

**INVESTIGATION OF PROTEOMIC AND LIPIDOMIC
MYCOBACTERIOPHAGE MASS SPECTROMETRY DATA**

by

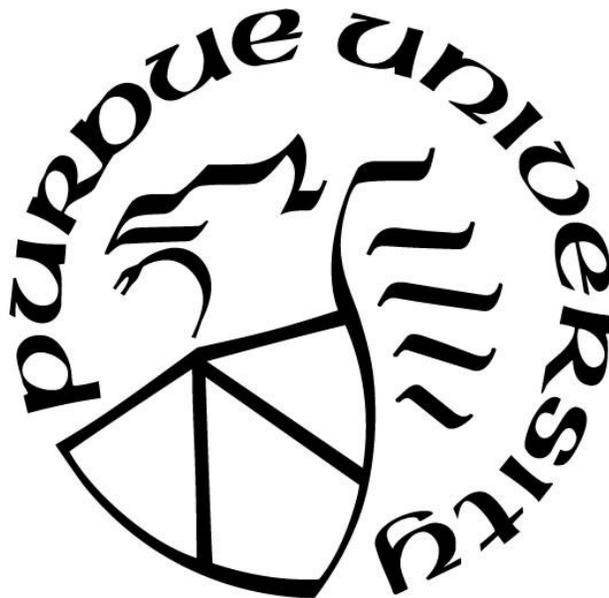
Gillian Smith

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science



Department of Agricultural and Biological Engineering

West Lafayette, Indiana

August 2021

**THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL**

Dr. Kari Clase, Chair

School of Agricultural and Biological Engineering

Dr. Jackeline Marmolejo

Proteomics Senior Research Associate

Dr. Ala Samarapungavan

School of Educational Studies

Approved by:

Dr. Nathan S. Mosier

*To my parents, for always encouraging me to try my hardest, do my best,
and believe in my own abilities.*

ACKNOWLEDGMENTS

I would first like to thank my committee chair Dr. Kari Clase for her guidance and support throughout my time as a master's student. I am sincerely grateful for the opportunity to work for her and for the many opportunities she has bestowed upon me that have fostered my passion for research and leadership. I would also like to thank my committee member Dr. Jackeline Marmolejo and the Purdue Proteomics Facility for their constant advice, troubleshooting, and patience. The expertise they have imparted on me and my project is invaluable. I would also like to thank my committee member Dr. Ala Samarapungavan for her support as I navigated new research spaces.

Special thanks to Bruce Cooper and the Purdue Metabolite Profiling Facility for their expertise when processing my lipid data. I would also like to thank the Biotechnology Innovation and Regulatory Science (BIRS) Center for providing funding for my research and to the department of Agricultural and Biological Engineering for providing me with funding as a teaching assistant to a class that I have greatly enjoyed.

Lastly, I would like to thank my family and friends for their constant support and encouragement, for which I am extremely grateful.

TABLE OF CONTENTS

LIST OF TABLES.....	8
LIST OF FIGURES.....	10
ABSTRACT.....	14
CHAPTER 1. INTRODUCTION.....	15
1.1 Statement of Purpose.....	15
1.2 Research Questions.....	17
1.3 Scope.....	17
1.4 Significance of Study.....	17
1.5 Assumptions.....	18
1.6 Limitations.....	18
1.7 Delimitations.....	18
CHAPTER 2. REVIEW OF RELEVANT LITERATURE.....	19
2.1 Bacteriophage Morphotypes and Life Cycles.....	19
2.2 Mycobacteriophage Genomes.....	21
2.3 Purdue University Phages.....	23
2.3.1 Cluster A3.....	25
2.3.2 Cluster B1.....	26
2.3.3 Cluster C1.....	26
2.4 Applications of Phages.....	26
2.5 Bacteriophage-Host Interaction.....	27
2.6 Mycobacteriophage Proteomics.....	29
2.7 Mycobacteriophage Lipidomics.....	29
CHAPTER 3. METHODOLOGY.....	31
3.1 Growing <i>M. smegmatis</i> Cell Cultures and Determining <i>M. smegmatis</i> Growth Curve.....	31
3.2 Optimizing Protein and Lipid Extraction Protocols for <i>M. smegmatis</i>	32
3.2.1 Sample Preparation and BCA Protein Assay.....	32
3.2.2 Acetone Extraction.....	33
3.2.3 Bligh-Dyer Extraction.....	33
3.2.4 Methyl- <i>tert</i> -Butyl Ether (MTBE) Extraction.....	33

3.2.5	Protein Reduction, Alkylation, and Digestion	34
3.2.6	Protein Sample Clean-up.....	34
3.3	Preparation of Mycobacteriophage Lysates	35
3.4	Phage Inoculation of <i>M. smegmatis</i> and Measuring Growth Curve	35
3.5	Using Optimal Protein and Lipid Extraction Methods on Phage-treated Samples.....	37
3.5.1	Sample Preparation and BCA Protein Assay	37
3.5.2	Lipid Extraction	38
3.5.3	Protein Extraction	38
3.6	Liquid Chromatography and Tandem Mass Spectrometry (LC-MS/MS)	39
3.7	Data Analysis.....	41
3.7.1	Protein Data Analysis	41
3.7.2	Lipid Data Analysis	42
3.8	Clean-up and Archival of Previously Discovered Mycobacteriophages	43
3.9	Using Simulations to Investigate the Mycobacteriophage Cluster B1 Frameshift	44
3.9.1	Choosing Proteins of Interest	44
3.9.2	Structural Comparisons	45
3.9.3	Molecular Dynamics Simulations.....	46
CHAPTER 4.	RESULTS.....	48
4.1	<i>M. smegmatis</i> and Phage-treated Growth Curves	48
4.2	Testing Multiple Protein Extraction Methods	50
4.3	Testing Lipid Extraction Methods	53
4.4	Examining Lipids from Phage-Treated Samples	54
4.5	Proteomic Data Analysis FrenchFry Case Study	61
4.6	Archived Phage Samples	68
4.7	Mycobacteriophage Cluster B1 Frameshift Investigation	70
4.7.1	Structural Comparisons	70
4.7.2	Molecular dynamics simulations	75
CHAPTER 5.	DISCUSSION	84
5.1	<i>M. smegmatis</i> Growth Curves	84
5.2	Testing Multiple Methods of Protein Extraction.....	85
5.3	Testing Multiple Methods of Lipid Extraction.....	87

5.4	Investigation of Lipids from Phage-treated <i>M. smegmatis</i>	89
5.5	Proteomic Data Analysis Case Study.....	92
5.6	Archival of Purdue University Phages.....	93
5.7	Investigation of Cluster B1 frameshift proteins	94
5.7.1	Structural comparisons	94
5.7.2	Molecular Dynamics Simulations.....	96
CHAPTER 6. CONCLUSION.....		99
APPENDIX.....		101
REFERENCES		141

LIST OF TABLES

Table 2-1. List of all mycobacteriophages isolated and sequenced at Purdue University as of 2020 alongside the year they were isolated, the year they were annotated, and their cluster.	25
Table 3-1. Two phages from diverse clusters selected to infected <i>M. smegmatis</i> cell cultures. ...	35
Table 4-1. PANTHER classifications of significant proteins in each extraction method according to the GO term cellular components.	51
Table 4-2. List of fold change significant lipids for <i>M. smegmatis</i> in each lipid extraction method categorized by lipid class by MS-DIAL nomenclature standards. A fold change greater than 1.5 was used to determine statistical significance. The term upregulated denotes which test the lipid group was concentrated in and was derived from MetaboAnalyst.	53
Table 4-3. List of significant lipids in mycobacteriophage Zalkecks treated samples categorized by their abundance over time. To determine significance, ANOVA with an adjusted p-value (FDR) cutoff of 0.05 was used and Fisher's LSD was used for post-hoc analysis. The list of each pattern over time was cross-referenced with the control samples of just <i>M. smegmatis</i> over time and any lipids found in both groups were removed.	58
Table 4-4. List of significant lipids in mycobacteriophage PotatoSplit treated samples categorized by their abundance over time. To determine significance, ANOVA with an adjusted p-value (FDR) cutoff of 0.05 was used and Fisher's LSD was used for post-hoc analysis. The list of each pattern over time was cross-referenced with the control samples of just <i>M. smegmatis</i> over time and any lipids found in both groups were removed.	59
Table 4-5. List of significant lipids categorized by lipid class by MS-DIAL nomenclature standards over a range of statistical tests and indicating which are upregulated in each group when applicable. Significance was determined by a fold change greater than 1.5 or in the ANOVA tests, an adjusted p-value (FDR) cutoff of 0.05 was used and Fisher's LSD was used for post-hoc analysis. M indicates the control samples only containing <i>M. smegmatis</i> , P indicates the mycobacteriophage PotatoSplit treated samples, and Z indicates the mycobacteriophage Zalkecks treated samples. .	60
Table 4-6. List of FrenchFry peptides found in mass spectrometry data along with their corresponding gene number, annotated function, and the frequency of how often they were counted in the data. The protein ID indicates the original label from the MaxQuant search.	61
Table 4-7. Number of fold change significant (fold change greater than 1.5) and t-test significant (p-value lower than 0.05) proteins in each statistical test. E indicates the proteins present in the samples infected at the exponential phase of bacterial growth while S indicates the proteins present in the samples infected at the station phase of bacterial growth. Hour 4 indicates the samples taken after four hours of inoculation with phage FrenchFry and hour 10 indicates the samples taken after ten hours of inoculation with phage FrenchFry. FF indicates the FrenchFry-treated samples while NC indicates the negative control of just <i>M. smegmatis</i>	64
Table 4-8. List of all mycobacteriophages that were either re-archived for Purdue University or sent to the HHMI Science Education Alliance-Phage Hunters Advancing Genomics and	

Evolutionary Science (SEA-PHAGES) program at the University of Pittsburgh for long-term storage. Phages in this list that were not sent into HHMI had already been previously archived with the program.68

Table 4-9. Top 5 PDB results for tail assembly chaperone 1. RMSD is the root-mean-square deviation and is the measure of the average distance between the atoms of superimposed proteins. Identity is the percentage sequence identity in the structurally aligned region. Coverage is the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by the length of the query protein.70

Table 4-10. The consensus prediction of GO terms for tail assembly chaperone 1 among the top scoring templates. The GO-Score associated with each prediction is defined as the average weight of the GO term, where the weights are assigned based on Cscore^{GO} of the template.70

Table 4-11. Top 5 PDB results for tail assembly chaperone 2. RMSD is the root-mean-square deviation and is the measure of the average distance between the atoms of superimposed proteins. Identity is the percentage sequence identity in the structurally aligned region. Coverage is the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by the length of the query protein.71

Table 4-12. The consensus prediction of GO terms for tail assembly chaperone 2 among the top scoring templates. The GO-Score associated with each prediction is defined as the average weight of the GO term, where the weights are assigned based on Cscore^{GO} of the template.72

Table 4-13. Top 5 PDB results for the supposed frameshift tail assembly chaperone. RMSD is the root-mean-square deviation and is the measure of the average distance between the atoms of superimposed proteins. Identity is the percentage sequence identity in the structurally aligned region. Coverage is the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by the length of the query protein.73

Table 4-14. The consensus prediction of GO terms for supposed frameshift tail assembly chaperone among the top scoring templates. The GO-Score associated with each prediction is defined as the average weight of the GO term, where the weights are assigned based on Cscore^{GO} of the template.73

LIST OF FIGURES

Figure 1-1. Process flow diagram of the bulk of experiments in this research project.	16
Figure 2-1. General structure of bacteriophages showing the capsid head, tail and tail fibers [5].	19
Figure 2-2. The bacteriophage lytic life cycle [5].	20
Figure 2-3. The bacteriophage lysogenic life cycle [5].	21
Figure 2-4. Cluster and subcluster visualization of all mycobacteriophages [12].	22
Figure 2-5. Example of a translational frameshift in bacteriophages [5].	23
Figure 2-6. Different bacterial host strategies to defend against multiple stages of the bacteriophage lytic life cycle represented in red [26].	27
Figure 2-7. Schematic of the cell envelope of <i>Mycobacteria</i> [47].	30
Figure 3-1. Overall methodology for testing different protein and lipid extraction protocols on <i>M. smegmatis</i> . Each protein and lipid extraction contained three technical replicates for statistical significance.	32
Figure 3-2. Example calculation of finding the volume needed to achieve an MOI of 10.	36
Figure 3-3. Overall methodology for testing optimal protein and lipid extraction protocols on phage-treated samples. Each protein and lipid extraction contained three biological replicates for statistical significance.	37
Figure 3-4. The DNA sequence of the suspected frameshift in the B1 Cluster, shown using DNA Master [32]. Tail assembly chaperone one is highlighted in yellow, in reading frame +1. Tail assembly chaperone two is highlighted in blue, in reading frame +3. The suspected area of the shift is boxed in blue, with the suspected slippery sequence boxed in red.	45
Figure 4-1. The average OD600 of <i>M. smegmatis</i> taken every two hours for 34 hours total with error bars indicating one standard deviation away from the average.	48
Figure 4-2. The average OD600 of each sample taken every four hours for 24 hours total with error bars indicating one standard deviation away from the average. Blue represents the control of only the host <i>M. smegmatis</i> , grey represents the sample of <i>M. smegmatis</i> treated with mycobacteriophage PotatoSplit, and orange represents the sample of <i>M. smegmatis</i> treated with mycobacteriophage Zalkecks. Each sample has three biological replicates for statistical significance.	49
Figure 4-3. Venn diagram indicating how many proteins of <i>M. smegmatis</i> were present in each of the three extraction methods.	50
Figure 4-4. Principal component analysis plots comparing the phage-treated samples (mycobacteriophage Zalkecks, Z, on the left and mycobacteriophage PotatoSplit, P, on the right) to the control sample of <i>M. smegmatis</i> host. M represents the dataset of <i>M. smegmatis</i> host. Each	

plot represents the first principal component on the x-axis and the second principal component on the y-axis each with their corresponding percentage of variance.	54
Figure 4-5. Principal component analysis plots comparing different time points of the phage-treated samples (mycobacteriophage Zalkecks on the left and mycobacteriophage PotatoSplit on the right). Bacteria heavy indicates the time points 0 and 4 hours which closely followed the OD600 of the control, <i>M. smegmatis</i> , while phage heavy indicates the time points 12 and 24 hours in which the presence of phage had a significant impact on OD600 compared to the control. Each plot represents the first principal component on the x-axis and the second principal component on the y-axis each with their corresponding percentage of variance	55
Figure 4-6. Heatmap of all significant ANOVA lipids found in mycobacteriophage PotatoSplit treated samples at each time point taken.	56
Figure 4-7. Boxplots of a significant lipid in the Zalkecks treated samples (left) and PotatoSplit treated samples (right).	57
Figure 4-8. Number of significant FrenchFry proteins expressed in the statistical comparison of mycobacteriophage FrenchFry treated samples to the negative control of just <i>M. smegmatis</i> grouped by functional classifications. The FrenchFry proteins represented in blue are upregulated in the FrenchFry data while those represented in orange are downregulated. NKF indicates no known function, DNA R & T indicates those involved in DNA replication and translation.	63
Figure 4-9. PANTHER pathway analysis of statistically significant proteins from the comparison of the exponential and stationary phases of inoculation based on the Gene Ontology terms corresponding to molecular function. The largest categories of binding and catalytic activity were broken into subcategories. Each category and subcategory listed contains the number of sequences applicable to the category.....	65
Figure 4-10. PANTHER pathway analysis of statistically significant proteins from the comparison of the exponential and stationary phases of inoculation based on the Gene Ontology terms corresponding to biological process. The largest categories of cellular process and metabolic process were broken into subcategories. Each category and subcategory listed contains the number of sequences contained within it.	66
Figure 4-11. PANTHER pathway analysis of statistically significant proteins from the comparison of the exponential and stationary phases of inoculation based on the Gene Ontology terms corresponding to protein class. The largest category of metabolite interconversion enzyme was broken into subcategories. Each category and subcategory listed contains the number of sequences contained within it.	67
Figure 4-12. The fold enrichment of the DAVID pathway analysis functional annotation results based off the KEGG pathway terms of the statistically significant proteins determined by MetaboAnalyst in the comparison of the exponential and stationary phases of inoculation.	68
Figure 4-13. The secondary structure of tail assembly chaperone 1 superimposed on the secondary structure of bacteriophage HK97 tail assembly chaperone (PDB ID: 2OB9). The structure in cyan is the I-TASSER predicted structure while the rest is the HK97 tail assembly chaperone coloring for different secondary structures.	71

Figure 4-14. The secondary structure of tail assembly chaperone 2 superimposed on the secondary structure of bacteriophage HK97 tail assembly chaperone (PDB ID: 2OB9). The structure in magenta is the I-TASSER predicted structure while the rest is the HK97 tail assembly chaperone coloring for different secondary structures. 72

Figure 4-15. The secondary structure of the supposed annotated frameshift superimposed on the secondary structure of bacteriophage HK97 tail assembly chaperone (PDB ID: 2OB9). The structure in cyan is the I-TASSER predicted structure while the rest is the HK97 tail assembly chaperone coloring for different secondary structures. 74

Figure 4-16. The secondary structure of the supposed annotated frameshift superimposed on the predicted structure of the annotated frameshift from *Gordonia terrae* phage CloverMinnie. The structure in blue is the I-TASSER predicted structure while the rest is the CloverMinnie frameshift, coloring for different types of secondary structures. 74

Figure 4-17. The potential energy curve for the energy minimization step of the simulation when running the Tail Assembly Chaperone 1 protein. The energy drops and reaches a plateau, meaning the energy has been properly minimized. 75

Figure 4-18. The temperature curves for the Tail Assembly Chaperone 1 protein during the Autoclave simulation (blue) and the Cell simulation (orange). The autoclave reference temperature is 394.15 K, while the cell reference temperature is 310.15 K. 76

Figure 4-19. The pressure curves for the Tail Assembly Chaperone 1 protein during the Autoclave simulation and the Cell simulation. The Autoclave Data is shown in blue while the 10 ps running average is in green. The Cell Data is shown in yellow while the 10 ps running average is in purple. The reference pressure is shown in red. 78

Figure 4-20. The RMSD of the Tail Assembly Chaperone 1 protein over 0.1 ns. Both simulations, Autoclave and Cell, are compared with the original crystal structure and the equilibrated structure. The Autoclave simulation has higher RMSD because the protein deteriorates at this temperature. 79

Figure 4-21. (A) the initial structure of the Tail Assembly Chaperone 1 protein. (B) The structure after 1.0 ns of the Cell simulation. (C) The structure after 0.1 ns of the Autoclave simulation. ... 79

Figure 4-22. The RMSD of the Tail Assembly Chaperone 2 protein over 1.0 ns. Both simulations, Autoclave and Cell, are compared with the original crystal structure and the equilibrated structure. The Autoclave simulation has higher RMSD because the protein deteriorates at this temperature. 80

Figure 4-23. (A) the initial structure of the Tail Assembly Chaperone 2 protein. (B) The structure after 1.0 ns of the Cell simulation. (C) The structure after 1.0 ns of the Autoclave simulation. ... 80

Figure 4-24. The RMSD of the Tail Assembly Chaperone 3 protein over 0.1 ns. Both simulations, Autoclave and Cell, are compared with the original crystal structure and the equilibrated structure. The Autoclave simulation has higher RMSD because the protein deteriorates at this temperature. 81

Figure 4-25. (A) the initial structure of the Tail Assembly Chaperone 3 protein. (B) The structure after 1.0 ns of the Cell simulation. (C) The structure after 0.1 ns of the Autoclave simulation. ... 81

Figure 4-26. The root mean square fluctuation of the Tail Assembly Chaperone 3 protein after 100 ns of simulation across the entirety of the protein. Blue represents the Autoclave simulation while orange represents the Cell simulation. The red circle indicates the area of the protein containing the translational frameshift. 82

Figure 4-27. The root mean square fluctuation of the Tail Assembly Chaperone 3 protein after 100 ns of simulation within the beginning of the frameshifted region of the protein. Blue represents the Autoclave simulation while orange represents the Cell simulation. The datapoints in yellow represent the amino acid of the frameshift. 83

ABSTRACT

Antibiotic resistance has been an increasing threat to humans since the inception of antibiotics. It can affect anyone, and the issue gets increasingly worse the more antibiotics are used over time. Bacteriophages are one potential way to fight back by harnessing their ability to infect and kill specific bacterial hosts, however, more needs to be studied about them before potential medical applications can be implemented.

The purpose of this research is to analyze the proteins and lipids being produced in the bacteriophage-host interaction to better understand their relationship through the use of mass spectrometry and bioinformatics tools. In this study, the bacterial growth curve of *Mycobacterium smegmatis* was measured to determine the time of ideal bacteriophage inoculation. Methods of protein and lipid extraction were then tested on *M. smegmatis* to determine the most effective protocols applicable to the bacterial host and thereby mycobacteriophages. With the ideal extraction protocols and time at which to inoculate the host, two phages were chosen, extraction was implemented, and mass spectrometry was performed on the proteins and lipids present. This exploratory study is based on the data analysis, showing what proteins and lipids are produced as a result of bacteriophage inoculation over time and what that illuminates about the bacteriophage-host interaction. Through the use of modern methods of untargeted proteomics and lipidomics, one has the capability to fill these gaps of what is being produced by the bacteriophage and host in this interaction and expand upon potential bacteriophage functions to provide a more comprehensive understanding of the pathogenesis of the infection.

Proteomic analysis determined that the acetone method of extraction was the most applicable to *M. smegmatis* and was used for further phage-treated samples and the method of proteomic data analysis tested on readily available mycobacteriophage data. Lipidomic analysis determined that the Bligh Dyer method of extraction was the most applicable to *M. smegmatis* and was used for further phage-treated samples. The lipids extracted from the phage-treated samples were categorized according to classification and showed functions relating to the cell membrane and energy utilization. Specific lipids from the phage-treated samples also indicated involvement in the abortive infection mechanisms of the phage-host interaction.

CHAPTER 1. INTRODUCTION

This chapter introduces the investigation of a proteomic and lipidomic mycobacteriophage mass spectrometry data research study. This chapter provides a statement of purpose of the research, research questions, scope, significance of the study, assumptions, limitations, and delimitations.

1.1 Statement of Purpose

Antibiotic resistance is a huge medical issue facing people today. It can infect anyone, and the issue gets increasingly worse the more antibiotics are used over time. Bacteriophages are one potential way to fight back by harnessing their ability to infect and kill specific bacterial hosts. The purpose of this research is to analyze the proteins and lipids being produced in the bacteriophage-host interaction to better understand their relationship through the use of cutting-edge technology. As outlined in Figure 1-1. Process flow diagram of the bulk of experiments in this research project., the bacterial growth curve of *Mycobacterium smegmatis* was measured to determine the time of ideal bacteriophage inoculation. The method of protein and lipid extraction was then tested on *M. smegmatis* to determine the most effective protocols applicable to the bacterial host and thereby mycobacteriophages, the viruses that infect the specific host. With the ideal extraction protocols and time at which to inoculate the host, two phages were chosen, extraction was implemented, and mass spectrometry was performed on the proteins and lipids present. This exploratory study is then based on the data analysis done of the data, showing what proteins and lipids are produced as a result of bacteriophage inoculation over time and what that illuminates about the bacteriophage-host interaction.

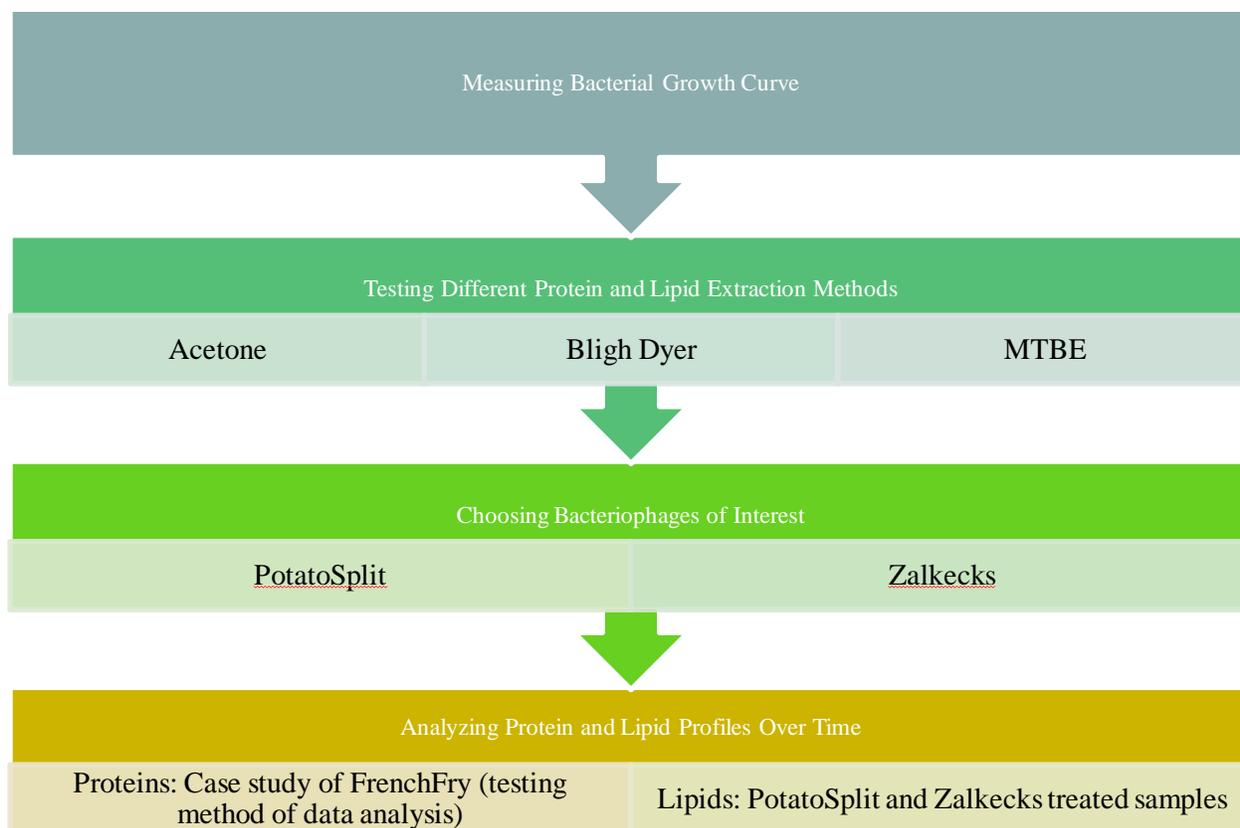


Figure 1-1. Process flow diagram of the bulk of experiments in this research project.

Another goal of this research project is to present deliverables pertaining to the archival of Purdue University phages. Novel mycobacteriophages isolated from Purdue University students in the years 2018 and 2019 were contaminated and of low titer. Through the course of this study, they were cleaned of contamination, their titers amplified, and were properly archived for long-term storage.

The final goal of this research project is to investigate the bacteriophage cluster B1 frameshift using structural comparisons and simulations. Multiple bioinformatics programs were utilized to predict and compare protein structure as well as run simulations of protein behavior at two different temperatures and pressures. Through comparing the protein structure and behavior to knowns, one can speculate about the location of the cluster B1 frameshift.

1.2 Research Questions

- Can one extraction method be used to multiplex *Mycobacteria smegmatis* samples?
- Are there significant proteins and lipids being produced in the phage-host interaction?

1.3 Scope

The scope of this study was a proteomic and lipidomic analysis of mycobacteriophages and their bacterial host using high-performance liquid chromatography-electrospray ionization-tandem mass spectrometry (HPLC-ESI-MS/MS) and various data analysis tools. The research focused on the characterization of proteins and classifications of lipids to determine features of the bacteriophage-host relationship of a host that has relevant medical applications. Statistical significance of the data was determined through the bioinformatics tool MetaboAnalyst and significant proteins were further analyzed in the protein pathway analysis programs PANTHER and DAVID while lipids were grouped based on their lipid classes.

1.4 Significance of Study

Antibiotic resistance has been an increasing threat to humans since the inception of antibiotics in the 1940s. It is estimated that over 35,000 people in the United States die due to antibiotic resistant infections every year. One potential way to combat this widespread problem is with bacteriophages: viruses that infect and kill their bacterial hosts. Bacteriophages infect their host, use the host cell's metabolic machinery to propagate, and kill the cell in the process. They outnumber bacteria in population 10:1, however, not everything is known about bacteriophages. Before they are used in a wide range of applications, especially medical applications, the entirety of their genome needs to be understood to ensure safe and effective use. Through the use of modern methods of untargeted proteomics and lipidomics, one has the capability to fill these gaps of what is being produced by the bacteriophage and host in this interaction and expand upon potential bacteriophage functions to provide a more comprehensive understanding of the pathogenesis of the infection.

1.5 Assumptions

The following assumptions are inherent to the pursuit of the study:

- *Mycobacterium smegmatis* and mycobacteriophage samples do not contain contaminants.
- Protein and lipid profile data generated by mass spectrometry is accurate.
- The lipid database reliably matches the MS2 spectra of the lipids in the bacterial host and bacteriophage-treated samples.
- The method of protein and lipid extraction most applicable to the bacterial host is also the most applicable to the phages that infect it.

1.6 Limitations

The following limitations are inherent to the pursuit of the study:

- MS-DIAL was used to identify the lipids in the mass spectrometry samples.
- The length of time for the protein simulations was dependent upon the computing network. The size of the protein being run affected how long the simulation could be run without crashing.

1.7 Delimitations

The following delimitations are inherent to the pursuit of the study:

- MetaboAnalyst was the only program used to calculate statistical significance in the protein and lipid mass spectrometry samples.
- Only the programs PANTHER and DAVID were used for protein pathway analysis.

CHAPTER 2. REVIEW OF RELEVANT LITERATURE

2.1 Bacteriophage Morphotypes and Life Cycles

Bacteriophages (phages) are viruses that infect and kill bacterial hosts [1]. Currently, bacteriophages outnumber bacteria 10:1 and represent the majority of all organisms on the planet [2]. They have a basic structure of a capsid head containing its genome, a tail used for motility and transporting the phage DNA into that of the host, and tail fibers used for the attachment of the phage to the host as seen in Figure 2-1. General structure of bacteriophages showing the capsid head, tail and tail fibers [5]. Phages are often host specific, meaning they often only infect one species of bacterial host, though there are some that can infect different species or genera [2]. There is a great interest in mycobacteriophages, which infect the host *Mycobacteria*, because exploring their method of infection has led to a greater understanding of pathogenic hosts like *Mycobacterium tuberculosis* and *Mycobacterium lepre* [3]. Mycobacteriophages are also categorized into three morphotypes: *Myoviridae*, *Podoviridae*, and *Siphoviridae* with the distinction between being between the length of the tail and whether the tail is contractile [4].

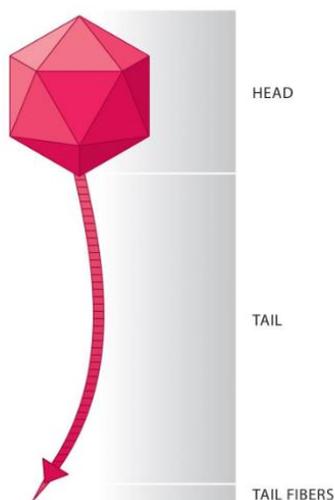


Figure 2-1. General structure of bacteriophages showing the capsid head, tail and tail fibers [5].

Bacteriophages have two different life cycles. In the lytic life cycle, shown in Figure 2-2. The bacteriophage lytic life cycle [5]., the phage attaches to the cell wall of the host and inserts its

DNA inside. Once the lytic phage infects its selected host, the phage DNA circularizes to avoid detection from the host and uses the host's machinery to replicate pieces of itself. These replicated pieces are then assembled into more phages and cell lysis occurs when the phages produce an enzyme to break the cell wall, bursting the cell membrane [6].

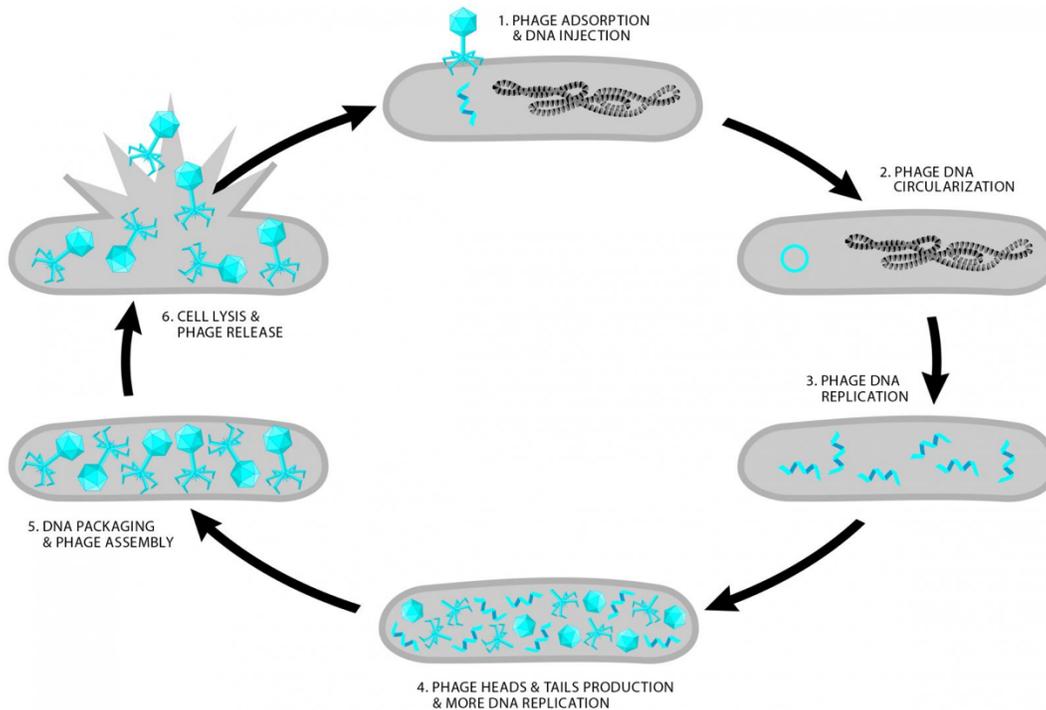


Figure 2-2. The bacteriophage lytic life cycle [5].

In the lysogenic life cycle, used by temperate phages, the genes required for the lytic cycle are repressed, and instead the phage's genetic information is maintained within the host cell normally through homologous recombination [7]. Once integrated into the host cell's DNA, the phage DNA will be passed down to daughter cells. As seen in Figure 2-3. The bacteriophage lysogenic life cycle [5]., temperate phages also utilize the lytic life cycle when the lysogen is in the presence of a stressor, causing the host cell to eventually burst. In single-cell infections, phage will utilize the lytic life cycle 80-90% of the time [5].

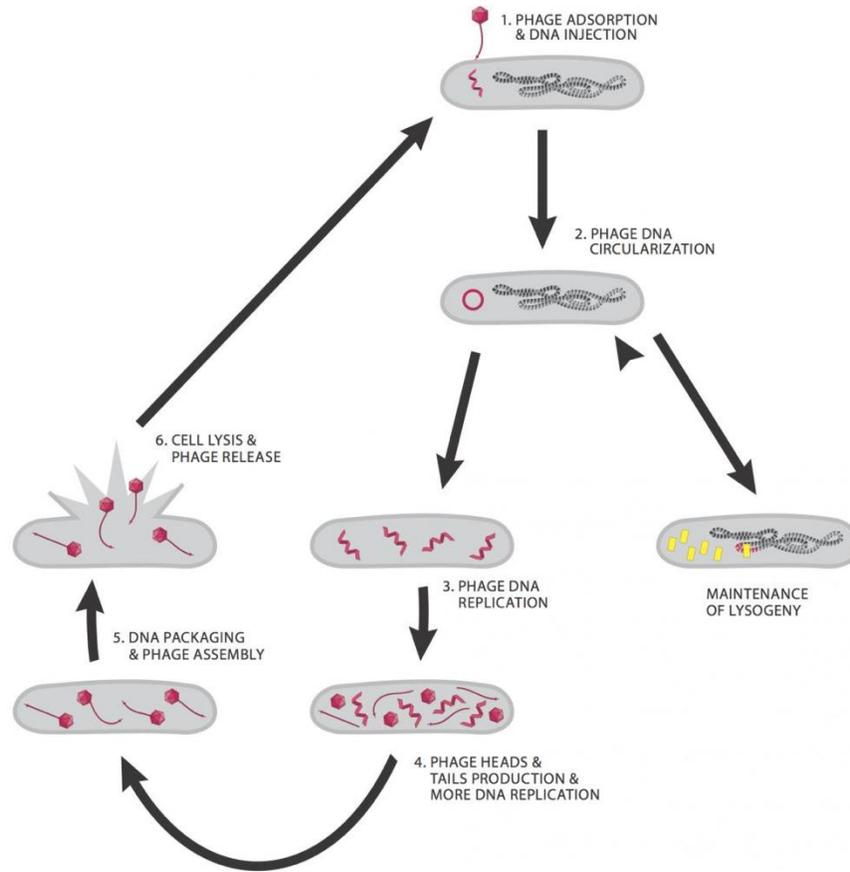


Figure 2-3. The bacteriophage lysogenic life cycle [5].

2.2 Mycobacteriophage Genomes

Mycobacteriophages display a wide range of genetic diversity. Their genomes are relatively small, ranging from 15 kb to 190 kb and densely packed with coding regions as well as display mosaic relationships between other phages [8]. Due to this relationship, phages are sorted into clusters and subclusters based on overall nucleotide similarity, the host that they infect, and physical characteristics [9]. These clusters have different unique characteristics that can be studied individually. Phages that belong to the same cluster or subcluster have a high level of genomic similarity and often share the same genome organization as seen in Figure 2-4. Cluster and subcluster visualization of all mycobacteriophages [12]. The architecture of the phage genome architectures is mosaic, meaning that it is assembled through horizontal gene transfer from segments of distinct evolutionary histories over time [10]. This leads to the creation of cassettes,

or regions of genes that are grouped typically by similar function. In many clusters, cassettes have a specific gene order, and these cassettes are conserved throughout most phages due to their necessity and their locations within the genome are extremely conserved within clusters [11]. A few examples would be a cassette containing all the genes needed for DNA replication, or all the genes necessary for lysing a host cell [8]. Knowledge of where a protein coding gene is located in relation to others with known functions can help identify what its possible function is.

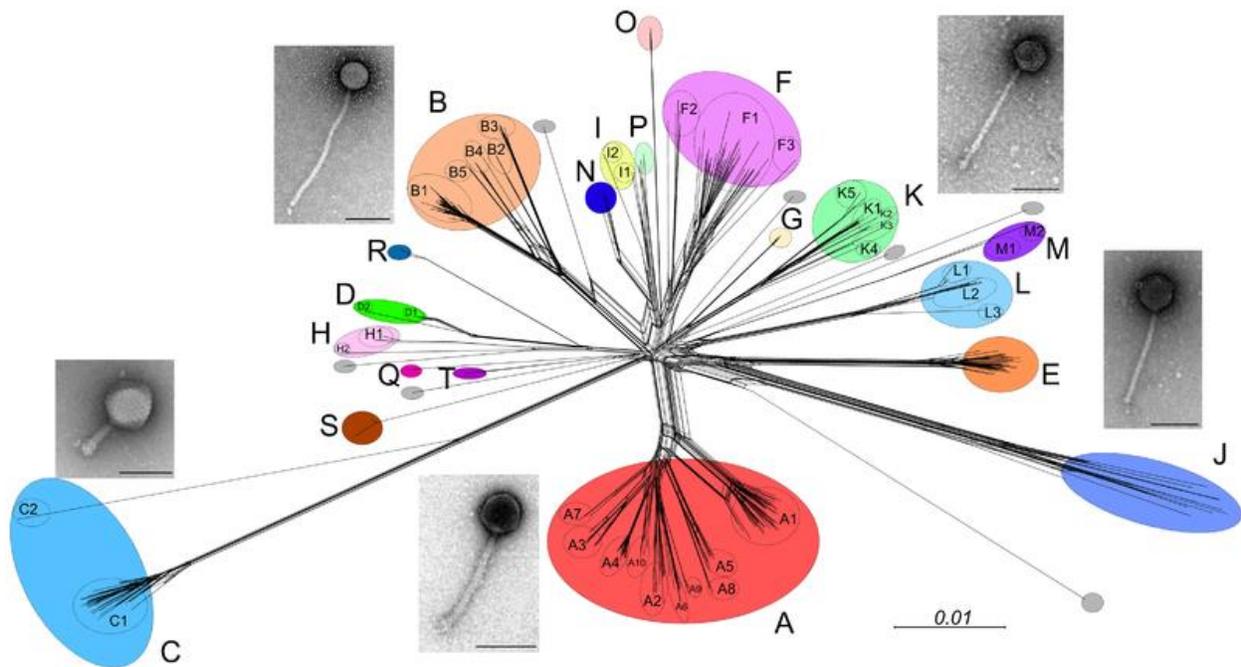


Figure 2-4. Cluster and subcluster visualization of all mycobacteriophages [12].

Individual genes are also grouped together into phams based on basepair sequence. Because phage genomes are highly conserved, many genes are repeated throughout different phages and therefore some phams can contain over 50 members and these members can span over different clusters. If a gene is not placed into a pham due to the lack of similarity with other phams, it is considered an orpham until the database grows to include more similar genes like it [11]. The members of a pham have a highly conserved function due to their genomic similarity.

Though phams may have a conserved function, much more needs to be learned about the scope of bacteriophage functions before they can be used in wider applications. Currently, only

0.0001% of their genetic information is known [13]. More bacteriophages are studied every year, but many of the proteins found to have no known function.

Another feature in common with almost all clusters of phages is a gene that contains a frameshift. Translational frameshifting occurs when the ribosome encounters a “slippery” sequence in the mRNA, such as GGAAAA, and loses track of how to count to three [14]. It also occurs in the two tail assembly chaperone genes of most flexible non-contractile tailed phages and is intentionally programmed into the phage genome [11]. An example of frameshifting is shown in Figure 2-5. Example of a translational frameshift in bacteriophages [5]. The bacteriophage can therefore utilize the protein product of either the shorter or longer DNA sequence and it is a way for the phage to maximize the production of protein products in a small base pair region. These frameshifts have been confirmed in wet lab settings; however, the B1 Cluster has no known frameshift [11].

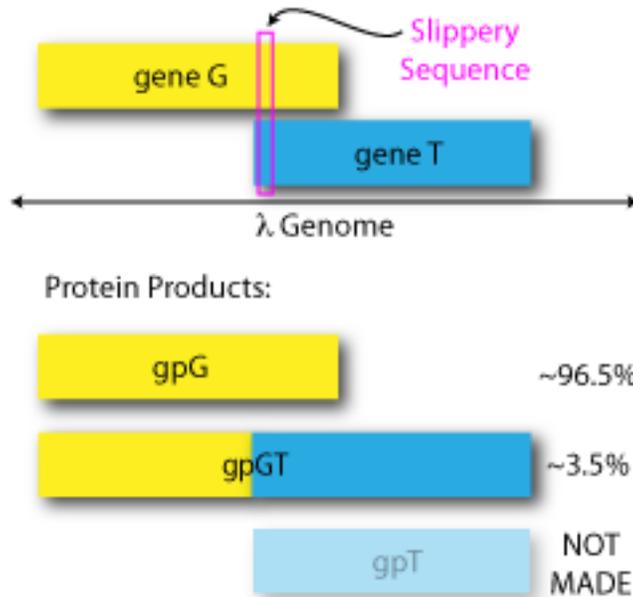


Figure 2-5. Example of a translational frameshift in bacteriophages [5].

2.3 Purdue University Phages

Every year, Purdue University’s bacteriophage archive is updated every year to include more novel mycobacteriophages found by Purdue University students as a part of the HHMI

Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) program. Those that are sequenced, as seen in Table 2-1, are annotated and their genomes are uploaded to GenBank.

Table 2-1. List of all mycobacteriophages isolated and sequenced at Purdue University as of 2020 alongside the year they were isolated, the year they were annotated, and their cluster.

Phage Name	Year Found	Year Annotated	Cluster
AFIS	2014	2019	A1
Petp2012	2012	2019	A1
MrGordo	2010	2011	A1
Fibonacci	2012	2019	A11
PotatoSplit	2012	2019	A3
Lemur	2012	2019	A4
JewelBug	2012	2019	A6
Cosmolli16	2015	2019	B1
Grand2040	2014	2019	B1
Mesh1	2014	2019	B1
Sophia	2012	2019	B1
Waterdiva	2012	2019	B1
Maru	2019	2020	B1
FrenchFry	2015	2019	B2
Zalkecks	2017	2019	C1
Czyszczo1	2010	2019	E
RiverMonster	2010	2018	E
Cactus	2011	2020	E
MilleniumForce	2012	2019	F1
Bobi	2010	2013	F1
Ochi17	2017	2019	F1
Royals2015	2015	2020	F1
EricMillard	2012	2019	J
Hughesyang	2014	2019	J
Krili	2018	2019	O
NiebruSaylor	2019	2020	O
VasuNzinga	2014	2018	S

2.3.1 Cluster A3

Cluster A3 is a subsection of Cluster A, which is the largest cluster of Actinobacteriophages, mainly consisting of mycobacteriophages [15]. Cluster A3 phages have a broad host range, which makes them of interest for medical applications, including *M. tuberculosis*, BCG, *Mycobacterium scrofulaceum*, *Mycobacterium fortuitum*, *Mycobacterium chelonae*, and some strains of both *Mycobacterium ulcerans* and *Mycobacterium avium*, in

addition to *M. smegmatis* [16]. Phages in this cluster are mostly temperate and are siphoviridae, with short-to-medium length flexible tails. There is currently only one A3 mycobacteriophage within the Purdue University sequenced archive as of 2020: PotatoSplit.

