A multifaceted approach to gain insight into the mechanisms underlying novel phage-host interactions

by

Saeed Binsabaan

B.S., King Saud University, 2002

M.S., Michigan Technological University, 2014

Submitted to the Graduate Faculty of the

Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2023

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Saeed Binsabaan

It was defended on

May 25, 2023

and approved by

Graham Hatfull, Ph.D., Professor, Biological Sciences

Andrea Berman, Ph.D., Associate Professor, Biological Sciences

Robert Duda, Ph.D., Assistant Professor, Biological Sciences

Daria Van Tyne, Ph.D., Assistant Professor, Department of Medicine

Dissertation Director: Andrew VanDemark, Ph.D., Associate Professor, Biological Sciences

A multifaceted approach to gain insight into the mechanisms underlying novel phage-host interactions

Saeed Binsabaan, PhD

University of Pittsburgh, 2023

Mycobacteriophages are a group of bacteriophages that infect mycobacterial species. Metagenomic studies suggest that mycobacteriophage genomes are remarkably diverse. As of March 2023, 12,381 phages have been isolated from a single mycobacterial species, *Mycobacterium smegmatis* mc^2 155, and 2184 genomes have been completely sequenced. These genomes harbor an extensive reservoir of genes coding for proteins of unknown function, many of which have unique sequences that do not match any protein sequence in the available protein database. This vast array of unexplored proteins represents a great source to gain insight into the potential novel pathways by which phage proteins mediate the interaction with the host cell.

This thesis explores new mechanisms regulating phage-mycobacterial host interactions by investigating selected novel mycobacteriophage proteins. We selected Phaedrus gp82 as a candidate protein with no known function or structural homolog but is predicted to crystallize. We showed that Phaedrus gp82 is a toxic protein that severely reduces colony size when overexpressed in Mycobacterium smegmatis. This effect arises from the interaction of Phaedrus gp82 with an essential *M. smegmatis* protein, MoxR, a multifunctional ATPase known to have chaperone function. The structure of Phaedrus gp82 was solved using x-ray crystallography at 1.4 Å resolution revealing that the protein consists of two domains, the base and wing domain. The electron density map revealed that Phaedrus gp82 contains a disordered loop, which was critical

for MoxR binding. A mutation in the disordered loop, wing domain deletion, or in the acidic residue D38, E43, and D45 abolishes small colony phenotype.

We propose that Phaedrus gp82 functions by reducing the ATPase activity of MoxR, thereby reducing the levels of properly folded MoxR clients. The complete collection of proteins that require MoxR activity to achieve their folded state is unknown, but one example, the essential protein RipA is known. Therefore, we speculate that the toxicity of Phaedrus gp82 is due to the resulting decrease in properly folded MoxR clients. This work provides insight into potential new pathways governing phage-host interactions; and underlines the importance of mycobacteriophage genomics as a promising tool to identify new drug targets for the treatment of mycobacterial pathogens.

Table of Contents

1.0 Introduction 1

1.1 Bacteriophages 1

1.1.1 Phage life cycle 2

1.1.2 Phage-bacteria co-evolution 4

1.1.3 Phage-host interactions 7

1.1.4 Phage-host range 10

1.1.5 Phage Research – Applications 12

1.2 Mycobacteriophage 13

1.2.1 Genetic diversity 14

1.2.2 Genome organization 15

1.2.3 Mycobacteriophage genomes encode for cytotoxic proteins 18

1.3 Tuberculosis 21

1.3.1 An Overview of the current situation 21

1.3.2 Tuberculosis challenges 22

1.3.3 Mycobacteriophage-guided Identification of potential novel drug targets 24

1.4 Superinfection exclusion 25

1.5 MoxR family ATPases 27

1.6 Thesis plan 30

2.0 Phaedrus gp82: A novel mycobacteriophage protein interacts with the host multifunctional MoxR ATPase 32

2.1 introduction 32

2.2 Results 34

2.2.1 Selection of Phaedrus gp82 34

2.2.2 Overexpression of Phaedrus gp82 reduces *M. smegmatis* colony size 36

2.2.3 The crystal structure of Phaedrus gp82 reveals two distinct domains 38

2.2.4 Phaedrus gp82 is a trimer in solution 42

2.2.5 Identifying residues and regions within P82 that are important for the small colony phenotype. 44

2.2.6 Phaedrus gp82 interacts with *M. smegmatis* MoxR ATPase 47

2.2.7 Connection between MoxR function and Phaedrus gp82 49

2.3 Discussion 51

2.4 Material and Methods 57

2.4.1 Bioinformatics 57

2.4.2 Protein Expression and Purification 58

2.4.3 Crystallization and structure determination 61

2.4.4 Cellular localization 62

2.4.5 Cytotoxicity plate assay and quantification 63

2.4.6 Chemical crosslinking 64

2.4.7 Growing mycobacteria and media 64

2.4.8 Pull-down assay and mass spec identification of binding partners 65

2.4.9 Colorimetric Determination of ATPase Activity 66

2.5 Supplemental data 67

3.0 Adephagia gp73: A potential anti-sigma factor 71

3.1 Introduction 71

3.2 Results 73

3.2.1 Crystal structure of Adephagia gp73 73

3.2.2 Adephagia gp73 is a dimer in solution 76

3.2.3 AlphaFold model of Adephagia gp73-host sigma factor 79

3.3 Discussion 82

3.4 Material and Methods 84

3.4.1 Protein purification 84

3.4.2 Crystallization and structure determination 85

3.4.3 Chemical Crosslinking 86

3.4.4 Analytical Size Exclusion Chromatography 86

4.0 Discussion and future directions 87

4.1 Phaedrus gp82 project 88

4.2 Adephagia gp73 project 92

Bibliography 93

List of Tables

Table 1. Data collection and refinement statistics for P82 41

Table 2. The other 11 highest protein candidates resulted from the bioinformatic pipeline57

Table 3. Data collection and refinement statistics (molecular replacement)..... 76

Table 4. The calculated physicochemical properties of the predicted AphaFold models of Adephagia gp73 in complex with 7 different host sigma factors. 81

List of Figures

Figure 1. Phage life cycle 3

Figure 2. Genome organization of mycobacteriophage Pipefish 18

Figure 3. Hexameric structure of MoxR ATPase 28

Figure 4. Gene location of MoxR ATPase in *M. smegmatis* 29

Figure 5. Selection of the candidate protein and the gene location within the phage genome 36

Figure 6. Overexpression of Phaedrus gp82 reduces colony size in *M. smegmatis* 38

Figure 7. Structure and sequence conservation of Phaedrus gp82. 40

Figure 8. Phaedrus gp82 is a trimer both in solution and within the crystal. 43

Figure 9. Identification of functionally important surfaces on Phaedrus gp82. 46

Figure 10. Phaedrus gp82 interacts with MoxR ATPase. 49

Figure 11. Phaedrus gp82 interacts with MoxR ATPase. 51

Figure 12. Predicted AlphaFold structure of Phaedrus gp82-MoxR monomer-monomer interaction 56

Figure 13. Structure of Phaedrus gp82 trimer and predicted MoxR hexamer and their relative sizes 56

Figure 14. Purification of Phaedrus gp82 - First nickel column 59

Figure 15. Purification of Phaedrus gp82 - Second nickel column 59

Figure 16. Purification of Phaedrus gp82 - Size exclusion chromatography column using Hiload S200 16/600 60

Figure 17. Phaesus gp82 - Pre and post induction and purified protein 60

Figure 18. Phaedrus gp82 crystal 62

Figure S1. Phaedrus gp82 localized to the host cytoplasm..... 67

Figure S2. Phaedrus gp82 disordered loop (DL) mutant is a trimer. 68

Figure S3. Phaedrus gp82 interacts with MoxR ATPase. 69

Figure S4. Identification of proteins that bind Phaedrus gp82. 70

Figure S5. Standard curve for the detection of orthophosphate using the malachite green assay. 70

Figure 19. Crystal structure of Adephagia gp73..... 74

Figure 20. Electrostatic surface potential of Adephagia gp73 75

Figure 21. Superposition of Adephagia gp73 with anti-sigma factor PupR..... 75

Figure 22. Crystal packing of Adephagia gp73..... 78

Figure 23. Adephagia gp73 oligomerization evaluated by chemical crosslinking 78

Figure 24. Adephagia gp73 oligomerization evaluated by analytical size exclusion chromatography..... 79

Figure 25. Superposition of Adephagia gp73 with anti-sigma factor RsiW from Bacillus subtilis. 81

1.0 Introduction

1.1 Bacteriophages

Bacteriophages are viruses that are capable of infecting and multiplying within bacterial cells. They are ubiquitous, living in a variety of environments wherever the bacterial host exists, including soil and marine environments (Ashelford et al., 2003; Wichels et al., 1998), as well as in extreme environmental conditions (Nuttall & Dyall-Smith, 1993; Prigent et al., 2005). Bacteriophages are incredibly abundant; it has been estimated that there are ~10^{31} total phage particles on our planet, which outnumbers the estimated total of microbial cells by 10-fold (Bergh et al., 1989; Wommack & Colwell, 2000). They exist in various morphologies, but the vast majority (~ 96%) are tailed double-stranded DNA bacteriophages belonging to the Caudovirales order, divided into three families: Myoviridae, Podoviridae, and Siphoviridae (Ackermann, 2007; Fokine & Rossmann, 2014). Bacteriophages play a vital role in the global ecosystem (Suttle, 2007), evolution and bacterial diversity (Canchaya et al., 2004; Fortier & Sekulovic, 2013; Ohnishi et al., 2001), and can mediate horizontal gene transfer between bacteria (Brussow et al., 2004; Kidambi et al., 1994). These viruses have contributed to the development of many molecular biology techniques and improved our understanding of many cellular processes (Keen, 2015; Kropinski, 2018).

1.1.1 Phage life cycle

Phages are obligate parasites that can adopt two major lifestyles: lytic and lysogenic. The Phage infection process involves five stages: adsorption, penetration, hijacking the host metabolism, virions assembly, and cell lysis (Guttman et al., 2005). To initiate the infection of the bacterial host, the phage will first recognize and bind to specific receptors on the host cell's surface (Dennehy & Abedon, 2020; Letarov & Kulikov, 2017; Moldovan et al., 2007). Successful binding to these receptors leads to irreversible phage adhesion followed by penetrating the cell envelope (Letarov & Kulikov, 2017). The penetration process is facilitated by depolymerases (hydrolases and lyases) residing in the phage tail. These degrade the peptidoglycan layer and subsequently form pores in the cell wall (Fernandes & Sao-Jose, 2018; Pires et al., 2016; Yan et al., 2014). This process eventually allows the phage to release its genome into the host cytoplasm. Afterward, replication of the phage genome will depend on the phage type, lytic or temperate. The host metabolism will be redirected to the phage's benefit in the case of lytic phage. This switch typically occurs when the host RNA polymerase recognizes the phage genome (Guttman et al., 2005; Miller et al., 2003). This allows the expression of early phage proteins, which redirect host cell machinery into replicating phage genome and phage protein synthesis. The new phage particles are then assembled into mature phage. The tailed phages encode for a cell lysis system mediated by proteins called holins that destroy the cell membrane, allowing the endolysins to reach and hydrolyze the peptidoglycan layer to lyse the cell (Ackermann, 1998). New phage progenies are then released from the lysed cell (Weinbauer, 2004).

Temperate phage, conversely, can enter either the lytic cycle or integrate their genome into the host chromosome to enter the lysogenic life cycle (Howard-Varona et al., 2017). The phage genome in lysogeny, known as a prophage, is replicated together with the host genome. However,

under stress conditions, the prophage can excise from the host chromosome and enter the lytic cycle, producing new progenies and eventually the host cell death. The steps of lytic and lysogenic cycles are depicted in (Figure 1).

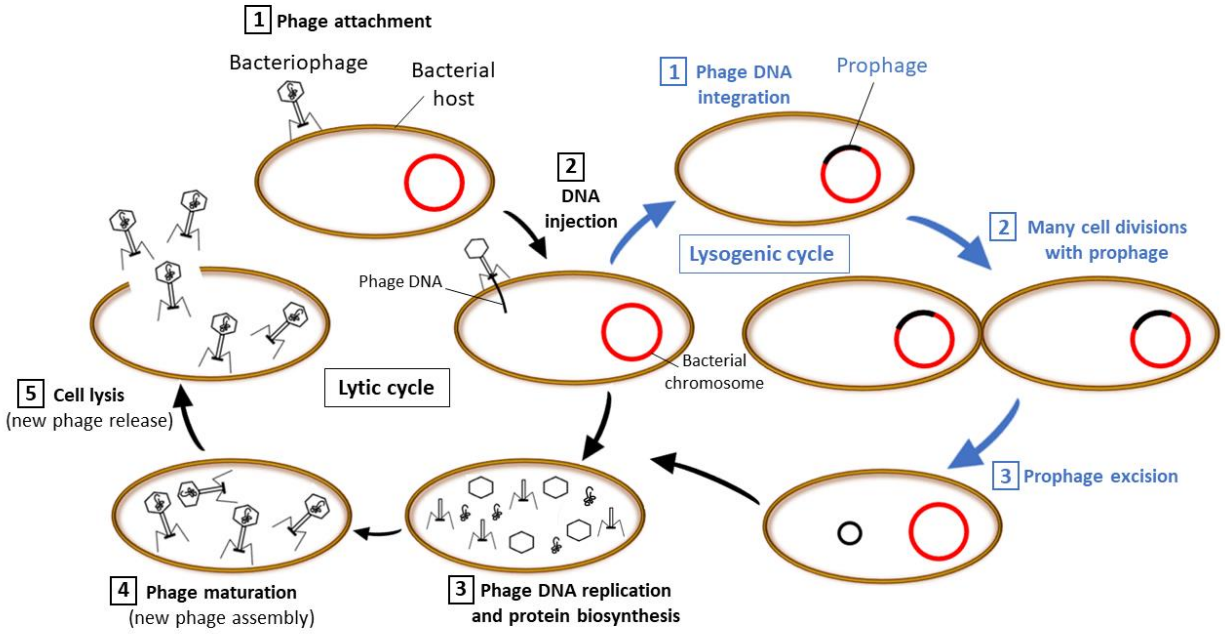

Figure 1. Phage life cycle

Phage life cycle starts with a successful attachment into the host cell receptors. After injecting the DNA into the host cell, the lytic phage takes over the host cell machineries to replicate the phage DNA and synthesize phage proteins. The new phage particles are assembled, and the host cell is lysed to release the new phage progenies. Temperate phage can enter either lytic or lysogenic cycle. In the lysogenic cycle, the phage DNA is integrated into the host chromosome which termed prophage. The prophage replicated along with the host chromosome. Under stressful conditions, prophage is excised from the host chromosome and enter the lytic cycle.

1.1.2 Phage-bacteria co-evolution

The temperate phage can either enter the lytic cycle or integrate its DNA into the host genome and becomes a prophage in the lysogenic cycle. The prophage remains integrated with the host genome and replicates with it as the bacterial cell divides. The lytic gene expression of the prophage is silent, repressed by CI repressor (CI repressor proteins in phage λ), until an environmental signal induces the lytic cycle (Canchaya et al., 2003). However, some prophage genes can be expressed, which may provide beneficial phenotypes to the bacterial host (Brussow et al., 2004). Prophages largely contribute to the evolution and virulence of many bacterial pathogens, such as *E. coli* (Ohnishi et al., 2001), *Salmonella enterica* (Cooke et al., 2007), and *Staphylococcus aureus* (Bae et al., 2006). For example, many prophages code for virulence factors, including cholera toxins (Waldor & Mekalanos, 1996), Diphtheria toxins (Gill et al., 1972), and botulism (Eklund et al., 1971). Other prophages harbor genes encoding for fitness factors that enhance the adaptation of the bacteria to the host, including effector proteins involved in bacterial invasion, enzymes such as superoxide dismutase and phospholipase, serum resistance proteins, and adhesion factors (Brussow et al., 2004). It was estimated that half of the bacterial genomes contain a minimum of one prophage and sometimes up to 20 prophages (Touchon et al., 2016), which may suggest that lysogeny significantly impacts bacterial evolution.

Horizontal gene transfer (HGT) is a process by which genetic information is transferred from one cell to another (Ochman et al., 2000). Phages continuously mediate HGT with an estimated global rate of 2×10^{16} gene transfer events per second, further underlining phages' influence on bacterial evolution (Bushman, 2002). Temperate phages mediate HGT between bacteria through two primary mechanisms, lysogenic conversion and transduction (Touchon et al., 2017). The lysogenic conversion allows bacteria to acquire genes that enhance their survival fitness

(Hartley et al., 2012). This process facilitates the dissemination of virulence factors that contribute to bacterial pathogenicity (Brussow et al., 2004). Transduction, on the other hand, occurs when a piece of bacterial DNA is packaged into the phage particle, which then can be delivered to another bacteria (Zinder & Lederberg, 1952). Transduction is regarded as one of the major processes contributing to the spread of antibiotic-resistance genes (von Wintersdorff et al., 2016).

As a natural predator of bacteria, phage infection imposes tremendous pressure on the bacterial host, enhancing natural selection. The Red Queen Hypothesis suggests that organisms should continuously evolve to ensure their fitness to improve their chances of survival when encountering predators (Valen, 2014). This implies that bacteria constantly evolve anti-phage defense systems while phages adapt to circumvent the bacterial defense to ensure successful infection. This relationship is termed the "evolutionary arms race" and has generated a variety of defense mechanisms through a very long period of coevolution. Therefore, bacteria-phage co-evolution is motivated by co-adaptation that promotes mutual proliferation.

Bacteria have developed defense mechanisms that can interfere with all stages of phage infection. These mechanisms include preventing phage adsorption on the surface of the host cell. Bacteria use different strategies to avoid phage adsorption, such as mutation in the phage receptors or blocking the phage receptors by producing an extracellular matrix (Labrie et al., 2010). For example, *Pseudomonas aeruginosa* modifies its surface by glycosylation of type IV pilus which blocks the adsorption of pilus-specific phages (Harvey et al., 2018). Blocking the injection of phage DNA into the host cytoplasm is another bacterial mechanism to prevent phage infection (Labrie et al., 2010). For example, *E. coli* prophage HK97 encodes a superinfection exclusion protein (gp15) that blocks the injection of DNA from phages HK97 and HK75 (Cumby et al., 2012). Bacteria also use restriction-modification (RM) systems to cleave the phage genome (Tock

& Dryden, 2005). RM system involves bacterial restriction endonuclease activity to cleave the phage DNA and methyltransferase activity to methylate its genome to avoid being recognized by endonucleases. Moreover, bacteria protect themselves against invading phages through the CRISPR-Cas (clustered regularly interspaced short palindromic repeats–CRISPR-associated proteins) system, which uses the memory of past infections to efficiently cleave the phage DNA (Barrangou et al., 2007). CRISPR-Cas system is classified into two classes and six types and involves different Cas nucleases with various mechanisms of action (Koonin et al., 2017). When these defense mechanisms fail to prevent the phage infection, the abortive infection (Abi) system will be activated as the last resort of defense, which triggers the cell death pathway to restrict phage replication and thereby protect the local bacterial population (Fineran et al., 2009). All the mechanisms mentioned above belong to the innate immune systems except CRISPR-Cas, considered the only adaptive immune system known in bacteria (Bikard & Marraffini, 2012).

However, phages have shown a unique adaptability to evade bacterial defense systems. This includes strategies that allow phage to gain access to the host cell receptors to avoid adsorption prevention (Samson et al., 2013). For example, the λ phage initiates infection through binding of its tail fiber (protein J) to the outer membrane receptor protein (LamB) of *E. coli*. The *E. coli* prevents the infection by modifying LamB; however, a single mutation in the protein J allows the λ phage infection by binding to another receptor, OmpF. (Meyer et al., 2012). Phages evolved strategies to evade bacterial RM systems (Tock & Dryden, 2005). This can be seen in the anti-RM system used by phage T4. The phage T4 replaces the cytosine in the restriction recognition sites with an unusual base hydroxymethyl cytosine (HMC) to be undetected by the bacterial RM system (O'Farrell et al., 1980). Phages can bypass the CRISPR-Cas system using anti-CRISPR (Acr) proteins. For example, five Acr (AcrIF1-5) proteins that inhibit the type I-F

CRISPR-Cas system were found in phages that infect *Pseudomonas aeruginosa* (Bondy-Denomy et al., 2013). Phages can escape the bacterial abortive infection system by, for example, producing antitoxin proteins (Samson et al., 2013), such as T4 phage antitoxin (Dmd) that directly interact with *Escherichia coli* (LsoA) and (RnlA) toxins and inhibits their toxicity (Otsuka & Yonesaki, 2012). Altogether, the evolution and diversity of defense and counter-defense mechanisms have been driven by selection pressure; potentially, many more defense systems remain undiscovered.

1.1.3 Phage-host interactions

The infection initiation of tailed phages requires primary contact with the bacterial host, which is made through the phage adsorption to the host cell surface. This interaction occurs between the phage receptor binding proteins (RBPs) located on the tail fibers and specific receptors on the bacterial surface. Various components of bacterial surface can serve as receptors for phage adsorption such as proteins, polysaccharides, lipopolysaccharide (LPS), pili, and flagella (Juliano Bertozzi Silva et al., 2016; Rakhuba et al., 2010). Once the phage tail fibers successfully attach to bacterial receptors, phage depolymerases start to degrade the cell wall polysaccharide, followed by the injection of phage DNA into the host cytoplasm (Fernandes & Sao-Jose, 2018; Pires et al., 2016; Yan et al., 2014). The lytic phage DNA will then be recognized by the host RNA polymerase, and the host metabolism will be redirected to the phage DNA replication and protein synthesis.

The phage genome encodes for a variety of proteins that interfere with host cell machinery processes including replication, transcription, translation, and cell division (Drulis-Kawa et al., 2012). Phage-host protein-protein interactions occur in all stages of the phage life cycle; however, the majority of these interactions take place during the early stages of phage infection (Miller et

al., 2003). It was estimated that the phage early proteins represent approximately 64% of phage-host protein-protein interactions (Roucourt & Lavigne, 2009). Most of the phage early proteins are relatively small, less than 250 amino acids in length, and many of them have little or no sequence homology to any known protein (Miller et al., 2003). Moreover, some of these proteins are highly toxic to the host cell, which emphasizes the importance of unraveling their functional mechanisms to develop novel antibacterial strategies (Liu et al., 2004).

Phage early proteins can directly interact with the host replication proteins such as DNA replication initiation protein DnaG, helicase loader DnaI, and the sliding clamp DnaN (Liu et al., 2004). In contrast to the temperate phages which depend on the host cell replication machinery (Friedman et al., 1984), the genome of lytic phages generally encodes for their replication proteins such as phage T4 (Mueser et al., 2010). However, in both scenarios, the host replication machinery will be taken over by the phage to shut it off or switch it to its own genome replication. For example, gp025, gp168, and gp240 of *Staphylococcus aureus* phage have been shown to interact with the host DnaN (Liu et al., 2004). The phage gp016 and gp104 interact with DnaI, while gp078 interacts with DnaG (Liu et al., 2004). These interactions with the essential host proteins DnaN, DnaI, and DnaG lead to the inhibition of DNA replication machinery. The temperate λ phage protein P interacts with an essential replication protein DnaB of *E. coli* (Mallory et al., 1990). This interaction leads to blocking the host replication machinery via inhibiting the ATPase activity of DnaB and suppressing its ability to bind DNA primase. The host transcription machinery is also a target for phage early proteins. Through the phage-host co-evolution, phages have developed different mechanisms to shut off, inhibit, or redirect the bacterial RNA polymerase complex (RNAP). Some phage genomes encode for σ factor that competes with and eventually replaces the host σ factor in the RNAP complex (Drulis-Kawa et al., 2012). Thus, the transcription of the phage

early genes will be initiated by the host RNAP complex that incorporates the phage σ factor subunit. The whole transcription machinery can also be taken over by the phage, such as phage T4, and redirects it to its own gene transcription. Moreover, some phages can shut off the host transcription machinery and use their encoded RNAP to transcribe their genes. For example, the phage T7 depends on the host RNAP to transcribe a few early genes that later in the infection cycle shut off the host RNAP. The T7 protein kinase gp0.7 was found to phosphorylate the β' subunit of the RNAP and many other host proteins (Robertson & Nicholson, 1990). This phosphorylation together with the direct binding of T7 gp2 to the host RNAP leads to an efficient transcriptional termination of the host RNAP and switches it to phage T7-dependent transcription for the rest of the genes (Nechaev & Severinov, 1999; Robertson & Nicholson, 1990).

Translation of phage mRNA typically depends on highly abundant host ribosomes (Nechaev & Severinov, 2008). However, evidence indicates that the phage proteins can interact with and influence the host proteins that are required for translation. For example, the genome of phage T4 encodes for the ADP-ribosyltransferases Alt and ModB, that ADP-ribosylate 27 and 8 host proteins, respectively (Depping et al., 2005). Some of these modified proteins are required for the host cell translation. Another example of host protein modification is the protein kinas gp0.7 of phage T7. The gp0.7 was shown to phosphorylate 90 bacterial host proteins, and 7 of these proteins are implicated in the translation process of the host cell (Robertson et al., 1994; Robertson & Nicholson, 1990, 1992). However, the consequences of this modification on the host cell are not well understood.

Phages also use the host cell division machinery as a target through protein-protein interactions. FtsZ is an essential bacterial protein that requires a proper assembly, a ring-shaped structure, to achieve its function in cell division (Bi & Lutkenhaus, 1991; de Boer, 2010). The

expression of Kil protein from prophage λ has been shown to inhibit bacterial cell division by preventing the ring formation of FtsZ (Haeusser et al., 2014). Many other metabolic pathways can be influenced by phage-encoded proteins. The SPO1 phage of *Bacillus subtilis* encodes for the Gp60 protein which was identified as a phage-encoded enolase inhibitor protein (PEIP) (Zhang et al., 2022). Expression of PEIP in the host cell disrupts the proper oligomerization of the host enolase leading to loss of enolase catalytic activity. Enolase catalyzes the reaction in the penultimate step of glycolysis which converts 2-phosphoglycerate (2-PG) to phosphoenolpyruvate (PEP). This inhibits the production of phosphoenolpyruvate (PEP) which is involved in peptidoglycan biosynthesis (Barreteau et al., 2008). These metabolic defects resulted in a growth reduction of PEIP-expressing cells and thinner cell walls.

1.1.4 Phage-host range

The genetic material of a bacteriophage is packaged into the capsid which is attached to its tail (Ackermann & Prangishvili, 2012). The tailed bacteriophages belong to the *Caudovirales* order have a genome of a double-stranded DNA and are predominant among other phages (Ackermann, 2007). The tail is a special part of the phage and plays a crucial role during the phage infection process. It involves the recognition of the bacterial host, penetration of the cell envelope, and providing access to the phage genome to transfer into the host cytoplasm (Nobrega et al., 2018). The recognition of the host cell is mediated by a specific interaction between the tail fiber or tail spike proteins and specific receptors on the surface of the bacterial cell (J. Bertozzi Silva et al., 2016; Dunne et al., 2018; Letarov & Kulikov, 2017). The phage-host range (host specificity) is mainly determined by the tail fibers; and the variety of tail fibers enables the phage to be efficiently recognized and adsorbed by diverse bacterial hosts (Arnaud et al., 2017; Hu et al., 2015).

Because the interaction between the tail fibers and the bacterial receptors is highly specific, phage infection will be limited to a narrow host range. Studies of phage-bacteria co-evolution demonstrated that the phage-host range could be changed by the occurrence of mutations in either the receptor binding proteins (RBPs) located on the phage tail or in the phage receptors on the bacterial surface (Duplessis et al., 2006; Perry et al., 2015). In fact, a point mutation can be sufficient to alter the phage-host range. For example, a mutant of mycobacteriophage Rosebush enabled an efficient infection of *Mycobacterium smegmatis* Jucho as compared to wild-type phage (Jacobs-Sera et al., 2012). Expanding the host range here was facilitated by a point mutation that led to a substitution of leucine for arginine residue at position 297 of gp32, which was identified as a tail fiber protein (Jacobs-Sera et al., 2012).

In the field of phage therapy, high specificity in phage infection is desirable. Phages that infect only a specific species of bacteria, in principle, are harmless to the benign commensal bacteria as compared to antibiotics which can be active against a broad range of bacterial species. However, the limited host range bears the disadvantage that the specific strain within the infecting bacterial species must be determined before proceeding with bacteriophage monotherapy trials. This can be overcome by developing a combination of well-characterized phages, each infecting a specific strain. This strategy allows us to determine which phage will be efficient for the specified infection (Kingwell, 2015; Merril et al., 2003). Furthermore, a cocktail of phages that infect a wide range of bacterial strains within one species is another approach to overcome the limited host range. Although this approach has proven successful, the major drawback is the complexity surrounding the isolation and production process (Ling et al., 2022). Thus, there is a growing interest in bioengineering phages with modified receptor-binding proteins to control the phage

specificity and avoid the complications related to the phage cocktail approach (Ando et al., 2015; Chen et al., 2017).

1.1.5 Phage Research – Applications

The long and distinguished history of bacteriophage research has contributed to advances in molecular biology and biotechnology. Bacteriophages have been used for decades as the primary tool to answer and clarify many questions in different fields of science and technology. After their discovery in the early 1900s, many scientists speculated about the potential implications of bacteriophage as antimicrobial therapeutic agents. Bacteriophages were successfully used for the first therapeutical attempt soon after their discovery to treat a 12-year-old boy suffering from dysentery (Sulakvelidze et al., 2001). The recent advancements in our understanding of bacteriophage diversity and evolution have led to a great interest in exploiting these particles in many aspects of medical and biological research.

The predatory ability of phages to infect and lyse a specific bacterial host represents an advantage in combating certain bacterial infections in humans or in treating environments contaminated with harmful bacteria. Although using phages as therapeutic agents were reported in the early 20th century, this strategy was generally abandoned after the discovery of antibiotics in the 1940s. However, the emergence of antibiotic resistance in bacteria in the last two decades encouraged the scientific community to return to phage therapy as an alternative to antibiotics (Hermoso et al., 2007). As a result, phage therapy is flourishing again and undergoing continuous development in many laboratories and medical institutions. Although many phage treatment successes have been reported (Dedrick et al., 2019; Djebara et al., 2019; Ferry, Boucher, et al., 2018; Ferry, Leboucher, et al., 2018; Jennes et al., 2017; Strathdee et al., 2019), some limitations

are present which could potentially impede the progress in the field of phage therapy (Lin et al., 2022; Loc-Carrillo & Abedon, 2011; Nilsson, 2019; Thiel, 2004).

Besides phage therapy, bacteriophages have been used in many applications, including biocontrol agents in food production and processing (Flaherty et al., 2001; Goode et al., 2003; Leverentz et al., 2003; Modi et al., 2001), phage typing (Mohammed, 2017; Rabsch, 2007), phage display (Kehoe & Kay, 2005; Smith & Petrenko, 1997), and vaccine delivery system (Clark & March, 2004; Irving et al., 2001; Piekarowicz et al., 2022). Similar to phage therapy, concerns exist about using phages in these applications, specifically safety and efficiency concerns. However, tremendous progress in molecular biology and biotechnology might bring about solutions that resolve such concerns in the near future.

1.2 Mycobacteriophage

Mycobacteriophages are a group of bacteriophages that are known to infect mycobacteria. Over 12,381 mycobacteriophages have been isolated from a single mycobacterial species, *Mycobacterium smegmatis*, while a small number of mycobacteriophages have also been isolated from other mycobacterial hosts (Russell & Hatfull, 2017). Approximately 18% (2230) of the isolated phages have been completely sequenced and grouped into clusters and singletons (Pope et al., 2015a; Russell & Hatfull, 2017). The sequenced genomes suggest that mycobacteriophages are genetically diverse and harbor a large repertoire of genes of unknown function. Thus, mycobacteriophages represent a valuable resource for understanding phage evolution and to gain deeper insight into the physiology of their mycobacterial host.

1.2.1 Genetic diversity

The first report of a completely sequenced mycobacteriophage genome was released in 1993 when a comprehensive study of phage L5 was carried out (Hatfull & Sarkis, 1993). Similarly, subsequent studies were conducted on phages D29, TM4, and Bxb1, and the complete genome was sequenced for each (Ford, Sarkis, et al., 1998; Ford, Stenstrom, et al., 1998; Mediavilla et al., 2000). These studies provided a glimpse into the genome structure and evolution of mycobacteriophages, but little is known about their diversity. However, the last 15 years have witnessed significant progress in phage genomics, which have dramatically augmented our understanding of these phages and their diversity. Such progress is reflected in the number of phages that have been isolated in addition to the number of genomes that have been fully sequenced. These data, as well as other related information can be found in the phage database website at phagesdb.org.

As of March 2023, 22,574 phages have been isolated from different bacterial genera within the phylum *Actinobacteria*. Among this number, 12,381 phages have been isolated from a single mycobacterial species, *Mycobacterium smegmatis* mc^2 155, and 2184 genomes have been completely sequenced (Pope et al., 2015a; Russell & Hatfull, 2017). This collection of sequenced mycobacteriophage genomes exhibited remarkable genetic diversity and provided a clearer insight into genome evolution. Comparison of the whole genome sequences suggests that many phage genomes share patches of closely related or identical sequences. Phage genomes were thus sorted into clusters according to their nucleotide sequence similarity (Hatfull et al., 2010; Pope et al., 2015; Russell & Hatfull, 2017). The number of newly sequenced phage genomes is increasing continuously. As compared to when 30 sequenced genomes are available (Hatfull et al., 2006), it became evident that many clusters can be further divided into subclusters according to the average

nucleotide identity (ANI) comparisons. Phage genomes with no sequence or very low similarity to others are called singletons. This grouping resulted so far in a total of 31 clusters and 10 singletons (Pope et al., 2015; Russell & Hatfull, 2017).

