Cell* **113**: 171-182.

Pei, J. & N. V. Grishin, (2005) The P5 protein from bacteriophage phi-6 is a distant homolog of lytic transglycosylases. *Protein Sci* **14**: 1370-1374.

Pham, T. T., D. Jacobs-Sera, M. L. Pedulla, R. W. Hendrix & G. F. Hatfull, (2007) Comparative genomic analysis of mycobacteriophage Tweety: evolutionary insights and construction of compatible site-specific integration vectors for mycobacteria. *Microbiology (Reading, England)* **153**: 2711-2723.

Pisabarro, A. G., M. A. de Pedro & D. Vazquez, (1985) Structural modifications in the peptidoglycan of Escherichia coli associated with changes in the state of growth of the culture. *Journal of bacteriology* **161**: 238-242.

Piuri, M. & G. F. Hatfull, (2006) A peptidoglycan hydrolase motif within the mycobacteriophage TM4 tape measure protein promotes efficient infection of stationary phase cells. *Molecular microbiology* **62**: 1569-1585.

Rachman, H., M. Strong, T. Ulrichs, L. Grode, J. Schuchhardt, H. Mollenkopf, G. A. Kosmiadi, D. Eisenberg & S. H. Kaufmann, (2006) Unique transcriptome signature of Mycobacterium tuberculosis in pulmonary tuberculosis. *Infection and immunity* **74**: 1233-1242.

Raman, S., R. Hazra, C. C. Dascher & R. N. Husson, (2004) Transcription regulation by the Mycobacterium tuberculosis alternative sigma factor SigD and its role in virulence. *Journal of bacteriology* **186**: 6605-6616.

Ravagnani, A., C. L. Finan & M. Young, (2005) A novel firmicute protein family related to the actinobacterial resuscitation-promoting factors by non-orthologous domain displacement. *BMC genomics* **6**: 39.

Raymond, J. B., S. Mahapatra, D. C. Crick & M. S. Pavelka, Jr., (2005) Identification of the namH gene, encoding the hydroxylase responsible for the N-glycolylation of the mycobacterial peptidoglycan. *The Journal of biological chemistry* **280**: 326-333.

Rickman, L., C. Scott, D. M. Hunt, T. Hutchinson, M. C. Menendez, R. Whalan, J. Hinds, M. J. Colston, J. Green & R. S. Buxton, (2005) A member of the cAMP receptor protein family of transcription regulators in Mycobacterium tuberculosis is required for virulence in mice and controls transcription of the rpfA gene coding for a resuscitation promoting factor. *Molecular microbiology* **56**: 1274-1286.

Roessner, C. A. & G. M. Ihler, (1984) Proteinase sensitivity of bacteriophage lambda tail proteins gpJ and pH in complexes with the lambda receptor. *Journal of bacteriology* **157**: 165-170.

Romantschuk, M., V. M. Olkkonen & D. H. Bamford, (1988) The nucleocapsid of bacteriophage phi 6 penetrates the host cytoplasmic membrane. *The EMBO journal* **7**: 1821-1829.

Romeis, T. & J. V. Holtje, (1994) Specific interaction of penicillin-binding proteins 3 and 7/8 with soluble lytic transglycosylase in Escherichia coli. *The Journal of biological chemistry* **269**: 21603-21607.

Rossmann, M. G., V. V. Mesyanzhinov, F. Arisaka & P. G. Leiman, (2004) The bacteriophage T4 DNA injection machine. *Current opinion in structural biology* **14**: 171-180.

Rost, B., (1996) PHD: predicting one-dimensional protein structure by profile-based neural networks. *Methods in enzymology* **266**: 525-539.

Rost, B., P. Fariselli & R. Casadio, (1996) Topology prediction for helical transmembrane proteins at 86% accuracy. *Protein Sci* **5**: 1704-1718.

Rost, B. & C. Sander, (1993) Prediction of protein secondary structure at better than 70% accuracy. *Journal of molecular biology* **232**: 584-599.

Rost, B., G. Yachdav & J. Liu, (2004) The PredictProtein server. *Nucleic acids research* **32**: W321-326.