2.3.2 Cluster B1

Cluster B1 is a subsection of Cluster B and consists of lytic phage, the results for which are experimental due to that any attempts to isolate lysogens have been unsuccessful [17]. Cluster B virions often contain a linear genome with terminally redundant and circularly permuted ends [11]. Their translational frameshift has also not been confirmed in wet lab [16].

2.3.3 Cluster C1

Cluster C1 is a subsection of Cluster C and also consists of lytic phage, the results for which are experimental due to that any attempts to isolate lysogens have been unsuccessful [18]. All mycobacteriophages with the *Myoviridae* morphotype are contained within cluster C [11].

2.4 Applications of Phages

Due to their ability to infect a wide range of bacterial hosts, bacteriophages have been of increasing interest in multiple fields for their potential applications including medical, food safety, and wastewater treatment. In terms of medical applications, there has been renewed interest in phages being used as potential treatments of antibiotic bacterial resistant infections. According to the World Health Organization, “antibiotic resistance is one the of the biggest threats to global health, food security, and development today” because as the use of antibiotics to treat non-lethal infections increases, the less effective they become at treating them [19]. Phage therapy has been one way to combat this problem and capitalizes on the use of lytic phage particles which are devoid of any metabolic machinery and do not possess an affinity for eukaryotic cells [20]. In recent years, phage therapy has been authorized by the FDA in a few cases as a last resort for bacterial infections that did not respond to antibiotics including *Mycobacterium abscessus* and *Mycobacterium tuberculosis* [21]. Phages are also being used in food safety applications to protect consumers from food-borne illnesses like *Salmonella enterica* and *Escherichia coli* [22]. They have been used to treat domesticated livestock to reduce intestinal colonization, to decontaminate surfaces in food-

processing facilities, and in direct sprays or washes of harvested foods [23]. Lastly, phages have been used to improve sludge and effluent emissions into the environment at wastewater treatment facilities by attacking *Salmonella enterica* and *Escherichia coli* [24]. In summary, phages have the potential to be utilized in a myriad of applications to manage bacterial populations.

2.5 Bacteriophage-Host Interaction

The interaction between phages and their bacterial hosts is dynamically evolving. Bacteria are under enormous evolutionary pressure since they are outnumbered 10:1 and phages rely upon their hosts to propagate and must circumvent the many antiviral barriers the hosts put in place to do so [25]. Host bacteria have developed strategies at multiple stages of the bacteriophage lytic life cycle to defend against foreign invaders and often layer multiple for increased efficiency, as seen in Figure 2-6. Different bacterial host strategies to defend against multiple stages of the bacteriophage lytic life cycle represented in red [26]., while phages develop counter-strategies to overcome the blocks set forth by the host.

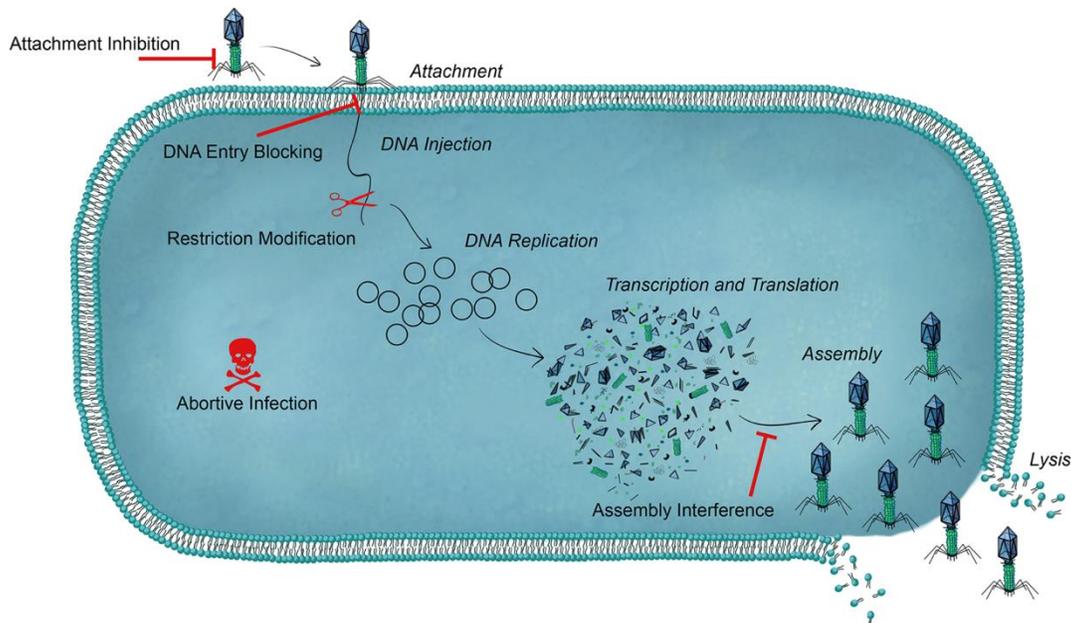


Figure 2-6. Different bacterial host strategies to defend against multiple stages of the bacteriophage lytic life cycle represented in red [26].

Bacteria have learned to develop strategies to defend against phage invaders and often layer multiple defense strategies to increase the likelihood that the bacterial lineage survives post phage

exposure [27]. One of the first lines of defense includes extracellular blocks like capsules that inhibit the phage from approaching the bacterial surface [27]. Bacteria also utilize adsorption resistance strategies that prevent the phage from binding to the surface of the bacteria cell by changing the surface receptors that phages target [28]. There are other intracellular blocks that the host can utilize including homoimmunity, the ability to recognize specific phage motifs, which enables blocking phage replication. Bacteria can also employ restriction-modification systems to detect restriction enzyme recognition sequences to cut the inserted phage genome [26]. This strategy is potentially risky in that if there are too many restriction enzyme recognition sequences being targeted, it increases the likelihood that the phage DNA is modified and can therefore evade recognition [27]. However, once recognized, bacteria can employ CRISPR/Cas systems to become phage resistant through attaining novel-to-host DNA sequences [29]. Lastly, if the host is infected, it can employ an abortive infection to kill phages even at the cost of its death [25]. By layering different phage-resistant strategies at different areas of the bacteriophage lytic life cycle, the host cell is able to increase its ability to detect and evade harmful pathogens.

Phages rely on counter-strategies to overcome the blocks put forth by the host to propagate. If bacterial hosts have employed extracellular blocks to mask their receptors with capsular polysaccharides, many phages contain degrading enzymes that can hydrolyze the capsule to access the binding receptor [28]. Phages also have the ability to modify their receptor-binding protein so that they can adsorb to the evolving host cell binding receptors and use diversity-generating retroelements to increase the amount of genetic diversity in the receptor-binding proteins to increase the likelihood of adsorbing to multiple host binding receptors [25]. Phages also deploy strategies for intracellular blocks from the host which mainly pertain to the host being able to target foreign DNA. To protect themselves from restriction-modification systems, phage with little amounts of restriction enzyme sites have a selective advantage [30]. The specific placement of restriction enzyme sites in the bacteriophage genome can play a role in how well the host can recognize foreign DNA as well. In some host systems, if the two sequences are not in a specific orientation or if the restriction enzyme sites are too far apart, the host will be unable to recognize them [31]. To protect themselves from CRISPR/Cas systems, phage can substitute a specific single nucleotide within the protospacer sequences to evade detection or interfere with the CRISPR/Cas system [28]. Phage can also contain their own CRISPR/Cas systems to allow the completion of the phage lytic life cycle and give the phage the ability to hijack the bacterial CRISPR/Cas systems

to promote their own propagation [28]. Phages also contain countermeasures to defend against abortive infection mechanisms which normally utilize toxins to kill both the invading phage and host. Phages mainly circumvent this hurdle by mutating specific genes to hijack the production of antitoxins that neutralize the bacterial toxin [28]. Overall, phages are constantly evolving ways to overcome the blocks set forth by the host to use the cell to propagate.

2.6 Mycobacteriophage Proteomics

Currently, bacteriophage genomes are sequenced and annotated using bioinformatic tools and software such as DNA Master [32] and PECAAN [33] to gather information on the proteins being produced by the phage. By comparing the amino acid sequences of known coding regions to databases like pFam [34], PDB [35], Conserved Domain [36], SCOPe [37], and Non-Redundant [38], one can discern a potential function or general classification of function. However, using this method, a majority of bacteriophage functions are unknown; in all of the genomes annotated at Purdue University, about 69% of genes have no known function. Mass spectrometry has tried to address this gap by confirming the genome annotation, accurately identifying proteins through comparing to databases of protein sequences, and therefore giving more insight into potential functions [39]. Most of the known functions pertain to cassettes that perform well-known processes like lysis, replication, and structure [8,16].

2.7 Mycobacteriophage Lipidomics

The field of lipidomics has been rapidly developing in recent years due to its importance in metabolism [40]. By applying methods of untargeted lipidomics to bacteriophages and their hosts, it can lead to a greater understanding of the role of lipid metabolism in infection pathogenesis [41], and lipids have already been found in mycobacteriophages D29 [42], DS6A [43], and R1 [44]. One main target area in bacteriophage lipidomics is to identify and study cellular receptors since not many are known. A specific peptidoglycolipid, mycoside C(sm), has been found in *M. smegmatis* lipidomic data to play a role in the binding of phage D4 [45] and it also been found that glycolipids may act as receptors for adsorption as found in mycobacteriophage Phlei [46]. It is also worthwhile to examine the lipids being produced within the host, many of which can be found in the cell envelope for *Mycobacterium smegmatis*, to see if levels of production change due to phage

infection to better understand the phage-host interaction. As seen in Figure 2-7. Schematic of the cell envelope of *Mycobacteria* [47]., the cell envelope of the host is a complex space with many types of lipids present, yet a high number of glycolipids can be seen [47]. Though there may not be information about individual lipids being produced in these interactions, classifying groups of lipids can lead to a greater understanding of the system.

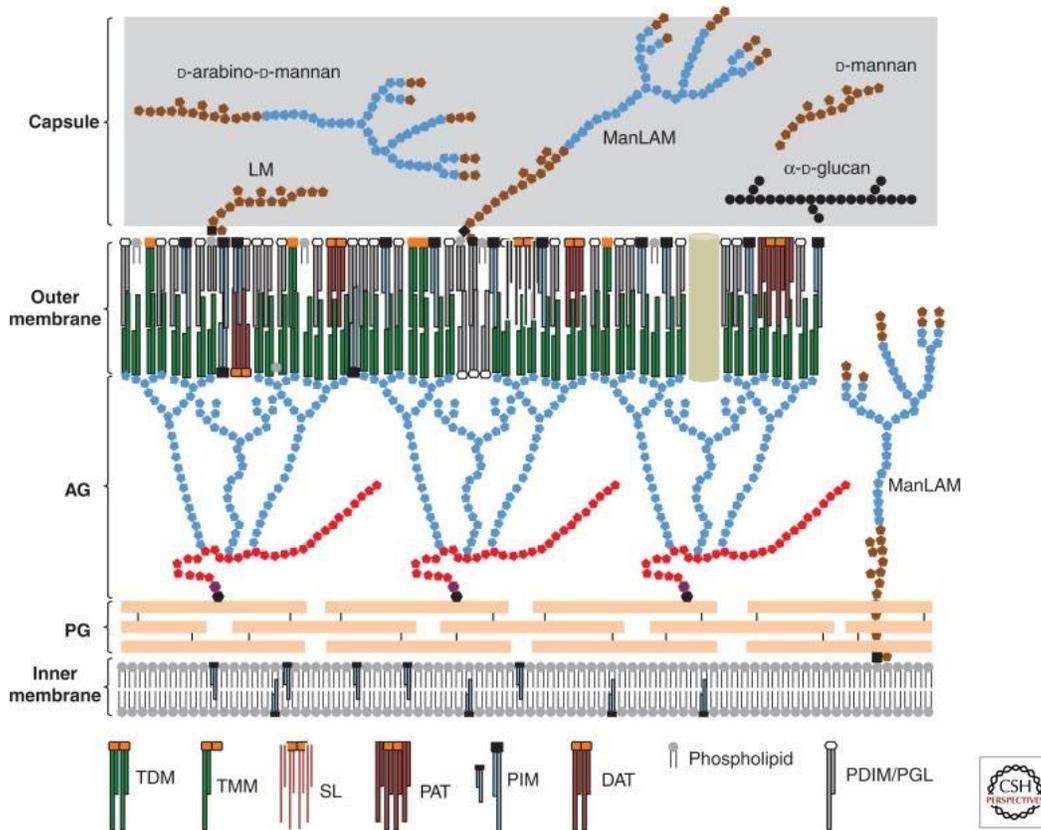


Figure 2-7. Schematic of the cell envelope of *Mycobacteria* [47].

CHAPTER 3. METHODOLOGY

3.1 Growing *M. smegmatis* Cell Cultures and Determining *M. smegmatis* Growth Curve

The *M. smegmatis* strain mc² 155 cell culture was reconstituted from -80°C stock by streak plating on an LB agar plate and incubating at 37°C for 72 hours. A single colony was then used to create a Passage 1 From Frozen (P1FF) stock by inoculating 7H9 liquid medium containing 50 mL 7H9 Middlebrook broth supplemented with 0.05% Tween[®]80, 1 mM calcium chloride, 10% AD supplement, 0.02% glycerol, 50 ug/ml carbenicillin, and 10 ug/ml cycloheximide which was incubated for 72 hours at 37°C with constant agitation at 250 rpm. Then a 250 ml Passage 2 From Frozen (P2FF) stock was created by diluting the P1FF stock in 7H9 liquid medium without Tween[®]80 (P1FF:7H9 liquid medium = 1:1000) and incubated at 37°C with shaking at 250 rpm for 72 hours. The P2FF culture was also used for the preparation of mycobacteriophage lysates.

The OD₆₀₀ of the *M. smegmatis* P2FF culture was determined every 2 hours for a total of 34 hours. The OD₆₀₀ values were then plotted and the growth curve of *M. smegmatis* was drawn. This P2FF culture at 48 hours of growth after the initial P1FF inoculation was used to test different protein and lipid extraction methods. Once the ideal extraction method for proteins and lipids was determined, another P2FF stock was created in the same process to be inoculated with different phages.

3.2 Optimizing Protein and Lipid Extraction Protocols for *M. smegmatis*

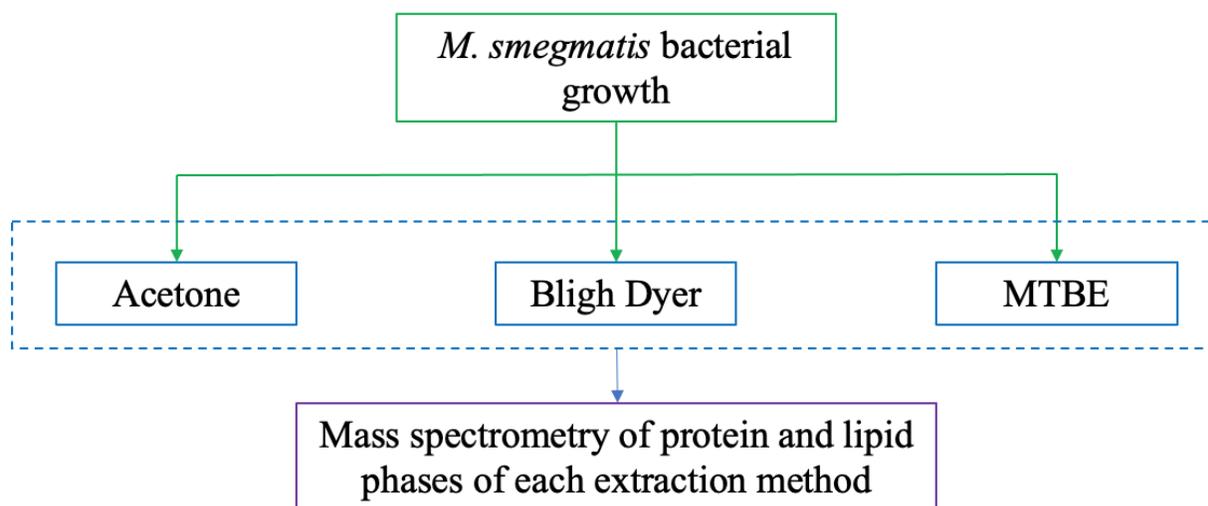


Figure 3-1. Overall methodology for testing different protein and lipid extraction protocols on *M. smegmatis*. Each protein and lipid extraction contained three technical replicates for statistical significance.

3.2.1 Sample Preparation and BCA Protein Assay

1 ml of the aforementioned *M. smegmatis* P2FF culture was collected, the cells were pelleted at 14,000 rpm for 10 min at 4°C, and the supernatant was removed. The pellet was then washed three times by resuspending it in 850 µl of Phosphate Buffer Saline (PBS) pH 7.4, centrifuging at 14,000 rpm for 10 min, and then removing the supernatant. After the final wash, the pellet was resuspended in 1400 µl of 100 mM Ammonium Bicarbonate (ABC) and was split into four precellys tubes [48] of equal volume due to the high concentration of cells. The precellys tubes were run at 6200 rpm for three rounds of 20 seconds to lyse the cells. They were then spun down, and the samples were taken out of the precellys tube and placed into microcentrifuge tubes. The BCA protein assay [49] was then used to determine the quantification of total protein in the samples. This was achieved by diluting each tube of the sample with double distilled water with a 1:10 dilution. 10 µl of the diluted samples were placed in wells alongside the BCA standards. 200 µl of BCA reaction mix (a 50:1 mixture of solutions A and B) was then added to each well of samples and standards. The plate was left to incubate for 30 minutes at 37°C before reading. With the reading of concentration, the volume of each tube to achieve 50 µg of protein was found and

was used to test three different protein and lipid extraction methods: acetone, Bligh-Dyer, and Methyl-*tert*-Butyl Ether (MTBE).

3.2.2 Acetone Extraction

For the acetone extraction method, the volume to achieve 50 ug of protein from the BCA protein assay was added to a 1.5 ml microcentrifuge tube. A volume of -20°C 100% acetone four times the amount to achieve 50 ug of protein was added to each replicate. Each sample was left for 12 hours at -20°C. The samples were then pelleted at 14,000 rpm for 10 minutes at 4°C. The supernatant containing lipids was removed and placed into a secondary microcentrifuge tube. Both the microcentrifuge tubes containing the protein pellet and the lipid supernatant were dried in a speedvac with no heat for 1-2 hours. The dried lipids were stored at -80°C until mass spectrometry could be performed on the lipids while the proteins needed further preparation. This process was performed three times to obtain technical replicates.

3.2.3 Bligh-Dyer Extraction

For the Bligh-Dyer extraction method, the volume to achieve 50 ug of protein from the BCA protein assay was first added to a 1.5 ml microcentrifuge tube. Double distilled water was added to bring the total volume up to 200 µl and the mixture was homogenized. The sample was then mixed for 10 minutes with 250 µl of chloroform and 500 µl of methanol. 200 µl of double distilled water was added and mixed gently. The sample was then centrifuged at 4000 rpm for 10 minutes and the three phases were collected separately into different microcentrifuge tubes. The upper phase (metabolites), middle phase (proteins), and bottom phase (lipids) were placed in a speedvac to dry with no heat for 1-2 hours. The dried metabolites and lipids were stored at -80°C until mass spectrometry could be performed on the lipids while the proteins needed further preparation. This process was performed three times to obtain technical replicates.

3.2.4 Methyl-*tert*-Butyl Ether (MTBE) Extraction

For the MTBE extraction method, the volume to achieve 50 ug of protein from the BCA protein assay was first added to a 1.5 ml microcentrifuge tube. The sample was then mixed for 10 minutes with 375 µl of methanol and 1250 µl MTBE. 312.5 µl of double distilled water was added

and mixed gently. The sample was then centrifuged at 4000 rpm for 10 minutes and the three phases were collected separately into different microcentrifuge tubes. The upper phase (lipids), middle phase (metabolites), and bottom pellet (proteins) were placed in a speedvac to dry with no heat for 1-2 hours. The dried metabolites and lipids were stored at -80°C until mass spectrometry could be performed on the lipids while the proteins needed further preparation. This process was performed three times to obtain technical replicates.

3.2.5 Protein Reduction, Alkylation, and Digestion

10 µl of 8 M urea and 10 mM dithiothreitol (DTT) was added to each protein pellet from the three different extraction methods and was incubated for 1 hour at 37°C with agitation at 800 rpm in a thermomixer. 10 µl of TEP mix (97.5% Acetonitrile (ACN), 2% Iodoethanol, and 0.5% Triethylphosphine (TEP)) was added to each protein subsample and incubated 1 hour at 37°C with agitation at 800 rpm in a thermomixer. The samples were then dried in a speedvac without heat for 2 hours.

Barocycler tubes were then rinsed with ACN and left to dry. For the trypsin digestion, trypsin [50] was dissolved in 25 mM ABC to make a 0.05 µg/µl concentration and 20 µl of enzyme mixture was added to each protein subsample to achieve an enzyme-to-substrate ratio of 1:50. The mixture was transferred to the clean barocycler tubes, which were then capped and loaded into the barocycler. A barocycler NEP2320 [51] ran the samples at 50°C for 60 cycles of 50 seconds at 20 kpsi and 10 seconds at atmospheric pressure.

3.2.6 Protein Sample Clean-up

Nest columns [52] for each protein sample were conditioned by adding 100 µl of ACN and centrifuging for 1 minute at 800 rpm, adding 100 µl of double distilled water and centrifuging for 1 minute at 800 rpm, and adding 100 µl of double distilled water and centrifuging for 1 minute at 1200 rpm. The protein subsamples were then loaded into their corresponding Nest columns [52] and centrifuged for 1 minute at 800 rpm, checking to see that the sample had run through the columns. The columns were then washed by adding 100 µl of 0.1% formic acid in double-distilled water and centrifuging for 1 minute at 800 rpm, adding another 100 µl of 0.1% formic acid in double-distilled water and centrifuging for 1 minute at 1200 rpm, and lastly centrifuging at 3000

rpm to remove any remaining liquid. The protein subsamples were then eluted out of the nest columns into new microcentrifuge tubes by adding 3 washes of 50 μ l of 80% ACN and 0.1% formic acid and centrifuging for 1 minute at 800, 1200, and 3000 rpm respectively. The protein subsamples were then dried at 45°C in the speedvac for about 3 hours. After the samples had dried, they were stored at -80°C until mass spectrometry was performed.

3.3 Preparation of Mycobacteriophage Lysates

Two phages (Table 3-1) isolated at Purdue University were selected to infect *M. smegmatis* cell cultures based off their differing subclusters, life cycles, morphotypes, and potential proteins of interest in their genomes. Streaks of each phage lysate from -80°C stocks were plated on LB agar plates and were incubated for 48 hours at 37°C. A single plaque from each plate was used in a serial dilution using prepared phage buffer. 10 μ l of each dilution was mixed with 250 μ l of *M. smegmatis* P2FF culture and left to inoculate on the bench for 10 minutes. The mixture was then mixed with top agar, poured onto an LB agar plate, and incubated at 37°C for 48 hours. The dilution that created a webbed plate was selected and soaked with 5 ml of phage buffer and left for 12 hours at 4°C. The phage lysate was then collected and filtered with a 0.22 μ m filter. Another serial dilution of each phage lysate was performed to calculate the phage forming unit per ml (pfu/ml) using the formula: $\frac{\text{number of plaques}}{\text{volume used}} * 10^3 * \text{dilution factor}$ [5].

Table 3-1. Two phages from diverse clusters selected to infected *M. smegmatis* cell cultures.

Phage	Cluster	GC% Content	Life Cycle	Morphotype
PotatoSplit	A3	64.0	Temperate	Siphoviridae
Zalkecks	C1	64.7	Lytic	Myoviridae

3.4 Phage Inoculation of *M. smegmatis* and Measuring Growth Curve

The growth curve of *M. smegmatis* was determined and an OD600 of 1.0 was chosen as the indicator of the stationary growth phase. P1FF stock was added to 25 ml of 7H9 liquid medium without Tween®80 (P1FF:7H9 liquid medium = 1:1000) and incubated at 37°C till it reached an OD600 of around 0.7 for each subsample before it was inoculated with phage lysate. For the phage-treated subsamples, phage lysate was added into each cell culture subsample with a multiplicity of

infection (MOI) of 10, was mixed gently for 5 minutes to initiate adsorption of the phage to the host cell and was incubated for 24 hours at 37°C with agitation at 250 rpm. The calculations for which can be found in Figure 3-2. Example calculation of finding the volume needed to achieve an MOI of 10. and the concentration of *M. smegmatis* cells per volume was taken from study that measured the population over time [53]. For the control samples, the cell culture subsamples were inoculated with phage buffer. The OD600 of the phage-bacteria and control mixtures were measured at hours 0, 1, 2, 4, 8, 12, 16, 20, and 24 using a nanophotometer NP80 [54] and plotted to obtain the growth curves. 7H9 liquid medium without Tween®80 was used as the standard. After 0, 4, 12, and 24 hours of phage infection, 1 ml of the subsamples were harvested and transferred into sterile microcentrifuge tubes for further protein and lipid extraction and 10 µl of the subsamples were used in a serial dilution to validate phage infection. This was performed three times for each phage and for the control of just *M. smegmatis* to obtain biological triplicates.

$$MOI = \left(\frac{\text{phage titer} \left(\frac{pfu}{ml} \right) * \text{volume of phage titer}}{\left(\frac{M. smegmatis \text{ cells}}{ml} \right) * \text{volume of bacteria}} \right)$$

$$10 = \frac{6.44E10 * x}{7.09E6 * 25} \rightarrow x = 27.5 \mu l$$

Figure 3-2. Example calculation of finding the volume needed to achieve an MOI of 10.

3.5 Using Optimal Protein and Lipid Extraction Methods on Phage-treated Samples

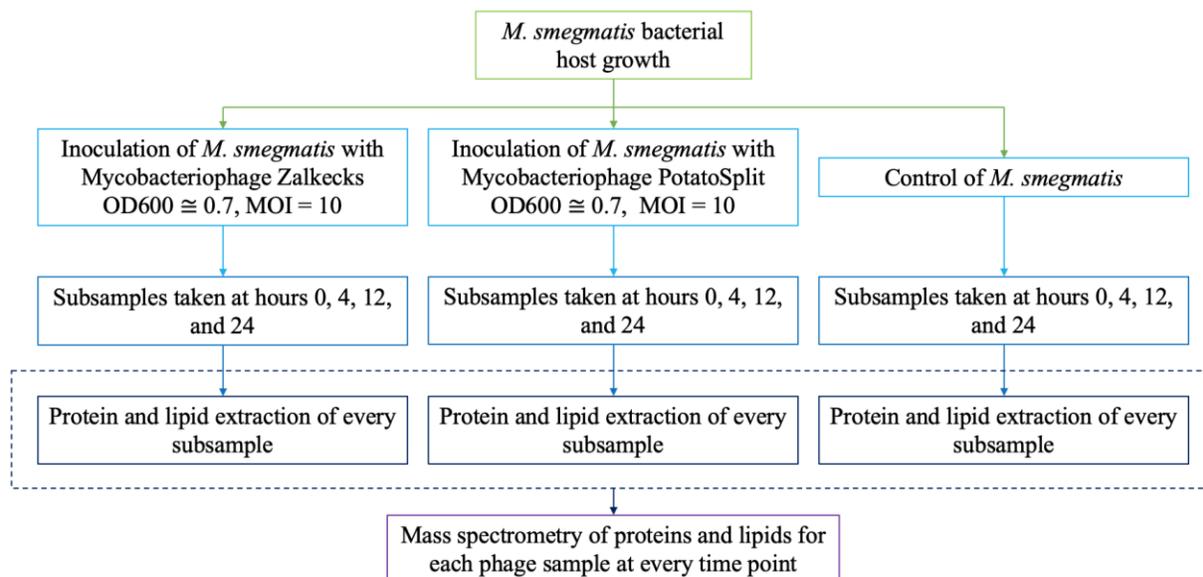


Figure 3-3. Overall methodology for testing optimal protein and lipid extraction protocols on phage-treated samples. Each protein and lipid extraction contained three biological replicates for statistical significance.

3.5.1 Sample Preparation and BCA Protein Assay

1 ml of the subsamples from the two different phage-bacteria mixtures and control at hours 0, 4, 12, and 24 were pelleted at 14,000 rpm for 10 min at 4°C and the supernatant were removed. The pellets were then washed three times by resuspending it in 850 µl of Phosphate Buffer Saline (PBS) pH 7.4, centrifuging at 14,000 rpm for 10 min, and then removing the supernatant. After the final wash, the pellets were resuspended in 300 µl of 100 mM Ammonium Bicarbonate (ABC) and sonicated for 3 min while the samples were on ice to lyse the cells. The BCA protein assay [49] was then used to determine the quantification of total protein in each of the subsamples. This was achieved by diluting each subsample with double distilled water with a 1:10 dilution. 10 µl of the diluted samples were placed in wells alongside the BCA standards. 200 µl of BCA reaction mix (a 50:1 mixture of solutions A and B) was then added to each well of samples and standards. The plate was left to incubate for 30 minutes at 37°C before reading. With the reading of concentration, the volume of each subsample to achieve 50 µg of protein was found. The volume to achieve 40 µg of protein was used for further protein extraction of the subsamples and the

volume to achieve 120 ug of protein was used for further lipid extraction of the subsamples, which was based on the results of the different protein and lipid extraction methods and was constrained by the total volume of the subsamples.

3.5.2 Lipid Extraction

For lipid extraction, the Bligh-Dyer extraction method was used, which was based on the previous findings of testing multiple lipid extraction methods. The volume to achieve 120 ug of protein from the BCA protein assay for each subsample was first added to 1.5 ml microcentrifuge tubes. Double distilled water was added to bring the total volume of each subsample up to 200 μ l and the mixtures were homogenized. The subsamples were then mixed for 10 minutes with 250 μ l of chloroform and 500 μ l of methanol. 200 μ l of double distilled water was added to each subsample and mixed gently. The subsamples were then centrifuged at 4000 rpm for 10 minutes and the bottom phase of lipids were collected separately into different microcentrifuge tubes and were placed in a speedvac to dry with no heat for 1-2 hours. The dried lipids were stored at -80°C until mass spectrometry could be performed. This process was performed for every phage-infected and uninfected subsample for a total of 36 subsamples.

3.5.3 Protein Extraction

For protein extraction, the acetone extraction method was used, which was based on the previous findings of testing multiple protein extraction methods. The volume to achieve 40 ug of protein from the BCA protein assay was added to a 1.5 ml microcentrifuge tube for every subsample. A volume of -20°C 100% acetone four times the amount to achieve 40 ug of protein was added to each subsample and was left for 12 hours at -20°C. The subsamples were then pelleted at 14,000 rpm for 10 minutes at 4°C. The supernatant containing lipids was removed and the protein pellets were dried in a speedvac with no heat for 1-2 hours. This process was performed for every phage-infected and uninfected subsample for a total of 36 subsamples.

10 μ l of 8 M urea and 10 mM dithiothreitol (DTT) was added to each protein pellet from the three different extraction methods and was incubated for 1 hour at 37°C with agitation at 800 rpm in a thermomixer. 10 μ l of TEP mix (97.5% Acetonitrile (ACN), 2% Iodoethanol, and 0.5% Triethylphosphine (TEP)) was added to each protein subsample and incubated 1 hour at 37°C with

agitation at 800 rpm in a thermomixer. The samples were then dried in a speedvac without heat for 2 hours.

Barocycler tubes were then rinsed with ACN and left to dry. For the trypsin digestion, (Promega, 2021) Trypsin was dissolved in 25 mM ABC to make a 0.05 ug/ μ l concentration and 20 μ l of enzyme mixture was added to each protein subsample to achieve an enzyme-to-substrate ratio of 1:50. The mixture was transferred to the clean barocycler tubes, which were then capped and loaded into the barocycler. A barocycler NEP2320 [51] ran the samples at 50°C for 60 cycles of 50 seconds at 20 kpsi and 10 seconds at atmospheric pressure.

Nest columns [52] for each protein sample were conditioned by adding 100 μ l of ACN and centrifuging for 1 minute at 800 rpm, adding 100 μ l of double distilled water and centrifuging for 1 minute at 800 rpm, and adding 100 μ l of double distilled water and centrifuging for 1 minute at 1200 rpm. The protein subsamples were then loaded into their corresponding nest columns and centrifuged for 1 minute at 800 rpm, checking to see that the sample had run through the columns. The columns were then washed by adding 100 μ l of 0.1% formic acid in double-distilled water and centrifuging for 1 minute at 800 rpm, adding another 100 μ l of 0.1% formic acid in double-distilled water and centrifuging for 1 minute at 1200 rpm, and lastly centrifuging at 3000 rpm to remove any remaining liquid. The protein subsamples were then eluted out of the nest columns into new microcentrifuge tubes by adding 3 washes of 50 μ l of 80% ACN and 0.1% formic acid and centrifuging for 1 minute at 800, 1200, and 3000 rpm respectively. The protein subsamples were then dried at 45°C in the speedvac for about 3 hours. After the samples had dried, they were stored at -80°C until mass spectrometry was performed.

3.6 Liquid Chromatography and Tandem Mass Spectrometry (LC-MS/MS)

The proteins in the *M. smegmatis* extraction method samples and the phage-treated samples were analyzed through reverse-phase high-performance liquid chromatography–electrospray ionization–tandem mass spectrometry (HPLC–ESI–MS/MS) using the Dionex UltiMate 3000 RSLC nano System [55] to the Q-Exactive High-Field (HF) Hybrid Quadrupole Orbitrap MS [56] and a Nano-electrospray Flex ion source [57]. Reverse phase peptide separation was accomplished using a trap column (300 μ m ID \times 5 mm) packed with 5 μ m 100 Å PepMap C18 medium, and then separated on a reverse phase column (50-cm long \times 75 μ m ID) packed with 2 μ m 100 Å PepMap C18 silica [58]. The column temperature was maintained at 50 °C.

Mobile phase solvent A was 0.1% formic acid (FA) in water and solvent B was 0.1% FA in 80% acetonitrile (ACN). Loading buffer was 98% water/2% ACN/0.1% FA. Peptides were separated by reverse phase by loading into the trap column in a loading buffer for 5-min at 5 $\mu\text{L}/\text{min}$ flow rate and eluted from the analytical column with a linear 82-min linear gradient of 6.5-27% of buffer B, then changing to 40% of B at 90 min, 100% of B at 97-min at which point the gradient was held for 7 min before reverting to 2% of B at 104-min. Peptides were separated from the analytical column at a flow rate of 300 nL/min. The mass spectrometer was operated in positive ion and standard data-dependent acquisition mode with Advanced Peak Detection function activated. The fragmentation of precursor ion was accomplished by higher energy collision dissociation at a normalized collision energy setting of 30%. The resolution of Orbitrap mass analyzer was set to 120,000 and 15,000 at 200 m/z for MS1 and MS2, respectively, with maximum injection time of 50 ms for MS1 and 20 ms for MS2. The dynamic exclusion was set at 60s to avoid repeated scanning of identical peptides and charge state was set at 2-7 with 2 as a default charge and mass tolerance of 10 ppm for both high and low masses. The full scan MS1 spectra were collected in the mass range of 375-1,500 m/z and MS2 in 300-1250 m/z. The spray voltage was set at 2 and Automatic Gain Control (AGC) target of 4e5 for MS1 and 5e4 for MS2, respectively. Three biological sample replicates from each treatment were utilized for LC-MS/MS, which was sufficient for good statistical power. Instrument optimization and recalibration was carried out at the start of each batch run using the Pierce calibration solution.

The lipids from the *M. smegmatis* extraction method samples and the phage-treated samples were dissolved in 50 μl mixture which composed of 50% Eluent A (water, 10 mM ammonium acetate, 0.1% formic acid) and 50% Eluent B (isopropyl alcohol: acetonitrile, 10 mM ammonium acetate, 0.1% formic acid). After centrifuging to remove the remaining particles, the samples (8 $\mu\text{l}/\text{sample}$) were loaded to Waters ACQUITY UPLC® BEH C18 1.7 μm columns in Agilent 6545 Q-TOF [59] for mass spectrometry analysis. The raw data was analyzed using MS-DIAL [60]. The mass, retention time, and intensity of the compounds' positive ions $[\text{M}+\text{H}]^+$ were obtained for both the extraction method samples and the phage-treated samples. The compounds' negative ions $[\text{M}-\text{H}]^-$ were obtained for the extraction method samples only.

3.7 Data Analysis

3.7.1 Protein Data Analysis

For the protein data analysis, the raw MS/MS data (.raw files) were processed using MaxQuant (v1.6.0.16) [61] with its integrated Andromeda search engine searched the spectra against the Uniprot *M. smegmatis* FASTA file [62] in conjunction with a common contaminants database and a reverse-decoy database [63]. Data were searched using trypsin/P enzyme digestion allowing for up to 2 missed cleavages. MaxQuant search was set to 1% FDR (False Discovery Rate) both at the peptide and protein levels. The minimum peptide length required for database search was set to seven amino acids. Precursor mass tolerance of ± 10 ppm, MS/MS fragment ions tolerance of ± 20 ppm, alkylation of cysteine and oxidation of methionine was set as fixed and variable modifications, respectively. The “unique plus razor peptides” were used for peptide quantitation. Razor peptides are the non-redundant, non-unique peptides assigned to the protein group with most other peptides. LFQ intensity values were used for relative protein abundance measurement. Proteins detected with at least 1 unique peptide and at least 2 MS/MS counts were only included for the final analysis.

When filtering the MaxQuant data, proteins with reverse or contamination identification were first removed. Any MS/MS counts of one in any sample or counts of two if only appearing in one of the biological triplicates were removed. Then, proteins with no total MS/MS counts were filtered out of the dataset. Any proteins with multiple protein IDs were also limited to one to limit redundancy. When examining the proteins associated with testing different method of protein extraction, the MS/MS count was used for data analysis as a measure of quantitative abundance while the LFQ intensity was used as the primary quantitative measure of abundance for testing the phage-treated samples. The datasets were then normalized and Metaboanalyst (v5.0) was used to compute statistical analysis [64]. Depending on how many statistical groups were being tested, fold changes and t-tests or ANOVA was used to determine significant proteins. For fold change unpaired analysis, FCs are calculated as the ratios between two group means using data before column-wise normalization was applied and a fold change threshold of 1.5 was used. For t-test unpaired analysis, an equal group variance and a p-value threshold of 0.05 was used to compute significance. For ANOVA analysis, an adjusted p-value (FDR) cutoff of 0.05 was used and Fisher’s LSD was used for post-hoc analysis.

For the investigation into the phage-treated samples, a case study was performed on previously isolated mycobacteriophage FrenchFry treated samples to explore the method of data analysis. First, MetaboAnalyst was used to determine significant proteins, using a fold change threshold of 1.5 and a t-test unpaired analysis with an equal group variance and a p-value threshold of 0.05. Significant proteins were analyzed through the PANTHER pathway analysis program [65]. PANTHER couples Gene Ontology terms to give an overall scope of function to the datasets [66]. and used the PANTHER pathway analysis program. Gene ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of significantly regulated proteins was also performed using DAVID (v6.8) [67,68]. The peptides in this case study were further analyzed by creating a Python (v3.9.5) script that searched the peptide sequence within the amino acid sequences of the FrenchFry coding regions to determine which proteins from its genome annotation were being expressed. Though the FrenchFry case study examined different research objectives, the method of data analysis shown can be applied to the phage-treated Zalkecks and PotatoSplit samples and is an effective method for using multiple bioinformatics tools to accumulate data.

3.7.2 Lipid Data Analysis

For the lipid data analysis, the raw MS/MS data (.raw files) were processed using MS-DIAL (v4) [60]. When testing the methods of extraction, both positive and negative ion modes were selected, while only positive ion mode was selected when analyzing the phage-treated samples. In data collection, an MS1 tolerance of 0.01 Daltons and an MS2 tolerance of 0.025 Daltons was used. The retention time was set to begin at 0.5 minutes and end at 16 minutes. The MS1 mass range was set to begin at 50 Daltons and end at 1200 Daltons. A max number of two charged molecules was used and the number of threads was set to four. In peak detection, the minimum peak height was set to an amplitude of 250. A mass slice width of 0.1 Daltons, a linear weight moving average smoothing method with a smoothing level of 3 and a minimum peak width of 5 was selected. In MS2 detection, a sigma window value of 0.1 and an MS/MS abundance cut off of 5 was used. In identification, all default setting were used. The following adducts were also selected for testing extraction methods: [M-H]⁻, [M-2H]²⁻, [M+CH₃COO]⁻, [M+2H]²⁺, [M+H-H₂O]⁺, [M+H]⁺, [M+Na]⁺, and [M+NH₄]⁺. The following adducts were selecting when analyzing the phage-treated samples: [M+2H]²⁺, [M+H-H₂O]⁺, [M+H]⁺, [M+Na]⁺, and

[M+NH₄]⁺. The data was searched against an MS-DIAL internal lipid databases and an identification score cut off of 80% was set.

Results of MS-DIAL were then filtered further. Blanks and any result without MS₂ spectra data was filtered out. Any result with an S/N below 20 was also filtered out. The rest of the peak data was cleaned so that the area under the peak was accurately represented for every mass spectrometry sample. For the phage-treated lipids samples only, any result without a reference match to the MS₂ spectra data was further analyzed by MS-Finder to predict a chemical formula of the unknown lipid [69,70].

With the filtered lipid data, the area under the peak was used to determine significance through MetaboAnalyst (v5.0) [64]. Depending on how many statistical groups were being tested, fold changes and t-tests or ANOVA was used to determine significant lipids. For fold change unpaired analysis, FCs are calculated as the ratios between two group means using data before column-wise normalization was applied and a fold change threshold of 1.5 was used. For t-test unpaired analysis, an equal group variance and a p-value threshold of 0.05 was used to compute significance. For ANOVA analysis, an adjusted p-value (FDR) cutoff of 0.05 was used and Fisher's LSD was used for post-hoc analysis. Significant lipids were then grouped by their lipid subclass according to the MS-DIAL nomenclature to report what lipids are relevant in each of the different methods of extraction and potential insight into more comprehensive understanding of the pathogenesis of the infection.

3.8 Clean-up and Archival of Previously Discovered Mycobacteriophages

As a part of the SEA-PHAGES program, Purdue University is required to send archive samples to the University of Pittsburgh and none of from 2018 and 2019 had been sent in. Purdue University phage lysates from 2018 and 2019 were therefore cleaned of contaminants, had their titers amplified to at least 5.0E9, and were re-archived as frozen to be stored at -80°C. A team of nine Purdue University undergraduates worked under supervision on 50 novel bacteriophages to archive or replace their archive samples at Purdue University and 16 novel bacteriophages to be archived for the SEA-PHAGES program. The common protocols used in this process were phage purification (protocol 6.1), serial dilutions (protocol 6.2), collecting plate lysates (protocol 6.3), making webbed plates from a known titer (protocol 7.1), and archiving a phage sample (protocol 7.3) all of which are contained within the SEA-PHAGES Discovery Guide [5].

3.9 Using Simulations to Investigate the Mycobacteriophage Cluster B1 Frameshift

3.9.1 Choosing Proteins of Interest

The bacteriophages used in this investigation were narrowed down to those isolated by Purdue University. Out of all bacteriophages isolated at Purdue University, six of the 27 sequenced phages are categorized as subcluster B1, as seen in Table 2-1. Maru's annotation was not finalized at the time of the experiment and was therefore not included in this investigation. The remaining five bacteriophages were chosen to be analyzed because they are the largest grouping of finalized, Purdue isolated phage within the same subcluster. They therefore have high genomic similarity between one another and findings about the chosen proteins will be applicable to the largest number of Purdue isolated phage.

With these five bacteriophages of interest, specific conserved proteins of no known function were chosen to be analyzed based on their location in the genome and knowledge about the B1 subcluster specifications. By targeting genes that are close to genes with known functions, there is a higher likelihood of finding relevant functional information about the gene [2]. The known and suspected tail assembly chaperones, as well as the frameshift, of the B1 cluster were investigated because cluster B is commonly known to not have its frameshift confirmed [11].

Within the B1 cluster, only one of the two tail assembly chaperones making up the translational frameshift has been confirmed. As seen in Figure 2-5. Example of a translational frameshift in bacteriophages [5]., the first tail assembly chaperone was analyzed in this investigation. The genes before and after the known tail assembly chaperone were run through the BLAST database in order to determine which would likely be the second tail assembly chaperone. The gene downstream of the known tail assembly chaperone was found to be related to different tail assembly chaperones with higher scores and therefore was the second tail assembly chaperone analyzed in this investigation.

