DROPPING ACID MAKES YOU SEE STARS: SAMBA VIRUS AS A MODEL SYSTEM FOR STUDYING GIANT VIRUS GENOME RELEASE

By

Jason Robert Schrad

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Biochemistry and Molecular Biology - Doctor of Philosophy

2019

ABSTRACT

DROPPING ACID MAKES YOU SEE STARS:
SAMBA VIRUS AS A MODEL SYSTEM FOR STUDYING GIANT VIRUS GENOME RELEASE

By

Jason Robert Schrad

As their name implies, giant viruses (GV) are viruses of immense size. These viruses tend to have capsids larger than 300 nm and genomes that encode for over 1000 open reading frames. These viruses dwarf more common viruses, such as the human rhinovirus (common cold) that has a particle size of 30 nm and encodes for only 11 proteins. Some GV genomes even contain introns, a feature not typically associated with viruses as they were thought to have evolved towards simplicity. The discovery of these viruses challenged the canonical view of the virus as a small and simple biological entity and has cast some doubt on our current understanding of the definitions of life.

GV have been isolated from every continent on the planet, yet most share several conserved structural features. These conserved features include an internal lipid membrane that contains the dsDNA genome as well as a seal complex that closes the capsid prior to genome release. In icosahedral GV (Mimivirus-like GV), this seal complex sits atop the capsid at one specific vertex, the stargate vertex, which opens to facilitate genome release. The mechanisms that trigger release of the seal complex *in vivo* remain unknown. To fill some of the gaps in our knowledge of the GV life cycle, I have developed an *in vitro* system for studying GV genome release using Samba virus (SMBV), an icosahedral GV isolated from a tributary of the Amazon River in Brazil.

First, I developed a method to visualize SMBV using cryo-electron microscopy (cryo-EM), cryo-electron tomography (cryo-ET), and scanning electron microscopy (SEM). I then investigated the molecular forces responsible for maintaining the structural integrity of the SMBV external seal complex, treating SMBV particles with conditions known to disrupt viral capsids. Following each treatment, we determined the percentage of open SMBV particles, looking for conditions that induced a marked increase in open SMBV capsids. Both low pH (at or below pH 3) and high temperature (100 °C) triggered an increase in open SMBV particles, suggesting that electrostatic interactions and entropy, respectively, play a role in maintaining the structural integrity of the SMBV external seal complex. The role of these forces in maintaining external seal complex integrity is conserved throughout the icosahedral GV as three other GV shared similar structural responses to these conditions.

Following low pH treatment small cracks appear in the GV capsid, mimicking the initiation of the genome release process and facilitating release of infection-related proteins. I separated the released proteins from the remaining capsid via centrifugation and analyzed the two populations via differential mass spectrometry. Through these analyses we identified ~300 proteins that are released from SMBV and/or Tupanvirus soda lake, a GV isolated from an alkaline lake in Brazil, capsids during the initial stages of the infection process. These findings provide some of the first molecular information on the GV genome release process and hint at what triggers this process *in vivo*. This work also provides the first *in vitro* system capable of mimicking stages of the GV infection process, paving the way for future structural and biochemical studies of the GV life cycle.

Copyright by
JASON ROBERT SCHRAD
2019

This Dissertation is dedicated to my family, who always supported me and fueled my passion for learning. Thank you for pushing me and believing that this nose guard could make it in the academic world.

ACKNOWLEDGMENTS

It takes a village to raise a scientist. Without the help and support of many people the completion of this dissertation would not have been possible. First of all, thank you to my mentor Dr. Kristin Parent. It has been a joy to work in your lab these last years. You have taught me so much about being a scientist. The skills that you have taught me will continue to help me throughout my career. Thank you for all of your support and your guidance.

Many past and present members of the Parent lab have also played a major role in my development as a scientist. Thank you John for making sure the lab runs smoothly. Thank you Natalia for the help and support. Thank you to Sarah and Sundhar for your great conversations and for your advice as I neared the end of my dissertation. I would also like to thank all of the undergrads that I had the opportunity to mentor and interact with: Sophia, Madeline, Kaitlynne, Will, Hailee, Kendall, Neeraj, and Mitch.

I would also like to thank my committee members, Dr. Terje Dokland, Dr. Michael Feig, Dr. Michael Garavito, and Dr. Gemma Reguera. Thank you for your support and guidance and for pushing me to look beyond the microscopy and into the biology.

Thank you to the Department of Biochemistry and Molecular Biology for serving as my departmental home and for fostering a collaborative and nurturing space for science. Special thanks go out to the BMB departmental staff, especially Jessica Lawrence for her assistance in the administrative side of graduate school, as well as Ashley Parks and Mary Thompson for their assistance as we scrambled with the EM facility. Additionally, huge thanks go out to Pappan who was always around when I needed help with one of the computers or I just needed a laugh.

Thank you also to the department and to the family of Dr. Watson for the Jack Throck Watson Graduate Fellowship in Biochemistry.

Thank you to all of the friends that I have made throughout my time in East Lansing. Whether it was grabbing a coffee, having a few beers, or simply getting together to watch a game, you all helped me get through the slog that is graduate school.

Finally, I would like to give a special thank you to my family. Thank you to my parents, Kelly and Kevin, for fostering my love of science (I will never forget that little red microscope) and helping me reach this point of my career. Thank you to my grandparents, Bob and Joyce, Bob and Jerri, your love and support has kept me going throughout this process and I have learned so much from each of you. Thank you to my siblings, Alec and Ryan, for being there for me and for keeping me grounded when I needed it. I love you all and I could not have completed this journey without you.

TABLE OF CONTENTS

LIST OF TABLES ..... **xi**

LIST OF FIGURES ..... **xii**

KEY TO SYMBOLS AND ABBREVIATIONS ..... **xiii**

CHAPTER 1 INTRODUCTION **1**
- WHY STUDY VIRUSES? 2
- WHAT IS A VIRUS? 6
- GIANT VIRUSES 9
  - Giant Virus Discovery 9
  - What are Giant Viruses? 10
  - Giant Virus Pathogenicity 11
  - With Great Size Comes Great Stability 13
- VIRAL GENOME RELEASE 16
  - Common Viral Genome Release Strategies 16
- GIANT VIRUS GENOME RELEASE 22
  - Stages of the Giant Virus Genome Release Process 22
- THE SYSTEM: SAMBA VIRUS 26
  - Samba Virus as a Model System for Studying Giant Viruses 26
- CHALLENGES IN STUDYING GIANT VIRUSES 28
  - Biological Challenges in Giant Virus Research 28
  - Challenges in Giant Virus Structural Biology Research 29
  - Recent Advances in Cryo-EM Ease Giant Virus Structural Biology 32
- QUESTIONS ASKED AND ANSWERED IN THIS THESIS 35

CHAPTER 2 MICROSCOPIC CHARACTERIZATION OF THE BRAZILIAN GIANT SAMBA VIRUS **39**
- ABSTRACT 40
- INTRODUCTION 41
- MATERIALS AND METHODS 44
  - Virus Preparation 44
  - Preparation of Cryo Specimens 44
  - Low Dose Imaging Conditions 45
  - Cryo-Electron Tomography 45
  - Fluorescence Microscopy 46
  - Scanning Electron Microscopy 46
  - Capsid and Nucleocapsid Measurements 47
- RESULTS 49
  - Cryo-Electron Microscopy (Cryo-EM) Revealed the Size and Morphology of Samba Virus Particles 49

A Comparison of Mimivirus and Samba Virus Particles Through the use of Cryo-Electron Microscopy 54
Three-Dimensional Structural Information of the Entire Samba Virus Virion was Obtained Through the use of Cryo-Electron Tomography (Cryo-ET) 55
A Comparison of Samba Virus and Mimivirus Particles via Scanning Electron Microscopy (SEM) Revealed Differences in Capsid Regularity and Potential Viral Ultrastructure 58
Fluorescence Light Microscopy Revealed Biomolecular Composition and Ultrastructural Lattice Formation of Samba Virus and Mimivirus Particles 60
DISCUSSION 66
ACKNOWLEDGMENTS 67
SUPPLEMENTARY MATERIALS 68

CHAPTER 3 BOILING ACID MIMICS INTRACELLULAR GIANT VIRUS GENOME RELEASE 69
SUMMARY 70
INTRODUCTION 71
RESTULTS AND DISCUSSION 75
Samba Virus is Resistant to the Vast Majority of Chemical Treatments 75
Electrostatic Interactions are Critical for Samba Virus Starfish Stability 77
Increased Thermal Energy is Required for Nucleocapsid Release 82
A Combination of Low pH and High Temperature Results in Complete Samba Virus Genome Release 85
Molecular Forces That Stabilize the Samba Virus Stargate Vertex are Conserved Amongst Diverse Giant Viruses 86
Numerous Proteins Released From Giant Virus Capsids During Stargate Opening 89
Identifying the Proteins Released From Samba Virus and Tupanvirus Virions at the Initiation of Infection 93
Expected Protein Types are Released From Samba Virus and Tupanvirus Virions During Genome Release 120
Samba Virus and Tupanvirus Also Release Novel Proteins During Stargate Opening 122
Making Some Sense of the Myriad Hypothetical Proteins in the Samba Virus and Tupanvirus Proteome 124
Opening the Stargate to New Avenues of Giant Virus Research 126
ACKNOWLEDGEMENTS 129
AUTHOR CONTRIBUTIONS 130
STAR METHODS 131
Contact for Reagent and Resource Sharing 131
Experimental Model and Subject Details 131
Acanthamoeba castellanii 131
Giant Viruses 131
Method Details 132
Treatment of SMBV Particles and Image Analysis 132
Determining the Percentage of Open SMBV Particles 132
Conditions That Did Not Increase POP 132
pH Titration of SMBV Particles 133

High Temperature Incubation 133
Combining High Temperature and Low pH 133
Cryo-Electron Microscopy (Cryo-EM) and Cryo-Electron Tomography (Cryo-ET) 134
Sample Preparation 134
Single Particle Cryo-Electron Microscopy 134
Cryo-Electron Tomography 135
Scanning Electron Microscopy 135
SEM Preparation and Imaging 135
Differential Mass Spectrometry 136
Sample Preparation 136
Proteolytic Digestion 136
LC/MS/MS and Data Analysis 137
Mass Spectrometry Data Synthesis 137
Classification/Functional Annotation of Proteins Identified via MS 138
Quantification and Statistical Analysis 138
Mass Spectrometry Analysis 138
Data and Software Availability 138
SUPPLEMENTARY MATERIALS 139

CHAPTER 4 DISCUSSION AND CONCLUSIONS 140
SIGNIFICANCE 141
SUMMARY 142
Chapter 2: Microscopic Characterization of the Brazilian Giant Samba Virus 142
How Does One Visualize a Biological Entity as Large as SMBV? 142
Does SMBV Utilize a Stargate Vertex to Facilitate Genome Release? 143
How Structurally Similar are SMBV and APMV? 143
Chapter 3: Boiling Acid Mimics Intracellular Giant Virus Genome Release 144
What Molecular Forces Promote SMBV Starfish Seal Complex Stability? 144
Are These Molecular Forces Conserved Across *Mimiviridae*? 145
What Stages of the GV Genome Release Process can be Mimicked *In Vitro*? 145
What is the Fate of the External Seal Complex? 145
Which Proteins are Released From SMBV and TV Capsids at the Initiation of Infection? 146
CONCLUSIONS 148
FUTURE DIRECTIONS 150

APPENDICES 152
APPENDIX A: SAMBA VIRUS AND TUPANVIRUS SODA LAKE MASS SPECTROMETRY 153
APPENDIX B: SUPPLEMENTARY VIDEOS 201

REFERENCES 203

LIST OF TABLES

Table 1.1 Viral Genome Release Structures on the Electron Microscopy Databank (EMDB). ...19

Table 3.1 Conditions That SMBV Particles Resist 76

Table 3.2 Identification of Proteins Released from SMBV and TV Capsids 94

Table 3.3 SMBV and TV Proteins with LFQ Percentages and Comparison Between Supernatant and Pellet Levels 98

Table 3.4 Homology Predictions of SMBV and TV Released Proteins 109

Table 3.5 Homology Pairings for Released SMBV Proteins 110

Table 3.6 SMBV and TV Released Protein Homologues 113

Table 3.7 Identity of Proteins Released by SMBV or TV and Their Homologues....115

Table 3.8 SMBV and TV Released Hypothetical Proteins with Predicted Functionalities 125

Table A.1 SMBV Mass Spectrometry Intensities 154

Table A.2 SMBV Mass Spectrometry LFQ Intensities....161

Table A.3 SMBV Peptide Counts and Sequence Coverage....169

Table A.4 TV Mass Spectrometry Intensities 177

Table A.5 TV Mass Spectrometry LFQ Intensities....185

Table A.6 TV Peptide Counts and Sequence Coverage....193

LIST OF FIGURES

Figure 1.1 Cryo-Electron Micrograph of SMBV and Bacteriophage L 14

Figure 1.2 Unique Structural Features Associated with Viral Genome Release 18

Figure 1.3 Cartoon Representation of the GV Life Cycle 23

Figure 2.1 Cryo-Electron Microscopy Data From SMBV Particles.....48

Figure 2.2 Comparison of APMV and SMBV via Cryo-Electron Microscopy Reveals that SMBV is not a Rigid Quasi-Icosahedron Like APMV, and Displays a Larger Degree of Structural Variation.....52

Figure 2.3 Cryo-Electron Tomograms of SMBV Particles.....53

Figure 2.4 Scanning Electron Micrographs of SMBV and APMV Particles.....59

Figure 2.5 Fluorescence Light Microscopy of SMBV and APMV Particles.....63

Figure 3.1 Low pH and High Temperature Triggered an Increase in SMBV POP and Changed the Star-Shaped Radiation Damage Pattern.....78

Figure 3.2 Electron Microscopy of SMBV Genome Release Stages 80

Figure 3.3 Percentage of Fiberless SMBV Particles at Varying Temperatures.....84

Figure 3.4 Post Genome Release Particles From Four GV.....88

Figure 3.5 SDS-PAGE of pH 2-Treated SMBV and TV.....91

Figure 3.6 Sample Preparation for SDS-PAGE and LC/MS/MS Experiments 92

Figure 3.7 Comparison of Proteins Released by SMBV and TV 106

Figure 3.8 Homology Prediction of Proteins Released by SMBV and TV 108

Figure 3.9 Cartoon Model of Giant Virus Genome Release Stages 127

KEY TO SYMBOLS AND ABBREVIATIONS

Å	Angstrom
SMBV	Samba virus
APMV	Acanthamoeba polyphaga mimivirus
TV	Tupanvirus soda lake
GV	Giant virus
Cryo-EM	Cryo-electron Microscopy
Cryo-ET	Cryo-electron Tomography
TEM	Transmission Electron Microscopy
SEM	Scanning Electron Microscopy
RMC	Random Model Computation
GUI	Graphical User Interface
POP	Percentage of Open Particles
UPP	Ubiquitin-Proteasome Degradation Pathway
CCD	Charge Coupled Device
SIRT	Simultaneous Iterative Reconstruction
WBP	Weighted Back Projection
MS	Mass Spectrometry
RNAP	RNA Polymerase
LFQ	Label Free Quantification

CHAPTER 1

INTRODUCTION

Portions of this work were adapted from material originally published in the open access journal *Viruses* and are adapted and/or reproduced here under the auspices of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/)

Parent, K.N., Schrad, J.R., Cingolani, G. 2018. Breaking Symmetry in Viral Icosahedral Capsids as Seen Through the Lenses of X-ray Crystallography and Cryo-Electron Microscopy. *Viruses* doi: 10.3390/v10020067.

WHY STUDY VIRUSES?

Viruses are the most abundant biological entities on the planet, with an estimated particle count between 10^{30} and 10^{31} (*1*). By their simplest definition, these particles consist of genetic material (single stranded or double stranded, DNA or RNA) surrounded by a protein shell and are only able to propagate within a host cell (*2*). Viruses are ubiquitous, with species discovered on all seven continents and in extreme environments such as Brazilian soda lakes (*3*), the depths of the ocean (*3*), and the wind-blasted deserts of Antarctica (*4*). Viruses infect all three domains of life, Eukaryotes (*5, 6*), Bacteria (*7-9*), and Archaea (*10, 11*), and there are even viruses that hijack other viruses in order to replicate (*12-14*). All told, the mass of all of the viruses on the planet is estimated to be greater than one million adult blue whales (*15*).

Alongside their ubiquity, or perhaps because of it, viruses play a role in many aspects of modern society. When most people think of viruses, they think of times when themselves or a loved one had contracted a virus and gotten sick. Indeed, the traditional view of viruses has painted them as antagonists of human health. Viral outbreaks amongst humans are thought to date back to our first forays into settling down and building civilization (*16, 17*). By settling down and concentrating in prime locales, our ancestors unwittingly provided viruses with much easier routes of transmission. Some of the earliest documented cases of viral outbreaks have been traced back to ancient Egypt (polio and smallpox) (*18*) and ancient Greece (smallpox) (*19*). Prominent historical viral outbreaks include the introduction of smallpox (Variola major and Variola minor) to the Americas (*20*) and the global flu pandemic of 1918 (Influenza H1N1) (*21*). More recently, prominent viral outbreaks include the 2015 Zika outbreak (*22, 23*), the relatively frequent Ebola outbreaks of this decade (*24*), and the ongoing HIV epidemic that is still estimated to infect 1.7 million new people every year, as of 2018 (UNAIDS).

As mentioned previously, viruses do not only infect humans; they infect all three domains of life (*5, 10, 11, 25, 26*). There are many viruses that infect livestock, including Marek's disease in chickens (Marek's Disease Virus (*27*)), bluetongue disease in sheep and cattle (Bluetongue virus (*28*)), and foot and mouth disease (FMD virus (*29*)) in many ruminants. Diseases, such as those caused by African cassava mosaic virus (cassava (*30*)) and Brome mosaic virus (soybeans (*31*)), cause billions of dollars of lost crops every year (*32*). Viruses are also capable of devastating industries that rely on microorganisms, most notably bacteriophages (phages) killing off bacteria in the dairy industry (*33, 34*).

Although viruses are capable of causing catastrophic harm to both humans and other organisms, they are not always deadly or debilitating. In humans for example, rhinovirus and adenovirus each cause the common cold (*35*) and a *herpesvirus* (herpes simplex virus 2) is one of the primary causes of cold sores (*36*). Other, less severe human viruses include norovirus (diarrhea, not typically fatal (*37*)), varicella zoster virus (chicken pox, shingles (*38*)), and numerous viruses present in the human virome that have not been associated with disease (*39, 40*). Although these diseases do not usually result in death, they do represent a significant economic cost for modern society. Indeed, influenza virus alone was estimated to have an economic burden of $10.4 billion in direct medical costs and $16.3 billion in lost earnings, annually (*41*).

While viruses are rarely beneficial to their hosts, with some exceptions being transducing phages that can transfer antibiotic resistance and pathogenicity genes (*42, 43*), they are not always harmful to their environments or to human society. For example, in the ocean bacteriophages and cyanophages are estimated to kill 20-40% of the bacteria and cyanobacteria each day (*44, 45*). This mass parasitism prevents overpopulation of the oceans and results in a

yearly carbon turnover of 145 gigatons (*46*). Some viruses have been used to protect plants from parasites, including *Baculoviruses* that have been developed as an insecticide to prevent crop devastation from insects such as worms and moths (*32, 47*). Similarly, mycoviruses (fungus-infecting viruses) have been employed to eliminate devastating fungal crop diseases

In terms of human health, there have been numerous instances of viruses being used for the greater good. A prime example of a beneficial virus is Vaccinia virus, the virus that was used to create a vaccine against the Variola (smallpox) virus (*48*). Adeno-associated virus (AAV), among others, has been developed as a candidate for gene therapy treatments (*49, 50*) and phages are making a resurgence in the United States as a viable human therapy (*51*). There are a few commercially available phage applications (SalmoFresh, ShiggaShield, etc. (*Intralytix*)) that are in use throughout the country to prevent bacterial growth. One of the most widespread uses, and potentially the use that the most people come into contact with, is the use of products like ListShield (*Intralytix*) to prevent the growth of *Listeria spp.* on deli meats. Phages have been used in Eastern Europe for decades to treat and prevent bacterial infections (*52*), although their use has not yet become widespread in the United States. There have been a few recent cases in the US where critically ill patients have been granted permission to use phage therapy under the auspices of the Compassionate Care Act (*51, 53*) and some phage treatments are currently under clinical trial (*54*). With the ever-increasing threat (and reality) of antibiotic-resistant bacteria, so-called phage therapy is likely to explode within the US medical field.

Apart from medical and commercial applications, many basic biological principles and techniques used in microbiological/biochemical/biological research were discovered, tested, and/or pioneered in viruses. For example, our current understanding of DNA as the genetic material of an organism, as opposed to its protein, was originally derived from the Hershey-

Chase experiment that utilized bacteriophage and a sophisticated separation strategy (*55*). Many common practices and techniques in molecular biology labs, including transduction (*42*), restriction enzyme digestion (*56*), and the T7 promoter (*57*), were either developed during early virus research or utilize biological systems designed to boost or prevent virus infection. Even the CRISPR-Cas system that is currently being deployed in a myriad of fields and research avenues (*58, 59*) evolved as a defense against phage infection; a pseudo-immune system that recognizes and destroys small fragments of viral DNA.

While it is abundantly clear that viruses play a crucial role in many aspects of our lives, we still lack a fundamental understanding of most viruses and their lifecycles. Understanding these viruses, as well as the interactions between viruses and their hosts, is critical for creating efficient, and cost effective, treatments and preventions for serious viral diseases (*60*) as well as developing new techniques and tools for the laboratory. This knowledge may also lead to continuing paradigm shifts within the scientific community. There may not be another CRISPR-Cas9-esque leap without continued study of viruses.

WHAT IS A VIRUS?

As touched upon briefly above, by the simplest definition, a virus is a segment of genetic material that is encased within a proteinaceous shell and that is able to generate more copies of itself once inside a suitable host cell (*2*). Viruses, unlike most other biological entities, can utilize either DNA or RNA as their transmissible biological material. In fact, the most common classification system for viruses, the Baltimore classification system (*25*), categorizes viruses based on their genetic material and their path to mRNA. Viruses can have DNA or RNA genomes, with each nucleic acid having both single stranded (ss) and double stranded (ds) varieties. Some viruses encode for very few of their own proteins, requiring them to rely heavily upon the host replication factors to produce progeny (*61*). Other viruses encode for nearly all of the machinery of life, only lacking ribosomes and some metabolic proteins to complete the requirements of being alive (*3, 12, 62-65*). This discrepancy in the level of reliance on the host cell highlights the immense diversity that is on display within the virosphere. Viruses differ in everything from their physical size and the size of their genomes all the way down to the makeup of their genetic material and how they produce mRNA.

The most abundant, or at least the most commonly isolated, viruses are the dsDNA viruses (*25*) (Baltimore class I) and they include the tailed bacteriophages (*Caudovirales*) as well as human-infecting viruses such as Adenovirus and *Herpesviruses*. These viruses follow the traditional Central Dogma informational highway (DNA-(m)RNA-protein) throughout their lifecycles. Class II viruses are ssDNA viruses including some bacteriophages (PhiX174, M13) and Parvoviruses. These viruses encode for DNA-dependent DNA polymerases that allow the virus to produce dsDNA and then mRNA. Class III viruses are the dsRNA viruses that include the Reoviruses and the Rotavirsuses.

Classes IV and V both encompass ssRNA viruses, although they differ in the sense of their RNA in relation to their mRNA. Positive sense ssRNA viruses (IV) have their genome in the same sense as their eventual mRNA, and they must make a negative sense RNA strand to make additional positive sense strands (RNA-dependent RNA polymerases build off of the existing RNA and cannot make a positive sense strand directly from a positive sense strand). Class IV viruses include *Picornaviruses* such as human rhinoviruses and *Togaviruses* such as Eastern equine encephalitis virus (EEEV). Negative sense ssRNA viruses (V) have to make a positive sense copy of their genome for replication and they are able to use this copy as their mRNA. Notable members of Class V include the influenza viruses (*Orthomyxoviridae*) as well as rabies virus (*Rhabdoviridae*).

Class VI viruses are retroviruses, like HIV, that contain a ssRNA (+) genome but have evolved a RNA-dependent DNA polymerase to reverse transcribe their genomes into ssDNA. From there, they utilize a DNA-dependent DNA polymerase to create dsDNA that can be used to create mRNA through the usual channels. These viruses typically encode for one of more integration proteins, allowing them to invade the host cell genome and wait for the proper time to activate and propagate. The final class of viruses (Class VII) utilizes a gapped dsDNA genome that uses ssRNA as a template for reverse transcription of the missing DNA. The most notable Class VII virus is hepatitis B virus (HBV).

Viruses also differ greatly in terms of their genome size and the number of proteins they encode for. In theory, the smallest virus would be composed of a single protein surrounding a ssRNA (+) gene that encodes for that protein. In practice, however, even the smallest viruses utilize more than one protein. The smallest known virus, porcine circovirus, encodes for four proteins within ~2000 bases of genetic material (*66*). Some viruses, including the Human

Rhinovirus (one of the smallest known human-infecting viruses), encode for a single gene product that forms a polyprotein. This polyprotein is then cleaved into the protein subunits required for viral replication and assembly (11 in the case of rhinovirus) via posttranslational modification (*67*).

Although they differ on the specifics, all viruses undergo similar stages throughout their lifecycle: 1) Host Recognition and Attachment, 2) Entry and Genome Release, 3) Replication, 4) Packaging and Assembly, and 5) Exit (*68-70*). Viruses have evolved various mechanisms to carry out these processes. For example, some viruses have coupled transcription and genome release, utilizing the energy generated by this process to draw the last of the genome out of the capsid (*71*). Other viruses have combined the Packaging/Assembly and Exit stages, building outer capsid layers/capsules right at the cell surface and releasing as assembly occurs (*72*).

GIANT VIRUSES

Giant Virus Discovery

Traditionally, viruses have been viewed as physically small entities, not visible through optical light microscopy. This convention stems from the discovery of viruses in the 1890's (*73, 74*). In these experiments, sap from tobacco plants infected with a mosaic disease was passed through a "sterile" 0.2 μm filter to remove anything as large or larger than a bacterium. The filtered sap retained its infectivity, suggesting that the infectious agent was small enough to pass through the filter. Through this work, tobacco mosaic virus (TMV) was discovered and the term virus was coined. Although the actual size and structure of the TMV particles would not be determined until 80 years later (*75*), the method of its discovery would set a standard for viral sizes that would last for over a century.

Prior to the dawn of the 21st century, only a single virus was discovered that exceeded the 200 nm size limit. This virus, Cafeteria roenbergensis virus (CroV) has a capsid size of 300 nm (*76*). At least two other viruses, Paramecium bursaria chlorella virus 1 (PBCV-1) (*77*) and Chilo iridescence virus (CIV) (*78*) abutted this size limitation with 190 and 185 nm capsid diameters, respectively. This arbitrary viral size limitation was shattered in 2003, however, following the discovery of *Acanthamoeba polyphaga* mimivirus (APMV), the first truly giant virus (*79*).

In 1992, a pneumonia outbreak occurred in Bradford, England. The causative agent of this outbreak was isolated from a water-cooling tower (*i.e.* an industrial air conditioner) and was originally identified as a bacterium. This "bacterium", dubbed the Bradford coccus due to its apparent shape in the light microscope, it was not able to pass through a 0.2 μm filter and stained Gram positive (*79*). This organism lacked a 16S RNA sequence, suggesting that it was viral as opposed to bacterial, although at over 400 nm in diameter, it was judged much too large to be a

virus (by contemporary standards). Electron micrographs of these particles revealed an icosahedral particle surrounded by a layer of fibers, reminiscent of viral particles. Eventually, this organism was identified as a virus (APMV) and the order and family of *Megavirales* and *Mimiviridae* were founded (*79*). This discovery, or rather this classification, rocked the foundations of virology and biology (*80, 81*) and lead to the ever-expanding field of giant virus research (*13, 82-84*).

What are Giant Viruses?

As the name implies, giant viruses (GV) possess giant capsids. GV tend to have capsids larger than 300 nm and can have genomes over 2 Mbp (*3, 83, 85, 86*). These viruses tend to encode for over 900 proteins (*79, 81, 83, 87*) and some of their genomes even contain introns, a rarity for viruses, as they are thought to evolve towards simplicity. Some GV encode for translational proteins (*88, 89*), tRNAs and their synthetases (*80, 90*), and even ribosomal proteins (*91, 92*). These proteins have rarely, if ever, been seen in the virosphere prior to the characterization of GV and have sparked renewed debate on the origins of viruses and their status as living organisms and even as a potential fourth domain of life (*6, 80, 89, 93, 94*).

The most common delineation between giant and non-giant viruses is that GV are visible through traditional optical microscopy (*4, 92, 95*). This cutoff can be rather nebulous; indeed, there are two schools of thought on the size limit of GV capsids. One school of thought sets 300 nm as the lower limit for the GV classification whereas the other school classifies any virus with a capsid larger than 200 nm as a GV (*83, 85, 87, 96*). Throughout this dissertation, we will use the 300 nm cutoff for the limitation of GV. While this cutoff does exclude several important viruses, including PBCV-1 (*97*), CroV (*76*), and Faustovirus (*98, 99*), these near-giant viruses

tend to utilize a very different genome release strategy than their giant siblings. We will discuss GV genome release mechanisms in greater detail below. Briefly, the GV with capsids larger than 300 nm tend to release their genomes through unique capsid vertices (*92, 96, 100*) whereas the smaller viruses do not (*76, 97, 98*). There is a third definition of GV, based on the number of annotated proteins in GenBank (*101*), but this cutoff is even more restrictive and is predicated on the presence of previous biological studies of the viruses, which are lacking for many GV.

Regardless of the physical size used to determine which viruses are GV, these large viruses dwarf their smaller counterparts in both size and complexity. Almost 70% of known viruses encode for less than 10 proteins (*102*). In contrast, the smallest known viruses, the porcine circoviruses, contain their ~2000 base (ssDNA) genomes inside of ~17 nm capsids and encode for only 4 proteins (*66*). The smallest human-infecting viruses, the human rhinoviruses that are one cause of the common cold (*67, 103*), have ~30 nm capsids and contain ~7200 base genomes (ssRNA) that ultimately encodes for 11 proteins (*67*). Even the *Herpesviruses*, thought to be large viruses prior to the discovery of GV, only have capsid sizes of ~130 nm (*104*) and encode for less than 100 proteins (*103*).

Giant Virus Pathogenicity

The majority of currently isolated GV infect amoebal hosts (*83, 85*). While it may appear strange that this diverse class of viruses tends to infect the same type of organism, this trend may have more to do with the isolation of GV than with their inherent biology. Indeed, many of these viruses were isolated from environmental and clinical samples using amoebas as "bait" (*4, 100, 105*). In these studies, the potential GV-containing samples were introduced to amoebal culture that was then observed for production of viral progeny and resultant cell lysis. Whether amoebas

are the natural hosts for these viruses remains a point of contention. GV have demonstrated the ability to infect all types of professionally phagocytic cells including amoebas (*79, 83*), mouse macrophages (*106, 107*), human macrophages (*108*). As these viruses can infect phagocytic cells of various organisms, it may be that the barrier to GV infection is cell entry (via phagocytosis) as opposed to the ability to hijack the host machinery (*109*).

Aside from amoebas, GV have been isolated from many multicellular organisms. These organisms include leeches (*110*), oysters and other shellfish (*111*), cattle (*4*), and even humans (*14, 79, 112-114*). In humans these viruses have been linked to several conditions, most commonly respiratory conditions such as pneumonia (*79, 113, 114*). Mice that had been given an intracardiac inoculation of mimivirus particles developed pneumonia-like symptoms (*107*). An unfortunate laboratory technician was also accidentally inoculated with mimivirus particles and developed similar symptoms (*115*). Additionally, GV have been linked to several other conditions and diseases. GV have been shown to induce various inflammatory conditions in humans including lymphadenitis (*116*), arthritis (*117*), and an increased interferon immune response (*118*), although this last may simply be an immunogenic response and not a direct result of viral pathogenesis. GV, especially the icosahedral Marseillevirus (*65*), have been linked to various cancers, including lymphoma (*112, 119*).

Many of the diseases and conditions that are thought to be caused by GV could also be symptoms caused by the presence of amoebas, hence the debate over causality versus correlation. Amoeba can cause pneumonia in many animals (*120*) and many of the GV-associated inflammatory conditions may simply be immunogenic responses to the virus or the amoebal hosts. Many of the hosts that have yielded GV are also reservoirs for amoebas, shedding some doubt on the true hosts of these viruses. There is current debate within the GV field as to

whether the viruses actually are pathogenic to mammals or if their amoebal hosts cause the observed symptoms. Even in experiments using isolated GV, such as the inoculation of the mice (*106, 107*), it is difficult to rule out residual amoebal contamination within the viral sample. Although there is debate regarding their infectivity, to be on the safe side, GV should be considered potential human pathogens until further studies can determine that they are not.

With Great Size Comes Great Stability

GV have demonstrated a remarkable level of capsid stability, surviving and thriving in extreme environments. These environments include highly alkaline (pH 9-12) lakes (*3*), up to 3 km deep in the ocean (~300 x atmospheric pressure) (*3*), dry valleys in Antarctica (cold deserts) (*4*), and the Siberian permafrost (*62, 63*). To survive in these environments, GV have evolved extreme particle stability. Some of these viruses are so stable that they can survive inside of 30,000-year-old ice cores and emerge as infectious particles (*62, 121*). Many human-infecting viruses, such as influenza (*122*) and Zika virus (*123*) are not able to survive for even a week when dried onto objects at room temperature. Other human viruses can withstand a few hours dried onto stainless steel (*122, 123*), but over time their particles desiccate and degrade. GV, on the other hand, are able to persist for months on hospital equipment (*14, 124*) and even on research laboratory equipment such as cryo-EM tweezers (Figure 1.1).