The tremendous genetic diversity of mycobacteriophages is further manifested by comparing the sequences of the genome-encoded proteins. This was achieved by sorting genes into phamilies (Hatfull et al.) according to the similarity of the gene product sequences using Phamerator software (Cresawn et al., 2011). The available sequenced mycobacteriophage genomes include 246,272 genes that have been sorted into 7798 protein Pham (Cresawn et al., 2011). Interestingly, among these protein phams, there are 171267 (69.5%) proteins of unknown functions (Pope et al., 2015). This suggests that mycobacteriophage genomes harbor a huge reservoir of gene products with no known function and no or very low homology to proteins of known structure in the available protein database. It is, perhaps, the largest collection of unknown function proteins reported so far, not to mention the fact that the sequenced genomes here represent only a small proportion of the entire phage population. Thus, this huge number of unexplored phage proteins represents a valuable platform to identify potentially novel protein families, novel protein folds, and novel mechanisms by which phages regulate the bacterial host.

1.2.2 Genome organization

The genome size of Mycobacteriophages is diverse, ranging from 38 – 165 kbp, with an average of 68 kbp and an average of 100 open reading frames per genome (Cresawn et al., 2011; Russell & Hatfull, 2017). Despite the extreme diversity of mycobacteriophages, their genomes share common features in the way they are organized. All mycobacteriophages have siphoviral morphology except phages in cluster C which have myoviral morphology. The virion structure

and assembly, genes of all Siphoviral mycobacteriophages, are syntenic with a conserved order and located in the left arm of the genome (Hatfull, 2014; Graham F. Hatfull, 2022). These structural genes occupy a region that spans 20 – 25 kbp of the genome and is transcribed rightward (Hatfull, 2018). The virion structure and assembly genes are similar in length to the equivalent bacterial genes. However, the average gene length of mycobacteriophages is about 600 bp which is much smaller than the bacterial gene (Hatfull et al., 2010; Russell & Hatfull, 2017). This suggests that mycobacteriophage genomes are replete with small genes coding for non-structural proteins.

The non-structural genes are located in the right arm of the genome. This part of the genome has an abundance of relatively small genes, most of which have no known functions (Hatfull, 2018). A few genes with predictable functions can be seen among these small genes, and those usually involved in nucleotide metabolism or DNA replication (Hatfull, 2014). The number of genes is dependent on the genome size, ranging from 25-30 genes in the smallest genomes, such as clusters G, N, and T, to 150 genes in the large genomes like cluster J (Hatfull, 2018). As previously demonstrated in phage Giles, many of these genes are probably not required for lytic growth (Dedrick et al., 2013). However, some of these genes are toxic when expressed in the bacterial host. This has been shown, for example, in gp2 and gp0.7 of phage T7, which are both involved in the shut off of the host RNA polymerase activity (Hesselbach & Nakada, 1977) and many other examples that will be discussed in the following section.

Lytic gene expression of mycobacteriophages have been characterized for many phages in different Clusters (Dedrick, Jacobs-Sera, et al., 2017; Dedrick et al., 2013; Dedrick, Mavrich, et al., 2017; Fan et al., 2016; Halleran et al., 2015; Ko & Hatfull, 2018). Typically, transcription occurs in two stages of infection, early and late gene transcription. The early genes are transcribed in the first 30 minutes of infection. The non-structural genes are generally transcribed in this stage.

However, the virion structure and assembly genes are transcribed late, approximately 30 minutes post-infection, and proceed for 3 hours (Hatfull, 2018).

One distinct feature of mycobacteriophage genomes is their mosaic nature (Pedulla et al., 2003). Each genome is built from a collection of specific units, and each unit is composed of a cluster of genes or can be as little as one gene (Hatfull, 2018). Given the extreme diversity of mycobacteriophages, these units can be compiled in a wide variety of ways to form a unique genome. Comparing unrelated genomes at their amino acids sequence level revealed that those genomes share some genes in common. However, the related genes are positioned in different locations within the genome and flanked by different genes. While the exact mechanism behind genome mosaicism is not completely clear, this phenomenon potentially originated from illegitimate recombination between genomes (Hendrix et al., 1999; Pedulla et al., 2003).

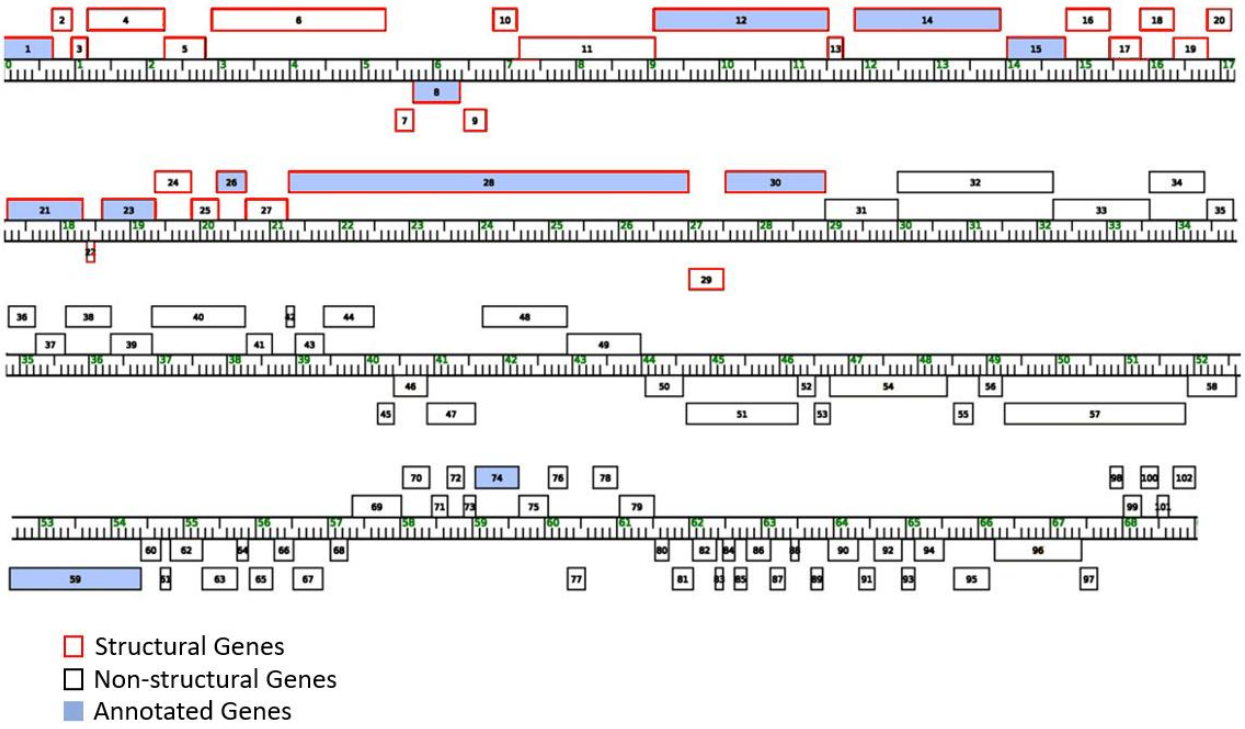

Figure 2. Genome organization of mycobacteriophage Pipefish

The genome map of mycobacteriophage Pipefish of subcluster B3 as an example of how mycobacteriophage genome organized. Each box represents a gene with its designated number. The structural genes are located in the left arm of the genome while the non-structural genes are in the right arm. Annotated genes are shown in light blue boxes. The right arm is enriched with small genes most of which have unknown function.

1.2.3 Mycobacteriophage genomes encode for cytotoxic proteins

Through billions of years of evolution, bacteriophage have developed a variety of mechanisms to ensure a successful infection. Phage genomes appear to encode non-structural specific proteins that can bind the host target and inhibit or shut off essential cellular processes. Indeed, some phage-encoded proteins can be lethal when expressed in the host cell. Therefore, characterizing phage genomes is a useful strategy to gain information about the encoded proteins

and identify their bacterial host partners to develop novel drugs against pathogens. This strategy has been utilized to develop antimicrobial drugs against *Staphylococcus aureus* by investigating protein-protein interactions between *S. aureus* phage and the host cell (Liu et al., 2004). From 964 open reading frames of *S. aureus* phage genomes, 31 novel proteins have been identified as growth inhibitors to the host cell.

Cytotoxic genes have also been reported in mycobacteriophages. Three gene products, gp77, gp78, and gp79 of phage L5 were identified as cytotoxic when expressed in the host cell, *M. smegmatis* (Rybniker et al., 2008). The gene product of gp77 was demonstrated to interact with MSMEG_3532, pyridoxal-5′-phosphate-dependent L-serine dehydratase (SdhA), which catalyzes the conversion of L-serine to pyruvate (Rybniker et al., 2011). However, the effect of this interaction on the host cell is not clear. Overexpression of gp79 in *M. smegmatis* induced cell filamentation and appeared to influence cell division (Rybniker et al., 2008). Another study showed that 32 genes of mycobacteriophage Waterfoul (Cluster K) are cytotoxic when expressed in the host cell *M. smegmatis* (Heller et al., 2022). This number of genes represents 34% of the phage Waterfoul genome, and 17 genes out of 32 are of unknown function. The effect of expression of these genes varied between the reduction of colony sizes and moderate or severe growth inhibition.

A large screening study has been conducted on 193 proteins encoded by 13 different mycobacteriophage genomes to examine their impact on expression in the host *M. smegmatis* (Ko & Hatfull, 2020). The study showed that 45 proteins (23.3%) have cytotoxic effects when expressed in the host cell. The majority of these toxic proteins (30 proteins) are of unknown function. The severity of the toxic effect on the host cell growth ranged from mild to severe, with 22 proteins out of 45 scored as highly toxic. Most of the highly toxic proteins gave rise to cellular

morphological alterations such as filamentation, bulging at cell poles, and curving. This large screen suggests that about 20% of phage-encoded proteins are toxic when expressed in the host cell.

The toxicity conferred by the overexpression of phage proteins does not adequately represent the real conditions in the phage environmental niches when phage infection occurs. Phages need the host cell machinery to complete a successful infection and, in any case, lytic growth will result in cell lysis. Thus, it was suggested that this toxicity results from the interaction with and inactivation of specific host proteins involved in cellular pathways that are essential for other phages to infect the same bacteria (Graham F. Hatfull, 2022; Ko & Hatfull, 2018, 2020). A typical example of such a mechanism is gp52 of mycobacteriophage Fruitloop (Cluster F) (Ko & Hatfull, 2018). Gp52 interacts with and inactivates an essential mycobacterial host, *M. smegmatis*, cell wall biosynthesis protein Wag31 (DivIVA), causing a severe reduction in the host cell viability. Both gp52 and Wag31 localize to the cell pole (Ko & Hatfull, 2018; Meniche et al., 2014). This interaction between gp52 and Wag31 prevents superinfection by other phages in Subcluster B2, such as Rosebush and Hedgerow, which depend on Wag31 for their infection. This study proposed that Wag31 plays a role in the DNA injection process of phages in Subcluster B. Therefore, this may suggest that the toxicity resulting from inactivating Wag31 is not as advantageous as excluding the other phages from superinfection, given that gp52 is not essential for the lytic growth of Fruitloop (Ko & Hatfull, 2018). This may support the hypothesis that many bacteriophage early proteins are required only in specific environmental conditions by providing a selective advantage for the phage (Miller et al., 2003).

1.3 Tuberculosis

1.3.1 An Overview of the current situation

In 1882, Robert Koch discovered the causative agent of human tuberculosis (TB), *Mycobacterium tuberculosis* (MTB) (Koch, 1982). Although more than a century has passed since this discovery was reported, TB still represents a major threat to human health globally. Human TB is an ancient disease that has coexisted with humans for perhaps thousands of years. Evidence showed that 9000-year-old MTB was detected in the skeletal remains of a woman and infant found in a Neolithic settlement in the Eastern Mediterranean (Hershkovitz et al., 2008) Tuberculosis is generally caused by a group of genetically related mycobacterium species termed *Mycobacterium tuberculosis* complex (MTBC). The MTBC species include *M. tuberculosis, M. africanum, M. canettii, M. pinnipedii, M. microti, M. mungi, M. caprae, and M. bovis* (Smith et al., 2009). Among these species, *M. tuberculosis* and *M. africanum* are the main species known to infect humans with TB, although the infection with *M. africanum* is rare, and most cases occurred in Western Africa (Gehre et al., 2016; Tientcheu et al., 2016). The rest are mainly animal-adapted species; however, humans can be infected by *M. bovis* through meat consumption or dairy products of the infected animals (Cosivi et al., 1998).

Until the emergence of coronavirus (COVID-19), an infectious disease caused by the SARS-CoV-2 virus, TB was the leading cause of global mortality due to a single infectious pathogen, ranking above that caused by HIV/AIDS (Global Tuberculosis Report, 2022). According to the World Health Organization (WHO) global tuberculosis report of 2022, the estimated global number of people who developed TB in 2021 is 10.6 million, and 1.6 million died of the disease in the same year. The WHO Global Tuberculosis Program has made progress toward ending TB as a global

health epidemic through the End TB strategy. This strategy aims to end the global TB problem. By the year 2035, the strategy aims to achieve a 95% reduction in the number of TB deaths and a 90% reduction in the number of new TB cases as compared to 2015. In recent years, however, COVID-19 has negatively affected TB diagnosis and treatment, and the progress towards achieving the goal of the End TB strategy is off track.

1.3.2 Tuberculosis challenges

Tubercle bacilli are transmitted between people through the air via droplets of respiratory secretions of the infected person coughing, sneezing, or speaking. MTB can then reach the lung and eventually reside in the alveolar macrophages (Houben et al., 2006; Warner & Mizrahi, 2007). Macrophages are phagocytic immune cells that can take up pathogens and destroy them (Galli & Saleh, 2020). However, MTB prefers to inhabit these cells. Thus, the success of MTB as a pathogen is due to its remarkable ability to replicate in macrophages and escape their powerful capability of destroying microbes. Typically, the phagocytic process is activated by recognizing and interacting with microbes through specific receptors on the cell surface (DesJardin et al., 2002; Rajaram et al., 2010). Phagocytic cells' structure changes once they engulf the pathogen to form an early phagosome. A maturation process then occurs in which the phagosome fuses with a lysosome to generate a mature phagolysosome (Welin & Lerm, 2012). MTB, however, interferes with the phagosome maturation process and can block phagosome-lysosome fusion (Meena & Rajni, 2010). The prevention of phagosome maturation is a central strategy for MTB to survive and replicate in the macrophages.

In macrophages, the pathogens are exposed to an aggressive environment such as acidic pH, reactive oxygen intermediate ROI, and reactive nitrogen intermediate RNI, which is optimal

for pathogen destruction (Ehrt & Schnappinger, 2009). However, MTB uses different mechanisms to circumvent the macrophage defenses by deploying a variety of factors that counteract ROI, RNI, and acidic macrophage environment. For example, MTB catalase-peroxidase encoded by KatG gene that decomposes H_2O_2 into H_2O and O_2 (Ng et al., 2004), superoxide dismutases that convert superoxide to hydrogen peroxide (Piddington et al., 2001), an antioxidant mycothiol (MSH) that protects against oxidative stress (Buchmeier et al., 2006; Newton et al., 2008), methionine sulphoxide reductase (Msr) as a repair system that converts methionine sulfoxide generated by ROI and RNI back to methionine (Lee et al., 2009), and both a putative porin OmpATb (Raynaud et al., 2002) and a membrane-associated protein *Rv3671c* (Vandal et al., 2008) that protect MTB against macrophage acidic environment.

The MTB replication in macrophages leads to active disease in approximately 10% of cases (Dye et al., 1999), while the immune response restricts the growth of the bacilli in the remaining cases. However, the bacilli are destroyed in only about 10% of the infected people who manifest restricted growth of MTB, while the bacilli enter a dormant state in the remaining cases (Al-Humadi et al., 2017). The dormant MTB is reactivated with any deficiency in the immune system of the infected individuals, such as individuals with HIV infection, diabetes, AIDS, malnutrition, or other causes (Corbett et al., 2003; Dooley & Chaisson, 2009; Frieden et al., 2003; Wells et al., 2007). Therefore, the deficiency of the immune system represents a major factor in controlling TB infection. For example, due to the opportunistic nature of MTB, HIV patients co-infected with TB are 18 times more likely to develop an active TB disease than non-HIV-infected individuals (Holzheimer et al., 2021). This situation makes TB disease a major threat to human health, requiring different diagnosis and treatment strategies (Montales et al., 2015).

The TB problem is aggravated by the emergence of multi-drug resistant tuberculosis (MDR-TB) and even worsened by extensively drug-resistant tuberculosis (XDR-TB). Both MDR-Tb and XDR-Tb complicate the challenges of coping with the TB pandemic and limit successful treatment (Pontali et al., 2019). Moreover, the MTB has a unique cell wall structure consisting of a thick layer of glycolipids, polysaccharides, and peptidoglycans which provides a barrier against TB drugs (Britton & Triccas, 2008; Ghazaei, 2018; Jankute et al., 2015; Kaur et al., 2009). Thus, besides the ability of MTB to remain dormant within macrophages, the TB treatment regimen of MDR-Tb and XDR-Tb can take several years. This prolonged treatment may give rise to non-compliance with therapy which represents an additional challenge for effective global TB control.

1.3.3 Mycobacteriophage-guided Identification of potential novel drug targets

Bacteriophages have played an important role in our understanding of cell biology on a molecular level. Research into bacteriophages has led to significant discoveries including random mutations of bacteria (Luria & Delbrück, 1943), a demonstration that DNA is the genetic material (Hershey & Chase, 1952), the understanding of gene regulation mechanisms (Jacob & Monod, 1961), as well as their significant role in the advancement of the biotechnology field. Despite such important contributions to our understanding of many biological processes, only a few studies have been conducted to investigate how bacteriophages might reveal new pathways for the development of new antibacterial drugs.

The emergence of MDR-Tb and XDR-Tb as well as the other challenges that hinder effective global TB management have urgently called for alternative ways to develop strategies and innovative solutions to cope with the TB pandemic. As a natural mycobacterial killer, mycobacteriophage could potentially be used to guide the search for the development of new anti-

mycobacterial agents. Given the enormous genetic diversity of mycobacteriophages and the large reservoir of their gene products that have unknown functions, the chances to identify new drug targets are very high. Many proteins encoded by mycobacteriophage genomes are toxic or have an inhibitory impact when overexpressed in the mycobacterial host (Ko & Hatfull, 2018, 2020). The interaction partners of these phage proteins have been also identified; however, very few have been subjected to detailed investigation. It should be noted that these phage proteins represent only a tiny fraction of the large pool of unknown function proteins. This emphasizes the importance of mycobacteriophage as a promising tool for understanding of mycobacterial physiology and identifying potential novel pathways that can be utilized as drug targets.

1.4 Superinfection exclusion

Phage infection represents a major threat to bacterial population. Phage-host evolution has led bacteria to develop different mechanisms to fight against phage infection. These mechanisms provide immunity against phages and work at different stages of the infection cycle. Bacteriophages, on the other hand, have evolved multiple counter-defense systems to combat bacterial defense and ensure successful infection.

Many bacteriophages have developed mechanisms that provide immunity to the infected cell from subsequent infection via superinfection exclusion (Sie) system. Superinfection exclusion can be defined as the process by which the infecting phage blocks secondary infection by the same or closely related phage, therefore preventing the competition for the host cell resources (Abedon, 2015; Folimonova, 2012). This mechanism can protect phage genome by preventing the replication of other phage genomes withing the same bacterial cell, thus decreasing the chance of

recombination or reassortment events of phage genes (Folimonova, 2012). Superinfection exclusion system is achieved through proteins encoded by bacteriophage genome; and usually encoded by prophages. Most of the reported superinfection exclusion systems are proteins encoded by prophages; however, lytic phages , such as T4, are also able to encode proteins involved in this system (van Houte et al., 2016).

Phages use this mechanism either in the early stage of infection by blocking phage adsorption, or in the subsequent stage by preventing the transfer of phage DNA into the host cytoplasm (Labrie et al., 2010). For example, phage T5 encodes a lipoprotein (Llp) that prevents subsequent phage adsorption event by blocking phage receptor, ferrichrome-iron receptor (FhuA) on the outer membrane (Labrie et al., 2010). This mechanism is also advantageous for the newly synthesized T5 virions by preventing their binding to the free receptors from lysed cells, and thus protect virions from inactivation. *Escherichia coli* phage T4 encodes two proteins, Imm and Sp, that are involved in superinfection exclusion system by blocking DNA injection into the host cell, thus preventing superinfection by the same or closely related phages (Lu & Henning, 1994). This system inhibits the degradation of the cell peptidoglycan layer which impedes DNA access to the cell cytoplasm.

Superinfection exclusion system is also found in mycobacteriophages. It was discovered through screening for mycobacteriophage proteins that are toxic when overexpressed in the host cell, *M. smegmatis* (Hatfull, 2018). The study revealed that mycobacteriophage Fruitloop gp52 of Cluster F is lethal when overexpressed in the host. Further investigation demonstrated that Fruitloop gp52 interacts with and inactivates an essential host protein, Wag31 (DivIVA) (Ko & Hatfull, 2018). Wag31 has been shown to localize to the cell pole (Meniche et al., 2014). Fruitllop gp52 expression inhibits the host cell infection by two phages that belong to Subcluster B2,

Hedgerow and Rosebush (Ko & Hatfull, 2018). The study suggests that Hedgerow and Rosebush phages depend on the host Wag31 for their efficient infection, and that Fruitloop gp52 excludes superinfection by the inactivation of Wag31.

The WhiB-like protein encoded by mycobacteriophage TM4 has been shown to be involved in superinfection exclusion. Overexpression of WhiB in *M. smegmatis* is lethal, however, basal expression from leaky vector showed TM4 resistance phenotype, suggesting its role in superinfection exclusion (Rybniker et al., 2010). Another example is the putative membrane protein gp32 encoded by mycobacteriophage Charlie of cluster N. It has been proposed that gp32 confers heterotypic superinfection exclusion of phage Che9c by preventing DNA entry into the host cell (Dedrick, Jacobs-Sera, et al., 2017).

1.5 MoxR family ATPases

MoxR is a member of a large family of ATPases Associated with various cellular Activities (AAA+ proteins), which belong to the P-loop NTPases superfamily. AAA+ proteins are involved in a variety of biological processes, including protein folding, protein degradation, DNA repair, and replication (Iyer et al., 2004). AAA+ proteins contain AAA+ modules: this module is 200 to 250 residues and includes conserved sequence motifs, Walker A and Walker B, that bind and hydrolyze ATP molecules to generate the energy required to achieve the molecular remodeling of the substrate (Neuwald et al., 1999; Ogura & Wilkinson, 2001). AAA+ ATPases generally operate as an oligomer, usually as a hexameric ring (Hanson & Whiteheart, 2005) (Figure 3).

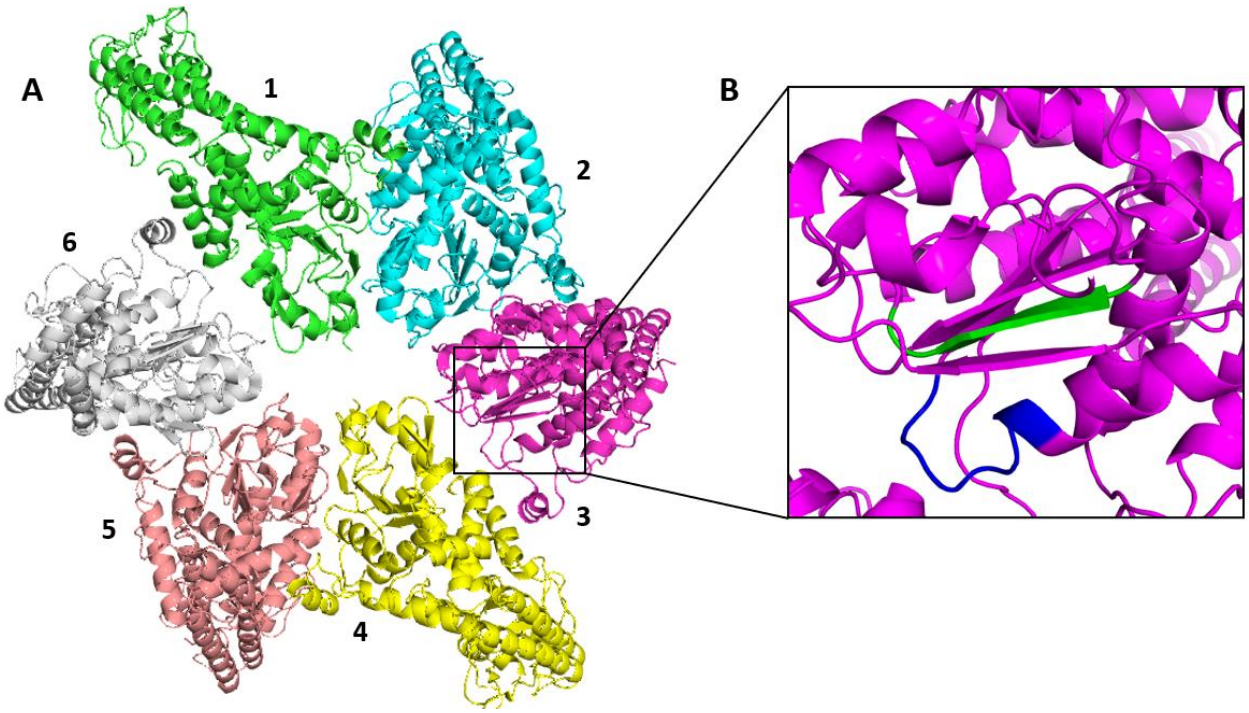

Figure 3. Hexameric structure of MoxR ATPase

(A) An example of MoxR ATPase (RavA) from *E. coli* shows the assembly of 6 subunits to form a hexameric ring-shaped structure. **(B)** Walker A (blue color) and Walker B (green color) motifs that bind and hydrolyze ATP molecule.

MoxR ATPases are prevalently found in bacteria and archaea; however, only a few members of this family have been characterized, and limited information is available about their biological function (Iyer et al., 2004; Snider & Houry, 2006). Bioinformatic analysis of 596 MoxR protein sequences showed that they can be classified into 7 subfamilies (Snider & Houry, 2006). These subfamilies include MoxR Proper (MRP), RavA, TM0930, APE2220, CGN, PA2707, and YehL. MoxR genes are commonly found near genes encoding for proteins that contain Von Willebrand Factor Type A (VWA) domain (Snider & Houry, 2006) (Figure. 4). The VWA-domain containing proteins are not well characterized in prokaryotes; however, these proteins are known to mediate protein-protein interactions (Springer, 2006; Whittaker & Hynes, 2002). A key element

of the VWA domain is the presence of a conserved sequence motif (DxSxS….T….D) called the metal ion-dependent adhesion site (MIDAS). This motif binds a metal ion, usually Mg^{2+}, which is critical for the function of the VWA domain (Whittaker & Hynes, 2002). Due to the proximity of the genes encoding for MoxR and VWA domain-containing protein, it was suggested that they potentially function together (Snider & Houry, 2006).

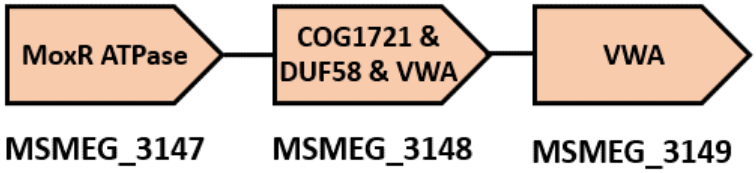

Figure 4. Gene location of MoxR ATPase in *M. smegmatis*

A representation of gene location of MoxR ATPase (MSMEG_3147) in *M. smegmatis* which belongs to MRP subfamily. MoxR gene is adjacent to two genes coding for Von Willebrand Factor Type A (VWA) domain-containing proteins, MSMEG_3148 and MSMEG_3149.

Based on the functionally characterized MoxR ATPases and their relationship with VWA domain-containing protein, it appears that VWA domain-containing protein functions as an adaptor that assists the localization of MoxR to its target protein. For example, MoxR in *E. coli* (RavA) and its adaptor VWA-containing protein (ViaA) (Wong et al., 2017); MoxR in *Paracoccus denitrificans* (NorQ) and its adaptor VWA-containing protein (NorD) (Kahle et al., 2018); and MoxR in *Acidithiobacillus ferrooxidans* (CbbQ) and it adaptor VWA-containing protein (CbbO) (Tsai et al., 2020).

Several previous studies suggest that MoxR ATPases have a chaperone-like function that is important for the assembly of protein complexes and maturation of specific proteins. For example, a member of the MRP subfamily in *Paracoccus denitrificans* is involved in the

maturation process of methanol dehydrogenase (MDH) (Van Spanning et al., 1991), while MoxR in *Methylobacterium extorquens* was also found to have a function associated with the maturation of MDH by inserting Ca^{2+} into the enzyme (Richardson & Anthony, 1992). Another example is MoxR1 in *Mycobacterium tuberculosis* which is involved in the correct folding of resuscitation-promoting factor (Rpf)-interaction protein A (RipA), which results in the secretion of RipA through the twin-arginine translocation (TAT) secretion system (Bhuwan et al., 2016).

1.6 Thesis plan

The main theme of my thesis is to provide experimental evidence that shows that mycobacteriophage genomes are potentially replete with genes encoding proteins that have novel folds and mediate the interaction with the host cell in a novel mechanism. The findings in this study do not necessarily generalize to the entire phage genomes; however, considering the remarkable genetic diversity of mycobacteriophages and the fact that they harbor a reservoir of genes of unknown functions, it would be rational to hypothesize that phage-host interactions are regulated by a variety of mechanisms, many of which are waiting to be discovered. In this context, mycobacteriophage genomes represent a treasure for researchers to discover new pathways governing phage-host interaction and learn about their potential applications in different fields of science and technology.

In this document, I included four chapters. In the first chapter, I provided a brief background about bacteriophages, focusing on mycobacteriophages. The second chapter discusses my major study, my pilot study with Phaedrus gp82. This used a bioinformatics pipeline designed by Dr. Andrew VanDemark to filter 214,000 actinobacteriophage proteins. Among this number of

proteins, we selected Phaedrus gp82, which has a novel sequence that does not match any sequence in the available protein database but was predicted to crystallize. Furthermore, Phaedrus gp82 is encoded by the genome of mycobacteriophage Phaedrus of Cluster B. I used an approach that includes biochemical, structural, and cell biology methods to investigate how such a novel protein might interact with the host cell.

The third chapter covers the findings on the crystal structure of Adephagia gp73 and its related biochemical and bioinformatics analysis. Adephagia gp73 is encoded by the mycobacteriophage Adephagia genome of Cluster K. Phage Adephagia has been shown to efficiently infect *Mycobacterium tuberculosis*. A large set of proteins encoded by phage Adephagia have been investigated for their cytotoxicity when expressed in the host cell, *Mycobacterium smegmatis* MC^2 155. No conclusive finding about the toxicity of Adephagia gp73 since the protein was not expressed in the host cell. However, our bioinformatics analysis suggests that Adephagia gp73 is potentially an anti-sigma factor. In the last chapter, I provide a discussion about this study's findings and future directions.

2.0 Phaedrus gp82: A novel mycobacteriophage protein interacts with the host multifunctional MoxR ATPase

Work in this section will be submitted for publication after my defense.

2.1 introduction

As multidrug-resistant bacterial infections rise worldwide, it is critical to develop novel antimicrobial strategies. Although conventional methods of antibiotic development have stalled in recent years, breakthrough discoveries in bacteriophage research are driving new innovations. Over billions of years of co-evolution with bacteria, these viruses have fine-tuned a staggeringly large and mechanistically diverse set of strategies to infect and kill their hosts. In addition to straightforward infection and lysis of their hosts, phages also express cytotoxic proteins which kill or slow the growth of bacteria when endogenously expressed. Identifying these proteins and their host interaction targets and elucidating their mechanisms of toxicity can shape the development of new antimicrobial approaches. This strategy has been employed to identify antimicrobial proteins in bacteriophages, including the mycobacteriophages (Ko & Hatfull, 2020; Liu et al., 2004; Shibayama & Dabbs, 2011).

Mycobacteriophages currently comprise the world's largest collection of isolated phages infecting a single bacterial host: *Mycobacterium smegmatis*. Over 2,000 distinct mycobacteriophage genomes have been sequenced and annotated and are publicly available at phagesdb.org (Russell & Hatfull, 2017). These phages have been bioinformatically classified into

dozens of *clusters* with nucleotide similarity. Their gene products are sorted into tens of thousands of *phamilies* (groups of phage proteins with amino acid similarity). Functions are known for less than half of these protein phamilies, with the remainder forming a vast reservoir of potential new antimicrobial proteins waiting to be discovered. In this context, we report Phaedrus gp82, a novel mycobacteriophage-encoded protein with an inhibitory effect on *Mycobacterium smegmatis* mc^2 155.