After determining the suspected frameshift protein of the B1 Cluster, it had to be annotated to determine the shifted form of the protein. The slippery sequence was located by comparing to known translational frameshift sequences [71]. In Figure 3-4. The DNA sequence of the suspected frameshift in the B1 Cluster, shown using DNA Master [32]. Tail assembly chaperone one is highlighted in yellow, in reading frame +1. Tail assembly chaperone two is highlighted in blue, in reading frame +3. The suspected area of the shift is boxed in blue, with the suspected slippery

sequence boxed in red., the DNA sequence of the likely shift is showing the proteins moving from the +1 reading frame to the +3 reading frame nearing the end of the first protein. In the F1 and O clusters, the common slippery sequence was CTCGAAAA. In the F1 phage, there was a sequence of CTCGGG. While this was not a perfect match, it was the most likely option. All Purdue phage with +1 to +3 shifts had repeated a basepair, so the decision was made to repeat a basepair here. A “G” was repeated in the triplet code.

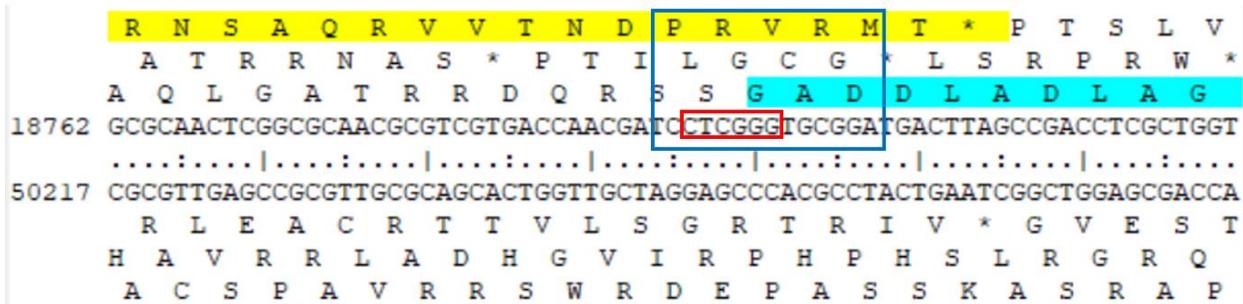


Figure 3-4. The DNA sequence of the suspected frameshift in the B1 Cluster, shown using DNA Master [32]. Tail assembly chaperone one is highlighted in yellow, in reading frame +1. Tail assembly chaperone two is highlighted in blue, in reading frame +3. The suspected area of the shift is boxed in blue, with the suspected slippery sequence boxed in red.

Once annotated, the protein sequence was run through NCBI BlastP. A match was found with a *Gordonia terrae* bacteriophage CloverMinnie [72]. CloverMinnie made the same frameshift annotation decisions that were made in this F1 annotation. Hereafter, Tail Assembly Chaperone 1 will refer to the known tail assembly chaperone, Tail Assembly Chaperone 2 will refer to the experimental found tail assembly chaperone that is not biologically produced by the phage and Tail Assembly Chaperone 3 will refer to the tail assembly chaperone containing the experimentally found translational frameshift.

3.9.2 Structural Comparisons

The sequences for both the individual tail assembly chaperones and the proposed frameshifted protein were then submitted to the software I-TASSER that predicts structures based on protein sequences [73]. I-TASSER outputs the top five model predictions with scores to help determine the best option. It also compares that structure with Protein Data Bank structures, locates possible ligand binding sites, and predicts Gene Ontology terms [74]. These results were analyzed

to determine a possible function classification for each of the NKF proteins selected, which was used to select matching PDB files to be used for structural comparison.

Using PyMOL, the top models of each protein were chosen, and structures examined [75]. The matching PDB files were superimposed on the NKF proteins to examine structural similarities. The MatchAlign score and RMSD were used to evaluate the best matches and multiple models were chosen in order to fully examine this annotation.

3.9.3 Molecular Dynamics Simulations

Simulations were carried out using GROMACS software [76]. GROMACS was used to simulate each protein in a box of water at two sets of temperatures and pressures. The values picked were based on the highest temperature in an autoclave and the ideal temperature for growing *M. smegmatis*. These temperatures are 394.1 K, hereafter referred to as Autoclave, and 310.15, hereafter referred to as Cell. The accompanying pressures are 1.03421 bar and 1.01325 bar respectively.

The process of the simulations began with the coordinate files produced by I-TASSER [73]. With these files, a GROMACS function was used to create a box with the protein placed in the center at least 1.0 nm from the edge. The size of the boxes varies for each protein based on the protein size. The box was then solvated and filled with water molecules. In order to use this software, the net charge of the system had to be neutral. Ions were added to achieve a net charge of 0. Na⁺ ions were added to raise charge and Cl⁻ ions were added to lower charge.

After a box was created and solvated, the energy was minimized to ensure the structure had no inappropriate geometry or steric clashes. If the final energy was negative and stable, the simulation was continued. The step size was 0.01, with a maximum number of steps at 50,000. After energy minimization, the system was equilibrated.

The first equilibration step used was the NVT ensemble, which holds constant the Number of Particles, Volume, and Temperature. The temperature coupling method used was a Berendsen thermostat with a heat bath at the temperature for Autoclave or Cell. The Particle Mesh Ewald method was used for electrostatics, and the Verlet cutoff scheme used for buffered neighbor searching. The cut off value for the radius was adjusted for each box size. This was run for 100 picoseconds for each protein, then the temperature was graphed to ensure it had reached a plateau. If it had not, this step was run for another 100 picoseconds with velocity generation turned off.

The second equilibration used was NPT, which holds constant the Number of Particles, Pressure, and Temperature. Again, the Berendsen thermostat was used, along with PME and Verlet. The barostat used was Parrinello-Rahman, and velocity generation was again off. This was run for 100 picoseconds, then the average pressure was checked to make sure the system was equilibrated properly. If it was not within a close range of the desired pressure, this step was repeated for another 100 picoseconds.

Once the system was equilibrated, a production run of the simulation was run for one nanosecond and data collected. If the protein was too large to run for 1 nanosecond, it was run for 0.5 ns or 0.1 ns. The trajectory files were loaded into PyMOL to capture images and videos of the simulations, and the root mean square distance data was used to examine how the proteins changed over the course of the simulation. The RMSD was compared with the backbone of the molecule for equilibrated structure after the NPT step and the crystal structure.

Tail assembly chaperone 3 was further investigated and was run at the same simulation process listed above for 100 nanoseconds. The root mean square fluctuation (RMSF), which is the average RMSD per each atom of the protein, compared with the alpha carbons of the molecule was then plotted to examine potential stable and unstable portions of the protein.

CHAPTER 4. RESULTS

4.1 *M. smegmatis* and Phage-treated Growth Curves

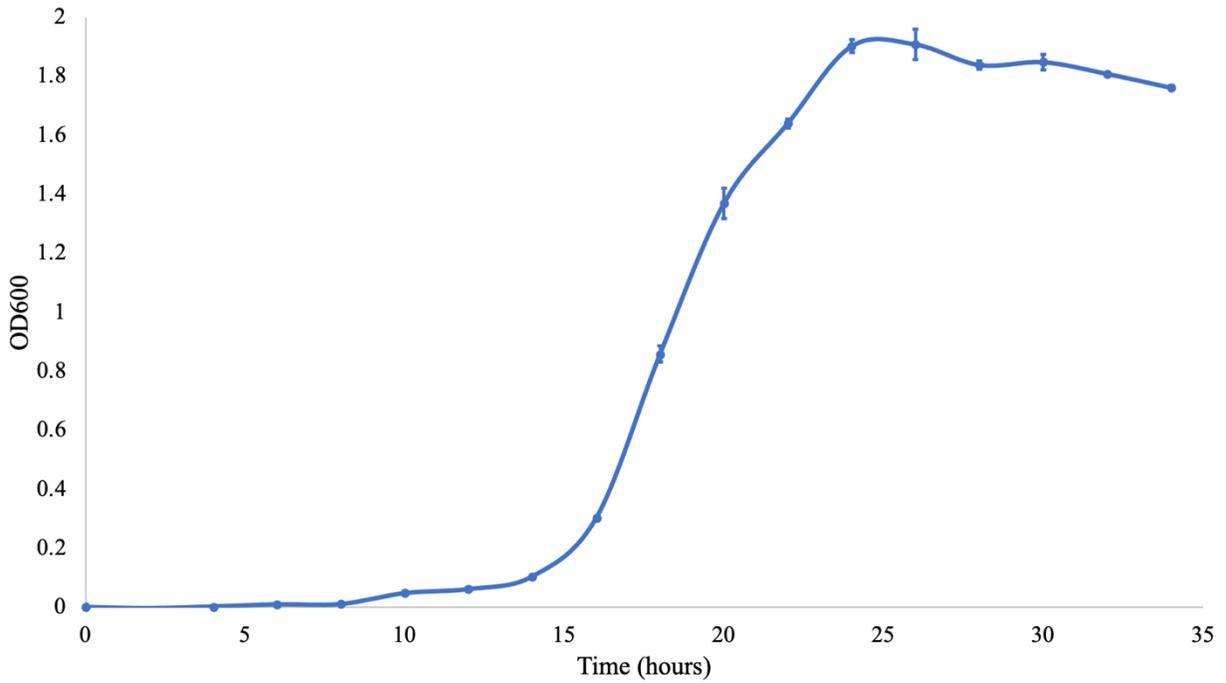


Figure 4-1. The average OD600 of *M. smegmatis* taken every two hours for 34 hours total with error bars indicating one standard deviation away from the average.

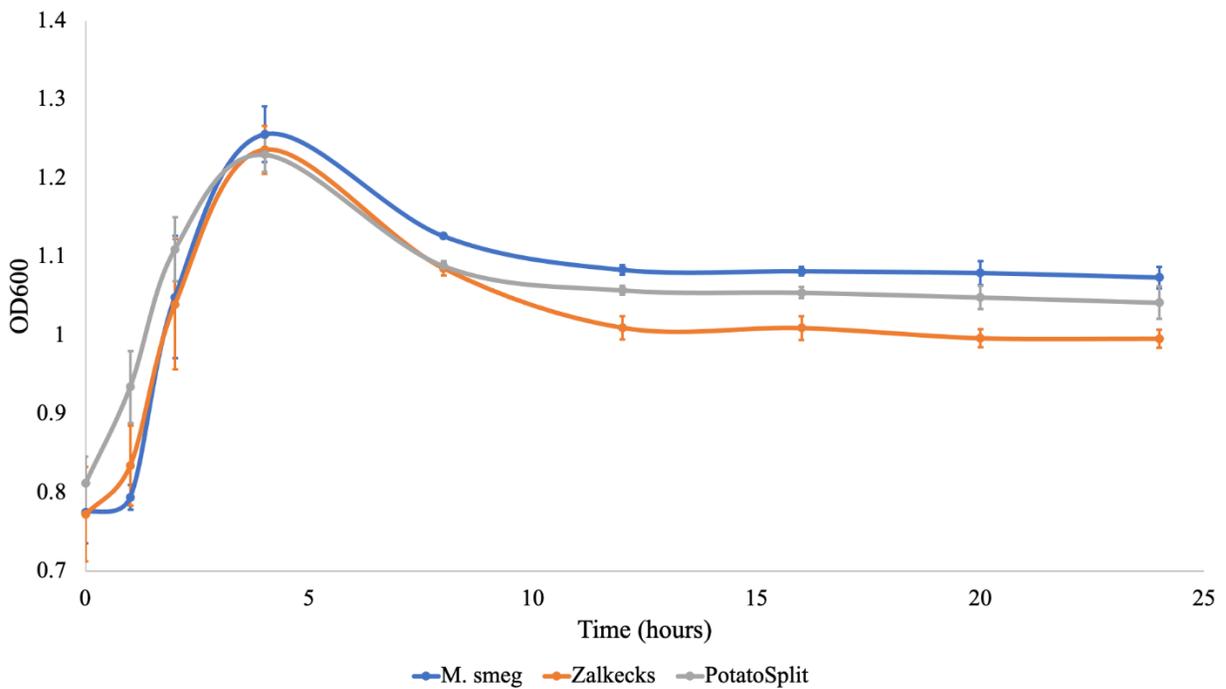


Figure 4-2. The average OD600 of each sample taken every four hours for 24 hours total with error bars indicating one standard deviation away from the average. Blue represents the control of only the host *M. smegmatis*, grey represents the sample of *M. smegmatis* treated with mycobacteriophage PotatoSplit, and orange represents the sample of *M. smegmatis* treated with mycobacteriophage Zalkecks. Each sample has three biological replicates for statistical significance.

4.2 Testing Multiple Protein Extraction Methods

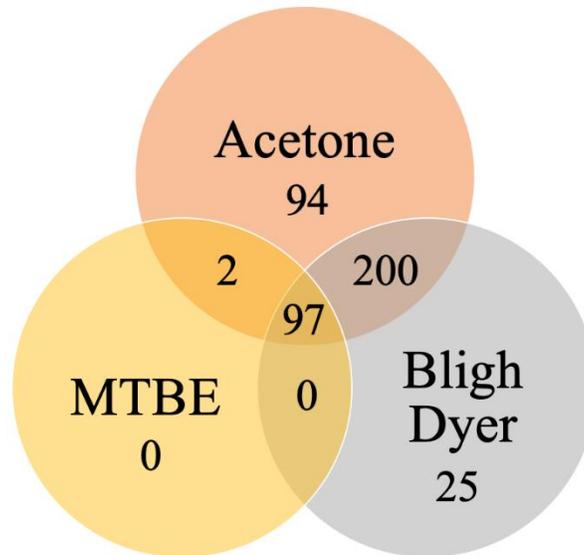


Figure 4-3. Venn diagram indicating how many proteins of *M. smegmatis* were present in each of the three extraction methods.

Table 4-1. PANTHER classifications of significant proteins in each extraction method according to the GO term cellular components.

Level 1	Level 2	Level 3	Acetone Specific		Bligh Dyer Specific		MTBE Specific	
			Sequences	Percentage	Sequences	Percentage	Sequences	Percentage
cellular anatomical entity (GO:0110165)			39	45.35	27	44.26	8	47.06
	cytoplasm (GO:0005737)		28	34.15	25	42.37	7	35
		plastid (GO:0009536)	1	3.45	2	11.76	1	12.5
		mitochondrion (GO:0005739)	11	37.93	4	23.53	2	25
		microbody (GO:0042579)	1	3.45	0	0	0	0
		cytosol (GO:0005829)	16	55.17	11	64.71	5	62.5
	cell periphery (GO:0071944)		3	3.66	2	3.39	1	5
		plasma membrane (GO:0005886)	3	100	2	100	1	100
	membrane (GO:0016020)		5	6.10	3	5.08	1	5
		membrane protein complex (GO:0098796)	1	16.67	1	20	0	0
		plasma membrane (GO:0005886)	3	50	2	40	1	100
		intrinsic component of membrane (GO:0031224)	2	33.33	0	0	0	0
		photosynthetic membrane (GO:0034357)	0	0	1	20	0	0
		organelle membrane (GO:0031090)	0	0	1	20	0	0
	periplasmic space (GO:0042597)		2	2.44	1	1.69	1	5
		outer membrane-bounded periplasmic space (GO:0030288)	2	100	1	100	1	100
	membrane-enclosed lumen (GO:0031974)		4	4.88	2	3.39	1	5
		organelle lumen (GO:0043233)	4	100	2	100	1	100
	intrinsic component of membrane (GO:0031224)		2	2.44	0	0	0	0
		integral component of membrane (GO:0016021)	2	66.67	0	0	0	0
		intrinsic component of plasma membrane (GO:0031226)	1	33.33	0	0	0	0
	envelope (GO:0031975)		2	2.44	2	3.39	1	5
		cell envelope (GO:0030313)	2	100	1	50	1	100
		organelle envelope (GO:0031967)	0	0	1	50	0	0
	chloroplast stroma (GO:0009570)		1	1.22	0	0	0	0
	organelle (GO:0043226)		19	23.17	11	18.64	3	15
		organellar ribosome (GO:0000313)	3	6.12	0	0	0	0
		organelle lumen (GO:0043233)	4	8.16	2	8.33	1	14.29
		intracellular organelle (GO:0043229)	19	38.78	11	45.83	3	42.86
		non-membrane-bounded organelle (GO:0043228)	11	22.45	3	12.5	1	14.29
		membrane-bounded organelle (GO:0043227)	12	24.49	8	33.33	2	28.57
	cytosol (GO:0005829)		16	19.51	11	18.64	5	25
		cytosolic ribosome (GO:0022626)	2	100	3	100	1	100
	organelle subcompartment (GO:0031984)		0	0	1	1.69	0	0
		plastid thylakoid (GO:0031976)	0	0	1	100	0	0
	thylakoid (GO:0009579)		0	0	1	1.69	0	0
		photosynthetic membrane (GO:0034357)	0	0	1	50	0	0
		plastid thylakoid (GO:0031976)	0	0	1	50	0	0

protein-containing complex (GO:0032991)		12	13.95	8	13.11	2	11.76
mitochondrial protein complex (GO:0098798)		3	21.43	2	22.22	0	0
mitochondrial large ribosomal subunit (GO:0005762)		3	100	0	0	0	0
inner mitochondrial membrane protein complex (GO:0098800)		0	0	1	100	0	0
ribonucleoprotein complex (GO:1990904)		10	71.43	3	33.33	1	100
ribosomal subunit (GO:0044391)		10	100	3	100	1	100
membrane protein complex (GO:0098796)		1	7.14	1	11.11	0	0
proton-transporting two-sector ATPase complex (GO:0016469)		1	50	1	50	0	0
proton-transporting two-sector ATPase complex, proton-transporting domain (GO:0033177)		1	50	0	0	0	0
inner mitochondrial membrane protein complex (GO:0098800)		0	0	1	50	0	0
catalytic complex (GO:1902494)		0	0	3	33.33	0	0
peptidase complex (GO:1905368)		0	0	1	50	0	0
transferase complex (GO:1990234)		0	0	1	50	0	0
intracellular (GO:0005622)		35	40.70	26	42.62	7	41.18
cytoplasm (GO:0005737)		28	58.33	25	64.10	7	70
plastid (GO:0009536)		1	3.45	2	11.76	1	12.5
mitochondrion (GO:0005739)		11	37.93	4	23.53	2	25
microbody (GO:0042579)		1	3.45	0	0	0	0
cytosol (GO:0005829)		16	55.17	11	64.71	5	62.5
proton-transporting ATP synthase complex (GO:0045259)		1	2.08	1	2.56	0	0
proton-transporting ATP synthase complex, coupling factor F(o) (GO:0045263)		1	100	0	0	0	0
mitochondrial proton-transporting ATP synthase complex (GO:0005753)		0	0	1	100	0	0
intracellular organelle (GO:0043229)		19	39.58	11	28.21	3	30
intracellular organelle lumen (GO:0070013)		4	14.81	2	13.33	1	25
intracellular non-membrane-bounded organelle (GO:0043232)		11	40.74	3	20	1	25
intracellular membrane-bounded organelle (GO:0043231)		12	44.44	8	53.33	2	50
organelle envelope (GO:0031967)		0	0	1	6.67	0	0
organelle subcompartment (GO:0031984)		0	0	1	6.67	0	0
proteasome complex (GO:0000502)		0	0	1	2.56	0	0
thylakoid (GO:0009579)		0	0	1	2.56	0	0
photosynthetic membrane (GO:0034357)		0	0	1	50	0	0
plastid thylakoid (GO:0031976)		0	0	1	50	0	0

4.3 Testing Lipid Extraction Methods

Table 4-2. List of fold change significant lipids for *M. smegmatis* in each lipid extraction method categorized by lipid class by MS-DIAL nomenclature standards. A fold change greater than 1.5 was used to determine statistical significance. The term upregulated denotes which test the lipid group was concentrated in and was derived from MetaboAnalyst.

Lipid Class	Acetone vs MTBE			Acetone vs Bligh Dyer			Bligh Dyer vs MTBE		
	Total	Acetone	MTBE	Total	Acetone	Bligh Dyer	Total	Bligh Dyer	MTBE
		Upregulated	Upregulated		Upregulated	Upregulated		Upregulated	
FA01: Fatty acids and Conjugates	8	2	6	8	5	3	10	8	2
FA07: Fatty esters	2	0	2	2	0	2	2	2	0
FA08: Fatty amides	8	5	3	13	4	9	12	7	5
GL00: Other Glycerolipids	17	10	7	17	9	8	19	10	9
GL01: Monoradylglycerols	1	0	1	2	0	2	2	2	0
GL02: Diradylglycerols	11	9	2	19	10	9	18	8	10
GL03: Triradylglycerols	20	19	1	31	14	17	31	17	14
GL05: Glycosyldiradylglycerols	5	3	2	7	3	4	7	4	3
GP00: Other Glycerophospholipids	0	0	0	1	1	0	1	0	1
GP01: Glycerophosphocholines	4	3	1	12	8	4	13	5	8
GP02: Glycerophosphoethanolamines	27	24	3	48	1	47	49	47	2
GP04: Glycerophosphoglycerols	17	12	5	23	12	11	24	12	12
GP06: Glycerophosphoinositols	8	5	3	11	9	2	9	1	8
GP10: Glycerophosphates	0	0	0	1	1	0	1	0	1
GP12: Glycerophosphoglycerophosphoglycerols	2	0	2	2	2	0	2	0	2
PR01: Isoprenoids	0	0	0	2	2	0	2	0	2
PR02: Quinones and hydroquinones	1	1	0	1	0	1	1	1	0
SP01: Sphingoid bases	16	15	1	21	9	12	21	10	11
SP02: Ceramides	13	7	6	21	12	9	19	9	10
SP03: Phosphosphingolipids	10	6	4	17	15	2	17	1	16
SP05: Neutral glycosphingolipids	7	5	2	8	5	3	7	2	5
SP06: Acidic glycosphingolipids	5	2	3	9	8	1	9	1	8
ST00: Other Sterol lipids	5	4	1	7	6	1	7	1	6

4.4 Examining Lipids from Phage-Treated Samples

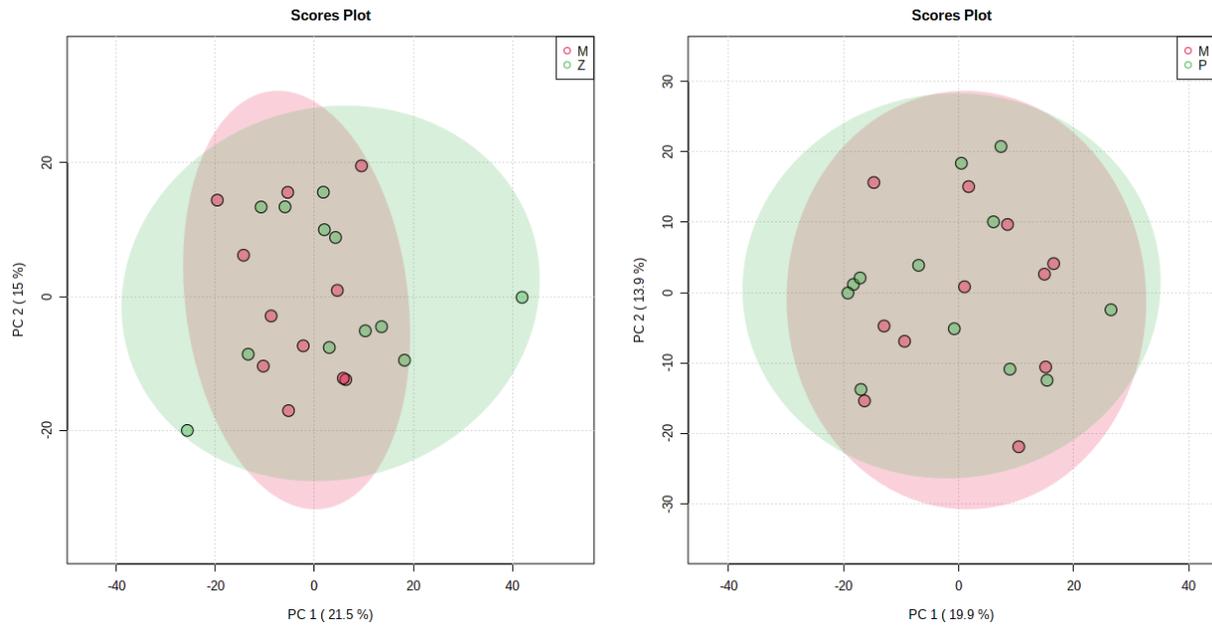


Figure 4-4. Principal component analysis plots comparing the phage-treated samples (mycobacteriophage Zalkecks, Z, on the left and mycobacteriophage PotatoSplit, P, on the right) to the control sample of *M. smegmatis* host. M represents the dataset of *M. smegmatis* host. Each plot represents the first principal component on the x-axis and the second principal component on the y-axis each with their corresponding percentage of variance.

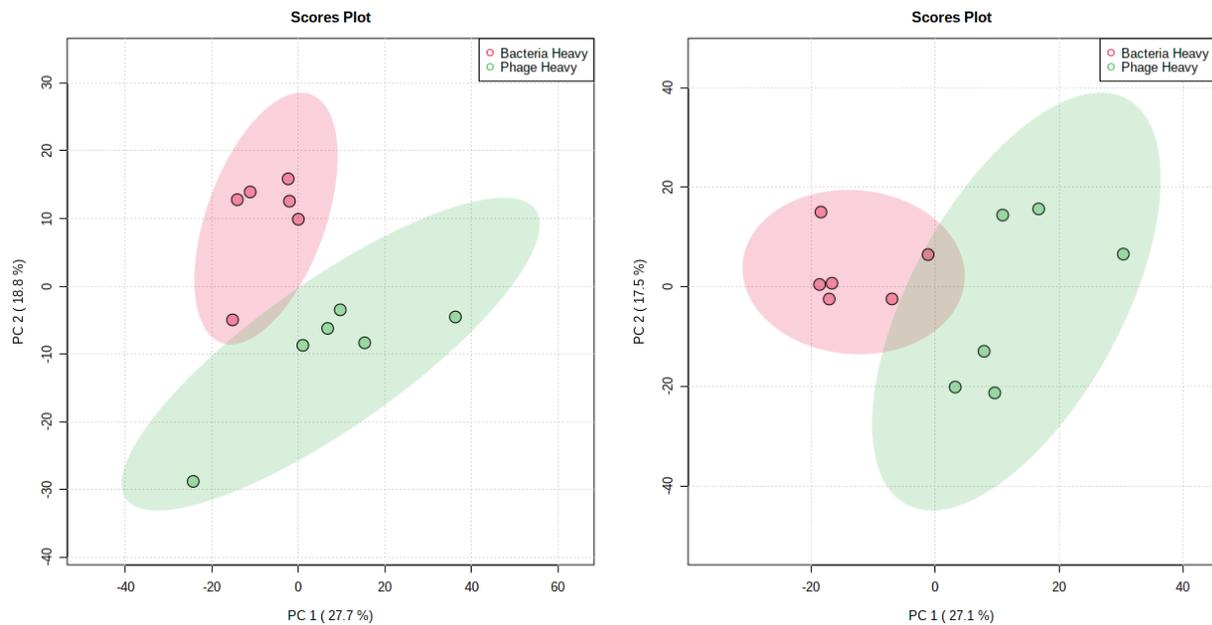


Figure 4-5. Principal component analysis plots comparing different time points of the phage-treated samples (mycobacteriophage Zalkecks on the left and mycobacteriophage PotatoSplit on the right). Bacteria heavy indicates the time points 0 and 4 hours which closely followed the OD600 of the control, *M. smegmatis*, while phage heavy indicates the time points 12 and 24 hours in which the presence of phage had a significant impact on OD600 compared to the control. Each plot represents the first principal component on the x-axis and the second principal component on the y-axis each with their corresponding percentage of variance

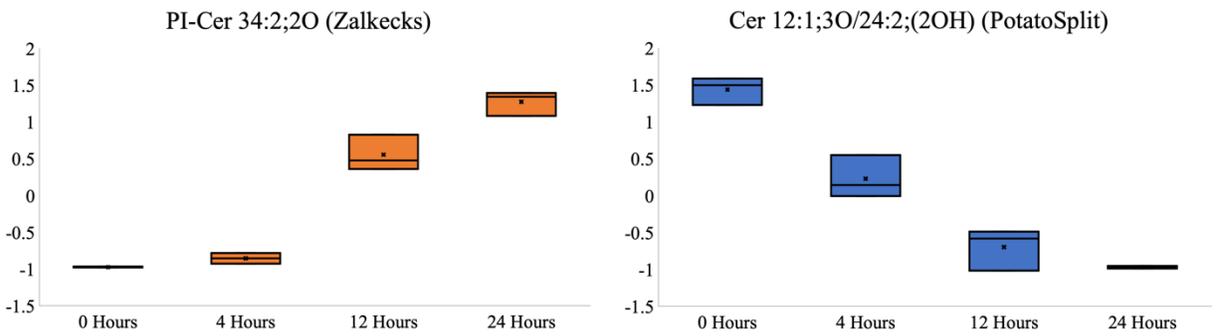


Figure 4-7. Boxplots of a significant lipid in the Zalkecks treated samples (left) and PotatoSplit treated samples (right).

Table 4-3. List of significant lipids in mycobacteriophage Zalkecks treated samples categorized by their abundance over time. To determine significance, ANOVA with an adjusted p-value (FDR) cutoff of 0.05 was used and Fisher's LSD was used for post-hoc analysis. The list of each pattern over time was cross-referenced with the control samples of just *M. smegmatis* over time and any lipids found in both groups were removed.

Zalkecks Lipids			
<i>Increases at Each Time Point</i>	<i>Decreases at Each Time Point</i>	<i>Increases in 0 and 4 hours</i>	<i>Increases in 12 and 24 hours</i>
RIKEN P-VS1 ID-10248	RIKEN P-VS1 ID-5826	RIKEN P-VS1 ID-5826	RIKEN P-VS1 ID-10784
RIKEN P-VS1 ID-7426	PG 34:2 PG 16:1_18:1	RIKEN P-VS1 ID-8424	RIKEN P-VS1 ID-10248
RIKEN P-VS1 ID-8167	TG 15:1_15:1_34:9	PG 34:2 PG 16:1_18:1	RIKEN P-VS1 ID-7426
RIKEN P-VS1 ID-8905		TG 15:1_15:1_34:9	RIKEN P-VS1 ID-8167
RIKEN P-VS1 ID-9219		Cer 16:0;2O/16:0	RIKEN P-VS1 ID-8905
RIKEN P-VS1 ID-9717		Cer 34:0;4O	RIKEN P-VS1 ID-9219
CerP 20:1;2O/28:6		DG 40:6	RIKEN P-VS1 ID-9717
DGGA 11:0_17:0		DG 64:17	PG 30:1 PG 14:0_16:1
DGGA 12:0_18:1		PE 32:1 PE 16:0_16:1	PG 35:2 PG 16:1_19:1
DGGA 17:0_14:1		PE 32:2 PE 16:1_16:1	DG 32:1
PE 14:0_24:5		PE 34:2 PE 16:1_18:1	MGDG O-15:4_17:4
PE 31:1 PE 14:0_17:1		PE 36:2 PE 18:1_18:1	PI 54:8
PE 33:1		PE 38:0	PI-Cer 35:2;2O
PE 33:1 PE 16:0_17:1		PI 55:8	SL 13:2;O/30:6;O
PE 36:1		PI 55:9	CerP 20:1;2O/28:6
PE 36:2 PE 17:1_19:1			DGGA 11:0_17:0
PG 36:2			DGGA 12:0_18:1
PI-Cer 33:2;2O			DGGA 17:0_14:1
PI-Cer 35:3;2O			PE 14:0_24:5
SL 12:1;O/28:5;O			PE 31:1 PE 14:0_17:1
SL 12:1;O/30:6;O			PE 33:1
SL 13:1;O/28:6;O			PE 33:1 PE 16:0_17:1
SL 15:3;O/30:6;O			PE 36:1
SL 16:3;O/30:6;O			PE 36:2 PE 17:1_19:1
SM 12:1;2O/28:6			PG 36:2
SM 39:8;3O			PI-Cer 33:2;2O
TG 15:4_15:4_15:4			PI-Cer 35:3;2O
TG 8:0_14:1_36:10			SL 12:1;O/28:5;O
			SL 12:1;O/30:6;O
			SL 13:1;O/28:6;O
			SL 15:3;O/30:6;O
			SL 16:3;O/30:6;O
			SM 12:1;2O/28:6
			SM 39:8;3O
			TG 15:4_15:4_15:4
			TG 8:0_14:1_36:10

Table 4-4. List of significant lipids in mycobacteriophage PotatoSplit treated samples categorized by their abundance over time. To determine significance, ANOVA with an adjusted p-value (FDR) cutoff of 0.05 was used and Fisher's LSD was used for post-hoc analysis. The list of each pattern over time was cross-referenced with the control samples of just *M. smegmatis* over time and any lipids found in both groups were removed.

PotatoSplit Lipids			
<i>Increases at Each Time Point</i>	<i>Decreases at Each Time Point</i>	<i>Increases in 0 and 4 hours</i>	<i>Increases in 12 and 24 hours</i>
RIKEN P-VS1 ID-8167 PE 14:0_24:5 TG 8:0_15:4_18:5	RIKEN P-VS1 ID-5423 RIKEN P-VS1 ID-5453 RIKEN P-VS1 ID-5826 RIKEN P-VS1 ID-5903 RIKEN P-VS1 ID-8424 RIKEN P-VS1 ID-9165 BMP 9:0_26:4 CAR 27:1 DG 42:6 DGTS 18:5_18:5 Hex2Cer 16:1;2O/16:1 HexCer 17:1;3O/20:4;(2OH) MGDG O-15:4_7:0 PC 19:3_19:3 PE 32:1 PE 34:2 PE 16:1_18:1 PE 34:3 PE 36:2 PE 18:1_18:1 PE 36:3 PE 18:1_18:2 PE P-37:2 PE P-16:1_21:1 PG 32:0 PG 16:0_16:0 PG 34:2 PG 16:1_18:1 PS 33:1	RIKEN P-VS1 ID-5453 RIKEN P-VS1 ID-5826 RIKEN P-VS1 ID-5423 RIKEN P-VS1 ID-5903 RIKEN P-VS1 ID-8424 RIKEN P-VS1 ID-9165 BMP 9:0_26:4 CAR 27:1 PG 32:0 PG 16:0_16:0 DG 42:6 DGTS 18:5_18:5 Hex2Cer 16:1;2O/16:1 HexCer 17:1;3O/20:4;(2OH) MGDG O-15:4_7:0 PC 19:3_19:3 PE 32:1 PE 34:2 PE 16:1_18:1 PE 34:3 PE 36:2 PE 18:1_18:1 PE 36:3 PE 18:1_18:2 PE P-37:2 PE P-16:1_21:1 PG 34:2 PG 16:1_18:1 PS 33:1 DG 40:6 PE 32:1 PE 16:0_16:1 PE 34:0 PE 16:0_18:0	RIKEN P-VS1 ID-8167 RIKEN P-VS1 ID-10784 PI-Cer 35:2;2O TG 8:0_15:4_18:5 PE 14:0_24:5 PG 30:1 PG 14:0_16:1 PI-Cer 33:2;2O PI-Cer 33:2;2O

Table 4-5. List of significant lipids categorized by lipid class by MS-DIAL nomenclature standards over a range of statistical tests and indicating which are upregulated in each group when applicable. Significance was determined by a fold change greater than 1.5 or in the ANOVA tests, an adjusted p-value (FDR) cutoff of 0.05 was used and Fisher's LSD was used for post-hoc analysis. M indicates the control samples only containing *M. smegmatis*, P indicates the mycobacteriophage PotatoSplit treated samples, and Z indicates the mycobacteriophage Zalkecks treated samples.

Statsitcal Test	FA07	FA08	GL00	GL01	GL02	GL03	GL05	GP01	GP02	GP04	GP06	PR01	PR02	SP01	SP02	SP03	SP05	SP06	ST00
MvP all hours	0	5	1	0	2	3	1	0	10	1	1	1	0	2	1	7	2	3	0
M upregulated	0	4	0	0	0	2	0	0	0	0	1	1	0	1	0	3	0	1	0
P upregulated	0	1	1	0	2	1	1	0	10	1	0	0	0	1	1	4	2	2	0
MvP 0-4 hours	0	2	1	0	4	6	0	0	9	1	1	1	0	0	2	5	1	2	1
M upregulated	0	1	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0
P upregulated	0	1	1	0	3	4	0	0	9	1	1	1	0	0	2	5	1	2	1
MvP 12-24 hours	0	5	3	0	3	2	1	4	9	3	3	1	0	5	5	5	3	3	0
M upregulated	0	4	2	0	1	1	0	4	0	2	2	1	0	3	1	3	2	1	0
P upregulated	0	1	1	0	2	1	1	0	9	1	1	0	0	2	4	2	1	2	0
P Hours ANOVA	1	0	1	0	6	1	1	1	10	7	2	0	0	0	1	3	2	0	0
P 0-4 hours v 12-24 hours	1	3	7	0	12	9	3	7	35	12	9	0	0	15	11	13	4	8	0
0-4 upregulated	1	1	2	0	7	4	1	7	21	7	5	0	0	6	6	6	3	5	0
12-24 upregulated	0	2	5	0	5	5	2	0	14	5	4	0	0	9	5	7	1	3	0
MvZ all hours	1	4	7	1	3	4	3	1	8	2	3	1	1	3	3	4	2	2	0
M upregulated	0	3	0	0	2	2	0	1	3	0	3	1	1	0	0	1	0	0	0
Z upregulated	1	1	7	1	1	2	3	0	5	2	0	0	0	3	3	3	2	2	0
MvZ 0-4 hrs	0	0	0	0	1	6	1	4	3	0	1	0	1	0	0	1	0	0	0
M upregulated	0	0	0	0	1	3	0	4	2	0	1	0	1	0	0	1	0	0	0
Z upregulated	0	0	0	0	0	3	1	0	1	0	0	0	0	0	0	0	0	0	0
MvZ 12-24hrs	1	5	9	1	7	3	3	1	10	5	6	1	1	4	5	5	2	2	0
M upregulated	0	4	1	0	4	1	0	1	5	0	6	1	1	0	1	1	0	0	0
Z upregulated	1	1	8	1	3	2	3	0	5	5	0	0	0	4	4	4	2	2	0
Z Hours ANOVA	0	1	3	0	8	6	2	0	18	8	5	0	0	8	4	6	0	0	0
Z 0-4 hours v 12-24 hours	1	1	15	1	13	8	4	2	37	9	9	0	0	11	11	12	5	5	0
0-4 upregulated	0	0	3	0	7	3	1	2	21	3	5	0	0	1	6	4	3	2	0
12-24 upregulated	1	1	12	1	6	5	3	0	16	6	4	0	0	10	5	8	2	3	0

4.5 Proteomic Data Analysis FrenchFry Case Study

Table 4-6. List of FrenchFry peptides found in mass spectrometry data along with their corresponding gene number, annotated function, and the frequency of how often they were counted in the data. The protein ID indicates the original label from the MaxQuant search.

Protein ID	Gene Number	Annotated Function	Peptide Frequency
FrenchFry	1	helix-turn-helix DNA binding domain protein	24
FrenchFry	2	queuine-tRNA ribosyltransferase	26
FrenchFry	3	QueC-like queosine biosynthesis protein	4
FrenchFry	4	QueD-like queosine biosynthesis protein	6
FrenchFry	5	QueE-like queosine biosynthesis protein	11
FrenchFry	6	GTP cyclohydrolase I	7
FrenchFry	7	terminase	16
FrenchFry	14	hypothetical protein	2
FrenchFry	15	major capsid protein	29
FrenchFry	17	hypothetical protein	11
FrenchFry	19	hypothetical protein	2
FrenchFry	20	hypothetical protein	5
FrenchFry	21	major tail protein	10
FrenchFry	24	head-to-tail adaptor	6
FrenchFry	25	hypothetical protein	3
FrenchFry	26	hypothetical protein	8
FrenchFry	27	tail assembly chaperone	5
FrenchFry	28	hypothetical protein (tail assembly chaperone)	8
FrenchFry	29	tape measure protein	59
FrenchFry	30	major tail protein	13
FrenchFry	31	minor tail protein	13
FrenchFry	32	minor tail protein	17
FrenchFry	33	minor tail protein	6
FrenchFry	34	hypothetical protein	4
FrenchFry	35	hypothetical protein	2
FrenchFry	36	hypothetical protein	5
FrenchFry	37	hypothetical protein	1
FrenchFry	38	hypothetical protein	9
FrenchFry	39	hypothetical protein (minor tail protein)	17
FrenchFry	40	hypothetical protein	4
FrenchFry	42	hypothetical protein	5
FrenchFry	43	hypothetical protein	12

Table 4-6 continued

Mycobacterium:False	44	helix-turn-helix DNA binding domain protein	5
Mycobacterium:False	45	helix-turn-helix DNA binding domain protein	6
Mycobacterium:False	46	hypothetical protein	5
FrenchFry	47	lysine A	23
FrenchFry	48	holin	2
Mycobacterium:False	49	hypothetical protein	10
Mycobacterium:False	50	hypothetical protein	11
Mycobacterium:False	51	DNA helicase	14
Mycobacterium:False	54	hypothetical protein	13
Mycobacterium:False	55	DNA primase/helicase	34
Mycobacterium:False	57	DNA polymerase I	20
Mycobacterium:False	60	hypothetical protein	7
Mycobacterium:False	62	hypothetical protein	3
FrenchFry	64	hypothetical protein	9
FrenchFry	66	hypothetical protein (helix-turn-helix DNA binding protein)	3
FrenchFry	67	DNA binding protein	8
FrenchFry	68	hypothetical protein	4
FrenchFry	69	hypothetical protein	3
FrenchFry	70	hypothetical protein (DNA binding protein)	4
Mycobacterium:False	76	hypothetical protein	4
Mycobacterium:False	77	hypothetical protein	3
Mycobacterium:False	78	hypothetical protein	2
Mycobacterium:False	83	hypothetical protein	10
Mycobacterium:False	85	hypothetical protein	8
Mycobacterium:False	86	hypothetical protein	8
Mycobacterium:False	90	hypothetical protein	5
Mycobacterium:False	91	hypothetical protein	3
Mycobacterium:False	93	hypothetical protein	9

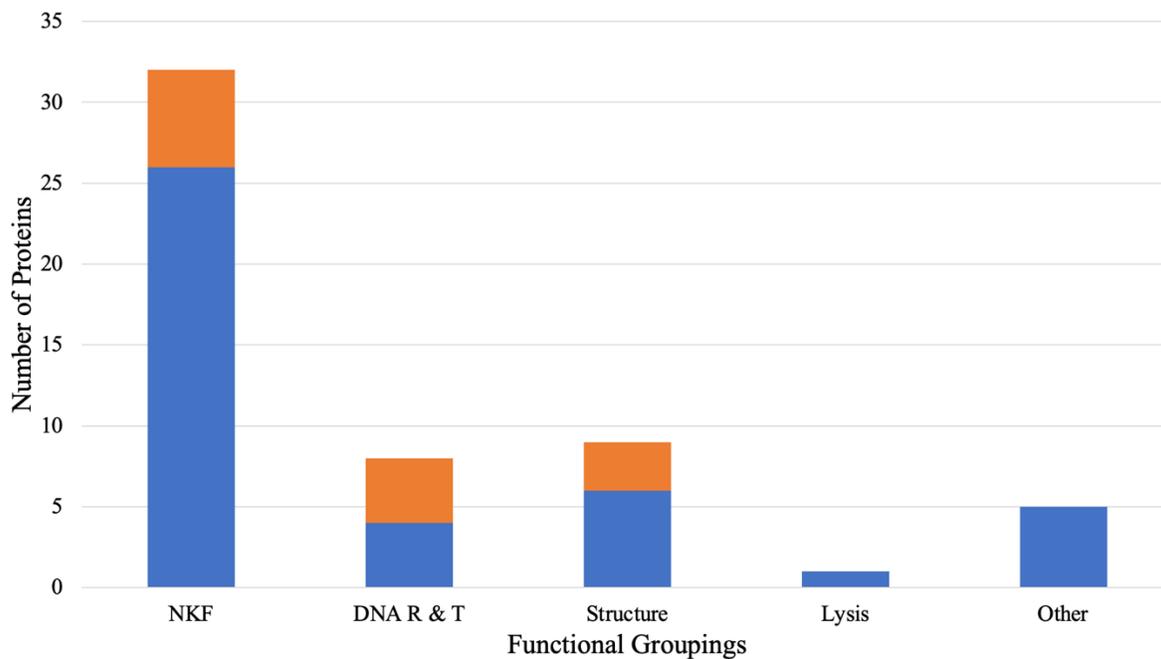


Figure 4-8. Number of significant FrenchFry proteins expressed in the statistical comparison of mycobacteriophage FrenchFry treated samples to the negative control of just *M. smegmatis* grouped by functional classifications. The FrenchFry proteins represented in blue are upregulated in the FrenchFry data while those represented in orange are downregulated. NKF indicates no known function, DNA R & T indicates those involved in DNA replication and translation.

Table 4-7. Number of fold change significant (fold change greater than 1.5) and t-test significant (p-value lower than 0.05) proteins in each statistical test. E indicates the proteins present in the samples infected at the exponential phase of bacterial growth while S indicates the proteins present in the samples infected at the stationary phase of bacterial growth. Hour 4 indicates the samples taken after four hours of inoculation with phage FrenchFry and hour 10 indicates the samples taken after ten hours of inoculation with phage FrenchFry. FF indicates the FrenchFry-treated samples while NC indicates the negative control of just *M. smegmatis*.