Rybalchenko, N., E. I. Golub, B. Bi & C. M. Radding, (2004) Strand invasion promoted by recombination protein beta of coliphage lambda. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 17056-17060.

Rydman, P. S. & D. H. Bamford, (2000) Bacteriophage PRD1 DNA entry uses a viral membrane-associated transglycosylase activity. *Molecular microbiology* **37**: 356-363.

Rydman, P. S. & D. H. Bamford, (2002) The lytic enzyme of bacteriophage PRD1 is associated with the viral membrane. *Journal of bacteriology* **184**: 104-110.

Rydman, P. S., J. Caldentey, S. J. Butcher, S. D. Fuller, T. Rutten & D. H. Bamford, (1999) Bacteriophage PRD1 contains a labile receptor-binding structure at each vertex. *Journal of molecular biology* **291**: 575-587.

Salas, M., R. P. Mellado & E. Vinuela, (1978) Characterization of a protein covalently linked to the 5' termini of the DNA of Bacillus subtilis phage phi29. *Journal of molecular biology* **119**: 269-291.

Sambrook, J., E. F. Fritsch & T. Maniatis, (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sarkis, G. J. & G. F. Hatfull, (1998) Mycobacteriophages. *Methods in molecular biology (Clifton, N.J* **101**: 145-173.

Scandella, D. & W. Arber, (1976) Phage lambda DNA injection into Escherichia coli pel- mutants is restored by mutations in phage genes V or H. *Virology* **69**: 206-215.

Schaffer, C. & P. Messner, (2005) The structure of secondary cell wall polymers: how Gram-positive bacteria stick their cell walls together. *Microbiology (Reading, England)* **151**: 643-651.

Schagger, H. & G. von Jagow, (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Analytical biochemistry* **166**: 368-379.

Schroeckh, V. & K. Martin, (2006) Resuscitation-promoting factors: distribution among actinobacteria, synthesis during life-cycle and biological activity. *Antonie van Leeuwenhoek* **89**: 359-365.

Sherman, D. R., M. Voskuil, D. Schnappinger, R. Liao, M. I. Harrell & G. K. Schoolnik, (2001) Regulation of the Mycobacterium tuberculosis hypoxic response gene encoding alpha -crystallin. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 7534-7539.

Shleeva, M., G. V. Mukamolova, M. Young, H. D. Williams & A. S. Kaprelyants, (2004) Formation of 'non-culturable' cells of Mycobacterium smegmatis in stationary phase in response to growth under suboptimal conditions and their Rpf-mediated resuscitation. *Microbiology (Reading, England)* **150**: 1687-1697.

Shleeva, M. O., K. Bagramyan, M. V. Telkov, G. V. Mukamolova, M. Young, D. B. Kell & A. S. Kaprelyants, (2002) Formation and resuscitation of "non-culturable" cells of Rhodococcus rhodochrous and Mycobacterium tuberculosis in prolonged stationary phase. *Microbiology (Reading, England)* **148**: 1581-1591.

Siboo, I. R., B. A. Bensing & P. M. Sullam, (2003) Genomic organization and molecular characterization of SM1, a temperate bacteriophage of Streptococcus mitis. *Journal of bacteriology* **185**: 6968-6975.

Snapper, S. B., L. Lugosi, A. Jekkel, R. E. Melton, T. Kieser, B. R. Bloom & W. R. Jacobs, Jr., (1988) Lysogeny and transformation in mycobacteria: stable expression of foreign genes. *Proceedings of the National Academy of Sciences of the United States of America* **85**: 6987-6991.

Snapper, S. B., R. E. Melton, S. Mustafa, T. Kieser & W. R. Jacobs, Jr., (1990) Isolation and characterization of efficient plasmid transformation mutants of Mycobacterium smegmatis. *Molecular microbiology* **4**: 1911-1919.

Stapleton, M. R., M. J. Horsburgh, E. J. Hayhurst, L. Wright, I. M. Jonsson, A. Tarkowski, J. F. Kokai-Kun, J. J. Mond & S. J. Foster, (2007) Characterization of IsaA and SceD, two putative lytic transglycosylases of Staphylococcus aureus. *Journal of bacteriology* **189**: 7316-7325.