Phaedrus gp82 was discovered through a mycobacteriophage protein screening pipeline, which started with bioinformatic selection to enrich for phages proteins likely to crystallize. Crystallization and collection of diffraction data facilitated generation of a *de novo* atomic model of the protein and analysis of the resulting structure. In parallel, microbiological investigation revealed that endogenous expression of Phaedrus gp82 results in a tiny colony phenotype in *M. smegmatis* mc^2 155. Co-immunoprecipitation in *M. smegmatis* revealed that Phaedrus gp82 binds to host protein MoxR (MSMEG_3147), a multifunctional ATPase family protein whose ortholog in *M. tuberculosis* is essential for the proper folding of TB virulence factor and cell division protein RipA (Bhuwan et al., 2016; Healy et al., 2020). We suggest that Phaedrus gp82 protein functions to deplete levels of *M. smegmatis* MoxR protein, indirectly reducing levels of essential proteins such as RipA that depend on MoxR's chaperone activity to achieve their folded state, and thereby decreasing cellular growth. This growth defect can be overcome either by plasmid-driven overexpression of MoxR or by structural mutation of Phaedrus gp82 to prevent binding. Since RipA is essential for efficient cell division in *M. tuberculosis* and its depletion increases the bacterium's sensitivity to multiple cell-wall targeting drugs (Healy et al., 2020), indirect depletion of RipA by Phaedrus gp82-mediated MoxR inactivation could present a pathway for TB

therapeutics. This study is therefore an example of the power of structural and microbiological analyses of phage-encoded proteins for drug discovery and innovation.

2.2 Results

2.2.1 Selection of Phaedrus gp82

In an effort to increase the probability that we would successfully be able to characterize a novel phage protein using both structural and functional assays, we developed a sequence selection pipeline to enrich for sequences likely to crystallize that did not have any homologs of known function or structure. First, we used successive screening of the entire proteome using 1) the pham report on phageDB (Russell & Hatfull, 2017); 2) BLASTP on phagesDB; and 3) BLASTP on NCBI (Camacho et al., 2009) to remove proteins with already annotated functions. These include structural proteins such as capsid and portal, but also regulatory proteins such as integrase, cro, etc. After culling already annotated sequences, the remaining candidates were screened with HHPred (Hildebrand et al., 2009; Zimmermann et al., 2018) to find remote homologs; and we excluded those proteins with E-value $<10^{-3}$ and probability 70%. ~148,000 of our ~214,000 protein sequences remained after this step.

To further enrich for sequences more likely to crystalize and to facilitate a common purification scheme, we discarded proteins with biochemical properties as predicted by ProtParam (Wilkins et al., 1999) outside the following ranges: $80 \leq$ amino acids ≤ 400, $2.0 \leq$ pI ≤ 6.5, $-0.55 \leq$ GRAVY index ≤ -0.1, instability index ≤ 35, % coiled-coils $\leq 5\%$, # cysteine residues ≤ 6, and $15\% \leq$ % charged residues $\leq 30\%$. We screened the remaining sequences for transmembrane helices using

TMHMM (Krogh et al., 2001; Sonnhammer et al., 1998) and for signal peptides (Petersen et al., 2011), excluding protein sequences with either of these features. These parameters were designed to optimize the likelihood of the protein being highly expressed, soluble in neutral aqueous media, stable, and amenable to crystallization. To ensure maximum flexibility in crystallographic structure determination process, we selected proteins with ≥ 1% methionine in their primary sequence to facilitate phasing via selenomethionine should it become necessary. 2,095 proteins met all these criteria. From there we further reduced the number of candidates by removing orphams (proteins with no homologs in the phage protein collection which are enriched in protein sequences containing sequencing errors), and removing left-arm genes (which, while unannotated, are likely to be structural). To further enrich for crystallizable sequences, we next required surface entropy tofall within -1.25 to -1.00, as calculated by XtalPred3 (Slabinski et al., 2007a; Slabinski et al., 2007b) and choosing a single representative from each protein phamily. (99 candidates), manually verifying lack of sequence-based homology with BLASTP (Camacho et al., 2009) and HHPred (Zimmermann et al., 2018) searches (71 candidates), eliminating proteins from phages infecting non-mycobacterial hosts (44 candidates), and finally choosing only proteins with a PPCpred (Mizianty & Kurgan, 2011; Mizianty & Kurgan, 2012) crystallization propensity > 0.8 (12 targets). Among this list of 12 targets is Phaedrus gp82, a protein that has no homolog of known function, but was strongly predicted to crystalize. The other 11 highest candidates are shown in (Table 2). The bioinformatic pipeline for selecting Phaedrus gp82 is depicted in (Figure 5A).

Phage Phaedrus is a lytic mycobacteriophage belonging to B3 sub-cluster and is a member of *Siphoviridae* family (Pope et al., 2015a; Pope et al., 2015b). It has a genome size of 68,090 bp and contains 98 genes, 79 of which encode a protein without a currently known function (Russell &

Hatfull, 2017). As shown in (Figure5), these genes of unknown function are enriched within the right hand region of the genome (Figure 5B), and are typically smaller than their counterparts of known function. Phaedrus gp82 is located at the extreme right end of the genome among a cluster of 25 small genes whose functions are nearly all unknown (Figure 5B). While not a part of our selection criteria, we note here that there is only 1 gene with an annotated function within proximity of Phaedrus gp82 and thus it would be very difficult to predict the function of Phaedrus gp82 *a priori*. Therefore, we conclude that Phaedrus gp82 is an excellent candidate to test the hypothesis that we can use functional and structural analysis to reveal new host-phage interactions.

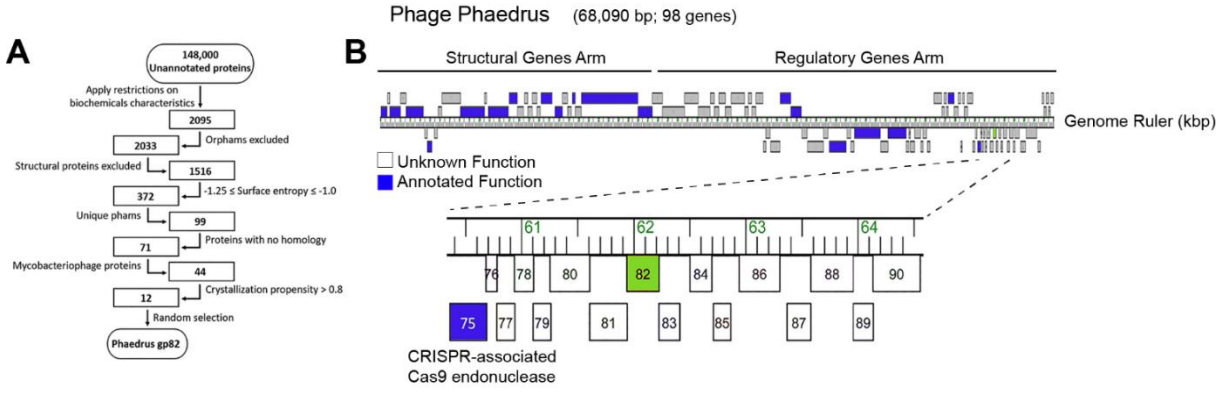

Figure 5. Selection of the candidate protein and the gene location within the phage genome

Flowchart describing the bioinformatic pipeline for the selection of Phaedrus gp82. The bioinformatic procedures used to select Phaedrus gp82 are indicated at each step. B) Phage Phaedrus genome and the location of Phaedrus gp82. The gene encoding Phaedrus gp82 is located at the extreme right arm of the genome among a cluster of small genes of unknown function.

2.2.2 Overexpression of Phaedrus gp82 reduces *M. smegmatis* colony size

To gain more insight into the biological function of P82, we first asked where the P82 protein localizes within its mycobacterial host. To do this, we inserted a fusion gene of mCherry

tagged Phaedrus gp82 into pKF8 and transformed the resulting plasmid into *M. smegmatis* mc^2 155. Cultures were grown to OD_{600}~0.6 and then induced and time points were taken over a 24-hour induction period. mCherry-P82 was found to be homogeneously diffuse throughout the host cytoplasm (Figure S1). In parallel, we performed the same experiments on the following control proteins: mCherry-Cuco gp50 fusion (UniProt ID G1JUM5), known to us to localize to the host cell pole (unpublished data), and mCherry alone, known to localize to the host cytoplasm (Ramesh et al., 2021). Altogether, our data suggest that Phaedrus gp82 is a cytoplasmic-localized protein.

To evaluate the potential cytotoxic effects of P82, we overexpressed the untagged protein in *M. smegmatis* mc^2 155, along with controls for empty vector expression and expression of mCherry and Fruitloop gp52, a mycobacteriophage protein known to be lethal to *M. smegmatis* mc^2 155 (Ko & Hatfull, 2018). As expected, there is no difference in either colony count or colony size for the empty vector, while overexpression of mCherry resulted in a modest but statistically significant decrease in colony size (Figure 6). In contrast, overexpression of Fruitloop gp52 results in an extreme reduction in colony counts between ATc-induced and uninduced plates (Figure 6A). Overexpression of P82 resulted in only a small reduction in colony count, indicating that overexpression is not lethal. Quantification of colony size however demonstrates that while colony numbers are similar, colony sizes are significantly reduced when overexpressing P82 (Figure 6B, 6C). The overexpression of P82 resulted in a 67% reduction in the colony size, from 2.24 mm^2 to 0.74 mm^2, after 4 days of growth. (Figure 6C).

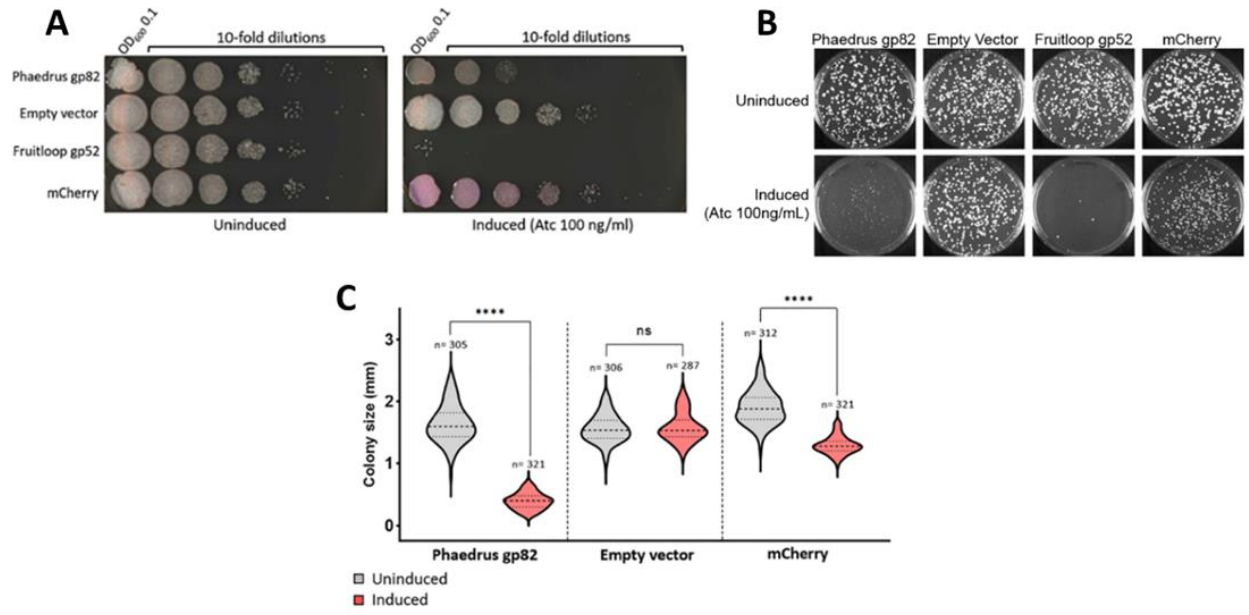

Figure 6. Overexpression of Phaedrus gp82 reduces colony size in *M. smegmatis*

A) Spot dilutions from cultures of *M. smegmatis* transformed with the inducible plasmid pKF8 expressing Phaedrus gp82 were plated on inducing and uninducing solid media. The empty vector was used as a negative control, mCherry was expressed as an expression control, and Fruitloop gp52 was included as a control for a toxic protein. **B)** representative plates illustrate the small colony phenotype observed with Phaedrus gp82 overexpression. *M. smegmatis* containing the indicated gene within the pKF8 plasmid were plated on inducing and uninducing solid media and incubated for 96 hours at 37 ̊C. **C)** Quantification of colony size resulting from the overexpression of Phaedrus gp82. Colonies expressing Phaedrus gp82 show a marked reduction in their size, while the empty vector does not. Overexpression of mCherry, which is not toxic, shows only a small reduction in colony size. The data were statistically analyzed using an unpaired t-test. The observed mean difference with a p-value <0.05 is considered statistically significant. (**** denotes p-value <0.0001, ns: not significant, n: colony numbers)

2.2.3 The crystal structure of Phaedrus gp82 reveals two distinct domains

In an effort to gain insight into potential functions for P82, we crystallized and determined the structure of P82 using X-ray crystallography. Small cubic crystals of ~40 microns on each edge grew at room temperature over a week using the vapor diffusion method. The crystals belong

to space group I23 and contain one molecule of P82 within the asymmetric unit. We obtained phasing information using anomalous dispersion from 4 native sulfur atoms (S-SAD) from data collected at our home source. The structure was then improved by refinement against native data at 1.2 Å resolution collected at APS beamline 31-IDD. The model was refined to an R-work/ R-free of 15.62% and 16.25% respectively; and shows excellent geometry. The final model contains all residues of the P82 protein except for a portion of the loop between helix α1 and α2 (residues 17-23) (Figure 7A), which was disordered in our electron density maps. Data collection and refinement statistics can be found in Table 1, and a complete description of the structure determination process can be found in the materials and methods.

The structure of P82 is organized into two distinct domains which we have named the base and wing domains (Figure 7B). The base domain folds into a compact structure anchored by helices α1 and α3 which are located underneath a small beta-sheet formed by strands β1, β4, and β5. The resulting configuration places the disordered loop extended away from the base and into solvent. The wing domain, formed by strands β2 and β3, also extends away from the base domain but is pointed in the opposite direction. We observe what we presume to be a magnesium ion at the junction between base and wing domains. This ion is well-ordered and is coordinated by E60 where it appears to be important for crystal packing. While magnesium is known to play a role in phage infection, a biological role for the magnesium that we observed here is unclear. The wing and one face of the protein is largely acidic, while there is a prominent basic patch at the bottom of the base domain. We used iterative searches of the sequence database using PSI-BLAST (Altschul et al., 1997) to identify potential P82 homologs. Interestingly, all homologs are found within phage or prophage genomes (Figure 7A). Mapping sequence conservation onto the surface of P82 reveals a large patch of residues in the middle of the base domain which is highly conserved

(Figure 7C). This conserved patch is largely hydrophobic while the opposite face of the molecule has a smaller patch of residues including P77, W79, L85, and R86 which is also hydrophobic but is less conserved. A search for structural homologs of P82 using the DALI server (L. Holm, 2022) revealed similarity with a subdomain within the F_{420}-reducing [NiFe] hydrogenase, gamma subunit (PDB: 6QGR) (Ilina et al., 2019) with an all-atom r.m.s.d. of 1.6 Å. In this hydrogenase, the subdomain with structural similarity to P82 forms a cap or lid over the central binding cavity which houses the FAD cofactor. Binding assays between P82 and FAD have not revealed an interaction, suggesting that P82 either binds a different cofactor or instead that it may interact with another protein instead of a ligand.

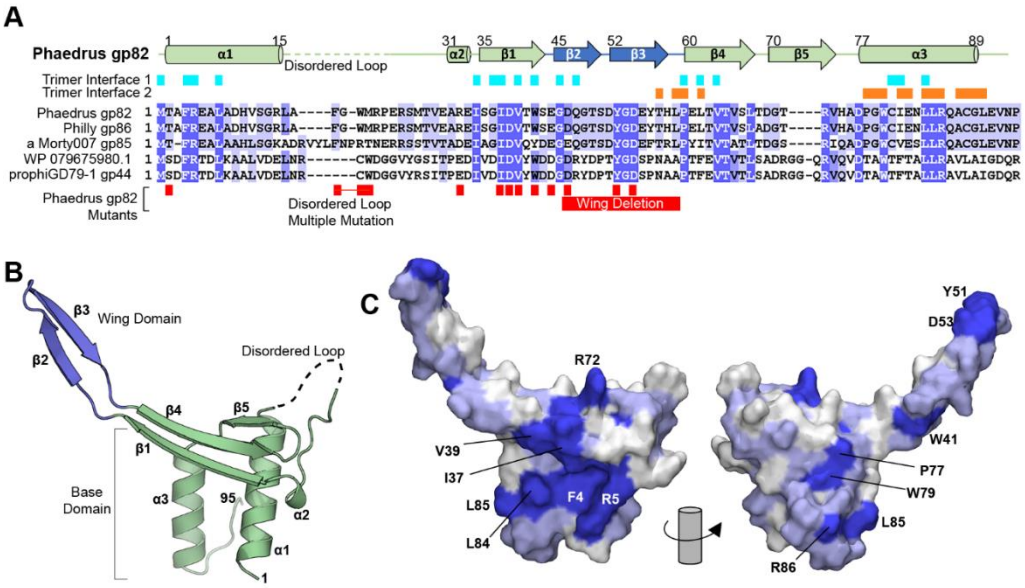

Figure 7. Structure and sequence conservation of Phaedrus gp82.

Diagram of the P82 protein including secondary structural element derived from the structure; interfaces between P82 subunits within the trimer; sequence conservation from a selection of P82 homologs; the positions of mutants within this study, **B)** Structure of the P82 monomer. The Base and Wing domains are indicated as well as the location of the disordered loop (aa 17-23). **C)** Sequence conservation from panel A mapped onto the surface of the P82 monomer. This view shows a highly conserved surface with several conserved residues on the opposite face.

Table 1. Data collection and refinement statistics for P82

	Sulfur anomalous	Native
Data collection		
Space group	I23	I23
Cell dimensions		
$a = b = c$ (Å)	83.75	83.73
Resolution (Å)	50 - 1.70 (1.73 - 1.70)	59.21 - 1.40 (1.42 - 1.40)
Unique reflections	10909 (993)	17487 (974)
R_{merge}	0.054 (0)	0.064 (0.24)
$I/\sigma I$	130.96 (4.62)	32.5 (9.9)
Completeness (%)	99.99 (100)	90.1 (100)
Redundancy	41.1 (31.2)	18.5 (18.9)
Refinement		
Resolution (Å)		14.80 - 1.40 (1.44 - 1.40)
R_{work} / R_{free} (%)		15.38/18.43
No. atoms		749
Protein		686
Water, Ions		62, 1
Avg B-factors (Å^2)		24.56
Protein		23.93
Ion		28.62
Water		31.51
R.m.s. deviations		
Bond lengths (Å)		0.004
Bond angles (°)		0.620
Ramachandran (%) Favored		100

Values in parentheses represent the outer resolution shell.
[a] $R_{merge} = (|(\Sigma I - \langle I \rangle)|)/ (\Sigma I)$, where $\langle I \rangle$ is the average intensity of multiple measurements.
[b] $R_{work} = \Sigma_{hkl}||F_o(hkl)|| - F_c(hkl)||/\Sigma_{hkl}|F_o(hkl)|$.
[c] R_{free} represents the cross-validation R factor for 10 % of the reflections against which the model was not refined.

2.2.4 Phaedrus gp82 is a trimer in solution

Packing of P82 within the crystal lattice revealed an assembly of three P82 proteins situated around a crystallographic three-fold symmetry axis (Figure 8B-C). Analysis of this trimeric assembly using PISA (Krissinel & Henrick, 2007), demonstrated that each interface contained 820-831 $Å^2$ of surface within it and a surface complementary score of 0.828 suggests that the trimeric packing may represent a biologically relevant interfaces. The trimeric P82 assembly adopts a conical shape driven primarily by packing of the base domains, specifically helix 3 which packs against β1 from the adjacent subunit (Figure 8A). This interface utilizes the large conserved patch noted on each subunit, further supporting the notion that the trimeric packing we observe is biologically relevant. There is an extensive network of van der Waals interactions within this interface. Interactions via I37, V39, and W41 of β1 were especially prominent (Figure 8A). Mutation of these residues was uniformly destabilizing and the resulting proteins could not be purified, further suggesting that in isolation, P82 strongly prefers the trimeric state. The top of the P82 cone is formed by the wing domain from each of the respective monomers. These fit together sterically but the packing interactions are far less extensive, containing only 166 $Å^2$ of interface area.

Next, we examined whether P82 could form trimers in solution biochemically. We performed chemical crosslinking using 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) and monitored the resulting species on an SDS-PAGE gel. As shown in (Figure 8D), we found that crosslinked trimers of P82 are formed readily, while tetramers or higher order species of P82 are largely not observed, suggesting P82 forms trimers in solution. We confirmed this via analytical size exclusion chromatography (SEC) using a Superdex 200 10/300 GL to monitor the retention volume of P82 and comparing against several control proteins including Bovine serum

albumin (BSA, 66 kDa), Tobacco Etch Virus protease (TEV, 27 kDa), and lysozyme (14 kDa). The SEC profile of P82 showed a single peak with a retention volume much lower than expected for a monomeric globular protein of 95 amino acids. In addition, the observed retention volume was consistent with an apparent molecular weight that is somewhat larger than that of TEV protease (Figure 8E). These results all suggest that P82 is a homotrimer, consistent with the crystallographic structure.

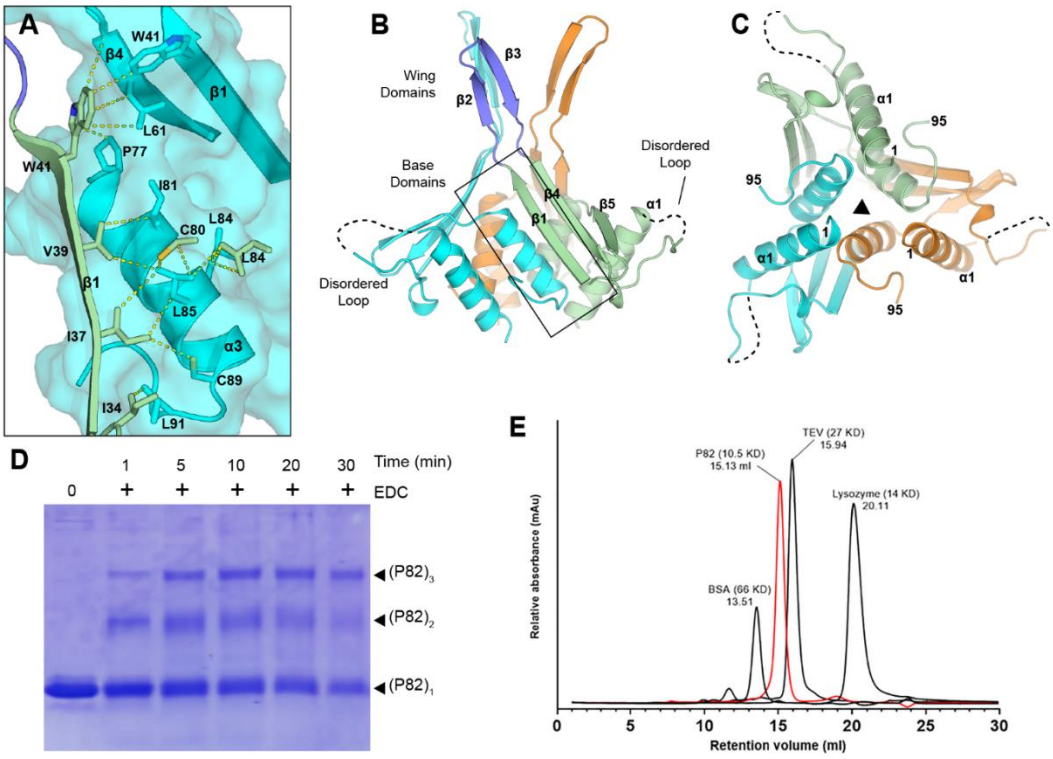

Figure 8. Phaedrus gp82 is a trimer both in solution and within the crystal.

A) The P82 trimerization interface is formed by packing between β1 and the large conserved hydrophobic surface on the adjacent molecule. Dashed yellow lines indicate van der Waals interactions. **B-C)** Packing of the P82 around a crystallographic 3-fold axis positions base domains adjacent to each other while wing domains are somewhat splayed out from the central axis. The disordered loop positions are highly exposed. **D)** Chemical crosslinking of P82 using EDC shows the rapid formation of P82 trimers. **E)** Size exclusion chromatography of P82 (red). Also shown are a variety of standards including BSA (66kDa), TEV protease (27kDa) and lysozyme (14kDa).

2.2.5 Identifying residues and regions within P82 that are important for the small colony phenotype.

The orientation of the wing domain in our crystal provides a large concave and acidic surface (Figure 9A). We hypothesized that disruption of this charged surface, or a putative protein-protein interaction surface, could block or limit the small colony phenotype we observe with WT P82 overexpression. To identify locations on the surface of P82 that might mediate protein-protein interactions, we generated a composite structural model in which the positions of disordered loop domains are those positions observed via crystallography. This composite model was analyzed via surface triplet propensity (Mehio et al., 2010) and PPISP (Kang et al., 2022; Qin & Zhou, 2007) to identify surfaces that may mediate protein interactions. Within the base domain, the disordered loop and nearby residues on the concave surface were identified. These include R72, H74, and D76. The STP analysis also identified residues within the wing, including D45, Y51, H57, and E60 (Figure 9B).

Guided by this analysis, we next sought to identify amino acids on the surface of the P82 trimer that are functionally important, hypothesizing that their disruption may relieve the small colony phenotype we observe with wild-type P82. We tested several regions within the protein in an attempt to sample all of the surfaces that might be important. Mutants tested include D45A, Y51A, and D53A, which constitute most of the putative protein-protein interaction sight identified by STP analysis (Figure 9B). E43 was chosen because it is at the junction between the base and wing domains and is also a main component of the acidic patch in this region. A D38A mutant was selected to probe an acidic residue within the base domain, while R32A was selected as it probes a different face of the molecule. T2A was selected to test whether the pore-like feature at the bottom of the trimer was important. Lastly, we generated a triple mutant within the disordered

loop (F18A/W20A/M21A) and a deletion of the entire wing (residues 45-58 replaced with a TG linker to avoid impacting the protein fold). The coding sequence for P82 containing the relevant mutation (or deletion) was introduced into pKF8 for inducible expression in *M. smegmatis* mc^2 155 without any tags. Using this, we asked whether overexpression of mutant P82 still generated the small colony phenotype or alternatively, whether the mutant suppressed the phenotype. We found that T2A and R32A mutants each had little or no effect, leading us to conclude that this face of P82 does not contribute to the phenotype. In contrast, the disordered loop triple mutant and the wing deletion completely suppressed the phenotype, suggesting these large regions on the P82 surface are connected to the overexpression phenotype and perhaps P82's biological function (Figure 9C). A single point mutant at D45A located closer to the base domain had the same impact as the wing deletion, while Y51A and D53A located near the wing's tip had a mild or no effect (Figure 9D). This suggests that the lower portion of the wing contributes more to this assay. Another mutant in this region, D38A, also had a large impact on the small colony phenotype. Together, these suggest that the central portion of P82 extending from the bottom of the wing near D45 down to the disordered loop is critical to the small colony phenotype.

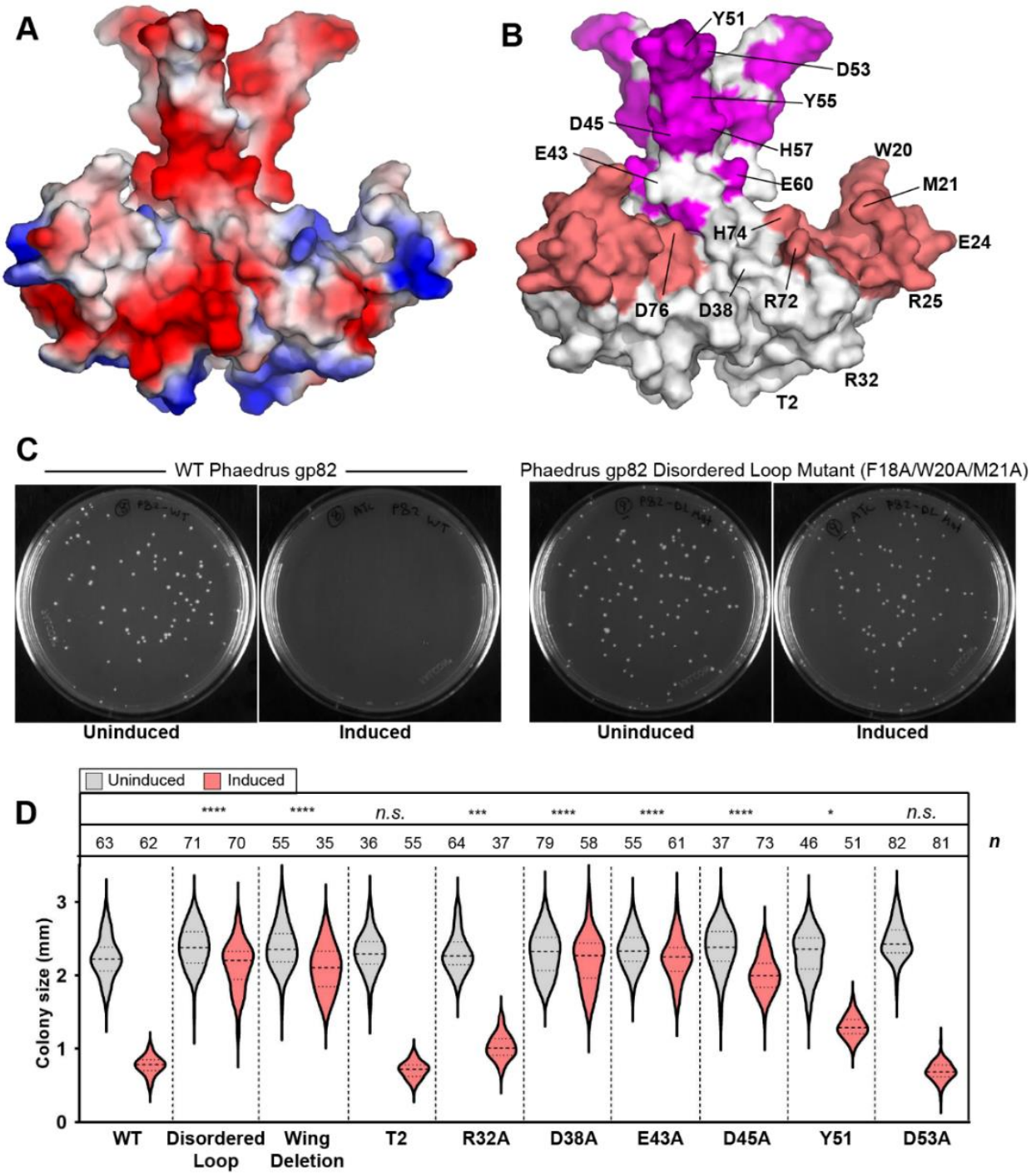

Figure 9. Identification of functionally important surfaces on Phaedrus gp82.

A) Electrostatic surface generated from the composite model of P82 shows that the wing and a portion of the base domain are highly acidic. **B)** Putative interaction sites as predicted by surface triplet propensity (magenta) and the PPISP website (salmon) suggest that both the base and wing domains may mediate protein interactions. Prominent residues from the predictions are indicated as are the positions of residues tested in subsequent mutagenesis. **C)**

Mutation of the disordered loop (F18A/W20A/M21A) blocked the small colony phenotype observed with P82 overexpression. **D)** Quantification of the small colony phenotype observed with WT P82 and a series of truncations and mutations. The number of colonies measured on each plate is indicated and their size measured via Fiji and the distribution visualized using a violin plot. The difference in colony size when grown on inducing and uninducing medium was measured and compared to the difference observed with the wild-type. Statistical significance is indicated with P<0.01=*, P<0.001=***, and P<0.0001=****. N.S. indicates that there is not a statistically significant difference between the groups with that mutant.

2.2.6 Phaedrus gp82 interacts with *M. smegmatis* MoxR ATPase

Next, we performed a pulldown experiment to identify a host protein that binds P82 and might be associated with the small colony phenotype. To accomplish this, we transformed *M. smegmatis* mc^2 155 with the pKF41 plasmid driving the expression of a C-terminally HA-tagged Phaedrus gp82 and then performed a pulldown assay using anti-HA agarose beads. Immunoprecipitated proteins were separated on SDS-PAGE, and two bands were observed in the pulldown with HA-tagged P82 that were not observed in an untagged P82 control pulldown (Figure 10A, Figure S3). These bands were excised from the gel, and digested with trypsin, and their identities were determined by LC/MS. The highest intensity band was identified as the AAA+ ATPase MoxR (MSMEG_3147, 82% coverage), while the lower intensity band was MSMEG_0970 which is a member of the phosphoglycerate mutase family (40% coverage) (Figure. S4). Subsequent analysis focused on the interaction with MoxR.

Next, we asked mutants known to affect the P82 overexpression phenotype alter interactions with MoxR. We began by repeating the anti-HA pulldown after overexpression of HA-tagged P82 containing the disordered triple mutant (F18A/W20A/M21A) or an untagged control containing the same mutant, finding that the interaction with MoxR appears to be lost with

the DL mutant (Figure 10B). Next, we used an integrating vector to introduce an expression cassette coding for the inducible expression of FLAG-tagged MoxR into the genome of *M. Smegmatis*. We then transformed these cells with plasmids expressing selected HA-tagged P82 variants or untagged controls. After inducing the expression of both FLAG-MoxR and HA-tagged P82, we repeated the pulldown and analyzed the immunoprecipitated proteins via Western blot using anti-FLAG or anti-HA primary antibodies. We found no anti-FLAG signal in the controls containing untagged P82 WT or DL mutants as expected while observing robust signal for MoxR pulldown with the HA-tagged version of WT P82. FLAG-MoxR signal was completely lost with the HA-tagged DL triple mutant (Figure 10C), indicating these residues play an important role in mediating MoxR interactions. Interestingly, we detected a robust signal when pulling down with HA-tagged P82 containing the wing deletion, suggesting that this portion of P82 does not interact with MoxR directly or plays a secondary role in binding (Figure 10C). Similarly, the E43A mutation can still support the pulldown of FLAG MoxR, as can HA-tagged P82 T2A, which served as a negative control. From these data, we conclude that the residues within the disordered loop play a primary role in mediating the physical interaction between P82 and MoxR.