Test	Number of Fold Change Significant Proteins	Number of T-test Significant Proteins
E vs S	1429	909
E vs S (hour 4)	1358	1035
E vs S (hour 10)	1415	779
4 vs 10 (exponential)	1244	61
4 vs 10 (stationary)	567	1
FF vs NC	317	0

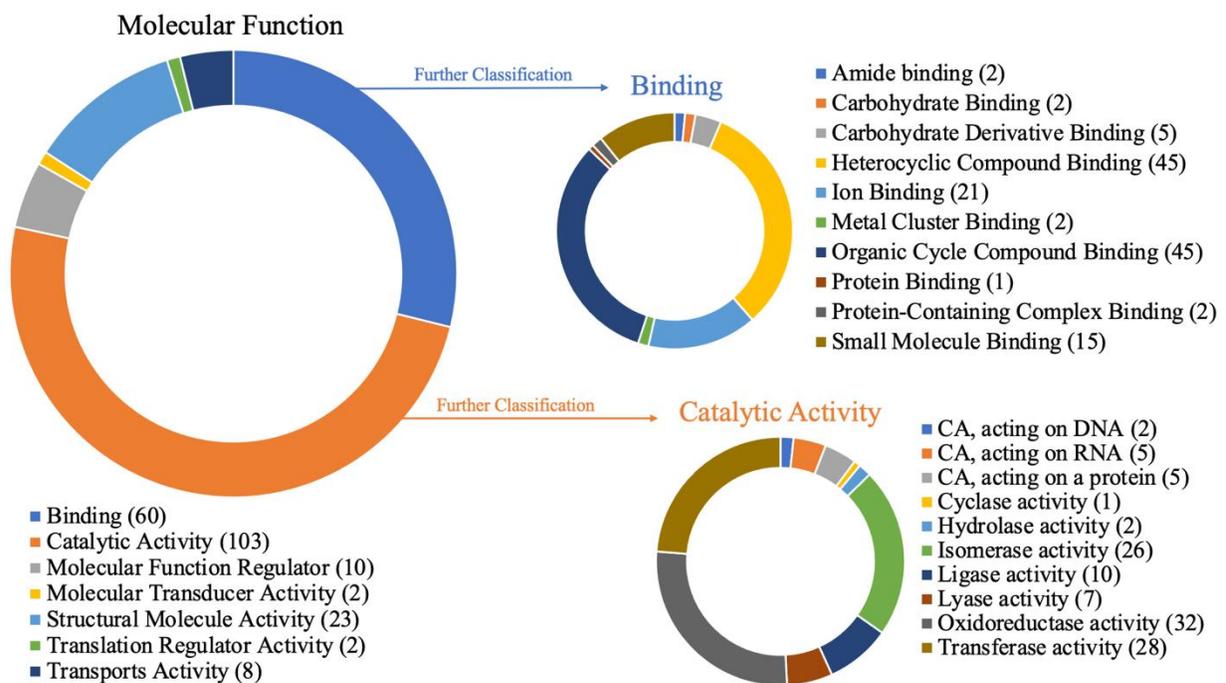


Figure 4-9. PANTHER pathway analysis of statistically significant proteins from the comparison of the exponential and stationary phases of inoculation based on the Gene Ontology terms corresponding to molecular function. The largest categories of binding and catalytic activity were broken into subcategories. Each category and subcategory listed contains the number of sequences applicable to the category.

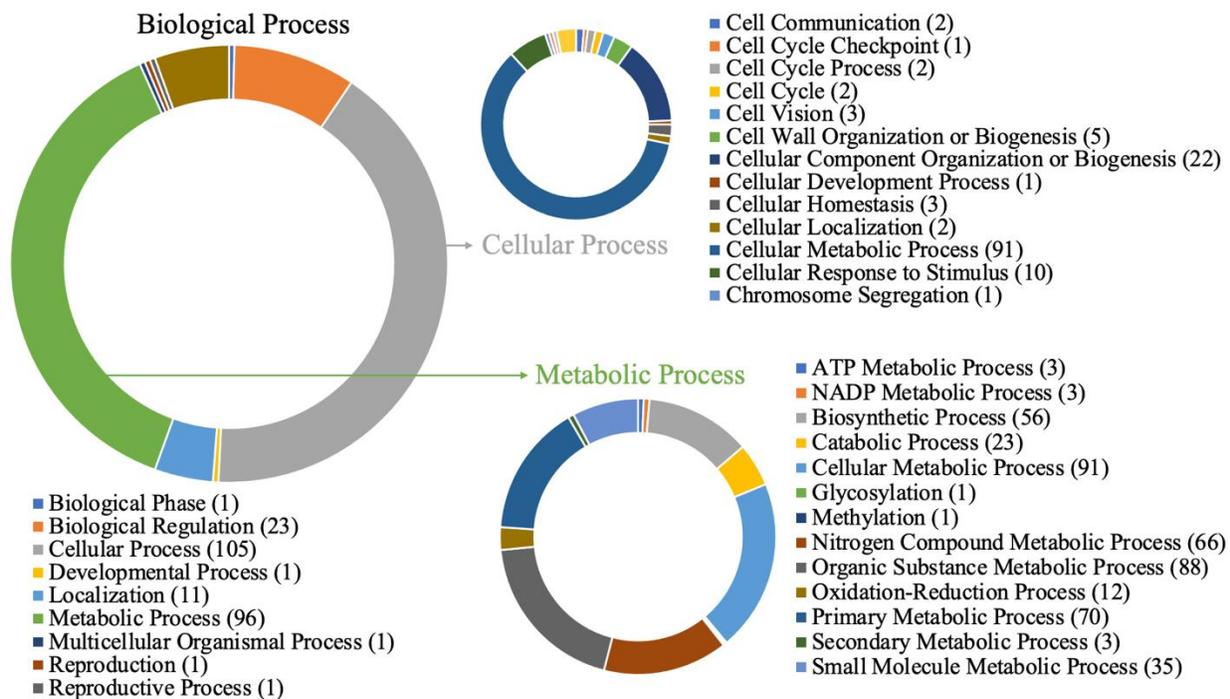


Figure 4-10. PANTHER pathway analysis of statistically significant proteins from the comparison of the exponential and stationary phases of inoculation based on the Gene Ontology terms corresponding to biological process. The largest categories of cellular process and metabolic process were broken into subcategories. Each category and subcategory listed contains the number of sequences contained within it.

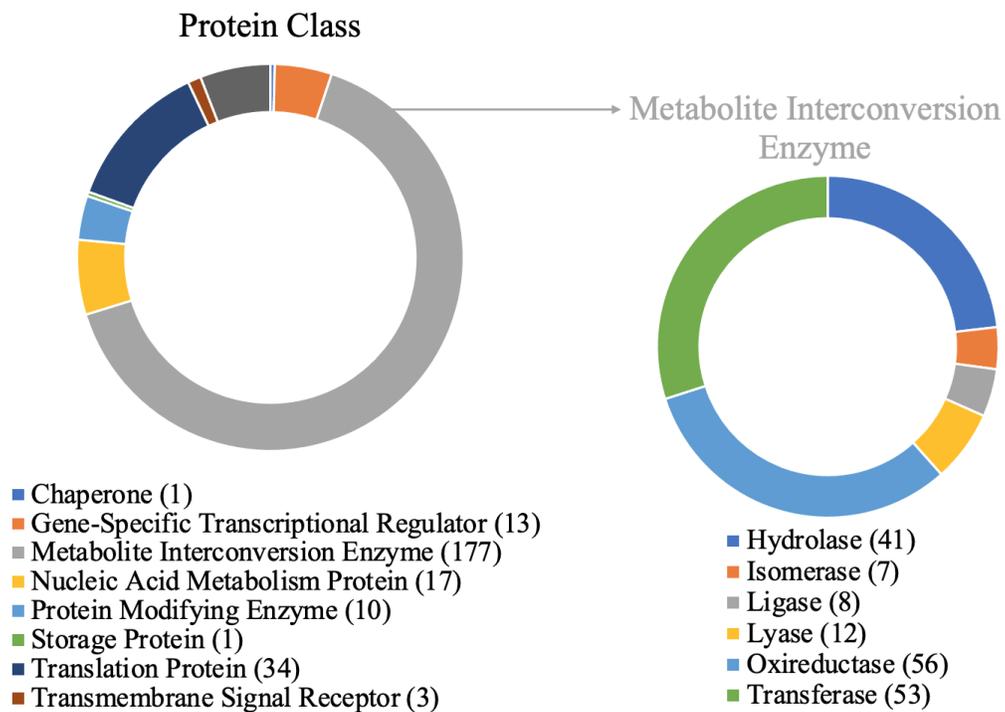


Figure 4-11. PANTHER pathway analysis of statistically significant proteins from the comparison of the exponential and stationary phases of inoculation based on the Gene Ontology terms corresponding to protein class. The largest category of metabolite interconversion enzyme was broken into subcategories. Each category and subcategory listed contains the number of sequences contained within it.

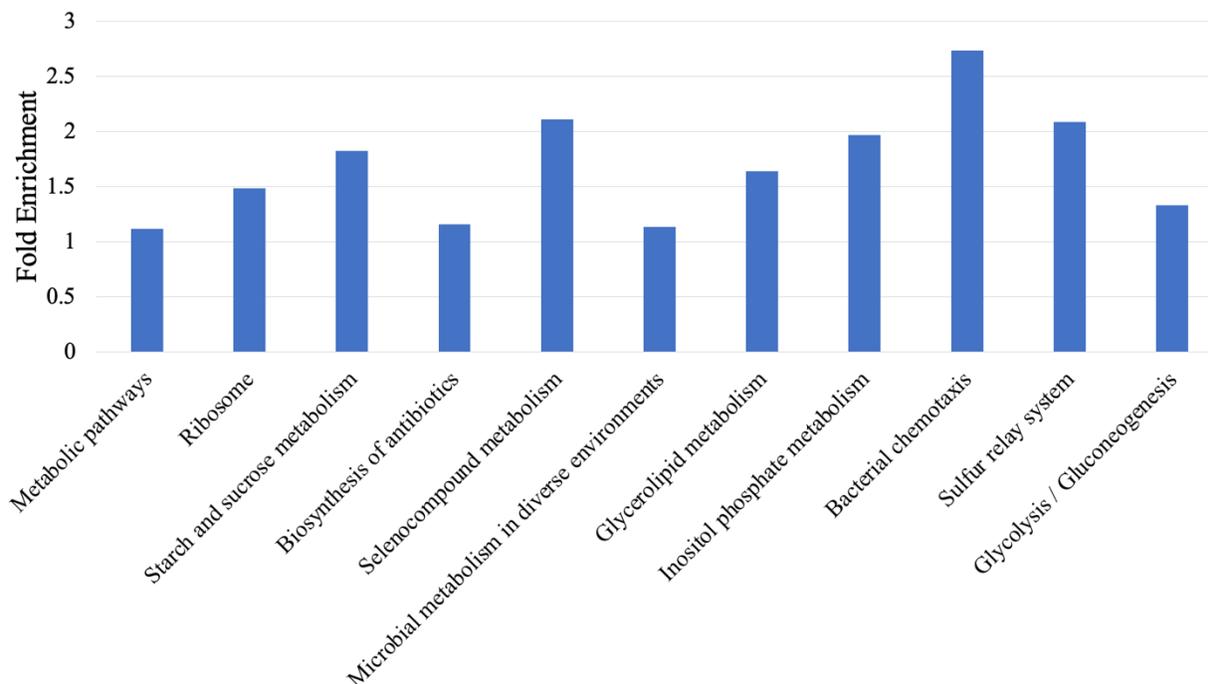


Figure 4-12. The fold enrichment of the DAVID pathway analysis functional annotation results based off the KEGG pathway terms of the statistically significant proteins determined by MetaboAnalyst in the comparison of the exponential and stationary phases of inoculation.

4.6 Archived Phage Samples

Table 4-8. List of all mycobacteriophages that were either re-archived for Purdue University or sent to the HHMI Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) program at the University of Pittsburgh for long-term storage. Phages in this list that were not sent into HHMI had already been previously archived with the program.

Phage	Year Isolated	Re-archived at Purdue	Sent for HHMI Archival
AlpineSix	2019	Yes	Already Archived
Apartment5	2019	Yes	Already Archived
Astoria	2019	Yes	Already Archived
Cloudy	2018	Yes	Yes
CoffeeAlways	2019	Yes	Already Archived
ColdWork	2019	Yes	Already Archived
Devastated	2018	Yes	Yes
Elsie	2018	Yes	Already Archived
Fliddinger	2019	Yes	Already Archived
Gem1	2018	Yes	Yes
Gillian	2018	Yes	Yes

Table 4-8 continued

GTdaves	2019	Yes	Already Archived
Gylpye	2018	Yes	Yes
HapyZ	2019	Yes	Already Archived
HonkyTonkAngel	2019	Yes	Already Archived
Izajani	2018	Yes	Yes
JakeTheDog	2018	Yes	Yes
JernigCza	2018	Yes	Already Archived
JuliusCaesar	2019	Yes	Already Archived
Krili	2018	Yes	Already Archived
Levi	2019	Yes	Already Archived
Lizzokovich	2018	Yes	Already Archived
Maru	2019	Yes	Already Archived
Meemlordius	2019	Yes	Already Archived
MegsJr	2019	Yes	Already Archived
Murai	2019	Yes	Already Archived
NashYanlok	2018	Yes	Already Archived
NiebruSaylor	2019	Yes	Already Archived
NiQu	2018	Yes	Already Archived
Orca	2018	Yes	Yes
Oromis	2018	Yes	Already Archived
Possibility	2019	Yes	Already Archived
Poyo	2018	Yes	Already Archived
Prongs	2019	Yes	Yes
PVRamachandran	2019	Yes	Already Archived
Raid51	2019	Yes	Already Archived
Redacted	2019	Yes	Already Archived
Redwood	2019	Yes	Already Archived
RetimsGems	2018	Yes	Already Archived
RomeoNJuliet	2019	Yes	Yes
SilverDipper	2019	Yes	Already Archived
Squidward	2018	Yes	Yes
Sydolivia	2019	Yes	Yes
Tanming	2018	Yes	Yes
Thalatta	2018	Yes	Already Archived
TLMidnight	2019	Yes	Already Archived
TreeDirt	2019	Yes	Already Archived
WALC	2018	Yes	Yes
Wendell	2018	Yes	Yes
Yoosorrell	2018	Yes	Yes

4.7 Mycobacteriophage Cluster B1 Frameshift Investigation

4.7.1 Structural Comparisons

Table 4-9. Top 5 PDB results for tail assembly chaperone 1. RMSD is the root-mean-square deviation and is the measure of the average distance between the atoms of superimposed proteins. Identity is the percentage sequence identity in the structurally aligned region. Coverage is the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by the length of the query protein.

Rank	Classification	RMSD	Identity	Coverage
1	Metal Binding Protein	3.01	0.084	0.757
2	Unknown Function	3.01	0.067	0.736
3	Metal Binding Protein	2.9	0.07	0.707
4	Unknown Function	3.32	0.037	0.757
5	Transport Protein	3.32	0.058	0.729

Table 4-10. The consensus prediction of GO terms for tail assembly chaperone 1 among the top scoring templates. The GO-Score associated with each prediction is defined as the average weight of the GO term, where the weights are assigned based on Cscore^{GO} of the template.

Type	Function	GO Score
Molecular Function	electron transfer activity	0.08
	cytochrome-c oxidase activity	0.08
	heme binding	0.08
	NADH dehydrogenase (quinone) activity	0.07
	ferric iron binding	0.07
Biological Process	transport	0.08
	electron transport chain	0.08
	aerobic respiration	0.08
	cellular iron ion homeostasis	0.07
	protein-containing complex assembly	0.07
Cellular Component	mitochondrial inner membrane	0.08
	respirasome	0.08
	integral component of membrane	0.08
	plasma membrane	0.07
	cytosol	0.07

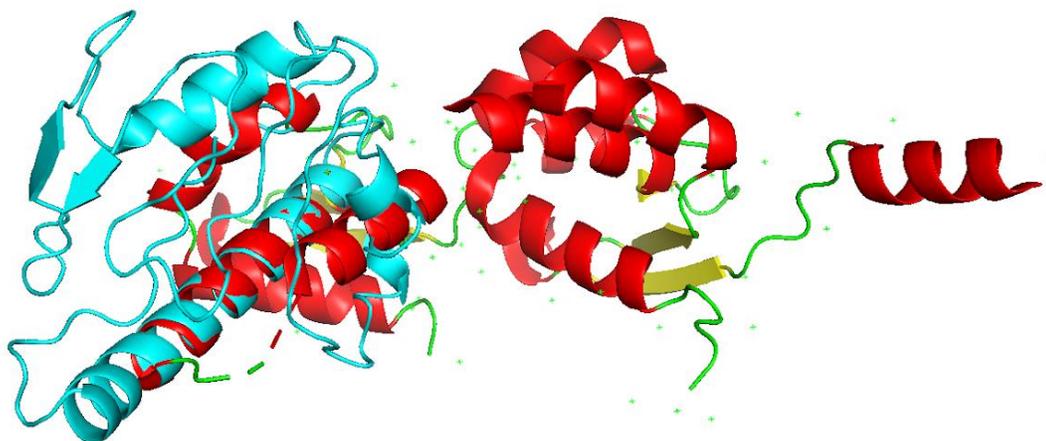


Figure 4-13. The secondary structure of tail assembly chaperone 1 superimposed on the secondary structure of bacteriophage HK97 tail assembly chaperone (PDB ID: 2OB9). The structure in cyan is the I-TASSER predicted structure while the rest is the HK97 tail assembly chaperone coloring for different secondary structures.

Table 4-11. Top 5 PDB results for tail assembly chaperone 2. RMSD is the root-mean-square deviation and is the measure of the average distance between the atoms of superimposed proteins. Identity is the percentage sequence identity in the structurally aligned region. Coverage is the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by the length of the query protein.

Rank	Classification	RMSD	Identity	Coverage
1	Membrane Protein	2.4	0.075	0.827
2	Lyase	2.46	0.121	0.827
3	Cold-Activity	2.75	0.106	0.867
4	Membrane Protein	3.38	0.097	0.92
5	Lyase	2.51	0.061	0.827

Table 4-12. The consensus prediction of GO terms for tail assembly chaperone 2 among the top scoring templates. The GO-Score associated with each prediction is defined as the average weight of the GO term, where the weights are assigned based on $Cscore^{GO}$ of the template.

Type	Function	Go Score
Molecular Function	signaling receptor activity	0.23
	chloride ion binding	0.13
	solute:proton antiporter activity	0.13
	voltage-gated chloride channel activity	0.13
	transferase activity, transferring acyl groups, acyl groups converted into alkyl on transfer	0.12
Biological Process	multi-organism process	0.46
	obsolete intracellular part	0.46
Cellular Component	integral component of membrane	0.33

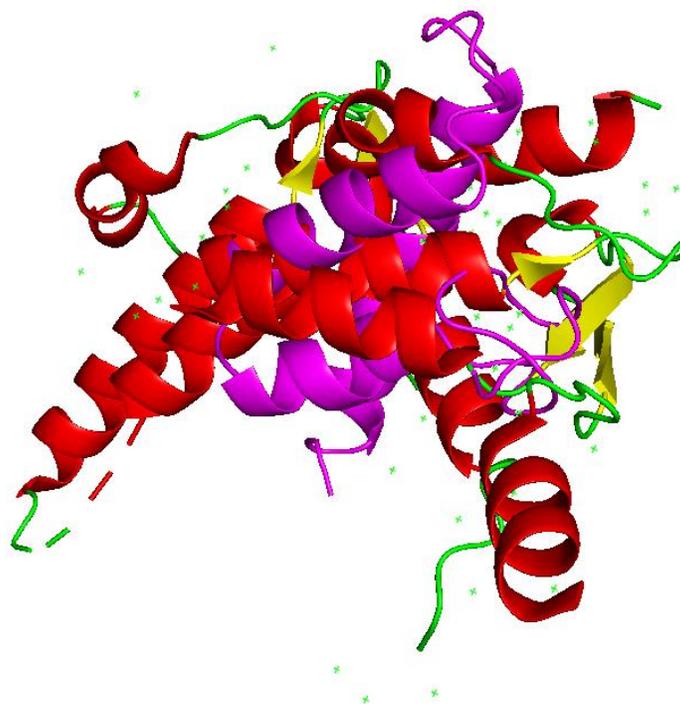


Figure 4-14. The secondary structure of tail assembly chaperone 2 superimposed on the secondary structure of bacteriophage HK97 tail assembly chaperone (PDB ID: 2OB9). The structure in magenta is the I-TASSER predicted structure while the rest is the HK97 tail assembly chaperone coloring for different secondary structures.

Table 4-13. Top 5 PDB results for the supposed frameshift tail assembly chaperone. RMSD is the root-mean-square deviation and is the measure of the average distance between the atoms of superimposed proteins. Identity is the percentage sequence identity in the structurally aligned region. Coverage is the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by the length of the query protein.

Rank	Classification	RMSD	Identity	Coverage
1	Transferase	2.45	0.086	0.907
2	Transferase	2.78	0.07	0.902
3	Transferase	3.21	0.101	0.803
4	Transferase	3.3	0.078	0.806
5	Transferase	3.69	0.072	0.78

Table 4-14. The consensus prediction of GO terms for supposed frameshift tail assembly chaperone among the top scoring templates. The GO-Score associated with each prediction is defined as the average weight of the GO term, where the weights are assigned based on Cscore^{GO} of the template.

Type	Function	Go Score
Molecular Function	adenyl ribonucleotide binding	0.35
	purine ribonucleoside triphosphate binding	0.35
	nucleotidyltransferase activity	0.31
Biological Process	glucose metabolic process	0.1
	immune response	0.09
	transcription, DNA-templated	0.09
	mRNA polyadenylation	0.09
	snoRNA polyadenylation	0.09
Cellular Component	intracellular membrane-bounded organelle	0.35

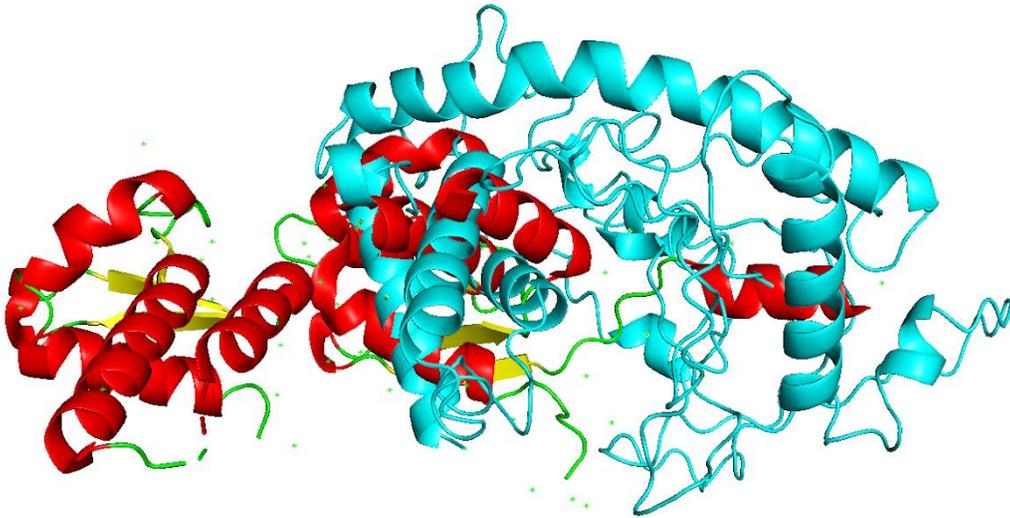


Figure 4-15. The secondary structure of the supposed annotated frameshift superimposed on the secondary structure of bacteriophage HK97 tail assembly chaperone (PDB ID: 2OB9). The structure in cyan is the I-TASSER predicted structure while the rest is the HK97 tail assembly chaperone coloring for different secondary structures.

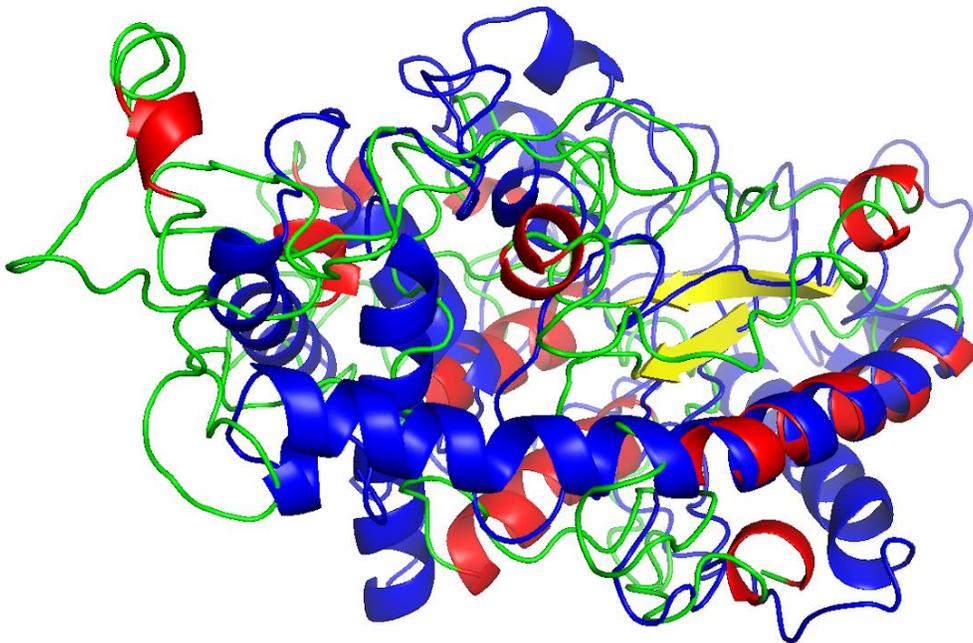


Figure 4-16. The secondary structure of the supposed annotated frameshift superimposed on the predicted structure of the annotated frameshift from *Gordonia terrae* phage CloverMinnie. The structure in blue is the I-TASSER predicted structure while the rest is the CloverMinnie frameshift, coloring for different types of secondary structures.

4.7.2 Molecular dynamics simulations

Simulation validation

The energy minimization step of each protein was evaluated by graphing potential energy against time. Figure 4-17. The potential energy curve for the energy minimization step of the simulation when running the Tail Assembly Chaperone 1 protein. The energy drops and reaches a plateau, meaning the energy has been properly minimized. shows the potential energy curve of the Tail Assembly Chaperone 1 protein. It reaches a plateau, meaning the energy has been properly minimized. This step was repeated for the other two proteins and those graphs can be seen in Figures 1 and 2 in the Appendix. Energy minimization only had to be done once for each structure, as the same minimized structure was used for the Autoclave and Cell simulations.

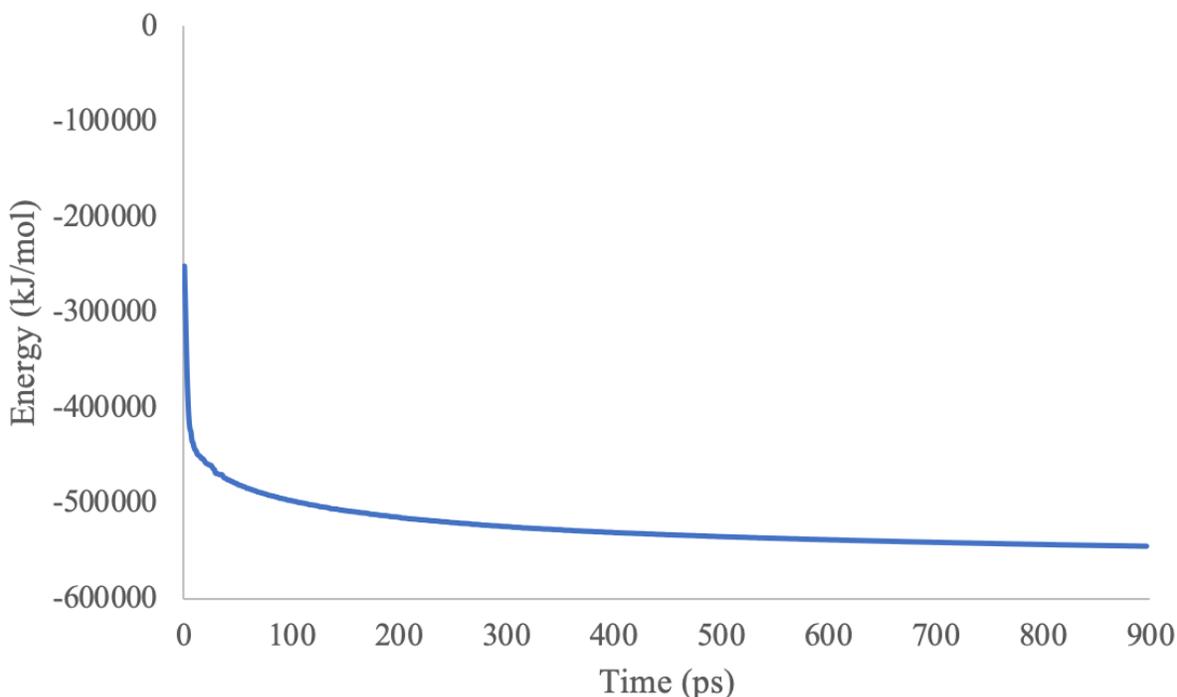


Figure 4-17. The potential energy curve for the energy minimization step of the simulation when running the Tail Assembly Chaperone 1 protein. The energy drops and reaches a plateau, meaning the energy has been properly minimized.

After the NVT step, the temperatures were plotted to ensure the system was stable on the correct reference temperature. Figure 4-18. The temperature curves for the Tail Assembly Chaperone 1 protein during the Autoclave simulation (blue) and the Cell simulation (orange). The

autoclave reference temperature is 394.15 K, while the cell reference temperature is 310.15 K. shows the temperature graphs for Autoclave and Cell temperatures for the Tail Assembly Chaperone 1 protein. The temperature fluctuates around the reference values, but the average temperatures were always within 10 K of the expected value. For the Autoclave, the temperature of the heat bath was set to 394 K. For the Cell, the temperature was set to 310 K. Figures 3 and 4 in the Appendix shows the temperature curves for the other two proteins simulated.

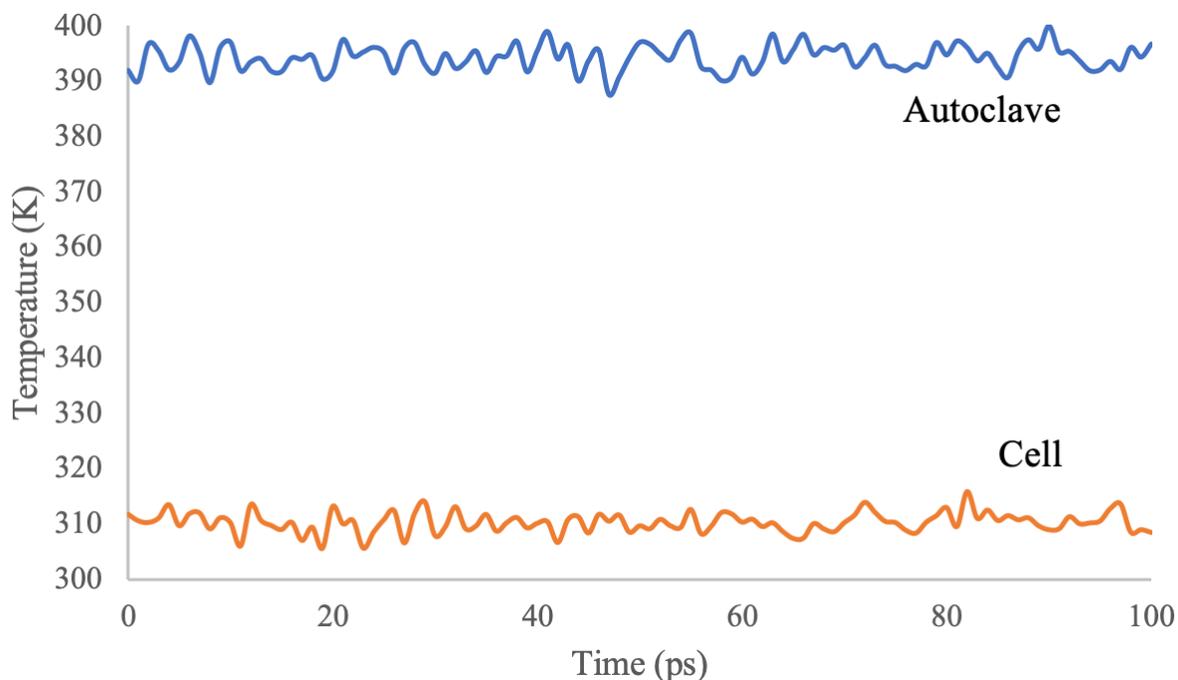


Figure 4-18. The temperature curves for the Tail Assembly Chaperone 1 protein during the Autoclave simulation (blue) and the Cell simulation (orange). The autoclave reference temperature is 394.15 K, while the cell reference temperature is 310.15 K.

After the NVT step, the pressure was analyzed to ensure it remained around the constant value. There is expected variation in the pressure at each time step, but the average was within 10 bar of the expected value. Figure 4-19 shows the pressures for each simulation of the Tail Assembly Chaperone 1 protein, as well as the running 10 picosecond average. The reference pressure is also shown in red, as it was 1.03 for both systems. Figures 5 and 6 in the Appendix contains the graphs for the other two proteins.

The average density was checked as well, but not plotted. The expected density for the Cell Simulation was near 1,000 kg/m³, however it was lower for the higher temperature. This is expected because temperature and density are inversely related.

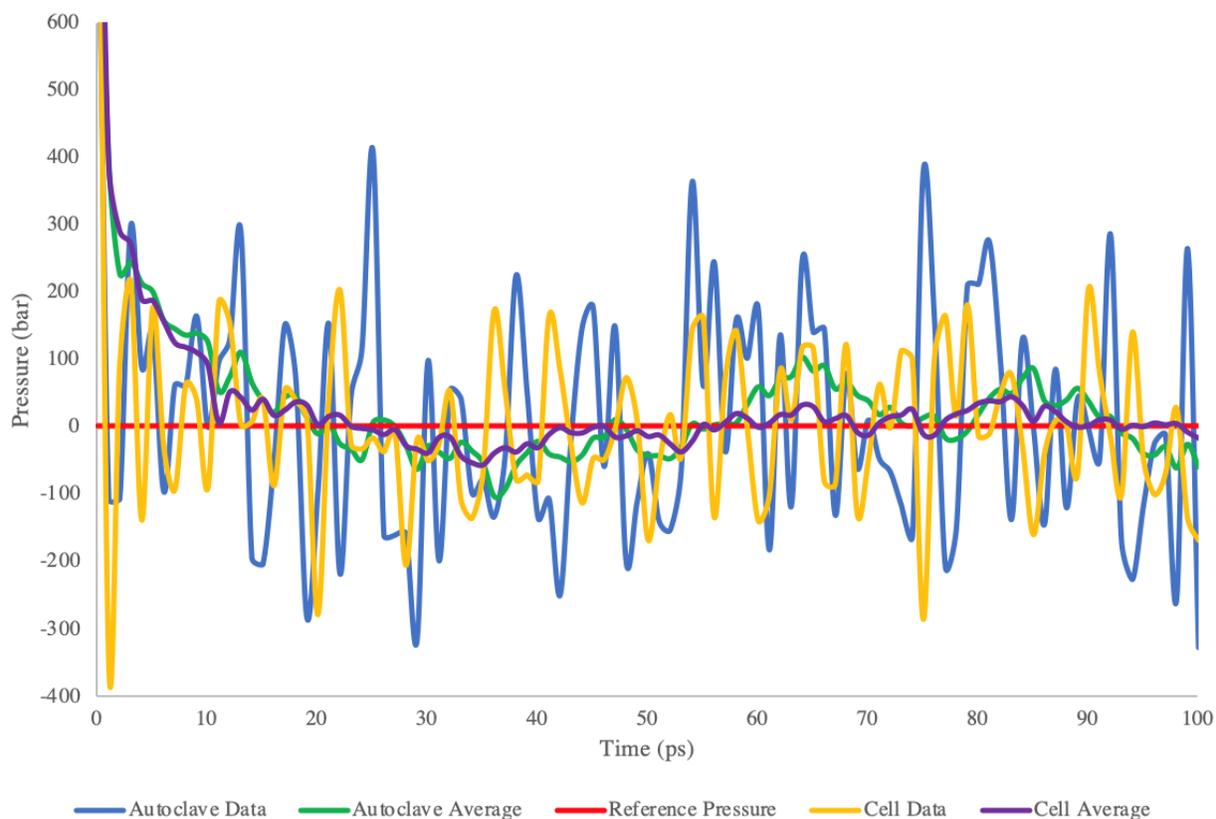


Figure 4-19. The pressure curves for the Tail Assembly Chaperone 1 protein during the Autoclave simulation and the Cell simulation. The Autoclave Data is shown in blue while the 10 ps running average is in green. The Cell Data is shown in yellow while the 10 ps running average is in purple. The reference pressure is shown in red.

Simulation results

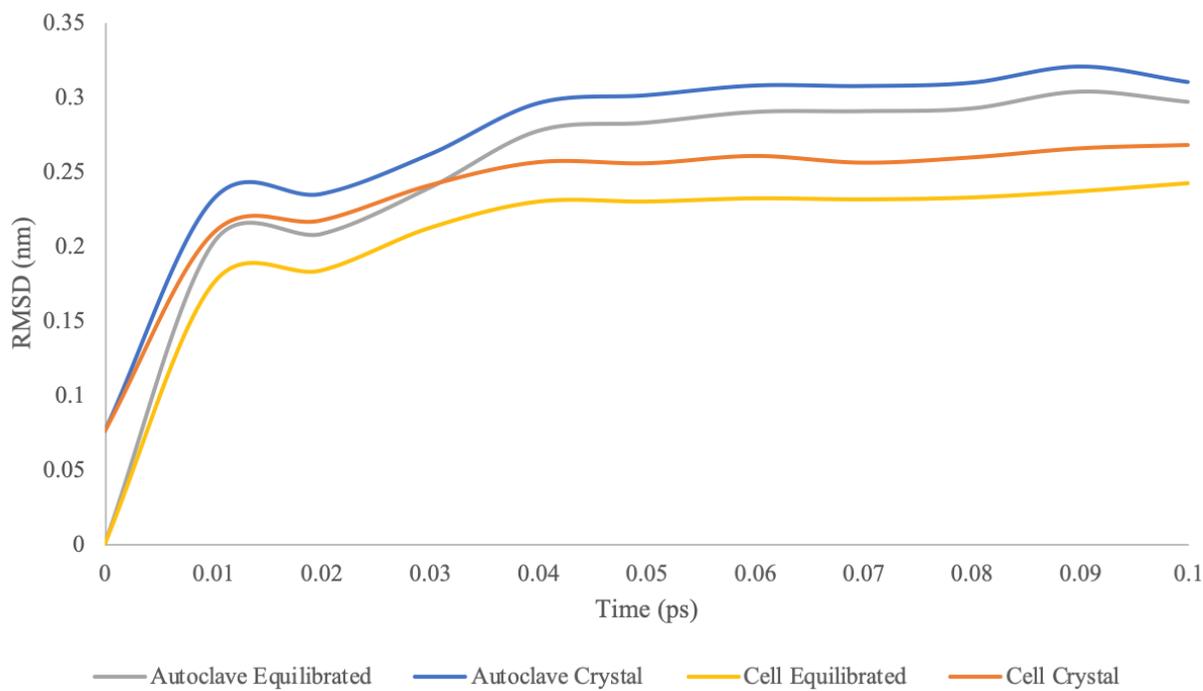


Figure 4-20. The RMSD of the Tail Assembly Chaperon 1 protein over 0.1 ns. Both simulations, Autoclave and Cell, are compared with the original crystal structure and the equilibrated structure. The Autoclave simulation has higher RMSD because the protein deteriorates at this temperature.

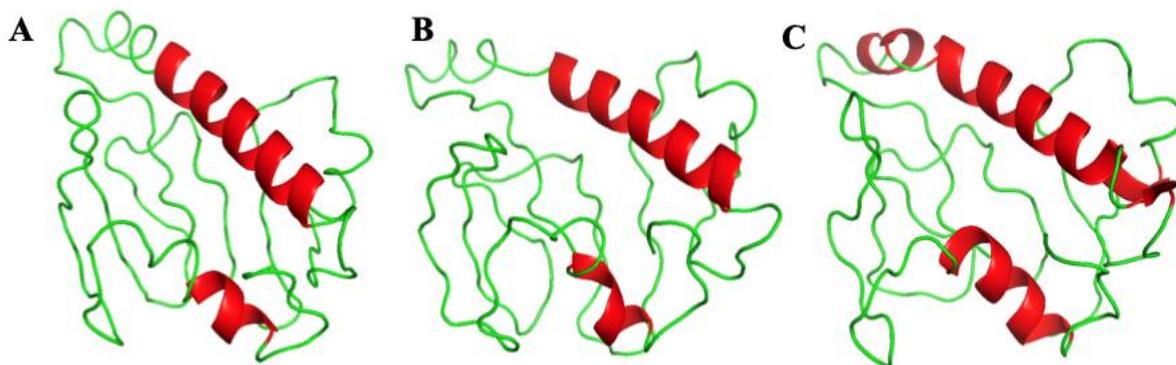


Figure 4-21. (A) the initial structure of the Tail Assembly Chaperone 1 protein. (B) The structure after 1.0 ns of the Cell simulation. (C) The structure after 0.1 ns of the Autoclave simulation.

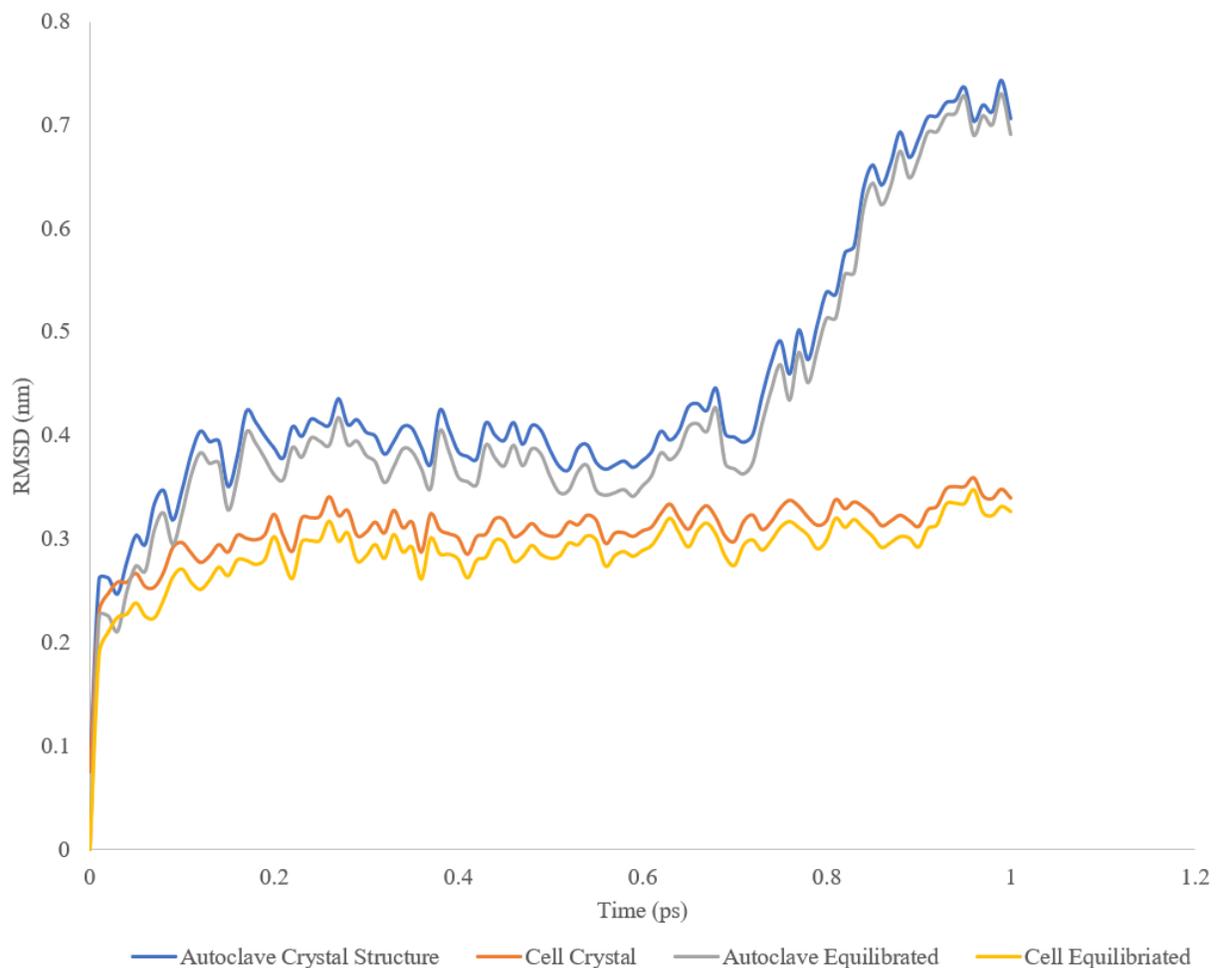


Figure 4-22. The RMSD of the Tail Assembly Chaperone 2 protein over 1.0 ns. Both simulations, Autoclave and Cell, are compared with the original crystal structure and the equilibrated structure. The Autoclave simulation has higher RMSD because the protein deteriorates at this temperature.