Figure 1.1

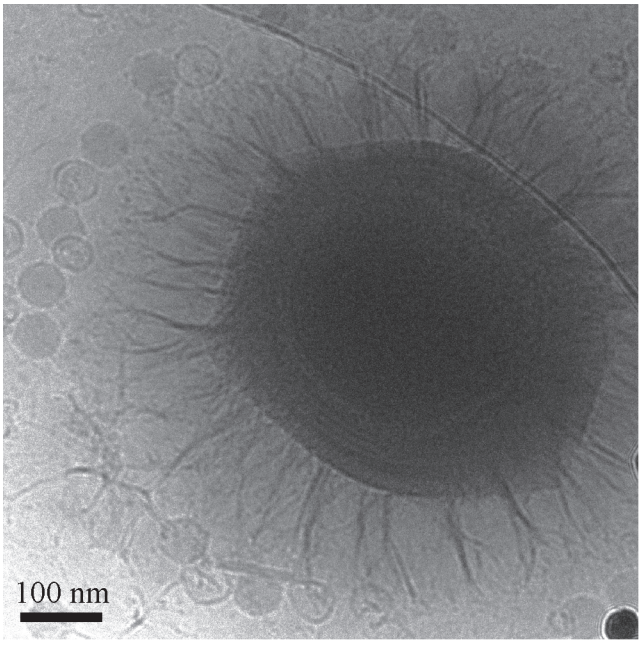

Figure 1.1 Cryo-Electron Micrograph of Samba Virus and Bacteriophage L. Cryo-electron micrograph depicting the size difference between Samba virus and a bacteriophage; phage L. Phage L is a *Podovirus* with an ~60 nm capsid. SMBV particles had adhered to the cryo-EM tweezers from a previous experiment (carried out nearly a month prior) and were resuspended by the addition of the 5 μL phage L particle droplet.

Particle stability can be beneficial to viruses, allowing them to persist in the environment as they await new host cells, however, it also presents a thermodynamic barrier that the viruses must overcome to initiate infection. Viruses encapsidate their genetic material within a proteinaceous shell, and, by definition, they cannot replicate within their own particles. For replication to occur, most viruses must break their capsid stability and release their genomes into the host cell. Viruses have evolved several mechanisms for overcoming this thermodynamic barrier and these structures and mechanisms tend to be conserved across viral families (*125, 126*). Examples include the structural changes in bacteriophage tail proteins that trigger genome release (*8, 127-129*), as well as conformational changes in fusion proteins in both influenza and Zika virus (*130-132*). Some viruses, however, have developed mechanisms to avoid releasing their genome into the host cell, producing new ssRNA molecules from within their capsids (*133, 134*). These viruses are largely dsRNA viruses such as *Rotaviruses* or *Reoviruses* and they are much less common than the viruses that break their capsids to facilitate replication (*25*).

Viral Genome Release

Common Viral Genome Release Strategies

Non-giant virus genome release strategies tend to fall into two categories; structural changes at a unique capsid vertex or more general structural rearrangements throughout the viral particles. Not all viruses fit into these two categories, however. Notable exceptions include syncytial viruses that force the host cell to fuse with nearby healthy cells, continuing the infection cycle without leaving the cellular environment (*135, 136*).

Many viral particles that utilize a unique capsid vertex trigger the necessary structural changes following interaction with one or more host-associated molecules (receptors) (*125, 126*). Tailed dsDNA bacteriophages (*Caudovirales*) represent some of the most well studied viruses that utilize unique capsid vertices. These viruses possess quasi-icosahedral capsids whose symmetry is disrupted at a unique vertex by the tail machinery (*125, 126*). Prior to genome release the tail complexes seal the capsid, preventing premature loss of DNA. Once a suitable host is found, the virus interacts with one or more host receptors, usually cell surface proteins or sugars (reviewed in (*137*)), leading to structural changes throughout the tail (*127, 138, 139*). This interaction is hypothesized to lead to a cascade of conformational changes, starting with the tail proteins and continuing into the portal complex that connects most bacteriophage tails to their capsids. These structural changes eventually trigger genome release.

Viruses that opt for more general structural changes, on the other hand, tend to use changes in the local environment (*e.g.* pH changes associated with internalization into the host cell (*140*)) to trigger conformational changes or cleavages in capsid-associated proteins (*130, 132*). Primary examples include HIV, which cleaves its Gag protein into capsid and nucleocapsid proteins as a precursor to infection (*141, 142*), influenza, which rearranges its H and NA proteins

following engulfment (*131, 143*), and Zika virus, which relies on conformational changes in its fusion peptides to trigger infection (*22*).

Regardless of the genome release mechanism utilized by a virus, the structures that are utilized in these processes tend to be conserved across viral families (Table 1.1, Figure 1.2, each adapted from (*126*)). Within the *Caudovirales* there are three structurally conserved tail morphologies; long contractile tails (*Myoviridae*), long non-contractile tails (*Siphoviridae*), and short tails (*Podoviridae*) (*7-9, 144*). Herpesviruses contain portal proteins that are structurally conserved with bacteriophage portal complexes (*138*) and these proteins are utilized in an analogous role during HSV-1 genome release (*145*). Similarly, many viral fusion proteins, utilized by many enveloped viruses to initiate infection, tend to take on one of three structures (*132*). This structural homology is even found across viral classes. For example, adenovirus spike proteins share structural homology with the tail needle knob of bacteriophage Sf6 (*129*).

Figure 1.2

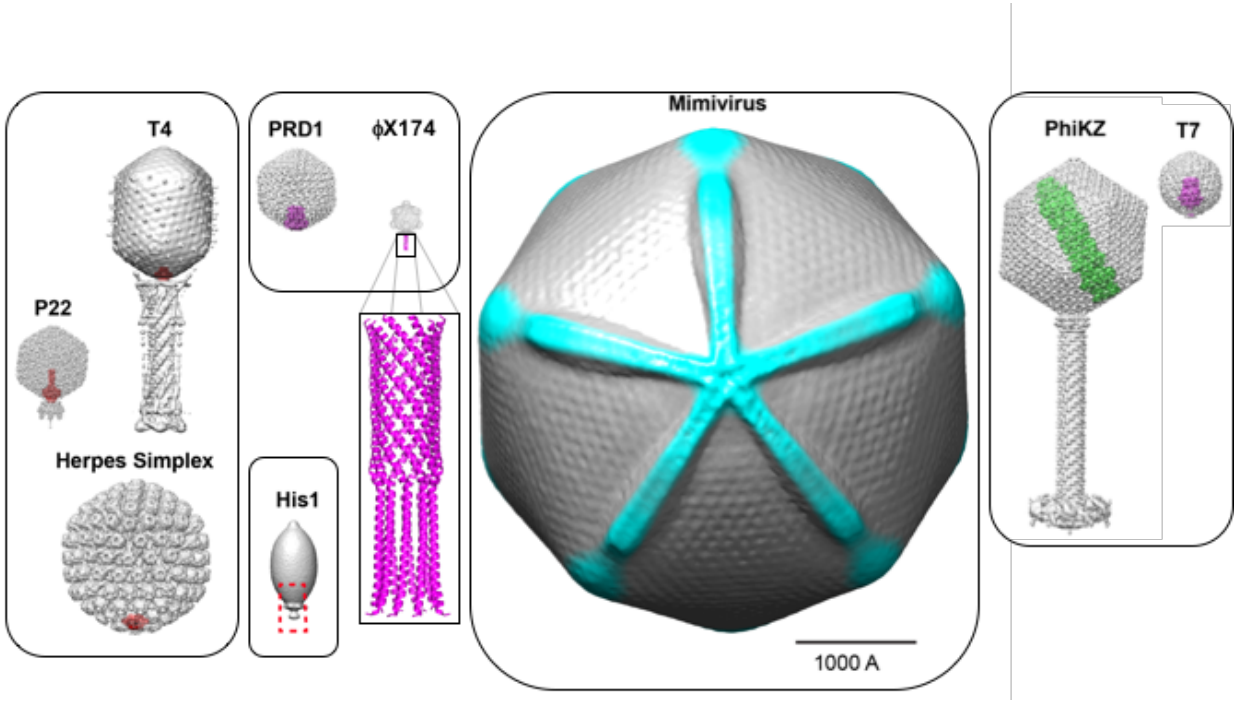

Figure 1.2 Unique Structural Features Associated With Viral Genome Release. Three-dimensional reconstructions of viral particles demonstrating the structural conservation of genome release structures. Bacteriophage and *Herpesvirus* portal proteins (PRD1, T7, T4, P22, Herpes Simplex; Purple) share structural homology and are grouped together in the left-most box. PRD1 and ϕX174 each utilize structural proteins that are released from the capsid upon genome release but that are hidden inside of the capsid prior to this event. His1 provides an example of a portal complex used by archaeal viruses. The long tail of PhiKZ is representative of the *Myoviridae* (bacteriophages with long contractile tails, but it also contains an inner body (Green) that is used during genome release. Mimivirus is the representative *Mimivirus* species and the reconstruction displayed here clearly demonstrates the starfish-shaped external seal complex. All viruses are to scale. This figure was adapted from (*126*) and is reused here under the auspices of the Creative Commons Attribution License. The EMDB ID's for the reconstructions are as follows: PRD1: EMD-5984, T7: EMD-5568, P22: EMD-8005; T4: EMD-2774; Herpes Simplex (HSV-1): EMD-5255, ϕX174: EMD-7033, His1: EMD-6223, Mimivirus: EMD-5039, PhiKZ: EMD-1415/EMD-1996.

Table 1.1

	Virus	*EMDB ID(s)*	*Cryo-EM*	*Cryo-ET*	*Year*
Podoviridae	Phi 29	1506, 5010	+++		2009
	Phi 29	1419, 1420	+++		2008
	Phi 29	6560	+++		2016
	CUS-3	5946	+++		2014
	Sf6	5730	+++		2013
	T7	5566-5573	+++		2013
	T7	5534-5537		+++	2013
	C1	5446	+++		2012
	P22	1119	+++		2005
	P22	1220	+++		2006
	P22	12222	+++		2006
	P22	1827	+++		2011
	P22	5348, 5231	+++		2011
	P22	8258-6261	+++		2016
	P22	8005	+++		2016
	P22	9010		+++	2018
	P22	7316	+++		2018
	P-SSP7	1707		+++	2010
	P-SSP7	6427		+++	2016
	P-SSP7	1714, 1715	+++		2010
	P-SSP7	3131		+++	TBP

Table 1.1 Viral Genome Release Structures on the Electron Microscopy Databank (EMDB). A tabulation of the viral structures available on the EMDB that are used during the genome release process (as of October 5th, 2019). These structures include phage tails, portal proteins, and other forms of unique viral vertices. The technique used to determine the structure, cryo-electron microscopy (cryo-EM) or cryo-electron tomography (cryo-ET), as well as the EMDB accession IDs are listed. This table is adapted from (*126*) under the auspices of the Creative Commons Attribution License.
*Non-icosahedral virus
**Giant Virus

Table 1.1 (cont'd)

	Virus	***EMDB ID(s)***	***Cryo-EM***	***Cryo-ET***	***Year***
Podoviridae	N4	1475	+++		2009
	Syn5	5743-5746		+++	2013
	ε15	1175	+++		2005
	ε15	5203, 5204		+++	2010
	ε15	5207-5209	+++		2010
	ε15	5216-5219		+++	2010
	BPP-1	1619		+++	2010
	K1E	1336	+++		2007
	K1-5	1337	+++		2007
Tectiviridae	PRD-1	3548-3550		+++	2017
	PRD-1	2438-2440		+++	2013
Myoviridae	PhiKZ	1415	+++		2007
	T4	1572, 1573	+++		2008
	T4	6323	+++		2015
	T4	2774, 6078-6083		+++	2015
Siphoviridae	P2	2463, 2464	+++		2013
	Araucaria	2335-2338	+++		2013
	1358	2820	+++		2016
	TW1	7070, 8854, 8867, 8868	+++		2017
ssDNA	ϕX174	7033, 8862	+++		2017
ssRNA	MS2	0338		+++	2019
	MS2	0448-0451	+++		2019
Archaeal	APBV1*	3857-3859	+++		2017
	His1*	6220-6222		+++	2015
Eukaryotic	HSV-1	5452, 5453		+++	2012
	HSV-1	5255, 5260, 5261	+++		2011
	HSV-1	1035-1038		+++	2007

Table 1.1 (cont'd)

	Virus	***EMDB ID(s)***	***Cryo-EM***	***Cryo-ET***	***Year***
Eukaryotic	HSV-1	4347	+++		2018
	HSV-1	9864	+++		2019
	KSHV	1320		+++	2007
	AAV2	622		+++	2019
	Epstein Barr Virus	10010	+++		2019
	Canine Parvovirus	20002	+++		2019
	Faustovirus	8144, 8145	+++		2016
	PBCV-1	1597	+++		2009
	PBCV-1	5384	+++		2012
	CroV**	8748	+++		2017
	Mimivirus**	5039	+++		2009
	Samba Virus**	8599		+++	2017
	CIV**	1580	+++		2009

Giant Virus Genome Release

Stages of the Giant Virus Genome Release Process

Similar to their smaller cousins, GV appear to also share conserved genome release mechanisms and structures. GV tend to combine the two common approaches found in smaller viruses, releasing their genomes through a unique vertex following phagocytosis and the resultant environmental changes that process entails. There are at least six stages of the GV genome release process: 1) Attachment/Host Recognition, 2) Phagocytosis, 3) Unique Vertex Opening, 4) Nucleocapsid Release and Fusion, 5) Viral Factory Formation and Replication, and 6) Release of Progeny (Figure 1.3).

Figure 1.3

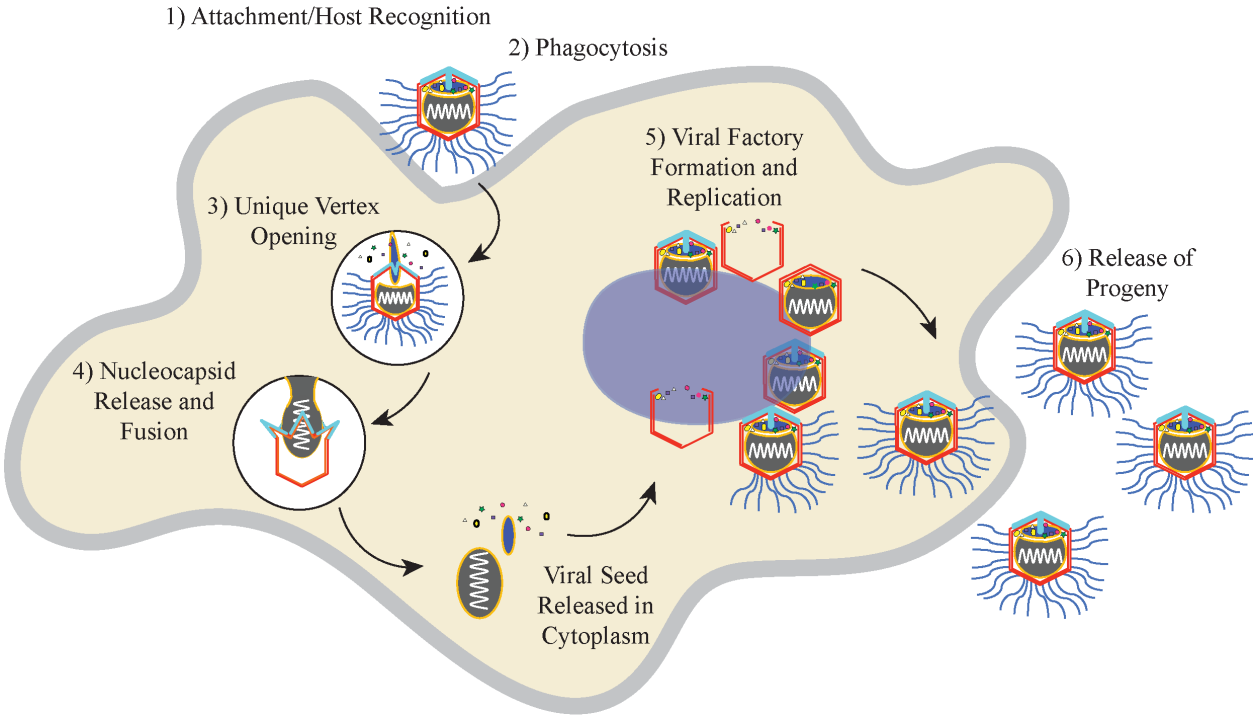

Figure 1.3 Cartoon Representation of the GV Life Cycle. Cartoon schematic of the known stages of the GV life cycle. These stages include 1) Attachment/Host Recognition, 2) Phagocytosis, 3) Unique Vertex Opening (disruption of the starfish seal complex or release of the cork-like seal), 4) Nucleocapsid Release and Fusion (with accompanying release of the viral seed into the cytoplasm), 5) Viral Factory Formation and Replication, and 6) Release of Progeny.

For host attachment and/or recognition, it is thought that the viruses use their external fiber layers, which are composed of a combination of protein and sugars (*109, 146-148*), to mimic bacterial cells (*109*). The host cell, believing that it has found a meal, engulfs the viral particle via phagocytosis. Once inside of the phagosome unknown triggers lead to seal complex disruption and opening of the unique vertex. Once the stability of the capsid has been bypassed, the genome containing lipid membrane (nucleocapsid) exits the capsid and fuses with the phagosomal membrane. This fusion releases the genome into the host cytoplasm where formation of the viral factory and production of GV progeny begins.

Prior to genome release, the unique capsid vertices are sealed by proteinaceous seal complexes (*63, 95, 149, 150*). GV have developed at least two types of seal complexes, either internal or external, and these complexes must be disrupted to facilitate genome release. Icosahedral GV, such as APMV and the newly discovered *Tupanviruses*, seal their unique vertices with star-shaped seal complexes, called starfish complexes (*92, 96, 150, 151*). These seals sit at a unique vertex on the icosahedral capsid, termed the stargate vertex due to its five-fold symmetry, which opens to facilitate genome release (*151*). Non-icosahedral GV, on the other hand, utilize seal complexes that resemble corks, sitting within the plane of the capsid as opposed to sitting on atop the capsid like the starfish complexes (*63, 149*).

Unlike many bacteriophages and other viruses with identified host receptors, the molecular triggers of GV seal complex disruption remain unknown. Indeed, although the stages that a GV must complete throughout its lifecycle are known (Figure 1.3), there is little data on the molecular/biomechanical changes that govern these stages. As mentioned previously, these viruses are incredibly complex and their sheer physical size has proven to be a challenge for structural studies. Some of the GV genome release stages have been visualized through

negatively stained, thin section transmission electron microscopy (TEM), but this technique is prone to structural artifacts (*68*). These artifacts can include structural damage from the massive changes in pH and salt concentration associated with negative staining as well as shearing marks from the sectioning process. Recent advances in cryo-electron microscopy (cryo-EM) have provided an avenue to study these viruses structurally (detailed below), although even with these advances GV are pushing the boundaries of the technique. The biological complexity of these viruses, as well as that of their hosts, has presented challenges in establishing model biological systems for these viruses. Throughout my dissertation, we have developed a new model system for studying GV infection using Samba virus (SMBV), a Brazilian GV.

The System: Samba Virus

Samba Virus as a Model System for Studying Giant Viruses

SMBV is a *Mimivirus* originally isolated from a tributary of the Amazon River in Brazil (*95*). This virus was isolated from surface water samples of the Rio Negro, a river with famously dark waters caused by the degradation of forest vegetation, near the city of Manaus. SMBV possesses an ~1.2 Mbp genome contained within an icosahedral capsid first thought to be ~350 nm in diameter. Using phylogenetic analyses of GV RNA reductase proteins, SMBV was placed within *Mimivirus* lineage A, the same lineage as APMV, the original GV (*82, 95*). SMBV encodes for over 900 ORFs, 91% of which are orthologous to APMV proteins. Almost half (~47%) of the SMBV ORFs shared homology with only other GV proteins and not with known proteins from other organisms, resulting in their annotation as hypothetical proteins.

Thin section TEM studies demonstrated that SMBV also shares many structural features with APMV (*95, 151, 152*). These features include a multi-layered capsid, an internal, genome-containing lipid membrane (the nucleocapsid), and a layer of external fibers. Initial TEM studies placed the SMBV capsid size at ~350 nm with an additional ~110 nm of external fibers, leaving SMBV slightly smaller in size than APMV. The sample preparation techniques used in these studies, namely fixation in plastic resin and the dehydration associated with negative staining (*68*), resulted in particle shrinkage. SMBV is, in fact, slightly larger that APMV under native conditions (see Chapter 2) (*95, 96*). Morphological characterization of SMBV particles also indicated the potential presence of a stargate/starfish vertex. Additional characterization of this unique vertex, and its function during the genome release process, can be found in Chapters 2 and 3 of this Thesis.

While it could be argued that a smaller, less complex virus could be used as a model system for studying GV genome release, the potential alternatives present their own set of limitations and challenges. The most obvious candidates for a simpler model system are the near-giant viruses such as PBCV-1, the Iridoviruses, or Faustovirus. These viruses are smaller than the mimivirus-like GV and do have less complicated genomes, however, these viruses do not contain stargate vertices and necessarily utilize a different genome release mechanism (*97-99, 153*) than the icosahedral GV. Studying these viruses would provide information about the biology and life cycles of *Mimiviridae*, but extrapolating information gleaned about their genome release to larger viruses would require more assumptions to be made than simply using SMBV.

Similarly, there are smaller viruses that release a lipid membrane during genome release. These viruses include Vaccinia and African Swine Fever Virus. Like the not-quite-giant viruses described above, these viruses do not utilize a stargate vertex during genome release (*154, 155*). Studying genome release in these viruses could provide insights into the GV genome release process, but application of this data to GV would require more assumptions than simply utilizing a GV in these studies.

SMBV is a prime candidate as a model system for studying GV. It shares many structural and genomic features with APMV and other lineage A *Mimiviruses* (*4, 64, 95, 96*). Crucially, SMBV utilizes the same genome release mechanism (the stargate/starfish vertex) as other *Mimiviruses*, providing opportunities for studying GV genome release. Unlike APMV and other GV, however, SMBV has not been associated with human disease (*84, 156, 157*), situating SMBV as an ideal candidate for studying GV in the laboratory.

Challenges in Studying Giant Viruses

Biological Challenges in Giant Virus Research

GV are incredibly complex for viruses. Their genomes are, by definition, orders of magnitude larger than their smaller cousins and many encode for around 1000 ORFs (*3, 79, 90, 95*). Many of these ORFs encode for proteins that do not share significant homology with known proteins from other organisms, including viruses. For example, APMV is predicted to encode for ~900 proteins. During the initial characterization of the APMV genome only ~300 of these proteins were assigned functional annotations, leaving the remaining two thirds of the predicted protein-encoding ORFs as hypothetical proteins of unknown function (*90*). Also, with the abundance of proteins utilized by these viruses, biochemical studies can become muddled. Separation of individual GV proteins can be challenging, as evidenced by the number of individual proteins (Table 3.2) identified from only five gel bands (Figure 3.5).

Similarly, the newly discovered *Tupanviruses* encode for ~1300 proteins. 775 *Tupanvirus* proteins have not appeared in other GV genomes and 375 of these proteins have not been seen in any the genome of any organism (termed ORFans) (*3*). *Tupanviruses* also encode for some of the most complete translational machinery of the virosphere (70 tRNAs, 20 tRNA synthetases, and at least 11 translation factors), and even encode for a mimic of an 18S RNA sequence (*3*). Pandoravirus salinus, the largest GV yet discovered, contains a 2.5 Mbp genome and is predicted to encode for over 2000 proteins (*86*).

Not only are GV the most complex entities in the virosphere, relatively little is known about the processes that govern the GV lifecycle. For example, no GV host receptor proteins have been discovered, leaving the molecular interactions that trigger genome release a mystery. While much of the lack of information on the GV life cycle can be attributed to the complexity

of the viruses themselves, the complexity of the amoebal hosts has presented its own challenges when studying GV. These amoebal hosts tend to be human pathogens (*120*) complicating GV research. Additionally, amoebas are relatively complex organisms, compared to the bacterial hosts of bacteriophages, further complicating the system.

Many of the challenges in studying GV could be alleviated through additional GV research. GV were identified in 2003 (*79*) meaning there has been less than 20 years of study on these viruses. Since the initial classification of APMV many new GV have been discovered (*3, 62, 63, 65, 86, 95, 113, 114, 149, 158*) and numerous studies have been performed on these viruses (reviewed in (*83, 100, 102*)). Each of these studies has resolved pieces of the jumbled puzzle that is the GV lifecycle.

Challenges in Giant Virus Structural Biology Research

The immense GV particle size makes structural studies incredibly challenging. Indeed, only one high resolution three-dimensional structure of a GV (APMV) has been published, to date (*150, 151*). Although it may seem counterintuitive, larger particles are more challenging to image through TEM (*68, 159*). This challenge stems from the nature of TEM and of cryo-EM.

TEM is a variant of electron microscopy that utilizes electron that have passed through a specimen to generate structural information about said specimen. Unlike other electron microscopy techniques (such as SEM) that visualize the specimen surface by detecting the electrons that have bounced back off of the sample, TEM is able to generate a 2D projection of the entire 3D structure of the specimen. As the sample is irradiated by the electron beam, the electrons pass through the sample and down to the electron detector. As the electrons interact

with the sample they can become scattered, changing the localization of the electrons on the detector and producing a scattering pattern. This pattern is then read by the electron detector (camera) to produce a micrograph.

As the sample is illuminated by the electron beam, it is being bombarded by thousands of electrons, typically on the order of ~40-50 electrons per square angstrom per exposure (*68*). This dose rate is roughly equivalent to the energy of an atomic bomb detonating across the street (*68*). Given this level of radiation, it should come as no surprise that biological samples are obliterated by TEM imaging without preservation and protection. Traditionally, this preservation has been provided by coating the samples with a heavy metal (*i.e.* uranium or tungsten) (*68*).

While it provides protection from the electron beam, as well as the high vacuum of the microscope, this preservation technique is prone to the generation of structural artifacts. The process of coating the particles with a heavy metal, called negative staining, involves large changes in salt concentration and pH as excess stain is wicked away. These rapid changes, alongside dehydration of the sample can lead to structural artifacts (*68*). To avoid the generation of these artifacts a novel method of sample preparation, rapidly freezing the samples (cryo-EM), was developed (*160*). In this technique, the sample is plunged into liquid ethane (-189 °C), freezing the sample so quickly that the water in the sample buffer does not have enough time to form crystalline lattices. The ice layer protects the sample from the vacuum of the microscope as well as from some of the radiation of the electron beam. The amorphous nature of the ice, however, does not produce a coherent scattering pattern, appearing transparent to the electron beam. This technique has been utilized to image all manner of microscopic organisms and processes (reviewed in (*68, 126, 160, 161*), among many others) and pioneers in its development were awarded with the 2017 Nobel Prize in Chemistry (*160*).

As mentioned above, TEM is based on the principle of localizing electrons that have passed through a sample. The more electrons that reach the detector in a specific area, the brighter the area of the resultant image becomes. As electrons interact with the sample, they become scattered and potentially lose energy (as opposed to simply being deflected). The larger the particle being imaged (*i.e.* the more atoms there are in the sample), the more electrons interact with the sample and deflect away from the detector. Additionally, as sample thickness increases, the likelihood of multiple scattering events increases. These events consist of electrons scattering off of multiple atoms within the sample, confusing the location of the atom(s) responsible for the scattering.

In addition to the sample, the vitreous ice that protects cryo-EM samples also scatters the electron beam. The amorphous nature of this ice layer typically prevents the electrons from scattering in a regular pattern, limiting the amount of signal caused by ice alone. When the ice layer becomes thicker, however, more scattering events occur, resulting in progressively darker images. For high-resolution data collection on small proteins, it is recommended that ice thickness is limited to 100 nm and below (*68*) with the ideal sample having only enough ice to cover the sample. A 100 nm ice layer would be impractical for GV cryo-EM as it would leave two thirds of the particle vulnerable to the vacuum of the microscope. For the larger GV the minimum ice thickness to fully engulf the particles is ~1 μm (*96, 151, 152*), ten times larger than the recommended thickness. Recent advances in cryo-EM imaging and sample preparation have mitigated some of these issues with imaging large specimens.

Recent Advances in Cryo-EM Ease Giant Virus Structural Biology

One of the most important pieces of equipment for imaging large samples, such as GV, is the presence of an energy filter on the microscope. These filters consist of an additional set of magnetic lenses placed between the specimen and the camera. These magnets are configured to only let electrons within a specified energy threshold pass through to the camera. In practice, energy filters are used to remove any electrons that have lost energy while passing through a sample. These electrons can lose energy by interacting with single atoms (inelastically scattered electrons) or through multiple scattering. By blocking these lower-energy electrons, the energy filter increases the inherent signal to noise ratio (SNR) of the micrographs. For samples as large as GV, an increased SNR is critical for visualizing structural features of the specimen.

In theory, another way to increase the SNR in cryo-electron micrographs of large specimens would be to increase the penetration of the electrons into the specimen by increasing the accelerating voltage of the microscope. With a higher accelerating voltage, the electrons are traveling faster through the sample and have less time to interact with the sample itself. To test this hypothesis, SMBV images were collected on a JEOL 2200-FS (200 keV accelerating voltage) at Michigan State University with an Omega energy filter and on JEOL 3200-FS (300 keV accelerating voltage) at Indiana University without a functional energy filter. The images taken at Indiana had a lower SNR than the images collected at Michigan State, (data not shown), demonstrating that the presence of an energy filter is more important for imaging GV than using a higher accelerating voltage.

A second advance in cryo-EM imaging of GV is the advent and improvement of direct electron detectors (*162*). These detectors directly detect electrons as opposed to using scintillators to convert the electrons into photons (used in older Charge Coupled Device (CCD)

camera imaging). Directly measuring the electrons that pass through the sample and onto the detector allows for a finer localization of the electron. These detectors are also capable of imaging in "movie mode," taking multiple images per second and stitching them together to create a final image. Collecting EM movies provides a means to correct for particle drift, the movement of the particles in ice caused by the interaction of the sample and the electron beam. Drift-corrected images appear sharper than non-corrected images as the blurring effect of particle motion has been removed (*163, 164*).

Drift correction allows for increased exposure times, which are critical for GV cryo-EM. With such large particles, GV must be imaged at relatively low nominal magnifications (*96, 151, 152*). At these lower magnifications the electron beam has a lower intensity, providing fewer electrons per $Å^2$ per second than at higher magnifications. With a lower dose rate, GV must be imaged for greater lengths of time (and potentially vulnerable to greater particle drift) to reach the standard total dose for cryo-EM of viral particles (30-50 $e^-/ Å^2$ (*68*)).

The combination of energy filters and direct detectors provides an opportunity for using non-standard cryo-EM techniques when studying GV. Cryo-electron tomography (cryo-ET) is a technique that can generate a three-dimensional structure of a single viral particle (*68, 159, 165*). In this technique, micrographs are collected along a range of specimen tilt angles, generating projections of the particles from various angles. These projections are then aligned and used to generate a three-dimensional volume of the particle(s) being imaged. Prior to the advent of direct electron detectors, tomography was not a feasible technique for GV as the relatively long exposure times, even when dividing the total electron dose amongst the tilt images, produced too much drift for accurate alignment (*166, 167*). Many GV have heterogeneous particle morphologies (*3, 62, 63, 96*), limiting the efficacy of single particle cryo-EM reconstructions.

Through tomography, structural information of individual GV particles can be generated, providing a three-dimensional glimpse at GV structural features.

Direct detector movie mode is also beneficial for generating "bubblegram" image series for GV particles. In this technique, the specimen is repeatedly exposed to the electron beam until radiation damage begins to build up (*168-170*). This damage, visualized via the build-up of H_2 gas through the interaction of the electron beam and proteins within the sample, can be used to locate unique features within viral capsids. Bubblegram imaging has been used to locate the inner body inside of the bacteriophage phiKZ capsid (*170*) and the ejection proteins in the bacteriophage P22 virion (*168*). Through the use of movie mode, individual frames throughout the bubblegram series can be combined to create a movie demonstrating the build-up of radiation damage over time and the location of unique structural features within GV capsids. An example of just such a movie demonstrating the star-shaped radiation damage pattern corresponding to the SMBV starfish seal complex can be seen in Supplemental Movie 2.

Taking advantage of these advances in cryo-EM imaging technology and techniques, we were able to visualize SMBV particles and fill in some of the gaps in the GV life cycle, specifically answering questions related to the GV genome release process.

Questions Asked and Answered in This Thesis

Many of the questions posed and answered within this Thesis revolve around the stages of the GV lifecycle. These questions revolve around SMBV and establishing it as a model system for studying GV. The questions posed throughout this work, as well as the Chapter of this Thesis in which these questions are answered (indicated in parentheses following the question) are as follows:

How does one visualize a biological entity as large as SMBV? (Chapter 2) Giant viruses have incredibly large particle sizes that, somewhat counterintuitively, make these viruses difficult to visualize through TEM (*68*). In order to answer biological questions about the GV lifecycle using structural biology, we had to first develop a system for visualizing these large particles and for generating structural data.

Does SMBV utilize a stargate vertex to facilitate genome release? (Chapter 2/Chapter 3) At the time of its discovery, SMBV was only the third GV that presented evidence of a stargate vertex, used during genome release (*113, 151*). This evidence sprang from negatively stained thin section TEM experiments, a technique that is prone to the generation of structural artifacts (*68*). To confirm the presence of an SMBV stargate vertex, and it's supposed use in the genome release process, we used cryo-EM and bubblegram imaging to locate the unique vertex and its external seal complex.

How structurally similar are SMBV and APMV? (Chapter 2) SMBV and APMV, the original GV (*79*), share a high level of sequence homology (*95*). Despite the genomic similarity, initial TEM imaging of these two viruses suggested that these viruses do not share a similar degree of structural homology (*95, 152*). To determine the degree of structural similarity between these two closely related viruses we analyzed each virus through cryo-EM, SEM, and fluorescence microscopy.

What molecular forces promote SMBV starfish seal complex stability? (Chapter 3) Little information is known about the molecular triggers responsible for the disruption of the GV starfish seal complex during genome release. To shed some light on this process, we treated SMBV particles with conditions known to disrupt other viral capsids (*e.g.* urea, guanidinium hydrochloride, low pH, high temperature) and analyzed the percentage of open SMBV particles via cryo-EM. Conditions that increased SMBV particle opening likely disrupted molecular forces that are responsible for starfish seal complex stability. These forces would need to be subverted during infection to facilitate genome release.

Are these molecular forces conserved across *Mimiviridae*? (Chapter 3) Disrupting electrostatic interactions (low pH) and increasing the thermal energy of the system (high temperature) each resulted in disruption of the SMBV starfish seal complex. As mentioned previously, SMBV is closely related to APMV and other GV. To determine if the molecular forces that promote SMBV starfish seal complex stability are conserved amongst other *Mimiviridae*, we treated APMV, Antarctica virus, and Tupanvirus soda lake with low pH and

high temperature. SEM imaging reveals that under these conditions, all three of these GV open their stargate vertices and release their genomes.