Steinbacher, S., S. Miller, U. Baxa, N. Budisa, A. Weintraub, R. Seckler & R. Huber, (1997) Phage P22 tailspike protein: crystal structure of the head-binding domain at 2.3 A, fully refined structure of the endorhamnosidase at 1.56 A resolution, and the molecular basis of O-antigen recognition and cleavage. *Journal of molecular biology* **267**: 865-880.

Stover, C. K., V. F. de la Cruz, T. R. Fuerst, J. E. Burlein, L. A. Benson, L. T. Bennett, G. P. Bansal, J. F. Young, M. H. Lee, G. F. Hatfull & et al., (1991) New use of BCG for recombinant vaccines. *Nature* **351**: 456-460.

Struthers-Schlinke, J. S., W. P. Robins, P. Kemp & I. J. Molineux, (2000) The internal head protein Gp16 controls DNA ejection from the bacteriophage T7 virion. *Journal of molecular biology* **301**: 35-45.

Swaminathan, S., H. M. Ellis, L. S. Waters, D. Yu, E. C. Lee, D. L. Court & S. K. Sharan, (2001) Rapid engineering of bacterial artificial chromosomes using oligonucleotides. *Genesis* **29**: 14-21.

Thunnissen, A. M., A. J. Dijkstra, K. H. Kalk, H. J. Rozeboom, H. Engel, W. Keck & B. W. Dijkstra, (1994) Doughnut-shaped structure of a bacterial muramidase revealed by X-ray crystallography. *Nature* **367**: 750-753.

Thunnissen, A. M., N. W. Isaacs & B. W. Dijkstra, (1995) The catalytic domain of a bacterial lytic transglycosylase defines a novel class of lysozymes. *Proteins* **22**: 245-258.

Tsui, L. C. & R. W. Hendrix, (1983) Proteolytic processing of phage lambda tail protein gpH: timing of the cleavage. *Virology* **125**: 257-264.

Tufariello, J. M., W. R. Jacobs, Jr. & J. Chan, (2004) Individual Mycobacterium tuberculosis resuscitation-promoting factor homologues are dispensable for growth in vitro and in vivo. *Infection and immunity* **72**: 515-526.

Tufariello, J. M., K. Mi, J. Xu, Y. C. Manabe, A. K. Kesavan, J. Drumm, K. Tanaka, W. R. Jacobs, Jr. & J. Chan, (2006) Deletion of the Mycobacterium tuberculosis resuscitation-promoting factor Rv1009 gene results in delayed reactivation from chronic tuberculosis. *Infection and immunity* **74**: 2985-2995.

van Asselt, E. J., A. J. Dijkstra, K. H. Kalk, B. Takacs, W. Keck & B. W. Dijkstra, (1999) Crystal structure of Escherichia coli lytic transglycosylase Slt35 reveals a lysozyme-like catalytic domain with an EF-hand. *Structure* **7**: 1167-1180.

van Kessel, J. C. & G. F. Hatfull, (2007) Recombineering in Mycobacterium tuberculosis. *Nature methods* **4**: 147-152.

van Kessel, J. C. & G. F. Hatfull, (2008) Efficient point mutagenesis in mycobacteria using single-stranded DNA recombineering: characterization of antimycobacterial drug targets. *Molecular microbiology* **67**: 1094-1107.

Vegge, C. S., L. Brondsted, H. Neve, S. Mc Grath, D. van Sinderen & F. K. Vogensen, (2005) Structural characterization and assembly of the distal tail structure of the temperate lactococcal bacteriophage TP901-1. *Journal of bacteriology* **187**: 4187-4197.

Vollmer, W. & U. Bertsche, (2007) Murein (peptidoglycan) structure, architecture and biosynthesis in Escherichia coli. *Biochim Biophys Acta*.

Vollmer, W., D. Blanot & M. A. de Pedro, (2008a) Peptidoglycan structure and architecture. *FEMS Microbiol Rev*.

Vollmer, W. & J. V. Holtje, (2001) Morphogenesis of Escherichia coli. *Current opinion in microbiology* **4**: 625-633.

Vollmer, W., B. Joris, P. Charlier & S. Foster, (2008b) Bacterial peptidoglycan (murein) hydrolases. *FEMS Microbiol Rev* **32**: 259-286.