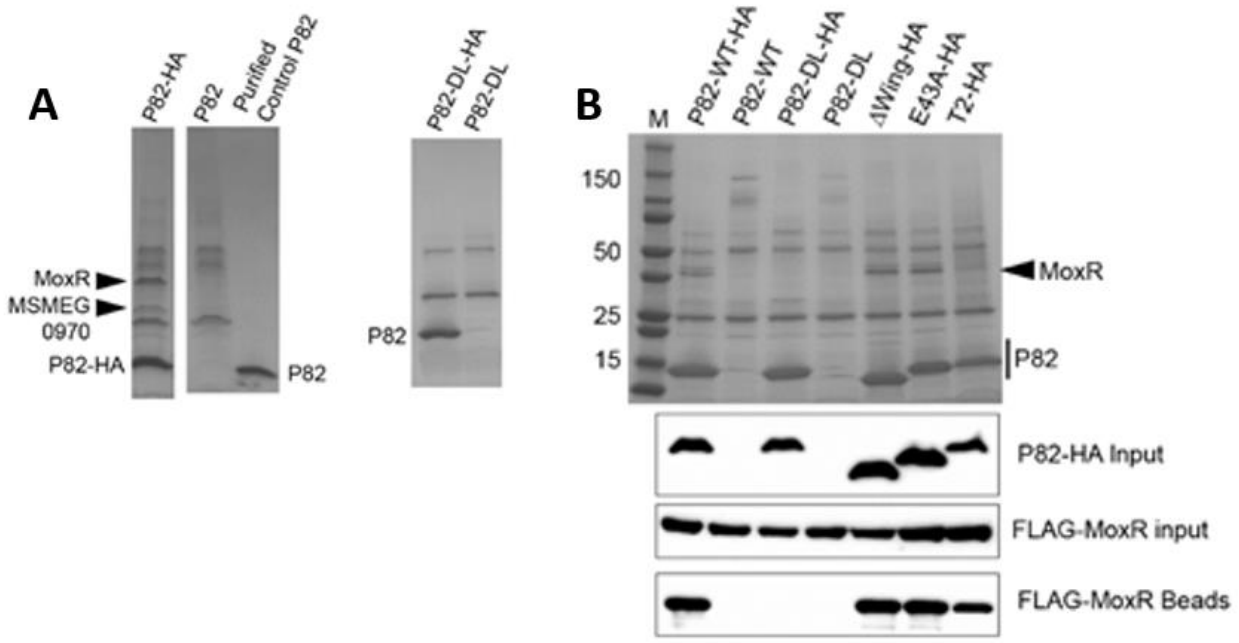

Figure 10. Phaedrus gp82 interacts with MoxR ATPase.

A) Anti-HA beads after pulldown from indicated lysate of *M.Smegmatis* expressing HA-tagged P82 or an untagged control. Purified recombinant P82 is included as a P82 standard. Lanes shown are from the gel. **B)** Anti-HA pulldown as in A, but expressing P82 containing the DL triple mutant, demonstrating that the interaction is lost with this variant of P82. **C)** Anti-HA pulldown of selected P82 mutants from *M.Smegmatis* transformed with FLAG-tagged MoxR as well as the indicated HA-P82 variant or its untagged control.

2.2.7 Connection between MoxR function and Phaedrus gp82

Next, we asked whether overexpression of MoxR could rescue the P82 small colony phenotype. Here we transformed with pKF7, which integrated an inducible expression cassette containing the MoxR coding sequence into *M. Smegmatis*, or a pKF7 empty vector control. Following integration, the appropriate strains were then transformed with pKF113, which drives the inducible expression of P82 or its empty vector control. We then tested the small colony phenotype as before, comparing colony size with and without induction. As shown in (Figure

11A-B), the expression of just MoxR does not generate a significant change in colony size, while expression of just P82 resulted in smaller colonies, as expected from our previous observations (Figure 11D). However, when both P82 and MoxR expression is induced, we see a significant reduction in the phenotype. Therefore, we conclude that MoxR induction can partially rescue the small colony phenotype resulting from P82 overexpression.

To test whether P82 could directly impact MoxR function, we used the malachite green assay to measure the effect of P82 on MoxR ATP hydrolysis. Malachite green molybdate forms a complex with orthophosphate after ATP hydrolysis that absorbs light at 650 nm (D'Angelo et al., 2001; O'Toole et al., 2007). We used this assay to measure the MoxR-mediated release of phosphate, finding it had a hydrolysis rate of 2.72 ± nmol/minute. The standard curve used to calculate phosphate release is shown in (Figure S5). Next, we introduced P82 at a concentration of 1.4 µM and observed a slight decrease in MoxR hydrolysis (Figure 11C). This effect was augmented upon introducing P82 at 14 µM (Figure 11D). From these results, we concluded that using purified components *in vitro*, P82, negatively impacts MoxR hydrolysis rates. Lastly, we asked whether P82-mediated inhibition of MoxR ATPase activity was affected by the disordered loop, as this region of P82 was required for MoxR interactions in our pulldown assay. Here, we find that using P82 with the nontoxic disordered loop triple mutant has a modest impact on MoxR ATP hydrolysis, requiring a 14.4 µM concentration to achieve the same level of inhibition as 1.4µM of wild-type P82 (Figure 11D). Similarly, ATP hydrolysis in the presence of 1.4µM P82 DL mutant was the same as MoxR alone (Figure 11B). These results are consistent with previous toxicity and pulldown results showing that the disordered loop is important for P82-MoxR interactions and further suggests that this interaction directly affects MoxR ATPase activity.

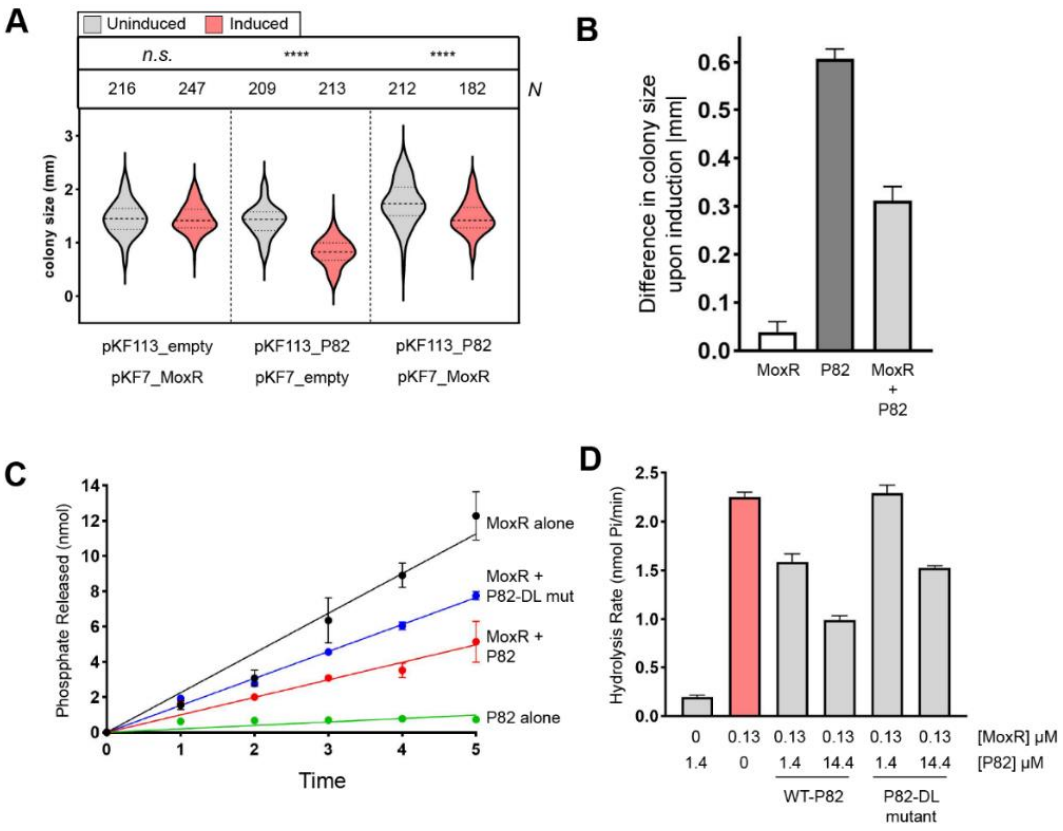

Figure 11. Phaedrus gp82 interacts with MoxR ATPase.

A) Violin plot of colony size with overexpression of MoxR, P82, or both MoxR and P82. Overexpression of MoxR partially rescues the small colony phenotype. **B)** Absolute differences in colony size after induction is shown from each population. **C-D)** Measurement of ATP hydrolysis from the MoxR AAA+ ATPase in vitro using the malachite green assay. We observe a decrease in hydrolysis rates in the presence of P82, which is lost when using the disordered loop triple mutant.

2.3 Discussion

The extraordinary levels of genetic diversity within mycobacteriophages reflect an evolutionary arms race between phages and their mycobacterial hosts. For their part, phages seek to assert control over a variety of host pathways to divert resources toward maintenance and

expansion of the phage population. Enriching small potentially regulatory genes of unknown function within the phage proteome has led many to speculate that this reservoir is replete with novel proteins mediating unexplored phage-host interactions. The efforts described here sought to test that hypothesis using P82.

We determined the structure of P82 using x-ray crystallography, finding that the protein adopts a trimer with two distinct domains: a base domain with extensive hydrophobic interactions that mediate trimerization, and a wing domain with a much smaller trimerization interface. Both domains impact the cytotoxic effects of P82 overexpression, as did residues within the acidic concave surface located in between these. Further, we were able to find two potential P82-interacting proteins from *M. smegmatis*: MoxR (MSMEG_3147), and a member of the phosphoglycerate mutase family of proteins, MSMEG_0970. MoxR is the founding member of a family of AAA+ proteins (ATPases Associated with various cellular Activities), called the MoxR Proper or MRP family (Snider & Houry, 2006). The MoxR family members are abundant in bacteria and archaea and have a chaperone activity to help the folding of proteins and the assembly of protein complexes (Bhuwan et al., 2016; Dieppedale et al., 2011; Pelzmann et al., 2009; Wong et al., 2017). Many MoxR clients are proteins with dedicated metabolic roles (Van Spanning et al., 1991), or are connected to a variety of functional roles including co-factor insertion (Leipe et al., 2003; Wong et al., 2014) and maturation of the RipA virulence factor in *M. tuberculosis* (Bhuwan et al., 2016). The diversity and functional importance of MoxR clients makes MoxR an especially attractive target for phage-derived therapeutics. It is not known from our data whether Phaedrus encodes gene 82 to target a specific MoxR client in *M. smegmatis*, but a disruption in MoxR-mediated ATPase activity could be expected to impact numerous cellular processes. For example, a disruption in metabolic pathways could reasonably manifest as a growth defect such as the small

colony phenotype we observe with P82 overexpression. Alternatively, the disruption of MoxR's chaperone activity by P82 binding could reduce the populations of properly folded essential proteins, for example those integral to cell wall growth and division.

An open-ended question resulting from our work is the nature of the interaction between MoxR and P82. A common feature of MRP family members is the presence of a von Willebrand factor A (Snider & Houry) domain in the gene immediately downstream of the MRP protein. In *M. smegmatis*, there are two proteins, MSMEG_3148 and MSMEG_3149 that match these criteria. A comparison of P82 with MSMEG_3148 and MSMEG_3149 does not reveal any sequence or structural similarities. In a similar vein, predictions of MoxR with P82 interactions using AlphaFold2 have not yielded plausible models that are supported by the experimental biochemical and phenotypic data described here. Therefore, a molecular understanding of the P82-MoxR interaction will require additional structural information or biochemical insight that would place constraints on the positioning of the two proteins relative to each other.

Sequence alignment suggests the presence of sequence variation in the disordered loop region (Figure 7A). This sequence variation may indicate that the level of disorder is different among these proteins, which may provide a mechanism for Phaedrus gp82 to bind and modulate MoxR ATPase activity. We propose that generating chimeric Phaedrus gp82 with swapped region covers the disordered loop segment could be used to investigate the functional and dynamic properties of the disordered loop. This might explain the evolutionary mechanism by which the sequence variation in the disordered loop region allows Phaedrus gp82 and its homolog with the same pham to gain new biological function.

Our data here supports the conclusion that the disordered loop is a critical sequence feature within P82 that mediates the interaction with MoxR. However, we note several other locations

outside the disordered loop motif, specifically the wing domain and individual residues such as E43, D38, and D45, where mutation/deletion of the residue altered the phenotype. Neither deletion of the Wing domain nor a E43A mutation altered the ability of P82 to pulldown MoxR from a cellular lysate. This raises the possibility that there are multiple points of contact between MoxR and P82 with the disordered loop being the primary contact. In this scenario, cells expressing the mutated P82 may still retain enough MoxR activity to support normal colony growth. Alternatively, these regions may represent a binding site to a factor other than MoxR that has not yet been identified. The phosphoglycerate mutase MSMEG_ 0970 may be a candidate for this interaction, as it was retained in our original pulldown. This possibility has not yet been tested, but it would explain why we were not able to generate stable complexes between P82 and MoxR when using purified components.

Phaedrus gp82-MoxR monomer-monomer interaction was predicted using AlphaFold (Figure 12). However, the predicted model has no significant physicochemical properties with a significant score (SCC) for the complex formation of 0.1 and solvation energy gain (ΔG) of -9.4 kcak/mol. Moreover, residues such as D38, E43, and D45 which block the small colony phenotype upon mutation are not in contact with MoxR in this model. This data suggests that AlphaFold model is not plausible.

Several reasons can contribute to the bacterial small colony phenotype. In *E. coli*, small colony variants can be attributed to mutations that occurred in protein-coding genes involved in electron transport chain (ETC) such as hemB (Roggenkamp et al., 1998), hemA (Hubbard et al., 2021), yigP (Xia et al., 2017), and lipA (Santos & Hirshfield, 2016). *M. smegmatis* deficient in mycolic acid biosynthesis (α- and epoxy-mycolic acid) has been shown to produce small colony phenotype (Di Capua et al., 2022; Lefebvre et al., 2018). The defect in α- and epoxy-mycolic acid

content is due to deficiency in HadD (MSMEG_0948) activity which is essential for biosynthesis of α- and epoxy-mycolic acid (Di Capua et al., 2022). A small colony phenotype resulting from the overexpression of phage protein could indicate that phage protein directly or indirectly interacts with and modulates the activity of the abovementioned host proteins.

Sequencing and comparative analysis of large numbers of mycobacteriophage genomes have identified many new gene functions and fundamentally improved our understanding of phage biology, the biology of their mycobacterial hosts, and the pathways targeted by phage-host interactions. The rise in antibiotic-resistant strains of *M. tuberculosis* and other pathogenic mycobacteria, as well as efforts to capitalize on the therapeutic potential of phage (Dedrick et al., 2022; G. F. Hatfull, 2022; Hatfull, 2023), highlights that there is a great need to identify many potentially therapeutic phage-host interactions as possible. While this pilot project represents just one protein and one pham, recent estimates have suggested that >20% of phage phams are predicted to be toxic to the bacterial host (Ko & Hatfull, 2020). Each of these has the potential to reveal a new phage-host interaction which could, in theory, be harnessed to negatively impact mycobacterial health. Our results suggest a pathway for identifying and characterizing these interactions and the molecular underpinnings of their toxicity.

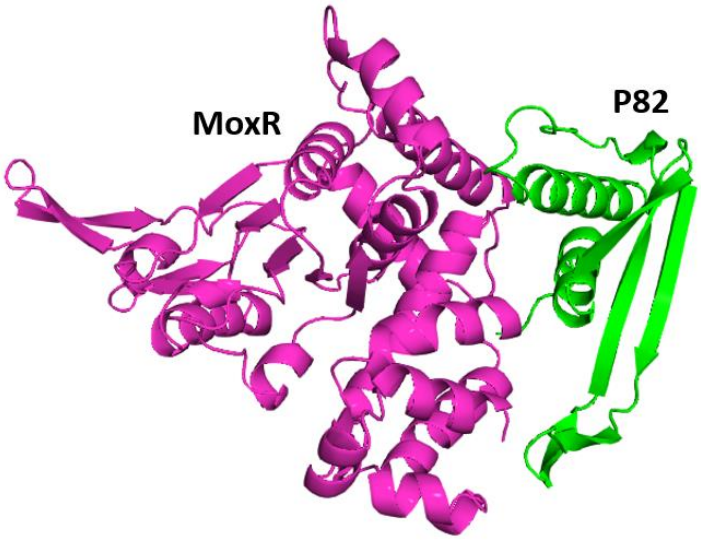

Figure 12. Predicted AlphaFold structure of Phaedrus gp82-MoxR monomer-monomer interaction

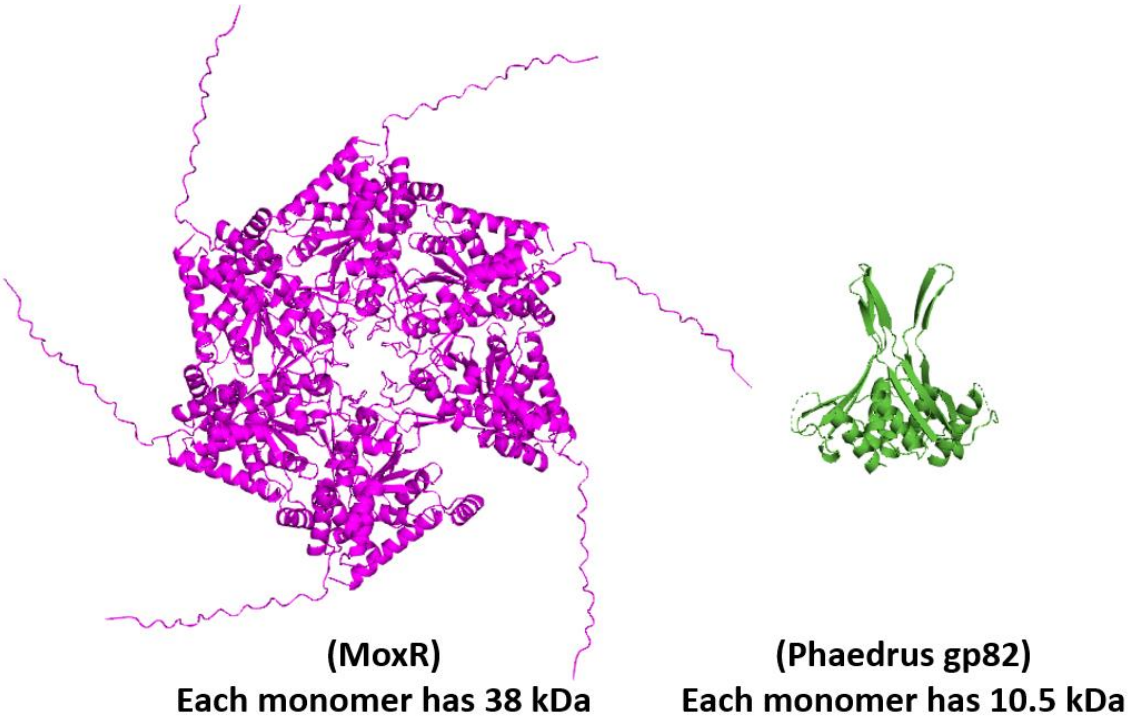

Figure 13. Structure of Phaedrus gp82 trimer and predicted MoxR hexamer and their relative sizes

Table 2. The other 11 highest protein candidates resulted from the bioinformatic pipeline

	Mycobacteriophage protein	Cluster	Subcluster
1	Gaia_106	X	None
2	Spud_94	C	C1
3	Severus_67	A	A10
4	L5_45	A	A2
5	Brujita_71	I	I1
6	Lamina13_48	A	A1
7	Cerasum_72	F	F1
8	Blue7_60	A	A6
9	Cuco_50	A	A5
10	Pharsalus_129	E	None
11	Theia_83	A	A5

2.4 Material and Methods

2.4.1 Bioinformatics

Our bioinformatic analysis utilized the phage proteome as of July 2017, which contained ~214,000 mycobacteriophage protein sequences. We first removed from our candidate pool proteins with an annotated function. We further proceeded to analyze the proteins of unknown function for their sequences and physical parameters to collect the information required to predict the propensity of protein production and their likelihood to crystallize, as described in the text. Physical parameters for sequences have been analyzed using ProtParam (Gasteiger et al., 2003). The shortlisted protein candidates were subjected to BLASTP (Altschul et al., 1990) and HHPred (Zimmermann et al., 2018) homology searches to ensure a lack of sequence-based homology to proteins of unknown function, and I-TASSER (Yang & Zhang, 2015; Zhang, 2008) and PSIPRED

(McGuffin et al., 2000) were used to predict the secondary structure of proteins. Both PPCpred (Mizianty & Kurgan, 2011) and XtalPred (Slabinski et al., 2007) have been used to predict protein crystallization propensity and a variety of sequence-based parameters.

2.4.2 Protein Expression and Purification

The codon-optimized coding sequence for Phaedrus gp82 (UniProtKB: B5A6J1) was cloned into the pKF3 plasmid (Googins et al., 2020) to create pKF87, driving expression of Phaedrus gp82 with N-terminal His_{10}-mRuby2 tags and a TEV protease cleavage site. The protein was expressed in LB media using Rosetta2 *E. coli*. Protein expression was induced at an optical density (OD_{600}) of 0.6 by adding 0.5 mM IPTG for 24 hours at room temperature. Cells were harvested by centrifugation and lysed by homogenization in lysis buffer containing (20 mM Tris, pH 8.0, 0.5 M NaCl, 10% glycerol, 20 mM Imidazole, and 1 mM β-mercaptoethanol). The cell lysate was clarified by centrifugation at 30,000 x *g* for 30 min. Phaedrus gp82 fusion protein was initially purified using nickel-affinity chromatography (Figure 12), followed by overnight TEV digestion to remove the His_{10}-mRuby2 tag. The resulting mixture was further purified by a second round of nickel affinity chromatography to remove the liberated His_{10}-mRuby2 tag (Figure 13). Anion exchange chromatography (HiTrap Q HP, GE Healthcare) followed by size exclusion chromatography (Sephacryl-200, GE Healthcare) in a buffer containing [20 mM Tris, pH 8, 150 mM NaCl, and 1 mM β-mercaptoethanol] completed the prep (Figure 14). Three non-native residues (GGS) at the N-terminus remain after TEV cleavage. The resulting protein fractions were evaluated by SDS-PAGE and the purity was > 99% (Figure 15). The protein fractions were subsequently pooled and concentrated in the same size exclusion chromatography buffer to approximately 8 mg/ml prior to crystallization.

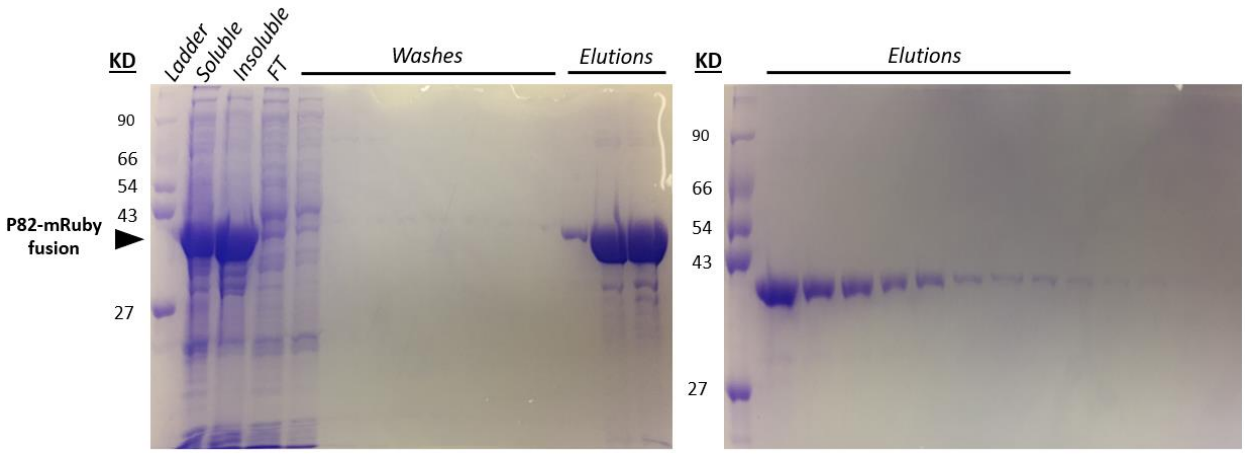

Figure 14. Purification of Phaedrus gp82 - First nickel column

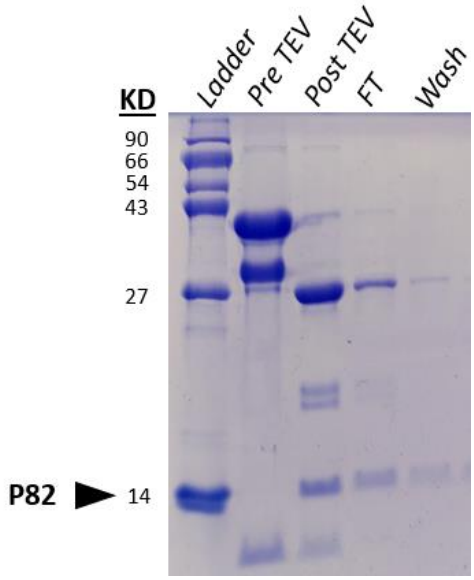

Figure 15. Purification of Phaedrus gp82 - Second nickel column

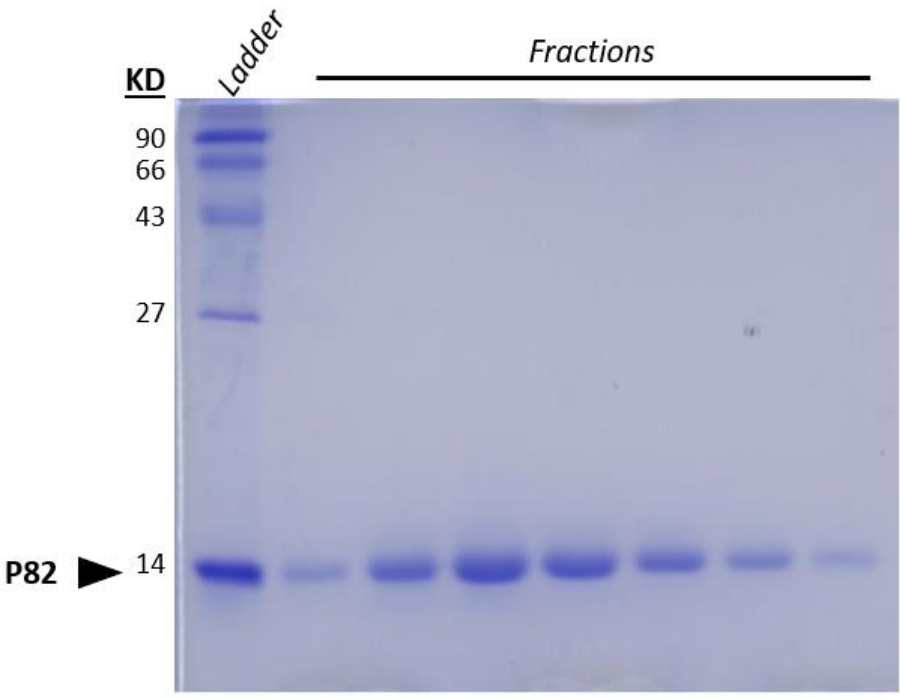

Figure 16. Purification of Phaedrus gp82 - Size exclusion chromatography column using Hiload S200 16/600

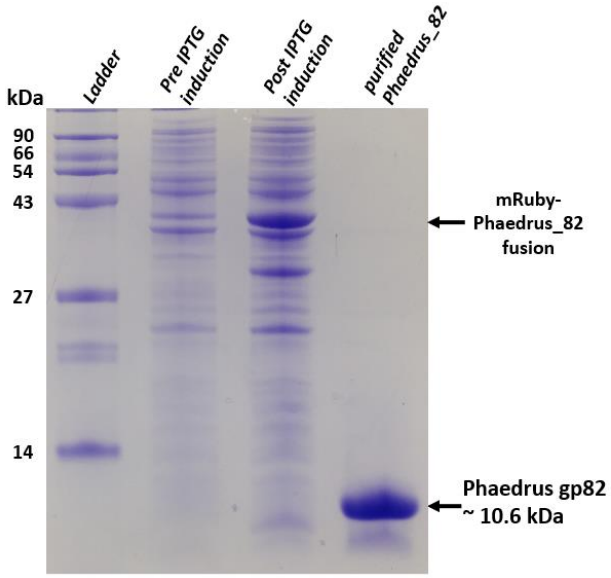

Figure 17. Phaesus gp82 - Pre and post induction and purified protein

Phaedrus gp82 was purified to using different purification methods to obtain highly pure protein for crystallization attempts.

2.4.3 Crystallization and structure determination

Initial crystals of Phaedrus gp82 were grown at 20 °*C using the* sitting-drop vapor diffusion method, mixing 1 µL of protein and 2 µL well solution containing 30% (v/v) PEG 550 MME, 0.05 M magnesium chloride hexahydrate, 0.1 M HEPES pH 7.5. Small cubic-shaped crystals (~50-100 microns on each edge) were grown over the course of several days (Figure 16). Crystals were soaked in cryoprotectant containing the reservoir solution supplemented with 30% glycerol and then flash-frozen in liquid nitrogen before data collection.

High-resolution diffraction data from native Phaedrus gp82 crystals were collected at beamline 31-IDD at the APS at Argonne National Laboratory (Chicago, Illinois). The data were then processed and scaled using the autoPROC toolbox (Vonrhein et al., 2011). Phasing information was obtained using single-wavelength anomalous dispersion of native sulfur atom (S-SAD) using data collected from a single crystal at our home source at Cu-Kα. Diffraction data from the home source were integrated, merged, and scaled with HKL2000 (Z. Otwinowski & W. Minor, 1997). Three of a possible four sulfur sites were identified using Phenix/HYSS (Adams et al., 2010; Grosse-Kunstleve & Adams, 2003), and phases were calculated using AutoSol (Terwilliger et al., 2009) as implemented within Phenix. Initial maps were improved by density modification using RESOLVE, resulting in readily interpretable electron density. An initial model was built into this map using COOT (Emsley et al., 2010), and the resulting model was refined against the high-resolution native data described above. The resulting model was further refined by positional and anisotropic B-factor refinement. Model quality throughout the refinement process was monitored and validated by MolProbity (Williams et al., 2018). Data collection and refinement statistics are shown in Table 1.

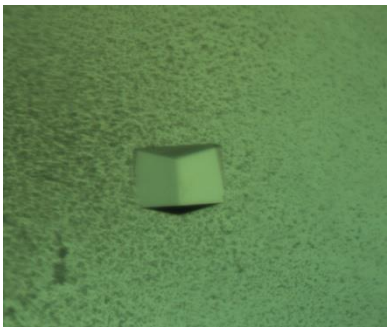

Figure 18. Phaedrus gp82 crystal

Cubic crystal was grown at 20 ℃ using the sitting-drop vapor diffusion method by mixing 1 µL of protein and 2 µL well solution containing 30% (v/v) PEG 550 MME, 0.05 M magnesium chloride hexahydrate, 0.1 M HEPES pH 7.5.

2.4.4 Cellular localization

Plasmids pKF2, pKF76, and pKF73, encoding Tet-inducible mCherry, mCherry-Cuco gp50 fusion, and mCherry-Phaedrus gp82 fusion, respectively, were created with Gibson Assembly using the same vector backbone. One hundred nanograms of each of these plasmids were electroporated into *M. smegmatis* mc^2155, followed by the addition of 7H9 and ADC and incubation at 37 °C for 3 hours to allow recovery of the cells. Cells were then plated on 7H10 solid medium containing hygromycin B (150 µg/ ml) and incubated at 37 ̊C for 72 hours. Transformants were grown in 7H9/ADC liquid culture containing Tween-80 (0.05%) with shaking at 37 ̊C until saturated. Saturated cultures were used to inoculate fresh media (7H9/ADC/Tween), and these samples were grown at 37 ̊C with shaking to an OD_{600} of 0.6. Cells from 100 µl of this culture were isolated by centrifugation at 11,000 *x g* for 1 minute, the supernatant removed, and the cells resuspended in 2 µl of SlowFade Diamond Antifade Mountant (ThermoFisher). The mixture was then placed on microscope glass slides as uninduced samples. Anhydrotetracycline (ATc) (100 ng/ ml) was then added to the rest of the culture at (OD_{600} = 0.6) to induce expression. The localization

of mCherry alone, mCherry-Phaedrus gp82 fusion, and mCherry-Cuco gp50 fusion were examined at different times between 3 and 24 hours of induction. Bright-field and fluorescent microscopy were performed using a Zeiss Axiostar Plus with an A-Plan 100x objective. Exposure times were set to automatic except for the uninduced samples, which were exposed for 20 seconds. Images were captured with AxioCam MRc 5 high-resolution camera and processed by AxioVision software.