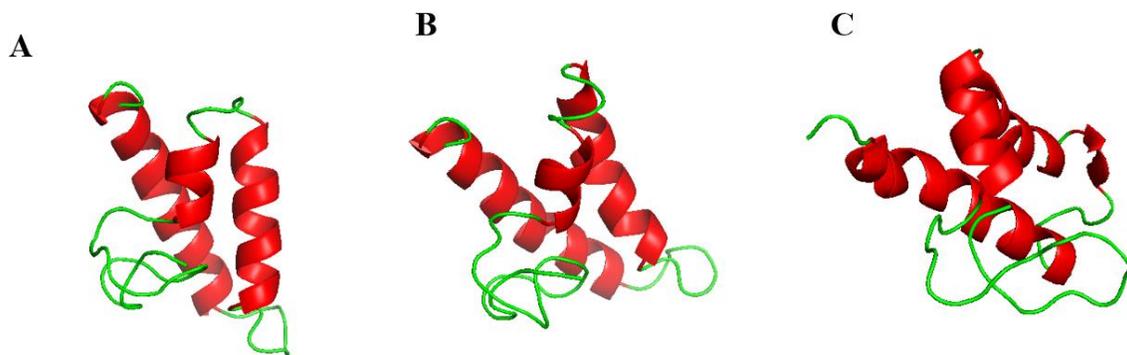


Figure 4-23. (A) the initial structure of the Tail Assembly Chaperone 2 protein. (B) The structure after 1.0 ns of the Cell simulation. (C) The structure after 1.0 ns of the Autoclave simulation.

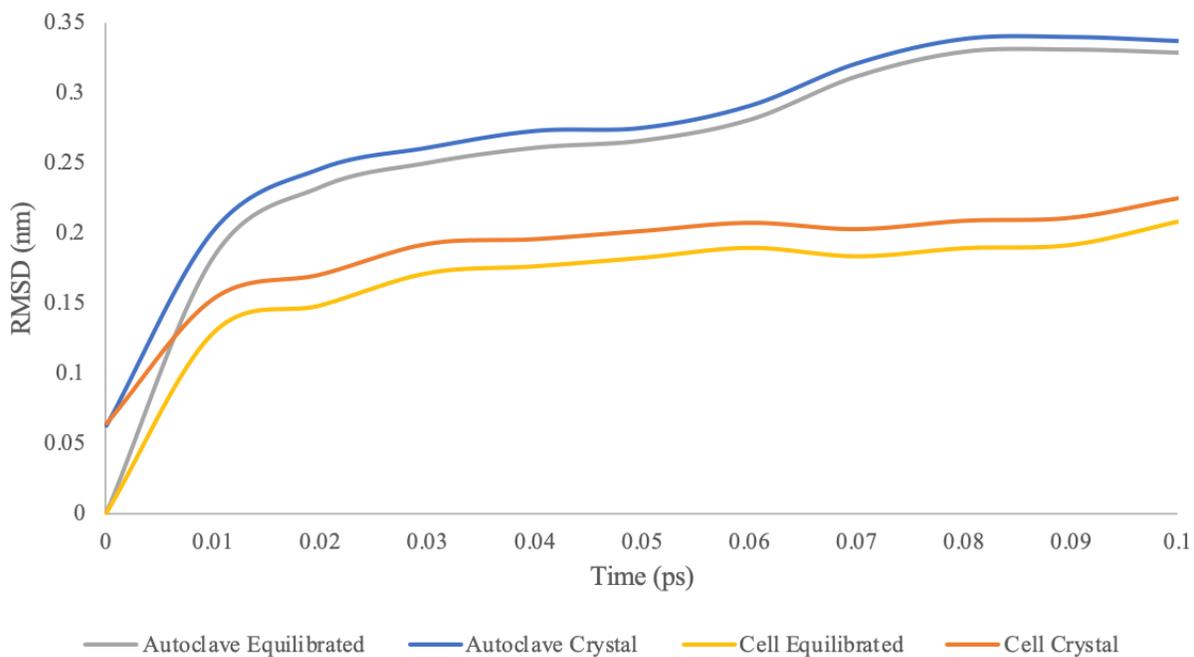


Figure 4-24. The RMSD of the Tail Assembly Chaperone 3 protein over 0.1 ns. Both simulations, Autoclave and Cell, are compared with the original crystal structure and the equilibrated structure. The Autoclave simulation has higher RMSD because the protein deteriorates at this temperature.

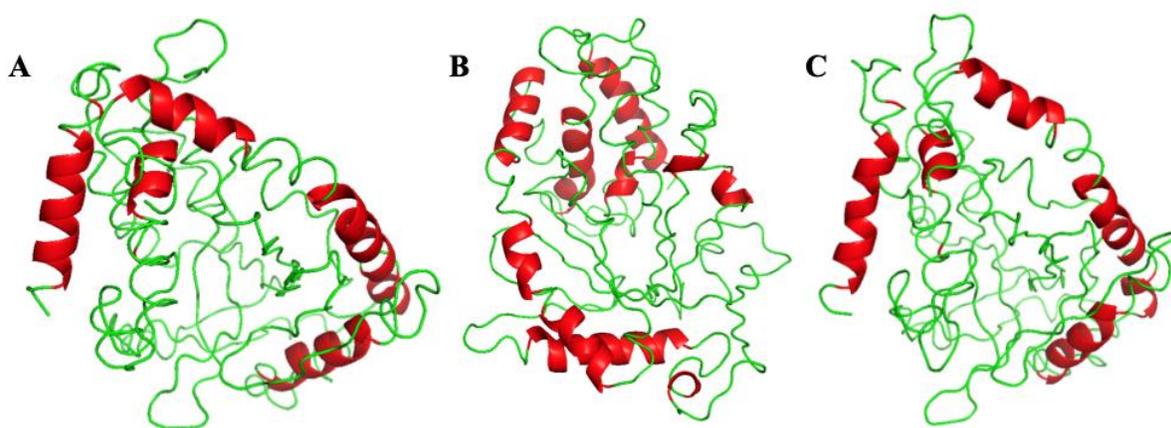


Figure 4-25. (A) the initial structure of the Tail Assembly Chaperone 3 protein. (B) The structure after 1.0 ns of the Cell simulation. (C) The structure after 0.1 ns of the Autoclave simulation.

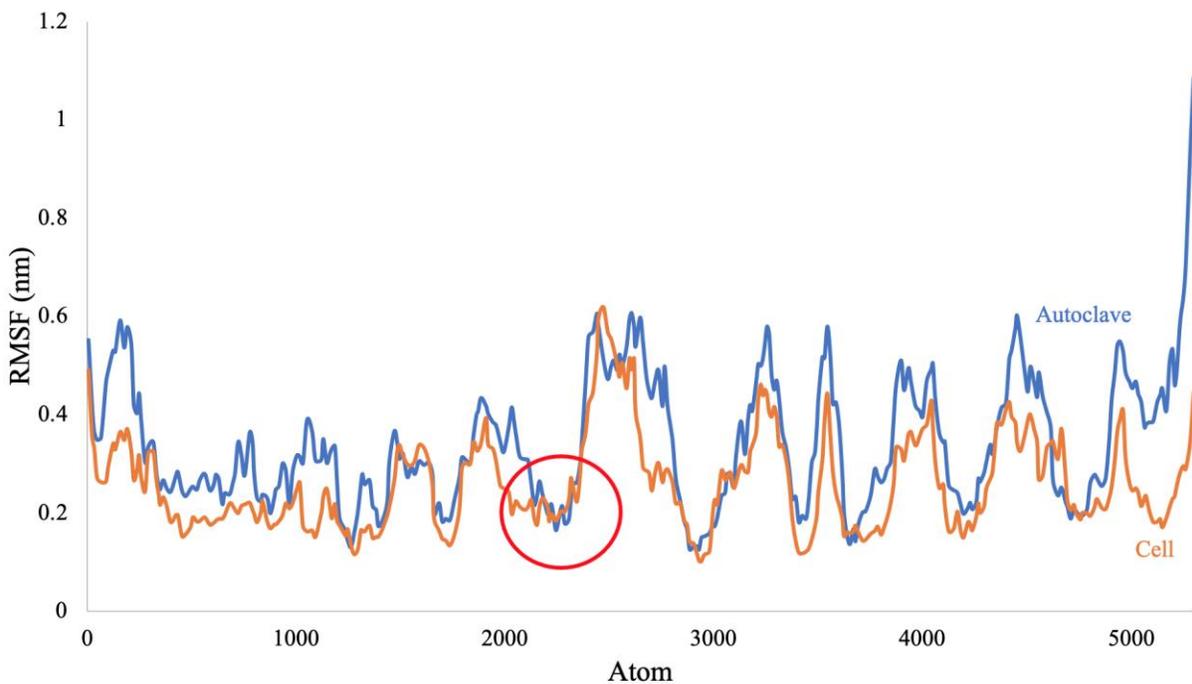


Figure 4-26. The root mean square fluctuation of the Tail Assembly Chaperone 3 protein after 100 ns of simulation across the entirety of the protein. Blue represents the Autoclave simulation while orange represents the Cell simulation. The red circle indicates the area of the protein containing the translational frameshift.

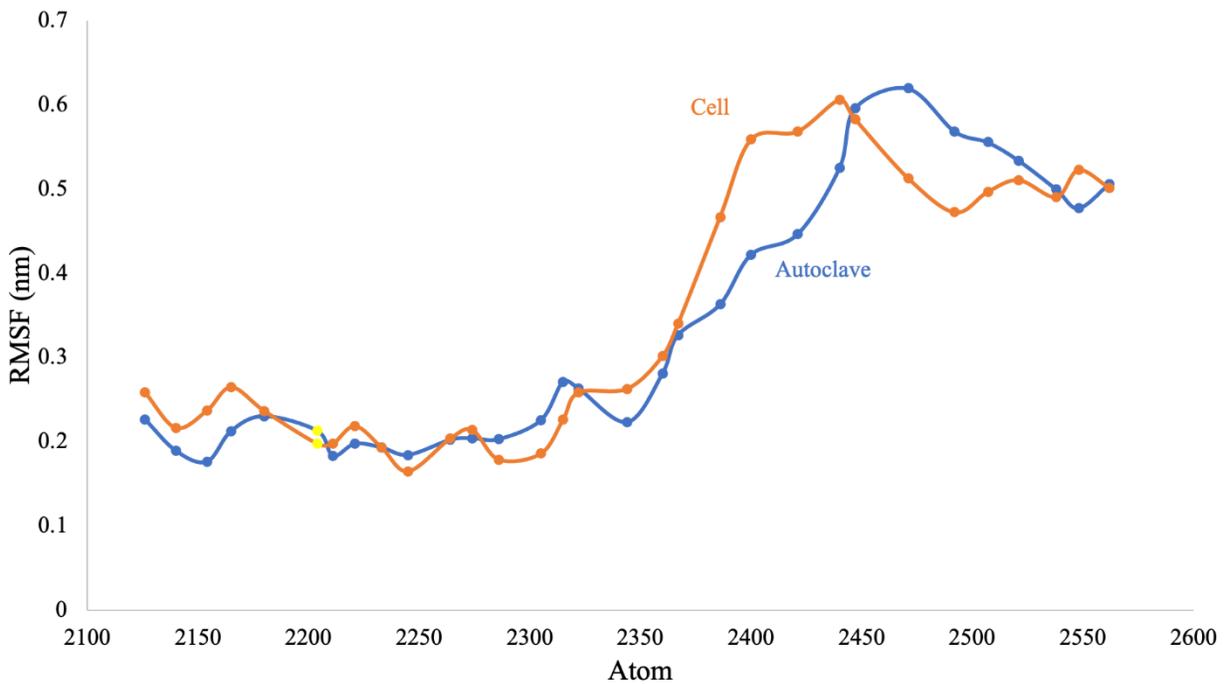


Figure 4-27. The root mean square fluctuation of the Tail Assembly Chaperone 3 protein after 100 ns of simulation within the beginning of the frameshifted region of the protein. Blue represents the Autoclave simulation while orange represents the Cell simulation. The datapoints in yellow represent the amino acid of the frameshift.

CHAPTER 5. DISCUSSION

5.1 *M. smegmatis* Growth Curves

The OD600 of *M. smegmatis* was plotted over the span of 34 hours to determine its bacterial growth curve, as seen in Figure 4-1. From this curve, *M. smegmatis* reaches the height of its peak around 24 hours and stays in its stationary phase thereafter. By plotting the bacterial growth curve, the ideal time at which to infect the host with mycobacteriophages can be determined. Wanting to capture the stationary phase in the mass spectrometry samples, the host should be inoculated before reaching the peak of its OD600 values. Therefore, it was found that an OD600 of 0.7 at the time of about 16 hours was ideal to inoculate the bacterial host with mycobacteriophages to ensure capturing the entirety of the stationary phase of the bacteria. By targeting the stationary phase of the bacteria, one can establish a strong population of bacteria for the phages to interact with to capture a robust sample for protein and lipid extraction.

When including mycobacteriophages Zalkecks and PotatoSplit, the OD600 was again measured for the control of *M. smegmatis* and each phage-treated sample, as seen in Figure 4-2. The amplitude of the bacterial growth curve significantly decreased in this measurement, most likely due to the shape and volume of the vessel containing the samples. A smaller volume of host bacteria was needed for these tests and therefore the incubation vessel changed from a 250 mL vented Erlenmeyer flask to a 50 mL vented conical tube. Though the proportions of the materials stayed the same, when agitated at 250 rpm, the surface area of the sample in contact with air decreased with the smaller vessel and therefore changed the growth rate of the bacteria. The bacteria samples were still inoculated around an OD600 of 0.7, yet the time at which this happened was at 13.5 hours of bacterial growth instead of the measured 16 hours from the previous experiment. From Figure 4-2, one can also see that the peaks of the samples at 4 hours and 14 hours of stable stationary phase were captured.

The phage-treated samples are also within the same OD600 range as the control, showing that the number of phage particles used to inoculate was not too high, which would kill the host bacteria too quickly before acquiring samples, and was not too low, ensuring that sufficient phage proteins and lipids are being measured. When comparing the phage-treated samples to the control of just bacterial host, the phage-treated samples start to deviate from the control at around 8 hours.

From there, they follow the trend of the control sample, but at a lower OD600. This indicates two distinct groups in the samples that were extracted for mass spectrometry: hours 0 and 4, which are similar to the bacterial growth curve, and hours 12 and 24 which deviate from the bacterial growth curve. From this observation, it is clear that the phages have a larger impact in the 12 and 24 hour samples, suggesting that these samples are more concentrated in phage proteins and lipids being expressed when compared to hours 0 and 4. The deviation from the bacterial growth curve at around 8 hours is most likely due to a combination of the phage killing the host bacteria, resulting in a drop in OD600, and due to the energy utilization of the nutrients supplied in the samples. The bacteria are using up more of the nutrients supplied in the samples to either create proteins and lipids to ward off against the phages or due to the phage infection, where the bacteria's metabolic machinery is being used to increase the production of phage particles. Because the existing cells present are using up more nutrients, the nutrients supplied in the samples were a limiting factor on how large the population was.

From Figure 4-2, one can see that mycobacteriophage Zalkecks had a lower OD600 stable stationary phase than mycobacteriophage PotatoSplit, which may be because of the way they interact with the host due to their different life cycles. Zalkecks is a lytic phage while PotatoSplit is temperate, meaning that it can switch between the lytic and lysogenic life cycles due to stressors. This may explain why its OD600 values of PotatoSplit were not as low as that of Zalkecks, either it wasn't as effective at killing or interacting with the host. Samples from the following time points were taken to perform protein and lipid extraction and eventually mass spectrometry: 0 hours, 4 hours, 12, hours, and 24 hours.

5.2 Testing Multiple Methods of Protein Extraction

By comparing the protein phases from the lipid extraction methods MTBE and Bligh Dyer to the proteins of the standard acetone, the purpose was to investigate if *M. smegmatis* samples could be multiplexed. In other words, if the protein phases from the traditionally lipid extraction methods were comparable to or outperformed the proteins from the standard acetone protein extraction method, then one extraction method could potentially be used to gather both proteins and lipid in future experiments. The criterion for comparison of the proteins from each method were which method was able to capture the widest range of significant proteins. Figure 4-3 illustrates the number of proteins found through each method of protein extraction. Acetone was

the method that captured the greatest number of proteins, only missing 25 that were exclusive to the Bligh Dyer extraction method. All the proteins present in the MTBE extraction method were also contained within the acetone extraction method. In terms of significant proteins, all are listed in Appendix Table 1. When comparing acetone vs Bligh Dyer, 128 significant proteins were more concentrated in acetone while only 54 significant proteins were more concentrated in Bligh Dyer. When comparing acetone to MTBE, 303 significant proteins were more concentrated in acetone while only 60 significant proteins were more concentrated in MTBE.

The significant proteins were then examined by the PANTHER Pathway Analysis program and their cellular component GO terms are listed in Table 4-1, which lists three levels of Gene Ontology terms of increasing specificity, how many sequences are contained within that category, and their percentage within each subcategory. For example, level 1 shows the cellular anatomical entity of the cellular component GO terms. Within this category there are 28 protein sequences specific to the acetone extraction method (which is ~34% of the cellular anatomical entities) that correspond to the cytoplasm. Furthermore, within this level 2 category there are 11 sequences specific to the acetone extraction method (or ~38% of the sequences in the cytoplasm) that correspond to the mitochondrion. By investigating the cellular components of the proteins present in the extraction methods, one can see if the methods of extraction target specific regions of the host cell, which could better inform future experiments that want to target specific areas of the host cell. It should be noted that Gene Ontology terms, key words that describe the knowledge of biological domains, are applied to a wide range of models and using them for a specific bacterial model to make direct comparisons of components would be misleading. However, analyzing the general classifications of the results can lead to a better understanding of the cellular locations of these proteins. From the results in Table 4-1, the acetone extraction method contains the most sequences present in almost all of the levels. There are a few level 2 subcategories pertaining specifically to the Bligh Dyer extraction method, such as organelle subcomponent, thylakoid, catalytic complex, and proteasome complex. However, these contain an extremely low number of sequences each, which does not make a significant impact on cellular location when compared to the acetone extraction method. The MTBE extraction method resulted in the least number of unique subcategories and was outperformed by both of the other extraction methods. Between the different extraction methods, many of the GO terms stayed consistent, but depending on the extraction method, the number of sequences reported for each differed. Based on this information,

all three of the extraction methods are targeting the same wide range of cellular locations but differ in how many proteins they extract from each place. This is mainly due to the simplistic nature of *M. smegmatis*, it does not produce an extensively wide range of types of proteins and therefore each extraction method was able to extricate proteins from most cellular locations.

Based on the number of proteins present and significant proteins analyzed in terms of their cellular location, the acetone extraction method was the most successful in capturing a wide range of significant *M. smegmatis* proteins. It contained the most proteins out of any extraction method and the significant proteins ranged over various cellular locations while reporting significantly more sequences when compared to Bligh Dyer and MTBE. Therefore, the protein phases from the MTBE and Bligh Dyer methods severely underperformed against the standard acetone extraction method and the only path to multiplex *M. smegmatis* samples was if the lipid phase of the acetone extraction method was comparable to the standard.

5.3 Testing Multiple Methods of Lipid Extraction

By comparing the lipid phases of the acetone protein extraction method and the MTBE lipid extraction method to the lipids of the standard Bligh Dyer lipid extraction method, the purpose was to investigate if *M. smegmatis* samples could be multiplexed. Specifically, if the lipids from the acetone and MTBE extraction methods were comparable to or outperformed the lipids from the Bligh Dyer extraction method, then one extraction method could potentially be used to collect proteins and lipids in future phage-treated experiments. The criterion for this comparison of lipids was to determine which extraction method supplied the widest range of significant lipids. In this investigation, positive and negative ESI-MS with tandem-MS was performed to see which were also more applicable to *M. smegmatis* and therefore mycobacteriophages. Based on the type of data, lipids could not be classified into specific extraction methods without statistical analysis due to that there is area under the peak present in every lipid for each extraction method. In terms of significant lipids, all are listed in Appendix Table 2. When comparing Bligh Dyer vs acetone, 184 significant lipids were more concentrated in Bligh Dyer while 171 significant lipids were more concentrated in acetone. When comparing Bligh Dyer to MTBE, 183 significant lipids were more concentrated in acetone while 170 significant lipids were more concentrated in MTBE. MS-DIAL was also not able to find reference MS2 spectra data for every lipid present in the samples and their retention time and mass to charge ratio can be found in Appendix Table 3. However, it should

be noted that many of these unnamed lipids came back as significant in the statistical tests performed, especially those significantly upregulated in the Bligh Dyer extraction method, showing the lack of database robustness for microbial lipids, especially those related to *M. smegmatis*.

The significant lipids were then grouped by their general structural lipid class, taken from the MS-DIAL nomenclature listed in Table 4-2. Table 4-2 lists how many significant lipids came back in each lipid class to better show the range of lipids present through each one of the extraction methods, and which out of the total were more concentrated in each extraction method. By investigating the groups of lipids present in the extraction methods, one can see which methods of extraction target specific groups of lipids, which are often present in specific areas of the cell, which could better inform future experiments that want to target specific areas in the host cell. The Bligh Dyer extraction method contained the most significant lipids especially in the categories of glycerophosphoethanolamines (GP02), triradylglycerols (GL03), and glycerophosphoinositols (GP06). Phosphatidylethanolamines (PE), a subcategory that was present in the glycerophosphoethanolamines, have been identified in *Mycobacterium* as polar lipids that are found within the cell envelope [77]. Lysophosphatidylethanolamines (LPE), also a subcategory that was present in the glycerophosphoethanolamines, have been found as components of the cell membrane [78]. Triacylglycerols (TG), a subcategory that was present in the triradylglycerols, have been found as main apolar intracellular lipids within *M. tuberculosis* and can serve as a long-term energy reserve [47]. Phosphatidylinositols (PI), a subcategory that was present in the glycerophosphoinositols, are major glycerophospholipids in both *M. tuberculosis* and *M. smegmatis* [47] and can be catabolized via lysophosphatidylinositol (LPI) [78].

Similar to the proteins, the lipid classes did not drastically differ between extraction methods. Lipid classes that were unique to their extraction method contained very few results. Among the different extraction methods, they targeted mainly the cell membrane and those involved in energy utilization. Again, this may be due to the simplistic nature of *M. smegmatis*, which does not produce an extensively wide range of classes of lipids and therefore each extraction method was able to extricate proteins from most cellular locations. However, the Bligh Dyer extraction method was able to get the most significant results back from the widest range of lipid classes.

MS-DIAL also compared the lipid mass spectrometry results to the RIKEN MetaDatabase as a part of its internal lipids package [79]. It should be noted that many of the RIKEN

MetaDatabase matches to MS2 spectra data were to models outside the scope of this research and using them for a specific bacterial model to make direct comparisons of components would be misleading. However, utilizing the match to analyzing general classifications of the results can lead to a better understanding of the lipids being produced by the bacterial host.

Based on the number of significant lipids analyzed, the Bligh Dyer extraction method was the most successful in capturing *M. smegmatis* lipids. It contained the most significant lipids out of any extraction method and the significant lipids ranged over many lipid classes when compared to Bligh Dyer and MTBE. Therefore, because the lipids gathered from the standard Bligh Dyer extraction method outperformed the lipids from the acetone and MTBE methods, one should not multiplex *M. smegmatis* samples because there is not one method that outperforms the others for both proteins and lipids. Instead, individual extraction methods should be used.

5.4 Investigation of Lipids from Phage-treated *M. smegmatis*

With the data retrieved from mass spectrometry and after being filtered, different statistical tests were run through MetaboAnalyst to determine significant lipids within the context of the testing groups, all of which can be seen in Table 4-5. Multiple testing groups were used due to the similarity of the samples being run. From Figure 4-4, when comparing the phage-treated samples to the control containing only the host *M. smegmatis*, there is significant overlap in the data due to that *M. smegmatis* is contained within both samples. However, by breaking down the testing groups into smaller subcategories, one can create distinct groups to draw conclusions upon, as seen in Figure 4-5. Testing groups analyzing different time points and phages were prioritized to draw conclusions about what is being produced, or not produced, at different time points after infection, by each phage. For example, Figure 4-5 represents the differences between 0 and 4 hours compared to 12 and 24 hours in the Zalkecks treated samples and the PotatoSplit treated samples respectively. The most distinctive groups in principal component analysis with the greatest percentage of variance came from comparing different time points of the phage-treated samples, indicating that at different times after infection there are distinct, specific lipids being, or not being, produced. Through all these tests, the amount of significant lipids increased in the tests containing only Zalkecks or PotatoSplit treated samples, reiterating that these specific testing groups are the most impactful to draw conclusions from. The significant lipids from each test were then grouped by their general lipid class, taken from the MS-DIAL nomenclature listed in Table 4-5, as well as

which lipids were upregulated by each group when applicable. Some areas to note are the lipid groups that are GP02: Glycerophosphoethanolamines, which were almost all exclusive to the host in comparison to the PotatoSplit treated samples but were significantly present in both the Zalkecks treated samples and the host in their respective comparison.

Relating to the OD600 values of the bacterial growth curves in Figure 4-2, the first two time points of the phage-treated samples closely followed the bacterial host while the last two time points deviated in their OD600 values, which is also reinforced by the statistical tests performed. When performing ANOVA tests on the different hours after infection of the individual phage-treated samples, clear patterns emerged. As seen in Figure 4-6, specific lipids only showed up in the time points that closely followed the bacterial growth curve (hours 0 and 4) while others showed up in the time points that were heavily infected with phages (hours 12 and 24). These lists of significant lipids in both Zalkecks and PotatoSplit also had distinct patterns within themselves. Most fell under the categories of increasing over every time point or decreasing over every time point, as seen in Figure 4-7, or increasing at the phage-heavy time points of 12 and 24 hours or increasing at the bacteria-heavy time points of 0 and 4 hours. These significant lipids were grouped into these four categories for both the mycobacteriophage Zalkecks and mycobacteriophage PotatoSplit treated samples and were then cross-referenced against the lipids with the same patterns within the control sample. The list of remaining lipids can be seen in Table 4-3 for Zalkecks treated samples and Table 4-4 for PotatoSplit treated samples. Within these two tables, one can see that the majority of significant Zalkecks treated lipids increased over each time point or increased at the phage-heavy time points while the majority of the significant PotatoSplit treated lipids decreased over time or decreased at the phage-heavy time points. This is a strong indicator of how each phage interacts with the host, which is heavily influenced by the life cycle of the phage. The impact of Zalkecks, a lytic phage, can be seen in specific lipids being produced at time points heavily influenced by phage infection most likely due to the host cell trying to defend against the phage infection or, having been hijacked by the phage, being used in phage propagation. This was reinforced by the general cellular location and function of the lipids found with these patterns being located in the cell membrane and those involved in energy utilization [47,78]. However, in the samples treated with PotatoSplit, many of the lipids within this interaction significantly occurred within the first four hours of inoculation. PotatoSplit, a temperate phage, interacts with its host by integrating its genetic information into the host's and lying dormant

within the host until triggered by a stressor, which may explain why there is little activity at the later time points of 12 and 24 hours. The lipids being produced in this interaction, mainly Phosphatidylethanolamines (PE), also have been identified in *Mycobacterium* as polar lipids that are found within the cell envelope and likely account for the initial confrontation of phage and host within the first four hours [77].

Delving into the specific results from these tests, potential functions can also be extrapolated from this data. Many of the lipids identified by MS-DIAL were grouped into classifications, which on their own cannot give a definitive function. However, those that were compared to the RIKEN MetaDatabase [79] were able to be investigated when the model matched the scope of this research. For example, RIKEN P-VS1 ID-5826 corresponded to Homodestruxin B, a fungal phytotoxin, and was shown as decreasing at every time point in both the Zalkecks and PotatoSplit treated samples [80]. This lipid production could be explained through the context of abortive infection mechanisms which normally utilize toxins to kill both the invading phage and host. Phages mainly circumvent this hurdle by mutating specific genes to hijack the production of antitoxins that neutralize the bacterial toxin, which could explain the decrease in its production over time in both phage samples [28]. Another lipid of note was RIKEN P-VS1 ID-8167 which corresponded to Ergokonin C, an antibiotic and antifungal agent that has been propagated in gram-positive *Bacillus subtilis* for its use as a steroid and was present as increasing with every time point in both the Zalkecks and PotatoSplit treated samples [81]. This may also have a role in abortive infection mechanisms where the phages are neutralizing the bacterial toxins to hijack the host cell. RIKEN P-VS1 ID-10784, identified as sturin, has been linked to T2 phage transfection of *E. coli* spheroplasts and was present in both the Zalkecks and PotatoSplit treated samples as increasing in the phage-heavy time points [82]. Other interesting lipids found in the Zalkecks samples at hours 12 and 24 were Cribrostatin 2 (RIKEN P-VS1 ID-9219) which has been linked to having antibacterial properties for gram-positive bacteria [83] and Altromycin D (RIKEN P-VS1 ID-7426) which also display gram-positive antibiotic properties [84]. Besides these specific cases, no other lipid identified through the RIKEN MetaDatabase that had a model relatable to the scope of the experiment, again showing the lack of database robustness for microbial lipids, especially those related to *M. smegmatis*.

5.5 Proteomic Data Analysis Case Study

The method for proteomic data analysis was tested upon previously acquired protein mass spectrometry data which analyzed different time points of mycobacteriophage FrenchFry inoculation of the *M. smegmatis* host. The purpose of using this data was to test the method of using alternative bioinformatics tools to analyze the mass spectrometry data and then apply it to the mycobacteriophage Zalkecks and PotatoSplit mass spectrometry proteins. In this method, the peptides were first searched using a Python script to validate the MaxQuant search, which originally assigned all proteins a protein ID of “FrenchFry”. The code searched all the peptides present in the raw mass spectrometry data against the genome file of FrenchFry and found alternative proteins that were associated with the FrenchFry genome, all of which were labeled as “Mycobacterium:False”. The Python script printed the gene number of the match, the associated annotated function, and counted how many peptides were present in the raw data, as seen in Table 4-6. From these results, 222 peptides show up with unknown function being listed as hypothetical protein, showing the lack of database knowledge surrounding phage proteins. One can also see other proteins like the major capsid head having a high number of peptides present and their high frequency may be due to that these proteins are very large and are necessary for assembling phage particles. The protein data was then filtered and ran through MetaboAnalyst to determine significant proteins of interest. Due to that there is bacterial host within the negative control and the phage-treated samples of this data, multiple statistical tests were run, as seen in Table 4-7, to create distinct groups from the data. The statistical test comparing the exponential and stationary points of phage inoculation covered all of the data and resulted in the most significant fold change proteins and was therefore used to further protein pathway analysis. The mycobacteriophage FrenchFry proteins that were significant in this test were plotted according to which functional classification they fell under in Figure 4-8, showing many being upregulated in the samples, which is expected due to the phage interacting with the host and therefore producing phage proteins. Specifically, the phage proteins pertaining to structure and DNA replication and translation were the largest groups of known function and show that the phage is infecting and using the host to propagate. In addition, the highest group of expressed proteins were of no known function, which are listed as NKF, showing the lack of database knowledge surrounding function of phage proteins.

The significant proteins from this statistical test were also examined through the pathway analysis programs, PANTHER and DAVID. By using the significant proteins as the input for the

pathway analysis programs, it decreases the amount of noise present in the data to focus on the important pathways present. PANTHER grouped the Gene Ontology terms associated with the significant proteins based on their molecular function (Figure 4-9), biological process (Figure 4-10), and protein class (Figure 4-11). By using PANTHER, the major functional pathways were found in the data and the proteins involved in each were noted. For example, many of the pathways under biological process are involved in energy utilization like ATP Metabolic process, NADP Metabolic Process, and Catabolic Process. DAVID examined the KEGG pathway terms associated with the proteins and reported the function annotation results while showing how significant they are, as seen in Figure 4-12. Highlighting a specific result, proteins involved in bacterial chemotaxis had the highest relevance to this data indicating potential functions of cell to cell signaling of toxic environments that may occur due to the host using abortive mechanisms to avoid phage propagation. To compare, the interface of PANTHER makes it an intuitive program that easily relays the important functional information while DAVID has access to a wider range of database information. By using multiple pathway analysis programs that have different types of outputs, one can widen the scope of function present possible in the data, which is necessary for an uncommon model like mycobacteriophage and *M. smegmatis*.

5.6 Archival of Purdue University Phages

Table 4-8 displays all of the mycobacteriophages from 2018 and 2019 that were re-archived for Purdue University and sent to the HHMI Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) program at the University of Pittsburgh for long-term archival. With the archival of these phages, Purdue University has no outstanding submissions that need to be sent in for archival.

5.7 Investigation of Cluster B1 frameshift proteins

5.7.1 Structural comparisons

Tail Assembly Chaperone 1

I-TASSER compared the amino acid sequence for the known tail assembly chaperone with those in the Protein Data Bank and deduced protein function based on ligand binding sites and Gene Ontology (GO) terms. As seen in Table 4-9, no tail assembly chaperones appear which is most likely due to that there are very few tail assembly chaperones within the PDB database. The function reported by COFACTOR and COACH ligand binding site programs on the biological annotations of the target protein only had one result with a confidence interval higher than 0.8, which was electron transport/photosynthesis. Lastly, as seen in Table 4-10, there is nothing directly related to tail assembly chaperones for the GO terms. This is most likely due to that there are no GO terms for tail assembly chaperones. However, it should be noted that these results are for an already confirmed tail assembly chaperone protein, so if the others come back with similar results, it alludes to them also being tail assembly chaperones.

The top structural results reported from I-TASSER had a C-score of -4.19, an estimated TM-score of 0.27 ± 0.08 , and an estimated RMSD of $14.5 \pm 3.7 \text{ \AA}$. This structure was then analyzed in PyMOL and superimposed onto the secondary structure of a known tail assembly chaperone taken from the PDB database (Figure 4-13. The secondary structure of tail assembly chaperone 1 superimposed on the secondary structure of bacteriophage HK97 tail assembly chaperone (PDB ID: 2OB9). The structure in cyan is the I-TASSER predicted structure while the rest is the HK97 tail assembly chaperone coloring for different secondary structures.). It has a MatchAlign score of 90.598 and an RMSD of 0.833, indicating a high level of structural similarity. While the I-TASSER results seemed to not be relevant to that of tail assembly chaperones, the structural comparison showed a distinct level of similarity. Knowing that this protein sequence is from a known tail assembly chaperone, this calibrated the understanding of this data to weigh the structural comparisons over than the I-TASSER predicted functions when it comes to predicting protein function.

Tail Assembly Chaperone 2

I-TASSER compared the amino acid sequence for the second half of the tail assembly chaperone with those in the Protein Data Bank and deduced protein function based on ligand binding sites and Gene Ontology (GO) terms. As seen in Table 4-11, no tail assembly chaperones appear in the PDB hits, which again is most likely due to that there are very few tail assembly chaperones within the PDB database. The function reported by COFACTOR and COACH ligand binding site programs on the biological annotations of the target protein did not have any results with a confidence interval higher than 0.8. Lastly, as seen in Table 4-12, there is nothing directly related to tail assembly chaperones for the GO terms, due to that there are no GO terms of tail assembly chaperones. Based on the comparison of these scores to those of the known tail assembly chaperone, there is very little evidence to prove this as a tail assembly chaperone. However, this amino acid sequence would not be produced in reality because it is missing the initial sequence of the annotation frameshift. Referring to Figure 2-5, the amino acid sequence is the equivalent to the third protein product that would not be made. Therefore, the results for this section are valid in that this protein in reality does not exist and the scores reflect that.

The top structural results reported from I-TASSER had a C-score of -2.74, an estimated TM-score of 0.40 ± 0.13 , and an estimated RMSD of $9.2 \pm 4.6 \text{ \AA}$. This structure was then analyzed in PyMOL and superimposed onto the secondary structure of a known tail assembly chaperone taken from the PDB database (Figure 4-14). It has a MatchAlign score of 131.694 and an RMSD of 8.739, indicating a low level of structural similarity. Again, due to that this protein sequence would not exist in reality, these scores seem valid.

Tail Assembly Chaperone 3

I-TASSER compared the amino acid sequence for the supposed frameshift tail assembly chaperone with those in the Protein Data Bank and deduced protein function based on ligand binding sites and Gene Ontology (GO) terms. As seen in Table 4-13, no tail assembly chaperones appear in the PDB hits, which again is most likely due to that there are very few tail assembly chaperones within the PDB database. The function reported by COFACTOR and COACH ligand binding site programs on the biological annotations of the target protein did not have any results with a confidence interval higher than 0.8. Lastly, as seen in Table 4-14, there is nothing directly related to tail assembly chaperones for the GO terms, due to that there are no GO terms of tail

assembly chaperones. Based on the comparison of these scores to those of the known tail assembly chaperone, there is very little evidence to prove this as a tail assembly chaperone.

The top structural results reported from I-TASSER had a C-score of -3.21, an estimated TM-score of 0.36 ± 0.12 , and an estimated RMSD of $14.3 \pm 3.8 \text{ \AA}$. This structure was then analyzed in PyMOL and superimposed onto the secondary structure of a known tail assembly chaperone taken from the PDB database (Figure 4-15). It has a MatchAlign score of 127.315 and an RMSD of 7.084, indicating a low level of structural similarity. However, it should be noted that there are very few bacteriophage tail assembly chaperones located in the PDB and those that are listed do not specify if they are the translational frameshift. Therefore, these results should not exclusively determine if the predicted frameshift has been located correctly.

This predicted protein structure for the supposed frameshift was also compared to the predicted I-TASSER structure of a known annotated frameshift from *Gordonia terrae* phage CloverMinnie. *Gordonia terrae* phage CloverMinnie, classified in the DR cluster, was chosen based on its initial high scores when running the previous tail assembly chaperones through the BLAST database. The supposed B1 frameshift was superimposed onto the predicted structure of *Gordonia terrae* phage CloverMinnie, as seen in Figure 4-16, and had a MatchAlign score of 97.686 and an RMSD of 0.798, indicating an extremely high level of structural similarity compared to previous comparisons. Based on the previous tail assembly chaperone results, the structural similarity was weighed higher than the I-TASSER scores in terms of significance. The comparison between the supposed B1 frameshift and *Gordonia terrae* phage CloverMinnie also are from phage that attacks different hosts with B1 infecting *Mycobacterium smegmatis* and CloverMinnie infecting *Gordonia terrae*, which may allude to some of the inconsistencies of the structural comparison. However, despite the fact that the proteins of these two phages infect different hosts, there is a strong structural comparison between the two proteins. Other studies would need to be done to definitively determine that this is the annotated frameshift, especially when it comes to determining the exact slippery sequence, but this gives strong evidence that this is the correct location of the frameshift in the B1 cluster.

5.7.2 Molecular Dynamics Simulations

The results from the simulation will be discussed by each protein. For every protein, the Root Mean Square Distance was plotted based on the original Equilibrated Structure or the Crystal

Structure. Additionally, simulations were analyzed in PyMOL. For all proteins considered, the structure deteriorates more at the high temperature of the Autoclave, which was the expected result. All proteins had average RMSD distances of near 0.5 nm for the Autoclave and 0.3 for the Cell.

Tail Assembly Chaperone 1

The first tail assembly chaperone had an average RMSD of 0.30 nm for the Autoclave simulations, and 0.25 nm for the Cell simulations over a time course of 0.1 ns. The results are shown in Figure 4-20. It is difficult to tell over time how the Autoclave simulation protein will continue to deteriorate but based on comparisons to the other phage proteins simulated for longer periods of time, this protein seems to follow its initial trend. The same principle applies to the Cell simulation, which is expected to remain relatively stable after the first 0.1 ns. The protein could not be simulated in either temperature for longer than 0.1 ns without the program crashing due to how large the protein was. The conformation changes of the tail assembly chaperone 1 protein can be seen in Figure 4-21. Their changes are not drastic in this protein and harder to notice visually. The alpha helices have turned outward, and the molecule has stretched in width, more so for the Autoclave simulation than the Cell simulation.

Tail Assembly Chaperone 2

The second tail assembly chaperone had an average RMSD of 0.43 nm for the Autoclave simulations, and 0.29 nm for the Cell simulations over a time course of 1.0 ns. The results are shown in Figure 4-22. While the Autoclave simulation continues to deteriorate over time, the Cell simulation remains relatively stable after the first 0.1 ns. This is expected, as the natural production of the protein would be at this temperature. The RMSD of 0.29 is lower than many of the other proteins in this study, which could be due to its small size. The conformation changes of the tail assembly chaperone two protein can be seen in Figure 4-23. Their changes are not drastic in this protein and harder to notice visually. The alpha helices have turned slightly in different directions from each and the molecule has stretched in width.

Tail Assembly Chaperone 3

The annotated B1 frameshift had an average RMSD of 0.33 nm for the Autoclave simulations, and 0.22 nm for the Cell simulations over a time course of 0.1 ns. The results are shown in Figure 4-24. It is difficult to tell over time how the Autoclave simulation protein will continue to deteriorate but based on comparisons to the other phage proteins simulated for longer periods of time, this protein seems to follow its initial trend. The same principle applies to the Cell simulation, which is expected to remain relatively stable after the first 0.1 ns. The protein could not be simulated in either temperature for longer than 0.1 ns without the program crashing due to how large the protein was. The conformation changes of the supposed frameshift tail assembly chaperone protein can be seen in Figure 4-25. Their changes are not drastic in this protein and harder to notice visually. The alpha helices have turned outward, and the molecule has stretched in width, more so for the Autoclave simulation than the Cell simulation.

When examining the further analysis of the frameshifted protein, the root mean square fluctuation shows distinct areas of stability within the protein, as shown in Figure 4-26. The increase in RMSF is expected at the Autoclave simulation because there is an increase in movement as temperature and pressure increases. However, when inspecting the area of the protein around the translational frameshift in Figure 4-27, the amino acid change resulting from the frameshift is highly stable in that there is little change in RMSF between the Cell and Autoclave simulations. The amino acids before the frameshift also result in minimal change in RMSF between the Cell and Autoclave simulations indicating high stability of this portion of the protein. High stability of the frameshifted area may have biological significance in that it is vital that this portion of the protein remain stable in order to correctly utilize the translational frameshift to produce both protein products.

CHAPTER 6. CONCLUSION

Through this research project, the relationship between bacteriophages and their host was examined using mass spectrometry and bioinformatics tools to analyze what significant proteins and lipids are being produced. In order to gather data for this exploratory study, multiple methods of protein and lipid extraction were explored to determine if *M. smegmatis* samples could be multiplexed in order to create a more efficient way to extract proteins and lipids. Through the use of modern methods of untargeted proteomics and lipidomics, one has the capability to fill these gaps of what is being produced by the bacteriophage and host in this interaction and expand upon potential bacteriophage functions to provide a more comprehensive understanding of the pathogenesis of the infection. Also in this research study, the archival of novel Purdue University mycobacteriophages and the investigation of the cluster B1 frameshift using molecular dynamic simulations is explored to add to the database knowledge currently known about bacteriophages.

Through this study, the proteomic analysis of mass spectrometry data determined that the acetone method of extraction was the most applicable to *M. smegmatis* and was used for further phage-treated samples. In addition, the lipidomic analysis of mass spectrometry determined that the Bligh Dyer method of extraction was the most applicable to *M. smegmatis* and was used for further phage-treated samples. Based on this information, *M. smegmatis* samples should not be multiplexed because there was not one extraction method that outperformed the others for both proteins and lipids. The lipids extracted from the phage-treated samples were categorized according to classification and showed functions relating to the cell membrane and to energy utilization. Specific lipids from the phage-treated samples also indicated involvement in the abortive infection mechanisms of the phage-host interaction. The mass spectrometry data analysis methodology was also applied to a case study of mycobacteriophage FrenchFry to show how the use of alternative bioinformatics tools can define function for an uncommon model of mycobacteria and mycobacteriophages. Also, two years' worth of novel Purdue University mycobacteriophages were sent in for archival and the potential translational frameshift was identified.

By investigating the products of the phage-host interaction, one can understand their potential risks and ensure safety in their applications. However, more needs to be studied in order for real world applications of phages to be used widely including further analysis of the protein

products and delving into discovery of potential function of the proteins of no known function. Specifically for this study, the application of the protein data analysis methodology should be implemented on the mycobacteriophage Zalkecks and PotatoSplit proteins retrieved through mass spectrometry. In the molecular dynamics simulations, the comparison to known proteins at simulation times longer than what were achieved through this project should be completed to create a deeper understanding of function in the system.

APPENDIX

Appendix Table 1. All significant proteins according to fold change and p-value, with their corresponding values, in the 3 methods of protein extraction when comparing to the standard: acetone. A fold change above 1.5 and a p-value less than 0.05 were considered significant. Log2 fold changes in pink indicate the lipid is upregulated in the acetone extraction method, blue indicates upregulation in the Bligh Dyer extraction method, green indicates upregulation in the MTBE extraction method, and black indicates no significance (NS). Proteins are listed by their UniProt Accession ID.

