12-15-2021

Characterization and Proteolytic Maturation of Head Proteins in a Giant Salmonella Virus

Aaron Scheuch
ads6855@rit.edu

Characterization and Proteolytic Maturation of Head Proteins in a Giant *Salmonella* Virus

by

Aaron Scheuch

A Thesis Submitted in Partial Fulfillment Of The Requirements for the

Degree of Master of Science in Bioinformatics

Department of Life Sciences

College of Science

Rochester Institute of Technology

Rochester, NY

December 15, 2021

Rochester Institute of Technology

Thomas H. Gosnell School of Life Sciences

Bioinformatics Program

To: Head, Thomas H. Gosnell School of Life Sciences

The undersigned state that Aaron Scheuch, a candidate for the Master of Science degree in Bioinformatics, has submitted his thesis and has satisfactorily defended it.

This completes the requirements for the Master of Science degree in Bioinformatics at Rochester Institute of Technology.

Thesis committee members:

Name	**Date**
_____	_____
Julie A. Thomas, Ph.D. Thesis Advisor	
_____	_____
Michael V. Osier, Ph.D.	
_____	_____
Gary R. Skuse, Ph.D.	
_____	_____
_____	_____

Abstract

Giant phages are of interest for biocontrol of pathogenic bacteria, however, there is limited knowledge about their biology. We study the *Salmonella* phage SPN3US to better understand virion structure and function. SPN3US has a contractile tail and a large T=27 icosahedral capsid that contains its 240-kb genome. Previous analyses showed the SPN3US virion is highly unusual because it contains >80 different proteins, a number that is highly unusual for a tailed phage. In addition, there is a mass (>40 MDa) of proteins ("ejection proteins") within the head that enter the *Salmonella* cell, possibly with roles in host takeover at the onset of infection. However, there is limited knowledge of the composition of the mature particle, the roles of individual proteins and how the SPN3US head and virion assemble. To address this gap in knowledge this research characterized both wild-type phage and a tailless mutant of SPN3US using high performance mass spectrometry to more accurately define the head proteome.

These data confirm the high structural complexity of the SPN3US virion with it containing 92 different proteins. The head was found to contain 54 proteins, of which 9 were determined to have undergone proteolytic cleavage by a phage-encoded protease. All of these processed proteins were cleaved *C*-terminal to the sequence motif AXE, including the protease responsible for these cleavages. These data provide new insight into head maturation events during virion assembly and form a strong foundation for future studies on the roles of individual head proteins.

Overall, these experiments illustrate that mass spectrometry is a powerful tool for defining the composition of highly complex viral particles, including the identification of post-translational

modifications indicative of maturation events during viral assembly, and could be more broadly implemented in the field.

Acknowledgements

I would like to thank my committee members Julie Thomas, Gary Skuse, and Michael Osier, RIT student Sara Mallory for the purification of SPN3US phage virions, former RIT student Andrea Denisse BenÍtez Quintana for sequencing the amber mutant am107, Dr. Ru-ching Hsia for transmission electron microscopy imaging of am107 at the UMB Electron Microscopy Core. I would also like to thank Dr. Susan Weintraub along with Sammy Pardo and Dana Molleur for correspondence regarding the mass spectral analyses.

Table of Contents

Abbreviations Used

CsCl- Cesium Chloride

PSM- Peptide Spectrum Match

gp- Gene Product

kb- Kilobase pair (kbp)

NCBI- National Center for Biotechnology Information (Bioinformatics Database)

BLAST- Basic Local Alignment Search Tool software

LCMS- Liquid Chromatography Mass Spectrometry

LB- Luria Broth

sup+- Permissive or Suppressor Strain of *Salmonella*

sup- Non-permissive or Non-suppressor Strain of *Salmonella*

vRNAP- virion RNA polymerase

TEM- Transmission Electron Microscope

Chapter 1: Introduction

Bacteriophages abundantly affect natural ecosystems, with estimates at over 10^{30} to 10^{31} currently on Earth. They are also a common nuisance in industrial fermentations, giving them the capacity to cause significant economic damage. Bacteriophages have been receiving heightened attention in recent years as an alternative to antibiotics, due to the increasing resistances being evolved against the latter. Early applications of bacteriophages as an antibacterial agent have often focused on agriculture (Clokie et al., 2011). Giant phage PhiEAH2 has been used to control fireblight (Figure 1.1), a disease affecting orchards caused by *Erwinia amylovora* (Grace et al., 2021). Phage therapy for medicinal use continues to develop, with current obstacles including the understanding of their gene functions for the purpose of genetics.

Figure 1.1: Erwiphage Plus Product for Fireblight Treatment

Bacteriophages possess a highly distinct structure. The top of the phage is an icosahedral-shaped capsid known as the head which stores the densely packed viral genome as well as several viral proteins. The proteins comprising the exterior shell structure of the head are assembled together like building blocks. The total number of these proteins required to assemble the head structure is a characteristic unique to each phage known as the copy number. This is a highly variable characteristic of phages, differing by several orders of magnitude even within a single phage. Proteases within

the head are responsible for processing of other viral proteins during assembly. Extending out from the bottom of the head is the tail which the genome passes through to enter the host cell (Marvin, 1998). Finally, at the bottom of the tail is the baseplate which facilitates binding to the host through the specific recognition of proteins on the host cell. This general structure is a hallmark of most bacteriophages.

Giant, or jumbo, phages are described as bacteriophages distinguished by genomes over 200 kilobases in length (Sharma et al., 2019) (see examples in Table 1.1). The first discovered and sequenced giant phage, PhiKZ, was originally discovered in 1978 (Krylov and Zhazykov, 1978). Giant phages in the past have been difficult to isolate, largely due to traditional methods of isolating bacteriophages. Giant phages can get caught up in traditional micron filters that smaller bacteriophages would be able to pass through. Giant phages also struggle to form plaques with standard agar concentrations and are now generally cultured in agar overlays with lower concentrations of agar. There are currently 56,573 tailed phage genomes listed on the NCBI. Of these genomes, only 489 are above 200kb in length, enough to qualify as giant phages (NCBI Virus (nih.gov)). The largest giant phage currently known, excluding metagenomic reconstructions, is *Lysinibacillus*-infecting Phage G, with a total genome length of 497 kb. There is still much progress to be made in the sequencing of giant phage genomes, and this extends to proteomic data.

Table 1.1: Summary of Giant Phage Genomic & Proteomic Data. Cells with '-' indicate data that were not available.

Phage Name	Genome Accession	Genome Size (kbp)	Genes	Host	Purification Method	Proteins Identified	Reference
SPN3US	NC_027402.1	240	264	*Salmonella enterica*	single CsCl	86	(Ali et al., 2017)
201phi2-1	EU197055	317	-	*Pseudomonas chlororaphis*	double CsCl	89	(Thomas et al., 2010)
PhiKZ	NC_004629.1	280	306	*P. aeruginosa*	double CsCl	62	(Lecoutere et al., 2009)
EL	AJ697969	-	-	*P. aeruginosa*	double CsCl	64	
BpSp	AKU29927.1	256	318	*Bacillus*	Sucrose density gradient centrifugation	23	(Yuan and Gao, 2016)
RAY	KU886224	271	-	*Erwinia amylovora*	-	32	(Sharma et al., 2019)
Deimos-Minion	KU886225	274	-	*Erwinia amylovora*	-	27	
AR9	NC_031039.1	251	292	*Bacillus*	Concentrated by PEG, Purified by CsCl	43	(Lavysh et al., 2016)
KTN4	KU521356.1	280	-	*P. aeruginosa*	-	111	(Danis-Wlodarczyk et al., 2016)

Structural proteins in giant phages are often conserved, but have diverged such that there is no sequence similarity at the DNA level. The major capsid protein of the *E. coli* virus HK97, gp5, has a structural fold which is conserved across most giant phages. However, the proteomic sequences comprising this fold have the ability to vary quite a bit (Krupovic and Koonin, 2017).

There are three primary classical structural families of bacteriophages: Myoviridae, Siphoviridae, and Podoviridae. All adhere to the structure of an icosahedral head with a tail and fibers which facilitate binding to the cell (Figure 1.2). The length of the tail can vary quite significantly however, with it being truncated significantly in Podoviridae.

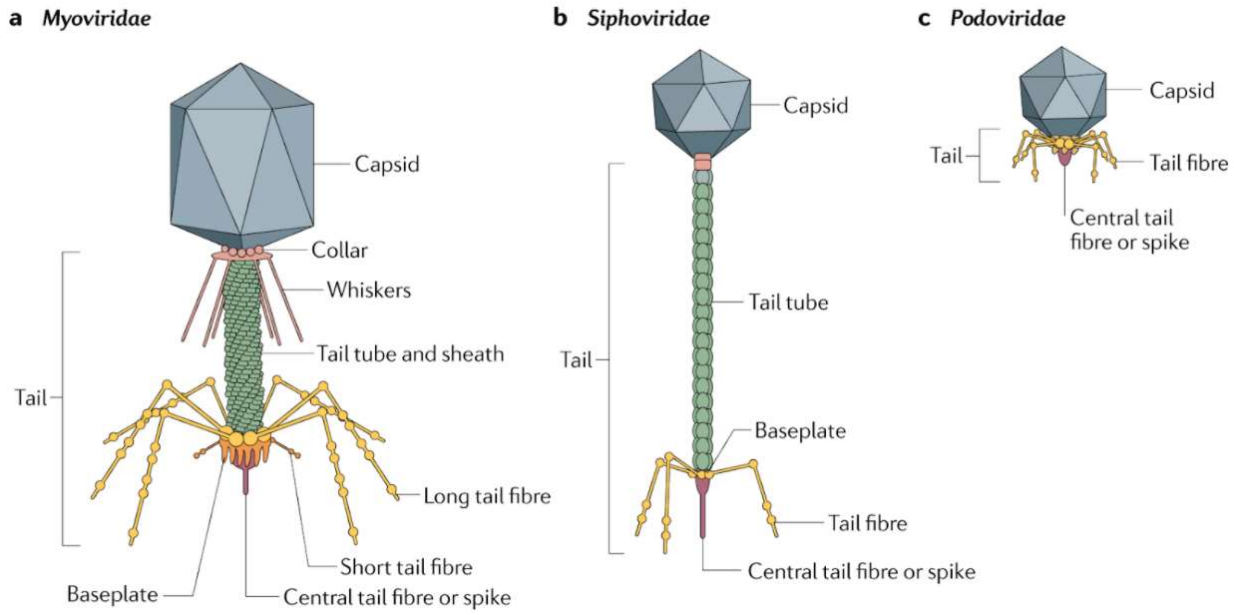


Figure 1.2: Structural Families of Bacteriophages Reproduced from *Nature Reviews Microbiology* Volume 16, pages 760-773 (2018) (Nobrega et al., 2018)

SPN3US is a giant phage which classically infects *Salmonella enterica*. It has close genetic similarity with a variety of other giant phages, including PhiKZ. The genome of SPN3US has also been fully sequenced with a length of 240kb and characterized into roughly 264 gene products (Lee et al., 2011). Only a few gene functions in SPN3US are known. Gp75 is the major capsid protein and by far the most abundant in the proteome, forming the bulk of the virion. Gp81 is the portal protein which joins the capsid to the tale. The protease, gp245, is responsible for proteolytic processing of structural proteins. Paralogs gp53 and gp54 are also highly abundant and suspected to play a role in capsid formation (Ali et al., 2017). SPN3US has an "inner body", which is a cylindrical protein core positioned down the center of the capsid, which contains the genome. The function of the inner body is not clear. The first intermediate structure in virion formation is the prohead, a rudimentary core which will develop into the mature head. The tail is then produced and finally attached to the mature head through head-tail joining proteins (Figure 1.3).

Because there are suppressive strains of *Salmonella*, it is possible to do genetics with phages that infect this host. The ability to do genetics allows for us to grow phages under conditions where one or more of their genes is prematurely truncated. If this is an essential gene, the virus will not have a productive infection, and we can study the role of that protein in infection. The proteome of SPN3US can therefore be well-manipulated and characterized. With better gene identification and functional characterization of the proteome of SPN3US, homology could be used to learn more about other, less well-characterized giant phages.

The goal of this thesis is to greatly improve the characterization of the proteome of SPN3US through the use of mass spectrometry and existing genomic and proteomic data. There are two core data sets used in this research: one from an Orbitrap mass spectrometer, and another from Lumos, a much higher sensitivity mass spectrometer. This data can be used to characterize myriad things, including structural proteins, processing sites, copy number, and total identified proteins in SPN3US. From this data, we have determined:

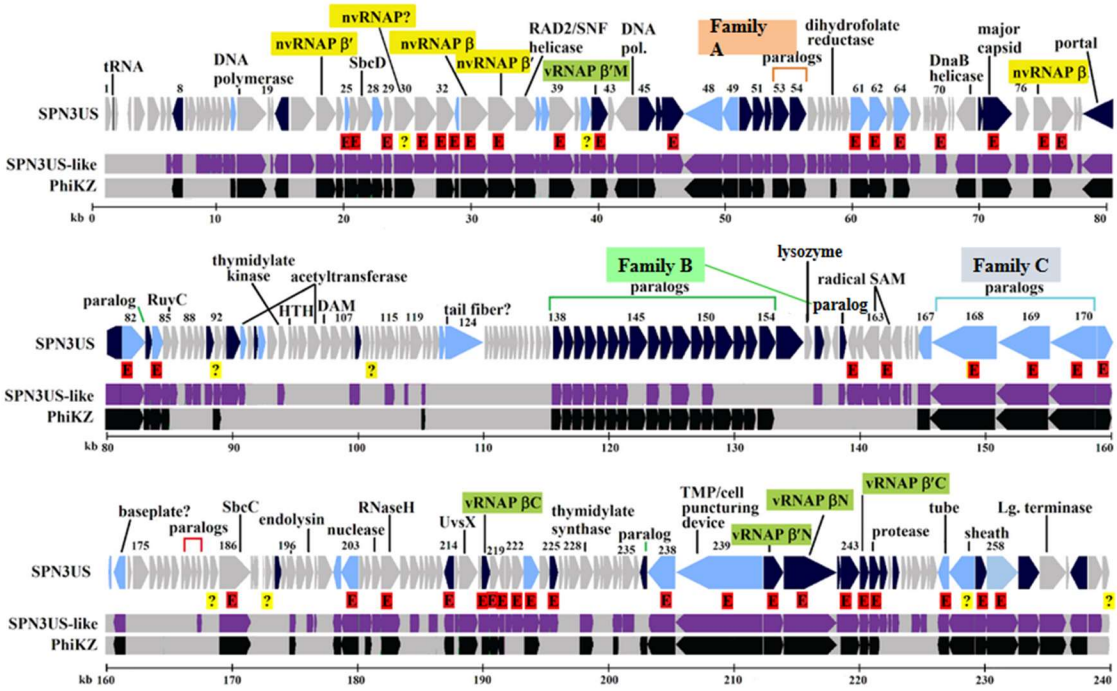
- Relative abundance of SPN3US proteins
- Head vs. Tail proteins
- Identification of low abundance proteins using the high sensitivity of Lumos
- Cleavage sites for all processed proteins, with corresponding molecular weight of the mature fragment
- Copy number estimation based on existing known copy number proteins (gp75, gp81, gp140, gp161, gp53, and gp64) extrapolated to remaining proteins
- The impact of *Salmonella* proteins on SPN3US infection: those which are differentially expressed in WT versus tailless samples, and which may be essential to infection

There are a variety of testing methodologies which have been used for this research. This includes the step gradient plus buoyant density gradient methods used to isolate SPN3US virions, SDS-page gels, bioinformatics tools such as PSI-BLAST and secondary structure prediction, mass spectral data of the experiments, and finally data analyses performed on that mass spectral data.

Chapter 2: Characterization of the SPN3US Wild-type Proteome

Giant phages possess very large capsids and extraordinarily complex genomes relative to other bacteriophages. The 240,413 bp genome of *Salmonella* phage SPN3US contains genes that encode a total of 264 hypothetical proteins (Lee et al., 2011) (Figure 2.1) and exhibits varying degrees of genetic similarity to many other giant phages (M Iyer et al., 2021). The increasing numbers of related giant phages being used for biotechnological applications indicates there is likely to be great value in understanding SPN3US's proteome further as the results can be used to understand the roles of homologous proteins in related giant phages. Of SPN3US's gene products, only a limited number of them have any hypothesized functions (Figure 2.1). Protein functions in this genome map were assigned using a combination of genetics, TEM, and mass spectroscopy. Similarly, almost all the giant phages related to SPN3US have poorly characterized gene functions, which has limited their ability to be used and genetically engineered for therapeutic use.

Figure 2.1: Genome Map of SPN3US Comparative with PhiKZ Genome



The majority of giant phages that have been isolated to date, including SPN3US, have contractile tails (i.e., they are myoviruses). There are four structural proteins which are always conserved in all myoviruses: the capsid, the sheath, the tube, and the portal. The capsid protein is typically by far the most abundant protein, forming the protective outer shell of the capsid or head. The sheath and tube proteins are also typically highly abundant proteins and are the main components of the tail. The sheath forms a helical structure on the exterior of the tail and is responsible for the tail contraction that is the identifying feature of all myoviruses. The tail tube forms a cylindrical tube within the sheath and upon infection and tail contraction has been observed to penetrate the host cell in some myoviruses (e.g. T4) (Hu et al., 2015). The portal

protein is present in a low copy number in the virion, and in all studied phages has been shown to form a ring of 12 subunits referred to as the portal or connector complex. The portal complex occupies one of the capsid vertices and is effectively the connection between the head and tail. The portal protein is multi-functional, having roles in head assembly, DNA packaging, tail assembly, and also helps facilitate the ejection of the genome from the head through the tail and into the host cell.

Historically, methods to identify phage virion proteins relied on SDS-PAGE migration and N-terminal sequencing. Mass spectrometry presents an alternative for the high-quality characterization of virion proteins that is being used more frequently to understand the composition of the virions of different phages. Mass spectrometry analyses is particularly helpful for the identification of the virion proteins of giant phages, as giant phage virions are complex and comprised of many different proteins relative to smaller phages. Typically, giant phage virion proteins will vary in number per virion by up to two orders of magnitude which makes the identification of all the virion components by SDS-PAGE alone impossible. Additionally, there are so many proteins that they cannot be visualized individually as bands on an SDS-PAGE gel.

Table 2.1: Examples of Identified Proteins of Various Giant Phages

Phage Name	Proteins Identified	Reference
SPN3US	86	(Ali et al., 2017)
201phi2-1	89	(Thomas et al., 2010)
PhiKZ	62	(Lecoutere et al., 2009)
EL	64	
BpSp	23	(Yuan and Gao, 2016)
RAY	32	(Sharma et al., 2019)
Deimos-Minion	27	
AR9	43	(Lavysh et al., 2016)
KTN4	111	(Danis-Wlodarczyk et al., 2016)

These data (Table 2.1) reflect the proteome sizes as they were understood at the time of their discovery. However, increases in mass spectral sensitivity could still potentially reveal additional proteins in these proteomes.

However, mass spectrometry has still not been utilized to its full potential for giant phage analyses. In fact, there are no standardized guidelines for such analyses, such as the recommended level of purity of the phage sample to be analyzed, suitable proteomic analyses and even the number of replicates that should be performed. For instance, no group has yet employed an analysis of highly purified virions using a high resolution, so-called “third generation”, mass spectrometer.

Similarly, no group has ever published replicate mass spectral analyses of their phage’s virion, meaning the reproducibility of the analyses is unknown. We have previously identified 86

different proteins in the virion of phage SPN3US (Ali et al., 2017), however these analyses were performed on a sample that had only undergone a short purification due to concerns as to virion stability during purification, no replicates were performed and the analyses were conducted on an Orbitrap mass spectrometer. Since then, our studies have shown that the phage is viable after more extensive purification. 86 different proteins have been identified in SPN3US. It is unclear if these proteins exist for virion structure, or to play a role in host takeover. To resolve whether all the 86 proteins (and possibly others) identified previously are truly part of the SPN3US virion, in these studies we sought to perform replicate mass spectral analyses of more highly purified SPN3US on an Orbitrap Fusion Lumos Tribrid mass spectrometer. Although the Orbitrap mass spectrometer is a relatively sensitive and commonly used mass spectrometer for proteomic analyses, the Lumos has up to twice the sensitivity in comparison to the Orbitrap mass spectrometer (Levy et al., 2018). The higher sensitivity of the Lumos mass spectrometer would be expected to facilitate identification of proteins present in the virion in very low abundance protein that would otherwise be missed.