2.4.5 Cytotoxicity plate assay and quantification

One hundred nanograms of plasmids pKF8 (empty vector) and its derivatives, pKF113 (encoding Tet-inducible untagged Phaedrus gp82), pKF114 (mCherry) and pKF115 (Fruitloop gp52), were individually transformed into 100 µl electrocompetent *M. smegmatis* mc^2 155. Fruitloop gp52 was used as a toxic control, mCherry was used as an expression and non-toxic control, and the empty vector was a background control. The transformed cells were allowed to recover for 3 hours at 37˚C before plating on 7H10 solid medium with hygromycin B (150 µg/ ml) and incubation at 37 ˚C for 72 hours. A liquid medium 7H9/ADC/Tween (0.05%) was then prepared, inoculated with a single colony from each plate, and incubated at 37˚C for 72 hours with shaking. These saturated cultures were used to inoculate fresh media, and samples were grown with shaking at 37˚C until reaching (OD_{600} ~ 0.3-0.7). These cultures, in exponential growth, were then normalized to OD_{600} 0.1, and ten-fold serial dilutions of each cell culture were prepared to reach a final dilution of 10^{-7}. From each dilution, 3 µl were spotted on ATc-induced and uninduced plates and incubated at 37 ˚C for 96 hours. To measure colony size, 10 µl from the 10^{-5} dilution was plated on ATc-induced and uninduced plates and incubated at 37˚C for 96 hours. Colony sizes were processed and analyzed using ImageJ-Fiji (Schindelin et al., 2012).

2.4.6 Chemical crosslinking

Chemical crosslinking was performed as previously described (Shi et al., 2014) with a few modifications. Briefly, the protein sample was incubated with 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) crosslinking buffer (100 mM MES pH 6, 100 mM NaCl, and 10% Glycerol). EDC and N-hydroxysuccinimide (NHS) were added to final concentrations of 20 mM and 0.4 mM, respectively, followed by incubation at room temperature with gentle agitation for 1 - 30 min. The reaction was quenched by adding (Tris-HCl, pH 8.0) to a final concentration of 50 mM and incubation at room temperature with agitation for 10 min. Samples were analyzed using SDS-PAGE and Coomassie blue stain.

2.4.7 Growing mycobacteria and media

M. smegmatis mc^2 155, an efficient plasmid transformation strain (Snapper et al., 1990), was grown as described previously (Payne et al., 2009). In brief, *M. smegmatis* mc^2 155 was grown at 37 °C in Middlebrook 7H9 liquid medium or on Middlebrook 7H10 agar. Both media were supplemented with 10% Albumin Dextrose Complex (ADC), 0.05% Tween 80, carbenicillin (CB) (50 μg/ml), and cycloheximide (CHX) (10 μg/ml). Media were supplemented with the appropriate antibiotics as needed. The final concentrations were as the following: Hygromycin B (150 μg/ml), Streptomycin (20 μg/ ml), and Kanamycin (20 μg/ ml).

2.4.8 Pull-down assay and mass spec identification of binding partners

M. smegmatis mc^2 155 was transformed with either pKF113 expressing Phaedrus gp82 or pKF226 expressing C-terminal HA-tagged Phaedrus gp82. Cell cultures were grown to OD_{600} ~0.5, then protein expression was induced by adding ATc (100 ng/ml) for ~18-24 hrs. Cells were then harvested by centrifugation and resuspended in TBS buffer (25 mM Tris-HCl PH 7.4, 150 mM NaCl). Cells were pelleted by centrifugation and resuspended in lysis buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% Triton X-100), including protease inhibitor cocktail, and lysed by sonication. Approximately 5 mg of cell lysate were incubated with anti-HA agarose beads (HA-Tag IP/Co-IP, Thermo Fisher Scientific) overnight at 4 °C with gentle agitation. Flowthrough samples were collected, and three washes were applied using lysis buffer, followed by the final wash with TBS buffer. Bound proteins were eluted by adding 50 µl of 2X non-reducing sample buffer followed by 5 min incubation at 95 °C and then centrifugation. Proteins were separated by 4-20 % gradient gel. The desired bands were excised for analysis. Samples were processed for protein identification at the Biomedical Mass Spectrometry Center, University of Pittsburgh.

Western blot was performed to demonstrate the presence of the binding partner in the elution from anti-HA agarose beads. Here, pSB10 is an integrated vector that confers kanamycin resistance and expresses FLAG-MoxR, while pKF226, pSB11, pSB12, pSB13, and pSB14 overexpress the C-terminal HA-tagged version of Phaedrus gp82, Phaedrus gp82 disordered loop, Δ wing domain, E43A, and T2A mutants, respectively. Pulldown experiments were performed as described above, and MoxR was detected via Western blot using an anti-FLAG antibody (ThermoFisher, MA1-91878).

2.4.9 Colorimetric Determination of ATPase Activity

The determination of ATPase activity was carried out by measuring the release of orthophosphate using the malachite green colorimetric assay as previously described (Lanzetta et al., 1979) with minor modifications. The color reagent buffer was prepared by mixing three parts of a malachite green solution (0.045% w/v in ddH_2O) with one part of ammonium molybdate (4.2% w/v in 4M HCl). The solution was agitated by stirring at room temperature until the solution turned a clear green-yellow color. NP-40 was then added to a final concentration of 0.1%. The color reagent buffer was filtered through a 0.45 μm syringe filter and stored at 4 °C. The reaction buffer contained 50 mM HEPES pH 7.5, 50 mM NaCl, 10 mM MgCl2, and 2 mM DTT. ATPase activity for MoxR was measured at 37 °C for 1-5 minutes by adding ATP (final concentration 2.5 mM) to the reaction buffer containing the indicated protein(s) to a final reaction volume of 200 μl. After incubation for the indicated times, 800 μl of color reagent was added. The solution was vortexed briefly and incubated at room temperature for 1 minute, followed by adding 100 μl of 34% (w/v) sodium citrate to quench the reaction. Absorbance was then measured at 650 nm. All reactions were performed in triplicate with calculated standard deviations. Serial dilutions of KH_2PO_4 were prepared to set up a standard curve to calculate the amount of phosphate released. Prism (GraphPad) was used to perform statistical analysis and generate plots of the data.

2.5 Supplemental data

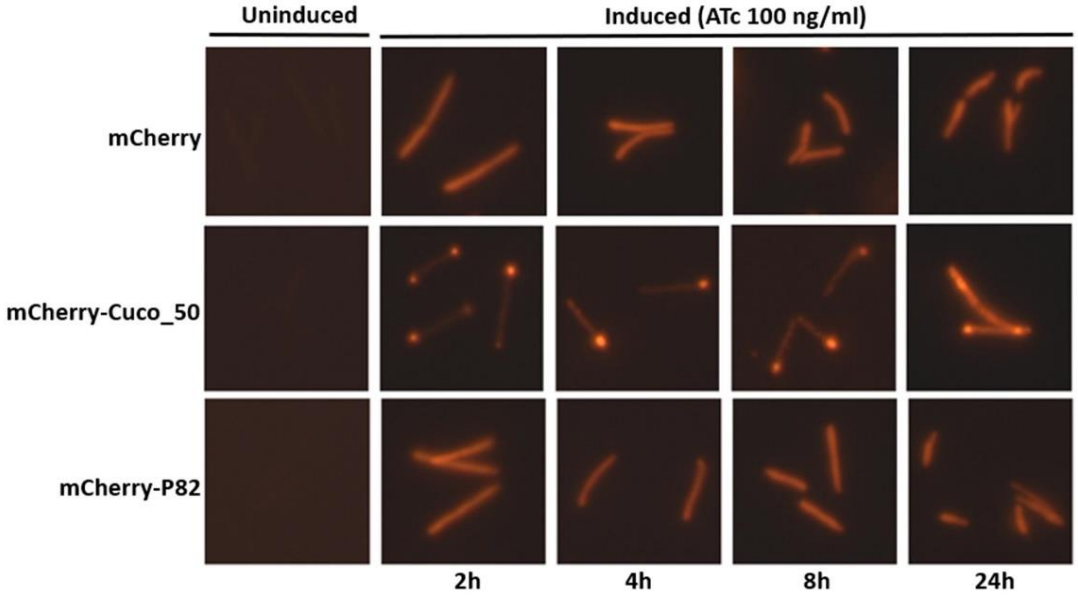

Figure S1. Phaedrus gp82 localized to the host cytoplasm.

An inducible Phaedrus gp82-mCherry fusion construct was transformed into *M. smegmatis* mc^2 155, and localization of the resulting fusion protein was monitored over 24 hours. Samples were taken at the indicated time points, and fluorescent microscopy performed. Mycobacteriophage protein Cuco gp50 (UniProt ID: G1JUM5) serves as a control as it is known to us to localize to the host cell pole. mCherry was also tested as a control for a protein known to localize to the cytoplasm and as a control for the fusion protein itself. From these data, we concluded that mCherry-P82 localizes to the cytoplasm.

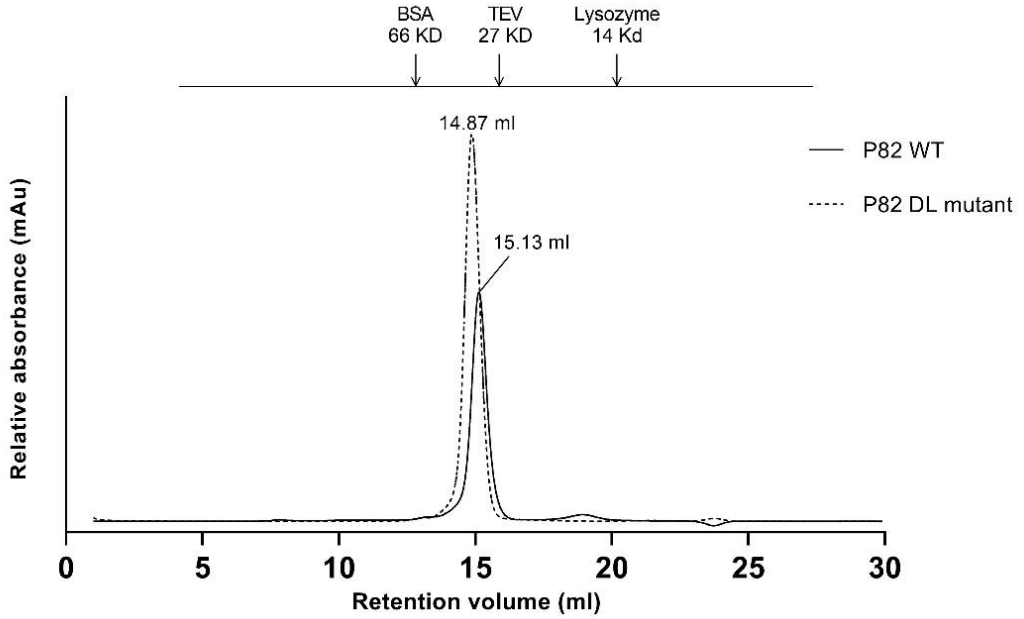

Figure S2. Phaedrus gp82 disordered loop (DL) mutant is a trimer.

Size Exclusion Chromatography of P82 WT vs. P82 disordered loop triple mutant. Analysis was performed using an analytical sizing column (Superdex S200 10/300 GL). The positions and molecular weights of the standards are indicated by arrows above the chromatogram. The elution profile suggests that Phaedrus gp82 DL mutant remains a trimer in solution.

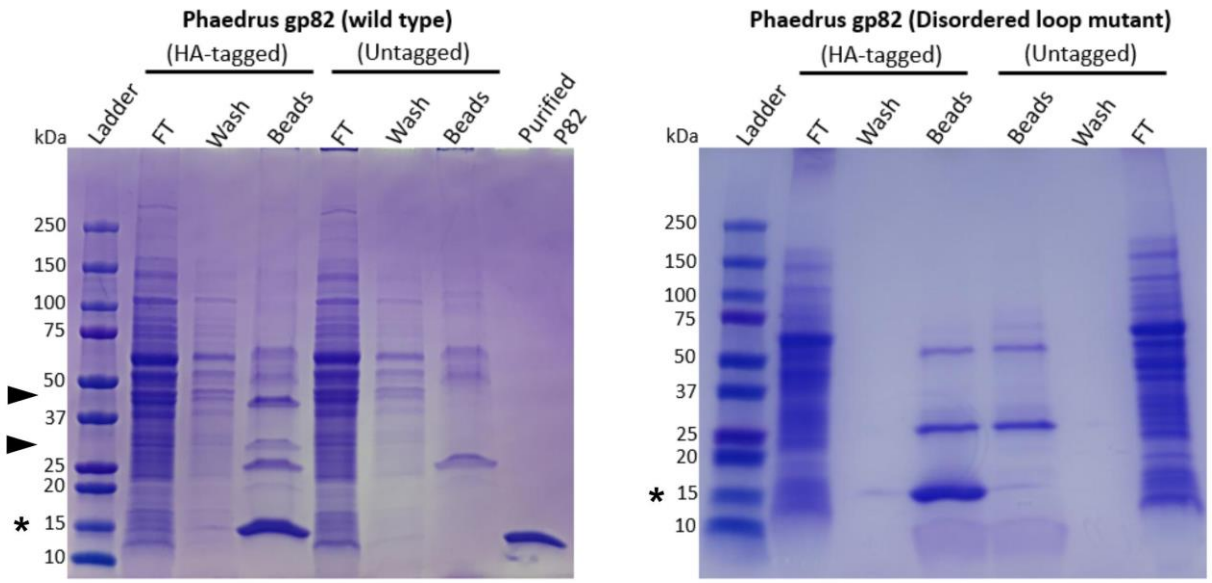

Figure S3. Phaedrus gp82 interacts with MoxR ATPase.

Cell lysates from *M. smegmatis* expressing either Phaedrus gp82 (C-terminal HA-tagged) or Phaedrus gp82 (untagged) were prepared and immunoprecipitated with anti-HA agarose beads. Flowthrough FT samples indicate unbound proteins, while the Wash sample is the last of the four washes performed. Proteins retained on the beads after washing were liberated by boiling with SDS loading buffer and are indicated as Bead samples. Samples were separated on 4–20% SDS-PAGE gel, followed by staining with Coomassie Blue. Arrows indicate the positions of proteins excised from the gel and sent for protein identification. LC/MS (Figure S4) identified the upper band as MSMEG_3147 (MoxR), while the lower band was identified as MSMEG_0970, a member of the phosphoglycerate mutase protein family. The asterisk indicates the migration of P82. On the right, the same experiment was run, loaded as indicated, but performed using HA-tagged P82 containing the disordered loop triple mutant.

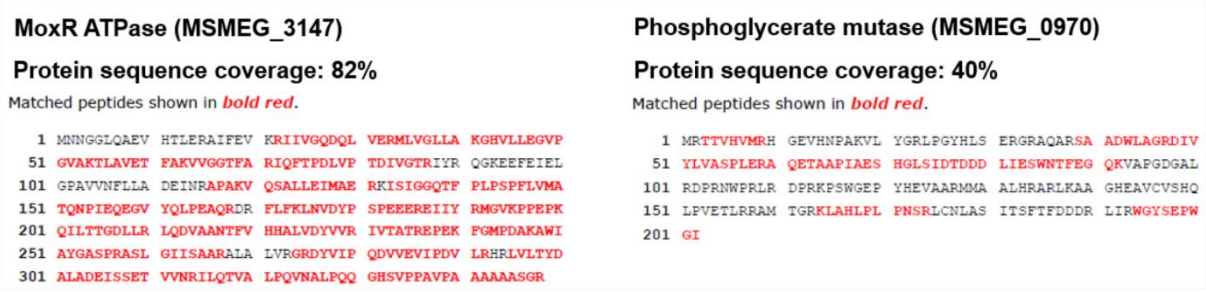

Figure S4. Identification of proteins that bind Phaedrus gp82.

Location of observed peptides from LC/MS within the primary sequences of MoxR and Phosphoglycerate mutase. Amino acids shown in red were contained within peptides observed by LC/MS with high confidence, while the black text indicates residues that were not observed in the analysis. Sequence coverage and strength were both much higher for MoxR.

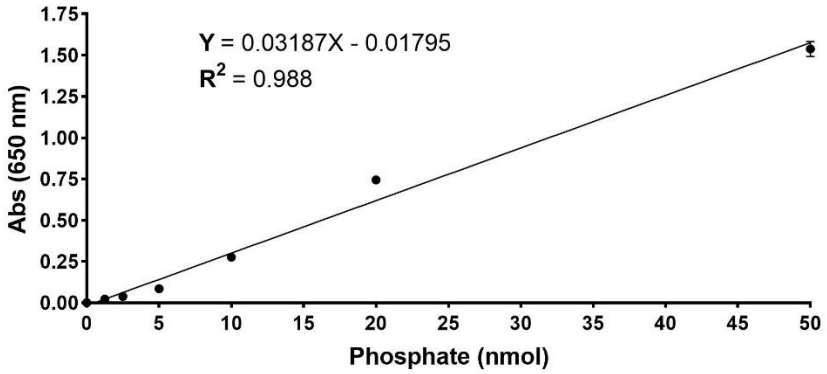

Figure S5. Standard curve for the detection of orthophosphate using the malachite green assay.

The standard curve was prepared using dilutions of KH_2PO_4. The error bars represent the standard deviation from three independent experiments.

3.0 Adephagia gp73: A potential anti-sigma factor

3.1 Introduction

Metagenomic studies have shown that mycobacteriophages are a highly diverse group of bacteriophages. Over 2230 mycobacteriophage genomes have been fully sequenced and grouped into 31 clusters and 10 singletons (Hatfull et al., 2010; Pope et al., 2015; Russell & Hatfull, 2017). The sequenced genomes include 246,272 genes sorted into 7798 protein phamilies (Cresawn et al., 2011). Approximately 171267 (69.5%) of these protein Pham are of unknown function (Pope et al., 2015). The unique diversity of these phages makes them fascinating organisms to learn more about phage biology and to gain insight into the potential novel mechanisms that regulate phage-host interactions. Given that mycobacteriophage genomes harbor a large pool of gene products of unknown function, there is excellent potential to identify unexplored pathways that regulate the phage-mycobacterial host interactions.

Phage infection can significantly alter biological processes of the host cell. Upon lytic infection, phage early gene products start interfering with host DNA replication, cell division, transcription, protein synthesis, and many other cellular pathways important for the host's normal growth (Drulis-Kawa et al., 2012). Billions of years of evolution between phage and bacterial hosts gave rise to the development of defense and counter-defense mechanisms which in turn keeps both entities in a constant evolutionary arms race (Dedrick, Jacobs-Sera, et al., 2017; Teklemariam et al., 2023). The number of mycobacteriophage proteins with unknown functions represents an extensive reservoir of unexplored proteins that could be exploited to discover potential novel mechanisms by which phage interacts with the host cell. Moreover, these proteins can potentially

guide us in identifying new mycobacterial targets for the development of antimycobacterial drugs. Not all these proteins may produce a specific phenotype when expressed in the host cell; however, those with cytotoxic or inhibitory effects are particularly interesting.

Some mycobacteriophages have been shown to infect the causative agent of tuberculosis, *Mycobacterium tuberculosis* (MTB). For example, Cluster K mycobacteriophage Angelica, CrimD, Adephagia, Anaya, TM4, and Pixie efficiently infect MTB (Pope et al., 2011). Their ability to infect MTB makes them of great research interest for exploring potential novel pathways used by these phages to infect their host. In this project, a set of non-structural proteins of phage Adephagia were determined and subjected to further investigation. The cytotoxicity assay was performed on 48 Adephagia proteins by expressing them individually in *M. smegmatis* mc^2 155. Moreover, the solubility of these proteins was examined by expressing each protein in *E. coli* as a fusion protein with N-terminal His_{10}-mRuby. Among these proteins, Adephagia gp73 was shown to be soluble and predicted to crystallize.

Adephagia gp73 was crystallized by a previous undergraduate student, Sterling Sanders. Here, we report the crystal structure of Adephagia gp73. Chemical crosslinking and size exclusion chromatography suggest that the protein is a dimer in solution. In addition, structural homology searches suggest that Adephagia gp73 is potentially an anti-sigma factor.

3.2 Results

3.2.1 Crystal structure of Adephagia gp73

To obtain information that might guide us to identify a potential function for Adephagia gp73, we crystallized and determined the protein structure using X-ray crystallography. Cubic crystals of around 100 microns on each edge grew at 4 °C over the course of several days using the vapor diffusion method. The crystals belong to space group P 43212 and contain one molecule of Adephagia gp73 within the asymmetric unit. We obtained phasing information using anomalous dispersion from 3 sulfur atoms (S-SAD) from data collected at our home source. The structure was then improved by refinement against native data at 1.0 Å resolution collected at APS beamline 31-IDD. The model was refined to an R-work/ R-free of 13.12% and 14.28%, respectively. Data collection and refinement statistics are shown in Table 2.

The electron density map shows all the residues of Adephagia gp73 except for the last 11 C-terminal residues, which are disordered. The secondary structural elements of Adephagia gp73 include four small helices (α1-α4), a small antiparallel β-sheet, and six random coils, as shown in (Figure. 12A-B). Adephagia gp73 is a small protein of 8.7 KD, and the overall structure is compact. The helices (α2- α4) form a 3-helical bundle that is connected to the β-sheet through α4, while the C-terminal coil is in contact with α1. Adephagia gp73 is an acidic protein with a theoretical pI of 4.8. The acidic residues, particularly glutamic acids, form negatively charged patches over the protein surface except for α1 and the C-terminal coil, where the positively charged patches of arginine residues are predominant. The electrostatic surface potential suggests that the negative and positive electrostatic surfaces are distributed over the entire protein surface (Figure. 13).

Structural homology search using DALI (Liisa Holm, 2022) and FATCAT (Ye & Godzik, 2004) servers generated several hits with good similarity to Adphagia gp73. However, the top hit was identified using PDBeFold (Krissinel & Henrick, 2004), which revealed significant similarity to the cytoplasmic domain of the *Pseudomonas putida* anti-sigma factor PupR (PDB: 5CAM). The anti-sigma factor is a small protein (9.5 KD) that folds into a 3-helical bundle. The structural alignment suggests that the 3-helical bundle of the two proteins is strikingly similar with an rmsd of 2.4 Å. (Figure. 14).

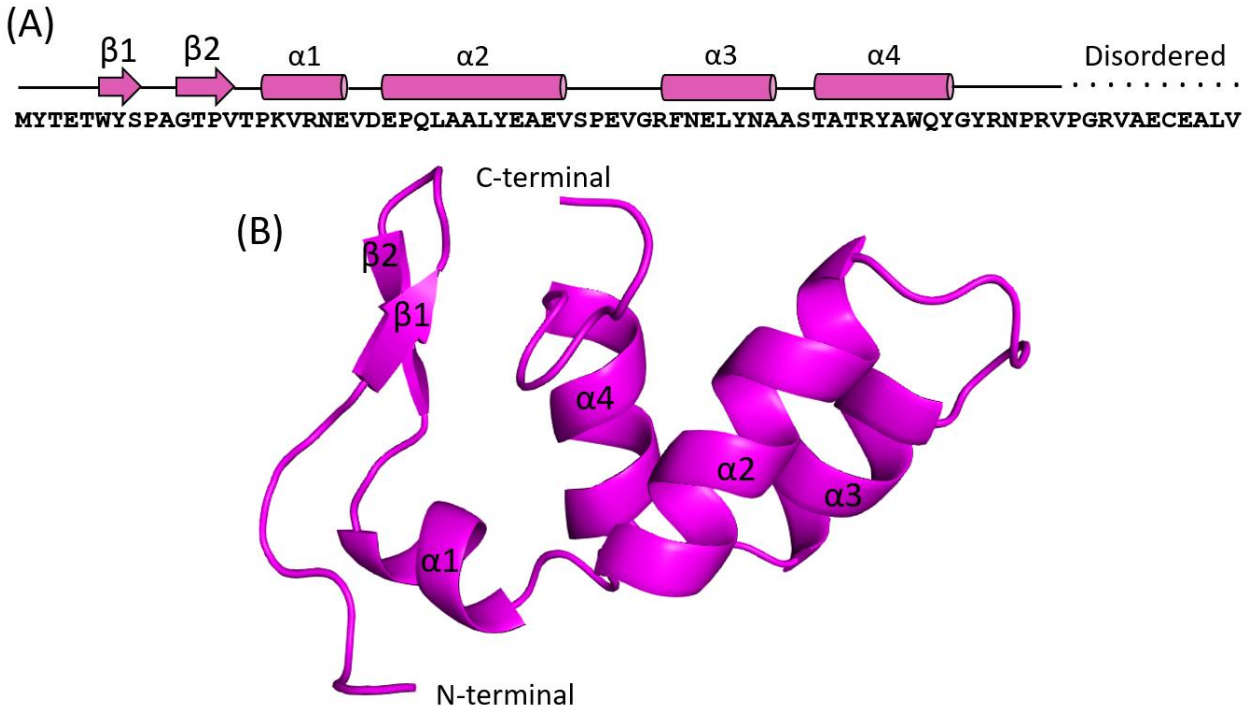

Figure 19. Crystal structure of Adephagia gp73

(A) Sequence and secondary structure representation of the crystal structure of Adephagia gp7. Dotted line represents the disordered region. **(B)** The overall structure of Adephagia gp73. The structure contains 3-helical bundle (α2-α4). The secondary structure elements are indicated. The protein contains a disordered region of eleven residues at the C-terminal end.

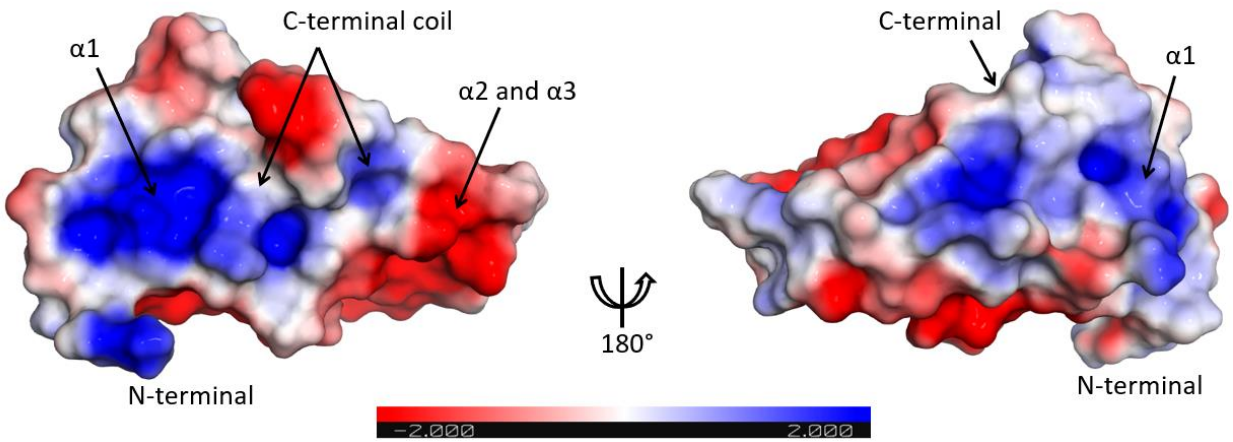

Figure 20. Electrostatic surface potential of Adephagia gp73

Electrostatic potential mapped onto the surface of Adephagia gp73. Negatively charged residues (red color) are distributed over the protein surface, while the positively charged (blue color) are predominant in the α1 and the C-terminal coil. The electrostatic potential was calculated with Pymol APBS plugin.

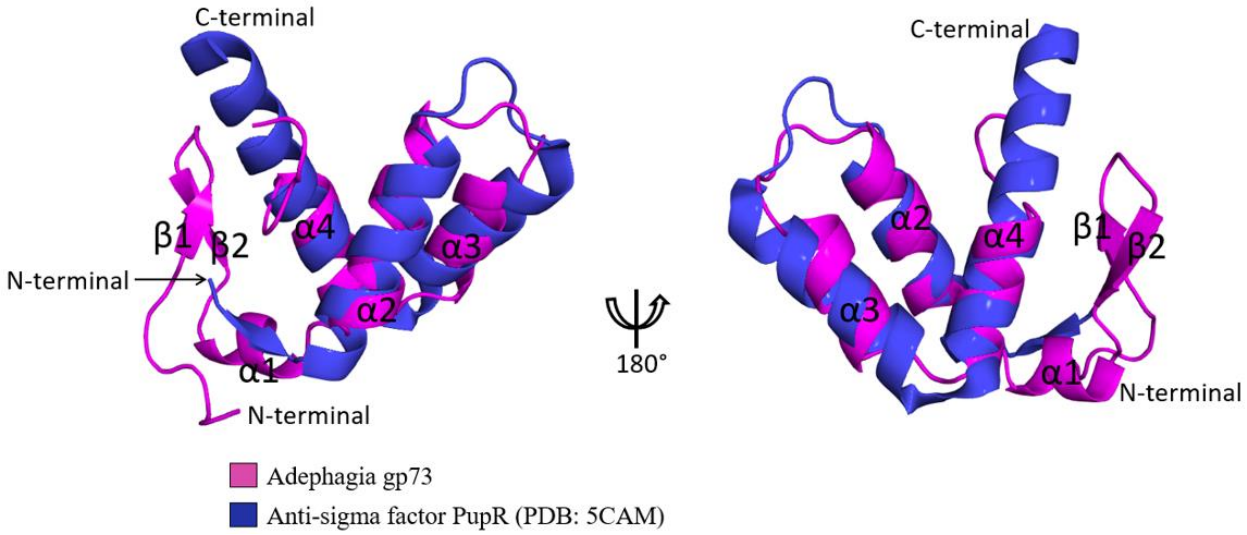

Figure 21. Superposition of Adephagia gp73 with anti-sigma factor PupR

Two views of superimposed structure of Adephagia gp 73 with anti-sigma factor PupR (PDB: 5CAM). Structural alignment shows that the 3-heical bundle of the two proteins are similar with an rmsd of 2.4 Å.

Table 3. Data collection and refinement statistics (molecular replacement)

	Sulfur anomalous	Native
Data collection		
Space group	P43212	P43212
Cell dimensions (Å)		
$a = b$	50.05	50.21
c	55.2	55.38
Resolution (Å)	37.1 - 1.70 (1.75 - 1.70)	37.20 - 1.14 (1.18-1.14)
Unique reflections	7566 (805)	24963 (2520)
R_{merge}	0.73 (0.85)	0.74 (0.66)
$I/\sigma I$	996.8 (149.7)	336.8 (64.4)
Completeness (%)	90.82 (100)	94.35 (97.22)
Redundancy	68.6 (55.10)	40.90 (27.15)
Refinement		
Resolution (Å)		21.80 - 1.01 (1.05 - 1.01)
R_{work} / $R_{free\ (\%)}$		13.12/14.28
No. atoms		666
Protein		546
Water		120
Avg B-factors		12.45
Protein		10.65
Water		20.63
R.m.s. deviations		
Bond lengths (Å)		0.008
Bond angles (°)		1.07
Ramachandran (%)		
Favored		100

Values in parentheses represent the outer resolution shell.
[a] $R_{merge} = (|(\Sigma I - \langle I \rangle)|)/ (\Sigma I)$, where $\langle I \rangle$ is the average intensity of multiple measurements.
[b] $R_{work} = \Sigma_{hkl}||F_o(hkl)|| - F_c(hkl)||/\Sigma_{hkl}|F_o(hkl)|$.
[c] R_{free} represents the cross-validation R factor for 10 % of the reflections against which the model was not refined.

3.2.2 Adephagia gp73 is a dimer in solution

We used PDBePISA (Krissinel & Henrick, 2007) to investigate how Adephagia gp73 assemblies. Crystal packing revealed that two molecules of Adephagia gp73 form a dimeric interface with a buried surface area of 533 Å^2. The monomers are assembled in such a way that

the 3-helical bundle of each monomer is in contact with the β-sheet and α1 of the other monomer (Figure. 15). However, the crystal packing calculation did not generate a significant score (CSS) for the complex formation mentioned above, meaning that the crystal packing calculation is uncertain. Therefore, we tested the ability of Adephagia gp73 to form an oligomer in solution. We carried out a protein crosslinking experiment using glutaraldehyde. We found that Adephagia gp73 forms a dimer within 1 minute of the addition of the crosslinker (Figure. 16). Higher oligomeric states start to form as the cross-linking reaction proceeds, but the highest intensity band was the one for the dimer. This suggests that Adephagia gp73 is potentially a dimer in solution. We confirmed this observation by using analytical size exclusion chromatography (SEC) Superdex 200 10/300 GL to monitor the retention volume of Adephagia gp73. We used Bovine serum albumin (BSA, 66 kDa), Tobacco Etch Virus (TEV, 27 kDa) protease, and lysozyme (14 kDa) as molecular weight standards. The SEC profile of Adephagia gp73 showed a single peak with a retention volume corresponding to an apparent molecular weight of a dimer (Figure. 17). Altogether, these experimental data may confirm that Adephagia gp73 forms a dimeric complex in solution.