Protein	Acetone vs Bligh Dyer		Acetone vs MTBE	
	<i>Log2 Fold Change</i>	<i>P-Val</i>	<i>Log2 Fold Change</i>	<i>P-Val</i>
A0QND6	NS	NS	2.6921	0.0022
A0QND7	3.0110	NS	3.0110	0.0109
A0QNE2	1.2224	NS	1.2224	NS
A0QNF5	-0.8672	NS	1.9619	NS
A0QNF6	NS	NS	-0.6738	NS
A0QNG7	NS	NS	2.9925	0.0181
A0QNJ5	-2.7890	NS	NS	NS
A0QNJ6	-2.3760	0.0000	NS	NS
A0QNJ7	NS	NS	2.6690	0.0149
A0QNQ9	1.8881	NS	1.8881	NS
A0QNZ3	NS	NS	2.6690	0.0149
A0QP06	NS	NS	2.6521	0.0015
A0QP11	NS	NS	2.4696	0.0002
A0QP20	NS	NS	-1.4888	NS
A0QP27	2.4381	0.0000	2.4381	0.0000
A0QP89	2.6662	NS	-1.4031	NS
A0QP93	NS	NS	2.9596	0.0177
A0QPE7	NS	NS	2.5052	0.0002
A0QPE8	0.6372	NS	3.0964	0.0158
A0QPH5	NS	NS	2.4230	0.0002
A0QPV4	2.4381	0.0000	2.4381	0.0000
A0QPV9	2.1534	NS	2.1534	NS
A0QPZ5	2.8670	NS	2.8670	0.0054
A0QQ61	NS	NS	-0.7994	NS
A0QQ62	2.4885	0.0012	2.4885	0.0012
A0QQ65	NS	NS	-1.7853	NS
A0QQ72	NS	NS	-2.6321	NS
A0QQC1	NS	NS	2.7578	0.0194

A0QQC8	NS	NS	2.3809	0.0000
A0QQF0	NS	NS	2.3018	NS
A0QQJ4	NS	NS	2.5567	0.0011
A0QQJ6	NS	NS	3.0894	NS
A0QQS3	2.0854	NS	2.0854	NS
A0QQU5	NS	NS	-1.1967	0.0005
A0QQW5	NS	NS	2.3890	0.0000
A0QQW8	-0.8397	NS	2.7600	0.0063
A0QQX7	NS	NS	2.6921	0.0022
A0QR00	NS	NS	2.5742	0.0005
A0QR29	NS	NS	-0.9954	NS
A0QR33	2.0697	NS	2.0697	NS
A0QR46	1.8881	NS	1.8881	NS
A0QR51	1.2224	NS	1.2224	NS
A0QR89	1.2427	NS	2.4287	0.0001
A0QRB0	-0.7357	NS	2.6690	0.0149
A0QRB1	NS	NS	2.4885	0.0012
A0QRD4	1.8881	NS	1.8881	NS
A0QRE7	NS	NS	2.6921	0.0022
A0QRN7	1.9619	NS	1.9619	NS
A0QRS0	NS	NS	2.5198	0.0006
A0QRX4	-0.9378	NS	-2.3185	NS
A0QRZ8	NS	NS	1.8881	NS
A0QS45	NS	NS	-2.0426	NS
A0QS46	NS	NS	-1.4140	NS
A0QS62	2.6706	NS	2.6706	0.0080
A0QS63	NS	NS	2.5305	0.0004
A0QS66	NS	NS	-0.8270	0.0238
A0QS72	NS	NS	2.6706	0.0080
A0QS90	NS	NS	2.8706	0.0146
A0QS98	NS	NS	-1.7055	NS
A0QSB1	NS	NS	1.9474	NS
A0QSD0	NS	NS	2.4381	0.0000
A0QSD1	2.6998	NS	2.6998	0.0062
A0QSD2	3.1972	NS	3.1972	0.0173
A0QSD3	NS	NS	2.4381	0.0000
A0QSD4	NS	NS	-2.2104	NS
A0QSD5	1.9619	NS	1.9619	NS
A0QSD6	1.4895	NS	2.5939	0.0007
A0QSD8	1.9619	NS	1.9619	NS

A0QSD9	2.2586	NS	2.2586	NS
A0QSE0	2.7377	0.0025	2.7377	0.0025
A0QSF9	1.2402	NS	-1.0392	NS
A0QSG0	2.6521	0.0015	-0.6667	NS
A0QSG1	NS	NS	2.4381	0.0000
A0QSG2	2.4885	0.0012	2.4885	0.0012
A0QSG3	NS	NS	2.6252	0.0009
A0QSG4	0.8932	NS	2.6061	0.0032
A0QSG6	NS	NS	-1.0930	NS
A0QSG7	2.6662	NS	2.6662	0.0064
A0QSG8	2.6706	NS	2.6706	0.0080
A0QSH8	-1.2224	NS	NS	NS
A0QSJ0	NS	NS	2.6690	0.0149
A0QSJ2	NS	NS	0.7349	NS
A0QSK7	2.4381	0.0000	2.4381	0.0000
A0QSL1	-0.6896	NS	-2.2083	NS
A0QSL5	NS	NS	2.4696	0.0002
A0QSL6	1.9474	NS	1.9474	NS
A0QSL7	NS	NS	-1.3480	NS
A0QSL9	NS	NS	2.4885	0.0012
A0QSN7	NS	NS	-0.6429	NS
A0QSN8	0.8536	NS	2.5972	0.0030
A0QSP0	2.8169	NS	2.8169	0.0054
A0QSP1	0.8108	NS	2.4321	0.0001
A0QSP2	2.8706	NS	2.8706	0.0146
A0QSP8	NS	NS	2.7349	0.0131
A0QSP9	0.6075	NS	2.5084	0.0004
A0QSR5	NS	NS	2.8551	NS
A0QSS3	NS	NS	-1.0088	0.0002
A0QSS4	NS	NS	-1.3631	NS
A0QSU3	NS	NS	-2.6276	NS
A0QSU4	NS	NS	2.2586	NS
A0QSX4	-0.9796	NS	1.2224	NS
A0QSZ1	NS	NS	2.5017	0.0003
A0QSZ3	NS	NS	NS	0.0080
A0QT01	0.7330	NS	2.5244	0.0021
A0QT04	NS	NS	2.4885	0.0012
A0QT08	NS	NS	2.3196	NS
A0QT14	2.4381	0.0000	2.4381	0.0000
A0QT17	NS	NS	1.0839	NS

A0QT18	NS	NS	-0.7885	NS
A0QT19	NS	NS	-0.7789	0.0132
A0QT20	NS	NS	2.4421	0.0001
A0QT21	NS	NS	-1.5898	NS
A0QT22	-0.8249	NS	2.0854	NS
A0QT42	1.4905	NS	2.5567	0.0011
A0QT50	-0.6899	NS	2.8657	0.0072
A0QT92	2.4381	0.0000	2.4381	0.0000
A0QT98	NS	NS	2.4230	0.0002
A0QTA4	-2.4165	NS	NS	NS
A0QTE1	NS	NS	-0.7110	NS
A0QTE3	NS	NS	2.6489	0.0025
A0QTE7	NS	NS	2.4918	0.0010
A0QTF4	NS	NS	2.4287	0.0001
A0QTK2	-1.2224	NS	NS	NS
A0QTS8	0.6221	NS	2.4381	0.0000
A0QTS9	NS	NS	-0.6462	NS
A0QTT7	-2.6230	NS	NS	NS
A0QTV1	NS	NS	0.6222	NS
A0QTV4	NS	NS	-0.6716	NS
A0QU00	1.9619	NS	1.9619	NS
A0QU07	1.9474	NS	1.9474	NS
A0QU51	NS	NS	-1.3172	NS
A0QU52	NS	NS	-1.1836	0.0015
A0QU54	-0.7368	NS	1.9474	NS
A0QUA6	NS	NS	2.4381	0.0000
A0QUG7	NS	NS	2.8294	NS
A0QUH9	NS	NS	-2.4445	NS
A0QUM7	NS	NS	2.8706	0.0146
A0QUV6	NS	NS	2.5337	0.0006
A0QUY2	-1.1097	NS	-1.1856	NS
A0QUY3	0.6221	NS	2.4381	0.0000
A0QUY7	NS	NS	2.2760	NS
A0QUZ0	NS	NS	2.4381	0.0000
A0QV09	-0.7017	NS	2.1866	NS
A0QV10	NS	NS	2.6648	0.0048
A0QV12	NS	NS	2.8933	NS
A0QV14	2.2586	NS	2.2586	NS
A0QV17	1.9474	NS	1.9474	NS
A0QV42	2.9596	NS	2.9596	0.0177

A0QV51	NS	NS	2.6557	0.0012
A0QV52	0.6818	NS	2.5354	0.0004
A0QVB8	0.6072	NS	-1.5137	NS
A0QVB9	NS	NS	2.5512	0.0039
A0QVC7	NS	NS	2.4381	0.0000
A0QVE0	NS	NS	2.5941	0.0022
A0QVK0	NS	NS	2.5337	0.0006
A0QVK3	-1.9456	NS	NS	NS
A0QVL2	NS	NS	2.3890	0.0000
A0QVM0	2.6921	0.0022	2.6921	0.0022
A0QVQ3	NS	NS	2.5512	0.0039
A0QVQ5	NS	NS	2.3983	0.0000
A0QVR3	2.0854	NS	2.0854	NS
A0QVU2	NS	NS	2.5017	0.0003
A0QVV5	2.6706	NS	2.6706	0.0080
A0QVX3	NS	NS	2.5182	0.0004
A0QVX4	NS	NS	-2.9397	NS
A0QVX6	NS	NS	2.0624	NS
A0QVY4	NS	NS	2.0697	NS
A0QVY9	NS	NS	2.6585	0.0015
A0QVZ3	NS	NS	1.9980	NS
A0QW02	NS	NS	-2.3092	0.0003
A0QWG2	1.2224	NS	1.2224	NS
A0QWH1	2.4381	0.0000	2.4381	0.0000
A0QWN3	NS	NS	-0.8712	NS
A0QWQ5	-0.6473	NS	1.9474	NS
A0QWS8	NS	NS	-1.5238	NS
A0QWT3	NS	NS	2.4823	0.0001
A0QWU8	2.0854	NS	2.0854	NS
A0QWV1	1.2224	NS	1.2224	NS
A0QWW2	NS	NS	-0.6196	NS
A0QWW3	NS	NS	2.4951	0.0002
A0QWW4	1.0202	NS	2.4885	0.0012
A0QWX6	NS	NS	2.5017	0.0003
A0QWX8	NS	NS	-0.8316	NS
A0QWY3	NS	NS	2.3890	0.0000
A0QX20	NS	0.0044	NS	NS
A0QX24	-1.2692	NS	1.9619	NS
A0QX32	NS	NS	1.2224	NS
A0QX35	NS	NS	2.9738	NS

A0QX36	NS	NS	2.8339	0.0122
A0QXA3	NS	NS	0.6982	NS
A0QXD0	2.5339	NS	2.5339	NS
A0QXS8	NS	NS	-1.3522	NS
A0QXX7	NS	NS	0.7309	NS
A0QXY0	2.4381	0.0000	2.4381	0.0000
A0QXZ5	NS	NS	2.4381	0.0000
A0QY21	NS	NS	2.8670	0.0054
A0QY55	2.6585	0.0015	2.6585	0.0015
A0QY58	2.8169	NS	2.8169	0.0054
A0QY79	2.5337	0.0006	2.5337	0.0006
A0QY95	NS	NS	3.4657	NS
A0QYA9	1.2224	NS	1.2224	NS
A0QYB5	0.9229	NS	2.7377	0.0025
A0QYD3	NS	NS	2.7112	0.0029
A0QYD5	NS	NS	2.8632	NS
A0QYD6	2.8706	NS	2.8706	0.0146
A0QYE0	2.4381	0.0000	2.4381	0.0000
A0QYE7	NS	NS	0.8046	NS
A0QYF5	NS	NS	3.6197	NS
A0QYF7	NS	NS	-0.7261	NS
A0QYG2	2.4885	0.0012	2.4885	0.0012
A0QYG3	2.8651	NS	2.8651	0.0102
A0QYL9	NS	NS	2.3196	NS
A0QYN8	3.3398	NS	3.3398	0.0247
A0QYN9	2.4381	0.0000	2.4381	0.0000
A0QYQ7	NS	NS	2.0697	NS
A0QYS6	NS	NS	2.6585	0.0015
A0QYU6	2.1622	NS	2.1622	NS
A0QYU8	-1.4521	NS	1.9619	NS
A0QYY6	NS	NS	-0.8977	NS
A0QZ34	-0.7360	NS	1.9474	NS
A0QZ37	0.6025	NS	3.1972	0.0173
A0QZ46	NS	NS	2.5354	0.0004
A0QZ47	NS	NS	2.8981	0.0063
A0QZ48	0.5955	NS	2.3683	0.0000
A0QZ96	NS	NS	1.9619	NS
A0QZB3	NS	NS	2.9596	0.0177
A0QZW2	NS	NS	-2.7836	NS
A0QZX6	NS	NS	-2.7836	NS

A0R006	NS	NS	2.8499	0.0055
A0R012	1.2224	NS	1.2224	NS
A0R048	0.9926	NS	2.5512	0.0039
A0R050	NS	NS	2.4381	0.0000
A0R057	-0.7386	NS	2.6521	0.0015
A0R066	NS	NS	2.8632	NS
A0R067	NS	NS	3.1972	0.0173
A0R069	NS	NS	-0.8760	NS
A0R072	NS	NS	-2.1678	NS
A0R083	-1.9456	NS	NS	NS
A0R0A1	NS	NS	2.4230	0.0002
A0R0B0	NS	NS	2.7359	0.0025
A0R0B3	NS	NS	-1.4525	NS
A0R0B4	NS	NS	2.4885	0.0012
A0R0B5	1.2369	NS	2.6286	0.0065
A0R0C7	NS	NS	2.6690	0.0149
A0R0C8	1.9474	NS	1.9474	NS
A0R0F4	1.9699	NS	2.4287	0.0001
A0R0G8	2.4885	0.0012	2.4885	0.0012
A0R0I8	NS	NS	2.5354	0.0004
A0R0R9	-1.2224	NS	NS	NS
A0R0W7	NS	NS	2.4709	0.0008
A0R0W9	NS	NS	1.9619	NS
A0R0X1	0.9716	NS	3.4797	NS
A0R102	-0.6923	NS	1.9619	NS
A0R111	1.9474	NS	1.9474	NS
A0R151	2.5972	0.0030	2.5972	0.0030
A0R183	2.6921	0.0022	2.6921	0.0022
A0R197	-2.3218	NS	NS	NS
A0R198	NS	NS	2.5017	0.0003
A0R1A7	-1.2224	NS	NS	NS
A0R1B3	-0.6269	NS	2.3442	0.0000
A0R1B5	NS	NS	2.5337	0.0006
A0R1B6	2.4381	0.0000	2.4381	0.0000
A0R1C3	NS	NS	2.8339	0.0122
A0R1D3	NS	NS	2.6845	0.0018
A0R1D4	-1.9456	NS	NS	NS
A0R1D7	1.2224	NS	1.2224	NS
A0R1D9	NS	NS	-1.3165	NS
A0R1H5	0.8353	NS	2.4230	0.0002

A0R1H7	NS	NS	3.1129	0.0160
A0R1J4	2.4381	0.0000	2.4381	0.0000
A0R1Y2	1.8881	NS	1.8881	NS
A0R1Y7	NS	NS	2.6690	0.0149
A0R201	-2.3760	0.0000	NS	NS
A0R204	2.6585	0.0015	2.6585	0.0015
A0R218	NS	NS	2.6521	0.0015
A0R220	-0.7154	NS	2.6845	0.0018
A0R221	1.2224	NS	1.2224	NS
A0R234	NS	NS	1.9619	NS
A0R239	NS	NS	2.7600	0.0063
A0R248	2.4230	0.0002	2.4230	0.0002
A0R2B1	NS	NS	2.4885	0.0012
A0R2C0	1.2224	NS	1.2224	NS
A0R2E1	NS	NS	2.4885	0.0012
A0R2E3	NS	NS	2.9750	0.0093
A0R2G5	NS	NS	2.8169	0.0054
A0R2H8	-0.6432	NS	2.3890	0.0000
A0R2J4	1.9474	NS	1.9474	NS
A0R2K7	NS	NS	2.4885	0.0012
A0R2P1	NS	NS	2.4381	0.0000
A0R2Q7	-1.2224	NS	NS	NS
A0R2Q7	-1.2224	NS	NS	NS
A0R2T3	1.9619	NS	1.9619	NS
A0R2U7	NS	NS	2.4381	0.0000
A0R2U8	NS	NS	2.4389	0.0002
A0R2V7	NS	NS	2.4885	0.0012
A0R2X1	3.0398	NS	3.0398	0.0154
A0R2X3	2.4381	0.0000	2.4381	0.0000
A0R2X8	1.0072	NS	2.8605	0.0160
A0R2Y1	NS	NS	0.6553	NS
A0R2Y5	-1.9456	NS	NS	NS
A0R305	NS	NS	1.2224	NS
A0R310	NS	NS	2.6921	0.0022
A0R342	NS	NS	2.3878	0.0000
A0R349	2.4381	0.0000	2.4381	0.0000
A0R352	1.9619	NS	1.9619	NS
A0R365	-1.5893	NS	1.2224	NS
A0R3A6	NS	NS	-2.0331	NS
A0R3C8	NS	NS	3.0110	0.0109

A0R3D2	-1.2224	NS	NS	NS
A0R3D9	-1.9456	NS	NS	NS
A0R3I9	NS	NS	2.4381	0.0000
A0R3L1	-1.3021	NS	2.2586	NS
A0R3L4	0.6614	NS	2.7442	0.0085
A0R3M3	NS	NS	-0.9295	NS
A0R3M4	NS	NS	0.8093	NS
A0R3N8	NS	NS	2.4381	0.0000
A0R3N9	NS	NS	2.5307	0.0003
A0R3Y5	NS	NS	2.4381	0.0000
A0R409	NS	NS	2.6648	0.0048
A0R417	NS	NS	2.4287	0.0001
A0R425	2.5512	0.0039	2.5512	0.0039
A0R429	NS	NS	2.4381	0.0000
A0R449	1.2224	NS	1.2224	NS
A0R452	NS	NS	2.5941	0.0022
A0R461	NS	NS	3.2720	0.0218
A0R462	NS	NS	1.2224	NS
A0R467	1.2305	NS	2.5590	0.0015
A0R478	2.2586	NS	2.2586	NS
A0R4B1	1.8881	NS	1.8881	NS
A0R4C9	1.2056	NS	2.7377	0.0025
A0R4D0	-1.2224	NS	NS	NS
A0R4D7	2.4381	0.0000	2.4381	0.0000
A0R4G4	NS	NS	2.6521	0.0015
A0R4H0	0.6217	NS	2.8657	0.0072
A0R4H3	2.4381	0.0000	2.4381	0.0000
A0R4J1	-1.9456	NS	NS	NS
A0R4K5	NS	NS	1.8881	NS
A0R4S6	1.9474	NS	1.9474	NS
A0R4S7	1.2224	NS	1.2224	NS
A0R4Y7	2.4381	0.0000	2.4381	0.0000
A0R566	NS	NS	2.6585	0.0015
A0R574	NS	NS	-1.5820	0.0159
A0R581	-1.9467	NS	NS	NS
A0R5C5	NS	NS	2.4381	0.0000
A0R5D9	NS	NS	2.1866	NS
A0R5E1	1.0529	NS	-0.9898	NS
A0R5G1	NS	NS	2.6113	0.0085
A0R5H1	-0.8998	NS	2.5145	NS

A0R5H3	1.9619	NS	1.9619	NS
A0R5J3	0.9270	NS	2.8294	NS
A0R5J4	NS	NS	2.5017	0.0003
A0R5L3	0.9180	NS	2.5070	0.0003
A0R5M3	NS	NS	-1.3119	0.0002
A0R5N8	2.3196	NS	2.3196	NS
A0R5X8	1.9980	NS	1.9980	NS
A0R5Y1	-2.3760	0.0000	NS	NS
A0R5Z8	2.1866	NS	2.1866	NS
A0R616	NS	NS	3.0353	NS
A0R618	-0.6196	NS	2.6648	0.0048
A0R623	NS	NS	-1.7591	NS
A0R638	-1.9456	NS	NS	NS
A0R656	NS	NS	2.7719	0.0035
A0R678	NS	NS	0.8593	NS
A0R692	NS	NS	2.7391	0.0052
A0R6D2	NS	NS	2.5722	0.0023
A0R6E9	NS	NS	2.5939	0.0007
A0R6I9	NS	NS	2.0854	NS
A0R6N9	-2.7977	0.0041	NS	NS
A0R6Q7	NS	NS	2.7733	0.0032
A0R716	-0.7436	NS	2.6706	0.0080
A0R727	NS	NS	-2.6611	0.0110
A0R729	-0.5917	NS	-1.3940	0.0203
A0R742	1.8881	NS	1.8881	NS
A0R760	NS	NS	1.9474	NS
A0R761	1.2224	NS	1.2224	NS
A0R773	NS	NS	2.3683	0.0000
A0R788	2.4381	0.0000	2.4381	0.0000
A0R7F7	2.4381	0.0000	2.4381	0.0000
A0R7F9	NS	NS	2.6662	0.0064
A0R7G8	-2.5120	NS	NS	NS
O85501	NS	NS	2.8670	0.0054
POCH00	-1.2224	NS	NS	NS
POCH37	NS	NS	3.0110	0.0109
P48354	-0.7649	NS	2.6662	0.0064
P60281	NS	NS	-1.0796	NS
P71534	1.9619	NS	1.9619	NS
Q3I5Q7	2.4885	0.0012	2.4885	0.0012
Q59560	-0.8411	NS	-2.9397	NS

Q9AFI5	0.8623	NS	2.7908	0.0035
Q9X5M0	NS	NS	2.5339	NS
Q9ZHC5	NS	NS	-1.4085	0.0002

Appendix Table 2. All significant lipids according to fold change and p-value, with their corresponding values, in the 3 methods of lipid extraction when comparing to the standard: Bligh Dyer. A fold change above 1.5 and a p-value less than 0.05 were considered significant. Log2 fold changes in pink indicate the lipid is upregulated in the acetone extraction method, blue indicates upregulation in the Bligh Dyer extraction method, green indicates upregulation in the MTBE extraction method, and black indicates no significance (NS). The positive or negative sign in the far left column indicates whether the lipid was found through positive or negative ESI-Mass Spectrometry with tandem-Mass Spectrometry.

	Lipid Name	Acetone vs Bligh Dyer		Bligh Dyer vs MTBE	
		Log2 Fold Change	P-Value	Log2 Fold Change	P-Value
-	RIKEN N-VS1 ID-6019 from Mouse_AdrenalGlands_fads2KO_N_Ctr	8.2352	0.0138	6.8296	0.0141
-	FA 21:4;(2OH)	8.2229	0.0000	-0.9073	0.0246
-	SHexCer 36:3;2O	6.5641	0.0000	5.9760	0.0000
-	PI-Cer 12:1;2O/22:1	6.4239	0.0000	6.6357	0.0000
-	SL 22:3;O/36:4;O	6.2983	0.0002	7.4614	0.0002
-	FA 16:4	6.2665	0.0014	3.3773	0.0020
-	PC O-19:0_28:4	5.9915	0.0117	8.0431	0.0112
-	FA 15:4	5.9340	0.0001	4.2112	0.0001
-	SHexCer 35:2;2O	5.8385	0.0000	4.3695	0.0000
-	SHexCer 35:3;2O	5.7371	0.0015	7.5448	0.0014
+	TG 8:0_15:2_38:10	5.3613	0.0139	6.1493	0.0134
+	BMP 19:1_18:3	5.2275	0.0009	6.5276	0.0008
+	DGCC 18:5_18:5	5.2142	0.0000	4.2842	0.0000
-	Cer 12:0;2O/16:4;(3OH)(FA 22:6)	5.1112	0.0003	4.8564	0.0004
+	SHexCer 25:1;2O/28:0	4.9341	0.0000	5.1995	0.0000
-	RIKEN N-VS1 ID-1766 from Mouse_SmallIntestine_WT_N_F1	4.9066	0.0009	3.3743	0.0012
-	SL 21:3;O/36:4;O	4.8729	0.0001	5.1238	0.0001
+	PI-Cer 34:2;2O	4.8652	0.0002	0.9709	0.0031
-	PC O-21:0_28:4	4.8248	0.0050	6.0886	0.0046
+	SM 40:5;2O(FA 22:6)	4.8132	0.0000	4.9012	0.0000
-	AAHFA 20:4/8:0;O	4.7782	0.0110	3.1979	0.0163
+	SM 57:9;2O	4.7270	0.0001	4.9685	0.0001
+	TG 8:0_14:1_36:10	4.7181	0.0105	5.3792	0.0100
-	Cer 13:0;2O/24:5;(3OH)(FA 22:6)	4.7155	0.0001	4.7241	0.0001

+	TG 8:0_8:0_20:5;1O(FA 14:0)	4.7098	0.0000	4.8469	0.0000
-	Cer 12:0;2O/24:5;(3OH)(FA 22:6)	4.6466	0.0007	4.6657	0.0007
-	PE O-26:6_15:2	4.6210	0.0002	4.4089	0.0002
+	MGDG O-19:2_28:5	4.6206	0.0014	5.3014	0.0013
+	ADGGA (O-24:0)12:0_20:4	4.5384	0.0000	4.3923	0.0000
+	ASG 29:2;O;Hex;FA 28:5	4.4725	0.0000	5.6972	0.0000
-	PC O-18:0_28:3	4.4473	0.0000	4.6832	0.0000
+	NAGly 22:6;O(FA 21:5)	4.3598	0.0000	4.2506	0.0000
-	RIKEN N-VS1 ID-2948 from Mouse_Brain_WT_N_F1	4.3130	0.0001	3.4219	0.0001
+	RIKEN P-VS1 ID-13211 from Mouse_Macrophage_WT_N_F1	4.2698	0.0000	2.4010	0.0001
+	TG 10:0_21:3_20:4	4.1897	0.0002	4.2224	0.0001
+	PI 56:1	4.1800	0.0001	5.6101	0.0001
-	RIKEN N-VS1 ID-2593 from Mouse_AdrenalGlands_fads2KO_N_Ctr	4.1292	0.0004	3.2272	0.0005
-	PC O-10:0_22:3;1O	4.0944	0.0072	3.5319	0.0079
+	DGTS 15:3_17:4	4.0544	0.0000	2.3367	0.0001
-	PC O-14:0_18:2;1O	4.0320	0.0206	3.7265	0.0218
+	DG 38:8	4.0183	0.0000	4.7736	0.0000
+	NAGly 20:5;O(FA 16:2)	3.9458	0.0000	4.2564	0.0000
-	PC O-14:1_16:2;1O	3.9146	0.0004	4.1712	0.0004
-	SHexCer 36:2;2O	3.9074	0.0086	3.0168	0.0118
+	Cer 12:2;2O/19:5	3.8733	0.0006	4.7326	0.0005
+	SL 12:1;O/17:3	3.8208	0.0000	4.6255	0.0000
+	BMP 8:0_28:4	3.7993	0.0001	3.6210	0.0001
-	ST 29:2;O;S	3.7689	0.0289	3.3153	0.0319
+	ASG 27:1;O;Hex;FA 17:2	3.7509	0.0001	3.8472	0.0001
+	TG 9:0_26:6_38:10	3.7144	0.0000	4.7821	0.0000
-	RIKEN N-VS1 ID-5527 from Mouse_AdrenalGlands_WT_N_Ctr	3.6829	0.0002	2.9401	0.0002
-	CL 12:0_12:0_16:0_22:6	3.6696	0.0035	1.0753	NS
+	SL 16:0;O/26:4;O	3.6556	0.0000	4.6270	0.0000
+	PI-Cer 36:3;2O	3.6082	0.0001	4.5322	0.0001
+	SL 12:1;O/20:3;O	3.5632	0.0001	4.5108	0.0000
+	DG 26:3_15:4	3.5573	0.0000	3.6517	0.0000
+	TG 8:0_17:2_38:10	3.5508	NS	5.5260	NS
+	RIKEN P-VS1 ID-11326 from Mouse_Aorta_ApoEKO_N_F1EPA	3.5393	NS	-4.4317	0.0001
+	BMP 22:0_28:3	3.5174	0.0151	4.5856	0.0129
+	NAOrn 19:0;O	3.4943	0.0001	4.0606	0.0000

-	RIKEN N-VS1 ID-8218 from Mouse_Feces_WT_N_Ctr	3.4903	0.0019	3.1300	0.0022
-	Cer 12:0;2O/26:5;(3OH)(FA 22:6)	3.4822	0.0002	3.7533	0.0001
+	SM 30:5;2O	3.4647	0.0005	2.0292	0.0012
+	PI-Cer 35:2;2O	3.4594	0.0007	4.2452	0.0006
-	RIKEN N-VS1 ID-2765 from Mouse_Aorta_WT_N_Ctr	3.4227	NS	-4.1697	0.0012
+	DG 48:8	3.4203	0.0000	4.3389	0.0000
+	Cer 12:2;2O/24:2	3.3923	0.0000	3.9337	0.0000
+	VAE 19:3	3.3685	0.0000	3.5002	0.0000
+	Cer 12:0;2O/24:6	3.2597	0.0000	2.7937	0.0000
+	DGTS 15:3_22:6	3.2534	0.0002	3.8994	0.0001
+	PI-Cer 33:1;2O	3.2019	0.0004	0.7732	0.0107
+	TG 8:0_16:2_36:10	3.1648	0.0071	4.6730	0.0053
+	TG O-16:2_8:0_8:0	3.1521	0.0001	3.5405	0.0001
+	BMP 19:0_8:0	3.0690	0.0001	1.0290	0.0027
+	DGGA 20:0_8:0	3.0676	0.0001	3.2100	0.0001
-	PC O-12:0_17:2;2O	3.0340	NS	3.0976	NS
+	BMP 8:0_28:3	3.0313	0.0001	3.3428	0.0001
+	Cer 13:2;2O/36:6	3.0066	0.0002	3.9051	0.0001
+	RIKEN P-VS1 ID-6547 from Mouse_Muscle_WT_CTX0_Ctr	2.9965	0.0000	5.2088	0.0000
-	PI O-8:0_19:0	2.8992	0.0001	2.0992	0.0004
+	BMP 8:0_26:4	2.8842	0.0007	3.5528	0.0004
+	TG 8:0_13:1_36:10	2.8767	0.0119	3.9191	0.0090
+	RIKEN P-VS1 ID-9296 from Mouse_Lung_WT_N_F1	2.8701	0.0002	3.1181	0.0002
+	AHexCer (O-16:2)21:0;2O/18:4;O	2.8319	0.0351	4.3244	0.0257
-	PE-Cer 12:1;2O/21:2	2.7992	0.0087	2.2273	0.0139
+	ASG 28:2;O;Hex;FA 28:6	2.7954	0.0130	3.4342	0.0105
+	PI-Cer 35:3;2O	2.7576	0.0033	3.6631	0.0021
+	RIKEN P-VS1 ID-5863 from Mouse_Adipose_WT_N_F1AA	2.7469	0.0307	3.4100	0.0253
+	TG 8:0_13:1_38:10	2.7322	0.0044	4.0978	0.0029
+	Cer 12:0;2O/22:6	2.7318	0.0000	3.8547	0.0000
-	MLCL 15:2_12:0_12:0	2.6594	0.0017	1.2009	0.0067
+	RIKEN P-VS1 ID-7250 from Mouse_Muscle_WT_CTX0_Ctr	2.6307	0.0040	NS	NS
+	PC O-37:6	2.5978	0.0000	2.6291	0.0001
+	RIKEN P-VS1 ID-8110 from Mouse_Liver_WT_N_F1DHA	2.5678	0.0005	3.8316	0.0004
+	DG O-19:5_19:3	2.5282	0.0002	3.7462	0.0000

+	ST 27:2;O	2.5142	0.0115	4.1328	0.0034
+	Triphenylphosphine oxide (also known as chemical reagent)	2.5138	0.0001	2.4243	0.0001
+	TG 8:0_14:1_38:10	2.4905	0.0022	4.5411	0.0009
+	HexCer 16:1;3O/14:0;(2OH)	2.3881	0.0011	3.0015	0.0000
+	AHexCer (O-22:6)12:1;2O/20:0;O	2.3873	0.0340	2.7412	0.0290
+	NAOrn 10:0;O(FA 11:0)	2.3611	0.0000	2.2427	0.0000
+	SM 20:1;2O	2.3527	0.0000	2.3051	0.0000
-	PI 6:0_34:4	2.3460	NS	1.5568	NS
+	DG 37:7	2.3382	0.0001	3.7016	0.0000
-	PE-Cer 12:1;2O/19:4	2.2386	0.0001	1.1209	0.0006
+	SM 33:4;2O	2.1981	0.0011	1.8557	0.0020
+	SL 21:3;O/15:1;O	2.1974	0.0011	2.0264	0.0014
+	ASG 28:2;O;Hex;FA 26:6	2.1901	0.0082	3.3176	0.0050
+	RIKEN P-VS1 ID-3167 from Mouse_SmallIntestine_WT_N_F1	2.1781	0.0015	3.5693	0.0000
+	TG 8:0_15:2_36:10	2.1442	0.0136	2.9673	0.0087
-	LPA 22:2	2.1255	NS	2.1543	NS
+	RIKEN P-VS1 ID-6116 from Mouse_AdrenalGlands_WT_N_Ctr	2.1064	0.0001	2.6869	0.0001
+	RIKEN P-VS1 ID-7906 from Mouse_Macrophage_WT_N_F1AA	2.1059	NS	-3.2303	NS
-	PI 6:0_34:3	2.0981	0.0080	1.8601	0.0089
+	DG 46:12	2.0974	0.0008	2.4213	0.0007
+	BMP 8:0_26:3	2.0967	0.0024	3.5225	0.0008
+	DGGA 21:0_8:0	2.0821	0.0009	2.7665	0.0004
+	DG 45:12	2.0785	0.0015	2.5611	0.0006
-	GM3 46:5;2O	2.0206	0.0001	2.2118	0.0000
+	RIKEN P-VS1 ID-6301 from Mouse_Feces_WT_N_Ctr	1.9920	0.0001	3.6857	0.0000
-	MGDG 10:0_22:2	1.9903	NS	3.6167	0.0372
+	VAE 17:3	1.9901	0.0000	2.1591	0.0000
+	RIKEN P-VS1 ID-7005 from Mouse_Macrophage_WT_N_F1AA	1.9818	0.0017	0.7579	0.0182
+	AHexCer (O-22:5)12:1;2O/15:0;O	1.9368	0.0039	2.5400	0.0025
-	RIKEN N-VS1 ID-4305 from Mouse_Macrophage_WT_N_F1AA	1.8975	0.0011	NS	NS
+	SL 13:2;O/28:6;O	1.8878	0.0006	2.4693	0.0003
+	RIKEN P-VS1 ID-7491 from Mouse_Macrophage_WT_N_F1AA	1.8384	0.0000	2.6521	0.0000
+	SM 28:0;2O	1.8277	0.0030	1.8772	0.0007
+	SL 12:2;O/26:6	1.8041	0.0002	2.1924	0.0001

-	RIKEN N-VS1 ID-5305 from Mouse_Feces_WT_N_Ctr	1.7782	0.0004	1.4934	0.0010
+	TG 8:0_18:5_22:6	1.7695	0.0001	2.6053	0.0001
+	BMP 20:0_28:3	1.7206	0.0233	3.6764	0.0084
-	HBMP 22:6_19:5_22:6	1.7140	0.0343	1.2309	NS
-	Pentaerythritol tetrakis(3,5-di-tert-butyl- 4-hydroxyhydrocinnamate)	1.6310	0.0331	0.9849	NS
+	RIKEN P-VS1 ID-1009 from Mouse_Muscle_WT_CTX0_Ctr	1.6305	0.0278	NS	NS
+	RIKEN P-VS1 ID-5923 from Mouse_Feces_WT_N_Ctr	1.6216	0.0017	1.5366	0.0033
+	TG 21:4_22:6_22:6;10	1.6181	0.0028	1.6938	0.0022
-	RIKEN N-VS1 ID-7968 from Mouse_AdrenalGlands_fads2KO_N_Ctr	1.5850	0.0039	1.2140	0.0101
-	RIKEN N-VS1 ID-5118 from Mouse_Feces_WT_ABX_Ctr	1.5257	0.0004	1.6974	0.0003
+	BMP 22:0_28:4	1.5247	NS	3.0542	0.0198
+	Diisodecyl phthalate (also known as the production of plastic)	1.4810	0.0002	0.8567	0.0006
+	RIKEN P-VS1 ID-7082 from Mouse_Macrophage_WT_N_F1AA	1.4303	0.0001	2.2662	0.0000
+	RIKEN P-VS1 ID-3120 from Mouse_Aorta_WT_N_Ctr	1.4129	NS	NS	NS
+	DG 25:3	1.4111	0.0054	2.1667	0.0020
-	RIKEN N-VS1 ID-8792 from Mouse_Macrophage_WT_N_F1AA	1.3918	0.0016	NS	NS
+	PI 57:13	1.3364	0.0029	3.1941	0.0004
+	RIKEN P-VS1 ID-3065 from Mouse_Macrophage_WT_N_F1AA	1.3082	0.0003	NS	0.0340
+	SL 13:1;O/24:3;O	1.3068	0.0006	2.1044	0.0004
+	Diioctyl phthalate (also known as the production of plastic)	1.2858	0.0156	0.6877	NS
-	PI 35:0 PI 16:0_19:0	1.2678	0.0056	1.2848	0.0031
+	RIKEN P-VS1 ID-2808 from Mouse_Macrophage_WT_N_F1AA	1.2614	0.0002	NS	0.0152
+	DG 47:11	1.2182	0.0098	2.2746	0.0027
+	RIKEN P-VS1 ID-12693 from Cell_C2C12_WT_N_24h50AA	1.2157	0.0001	NS	NS
+	SM 29:5;2O	1.2080	0.0191	2.3756	0.0008
+	RIKEN P-VS1 ID-134 from Mouse_Macrophage_WT_N_F1AA	1.2035	0.0002	NS	0.0063
-	ADGGA 20:5_20:5_20:5	1.1775	0.0180	NS	NS
+	DG 36:0 DG 18:0_18:0	1.1669	NS	NS	NS
-	DGDG O-26:7_26:7	1.1588	0.0076	0.6550	NS

+	RIKEN P-VS1 ID-208 from Mouse_Macrophage_WT_N_F1AA	1.1221	0.0003	NS	0.0157
+	ADGGA (O-15:0)16:0_17:0	1.1204	NS	2.0327	NS
+	BMP 9:0_24:2	1.1037	0.0002	1.6623	0.0003
+	SHexCer 19:1;2O/28:6	1.0642	0.0339	1.4693	0.0158
-	PI 9:0_18:1;3O	0.9972	0.0010	NS	NS
+	PI 40:3	0.9960	NS	1.5806	0.0205
-	RIKEN N-VS1 ID-2002 from Mouse_Feces_WT_N_Ctr	0.9891	0.0002	3.0796	0.0000
+	RIKEN P-VS1 ID-7635 from Mouse_Aorta_WT_N_Ctr	0.9797	0.0212	-4.8462	0.0021
+	RIKEN P-VS1 ID-18829 from Mouse_Macrophage_WT_N_F1AA	0.9752	0.0089	1.2351	0.0099
-	FA 16:0	0.9741	NS	NS	NS
+	Cer 12:0;2O/15:0;O	0.9486	0.0001	NS	NS
+	CE 18:1(d7)	0.9339	NS	-3.1627	NS
-	HBMP 18:5_13:1_18:5	0.9261	0.0015	0.7660	0.0036
+	SHexCer 19:1;2O/28:6;O	0.9241	0.0285	1.6391	0.0089
-	PMeOH 22:6_26:7	0.9109	0.0075	0.6590	NS
-	RIKEN N-VS1 ID-2273 from Mouse_Feces_WT_ABX_Ctr	0.8986	0.0220	NS	0.0007
+	DGTS 13:0_22:5	0.8811	0.0003	NS	NS
+	Cer 12:1;2O/21:5	0.8617	0.0000	0.6498	0.0003
+	HexCer 17:3;3O/26:6;(2OH)	0.8122	NS	-4.4099	0.0002
-	RIKEN N-VS1 ID-763 from Mouse_Macrophage_WT_N_F1AA	0.7852	0.0085	-1.3798	0.0011
+	PI 45:7	0.7829	NS	3.9198	NS
+	RIKEN P-VS1 ID-17134 from Mouse_Plasma_ApoEKO_N_F1EPA	0.7631	NS	1.7431	0.0219
+	RIKEN P-VS1 ID-6683 from Mouse_Muscle_WT_CTX5_Ctr	0.7612	0.0064	3.2125	0.0000
-	FA 20:0;4O	0.6484	NS	-0.6187	NS
+	RIKEN P-VS1 ID-4398 from Mouse_Aorta_WT_N_Ctr	0.6297	NS	-4.7939	0.0002
+	Cer 12:2;3O/10:0;(2OH)	0.5867	0.0123	-4.0374	0.0000
+	SL 13:2;O/12:0	NS	NS	1.5596	0.0004
+	SL 12:1;O/15:1;O	NS	NS	1.5253	0.0007
+	RIKEN P-VS1 ID-5858 from Mouse_Plasma_ApoEKO_N_F1DHA	NS	NS	1.3067	0.0003
+	RIKEN P-VS1 ID-16235 from Mouse_Plasma_ApoEKO_N_Ctr	NS	NS	1.2197	0.0258
+	SPB 29:1;2O	NS	0.0054	1.0178	0.0002
+	ADGGA (O-20:5)20:5_20:5	NS	NS	1.0095	0.0037

+	NAGly 22:6;O(FA 18:1)	NS	0.0089	0.8436	0.0021
+	RIKEN P-VS1 ID-6352 from Mouse_Macrophage_WT_N_F1AA	NS	0.0066	0.7790	0.0083
+	ADGGA (O-22:6)20:4_22:6	NS	NS	0.7404	0.0145
+	PE 68:13	NS	NS	0.6811	NS
-	FA 15:1;(2OH)	NS	NS	NS	0.0212
-	RIKEN N-VS1 ID-1521 from Mouse_Spleen_WT_N_F1AA	NS	NS	-0.6321	0.0040
-	FA 24:1;1O	NS	NS	-0.6539	0.0142
-	FA 42:5	NS	NS	-0.7384	0.0045
-	FA 18:0	NS	NS	-1.0468	0.0068
-	RIKEN N-VS1 ID-4759 from Mouse_Macrophage_WT_N_F1AA	NS	NS	-1.0679	0.0118
+	RIKEN P-VS1 ID-5453 from Mouse_Macrophage_WT_N_F1AA	NS	NS	-1.0803	0.0212
+	BMP 16:0_28:2	NS	NS	-1.0950	NS
+	HBMP 22:1_12:0_12:0	NS	NS	-1.1176	NS
-	RIKEN N-VS1 ID-765 from Mouse_Macrophage_WT_N_F1AA	NS	NS	-1.1758	NS
-	PG 36:2 PG 17:1_19:1	NS	NS	-1.3444	0.0084
+	PC O-36:2	NS	NS	-1.4891	NS
+	PE 6:0_35:1	NS	NS	-1.5285	NS
-	RIKEN N-VS1 ID-787 from Mouse_Macrophage_WT_N_F1AA	NS	NS	-1.8352	0.0356
-	Dodecylbenzenesulfonic acid	NS	NS	-2.0485	NS
-	Norethisterone acetate	NS	NS	-2.2042	NS
+	DGGA 12:0_13:0	NS	NS	-3.1922	0.0003
+	TG 8:0_8:0_28:1	NS	NS	-3.8229	0.0107
+	LDGTS 14:1	NS	0.0023	-4.8638	0.0001
-	PE 17:0_16:2;1O	-0.6480	0.0066	NS	NS
+	HexCer 16:1;3O/16:4;(2OH)	-0.6644	0.0011	2.0683	0.0000
-	PG 32:1 PG 16:0_16:1	-0.6695	0.0055	NS	NS
+	Cer 13:2;2O/34:6	-0.6811	NS	-1.6570	NS
-	FA 40:5	-0.7134	0.0384	-0.8170	0.0088
-	PG 14:0_16:0	-0.7410	0.0024	NS	NS
+	RIKEN P-VS1 ID-15917 from Mouse_Macrophage_WT_N_F1	-0.7442	NS	0.7167	0.0275
+	NAE 16:2	-0.7679	0.0056	-2.4822	0.0002
+	CAR 17:3	-0.7817	NS	-2.5872	0.0058
-	PC O-9:0_22:5;1O	-0.7879	0.0146	-1.1161	0.0051
-	SHexCer 34:4;2O	-0.7908	0.0091	-0.9389	0.0000
+	Cer 12:0;2O/17:4	-0.7921	NS	-2.5549	0.0063