Among the PhiKZ-related giant phages, there is an exclusive class of proteins which is conserved: virion RNA polymerases (vRNAPs). This is suggestive of these giant phages having a more independent life cycle relative to other phages, having the ability to transcribe their own genome (Ceyssens et al., 2014). These proteins consist of 5 subunits which assemble together to form the final polymerase.

Previous mass spectral analyses on the suppressed mutant am27 of SPN3US showed a total of 82 identified proteins. The most abundant was the major capsid gp75 at 1723 total peptide

spectrum matches (PSM), followed by gp53 at 708 total PSM, gp56 at 686 total PSM, and gp54 at 681 total PSM.

Methods

Preparation of Purified SPN3US Virions

A high titer stock of wild-type SPN3US was prepared from a single plaque that had been resuspended in SM buffer. SM buffer is composed of: 100mM NaCl, 8mM $MgSO_4 \bullet 7H_2O$, and 50mM Tris-Cl (1 M, pH 7.5). Briefly, ~10^5 pfu of the phage suspension was plated in a soft LB overlay containing 0.34 % agar and 1 mM of $MgCl_2$ and 1 mM of $CaCl_2$ on LB bottom plates. Each overlay also contained *Salmonella enterica* Typhimirium strain TT9079 that had been propagated to exponential phase in LB broth supplemented with nutrient broth (LB+N) and 1 mM of $MgCl_2$ and 1 mM of $CaCl_2$. Sixteen plates were made in total and incubated overnight at 30 °C. The next day the overlays were harvested and diluted approximately 2-fold in SM buffer with lysozyme (~2 mg/ml final concentration) and left overnight at 4 °C. The following day, the stock was clarified and concentrated by differential centrifugation (7500 rpm, 10 min and then 18,000 rpm, 30 min, at 4 °C in a Beckman JA25.50 rotor) and the phage pellet resuspended in SM buffer.

The following day, the resuspended phage sample underwent cesium chloride (CsCl) step gradient and then CsCl buoyant density gradient purification. Phage samples (800 µl) were layered onto CsCl step gradients composed of the following concentrations of CsCl: 1.59 g/ml (1 ml), 1.52 g/ml (1 ml), 1.41 g/ml (0.9 ml), 1.30 g/ml (0.9 ml), and 1.21 g/ml (0.9 ml). The buffer used throughout the gradient was 10 mM Tris-HCl (pH 7.5) and 1 mM $MgCl_2$. The tubes were

spun at 30,000 rpm for 3 h at 4°C in an SW50.1 rotor (Beckman Coulter ultracentrifuge), and the resulting bands were harvested by side tube puncture. The refractive index of each sample was measured using a refractometer, and then the sample was added to a freshly prepared solution of 10 mM Tris-HCl (pH 7.5) and 1 mM $MgCl_2$ containing CsCl at the refractive index of each sample. The buoyant-density gradients then underwent overnight centrifugation at 30,000 rpm at 4°C. Samples were again collected by side tube puncture, and the refractive index was recorded and then dialyzed against three changes of 50 mM Tris-Cl (pH 7.5), 200 mM NaCl, and 10 mM $MgCl_2$ to remove all remaining cesium chloride.

Samples were boiled for 10 minutes in SDS sample buffer prior to electrophoresis on Criterion XT MOPS 12 % or 4-12 % gradient SDS-PAGE reducing gels (Bio-Rad) and proteins visualized by staining with Coomassie blue. Gel lanes were divided into six slices for GeLCMS analyses of each sample. Samples were analyzed in triplicate for the wild-type phage. After de-staining, proteins in the gel slices were reduced with TCEP [tris(2-carboxyethyl)phosphine hydrochloride] and then alkylated with iodoacetamide in the dark before digestion with trypsin (Promega). HPLC-electrospray ionization-tandem mass spectrometry (HPLC-ESI-MS/MS) was performed on a Thermo Fisher Lumos mass spectrometer. Mascot (Matrix Science; London, UK) was used to search the MS files against a locally generated SPN3US and *Salmonella* protein database that had been concatenated with the SwissProt database (2012_11_170320; version 51.6). Subset searching of the Mascot output by X! Tandem, determination of probabilities of peptide assignments and protein identifications, and cross correlation of the Mascot and X! Tandem identifications were accomplished by Scaffold (Proteome Software). MS data files were either

processed individually or the files for an entire gel lane were combined via the “MudPIT” option in Scaffold.

The results for identified proteins, numbers of unique peptides, total spectra, and sequence coverage for each experiment were exported from Scaffold with the following quality filters: peptide, 95 %; protein, 99.9 %; minimum number of peptides, 2. An estimate of the relative abundance of SPN3US virion proteins was obtained by dividing the total number of spectra assigned to each protein (Peptide Spectrum Matches, PSM) identified by MudPIT analyses by its molecular mass (PSM/ M) as performed for SPN3US and related phages. For molecular weight calculations, the mature molecular weight (that is, the weight of the protein remaining after proteolytic processing, see chapter 4) was used. Our previous results have demonstrated that PSM/M provides a useful indicator of relative abundance of virion proteins. That is, proteins with similar PSM/M values are typically present in similar relative abundances in the virion; proteins with PSM/M ≤ 1 are likely to be present in only few copies, or even less than one copy, per virion.

For comparison purposes, Wild-type SPN3US data from an earlier experiment is referenced (Ali et al, 2017). In that experiment, the phage was purified via only a single CsCl step gradient and was analyzed on an Orbitrap mass spectrometer (no replicates were performed).

Various data manipulations were performed on the data. Summations of total PSM across whole experiments were done in order to compare relative sensitivities. These summations were filtered to show either exclusively SPN3US proteins, or exclusively *Salmonella* proteins in

each experiment. Tests across the three Lumos replicates of wild-type SPN3US were performed in order to determine in which proteins there is the greatest confidence.

Results and Discussion

In these studies, a high titer stock of SPN3US was purified through step and then buoyant density CsCl gradients (Figure 2.2).

Figure 2.2: Images of Viral Purification Steps

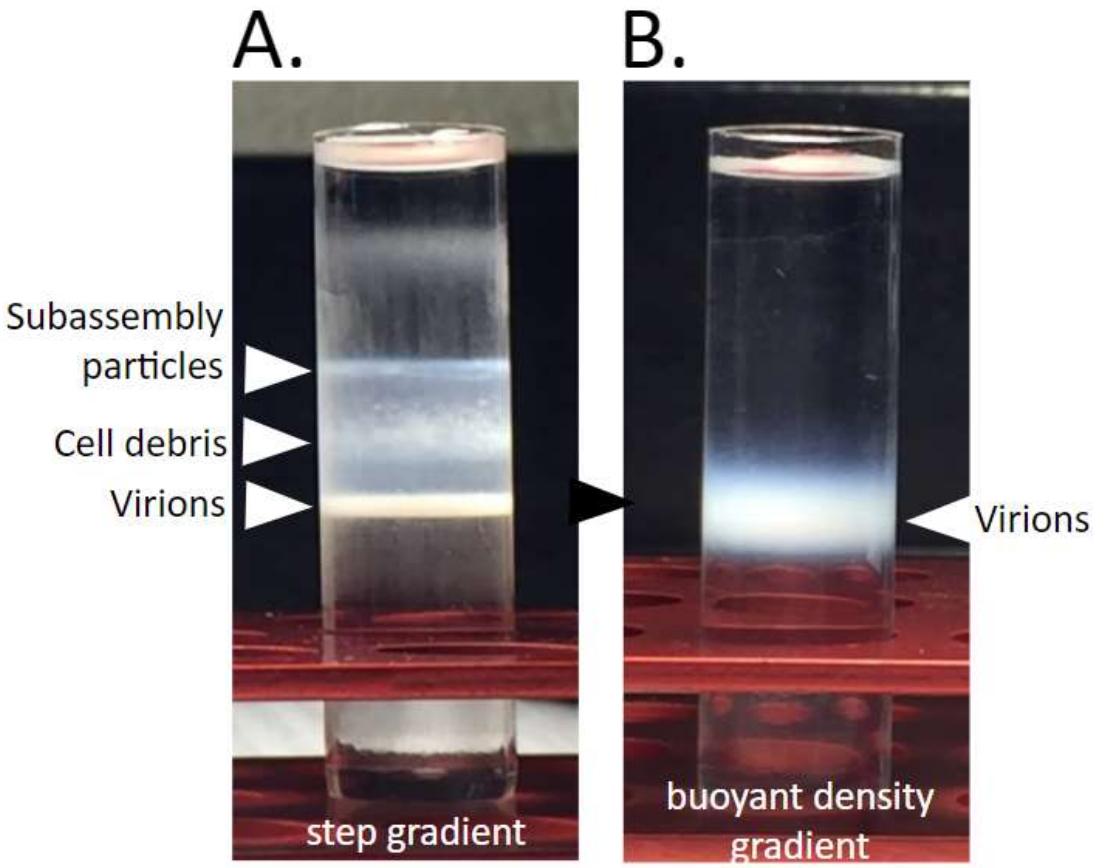


Table 2.2: Summary of SPN3US Proteins Detected in Mass Spectral Data

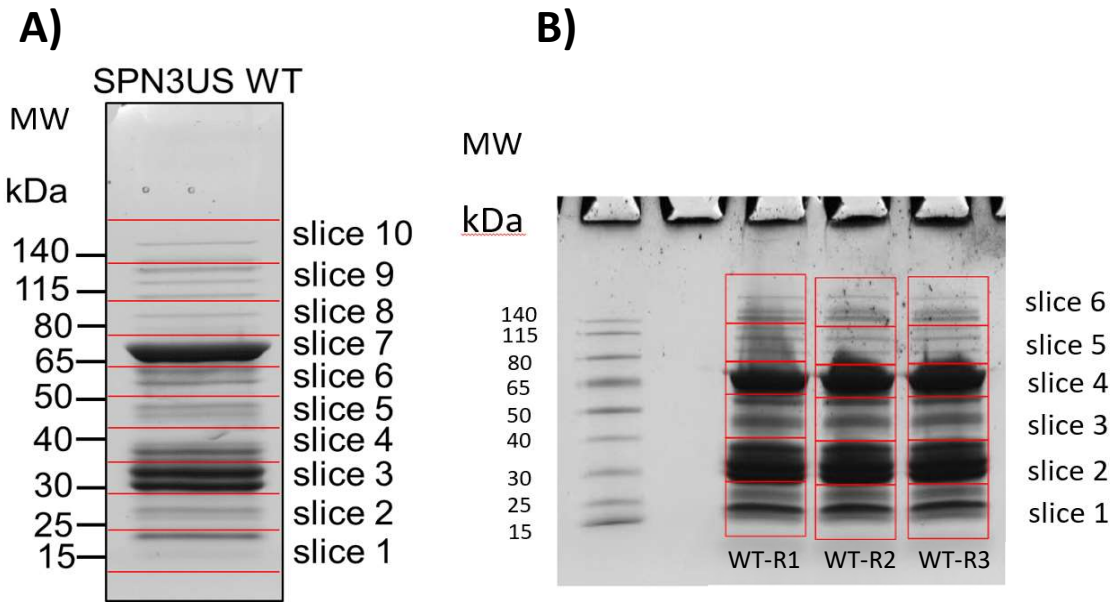
Purification Method	**Mass spectrometer/Experiment**	**Total PSM in Experiment**	**Total # of Different SPN3US Proteins Identified**
Single CsCl Step gradient	Orbitrap Wild-type	8666	86
Dual CsCl gradients	Lumos Wild-type Replicate A	19313	91
	Lumos Wild-type Replicate B	18563	91
	Lumos Wild-type Replicate C	19053	92

The higher sensitivity Lumos mass spectrometer allowed for almost twice the total detected PSMin each experiment and up to 6 additional identified proteins (Table 2.2).

Of the SPN3US proteins there are a limited number of SPN3US virion proteins with known functions. These include gp75, which was identified using bioinformatics as the major capsid protein and therefore makes up the exterior shell of the head. Consistent with this role, previous mass spectral studies showed gp75 to be the most abundant protein overall in the virion (Ali et al., 2017). A gel LCMS of SPN3US found the most highly abundant protein was gp75, consistent with its assignment as the major capsid protein. In addition, SPN3US also has

53, 54, 255, 256 as highly abundant proteins estimated to be present in the virion at several 100 copies. At least 270 copies of 255 and 256 are present as well. Gp75 is a high molecular weight protein, at 80.1 kDa. Gp81 is the portal protein, facilitating ejection of the genome into the host *Salmonella*. While highly essential to the virion, it is fairly low in abundance: comprising about 1% of total PSM in each experiment. This is much less abundant than gp75, which comprises about 18% of total PSM in each experiment. There are two paralog families: Family A consists of gp53 and gp54, which are highly abundant proteins. In the standard sensitivity Orbitrap experiments, gp53 and gp54 are in similar abundance. However, on the higher sensitivity Lumos, gp53 has roughly twice the total PSM relative to gp54 in all three experiments. Paralog family B consists of gp138 through gp154. These proteins have considerably lower abundances relative to Family A.

Figure 2.3: SDS-PAGE Electrophoresis of Purified Wild-type (WT) *Salmonella* phage SPN3US used for Mass Spectral Analyses. (A) SPN3US proteins after the phage was purified through a step gradient. In this experiment the gel lane was excised into ten slices. No replicate analyses were performed. (B) SPN3US purified through cesium chloride step gradient, followed by a buoyant density gradient, featuring 3 replicates (WT-R1, WT-R2, and WT-R3). Gel slices excised for mass spectral analyses are indicated.



Between the two gels from the wild-type experiments (Figure 2.3), there is overall much more protein loaded onto the gel in B, which becomes very apparent in mass spectral analyses. The overall banding pattern, however, is very similar between the two.

The lowest total PSM of any protein detected in the Orbitrap data, gp49 at 3 total PSM, had an average total PSM of 21 (Table 2.3) across the Lumos replicates. The protein with the highest total PSM, the major capsid protein gp75, had a total PSM of 1723 in the Orbitrap experiment versus an average of 4733 total PSM across the Lumos replicates.

The wild-type experiments on the Lumos mass spectrometer exhibited much greater sensitivity relative to the wild-type experiment on the Orbitrap mass spectrometer. This is most apparent in the total PSM and total identified proteins in each experiment, with each Lumos replicate having over twice the total PSM compared to the am27 Orbitrap experiment. 5 additional proteins were discovered in the am107 Lumos experiments: gp28, gp33, gp37, gp43, gp100, and gp122. All of these proteins are very low abundance, with total PSMs comprising less than 0.10% of all PSMs in the experiment. It's also made clear how high-quality the mass spectral data is: across all three replicates of the Lumos data, at least 90 proteins were consistent among all of them. This is indicative of the high quality of the mass spectral data. In future experiments, it is reasonable to conclude a single replicate would be sufficient for gathering reliable proteomic data.

The Lumos mass spectrometer was easily able to detect proteins across three orders of magnitude, with total PSM values ranging from the single digits, for very low abundance proteins, to the thousands, such as in the major capsid gp75.

A protein's copy number is the number of copies it possesses in the final structure of the virion. Use total PSM counts from mass spectral data as well as a protein with a known copy number, it is possible to estimate the copy numbers of unknown proteins (Table 2.4). For these calculations, the total PSM of each protein in the experiment was normalized by mature molecular weight. The known copy number for a single protein was then divided by its total PSM divided by its mature molecular weight, to get a multiplication factor. The PSM/MW of proteins could then be multiplied this factor to estimate a copy number. While this method does result in fairly unreliable copy numbers being estimated as shown, it can be useful for preliminary estimations. For future research, once more copy numbers have been confirmed experimentally, it could then be better determined which protein would be most appropriate as a template.

Including only proteins that appeared in at least 2 of 3 wild-type replicates, 92 unique proteins were found to exist in wild-type SPN3US.

Table 2.3: SPN3US Proteins Identified in Wild-Type (WT) Purified Virions. Phage samples were either purified through a single cesium chloride step gradient and analyzed on the Orbitrap mass spectrometer (single-gradient Orbitrap) or through consecutive step buoyant density cesium chloride gradients and analyzed on the Lumos mass spectrometer. Total peptide spectrum matches (PSM) are listed for each experiment and the average and standard deviation provided for the triplicate samples analyzed on the Lumos mass spectrometer. Only proteins identified in at least 2 of the 3 replicates in the Lumos double gradient-analyzed samples are included in this table (see Appendix Table A.1 for full list of proteins).

			Single-gradient Orbitrap	Double-gradient Lumos-a	Double-gradient Lumos-b	Double-gradient Lumos-c	Double-gradient Lumos Average	Double-gradient Lumos Standard Deviation	Putative function
gp	MW (kDa)	Mature MW (kDa)	Total PSM	Total PSM	Total PSM	Total PSM	Average PSM	SD PSM	
8	30.3	30.3	32	23	39	35	32.3	8.3	
17	15.3	15.3	13	31	26	27	28.0	2.6	
21	40.7	40.7	20	27	23	21	23.7	3.1	
25	14.6	14.6	6	7	10	9	8.7	1.5	
26	24.9	24.9	0	0	0	0	0.0	0.0	
28	32	32	0	5	3	5	4.3	1.2	
33	12.3	12.3	0	0	3	0	1.0	1.7	
37	14.6	14.6	0	4	6	7	5.7	1.5	new candidate virion protein
38	23.2	23.2	4	6	4	7	5.7	1.5	
41	31.8	31.8	8	7	19	11	12.3	6.1	
42	49.3	49.3	52	54	63	77	64.7	11.6	vRNAP β′M
43	19	19	0	0	2	0	0.7	1.2	
45	50.3	48.2	144	239	224	253	238.7	14.5	
46	16.2	16.2	17	49	45	49	47.7	2.3	
47	62.8	50.7	110	358	345	253	318.7	57.2	
48	111	111	125	290	280	318	296.0	19.7	

49	48.1	48.1	7	18	20	26	21.3	4.2	
50	39.4	25.6	56	212	182	200	198.0	15.1	
51	34.9	34.9	81	117	100	108	108.3	8.5	
52	21	21	30	133	130	136	133.0	3.0	
53	45.2	31.5	662	1987	1873	1967	1942.3	60.9	ejection proteins (paralogs)
54	45.1	31.9	643	1088	1021	1071	1060.0	34.8	
61	58.3	58.3	35	63	65	68	65.3	2.5	
62	52	52	44	79	65	70	71.3	7.1	
64	48.9	48.9	20	39	47	41	42.3	4.2	
73	58.6	58.6	0	0	0	0	0.0	0.0	
74	12.9	12.9	40	141	125	134	133.3	8.0	
75	80.1	70.4	1592	4806	4666	4726	4732.7	70.2	major capsid
81	100.2	72.3	93	177	176	180	177.7	2.1	portal
82	84.8	84.8	36	107	103	108	106.0	2.6	
83	20	20	30	86	96	94	92.0	5.3	
84	32.9	32.9	51	83	81	79	81.0	2.0	
91	23.6	23.6	13	26	33	31	30.0	3.6	
94	41.6	41.6	35	37	33	34	34.7	2.1	
95	17.6	17.6	15	26	21	26	24.3	2.9	
97	11.9	11.9	7	8	9	8	8.3	0.6	
98	22.4	22.4	10	0	2	4	2.0	2.0	
100	23.2	23.2	0	3	3	4	3.3	0.6	
103	24.1	24.1	0	0	0	0	0.0	0.0	
104	37.6	37.6	0	0	0	0	0.0	0.0	
108	10.6	10.6	0	0	0	0	0.0	0.0	
109	17.7	17.7	40	44	39	32	38.3	6.0	
111	11.9	11.9	0	2	0	2	1.3	1.2	
122	18.8	18.8	0	3	2	4	3.0	1.0	
123	16.5	16.5	7	11	9	16	12.0	3.6	