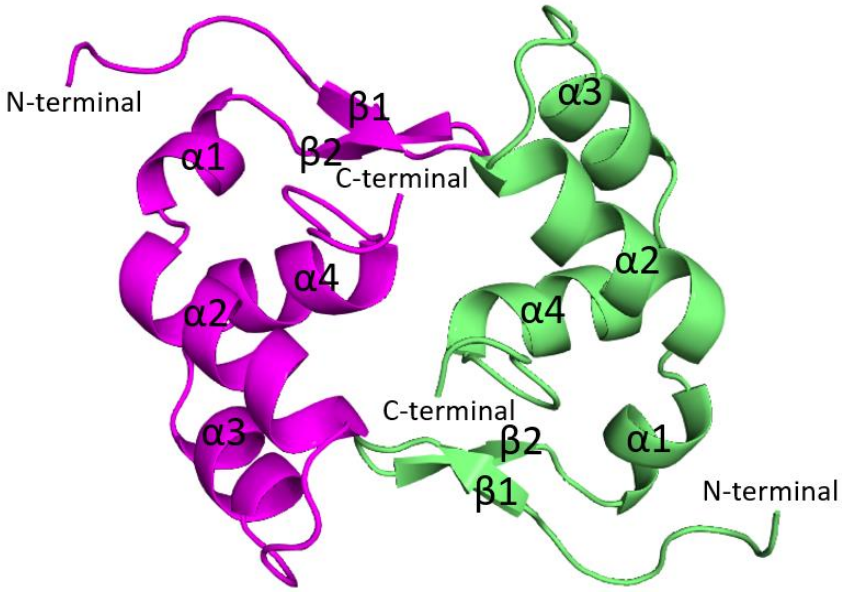

Figure 22. Crystal packing of Adephagia gp73

Crystal packing revealed that two molecules of Adephagia gp73 form a dimeric interface. The complex is stabilized by hydrophobic interactions. The secondary structure elements are indicated for each monomer.

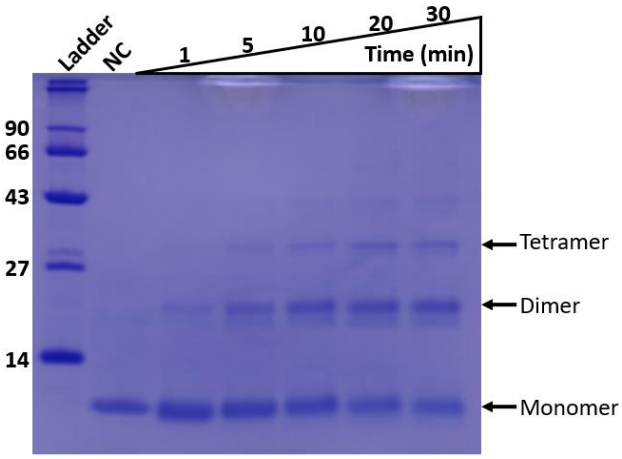

Figure 23. Adephagia gp73 oligomerization evaluated by chemical crosslinking

Adephagia gp73 oligomerization evaluated by chemical crosslinking using glutaraldehyde. SDS PAGE shows non-crosslinked (NC) Adephagia gp73 in the second lane, and crosslinked protein in the rest of the lanes. The crosslinking incubation time is indicated. The intensity of the dimer band is increasing as the crosslinking reaction proceeds, suggesting that Adephagia gp73 has a higher affinity to form a dimer.

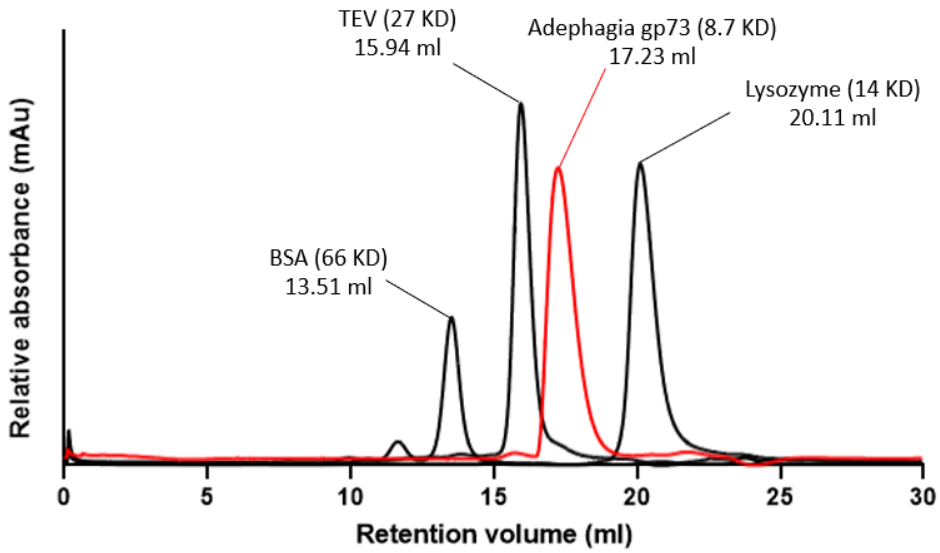

Figure 24. Adephagia gp73 oligomerization evaluated by analytical size exclusion chromatography

Superdex S200 10/300 chromatogram shows the elution volumes of Adephagia gp73 and the standard proteins bovine serum albumin (BSA), TEV protease, and lysozyme. The elution volume of Adephagia gp73 compared to the standard proteins suggests that Adephagia gp73 is a dimer.

3.2.3 AlphaFold model of Adephagia gp73-host sigma factor

Genomic analysis suggests that the genome of *Mycobacterium smegmatis* contains 26 sigma factors (Waagmeester et al., 2005). Seven of these sigma factors have been characterized experimentally (Fernandes et al., 1999; Predich et al., 1995; Wu et al., 1997). We then investigated the interaction of Adephagia gp73 with all the seven characterized sigma factors by generating AlphaFold predicted models. To ensure the validity of the predicted models, we used a control model where the interaction of the sigma-anti-sigma factor is known. Here the control was the predicted model of extracytoplasmic function (ECF) sigma factors SigH (MSMEG_1914) with its cognate anti-sigma factor RshA (MSMEG_1915). The significance of the predicted models was assessed using three physicochemical properties: complex formation significance score (CSS),

interface area, and solvation energy gain (ΔG) (Table 3). We then compare the data between generated models to evaluate how valuable the prediction process is. However, although the control (MSMEG_1915-MSMEG_1914) shows a convincing model, some other models were predicted with better physicochemical properties, such as MSMEG_1915-MSMEG_1690 and MSMEG_1915-MSMEG_0573. Moreover, the predicted models for Adephagia gp73 with all seven sigma factors had no significant physicochemical properties. This may suggest that using AlphaFold prediction is not an appropriate strategy to identify the potential sigma factor that binds Adephagia gp73.

We then sought to identify a crystal structure of the sigma factor with its cognate anti-sigma factor to identify the interface residues of the anti-sigma factor. Here, we aimed to find shared residues among the anti-sigma factor and Adephagia gp73 involved in the interaction with sigma factor. We identified the crystal structure of sigma factor SigW in complex with its cognate anti-sigma factor RsiW from *Bacillus subtilis* (PDB: 5WUR). We found that Adephagia gp73 has strong structural similarity to the anti-sigma factor RsiW (Figure 18). However, there is no evidence of shared residues involved in the protein interface between RsiW and Adephagia gp73. Thus, our strategies suggest that the sigma factor that potentially binds Adephagia gp73 should be determined experimentally.

Table 4. The calculated physicochemical properties of the predicted AphaFold models of Adephagia gp73 in complex with 7 different host sigma factors.

Anti-sigma factor (Structure 1) (CONTROL)	Sigma factor (Structure 2)	Complex Formation Significance Score (CSS)	Interface area (Å^2)	ΔG (kcal/mol)
MSMEG_1915	MSMEG_1914	1.00	1977.8	-15.1
MSMEG_1915	MSMEG_2752	1.00	1170.2	-14.6
MSMEG_1915	MSMEG_4405	1.00	2181.4	-14.6
MSMEG_1915	MSMEG_5072	1.00	1978.7	-12.7
MSMEG_1915	MSMEG_1690	1.00	1985.5	-17.4
MSMEG_1915	MSMEG_2758	0.00	580.8	-4.6
MSMEG_1915	MSMEG_0573	1.00	2166.3	-20.9
Adephagia gp73	MSMEG_1914	0.00	1161.4	-6.3
Adephagia gp73	MSMEG_2752	0.00	913.8	-1.5
Adephagia gp73	MSMEG_4405	0.00	1265.9	-4.4
Adephagia gp73	MSMEG_5072	0.00	969.0	-0.9
Adephagia gp73	MSMEG_1690	0.10	1047.8	-5.1
Adephagia gp73	MSMEG_2758	0.00	580.8	-4.6
Adephagia gp73	MSMEG_0573	0.00	1123.9	-4.2

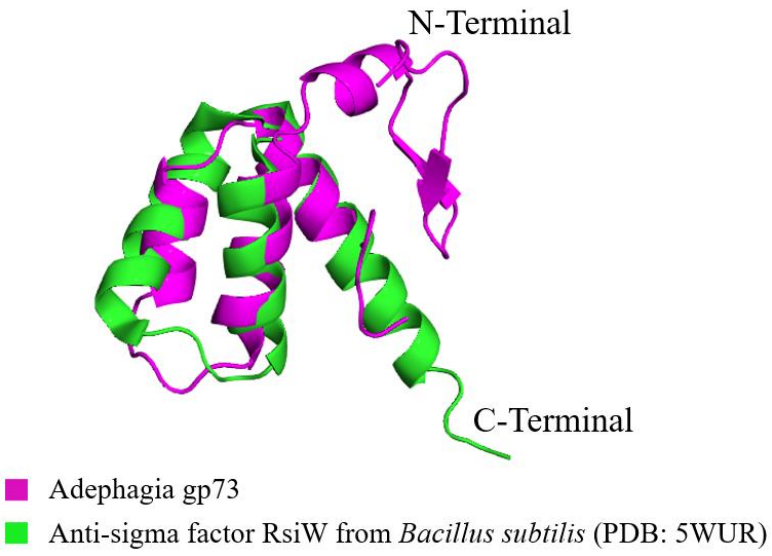

Figure 25. Superposition of Adephagia gp73 with anti-sigma factor RsiW from Bacillus subtilis.

Superimposed structure of Adephagia gp 73 with anti-sigma factor RsiW (PDB: 5WUR). Structural alignment of the two proteins yielded an rmsd of 2.5 Å.

3.3 Discussion

The sequenced genomes of mycobacteriophages provided clear evidence that their genomes are remarkably diverse. This diversity arose from billions of years of an evolutionary arms race between phages and their mycobacterial hosts (Hendrix, 2002). As bacteria defend themselves against phage attack, phages respond to the bacterial defense by developing counterattack strategies to ensure their survival (Shabbir et al., 2016).

Mycobacteriophage evolution enabled them to expand their host range to infect different species (Jacobs-Sera et al., 2012). Mycobacteriophage adephagia has been shown to infect the causative agent of tuberculosis, *Mycobacterium tuberculosis* (MTB). The emergence of MDR-MTB and XDR-MTB hampered the effective management of this deadly pathogen. This situation entails developing new strategies to cope with MTB. The ability of mycobacteriophage adephagia to infect MTB makes it of great interest to investigate the proteins encoded by its genome that could modulate the host metabolism. This may allow us to explore new pathways to cope with MTB, specifically to identify a potential target that can be used for the development of new anti-MTB drugs.

Here, Adephagia gp73 is among many other candidate proteins encoded by the genome of phage Adephagia, which have been selected for further investigation. Initially, we attempted to extract useful information about this protein by solving its structure. We successfully crystallized the proteins and solved the structure at 1.0 Å resolution. We experimentally showed that Adephagia gp 73 is a dimer in solution. Many efforts have been made to obtain useful information about the host binding partner through pulldown assay; however, our data suggest that Adephagia gp73 was not expressed in the host cell.

Structural homology search suggests that Adephagia gp73 bears similarity to the cytoplasmic domain of the *Pseudomonas putida* anti-sigma factor PupR (PDB: 5CAM). Adephagia gp73 folds into a 3-helical bundle found in many anti-sigma factors. We, therefore, proposed that Adephagia gp73 is an anti-sigma factor. We used AlphaFold prediction to identify the potential sigma factor that binds Adephagia gp73. Among the 26 predicted sigma factors encoded by *M. smegmatis* genome, 7 sigma factors have been experimentally characterized, which were used to generate AlphaFold prediction models. However, the AlphaFold models of Adephagia gp73 with all 7 experimentally revealed no significant physicochemical properties, suggesting that model prediction may not be a valuable strategy to identify the potential sigma factor. Moreover, although Adephagia gp73 is similar to the anti-sigma factor RsiW from *Bacillus subtilis*, we could not identify any shared residues between RsiW and Adephagia gp73 that are involved in the protein interface. It has been reported that anti-sigma factors share minimal sequence similarity making them difficult to identify through sequence homology search (Campbell et al., 2007). However, Adephagia gp73 shares the structural motif of a 3-helical bundle with other anti-sigma factors. This helical bundle is termed the anti-sigma domain (ASD), found in many anti-sigma factors (Campbell et al., 2007; Campbell et al., 2003). The ASD was suggested to regulate the activity of their cognate sigma factors in response to different environmental signals (Campbell et al., 2007). Adephagia gp73 was not expressed using both extrachromosomal and integrated plasmids. Further investigation will be needed to identify the optimal approach to express Adephagia gp73 in *M. smegmatis*.

3.4 Material and Methods

3.4.1 Protein purification

The coding sequence for Adephagia gp73 (UniProtKB: G1BPS7) was cloned into a modified pET28a vector, driving the expression of Adephagia gp73 with N-terminal His_{10}-mRuby2 tags and TEV protease cleavage site. The protein was expressed in LB media using Rosetta2 *E. coli*. Protein expression was induced at an optical density (OD_{600}) of 0.6 by adding 0.5 mM IPTG for 24 hours at room temperature. Cells were harvested by centrifugation and lysed by homogenization in lysis buffer containing (20 mM Tris, pH 8.0, 0.5 M NaCl, 10% glycerol, 20 mM Imidazole, and 1mM β-mercaptoethanol). The cell lysate was clarified by centrifugation at 30,000 x *g* for 30 min. Adephagia gp73 fusion protein was initially purified using nickel-affinity chromatography, followed by overnight TEV digestion to remove the His_{10}-mRuby2 tag. The resulting mixture was further purified by a second round of nickel affinity chromatography to remove the His_{10}-mRuby2 tag. Anion exchange chromatography (HiTrap Q HP, GE Healthcare) followed by size exclusion chromatography (Sephacryl-200, GE Healthcare) in a buffer containing [20 mM Tris, pH 8, 150 mM NaCl, and 1 mM β-mercaptoethanol] completed the prep. Three non-native residues (GGS) at the N-terminus remain after TEV cleavage. The resulting protein fractions were evaluated by SDS-PAGE, and the purity was > 99%. The protein fractions were subsequently pooled and concentrated in the same size exclusion chromatography buffer to approximately 10 mg/ml before crystallization.

3.4.2 Crystallization and structure determination

The crystals of Adephagia gp73 were grown at 4 °C using the sitting-drop vapor diffusion method, mixing 1µL of protein and 1µL well solution containing 30% (v/v) PEG 8000, 0.1 M Tris pH 8. Cubic-shaped crystals (~100 microns on each edge) were grown over the course of several days. Crystals were soaked in cryoprotectant containing the reservoir solution supplemented with 30% glycerol and then flash-frozen in liquid nitrogen before data collection.

High-resolution diffraction data from native Adephagia gp73 crystals were collected at beamline (31-IDD) at the APS at Argonne National Laboratory (Chicago, Illinois). The data were then processed and scaled using the autoPROC toolbox (Vonrhein et al., 2011). Phasing information was obtained using single-wavelength anomalous dispersion of native sulfur atom (S-SAD) using data collected from a single crystal at our home source at Cu-Kα. Diffraction data were then integrated, merged, and scaled with HKL2000 (Zbyszek Otwinowski & Wladek Minor, 1997). Three sulfur sites were identified using Phenix/HYSS (Adams et al., 2010; Grosse-Kunstleve & Adams, 2003), and phases were calculated using AutoSol (Terwilliger et al., 2009) as implemented within Phenix. Initial maps were improved by density modification using RESOLVE, resulting in readily interpretable electron density. Using COOT, an initial model was built into this map, and the resulting model was refined against the high-resolution native data described above. The resulting model was further refined by positional and anisotropic B-factor refinement. Model quality throughout the refinement process was monitored and validated by MolProbity. Data collection and refinement statistics are shown in Table 1.

3.4.3 Chemical Crosslinking

A chemical crosslinking experiment was performed as previously described (Slavin et al., 2020) with a few modifications. First, the protein sample at a final concentration of 1 mg/ml was incubated with glutaraldehyde crosslinking buffer containing (50 mM HEPES pH7, 100 mM NaCl). Next, the glutaraldehyde was added at a final concentration of 1%, and the final reaction volume was 80 µl. Next, the reaction mix was incubated at room temperature with agitation for 30 minutes. Samples of 10 µl each were taken after 1, 5, 10, 20, and 30 minutes of incubation, and the reaction was quenched in each sample by adding (Tris-HCl, pH8.0) to a final concentration of 50 mM. Protein samples were then separated by 15% SDS-PAGE and visualized by staining with Coomassie blue staining.

3.4.4 Analytical Size Exclusion Chromatography

Analytical size exclusion was performed using Superdex 200 10/300 column. The column was equilibrated with the running buffer containing (20 mM Tris pH 7.5, 150 mM NaCl, and 1 mM BME). Approximately 500 µl of the purified Adephagia gp73 was loaded into the column at a concentration of 2 mg/ml and running at a flow rate of 0.3 ml/min. Bovine serum albumin (BSA), TEV protease, and Lysozyme were used as protein standards. The same buffer and flow rate were used for running the protein standards.

4.0 Discussion and future directions

The sequenced mycobacteriophage genomes are, perhaps, the most extensive collection of genomes that harbor such a reservoir of unknown function genes. Although we know the fundamentals and many facts about phage biology, mycobacteriophage genomics suggests learning more about phage-host relationships. Therefore, the phage database represents a unique platform to gain a closer insight into the novel mycobacteriophage gene products and how they might interact with the host cell. However, given that the number of protein Phams is currently 7798, it is impractical to experimentally identify the function of each unknown function protein even if we select a representative protein of each Pham, not to mention the fact that the number of new Pham is constantly increasing.

Mycobacteriophages have been assorted into 31 clusters. Among these clusters, 12 have been further divided into 73 subclusters. This will make a little less than 100 clusters/subclusters (or groups) of mycobacteriophages. Given the abovementioned data, research should be focused on those phages that can infect the pathogenic *M. tuberculosis*. Therefore, identifying the phages that can infect *M. tuberculosis* would be important in determining the candidate proteins for subsequent investigation.

Superinfection exclusion seems to be a common phenomenon in a wide range of bacteriophages. To complete a successful infection, phages will need the host cell machinery; and switching to lytic growth will typically result in cell lysis. It was proposed that the toxic impact resulting from overexpression of phage protein in the host cell is due to an inactivation of specific host proteins involved in cellular pathways that are essential for other phages to infect the same bacteria (Ko & Hatfull, 2018, 2020). For example, gp52 of mycobacteriophage Fruitloop which

interacts with and inactivates the host protein Wag31. This interaction prevents superinfection by other phages in Subcluster B2, such as Rosebush and Hedgerow, which depend on Wag31 for their infection (Ko & Hatfull, 2018). It has been reported that approximately 23% of mycobacteriophage proteins have toxic effects when overexpressed in the host cell, *M. smegmatis* (Ko & Hatfull, 2020). It has been also hypothesized that many bacteriophages early proteins are required only in specific environmental conditions by providing a selective advantage for the phage (Miller et al., 2003). We therefore expect that superinfection exclusion is a common defense strategy used by phages to prevent secondary infection. The toxic effect resulting from overexpression of phage proteins in the host cell is potentially due to interfering with cellular pathways essential for other phages to infect the same bacteria.

4.1 Phaedrus gp82 project

Deriving useful information about Phaedrus gp82 through a sequence homology search was challenging. I think many thousands of mycobacteriophage proteins have no sequence homologs outside their assigned phams, and thus Phaedrus gp82 is not a unique case. However, I believe that the strategy we used to approach this project was successful. We used several methods, including biochemical, structural, and cell biology techniques, to extract information that led us to answer many biological questions relevant to this project. We were able to identify the potential function of 41 proteins belonging to pham number 2312, including Phaedrus gp82 as a representative of this pham. The 41 protein sequences in this pham are novel and of unknown function.

We identified how Phaedrus gp82 folds, and we experimentally determined segments of the protein that are important for the protein interactions and the critical residues at the binding site. We showed that the overexpression of Phaedrus gp82 in the host cell results in a small colony phenotype. In addition, we demonstrated that Phaedrus gp82 interacts with the host protein MoxR ATPase, and we showed that the disordered loop is critical for Phaedrus gp82 binding. MoxR ATPase is proposed to perform a chaperone function (Snider & Houry, 2006). Like other P-loop NTPases, MoxR ATPases bind and hydrolyze ATP molecules to gain the energy required to help proteins fold and assemble into protein complexes (Bhandari et al., 2022). We showed that the ATP hydrolysis activity of MoxR was reduced upon Phaedrus gp82 binding. This suggests that Phaedrus gp82 either blocks the ATP binding site or introduces conformational changes in MoxR that prevent ATP binding.

However, several aspects of the Phaedrus gp82 project need to be investigated further to gain deeper insight into the mechanism by which Phaedrus gp82 mediates the interaction with the host cell. We demonstrated that the disordered loop of Phaedrus p82 mediates the interaction with MoxR, and mutation in this loop blocks the small colony phenotype. The wing deletion, D38A, E43A, and D45A mutants block the small colony phenotype but they still bind MoxR. Therefore, we hypothesize that there is potentially another protein involved in the observed small colony phenotype. According to our pull-down experiment, one protein candidate would be phosphoglycerate mutase (MSMEG_0970), which was co-eluted with HA-tagged Phaedrus gp82 besides MoxR. It is possible that Phaedrus gp82, MoxR, and MSMEG_0970 form a complex and that a stable complex requires the presence of all three proteins. MSMEG_0970 has been expressed in *E. coli* and purified. However, during the purification, the protein was prone to degradation and seemed to be misfolded. Therefore, I was not able to test this hypothesis in vitro. This hypothesis

can be tested by performing two experiments of co-immunoprecipitation followed by western blot. The first experiment examines whether MoxR binds MSMEG_0970 in the absence of Phaedrus gp82. This can be achieved by overexpressing both HA-tagged MoxR and FLAG-tagged MSMEG_0970 in *M. smegmatis* and performing a pulldown experiment followed by western blot. If the binding is confirmed, then this suggests that either: 1) Phaedrus gp82 binds MoxR and that MSMEG_0970 co-eluted in our pulldown experiment, or 2) Phaedrus gp82 binds MSMEG_0970 and that MoxR co-eluted. If there is no binding, this may suggest that Phaedrus gp82 mediates the interaction between MSMEG_0970 and MoxR. However, we previously confirmed that Phaedrus gp82 binds MoxR in vitro by observing a reduction in MoxR ATP hydrolysis. Thus, this might suggest that Phaedrus gp82, MoxR, and MSMEG_0970 form a complex and that Phaedrus gp82 potentially mediates the interaction between MoxR and MSMEG_0970.

Another protein candidates for the same hypothesis would be Von Willebrand Factor Type A (VWA) domain-containing proteins (MSMEG_3148 and MSMEG_3149). MoxR genes are commonly found near genes encoding for proteins that contain VWA domain. Although not well characterized in prokaryotes, the VWA domain-containing proteins are known to mediate protein-protein interactions (Springer, 2006; Whittaker & Hynes, 2002). Due to the proximity of the genes encoding for MoxR and VWA domain-containing protein, they were suggested to function together (Snider & Houry, 2006). In *M. smegmatis*, the MoxR gene is followed by two genes encoding for VWA domain-containing protein, MEMEG_3148 and MEMEG_3149. Two individual in vitro experiments can be performed to investigate the interaction between 1) MoxR and MEMEG_3148 and 2) MoxR and MEMEG_3149. First, these two proteins can be overexpressed in *E. coli* and then purified to homogeneity. Then investigate the interaction of each protein with MoxR using native-PAGE. If the interaction is confirmed, the next step is adding

Phaedrus gp82 to the complex in a titration experiment. The outcome of this experiment might provide a novel insight into the function of both MoxR and VWA domain-containing proteins and how Phaedrus gp82 might interfere with their mechanism of action.

Another hypothesis would be that there is a ligand mediates a stable interaction between Phaedrus gp82 and MoxR, or that MoxR undergoes posttranslational modification in *M. smegmatis* which may facilitate Phaedrus gp82 binding. This hypothesis can be tested by overexpressing FLAG-MoxR in *M. smegmatis* and then purifying the FLAG-tagged MoxR from the cell lysate using anti-FLAG affinity agarose beads. After washing impurities, the interaction between Phaedrus gp82 and FLAG-tagged MoxR can be examined on native PAGE. If the interaction is confirmed, a sample of MoxR can be analyzed by mass spectrometry to detect the potential posttranslational modification. In this case, a sample of FALG-tagged MoxR overexpressed in *E. coli* will be used as a control (unmodified) to compare the difference in mass between modified and unmodified MoxR.

Finally, MoxR has an ortholog in *M. tuberculosis* which has been demonstrated to be required for the proper folding of resuscitation-promoting factor (Rpf)-interaction protein A (RipA) (Bhuwan et al., 2016). Furthermore, RipA has been shown to interact with lytic transglycosylases, RpfB, and RpfE, and all are required for proper cell division (Hett et al., 2008; Hett et al., 2007). We hypothesize that the interaction between Phaedrus gp82 and MoxR impedes the proper function of MoxR as a chaperone, which may result in incorrectly folded RipA and consequently reduce cellular growth. To test this hypothesis, first we need to confirm the interaction between MoxR and RipA. Similar approach used in Bhuwan et al can be performed with a little modification. Both HA-tagged MoxR and FLAG-tagged RipA are overexpressed (using integrated vectors) in *M. smegmatis* and perform pulldown experiments followed by

western blot. If the interaction is confirmed, the follow-up experiment would be to overexpress HA-tagged MoxR, FLAG-tagged RipA, and untagged Phaedrus gp82 (extrachromosomal vector) followed by pulldown and western blot. If no interaction is detected between MoxR and RipA, it may strongly suggest that Phaedrus gp82 blocks the interaction between MoxR and RipA and that the small colony phenotype could result from incorrectly folded RipA.

4.2 Adephagia gp73 project

Adephagia gp73 has no sequence homology to any protein outside the pham where it belongs. However, its structure suggests that it has an anti-sigma domain (ASD), a structural motif found in many anti-sigma factors. Many attempts have been made to identify the binding partner using pull-down assays. However, the protein was not expressed in *M. smegmatis* using both extrachromosomal and integrated plasmids under optimal environmental conditions. Upon determining the optimal approach to express Adephagia gp73, the binding partner (the host sigma factor) can be identified using pull-down experiment. The next step would be to perform RNAseq to identify genes whose transcription might be regulated by Adephagia gp73. This can be achieved by comparing the total RNA from *M. smegmatis* expressing Adephagia gp73 to *M. smegmatis* wild type.

Bibliography

Abedon, S. T. (2015). Bacteriophage secondary infection. *Virol Sin*, *30*(1), 3-10. https://doi.org/10.1007/s12250-014-3547-2

Ackermann, H. W. (1998). Tailed bacteriophages: the order caudovirales. *Adv Virus Res*, *51*, 135-201. https://doi.org/10.1016/s0065-3527(08)60785-x

Ackermann, H. W. (2007). 5500 Phages examined in the electron microscope. *Arch Virol*, *152*(2), 227-243. https://doi.org/10.1007/s00705-006-0849-1

Ackermann, H. W., & Prangishvili, D. (2012). Prokaryote viruses studied by electron microscopy. *Arch Virol*, *157*(10), 1843-1849. https://doi.org/10.1007/s00705-012-1383-y

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., & Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*, *66*(Pt 2), 213-221. https://doi.org/10.1107/s0907444909052925

Al-Humadi, H. W., Al-Saigh, R. J., & Al-Humadi, A. W. (2017). Addressing the Challenges of Tuberculosis: A Brief Historical Account. *Front Pharmacol*, *8*, 689. https://doi.org/10.3389/fphar.2017.00689

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol*, *215*(3), 403-410. https://doi.org/10.1016/s0022-2836(05)80360-2

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, *25*(17), 3389-3402. https://doi.org/10.1093/nar/25.17.3389

Ando, H., Lemire, S., Pires, D. P., & Lu, T. K. (2015). Engineering Modular Viral Scaffolds for Targeted Bacterial Population Editing. *Cell Syst*, *1*(3), 187-196. https://doi.org/10.1016/j.cels.2015.08.013

Arnaud, C.-A., Effantin, G., Vivès, C., Engilberge, S., Bacia-Verloop, M., Boulanger, P., Girard, E., Schoehn, G., & Breyton, C. (2017). Bacteriophage T5 tail tube structure suggests a trigger mechanism for Siphoviridae DNA ejection. *Nature Communications*, *8*(1), 1953. https://doi.org/10.1038/s41467-017-02049-3

Ashelford, K. E., Day, M. J., & Fry, J. C. (2003). Elevated abundance of bacteriophage infecting bacteria in soil. *Appl Environ Microbiol*, *69*(1), 285-289. https://doi.org/10.1128/AEM.69.1.285-289.2003

Bae, T., Baba, T., Hiramatsu, K., & Schneewind, O. (2006). Prophages of Staphylococcus aureus Newman and their contribution to virulence. *Mol Microbiol*, *62*(4), 1035-1047. https://doi.org/10.1111/j.1365-2958.2006.05441.x

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, *315*(5819), 1709-1712. https://doi.org/10.1126/science.1138140

Barreteau, H., Kovac, A., Boniface, A., Sova, M., Gobec, S., & Blanot, D. (2008). Cytoplasmic steps of peptidoglycan biosynthesis. *FEMS Microbiol Rev*, *32*(2), 168-207. https://doi.org/10.1111/j.1574-6976.2008.00104.x

Bergh, O., Borsheim, K. Y., Bratbak, G., & Heldal, M. (1989). High abundance of viruses found in aquatic environments. *Nature*, *340*(6233), 467-468. https://doi.org/10.1038/340467a0

Bertozzi Silva, J., Storms, Z., & Sauvageau, D. (2016). Host receptors for bacteriophage adsorption. *FEMS Microbiol Lett*, *363*(4). https://doi.org/10.1093/femsle/fnw002

Bertozzi Silva, J., Storms, Z., & Sauvageau, D. (2016). Host receptors for bacteriophage adsorption. *FEMS Microbiology Letters*, *363*(4). https://doi.org/10.1093/femsle/fnw002

Bhandari, V., Van Ommen, D. A. J., Wong, K. S., & Houry, W. A. (2022). Analysis of the Evolution of the MoxR ATPases. *The Journal of Physical Chemistry A*, *126*(29), 4734-4746. https://doi.org/10.1021/acs.jpca.2c02554

Bhuwan, M., Arora, N., Sharma, A., Khubaib, M., Pandey, S., Chaudhuri, T. K., Hasnain, S. E., & Ehtesham, N. Z. (2016). Interaction of Mycobacterium tuberculosis Virulence Factor RipA with Chaperone MoxR1 Is Required for Transport through the TAT Secretion System. *mBio*, *7*(2), e02259. https://doi.org/10.1128/mBio.02259-15

Bi, E. F., & Lutkenhaus, J. (1991). FtsZ ring structure associated with division in Escherichia coli. *Nature*, *354*(6349), 161-164. https://doi.org/10.1038/354161a0

Bikard, D., & Marraffini, L. A. (2012). Innate and adaptive immunity in bacteria: mechanisms of programmed genetic variation to fight bacteriophages. *Curr Opin Immunol*, *24*(1), 15-20. https://doi.org/10.1016/j.coi.2011.10.005

Bondy-Denomy, J., Pawluk, A., Maxwell, K. L., & Davidson, A. R. (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*, *493*(7432), 429-432. https://doi.org/10.1038/nature11723

Britton, W. J., & Triccas, J. A. (2008). The Constituents of the Cell Envelope and Their Impact on the Host Immune System. In *The Mycobacterial Cell Envelope* (pp. 249-270). https://doi.org/https://doi.org/10.1128/9781555815783.ch16

Brussow, H., Canchaya, C., & Hardt, W. D. (2004). Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev*, *68*(3), 560-602, table of contents. https://doi.org/10.1128/MMBR.68.3.560-602.2004

Buchmeier, N. A., Newton, G. L., & Fahey, R. C. (2006). A mycothiol synthase mutant of Mycobacterium tuberculosis has an altered thiol-disulfide content and limited tolerance to stress. *J Bacteriol*, *188*(17), 6245-6252. https://doi.org/10.1128/jb.00393-06

Bushman, F. (2002). *Lateral DNA Transfer: Mechanisms and Consequences*. Cold Spring Harbor Laboratory Press. https://books.google.com/books?id=8oIaOGHURroC

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*, 421. https://doi.org/10.1186/1471-2105-10-421

Campbell, E. A., Greenwell, R., Anthony, J. R., Wang, S., Lim, L., Das, K., Sofia, H. J., Donohue, T. J., & Darst, S. A. (2007). A conserved structural module regulates transcriptional responses to diverse stress signals in bacteria. *Mol Cell*, *27*(5), 793-805. https://doi.org/10.1016/j.molcel.2007.07.009

Campbell, E. A., Tupy, J. L., Gruber, T. M., Wang, S., Sharp, M. M., Gross, C. A., & Darst, S. A. (2003). Crystal structure of Escherichia coli sigmaE with the cytoplasmic domain of its anti-sigma RseA. *Mol Cell*, *11*(4), 1067-1078. https://doi.org/10.1016/s1097-2765(03)00148-5

Canchaya, C., Fournous, G., & Brussow, H. (2004). The impact of prophages on bacterial chromosomes. *Mol Microbiol*, *53*(1), 9-18. https://doi.org/10.1111/j.1365-2958.2004.04113.x