Votyakova, T. V., A. S. Kaprelyants & D. B. Kell, (1994) Influence of Viable Cells on the Resuscitation of Dormant Cells in Micrococcus luteus Cultures Held in an Extended Stationary Phase: the Population Effect. *Appl Environ Microbiol* **60**: 3284-3291.

Walderich, B. & J. V. Holtje, (1991) Subcellular distribution of the soluble lytic transglycosylase in Escherichia coli. *Journal of bacteriology* **173**: 5668-5676.

Weigel, C. & H. Seitz, (2006) Bacteriophage replication modules. *FEMS microbiology reviews* **30**: 321-381.

Weigle, J., (1966) Assembly of phage lambda in vitro. *Proceedings of the National Academy of Sciences of the United States of America* **55**: 1462-1466.

Wietzerbin, J., B. C. Das, J. F. Petit, E. Lederer, M. Leyh-Bouille & J. M. Ghuysen, (1974) Occurrence of D-alanyl-(D)-meso-diaminopimelic acid and meso-diaminopimelyl-meso-diaminopimelic acid interpeptide linkages in the peptidoglycan of Mycobacteria. *Biochemistry* **13**: 3471-3476.

Wilhelm, S. W., S. M. Brigden & C. A. Suttle, (2002) A dilution technique for the direct measurement of viral production: a comparison in stratified and tidally mixed coastal waters. *Microbial ecology* **43**: 168-173.

Wommack, K. E. & R. R. Colwell, (2000) Virioplankton: viruses in aquatic ecosystems. *Microbiol Mol Biol Rev* **64**: 69-114.

Xu, J., (2001) A Conserved Frameshift Strategy in dsDNA Long Tailed Bacteriophages. In: Department of Biological Sciences. Pittsburgh: University of Pittsburgh, pp.

Xu, J., R. W. Hendrix & R. L. Duda, (2004) Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. *Molecular cell* **16**: 11-21.

Yeremeev, V. V., T. K. Kondratieva, E. I. Rubakova, S. N. Petrovskaya, K. A. Kazarian, M. V. Telkov, S. F. Biketov, A. S. Kaprelyants & A. S. Apt, (2003) Proteins of the Rpf family: immune cell reactivity and vaccination efficacy against tuberculosis in mice. *Infection and immunity* **71**: 4789-4794.

Yu, D., H. M. Ellis, E. C. Lee, N. A. Jenkins, N. G. Copeland & D. L. Court, (2000) An efficient recombination system for chromosome engineering in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 5978-5983.

Yu, D., J. A. Sawitzke, H. Ellis & D. L. Court, (2003) Recombineering with overlapping single-stranded DNA oligonucleotides: testing a recombination intermediate. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 7207-7212.

Zhang, Y., F. Buchholz, J. P. Muyrers & A. F. Stewart, (1998) A new logic for DNA engineering using recombination in Escherichia coli. *Nature genetics* **20**: 123-128.

Zhang, Y., J. P. Muyrers, J. Rientjes & A. F. Stewart, (2003) Phage annealing proteins promote oligonucleotide-directed mutagenesis in Escherichia coli and mouse ES cells. *BMC molecular biology* **4**: 1.

Zhang, Y., Y. Yang, A. Woods, R. J. Cotter & Z. Sun, (2001) Resuscitation of dormant Mycobacterium tuberculosis by phospholipids or specific peptides. *Biochemical and biophysical research communications* **284**: 542-547.

Zimmer, M., E. Sattelberger, R. B. Inman, R. Calendar & M. J. Loessner, (2003) Genome and proteome of Listeria monocytogenes phage PSA: an unusual case for programmed + 1 translational frameshifting in structural protein synthesis. *Molecular microbiology* **50**: 303-317.

Zweig, M. & D. J. Cummings, (1973a) Cleavage of head and tail proteins during bacteriophage T5 assembly: selective host involvement in the cleavage of a tail protein. *Journal of molecular biology* **80**: 505-518.

Zweig, M. & D. J. Cummings, (1973b) Structural proteins of bacteriophage T5. *Virology* **51**: 443-453.