124	113.7	113.7	73	103	117	102	107.3	8.4	
126	10.7	10.7	0	0	0	0	0.0	0.0	
138	29.2	29.2	35	70	57	77	68.0	10.1	
139	29.8	29.8	37	56	55	58	56.3	1.5	
140	31.4	31.4	41	54	40	42	45.3	7.6	
141	32.6	32.6	306	561	487	521	523.0	37.0	
142	30.8	30.8	65	83	80	83	82.0	1.7	
143	31.9	31.9	118	248	249	246	247.7	1.5	
144	30	30	83	119	118	103	113.3	9.0	
145	50.8	50.8	46	37	39	42	39.3	2.5	
146	36.9	36.9	31	29	25	34	29.3	4.5	
147	33.7	33.7	17	33	31	31	31.7	1.2	
148	53.7	53.7	53	105	108	130	114.3	13.7	
149	36.2	36.2	78	125	111	103	113.0	11.1	
150	33.8	33.8	29	38	32	37	35.7	3.2	
151	52.2	52.2	25	58	55	55	56.0	1.7	
152	36.8	36.8	69	103	75	90	89.3	14.0	
153	33.8	33.8	26	40	46	39	41.7	3.8	
154	50	50	45	69	79	88	78.7	9.5	
155	78.3	78.3	55	154	153	162	156.3	4.9	
156	23.6	23.6	0	0	0	0	0.0	0.0	
157	24.5	24.5	9	2	2	2	2.0	0.0	
158	19.9	19.9	4	3	6	8	5.7	2.5	
160	18.5	18.5	175	539	479	499	505.7	30.6	
161	25.9	25.9	0	0	0	0	0.0	0.0	
162	21.9	21.9	0	0	0	0	0.0	0.0	
163	50.3	50.3	0	0	0	0	0.0	0.0	
167	6.9	6.9	85	171	167	210	182.7	23.8	
168	188.1	188.1	203	329	366	374	356.3	24.0	
169	149	149	210	377	395	419	397.0	21.1	

170	135.4	135.4	135	246	236	246	242.7	5.8	
171	47.6	47.6	35	54	56	53	54.3	1.5	
172	7	7	6	0	0	0	0.0	0.0	
173	34.6	34.6	8	5	8	6	6.3	1.5	
174	12	12	0	2	0	2	1.3	1.2	
176	22.3	22.3	0	0	0	0	0.0	0.0	
178	20	20	0	0	0	0	0.0	0.0	
193	19.9	19.9	15	45	32	40	39.0	6.6	
199	22.4	22.4	0	0	0	2	0.7	1.2	
202	23.5	23.5	32	43	43	52	46.0	5.2	
203	51.9	51.9	72	164	157	160	160.3	3.5	
214	28.1	28.1	79	118	154	137	136.3	18.0	
218	25.2	25.2	27	42	47	45	44.7	2.5	vRNAP βC
223	45.3	45.3	29	48	44	54	48.7	5.0	
225	25.1	22.3	14	75	78	67	73.3	5.7	
229	39	39	0	0	0	0	0.0	0.0	
233	17.8	17.8	0	0	0	0	0.0	0.0	
237	19.9	19.9	50	54	68	54	58.7	8.1	
238	82.1	82.1	46	84	90	72	82.0	9.2	
239	259.1	259.1	200	362	416	400	392.7	27.7	
240	59.6	59.6	33	103	114	117	111.3	7.4	vRNAP β'N
241	159.1	159.1	189	285	278	315	292.7	19.7	vRNAP βN
242	10.5	10.5	7	16	17	18	17.0	1.0	
243	54.6	54.6	251	471	419	448	446.0	26.1	
244	27	27	12	15	27	23	21.7	6.1	vRNAP β'C
245	30.7	23.4	25	63	73	65	67.0	5.3	
246	23.9	23.9	22	80	67	68	71.7	7.2	
248	21	21	17	25	24	24	24.3	0.6	
249	15.3	15.3	0	0	0	0	0.0	0.0	
255	32.7	32.7	209	406	369	370	381.7	21.1	tube

256	75.7	75.7	669	1847	1686	1686	1739.7	93.0	sheath
257	34.2	34.2	97	145	123	132	133.3	11.1	
258	96.4	96.4	76	196	209	199	201.3	6.8	
259	61	61	115	348	305	316	323.0	22.3	
262	52.7	52.7	46	69	58	86	71.0	14.1	

Table 2.4: Copy Number Estimation Calculations

2019 Lumos Wild-type Data, 99.9% Protein Threshold, 2 minimum peptides, 95% Protein Threshold

gp	Function	Expected Copy Number	Copy Number major capsid (gp75, 24)	Copy Number portal (gp81, 5.43)	Copy Number sheath (gp255, 23.12)	Copy Number tube (gp256, 11.74)
75	Major Capsid	1615	1613.29	365.01	1554.14	789.17
81	Portal	12	53.13	12.02	51.19	25.99
255	Sheath	270	280.37	63.43	270.09	137.15
256	Tube	270	551.65	124.81	531.42	269.85

The Lumos mass spectrometer is valuable at revealing low abundance proteins, but can also result in unexpected behavior. One example is seen comparing gp53 and gp54, two paralogues known to exhibit the same copy number in the virion. On the Orbitrap mass spectrometer, the two have comparable total PSM values of 662 and 643 respectively. In all of the Lumos replicates, however, gp53 exhibits roughly twice the total PSM of gp54 (1987 vs. 1088 in Replicate 1, 1873 vs. 2021 in Replicate 2, and 1967 vs. 1071 in Replicate 3).

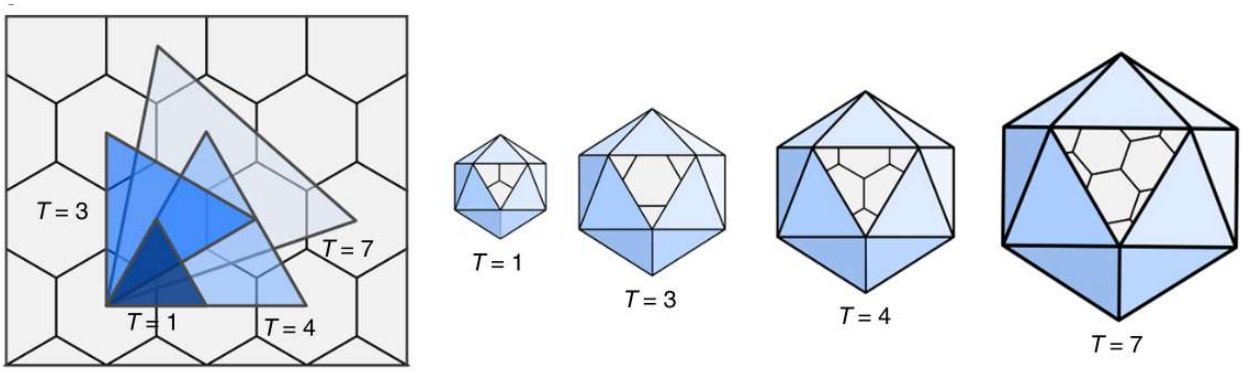
As a whole, this research indicates the value of using a higher sensitivity mass spectrometer in proteomic analyses, enabling higher overall coverage and the detection of more unique proteins.

Chapter 3: Identification of SPN3US Head Proteins

Introduction to Phage Head Structure

The head or capsid of all tailed phages is a protein shell structure of icosahedral base symmetry that contains the double-stranded DNA genome. The exterior shell of the heads of different phages is characterized by the number of planar faces composing its structure (Prasad and Schmid, 2012) (e.g., Figure 3.1). This is a hexagonal lattice with triangles overlaid. These triangles illustrate the shape used to fold the triangles into a 3, when the lattice is folded into a 3D icosahedron. Various asymmetrical shapes are produced when the triangles are stacked on one another. The total number of triangles, or protein copies in nature, composing each face of the icosahedron is known as the triangulation number (T) (Twarock and Luque, 2019).

Figure 3.1: Viral Capsid Architecture based on Caspar and Klug Theory. Reproduced from Twarock and Luque (2019).



Homology in Phage Capsid Structures

Despite great variability in the overall dimensions of the heads of a wide variety of phages, the subunits which form the outer shell have been all shown to exhibit structural homology to the major capsid protein, known as gp5, of HK97, an *Escherichia* virus (Pietilä et al., 2013).

Sometimes, the icosahedral symmetry is not perfectly intact. In some, the major capsid proteins also form an arrangement which is referred to as “chain-mail” like. This surface consists of covalently cross-linked capsid subunits that form interlinked pentamers and hexamers of the phage’s major capsid protein. Many studies have highlighted the reach of the HK97 capsid fold structure with tailed phages with every morphotype having been studied all having some form of this fold (Duda and Teschke, 2019). While the HK97 fold is highly conserved, the interlinking chainmail structure is not as strongly conserved. Furthermore, as genome length increases, the triangulation number of the capsid increases. The HK97 structure is highly conserved, scaling to different triangulation numbers. Finally, all giant phages tested possessed a contractile tail which facilitates injection of the genome into the host (Hua et al., 2017).

Phage Capsids Are Incredibly Stable Structures

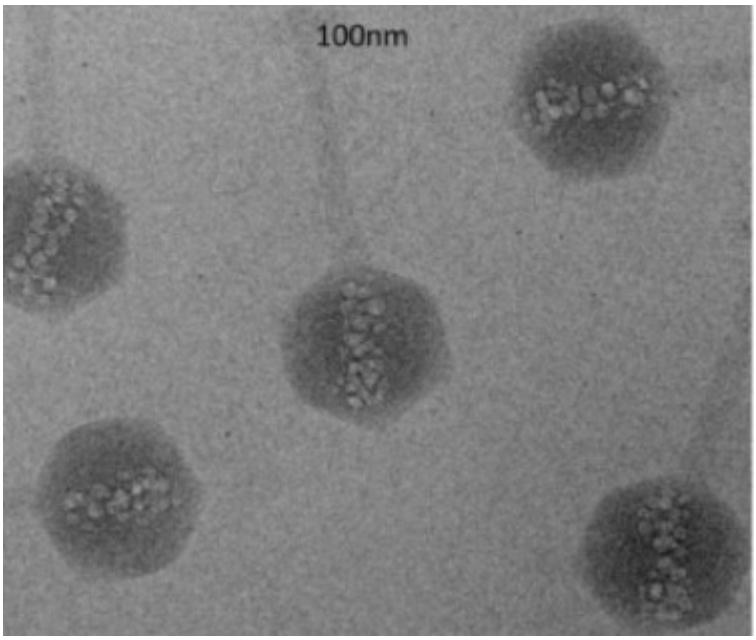
Across all bacteriophages, how the capsid structure is stabilized has been a point of curiosity. Relative to the size of the genome, the capsid forms a large volume, and is also relatively thin compared to its size. It has long been suspected that there are mechanisms in place to help stabilize the structure of the capsid. One conserved protein suspected to function in this regard is the auxiliary protein gpD, originating in lambda phage (Yang et al., 2008). The pressure inside a Lambda shell can reach 60 atm (Lander et al., 2008). This is due to the bacteriophage genome being very densely packaged, and necessitates a durable capsid to contain it. gpD, functioning as a capsid stabilization protein, attaches to the three-fold vertices to provide greater support.

Capsids of Giant Phages

Like all other tailed phages, giant phage capsids are formed by many copies of a subunit with some version of the HK97 fold that is arranged into structures with extremely high triangulation numbers relative to smaller phages. The head capsid of SPN3US has been determined through Cryo-EM to have a T=27 architecture, indicating that it has copies of its major capsid, gp75, comprising each of its planar faces (Reilly et al., 2020).

The inner body structured was first discovered in the bacteriophage PhiKZ (Wu et al., 2012). SPN3US possesses an inner body counterpart within its capsid whose structure is not as regular as that of PhiKZ whose major components likely are gp53 and gp54 (Figure 3.2).

Figure 3.2: Electron micrograph of purified wild-type PhiKZ virions. The inner body is visible as the cylindrical mass of ejection proteins along the length of the capsid.



Within the fully matured virion head, protein clusters along the inner vertices are also observed, known as “antlers”. The function of these is yet to be understood (Heymann et al., 2020).

The ease with which the proteomes of bacteriophages can be structurally characterized is often dependent on the host. SPN3US, with its host *Salmonella* allowing for nonpermissive strains, allows for more detailed proteomic analyses to take place by allowing nonviable mutants to be propagated. Tailless mutants contain one or more mutations which result in virus particles containing only the head.

Two tailless mutants isolated, am27 and am107, were used in combination with permissive and nonpermissive strains of *Salmonella* to help characterize the virion of SPN3US. Using mass spectrometry, replicates can be separated into wild-type and mutant phage particles. Gene products which are absent from the mutant, but present in the wild-type, must therefore be exclusive to the tail, while those present in both must be head proteins.

Methods

Isolation and Genome Sequencing of a SPN3US Tail Tube Protein Mutant Phage

The SPN3US tailless mutant am107 was isolated by Andrea Denisse Quintana Benitez [reference thesis], the aim of which was to achieve SPN3US mutants of unique phenotypes. Mutant am107 mutant collection was generated using random mutagenesis (hydroxylamine mutagenesis), as described previously (Thomas et al., 2016), since a targeted gene editing system has not yet been developed for this phage. In brief, wild-type SPN3US was treated with 0.4 M hydroxylamine at 37 °C for ~24 h and then plated to obtain single plaques. Mutant am107 was isolated via the screening of plaques for detection of amber mutant candidates with a conditional-lethal phenotype (growth on permissive or suppressor hosts versus no growth on non-permissive or non-amber suppressing hosts). The bacterial strains used to isolate and propagate am107 were the strains TT6675 (supD) and TT9079 (sup0), which were provided by Dr. John Roth, University of California Davies (these strains and others in the Roth strain collection can be searched at the URL http://rothlab.ucdavis.edu/textStrainer).

Genomic DNA was extracted from a high titer stock of am107 using the Norgen Phage DNA extraction kit following the manufacturer's protocol. Mutant DNA underwent a NexteraXT workflow and were sequenced on an Illumina MiSeq (2x200 bp) at the University of Rochester Genomics Facility. Mutant genomes were assembled using DNASTAR SeqMan and single nucleotide polymorphisms (SNPs) were identified using SeqMan Pro. The wild-type (WT) phage genome was used for reference-based alignments.

Purification and Transmission Electron Microscopy (TEM) of SPN3US Tailless Particles

Liquid infections of SPN3US mutant am107 in the non-permissive host, TT9079, were performed in LB + N broth at 33–34 °C at an MOI of 10 for 3 h. At ~25–30 min post infection cells were spun (5000 rpm, room temperature), and resuspended in fresh media to remove input phage. At the end of infection, cultures were treated with lysozyme (2 mg/mL) for 30 min at room temperature and underwent differential centrifugation to remove large debris and concentrate the remaining particles. The differential centrifugation involved a low speed spin (7000 g, 10 min, 4 °C) and a high speed spin (39,000 g, 40 min, 4 °C). Pellets were resuspended overnight in SM buffer at 4 °C.

The resulting suspension underwent purification by sequential CsCl step and buoyant density gradient ultracentrifugation, as described in Chapter 2 methods. After CsCl gradient purification, gradient bands were harvested by side tube puncture and refractive indices measured on an Abbe refractometer. To remove CsCl, the harvested gradient bands were dialyzed against three changes of 0.2 Tm buffer (50 mM Tris-Cl, pH 7.5, 200 mM NaCl and 10 mM $MgSO_4$) (3X 30 min, 4 °C). Purified particles were adsorbed to 400 mesh carbon-coated grids, negatively stained with 1% uranyl acetate and examined at 80.0 kV using a Tecnai T12 transmission electron microscope (TEM was performed at the University of Maryland Dental School Electron Microscopy Core Facility by Dr. Ru-ching Hsia).

Purified particles were boiled for 10 min in SDS sample buffer prior to a short electrophoresis (~20 min) on a Criterion XT MOPS 12% SDS-PAGE reducing gel (Bio- Rad) and proteins visualized by staining with Coomassie blue. Three replicate gel lanes were each divided into six slices for

GeLCMS analyses of the am107 mutant. After de-staining, proteins in the gel slices were reduced with TCEP (tris(2-carboxyethyl)phosphine hydrochloride) and then alkylated with iodoacetamide in the dark before digestion with trypsin (Promega). HPLC-electrospray ionization-tandem mass spectrometry (HPLC-ESI-MS/MS) was performed on the Mascot (Matrix Science; London, UK) was used to search the MS files against a locally generated SPN3US and *Salmonella* protein database that had been concatenated with the SwissProt database (2012_11_170320; version 51.6). Subset searching of the Mascot output by X! Tandem, determination of probabilities of peptide assignments and protein identifications, and cross correlation of the Mascot and X! Tandem identifications were accomplished by Scaffold (Proteome Software). MS data files were either processed individually or the files for an entire gel lane were combined via the "MudPIT" option in Scaffold.

Results/Discussion

Sequencing of an SPN3US amber mutant candidate as a tailless mutant

Genome sequencing of SPN3US amber mutant 107 (am107) showed that it has a nonsense mutation from a cytosine to a thymine at base 448 in gene product 255, the tail tube protein. We hypothesized that this mutation would prevent formation of the contractile sheath, and particles with a tailless phenotype. Mutagenesis using hydroxylamine results in C->T or G->A

transition mutations (Stolarski et al., 1987). In addition to the amber mutation, am107 acquired 17 other mutations from the amber mutagenesis, although these are not expected to have a meaningful impact on phenotype, based on it having similar mass spectral coverage to the wild-type.

Table 3.1: SPN3US Mutant am107 Substitution Report

Mutations identified in the genome of SPN3US mutant am107. Referenced position refers to the base position in the SP3US genome, Referenced base refers to the wild-type base, called base refers to the mutated base. The substitution mutation at base 226889, in bold, is responsible for the tailless phenotype of am107. The insertion mutations at bases 200944 and 239933 are derived from the laboratory's wild-type SPN3US.

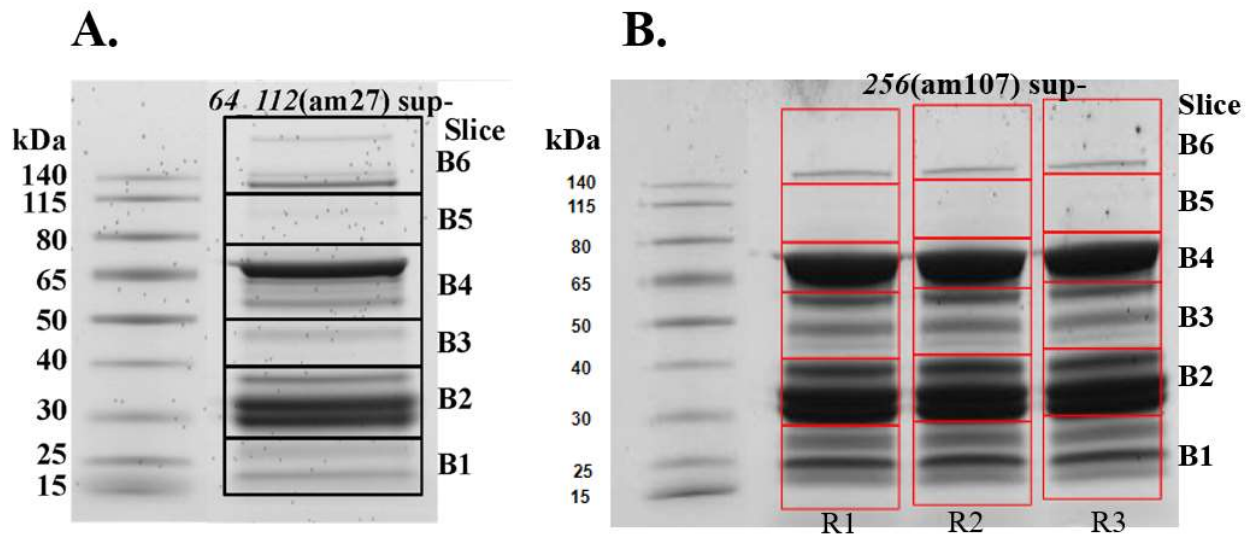
Referenced Position	**Referenced Base**	**Called Base**	**Impact**	**Feature**	**DNA Change**	**Amino Acid Change**	**ORF (gp)**
27612	C	T	Non-synonymous	CDS	394C>T	P132S	32
53716	G	A	Synonymous	CDS	111G>A	Q37Q	53
61708	G	A	Synonymous	CDS	519G>A	L173L	62
63491	G	A	Synonymous	CDS	405G>A	E135E	64
69613	G	A	Non-synonymous	CDS	46G>A	D16N	74

82224	C	T	Non-synonymous	CDS	1498C>T	R500C	82
109100	C	T	Non-synonymous	CDS	2275C>T	L759F	124
119591	G	A	Non-synonymous	CDS	690G>A	P230P	142
126818	G	A	Synonymous	CDS	423G>A	L141L	149
130321	G	A	Synonymous	CDS	570G>A	G190G	152
136018	G	A	Synonymous	CDS	594G>A	V198V	156
153287	C	T	Synonymous	CDS	1659G>A	A553A	169
185132	G	A	Non-synonymous	CDS	292G>A	G98S	210
200944	-	T	Insertion				
223871	C	A	Non-synonymous	CDS	70C>A	R24S	250
226889	**G**	**A**	**Nonsense**	**CDS**	**448C>T**	**Q150**	**255**
231556	G	A	Synonymous	CDS	1398G>A	V466V	258
239933	-	T	Insertion				

As expected, a large number of mutations (16) were acquired from mutagenesis (Table 3.1). These are overwhelmingly transition mutations. None of these mutations were found to have an effect on phenotype, with the exception of the nonsense mutation at base 226889. This mutation interferes with production of the tube protein, resulting in an expected tailless phenotype when propagated in a non-permissive strain of *Salmonella*. There are also two insertion mutations; these, however, are common to the wild-type SPN3US used in the laboratory.