Canchaya, C., Proux, C., Fournous, G., Bruttin, A., & Brüssow, H. (2003). Prophage genomics. *Microbiol Mol Biol Rev*, *67*(2), 238-276, table of contents. https://doi.org/10.1128/mmbr.67.2.238-276.2003

Chen, M., Zhang, L., Abdelgader, S. A., Yu, L., Xu, J., Yao, H., Lu, C., & Zhang, W. (2017). Alterations in gp37 Expand the Host Range of a T4-Like Phage. *Appl Environ Microbiol*, *83*(23). https://doi.org/10.1128/aem.01576-17

Clark, J. R., & March, J. B. (2004). Bacterial viruses as human vaccines? *Expert Review of Vaccines*, *3*(4), 463-476. https://doi.org/10.1586/14760584.3.4.463

Cooke, F. J., Wain, J., Fookes, M., Ivens, A., Thomson, N., Brown, D. J., Threlfall, E. J., Gunn, G., Foster, G., & Dougan, G. (2007). Prophage sequences defining hot spots of genome variation in Salmonella enterica serovar Typhimurium can be used to discriminate between field isolates. *J Clin Microbiol*, *45*(8), 2590-2598. https://doi.org/10.1128/jcm.00729-07

Corbett, E. L., Watt, C. J., Walker, N., Maher, D., Williams, B. G., Raviglione, M. C., & Dye, C. (2003). The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. *Arch Intern Med*, *163*(9), 1009-1021. https://doi.org/10.1001/archinte.163.9.1009

Cosivi, O., Grange, J. M., Daborn, C. J., Raviglione, M. C., Fujikura, T., Cousins, D., Robinson, R. A., Huchzermeyer, H. F., de Kantor, I., & Meslin, F. X. (1998). Zoonotic tuberculosis

due to Mycobacterium bovis in developing countries. *Emerg Infect Dis*, *4*(1), 59-70. https://doi.org/10.3201/eid0401.980108

Cresawn, S. G., Bogel, M., Day, N., Jacobs-Sera, D., Hendrix, R. W., & Hatfull, G. F. (2011). Phamerator: a bioinformatic tool for comparative bacteriophage genomics. *BMC Bioinformatics*, *12*(1), 395. https://doi.org/10.1186/1471-2105-12-395

Cumby, N., Edwards, A. M., Davidson, A. R., & Maxwell, K. L. (2012). The bacteriophage HK97 gp15 moron element encodes a novel superinfection exclusion protein. *J Bacteriol*, *194*(18), 5012-5019. https://doi.org/10.1128/jb.00843-12

D'Angelo, E., Crutchfield, J., & Vandiviere, M. (2001). Rapid, sensitive, microscale determination of phosphate in water and soil. *J Environ Qual*, *30*(6), 2206-2209. https://doi.org/10.2134/jeq2001.2206

de Boer, P. A. (2010). Advances in understanding E. coli cell fission. *Curr Opin Microbiol*, *13*(6), 730-737. https://doi.org/10.1016/j.mib.2010.09.015

Dedrick, R. M., Freeman, K. G., Nguyen, J. A., Bahadirli-Talbott, A., Cardin, M. E., Cristinziano, M., Smith, B. E., Jeong, S., Ignatius, E. H., Lin, C. T., Cohen, K. A., & Hatfull, G. F. (2022). Nebulized Bacteriophage in a Patient With Refractory Mycobacterium abscessus Lung Disease. *Open Forum Infect Dis*, *9*(7), ofac194. https://doi.org/10.1093/ofid/ofac194

Dedrick, R. M., Guerrero-Bustamante, C. A., Garlena, R. A., Russell, D. A., Ford, K., Harris, K., Gilmour, K. C., Soothill, J., Jacobs-Sera, D., Schooley, R. T., Hatfull, G. F., & Spencer, H. (2019). Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant Mycobacterium abscessus. *Nat Med*, *25*(5), 730-733. https://doi.org/10.1038/s41591-019-0437-z

Dedrick, R. M., Jacobs-Sera, D., Bustamante, C. A., Garlena, R. A., Mavrich, T. N., Pope, W. H., Reyes, J. C., Russell, D. A., Adair, T., Alvey, R., Bonilla, J. A., Bricker, J. S., Brown, B. R., Byrnes, D., Cresawn, S. G., Davis, W. B., Dickson, L. A., Edgington, N. P., Findley, A. M., . . . Hatfull, G. F. (2017). Prophage-mediated defence against viral attack and viral counter-defence. *Nat Microbiol*, *2*, 16251. https://doi.org/10.1038/nmicrobiol.2016.251

Dedrick, R. M., Marinelli, L. J., Newton, G. L., Pogliano, K., Pogliano, J., & Hatfull, G. F. (2013). Functional requirements for bacteriophage growth: gene essentiality and expression in mycobacteriophage Giles. *Mol Microbiol*, *88*(3), 577-589. https://doi.org/10.1111/mmi.12210

Dedrick, R. M., Mavrich, T. N., Ng, W. L., & Hatfull, G. F. (2017). Expression and evolutionary patterns of mycobacteriophage D29 and its temperate close relatives. *BMC Microbiol*, *17*(1), 225. https://doi.org/10.1186/s12866-017-1131-2

Dennehy, J. J., & Abedon, S. T. (2020). Adsorption: Phage Acquisition of Bacteria. In *Bacteriophages* (pp. 1-25). https://doi.org/10.1007/978-3-319-40598-8_2-1

Depping, R., Lohaus, C., Meyer, H. E., & Rüger, W. (2005). The mono-ADP-ribosyltransferases Alt and ModB of bacteriophage T4: target proteins identified. *Biochem Biophys Res Commun*, *335*(4), 1217-1223. https://doi.org/10.1016/j.bbrc.2005.08.023

DesJardin, L. E., Kaufman, T. M., Potts, B., Kutzbach, B., Yi, H., & Schlesinger, L. S. (2002). Mycobacterium tuberculosis-infected human macrophages exhibit enhanced cellular adhesion with increased expression of LFA-1 and ICAM-1 and reduced expression and/or function of complement receptors, FcgammaRII and the mannose receptor. *Microbiology (Reading)*, *148*(Pt 10), 3161-3171. https://doi.org/10.1099/00221287-148-10-3161

Di Capua, C. B., Belardinelli, J. M., Carignano, H. A., Buchieri, M. V., Suarez, C. A., & Morbidoni, H. R. (2022). Unveiling the Biosynthetic Pathway for Short Mycolic Acids in Nontuberculous Mycobacteria: Mycobacterium smegmatis MSMEG_4301 and Its Ortholog Mycobacterium abscessus MAB_1915 Are Essential for the Synthesis of α'-Mycolic Acids. *Microbiol Spectr*, *10*(4), e0128822. https://doi.org/10.1128/spectrum.01288-22

Dieppedale, J., Sobral, D., Dupuis, M., Dubail, I., Klimentova, J., Stulik, J., Postic, G., Frapy, E., Meibom, K. L., Barel, M., & Charbit, A. (2011). Identification of a putative chaperone involved in stress resistance and virulence in Francisella tularensis. *Infect Immun*, *79*(4), 1428-1439. https://doi.org/10.1128/iai.01012-10

Djebara, S., Maussen, C., De Vos, D., Merabishvili, M., Damanet, B., Pang, K. W., De Leenheer, P., Strachinaru, I., Soentjens, P., & Pirnay, J. P. (2019). Processing Phage Therapy Requests in a Brussels Military Hospital: Lessons Identified. *Viruses*, *11*(3). https://doi.org/10.3390/v11030265

Dooley, K. E., & Chaisson, R. E. (2009). Tuberculosis and diabetes mellitus: convergence of two epidemics. *Lancet Infect Dis*, *9*(12), 737-746. https://doi.org/10.1016/s1473-3099(09)70282-8

Drulis-Kawa, Z., Majkowska-Skrobek, G., Maciejewska, B., Delattre, A. S., & Lavigne, R. (2012). Learning from bacteriophages - advantages and limitations of phage and phage-encoded protein applications. *Curr Protein Pept Sci*, *13*(8), 699-722. https://doi.org/10.2174/138920312804871193

Dunne, M., Hupfeld, M., Klumpp, J., & Loessner, M. J. (2018). Molecular Basis of Bacterial Host Interactions by Gram-Positive Targeting Bacteriophages. *Viruses*, *10*(8). https://doi.org/10.3390/v10080397

Duplessis, M., Lévesque, C. M., & Moineau, S. (2006). Characterization of Streptococcus thermophilus host range phage mutants. *Appl Environ Microbiol*, *72*(4), 3036-3041. https://doi.org/10.1128/aem.72.4.3036-3041.2006

Dye, C., Scheele, S., Dolin, P., Pathania, V., & Raviglione, M. C. (1999). Consensus statement. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project. *Jama*, *282*(7), 677-686. https://doi.org/10.1001/jama.282.7.677

Ehrt, S., & Schnappinger, D. (2009). Mycobacterial survival strategies in the phagosome: defence against host stresses. *Cell Microbiol*, *11*(8), 1170-1178. https://doi.org/10.1111/j.1462-5822.2009.01335.x

Eklund, M. W., Poysky, F. T., Reed, S. M., & Smith, C. A. (1971). Bacteriophage and the toxigenicity of Clostridium botulinum type C. *Science*, *172*(3982), 480-482. https://doi.org/10.1126/science.172.3982.480

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*, *66*(Pt 4), 486-501. https://doi.org/10.1107/s0907444910007493

Fan, X., Duan, X., Tong, Y., Huang, Q., Zhou, M., Wang, H., Zeng, L., Young, R. F., 3rd, & Xie, J. (2016). The Global Reciprocal Reprogramming between Mycobacteriophage SWU1 and Mycobacterium Reveals the Molecular Strategy of Subversion and Promotion of Phage Infection. *Front Microbiol*, *7*, 41. https://doi.org/10.3389/fmicb.2016.00041

Fernandes, N. D., Wu, Q. L., Kong, D., Puyang, X., Garg, S., & Husson, R. N. (1999). A mycobacterial extracytoplasmic sigma factor involved in survival following heat shock and oxidative stress. *J Bacteriol*, *181*(14), 4266-4274. https://doi.org/10.1128/jb.181.14.4266-4274.1999

Fernandes, S., & Sao-Jose, C. (2018). Enzymes and Mechanisms Employed by Tailed Bacteriophages to Breach the Bacterial Cell Barriers. *Viruses*, *10*(8). https://doi.org/10.3390/v10080396

Ferry, T., Boucher, F., Fevre, C., Perpoint, T., Chateau, J., Petitjean, C., Josse, J., Chidiac, C., L'Hostis, G., Leboucher, G., & Laurent, F. (2018). Innovations for the treatment of a complex bone and joint infection due to XDR Pseudomonas aeruginosa including local application of a selected cocktail of bacteriophages. *J Antimicrob Chemother*, *73*(10), 2901-2903. https://doi.org/10.1093/jac/dky263

Ferry, T., Leboucher, G., Fevre, C., Herry, Y., Conrad, A., Josse, J., Batailler, C., Chidiac, C., Medina, M., Lustig, S., & Laurent, F. (2018). Salvage Debridement, Antibiotics and Implant Retention ("DAIR") With Local Injection of a Selected Cocktail of Bacteriophages: Is It an Option for an Elderly Patient With Relapsing Staphylococcus aureus Prosthetic-Joint Infection? *Open Forum Infect Dis*, *5*(11), ofy269. https://doi.org/10.1093/ofid/ofy269

Fineran, P. C., Blower, T. R., Foulds, I. J., Humphreys, D. P., Lilley, K. S., & Salmond, G. P. (2009). The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *Proc Natl Acad Sci U S A*, *106*(3), 894-899. https://doi.org/10.1073/pnas.0808832106

Flaherty, J. E., Harbaugh, B. K., Jones, J. B., Somodi, G. C., & Jackson, L. E. (2001). H-mutant Bacteriophages as a Potential Biocontrol of Bacterial Blight of Geranium. *HortScience HortSci*, *36*(1), 98-100. https://doi.org/10.21273/HORTSCI.36.1.98

Fokine, A., & Rossmann, M. G. (2014). Molecular architecture of tailed double-stranded DNA phages. *Bacteriophage*, *4*(1), e28281. https://doi.org/10.4161/bact.28281

Folimonova, S. Y. (2012). Superinfection exclusion is an active virus-controlled function that requires a specific viral protein. *J Virol*, *86*(10), 5554-5561. https://doi.org/10.1128/jvi.00310-12

Ford, M. E., Sarkis, G. J., Belanger, A. E., Hendrix, R. W., & Hatfull, G. F. (1998). Genome structure of mycobacteriophage D29: implications for phage evolution. *J Mol Biol*, *279*(1), 143-164. https://doi.org/10.1006/jmbi.1997.1610

Ford, M. E., Stenstrom, C., Hendrix, R. W., & Hatfull, G. F. (1998). Mycobacteriophage TM4: genome structure and gene expression. *Tuber Lung Dis*, *79*(2), 63-73. https://doi.org/10.1054/tuld.1998.0007

Fortier, L. C., & Sekulovic, O. (2013). Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence*, *4*(5), 354-365. https://doi.org/10.4161/viru.24498

Frieden, T. R., Sterling, T. R., Munsiff, S. S., Watt, C. J., & Dye, C. (2003). Tuberculosis. *Lancet*, *362*(9387), 887-899. https://doi.org/10.1016/s0140-6736(03)14333-4

Friedman, D. I., Olson, E. R., Georgopoulos, C., Tilly, K., Herskowitz, I., & Banuett, F. (1984). Interactions of bacteriophage and host macromolecules in the growth of bacteriophage lambda. *Microbiol Rev*, *48*(4), 299-325. https://doi.org/10.1128/mr.48.4.299-325.1984

Galli, G., & Saleh, M. (2020). Immunometabolism of Macrophages in Bacterial Infections. *Front Cell Infect Microbiol*, *10*, 607650. https://doi.org/10.3389/fcimb.2020.607650

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*, *31*(13), 3784-3788. https://doi.org/10.1093/nar/gkg563

Gehre, F., Kumar, S., Kendall, L., Ejo, M., Secka, O., Ofori-Anyinam, B., Abatih, E., Antonio, M., Berkvens, D., & de Jong, B. C. (2016). A Mycobacterial Perspective on Tuberculosis in West Africa: Significant Geographical Variation of M. africanum and Other M. tuberculosis Complex Lineages. *PLoS Negl Trop Dis*, *10*(3), e0004408. https://doi.org/10.1371/journal.pntd.0004408

Ghazaei, C. (2018). Mycobacterium tuberculosis and lipids: Insights into molecular mechanisms from persistence to virulence. *J Res Med Sci*, *23*, 63. https://doi.org/10.4103/jrms.JRMS_904_17

Gill, D. M., Uchida, T., & Singer, R. A. (1972). Expression of diphtheria toxin genes carried by integrated and nonintegrated phage beta. *Virology*, *50*(3), 664-668. https://doi.org/https://doi.org/10.1016/0042-6822(72)90420-5

Goode, D., Allen, V. M., & Barrow, P. A. (2003). Reduction of experimental Salmonella and Campylobacter contamination of chicken skin by application of lytic bacteriophages. *Appl Environ Microbiol*, *69*(8), 5032-5036. https://doi.org/10.1128/aem.69.8.5032-5036.2003

Googins, M. R., Woghiren-Afegbua, A. O., Calderon, M., St Croix, C. M., Kiselyov, K. I., & VanDemark, A. P. (2020). Structural and functional divergence of GDAP1 from the glutathione S-transferase superfamily. *Faseb j*, *34*(5), 7192-7207. https://doi.org/10.1096/fj.202000110R

Grosse-Kunstleve, R. W., & Adams, P. D. (2003). Substructure search procedures for macromolecular structures. *Acta Crystallogr D Biol Crystallogr*, *59*(Pt 11), 1966-1973. https://doi.org/10.1107/s0907444903018043

Guttman, B., Raya, R., & Kutter, E. (2005). Basic Phage Biology. In K. E & S. A (Eds.), *Bacteriophages: Biology and Application* (pp. 29-66). https://doi.org/10.1201/9780203491751.CH3

Haeusser, D. P., Hoashi, M., Weaver, A., Brown, N., Pan, J., Sawitzke, J. A., Thomason, L. C., Court, D. L., & Margolin, W. (2014). The Kil peptide of bacteriophage λ blocks Escherichia coli cytokinesis via ZipA-dependent inhibition of FtsZ assembly. *PLoS Genet*, *10*(3), e1004217. https://doi.org/10.1371/journal.pgen.1004217

Halleran, A., Clamons, S., & Saha, M. (2015). Transcriptomic Characterization of an Infection of Mycobacterium smegmatis by the Cluster A4 Mycobacteriophage Kampy. *PLoS One*, *10*(10), e0141100. https://doi.org/10.1371/journal.pone.0141100

Hanson, P. I., & Whiteheart, S. W. (2005). AAA+ proteins: have engine, will work. *Nature Reviews Molecular Cell Biology*, *6*(7), 519-529. https://doi.org/10.1038/nrm1684

Hartley, M. A., Ronet, C., & Fasel, N. (2012). Backseat drivers: the hidden influence of microbial viruses on disease. *Curr Opin Microbiol*, *15*(4), 538-545. https://doi.org/10.1016/j.mib.2012.05.011

Harvey, H., Bondy-Denomy, J., Marquis, H., Sztanko, K. M., Davidson, A. R., & Burrows, L. L. (2018). Pseudomonas aeruginosa defends against phages through type IV pilus glycosylation. *Nature Microbiology*, *3*(1), 47-52. https://doi.org/10.1038/s41564-017-0061-y

Hatfull, G. F. (2014). Molecular Genetics of Mycobacteriophages. *Microbiol Spectr*, *2*(2), 1-36.

Hatfull, G. F. (2018). Mycobacteriophages. *Microbiol Spectr*, *6*(5). https://doi.org/10.1128/microbiolspec.GPP3-0026-2018

Hatfull, G. F. (2022). Mycobacteriophages: From Petri dish to patient. *PLOS Pathogens*, *18*(7), e1010602. https://doi.org/10.1371/journal.ppat.1010602

Hatfull, G. F. (2022). Mycobacteriophages: From Petri dish to patient. *PLoS Pathog*, *18*(7), e1010602. https://doi.org/10.1371/journal.ppat.1010602

Hatfull, G. F. (2023). Phage Therapy for Nontuberculous Mycobacteria: Challenges and Opportunities. *Pulm Ther*, *9*(1), 91-107. https://doi.org/10.1007/s41030-022-00210-y

Hatfull, G. F., Jacobs-Sera, D., Lawrence, J. G., Pope, W. H., Russell, D. A., Ko, C. C., Weber, R. J., Patel, M. C., Germane, K. L., Edgar, R. H., Hoyte, N. N., Bowman, C. A., Tantoco, A. T., Paladin, E. C., Myers, M. S., Smith, A. L., Grace, M. S., Pham, T. T., O'Brien, M. B., . . . Hendrix, R. W. (2010). Comparative genomic analysis of 60 Mycobacteriophage genomes: genome clustering, gene acquisition, and gene size. *J Mol Biol*, *397*(1), 119-143. https://doi.org/10.1016/j.jmb.2010.01.011

Hatfull, G. F., Pedulla, M. L., Jacobs-Sera, D., Cichon, P. M., Foley, A., Ford, M. E., Gonda, R. M., Houtz, J. M., Hryckowian, A. J., Kelchner, V. A., Namburi, S., Pajcini, K. V., Popovich, M. G., Schleicher, D. T., Simanek, B. Z., Smith, A. L., Zdanowicz, G. M., Kumar, V., Peebles, C. L., . . . Hendrix, R. W. (2006). Exploring the Mycobacteriophage Metaproteome: Phage Genomics as an Educational Platform. *PLOS Genetics*, *2*(6), e92. https://doi.org/10.1371/journal.pgen.0020092

Hatfull, G. F., & Sarkis, G. J. (1993). DNA sequence, structure and gene expression of mycobacteriophage L5: a phage system for mycobacterial genetics. *Mol Microbiol*, *7*(3), 395-405. https://doi.org/10.1111/j.1365-2958.1993.tb01131.x

Healy, C., Gouzy, A., & Ehrt, S. (2020). Peptidoglycan Hydrolases RipA and Ami1 Are Critical for Replication and Persistence of Mycobacterium tuberculosis in the Host. *mBio*, *11*(2). https://doi.org/10.1128/mBio.03315-19

Heller, D., Amaya, I., Mohamed, A., Ali, I., Mavrodi, D., Deighan, P., & Sivanathan, V. (2022). Systematic overexpression of genes encoded by mycobacteriophage Waterfoul reveals novel inhibitors of mycobacterial growth. *G3 (Bethesda)*, *12*(8). https://doi.org/10.1093/g3journal/jkac140

Hendrix, R. W. (2002). Bacteriophages: evolution of the majority. *Theor Popul Biol*, *61*(4), 471-480. https://doi.org/10.1006/tpbi.2002.1590

Hendrix, R. W., Smith, M. C. M., Burns, R. N., Ford, M. E., & Hatfull, G. F. (1999). Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. *Proceedings of the National Academy of Sciences*, *96*(5), 2192-2197. https://doi.org/doi:10.1073/pnas.96.5.2192

Hermoso, J. A., García, J. L., & García, P. (2007). Taking aim on bacterial pathogens: from phage therapy to enzybiotics. *Curr Opin Microbiol*, *10*(5), 461-472. https://doi.org/10.1016/j.mib.2007.08.002

Hershey, A. D., & Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J Gen Physiol*, *36*(1), 39-56. https://doi.org/10.1085/jgp.36.1.39

Hershkovitz, I., Donoghue, H. D., Minnikin, D. E., Besra, G. S., Lee, O. Y., Gernaey, A. M., Galili, E., Eshed, V., Greenblatt, C. L., Lemma, E., Bar-Gal, G. K., & Spigelman, M. (2008). Detection and molecular characterization of 9,000-year-old Mycobacterium tuberculosis

from a Neolithic settlement in the Eastern Mediterranean. *PLoS One*, *3*(10), e3426. https://doi.org/10.1371/journal.pone.0003426

Hesselbach, B. A., & Nakada, D. (1977). "Host Shutoff" Function of Bacteriophage T7: Involvement of T7 Gene 2 and Gene 0.7 in the Inactivation of Escherichia coli RNA Polymerase. *Journal of Virology*, *24*(3), 736-745. https://doi.org/doi:10.1128/jvi.24.3.736-745.1977

Hett, E. C., Chao, M. C., Deng, L. L., & Rubin, E. J. (2008). A mycobacterial enzyme essential for cell division synergizes with resuscitation-promoting factor. *PLoS Pathog*, *4*(2), e1000001. https://doi.org/10.1371/journal.ppat.1000001

Hett, E. C., Chao, M. C., Steyn, A. J., Fortune, S. M., Deng, L. L., & Rubin, E. J. (2007). A partner for the resuscitation-promoting factors of Mycobacterium tuberculosis. *Mol Microbiol*, *66*(3), 658-668. https://doi.org/10.1111/j.1365-2958.2007.05945.x

Hildebrand, A., Remmert, M., Biegert, A., & Söding, J. (2009). Fast and accurate automatic structure prediction with HHpred. *Proteins*, *77 Suppl 9*, 128-132. https://doi.org/10.1002/prot.22499

Holm, L. (2022). Dali server: structural unification of protein families. *Nucleic Acids Res*, *50*(W1), W210-w215. https://doi.org/10.1093/nar/gkac387

Holm, L. (2022). Dali server: structural unification of protein families. *Nucleic Acids Research*, *50*(W1), W210-W215. https://doi.org/10.1093/nar/gkac387

Holzheimer, M., Buter, J., & Minnaard, A. J. (2021). Chemical Synthesis of Cell Wall Constituents of Mycobacterium tuberculosis. *Chem Rev*, *121*(15), 9554-9643. https://doi.org/10.1021/acs.chemrev.1c00043

Houben, E. N., Nguyen, L., & Pieters, J. (2006). Interaction of pathogenic mycobacteria with the host immune system. *Curr Opin Microbiol*, *9*(1), 76-85. https://doi.org/10.1016/j.mib.2005.12.014

Howard-Varona, C., Hargreaves, K. R., Abedon, S. T., & Sullivan, M. B. (2017). Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *ISME J*, *11*(7), 1511-1520. https://doi.org/10.1038/ismej.2017.16

Hu, B., Margolin, W., Molineux, I. J., & Liu, J. (2015). Structural remodeling of bacteriophage T4 and host membranes during infection initiation. *Proc Natl Acad Sci U S A*, *112*(35), E4919-4928. https://doi.org/10.1073/pnas.1501064112

Hubbard, A. T. M., Bulgasim, I., & Roberts, A. P. (2021). A novel hemA mutation is responsible for a small-colony-variant phenotype in Escherichia coli. *Microbiology (Reading)*, *167*(3). https://doi.org/10.1099/mic.0.000962

Ilina, Y., Lorent, C., Katz, S., Jeoung, J. H., Shima, S., Horch, M., Zebger, I., & Dobbek, H. (2019). X-ray Crystallography and Vibrational Spectroscopy Reveal the Key Determinants of

Biocatalytic Dihydrogen Cycling by [NiFe] Hydrogenases. *Angew Chem Int Ed Engl*, *58*(51), 18710-18714. https://doi.org/10.1002/anie.201908258

Irving, M. B., Pan, O., & Scott, J. K. (2001). Random-peptide libraries and antigen-fragment libraries for epitope mapping and the development of vaccines and diagnostics. *Curr Opin Chem Biol*, *5*(3), 314-324. https://doi.org/10.1016/s1367-5931(00)00208-8

Iyer, L. M., Leipe, D. D., Koonin, E. V., & Aravind, L. (2004). Evolutionary history and higher order classification of AAA+ ATPases. *Journal of Structural Biology*, *146*(1), 11-31. https://doi.org/https://doi.org/10.1016/j.jsb.2003.10.010

Jacob, F., & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol*, *3*, 318-356. https://doi.org/10.1016/s0022-2836(61)80072-7

Jacobs-Sera, D., Marinelli, L. J., Bowman, C., Broussard, G. W., Guerrero Bustamante, C., Boyle, M. M., Petrova, Z. O., Dedrick, R. M., Pope, W. H., Modlin, R. L., Hendrix, R. W., & Hatfull, G. F. (2012). On the nature of mycobacteriophage diversity and host preference. *Virology*, *434*(2), 187-201. https://doi.org/10.1016/j.virol.2012.09.026

Jankute, M., Cox, J. A., Harrison, J., & Besra, G. S. (2015). Assembly of the Mycobacterial Cell Wall. *Annu Rev Microbiol*, *69*, 405-423. https://doi.org/10.1146/annurev-micro-091014-104121

Jennes, S., Merabishvili, M., Soentjens, P., Pang, K. W., Rose, T., Keersebilck, E., Soete, O., François, P. M., Teodorescu, S., Verween, G., Verbeken, G., De Vos, D., & Pirnay, J. P. (2017). Use of bacteriophages in the treatment of colistin-only-sensitive Pseudomonas aeruginosa septicaemia in a patient with acute kidney injury-a case report. *Crit Care*, *21*(1), 129. https://doi.org/10.1186/s13054-017-1709-y

Kahle, M., Ter Beek, J., Hosler, J. P., & Ädelroth, P. (2018). The insertion of the non-heme Fe(B) cofactor into nitric oxide reductase from P. denitrificans depends on NorQ and NorD accessory proteins. *Biochim Biophys Acta Bioenerg*, *1859*(10), 1051-1058. https://doi.org/10.1016/j.bbabio.2018.05.020

Kang, Y., Xu, Y., Wang, X., Pu, B., Yang, X., Rao, Y., & Chen, J. (2022). HN-PPISP: a hybrid network based on MLP-Mixer for protein–protein interaction site prediction. *Briefings in Bioinformatics*, *24*(1). https://doi.org/10.1093/bib/bbac480

Kaur, D., Guerin, M. E., Skovierová, H., Brennan, P. J., & Jackson, M. (2009). Biogenesis of the cell wall and other glycoconjugates of Mycobacterium tuberculosis. *Adv Appl Microbiol*, *69*, 23-78. https://doi.org/10.1016/s0065-2164(09)69002-x

Keen, E. C. (2015). A century of phage research: bacteriophages and the shaping of modern biology. *Bioessays*, *37*(1), 6-9. https://doi.org/10.1002/bies.201400152

Kehoe, J. W., & Kay, B. K. (2005). Filamentous phage display in the new millennium. *Chem Rev*, *105*(11), 4056-4072. https://doi.org/10.1021/cr000261r

Kidambi, S. P., Ripp, S., & Miller, R. V. (1994). Evidence for phage-mediated gene transfer among Pseudomonas aeruginosa strains on the phylloplane. *Appl Environ Microbiol*, *60*(2), 496-500. https://doi.org/10.1128/aem.60.2.496-500.1994

Kingwell, K. (2015). Bacteriophage therapies re-enter clinical trials. *Nat Rev Drug Discov*, *14*(8), 515-516. https://doi.org/10.1038/nrd4695

Ko, C. C., & Hatfull, G. F. (2018). Mycobacteriophage Fruitloop gp52 inactivates Wag31 (DivIVA) to prevent heterotypic superinfection. *Mol Microbiol*, *108*(4), 443-460. https://doi.org/10.1111/mmi.13946

Ko, C. C., & Hatfull, G. F. (2020). Identification of mycobacteriophage toxic genes reveals new features of mycobacterial physiology and morphology. *Sci Rep*, *10*(1), 14670. https://doi.org/10.1038/s41598-020-71588-5

Koch, R. (1982). Classics in infectious diseases. The etiology of tuberculosis: Robert Koch. Berlin, Germany 1882. *Rev Infect Dis*, *4*(6), 1270-1274.

Koonin, E. V., Makarova, K. S., & Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol*, *37*, 67-78. https://doi.org/10.1016/j.mib.2017.05.008

Krissinel, E., & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica Section D*, *60*(12 Part 1), 2256-2268. https://doi.org/doi:10.1107/S0907444904026460

Krissinel, E., & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J Mol Biol*, *372*(3), 774-797. https://doi.org/10.1016/j.jmb.2007.05.022

Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*, *305*(3), 567-580. https://doi.org/10.1006/jmbi.2000.4315

Kropinski, A. M. (2018). Bacteriophage research - What we have learnt and what still needs to be addressed. *Res Microbiol*, *169*(9), 481-487. https://doi.org/10.1016/j.resmic.2018.05.002

Labrie, S. J., Samson, J. E., & Moineau, S. (2010). Bacteriophage resistance mechanisms. *Nature Reviews Microbiology*, *8*(5), 317-327. https://doi.org/10.1038/nrmicro2315

Lanzetta, P. A., Alvarez, L. J., Reinach, P. S., & Candia, O. A. (1979). An improved assay for nanomole amounts of inorganic phosphate. *Anal Biochem*, *100*(1), 95-97. https://doi.org/10.1016/0003-2697(79)90115-5

Lee, W. L., Gold, B., Darby, C., Brot, N., Jiang, X., de Carvalho, L. P., Wellner, D., St John, G., Jacobs, W. R., Jr., & Nathan, C. (2009). Mycobacterium tuberculosis expresses methionine sulphoxide reductases A and B that protect from killing by nitrite and hypochlorite. *Mol Microbiol*, *71*(3), 583-593. https://doi.org/10.1111/j.1365-2958.2008.06548.x

Lefebvre, C., Boulon, R., Ducoux, M., Gavalda, S., Laval, F., Jamet, S., Eynard, N., Lemassu, A., Cam, K., Bousquet, M.-P., Bardou, F., Burlet-Schiltz, O., Daffé, M., & Quémard, A. (2018). HadD, a novel fatty acid synthase type II protein, is essential for alpha- and epoxy-mycolic acid biosynthesis and mycobacterial fitness. *Scientific Reports*, *8*(1), 6034. https://doi.org/10.1038/s41598-018-24380-5

Leipe, D. D., Koonin, E. V., & Aravind, L. (2003). Evolution and classification of P-loop kinases and related proteins. *J Mol Biol*, *333*(4), 781-815. https://doi.org/10.1016/j.jmb.2003.08.040

Letarov, A. V., & Kulikov, E. E. (2017). Adsorption of Bacteriophages on Bacterial Cells. *Biochemistry (Mosc)*, *82*(13), 1632-1658. https://doi.org/10.1134/S0006297917130053

Leverentz, B., Conway, W. S., Camp, M. J., Janisiewicz, W. J., Abuladze, T., Yang, M., Saftner, R., & Sulakvelidze, A. (2003). Biocontrol of Listeria monocytogenes on fresh-cut produce by treatment with lytic bacteriophages and a bacteriocin. *Appl Environ Microbiol*, *69*(8), 4519-4526. https://doi.org/10.1128/aem.69.8.4519-4526.2003

Lin, J., Du, F., Long, M., & Li, P. (2022). Limitations of Phage Therapy and Corresponding Optimization Strategies: A Review. *Molecules*, *27*(6). https://doi.org/10.3390/molecules27061857

Ling, H., Lou, X., Luo, Q., He, Z., Sun, M., & Sun, J. (2022). Recent advances in bacteriophage-based therapeutics: Insight into the post-antibiotic era. *Acta Pharmaceutica Sinica B*, *12*(12), 4348-4364. https://doi.org/https://doi.org/10.1016/j.apsb.2022.05.007

Liu, J., Dehbi, M., Moeck, G., Arhin, F., Bauda, P., Bergeron, D., Callejo, M., Ferretti, V., Ha, N., Kwan, T., McCarty, J., Srikumar, R., Williams, D., Wu, J. J., Gros, P., Pelletier, J., & DuBow, M. (2004). Antimicrobial drug discovery through bacteriophage genomics. *Nature Biotechnology*, *22*(2), 185-191. https://doi.org/10.1038/nbt932

Loc-Carrillo, C., & Abedon, S. T. (2011). Pros and cons of phage therapy. *Bacteriophage*, *1*(2), 111-114. https://doi.org/10.4161/bact.1.2.14590

Lu, M. J., & Henning, U. (1994). Superinfection exclusion by T-even-type coliphages. *Trends Microbiol*, *2*(4), 137-139. https://doi.org/10.1016/0966-842x(94)90601-7

Luria, S. E., & Delbrück, M. (1943). Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics*, *28*(6), 491-511. https://doi.org/10.1093/genetics/28.6.491

Mallory, J. B., Alfano, C., & McMacken, R. (1990). Host virus interactions in the initiation of bacteriophage lambda DNA replication. Recruitment of Escherichia coli DnaB helicase by lambda P replication protein. *J Biol Chem*, *265*(22), 13297-13307.