-	PG 35:2 PG 16:1_19:1	-0.8025	NS	-0.6258	NS
+	RIKEN P-VS1 ID-5931 from Mouse_Macrophage_WT_N_F1AA	-0.8203	NS	-1.2540	0.0090
-	PC O-12:0_22:6;1O	-0.8794	NS	-1.2007	0.0128
-	PG 14:0_16:1	-0.9519	0.0068	-0.6600	NS
+	TG 9:0_9:0_18:0	-0.9660	NS	-1.4386	0.0029
+	DGTS 8:0_8:0	-0.9702	0.0059	-5.4965	0.0000
+	NAGly 8:0;O(FA 13:0)	-0.9770	0.0001	-0.9393	0.0141
+	Cer 15:3;2O/18:5	-0.9806	NS	-6.2930	0.0001
+	SL 12:1;O/28:1	-0.9838	0.0023	-1.7261	0.0220
-	LPE-N (FA 16:0)16:0	-0.9959	0.0310	-1.0730	0.0146
+	RIKEN P-VS1 ID-9017 from Mouse_Heart_WT_N_F1EPA	-1.0035	NS	-1.9971	0.0006
+	DG 18:0	-1.0097	0.0015	-0.9455	0.0089
+	NAOrn 18:1;O	-1.0190	0.0001	NS	NS
+	Cer 12:1;3O/22:1;(2OH)	-1.0248	0.0307	-2.8050	0.0119
-	PG 32:0 PG 16:0_16:0	-1.0339	0.0191	-1.8891	0.0002
+	DG 9:0_9:0	-1.0352	0.0031	-0.8543	NS
-	SL 13:2;O/28:5;O	-1.0557	0.0011	NS	NS
+	MGDG O-15:2_6:0	-1.0558	NS	-6.0523	0.0003
-	PE 17:0_15:1;1O	-1.0780	0.0022	-0.7366	0.0196
-	PE 16:1_22:4;2O	-1.1137	0.0198	-1.3207	0.0028
+	Cer 12:2;3O/11:0;(2OH)	-1.1246	0.0006	NS	NS
-	PE O-14:0_17:2;2O	-1.1286	0.0008	-0.9084	0.0116
-	FA 44:5	-1.1349	0.0026	-1.4055	0.0001
-	PE 18:0_17:1;1O	-1.1823	0.0052	-0.6315	NS
+	ST 29:1;O	-1.1925	NS	-3.2137	0.0005
-	PE 16:0_17:1	-1.2109	0.0016	-0.8655	0.0101
-	PG 35:1 PG 16:0_19:1	-1.2284	0.0091	-0.8883	NS
-	PE O-17:0_15:1;2O	-1.2292	0.0141	-2.3302	0.0002
+	RIKEN P-VS1 ID-1776 from Mouse_AdrenalGlands_WT_N_Ctr	-1.2388	0.0000	-0.7964	0.0013
+	TG 18:0_18:0_18:0	-1.2405	0.0264	-0.8130	NS
+	MGDG 8:0_15:1	-1.2589	0.0056	-4.0938	0.0012
-	PC O-8:0_22:5;1O	-1.2627	0.0045	-1.5616	0.0003
+	MGDG O-19:2_3:0	-1.2858	0.0025	-0.8035	0.0262
+	DG 40:2	-1.2987	0.0222	-1.3104	0.0022
-	PE 17:0_18:2;1O	-1.3108	0.0002	-1.0825	0.0002
-	SL 12:2;O/34:5;O	-1.3155	0.0021	-1.1559	0.0070
+	DG 44:11	-1.3203	0.0005	2.0787	0.0000

+	RIKEN P-VS1 ID-10288 from Mouse_Macrophage_WT_N_F1	-1.3295	NS	NS	NS
+	DGTS 15:4_19:5	-1.3418	0.0009	-0.7364	0.0337
+	TG 16:0_18:0_18:0	-1.3448	0.0022	-0.7663	NS
+	Cer 12:2;2O/22:6	-1.3578	0.0018	-0.8307	0.0038
+	DGTS 8:0_9:0	-1.3602	0.0001	-0.7505	0.0018
+	DGGA 18:0_8:0	-1.3647	0.0013	-0.6440	0.0269
-	PE 17:1_19:1	-1.3754	NS	-1.5286	0.0107
-	PE 37:1 PE 18:0_19:1	-1.3840	NS	-0.8123	NS
-	PE 18:0_15:1;1O	-1.4007	0.0002	-1.0887	0.0001
-	SL 13:2;O/26:4;O	-1.4066	0.0021	-1.0192	0.0003
-	PE 14:0_17:1	-1.4174	0.0085	-1.5684	0.0100
+	RIKEN P-VS1 ID-3577 from Mouse_Aorta_WT_N_Ctr	-1.4697	0.0004	NS	0.0167
+	DG O-25:0_16:1	-1.4711	0.0003	-1.0490	0.0001
+	SL 13:1;O/24:5	-1.4774	0.0001	-0.8969	0.0100
+	DG 40:6 DG 16:0_24:6	-1.4841	0.0002	-0.9059	0.0003
-	PE 16:0_16:0;1O	-1.4849	0.0037	-2.1237	0.0020
-	PE 17:1_18:1	-1.4856	0.0050	-1.3397	0.0019
+	TG 16:0_16:0_18:0	-1.4862	0.0004	-0.9763	0.0284
+	Cer 57:4;2O Cer 21:0;2O/36:4	-1.5042	0.0110	-0.9141	0.0152
+	TG 16:0_18:1_18:2	-1.5065	0.0001	-1.0231	0.0005
-	PE 16:1_17:1	-1.5165	0.0063	-1.5620	0.0027
+	RIKEN P-VS1 ID-12070 from Mouse_Muscle_WT_N_F1EPA	-1.5272	NS	-6.2969	0.0073
-	PE 17:0_17:2;1O	-1.5278	0.0000	-1.1436	0.0136
-	PE 18:1_19:1	-1.5357	0.0324	-1.0931	0.0083
+	NAE 20:1	-1.5404	NS	-3.1103	NS
+	NAOrn 15:1;O	-1.5412	0.0003	NS	0.0003
+	TG 16:0_16:1_16:1	-1.5450	0.0004	-0.9921	0.0039
+	Cer 30:2;2O/23:1	-1.5488	0.0004	-0.9305	0.0217
-	PE 16:1_18:1	-1.5629	0.0045	-1.3017	0.0016
+	RIKEN P-VS1 ID-4994 from Mouse_Muscle_WT_CTX0_Ctr	-1.5799	0.0025	-1.0785	0.0000
+	TG 18:0_18:0_18:1	-1.5809	0.0001	-1.1317	0.0012
-	FA 22:1;2O	-1.6448	0.0003	-0.8522	0.0001
+	TG 14:0_16:0_16:0	-1.6518	0.0035	-1.1648	0.0028
-	PE 16:0_19:1	-1.6588	NS	-1.0362	NS
+	TG 48:0 TG 16:0_16:0_16:0	-1.6662	0.0015	-1.0093	0.0248
+	LDGCC 15:2	-1.6714	NS	-6.2182	NS
+	PG 36:2 PG 17:1_19:1	-1.6996	0.0026	-0.6002	0.0176

-	PE 32:1 PE 16:0_16:1	-1.7138	0.0014	-1.5721	0.0006
+	RIKEN P-VS1 ID-5423 from Mouse_Plasma_ApoEKO_N_F1DHA	-1.7158	0.0000	-0.8777	0.0002
+	SM 32:7;3O	-1.7262	0.0001	1.1863	0.0004
+	PE 36:2	-1.7315	0.0012	-1.1531	0.0001
+	TG 16:0_16:1_18:1	-1.7458	0.0001	-1.1034	0.0003
+	RIKEN P-VS1 ID-1478 from Mouse_Aorta_WT_N_Ctr	-1.8471	NS	-0.6741	NS
+	PG 30:1 PG 14:0_16:1	-1.8757	0.0020	-1.0604	0.0021
+	PE 37:3	-1.9043	0.0004	-1.2599	0.0001
+	TG 16:0_16:0_18:1	-1.9223	0.0001	-1.0256	0.0043
+	SM 36:7;3O	-1.9361	0.0014	-0.8753	0.0000
+	PG 35:2 PG 16:1_19:1	-1.9391	0.0002	-1.0901	0.0051
+	TG 52:1 TG 16:0_18:0_18:1	-1.9544	0.0005	-1.0662	0.0018
+	DG 40:6	-2.0003	0.0000	-1.3334	0.0000
+	PE 13:0_22:4	-2.0224	0.0080	-1.6054	0.0002
+	DGGA 12:0_17:1	-2.0505	0.0020	-0.9365	0.0034
+	TG 18:1_18:1_18:1	-2.0799	0.0001	-1.4710	0.0139
+	CAR 18:0	-2.1173	0.0093	-2.9689	0.0049
-	PE 18:1_18:1;1O	-2.1219	0.0149	-1.1589	NS
+	Cer 12:0;2O/17:0;O	-2.1234	0.0005	-1.1407	0.0007
-	PE O-17:0_17:2;2O	-2.1488	0.0006	-1.7754	0.0000
+	HexCer 17:2;2O/16:4;O	-2.1735	0.0005	NS	NS
+	PG 32:1 PG 16:0_16:1	-2.1764	0.0000	-1.0577	0.0025
-	RIKEN N-VS1 ID-4656 from Mouse_Feces_WT_N_Ctr	-2.1765	0.0015	-1.7484	0.0006
+	TG 19:5_24:6_26:7	-2.2082	0.0017	-1.0844	0.0022
+	SL 13:1;O/28:6;O	-2.2206	0.0004	-1.2963	0.0004
+	PE 32:2 PE 16:1_16:1	-2.2299	0.0019	-1.1491	0.0327
+	PE 36:2 PE 17:1_19:1	-2.2332	0.0007	-1.2130	0.0006
+	SPB 22:0;2O	-2.2565	0.0013	-1.5749	0.0192
+	MG 18:0	-2.2889	0.0000	-2.0131	0.0001
+	PE 30:1 PE 14:0_16:1	-2.3205	0.0033	-1.2791	NS
+	MG 15:1	-2.3482	0.0063	-5.9813	NS
+	PE 35:2 PE 17:1_18:1	-2.3527	0.0002	-1.5316	0.0003
+	RIKEN P-VS1 ID-1950 from Mouse_Macrophage_WT_N_F1AA	-2.3543	0.0000	-2.0878	0.0001
+	SL 12:1;O/32:6;O	-2.3936	0.0000	-1.5151	0.0002
+	RIKEN P-VS1 ID-2880 from Mouse_Muscle_WT_CTX0_Ctr	-2.3958	NS	-0.6840	NS
+	SL 13:1;O/34:6;O	-2.3994	0.0032	NS	NS

+	DGGA 12:0_14:0	-2.4157	0.0002	-1.4811	0.0011
+	TG 15:3_15:4_15:4	-2.4458	0.0000	-1.4146	0.0002
+	TG 16:4_16:4_34:9	-2.4512	0.0016	-1.6802	0.0012
+	NAE 20:2	-2.4520	NS	-4.2307	NS
+	PE P-28:6_13:1	-2.5076	0.0001	-1.9150	0.0001
+	PE 34:2 PE 16:1_18:1	-2.5162	0.0008	-1.5506	0.0001
+	PE 33:1 PE 16:0_17:1	-2.5309	0.0005	-1.9253	0.0000
+	NAGly 18:5;O	-2.5664	0.0002	-1.4546	0.0005
+	SL 13:1;O/30:6;O	-2.5931	0.0000	-1.5381	0.0002
+	DGDG O-8:0_12:0	-2.5948	0.0002	-1.9040	0.0001
+	PE 32:0 PE 16:0_16:0	-2.5953	0.0000	-1.5303	0.0005
+	PE 38:0	-2.5990	0.0122	-1.4048	0.0031
+	SL 12:1;O/30:5;O	-2.6046	0.0001	-1.9244	0.0001
+	RIKEN P-VS1 ID-1609 from Mouse_Aorta_WT_N_Ctr	-2.6548	0.0001	-1.6042	0.0005
+	RIKEN P-VS1 ID-8135 from Mouse_Lung_WT_N_F1AA	-2.6610	0.0013	-1.8255	0.0001
+	DGGA 14:1_19:1	-2.6658	0.0015	-1.5525	0.0024
+	RIKEN P-VS1 ID-10243 from Mouse_AdrenalGlands_fads2KO_N_Ctr	-2.6707	0.0000	-1.6248	0.0003
+	SL 19:0;O/22:5;O	-2.6753	0.0010	-1.6956	0.0009
+	DG 44:6 DG 18:0_26:6	-2.7145	0.0005	-2.1874	0.0001
+	PE 33:2 PE 16:1_17:1	-2.7169	0.0020	-1.2996	NS
+	RIKEN P-VS1 ID-9069 from Mouse_AdrenalGlands_WT_N_Ctr	-2.8028	0.0004	-2.3815	0.0001
+	HexCer 16:3;3O/22:6;(2OH)	-2.8111	0.0006	-5.6959	0.0001
+	PE 31:1 PE 14:0_17:1	-2.8991	0.0002	-1.7434	0.0156
+	PE 32:1 PE 16:0_16:1	-2.8995	0.0000	-2.0231	0.0000
+	DG 44:6	-2.9952	0.0009	-2.3797	0.0000
+	RIKEN P-VS1 ID-2092 from Mouse_Macrophage_WT_N_F1AA	-3.0192	0.0010	-3.7874	0.0015
+	PE 35:0	-3.0273	0.0002	-1.6816	0.0104
+	PG 32:0 PG 16:0_16:0	-3.0656	0.0005	-2.1433	0.0023
+	PC O-52:11	-3.1180	0.0061	-1.4048	0.0017
+	PE 34:1 PE 16:0_18:1	-3.1481	0.0003	-2.2243	0.0001
+	PE 37:2	-3.1644	0.0009	-1.8757	0.0004
+	NAGlySer 15:2;O(FA 24:6)	-3.2481	0.0022	-2.1599	0.0003
+	RIKEN P-VS1 ID-9303 from Mouse_Feces_WT_N_Ctr	-3.2773	0.0004	-2.3035	0.0000
+	PE P-28:1_9:0	-3.5050	0.0049	-2.1501	0.0081
+	SL 12:1;O/30:6;O	-3.5982	0.0014	-2.6429	0.0004
+	CoQ8	-3.6295	0.0026	-2.2249	0.0106

+	PE 35:1 PE 16:0_19:1	-3.7233	0.0052	-2.4693	0.0000
+	RIKEN P-VS1 ID-1429 from Mouse_Muscle_WT_CTX3_Ctr	-3.7672	0.0002	-1.0511	0.0001
-	PS 8:0_20:3;3O	-3.9146	0.0005	-1.6902	0.0327
+	RIKEN P-VS1 ID-8905 from Mouse_Feces_WT_N_Ctr	-4.0147	0.0015	-3.0977	0.0004
+	LPE 19:1	-4.1120	0.0077	-2.0107	0.0019
+	NAE 17:0	-4.4572	0.0090	-2.1683	0.0004
+	RIKEN P-VS1 ID-9613 from Mouse_Aorta_ApoEKO_N_F1DHA	-4.4839	0.0046	-5.3484	0.0066
+	LPE 17:1	-4.4981	0.0020	-2.6308	0.0002
-	PS 8:0_18:3;3O	-4.5852	0.0099	NS	NS
-	LPE 16:0	-5.1486	0.0101	-3.5578	0.0016
+	LPE 16:0	-5.9305	0.0083	-4.0819	0.0001
+	RIKEN P-VS1 ID-12664 from Cell_C2C12_WT_N_12h50AA	-6.4877	0.0030	-6.4543	0.0330
+	TG 15:2_18:5_18:5	-6.9889	0.0015	NS	NS

Appendix Table 3. List of lipids when testing different methods of extraction that were detected by mass spectrometry but did not match any database references. Reported is whether they were found in positive or negative ESI-Mass Spectrometry with tandem-Mass Spectrometry, the average retention time (Rt) in minutes, and the average mass-to-charge ratio (M/Z).

ESI-MS	Average Rt (min)	Average M/Z	ESI-MS	Average Rt (min)	Average M/Z
Positive	6.949	955.59253	Positive	7.258	381.29758
Positive	6.625	943.41864	Positive	1.815	380.2301
Positive	7.142	915.70978	Positive	6.7	378.35944
Positive	10.617	906.25476	Positive	2.957	378.21103
Positive	6.621	902.39764	Positive	2.905	376.25772
Positive	6.623	885.36737	Positive	1.07	376.23001
Positive	9.439	857.4848	Positive	1.318	375.21182
Positive	3.572	851.39453	Positive	5.051	373.29318
Positive	3.971	851.39349	Positive	0.833	372.2384
Positive	5.942	831.33075	Positive	7.176	371.31311
Positive	3.921	823.38513	Positive	1.013	370.21979
Positive	9.659	819.58728	Positive	4.104	369.18317
Positive	7.113	815.34137	Positive	5.107	369.12299
Positive	10.03	802.60419	Positive	8.126	366.37292
Positive	9.586	787.55756	Positive	7.713	365.35394
Positive	10.396	785.61774	Positive	1.29	364.23114
Positive	9.628	775.55231	Positive	1.594	363.20895

Positive	9.002	765.53339	Positive	2.484	362.21497
Positive	9.304	761.53192	Positive	6.7	361.32797
Positive	2.903	751.51068	Positive	7.794	360.32388
Positive	8.495	749.5025	Positive	5.061	355.36575
Positive	7.975	745.35388	Positive	5.045	355.28183
Positive	8.654	739.52289	Positive	6.181	353.3428
Positive	6.073	727.38068	Positive	6.539	353.2648
Positive	8.081	723.48853	Positive	8.28	352.35547
Positive	9.435	721.51013	Positive	4.273	343.15656
Positive	9.138	719.49463	Positive	7.252	341.30246
Positive	6.521	716.25458	Positive	5.053	339.37521
Positive	8.074	711.49622	Positive	4.92	339.28696
Positive	7.735	711.34918	Positive	5.776	336.32657
Positive	6.452	711.33154	Positive	7.197	336.32635
Positive	8.201	707.45355	Positive	6.282	336.32605
Positive	7.38	702.21521	Positive	6.734	334.31357
Positive	6.395	701.35388	Positive	1.349	334.18597
Positive	6.521	699.23181	Positive	5.149	332.33145
Positive	5.734	697.41217	Positive	6.542	331.2861
Positive	8.495	695.45251	Positive	2.838	329.18756
Positive	5.942	695.36121	Positive	7.806	326.34052
Positive	6.077	691.31024	Positive	2.393	324.14523
Positive	10.743	681.60095	Positive	7.795	321.31409
Positive	8.647	681.474	Positive	1.152	320.20453
Positive	7.108	679.36523	Positive	5.735	319.28342
Positive	6.073	669.3371	Positive	1.036	318.19107
Positive	8.07	653.44019	Positive	1.807	316.17786
Positive	6.452	653.28235	Positive	1.395	315.10956
Positive	8.208	649.40002	Positive	4.941	313.35663
Positive	8.564	637.40527	Positive	6.542	313.27252
Positive	7.12	621.30975	Positive	3.385	313.2355
Positive	7.982	609.38116	Positive	4.111	313.1185
Positive	5.214	605.3255	Positive	7.658	312.32364
Positive	8.779	605.3175	Positive	5.432	311.29276
Positive	6.165	597.30664	Positive	7.089	310.31134
Positive	9.931	593.55231	Positive	1.208	306.18958
Positive	9.784	591.5332	Positive	4.376	304.29898
Positive	9.293	589.51642	Positive	1.431	304.1723
Positive	5.678	589.32404	Positive	1.031	299.14795
Positive	6.953	587.39539	Positive	4.826	298.34698

Positive	4.497	583.23059	Positive	3.578	295.11963
Positive	8.878	575.49866	Positive	0.909	295.09512
Positive	4.673	570.21344	Positive	4.199	294.20569
Positive	7.979	568.35748	Positive	4.421	294.19366
Positive	6.337	567.45972	Positive	0.931	294.18713
Positive	8.851	565.48492	Positive	0.871	290.1575
Positive	7.114	565.24536	Positive	1.074	288.18222
Positive	5.208	564.29315	Positive	2.566	286.24969
Positive	9.32	563.50458	Positive	5.488	284.32852
Positive	8.774	561.48737	Positive	6.449	282.28049
Positive	6.133	553.44043	Positive	5.035	281.24734
Positive	3.967	553.24561	Positive	3.578	281.13708
Positive	8.382	551.55414	Positive	0.888	278.15945
Positive	9.472	551.5061	Positive	4.34	277.17999
Positive	7.982	551.32825	Positive	0.937	274.16617
Positive	12.248	549.48694	Positive	1.383	274.08099
Positive	8.946	549.48645	Positive	0.923	272.15018
Positive	5.678	548.30444	Positive	4.514	269.25
Positive	8.355	547.46979	Positive	0.896	269.13626
Positive	8.727	547.39777	Positive	7.252	267.26843
Positive	5.235	547.26874	Positive	1.37	267.16052
Positive	4.498	542.21033	Positive	4.421	264.17993
Positive	6.27	539.42883	Positive	5.678	263.13062
Positive	6.165	539.26001	Positive	0.899	260.14938
Positive	9.102	538.55249	Positive	1.39	257.05856
Positive	6.813	537.44745	Positive	4.688	256.30081
Positive	8.326	535.51642	Positive	6.337	256.26157
Positive	8.77	535.46844	Positive	2.66	255.1945
Positive	6.651	531.33331	Positive	0.88	255.12251
Positive	5.688	531.27826	Positive	2.25	244.19273
Positive	5.652	525.41388	Positive	6.641	244.189
Positive	4.498	525.17633	Positive	4.835	242.28485
Positive	7.779	524.51825	Positive	4.421	234.16975
Positive	6.521	524.45569	Positive	4.1	230.24498
Positive	8.943	523.47357	Positive	2.548	228.19395
Positive	8.376	521.45947	Positive	6.558	228.19337
Positive	7.926	514.51855	Positive	3.277	227.20218
Positive	5.407	511.39587	Positive	2.265	226.18147
Positive	7.195	511.3649	Positive	6.497	226.1769
Positive	6.598	510.40375	Positive	2.57	225.14868

Positive	6.191	509.42014	Positive	3.863	221.11617
Positive	7.739	508.45264	Positive	4.177	221.11607
Positive	5.113	497.38208	Positive	7.437	219.17235
Positive	8.446	496.51166	Positive	0.952	211.13037
Positive	7.668	495.48755	Positive	6.497	209.15062
Positive	5.954	495.40359	Positive	1.632	205.08102
Positive	6.293	495.4035	Positive	1.332	195.13948
Positive	3.967	495.18784	Positive	0.982	183.09976
Positive	5.76	492.32095	Positive	0.833	173.08986
Positive	1.864	485.34375	Positive	0.949	169.08749
Positive	5.603	481.38394	Positive	0.941	167.10664
Positive	6.817	480.42175	Positive	0.983	165.08888
Positive	3.921	480.18689	Positive	4.267	163.07405
Positive	8.439	479.48218	Positive	1.901	163.07365
Positive	8.322	477.46274	Positive	4.298	161.09798
Positive	6.631	473.28012	Positive	1.383	159.08109
Positive	3.582	473.2663	Positive	7.097	149.02155
Positive	4.855	471.3342	Positive	4.659	149.0215
Positive	5.568	467.3725	Positive	3.578	147.06532
Positive	6.507	466.409	Positive	3.578	129.0522
Positive	3.146	466.17783	Positive	1.078	123.07935
Positive	6.27	464.39594	Positive	9.053	1193.77844
Positive	5.855	461.35992	Positive	1.14	113.05937
Positive	6.473	459.2623	Positive	0.949	109.06334
Positive	4.367	457.24814	Positive	5.695	1078.57214
Positive	7.051	455.33606	Positive	8.495	1074.63599
Positive	4.331	455.25589	Positive	9.535	1073.73938
Positive	3.575	453.17102	Positive	3.562	107.08674
Positive	2.716	452.25027	Positive	9.725	1061.73572
Positive	7.058	450.37857	Negative	8.941	159.08627
Positive	4.673	449.14697	Negative	2.705	166.06165
Positive	3.136	449.14676	Negative	0.926	168.98695
Positive	3.925	445.24353	Negative	2.867	193.08559
Positive	7.672	437.43591	Negative	2.705	223.13161
Positive	3.585	437.19061	Negative	3.35	239.12907
Positive	3.909	437.19009	Negative	2.262	242.17421
Positive	6.885	436.40182	Negative	1.774	291.12234
Positive	6.337	434.34677	Negative	1.099	297.13239
Positive	1.251	433.25104	Negative	1.627	298.94058
Positive	3.941	429.22549	Negative	2.888	327.17294

Positive	4.807	427.30771	Negative	1.11	343.13623
Positive	4.104	427.23605	Negative	5.398	347.16931
Positive	5.632	423.34482	Negative	6.512	365.24319
Positive	1.795	421.25464	Negative	2.973	369.98154
Positive	2.602	419.2421	Negative	6.618	395.24161
Positive	2.467	418.24451	Negative	3.694	395.24173
Positive	3.79	416.21201	Negative	3.921	421.15302
Positive	7.058	415.34479	Negative	7.105	431.14136
Positive	4.323	414.22556	Negative	3.916	431.18466
Positive	5.724	413.37445	Negative	2.455	435.95471
Positive	6.263	412.39093	Negative	7.112	447.33014
Positive	2.33	412.15894	Negative	3.921	448.17136
Positive	6.019	411.16953	Negative	6.349	449.31271
Positive	5.815	410.37219	Negative	3.684	459.20517
Positive	6.216	408.37125	Negative	7.045	475.32828
Positive	1.278	408.22369	Negative	7.048	477.34232
Positive	6.338	407.34418	Negative	2.921	488.23892
Positive	5.933	406.35431	Negative	3.922	499.168
Positive	5.116	403.23193	Negative	4.492	569.1698
Positive	3.299	401.19299	Negative	8.013	580.2135
Positive	4.367	399.194	Negative	9.763	591.57452
Positive	7.874	398.40085	Negative	8.584	608.24139
Positive	5.782	397.38016	Negative	6.284	615.35486
Positive	4.91	397.33932	Negative	9.132	636.27765
Positive	5.676	397.20172	Negative	5.011	639.14917
Positive	3.572	397.20172	Negative	3.79	645.11377
Positive	6.885	395.41223	Negative	3.794	655.15186
Positive	6.937	395.39075	Negative	9.644	664.30365
Positive	2.33	395.13684	Negative	9.61	664.31256
Positive	2.91	393.28381	Negative	5.948	671.26086
Positive	1.299	392.22757	Negative	3.794	672.13922
Positive	6.309	391.3396	Negative	5.918	681.29364
Positive	1.145	389.22839	Negative	5.945	698.284
Positive	6.438	388.39648	Negative	9.498	748.51282
Positive	3.923	387.19302	Negative	3.794	749.15527
Positive	4.104	386.20642	Negative	6.089	771.98523
Positive	4.936	383.27762			

Appendix Table 4. List of lipids when testing bacteriophage-treated *M. smegmatis* that were detected by mass spectrometry but did not match any database references. Reported is the average retention time (Rt) in minutes, and the average mass-to-charge ratio (M/Z), and the chemical formula predicted by MS-Finder. All lipids were found through positive ESI-Mass Spectrometry with tandem-Mass Spectrometry.

Average Rt (min)	Average M/Z	Chemical Formula	Average Rt (min)	Average M/Z	Chemical Formula
5.661	1078.5695		5.104	453.3544	C27H48O5
4.816	242.3040		6.363	510.4041	C27H51N5O4
4.277	163.0757	C10H10O2	13.137	563.4984	C27H62N8O4
1.979	163.0759	C10H10O2	11.055	1021.7352	C27H89N32O8P
2.053	163.0762	C10H10O2	10.553	1191.8278	C27H96N46O4P2
1.384	257.0568	C10H12N2O4S	5.911	655.3367	C28H107N28O16P7
1.384	274.0851	C10H12N2O4S	3.612	437.1909	C28H24N2O3
0.932	165.0916	C10H12O2	7.14	509.1871	C28H28O9
0.961	165.0922	C10H12O2	4.094	427.2369	C28H30N2O2
0.939	167.1060	C10H14O2	4.456	429.2541	C28H32N2O2
0.989	183.1037	C10H14O3	3.862	523.2372	C28H34N4O4S
0.749	261.1027	C10H16N2O6	5.644	531.2714	C28H38N2O8
1.384	159.0787	C11H10O	3.558	531.2725	C28H38N2O8
1.086	179.1054	C11H14O2	8.108	551.3062	C28H38N8O4
1.094	179.1066	C11H14O2	5.615	423.3433	C28H39NO
1.214	238.1196	C11H15N3O3	2.264	469.3108	C28H40N2O4
0.893	213.1121	C11H16O4	5.449	423.3436	C28H42N2O
6.241	239.1497	C11H18N4O2	5.742	475.2988	C28H42O6
5.875	239.1501	C11H18N4O2	5.742	492.3236	C28H42O6
5.537	239.1510	C11H18N4O2	4.9	397.3447	C28H44O
0.736	239.1513	C11H18N4O2	5.363	591.3408	C28H48N9P
0.736	256.1772	C11H18N4O2	6.341	448.3713	C28H49NO3
0.678	214.1432	C11H19NO3	5.423	467.3726	C28H50O5
6.047	251.0453	C12H11O4P	6.336	481.4385	C28H53N3O2
0.938	209.1200	C12H16O3	7.032	529.4310	C28H56N4O5
1.35	195.1382	C12H18O2	6.109	553.4423	C28H57NO8
1.373	211.1330	C12H18O3	7.18	915.7136	C28H87N26O6P
0.938	211.1346	C12H18O3	3.864	465.1778	C29H24N2O4
0.923	227.1292	C12H18O4	3.867	482.2038	C29H24N2O4
2.857	273.1252	C12H20N2O3S	3.956	553.2466	C29H36N4O5S
1.362	213.1475	C12H20O3	4.787	496.2652	C29H37NO6
0.699	261.1309	C12H20O6	4.851	513.2981	C29H40N2O6

0.867	260.1480	C12H23NO6	5.027	561.2764	C29H40N2O9
0.855	278.1600	C12H23NO6	4.569	561.2858	C29H40N2O9
6.613	228.2071	C12H25N3O	1.877	485.3445	C29H41NO4
6.185	283.1734	C12H26O7	4.434	515.3117	C29H42N2O6
0.741	300.1999	C12H26O7	6.759	422.3488	C29H43NO
0.964	183.0836	C13H10O	4.211	518.3154	C29H43NO7
5.182	189.1300	C13H16O	6.961	529.3444	C29H46N4O6
4.162	221.1187	C13H16O3	5.763	397.3799	C29H50O
1.386	315.1083	C13H18N2O5S	5.697	481.3848	C29H52O5
1.098	223.1352	C13H18O3	5.075	497.3824	C29H52O6
0.98	239.1297	C13H18O4	5.115	559.3944	C29H54N2O8
0.908	255.1225	C13H18O5	7.62	435.4211	C29H54O2
0.908	272.1485	C13H18O5	7.718	437.4351	C29H56O2
6.47	226.1811	C13H20O2	6.504	524.4554	C29H57N5O3
1.114	258.1726	C13H20O4	10.344	563.5068	C29H74N2OS3
5.843	283.1743	C13H22N4O3	6.616	473.2814	C30H36N2O3
6.47	228.1976	C13H22O2	6.308	434.3466	C30H43NO
6.627	244.1927	C13H22O3	5.982	451.3761	C30H43NO
0.831	290.1620	C13H23NO6	6.308	451.3766	C30H43NO
1.872	401.1103	C13H25N2O8PS	5.575	467.3732	C30H43NO2
2.558	228.1969	C13H25NO2	4.914	529.3279	C30H44N2O6
0.944	316.1786	C13H27N5O3S	6.249	452.3969	C30H46N2
0.711	262.1983	C13H27NO4	9.482	605.3176	C30H46N4O10
2.432	369.0890	C14H16N4O6S	4.788	484.3843	C30H49N3O2
2.435	386.1103	C14H16N4O6S	7.035	472.4235	C30H53N3O
4.397	249.1099	C14H16O4	5.394	511.3975	C30H54O6
4.093	313.1179	C14H20N2O4S	9.056	535.4703	C30H58N6O2
1.048	253.1443	C14H20O4	9.424	765.5401	C30H82N10O2S4
0.908	269.1418	C14H20O5	4.945	593.1671	C31H28O12
0.705	236.1662	C14H21NO2	4.945	610.1940	C31H28O12
1.276	271.1522	C14H22O5	5.067	507.2386	C31H30N4O3
0.94	294.1942	C14H23N5O2	3.568	609.2387	C31H36N4O7S
2.64	237.2029	C14H24N2O	4.919	572.2887	C31H41NO9
0.95	301.1718	C14H24N2O5	8.114	551.3288	C31H42N4O5
3.3	227.2005	C14H26O2	5.451	463.3163	C31H42O3
1.186	306.1887	C14H27NO6	4.218	559.3377	C31H46N2O7
4.102	230.2481	C14H31NO	6.616	531.3342	C31H46O7
2.047	261.1085	C15H16O4	5.196	605.3188	C31H48N4O4S2
4.093	263.1270	C15H18O4	9.134	605.3194	C31H48N4O4S2
4.537	263.1278	C15H18O4	7.94	519.3704	C31H50O6

4.296	263.1280	C15H18O4	8.231	519.3706	C31H50O6
3.981	263.1286	C15H18O4	6.455	654.3736	C31H51N5O10
5.641	263.1292	C15H18O4	9.652	705.4737	C31H64N10O6S
3.559	295.1173	C15H18O6	6.146	539.2611	C32H34N4O4
3.328	295.1174	C15H18O6	6.275	478.3714	C32H47NO2
4.015	295.1175	C15H18O6	6.831	480.3920	C32H49NO2
3.559	281.1349	C15H20O5	5.075	603.4232	C32H54N6O5
3.434	297.1338	C15H20O6	6.113	584.4757	C32H57N9O
2.121	219.1749	C15H22O	8.487	477.4638	C32H60O2
9.303	219.1750	C15H22O	8.045	514.5228	C32H67NO3
7.498	219.1761	C15H22O	6.36	565.2467	C33H32N4O5
2.094	251.1659	C15H22O3	7.139	565.2486	C33H32N4O5
1.012	283.1579	C15H22O5	5.911	581.2463	C33H32N4O6
1.079	299.1462	C15H22O6	5.194	564.2943	C33H38O7
2.4	253.1814	C15H24O3	7.357	631.3216	C33H46N2O10
1.491	267.1581	C15H24O5	8.114	568.3582	C33H46N2O3S
2.647	237.1823	C15H26O3	8.114	609.3824	C33H48N6O5
2.649	255.1983	C15H26O3	8.936	547.3997	C33H54O6
1.935	271.1911	C15H26O4	15.683	521.4511	C33H60O4
0.693	310.2253	C15H27N5O2	15.874	521.4594	C33H60O4
1.449	302.1972	C15H27NO5	8.604	521.4595	C33H60O4
2.753	341.2303	C15H28N6O3	7.867	507.4797	C33H62O3
3.954	240.2327	C15H29NO	9.303	523.4745	C33H62O4
4.896	228.2690	C15H33N	9.324	550.4949	C33H63N3O3
4.615	279.1567	C16H22O4	2.576	579.1637	C34H26O9
1.369	334.1857	C16H23N5O3	4.336	565.2810	C34H36N4O4
0.931	313.1628	C16H24O6	6.512	523.4346	C34H51NO2
7.679	1147.8740	C16H25N11O38P6	4.732	575.3941	C34H54O7
0.638	332.2041	C16H29NO6	9.184	607.3875	C34H54O9
6.327	256.2650	C16H33NO	5.069	542.4278	C34H55NO4
4.824	242.2829	C16H35N	4.429	595.2903	C35H38N4O5
4.996	258.2782	C16H35NO	6.445	670.2797	C35H43NO12
4.03	327.0769	C17H14N2O3S	5.642	589.3260	C35H44N2O6
4.676	313.1176	C17H16N2O4	5.211	655.3187	C35H46N2O10
1.134	345.0859	C17H16N2O4S	5.027	619.3306	C35H46N4O4S
3.911	331.1279	C17H18N2O5	7.304	653.3377	C35H48N4O6S
5.096	369.1241	C17H20O9	8.385	611.3666	C35H50N2O7
1.09	318.1899	C17H23N3O3	9.978	607.3876	C35H50N4O5
4.162	294.2037	C17H24O3	6.846	537.4484	C35H53NO2
1.131	320.2044	C17H25N3O3	7.704	651.3768	C35H54O11

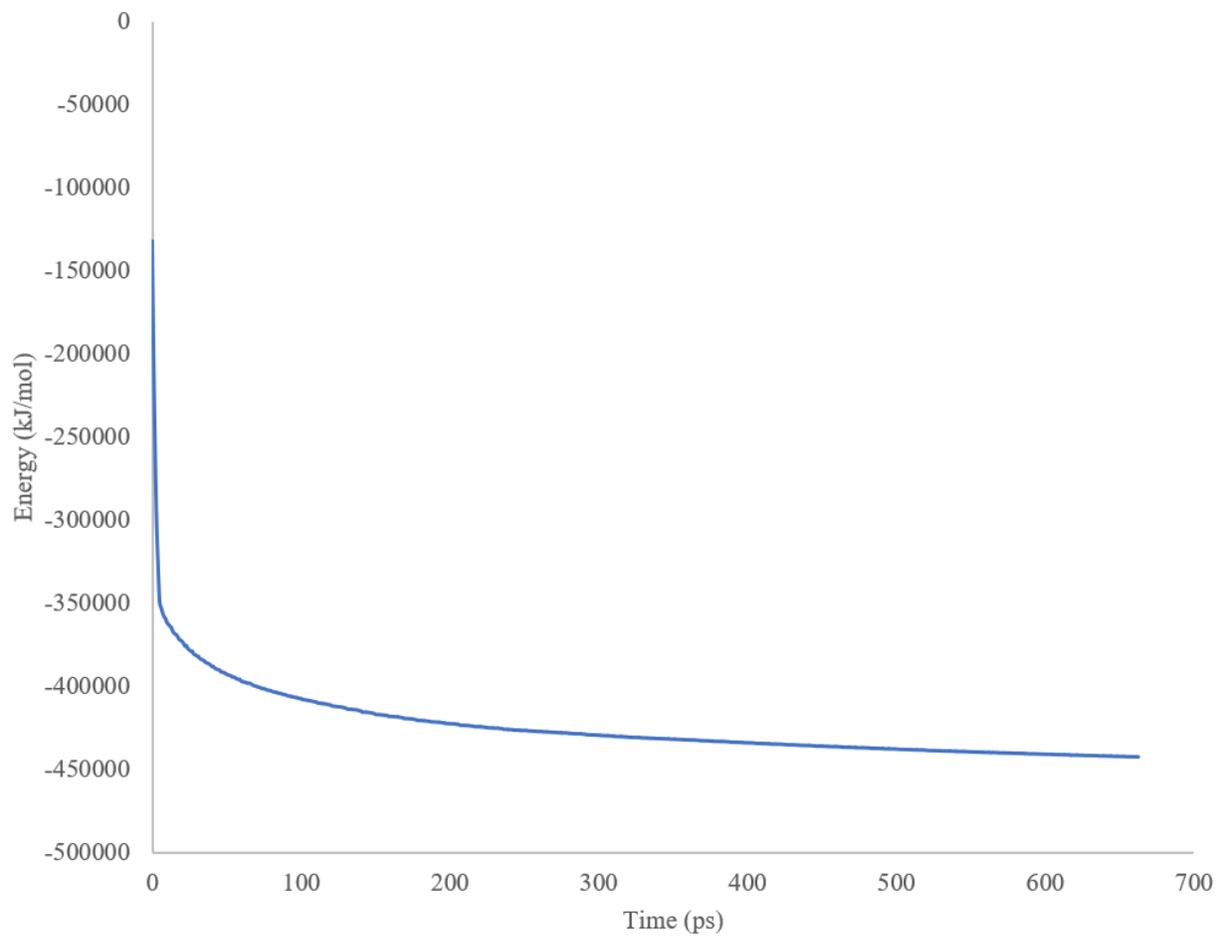
1.327	311.1870	C17H26O5	6.959	587.3970	C35H54O7
3.133	326.1950	C17H27NO5	6.149	540.4490	C35H57NO3
4.493	269.2485	C17H32O2	8.658	603.4330	C35H58N2O6
4.85	268.2616	C17H33NO	5.779	570.4614	C35H59N3O3
3.974	301.2854	C17H36N2O2	15.499	547.4720	C35H62O4
4.675	256.2978	C17H37N	8.625	547.4725	C35H62O4
6.143	509.4215	C17H45N15O2	9.176	547.4732	C35H62O4
3.118	393.0844	C18H16O10	15.608	549.4877	C35H64O4
3.351	315.1207	C18H18O5	9.891	549.4877	C35H64O4
3.328	453.1709	C18H24N6O8	8.886	565.4862	C35H64O5
1.068	320.1908	C18H25NO4	9.34	565.4863	C35H64O5
2.756	309.2083	C18H28O4	10.016	551.4990	C35H66O4
3.346	313.2402	C18H32O4	10.557	551.5001	C35H66O4
2.648	313.2459	C18H32O4	8.938	681.4688	C35H69O10P
6.453	303.2914	C18H38O3	6.553	659.2897	C36H42N4O6S
5.764	336.3112	C18H38O4	10.136	605.3111	C36H44O8
4.081	270.3154	C18H39N	6.342	669.3317	C36H48N2O10
5.499	283.1724	C19H22O2	8.807	637.3991	C36H52N4O6
2.324	395.1296	C19H22O9	6.31	567.4604	C36H55NO3
2.335	412.1588	C19H22O9	8.401	649.4022	C36H56O10
2.855	329.1869	C19H24N2O3	9.137	561.4896	C36H64O4
1.06	376.2305	C19H29N5O3	9.662	561.4902	C36H64O4
0.756	432.2812	C19H37N5O6	9.841	563.5067	C36H66O4
5.486	379.3196	C19H39NO5	10.014	565.5201	C36H68O4
4.887	329.3179	C19H40N2O2	6.504	716.2585	C37H38N4O6S2
5.998	333.3008	C19H40O4	10.223	621.3105	C37H40N4O5
6.005	350.3261	C19H40O4	7.137	638.3403	C37H43N5O5
7.89	411.3805	C19H47N5OS	6.039	669.3309	C37H48O11
6.51	719.2593	C19H66N4O5S9	6.44	637.3362	C37H48O9
3.446	469.0328	C20H12N4O6S2	5.914	831.3345	C37H50N8O12S
3.465	486.0600	C20H12N4O6S2	5.711	639.3539	C37H50O9
2.57	337.1486	C20H20N2O3	5.36	555.4209	C37H56O2
2.712	452.1035	C20H20O7P2	6.68	581.4722	C37H57NO3
4.595	339.1706	C20H22N2O3	5.283	617.4388	C37H60O7
5.811	327.2003	C20H26N2O2	9.316	575.4998	C37H66O4
6.146	327.2007	C20H26N2O2	9.293	575.5029	C37H66O4
5.456	327.2010	C20H26N2O2	15.328	575.5103	C37H66O4
3.461	375.1915	C20H26N2O5	9.52	593.5118	C37H68O5
4.337	451.1643	C20H26N4O6S	6.44	653.2850	C38H40N2O8
4.82	314.2141	C20H27NO2	7.911	621.3445	C38H44N4O4

4.72	337.2762	C20H33NO2	6.555	695.3554	C38H50N2O10
1.128	352.2313	C20H33NO2S	5.931	695.3557	C38H50N2O10
2.278	352.2315	C20H33NO2S	7.156	815.3362	C38H54O19
4.726	399.2496	C20H34N2O6	9.594	661.4394	C38H60O9
5.1	403.2312	C20H34O8	9.801	589.5198	C38H68O4
2.962	410.2733	C20H35N5O4	10.431	591.5355	C38H70O4
1.437	427.2997	C20H35N5O4	9.582	719.4904	C38H71O10P
5.405	311.2968	C20H38O2	10.638	593.5546	C38H72O4
4.946	313.3563	C20H41N	10.542	593.5550	C38H72O4
7.745	312.3294	C20H41NO	5.393	865.6003	C39H100N4OS7
6.324	347.3171	C20H42O4	7.929	679.4030	C39H54N2O8
6.335	364.3424	C20H42O4	8.678	607.5749	C39H74O4
5.598	363.3076	C20H42O5	5.502	468.3884	C39H99N16O7P
5.585	380.3381	C20H42O5	3.561	107.0838	C3H10N2O2
4.351	457.2499	C20H42O7P2	3.33	107.0842	C3H10N2O2
4.82	298.3457	C20H43N	4.359	105.0687	C3H8N2O2
7.144	335.1776	C21H22N2O2	9.185	665.4459	C40H60N2O6
2.435	427.1425	C21H22N4O4S	7.597	563.5132	C40H66O
4.094	369.1833	C21H24N2O4	9.283	747.5222	C40H75O10P
4.093	386.2067	C21H24N2O4	6.442	695.3930	C41H54N6S2
4.679	386.2069	C21H24N2O4	8.228	653.4445	C41H56N4O3
2.712	493.1280	C21H24N4O6S2	5.392	600.4725	C41H61NO2
4.362	405.1570	C21H24O8	8.186	711.4644	C41H62N2O8
4.453	371.1985	C21H26N2O4	8.733	749.5069	C41H75O8P
3.911	387.1923	C21H26N2O5	9.702	773.5466	C41H76N2O11
3.91	404.2182	C21H26N2O5	9.323	549.4896	C42H112N32O2
5.031	371.1987	C21H28N2O5	11.339	1093.7127	C42H132N4OS12
1.592	363.2161	C21H30O5	6.504	699.2263	C42H34O10
5.568	509.1869	C21H32O14	3.512	659.4309	C42H58O6
5.642	397.2015	C21H32O5S	9.794	735.5196	C42H66N6O5
3.557	397.2022	C21H32O5S	4.301	414.2241	C42H66O16
3.659	413.1983	C21H32O6S	5.091	767.5374	C42H73N3O30P14
2.903	393.2840	C21H36N4O3	10.016	551.4998	C43H116N30O3
4.362	304.2978	C21H37N	8.612	723.4177	C43H54N4O6
5.126	409.3286	C21H37N5O2	4.297	397.2001	C43H56N2O12
4.899	321.2806	C21H38O3	7.281	903.5777	C43H74N12O9
4.897	339.2914	C21H38O3	7.034	903.5794	C43H74N12O9
4.42	355.2868	C21H38O4	3.911	795.3607	C44H52N4O11
7.214	371.3183	C21H39NO3	6.402	853.3885	C44H52N8O10
6.679	361.3348	C21H44O4	5.711	697.4063	C44H56O7