SDS-PAGE Gel Comparisons

Figure 3.3: SDS-PAGE gels of tailless particles derived from SPN3US mutants propagated in a non-suppressor strain (sup-). R1, R2, R3 refer to the three mutant replicates in the am107 gel. B1-B6 identify unique gel slices that are grouped together for the purpose of mass spectrometry. Molecular weight ladder is consistent between the two gels.

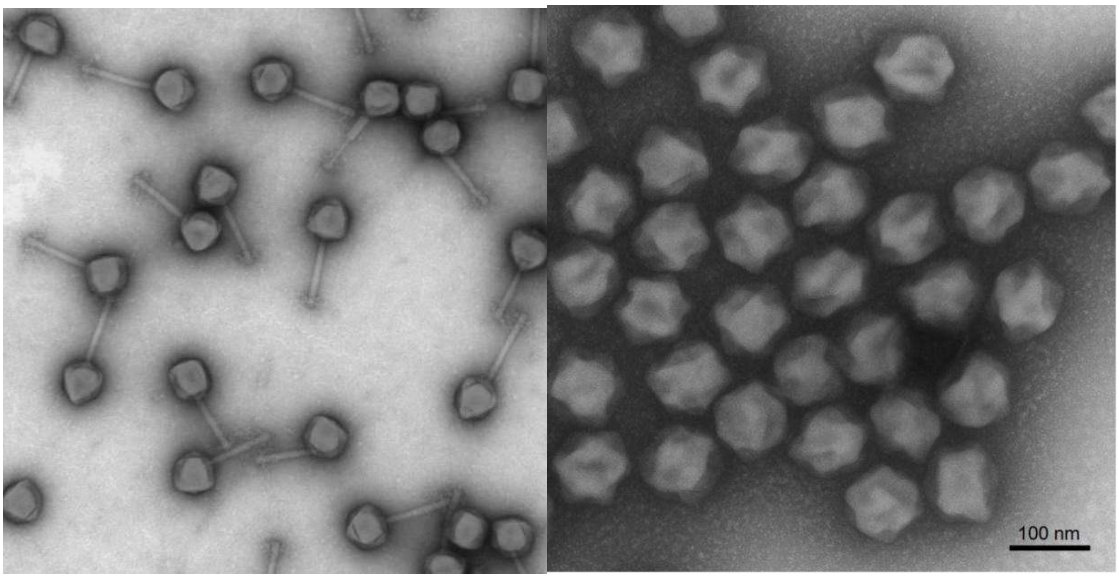


.

Looking exclusively at mutant lanes (Figure 3.3), the gels for am27 and am107 manifest similarly. However, greater concentrations of protein overall are observed in the am107 gel. The thickest bands are the major capsid protein gp75 at 80 kDa, gp53 and gp54 at 45 kDa, and gp141 at 32 kDa.

Transmission Electron Microscopy

Figure 3.4: Transmission electron micrograph of am107 on permissive (A) vs. non-permissive (B) Strains of *Salmonella*



A B

These tailless particles were confirmed through electron microscopy (Figure 3.4). On a permissive strain of *Salmonella* (left), standard virus particles with sheathed tails are seen, comparable to wild-type SPN3US. Non-permissive *Salmonella* (right), however, confirms that the am107 mutation in base 448 results in a tailless phenotype.

Table 3.2: Summation of Mass Spectral Analyses of Tailless Mutants

Sample	Total # of Unique Proteins Identified	Total PSM in Sample
am27 (Orbitrap)	53	5802
am107 Replicate 1 (Lumos)	66	14036
am107 Replicate 2 (Lumos)	65	14036
am107 Replicate 3 (Lumos)	64	13387

For the purpose of this study, proteins which were seen in all three replicates of am107 can be confidently categorized as truly head proteins. Proteins found in only one or two am107 replicates were extremely low abundance (<5 total PSM) and therefore they cannot be classified as truly head proteins with high confidence. In total, 55 unique proteins were found.

Table 3.3: List of SPN3US Proteins Identified in Tailless Particles. Tailless particles were purified and derived from SPN3US amber mutant am107 after propagation on the nonpermissive strain of *Salmonella*. Only proteins identified in at least 2 of 3 replicates are provided below, see Appendix table A.2 for the full data set. Bolded rows indicate proteins known to be processed.

			am107 Double-gradient Lumos 6-slice-a	am107 Double-gradient Lumos 6 slice-b	am107 Double-gradient Lumos 6 slice-c	am107 Double-gradient Lumos Average	am107 Double-gradient Lumos Standard Deviation	Putative function
gp	MW (kDa)	Mature MW (kDa)	Total PSM	Total PSM	Total PSM	Average PSM	SD PSM	
8	30.3	30.3	38	43	31	37.3	6.0	
17	15.3	15.3	0	0	0	0.0	0.0	
21	40.7	40.7	23	25	26	24.7	1.5	
28	32	32	5	5	3	4.3	1.2	
38	23.2	23.2	2	0	3	1.7	1.5	
41	31.8	31.8	16	11	15	14.0	2.6	
42	49.3	49.3	89	75	71	78.3	9.5	vRNAP β′M

45	**50.3**	**48.2**	**255**	**235**	**226**	238.7	14.8	
46	16.2	16.2	39	32	38	36.3	3.8	
47	**62.8**	**50.7**	**255**	**235**	**226**	238.7	14.8	
48	111	111	2	3	2	2.3	0.6	
50	39.4	25.6	213	196	198	202.3	9.3	
51	34.9	34.9	123	117	103	114.3	10.3	
52	21	21	128	115	74	105.7	28.2	
53	**45.2**	**31.5**	**2225**	**2240**	**2141**	2202.0	53.4	ejection proteins
54	**45.1**	**31.9**	**1361**	**1238**	**1157**	1252.0	102.7	(paralogs)
62	52	52	0	5	4	3.0	2.6	
74	12.9	12.9	126	132	135	131.0	4.6	
75	**80.1**	**70.4**	**4864**	**4799**	**4621**	4761.3	125.8	major capsid
81	**100.2**	**72.3**	**170**	**182**	**175**	175.7	6.0	portal
83	20	20	76	65	70	70.3	5.5	
94	41.6	41.6	21	25	18	21.3	3.5	

95	17.6	17.6	17	22	26	21.7	4.5	
97	11.9	11.9	6	9	10	8.3	2.1	
100	23.2	23.2	6	5	4	5.0	1.0	
109	17.7	17.7	44	52	53	49.7	4.9	
111	11.9	11.9	0	0	0	0.0	0.0	
124	113.7	113.7	4	4	3	3.7	0.6	
138	29.2	29.2	78	86	74	79.3	6.1	
139	29.8	29.8	48	56	55	53.0	4.4	
140	31.4	31.4	46	36	43	41.7	5.1	
141	32.6	32.6	542	573	533	549.3	21.0	
142	30.8	30.8	93	98	65	85.3	17.8	
143	31.9	31.9	275	320	251	282.0	35.0	
144	30	30	104	121	100	108.3	11.2	
145	50.8	50.8	46	46	41	44.3	2.9	
146	36.9	36.9	31	28	25	28.0	3.0	

147	33.7	33.7	42	40	42	41.3	1.2	
148	53.7	53.7	93	102	94	96.3	4.9	
149	36.2	36.2	92	103	90	95.0	7.0	
150	33.8	33.8	34	33	29	32.0	2.6	
151	52.2	52.2	50	55	44	49.7	5.5	
152	36.8	36.8	104	119	120	114.3	9.0	
153	33.8	33.8	41	46	38	41.7	4.0	
154	50	50	65	56	61	60.7	4.5	
155	78.3	78.3	94	94	93	93.7	0.6	
158	19.9	19.9	0	7	7	4.7	4.0	
160	18.5	18.5	411	517	593	507.0	91.4	
169	149	149	4	3	0	2.3	2.1	
193	19.9	19.9	26	44	25	31.7	10.7	
214	28.1	28.1	143	170	135	149.3	18.3	
218	25.2	25.2	46	57	39	47.3	9.1	vRNAP βC

225	**25.1**	**22.3**	**61**	**53**	**45**	53.0	8.0	
237	19.9	19.9	2	5	0	2.3	2.5	
240	59.6	59.6	133	124	115	124.0	9.0	vRNAP β'N
241	159.1	159.1	233	241	246	240.0	6.6	vRNAP β'N
242	10.5	10.5	15	12	15	14.0	1.7	
243	54.6	54.6	528	482	489	499.7	24.8	
244	27	27	21	18	9	16.0	6.2	vRNAP β'C
245	**30.7**	**23.4**	**61**	**62**	**57**	60.0	2.6	
246	23.9	23.9	79	67	87	77.7	10.1	
248	21	21	26	30	24	26.7	3.1	
256	75.7	75.7	30	31	27	29.3	2.1	sheath
257	34.2	34.2	138	145	150	144.3	6.0	
258	96.4	96.4	0	0	2	0.7	1.2	
262	52.7	52.7	83	80	87	83.3	3.5	

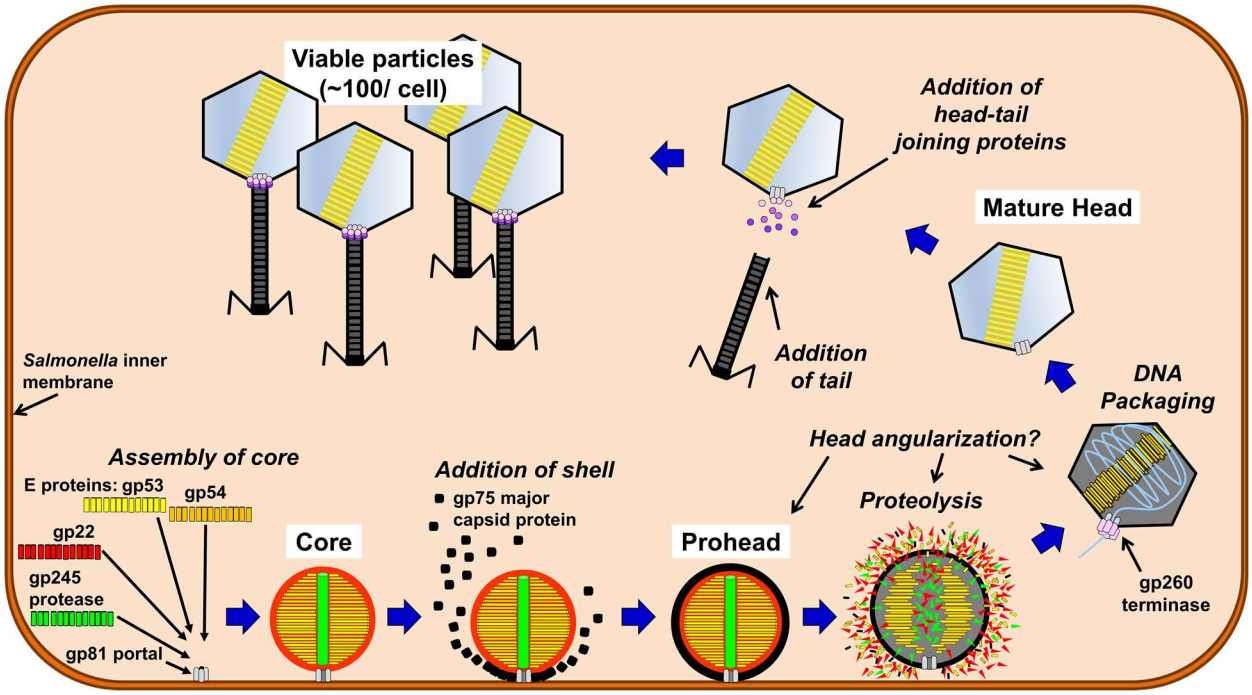
Based on the mass spectral results found, there are a variety of trends seen with respect to head proteins identified using the am27 mutant, on the less sensitive Orbitrap mass spectrometer, relative to the am107 data on the more sensitive Lumos (Table 3.2).

Mass spectral studies of SPN3US in wild-type SPN3US versus am107 grown on a permissive strain of *Salmonella* make it quite clearly defined which proteins are localized to the head (Table 3.3). Notably, all proteins known to be processed by the prohead protease are found to be located in the head. This is to be expected as the prohead protease processes proteins within the head during prohead maturation. A previous experiment using am27 had also been performed, however this was done on a less sensitive Orbitrap mass spectrometer, and with only a single replicate. Some proteins were found in am107 but not am27 and vice versa. It is a possibility that the different mutations present in each mutant could be contributing to these differences. The increased sensitivity of the Lumos could also be responsible for uncovering lower abundance proteins that could not be detected on the Orbitrap. Of the proteins discovered between am27 and am107, it is likely those with the greatest confidence as head proteins would be the proteins seen in both mutants. Head proteins only seen in one of the mutants would instill less confidence, and could be a direction of further research. Ultimately, these tailless mutant studies have provided a reasonable foundation in characterizing the proteome of the SPN3US head.

Chapter 4: Proteolytic Processing of Virion Proteins

Viral Capsid Processing & Maturation

Figure 4.1: Diagram of SPN3US Head Maturation



One of the key principles to understanding maturation of the virion structure is the degree of processing that is occurring within the proteome of the virus. The initial structure in head maturation is a prohead structure consisting solely of proteins (Figure 4.1). Following prohead formation, proteolysis is performed by the viral protease (gp245 in the case of SPN3US), whereby the prohead then expands. The protease is located within the interior of the capsid, therefore all processed proteins can be expected to be head proteins. The capsid is then expanded enough for DNA to be packaged inside, and finally the tail and its accompanying elements (the sheath, portal, and baseplate) are attached. In many tailed bacteriophages, head proteins are cleaved at a motif by an endogenous protease before assuming their eventual structure in the

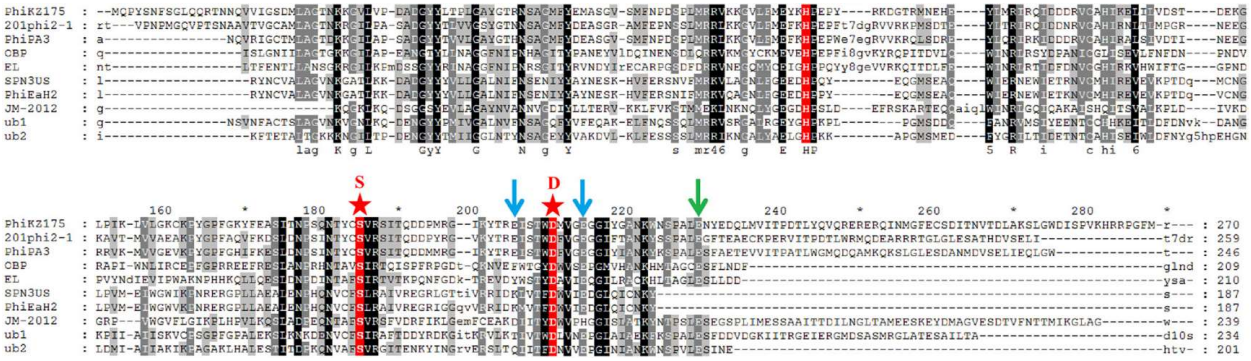
capsid. T4 head proteins undergo proteolysis by its protease gp21. The model giant phage PhiKZ also undergoes proteolytic processing during prohead maturation, facilitated by its own protease gp175.

In PhiKZ, gp175 cleaves 19 head proteins at the sequence motif 'S/A/G-X-E', 6 of which are expected to be structural components of PhiKZ's inner body (Thomas et al., 2012).

Across all tailed phages, there is some highly diverged homology between prohead proteases. Giant phages related to PhiKZ and SPN3US all have protease with homology to PhiKZ's protease gp175 (Thomas and Black, 2013).

Figure 4.2: Sequence Conservation Across Homologous Proteases

Alignment of the ?KZ protease, gp175, with homologs in Pseudomonas phages 201-Phi-2-1 (gp268), Phi-PA3 (gp205), OBP (gp283), and EL (gp192) and homologs in Salmonella phage SPN3US (gp245), Erwinia phage PhiEaH2 (gp165), and Halocynthia phage JM-2012 (gp80). Homologs in uncultured bacteria are indicated by ub; the GenBank identifier for ub1 is EKD22713, and the GenBank identifier for ub2 is EKD89709. Catalytic histidine and serine residues are indicated in red and marked with stars, as is the candidate for the third catalytic residue (D-191). The PhiKZ gp175 maturation cleavage site (E-210) is indicated with a green arrow. Putative cleavage sites for PhiKZ gp175 self-inactivation (E-186 and E-195) are indicated with blue arrows. Reproduced from (Thomas and Black, 2013).



There is a considerable amount of sequence conservation across the proteases of giant phages (Figure 4.2). Such conservation suggests a single evolutionary event from which proteolysis across giant phages arose.

The SPN3US Protease, gp245

Proteolytic maturation has been shown using genetics to be essential for SPN3US head formation, specifically head maturation. A study using a nonfunctional amber mutant of SPN3US, am59 which has an amber mutation in the gp245 gene, demonstrated that when gp245 is knocked no viable particles of SPN3US are formed. Using TEM of am59 infected cells that had undergone thin-sectioning, it was shown that in the absence of gp245 proheads assemble on the *Salmonella* inner membrane and are never released. Mass spectral analyses of immature SPN3US heads isolated from lysed cells confirmed that there was no gp245 in those particles, and that none of the head proteins known to be processed from an earlier study had been cleaved.

That is, gp245 is responsible for performing all the proteolytic maturation cleavages in the SPN3US head and this proteolytic maturation is critical for the transition of SPN3US proheads from protein-only particles anchored to the cell wall to expanded particles containing considerably less internal protein that are no longer anchored to the cell wall, thus allowing for the expansion of the capsid and packaging of the genome.

Further cryo-EM analyses of am59 particles showed them to form “mottled capsids”, with an internal structure consisting of a lattice of the paralogs gp53 and gp54, lacking the DNA present in the mature capsid (Heymann et al., 2020). Furthermore, knocking out gp245 results in nonviable virus particles.

The SPN3US Cleavage Motif

The motif at which the endogenous protease will cleave proteins varies in different phages. In SPN3US, the motif has been found to be 'AXE', with all processed proteins adhering to this processing site. 'X' here refers to an amino acid that can vary with the motif still being recognized by the protease. The mass spectral data used for SPN3US in this thesis has been used to help determine processed proteins in the SPN3US proteome, in an experiment cleaved by trypsin and with the heightened sensitivity of the Lumos mass spectrometer. Mass spectral coverage beginning immediately after the glutamate residue of the 'AXE' motif is indicative of processing. Furthermore, the gap in coverage between the propeptide (the peptide cleaved off) and the mature protein (that remaining after cleavage) can also be seen in mass spectral data. In these cases, the N-terminal region of the protein will be seen lacking coverage, with coverage starting following the 'AXE' motif in the sequence of residues.

Despite these previous studies there is much to be learned about proteolytic maturation in SPN3US – as well as other giant phages – including identification of all the proteins that undergo proteolytic cleavage, and where those cleavage sites are located in each cleaved protein. This is in part, because with so many proteins in the virion, other than a handful of proteins, it is not clear which proteins are in which band. This means it is not clear if proteins have migrated to a position that is consistent with their mature molecular weight or slower or faster, the latter being indicative of proteolytic processing. Similarly, it is not at all clear that all the proteins in the SPN3US head have been identified. This is because proteolytic processing must be confirmed either by intensive experimental work in vitro using purified recombinant substrates and protease (which with all the proteins in the virion would take years to complete and is therefore not a viable option) or via the detection of specific peptides at the cleavage site by examination of the

mass spectral sequence coverage. Since those peptides (referred to as "semi-tryptic" peptides as their termini that were not produced by trypsin digestion) are often short and/or underrepresented in the protein population (especially in low abundance proteins), it is normal for multiple mass spectral experiments of the same sample to need to be analyzed to identify those cleavages.