Which stages of the GV genome release process can be mimicked *in vitro*? (Chapter 3) Many GV infect amoebal hosts. While not necessarily as complex as human cell lines, amoebae are significantly larger and more complex than bacteria. Due to this complexity, along with the complexity of the viruses themselves, little information is available concerning the stages of the GV infection process. Amoebas are so large that structural studies of this process *in vivo* are impossible without Focused Ion Beam (FIB) milling (thinning thick cryo-EM samples using an ion beam) or thin sectioning. To study this relatively unknown process, we developed an *in vitro* system that mimics four distinct stages of the GV genome release process: 1) Native particles (**Pre-Release**), 2) **Initiation of Infection**, 3) **Nucleocapsid Release**, and 4) **Completion** (fully released).

What is the fate of the external starfish seal complex? (Chapter 3) While it is known that the external seal complex must be disrupted to facilitate GV genome release, the ultimate fate of this structure is unknown. There are two possibilities for the seal complex's fate: a) removal from the capsid *en masse* (like a star-shaped hat), or b) unzipping of the seal complex while maintaining contact with the stargate vertex. Through SEM, we provide evidence that the APMV, SMBV, and Antarctica virus external seal complexes unzip to facilitate genome release, but the Tupanvirus seal complex may release from the capsid *en masse*.

Which proteins are released from SMBV and Tupanvirus capsids at the Initiation of Infection? (Chapter 3) At the initiation of infection, the GV starfish seal complex unzips, facilitating release of the extra membrane sac and any free floating proteins within the capsid. To identify the proteins that are released at this stage of the GV genome release process, we separated free (released) and capsid-associated (not released) proteins and identified them via differential mass spectrometry. 86 proteins are released from the SMBV capsid and 56 proteins are released from the Tupanvirus soda lake capsid.

CHAPTER 2

MICROSCOPIC CHARACTERIZATION OF THE BRAZILIAN GIANT SAMBA VIRUS

This work was originally published in *Viruses* and is reused here under the auspices of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/).

Schrad, J.R., Young, E.J., Abrahão, J.S., Cortines, J.R., Parent, K.N. 2017. Microscopic Characterization of the Brazilian Giant Samba Virus. *Viruses* doi:10.3390/v9020030.

Minor edits have been made to this manuscript to conform to dissertation requirements.

ABSTRACT

Prior to the discovery of mimivirus in 2003, viruses were thought to be physically small and genetically simple. Mimivirus, with its ~750 nm particle size and its ~1.2 Mbp genome, shattered these notions and changed what it meant to be a virus. Since this discovery, the isolation and characterization of giant viruses has exploded. One of the more recently discovered giant viruses, Samba Virus, is a *Mimivirus* that was isolated from the Rio Negro in the Brazilian Amazon. Initial characterization of Samba revealed some structural information, although the preparation techniques used are prone to the generation of structural artifacts. To generate more native-like structural information for Samba, we analyzed the virus through cryo-electron microscopy, cryo-electron tomography, scanning electron microscopy, and fluorescence microscopy. These microscopy techniques demonstrated that Samba particles have a capsid diameter of ~527 nm and a fiber length of ~155 nm, making Samba the largest *Mimivirus* yet characterized. We also compared Samba to a fiberless mimivirus variant. Samba particles, unlike those of mimivirus, do not appear to be rigid, and quasi-icosahedral, although the two viruses share many common features, including a multi-layered capsid and an asymmetric nucleocapsid, which may be common amongst the *Mimiviruses*.

INTRODUCTION

Historically, following the discovery that the causative agent of tobacco mosaic disease could pass through sterile (0.22 μm) filters (*73*), viruses were thought to be small and simple; containing only a few genes (*171*). However, the re-classification of Acanthamoeba polyphaga mimivirus (APMV) (*79, 90*) fundamentally changed our understanding of viral life (*172*). Originally isolated in 1993 from a water cooling tower in Bradford, UK following a pneumonia outbreak, the so-called "Bradford coccus" was initially classified as a bacterium. These supposed cocci were visible under the light microscope and appeared to stain Gram positive (*79*). At the time, APMV was thought to be too large to be a virus (~700 nm particle diameter). It was not until 2003 that the inability to culture the Bradford coccus, and its lack of a 16S rDNA sequence, lead to the re-classification of this organism as the *mi*crobe *mi*micking (Mimi) virus (*79*).

Since then, dozens of other "giant" viruses, defined as viruses that are readily visible through light microscopy (*87*), have been discovered through co-culturing with *Acanthamoeba spp.* (*79, 87, 173*). Some of these newly discovered giant viruses fall into the viral families *Mimiviridae* (*82, 87*) and *Marseilleviridae* (*65, 173*), and many more remain unclassified, including Pandoravirus (*86, 174*), Pithovirus (*63*), Faustovirus (*98, 99*), and Mollivirus (*62*). Of these, *Mimiviridae* has been the most well-studied (*13, 82, 83*), and APMV is the only *Mimivirus* with detailed structural information available (*150-152, 175*). A three-dimensional reconstruction of APMV (EMD-5039) (*151*) shows that these viruses are comprised of a multi-layered capsid, an external layer of fibers, and an internal, genome-containing nucleocapsid (*151, 152*). In addition, structural data has elucidated that APMV releases its genome through a unique vertex, initially termed the "stargate" (*176*), which is closed by a protein complex called the "starfish"

seal (*151*). This unique vertex opens the capsid and releases the genome-containing nucleocapsid from within the virion.

One of the newer members of *Mimiviridae*, Samba virus (SMBV) was originally isolated from the Rio Negro, a tributary of the Amazon River, in Brazil (*95*). SMBV contains a ~1.2-Mbp double-stranded DNA genome, encoding for ~971 putative open reading frames (*85*). All known members of *Mimiviridae* infect *Acanthamoeba spp.* (*82*), and SMBV, specifically, infects *Acanthamoeba castellanii*. Once the viral infection process has begun, SMBV takes over the *A. castellanii* cellular machinery and creates a viral factory within the host cytoplasm (*177, 178*). Similar to APMV and its Sputnik virophage (*12*), SMBV has an associated virophage, Rio Negro virus (*95*).

As the prototypical member of *Mimiviridae*, and the first giant virus to be characterized, APMV has become the standard to which all subsequent members of this viral family are compared. As SMBV and APMV are both members of *Mimiviridae*, it is likely that the two viruses share some structural features. Some of these common features, including a multi-layered capsid, external fibers, etc., were observed during the initial isolation and characterization of SMBV particles (*95*). This original study utilized thin-section transmission electron microscopy (TEM) to generate a first glimpse of the structural features of the SMBV virion, estimating the total particle size (capsid + fibers) at ~575 nm. While this initial characterization provided invaluable structural and biological information about SMBV, the sample preparation techniques used during sectioning of biological samples are prone to the generation of structural artifacts (*68, 179*).

To obtain a more native-like view of the structural features present in the SMBV virion, we analyzed SMBV particles through the use of cryo-electron microscopy (cryo-EM), cryo-

electron tomography (cryo-ET), scanning electron microscopy (SEM), and fluorescence light microscopy. The vitrification process utilized during sample preparation for cryo-EM and cryo-ET (*68*) preserve the viral particles in a near-native state, limiting the generation of structural artifacts. While not as artifact-free as vitrification, the critical point drying technique used during the preparation of SEM samples avoids dehydration and physical shearing of particles that accompanies thin section sample preparation, providing more native-like structural information. Fluorescence light microscopy relies on the addition of fluorescent dyes, which may result in the generation of some structural artifacts, but this process retains specimens in a fully hydrated state.

To compare SMBV and APMV, we analyzed a fiberless variant of APMV (*64*) through the use of cryo-EM, SEM, and fluorescence microscopy. Two differences were readily apparent between SMBV and APMV; SMBV appeared to be less structurally rigid than the quasi-icosahedral particles of APMV, and the SMBV virion was larger than that of APMV. SMBV particles displayed a high level of structural heterogeneity and appeared to deviate from quasi-icosahedral symmetry in the cryo-electron and scanning electron micrographs. SMBV had a larger capsid (by ~27 nm), and longer fibers (by ~30 nm) than those of APMV (500-nm capsid diameter, 125-nm fiber length) (*152*), making SMBV the largest known *Mimivirus*. Aside from these readily visible differences, SMBV and APMV shared many common features, including the presence of multiple layers of the viral capsid, an external layer of fibers, etc. Given the relatedness of SMBV and APMV, we propose that the structural characteristics demonstrated here may be common amongst *Mimiviridae*.

MATERIALS AND METHODS

Virus Preparation

The giant viruses were both propagated following the same protocol. *A. castellanii* cells were cultured in 712 PYG w/Additives (ATCC), at pH 6.5, in the presence of the antibiotics gentamicin and penicillin/streptomycin, with final working concentrations of 15 μg/mL and 100 U/mL, respectively, to reach a 90% confluence. Cells were then counted using a Newbauer chamber and a solution of APMV or SMBV (diluted in PBS (phosphate buffered saline), just enough to cover the cell monolayer) was added to a multiplicity of infection (M.O.I.) of 10 for 1 h at room temperature. After the incubation was finished, PYG media was added in the presence of the antibiotics (above) and culture flasks were incubated at 28 °C for 48 h, when most of the amoebal cells were lysed as a result of the infection. The suspension containing cell debris and cell particles were centrifuged at 900× *g*; the resulting supernatant was carefully filtered using a 2-μm filter and then was immediately applied over a 22% sucrose cushion (*w*/*w*) at 15,000× *g* for 30 min. Visible white viral particle pellets were resuspended in PBS and stored at −80 °C. Viruses were titered using the Reed–Muench protocol (*180*). On average, virus isolation yielded 10^{10} $TCID_{50}$/mL (TCID = tissue culture infective dose).

Preparation of Cryo Specimens

Small (5 μL) aliquots of purified virus particles (either APMV or SMBV) were vitrified using established procedures (*68*). Samples were applied to holey Quantifoil grids (R3.5/1), which had been plasma cleaned for 20 s in a Fischione model 1020 plasma cleaner. Grids were blotted for 7–10 s using Whatman filter paper to remove excess sample, plunged into liquid

ethane for vitrification, and then transferred to a pre-cooled Gatan 914 specimen holder, which maintained the specimen at liquid nitrogen temperature.

Low-Dose Imaging Conditions

Virus particles were imaged in a JEOL JEM-2200FS TEM operating at 200 keV, using low-dose conditions controlled by SerialEM (v3.5.0_beta) (*167*) with the use of an in-column Omega Energy Filter, operating at a slit width of 35 eV. Micrographs were recorded using a Direct Electron DE-20 camera (Direct Electron, LP, San Diego, CA, USA), cooled to −40 °C. Movie correction was performed on whole frames using the Direct Electron software package, v2.7.1 (*181*). Micrographs used for single particle analysis were recorded on the DE-20 using a capture rate of 25 frames per second for a total exposure ranging from 75 to 300 frames (~35 e^- /Å^2 total dose recorded at the DE-20 sensor). Cryo-EM images were acquired between 4000 and 20,000× nominal magnifications (14.7–2.61 Å/pixel, respectively). The objective lens defocus settings for single particle images ranged from 15 to 25 μm underfocus.

Cryo-Electron Tomography

After plasma cleaning, but prior to the addition of SMBV particles, 5 μL of a solution of 10 nm nanogold fiducial markers were air-dried onto holey carbon grids. Tilt series projections were acquired using SerialEM (v3.5.0_beta) (*167*) at a capture rate of 15 frames per second for 45 frames per tilt angle, along a tilt range of ±55° with tilt increments of 1–2° and 0.7 electrons per square angstrom per tilt image. Tilt series were acquired at 4000 or 8000× nominal magnification (14.7 or 6.87 Å/pixel). Tilt series alignment was performed using IMOD (v4.7.15) (*182*) and standard tomographic reconstruction practices, using both the SIRT (simultaneous

iterative reconstruction) and WBP (weighted back projection) reconstruction strategies. The contrast in the tomograms generated using SIRT was far better than the contrast in the tomograms generated using the WBP reconstruction strategy, therefore we have presented the SIRT data here. Contrast was increased in the tomograms through median (x3) and Gaussian (1.5 pixels) filtering. Key features of the tomograms were traced using the drawing tools functionality in IMOD (3dmod).

Fluorescence Microscopy

APMV and SMBV particles were stained with 1 μg/mL 4',6-diamino-phenylindole (DAPI, DNA) and 0.1 μg/mL fluorescein isothiocyanate (FITC, protein) overnight. Virus particles were then imaged using a Zeiss Axio Observer A1 microscope (100×, 1.45 NA) outfitted with an Axiocam ICc5 camera. DAPI fluorescence was imaged with Zeiss filter set 49 and FITC fluorescence was imaged with Zeiss filter set 38 HE. Micrographs were then processed using Zeiss Zen software.

Scanning Electron Microscopy

SMBV particles were imaged using the in-lens detector of a JEOL JSM-7500F (SMBV) or a FEG Quanta 200 FEI (APMV) scanning electron microscope; operating at 5 kV (JSM-7500F) or 15kV (Quanta 200). Prior to imaging, virus particles were desiccated using an EM CPD300 critical point dryer, fixed with glutaraldehyde in PBS buffer at pH = 7.4 onto poly-l-Lysine treated SEM slides, and sputter coated with a ~2.7-nm layer of iridium using a Q150T Turbo Pumped Coater. Particles were imaged between 7000× and 50,000× nominal magnification.

Capsid and Nucleocapsid Measurements

Capsid and total particle diameters of APMV (274 particles from 94 micrographs) and SMBV (500 particles from 226 micrographs) were measured from two-dimensional projections of cryo-electron micrographs. Capsid diameter and total particle diameter measured across three axes (putative five-fold to five-fold) for each viral particle were analyzed (Figure 2.1A). The length of the SMBV fibers was determined by subtracting the capsid diameter from the total particle diameter and dividing by two. All other measurements were taken using three-dimensional volumes resulting from cryo-electron tomograms. The spacing of the SMBV capsid layers was measured (11 total tomograms). Nucleocapsid dimensions could only be conclusively measured in 8 out of the 34 total tomograms, owing to contrast limitations. Nucleocapsid diameter was measured along four axes with one axis bisecting the portion of the nucleocapsid that is pulled away from the capsid, and another axis normal to the bisecting axis. The distance from the nucleocapsid to the innermost layer of the capsid was measured at the pulled away region. Capsid spacing and the distance between the nucleocapsid and the capsid were measured at ten locations throughout the remainder of the virion, in order to obtain average values throughout the SMBV capsid. All measurements were taken using the measure tool in EMAN2's e2display.py GUI (*183*).

Figure 2.1

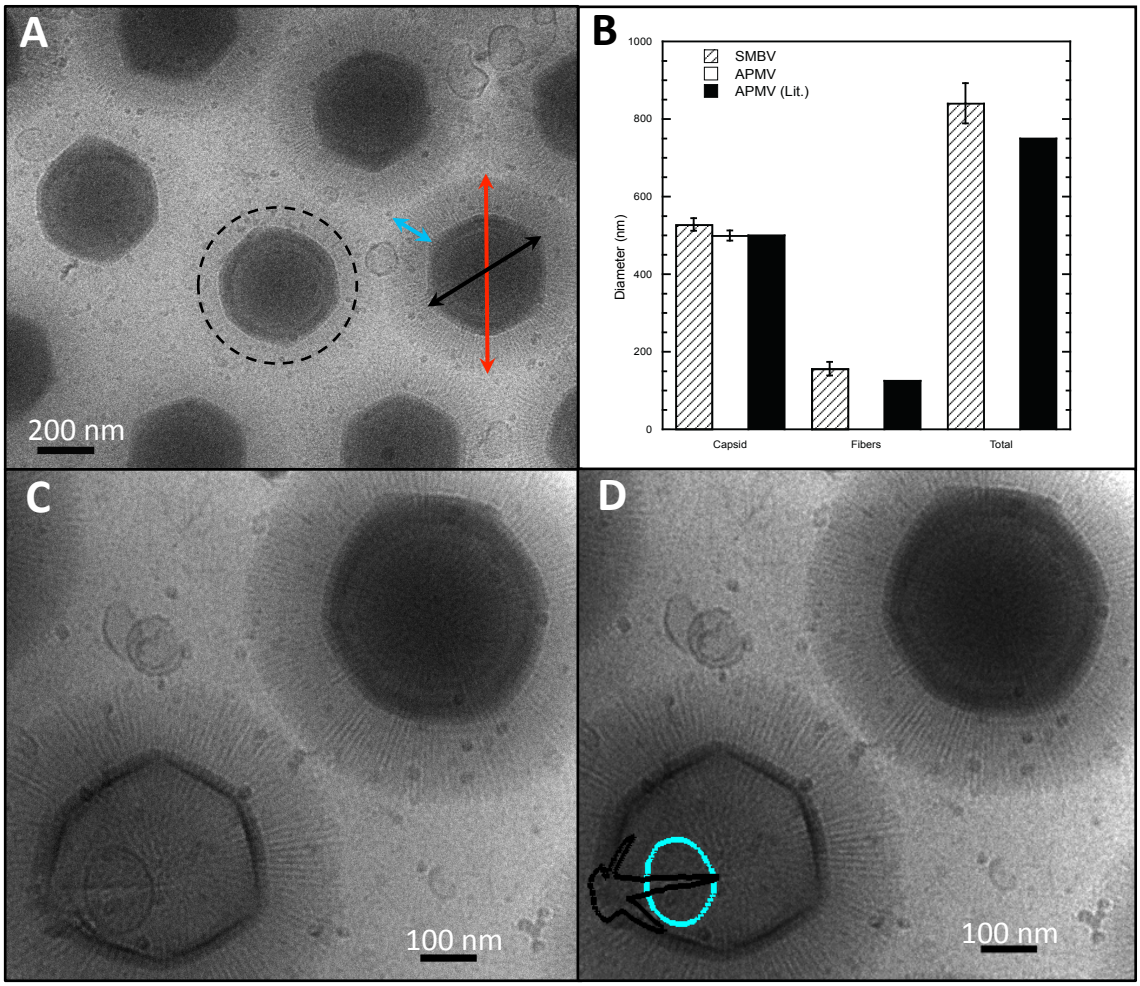

Figure 2.1 Cryo-Electron Microscopy Data From SMBV Particles. A) Representative micrograph depicting "fibered" and "fiberless" (circled) SMBV particles. Arrows provide an example of how the total particle diameter (red arrow), capsid diameter (black arrow), and the fiber length (cyan arrow) were measured for SMBV and APMV particles. B) Capsid diameter, fiber length, and total particle diameter of SMBV (Striped) and APMV (White) particles from this study, as well as APMV particles from Xiao, *et al.*, 2005 (*152*) (Black). C) Cryo-electron micrograph of "fibered" and "open/empty" SMBV particles. The star-shaped capsid opening (black) and the membrane sac that remains within "open/empty" particles (cyan) are highlighted in D.

RESULTS

Cryo-Electron Microscopy (Cryo-EM) Revealed the Size and Morphologies of Samba Virus Particles

SMBV particles, like those of all members of *Mimiviridae*, are very large, requiring a thick layer of vitreous ice (> 1 μm) to preserve the specimen in a near-native state for cryo-EM imaging. The thickness of the ice layer detracted from the contrast of SMBV cryo-EM images, especially while using a 200-keV TEM. With the use of an in-column Omega Energy Filter (JEOL 2200-FS) and a DE-20 direct detection device (Direct Electron, LP, San Diego, CA, USA), contrast in the cryo-electron micrographs was improved. SMBV particles were also imaged using a 300-keV TEM (JEOL 3200, data not shown), but these images displayed no appreciable difference in quality from the micrographs collected at 200 keV using the Omega Energy Filter. We were able to generate two-dimensional projection images of vitrified SMBV particles with sufficient contrast to accurately measure and describe several structural features of interest. The cryo-electron micrographs revealed external fibers, at least two capsid layers, and an internal genome-containing nucleocapsid within the SMBV virion (Figure 2.1).

Within the cryo-EM images, three distinct particle morphologies were visible, the most abundant of which were "fibered" SMBV particles (Figure 2.1A-B). These particles, comprising ~81% of the ~2800 particles imaged via single-particle cryo-EM, were surrounded by a layer of external fibers, which are thought to be important for host attachment. "Fiberless" particles represented the second most abundant particle morphology, at ~13.5%, (Figure 2.1A, indicated by a dashed circle). These do not contain external fibers. The ability of these particles to infect *A. castellanii* is currently unknown. In *Mimiviridae,* fibers are hypothesized to play a role in cell attachment and entry via phagocytosis (*109*), and the same may also be true for SMBV. However, a fiberless variant, "M4", was shown to enter and propagate inside cells (*64*). The least

abundant particle morphology, at ~5.5% of particles, were "open/empty" SMBV particles (Figure 2.1B). These particles contained neither the nucleocapsid nor the double-stranded DNA genome, and were visually represented in the cryo-electron micrographs as lighter particles, due to the absence of the electron-dense material within the capsid (Figure 2.1C). It was hypothesized that these particles reflect a post-genome ejection stage and have opened their capsids at a unique capsid vertex (Figure 2.1D, highlighted in black), reminiscent of the starfish vertex seen in mimivirus (*150-152, 176*). The open/empty particles appeared to have a residual membrane component, which remained associated with the capsid after genome release (Figure 2.1D, highlighted in cyan). A similar residual membrane can be seen in two-dimensional projections of open APMV particles (*151*).

Even with low contrast, the cryo-EM images provided an accurate determination of the native size of the SMBV capsid and external fibers. The initial characterization of SMBV utilized plastic-embedded thin sections of infected amoeba and reported a capsid diameter of 352 nm, a fiber length of 112 nm, and a total particle diameter of 574 nm (*95*). As mentioned previously, the sample preparation techniques used to generate thin sections of biological samples can lead to the generation of artifacts; in particular, the dehydration steps can lead to shrunken particles (*68, 179*). Since specimens in cryo-EM remain fully hydrated, we measured the diameter of the capsid and the total particle diameter of 500 SMBV particles to determine the size of the SMBV virion (Figure 2.1A, C). Averaging these measurements yielded a capsid diameter of ~527 nm (Figure 2.1A, black arrow) and a total particle diameter of ~834 nm (Figure 2.1A, red arrow), which is significantly larger than previously reported (*95*). The size discrepancy between the particles visualized by cryo-EM and by thin-section TEM is most likely due to dehydration-linked particle shrinkage during the thin-section preparation steps. We were

able to subtract the measured capsid diameter from the measured total particle diameter of each particle to estimate the "diameter" of the external fiber layer (assumed to be twice the fiber length). For the 500 SMBV particles measured in this study, the average fiber length (Figure 2.1A, cyan arrow) measured ~155 nm.

The structure of APMV, previously determined by cryo-EM (EMD-5039, (*151*)), demonstrated that APMV particles are quasi-icosahedral with one unique vertex housing the "starfish" structure used to release the nucleocapsid during genome release. Three-dimensional image reconstructions of APMV, imposing icosahedral symmetry, and/or 5-fold symmetry yielded maps clearly displayed the APMV structural features (*150, 151*). As SMBV is closely related to APMV (*95*), it was hypothesized that SMBV particles would share a similar quasi-icosahedral nature. Therefore, we attempted single-particle reconstructions of ~2800 SMBV particles using a random model computation (RMC) (*184*) and Auto3dem (*185*), as well as EMAN2 (*186*). SMBV particles displayed a high degree of structural heterogeneity, as evidenced by visual inspection (Figure 2.1 and Figure 2.2), failure to obtain consistent classes using the EMAN2 classification procedure (data not shown), and results from cryo-tomography (below, Figure 2.3). To eliminate the external fibers as a confounding factor for the three-dimensional reconstruction, we also attempted an RMC on fiberless SMBV particles that were present in the two-dimensional projection images. In total, we tried 100 RMCs for both the complete particle set and the subset of fiberless particles. All RMCs failed to produce a coherent icosahedral structure, suggesting that either SMBV is unlike APMV, and not quasi-icosahedral, or that we had a mixed population of icosahedral and non-icosahedral particles and were unable to distinguish between these particle types in our micrographs. If rigid, quasi-icosahedral SMBV particles are indeed present; the frequency was too low to detect them in this sample.

Figure 2.2

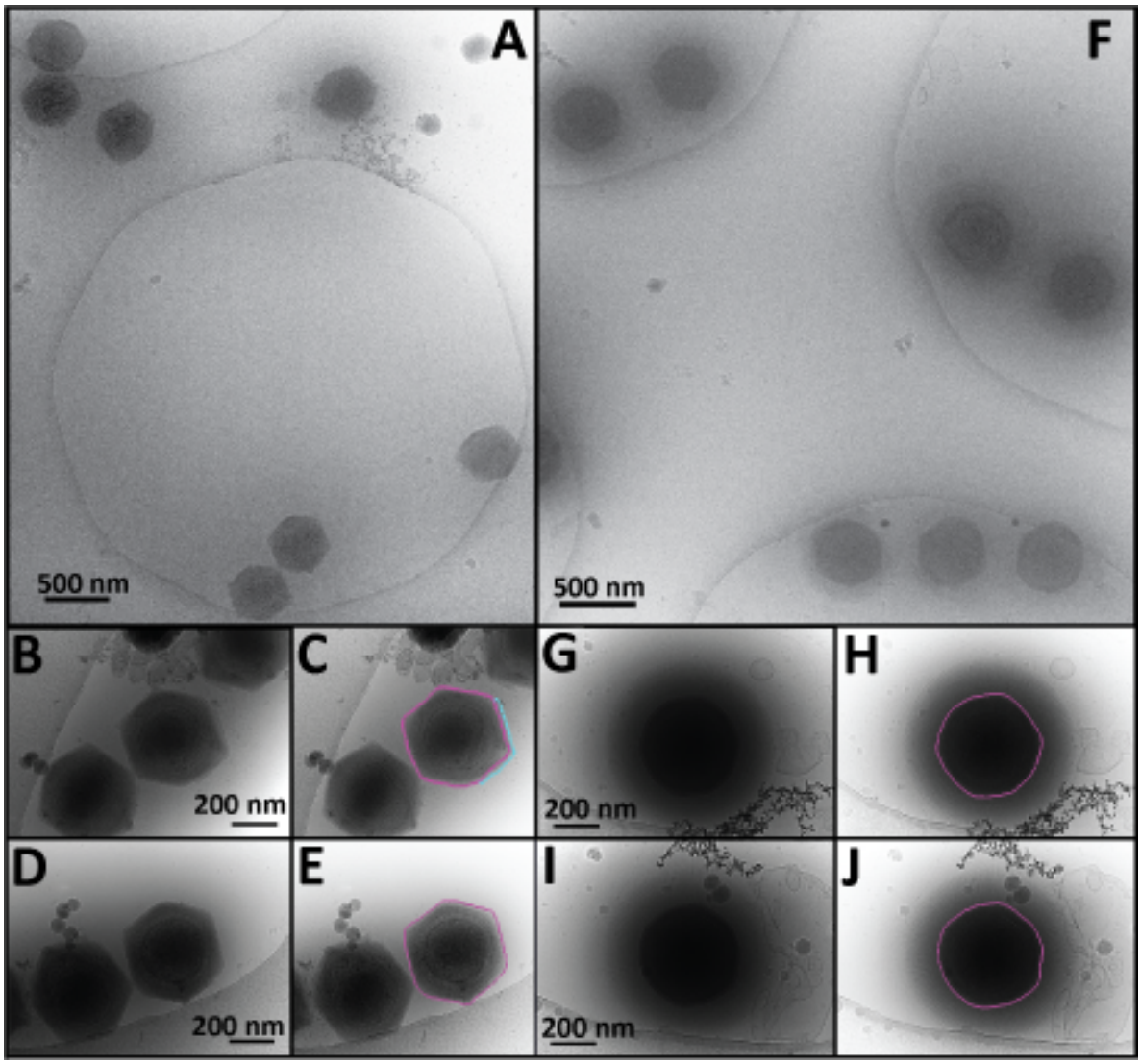

Figure 2.2 Comparison of APMV and SMBV via Cryo-EM Reveals that SMBV is a not a Rigid Quasi-Icosahedron Like APMV, and Displays a Larger Degree of Structural Variation Than APMV. A) Low magnification (4,000 X) micrograph of APMV particles. B and D) Higher magnification (20,000 X) micrographs of APMV particles with features highlighted in C and E, respectively. F) Low magnification (4,000 X) micrograph of SMBV particles. G and I) Higher magnification (20,000 X) micrographs of SMBV particles with features highlighted in H and J, respectively. For panels C, E, H, J: Outer capsid layers are highlighted in magenta. The presumed starfish seal complex in panel C is highlighted in cyan.

Figure 2.3

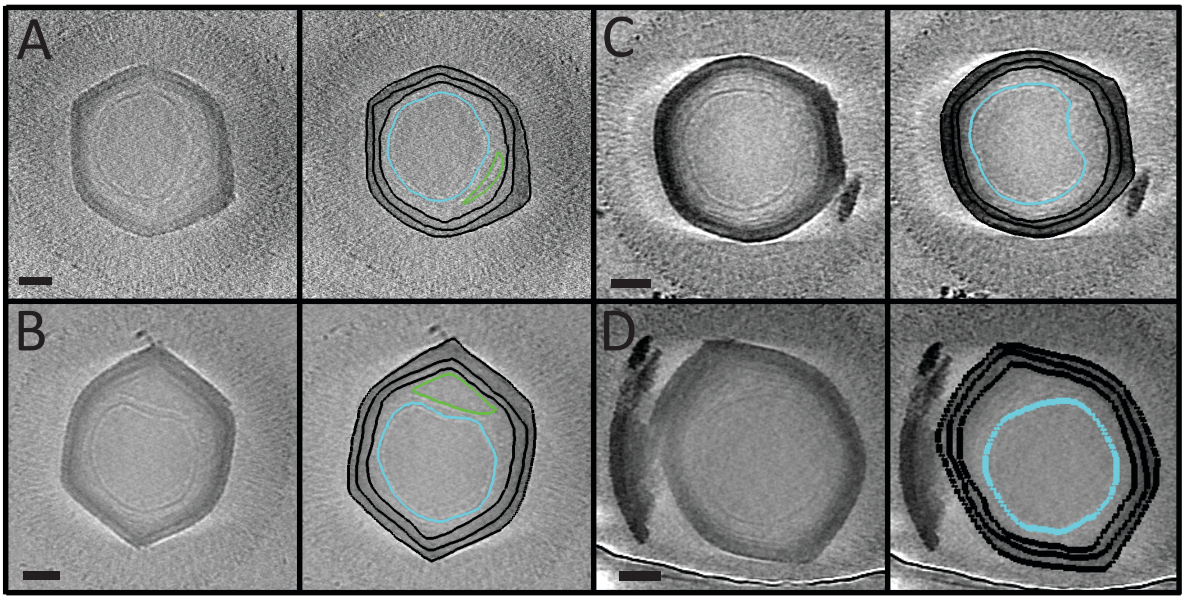

Figure 2.3 Cryo-Electron Tomograms of SMBV Particles. These micrographs depict two-dimensional projections of three-dimensional data from four representative SMBV tomograms. Projections represent 10 slices (14.7 nm thick for A & B and 6.9 nm thick for C & D) computationally combined using the Slicer functionality in IMOD (3dmod). Capsid layers (black), nucleocapsid (cyan), and membrane sac (green) within the SMBV virions are highlighted in the right-hand panels. Tilt series were acquired along a tilt range of ± 55° with tilt increments of 1-2°. Tomograms were generated using IMOD v4.7.15. Tomograms in A & B were collected at 4,000 X, and C & D were collected at 8,000 X nominal magnification. Scale bars represent 100 nm.

A Comparison of Mimivirus and Samba Virus Particles Through the use of Cryo-Electron Microscopy

Since SMBV did not display a rigid, quasi-icosahedral capsid structure as seen in APMV, we also analyzed cryo-electron micrographs of a fiberless variant of APMV (*64*) (Figure 2.2A–C). Since a plethora of structural information is available for APMV (*68, 150, 151, 175, 187*), we felt that using the same experimental setup to analyze the two viruses would provide a good control to compare the shape of APMV and SMBV capsids, and to confirm that the plasticity observed in SMBV is not a result of preparation techniques. Comparing SMBV and APMV particles in the same state (both fibered or both fiberless) would be ideal, however we did not have access to identical samples. We did not have a sample of fibered APMV, and the only process, to our knowledge, which is known to defiber giant virus particles (*151*) is treatment with proteinase K, lysozyme, and bromelain, which does not remove the SMBV fibers. This preliminary result suggests that the composition of SMBV fibers differs from that of other members of *Mimiviridae.*

An average of the measured capsid diameters of 274 APMV particles resulted in a capsid diameter of ~499 nm, which matches the previously reported value (*152*) (Figure 2.1B). A small percentage of both APMV and SMBV particles displayed a notch-like structure at a unique vertex within the capsid (Figure 2.2D). This feature has been reported previously in APMV (*152*), although its biological function is currently unknown. APMV particles within the cryo-EM images appeared to have a much higher degree of structural homogeneity than that seen in the SMBV particles (Figure 2.1 and Figure 2.2). APMV particles within the cryo-electron micrographs were clearly quasi-icosahedral with rigid facets, consistent with the published

structure (*151*). SMBV particles, on the other hand, exhibited a high degree of structural plasticity (Figure 2.2).

Three-Dimensional Structural Information of the Entire Samba Virus Virion was Obtained Through the use of Cryo-Electron Tomography (Cryo-ET)

With the large degree of heterogeneity displayed in the SMBV particles (see above), we were unable to generate a three-dimensional structure of the SMBV virion through the use of single particle cryo-electron microscopic analysis. Cryo-electron tomography (cryo-ET) eliminates the need to average many particles, allowing us to circumvent the heterogeneity of the SMBV particles. With a total particle diameter of ~834 nm SMBV is, to our knowledge, the largest specimen successfully imaged using cryo-ET without the use of focused ion beam (FIB)-milling(*188, 189*), freeze fracturing (*176*), cryo-sectioning (*190*), or other techniques which are used to reduce sample thickness (*159, 191*).