McGuffin, L. J., Bryson, K., & Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics*, *16*(4), 404-405. https://doi.org/10.1093/bioinformatics/16.4.404

Mediavilla, J., Jain, S., Kriakov, J., Ford, M. E., Duda, R. L., Jacobs, W. R., Jr., Hendrix, R. W., & Hatfull, G. F. (2000). Genome organization and characterization of mycobacteriophage Bxb1. *Mol Microbiol*, *38*(5), 955-970. https://doi.org/10.1046/j.1365-2958.2000.02183.x

Meena, L. S., & Rajni. (2010). Survival mechanisms of pathogenic Mycobacterium tuberculosis H37Rv. *Febs j*, *277*(11), 2416-2427. https://doi.org/10.1111/j.1742-4658.2010.07666.x

Mehio, W., Kemp, G. J., Taylor, P., & Walkinshaw, M. D. (2010). Identification of protein binding surfaces using surface triplet propensities. *Bioinformatics*, *26*(20), 2549-2555. https://doi.org/10.1093/bioinformatics/btq490

Meniche, X., Otten, R., Siegrist, M. S., Baer, C. E., Murphy, K. C., Bertozzi, C. R., & Sassetti, C. M. (2014). Subpolar addition of new cell wall is directed by DivIVA in mycobacteria. *Proc Natl Acad Sci U S A*, *111*(31), E3243-3251. https://doi.org/10.1073/pnas.1402158111

Merril, C. R., Scholl, D., & Adhya, S. L. (2003). The prospect for bacteriophage therapy in Western medicine. *Nature Reviews Drug Discovery*, *2*(6), 489-497. https://doi.org/10.1038/nrd1111

Meyer, J. R., Dobias, D. T., Weitz, J. S., Barrick, J. E., Quick, R. T., & Lenski, R. E. (2012). Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science*, *335*(6067), 428-432. https://doi.org/10.1126/science.1214449

Miller, E. S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T., & Ruger, W. (2003). Bacteriophage T4 genome. *Microbiol Mol Biol Rev*, *67*(1), 86-156, table of contents. https://doi.org/10.1128/MMBR.67.1.86-156.2003

Mizianty, M. J., & Kurgan, L. (2011). Sequence-based prediction of protein crystallization, purification and production propensity. *Bioinformatics*, *27*(13), i24-33. https://doi.org/10.1093/bioinformatics/btr229

Mizianty, M. J., & Kurgan, L. A. (2012). CRYSpred: accurate sequence-based protein crystallization propensity prediction using sequence-derived structural characteristics. *Protein Pept Lett*, *19*(1), 40-49. https://doi.org/10.2174/092986612798472910

Modi, R., Hirvi, Y., Hill, A., & Griffiths, M. W. (2001). Effect of phage on survival of Salmonella enteritidis during manufacture and storage of cheddar cheese made from raw and pasteurized milk. *J Food Prot*, *64*(7), 927-933. https://doi.org/10.4315/0362-028x-64.7.927

Mohammed, M. (2017). Phage typing or CRISPR typing for epidemiological surveillance of Salmonella Typhimurium? *BMC Res Notes*, *10*(1), 578. https://doi.org/10.1186/s13104-017-2878-0

Moldovan, R., Chapman-McQuiston, E., & Wu, X. L. (2007). On kinetics of phage adsorption. *Biophys J*, *93*(1), 303-315. https://doi.org/10.1529/biophysj.106.102962

Montales, M. T., Chaudhury, A., Beebe, A., Patil, S., & Patil, N. (2015). HIV-Associated TB Syndemic: A Growing Clinical Challenge Worldwide. *Front Public Health*, *3*, 281. https://doi.org/10.3389/fpubh.2015.00281

Mueser, T. C., Hinerman, J. M., Devos, J. M., Boyer, R. A., & Williams, K. J. (2010). Structural analysis of bacteriophage T4 DNA replication: a review in the Virology Journal series on bacteriophage T4 and its relatives. *Virology Journal*, *7*(1), 359. https://doi.org/10.1186/1743-422X-7-359

Nechaev, S., & Severinov, K. (1999). Inhibition of Escherichia coli RNA polymerase by bacteriophage T7 gene 2 protein. *J Mol Biol*, *289*(4), 815-826. https://doi.org/10.1006/jmbi.1999.2782

Nechaev, S., & Severinov, K. (2008). The elusive object of desire--interactions of bacteriophages and their hosts. *Curr Opin Microbiol*, *11*(2), 186-193. https://doi.org/10.1016/j.mib.2008.02.009

Neuwald, A. F., Aravind, L., Spouge, J. L., & Koonin, E. V. (1999). AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res*, *9*(1), 27-43.

Newton, G. L., Buchmeier, N., & Fahey, R. C. (2008). Biosynthesis and functions of mycothiol, the unique protective thiol of Actinobacteria. *Microbiol Mol Biol Rev*, *72*(3), 471-494. https://doi.org/10.1128/mmbr.00008-08

Ng, V. H., Cox, J. S., Sousa, A. O., MacMicking, J. D., & McKinney, J. D. (2004). Role of KatG catalase-peroxidase in mycobacterial pathogenesis: countering the phagocyte oxidative burst. *Mol Microbiol*, *52*(5), 1291-1302. https://doi.org/10.1111/j.1365-2958.2004.04078.x

Nilsson, A. S. (2019). Pharmacological limitations of phage therapy. *Ups J Med Sci*, *124*(4), 218-227. https://doi.org/10.1080/03009734.2019.1688433

Nobrega, F. L., Vlot, M., de Jonge, P. A., Dreesens, L. L., Beaumont, H. J. E., Lavigne, R., Dutilh, B. E., & Brouns, S. J. J. (2018). Targeting mechanisms of tailed bacteriophages. *Nat Rev Microbiol*, *16*(12), 760-773. https://doi.org/10.1038/s41579-018-0070-8

Nuttall, S. D., & Dyall-Smith, M. L. (1993). HF1 and HF2: novel bacteriophages of halophilic archaea. *Virology*, *197*(2), 678-684. https://doi.org/10.1006/viro.1993.1643

O'Farrell, P. H., Kutter, E., & Nakanishi, M. (1980). A restriction map of the bacteriophage T4 genome. *Mol Gen Genet*, *179*(2), 421-435. https://doi.org/10.1007/bf00425473

O'Toole, M., Lau, K. T., Shepherd, R., Slater, C., & Diamond, D. (2007). Determination of phosphate using a highly sensitive paired emitter-detector diode photometric flow detector. *Anal Chim Acta*, *597*(2), 290-294. https://doi.org/10.1016/j.aca.2007.06.048

Ochman, H., Lawrence, J. G., & Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature*, *405*(6784), 299-304. https://doi.org/10.1038/35012500

Ogura, T., & Wilkinson, A. J. (2001). AAA+ superfamily ATPases: common structure--diverse function. *Genes Cells*, *6*(7), 575-597. https://doi.org/10.1046/j.1365-2443.2001.00447.x

Ohnishi, M., Kurokawa, K., & Hayashi, T. (2001). Diversification of Escherichia coli genomes: are bacteriophages the major contributors? *Trends Microbiol*, *9*(10), 481-485. https://doi.org/10.1016/s0966-842x(01)02173-4

Otsuka, Y., & Yonesaki, T. (2012). Dmd of bacteriophage T4 functions as an antitoxin against Escherichia coli LsoA and RnlA toxins. *Mol Microbiol*, *83*(4), 669-681. https://doi.org/10.1111/j.1365-2958.2012.07975.x

Otwinowski, Z., & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*, *276*, 307-326. https://doi.org/10.1016/s0076-6879(97)76066-x

Otwinowski, Z., & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology* (Vol. 276, pp. 307-326). Academic Press. https://doi.org/https://doi.org/10.1016/S0076-6879(97)76066-X

Payne, K., Sun, Q., Sacchettini, J., & Hatfull, G. F. (2009). Mycobacteriophage Lysin B is a novel mycolylarabinogalactan esterase. *Mol Microbiol*, *73*(3), 367-381. https://doi.org/10.1111/j.1365-2958.2009.06775.x

Pedulla, M. L., Ford, M. E., Houtz, J. M., Karthikeyan, T., Wadsworth, C., Lewis, J. A., Jacobs-Sera, D., Falbo, J., Gross, J., Pannunzio, N. R., Brucker, W., Kumar, V., Kandasamy, J., Keenan, L., Bardarov, S., Kriakov, J., Lawrence, J. G., Jacobs, W. R., Jr., Hendrix, R. W., & Hatfull, G. F. (2003). Origins of highly mosaic mycobacteriophage genomes. *Cell*, *113*(2), 171-182. https://doi.org/10.1016/s0092-8674(03)00233-2

Pelzmann, A., Ferner, M., Gnida, M., Meyer-Klaucke, W., Maisel, T., & Meyer, O. (2009). The CoxD protein of Oligotropha carboxidovorans is a predicted AAA+ ATPase chaperone involved in the biogenesis of the CO dehydrogenase [CuSMoO2] cluster. *J Biol Chem*, *284*(14), 9578-9586. https://doi.org/10.1074/jbc.M805354200

Perry, E. B., Barrick, J. E., & Bohannan, B. J. (2015). The Molecular and Genetic Basis of Repeatable Coevolution between Escherichia coli and Bacteriophage T3 in a Laboratory Microcosm. *PLoS One*, *10*(6), e0130639. https://doi.org/10.1371/journal.pone.0130639

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*, *8*(10), 785-786. https://doi.org/10.1038/nmeth.1701

Piddington, D. L., Fang, F. C., Laessig, T., Cooper, A. M., Orme, I. M., & Buchmeier, N. A. (2001). Cu,Zn superoxide dismutase of Mycobacterium tuberculosis contributes to survival in activated macrophages that are generating an oxidative burst. *Infect Immun*, *69*(8), 4980-4987. https://doi.org/10.1128/iai.69.8.4980-4987.2001

Piekarowicz, A., Kłyż, A., & Stein, D. C. (2022). A New Vaccination Method Based on Phage NgoΦ6 and Its Phagemid Derivatives. *Front Microbiol*, *13*, 793205. https://doi.org/10.3389/fmicb.2022.793205

Pires, D. P., Oliveira, H., Melo, L. D., Sillankorva, S., & Azeredo, J. (2016). Bacteriophage-encoded depolymerases: their diversity and biotechnological applications. *Appl Microbiol Biotechnol*, *100*(5), 2141-2151. https://doi.org/10.1007/s00253-015-7247-0

Pontali, E., Raviglione, M. C., & Migliori, G. B. (2019). Regimens to treat multidrug-resistant tuberculosis: past, present and future perspectives. *Eur Respir Rev*, *28*(152). https://doi.org/10.1183/16000617.0035-2019

Pope, W. H., Bowman, C. A., Russell, D. A., Jacobs-Sera, D., Asai, D. J., Cresawn, S. G., Jacobs, W. R., Hendrix, R. W., Lawrence, J. G., & Hatfull, G. F. (2015). Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. *Elife*, *4*, e06416. https://doi.org/10.7554/eLife.06416

Pope, W. H., Bowman, C. A., Russell, D. A., Jacobs-Sera, D., Asai, D. J., Cresawn, S. G., Jacobs, W. R., Jr., Hendrix, R. W., Lawrence, J. G., Hatfull, G. F., Science Education Alliance Phage Hunters Advancing, G., Evolutionary, S., Phage Hunters Integrating, R., Education, & Mycobacterial Genetics, C. (2015a). Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. *Elife*, *4*, e06416. https://doi.org/10.7554/eLife.06416

Pope, W. H., Carbonara, M. E., Cioffi, H. M., Cruz, T., Dang, B. Q., Doyle, A. N., Fan, O. H., Gallagher, M., Gentile, G. M., German, B. A., Farrell, M. E., Gerwig, M., Hunter, K. L., Lefever, V. E., Marfisi, N. A., McDonnell, J. E., Monga, J. K., Quiroz, K. G., Pong, A. C., . . . Hatfull, G. F. (2015b). Genome Sequences of Mycobacteriophages AlanGrant, Baee, Corofin, OrangeOswald, and Vincenzo, New Members of Cluster B. *Genome Announc*, *3*(3). https://doi.org/10.1128/genomeA.00586-15

Pope, W. H., Ferreira, C. M., Jacobs-Sera, D., Benjamin, R. C., Davis, A. J., DeJong, R. J., Elgin, S. C., Guilfoile, F. R., Forsyth, M. H., Harris, A. D., Harvey, S. E., Hughes, L. E., Hynes, P. M., Jackson, A. S., Jalal, M. D., MacMurray, E. A., Manley, C. M., McDonough, M. J., Mosier, J. L., . . . Hatfull, G. F. (2011). Cluster K mycobacteriophages: insights into the evolutionary origins of mycobacteriophage TM4. *PLoS One*, *6*(10), e26750. https://doi.org/10.1371/journal.pone.0026750

Predich, M., Doukhan, L., Nair, G., & Smith, I. (1995). Characterization of RNA polymerase and two sigma-factor genes from Mycobacterium smegmatis. *Mol Microbiol*, *15*(2), 355-366. https://doi.org/10.1111/j.1365-2958.1995.tb02249.x

Prigent, M., Leroy, M., Confalonieri, F., Dutertre, M., & DuBow, M. S. (2005). A diversity of bacteriophage forms and genomes can be isolated from the surface sands of the Sahara Desert. *Extremophiles*, *9*(4), 289-296. https://doi.org/10.1007/s00792-005-0444-5

Programme, G. T. (2022). *Global tuberculosis report*. https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022

Qin, S., & Zhou, H. X. (2007). meta-PPISP: a meta web server for protein-protein interaction site prediction. *Bioinformatics*, *23*(24), 3386-3387. https://doi.org/10.1093/bioinformatics/btm434

Rabsch, W. (2007). Salmonella Typhimurium Phage Typing for Pathogens. In H. Schatten & A. Eisenstark (Eds.), *Salmonella: Methods and Protocols* (pp. 177-211). Humana Press. https://doi.org/10.1007/978-1-59745-512-1_10

Rajaram, M. V., Brooks, M. N., Morris, J. D., Torrelles, J. B., Azad, A. K., & Schlesinger, L. S. (2010). Mycobacterium tuberculosis activates human macrophage peroxisome proliferator-activated receptor gamma linking mannose receptor recognition to regulation of immune responses. *J Immunol*, *185*(2), 929-942. https://doi.org/10.4049/jimmunol.1000866

Rakhuba, D. V., Kolomiets, E. I., Dey, E. S., & Novik, G. I. (2010). Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell. *Pol J Microbiol*, *59*(3), 145-155.

Ramesh, M., Nitharwal, R. G., Behra, P. R. K., Fredrik Pettersson, B. M., Dasgupta, S., & Kirsebom, L. A. (2021). Intracellular localization of the mycobacterial stressosome complex. *Scientific Reports*, *11*(1), 10060. https://doi.org/10.1038/s41598-021-89069-8

Raynaud, C., Papavinasasundaram, K. G., Speight, R. A., Springer, B., Sander, P., Böttger, E. C., Colston, M. J., & Draper, P. (2002). The functions of OmpATb, a pore-forming protein of Mycobacterium tuberculosis. *Mol Microbiol*, *46*(1), 191-201. https://doi.org/10.1046/j.1365-2958.2002.03152.x

Richardson, I. W., & Anthony, C. (1992). Characterization of mutant forms of the quinoprotein methanol dehydrogenase lacking an essential calcium ion. *Biochem J*, *287 (Pt 3)*(Pt 3), 709-715. https://doi.org/10.1042/bj2870709

Robertson, E. S., Aggison, L. A., & Nicholson, A. W. (1994). Phosphorylation of elongation factor G and ribosomal protein S6 in bacteriophage T7-infected Escherichia coli. *Mol Microbiol*, *11*(6), 1045-1057. https://doi.org/10.1111/j.1365-2958.1994.tb00382.x

Robertson, E. S., & Nicholson, A. W. (1990). Protein kinase of bacteriophage T7 induces the phosphorylation of only a small number of proteins in the infected cell. *Virology*, *175*(2), 525-534. https://doi.org/10.1016/0042-6822(90)90437-v

Robertson, E. S., & Nicholson, A. W. (1992). Phosphorylation of Escherichia coli translation initiation factors by the bacteriophage T7 protein kinase. *Biochemistry*, *31*(20), 4822-4827. https://doi.org/10.1021/bi00135a012

Roggenkamp, A., Sing, A., Hornef, M., Brunner, U., Autenrieth, I. B., & Heesemann, J. (1998). Chronic prosthetic hip infection caused by a small-colony variant of Escherichia coli. *J Clin Microbiol*, *36*(9), 2530-2534. https://doi.org/10.1128/jcm.36.9.2530-2534.1998

Roucourt, B., & Lavigne, R. (2009). The role of interactions between phage and bacterial proteins within the infected cell: a diverse and puzzling interactome. *Environ Microbiol*, *11*(11), 2789-2805. https://doi.org/10.1111/j.1462-2920.2009.02029.x

Russell, D. A., & Hatfull, G. F. (2017). PhagesDB: the actinobacteriophage database. *Bioinformatics*, *33*(5), 784-786. https://doi.org/10.1093/bioinformatics/btw711

Rybniker, J., Krumbach, K., van Gumpel, E., Plum, G., Eggeling, L., & Hartmann, P. (2011). The cytotoxic early protein 77 of mycobacteriophage L5 interacts with MSMEG_3532, an L-serine dehydratase of Mycobacterium smegmatis. *J Basic Microbiol*, *51*(5), 515-522. https://doi.org/10.1002/jobm.201000446

Rybniker, J., Nowag, A., van Gumpel, E., Nissen, N., Robinson, N., Plum, G., & Hartmann, P. (2010). Insights into the function of the WhiB-like protein of mycobacteriophage TM4--a transcriptional inhibitor of WhiB2. *Mol Microbiol*, *77*(3), 642-657. https://doi.org/10.1111/j.1365-2958.2010.07235.x

Rybniker, J., Plum, G., Robinson, N., Small, P. L., & Hartmann, P. (2008). Identification of three cytotoxic early proteins of mycobacteriophage L5 leading to growth inhibition in Mycobacterium smegmatis. *Microbiology (Reading)*, *154*(Pt 8), 2304-2314. https://doi.org/10.1099/mic.0.2008/017004-0

Samson, J. E., Magadán, A. H., Sabri, M., & Moineau, S. (2013). Revenge of the phages: defeating bacterial defences. *Nat Rev Microbiol*, *11*(10), 675-687. https://doi.org/10.1038/nrmicro3096

Santos, V., & Hirshfield, I. (2016). The Physiological and Molecular Characterization of a Small Colony Variant of Escherichia coli and Its Phenotypic Rescue. *PLoS One*, *11*(6), e0157578. https://doi.org/10.1371/journal.pone.0157578

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J. Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nat Methods*, *9*(7), 676-682. https://doi.org/10.1038/nmeth.2019

Shabbir, M. A., Hao, H., Shabbir, M. Z., Wu, Q., Sattar, A., & Yuan, Z. (2016). Bacteria vs. Bacteriophages: Parallel Evolution of Immune Arsenals. *Front Microbiol*, *7*, 1292. https://doi.org/10.3389/fmicb.2016.01292

Shi, Y., Fernandez-Martinez, J., Tjioe, E., Pellarin, R., Kim, S. J., Williams, R., Schneidman-Duhovny, D., Sali, A., Rout, M. P., & Chait, B. T. (2014). Structural characterization by cross-linking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex. *Mol Cell Proteomics*, *13*(11), 2927-2943. https://doi.org/10.1074/mcp.M114.041673

Shibayama, Y., & Dabbs, E. R. (2011). Phage as a source of antibacterial genes: Multiple inhibitory products encoded by Rhodococcus phage YF1. *Bacteriophage*, *1*(4), 195-197. https://doi.org/10.4161/bact.1.4.17746

Slabinski, L., Jaroszewski, L., Rodrigues, A. P., Rychlewski, L., Wilson, I. A., Lesley, S. A., & Godzik, A. (2007a). The challenge of protein structure determination--lessons from structural genomics. *Protein Sci*, *16*(11), 2472-2482. https://doi.org/10.1110/ps.073037907

Slabinski, L., Jaroszewski, L., Rychlewski, L., Wilson, I. A., Lesley, S. A., & Godzik, A. (2007). XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics*, *23*(24), 3403-3405. https://doi.org/10.1093/bioinformatics/btm477

Slabinski, L., Jaroszewski, L., Rychlewski, L., Wilson, I. A., Lesley, S. A., & Godzik, A. (2007b). XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics*, *23*(24), 3403-3405. https://doi.org/10.1093/bioinformatics/btm477

Slavin, M., Tayri-Wilk, T., Milhem, H., & Kalisman, N. (2020). Open Search Strategy for Inferring the Masses of Cross-Link Adducts on Proteins. *Analytical Chemistry*, *92*(24), 15899-15907. https://doi.org/10.1021/acs.analchem.0c03292

Smith, G. P., & Petrenko, V. A. (1997). Phage Display. *Chem Rev*, *97*(2), 391-410. https://doi.org/10.1021/cr960065d

Smith, N. H., Hewinson, R. G., Kremer, K., Brosch, R., & Gordon, S. V. (2009). Myths and misconceptions: the origin and evolution of Mycobacterium tuberculosis. *Nat Rev Microbiol*, *7*(7), 537-544. https://doi.org/10.1038/nrmicro2165

Snapper, S. B., Melton, R. E., Mustafa, S., Kieser, T., & Jacobs, W. R., Jr. (1990). Isolation and characterization of efficient plasmid transformation mutants of Mycobacterium smegmatis. *Mol Microbiol*, *4*(11), 1911-1919. https://doi.org/10.1111/j.1365-2958.1990.tb02040.x

Snider, J., & Houry, W. A. (2006). MoxR AAA+ ATPases: a novel family of molecular chaperones? *J Struct Biol*, *156*(1), 200-209. https://doi.org/10.1016/j.jsb.2006.02.009

Sonnhammer, E. L., von Heijne, G., & Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol*, *6*, 175-182.

Springer, T. A. (2006). Complement and the multifaceted functions of VWA and integrin I domains. *Structure*, *14*(11), 1611-1616. https://doi.org/10.1016/j.str.2006.10.001

Strathdee, S., Patterson, T., & Barker, T. (2019). *The Perfect Predator: A Scientist's Race to Save Her Husband from a Deadly Superbug: A Memoir*. Hachette Books. https://books.google.com/books?id=C9FsDwAAQBAJ

Sulakvelidze, A., Alavidze, Z., & Morris, J. G. (2001). Bacteriophage Therapy. *Antimicrobial Agents and Chemotherapy*, *45*(3), 649-659. https://doi.org/doi:10.1128/AAC.45.3.649-659.2001

Suttle, C. A. (2007). Marine viruses--major players in the global ecosystem. *Nat Rev Microbiol*, *5*(10), 801-812. https://doi.org/10.1038/nrmicro1750

Teklemariam, A. D., Al-Hindi, R. R., Qadri, I., Alharbi, M. G., Ramadan, W. S., Ayubu, J., Al-Hejin, A. M., Hakim, R. F., Hakim, F. F., Hakim, R. F., Alseraihi, L. I., Alamri, T., & Harakeh, S. (2023). The Battle between Bacteria and Bacteriophages: A Conundrum to Their Immune System. *Antibiotics (Basel)*, *12*(2). https://doi.org/10.3390/antibiotics12020381

Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Zwart, P. H., & Hung, L. W. (2009). Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr D Biol Crystallogr*, *65*(Pt 6), 582-601. https://doi.org/10.1107/s0907444909012098

Thiel, K. (2004). Old dogma, new tricks--21st Century phage therapy. *Nat Biotechnol*, *22*(1), 31-36. https://doi.org/10.1038/nbt0104-31

Tientcheu, L. D., Bell, A., Secka, O., Ayorinde, A., Otu, J., Garton, N. J., Sutherland, J. S., Ota, M. O., Antonio, M., Dockrell, H. M., Kampmann, B., & Barer, M. R. (2016). Association of slow recovery of Mycobacterium africanum-infected patients posttreatment with high content of Persister-Like bacilli in pretreatment sputum. *Int J Mycobacteriol*, *5 Suppl 1*, S99-s100. https://doi.org/10.1016/j.ijmyco.2016.09.033

Tock, M. R., & Dryden, D. T. (2005). The biology of restriction and anti-restriction. *Curr Opin Microbiol*, *8*(4), 466-472. https://doi.org/10.1016/j.mib.2005.06.003

Touchon, M., Bernheim, A., & Rocha, E. P. (2016). Genetic and life-history traits associated with the distribution of prophages in bacteria. *ISME J*, *10*(11), 2744-2754. https://doi.org/10.1038/ismej.2016.47

Touchon, M., Moura de Sousa, J., A, & Rocha, E., P. C. (2017). Embracing the enemy: the diversification of microbial gene repertoires by phage-mediated horizontal gene transfer. *Current Opinion in Microbiology*, *38*, 66-73. https://doi.org/10.1016/j.mib.2017.04.010

Tsai, Y.-C. C., Ye, F., Liew, L., Liu, D., Bhushan, S., Gao, Y.-G., & Mueller-Cajar, O. (2020). Insights into the mechanism and regulation of the CbbQO-type Rubisco activase, a MoxR AAA+ ATPase. *Proceedings of the National Academy of Sciences*, *117*(1), 381-387. https://doi.org/doi:10.1073/pnas.1911123117

Valen, L. V. (2014). 19. A New Evolutionary Law (1973). In A. S. Felisa, L. G. John, & H. B. James (Eds.), *Foundations of Macroecology* (pp. 284-314). University of Chicago Press. https://doi.org/doi:10.7208/9780226115504-022

van Houte, S., Buckling, A., & Westra, E. R. (2016). Evolutionary Ecology of Prokaryotic Immune Mechanisms. *Microbiol Mol Biol Rev*, *80*(3), 745-763. https://doi.org/10.1128/mmbr.00011-16

Van Spanning, R. J., Wansell, C. W., De Boer, T., Hazelaar, M. J., Anazawa, H., Harms, N., Oltmann, L. F., & Stouthamer, A. H. (1991). Isolation and characterization of the moxJ, moxG, moxI, and moxR genes of Paracoccus denitrificans: inactivation of moxJ, moxG,

and moxR and the resultant effect on methylotrophic growth. *J Bacteriol*, *173*(21), 6948-6961. https://doi.org/10.1128/jb.173.21.6948-6961.1991

Vandal, O. H., Pierini, L. M., Schnappinger, D., Nathan, C. F., & Ehrt, S. (2008). A membrane protein preserves intrabacterial pH in intraphagosomal Mycobacterium tuberculosis. *Nat Med*, *14*(8), 849-854. https://doi.org/10.1038/nm.1795

von Wintersdorff, C. J., Penders, J., van Niekerk, J. M., Mills, N. D., Majumder, S., van Alphen, L. B., Savelkoul, P. H., & Wolffs, P. F. (2016). Dissemination of Antimicrobial Resistance in Microbial Ecosystems through Horizontal Gene Transfer. *Front Microbiol*, *7*, 173. https://doi.org/10.3389/fmicb.2016.00173

Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T., & Bricogne, G. (2011). Data processing and analysis with the autoPROC toolbox. *Acta Crystallographica Section D*, *67*(4), 293-302. https://doi.org/doi:10.1107/S0907444911007773

Waagmeester, A., Thompson, J., & Reyrat, J. M. (2005). Identifying sigma factors in Mycobacterium smegmatis by comparative genomic analysis. *Trends Microbiol*, *13*(11), 505-509. https://doi.org/10.1016/j.tim.2005.08.009

Waldor, M. K., & Mekalanos, J. J. (1996). Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science*, *272*(5270), 1910-1914. https://doi.org/10.1126/science.272.5270.1910

Warner, D. F., & Mizrahi, V. (2007). The survival kit of Mycobacterium tuberculosis. *Nat Med*, *13*(3), 282-284. https://doi.org/10.1038/nm0307-282

Weinbauer, M. G. (2004). Ecology of prokaryotic viruses. *FEMS Microbiol Rev*, *28*(2), 127-181. https://doi.org/10.1016/j.femsre.2003.08.001

Welin, A., & Lerm, M. (2012). Inside or outside the phagosome? The controversy of the intracellular localization of Mycobacterium tuberculosis. *Tuberculosis (Edinb)*, *92*(2), 113-120. https://doi.org/10.1016/j.tube.2011.09.009

Wells, C. D., Cegielski, J. P., Nelson, L. J., Laserson, K. F., Holtz, T. H., Finlay, A., Castro, K. G., & Weyer, K. (2007). HIV infection and multidrug-resistant tuberculosis: the perfect storm. *J Infect Dis*, *196 Suppl 1*, S86-107. https://doi.org/10.1086/518665

Whittaker, C. A., & Hynes, R. O. (2002). Distribution and evolution of von Willebrand/integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere. *Mol Biol Cell*, *13*(10), 3369-3387. https://doi.org/10.1091/mbc.e02-05-0259

Wichels, A., Biel, S. S., Gelderblom, H. R., Brinkhoff, T., Muyzer, G., & Schutt, C. (1998). Bacteriophage diversity in the North Sea. *Appl Environ Microbiol*, *64*(11), 4128-4133. https://doi.org/10.1128/AEM.64.11.4128-4133.1998

Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., & Hochstrasser, D. F. (1999). Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol*, *112*, 531-552. https://doi.org/10.1385/1-59259-584-7:531

Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall, W. B., 3rd, Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S., & Richardson, D. C. (2018). MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci*, *27*(1), 293-315. https://doi.org/10.1002/pro.3330

Wommack, K. E., & Colwell, R. R. (2000). Virioplankton: viruses in aquatic ecosystems. *Microbiol Mol Biol Rev*, *64*(1), 69-114. https://doi.org/10.1128/MMBR.64.1.69-114.2000

Wong, K. S., Bhandari, V., Janga, S. C., & Houry, W. A. (2017). The RavA-ViaA Chaperone-Like System Interacts with and Modulates the Activity of the Fumarate Reductase Respiratory Complex. *J Mol Biol*, *429*(2), 324-344. https://doi.org/10.1016/j.jmb.2016.12.008

Wong, K. S., Snider, J. D., Graham, C., Greenblatt, J. F., Emili, A., Babu, M., & Houry, W. A. (2014). The MoxR ATPase RavA and its cofactor ViaA interact with the NADH:ubiquinone oxidoreductase I in Escherichia coli. *PLoS One*, *9*(1), e85529. https://doi.org/10.1371/journal.pone.0085529

Wu, Q. L., Kong, D., Lam, K., & Husson, R. N. (1997). A mycobacterial extracytoplasmic function sigma factor involved in survival following stress. *J Bacteriol*, *179*(9), 2922-2929. https://doi.org/10.1128/jb.179.9.2922-2929.1997

Xia, H., Tang, Q., Song, J., Ye, J., Wu, H., & Zhang, H. (2017). A yigP mutant strain is a small colony variant of E. coli and shows pleiotropic antibiotic resistance. *Can J Microbiol*, *63*(12), 961-969. https://doi.org/10.1139/cjm-2017-0347

Yan, J., Mao, J., & Xie, J. (2014). Bacteriophage polysaccharide depolymerases and biomedical applications. *BioDrugs*, *28*(3), 265-274. https://doi.org/10.1007/s40259-013-0081-y

Yang, J., & Zhang, Y. (2015). I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res*, *43*(W1), W174-181. https://doi.org/10.1093/nar/gkv342

Ye, Y., & Godzik, A. (2004). FATCAT: a web server for flexible structure comparison and structure similarity searching. *Nucleic Acids Res*, *32*(Web Server issue), W582-585. https://doi.org/10.1093/nar/gkh430

Zhang, K., Li, S., Wang, Y., Wang, Z., Mulvenna, N., Yang, H., Zhang, P., Chen, H., Li, Y., Wang, H., Gao, Y., Wigneshweraraj, S., Matthews, S., Zhang, K., & Liu, B. (2022). Bacteriophage protein PEIP is a potent Bacillus subtilis enolase inhibitor. *Cell Rep*, *40*(1), 111026. https://doi.org/10.1016/j.celrep.2022.111026

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, *9*, 40. https://doi.org/10.1186/1471-2105-9-40

Zimmermann, L., Stephens, A., Nam, S. Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A. N., & Alva, V. (2018). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J Mol Biol*, *430*(15), 2237-2243. https://doi.org/10.1016/j.jmb.2017.12.007

Zinder, N. D., & Lederberg, J. (1952). Genetic exchange in Salmonella. *J Bacteriol*, *64*(5), 679-699. https://doi.org/10.1128/jb.64.5.679-699.1952