To address these needs in these studies I obtained mass spectral data of purified SPN3US heads (or tailless particles) by using the same GeLCMS approach as used in the analyses in Chapter 3, but in this experiment the sample was run the full length of the protein gel to enable better separation of individual protein species and the entire gel lane excised into many slices to separate out individual bands. The goal of this was to facilitate identification of which protein species were the major components of each visualized band in the SPN3US SDS-PAGE profile. This information is a critical component of identifying which proteins undergo proteolytic maturation, in conjunction with protein sequence coverage data.

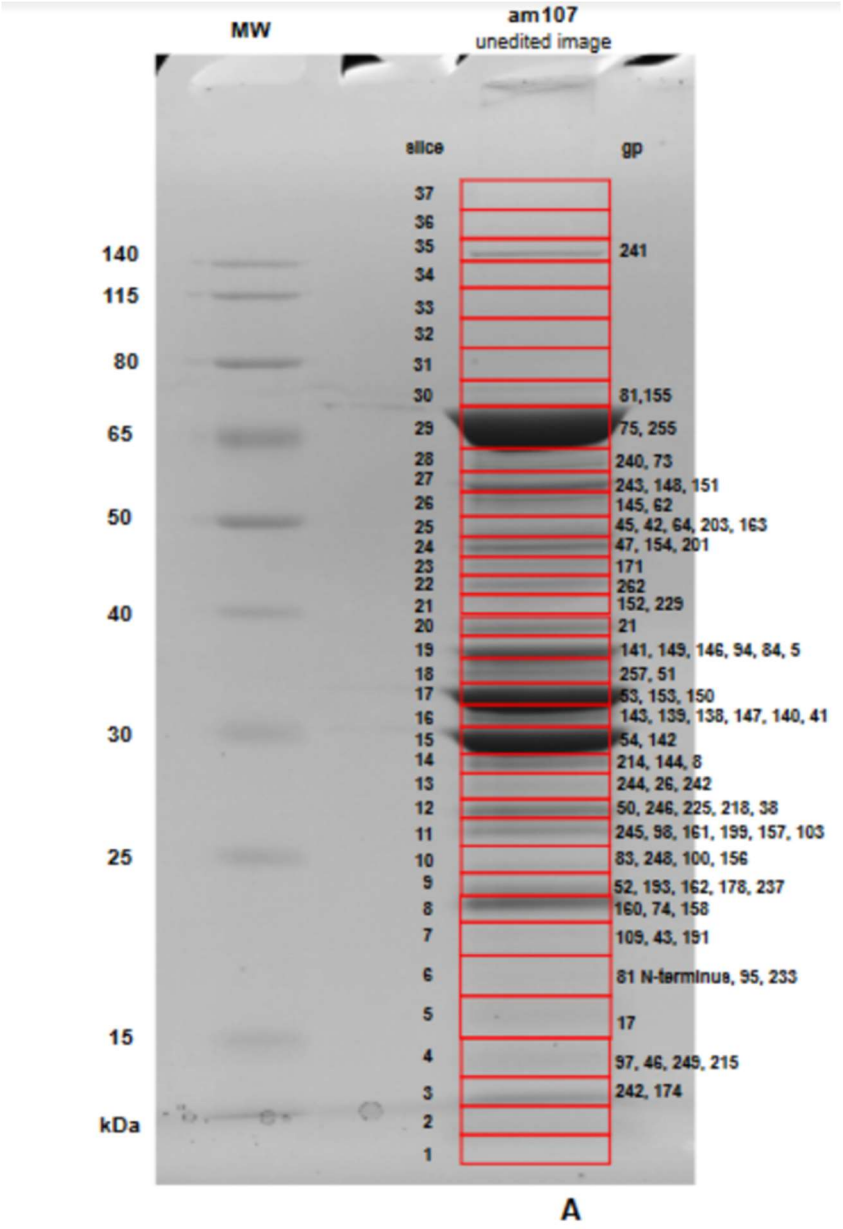
Methods

These analyses were based on the mass spectral analyses of a sample of purified SPN3US tailless particles derived from am107 grown under non-permissive conditions that was characterized in Chapter 2. Again, a GelCMS analysis was performed but this time the sample underwent electrophoresis to completion (i.e., the entire length of the gel lane was utilized). What protein gel conditions (Bis-Tris gel, with polyacrylamide concentration of 12 %) using MOPs running buffer. and electrophoresis conditions (see gel) Individual protein bands were visualized using ? stain (see gel). The entire gel lane was harvested into 37 slices, that were excised with care to ensure individual gel bands were separated as much as was feasible. The mass spectrometry data

was exported from Scaffold (Proteome Software) at 99.9% protein threshold, 2 minimum peptides, and 95% peptide threshold. Of the identified proteins, the dataset was condensed to proteins previously determined to be head proteins. The gel slice in which each head protein had its peak number of total PSM identified was identified. A plot was made of the Mw marker bands relative to Mw. The expected gel slice position for each head protein was predicted using this plot and then compared with the actual gel slice the protein was identified as having peaked in to. Proteins were categorized as having migrated as expected or slower or faster than expected. The sequence coverage of all head proteins was examined to identify which proteins had protein sequence coverage that was indicative of proteolytic processing (i.e., no sequence coverage on the *N* termini) and also for the presence of semi-tryptic peptides.

To create a logo for the SPN3US cleavage motif from all processing sites, a FASTA file was created containing all the regions 10 residues N-terminal and C-terminal to the glutamate residue at which cleavage had been observed. This file was then uploaded into the WebLogo website (Crooks et al., 2004). Processing data, including motifs and coverage, were gathered from the 37-slice gel mass spectral data.

Figure 4.3: SDS-PAGE gel of purified SPN3US tailless particles used for mass spectral analyses. The tailless particles were produced by the growth of SPN3US am107 under non-permissive conditions. The gel lane was cut into 37 slices, as indicated. SPN3US gene products are labelled beside the slice in which their PSM peaked, listed in order of most abundant.



In the 37-slice analyses – 82 total proteins were identified (Appendix A.2), more than previous studies, close to the total number of different proteins identified in the full virion, but these

analyses were only performed on tailless particles. This higher number of different proteins identified in the am107 sample was probably in large parts caused by the high sensitivity of the Lumos mass spectrometer (similarly to as observed in the previous analyses of am107 in chapter 3) as supported by the detection of a total PSM of 34419 in this sample, which was approximately 2-fold greater than identified in any of the replicate samples that underwent 6-slice GeLCMS (chapters 2 and 3).

Based on these results, the list of identified proteins was initially curated to 55, taking into consideration what had previously been deduced to be a "true" head protein in Chapter 3.

The proteins identified in the am107 sample were then analyzed with regard to their migration in the 37 slice gel (Figure 4.3), specifically in which gel slice the total PSM identified for each protein peaked (Table 4.1). Once the position of the gel in which the total PSM of each SPN3US head protein had peaked was determined, that observed gel slice number was compared to the expected or predicted gel slice that protein was expected to peak in based on its Molecular mass. 15 head proteins were noted to migrate faster in the gel than expected, making them excellent candidates for having undergone proteolytic maturation. This expectation is supported by the fact that the major head proteins, gp75, gp53 and gp54, that were all confirmed to be processed by gp245 previously, all migrated faster than expected (2, 6, and 8 slices, respectively). These analyses determined that 17 of the head proteins migrated close to where they would be expected, it should be noted that this observation may not preclude a protein having undergone proteolytic maturation that resulted in the removal of only a short propeptide. For example, in the instance of gp45, which migrated to slice 25 – the gel slice it was predicted to be identified in - but was shown previously to have 20 residues removed from its N-terminus.

Table 4.1: Expected vs. Observed Gel Migration of SPN3US Processed Proteins. An astericks indicates molecular weight for a polypeptide chain whose maturation site was confirmed.

gp	Expected Slice (by full protein M.W.)	Observed Slice	Expected MW (kDa)	Predicted MW (kDa) (full-length)	Predicted MW (kDa) (Cleaved)
Proteins whose processing or cleavage status has been confirmed:					
45	25	25	50	50	48*
47	28	24	42	62.8	54.3
50	22	24	42	48	26
53	23	17	46	33	32
54	23	15	45	29	30
75	32	29	84	68	70
81	29	31	100.2	65	82
225	10	12	25	27	22
245	9	12	24	27	13
Proteins which are potentially processed but still in need of confirmation:					
39	30	30	74	70	59

94	21	19	41.6	35	35.5

One tool for detection of processing is through the observation of migration patterns down the gel using mass spectral data. The slice in which a given proteins should be predictable based on its molecular weight in accordance with the protein ladder. A protein peaking in an unexpected slice is likely indicative of processing, as the cleaved off propeptide will result in it migrating farther along the gel than it otherwise would. This phenomenon is indeed seen in all processed proteins (Table 4.2), to varying degrees, although not in gp45.

The mass spectral protein sequence coverage was then used to more clearly infer or confirm processing sites in any proteins found to have unexpected migration patterns

Confirmation of proteolytic maturation using mass spectra sequence coverage data

A total of 9 SPN3US proteins were determined to have undergone proteolytic maturation (gp45, gp47, gp50, gp53, gp54, gp75, gp81, gp225, and the prohead protease itself, gp245) (Table 4.2). It's notable that the protease does in fact cleave itself – the only cleavage know to remove a C-terminal propeptide in SPN3US - likely with functional implications, losing about 24% of its mass in the process.

Of these, cleavages in gp225 and gp54 were newly discovered in this work. Gp53 and gp54, paralogs of one another, are cleaved in similar positions, resulting in highly similar mature molecular weights for both. Prior to these studies the maturation cleavage of gp54 was not

identified but inferred to be at X, however these studies identified that the maturation cleavage is at X and corrected the known gel position of gp53 relative to gp54.

Table 4.2: SPN3US Proteins cleaved by the prohead protease gp245

gp	Predicted Mass (kDa)	Mature Mass (kDa)	Maturation Cleavage Site(s)	Function/ Comment
45	50.3	48.2	ASE-20	Unknown function
47	62.8	54.3	AVE-75	Unknown function
50	39.4	25.6	ATE-127	Unknown function
53	45.2	31.5	AQE-125, AQE-95	Ejection protein
54	45.1	30.3	ARE-137	Ejection protein
75	83.9	70.4	ATE-130	Major capsid
81	100.2	72.3	ATE-161	Portal, expected maturation is AQE-254
225	25.1	22.3	ARE-24	Unknown function
245	30.7	23.4	AQE-203	Protease

The ‘AXE’ motif recognized by SPN3US’ protease is demonstrated to be strongly adhered to. All cleaved proteins adhere perfectly to it, being cleaved immediately following the glutamate residue. There is no consistent pattern to the ‘X’ residue, supporting its insignificance to the protease. Interestingly, a valine N-terminal to the adenine and serine C-terminal to the glutamate also appear to be fairly well conserved, however these are not seen in all processed proteins as the adenine and glutamate are.

Figure 4.4 *Salmonella* Phage SPN3US Prohead Protease Cleavage Motif Consensus

Logo was generated using WebLogo and the regions incorporating 10 amino acids upstream and downstream of the cleavage site of each SPN3US protein determined to be cleaved by the SPN3US prohead protease (see Table A.5). Cleavage always occurs C-terminal to the glutamate residue.

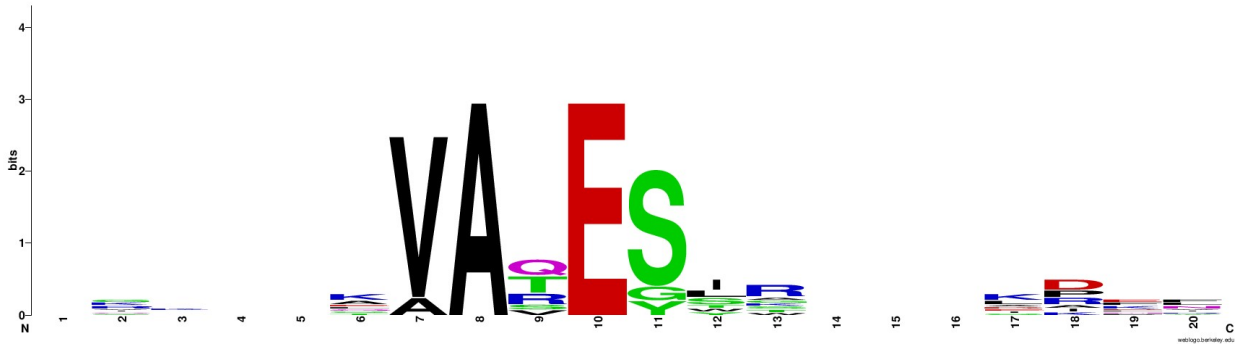


Figure 4.4 illustrates how processed proteins are clearly indicated in mass spectral data, visualized here using Scaffold software. Outlined in a red rectangle (A), the prohead shows distinct sequence coverage for the propeptide compared to the mature protein. (B) shows the total PSM of the most abundant peak in slice 12, which is consistent with its mature molecular weight. of 22.68 kDa. Sequence coverage begins at the peptide "SISTEKER', immediately N-terminal to the glutamate of the motif 'ARE'. This methodology was used to identify all processed proteins found in SPN3US.

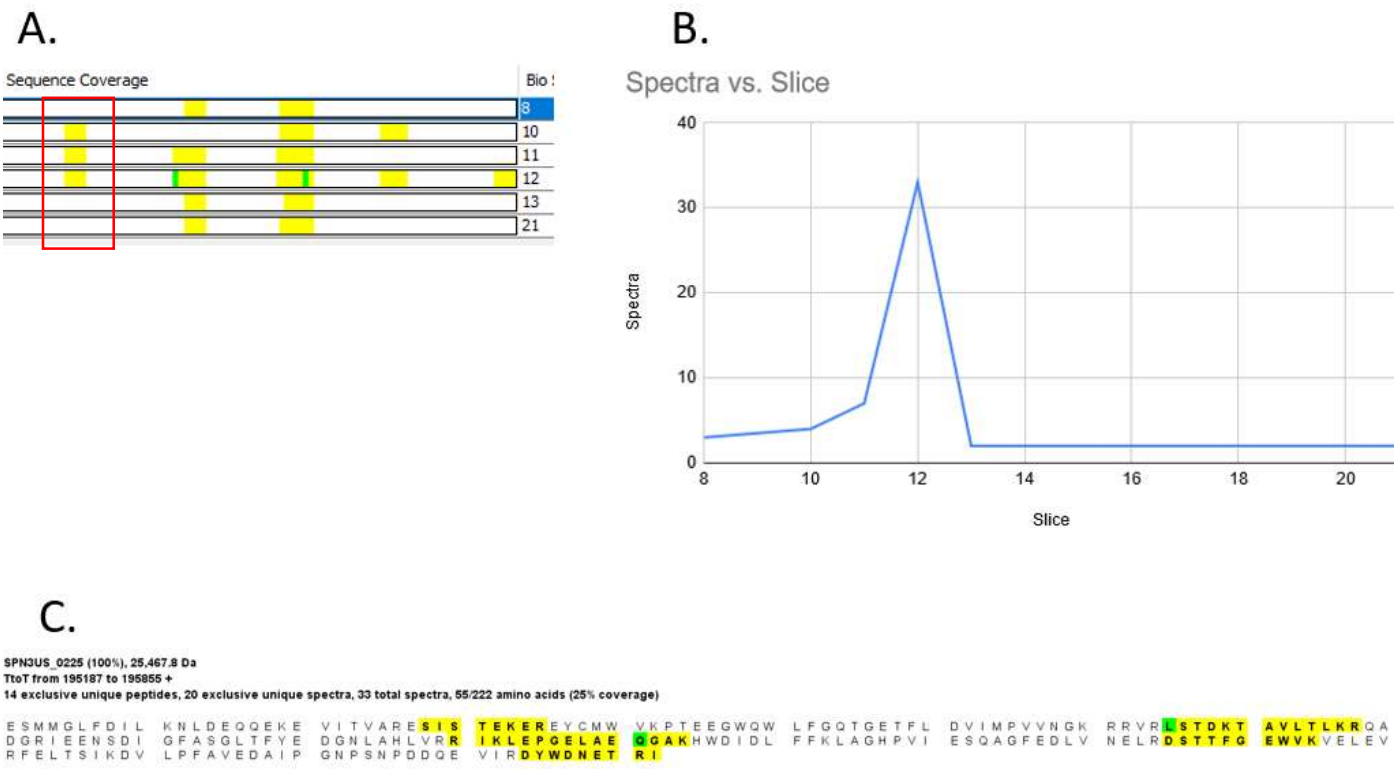


Figure 4.5: Spectral Coverage of SPN3US Processed Protein gp225. Mass spectrometry analyses of *Salmonella* myovirus SPN3US protein gp225 from purified heads. (A) Protein sequence coverage of gp225 identified in 37 gel slice analyses, prohead coverage outlined in red. (B) Plot of all mass spectra identified in each slice. (C) Protein sequence coverage of the peak slice.

Figure 4.6: Spectral Coverage of SPN3U3 Processed Paralog Protein gp53

Mass spectrometry analyses of *Salmonella* myovirus SPN3US protein gp53 from purified heads. (A) Summary of protein sequence coverage of gp53 identified in 37 gel slice analyses, prohead coverage outlined in red. (B) Plot of all mass spectra identified in each slice. (C) Protein sequence coverage of the peak slice.