As the most abundant particle morphology, and with the fibers thought to be important for attachment, we decided to focus our cryo-ET efforts on fibered particles. We generated 20 tomograms displaying 34 fibered SMBV particles. Four representative volumes are displayed in Figure 2.3 (Supplemental Video 1). A representative tomogram is accessible through the Electron Microscopy Data Bank (EMDB) with the following accession number: EMD-8599. These tomograms displayed the structural features of the SMBV virion in greater detail than the single particle cryo-electron micrographs (Figure 1).

The tomograms provided enough detail to visualize several layers within the SMBV capsid, and provided further confirmation of the heterogeneity observed in our 2D projection images. In APMV, the capsid is hypothesized to consist of two layers of protein surrounding a layer of lipid, resulting in at least three visible layers within the capsid (*152, 192*). Like in

APMV, the tomograms depicted at least three distinct layers within the SMBV capsid (Figure 2.2, highlighted in black in the right-hand panels), although the exact biochemical composition of these three layers is currently unknown. The average thickness of the SMBV capsid, measured at 10 locations around the capsid for 10 SMBV tomograms, was at most 43.3 ± 6.4 nm, with at least a 20.6 ± 3.6-nm separation between the outermost layers and at least 22.6 ± 3.9-nm separation between the two internal layers. Previous work has shown that the thickness of viral layers does not change according to defocus values ranging 1–8 μm (*193*). In this work, we used higher underfocus objective lens settings. Therefore, we present the inter-layer spacing and the thickness of the layers as lower and upper thresholds, respectively. Capsid thickness within SMBV particles appeared to have a high degree of variation in both the thickness of the complete capsid (even within individual particles) and variation in the separation between the various capsid layers, and likely explains why we were unable to obtain a three-dimensional reconstruction from single particle analysis. As a result of this heterogeneity, we were also unable to perform meaningful sub-tomogram averaging.

Cryo-ET also provided a more detailed view of the SMBV fibers than the two-dimensional cryo-EM projection images. The external fibers appeared to be evenly dispersed throughout the SMBV virion, but they did not appear to have a uniform, rigid structure. In an attempt to determine if SMBV fibers have a helical nature, we also boxed 163 fibers from two-dimensional projections of five SMBV particles. Power spectra of these boxed fibers were generated using SPIDER as a part of the IHRSR++ workflow (*194*), and failed to produce a recognizable helical diffraction pattern, suggesting that the fibers are either not helical in nature, or were too heterogeneous to produce a regular pattern. Results from the tomograms show that the fibers are rather flexible and it proved difficult to extract individual fibers as 3D volumes

since the fibers were very closely packed. Therefore, performing sub-tomogram averaging on extracted density from fibers was not possible with our current data set.

Like APMV, the SMBV genome is contained within an internal nucleocapsid. Sitting in the center of the virion, and containing relatively electron dense DNA, the SMBV nucleocapsid was visible within the two dimensional cryo-electron micrographs (Figure 2.1). However, the SMBV nucleocapsids were much easier to resolve in the cryo-electron tomograms (Figure 2.3, highlighted in cyan in the right-hand panels). Within the 34 SMBV particles analyzed via cryo-ET, 31 of the particles displayed clear nucleocapsid boundaries. The remainder displayed density that resembled the nucleocapsid but was not clearly discernable owing to the low contrast within the reconstructions. These nucleocapsids had an average diameter of 289.6 ± 27.8 nm, although this number is likely skewed, as some SMBV nucleocapsids were not spherical. In nine of the 31 SMBV particles with visible nucleocapsids (which corresponds to 29% of the particles with clear nucleocapsids), the nucleocapsid was deformed by ~15 nm, appearing to pull away from one capsid vertex. Where the nucleocapsid was pulled away the capsid, it resided ~75 nm away from the innermost capsid layer, as opposed to the remainder of the nucleocapsid, which was ~40 nm away from the capsid, on average. This phenomenon was also observed in APMV, with sufficient frequency to appear in the single particle reconstruction of the virus (*151*). In the APMV three-dimensional reconstruction, the capsid vertex that the nucleocapsid is pulling away from houses the starfish structure. The presence of similar asymmetry in the SMBV nucleocapsid may provide further evidence that the SMBV virion also contains this so-called starfish seal at a unique vertex. The absence of nucleocapsid asymmetry in some SMBV particles was likely a result of particle orientation and is consistent with the missing wedge effect inherent in cryo-ET (*165, 191*). This effect limits the region of three-dimensional information available in our

tomograms. In addition, three SMBV particles clearly exhibited the presence of an extra membrane sac within the virion (Figure 2.3A-B, highlighted in green in the right-hand panels), which was also seen in two-dimensional projections of empty capsids (Figure 2.1D, highlighted in cyan). The biochemical composition of this sac, and its biological function, is currently unknown. This extra membrane sac was observed in an empty APMV particle, yet was not resolved within the three-dimensional reconstruction (*151*), likely owing to the 5-fold averaging employed in that study.

A Comparison of Samba Virus and Mimivirus Particles via Scanning Electron Microscopy (SEM) Revealed Differences in Capsid Regularity and Potential Viral Ultrastructure

To obtain further structural information about the SMBV capsid, and to corroborate our observations from both cryo-EM and cryo-ET, we analyzed SMBV particles via scanning electron microscopy (Figure 2.4). To avoid dehydration of the particles, and the accompanied structural artifacts, the SMBV particles were dried using a critical point dryer prior to the sputter coating process. Low magnification SEM images revealed material stretching between the SMBV particles (Figure 2.4A). The composition of this material is currently unknown, but it appeared to form fibrous strings between SMBV particles. This material was consistently present, even when SEM samples of SMBV were prepared using various procedures (data not shown). It is unknown whether this material plays any role in SMBV biology.

Figure 2.4

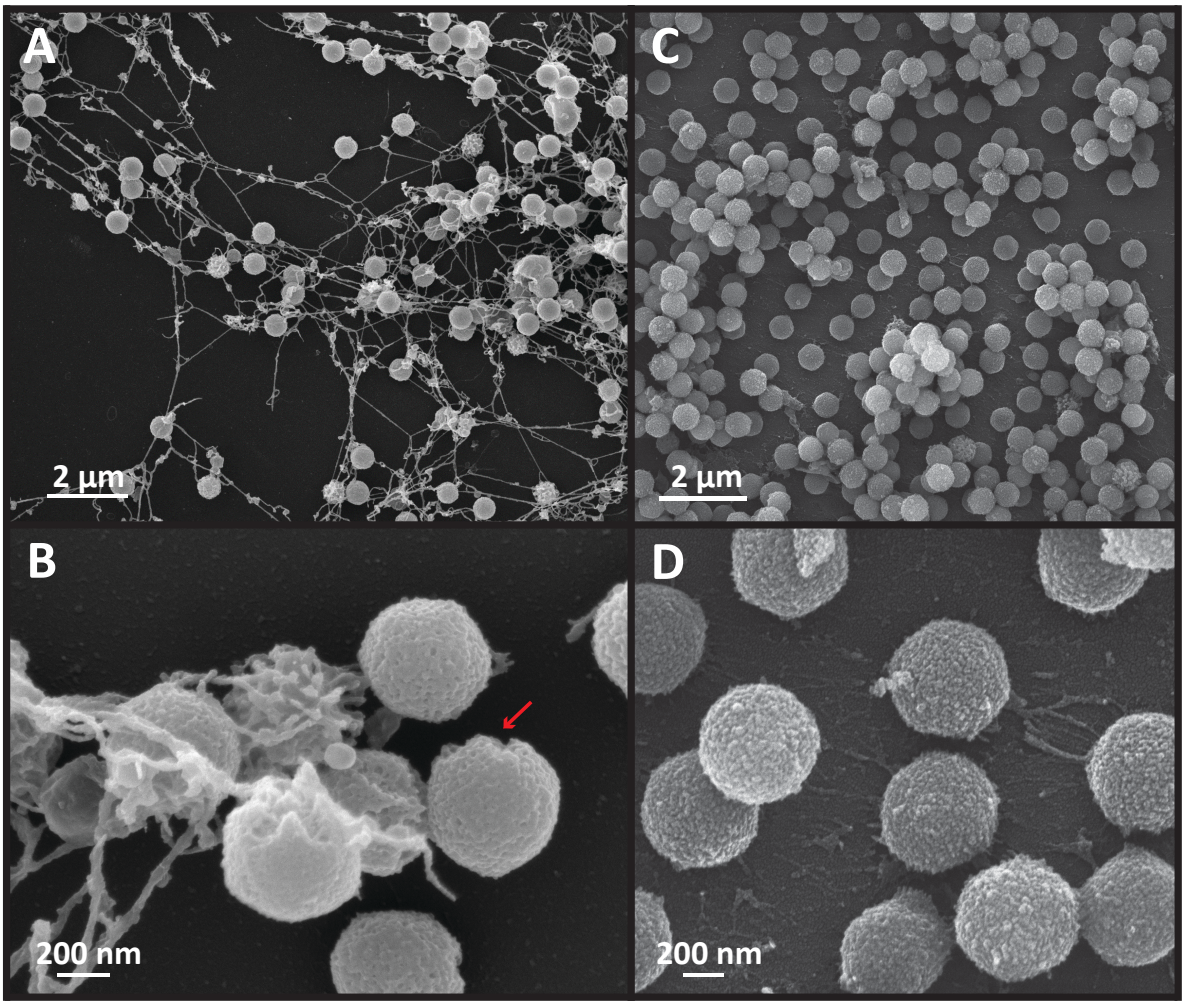

Figure 2.4 Scanning Electron Micrographs of SMBV and APMV Particles. A) Low magnification (7,000 X) field of view of SMBV particles. B) Higher magnification (50,000 X) image of SMBV particles. The red arrow points to a presumably fiberless region at a unique vertex of an SMBV particle, potentially revealing the location of the starfish seal. C) Low magnification (10,000 X) micrograph of a fiberless APMV variant (*64*). D) Higher magnification (50,000 X) micrograph of APMV particles.

The scanning electron micrographs also gave us some idea of the surface of the SMBV particles. Within the low magnification images, most of the SMBV particles appeared to be smooth, but a few of the particles appeared to be surrounded by a layer of "spikes" (Figure 2.4A-B). The "spikes" on these particles were likely external fibers that had clumped together during the critical point drying or the sputter coating processes, although this is currently impossible to determine, as we are unable to remove the SMBV fibers. Higher magnification micrographs of the SMBV particles (Figure 2.4B) provided greater detail of the surface of the virus and the fibrous strings. The surface of the SMBV particles did not appear to be regular when compared to that of APMV. Previous work on APMV using atomic force microscopy demonstrated a lack of fibers surrounding the starfish (*187*). It appears that this may be consistent in SMBV based on surface variation at unique vertices seen in SEM data (arrow in Figure 2.4B highlights one such vertex).

APMV scanning electron micrographs demonstrated some connective material (Figure 2.4C-D) but not nearly as much as in the SMBV sample. Higher magnification micrographs of APMV viral particles provided greater detail about the surface of the APMV and SMBV particles. While the APMV particles appeared to be regular in shape and had a uniform surface, SMBV particles appeared to have variable sizes and surface uniformities.

Fluorescence Light Microscopy Revealed Biomolecular Composition and Ultrastructural Lattice Formation of Samba Virus and Mimivirus Particles

Although techniques such as cryo-EM and cryo-ET possess near-atomic resolution in determining structures and visualizing surfaces, one can only speculate as to the exact biomolecular composition of the various virion components (fibers, capsid, etc.). Previous work has been successful in staining giant viruses using fluorescent dyes for flow cytometry (*195*).

Here, we took advantage of fluorescent dyes in microscopy experiments, which allowed for the differentiation of biomolecules and provided additional details of capsid architecture that we were unable to ascertain by cryo-EM alone. To determine the positions of the various components within *Mimivirus* virions, and to perform another comparison between APMV and SMBV, we dyed the viral particles with FITC (which is amine reactive and dyes proteins) and DAPI (selective for DNA), and then visualized the dye localization through the use of fluorescence light microscopy.

Although we were unable to visualize the viral particles in as great of structural detail as we were able to with cryo-EM, cryo-ET, and SEM, through the use of light microscopy, we were able to view comparative similarities and differences between SMBV and APMV particles. One of the most striking results of the bright field microscopy was the difference in organization between the two viruses. SMBV particles appeared to self-organize into large lattices, some of which were tens of microns in size (Figure 2.5A). This observation highlights an additional benefit to using fluorescence microscopy to visualize SMBV. In our cryo-electron experiments, we were unable to detect the presence of higher-order aggregates in the vitrified specimens as thicker areas of ice did not allow sufficient contrast in resulting micrographs, and thus were avoided during imaging. These lattices are reminiscent of the hexagonal lattices seen within bacterial cells during bacteriophage P22 infection (*196*). This observation contrasts sharply with APMV (Figure 2.5B), which appears to form loose aggregates, lacking the rigid organization that was seen in SMBV (Figure 5A). This difference in lattice organization may be a property of the viruses themselves, but it may also be due to the lack of fibers in the APMV samples. As mentioned previously, the *Mimivirus* fibers are thought to play a role in attachment (*109*), so it is possible that the fibers are responsible for the organization of SMBV particles and the lack of

organization within the APMV sample. The lack of organization, the abundant aggregation, and the smaller relative size of APMV particles combined to cause difficulty while imaging these particles.

Figure 2.5

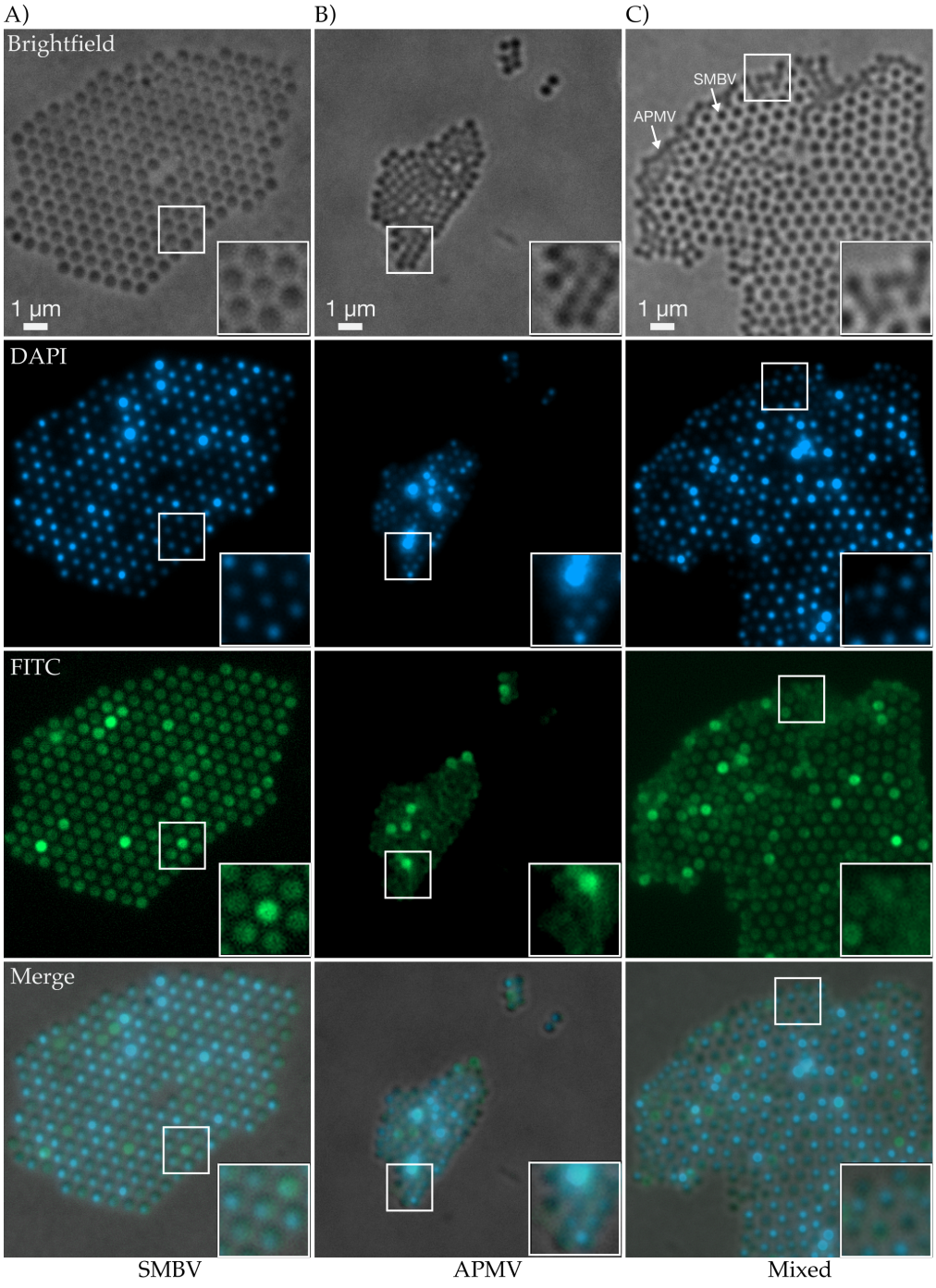

Figure 2.5 Fluorescence Light Microscopy of SMBV and APMV Particles. A) SMBV imaged via transmitted light, DAPI stain, and FITC stain which demonstrated defined particles and higher-order organizational characteristics B) APMV imaged via transmitted light, DAPI DNA stain, and FITC protein stain which highlighted a lack of particle definition and loose aggregation C) A mixed population of SMBV and APMV imaged with transmitted light, DAPI, and FITC stains distinctly showing SMBV lattice interruption from APMV particle association.

As noted previously, the strength of labeling specific biomolecules, and detailing relative location within particles, is one of the main attributes of fluorescent light microscopy. DAPI DNA staining demonstrated similar attributes between particles from both viruses. For both APMV and SMBV, some particles displayed dense, brightly fluorescent DAPI staining while the other particles appeared to be more punctate (Figure 2.5A–C). Enlarged views of some SMBV and APMV particles (Figure 2.5A–C, insets) demonstrated the asymmetrically localized DAPI fluorescence within the viral particles. The DAPI fluorescence signal appeared to be smaller and contained within the bright field and FITC signals (see below). This observation is what one would expect from a virus, with the nucleic acid genome contained within a proteinaceous capsid, and confirms that fluorescence microscopy can be used to localize virion components within giant viruses. The DAPI signal also appears to be asymmetrically localized within some SMBV capsids. This observation matches the nucleocapsid asymmetry observed in the two-dimensional projections of SMBV particles from both cryo-EM and cryo-ET.

While the DAPI fluorescence for APMV and SMBV particles appeared similar, the two viruses demonstrated stark differences when visualized for FITC fluorescence, which is amine reactive and binds proteins. The SMBV FITC fluorescence supported the bright field observation of conjoined, self-organized particles (Figure 2.5A). Also, across some individual SMBV particles, the signal was particulate, demonstrating small foci of brighter fluorescence (Figure 2.5A, inset). Again, due to the resolution limitations of fluorescent light microscopy, it is difficult to determine the true significance of this punctate patterning of SMBV particles without further experimentation and investigation. A heterogeneously stained population is consistent with the heterogeneity observed using cryo-EM and cryo-ET as described above. APMV particles, on the other hand, lacked any detailed features under FITC fluorescence. While some

APMV particles appeared to be more fluorescent than others, many of the particles lacked the clearly defined protein boundaries present in the SMBV particles, and these particles lacked the stippling feature of SMBV.

For a truly direct comparison of the APMV and SMBV samples, we combined the two viruses prior to addition of the fluorescent dyes. This mixture directly demonstrated the differences between particles within the APMV and SMBV samples, and allowed us to visualize the interaction between the two viruses. Bright field microscopy showed a mixed lattice-aggregate of SMBV and APMV particles. The APMV particles were interspersed within the SMBV lattice (Figure 2.5C), and appeared to perturb SMBV particle lattices. These observations were further supported by the FITC fluorescence. The protein dye demonstrated APMV particles, which lacked a defined FITC boundary, within the larger SMBV lattices. This interspersal of APMV particles within the SMBV lattice suggests that SMBV, and potentially all *Mimiviruses*, are able to interact with other virus particles within aggregated lattices. We speculate that the giant virus-associated virophages (e.g., Sputnik, Rio Negro virus) may also be able to interact in these lattices during *Mimivirus* infections.

DISCUSSION

In summation, the cross-platform techniques as described in this paper highlight similarities and differences between SMBV and APMV. SMBV has a larger capsid diameter (~527 nm), fiber length (~155 nm), and total particle diameter (~834 nm) than APMV (~500 nm, ~125 nm, ~750 nm, respectively), making SMBV the largest member of *Mimivirus* described to date. The major difference between APMV and SMBV appears to be the global structure of the viral capsid. APMV particles appear to be quasi-icosahedral, with rigid sides and a unique vertex that houses the starfish complex, consistent with previously published reports. SMBV, on the other hand, does not appear to share the same degree of rigidity and a quasi-icosahedral architecture with rigid facets is less obvious. Instead, SMBV exhibits a much higher degree of structural variance. For example, in the cryo-EM images, APMV particles appear to be more regular in shape and have fewer structural variations than the SMBV particles. In the SEM images, APMV particles appear to have a smoother capsid surface and fewer structural irregularities. SMBV particles form self-organized lattices within the fluorescence micrographs whereas APMV particles tend to randomly aggregate. These differences in ultrastructure are likely caused by the presence of the external fibers in SMBV and their absence in APMV. In cryo-EM, cryo-ET, and fluorescence micrographs SMBV particles show an asymmetrically-localized nucleocapsid, which varies in structure from particle to particle. Future work to make use of advanced light microscopy techniques (such as super-resolution microscopy) will help to elucidate if these are indeed common features among giant viruses and will provide additional insight that cannot be gained from electron microscopy alone.

There are over 50 *Mimiviruses* isolated and characterized to date. Recently, a pan-genome analysis compared SMBV, APMV, and others (*85*). Key results reveal that the genome

of SMBV is most similar to APMV, and retains high similarity with other *Mimiviruses* such as Oyster virus (OYTV) and Amazonian virus (AMAV). This pan-genome analysis of Brazilian *Mimivirus* group A showed that a total of 58 clusters consisting of 179 paralogous proteins were identified in SMBV, which is similar to APMV, and reciprocal best-hit analysis identified 917 orthologous proteins shared between these viruses. The four predicted capsid proteins in SMBV have 98–100% identity to those known in APMV. Previous predictions indicate that the APMV major capsid protein "L425" is likely to have a double jelly-roll structure (*151*). It is tempting to predict that the SMBV major capsid protein will have a similar structure. However, making structural predictions regarding the capsid protein based solely on the genetic material is difficult at best. For example, introns in the mimivirus capsid protein gene have been shown to complicate genomic predictions, and mass spectrometry and recombinant expression systems were required to fully characterize this gene product (*197*). The SMBV capsid protein gene has up to three introns (GenBank AHJ40114.2). We can conclude that there are sufficient differences in the global architecture of SMBV and APMV. Therefore, it follows logically that there will likely be some differences in the structural protein building blocks that form the native virions. Further detailed biochemical and structural experiments of the SMBV capsid proteins are needed to dissect these differences at the molecular level.

ACKNOWLEDGMENTS

We would like to thank Xudong Fan and Carol Flegler at the Michigan State University Center for Advanced Microscopy and Daniel Ducat at the Michigan State University Department of Biochemistry and Molecular Biology for their guidance and support for the TEM, SEM, and fluorescence light microscopy, respectively. We would like to thank Kaillathe "Pappan" Padmanabhan for his assistance in the setup and maintenance of our computational resources. We would like to thank Kit Pogliano for suggesting the idea of trying fluorescence microscopy. We would like to thank Direct Electron, especially Michael Spilman, for their support and assistance with the DE-20 camera and image processing. We would like to thank Centro de Microscopia da Universidade Federal de Minas Gerais. Thank you to David Gene Morgan and the electron microscopy center at Indiana University, Bloomington, for access to their 300 keV scope. This work was supported by the American Association for the Advancement of Science Marion Milligan Mason Award for Women in the Chemical Sciences to Kristin N. Parent., by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) to Juliana R. Cortines, by CNPq, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) to Jônatas S. Abrahão. Jônatas S. Abrahão is a CNPq/Mediterranee Infection researcher.

SUPPLEMENTARY MATERIALS

Supplementary Video 1: Z-slices of a representative SMBV tomogram (central section depicted in Figure 2.3B).

CHAPTER 3

BOILING ACID MIMICS INTRACELLULAR GIANT VIRUS GENOME RELEASE

This work was originally submitted as a preprint to bioRxiv and is currently in revision at Cell.

Schrad, J.R., Abrahão, J.S., Cortines, J.R., Parent, K.N. 2019. Boiling Acid Mimics Intracellular Giant Virus Genome Release. *Cell* (in revision, bioRxiv doi: https://doi.org/10.1101/777854).

Minor edits have been made to this manuscript to conform to dissertation requirements.

SUMMARY

Since their discovery, giant viruses have expanded our understanding of the principles of virology. Due to their gargantuan size and complexity, little is known about the life cycles of these viruses. To answer outstanding questions regarding giant virus infection mechanisms, we set out to determine biomolecular conditions that promote giant virus genome release. We generated four metastable infection intermediates in Samba virus (lineage A *Mimiviridae*) as visualized by cryo-EM, cryo-ET, and SEM. Each of these four intermediates reflects a stage that occurs *in vivo*. We show that these genome release stages are conserved in other, diverse giant viruses. Finally, we identified proteins that are released from Samba and newly discovered Tupanvirus through differential mass spectrometry. Our work revealed the molecular forces that trigger infection are conserved amongst disparate giant viruses. This study is also the first to identify specific proteins released during the initial stages of giant virus infection.

INTRODUCTION

A hallmark of newly discovered giant viruses (GV) is their incredibly complex biology, including gargantuan capsid sizes and large genomes. The sheer size and complexity of these viruses, especially the inclusion of "junk" DNA in the form of introns (*197, 198*), challenges the canonical view of viruses as small, streamlined, and efficient killing machines. For example, most GV are larger than 300 nm and many have genomes exceeding 1MB, containing an estimated 1000+ open reading frames (see Table 1 in (*100*)). By contrast, some of the smallest viruses include the porcine circoviris (17 nm capsid, ~2000 base genome, four proteins, (*66*)) and the human rhinovirus (~7200 base genome, 30 nm capsid, 11 proteins, (*67*)). ~69% of known viruses encode for less than 10 proteins (*102*), highlighting the complexity of GV and the true extent of our lack of knowledge concerning this new class of viruses.

GV have been isolated from a wide variety of hosts, including amoeba (*83*), animals (*4, 107, 111, 199*), as well as human and murine cells (*108, 200*). However, amoebas also infect these creatures, casting doubt on the true viral reservoir. Although GV have been associated with human diseases such as respiratory diseases (*107, 113, 114, 157*), inflammatory conditions (*116, 117*), and cancers (*112*), no direct link between GV and human disease has yet been established. Despite an unusually broad host range and pathogenicity, little information is available on how GV access their hosts. Host cell infection usually occurs via phagocytosis (*82, 108*). Once phagocytosed, a unique capsid vertex opens which promotes nucleocapsid release and fusion with the phagosomal membrane, ultimately releasing the genome into the host cytoplasm. A pseudo-organelle, called a viral factory, is then formed (*178*) and host replication factors are hijacked. The endpoint of GV infections is host cell death and release of new GV progeny into the environment.

GV are ubiquitous (*4, 83*) and maintain infectivity in harsh environments such as alkaline lakes (*3*), frozen permafrost (*63*), 3 km deep in the ocean (*3*) and dry valleys in Antarctica (*4, 105*). GV have retained infectivity following exposure to harsh chemicals (*201*), extreme pH and salinity (*3*), extreme temperatures (*4, 63*), and are able to persist on hospital equipment (*201, 202*). To survive such extremes, GV have developed incredible capsid stability. Some giant viral capsids can retain infectivity for 30,000 years in permafrost (*62, 63*).

Although capsid stability is beneficial for a virus to persist in harsh environments, it also creates a thermodynamic barrier that must be overcome once a suitable host cell is encountered. Traversing an energy barrier to promote infection and genome transfer into a host cell is not a problem unique to GV; all known viruses must do this to propagate. Strategies and structures used for genome translocation are conserved across viral families. Amongst the tailed dsDNA bacteriophages (*Caudovirales*), tail complexes interact with host receptor proteins to trigger conformational changes in the virion, leading to genome release (*126*). Similarly, many classes of eukaryotic viruses have conserved genome release mechanisms. Most enveloped viruses, including HIV, influenza, Zika virus, and herpesvirus, utilize one of three structurally conserved membrane fusion protein varieties (*132*). Non-enveloped viruses, such as rhinovirus, poliovirus, and adenovirus, utilize conserved capsid structures to interact with host receptors to trigger genome uncoating (*203*).

Morphologically, GV virions are either icosahedral, as exemplified by Acanthamoeba polyphaga mimivirus (*79*), or non-icosahedral typified by Mollivirus and Pithovirus (*62, 63*). Similar to their smaller cousins, GV also share conserved capsid structures that are used during infection. In many GV, the unique capsid vertex provides a gateway for the infection process, but they also provide a mechanism to prevent premature loss of their precious cargo. GV have

developed at least two distinct vertex structures to seal the unique vertex until the time is right for infection: “corks” and “starfish”. Non-icosahedral GV tend to utilize one or more cork-like structures to seal their unique capsid locations (*63, 86, 149*). These complexes are located flush with the capsid surface. A newly-discovered class of non-icosahedral GV, consisting of members such as Pandoravirus (*63*) and orpheovirus (*204*), contain an ostiole-like structure, distinct from the cork-like structure.

Mimivirus-like icosahedral GV utilize an external proteinaceous seal complex that resembles a five-pointed starfish (*150, 151*). These complexes sit at the outermost layer of the capsid at a unique five-fold vertex (called the stargate vertex due to its symmetry and appearance) and prevent it from opening (*151*). Traditionally, both the unique capsid vertex and the external seal complex have been packaged together and called either the “stargate” or the “starfish”. We will refer to the unique capsid vertex as the stargate and the seal complex as the starfish. Non-mimivirus-like icosahedral GV such as PBCV-1 (*97*), Faustovirus (*98*), and Pacmanvirus (*205*) do not utilize stargate vertices and have evolved alternative genome release strategies.

Starfish structures are found in diverse GV such as mimivirus (*150, 151*), Samba virus (SMBV, (*95, 96*)), and the newly discovered *Tupanviruses* (*3, 206*), and are more common than the cork-like seals amongst GV. Yet, relatively little is known about the mechanism governing the stargate. The molecular forces and biochemical trigger(s), such as receptor proteins or phagosomal transitions that facilitate stargate opening are unknown. Additionally, the ultimate fate of the starfish remains a mystery; is the complex removed from the capsid *en masse*, or does the complex simply unzip?

The general steps and macroscopic, gross morphological changes that accompany GV infection have been visualized via thin section transmission electron microscopy (TEM) of infected cells (*82, 206*). Following phagocytosis the stargate vertex begins to open between 1-3 hours post infection (*206*), yet, little is known about the specific proteins and biomechanical forces that mediate this process. This knowledge gap is largely due to two factors, the complexity of GV virions and the lack of a robust model system for detailed biochemical and/or biophysical studies. Here, we have created the first *in vitro* model system for studying the choreography that governs GV genome release using SMBV, a member of *Mimiviridae* lineage A (*95*). We were able to trap infection intermediates, identify specific proteins released during the initial stage of stargate opening, and test the efficacy of this technique on other icosahedral GV including a mimivirus variant, M4 (*64*), Tupanvirus soda lake (TV, (*3*)), and Antarctica virus (*4*). Additionally, our model reveals that members of *Mimiviridae* lineage A unzip their starfish complexes to initiate infection.

RESULTS AND DISCUSSION

Samba Virus is Resistant to the Vast Majority of Chemical Treatments

To probe the molecular forces that play a role in SMBV starfish complex stability, we exposed SMBV to treatments known to affect morphology and infectivity in other viruses (Table 3.1). The effect of each treatment on particle stability was assessed via cryo-EM. Treatments included the denaturants urea (up to 9 M) and guanidinium hydrochloride (up to 6 M), the detergent Triton X-100, organic solvents such as chloroform and DMSO, as well as enzymes including DNase I, bromelain, proteinase K, and lysozyme. Both urea and guanidinium hydrochloride denature proteins and have historically been used to disrupt bacteriophage capsids (*194, 207-210*). Triton X-100 is a detergent that we hypothesized could disrupt the two membranes inside of the GV capsid, the nucleocapsid and the extra membrane sac, if it could permeate the capsid. Additionally, chloroform and DMSO are organic solvents that disrupt lipid membranes and have been shown to disrupt viruses with internal lipid membranes (*207, 211-214*). The combination of bromelain, proteinase K, and lysozyme is the cocktail used to defiber mimivirus particles (*187*).

None of these treatments resulted in disruption of the SMBV virion, over the baseline of ~5% spontaneously open SMBV particles as observed under native conditions (*96*). Two treatments did lead to significantly increased disruption of the stargate vertex: low pH and high temperature (see following sections).

Table 3.1

Condition	*Concentration(s)*	*% Open SMBV*
Urea	9M	2.94
Guanidinium Hydrochloride	3M, 6M	2.90
DMSO	1% (v/v)	2.00
Triton X-100	1% (v/v)	0.00
Bromelain, Proteinase K, Lysozyme	14, 1, 10 mg/mL	0.00
Chloroform	20% (v/v)	4.17
DNase I	2 mg/mL	2.33

Table 3.1 Conditions That SMBV Particles Resist. Treatment conditions that did not produce a marked increase in the percentage of open SMBV particles.

Electrostatic Interactions are Critical for Samba Virus Starfish Stability

We hypothesized that pH changes occurring during and after phagocytosis may trigger SMBV stargate opening. Therefore, we dialyzed SMBV particles against different sodium phosphate buffer solutions, ranging in pH from 2-12 (Figure 3.1A). Particles were visualized via cryo-EM (Figure 3.2E) and the percent of open particles (POP) was calculated. At and above pH 4, there was no appreciable change in the POP, compared to native (pH 7.4) levels (Figure 3.2A-D). However, at and below pH 3, ~60% of the SMBV capsids had opened. While the conditions that produced an increase in SMBV POP ($pH \leq 3$) are more acidic than the environment predicted within the amoebal phagosome (*215-217*), they are similar. Thus, it demonstrates that our *in vitro* results reflect a relevant stage of the GV infection mechanism.