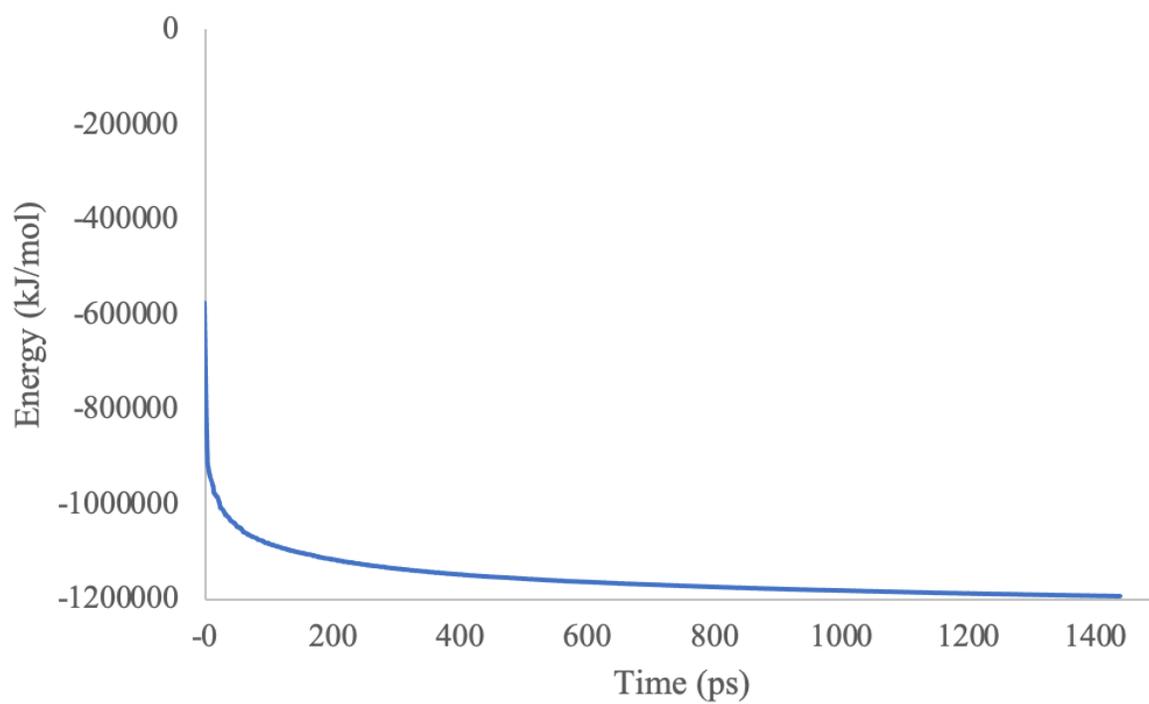
6.68	378.3587	C21H44O4	7.679	709.4360	C44H60N4S2
9.438	1193.7760	C21H47N7O8P4S17	8.337	842.4764	C44H67N5O9S
6.291	408.3702	C21H50N3O2P	9.986	857.4825	C44H73O14P
7.137	622.3131	C21H79N40O20P	9.389	791.5519	C44H81O8P
3.118	449.1466	C22H24O10	10.049	793.5698	C44H83O8P
4.266	433.1505	C22H24O9	10.634	853.5793	C44H85O13P
5.679	433.1521	C22H24O9	6.039	727.3843	C45H50N4O5
5.628	435.1691	C22H26O9	7.066	881.4235	C45H60N4O14
4.428	433.1813	C22H28N2O5S	9.224	759.5252	C45H77O8P
4.6	397.2208	C22H28N4O3	3.769	851.4002	C45H81N68O6P
4.009	416.2180	C22H29N3O5	3.309	473.2655	C46H72N8O13
4.475	525.1763	C22H29N4O9P	4.016	473.2667	C46H72N8O13
4.474	542.2024	C22H29N4O9P	10.48	1061.7311	C47H119N2O10P3S3
4.89	542.2033	C22H29N4O9P	10.345	705.5298	C47H70O3
4.435	450.2013	C22H31N3O5S	5.048	837.5648	C47H81O10P
5.971	394.3532	C22H40N4O	5.176	895.6071	C47H82N4O12
5.072	339.3745	C22H43N	2.903	376.2577	C48H66N2O5
6.074	406.3550	C22H47NO5	6.834	833.5398	C48H72N4O8
11.919	1121.7423	C22H48N3O2PS22	9.782	761.5374	C48H72O7
3.118	914.3127	C22H84N6O8S14	6.618	761.4893	C49H64N2O5
2.723	435.0742	C23H14O9	8.473	900.5158	C49H73NO14
1.871	418.1393	C23H19N3O5	5.644	119.0872	C4H10N2O2
3.794	447.1710	C23H26O9	0.758	107.0625	C4H10OS
4.422	481.1720	C23H28O11	0.829	198.9912	C4H6O7S
4.523	400.2169	C23H29NO5	4.364	133.0668	C4H8N2O3
5.241	447.1718	C23H30N2O3S2	2	105.0336	C4H8OS
0.754	388.2538	C23H30O4	3.561	851.3978	C51H54N4O8
3.913	445.2434	C23H32N4O5	10.453	1087.7527	C52H118N2O10P4S
0.671	347.2991	C23H38O2	3.913	823.3793	C52H54O9
4.813	443.3130	C23H39NO6	9.85	805.5708	C52H78O5
0.639	365.3005	C23H40O3	7.345	847.5576	C54H74N2O6
5.346	397.2985	C23H40O5	11.332	785.6222	C54H82O2
5.142	332.3302	C23H41N	5.846	482.4101	C55H110O12
7.781	365.3550	C23H44N2O	5.637	548.2980	C55H86N2O20
6.149	353.3422	C23H44O2	11.996	1095.7253	C55H98N8O14
8.506	352.3546	C23H45NO	10.727	1075.7588	C57H106N2O16
3.116	897.2918	C23H81N2O10PS10	4.037	851.3984	C57H54O7
10.756	1180.6950	C23H84N45O6P3	8.682	1016.5998	C57H85N5O7S2
3.717	867.3723	C23H86N4O8S10	8.009	917.5982	C58H80N2O7
5.062	449.1835	C24H24N4O5	7.42	917.6007	C58H80N2O7

2.479	362.2146	C24H27NO2	7.137	621.3131	C58H96O28
2.479	362.2152	C24H27NO2	5.884	438.3770	C58H98O5
2.945	378.2102	C24H27NO3	7.809	903.5799	C59H74N4O4
4.605	397.2006	C24H28O5	5.277	485.3463	C59H88N2O8
3.953	495.1883	C24H30O11	4.02	135.0962	C5H14N2S
4.145	447.2080	C24H30O8	3.316	135.0966	C5H14N2S
4.92	514.2387	C24H31N7O6	0.991	101.0590	C5H8O2
4.435	491.2331	C24H34N4O5S	0.952	101.0627	C5H8O2
5.559	491.2341	C24H36N4O6S	4.61	149.0232	C5H8O3S
4.457	377.3106	C24H40O3	2.725	197.0013	C5H8O4S2
7.875	360.3221	C24H41NO	0.738	181.0266	C5H9O5P
6.305	391.3386	C24H42N2O2	0.739	198.0512	C5H9O5P
1.859	528.3511	C24H42N6O6	10.092	1073.7372	C60H96N8O9
4.723	395.3160	C24H42O4	10.558	1165.8146	C64H112N2O16
5.542	429.3185	C24H44O6	10.411	1166.6754	C65H103N3O7S4
8.224	366.3727	C24H47NO	10.659	1101.7653	C65H104N4O8S
4.101	459.4916	C24H58N8	6.232	539.4272	C65H113N4O6P
7.009	498.4411	C24H59N5O3S	5.899	1116.6078	C67H81N5O10
1.885	459.1663	C25H22N4O5	10.395	1152.6694	C67H89N7O10
3.116	466.1778	C25H24N2O4S	8.795	1074.6360	C69H87N9O9
3.657	430.2216	C25H24N4O2	3.56	147.0623	C6H10O4
1.246	392.2288	C25H29NO3	0.748	133.0873	C6H12O3
4.981	612.1937	C25H33N5O11S	0.772	164.0926	C6H13NO4
4.567	520.2602	C25H34N4O5S	0.785	135.0923	C6H14O3
4.374	443.2542	C25H36N2O6	6.004	151.0972	C6H14O4
5.233	441.3202	C25H44O6	0.705	116.1070	C6H15NO2
6.332	465.3864	C25H45N5O2	0.919	105.0492	C6H6N2O
7.034	440.3832	C25H49N3O3	0.916	123.0590	C6H6N2O
6.895	453.4433	C25H57N2O2S	1.379	141.0673	C6H8N2O2
5.982	411.1698	C26H22N2O3	3.559	129.0533	C6H8O3
4.405	427.2142	C26H26N4O2	11.254	1119.7229	C70H94N4O8
3.822	437.1905	C26H28O6	10.379	1192.6934	C72H89N9O7
3.329	437.1906	C26H28O6	11.732	1107.7266	C78H96N2O4
4.017	437.1969	C26H28O6	6.334	151.0976	C7H10N4
4.3	455.2548	C26H34N2O5	0.959	127.0741	C7H12O3
4.336	524.2510	C26H34O10	0.802	144.1014	C7H13NO2
3.586	473.2666	C26H36N2O6	1.062	179.0920	C7H14O5
2.34	485.3057	C26H44O8	3.451	197.0013	C7H4N2O3S
6.351	407.3507	C26H46O3	0.925	109.0659	C7H8O
6.649	479.4079	C26H47N5O2	6.194	583.4561	C81H116N2O3

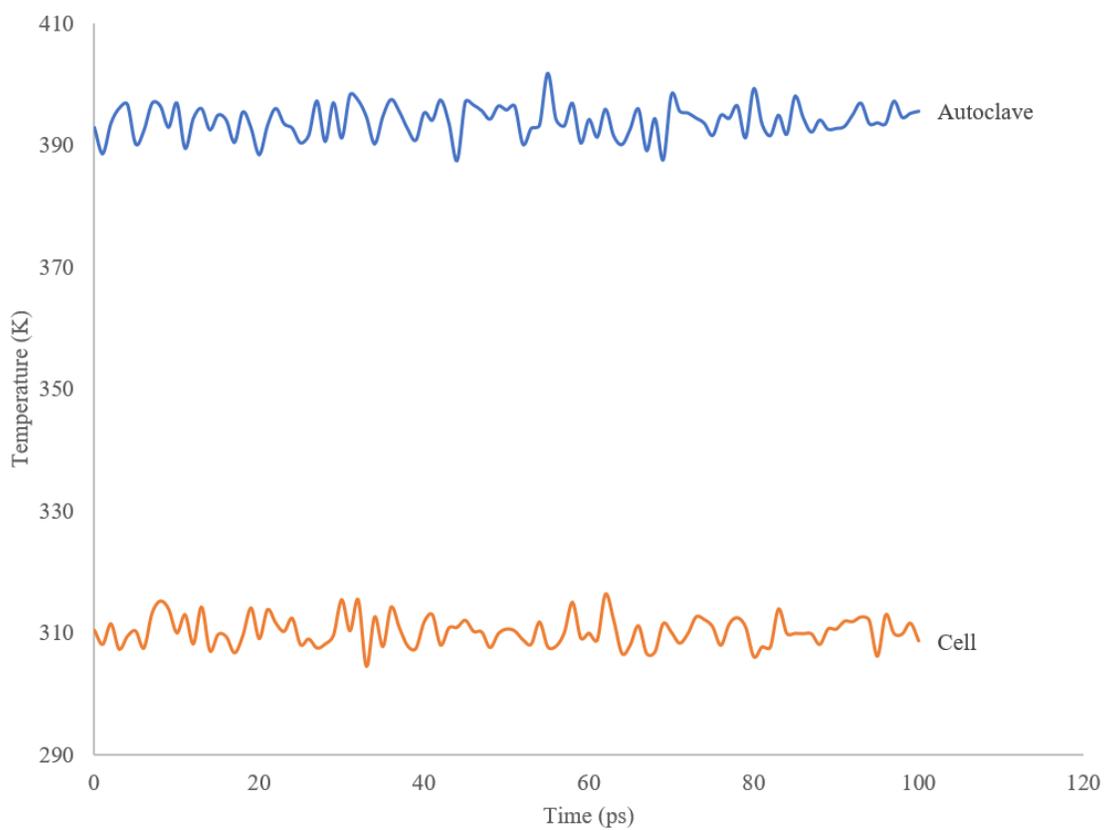
6.274	495.4041	C26H47N5O3	3.863	135.0971	C8H10N2
6.807	393.3719	C26H48O2	0.975	123.0850	C8H10O
6.258	464.3956	C26H49N5O2	0.916	151.0909	C8H12N2O2
5.541	424.3618	C26H49NOS	0.91	213.0980	C8H12N4O3
5.928	495.4032	C26H50N6O3	1.036	141.0918	C8H12O2
6.893	395.3912	C26H50O2	0.987	139.0765	C8H12O3
6.765	466.4167	C26H51N5O2	0.769	184.1143	C8H13N3O2
9.184	1179.7606	C26H53NO4S23	7.498	203.1435	C8H18N4O2
6.591	510.3970	C26H56NO6P	0.734	195.1240	C8H18O5
8.398	707.4583	C26H82N4O2S7	0.817	130.1598	C8H19N
6.778	815.5329	C26H90N10O3S7	7.093	149.0231	C8H6O4
7.164	679.3687	C27H106N38O13P6	2.577	201.0438	C8H8O6
3.074	518.0869	C27H12N6OS2	0.96	121.1024	C9H12
2.582	557.1798	C27H28N2O11	0.918	169.0860	C9H12O3
3.771	628.1808	C27H30O16	3.02	247.0789	C9H14N2O4S
10.466	1101.7618	C27H34N3O10P3S14	0.704	217.1025	C9H16N2O2S
4.572	503.2282	C27H34O9	0.922	274.1647	C9H16N6O3
4.859	472.2739	C27H37NO6	0.711	172.1335	C9H17NO2
6.328	466.4096	C27H48N2O3	1.373	171.1510	C9H18N2O
1.871	163.0407	C9H6O3	0.705	156.1373	C9H19NO2
3.348	149.0613	C9H8O2	0.702	192.1608	C9H21NO3



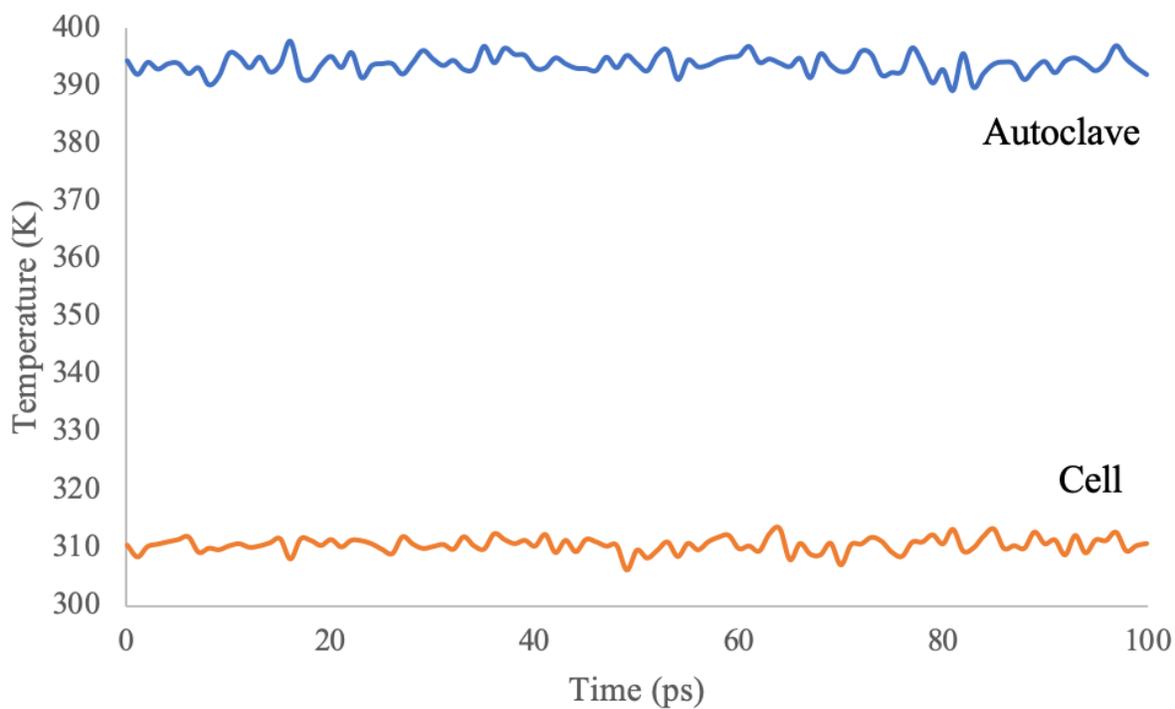
Appendix Figure 1. The potential energy curve for the Tail Assembly Chaperone 2.



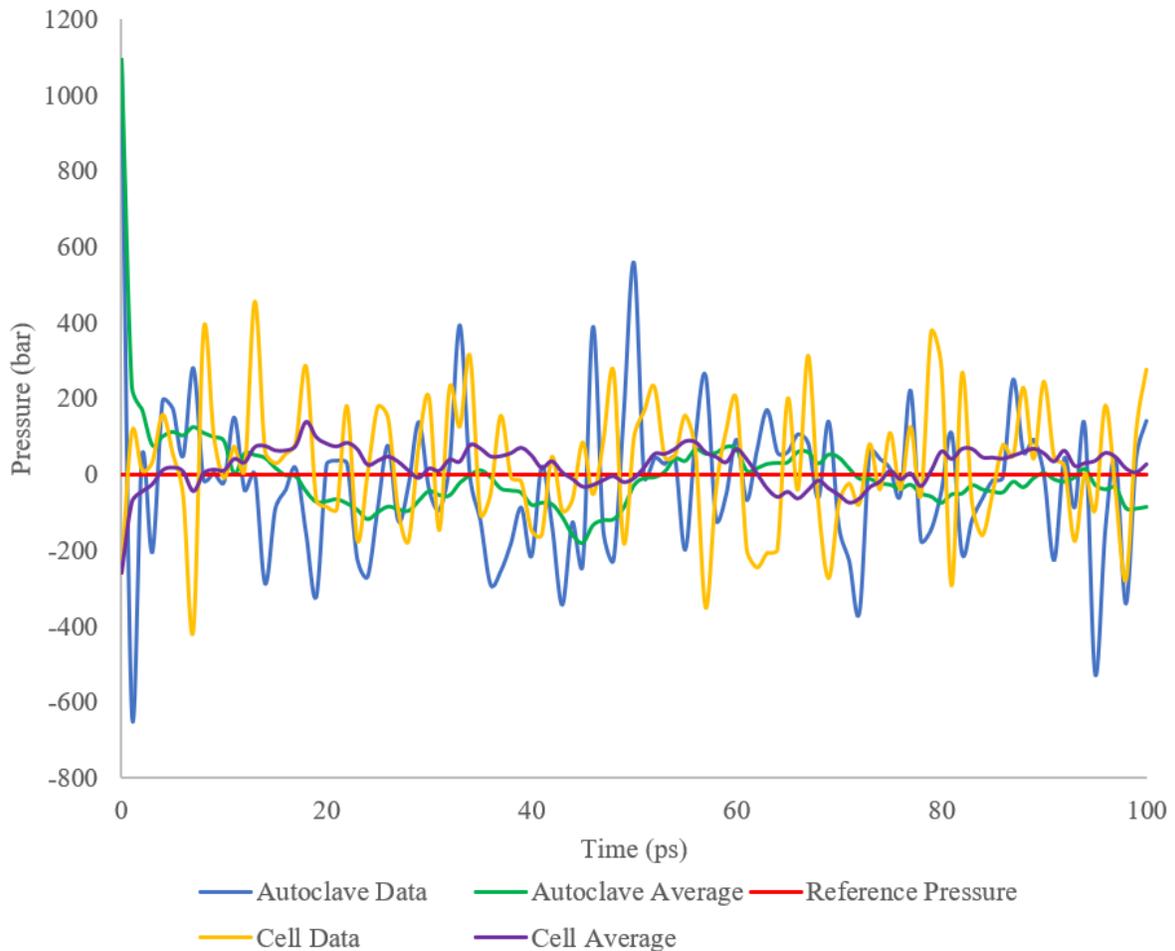
Appendix Figure 2. The potential energy curve for the Tail Assembly Chaperone 3.



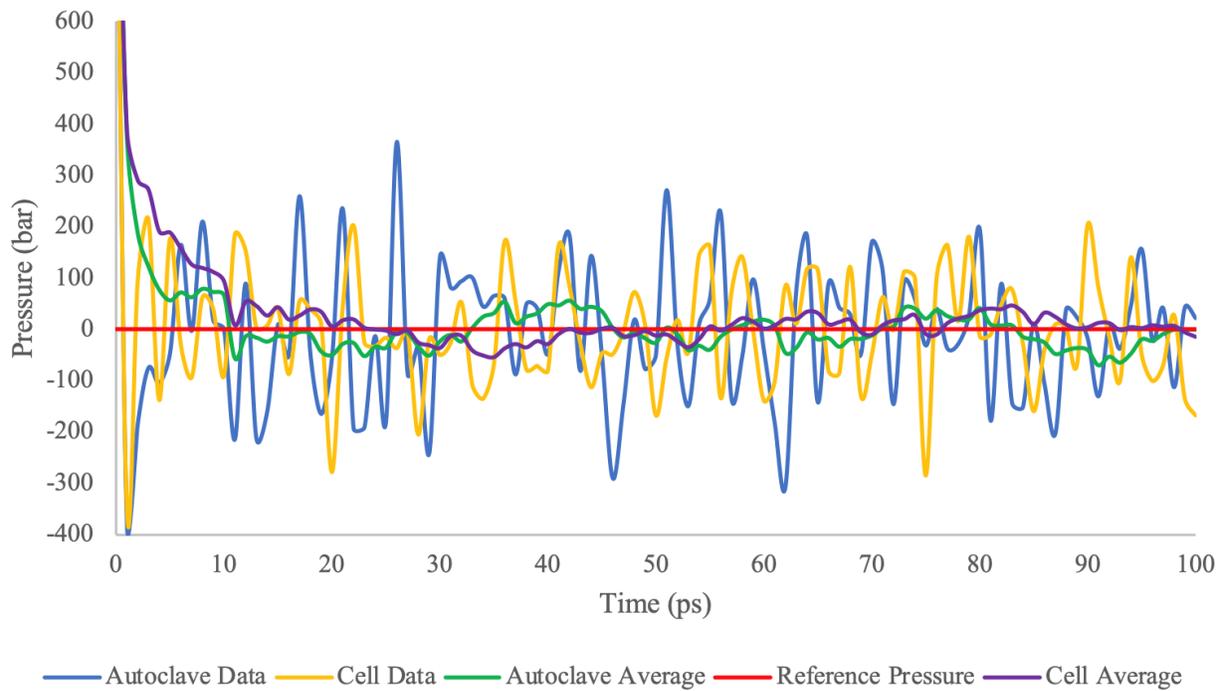
Appendix Figure 3. The temperature curves for Tail Assembly Chaperone 2 during the Autoclave simulation (blue) and the Cell simulation (orange). The autoclave reference temperature is 394.15 K, while the cell reference temperature is 310.15 K.



Appendix Figure 4. The temperature curves for Tail Assembly Chaperone 3 during the Autoclave simulation (blue) and the Cell simulation (orange). The autoclave reference temperature is 394.15 K, while the cell reference temperature is 310.15 K.



Appendix Figure 5. The pressure curves for Tail Assembly Chaperone 2 during the Autoclave simulation and the Cell simulation. The Autoclave Data is shown in blue while the 10 ps running average is in green. The Cell Data is shown in yellow while the 10 ps running average is in purple. The reference pressure is shown in red.



Appendix Figure 6. The pressure curves for Tail Assembly Chaperone 3 during the Autoclave simulation and the Cell simulation. The Autoclave Data is shown in blue while the 10 ps running average is in green. The Cell Data is shown in yellow while the 10 ps running average is in purple. The reference pressure is shown in red.

REFERENCES

1. Moineau S. Bacteriophage. Brenner's Encyclopedia of Genetics: Second Edition [Internet]. Elsevier Inc.; 2013 [cited 2021 May 14]. p. 280–3. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK493185/>
2. Hatfull GF, Hendrix RW. Bacteriophages and their genomes. *Current Opinion in Virology*. 2011.
3. Jacobs WR. Advances in Mycobacterial Genetics: New Promises for Old Diseases. *Immunobiology*. 1992;184.
4. Ackermann HW, Prangishvili D. Prokaryote viruses studied by electron microscopy. *Archives of Virology*. 2012.
5. Poxleitner M, Pope W, Jacobs-Sera D, Sivanathan V, Hatfull G. *Phage Discovery Guide*. Howard Hughes Medical Institute. 2018.
6. Clokie MRJ, Millard AD, Letarov A v., Heaphy S. Phages in nature. *Bacteriophage*. 2011;1.
7. Howard-Varona C, Hargreaves KR, Abedon ST, Sullivan MB. Lysogeny in nature: Mechanisms, impact and ecology of temperate phages. *ISME Journal*. 2017.
8. Hatfull GF. Mycobacteriophages: Genes and genomes. *Annual Review of Microbiology*. 2010.
9. Hatfull GF, Jacobs-Sera D, Lawrence JG, Pope WH, Russell DA, Ko CC, et al. Comparative Genomic Analysis of 60 Mycobacteriophage Genomes: Genome Clustering, Gene Acquisition, and Gene Size. *Journal of Molecular Biology*. Academic Press; 2010;397:119–43.
10. Hendrix RW. Bacteriophages: Evolution of the Majority. *Theoretical Population Biology*. Academic Press Inc.; 2002;61:471–80.
11. Hatfull GF. The Secret Lives of Mycobacteriophages. *Advances in Virus Research*. Academic Press Inc.; 2012. p. 179–288.
12. *Bioinformatics Guide* [Internet]. [cited 2021 May 14]. Available from: <https://seaphagesbioinformatics.helpdocsonline.com/home>
13. Bibby K. Improved Bacteriophage Genome Data is Necessary for Integrating Viral and Bacterial Ecology. *Microbial Ecology* [Internet]. Springer Science and Business Media, LLC; 2014 [cited 2021 May 15];67:242–4. Available from: <https://pubmed.ncbi.nlm.nih.gov/24253663/>

14. Xu J, Hendrix RW, Duda RL. Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. *Molecular Cell* [Internet]. *Mol Cell*; 2004 [cited 2021 May 15];16:11–21. Available from: <https://pubmed.ncbi.nlm.nih.gov/15469818/>
15. The Actinobacteriophage Database | Cluster A Phages [Internet]. [cited 2021 May 15]. Available from: <https://phagesdb.org/clusters/A/>
16. Hatfull GF. Molecular Genetics of Mycobacteriophages. *Molecular Genetics of Mycobacteria* [Internet]. wiley; 2015 [cited 2021 May 15]. p. 81–119. Available from: </pmc/articles/PMC4199240/>
17. The Actinobacteriophage Database | Cluster B Phages [Internet]. [cited 2021 May 15]. Available from: <https://phagesdb.org/clusters/B/>
18. The Actinobacteriophage Database | Cluster C Phages [Internet]. [cited 2021 May 15]. Available from: <https://phagesdb.org/clusters/C/>
19. Antibiotic resistance [Internet]. [cited 2021 May 17]. Available from: <https://www.who.int/news-room/fact-sheets/detail/antibiotic-resistance>
20. Harada LK, Silva EC, Campos WF, del Fiol FS, Vila M, Dąbrowska K, et al. Biotechnological applications of bacteriophages: State of the art [Internet]. *Microbiological Research*. Elsevier GmbH; 2018 [cited 2021 May 17]. p. 38–58. Available from: <https://pubmed.ncbi.nlm.nih.gov/29853167/>
21. Dedrick RM, Guerrero-Bustamante CA, Garlena RA, Russell DA, Ford K, Harris K, et al. Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant *Mycobacterium abscessus*. *Nature Medicine* [Internet]. Nature Publishing Group; 2019 [cited 2021 May 17];25:730–3. Available from: <https://doi.org/10.1038/s41591-019-0437-z>
22. Moyer ZD, Woolston J, Sulakvelidze A. Bacteriophage applications for food production and processing [Internet]. *Viruses*. MDPI AG; 2018 [cited 2021 May 17]. Available from: <https://pubmed.ncbi.nlm.nih.gov/29671810/>
23. Sulakvelidze A. Using lytic bacteriophages to eliminate or significantly reduce contamination of food by foodborne bacterial pathogens [Internet]. *Journal of the Science of Food and Agriculture*. John Wiley & Sons, Ltd; 2013 [cited 2021 May 17]. p. 3137–46. Available from: www.soci.org
24. Withey S, Cartmell E, Avery LM, Stephenson T. Bacteriophages - Potential for application in wastewater treatment processes [Internet]. *Science of the Total Environment*. Sci Total Environ; 2005 [cited 2021 May 17]. p. 1–18. Available from: <https://pubmed.ncbi.nlm.nih.gov/15740754/>
25. Hampton HG, Watson BNJ, Fineran PC. The arms race between bacteria and their phage foes [Internet]. *Nature*. Nature Research; 2020 [cited 2021 May 15]. p. 327–36. Available from: <https://doi.org/10.1038/s41586-019-1894-8>

26. Seed KD. Battling Phages: How Bacteria Defend against Viral Attack [Internet]. PLoS Pathogens. Public Library of Science; 2015 [cited 2021 May 15]. p. e1004847. Available from: <https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1004847>
27. Abedon ST. Bacterial ‘immunity’ against bacteriophages. Bacteriophage [Internet]. Informa UK Limited; 2012 [cited 2021 May 14];2:50–4. Available from: [/pmc/articles/PMC3357385/](https://pubmed.ncbi.nlm.nih.gov/23357385/)
28. Samson JE, Magadán AH, Sabri M, Moineau S. Revenge of the phages: Defeating bacterial defences [Internet]. Nature Reviews Microbiology. Nature Publishing Group; 2013 [cited 2021 May 14]. p. 675–87. Available from: www.nature.com/reviews/micro
29. Garneau JE, Dupuis MÈ, Villion M, Romero DA, Barrangou R, Boyaval P, et al. The CRISPR/cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature [Internet]. Nature Publishing Group; 2010 [cited 2021 May 15];468:67–71. Available from: <https://www.nature.com/articles/nature09523>
30. Tock MR, Dryden DTF. The biology of restriction and anti-restriction. Current Opinion in Microbiology. Elsevier Current Trends; 2005. p. 466–72.
31. Krüger DH, Barcak GJ, Reuter M, Smith HO. EcoRII can be activated to cleave refractory DNA recognition sites. Nucleic Acids Research [Internet]. Oxford Academic; 1988 [cited 2021 May 15];16:3997–4008. Available from: <https://academic.oup.com/nar/article/16/9/3997/2377751>
32. Lawrence J. DNA Master. 2007.
33. PECAAN [Internet]. [cited 2021 May 17]. Available from: <https://discover.kbrinsgd.org/evidence/summary>
34. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. Nucleic Acids Research [Internet]. Oxford University Press; 2021 [cited 2021 May 17];49:D412–9. Available from: <https://covid-19.oup.com/>
35. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank [Internet]. Nucleic Acids Research. Oxford University Press; 2000 [cited 2021 May 17]. p. 235–42. Available from: <http://www.rcsb.org/pdb/status.html>
36. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, et al. CDD: A Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Research [Internet]. Oxford Academic; 2011 [cited 2021 May 17];39:D225–9. Available from: https://academic.oup.com/nar/article/39/suppl_1/D225/2507607
37. Fox NK, Brenner SE, Chandonia JM. SCOPe: Structural Classification of Proteins - Extended, integrating SCOP and ASTRAL data and classification of new structures. Nucleic Acids Research [Internet]. Oxford Academic; 2014 [cited 2021 May 17];42:D304–9. Available from: <http://scop.berkeley.>

38. Pruitt KD, Tatusova T, Maglott DR. NCBI Reference Sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research* [Internet]. Oxford Academic; 2005 [cited 2021 May 17];33:D501–4. Available from: <http://www.ncbi.nlm.nih.gov/RefSeq/collaborators>.
39. Lavigne R, Ceysens PJ, Robben J. Phage proteomics: applications of mass spectrometry. *Methods in molecular biology* (Clifton, NJ) [Internet]. *Methods Mol Biol*; 2009 [cited 2021 May 17];502:239–51. Available from: <https://pubmed.ncbi.nlm.nih.gov/19082560/>
40. Stephenson DJ, Hoeflerlin LA, Chalfant CE. Lipidomics in translational research and the clinical significance of lipid-based biomarkers [Internet]. *Translational Research*. Mosby Inc.; 2017 [cited 2021 May 17]. p. 13–29. Available from: </pmc/articles/PMC5659874/>
41. Kumar NG, Contaifer D, Baker PRS, Ekroos K, Jefferson KK, Wijesinghe DS. Untargeted lipidomic analysis to broadly characterize the effects of pathogenic and non-pathogenic staphylococci on mammalian lipids. *PLoS ONE* [Internet]. Public Library of Science; 2018 [cited 2021 May 17];13:e0206606. Available from: <https://doi.org/10.1371/journal.pone.0206606>
42. Tokunaga T, Kataoka T, Suga K. Phage inactivation by an ethanol-ether extract of *Mycobacterium smegmatis*. *American Review of Respiratory Disease*. 1970;101.
43. Bowman BU, Newman HA, Moritz JM, Koehler RM. Properties of mycobacteriophage DS6A. II. Lipid composition. *American Review of Respiratory Disease* [Internet]. *Am Rev Respir Dis*; 1973 [cited 2021 May 17];107:42–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/4630313/>
44. Fay D, Bowman BU. Structure of native and chloroform-methanol-treated mycobacteriophage R1. *Journal of Virology* [Internet]. American Society for Microbiology; 1978 [cited 2021 May 17];27:432–5. Available from: <https://pubmed.ncbi.nlm.nih.gov/691117/>
45. Furuchi A, Tokunaga T. Nature of the receptor substance of *Mycobacterium smegmatis* for D4 bacteriophage adsorption. *Journal of Bacteriology*. 1972;111.
46. Bisso G, Castelnovo G, Nardelli MG, Orefici G, Arancia G, Lanéelle G, et al. A study on the receptor for a mycobacteriophage : phage phlei. *Biochimie*. 1976;58.
47. Jackson M. The mycobacterial cell envelope-lipids. *Cold Spring Harbor Perspectives in Medicine* [Internet]. Cold Spring Harbor Laboratory Press; 2014 [cited 2021 May 17];4. Available from: </pmc/articles/PMC4200213/>
48. Precellys Tissue Homogenizer, Lysing Kits [Internet]. [cited 2021 May 14]. Available from: <https://www.bertin-corp.com/tissue-homogenizer/39-precellys-lysing-kits-vk05.html>
49. Pierce™ BCA Protein Assay Kit [Internet]. [cited 2021 May 14]. Available from: <https://www.thermofisher.com/order/catalog/product/23225#/23225>

50. Trypsin Gold, Mass Spectrometry Grade | Trypsin Digestion [Internet]. [cited 2021 May 14]. Available from: <https://www.promega.com/products/mass-spectrometry/trypsin/trypsin-gold-mass-spectrometry-grade/?catNum=V5280>
51. Barocycler 2320EXT – Pressure BioSciences, Inc. [Internet]. [cited 2021 May 24]. Available from: <https://products.pressurebiosciences.com/products/barocycler-2320ext>
52. ID Biologics Kits Price List [Internet]. [cited 2021 May 14]. Available from: https://www.nestgrp.com/prices/id-bio_p/idbp.shtml
53. Gonzalez-y-Merchand JA, Zaragoza-Contreras R, Guadarrama-Medina R, Helguera-Repetto AC, Rivera-Gutierrez S, Cerna-Cortes JF, et al. Evaluation of the cell growth of mycobacteria using Mycobacterium smegmatis mc2 155 as a representative species. *Journal of Microbiology* [Internet]. *J Microbiol*; 2012 [cited 2021 May 15];50:419–25. Available from: <https://pubmed.ncbi.nlm.nih.gov/22752905/>
54. NP80 - Implen [Internet]. [cited 2021 May 14]. Available from: <https://www.implen.de/product-page/implen-nanophotometer-np80-microvolume-cuvette-spectrophotometer/>
55. UltiMate™ 3000 RSLCnano System [Internet]. [cited 2021 May 16]. Available from: <https://www.thermofisher.com/order/catalog/product/ULTIM3000RSLCNANO#/ULTIM3000RSLCNANO>
56. Thermo Scientific Q Exactive HF-X hybrid quadrupole-Orbitrap mass spectrometer Pushing the leading edge in protein analysis.
57. Nanospray Flex™ Ion Sources [Internet]. [cited 2021 May 16]. Available from: <https://www.thermofisher.com/order/catalog/product/ES071#/ES071>
58. Acclaim™ PepMap™ 100 C18 HPLC Columns [Internet]. [cited 2021 May 16]. Available from: <https://www.thermofisher.com/order/catalog/product/164567#/164567>
59. 6545 Quadrupole Time-of-Flight LC/MS, LC/Q-TOF MS | Agilent [Internet]. [cited 2021 May 24]. Available from: <https://www.agilent.com/en/product/liquid-chromatography-mass-spectrometry-lc-ms/lc-ms-instruments/quadrupole-time-of-flight-lc-ms/6545-lc-q-tof>
60. Tsugawa H, Cajka T, Kind T, Ma Y, Higgins B, Ikeda K, et al. MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nature Methods* [Internet]. Nature Publishing Group; 2015 [cited 2021 May 19];12:523–6. Available from: </pmc/articles/PMC4449330/>
61. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology* [Internet]. Nature Publishing Group; 2008 [cited 2021 May 24];26:1367–72. Available from: <http://www.nature.com/naturebiotechnology>

62. *Mycobacterium smegmatis* (strain ATCC 700084 / mc(2)155) (*Mycobacterium smegmatis*) [Internet]. [cited 2021 May 16]. Available from: <https://www.uniprot.org/proteomes/UP000000757>
63. Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature Protocols* [Internet]. Nature Publishing Group; 2016 [cited 2021 May 16];11:2301–19. Available from: <https://thermo.flexnetoperations.com/control/>
64. Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, et al. MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis. *Nucleic Acids Research* [Internet]. Oxford University Press; 2018 [cited 2021 May 16];46:W486–94. Available from: <http://old.metaboanalyst.ca>
65. Mi H, Ebert D, Muruganujan A, Mills C, Albu LP, Mushayamaha T, et al. PANTHER version 16: A revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Research* [Internet]. Oxford University Press; 2021 [cited 2021 May 16];49:D394–403. Available from: <http://www>.
66. Mi H, Thomas P. PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. *Methods in molecular biology* (Clifton, NJ) [Internet]. NIH Public Access; 2009 [cited 2021 May 16];563:123–40. Available from: </pmc/articles/PMC6608593/>
67. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* [Internet]. *Nucleic Acids Res*; 2009 [cited 2021 May 16];37:1–13. Available from: <https://pubmed.ncbi.nlm.nih.gov/19033363/>
68. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* [Internet]. *Nat Protoc*; 2009 [cited 2021 May 16];4:44–57. Available from: <https://pubmed.ncbi.nlm.nih.gov/19131956/>
69. Lai Z, Tsugawa H, Wohlgemuth G, Mehta S, Mueller M, Zheng Y, et al. Identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics. *Nature Methods* [Internet]. Nature Publishing Group; 2018 [cited 2021 May 20];15:53–6. Available from: <http://prime.psc.riken.jp/>
70. Tsugawa H, Kind T, Nakabayashi R, Yukihiro D, Tanaka W, Cajka T, et al. Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. *Analytical Chemistry* [Internet]. American Chemical Society; 2016 [cited 2021 May 20];88:7946–58. Available from: <http://prime.psc.riken.jp/>
71. Annotating a Gene with a Programmed Translational Frameshift [Internet]. [cited 2021 May 15]. Available from: <https://seaphagesbioinformatics.helpdocsonline.com/article-54>
72. *Gordonia* phage CloverMinnie, complete genome - Nucleotide - NCBI [Internet]. [cited 2021 May 15]. Available from: <https://www.ncbi.nlm.nih.gov/nucleotide/MN234196>

73. Roy A, Kucukural A, Zhang Y. I-TASSER: A unified platform for automated protein structure and function prediction. *Nature Protocols* [Internet]. Nature Publishing Group; 2010 [cited 2021 May 15];5:725–38. Available from: <http://zhanglab.ccmb.med>.
74. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER suite: Protein structure and function prediction [Internet]. *Nature Methods*. Nature Publishing Group; 2014 [cited 2021 May 15]. p. 7–8. Available from: <http://zhanglab.ccmb.med.umich.edu/I-TASSER/>
75. The PyMOL Molecular Graphics System. Schrödinger, LLC.;
76. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. Elsevier B.V.; 2015;1–2:19–25.
77. Eulenburg G, Higman VA, Diehl A, Wilmanns M, Holton SJ. Structural and biochemical characterization of Rv2140c, a phosphatidylethanolamine-binding protein from *Mycobacterium tuberculosis*. *FEBS Letters* [Internet]. FEBS Lett; 2013 [cited 2021 May 20];587:2936–42. Available from: <https://pubmed.ncbi.nlm.nih.gov/23907008/>
78. Robins DC. PHOSPHATIDYLETHANOLAMINE AND LYSOPHOSPHATIDYLETHANOLAMINE [Internet]. *Journal of Pharmacy and Pharmacology*. Oxford Academic; 1963 [cited 2021 May 20]. p. 701–22. Available from: <https://academic.oup.com/jpp/article/15/1/701/6216867>
79. Kobayashi N, Kume S, Lenz K, Masuya H. RIKEN MetaDatabase: A database platform for health care and life sciences as a microcosm of linked open data cloud. *International Journal on Semantic Web and Information Systems* [Internet]. IGI Global; 2018 [cited 2021 May 20];14:140–64. Available from: <https://dl.acm.org/doi/abs/10.4018/IJSWIS.2018010106>
80. Pedras MSC. In planta sequential hydroxylation and glycosylation of a fungal phytotoxin: Avoiding cell death and overcoming the fungal invader. *Proceedings of the National Academy of Sciences* [Internet]. *Proceedings of the National Academy of Sciences*; 2001 [cited 2021 May 23];98:747–52. Available from: </pmc/articles/PMC14659/>
81. New 13-carboxy-sterol cpd. ergokonin C - useful as antibiotic and antifungal agent, and as intermediate for steroid synthesis - Patent DE-4115490-A1 - PubChem [Internet]. [cited 2021 May 23]. Available from: <https://pubchem.ncbi.nlm.nih.gov/patent/DE-4115490-A1>
82. Henner WD, Kleber I, Benzinger R. Transfection of *Escherichia coli* Spheroplasts III. Facilitation of Transfection and Stabilization of Spheroplasts by Different Basic Polymers. *Journal of Virology*. American Society for Microbiology; 1973;12:741–7.
83. Pettit RK, Fakoury BR, Knight JC, Weber CA, Pettit GR, Cage GD, et al. Antibacterial activity of the marine sponge constituent cribrastatin 6. *Journal of Medical Microbiology* [Internet]. Microbiology Society; 2004 [cited 2021 May 23];53:61–5. Available from: <https://www.microbiologyresearch.org/content/journal/jmm/10.1099/jmm.0.05250-0>

84. Brill GM, McAlpine JB, Whittern DN, Buko AM. Altromycins, novel pluramycin-like antibiotics: II. Isolation and elucidation of structure. *The Journal of Antibiotics* [Internet]. *J Antibiot* (Tokyo); 1990 [cited 2021 May 23];43:229–37. Available from: <https://pubmed.ncbi.nlm.nih.gov/2324008/>