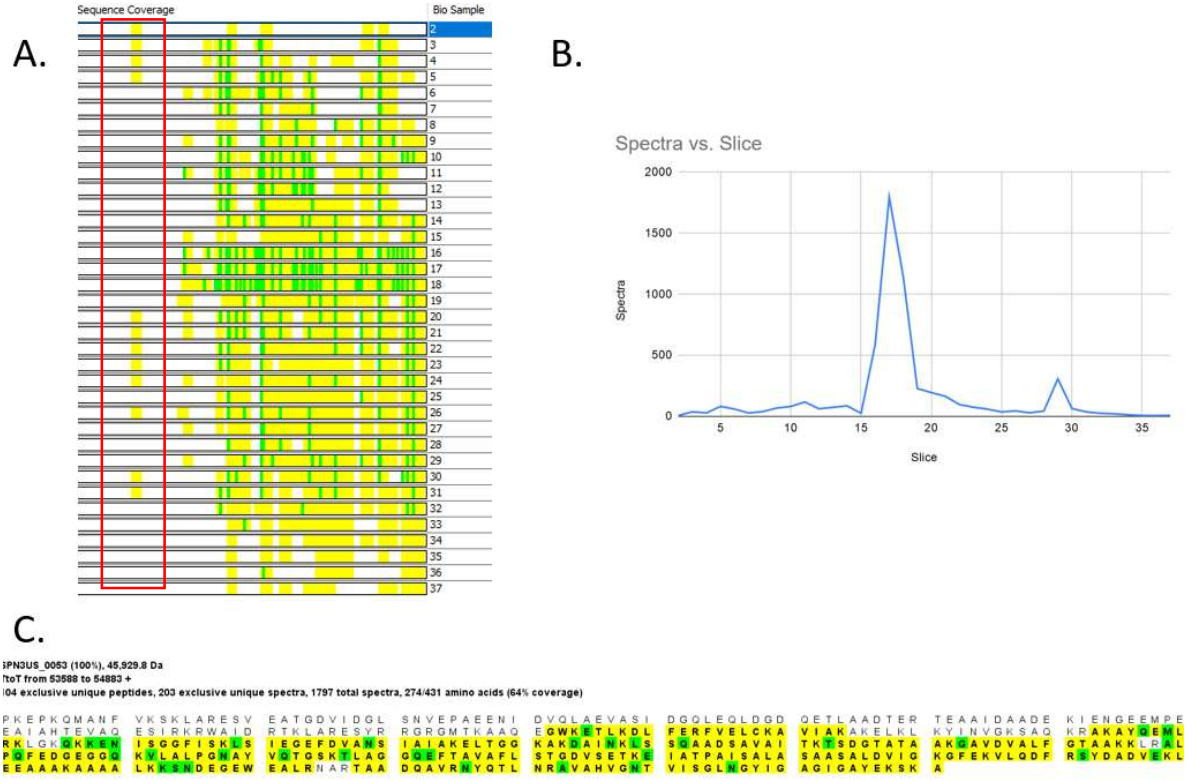
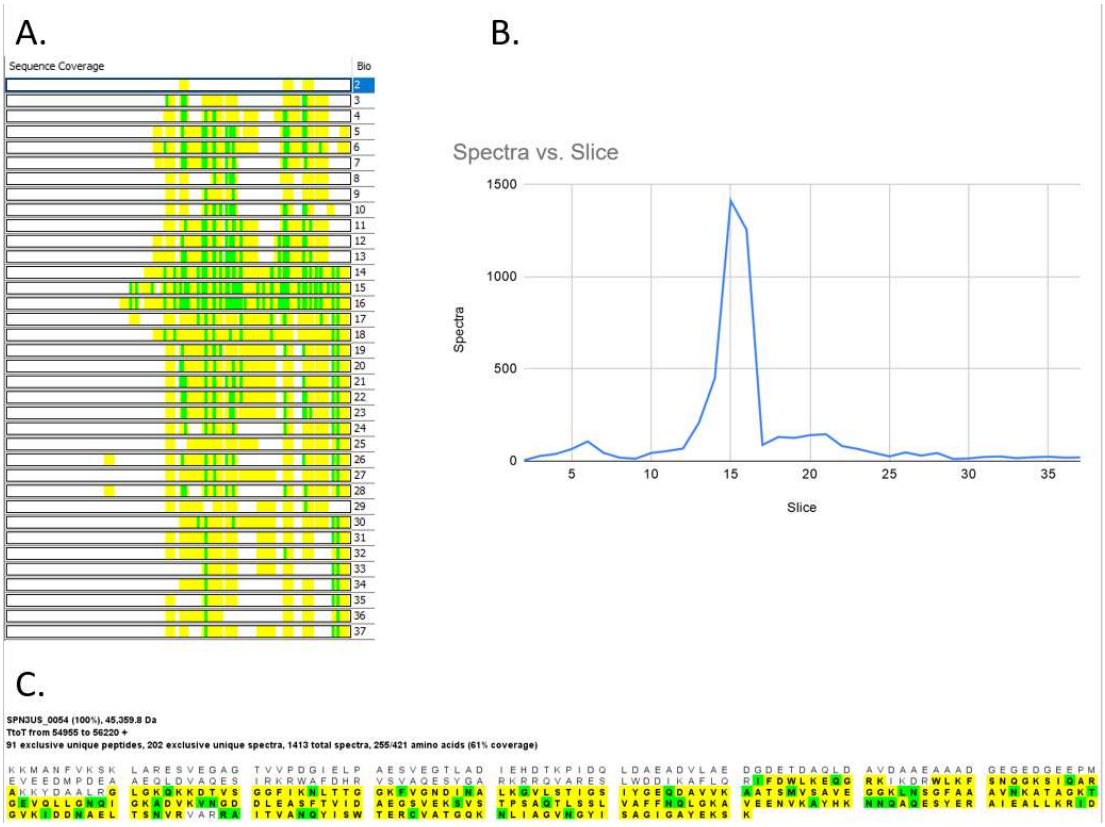


Figure 4.7: Spectral Coverage of SPN3US Processed Paralog Protein gp54

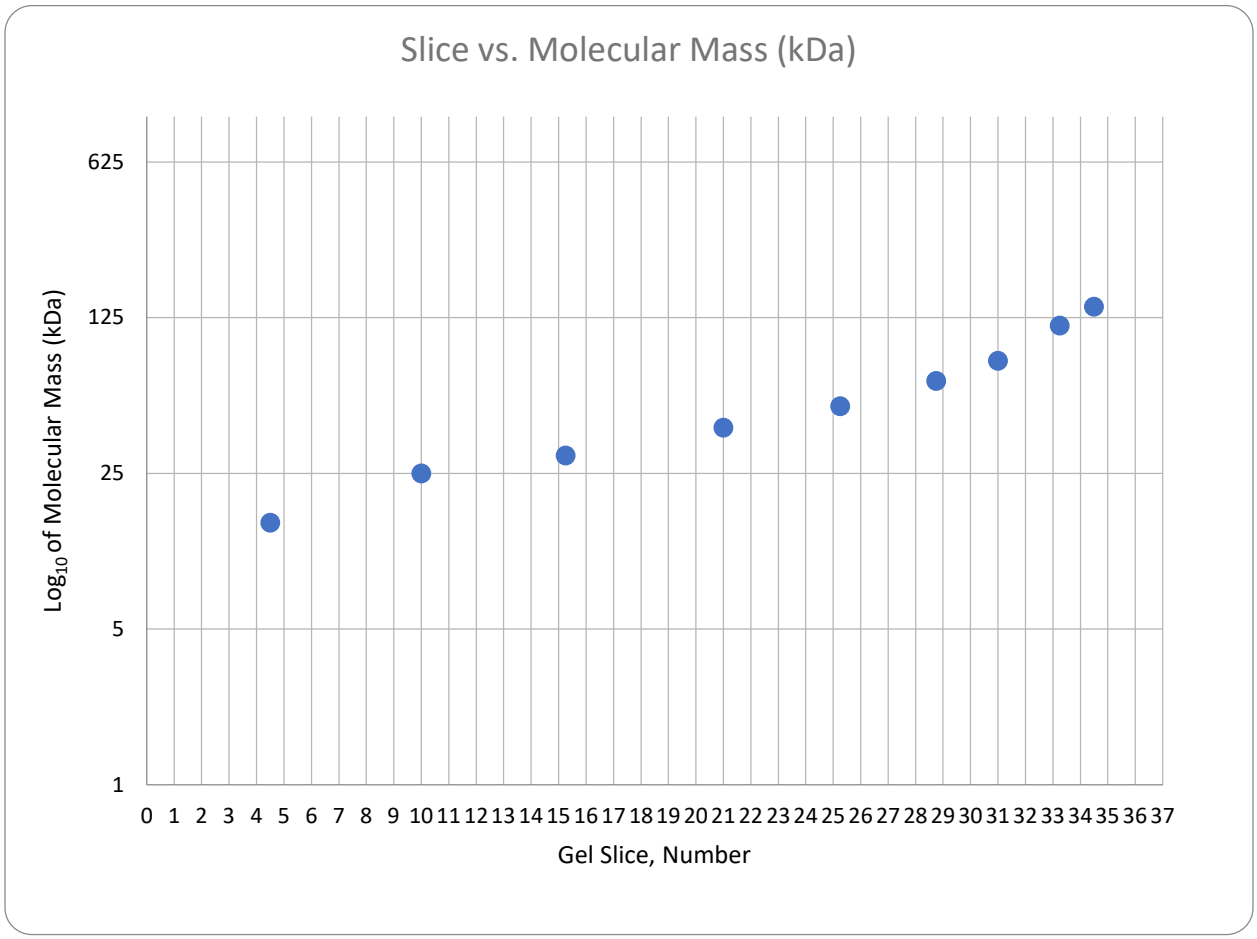


It is noteworthy that these studies revealed that remnants of the propeptide region of processed proteins often remain in the sample in addition to the mature protein. gp53 (Figure 4.6) shows clear evidence of incomplete processing, with a distinct secondary peak occurring in slice 29. There is also a small fragment of sequence coverage found in several slices of gp53 (outlined in red), absent from its paralog gp54 (Figure 4.7). It is unclear why these proteins show different sequence coverage, but it appears to be an artifact of the higher sensitivity of the Lumos mass spectrometer, as this trend is not seen on the standard Orbitrap. It could also have a biological cause, based on whether the protein with its propeptide region still attached is seen in the gel.

Migration Analyses

Figure 4.8: Molecular Weight – Slice Plot

Horizontal axis: Gel Slice, with slice 30 representing the end of the gel (i.e., the lightest proteins). Vertical axis: Molecular weight (kDA). Plot points represent the gel slices each band of the molecular weight ladder corresponded to.

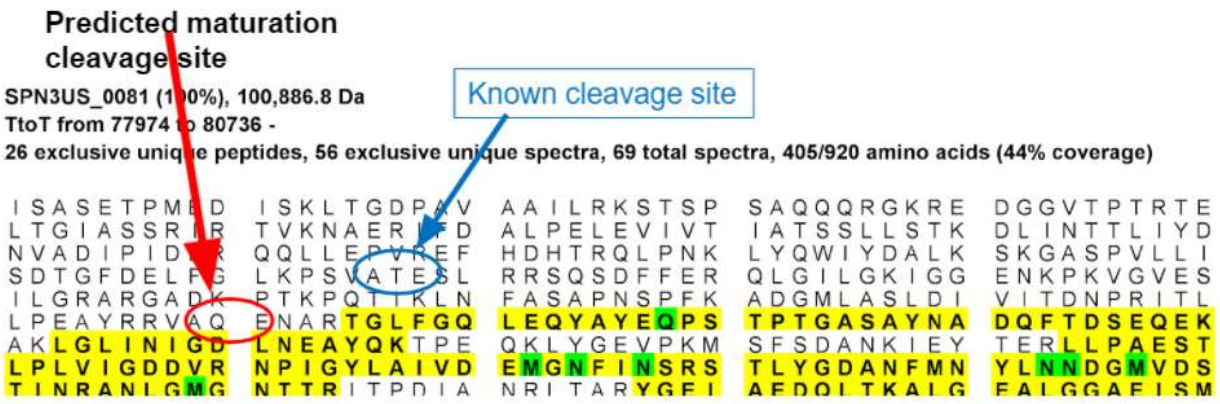


From the 37-slice SDS-PAGE gel, a plot was prepared based on the molecular weight ladder and the gel slices each of its bands corresponded to (Figure 4.8). Based on this plot, the expected slice a protein would end up in based on its full molecular weight can be estimated. A protein peaking in a gel inconsistent with what would be expected is a preliminary indicator of it being processed.

Of the 93 proteins with their migration patterns analyzed, all nine proteins confirmed to be processed migrated faster than would be expected for their full molecular weight. 47 proteins total were found to migrate slower than expected, 14, were found to migrate faster than expected, and the remaining 32 migrated precisely as expected (Table 4.3).

Beyond the nine proteins confirmed to be processed, a few are ambiguous. gp94, for example, migrates slower through the gel than would be expected for its molecular weight, seen in slice 19 when its full molecular weight would be expected to be seen in slice 21. gp94 also possesses the 'A-X-E' motif, with an 'AWE' seen close to the C-terminus.

Figure 4.9: Multiple Cleavage Sites of gp81



Some proteins, including gp81, exhibit multiple hypothetical cleavage sites (Figure 4.9). gp81 is known to be cleaved at its 'AXE' motif, but another 'AQE' motif is seen downstream of that. This reveals a limitation of the cleavage enzyme trypsin.

A promising path to take for future research would be trying the mass spectral experiment with an enzyme other than trypsin. Several ‘AXE’ motifs were seen that were not associated with the start of additional sequence coverage, such as those appearing in gp53. Repeating the 37-slice experiment with a new enzyme could potentially reveal the presence of processed proteins beyond those discussed in this paper.

Table 4.3: Migration Data For All Truly Head SPN3US Proteins. Expected slice is based on ladder slices corresponding to the full molecular weight protein.

gp	MW (kDa)	Mature MW (kDa)	Expected slice number with the peak amount of total PSM	Observed slice containing the peak number of total mass spectra	Gel Migration	Is cleavage indicated by sequence coverage?	Putative credible cleavage motif present?
8	30.3	30.3	15	14	1 slice faster than expected		
21	40.7	40.7	21	20	1 slice faster than expected		
28	32	32	15	15	as expected		
42	49.3	49.3	24.5	25	as expected		
45	50.3	45.2	25	25	as expected		
46	16.2	16.2	4.5	4	as expected		
47	62.8	54.3	28	24	4 slices faster than expected	Yes	Yes
50	39.4	39.4	21	24	3 slices slower than expected		
51	34.9	34.9	18	18	as expected		
52	21	21	7	9	2 slices slower than expected		
53	45.2	31.5	23	17	6 slices faster than expected	Yes	Yes
54	45.1	31.9	23	15	8 slices faster than expected	Yes	Yes
74	12.9	12.9	n/a (guess of 3)	8	as expected		
75	80.1	70.4	31	29	2 slices faster than expected	Yes	Yes
81	100.2	72.3	32	30	2 slices faster than expected	Yes	Yes
83	20	20	7	10	3 slices slower than expected		
94	41.6	41.6	21	19	2 slices faster than expected		Yes
95	17.6	17.6	6	6	as expected		

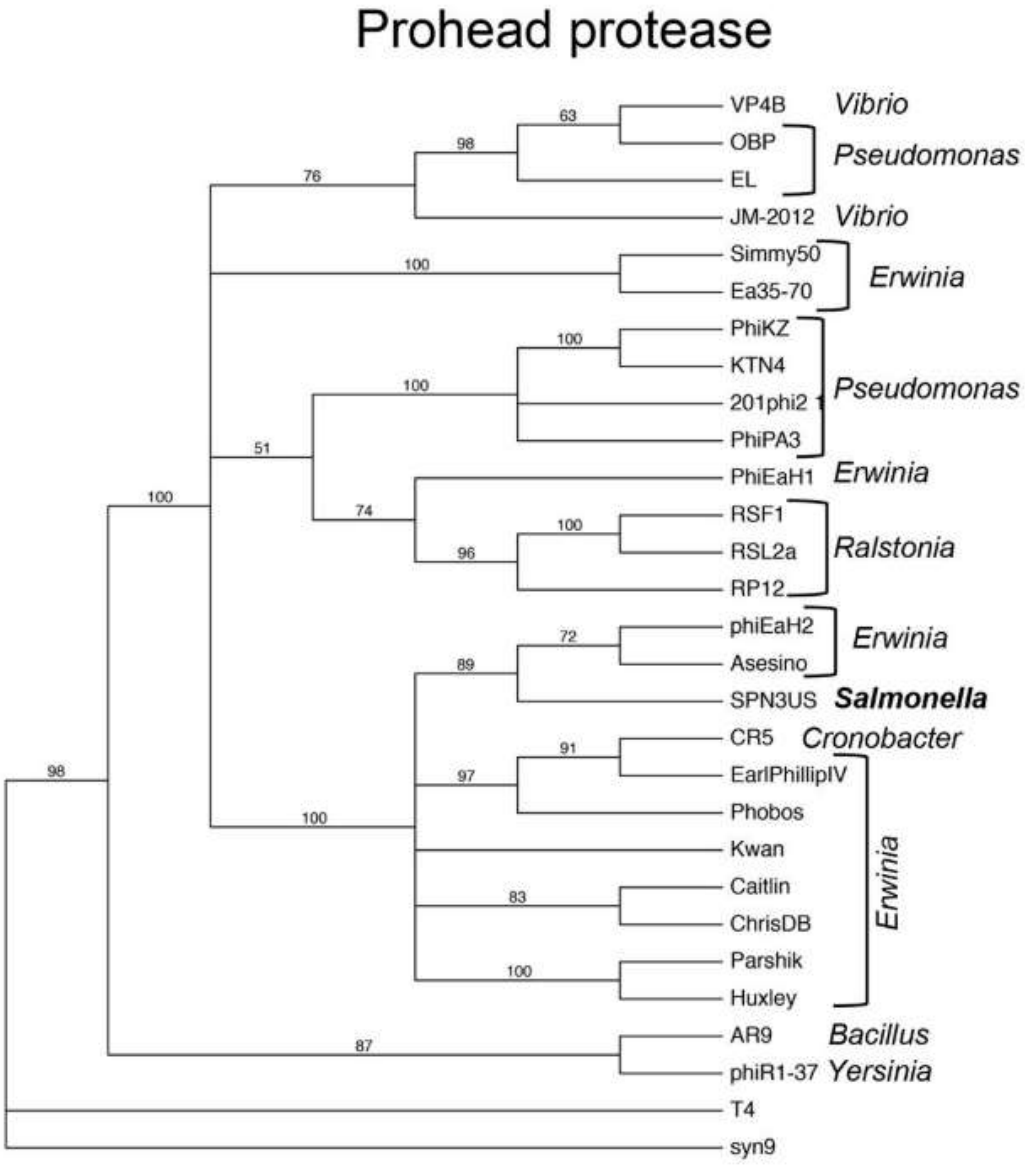
97	11.9	11.9	n/a (guess of 3)	4	as expected		
100	23.2	23.2	7.5	10	2.5 slices slower than expected		
109	17.7	17.7	5	7	2 slices slower than expected		
138	29.2	29.2	15	16	1 slice slower than expected		
139	29.8	29.8	15	16	1 slice slower than expected		
140	31.4	31.4	15	16	1 slice slower than expected		
141	32.6	32.6	15	19	4 slices slower than expected		
142	30.8	30.8	14.75	15	as expected		
143	31.9	31.9	15	16	1 slice slower than expected		
144	30	30	15.25	14	as expected		
145	50.8	50.8	25	26	1 slice slower than expected		
146	36.9	36.9	18	19	1 slice slower than expected		
147	33.7	33.7	16	16	as expected		
148	53.7	53.7	25.5	27	1.5 slices slower than expected		
149	36.2	36.2	20	19	1 slice faster than expected		
150	33.8	33.8	16	17	1 slice slower than expected		
151	52.2	52.2	25	27	2 slices slower than expected		
152	36.8	36.8	18.25	21	2.75 slices slower than expected		
153	33.8	33.8	16	17	1 slice slower than expected		
154	50	50	25	24	1 slice faster than expected		
155	78.3	78.3	30.5	30	as expected		
160	18.5	18.5	6.5	8	1.5 slices slower than expected		

193	19.9	19.9	7	9	2 slices slower than expected		
214	28.1	28.1	13	14	1 slice slower than expected		
218	25.2	25.2	10	12	2 slices slower than expected		
225	25.1		10	12	2 slices slower than expected	Yes	Yes
240	59.6	59.6	27	28	1 slice slower than expected		
241	159.1	159.1	n/a (guess of 35.5)	35	as expected		
242	10.5	10.5	n/a (guess of 3)	3	as expected		
243	54.6	54.6	26.5	27	as expected		
244	27	27	13	13	as expected		
245	30.7	23.4	15	11	4 slices faster than expected	Yes	Yes
246	23.9	23.9	7.25	12	4.75 slices slower than expected		
248	21	21	7	10	3 slices slower than expected		
257	34.2	34.2	18	18	as expected		
262	52.7	52.7	25	22	3 slices faster than expected		

There is a strong amount of conservation among the proteases of different giant phages (Fig 4.10). The protease of SPN3US, gp245, is especially closely related to two *Erwinia* phages.

Figure 4.10: Phylogenetic Tree of Prohead Proteases Across Different Giant Phages

Adapted from (Ali et al., 2017)



Chapter 5: *Salmonella* Proteins Identified in Mass Spectral Data

Introduction

Mass spectral proteomes of a virus can often also be used to glean information about the proteome of its host. Even well-purified virions will have remnants of host proteins remaining, and these will appear in the mass spectral data. While these are useful as a metric of contamination (as discussed in Chapter 2), they can also be used to determine which host proteins are essential for infection. In the case of SPN3US, *Salmonella* proteins remaining on the purified virions can indicate which of those proteins the virus is interacting with, and which are essential to host infection.

Methods

Viral purification and SDS-PAGE methods are as described in Chapter 2. The same mass spectral data is used as in other chapters, focusing instead on *Salmonella* proteins detected in the experiments. This includes one experiment done with a single CsCl gradient, and others done with a step gradient plus a buoyant density gradient. All host proteins mentioned are from the host organism *Salmonella enterica* typhimurium LT2.

Proteins that were not identified in all replicates and with total PSM of less than 5 were not considered in the data listed.