Figure 3.1

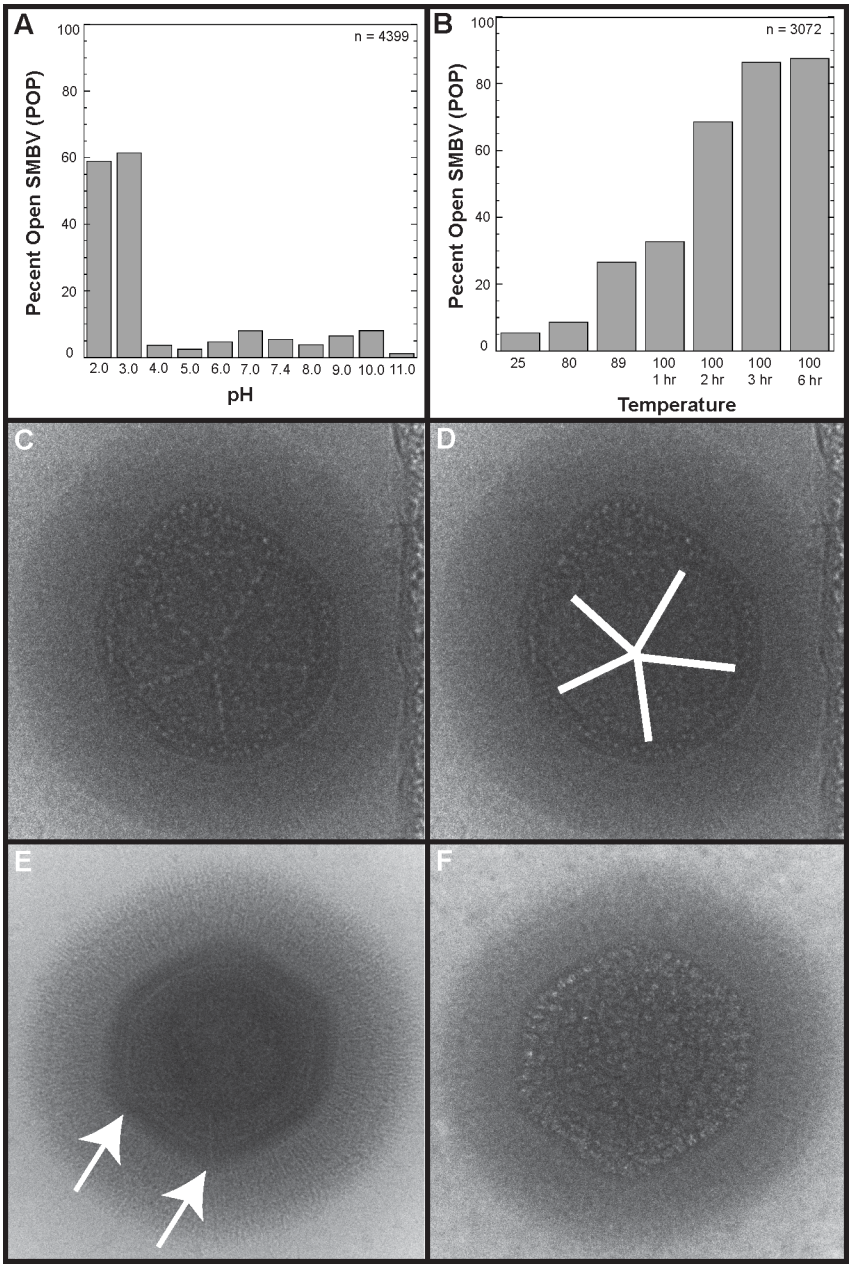

Figure 3.1 Low pH and High Temperature Triggered an Increase in SMBV POP and Changed the Star-Shaped Radiation Damage Pattern. A) The percentage of open SMBV particles (POP) following treatment at various pH (see Figure 3.3 and Table 3.1). B) The POP of SMBV particles incubated at elevated temperatures. C) "Bubblegram" image of a native SMBV particle with a clear star-shaped radiation damage pattern (highlighted in white in D, see Movie S2). E) First exposure in a bubblegram series of a pH 2-treated SMBV particle. The cracked stargate vertex lies in a top-down view. Arrows highlight the slight cracks in the SMBV capsid. F) Final exposure of the bubblegram series begun in E. Note the absence of the star-shaped radiation damage pattern following starfish disruption.

Unlike spontaneously opened GV capsids (*96, 151, 176*), these SMBV capsids were not fully open. Instead, the particles had small, noticeable cracks at one capsid vertex that assumed a star-shaped pattern. The opening of the stargate vertex at low pH is irreversible: SMBV particles returned to neutral pH still displayed star-shaped cracks in their capsids (data not shown). In some particles the extra membrane sac was caught in the process of leaving the capsid through the newly opened vertex (Figure 3.2E). In other particles, the sac is not visible, suggesting that it had escaped prior to imaging. Release of the sac, also referred to as the viral seed, has been hypothesized in other GV. The viral seed is thought to contain proteins responsible for the formation of the GV viral factory (*177, 178, 206*). To our knowledge, this is the first study to demonstrate release of the viral seed and to identify some of the proteins that may be released with this complex (below).

Figure 3.2

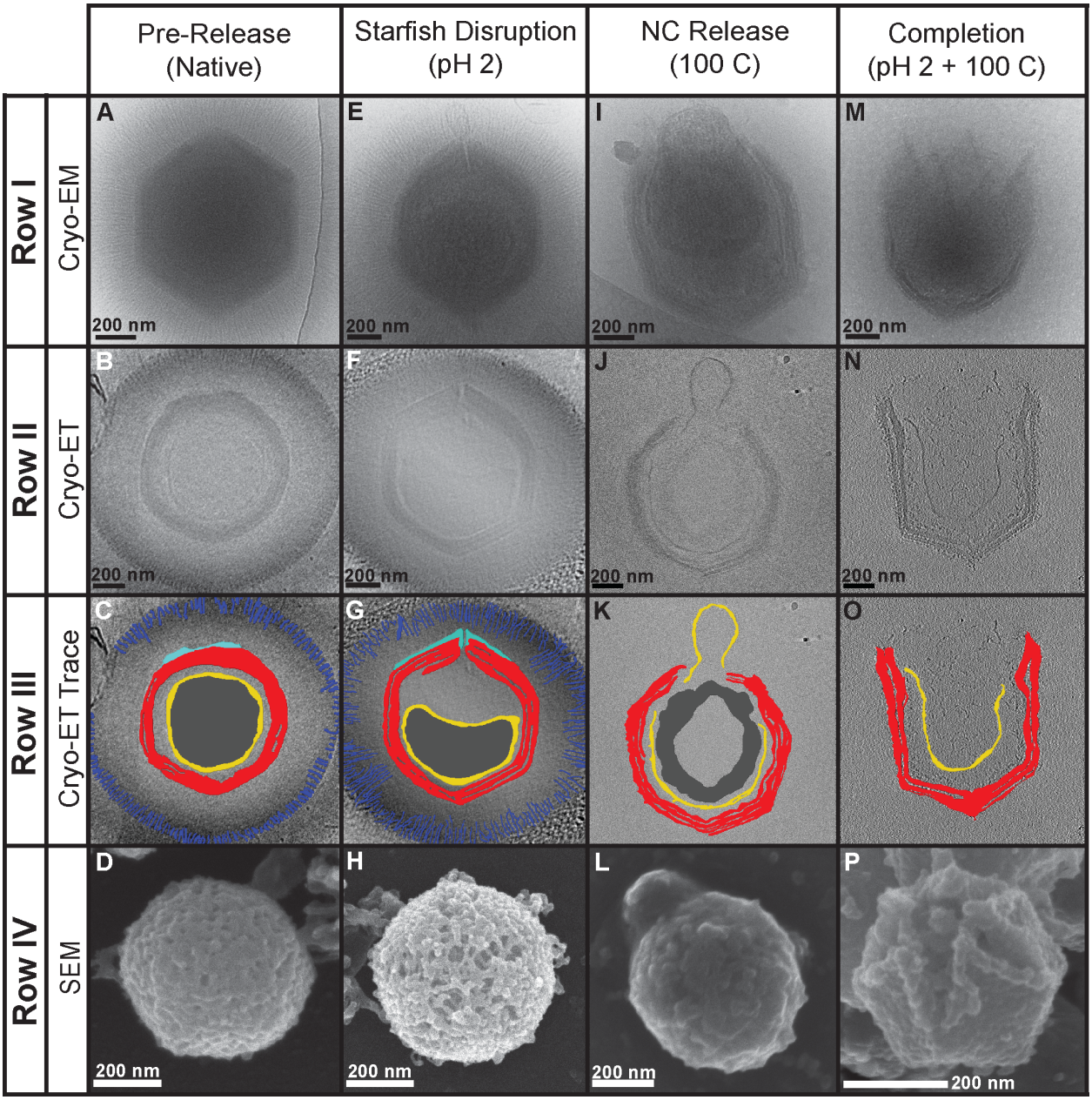

Figure 3.2 Electron Microscopy of SMBV Genome Release Stages. *Row I)* Two dimensional cryo-electron micrographs of particles following either no treatment (A), or post incubation with pH 2 (E), 100 °C (I), or both pH 2 + 100 °C (M). *Row II)* Central slices (z = 20) of cryo-electron tomograms of particles following either no treatment (B) or post incubation with pH 2 (F) 100 °C (J), or both pH 2 + 100 °C (N). *Row III)* Central slices of cryo-tomograms with key features highlighted. Blue = distal tips of the external fiber layer, Cyan = starfish seal complex, Red = capsid, Yellow = lipid membranes (nucleocapsid), Dark grey = dsDNA. Slices are shown for virions following either no treatment (C) or post incubation with pH 2 (G) 100 °C (K), or both pH 2 + 100 °C (O). *Row IV)* Scanning electron micrographs of particles in various stages of genome release following either no treatment (D) or post incubation with pH 2 (H) 100 °C (L) or both pH 2 + 100 °C (P). See Movies S3-S10 for videos of the tomograms and tilt series. See EMD-20745-20748 for tomogram volumes.

We could see that the particles had indeed opened following low pH treatment. Using 2D images alone we could not, however, determine if the starfish complex was released *en masse* or if it remained associated with the capsid. Therefore, we used scanning electron microscopy (SEM) to probe surface features. Unfortunately, SEM images of pH 2-treated SMBV particles (Figure 3.2H) also did not provide definitive evidence for the presence of the starfish seal as the layer of external fibers blocked access to the capsid surface. We next generated 3D reconstructions of opened SMBV particles through cryo-electron tomography (cryo-ET) (Figure 3.2F-G, Movie S4, EMD-20747). Tomograms confirm that the stargate vertex, and only the stargate vertex, is open in the pH 2-treated particles. Extra density corresponding to the starfish seal is clearly observed along the edges of the outer capsid layer at the stargate vertex (Movie S4). Therefore, it is likely that at least some, if not all, of the proteins that comprise the starfish seal complex remain attached to the capsid after low pH treatment. The presence of this density in our tomograms suggests that the SMBV starfish likely destabilizes through an "unzipping" mechanism rather than *en masse* release. As low pH treatment is able to trigger stargate vertex opening *in vitro*, we conclude that electrostatic interactions play a very important role in stabilizing this vertex prior to infection.

The increased concentration of H^+ ions at low pH is likely to change the protonation state of the amino acids within the starfish seal proteins. These changes in protonation state are likely to disrupt hydrogen bonding within and between proteins, potentially decreasing the stability of protein-protein interactions and/or protein folding states. These changes could be caused by side chain protonation of aspartic acid and glutamic acid residues (pKa = 3.65, 4.25, respectively). It is unlikely that protonation of the α-carboxyl groups is responsible for these structural changes.

The free pKa's for these carboxyl groups are around 2 and the morphological changes in the GV particles was visible at both pH 2 and pH 3.

We next turned to "bubblegram" imaging, a cryo-EM imaging technique used for localizing unique features within macromolecular complexes. In this technique, samples are intentionally overexposed to produce beam-induced radiation damage. If there is a unique feature within a complex, hydrogen (H_2) gas released as a result of the radiation damaging can become trapped and sometimes produces noticeable "bubbling" in the micrograph. This bubbling can be used to reveal the location and shape of the unique features in viral capsids (*126*) such as bacteriophage ΦKZ inner bodies (*170*) and also ejection proteins in bacteriophage P22 (*168*).

When untreated SMBV particles were exposed to excessive electron radiation many of the particles produced a star-shaped radiation damage pattern (Figure 3.1C-D, Movie S2). By contrast, pH 2-treated SMBV particles, displayed no star-shaped pattern (Figure 3.1E-F). As expected, the lack of a star-shaped radiation damage pattern is consistent with the hypothesis that the H_2 gas is no longer being trapped in the SMBV virion as the low pH treatment disrupted the stargate vertex seal.

Increased Thermal Energy is Required for Nucleocapsid Release

Lowering pH alone was insufficient to fully open SMBV particles, indicating that electrostatic interactions are not solely responsible for sealing the stargate. Therefore, we analyzed the effect of temperature on the stability of SMBV particles. We incubated the virions one hour at up to 100 °C, assayed the virions for morphological changes using cryo-EM, and then compared these data to images of particles that had been incubated at room temperature (25

°C). After 1 hour at 100 °C, the POP was ~33 % (Figure 3.1B). Following an additional incubation for up to five hours, the POP increased to a maximum of ~88%.

Unlike low pH, which simply cracks the stargate vertex, higher temperatures resulted in open stargate vertices with nucleocapsids in the process of exiting the virion (Figure 3.2I-L, Movie S5-S6, EMD-20748). Within these nucleocapsids the DNA appears to have reorganized leaving pockets of seemingly empty space (discussed in greater detail below.) Additionally, much of the external fiber layer is removed (Figure 3.2I-L, Figure 3.3) and the extra membrane sac is fully released from these particles. The use of high temperatures could be an alternative GV defibering method to that proposed in (*187*), especially as this previously described technique did not defiber SMBV particles (data not shown). High temperature induces a conformational change that closely mimics a stage of mimivirus infection seen *in vivo* (see Figure 2-III in (*82*)), where the nucleocapsid leaves the capsid and prepares to fuse with the amoebal phagosome membrane. As increased thermal energy induces stargate opening *in vitro*, we conclude that entropic barriers must be overcome during GV stargate opening *in vivo*. In the amoeba, these entropic barriers are likely lowered by interaction with a cellular receptor, although the identity of these receptors is currently not known for any GV.

Figure 3.3

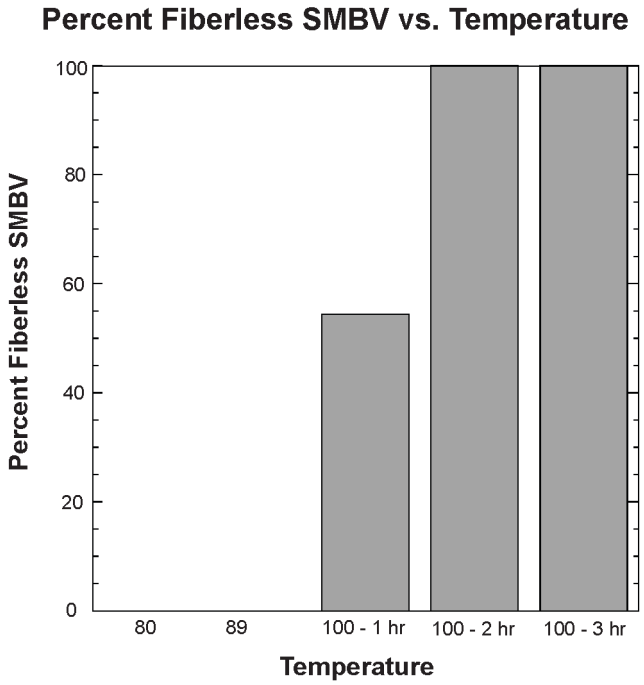

Figure 3.3 Percentage of Fiberless SMBV Particles at Varying Temperatures. Histogram of the percentage of fiberless (open or unopen) SMBV particles at various temperatures and incubation times.

Following both low pH and high temperature treatment (individually) there were pockets within the SMBV nucleocapsids that appear to be devoid of DNA (Figure 3.2J-K). These seemingly empty pockets are not visible in the untreated SMBV particles (Figure 3.2B-C). While it is possible that the void inside of SMBV nucleocapsids could be due to the extreme conditions used, it is more likely that this is biologically relevant. These pockets are only observed in SMBV particles that have begun releasing their genome, suggesting that the DNA may undergo reorganization during this process. The SMBV genome contains various chromosome condensation and histone-like proteins that could be used for this function. Mass spectrometry experiments (described below, and shown in Table 3.1) suggest that many of these proteins remain with the nucleocapsid after the initial opening stage. Genome reorganization is an important stage of many virus infection processes, including HIV (*218*) and Adenovirus (*219*). We hypothesize that genome rearrangement is also important for facilitating GV genome release into the host.

A Combination of Low pH and High Temperature Results in Complete Samba Virus Genome Release

Individually, low pH and high temperature had different physical effects on SMBV. These disparate treatments are affecting two different types of biomolecular interactions (electrostatic interactions and entropy, respectively) and each appears to contribute to SMBV virion stability. Therefore, we hypothesized that combining low pH and high temperature might have a compound effect on stargate opening. Again, following treatment the SMBV particle morphology was analyzed via cryo-EM (Figure 3.2M), cryo-ET (Figure 3.2N-O), and SEM (Figure 3.2P). These particles have completed the entire genome release process, as seen by the

absence of the nucleocapsid. Additionally, SMBV particles were completely defibered and the internal capsid layer(s) appeared to be less rigid than the outer capsid layer (Figure 3.2O, Movie S7-10, EMD-20745 & EMD-20746). Once disrupted, the capsid is more electron transparent and apparent connections between the two capsid layers were now visible in the tomograms (Figure 3.2N, Movies S8 & S10). Anchoring/tethering proteins that connect these two capsid layers may play a role in the extraordinary capsid stability of GV.

SEM of dual treated SMBV particles (Figure 3.2P) provides further evidence for the fate of the starfish seal. Particles treated with both low pH and high temperature clearly contain extra density around the edges of the stargate vertex, corresponding to the starfish seal. This extra density is consistent with our cryo-ET data described above where rather than completely dissociating from the capsid *en masse*, the starfish seal unzips to allow the stargate to open while still retaining contacts with the capsid.

Molecular Forces That Stabilize the SMBV Stargate Vertex are Conserved Amongst Diverse Giant Viruses

We tested the effects of a combination of pH and temperature on three other GV (from two distinct *Mimiviridae* lineages); Antarctica virus ((*4*), *Mimivirus A*), TV (*3*), and mimivirus M4 ((*64*), *Mimivirus A*)). Following treatment, each virus was characterized via cryo-EM (data not shown) and SEM (Figure 3.4). Similar to SMBV, all three GV had opened their stargate vertices and released their nucleocapsids after being boiled in acid. All three GV also appeared to lose the majority of their fibers during treatment. All four of the GV tested in this study had fully open stargate vertices following low pH and high temperature treatment. While all four viruses analyzed here are mimivirus-like icosahedral GV, these viruses encompass two separate GV

clades belonging to the *Mimiviridae* family: of the genus *Mimivirus* (SMBV, M4, Antarctica) and the proposed genus *Tupanvirus* (TV, (*101*)). These data strongly indicate that the general forces that stabilize virions and facilitate infection are conserved among distantly related amoeba-infecting members of *Mimiviridae*.

Although the general forces appear to be highly conserved, some specific mechanisms of starfish disruption are likely conserved only within distinct lineages. In our SEM data, Antarctica and mimivirus particles (Figure 3.4A & 3.4D, respectively) displayed density along the edges of the open stargate vertices, similar to the density seen in SMBV (Figure 3.2P, 3.4C). The presence of this extra density suggests that, like SMBV, the Antarctica and mimivirus starfish complexes unzip to facilitate stargate opening and genome release. TV, on the other hand, does not display this extra density (Figure 3.4B), suggesting that the TV starfish may completely dissociate from the capsid *en masse* during infection. TV particles also appear to fully open their stargate vertices following low pH treatment alone (data not shown). In total, our data suggest that the mechanism of seal complex unzipping may be conserved amongst *Mimiviridae* with slight deviations present between the *Mimiviruses* and the proposed *Tupanvirus* genus.

Figure 3.4

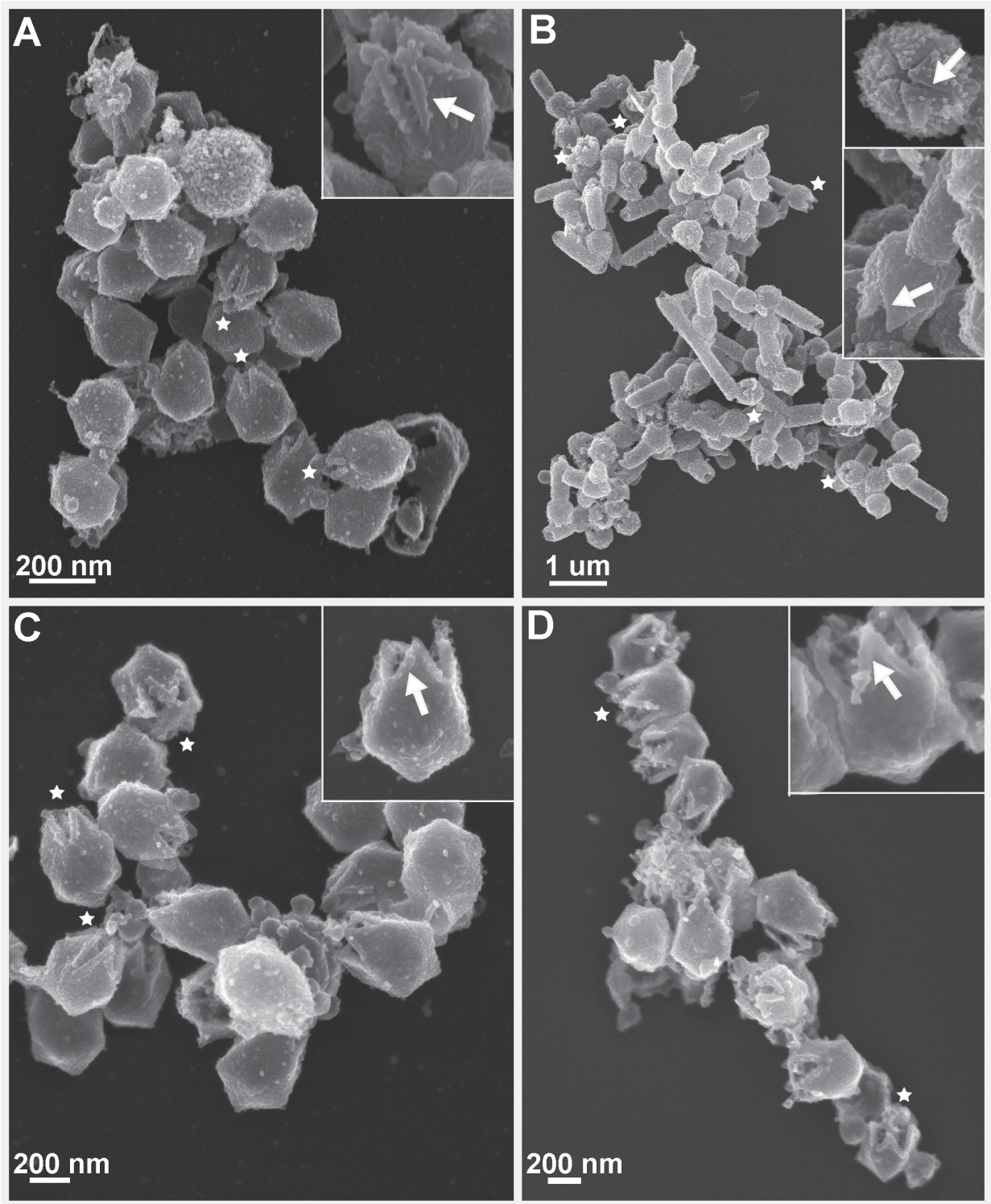

Figure 3.4 Post Genome Release Particles From Four GV. Scanning electron micrographs of low pH and high temperature-treated A) SMBV, B) TV, C) Antarctica virus, and D) mimivirus particles. Inserts demonstrate enlarged views highlighting capsids where either clear retention of the starfish seal can be seen in SMBV, mimivirus, and Antarctica particles or the lack of starfish seal retention can be seen in TV. Asterisks in the main panels depict selected particles with clearly visible open stargate vertices.

GV have changed our canonical view of virology, defying the previously known limits of capsid sizes and stabilities. Giantism is known to cause developmental and structural problems for higher organisms, such as humans (*220*), but icosahedral GV have evolved a common stargate vertex and accompanying stabilization mechanisms to counteract these issues. The description of a new GV genome release strategy signifies another paradigm shift in our understanding of virology. As mentioned previously, smaller viruses tend to share conserved genome release mechanisms. This conservation can be observed within viral families such as *Flaviviridae* (fusion proteins (*130*)), *Caudovirales* (tail complexes (*126*)), or *Orthomyxoviridae* and *Paramyxoviridae* glycoproteins (*221*). This conservation also occurs across viral kingdoms. The Herpesvirus portal complex shares structural similarity with many bacteriophage portal proteins (*145, 222*) and the Adenovirus spike protein is homologous with the bacteriophage Sf6 tail needle knob protein (*129*). GV have eschewed all of these known genome release structures and appear to have forged their own mechanisms, as exemplified by the common stargate mechanism.

Numerous Proteins are Released From Giant Virus Capsids During Stargate Opening

As obvious morphological changes occurred in the GV capsids during low pH and high temperature treatments, we hypothesized that proteins were likely released from the capsids at each of these stages. We analyzed proteins that remained within the SMBV and TV capsids and proteins liberated from the capsids after each treatment. We used four conditions, native virions (pH 7.4, room temperature), low pH (pH 2, room temperature), high temperature (pH 7, 100 °C), and combined (pH 2, 100 °C). We then performed pellet/supernatant separations to physically separate the virions and released proteins. Following separation, we analyzed the contents of

each sample via SDS-PAGE (Figure 3.5). A sample preparation scheme for these experiments can be seen in Figure 3.6. Antarctica virus and mimivirus both showed a similar banding pattern as SMBV (data not shown). We did not perform MS experiments with these viruses as there is no annotated Antarctica virus genome and mimivirus and SMBV are highly similar (*201*).

For both SMBV and TV, distinct proteins were released from the capsid following low pH treatment. Some of these proteins can be seen at the same apparent molecular weight as proteins in the native capsid (pellet) lane, suggesting they had been released from the capsid without significant modification/cleavage. Other proteins, especially in the TV sample, did not match proteins in the native capsid lane. These bands likely represent proteins that were cleaved during low pH treatment. For both viruses, the native supernatant lanes did not contain any visible protein bands. When the particles were incubated at 100 °C (with or without prior pH 2 treatment) it appeared that the majority of proteins were proteolytically cleaved and appeared as a continuous smear on the gel (data not shown) preventing detailed analysis of these samples.

Figure 3.5

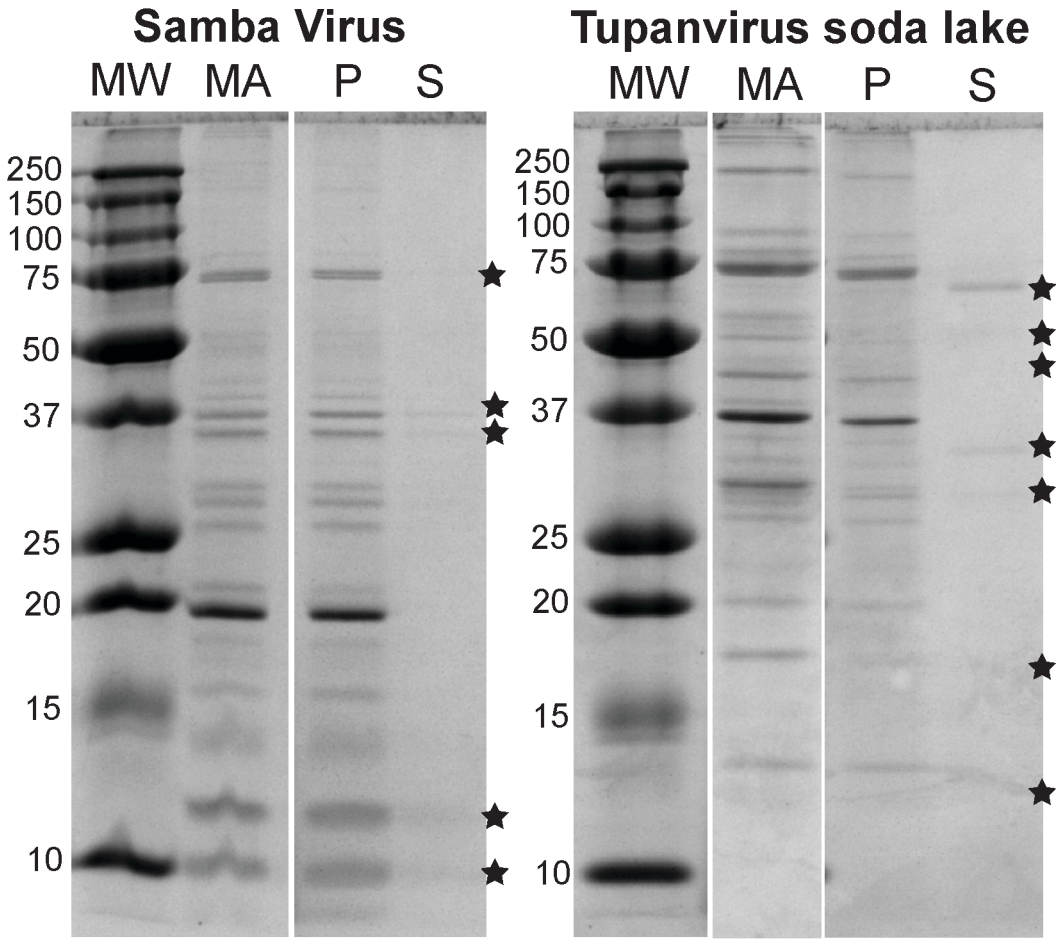

Figure 3.5 SDS-PAGE of pH 2-Treated SMBV and TV. SDS-PAGE Bands of SMBV and TV. MW = Molecular Weight standard, MA = Material Applied (untreated viral particles), P = pellets from pH 2-treated virions, S = supernatants from pH 2-treated virions. Visible bands of proteins released into the supernatant are highlighted with asterisks. See Figure S2 for the sample preparation scheme.

Figure 3.6

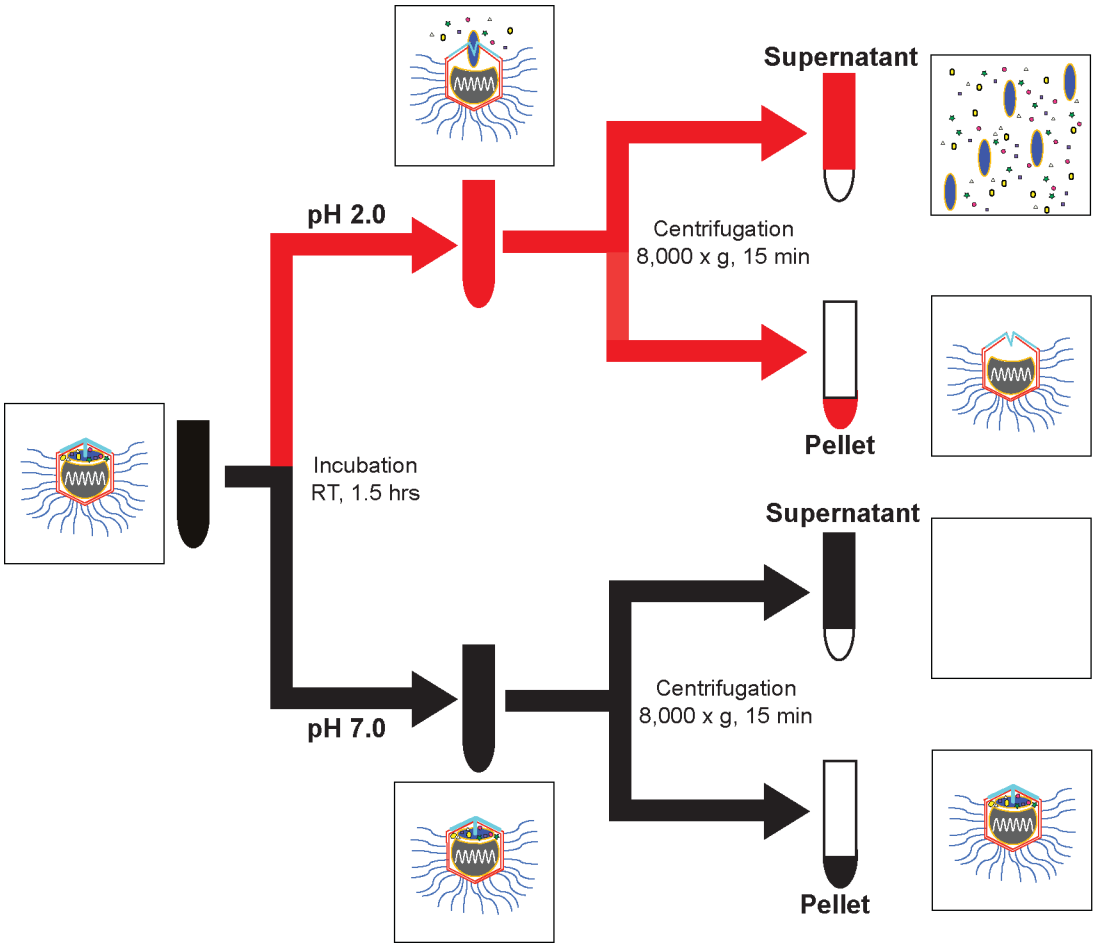

Figure 3.6 Sample Preparation for SDS-PAGE and LC/MS/MS Experiments. A cartoon of the workflow schematic used to prepare samples for both SDS-PAGE and LC/MS/MS experiments.

Identifying the Proteins Released From Samba Virus and Tupanvirus Virions at the Initiation of Infection.

To characterize the proteins released during the initial stages of GV infection, we used mass spectrometry (MS). Initially, we focused on in-gel digestion of bands from the pH 2-treated SMBV and TV supernatant samples. The low pH-treated particles mimic the beginning of the GV infection process, as the stargate vertex begins to open and the extra membrane sac leaves the capsid. Trypsinized fragments were analyzed via LC/MS/MS and the resultant peptides were compared to published SMBV and TV genome sequences (GenBank KF959826.2 & KY523104.1, respectively) as well as the *A. castellanii* genome (GenBank KB007974.1) to identify any contaminating host proteins from our analysis. The *A. castellanii* actin protein was retained within these results, as this protein is known to play a role in the infection and genome release processes of Iridoviruses (*223*). From this initial experiment, we identified 48 SMBV and 26 TV proteins that are released from the virion following low pH treatment. These proteins are labeled with a (+) in the "Band" column of Table 3.2.

Excising visible gel bands for MS analysis has the potential to miss proteins within the sample: some bands may be too faint to detect, some proteins may be too large or too small to be fully resolved or extracted, etc. Therefore, we also analyzed SMBV and TV samples using shotgun proteomics to maximize coverage in our study. We analyzed low pH pellet and supernatant samples, as well as the untreated virus using the sample preparation scheme shown in Figure 3.6. From this experiment we identified 43 SMBV proteins and 37 TV proteins ((+) in the "Shotgun" column of Table 3.2 and Table 3.3). Of these proteins, 5 SMBV proteins and 7 TV proteins were previously identified from analysis of the gel bands.

Table 3.2

Samba Virus						
Protein	***ID***	***Category***	***Presence***		***Ratio of Ratios***	
			Band	***Shotgun***	***Up in Supe***	***Down in Pellet***
Actin[a,e]	CAA23399.1	S	+	+	+	
Rpl7A, partial[a]	AAY21190.1	-		+	+	
amine oxidase	AHJ39955.1	M		+		+
DNA-dependent RNP subunit RPB9	AHJ39967.2	Tl		+		
ubiquitin-conjugating enzyme e2	AHJ39993.2	M	+	+		
WD repeat-containing protein	AHJ40002.1	-		+		
b-type lectin protein	AHJ40019.2	S		+		
protein phosphatase 2c	AHJ40032.1	Rg	+			
formamidopyrimidine-DNA glycosylase	AHJ40038.1	Ho	+			
hypothetical protein	AHJ40051.1	H/TM	+			
poly(A) polymerase catalytic subunit	AHJ40056.1	Tx	+			
hypothetical protein	AHJ40060.1	H		+		+
hypothetical protein	AHJ40061.1	H		+		+
thioredoxin domain-containing protein	AHJ40071.1	Ho	+			
mRNA-capping enzyme	AHJ40083.1	Tx		+		+
putative FtsJ-like methyltransferase	AHJ40084.1	Tx	+			
hypothetical protein	AHJ40087.2	H	+			
low complexity protein	AHJ40093.1	H		+		
core protein	AHJ40101.1	S	+			
hypothetical protein	AHJ40107.2	H		+		
capsid protein 1	AHJ40114.2	S	+			+
hypothetical protein	AHJ40128.1	H	+			
thioredoxin domain-containing protein	AHJ40129.2	Ho	+	+	+	+
hypothetical protein	AHJ40139.1	S		+	+	
hypothetical protein	AHJ40144.1	H		+		
DNA-directed RNA polymerase subunit l	AHJ40151.2	Tl	+			
hypothetical protein	AHJ40159.1	H	+			
hypothetical protein	AHJ40160.2	H		+		
hypothetical protein	AHJ40162.1	H	+			

Table 3.2 Identification of Proteins Released From SMBV and TV Capsids. Proteins released from SMBV and TV particles, whether they were identified in the excised gel band experiment (Bands) and/or in the shotgun proteomics experiment (Shotgun), and whether the proteins were overabundant in the supernatant shotgun sample (Up in Supe) or depleted in the pellet shotgun sample (Down in Pellet). Superscript designations represent the following:
[a]*Acanthamoeba castellanii* proteins
[b]Proteins involved in genome rearrangement
[c]Proteins directly involved in a putative Ubiquitin-Proteasome Degradation Pathway
[d]Metal-Conjugating Proteins
[e]Proteins similar to *Irivovirus* UPP-associated proteins

Table 3.2 (cont'd)

Samba Virus						
Protein	***ID***	***Category***	***Presence***		***Ratio of Ratios***	
			Band	***Shotgun***	***Up in Supe***	***Down in Pellet***
hypothetical protein	AHJ40169.1	H		+		
DNA-directed RNAP subunit 1	AHJ40172.1	Tl		+		+
hypothetical protein	AHJ40183.2	H		+	+	
alpha beta hydrolase/esterase/lipase[e]	AHJ40190.1	I/E	+			
hypothetical protein	AHJ40207.1	H	+			
hypothetical protein	AHJ40211.1	H	+			
hypothetical protein	AHJ40213.2	H	+	+	+	+
hypothetical protein	AHJ40220.1	H	+			
hypothetical protein	AHJ40230.1	H		+		
hypothetical protein	AHJ40243.1	H		+		+
mannose-6P isomerase	AHJ40247.1	M		+	+	
hypothetical protein	AHJ40254.1	H		+		
hypothetical protein	AHJ40271.2	H		+		
Tat pathway signal sequence domain protein[e]	AHJ40276.1	I	+		+	
collagen-like protein 7	AHJ40290.2	S		+	+	+
hypothetical protein	AHJ40316.2	H		+		+
hypothetical protein	AHJ40318.2	H		+		
hypothetical protein	AHJ40319.1	H	+			
hypothetical protein	AHJ40326.2	H		+	+	
low complexity protein	AHJ40329.1	H	+		+	
hypothetical protein	AHJ40333.1	H/TM		+	+	
chemotaxis protein	AHJ40337.1	I/S	+			
hypothetical protein	AHJ40339.1	H	+			+
hypothetical protein	AHJ40340.1	H	+			
ubiquitin thioesterase	AHJ40341.2	M	+			
hypothetical protein	AHJ40367.2	H/TM	+			
virion-associated membrane protein	AHJ40371.2	I	+			
lanosterol 14-alpha-demethylase	AHJ40393.1	I		+		+
hypothetical protein	AHJ40423.1	H	+		+	
collagen triple helix repeat containing protein	AMK61745.1	S		+		+
choline dehydrogenase-like protein	AMK61776.1	M	+			+
DNA topoisomerase 1b	AMK61799.1	Rg	+			
probable glutaredoxin	AMK61800.1	Rp/Ho	+			+
hypothetical protein	AMK61829.1	H	+	+	+	
hypothetical protein	AMK61837.1	H		+		+
hypothetical protein	AMK61849.1	H/TM		+		
hypothetical protein	AMK61856.1	H	+			+
regulator of chromosome condensation[b]	AMK61866.1	Rg/I	+			+
thiol protease	AMK61869.1	E	+			
hypothetical protein	AMK61892.1	H		+		

Table 3.2 (cont'd)

Samba Virus						
Protein	***ID***	***Category***	***Presence***		***Ratio of Ratios***	
			Band	***Shotgun***	***Up in Supe***	***Down in Pellet***
hypothetical protein	AMK61902.1	H	+			
anaerobic nitric oxide reductase transcription regulator NorR	AMK61903.1	Rg	+			
ankyrin repeat protein	AMK61918.1	-	+			
hypothetical protein	AMK61920.1	H	+			
hypothetical protein	AMK61935.1	H	+			
hypothetical protein	AMK61942.1	H		+		+
N-acetyltransferase	AMK61955.1	M	+			
prolyl 4-hydroxylase	AMK61959.1	Ho		+		
proline rich protein	AMK61968.1	-	+			+
hypothetical protein	AMK61977.1	H	+	+		
NHL repeat-containing protein	AMK61987.1	-		+		
hypothetical protein	AMK62013.1	H		+		
hypothetical protein	AMK62059.1	S	+			+
hypothetical protein	AMK62082.1	H	+			
choline dehydrogenase-like protein	AMK62096.1	M		+		+
Ubiquitin[a,c]	CAA53293.1	M		+		+

Tupanvirus Soda Lake						
Protein	***ID***	***Category***	***Presence***		***Ratio of Ratios***	
			Band	***Shotgun***	***Up in Supe***	***Down in Pellet***
hypothetical protein	AUL78681.1	E/H	+	+	+	+
hypothetical protein	AUL77600.1	E/H		+		
putative ORFan	AUL77729.1	H		+		
putative ORFan	AUL78088.1	H	+	+	+	+
hypothetical protein	AUL78481.1	H		+		
hypothetical protein	AUL78232.1	Rg		+		+
hypothetical protein	AUL78466.1	H	+	+	+	+
hypothetical protein	AUL77936.1	H		+		
hypothetical protein	AUL78214.1	H		+		+
hypothetical protein	AUL77907.1	H	+	+	+	+
hypothetical protein	AUL78468.1	H	+	+	+	+
hypothetical protein	AUL77723.1	H	+	+	+	+
hypothetical protein	AUL78464.1	H	+			
hypothetical protein	AUL78055.1	H		+	+	+
hypothetical protein	AUL77930.1	H		+	+	+
putative ORFan	AUL78635.1	H	+			+
hypothetical protein	AUL77752.1	H		+		
hypothetical protein	AUL78219.1	H		+		
hypothetical protein	AUL78093.1	H	+			+

Table 3.2 (cont'd)

Tupanvirus Soda Lake						
Protein	***ID***	***Category***	***Presence***		***Ratio of Ratios***	
			Band	***Shotgun***	***Up in Supe***	***Down in Pellet***
hypothetical protein	AUL78067.1	H		+		
hypothetical protein	AUL78191.1	H	+			+
hypothetical protein	AUL78287.1	H		+		+
hypothetical protein	AUL77694.1	H	+			
hypothetical protein	AUL77820.1	H	+			
hypothetical protein	AUL78135.1	H		+		+
hypothetical protein	AUL78143.1	H	+			+
hypothetical protein[b]	AUL78348.1	H/Rg	+		+	+
hypothetical protein	AUL77688.1	H/TM	+			+
hypothetical protein	AUL78288.1	H/TM		+		+
hypothetical protein	AUL77718.1	H/TM/E	+			
Cu-Zn superoxide dismutase[d]	AUL78503.1	Ho	+			
mg709 protein[d]	AUL77661.1	Ho		+		+
thioredoxin domain-containing protein	AUL77963.1	Ho		+		+
catalase HPII	AUL78097.1	Ho		+		+
Ig family protein	AUL78630.1	I	+			
phosphatidylethanolamine-binding protein-like protein	AUL77474.1	I	+			
putative N-acetyl transferase	AUL77680.1	M		+		
arylsulfatase	AUL78269.1	M		+		
ubiquitin domain-containing protein[c]	AUL78040.1	M	+	+	+	+
glyoxalase	AUL78134.1	M	+			
putative protein kinase	AUL78629.1	Rg		+		+
glutaredoxin	AUL78724.1	Rp/Ho		+	+	+
SNF2 family helicase	AUL77941.1	Rp/Rg	+			
capsid protein 1	AUL78147.1	S	+			
putative fibril associated protein	AUL78400.1	S		+		
kinesin-like protein[a]	AUL77838.1	S		+		+
major core protein	AUL78082.1	S	+			
putative pore coat assembly factor	AUL78211.1	S		+		+
mimivirus elongation factor aef-2	AUL78714.1	Tl	+			
DNA-directed RNAP subunit	AUL78016.1	Tl		+	+	
intein-containing DNA-directed RNAP subunit 2	AUL78362.1	Tl		+		
DNA-directed RNAP subunit 6	AUL78368.1	Tl		+	+	+
DNA-directed RNAP subunit 1	AUL78302.1	Tl		+		
putative ATP-dependent RNA helicase	AUL77829.1	Tx/Tl	+			
Actin[a,e]	CAA23399.1	S		+	+	

Table 3.3

Samba virus																	
Protein	*Accession ID*	*Pep*	*Material Applied*			*Pellet*			*Supernatant*			*Supernatant/MA*			*Pellet/MA*		
			2	*3*	*Avg*	*2*	*3*	*Avg*	*2*	*3*	*Avg*	*2*	*3*	*Avg*	*2*	*3*	*Avg*
			%	*%*	*%*	*%*	*%*	*%*	*%*	*%*	*%*						
actin	CAA23399.1	11	0.3	0.3	0.3	1.0	0.3	0.6	1.7	0.9	1.3	6.5	3.3	4.9	3.9	0.9	4.8
Rpl7A, partial	AAY21190.1	2	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.2	3.7	5.5	4.6	1.5	0.5	2.0
mannose-6P isomerase	AHJ40247.1	10	0.1	0.0	0.0	0.3	0.1	0.2	0.0	0.1	0.1	0.4	7.9	4.1	4.0	12.9	16.9
Tat pathway signal sequence domain protein	AHJ40276.1	9	0.1	0.1	0.1	0.2	0.4	0.3	0.2	0.3	0.2	2.1	4.7	3.4	2.1	6.4	8.5
hypothetical protein	AMK62013.1	8	0.1	0.1	0.1	0.0	0.2	0.1	0.0	0.6	0.3	0.0	5.9	3.0	0.0	2.3	2.3
collagen-like protein 7	AHJ40290.2	6	0.2	0.1	0.2	0.0	0.0	0.0	0.2	0.3	0.2	0.8	3.2	2.0	0.0	0.4	0.4
hypothetical protein	AMK61829.1	25	1.7	0.3	1.0	0.7	1.2	1.0	0.4	1.1	0.8	0.3	3.3	1.8	0.4	3.5	4.0
hypothetical protein	AHJ40139.1	25	0.1	0.0	0.1	0.1	0.2	0.2	0.0	0.1	0.1	0.0	3.2	1.6	1.0	6.4	7.4
hypothetical protein	AHJ40423.1	5	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.1	1.2	1.7	1.4	1.1	1.2	2.2
hypothetical protein	AHJ40333.1	7	0.8	0.8	0.8	0.6	2.3	1.4	0.5	1.5	1.0	0.6	2.0	1.3	0.7	2.9	3.7
hypothetical protein	AHJ40183.2	11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.3	1.2	0.0	4.5	4.5
hypothetical protein	AHJ40326.2	4	0.3	0.2	0.3	0.5	0.4	0.5	0.3	0.3	0.3	1.2	1.1	1.1	2.0	1.9	3.8
thioredoxin domain-containing protein	AHJ40129.2	10	0.7	0.7	0.7	0.3	0.6	0.4	0.0	1.3	0.7	0.0	2.0	1.0	0.4	0.9	1.3
low complexity protein	AHJ40329.1	22	16.1	18.2	17.2	56.2	28.9	42.5	19.4	14.4	16.9	1.2	0.8	1.0	3.5	1.6	5.1
Ubiquitin-60S ribosomal protein L40	CAA53293.1	6	0.2	0.3	0.2	0.1	0.1	0.1	0.2	0.3	0.2	0.9	1.0	0.9	0.5	0.4	0.8
hypothetical protein	AHJ40230.1	5	0.3	0.4	0.3	0.5	0.5	0.5	0.3	0.2	0.3	1.2	0.7	0.9	2.1	1.5	3.6
probable glutaredoxin	AMK61800.1	5	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	1.8	0.9	0.0	0.7	0.7
ubiquitin-conjugating enzyme e2	AHJ39993.2	6	1.0	1.1	1.1	1.3	1.0	1.2	0.8	0.9	0.8	0.8	0.8	0.8	1.3	0.9	2.2
proline rich protein	AMK61968.1	13	0.9	0.9	0.9	0.3	1.0	0.6	0.4	1.0	0.7	0.4	1.1	0.8	0.3	1.0	1.4
hypothetical protein	AHJ40169.1	3	0.9	0.7	0.8	0.9	1.4	1.1	0.4	0.6	0.5	0.4	0.7	0.6	1.0	1.8	2.8

Table 3.3 SMBV and TV Proteins With LFQ Percentages and Comparison Between Supernatant and Pellet Levels. Table of the proteins identified through the shotgun mass spectrometry experiments for SMBV and TV. Percentages for the Material Applied, Pellet, and Supernatant samples represent the percentage of the overall signal that each protein accounted for in the LFQ intensities. Supernatant/MA and Pellet/MA values represent the relative contribution of each protein to the given sample's spectral intensity as compared to the untreated particles (MA).

Table 3.3 (cont'd)

Samba virus																	
Protein	*Accession ID*	*Pep*	*Material Applied*			*Pellet*			*Supernatant*			*Supernatant/MA*			*Pellet/MA*		
			2	*3*	*Avg*	*2*	*3*	*Avg*	*2*	*3*	*Avg*	*2*	*3*	*Avg*	*2*	*3*	*Avg*
			%	*%*	*%*	*%*	*%*	*%*	*%*	*%*	*%*						
hypothetical protein	AHJ40087.2	13	0.7	0.3	0.5	0.5	1.4	1.0	0.0	0.4	0.2	0.0	1.4	0.7	0.8	5.6	6.4
lanosterol 14-alpha-demethylase	AHJ40393.1	3	0.1	0.4	0.3	0.1	0.3	0.2	0.1	0.1	0.1	1.1	0.2	0.7	1.2	0.6	1.8
b-type lectin protein	AHJ40019.2	4	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	1.3	0.6	1.2	6.3	7.5
thioredoxin domain-containing protein	AHJ40071.1	27	1.7	0.9	1.3	0.9	1.4	1.2	0.3	0.9	0.6	0.1	1.0	0.6	0.5	1.6	2.2
kinesin-like protein	AHJ40024.1	44	0.1	0.0	0.1	0.0	0.3	0.2	0.0	0.0	0.0	0.0	1.1	0.6	0.0	11.2	11.2
low complexity protein	AHJ40093.1	13	1.5	1.0	1.2	1.3	1.4	1.4	0.3	0.9	0.6	0.2	0.9	0.6	0.9	1.5	2.3
hypothetical protein	AHJ40162.1	11	1.4	1.7	1.6	0.9	0.8	0.9	0.9	0.8	0.8	0.6	0.5	0.6	0.6	0.5	1.1
hypothetical protein	AHJ40160.2	11	0.4	0.3	0.3	0.4	0.5	0.4	0.0	0.3	0.1	0.0	1.0	0.5	1.0	1.9	2.8
hypothetical protein	AHJ40213.2	24	1.0	1.0	1.0	0.7	0.9	0.8	0.5	0.6	0.5	0.5	0.6	0.5	0.7	0.8	1.5
anaerobic NOR transcription regulator NorR	AMK61903.1	7	0.3	0.2	0.2	0.2	0.4	0.3	0.0	0.1	0.1	0.2	0.7	0.4	0.7	2.0	2.6
core protein	AHJ40101.1	44	1.2	0.5	0.8	0.5	1.4	0.9	0.2	0.3	0.3	0.2	0.6	0.4	0.4	2.9	3.2
ubiquitin thioesterase	AHJ40341.2	10	0.5	0.5	0.5	0.3	0.3	0.3	0.2	0.2	0.2	0.3	0.4	0.4	0.6	0.6	1.2
hypothetical protein	AMK61920.1	30	3.6	2.7	3.2	1.1	2.1	1.6	0.3	1.7	1.0	0.1	0.6	0.4	0.3	0.8	1.1
hypothetical protein	AHJ40271.2	8	0.3	0.4	0.4	0.3	1.4	0.8	0.0	0.3	0.1	0.0	0.7	0.3	1.0	3.3	4.3
amine oxidase	AHJ39955.1	7	0.2	0.3	0.3	0.1	0.2	0.2	0.0	0.1	0.1	0.2	0.4	0.3	0.5	0.8	1.3
hypothetical protein	AMK62059.1	33	8.0	10.4	9.2	4.6	6.8	5.7	1.5	3.6	2.6	0.2	0.3	0.3	0.6	0.6	1.2
choline dehydrogenase-like protein	AMK62096.1	27	1.6	2.3	2.0	0.6	1.1	0.9	0.2	0.7	0.5	0.1	0.3	0.2	0.4	0.5	0.9
hypothetical protein	AHJ40128.1	94	1.8	0.8	1.3	0.1	2.0	1.0	0.1	0.3	0.2	0.0	0.4	0.2	0.0	2.5	2.5
hypothetical protein	AMK61902.1	9	0.6	0.7	0.6	0.7	0.5	0.6	0.0	0.3	0.1	0.0	0.4	0.2	1.2	0.8	2.0
choline dehydrogenase-like protein	AMK61776.1	56	25.3	33.0	29.1	9.1	19.6	14.3	3.7	8.5	6.1	0.1	0.3	0.2	0.4	0.6	1.0
hypothetical protein	AHJ40316.2	2	0.3	0.3	0.3	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.3	0.2	0.0	0.4	0.4
hypothetical protein	AHJ40339.1	15	0.1	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.3	0.1	0.4	0.4	0.8
WD repeat-containing protein	AHJ40002.1	17	0.3	0.3	0.3	0.1	0.3	0.2	0.0	0.1	0.0	0.0	0.3	0.1	0.2	0.9	1.1
hypothetical protein	AHJ40318.2	10	0.1	0.1	0.1	0.0	0.4	0.2	0.0	0.0	0.0	0.0	0.3	0.1	0.3	2.4	2.7

Table 3.3 (cont'd)

Samba virus																	
Protein	*Accession ID*	*Pep*	*Material Applied*			*Pellet*			*Supernatant*			*Supernatant/MA*			*Pellet/MA*		
			2	*3*	*Avg*	*2*	*3*	*Avg*	*2*	*3*	*Avg*	*2*	*3*	*Avg*	*2*	*3*	*Avg*
			%	*%*	*%*	*%*	*%*	*%*	*%*	*%*	*%*						
hypothetical protein	AMK61942.1	10	1.2	0.5	0.9	0.1	0.5	0.3	0.1	0.1	0.1	0.0	0.2	0.1	0.1	0.9	1.0
hypothetical protein	AMK61856.1	23	0.9	1.0	0.9	0.6	0.6	0.6	0.0	0.2	0.1	0.0	0.2	0.1	0.7	0.6	1.4
capsid protein 1	AHJ40114.2	45	1.5	1.0	1.2	0.1	1.6	0.8	0.0	0.1	0.1	0.0	0.1	0.1	0.1	1.6	1.6
collagen triple helix repeat containing protein	AMK61745.1	9	1.8	1.6	1.7	0.1	0.3	0.2	0.0	0.2	0.1	0.0	0.1	0.0	0.0	0.2	0.2
GMC-type oxidoreductase	AMK61775.1	3	0.0	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	1.6	1.4	3.0
glucose-methanol-choline oxidoreductase	AHJ40412.1	4	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.3
collagen triple helix repeat containing protein	AHJ40289.2	2	0.1	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.7
hypothetical protein	AHJ40232.2	3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.8
translocase of outer mitochondrial membrane 40	ADZ24223.1	2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.2	1.2
putative lipoxygenase	AMK61740.1	5	0.1	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.7
hypothetical protein	AMK61967.1	3	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.5	3.5
hypothetical protein	AMK61977.1	4	0.0	0.1	0.1	0.1	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	2.9	2.2	5.2
hypothetical protein	AMK61837.1	6	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.4
hypothetical protein	AHJ40107.2	3	0.1	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.4	2.4
DNA-dir. RNAP subunit RPB9	AHJ39967.2	7	0.0	0.1	0.1	0.2	0.1	0.2	0.0	0.0	0.0	0.0	0.0	0.0	6.9	1.1	8.0
mRNA-capping enzyme	AHJ40083.1	10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.9
DNA-dir. RNAP subunit 1	AHJ40172.1	16	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.5	1.5
prolyl 4-hydroxylase	AMK61959.1	6	0.0	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.1	1.1
hypothetical protein	AMK61849.1	6	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.5	2.5
hypothetical protein	AHJ40243.1	5	0.1	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.1	1.1
hypothetical protein	AHJ40051.1	8	0.2	0.1	0.1	0.1	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.4	2.0	2.3
hypothetical protein	AMK61892.1	17	0.1	0.0	0.1	0.0	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.2	3.2
NHL repeat-containing protein	AMK61987.1	7	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.1	2.1

Table 3.3 (cont'd)

Tupanvirus soda lake																	
Protein	*Accession ID*	*Pep*	*Material Applied*			*Pellet*			*Supernatant*			*Supernatant/MA*			*Pellet/MA*		
			2	*3*	*Avg*	*2*	*3*	*Avg*	*2*	*3*	*Avg*	*2*	*3*	*Avg*	*2*	*3*	*Avg*
			%	*%*	*%*	*%*	*%*	*%*	*%*	*%*	*%*						
alpha beta hydrolase/esterase/lipase	AHJ40190.1	10	0.1	0.1	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.1	1.1
hypothetical protein	AHJ40144.1	9	0.1	0.0	0.1	0.4	0.2	0.3	0.0	0.0	0.0	0.0	0.0	0.0	5.3	4.4	9.8
hypothetical protein	AHJ40220.1	18	0.2	0.1	0.2	0.3	0.2	0.3	0.0	0.0	0.0	0.0	0.0	0.0	1.3	2.7	4.0
hypothetical protein	AHJ40061.1	22	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.8	1.8
hypothetical protein	AHJ40254.1	5	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	1.1	1.7	2.8
hypothetical protein	AHJ40060.1	4	0.1	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.9
regulator of chromosome condensation	AMK61866.1	30	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.8

Tupanvirus soda lake																	
Protein	*Accession ID*	*Pep*	*Material Applied*			*Pellet*			*Supernatant*			*Supernatant/MA*			*Pellet/MA*		
			2	*3*	*Avg*	*2*	*3*	*Avg*	*2*	*3*	*Avg*	*2*	*3*	*Avg*	*2*	*3*	*Avg*
			%	*%*	*%*	*%*	*%*	*%*	*%*	*%*	*%*						
actin	CAA23399.1	7	0.1	0.0	0.0	0.1	0.0	0.0	77.0	0.8	38.9	1307.7	42.2	674.9	1.3	0.9	1.10
ubiquitin domain-containing protein	AUL78040.1	6	0.2	0.3	0.2	0.1	0.1	0.1	18.2	5.4	11.8	85.3	20.8	53.1	0.5	0.5	0.47
putative ORFan	AUL78088.1	7	0.1	0.2	0.2	0.0	0.1	0.0	0.0	6.9	3.5	0.0	29.9	14.9	0.0	0.3	0.16
hypothetical protein	AUL78468.1	3	0.1	0.1	0.1	0.0	0.0	0.0	0.0	1.9	0.9	0.0	20.4	10.2	0.4	0.5	0.43
glutaredoxin	AUL78724.1	2	0.1	0.1	0.1	0.0	0.0	0.0	0.0	1.1	0.5	0.0	19.3	9.7	0.2	0.0	0.12
hypothetical protein	AUL78348.1	3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	13.7	6.9	0.0	0.0	0.00
DNA-dir. RNAP. subunit	AUL78016.1	3	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.7	0.4	0.0	12.9	6.5	0.9	1.4	1.14
hypothetical protein	AUL78055.1	5	0.1	0.1	0.1	0.1	0.1	0.1	0.0	1.5	0.8	0.0	11.6	5.8	0.5	0.4	0.45
hypothetical protein	AUL77930.1	3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.2	0.0	9.8	4.9	0.0	0.0	0.00
hypothetical protein	AUL78681.1	2	0.1	0.2	0.2	0.0	0.0	0.0	0.0	2.0	1.0	0.0	9.6	4.8	0.2	0.1	0.17
DNA--dir. RNAP subunit 6	AUL78368.1	5	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.2	0.1	0.0	7.8	3.9	0.0	0.3	0.17
hypothetical protein	AUL77723.1	8	0.4	0.4	0.4	0.3	0.3	0.3	0.0	2.0	1.0	0.0	4.8	2.4	0.8	0.6	0.69
hypothetical protein	AUL77907.1	9	0.7	0.6	0.6	0.5	0.6	0.6	0.0	1.8	0.9	0.0	3.0	1.5	0.7	1.0	0.86
hypothetical protein	AUL78288.1	5	0.7	0.5	0.6	0.3	0.2	0.2	0.0	1.4	0.7	0.0	2.7	1.4	0.4	0.3	0.37

Table 3.3 (cont'd)

Tupanvirus soda lake																	
Protein	*Accession ID*	*Pep*	*Material Applied*			*Pellet*			*Supernatant*			*Supernatant/MA*			*Pellet/MA*		
			2	*3*	*Avg*	*2*	*3*	*Avg*	*2*	*3*	*Avg*	*2*	*3*	*Avg*	*2*	*3*	*Avg*
			%	*%*	*%*	*%*	*%*	*%*	*%*	*%*	*%*						
hypothetical protein	AUL78466.1	4	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.2	0.1	0.0	2.5	1.2	0.8	0.4	0.61
mg709 protein	AUL77661.1	4	0.4	0.3	0.4	0.1	0.2	0.1	0.0	0.6	0.3	0.0	1.9	1.0	0.2	0.6	0.40
hypothetical protein	AUL78191.1	31	8.7	9.7	9.2	5.4	8.8	7.1	0.0	18.6	9.3	0.0	1.9	1.0	0.6	0.9	0.76
putative pore coat assembly factor	AUL78211.1	11	0.3	0.4	0.3	0.2	0.2	0.2	0.0	0.6	0.3	0.0	1.8	0.9	0.7	0.6	0.66
catalase HPII	AUL78097.1	12	0.7	0.7	0.7	0.4	0.5	0.5	0.0	1.1	0.5	0.0	1.5	0.7	0.7	0.7	0.68
thioredoxin domain-containing protein	AUL77963.1	9	1.8	1.8	1.8	0.6	0.3	0.5	0.0	2.4	1.2	0.0	1.4	0.7	0.3	0.2	0.27
hypothetical protein	AUL77936.1	6	0.2	0.3	0.2	0.3	0.3	0.3	0.0	0.3	0.2	0.0	1.2	0.6	1.9	1.1	1.50
putative protein kinase	AUL78629.1	4	0.5	0.3	0.4	0.4	0.2	0.3	0.0	0.2	0.1	0.0	0.7	0.3	0.7	0.7	0.71
DNA-dir. RNAP subunit 1	AUL78302.1	32	0.2	0.6	0.4	0.3	0.8	0.6	0.0	0.4	0.2	0.0	0.6	0.3	1.4	1.3	1.36
arylsulfatase	AUL78269.1	9	0.4	0.4	0.4	0.5	0.3	0.4	0.0	0.3	0.1	0.0	0.6	0.3	1.2	0.8	1.00
kinesin-like protein	AUL77838.1	12	0.1	0.1	0.1	0.2	0.0	0.1	0.0	0.1	0.0	0.0	0.6	0.3	1.3	0.4	0.85
hypothetical protein	AUL77694.1	8	1.6	2.6	2.1	1.4	2.9	2.1	0.0	1.6	0.8	0.0	0.6	0.3	0.9	1.1	1.00
capsid protein 1	AUL78147.1	47	46.7	42.3	44.5	55.7	51.8	53.7	4.7	20.7	12.7	0.1	0.5	0.3	1.2	1.2	1.21
hypothetical protein	AUL78067.1	10	0.3	0.3	0.3	0.4	0.3	0.4	0.0	0.2	0.1	0.0	0.5	0.2	1.2	1.0	1.09
major core protein	AUL78082.1	35	5.4	6.3	5.9	7.4	4.8	6.1	0.0	2.9	1.5	0.0	0.5	0.2	1.4	0.8	1.07
hypothetical protein	AUL78214.1	10	1.3	1.2	1.3	0.9	1.3	1.1	0.0	0.6	0.3	0.0	0.5	0.2	0.7	1.1	0.88
putative fibril associated protein	AUL78400.1	17	5.7	4.0	4.8	5.0	5.0	5.0	0.0	1.7	0.9	0.0	0.4	0.2	0.9	1.2	1.07
hypothetical protein	AUL78287.1	14	1.8	2.4	2.1	0.8	1.2	1.0	0.0	1.0	0.5	0.0	0.4	0.2	0.4	0.5	0.46
hypothetical protein	AUL78232.1	3	0.3	0.2	0.2	0.0	0.1	0.1	0.0	0.1	0.0	0.0	0.4	0.2	0.1	0.4	0.26
hypothetical protein	AUL78219.1	8	1.6	2.2	1.9	1.9	2.5	2.2	0.0	0.8	0.4	0.0	0.4	0.2	1.2	1.1	1.16
putative ORFan	AUL77729.1	7	0.8	0.5	0.6	0.4	0.7	0.6	0.0	0.2	0.1	0.0	0.3	0.2	0.5	1.5	1.02
hypothetical protein	AUL77688.1	5	3.5	3.9	3.7	2.4	3.1	2.8	0.0	1.2	0.6	0.0	0.3	0.2	0.7	0.8	0.74
hypothetical protein	AUL78135.1	11	2.4	2.2	2.3	2.0	1.6	1.8	0.0	0.6	0.3	0.0	0.3	0.1	0.8	0.7	0.79
hypothetical protein	AUL78143.1	9	5.3	7.2	6.3	5.9	4.9	5.4	0.0	2.1	1.0	0.0	0.3	0.1	1.1	0.7	0.90
intein-containing DNA-dir. RNAP subunit 2	AUL78362.1	7	0.2	0.2	0.2	0.2	0.3	0.2	0.0	0.1	0.0	0.0	0.2	0.1	1.1	1.1	1.09
hypothetical protein	AUL77600.1	3	0.4	0.3	0.4	0.5	0.2	0.3	0.0	0.1	0.0	0.0	0.2	0.1	1.2	0.6	0.88

Table 3.3 (cont'd)

Tupanvirus soda lake																	
Protein	*Accession ID*	*Pep*	*Material Applied*			*Pellet*			*Supernatant*			*Supernatant/MA*			*Pellet/MA*		
			2	*3*	*Avg*	*2*	*3*	*Avg*	*2*	*3*	*Avg*	*2*	*3*	*Avg*	*2*	*3*	*Avg*
			%	*%*	*%*	*%*	*%*	*%*	*%*	*%*	*%*						
hypothetical protein	AUL78481.1	4	0.1	0.2	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.1	0.1	1.1	0.8	0.99
hypothetical protein	AUL77492.1	3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.1	0.7	0.92
hypothetical protein	AUL77863.1	2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.8	0.71
DNA-dir. RNAP subunit	AUL78244.1	3	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.3	1.0	2.11
thiol oxidoreductase E10R	AUL77655.1	2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.3	0.8	1.06
putative ankyrin repeat protein	AUL78278.1	6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.9	1.3	1.58
bifunctional metalloprotease ubiquitin-protein ligase	AUL78691.1	3	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.2	0.9	1.04
hypothetical protein	AUL78731.1	2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00
putative ORFan	AUL77532.1	2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.8	0.9	1.31
hypothetical protein	AUL78045.1	3	0.0	0.1	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.01
structural ppiase-like protein	AUL77649.1	4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.8	0.76
hypothetical protein	AUL77666.1	5	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	3.0	1.1	2.06
mg749 protein	AUL77517.1	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.5	0.44
hypothetical protein	AUL78068.1	7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.1	1.4	1.25
dna topoisomerase 1b	AUL78109.1	10	0.0	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.9	0.81
intein-containing DNA-dir. RNAP subunit 2	AUL78361.1	12	0.0	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	2.2	1.3	1.76
chemotaxis	AUL78637.1	4	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	1.2	0.9	1.10
phosphoesterase-like protein	AUL77796.1	4	0.1	0.1	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.7	0.81
hypothetical protein	AUL78280.1	2	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.4	1.17
hypothetical protein	AUL78198.1	2	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.28
hypothetical protein	AUL77647.1	7	0.1	0.0	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.9	1.1	0.99
hypothetical protein	AUL78155.1	2	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	1.6	0.7	1.12
hypothetical protein	AUL78061.1	3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00
putative protein phosphatase 2c	AUL77859.1	4	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.1	1.05

Table 3.3 (cont'd)

Protein	*Accession ID*	*Pep*	*Material Applied*			*Pellet*			*Supernatant*			*Supernatant/MA*			*Pellet/MA*		
			2	*3*	*Avg*	*2*	*3*	*Avg*	*2*	*3*	*Avg*	*2*	*3*	*Avg*	*2*	*3*	*Avg*
			%	*%*	*%*	*%*	*%*	*%*	*%*	*%*	*%*						
FtsJ-like methyl transferase	AUL78032.1	5	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	1.4	1.4	1.41
hypothetical protein	AUL77903.1	3	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.03
hypothetical protein	AUL77961.1	5	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.8	0.88
SNF2 family helicase	AUL77941.1	9	0.1	0.1	0.1	0.1	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	1.8	2.4	2.10
polyA polymerase catalytic subunit	AUL77929.1	12	0.1	0.2	0.1	0.1	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	1.1	1.0	1.08
hypothetical protein	AUL78093.1	4	0.1	0.2	0.2	0.1	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.7	0.73
hypothetical protein	AUL78319.1	4	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.00
thioredoxin domain-containing protein	AUL78192.1	2	0.1	0.1	0.1	0.2	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	1.5	0.0	0.75
DNA-dep. RNAP subunit Rpb9	AUL78739.1	4	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.7	0.83
hypothetical protein	AUL77933.1	6	0.1	0.2	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	1.2	0.9	1.04
mRNA capping enzyme	AUL78031.1	9	0.1	0.2	0.2	0.2	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	1.2	1.3	1.26
glycosyl hydrolase family 18	AUL77711.1	2	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.7	1.0	0.84
NTPase	AUL78021.1	14	0.2	0.2	0.2	0.2	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.9	1.1	1.02
putative oxireductase	AUL77599.1	8	0.2	0.3	0.2	0.1	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.6	0.66
hypothetical protein	AUL78246.1	15	0.2	0.2	0.2	0.3	0.3	0.3	0.0	0.0	0.0	0.0	0.0	0.0	1.6	1.4	1.48
putative ORFan	AUL78635.1	3	0.3	0.2	0.2	0.1	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.3	1.0	0.63
hypothetical protein	AUL78601.1	4	0.3	0.1	0.2	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.5	1.0	0.77
putative early transcription factor	AUL77899.1	19	0.3	0.4	0.4	0.3	0.4	0.4	0.0	0.0	0.0	0.0	0.0	0.0	1.2	0.9	1.06
putative ORFan	AUL78206.1	4	0.4	0.5	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00
hypothetical protein	AUL77752.1	9	0.5	0.4	0.5	0.3	0.7	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.6	1.6	1.13
hypothetical protein	AUL78292.1	3	0.7	0.5	0.6	0.1	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.3	0.23

In total, 86 SMBV proteins and 56 TV proteins were identified as having been released from the capsids at low pH. TV was isolated from an environment with high salinity and alkaline pH (9-12, (*3*)). SMBV, on the other hand, was isolated from a tributary of the Amazon River, a relatively neutral environment. Due to its location, TV had to evolve pH stability into its capsid to a greater extent than SMBV. While TV was originally isolated from a basic environment some of the strategies that the virus could have developed to stabilize its proteins, such as using a higher percentage of non-polar amino acids, could also stabilize the proteins at low pH.