Results/Discussion

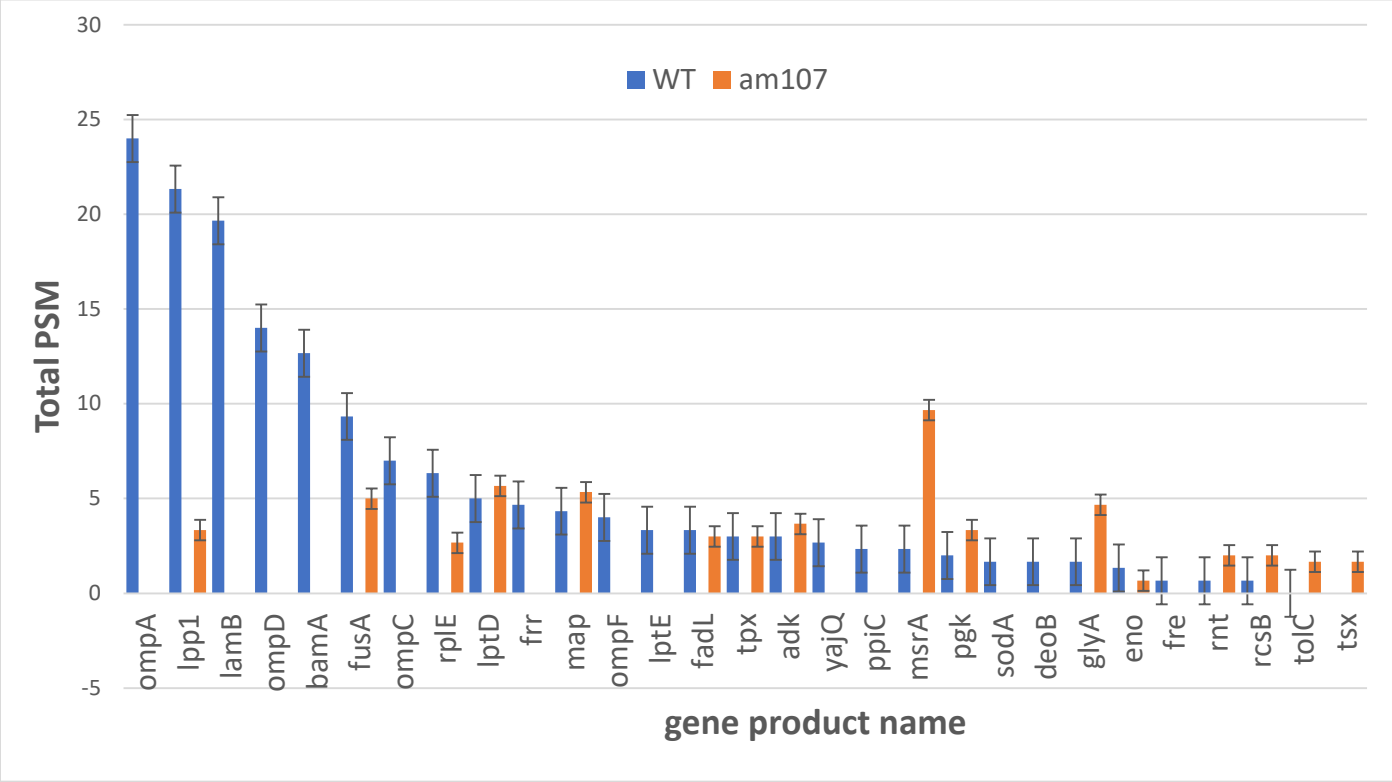
A total of 37 unique *Salmonella* proteins were discovered across all samples in both experiments. A total of 29 *Salmonella* proteins were present in 2 or more total PSM in at least 2 of 3 replicates of either the wild-type or tailless am107 experiment. Of these, 2 proteins, tolC and tsx, were found exclusively in am107 replicates. 13 proteins: ompA, lamB, ompD, bamA, ompC, frr, ompF, lptE, yajQ, ppiC, sodA, deoB, and fre were found exclusively in wild-type replicates.

The total PSM of *Salmonella* proteins was much higher on average in the wild-type virions (Figure 5.2), with only a few outliers, such as msrA.

Of these *Salmonella* proteins, 11 were enzymes, 10 were membrane proteins, and 3 were ribosomal. The remaining proteins were of miscellaneous functions (Full table is available in Appendix A.4). Membrane and ribosomal proteins being interacted with by the virion is consistent with their role in infection.

Future work regarding this should focus on examining the essentiality of each of these proteins to the infection of *Salmonella* by SPN3US. Any of these discovered genes which are non-essential to *Salmonella* could be knocked out in a mutant and infected. This would determine which proteins are mandatory for SPN3US infection, and what their roles could be in that regard. lamB, for example, is found in high amounts in SPN3US virions. This gene is non-essential and is known be the canonical receptor for lambda phage (Andrews and Fields, 2020). A knockout of this gene would be a good potential starting point for further research.

Figure 5.1: Comparison of Total PSM Identified In Purified Wild-type Phage versus. Tailless Mutant am107. Averages of 3 replicates are shown. Error bars represent standard deviation between replicates. See Appendix A.1/A.2 for full data set.



Chapter 6: Conclusions & Future Research

It is rapidly becoming essential to further our understanding of bacteriophage proteomics, from phage therapy applications to the ecological ramifications bacteriophages have. While mass spectrometry has not traditionally been used for phage proteomics, this research supports its increased adoption. This research focused on characterizing the virion of *Salmonella* phage SPN3US using high-performance mass spectrometry and comparing these data to those found previously using a standard Orbitrap. These analyses of SPN3US showed that phage proteins can be identified with high confidence scores across a wide dynamic range (based on total numbers of peptide spectrum matches for each protein identified) in a manner that is reproducible between replicates. The implementation of high-performance mass spectrometry was found to facilitate a more comprehensive characterization of the phage virion proteome as assessed by the detection of previously unidentified post-translational modification events in phage structural proteins.

Including the newly obtained Lumos mass spectrometer data, a total of 92 unique proteins were found to exist in the proteome of wild-type SPN3US. 54 unique proteins were found to be localized to the head of the virion, using the results from the tailless mutant am107. Newly discovered to have been processed was the protein gp225 (Figure 4.5). There were also some host *Salmonella* proteins that were found to be strongly associated with the virion, mostly consisting of membrane and ribosomal proteins.

This work serves as a foundation for improving functional characterization of the SPN3US proteome. S proteins had features suggestive but not confirmatory of processing, such as an

'AXE' motif or gel migration patterns that were inconsistent with what would be expected for their molecular weight, potentially indicating they were cleaved somewhere. One avenue of future work would be the usage of a different cleavage enzyme to potentially identify additional processed proteins. Ideally a variety of cleavage enzymes would be used in the experiment to identify all truly processed proteins across all of them. Processed proteins must be head proteins, and accessible to the prohead protease. Another avenue of potential future research could focus on the *Salmonella* proteins identified as being present in purified SPN3US samples. Although these proteins may simply be contaminants, it is also possible that some are associated with components of the virion in some way based on the degree of purification of each sample. Future studies could investigate whether SPN3US infection is impacted if each protein was knocked out in a *Salmonella* mutant. Any of these *Salmonella* proteins then found to be essential for infection could be targeted for further analyses of their specific roles in SPN3US infection/replication.

Future studies could focus on how best to implement mass spectrometry for phage proteomic analyses and the quantification of individual phage proteins. Overall, this supports the furthered use of mass spectrometry for phage proteomics, and is a step toward better understanding the proteome of SPN3US, which will help to understand the bacteriophages homologous to SPN3US as well.

Bibliography

Ali, B., Desmond, M.I., Mallory, S.A., Benítez, A.D., Buckley, L.J., Weintraub, S.T., Osier, M. V., Black, L.W., Thomas, J.A., 2017. To be or not to be T4: Evidence of a complex evolutionary pathway of head structure and assembly in giant Salmonella virus SPN3US. Front. Microbiol. 8, 1–21. https://doi.org/10.3389/fmicb.2017.02251

Andrews, B., Fields, S., 2020. Distinct patterns of mutational sensitivity for λ resistance and maltodextrin transport in escherichia coli lamb. Microb. Genomics 6. https://doi.org/10.1099/mgen.0.000364

Ceyssens, P.-J., Minakhin, L., Van den Bossche, A., Yakunina, M., Klimuk, E., Blasdel, B., De Smet, J., Noben, J.-P., Blasi, U., Severinov, K., Lavigne, R., 2014. Development of Giant Bacteriophage KZ Is Independent of the Host Transcription Apparatus. J. Virol. 88, 10501–10510. https://doi.org/10.1128/jvi.01347-14

Clokie, M.R.J., Millard, A.D., Letarov, A. V., Heaphy, S., 2011. Phages in nature. Bacteriophage 1, 31–45. https://doi.org/10.4161/bact.1.1.14942

Crooks, G., Hon, G., Chandonia, J., Brenner, S., 2004. NCBI GenBank FTP Site\nWebLogo: a sequence logo generator. Genome Res 14, 1188–1190. https://doi.org/10.1101/gr.849004.1

Danis-Wlodarczyk, K., Vandenheuvel, D., Jang, H. Bin, Briers, Y., Olszak, T., Arabski, M., Wasik, S., Drabik, M., Higgins, G., Tyrrell, J., Harvey, B.J., Noben, J.P., Lavigne, R., Drulis-Kawa, Z., 2016. A proposed integrated approach for the preclinical evaluation of phage therapy in Pseudomonas infections. Sci. Rep. 6, 1–13. https://doi.org/10.1038/srep28115

Duda, R.L., Teschke, C.M., 2019. The amazing HK97 fold: versatile results of modest differences. Curr. Opin. Virol. 36, 9–16. https://doi.org/10.1016/j.coviro.2019.02.001

Grace, E.R., Rabiey, M., Friman, V.P., Jackson, R.W., 2021. Seeing the forest for the trees: Use of phages

to treat bacterial tree diseases. Plant Pathol. 70, 1987–2004. https://doi.org/10.1111/ppa.13465

Heymann, J.B., Wang, B., Newcomb, W.W., Wu, W., Winkler, D.C., Cheng, N., Reilly, E.R., Hsia, R., Thomas, J.A., Steven, A.C., 2020. The Mottled Capsid of the Salmonella Giant Phage SPN3US , a Likely Maturation Intermediate with a Novel Internal Shell. Viruses.

Hu, B., Margolin, W., Molineux, I.J., Liu, J., 2015. Structural remodeling of bacteriophage T4 and host membranes during infection initiation. Proc. Natl. Acad. Sci. U. S. A. 112, E4919–E4928. https://doi.org/10.1073/pnas.1501064112

Hua, J., Huet, A., Lopez, C.A., Toropova, K., Pope, W.H., Duda, R.L., Hendrix, R.W., Conway, J.F., 2017. Capsids and Genomes of Jumbo-Sized Bacteriophages Reveal the Evolutionary Reach of the HK97 Fold. Am. Soc. Microbiol. 8, 1–15.

Krupovic, M., Koonin, E. V., 2017. Multiple origins of viral capsid proteins from cellular ancestors. Proc. Natl. Acad. Sci. U. S. A. 114, E2401–E2410. https://doi.org/10.1073/pnas.1621061114

Krylov, V.N., Zhazykov, Iz., 1978. [Pseudomonas bacteriophage phiKZ--possible model for studying the genetic control of morphogenesis]. Genetika 14, 678–685.

Lander, G.C., Evilevitch, A., Jeembaeva, M., Potter, C.S., Carragher, B., Johnson, J.E., 2008. Bacteriophage Lambda Stabilization by Auxiliary Protein gpD: Timing, Location, and Mechanism of Attachment Determined by Cryo-EM. Structure 16, 1399–1406. https://doi.org/10.1016/j.str.2008.05.016

Lavysh, D., Sokolova, M., Minakhin, L., Yakunina, M., Artamonova, T., Kozyavkin, S., Makarova, K.S., Koonin, E. V., Severinov, K., 2016. The genome of AR9, a giant transducing Bacillus phage encoding two multisubunit RNA polymerases. Virology 495, 185–196. https://doi.org/10.1016/j.virol.2016.04.030

Lecoutere, E., Ceyssens, P.J., Miroshnikov, K.A., Mesyanzhinov, V. V., Krylov, V.N., Noben, J.P., Robben,

J., Hertveldt, K., Volckaert, G., Lavigne, R., 2009. Identification and comparative analysis of the structural proteomes of ϕKZ and EL, two giant Pseudomonas aeruginosa bacteriophages. Proteomics 9, 3215–3219. https://doi.org/10.1002/pmic.200800727

Lee, J.-H., Shin, H., Kim, H., Ryu, S., 2011. Complete Genome Sequence of Salmonella Bacteriophage SPN3US. J. Virol. 85, 13470–13471. https://doi.org/10.1128/jvi.06344-11

Levy, M.J., Washburn, M.P., Florens, L., 2018. Probing the Sensitivity of the Orbitrap Lumos Mass Spectrometer Using a Standard Reference Protein in a Complex Background. J. Proteome Res. 17, 3586–3592. https://doi.org/10.1021/acs.jproteome.8b00269

M Iyer, L., Anantharaman, V., Krishnan, A., Burroughs, A.M., Aravind, L., 2021. Jumbo Phages: A Comparative Genomic Overview of Core Functions and Adaptions for Biological Conflicts. Viruses 13, 1–42. https://doi.org/10.3390/v13010063

Marvin, D., 1998. Filamentous phage structure, infection and assembly. Curr. Opin. Struct. Biol. 8, 150–158. https://doi.org/10.1016/S0959-440X(98)80032-8

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

Pietilä, M.K., Laurinmäki, P., Russell, D.A., Ko, C.C., Jacobs-Sera, D., Hendrix, R.W., Bamford, D.H., Butcher, S.J., 2013. Structure of the archaeal head-tailed virus HSTV-1 completes the HK97 fold story. Proc. Natl. Acad. Sci. U. S. A. 110, 10604–10609. https://doi.org/10.1073/pnas.1303047110

Reilly, E.R., Abajorga, M.K., Kiser, C., Mohd Redzuan, N.H., Haidar, Z., Adams, L.E., Diaz, R., Pinzon, J.A., Hudson, A.O., Black, L.W., Hsia, R.C., Weintraub, S.T., Thomas, J.A., 2020. A Cut above the Rest: Characterization of the Assembly of a Large Viral Icosahedral Capsid. Viruses 12, 1–22.

https://doi.org/10.3390/v12070725

Sharma, R., Pielstick, B.A., Bell, K.A., Nieman, T.B., Stubbs, O.A., Yeates, E.L., Baltrus, D.A., Grose, J.H., 2019. A Novel, Highly Related Jumbo Family of Bacteriophages That Were Isolated Against Erwinia. Front. Microbiol. 10, 1–16. https://doi.org/10.3389/fmicb.2019.01533

Thomas, J.A., Black, L.W., 2013. Mutational Analysis of the Pseudomonas aeruginosa Myovirus φKZ Morphogenetic Protease gp175. J. Virol. 87, 8713–8725. https://doi.org/10.1128/jvi.01008-13

Thomas, J.A., Weintraub, S.T., Hakala, K., Serwer, P., Hardies, S.C., 2010. Proteome of the large Pseudomonas myovirus 201ϕ2-1: Delineation of proteolytically processed virion proteins. Mol. Cell. Proteomics 9, 940–951. https://doi.org/10.1074/mcp.M900488-MCP200

Twarock, R., Luque, A., 2019. Structural puzzles in virology solved with an overarching icosahedral design principle. Nat. Commun. 10, 1–9. https://doi.org/10.1038/s41467-019-12367-3

Wu et al., 2013, 2012. Bubblegrams Reveal the Inner Body of Bacteriophage PhiKZ. Science (80-.). 23, 1–7. https://doi.org/10.1038/jid.2014.371

Yang, Q., Maluf, N.K., Catalano, C.E., 2008. Packaging of a Unit-Length Viral Genome: The Role of Nucleotides and the gpD Decoration Protein in Stable Nucleocapsid Assembly in Bacteriophage λ. J. Mol. Biol. 383, 1037–1048. https://doi.org/10.1016/j.jmb.2008.08.063

Yuan, Y., Gao, M., 2016. Proteomic analysis of a novel Bacillus Jumbo phage revealing glycoside hydrolase as structural component. Front. Microbiol. 7, 1–11. https://doi.org/10.3389/fmicb.2016.00745

Appendix

Table A.1: Complete Mass Spectra Data For All Wild-type SPN3US Proteins Identified In All Experiments

			WT Single-gradient Orbitrap	**WT Double-gradient Lumos-b**	**WT Double-gradient Lumos-b**	**WT Double-gradient Lumos-c**	**WT Double-gradient Lumos Average**	**WT Double-gradient Lumos Standard Deviation**	**am27 sup+**	**Comments (Head/tail/non-virion)**
gp	MW (kDa)	Mature MW (kDa)	Total PSM	Total PSM	Total PSM	Total PSM	Average PSM	SD PSM	Total PSM	
8	30.3	30.3	32	23	39	35	32.3	8.3	30	Head
17	15.3	15.3	13	31	26	27	28.0	2.6	21	Tail
21	40.7	40.7	20	27	23	21	23.7	3.1	19	Head
25	14.6	14.6	6	7	10	9	8.7	1.5	8	Tail

26	24.9	24.9	0	0	0	0	0.0	0.0	0	Non-virion
28	32	32	0	5	3	5	4.3	1.2	0	Head
33	12.3	12.3	0	0	3	0	1.0	1.7	0	Non-virion
37	14.6	14.6	0	4	6	7	5.7	1.5	4	Tail
38	23.2	23.2	4	6	4	7	5.7	1.5	0	Tail
41	31.8	31.8	8	7	19	11	12.3	6.1	11	Head
42	49.3	49.3	52	54	63	77	64.7	11.6	37	Head
43	19	19	0	0	2	0	0.7	1.2	0	Non-virion
45	50.3	48.2	144	239	224	253	238.7	14.5	124	Head
46	16.2	16.2	17	49	45	49	47.7	2.3	22	Head
47	62.8	50.7	110	358	345	253	318.7	57.2	205	Head
48	111	111	125	290	280	318	296.0	19.7	164	Tail
49	48.1	48.1	7	18	20	26	21.3	4.2	3	Tail
50	39.4	25.6	56	212	182	200	198.0	15.1	82	Head
51	34.9	34.9	81	117	100	108	108.3	8.5	90	Head

52	21	21	30	133	130	136	133.0	3.0	26	Head
53	45.2	31.5	662	1987	1873	1967	1942.3	60.9	708	Head
54	45.1	31.9	643	1088	1021	1071	1060.0	34.8	681	Head
61	58.3	58.3	35	63	65	68	65.3	2.5	33	Tail
62	52	52	44	79	65	70	71.3	7.1	35	Tail
64	48.9	48.9	20	39	47	41	42.3	4.2	9	Tail
73	58.6	58.6	0	0	0	0	0.0	0.0	0	Non-virion
74	12.9	12.9	40	141	125	134	133.3	8.0	45	Head
75	80.1	70.4	1592	4806	4666	4726	4732.7	70.2	1723	Head
81	100.2	72.3	93	177	176	180	177.7	2.1	125	Head
82	84.8	84.8	36	107	103	108	106.0	2.6	51	Tail
83	20	20	30	86	96	94	92.0	5.3	19	Head
84	32.9	32.9	51	83	81	79	81.0	2.0	66	Tail
91	23.6	23.6	13	26	33	31	30.0	3.6	10	Tail
94	41.6	41.6	35	37	33	34	34.7	2.1	12	Head

95	17.6	17.6	15	26	21	26	24.3	2.9	11	Head
97	11.9	11.9	7	8	9	8	8.3	0.6	6	Head
98	22.4	22.4	10	0	2	4	2.0	2.0	0	Possible contaminant
100	23.2	23.2	0	3	3	4	3.3	0.6	0	Head
103	24.1	24.1	0	0	0	0	0.0	0.0	0	Non-virion
104	37.6	37.6	0	0	0	0	0.0	0.0	0	Non-virion
108	10.6	10.6	0	0	0	0	0.0	0.0	0	Non-virion
109	17.7	17.7	40	44	39	32	38.3	6.0	28	Head
111	11.9	11.9	0	2	0	2	1.3	1.2	0	Non-virion
122	18.8	18.8	0	3	2	4	3.0	1.0	0	Tail
123	16.5	16.5	7	11	9	16	12.0	3.6	5	Tail
124	113.7	113.7	73	103	117	102	107.3	8.4	64	Tail
126	10.7	10.7	0	0	0	0	0.0	0.0	0	Non-virion
138	29.2	29.2	35	70	57	77	68.0	10.1	40	Head