187 and 169 total proteins were identified within the untreated mature virions of TV and SMBV, respectively (Figure 3.7). To identify proteins of interest (those that had been released), we calculated the percent of the total peptide signal for each protein. We compared these percentages across the three samples, specifically looking at the ratios of supe:MA (Material Applied) and pellet:MA. Proteins where the supe:MA > 1 were enriched in the treated supernatant sample, indicating that they had been released from the capsids. These proteins are identified with a (+) in the "Up in Supe" column of Table 1. Conversely, proteins with pellet:MA < 1 were less abundant in the treated pellet than the native particles, and likely also released. These proteins are identified with a (+) in the "Down in Pellet" column of Table 1. Proteins that are enriched in the supernatant samples are definitely released from the GV capsids, as no proteins were identified in the untreated supernatant samples (data not shown). Proteins that are depleted in the pellet samples are also likely released from the GV particles, although it is unlikely that any of these proteins are completely absent from the pellet samples (see POP in Figure 3.1A).

Figure 3.7

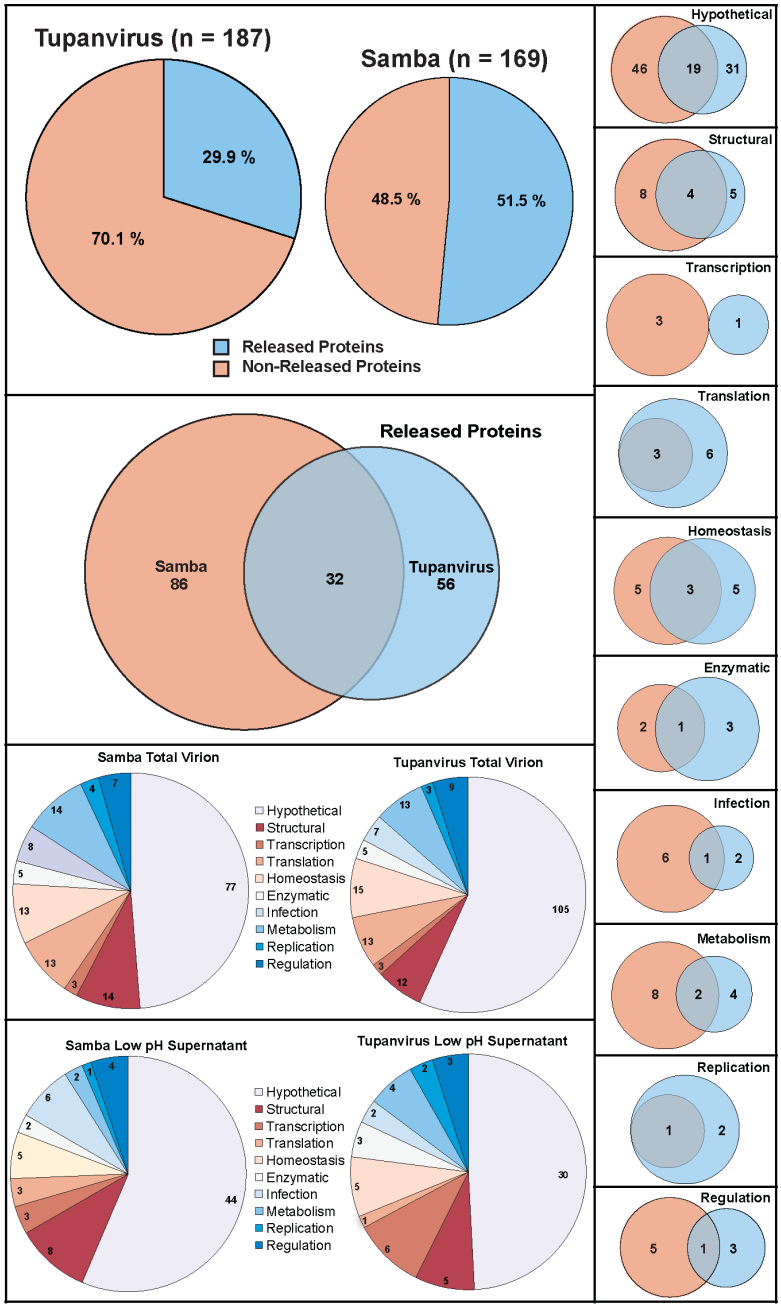

Figure 3.7 Comparison of Proteins Released by SMBV and TV. Venn diagrams comparing the total protein content (A-B) and proteins released following low pH treatment (C) of SMBV (Red) and TV (Blue) particles. The homology present within these protein sets is depicted in panels D-N. See Tables S2 for hypothetical proteins with predicted transmembrane domains and Table 3.3 for the relative abundance of individual proteins in each the untreated particles and the treated pellet and supernatant samples.

SMBV releases a higher number and percentage of these proteins (86, 51.5%) than TV particles (56, 29.9%). Putative functions for the released proteins were determined via 1) previous annotation (*3, 95*), 2) NCBI BLAST analysis, 3) HHBLITS analysis (*224*), 4) InterPro functional prediction (*225*), and 5) PSIPRED domain prediction using the DomPred functionality (*226, 227*). Released proteins for each virus were separated into the following 10 categories: Hypothetical (hypothetical proteins or ORFans), Structural, Transcription, Translation, Homeostasis, Enzymatic, Infection, Metabolism, Replication, and Regulation (Figure 3.7B-N).

For BLAST analysis, proteins sharing >35% sequence similarity were determined to share potential homology. The resultant homology pairs can be seen in Figure 3.8 and Tables 3.4 to 3.7. In Figure 3.8 and Table 3.5 the proteins released from SMBV and TV capsids were also compared to the entire predicted proteomes of each virus. From these analyses, we were able to identify putative functions for three SMBV hypothetical proteins and one TV hypothetical protein.

Figure 3.8

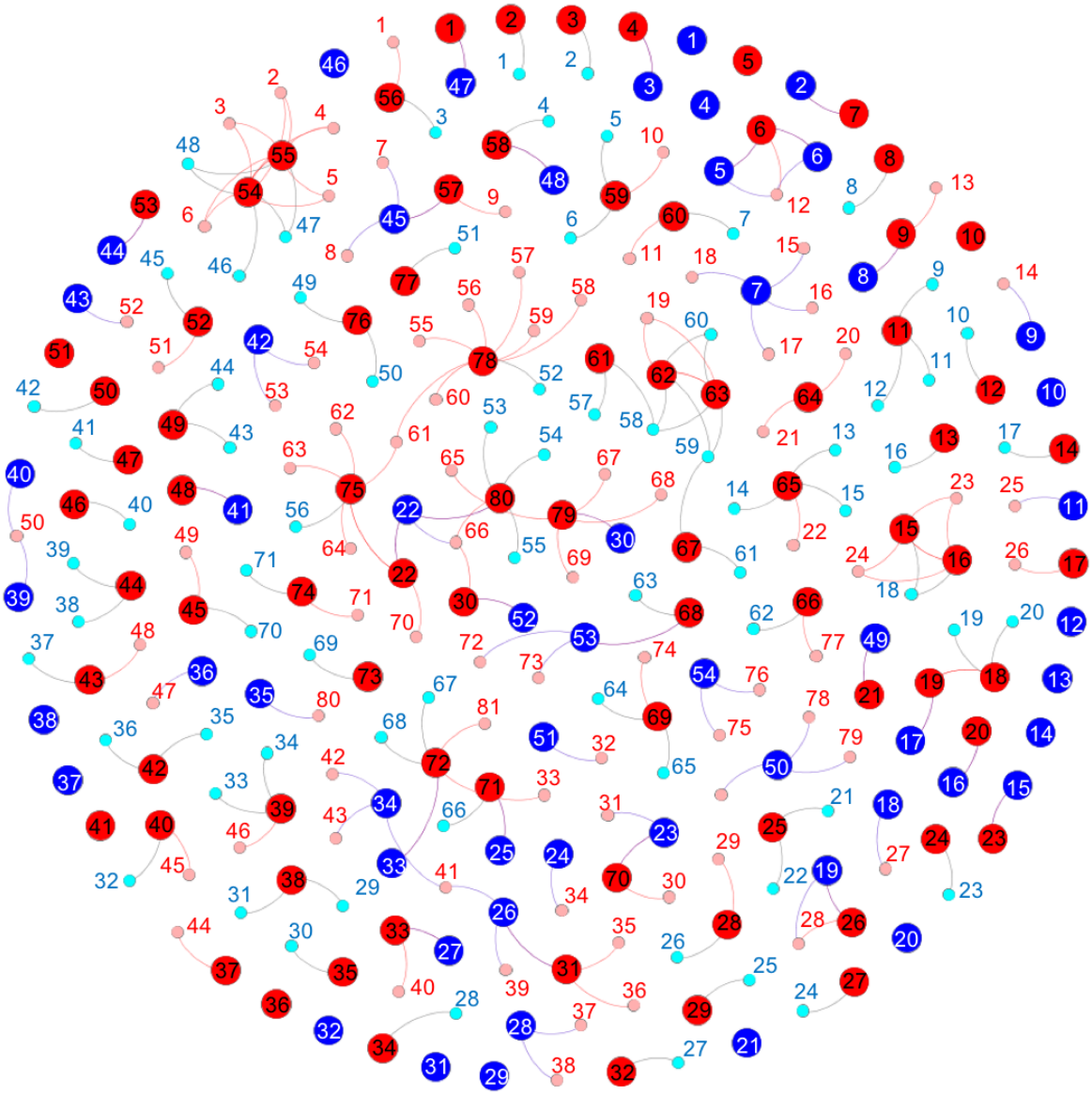

Figure 3.8 Homology Prediction of Proteins Released by SMBV and TV. Homology network of the proteins released from SMBV and TV virions during the initiation of infection. Released proteins are represented by large nodes (SMBV = Red, TV = Blue). Non-released proteins are represented by small nodes (SMBV = pink, TV = cyan). Homology was predicted using BLAST+ (*228*) with a 35 % sequence identity cutoff. Network creation was performed using Gephi (*229*). Identities of the proteins and analysis of the network can be seen in Tables 3.4-3.7.

Table 3.4

Samba Protein			***Tupan Protein***		
# (Fig. 3.8)	*ID*	*Protein*	*# (Fig 3.8)*	*ID*	*Protein*
1	AHJ40211.1	hypothetical protein	47	AUL77729.1	putative ORFan
4	AHJ40051.1	hypothetical protein	3	AUL77907.1	hypothetical protein
6	AHJ40144.1	hypothetical protein	5	AUL78219.1	hypothetical protein
6	AHJ40144.1	hypothetical protein	6	AUL78214.1	hypothetical protein
7	AHJ40107.2	hypothetical protein	2	AUL78093.1	hypothetical protein
9	AHJ40213.2	hypothetical protein	8	AUL77723.1	hypothetical protein
19	AMK61800.1	probable glutaredoxin	17	AUL78724.1	glutaredoxin
20	AHJ40220.1	hypothetical protein	16	AUL77694.1	hypothetical protein
21	AMK61920.1	hypothetical protein	49	AUL78287.1	hypothetical protein
22	AMK61942.1	hypothetical protein	22	AUL77718.1	hypothetical protein
23	AMK62059.1	hypothetical protein	15	AUL78400.1	putative fibril-associated protein
26	AHJ40071.1	thioredoxin domain-containing protein	19	AUL77963.1	thioredoxin domain-containing protein
30	AMK62013.1	hypothetical protein	52	AUL78681.1	hypothetical protein
31	AHJ40172.1	DNA-dir. RNAP subunit 1	26	AUL78302.1	DNA-dir. RNAP subunit 1
33	AMK61903.1	anaerobic nitric oxide reductase transcription factor regulator NorR	27	AUL78232.1	hypothetical protein
48	AMK61955.1	N-acetyltransferase	41	AUL77680.1	putative N-acetyl transferase
53	AHJ40061.1	hypothetical protein	44	AUL77936.1	hypothetical protein
57	AHJ40139.1	hypothetical protein	45	AUL78211.1	putative pore coat assembly factor
58	AMK61959.1	prolyl 4-hydroxylase	48	AUL77661.1	mg709 protein
68	AHJ40114.2	capsid protein 1	53	AUL78147.1	capsid protein 1
70	AHJ40128.1	hypothetical protein	23	AUL78191.1	hypothetical protein
71	AHJ40160.2	hypothetical protein	25	AUL78288.1	hypothetical protein
72	AHJ39993.2	ubiquitin-conjugating enzyme e2	33	AUL78348.1	hypothetical protein

Table 3.4 Homology Predictions of SMBV and TV Released Proteins. Predicted homology pairs of SMBV and TV proteins released at the initiation of infection. Homology is based on >35% sequence identity predicted using the NCBI BLAST+ software (*228*). Numbers (# (Fig 3.8)) for each protein represent the number of the corresponding node in the homology network in Figure 3.8.

Table 3.5

SMBV Proteins				
Released Protein	***Paired Protein***			
	SMBV	***SMBV - Not Released***	***TV***	***TV - Not Released***
AHJ39955.1		AHJ40046.2		AUL78382.1
AHJ39967.2				AUL78406.1, AUL78739.1
AHJ39993.2	AHJ40160.2	AMK61929.1	AUL78348.1	AUL77725.1, AUL77569.1
AHJ40002.1				AUL78316.1
AHJ40019.2				AUL77824.1
AHJ40032.1				AUL78531.1, AUL77859.1
AHJ40038.1		AHJ40336.2		AUL78098.1, AUL 77877.1
AHJ40051.1			AUL77907.1	
AHJ40056.1				AUL77929.1
AHJ40060.1				AUL77884.1
AHJ40061.1			AUL77936.1	
AHJ40071.1		AHJ40209.1	AUL77963.1	
AHJ40083.1	AMK61942.1	AMK61914.1, AMK61841.1, AHJ39847.1, AHJ39887.1		AUL78031.1
AHJ40084.1		AMK61735.1		AUL78032.1
AHJ40087.2				AUL78038.1
AHJ40093.1				AUL78061.1
AHJ40101.1	AHJ40367.2	AMK61830.1, AHJ40291.2, AHJ39870.1, AHJ39889.1		
AHJ40107.2			AUL78093.1	
AHJ40114.2			AUL78147.1	AUL78403.1
AHJ40128.1			AUL78191.1	AUL77471.1, AUL78575.1
AHJ40129.2	AMK61800.1			AUL78192.1, AUL78738.1
AHJ40139.1		AHJ40145.1	AUL78211.1	AUL78068.1, AUL77929.1
AHJ40144.1		AHJ40242.1	AUL78219.1, AUL78214.1	
AHJ40159.1		AHJ40191.2		AUL78286.1, AUL78710.1
AHJ40160.2	AHJ39993.2	AHJ40372.1	AUL78288.1	AUL78383.1
AHJ40162.1				AUL78292.1
AHJ40169.1				AUL78301.1
AHJ40172.1		AHJ39852.1, AMK61764.1	AUL78302.1	

Table 3.5 Homology Pairings for Released SMBV Proteins. SMBV or TV proteins that share >35% sequence homology with the released SMBV proteins. These connections are visually depicted in Figure 3.8 and the identity of each of these proteins can be found in Table 3.6.

Table 3.5 (cont'd)

Released Protein	*Paired Protein*			
	SMBV	***SMBV - Not Released***	***TV***	***TV - Not Released***
AHJ40183.2		AMK61929.1, AMK61737.1		
AHJ40190.1		AHJ40095.2		AUL78070.1
AHJ40211.1			AUL77729.1	
AHJ40213.2			AUL77723.1	AUL78251.1
AHJ40220.1			AUL77694.1	
AHJ40230.1		AHJ4001.2		
AHJ40236.2				AUL78092.1, AUL78633.1, AUL78623.1
AHJ40243.1				AUL78715.1, AUL78575.1
AHJ40247.1				AUL77553.1
AHJ40254.1				AUL77517.1
AHJ40271.2				AUL77471.1
AHJ40278.1				AUL77477.1, AUL77492.1
AHJ40290.2	AMK61745.1	AMK61801.1, AMK61821.1, AHJ40289.2, AMK61819.1, AMK61820.1		AUL77853.1
AHJ40316.2				AUL78505.1, AUL78575.1
AHJ40318.2	AMK61987.1	AMK61984.1		AUL77903.1, AUL77477.1, AUL78577.1
AHJ40319.1				AUL78600.1
AHJ40333.1				AUL78659.1
AHJ40337.1				AUL78687.1
AHJ40339.1				AUL78702.1, AUL77903.1
AHJ40340.1				AUL78037.1, AUL78707.1
AHJ40341.2		AMK61967.1		
AHJ40367.2	AHJ40101.1	AMK61919.1		AUL78282.1, AUL78281.1, AUL77531.1
AHJ40371.2		AHJ39877.1		AUL78251.1, AUL77677.1, AUL78501.1
AHJ40423.1		AHJ40389.1		AUL78670.1, AUL78587.1
AMK61745.1	AHJ40290.2	AMK61801.1, AMK61821.1, AHJ40289.2, AMK61819.1, AMK61820.1		AUL77475.1, AUL78470.1, AUL77531.1

Table 3.5 (cont'd)

Released Protein	*Paired Protein*			
	SMBV	***SMBV - Not Released***	***TV***	***TV - Not Released***
AMK61776.1	AMK62096.1	AHJ40412.1, AMK61775.1		AUL77599.1
AMK61799.1				AUL78109.1
AMK61800.1	AHJ40129.2		AUL78724.1	
AMK61829.1				AUL78347.1
AMK61849.1				AUL77856.1
AMK61856.1		AMK61982.1		AUL77896.1
AMK61866.1		AHJ40000.1		AUL77933.1
AMK61869.1		AHJ39939.2		AUL77949.1
AMK61892.1		AHJ40100.2		AUL78068.1
AMK61903.1		AMK61992.1	AUL78232.1	
AMK61918.1		AMK61707.1, AMK61751.1, AMK61915.1, AMK62066.1, AHJ40428.1, AHJ40355.2, AMK61914.1		AUL78278.1
AMK61920.1			AUL78287.1	
AMK61935.1		AHJ40248.1		AUL77772.1
AMK61942.1	AHJ40083.1	AMK62036.1	AUL77718.1	
AMK61955.1			AUL77680.1	
AMK61959.1			AUL77661.1	AUL78208.1
AMK61977.1		AHJ40409.1		AUL77492.1
AMK61987.1	AHJ40318.2	AMK61984.1		AUL77903.1, AUL77477.1, AUL78577.1
AMK62013.1		AMK61919.1	AUL78681.1	
AMK62059.1			AUL78400.1	
AMK62082.1				AUL78112.1, AUL78561.1
AMK62096.1	AMK1776.1	AHJ40412.1, AMK61775.1		AUL77599.1

Table 3.6

Protein Number	*SMBV - Released*	*SMBV - Not Released*	*TV - Released*	*TV - Not Released*
Color (Fig. 3.8)	*Red*	*Pink*	*Blue*	*Cyan*
1	AHJ40211.1	AHJ40000.1	AUL78635.1	AUL78109.1
2	AMK61799.1	AMK61801.1	AUL78093.1	AUL78659.1
3	AHJ40333.1	AMK61821.1	AUL77907.1	AUL77933.1
4	AHJ40051.1	AHJ40289.2	AUL78466.1	AUL78208.1
5	AMK61837.1	AMK61819.1	AUL78219.1	AUL78098.1
6	AHJ40144.1	AMK61820.1	AUL78214.1	AUL77877.1
7	AHJ40107.2	AHJ40117.2	AUL78067.1	AUL78068.1
8	AHJ40056.1	AMK61744.1	AUL77723.1	AUL77929.1
9	AHJ40213.2	AHJ40145.1	AUL78097.1	AUL78092.1
10	AMK61968.1	AHJ40336.2	AUL78468.1	AUL77856.1
11	AHJ40236.2	AHJ40100.2	AUL78629.1	AUL78633.1
12	AMK61849.1	AHJ40242.1	AUL78503.1	AUL78623.1
13	AHJ40247.1	AMK62030.1	AUL78134.1	AUL78251.1
14	AHJ40319.1	AMK61957.1	AUL77600.1	AUL77677.1
15	AMK61776.1	AMK61705.1	AUL78400.1	AUL78501.1
16	AMK62096.1	AHJ39959.1	AUL77694.1	AUL77553.1
17	AHJ40230.1	AHJ40199.2	AUL78724.1	AUL78600.1
18	AHJ40129.2	AMK62000.1	AUL77829.1	AUL77599.1
19	AMK61800.1	AMK61984.1	AUL77963.1	AUL78192.1
20	AHJ40220.1	AMK61929.1	AUL78143.1	AUL78738.1
21	AMK61920.1	AMK61737.1	AUL78464.1	AUL78406.1
22	AMK61942.1	AHJ39877.1	AUL77718.1	AUL78739.1
23	AMK62059.1	AHJ40412.1	AUL78191.1	AUL77517.1
24	AHJ40254.1	AKM61775.1	AUL78269.1	AUL78316.1
25	AHJ39967.2	AMK62004.1	AUL78288.1	AUL78687.1
26	AHJ40071.1	AHJ4001.2	AUL78302.1	AUL78070.1
27	AHJ40002.1	AHJ40021.2	AUL78232.1	AUL77824.1
28	AHJ40190.1	AHJ40209.1	AUL78368.1	AUL78347.1
29	AHJ40337.1	AHJ40095.2	AUL78088.1	AUL78505.1
30	AMK62013.1	AHJ39897.2	AUL78082.1	AUL77471.1
31	AHJ40172.1	AHJ40388.1	AUL78135.1	AUL78575.1
32	AHJ40019.2	AHJ40268.1	AUL78714.1	AUL78032.1
33	AMK61903.1	AHJ40372.1	AUL78348.1	AUL78670.1
34	AMK61829.1	AHJ39945.1	AUL78362.1	AUL78587.1
35	AHJ40271.2	AHJ39852.1	AUL77752.1	AUL78715.1
36	AHJ40329.1	AMK61764.1	AUL78016.1	AUL77650.1
37	AHJ40341.2	AHJ39981.1	AUL77820.1	AUL78382.1
38	AHJ40316.2	AHJ40349.1	AUL77930.1	AUL78037.1
39	AHJ40423.1	AHJ40170.1	AUL78055.1	AUL78707.1
40	AHJ40084.1	AMK61992.1	AUL78481.1	AUL77884.1
41	AHJ40393.1	AHJ39850.1	AUL77680.1	AUL78038.1
42	AHJ40243.1	AMK61823.1	AUL77688.1	AUL78292.1
43	AHJ39955.1	AHJ39988.2	AUL78630.1	AUL78531.1
44	AHJ40340.1	AMK61967.1	AUL77936.1	AUL77859.1

Table 3.6 SMBV and TV Released Protein Homologues. Non-released SMBV or TV proteins with predicted homology to proteins released by either of the viruses. These proteins are represented in Figure 3.8 by nodes of the color noted in Row 2. The specific homology pairs and the identities of these proteins can be found in Table 3.7.

Table 3.6 (cont'd)

Protein Number	*SMBV - Released*	*SMBV - Not Released*	*TV - Released*	*TV - Not Released*
Color (Fig. 3.8)	*Red*	*Pink*	*Blue*	*Cyan*
45	AMK61869.1	AMK61735.1	AUL78211.1	AUL77772.1
46	AHJ40060.1	AHJ40389.1	AUL78040.1	AUL77475.1
47	AHJ40087.2	AHJ40080.2	AUL77729.1	AUL78470.1
48	AMK61955.1	AHJ40046.2	AUL77661.1	AUL77853.1
49	AHJ40032.1	AHJ39939.2	AUL78287.1	AUL78112.1
50	AHJ40162.1	AHJ40296.1	AUL77941.1	AUL78561.1
51	AMK61902.1	AHJ40248.1	AUL77474.1	AUL78061.1
52	AMK61935.1	AHJ40429.1	AUL78681.1	AUL78278.1
53	AHJ40061.1	AHJ39883.1	AUL78147.1	AUL78282.1
54	AMK61745.1	AHJ40132.1	AUL77838.1	AUL78281.1
55	AHJ40290.2	AMK61707.1	-------	AUL77531.1
56	AMK61866.1	AMK61751.1	-------	AUL78031.1
57	AHJ40139.1	AMK61915.1	-------	AUL78702.1
58	AMK61959.1	AMK62066.1	-------	AUL77903.1
59	AHJ40038.1	AHJ40428.1	-------	AUL77477.1
60	AMK61892.1	AHJ40355.2	-------	AUL78577.1
61	AHJ40339.1	AMK61914.1	-------	AUL77492.1
62	AMK61987.1	AMK61841.1	-------	AUL77492.1
63	AHJ40318.2	AHJ39887.1	-------	AUL78403.1
64	AHJ40183.2	AHJ39847.1	-------	AUL78286.1
65	AHJ40371.2	AMK61830.1	-------	AUL78710.1
66	AMK61977.1	AMK61919.1	-------	AUL78383.1
67	AHJ40278.1	AHJ402919.2	-------	AUL77725.1
68	AHJ40114.2	AHJ39870.1	-------	AUL77569.1
69	AHJ40159.1	AHJ39889.1	-------	AUL78301.1
70	AHJ40128.1	AMK62036.1	-------	AUL77949.1
71	AHJ40160.2	AMK61982.1	-------	AUL77896.1
72	AHJ39993.2	AHJ40127.2	-------	-------
73	AHJ40169.1	AHJ40126.1	-------	-------
74	AMK61856.1	AHJ40191.2	-------	-------
75	AHJ40083.1	AHJ40024.1	-------	-------
76	AMK62082.1	AHJ40141.2	-------	-------
77	AHJ40093.1	AHJ40409.1	-------	-------
78	AMK61918.1	AHJ40201.1	-------	-------
79	AHJ40101.1	AHJ40063.1	-------	-------
80	AHJ40367.2	AMK61946.1	-------	-------
81	AMK61942.1	-------	-------	-------

Table 3.7

SMBV - Released			
Protein ID	***Protein***	***Protein ID***	***Protein***
AHJ39955.1	Amine Oxidase	AHJ40290.2	Collagen-Like Protein 7
AHJ39967.2	DNA-dep. RNAP Subunit RPB9	AHJ40316.2	Hypothetical Protein
AHJ39993.2	Ubiquitin-Conjugating Enzyme e2	AHJ40318.2	Hypothetical Protein
AHJ40002.1	WD Repeat-Containing Protein	AHJ40319.1	Hypothetical Protein
AHJ40019.2	B-Type Lectin Protein	AHJ40329.1	Low Complexity Protein
AHJ40032.1	Protein Phosphatase 2c	AHJ40333.1	Hypothetical Protein
AHJ40038.1	Formamidopyrimidine-DNA Glycosylase	AHJ40337.1	Chemotaxis Protein
AHJ40051.1	Hypothetical Protein	AHJ40339.1	Hypothetical Protein
AHJ40056.1	Poly (A) Polymerase Catalytic Subunit	AHJ40340.1	Hypothetical Protein
AHJ40060.1	Hypothetical Protein	AHJ40341.2	Ubiquitin Thioesterase
AHJ40061.1	Hypothetical Protein	AHJ40367.2	Hypothetical Protein
AHJ40071.1	Thioredoxin Domain-Containing Protein	AHJ40371.2	Virion-Associated Membrane Protein
AHJ40083.1	mRNA-Capping Enzyme	AHJ40393.1	Lansterol 14-Alpha-Demethylase
AHJ40084.1	Putative FtsJ-Like Methyltransferase	AHJ40423.1	Hypothetical Protein
AHJ40087.2	Hypothetical Protein	AMK61745.1	Collagen Triple Helix Repeat Containing Protein
AHJ40093.1	Low Complexity Protein	AMK61776.1	Choline Dehydrogenase-Like Protein
AHJ40101.1	Core Protein	AMK61799.1	DNA Topoisomerase 1b
AHJ40107.2	Hypothetical Protein	AMK61800.1	Probable Glutaredoxin
AHJ40114.2	Capsid Protein 1	AMK61829.1	Hypothetical Protein
AHJ40128.1	Hypothetical Protein	AMK61837.1	Hypothetical Protein
AHJ40129.2	Thioredoxin Domain-Containing Protein	AMK61849.1	Hypothetical Protein
AHJ40139.1	Hypothetical Protein	AMK61856.1	Hypothetical Protein
AHJ40144.1	Hypothetical Protein	AMK61866.1	Regulator of Chromosome Condensation
AHJ40159.1	Hypothetical Protein	AMK61869.1	Thiol Protease
AHJ40160.2	Hypothetical Protein	AMK61892.1	Hypothetical Protein
AHJ40162.1	Hypothetical Protein	AMK61902.1	Hypothetical Protein
AHJ40169.1	Hypothetical Protein	AMK61903.1	Anaerobic Nitric Oxide Reductase Transcription Factor Regulator NorR
AHJ40172.1	DNA-dir. RNAP Subunit 1	AMK61918.1	Ankyrin Repeat Protein

Table 3.7 Identity of Proteins Released by SMBV or TV and Their Homologues. The identification of the proteins released from either SMBV or TV particles and the proteins from either of the virus genomes that were predicted to share sequence homology. Homology pairs are listed in Table 3.4 for pairs where both proteins were released and in Table 3.5 where only one of the proteins was released from either of the GV.