139	29.8	29.8	37	56	55	58	56.3	1.5	44	Head
140	31.4	31.4	41	54	40	42	45.3	7.6	31	Head
141	32.6	32.6	306	561	487	521	523.0	37.0	420	Head
142	30.8	30.8	65	83	80	83	82.0	1.7	85	Head
143	31.9	31.9	118	248	249	246	247.7	1.5	234	Head
144	30	30	83	119	118	103	113.3	9.0	102	Head
145	50.8	50.8	46	37	39	42	39.3	2.5	30	Head
146	36.9	36.9	31	29	25	34	29.3	4.5	34	Head
147	33.7	33.7	17	33	31	31	31.7	1.2	23	Head
148	53.7	53.7	53	105	108	130	114.3	13.7	47	Head
149	36.2	36.2	78	125	111	103	113.0	11.1	65	Head
150	33.8	33.8	29	38	32	37	35.7	3.2	35	Head
151	52.2	52.2	25	58	55	55	56.0	1.7	34	Head
152	36.8	36.8	69	103	75	90	89.3	14.0	80	Head
153	33.8	33.8	26	40	46	39	41.7	3.8	27	Head

154	50	50	45	69	79	88	78.7	9.5	22	Head
155	78.3	78.3	55	154	153	162	156.3	4.9	80	Head
156	23.6	23.6	0	0	0	0	0.0	0.0	0	Non-virion
157	24.5	24.5	9	2	2	2	2.0	0.0	0	Tail
158	19.9	19.9	4	3	6	8	5.7	2.5	0	Tail
160	18.5	18.5	175	539	479	499	505.7	30.6	280	Head
161	25.9	25.9	0	0	0	0	0.0	0.0	0	Non-virion
162	21.9	21.9	0	0	0	0	0.0	0.0	0	Non-virion
163	50.3	50.3	0	0	0	0	0.0	0.0	0	Non-virion
167	6.9	6.9	85	171	167	210	182.7	23.8	79	Tail
168	188.1	188.1	203	329	366	374	356.3	24.0	292	Tail
169	149	149	210	377	395	419	397.0	21.1	252	Tail
170	135.4	135.4	135	246	236	246	242.7	5.8	189	Tail
171	47.6	47.6	35	54	56	53	54.3	1.5	30	Tail
172	7	7	6	0	0	0	0.0	0.0	0	Non-virion

173	34.6	34.6	8	5	8	6	6.3	1.5	8	Non-virion
174	12	12	0	2	0	2	1.3	1.2	0	Non-virion
176	22.3	22.3	0	0	0	0	0.0	0.0	0	Non-virion
178	20	20	0	0	0	0	0.0	0.0	0	Non-virion
193	19.9	19.9	15	45	32	40	39.0	6.6	9	Head
199	22.4	22.4	0	0	0	2	0.7	1.2	0	Non-virion
202	23.5	23.5	32	43	43	52	46.0	5.2	38	Tail
203	51.9	51.9	72	164	157	160	160.3	3.5	54	Tail
214	28.1	28.1	79	118	154	137	136.3	18.0	77	Head
218	25.2	25.2	27	42	47	45	44.7	2.5	23	Head
223	45.3	45.3	29	48	44	54	48.7	5.0	29	Tail
225	25.1	22.3	14	75	78	67	73.3	5.7	23	Tail
229	39	39	0	0	0	0	0.0	0.0	0	Tail
233	17.8	17.8	0	0	0	0	0.0	0.0	0	Tail
237	19.9	19.9	50	54	68	54	58.7	8.1	32	Tail

238	82.1	82.1	46	84	90	72	82.0	9.2	55	Tail
239	259.1	259.1	200	362	416	400	392.7	27.7	100	Tail
240	59.6	59.6	33	103	114	117	111.3	7.4	56	Head
241	159.1	159.1	189	285	278	315	292.7	19.7	203	Head
242	10.5	10.5	7	16	17	18	17.0	1.0	18	Head
243	54.6	54.6	251	471	419	448	446.0	26.1	221	Head
244	27	27	12	15	27	23	21.7	6.1	12	Head
245	30.7	23.4	25	63	73	65	67.0	5.3	28	Head
246	23.9	23.9	22	80	67	68	71.7	7.2	28	Head
248	21	21	17	25	24	24	24.3	0.6	14	Head
249	15.3	15.3	0	0	0	0	0.0	0.0	0	Non-virion
255	32.7	32.7	209	406	369	370	381.7	21.1	244	Tail
256	75.7	75.7	669	1847	1686	1686	1739.7	93.0	686	Tail
257	34.2	34.2	97	145	123	132	133.3	11.1	96	Head
258	96.4	96.4	76	196	209	199	201.3	6.8	124	Tail

259	61	61	115	348	305	316	323.0	22.3	153	Tail
262	52.7	52.7	46	69	58	86	71.0	14.1	41	Head
Sum of Total PSM			237	613	572	601			318	

Table A.2: Complete Mass Spectral Data For All Tailless Mutant SPN3US Proteins Identified In All Experiments

			am27 sup-	**am107 Double-gradient Lumos 6-slice-a**	**am107 Double-gradient lumos 6 slice-b**	**am107 Double-gradient lumos 6 slice-c**	**am107 Double-gradient Lumos Average**	**am107 Double-gradient Lumos Standard Deviation**	**am107 Double-gradient Lumos 37 slice**	**Comments (Head/tail/non-virion)**

gp	MW (kDa)	Mature MW (kDa)	Total PSM	Total PSM	Total PSM	Total PSM	Average PSM	SD PSM	Total PSM	
8	30.3	30.3	30	38	43	31	37.3	6.0	148	Head
17	15.3	15.3	0	0	0	0	0.0	0.0	2	Tail
21	40.7	40.7	22	23	25	26	24.7	1.5	95	Head
25	14.6	14.6	0	0	0	0	0.0	0.0	0	Tail
26	24.9	24.9	0	0	0	0	0.0	0.0	4	Non-virion
28	32	32	0	5	5	3	4.3	1.2	13	Head
33	12.3	12.3	0	0	0	0	0.0	0.0	0	Non-virion
37	14.6	14.6	0	0	0	0	0.0	0.0	0	Tail
38	23.2	23.2	0	2	0	3	1.7	1.5	8	Tail
41	31.8	31.8	0	16	11	15	14.0	2.6	80	Head
42	49.3	49.3	36	89	75	71	78.3	9.5	118	Head
43	19	19	0	2	0	0	0.7	1.2	5	Non-virion
45	50.3	48.2	128	255	235	226	238.7	14.8	749	Head

46	16.2	16.2	15	39	32	38	36.3	3.8	25	Head
47	62.8	50.7	108	255	235	226	238.7	14.8	634	Head
48	111	111	0	2	3	2	2.3	0.6	0	Tail
49	48.1	48.1	0	0	0	0	0.0	0.0	0	Tail
50	39.4	25.6	76	213	196	198	202.3	9.3	534	Head
51	34.9	34.9	86	123	117	103	114.3	10.3	367	Head
52	21	21	21	128	115	74	105.7	28.2	145	Head
53	45.2	31.5	703	2225	2240	2141	2202.0	53.4	4748	Head
54	45.1	31.9	645	1361	1238	1157	1252.0	102.7	4067	Head
61	58.3	58.3	0	0	2	0	0.7	1.2	2	Tail
62	52	52	0	0	5	4	3.0	2.6	8	Tail
64	48.9	48.9	0	0	0	0	0.0	0.0	3	Tail
73	58.6	58.6	0	0	0	0	0.0	0.0	3	Non-virion
74	12.9	12.9	35	126	132	135	131.0	4.6	231	Head
75	80.1	70.4	1691	4864	4799	4621	4761.3	125.8	7713	Head

81	100.2	72.3	105	170	182	175	175.7	6.0	365	Head
82	84.8	84.8	0	0	0	0	0.0	0.0	0	Tail
83	20	20	18	76	65	70	70.3	5.5	144	Head
84	32.9	32.9	0	4	0	0	1.3	2.3	14	Tail
91	23.6	23.6	0	0	0	0	0.0	0.0	0	Tail
94	41.6	41.6	17	21	25	18	21.3	3.5	64	Head
95	17.6	17.6	6	17	22	26	21.7	4.5	37	Head
97	11.9	11.9	3	6	9	10	8.3	2.1	23	Head
98	22.4	22.4	0	0	0	2	0.7	1.2	6	Possible contaminant
100	23.2	23.2	0	6	5	4	5.0	1.0	12	Head
103	24.1	24.1	0	0	0	0	0.0	0.0	3	Non-virion
104	37.6	37.6	0	0	0	0	0.0	0.0	3	Non-virion
108	10.6	10.6	0	0	2	0	0.7	1.2	0	Non-virion
109	17.7	17.7	38	44	52	53	49.7	4.9	74	Head

111	11.9	11.9	0	0	0	0	0.0	0.0	0	Non-virion
122	18.8	18.8	0	0	0	0	0.0	0.0	0	Tail
123	16.5	16.5	0	0	0	0	0.0	0.0	0	Tail
124	113.7	113.7	0	4	4	3	3.7	0.6	0	Tail
126	10.7	10.7	0	0	0	0	0.0	0.0	3	Non-virion
138	29.2	29.2	44	78	86	74	79.3	6.1	197	Head
139	29.8	29.8	36	48	56	55	53.0	4.4	225	Head
140	31.4	31.4	39	46	36	43	41.7	5.1	119	Head
141	32.6	32.6	326	542	573	533	549.3	21.0	2403	Head
142	30.8	30.8	63	93	98	65	85.3	17.8	415	Head
143	31.9	31.9	134	275	320	251	282.0	35.0	626	Head
144	30	30	91	104	121	100	108.3	11.2	422	Head
145	50.8	50.8	31	46	46	41	44.3	2.9	135	Head
146	36.9	36.9	31	31	28	25	28.0	3.0	103	Head
147	33.7	33.7	17	42	40	42	41.3	1.2	123	Head

148	53.7	53.7	52	93	102	94	96.3	4.9	284	Head
149	36.2	36.2	54	92	103	90	95.0	7.0	325	Head
150	33.8	33.8	20	34	33	29	32.0	2.6	112	Head
151	52.2	52.2	18	50	55	44	49.7	5.5	125	Head
152	36.8	36.8	52	104	119	120	114.3	9.0	381	Head
153	33.8	33.8	16	41	46	38	41.7	4.0	188	Head
154	50	50	30	65	56	61	60.7	4.5	117	Head
155	78.3	78.3	61	94	94	93	93.7	0.6	184	Head
156	23.6	23.6	0	0	0	0	0.0	0.0	4	Non-virion
157	24.5	24.5	0	0	0	0	0.0	0.0	0	Tail
158	19.9	19.9	0	0	7	7	4.7	4.0	7	Tail
160	18.5	18.5	149	411	517	593	507.0	91.4	348	Head
161	25.9	25.9	0	0	0	0	0.0	0.0	2	Non-virion
162	21.9	21.9	0	0	0	0	0.0	0.0	3	Non-virion
163	50.3	50.3	0	0	0	0	0.0	0.0	2	Non-virion

167	6.9	6.9	0	0	0	0	0.0	0.0	0	Tail
168	188.1	188.1	0	0	0	0	0.0	0.0	0	Tail
169	149	149	0	4	3	0	2.3	2.1	0	Tail
170	135.4	135.4	0	0	0	0	0.0	0.0	0	Tail
171	47.6	47.6	0	0	0	0	0.0	0.0	0	Tail
172	7	7	0	0	0	0	0.0	0.0	0	Non-virion
173	34.6	34.6	0	0	0	0	0.0	0.0	0	Non-virion
174	12	12	0	0	0	0	0.0	0.0	0	Non-virion
176	22.3	22.3	0	0	0	0	0.0	0.0	2	Non-virion
178	20	20	0	0	0	0	0.0	0.0	2	Non-virion
193	19.9	19.9	6	26	44	25	31.7	10.7	22	Head
199	22.4	22.4	0	0	0	2	0.7	1.2	3	Non-virion
202	23.5	23.5	0	0	0	0	0.0	0.0	0	Tail
203	51.9	51.9	0	0	0	0	0.0	0.0	0	Tail
214	28.1	28.1	68	143	170	135	149.3	18.3	446	Head

218	25.2	25.2	28	46	57	39	47.3	9.1	48	Head
223	45.3	45.3	0	0	0	0	0.0	0.0	0	Tail
225	25.1	22.3	15	61	53	45	53.0	8.0	44	Tail
229	39	39	0	0	0	0	0.0	0.0	2	Tail
233	17.8	17.8	0	0	0	0	0.0	0.0	4	Tail
237	19.9	19.9	0	2	5	0	2.3	2.5	2	Tail
238	82.1	82.1	0	0	0	0	0.0	0.0	2	Tail
239	259.1	259.1	0	0	2	0	0.7	1.2	0	Tail
240	59.6	59.6	38	133	124	115	124.0	9.0	179	Head
241	159.1	159.1	152	233	241	246	240.0	6.6	526	Head
242	10.5	10.5	10	15	12	15	14.0	1.7	47	Head
243	54.6	54.6	217	528	482	489	499.7	24.8	1426	Head
244	27	27	13	21	18	9	16.0	6.2	81	Head
245	30.7	23.4	26	61	62	57	60.0	2.6	89	Head
246	23.9	23.9	39	79	67	87	77.7	10.1	203	Head

248	21	21	10	26	30	24	26.7	3.1	46	Head
249	15.3	15.3	0	0	0	0	0.0	0.0	3	Non-virion
255	32.7	32.7	0	4	0	0	1.3	2.3	0	Tail
256	75.7	75.7	8	30	31	27	29.3	2.1	62	Tail
257	34.2	34.2	83	138	145	150	144.3	6.0	309	Head
258	96.4	96.4	0	0	0	2	0.7	1.2	0	Tail
259	61	61	0	0	0	0	0.0	0.0	0	Tail
262	52.7	52.7	42	83	80	87	83.3	3.5	123	Head
Sum of Total PSM			5802	14036	14036	13387			30971	

Table A.3: Summation of Copy Number Analyses. Double-gradient Lumos Wild-type Data- 99.9% protein threshold, 2 minimum peptides, 95% protein threshold. Total PSM listed is the average of 3 WT replicates. Uses expected copy number for one of the proteins to extrapolate the estimated copy numbers for the other proteins based on their total PSM.

gp	**Function**	**Expected Copy Number**	**Mature MW (kDa)**	**WT Total PSM**	**Total PSM/MW**	**Copy Number major capsid (gp75, 24)**	**Copy Number portal (gp81, 5.43)**	**Copy Number sheath (gp255, 23.12)**	**Copy Number tube (gp256, 11.74)**

75	Major Capsid	1615	70.41	4733	67.22	1613.29	365.01	1554.14	789.17
81	Portal	12	80.4	178	2.21	53.13	12.02	51.19	25.99
255	Sheath	270	32.7	382	11.68	280.37	63.43	270.09	137.15
256	Tube	270	75.7	1740	22.99	551.65	124.81	531.42	269.85
61	Neck	12	58.7	65	1.11	26.58	6.01	25.60	13.00
62	Neck	12	52	71	1.37	32.77	7.41	31.57	16.03
64	Neck	12	48.9	42	0.86	20.61	4.66	19.86	10.08

Table A.4: *Salmonella* Proteins Identified In Mass Spectral Data of SPN3US Virions

Identified Proteins	**Alternate ID**	**Accession Number**	**Functional Category**	**Molecular Weight**	**WT**	**WT**	**WT**	**am 107**	**am 107**	**am 107**
Phosphoglycerate kinase	pgk	PGK_SALTY	enzyme	41 kDa	0	3	4	11	9	9
Adenylate kinase	adk	KAD_SALTY	enzyme	23 kDa	2	5	6	6	5	5
Major outer membrane lipoprotein 1	lpp1	LPP1_SALTY	membrane	8 kDa	21	22	21	5	3	2
Ribosome-recycling factor	frr	RRF_SALTY	ribosomal	21 kDa	4	5	6	4	7	6
Serine hydroxymethyltransferase	glyA	GLYA_SALTY	enzyme	45 kDa	0	6	3	4	4	3
Probable thiol peroxidase	tpx	TPX_SALTY	enzyme	18 kDa	2	0	4	4	3	3
Elongation factor G	fusA	EFG_SALTY	elongation factor	78 kDa	7	12	9	3	6	6
Enolase	eno	ENO_SALTY	enzyme	46 kDa	0	2	3	3	5	6
Superoxide dismutase [Mn]	sodA	SODM_SALTY	enzyme	23 kDa	0	0	2	3	0	3
Methionine aminopeptidase	map	MAP1_SALTY	enzyme	29 kDa	4	4	2	3	3	3
Phosphopentomutase	deoB	DEOB_SALTY	enzyme	44 kDa	0	0	0	3	0	2
50S ribosomal protein L5	rplE	RL5_SALTY	ribosomal	20 kDa	6	6	7	3	5	0
Transcriptional regulatory protein RcsB	rcsB	RCSB_SALTY	miscellaneous	24 kDa	0	0	0	3	0	2

Ribonuclease T	rnt	RNT_SALTY	ribosomal enzyme	24 kDa	0	2	0	2	2	2
Superoxide dismutase [Fe]	sodB	SODF_SALTY	enzyme	21 kDa	0	0	0	2	0	0
UPF0234 protein YajQ	yajQ	YAJQ_SALTY	miscellaneous	18 kDa	2	4	3	2	4	3
Outer membrane protein A	ompA	OMPA_SALTY	membrane	38 kDa	24	19	29	0	0	0
Maltoporin	lamB	LAMB_SALTY	membrane	51 kDa	20	15	24	0	0	0
Outer membrane protein assembly factor BamA	bamA	BAMA_SALTY	membrane	90 kDa	8	12	18	0	0	0
Outer membrane porin protein OmpD	ompD	OMPD_SALTY	membrane	40 kDa	13	16	13	0	0	0
Outer membrane protein C	ompC	OMPC_SALTY	membrane	41 kDa	7	6	8	0	0	0
LPS-assembly protein LptD	lptD	LPTD_SALTY	miscellaneous	90 kDa	5	3	4	0	0	0
Long-chain fatty acid transport protein	fadL	FADL_SALTY	membrane	47 kDa	3	6	5	0	0	0
LPS-assembly lipoprotein LptE	lptE	LPTE_SALTY	miscellaneous	21 kDa	3	4	3	0	0	0
Peptidyl-prolyl cis-trans isomerase C	ppiC	PPIC_SALTY	enzyme	10 kDa	2	3	3	0	0	0
Phosphoenolpyruvate carboxykinase (ATP)	pckA	PCKA_SALTI	enzyme	60 kDa	0	0	2	0	0	0
Outer membrane protein TolC	tolC	TOLC_SALEN	membrane	54 kDa	0	2	3	0	0	0
Ribose-5-phosphate isomerase A	rpiA	RPIA_SALTY	enzyme	23 kDa	2	0	0	0	2	0
NAD(P)H-flavin reductase	fre	FRE_SALTY	enzyme	26 kDa	0	3	2	0	0	0
Peptide methionine sulfoxide reductase MsrA	msrA	MSRA_SALTY	enzyme	23 kDa	2	2	0	0	2	0
Adenine phosphoribosyltransferase	apt	APT_SALTY	enzyme	20 kDa	0	0	0	0	0	2

Outer membrane protein F	omp F	OMPF_SALTY	membrane	40 kDa	4	0	3	0	0	0
Virulence transcriptional regulatory protein PhoP	phoP	PHOP_SALTY	miscellaneous	26 kDa	2	0	0	0	2	0
TraT complement resistance protein	traT	TRAT_SALTM	miscellaneous	26 kDa	0	0	2	0	0	0
Nucleoside diphosphate kinase	ndk	NDK_SALTY	enzyme	16 kDa	2	0	0	0	0	0
Nucleoside-specific channel-forming protein tsx	tsx	TSX_SALTY	membrane	33 kDa	0	2	0	0	0	0
3-phosphoshikimate 1-carboxyvinyltransferase	aroA	AROA_SALTY	enzyme	46 kDa	0	0	0	0	0	2
				SUM OF TOTAL PSM	145	164	189	61	62	59

A.5: FASTA Source Text for WebLogo image (Figure 4.4)

```
>gp45
LADAKAVASESIGFQDKKKLA
>gp47
LSKDVAVAVESSRPGTIPEAV
>gp50
QKSDNKVATESSAGSVLPNNT
>gp53
GRGMTKAAQEGWKETLKDLFE
>gp54
ARKRRQVARESLWDDIKAFLQ
>gp75
YGKTPLVATEYYTNKDLDKNL
>gp81
FGLKPSVATESLRRSQSDFFE
>gp225
EKEVITVARESISTEKEREYC
>gp245
RGALQQLAQEYLQLAPKVSNL
```