Table 3.7 (cont'd)

SMBV - Released			
Protein ID	***Protein***	***Protein ID***	***Protein***
AHJ40243.1	Hypothetical Protein	AMK61987.1	NHL Repeat-Containing Protein
AHJ40247.1	Mannose-6P Isomerase	AMK62013.1	Hypothetical Protein
AHJ40183.2	Hypothetical Protein	AMK61920.1	Hypothetical Protein
AHJ40190.1	Alpha/Beta Hydrolase/Esterase/Lipase	AMK61935.1	Hypothetical Protein
AHJ40211.1	Hypothetical Protein	AMK61942.1	Hypothetical Protein
AHJ40213.2	Hypothetical Protein	AMK61955.1	N-Acetyltransferase
AHJ40220.1	Hypothetical Protein	AMK61959.1	Prolyl 4-Hydroxylase
AHJ40230.1	Hypothetical Protein	AMK61968.1	Proline Rich Protein
AHJ40236.2	Thiol Oxidoreductase e10r	AMK61977.1	Hypothetical Protein
AHJ40254.1	Hypothetical Protein	AMK62059.1	Hypothetical Protein
AHJ40271.2	Hypothetical Protein	AMK62082.1	Hypothetical Protein
AHJ40278.1	Glycosyltransferase Family 10 Protein	AMK62096.1	Choline Dehydrogenase-Like Protein

SMBV - Not Released			
Protein ID	***Protein***	***Protein ID***	***Protein***
AHJ39847.1	Hypothetical Protein	AHJ40336.2	Endonuclease VIII-Like Protein
AHJ39850.1	Hypothetical Protein	AHJ40349.1	Hypothetical Protein
AHJ39852.1	Hypothetical Protein	AHJ40355.2	5'-3'-deoxribonucleotidase
AHJ39870.1	Hypothetical Protein	AHJ40372.1	Hypothetical Protein
AHJ39877.1	Virion-Associated Membrane Protein	AHJ40388.1	Hypothetical Protein
AHJ39883.1	Hypothetical Protein	AHJ40389.1	Hypothetical Protein
AHJ39887.1	Hypothetical Protein	AHJ40409.1	Hypothetical Protein
AHJ39889.1	Hypothetical Protein	AHJ40412.1	Glucose-Methanol-Choline Oxidoreductase
AHJ39897.2	Hypothetical Protein	AHJ40428.1	Hypothetical Protein
AHJ39939.2	Hypothetical Protein	AHJ40429.1	Hypothetical Protein
AHJ39945.1	Hypothetical Protein	AMK61705.1	Hypothetical Protein
AHJ39959.1	Hypothetical Protein	AMK61707.1	Ankyrin Repeat Protein
AHJ39981.1	DNA-dir. RNAP Subunit 5	AMK61735.1	WD Repeat Family Protein
AHJ39988.2	DNA-dir. RNAP Subunit 2	AMK61737.1	Ankyrin Repeat Protein
AHJ40000.1	Hypothetical Protein	AMK61744.1	Hypothetical Protein
AHJ4001.2	Putative Glycosyltransferase	AMK61751.1	Hypothetical Protein
AHJ40021.2	ATP=Dependent RNA Helicase	AMK61764.1	Hypothetical Protein
AHJ40024.1	Kinesin-Like Protein	AMK61775.1	GMC-Type Oxidoreductase
AHJ40046.2	Hypothetical Protein	AMK61801.1	Collagen-Like Protein 2
AHJ40063.1	Transcription Termination Factor	AMK61819.1	Collagen Triple Helix Repeat Containing Protein
AHJ40080.2	DNA-dir. RNAP Subunit	AMK61820.1	Collagen Triple Helix Repeat Containing Protein

Table 3.7 (cont'd)

SMBV - Not Released			
Protein ID	***Protein***	***Protein ID***	***Protein***
AHJ40095.2	Alpha/Beta Hydrolase	AMK61821.1	Collagen Triple Helix Repeat Containing Protein
AHJ40100.2	Hypothetical Protein	AMK61823.1	DNA-dir. RNAP subunit 2
AHJ40117.2	Hypothetical Protein	AMK61830.1	Heat Shock Protein 70-Like Protein
AHJ40126.1	Hypothetical Protein	AMK61841.1	Oxoglytarate Malate Carrier Protein
AHJ40127.2	Capsid Protein 4	AMK61914.1	Ankyrin Repeat Protein
AHJ40132.1	Hypothetical Protein	AMK61915.1	Ankyrin Repeat Protein
AHJ40141.2	Hypothetical Protein	AMK61919.1	PAN Domain-Containing Protein
AHJ40145.1	Ubiquitin-Conjugating Enzyme e2	AMK61929.1	Hypothetical Protein
AHJ40170.1	DNA-dir. RNAP Subunit 1	AMK61946.1	ATP-Dependent RNA Helicase
AHJ40191.2	5'-3' Exonuclease 20	AMK61957.1	Hypothetical Protein
AHJ40199.2	Hypothetical Protein	AMK61967.1	Hypothetical Protein
AHJ40201.1	ATP-dep. RNA Helicase	AMK61982.1	Hypothetical Protein
AHJ40209.1	Thioredoxin-Like Protein	AMK61984.1	Hypothetical Protein
AHJ40242.1	Hypothetical Protein	AMK61992.1	Hypothetical Protein
AHJ40248.1	Histone Demethylase	AMK62000.1	Methylated-DNA-Protein-Cysteine Methyltransferase
AHJ40268.1	Phosphatidylethanolamine-Binding Protein	AMK62004.1	Protein Kinase-Like Protein
AHJ40289.2	Collagen Triple Helix Repeat Containing Protein	AMK62030.1	Homeobox Protein
AHJ402919.2		AMK62036.1	F-Box Protein
AHJ40296.1	Hypothetical Protein	AMK62066.1	Ankyrin Repeat Protein

TV - Released			
Protein ID	**Protein**	**Protein ID**	**Protein**
AUL77474.1	Phosphatidylethanolamine-Binding Protein	AUL78135.1	Hypothetical Protein
AUL77600.1	Hypothetical Protein	AUL78143.1	Hypothetical Protein
AUL77661.1	mg709 Protein	AUL78147.1	Capsid Protein 1
AUL77680.1	Putative N-Acetyltransferase	AUL78191.1	Hypothetical Protein
AUL77688.1	Hypothetical Protein	AUL78211.1	Putative Pore Coat Assembly Factor
AUL77694.1	Hypothetical Protein	AUL78214.1	Hypothetical Protein
AUL77718.1	Hypothetical Protein	AUL78219.1	Hypothetical Protein
AUL77723.1	Hypothetical Protein	AUL78232.1	Hypothetical Protein
AUL77729.1	Putative ORFan	AUL78269.1	Arylsulfatase
AUL77752.1	Hypothetical Protein	AUL78287.1	Hypothetical Protein
AUL77820.1	Hypothetical Protein	AUL78288.1	Hypothetical Protein
AUL77829.1	Putative ATP-Dependent RNA Helicase	AUL78302.1	DNA-dir. RNAP subunit 1

Table 3.7 (cont'd)

TV - Released			
Protein ID	**Protein**	**Protein ID**	**Protein**
AUL77838.1	Kinesin-Like Protein	AUL78348.1	Hypothetical Protein
AUL77907.1	Hypothetical Protein	AUL78362.1	Intein-Containing DNA-dir. RNAP Subunit 2
AUL77930.1	Hypothetical Protein	AUL78368.1	DNA-dir. RNAP Subunit 6
AUL77936.1	Hypothetical Protein	AUL78400.1	Putative Fibril-Associated Protein
AUL77941.1	SNF2 Family Helicase	AUL78464.1	Hypothetical Protein
AUL77963.1	Thioredoxin Domain-Containing Protein	AUL78466.1	Hypothetical Protein
AUL78016.1	DNA-dir. RNAP Subunit	AUL78468.1	Hypothetical Protein
AUL78040.1	Ubiquitin Domain-Containing Protein	AUL78481.1	Hypothetical Protein
AUL78055.1	Hypothetical Protein	AUL78503.1	Cu-Zn Superoxide Dismutase
AUL78067.1	Hypothetical Protein	AUL78629.1	Putative Protein Kinase
AUL78082.1	Major Core Protein	AUL78630.1	Ig Family Protein
AUL78088.1	Putative ORFan	AUL78635.1	Putative ORFan
AUL78093.1	Hypothetical Protein	AUL78681.1	Hypothetical Protein
AUL78097.1	Catalase HPII	AUL78714.1	Mimivirus Elongation Factor Aef-2
AUL78134.1	Glyoxylase	AUL78724.1	Glutaredoxin

TV - Not Released			
Protein ID	**Protein**	**Protein ID**	**Protein**
AUL77471.1	Putative ORFan	AUL78192.1	Thioredoxin Domain-Containing Protein
AUL77475.1	Hypothetical Protein	AUL78208.1	Putative ORFan
AUL77477.1	Hypothetical Protein	AUL78251.1	Putative Virion-Associated Membrane Protein
AUL77492.1	Hypothetical Protein	AUL78278.1	Putative Ankyrin Repeat Protein
AUL77517.1	mg749 Protein	AUL78281.1	Putative PAN Domain-Containing Protein
AUL77531.1	Hypothetical Protein	AUL78282.1	Hypothetical Protein
AUL77553.1	Mannose-6P Isomerase	AUL78286.1	Hypothetical Protein
AUL77569.1	Hypothetical Protein	AUL78292.1	Hypothetical Protein
AUL77599.1	Putative Oxidoreductase	AUL78301.1	Hypothetical Protein
AUL77650.1	Putative ORFan	AUL78316.1	Hypothetical Protein
AUL77677.1	Hypothetical Protein	AUL78347.1	Hypothetical Protein
AUL77725.1	mg437 Protein	AUL78382.1	Amino Oxidase Family Protein
AUL77772.1	Hypothetical Protein	AUL78383.1	Putative DNA Topoisomerase 2 Isoform X2
AUL77824.1	Putative B-Type Lectin	AUL78403.1	Putative Major Capsid Protein
AUL77853.1	Hypothetical Protein	AUL78406.1	UDP-Glucose 4-Epimerase GalE

Table 3.7 (cont'd)

TV - Not Released			
Protein ID	**Protein**	**Protein ID**	**Protein**
AUL77856.1	Hypothetical Protein	AUL78470.1	Collagen Alpha-1(XXVII) Chain Flags Precursor
AUL77859.1	Putative Protein Phosphatase 2c	AUL78501.1	Putative Virion-Associated Membrane Protein
AUL77877.1	Formamidopyrimidine-DNA Glycosylase	AUL78505.1	Hypothetical Protein
AUL77884.1	Hypothetical Protein	AUL78531.1	Hypothetical Protein
AUL77896.1	Hypothetical Protein	AUL78561.1	Hypothetical Protein
AUL77903.1	Hypothetical Protein	AUL78575.1	Hypothetical Protein
AUL77929.1	Poly (A) Polymerase Catalytic Subunit	AUL78577.1	Hypothetical Protein
AUL77933.1	Hypothetical Protein	AUL78587.1	Tlr 6Fp Protein
AUL77949.1	Thiol Protease	AUL78600.1	Hypothetical Protein
AUL78031.1	mRNA Capping Enzyme	AUL78623.1	Hypothetical Protein
AUL78032.1	FtsJ-Like Methyltransferase	AUL78633.1	Hypothetical Protein
AUL78037.1	Putative ORFan	AUL78659.1	Hypothetical Protein
AUL78038.1	Hypothetical Protein	AUL78670.1	Putative Glutamine Amidotransferase-Like Protein
AUL78061.1	Hypothetical Protein	AUL78687.1	Putative Chemotaxis Protein CheD
AUL78068.1	Hypothetical Protein	AUL78702.1	Hypothetical Protein
AUL78070.1	Alpha/Beta Hydrolase Family Protein	AUL78707.1	Hypothetical Protein
AUL78092.1	Hypothetical Protein	AUL78710.1	Mimivirus Peptide Chain Elongation Factor eRF1
AUL78098.1	Endonuclease VIII-Like Protein	AUL78715.1	Putative ORFan
AUL78109.1	DNA Topoisomerase 1b	AUL78738.1	Helicase III/VV D5-Type ATPase N-Terminus
AUL78112.1	Hypothetical Protein	AUL78739.1	DNA-dep. RNAP Subunit 6

The majority of the proteins released for each virus (53% for SMBV, 55% for TV) are hypothetical proteins or proteins with unknown function. 17 of the proteins released by the two viruses displayed obvious homology between SMBV and TV (BLAST results or functional homology prediction). All of the released SMBV proteins predicted to be involved in both Translation and Replication had homologues amongst the released TV proteins. The proteins predicted to be involved in Transcription and Regulation, on the other hand, did not show any readily apparent homology. The homology between the released TV and SMBV proteins in general and within each category can be found in Figure 3.7.

Expected Protein Types are Released from Samba Virus and Tupanvirus Virions During Genome Release

GV need to carry out the same basic stages of the viral life cycle as their smaller cousins to replicate. Common stages include genome translocation into the host cell, blocking host replication, hijacking host machinery to make viral proteins, and making new viral proteins (*68*). Both SMBV and TV likely release proteins that are predicted to perform these functions, as many smaller viruses release whole proteins or peptides to facilitate this function (*126, 168, 230*). Hypothetical or unknown function proteins released from GV particles likely aid in performing these critical functions as many of them are released during the initial phase of opening. Aside from identifying the putative functions for the hypothetical proteins discussed below, determining the specific function of these proteins lies beyond the scope of this study.

Before the virus is able to hijack the host machinery and begin replication, it must enter the host cell and translocate its genome across the phagosomal membrane into the cytoplasm. SMBV releases putative membrane proteins, such as a virion associated membrane protein (AHJ40731.1) and as well as hypothetical proteins with predicted transmembrane domains that

may play a role in membrane fusion ("H/TM" in Table 3.2). Therefore, the results of this study help to assign putative roles to many proteins with previously unknown function, highlighting the power of this new method.

Additionally, both SMBV and TV release proteins predicted to play a role in an Ubiquitin-Proteasome degradation pathway (UPP, delineated by [c] in Table 1). These proteins are known to facilitate genome release in other viruses including the large, but not quite giant Iridoviruses (*223*) and Herpesviruses (*231*). In Iridovirus infection, the UPP is coupled with metabolic, cytoskeletal, macromolecule biosynthesis, and signal transduction proteins to facilitate infection (*223*). Proteins predicted to carry out these functions are released from both the SMBV and TV virions alongside the UPP-related proteins ([e] in Table 3.2).

Following genome translocation, the virus forces the cell machinery to transition from making new cellular products to making viral components. Both SMBV and TV release various subunits of a DNA-dependent RNA polymerase (SMBV: AHJ39967.2, AHJ40151.2, AHJ40172.1; TV: AUL78016.1, AUL78362.1, AUL78368.1, AUL78302.1). This series of proteins is critical for the lifecycle of the virus as it directs the cellular machinery of the host to recognize viral DNA *in lieu* of cellular DNA. These proteins, especially the various DNA-dependent RNA polymerases, may play a role in transcription as hypothesized to occur following stargate opening but before nucleocapsid release (*232*) Additional proteins in this category likely include some of the metabolic proteins released by the viruses, especially the catabolic proteins that may play a role in degrading host defenses and machinery. These proteins include a SMBV thiol protease (AMK61869.1), a SMBV amine oxidase (AHJ39955.1), and a hypothetical TV protein with a predicted inosine/uridine-favoring nucleoside hydrolase domain (AUL71835.1). Aside from these RNA polymerase subunits, both TV and SMBV release proteins that facilitate

transcription. SMBV releases a poly (A) polymerase (AHJ40056.1), an mRNA-capping enzyme (AHJ40083.1), and an anaerobic transcription regulator (AMK61903.1). TV releases an SNF2 family helicase (AUL77941.1), an ATP-dependent RNA helicase (AUL77829.1), and a mimivirus-like elongation factor (AUL78714.1).

Many of the proteins we identified matched proteins that one would expect to be released during the initial stages of viral infection and greatly supports our hypothesis that the *in vitro* stages generated in this study are reflective of those that occur *in vivo*. These data provide new insights into GV biology and ultimately lead to our proposed model (see next sections).

Samba Virus and Tupanvirus Also Release Novel Proteins During Stargate Opening

SMBV and TV also release proteins that are relatively uncommon amongst viruses. These proteins include metal-binding homeostasis proteins as well as chemotaxis-regulating proteins.

Our mass spectrometry data conclusively show that both SMBV and TV release proteins that are predicted to play a role in maintaining homeostasis (Figure 3.7E). Many of these proteins are predicted to have redox activity, protecting the virus and its cargo from reactive oxygen species (ROS) that can be found in the host phagosome (*215*). These proteins include several thioredoxin-like or thioredoxin domain-containing proteins (SMBV: AHJ40071.1, AHJ40129.2; TV: AUL77963.1) and glutaredoxins (SMBV: AMK618100.1; TV: AUL78724.1). TV releases a catalase protein (AUL78097.1) as well as glyoxylase (AUL78134.1) while SMBV releases a prolyl 4-hydroxylase (AMK61959.1). These proteins are also projected to protect the GV from ROS during the infection process. Here we show that these proteins are indeed released very early in the infection process.

Redox-active proteins are also thought to play an important role in protecting the viruses from the harsh conditions present in the host phagosome. During phagocytosis amoebal phagosomes drop to ~pH 4 (not low enough to trigger stargate opening), but they are also inundated with metals (like Cu and Zn) and reactive oxygen species (*216, 217*). Both viruses release metal-binding proteins (identified by [d] in Table 1) including SMBV's lanosterol demethylase (AHJ40393.1) --a cytochrome p450-like protein-- and prolyl 4-hydroxylase (AMK61959.1) and TV's mg709 (AUL77661.1) --a putative prolyl 4-hydroxylase with iron ion binding capabilities-- and Cu-Zn superoxide dismutase (AUL78503.1). It is likely that these proteins, in conjunction with the ROS-mitigating proteins described above, allow these viruses to survive the onslaught of low pH, high ROS, and high metal concentration found inside of the host phagosomes. We also note that the low pH of the phagosomes is similar to the low pH used in our in vitro assay, likely reflecting a physiologically relevant stage that describes GV infection mechanisms.

While *Tupanvirus* infection is hypothesized to occur through phagocytosis (*206*), no biological data has yet been provided to substantiate said hypothesis. This proposal stems from visualization of phagocytosis of TV by *Vermamoeba vermiformis* and subsequent TV stargate opening via thin-section TEM. Thin section TEM, embedding biological samples within epoxy resin then slicing thin, electron translucent sections off of the block, is prone to structural artifacts (*68*). Therefore, it is critical that any hypotheses generated from thin section TEM imaging are supported by data from another technique. The release of proteins capable of mitigating the harsh environment of the amoebal phagosome provides biological evidence to support this hypothesis.

SMBV and TV also contain proteins that are predicted to regulate chemotaxis. SMBV releases a chemotaxis protein (AHJ40337.1) that shares homology with the putative chemotaxis protein CheD found in mimivirus (AKI80461.1) and TV (AUL78687.1). CheD proteins regulate chemotaxis via deamidation of chemotaxis receptors (InterPro). TV has been shown to shut down host chemotaxis (*92, 233*) and it is likely that these CheD-like chemotaxis regulation proteins are involved in this process. While TV does contain a CheD-like chemotaxis protein that was identified in the total virion MS data, this protein was not present following low pH treatment.

Making Some Sense of the Myriad Hypothetical Proteins in the Samba Virus and Tupanvirus Proteome

Of the 356 proteins identified in the total virion MS for both SMBV and TV, ~52% (Figure 3.7B) were annotated as being hypothetical proteins, low complexity proteins (SMBV), or ORFans (an open reading frame that is not found in other reported genomes). In SMBV, 77 of these proteins were released (46 proteins) and in TV 31 proteins are released following low pH treatment. As these proteins are released from the GV virion during the initial stages of the genome release process, we hypothesize that these proteins play a role in either the infection process (phagosome survival, membrane fusion, etc.) or in the beginning stages of replication. Hypothetical proteins with additional functional information predicted via BLAST, HHBLIST, PSIPRED, or InterPro are listed in Table 3.8. Interestingly, only four of the hypothetical proteins released by the two GV share homology when analyzed via BLAST and HHBLITS, suggesting that while related, SMBV and TV have significant evolutionary divergence.

Table 3.8

Virus	*Accession Number*	Structural/Functional Prediction	Structural/Functional Prediction
SMBV	*AHJ4005.1.1*	TM Helix	
SMBV	*AHJ40159.1*	TM Helix	
SMBV	*AHJ40213.2*	Methyl-Accepting Chemotaxis Protein	
SMBV	*AHJ40326.1*	Alpha L Rhamnidose Domain	
SMBV	*AHJ40333.1*	Beta galactosidase Domain	SpoIID/LytB Domain
SMBV	*AHJ40333.1*	TM Helix	
SMBV	*AHJ40367.2*	TM Helix	Apple Domain (Proteolysis)
SMBV	*AHJ40423.1*	Crp/Fnr Family Transcription Regulator	
SMBV	*AMK61829.1*	Coiled-Coil Domain-Containing Protein 180-like Isoform	
SMBV	*AMK61849.1*	TM Helix	
SMBV	*AMK61920.1*	Ankyrin Repeat Protein	
SMBV	*AMK61942.1*	Alpha/Beta Hydrolase	Zn-Finger Protein
SMBV	*AMK61942.1*	TM Helix	
SMBV	*AMK62013.1*	Coiled-Coil and C2 Domain-Containing Protein 1-like Isoform	
SMBV	*AMK62059.1*	LamG Superfamily (Incomplete Domain)	
TV	*AUL77600.1*	Alpha/Beta Hydrolase	
TV	*AUL77688.1*	TM Helix	
TV	*AUL77718.1*	Apple Domain (Proteolysis)	TM Helix
TV	*AUL77723.1*	Membrane Helix	
TV	*AUL77930.1*	Sugar O-Acetyltransferase	Outer Dense Fiber Protein 3-B-like Protein
TV	*AUL78055.1*	ATPase	Ion Channel
TV	*AUL78135.1*	(Ribo)Nuclease Hydrolase	Inosine-uridine Nucleoside Hydrolase
TV	*AUL78143.1*	Nuclear Transport Family 2 Protein	Sgc/EcaC Family Oxidoreductase
TV	*AUL78288.1*	TM Helix	
TV	*AUL78348.1*	MC1 Domain (DNA Protection)	Ubiquitin-Conjugating Enzyme
TV	*AUL78681.1*	Apple Domain (Proteolysis)	

Table 3.8 SMBV and TV Released Hypothetical Proteins with Predicted Functionalities. Functional predictions of SMBV and TV hypothetical proteins from BLAST, HHBLITS, PSIPRED, or InterPro Analysis.

Opening the Stargate to New Avenues of Giant Virus Research

By modulating temperature and pH we were able to mimic four unique, and metastable stages of the GV genome release process (Figure 3.9). GV particles that mimic these genome release stages have been seen in previous experiments (*3, 96, 151, 176*), although previous visualization of these particles relied on finding the "one-in-a-million" particle in the correct state. We are now able to mimic GV genome release stages reliably and with high frequency. Additionally, these conditions forgo the need to synchronize infection and trap GV particles in phagosomes at very specific times to generate the condition of interest. Eschewing the host cell may limit specific avenues of study, such as searching for a host receptor(s), but it dramatically simplifies any studies aimed at the virus and the changes it undergoes during the genome release process.

Figure 3.9

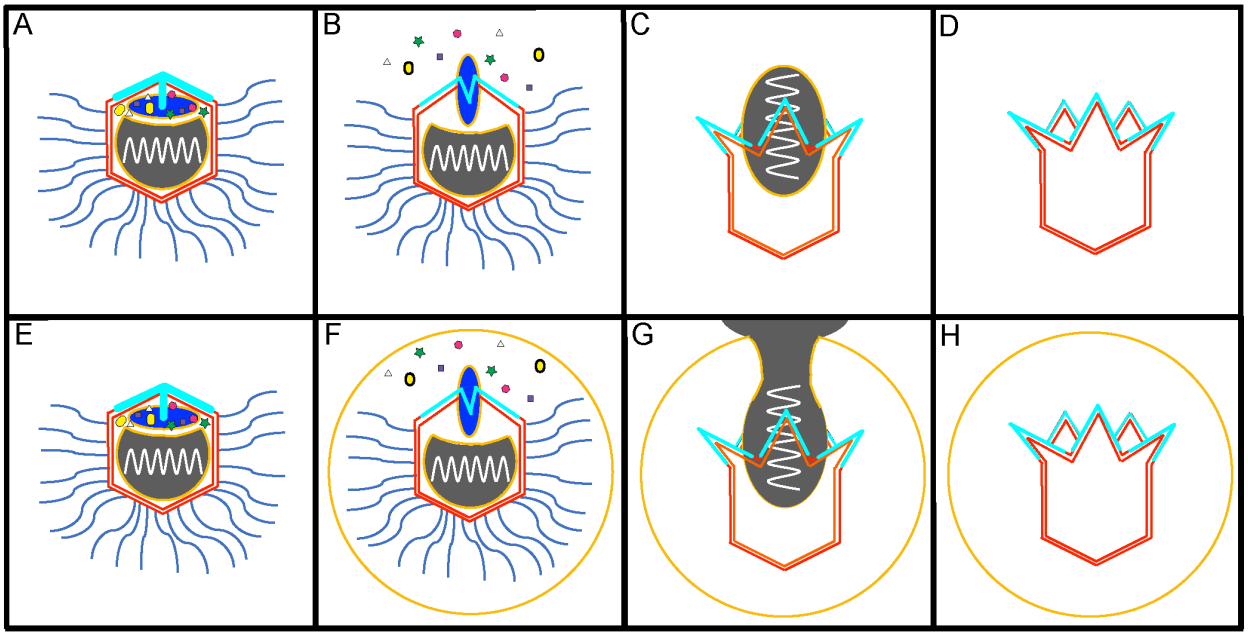

Figure 3.9 Cartoon Model of GV Genome Release Stages. Schematic of at least four distinct stages of the GV genome release process as identified in this study. A) Native, intact virions. B) Disruption of the Starfish Seal (Initiation of Infection): Particles with stargate vertices that are beginning to open. C) Nucleocapsid Release: Particles with fully open stargate vertices that are in the process of releasing the nucleocapsid from the capsid. D) Fully Released (Completion): Particles that have completed the genome release process. Coloration matches scheme used in Figure 2. The top row depicts particle states as induced *in vitro*. The bottom row corresponds to analogous structures seen in thin section micrographs of infected cells (*83*). The orange circles in panels F, G, and H correspond to the phagosomal membrane.

Additionally, we have identified proteins that are released during the initial stages of infection in two GV, SMBV and TV. Over half of the proteins released by these viruses are annotated as hypothetical, low complexity, or as an ORFan. We were able to provide functional predictions for some of these proteins through homology. Even so, an exact functional determination of these proteins remains elusive. The release of these proteins at the initiation of stargate opening suggests that these proteins play an important role in the early stages of GV infection (phagosome survival, genome translocation, early transcription, host defense suppression, etc.). The exact functions of these proteins, as well as how their interactions mediate and orchestrate GV infection, are prime candidates for future study. The importance of these potential future studies is enhanced by the fact that many GV appear to share similar strategies for genome release. All four of the GV tested in this study responded to the treatment conditions, suggesting that these GV utilize similar molecular forces during genome release, and likely similar proteins to counteract these forces.

ACKNOWLEDGEMENTS

The authors would like to thank the MSU Proteomics and Mass Spec. Facilities, especially Drs. D. Jones, C. Wilkerson, and D. Whiten, for their assistance with MS experiments. Additionally, we thank Drs. W. Jiang, T. Klose, and V. Bowman at Purdue University's Midwest Cryo-EM Consortium (NIH Consortium #U24GM116789-03). We would also like to thank C. Flegler at the MSU Center for Advanced Microscopy for her expertise with SEM experiments. The MSU High Performance Computation Cluster (HPCC) provided computational tools and support for cryo-EM image motion correction. Dr. K. Padmanabhan and Dr. M. Feig provided additional assistance with computational resources and expert consultation in protein homology and functional predictions. Funding for this project was provided by the AAAS Marion Mason Milligan award for Women in the Chemical Sciences (KNP), the JK Billman, Jr., MD Endowed Research Professorhip (KNP), Funding was provided by Fundo de Amparo à Pesquisa do Estado do Rio de Janeiro e Conselho Nacional de Pesquisa (JRC), and the Burroughs Wellcome Fund (KNP). JRS has been supported by the Jack Throck Waston Fellowship and the August and Ernest Frey Research Fellowship from MSU and NIH R01 GM110185 (KNP). JSA has been supported by CNPq, CAPES, MS, and FAPEMIG. Nvidia provided GPU support for cryo-EM and cryo-ET image processing.

AUTHOR CONTRIBUTIONS

JRS: Designed and carried out experiments, wrote the manuscript, data analysis

JRC: Viral sample preparation and manuscript review, data analysis

JSA: manuscript review

KNP: Designed and carried out experiments, manuscript review, data analysis

STAR METHODS

Contact for Reagent and Resource Sharing

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Kristin Parent (kparent@msu.edu).

Experimental Model and Subject Details

Acanthamoeba castellanii

Acanthamoeba castellanii cells were purchased from ATCC (ATCC 30010). *Acanthamoeba castellanii* (ATCC 30010) was cultivated in 712 PYG media w/Additives (ATCC recipe) at pH 6.5 in the presence of gentamicin (15 μg/mL) and penicillin/streptomycin (100 U/mL) at 28 °C to reach a 90% confluence.

Giant Viruses

Tupanvirus soda lake (TV), Antarctica virus, and Samba virus (SMBV) were isolated previously (*3, 4, 95*). M4 virus was kindly provided by Dr. Bernard La Scola and Dr. Thomas Klose (*64*). *Acanthamoeba castellanii* (ATCC 30010) was cultivated in 712 PYG media w/ Additives (ATCC recipe) at pH 6.5 in the presence of gentamicin (15 μg/mL) and penicillin/streptomycin (100 U/mL) at 28 °C to reach a 90% confluence. SMBV or TV virions were diluted in phosphate buffered saline (PBS) and added to the cells to a multiplicity of infection of 5 (TV) or 10 (SMBV). An initial incubation was carried out for one hour at room temperature. After the initial incubation, additional PYG media was added to the cells and the flasks were incubated at 28 °C for 48 hours. After 48 hours, more of the free amoebal cells had been lysed. Suspensions containing cell debris and cell particles were centrifuged at 900 x g to

pellet residual cells. The resulting supernatant was filtered using a 2 μm filter and was immediately applied to a 22% sucrose cushion (w/w) at 15,000 x g for 30 min. Viral pellets were resuspended in PBS and stored at -80 °C. Viruses were tittered using the Reed-Muench protocol (*234*). On average, virus isolation yielded 10^{10} $TCID_{50}$/mL (TCID = tissue culture infective dose).

Method Details

Treatment of SMBV Particles and Image Analysis

Determining the Percentage of Open SMBV Particles

For all treatments, the percentage of open SMBV particles (POP) was determined via single particle cryo-electron microscopy. These percentages were compared to the native (untreated) level of spontaneous SMBV particle opening, determined previously to be ~5% (*96*).

Conditions That Did Not Increase POP

SMBV particles were treated with various conditions that have been shown to disrupt/destroy other viruses. These conditions include urea, guanidinium hydrochloride, DMSO, Triton X-100, chloroform, DNase I, and an enzyme cocktail (lysozyme, bromelain, proteinase K) that was previously shown to remove APMV fibers (*187*). Treatments were applied for 1-2 hours prior to POP determination via cryo-EM. Concentrations for the various conditions, as well as the resultant POP values, can be found in Table 3.1.

pH Titration of SMBV Particles

25-50 μL of SMBV particles were added to Millipore VSWP Membrane Filter dialysis discs (0.025 μm cutoff) which were then floated onto ~25 mL of 20 mM sodium phosphate buffer, adjusted to the desired pH. The samples were allowed to equilibrate for 1.5-2 hours. For conditions where low pH would interfere with additional treatment (*e.g.* pH 2 + DNase I or pH 2 samples submitted for mass spectrometry) the particles were dialyzed for an additional 1.5-2 hours against pH 7.0 buffer to restore neutral pH.

High Temperature Incubation

GV particles were incubated in a BioRad T100 thermal cycler at 80, 89, and 100 °C for 1 hour. SMBV particles remained intact following 1 hours at 100 °C, so additional incubations at 100 °C were performed at 2, 3, or 6 hours. As a control, SMBV particles were also incubated for 1 hour at room temperature (25 °C).

Combining High Temperature and Low pH

To determine the effect of combining low pH and high temperature, GV particles were sequentially treated with pH 2 and 100 °C. First, SMBV particles were dialyzed against 20 mM sodium phosphate buffer, adjusted to pH 2, for 2 hours. Following dialysis, SMBV particles were incubated at 100 °C for 3 hours.

Cryo-Electron Microscopy (Cryo-EM) and Cryo-Electron Tomography (Cryo-ET)

Sample Preparation

Samples for cryo-EM and cryo-ET were prepared as described previously (*96*). Briefly, small (3-5 μL) aliquots of virus particles were applied to R2/2 (cryo-EM) or R 3.5/1 (cryo-ET) Quantifoil grids (Electron Microscopy Solutions) that had been plasma cleaned for 20 seconds in a Fischione model 1020 plasma cleaner. Prior to virus addition, 5-10 μL of 10 nm nanogold fiducial markers were applied to the R3.5/1 grids and were air dried to provide markers for fiducial alignment of the tilt series. The samples were plunge frozen in liquid ethane using a manual plunge-freezing device (Michigan State University Physics Machine Shop). Frozen-hydrated samples were stored, transferred, and imaged under liquid nitrogen temperatures.

Single Particle Cryo-Electron Microscopy

Single particle cryo-EM experiments were performed at Michigan State University. Virus particles were imaged in a JEOL 2200-FS TEM operating at 200 keV, using low dose conditions controlled by SerialEM (version 3.5.0-beta, (*167*)) with the use of an in-column Omega Energy Filter operating at a slit width of 35 eV. Micrographs were recorded at 25 frames per second using a Direct Electron DE-20 direct detector, cooled to -38 °C. Motion correction was performed using the Direct Electron software package (Direct Electron, LLC). Micrographs were collected between 8,000 and 10,000 X nominal magnification (6.87 and 5.30 Å/pixel, respectively). The objective lens defocus settings ranged from 10 to 15 μm underfocus. Micrographs were collected for 5 seconds, resulting in a total dose of ~35 e-/Å^2. For bubblegram imaging, the SMBV particles were imaged for an additional four exposures, resulting in a total dose of ~140 e-/Å^2

Cryo-Electron Tomography

Cryo-ET tilt series were collected using a Titan Krios TEM operating at 300 keV with a post-column GIF (20 eV slit width) under low dose conditions controlled by SerialEM or Leginon at Purdue University. Images were collected using a Gatan K2 direct electron detector operating at 100 milliseconds/frame. Images were collected in super resolution mode between 33,000 and 53,000 X nominal magnification (2.12 - 1.33 Å/pixel). Tilt series were carried out between +/- 50 ° with bidirectional image collection every 2 °. Images were collected for 5 seconds, resulting in ~2.5 electrons/$Å^2$ per tilt image (~125 electrons/$Å^2$, total exposure dose).

Individual micrographs were corrected for particle motion and binned by a factor of two using MotionCor2 v1.2.0 (*164*) and the corrected images were stitched back into a tilt series using the Newstack functionality in IMOD (*235*). Tilt series alignment, using fiducial markers, and tomogram generation was carried out using IMOD v4.7.5. Final tomogram volumes were generated using ten iterations of the SIRT reconstruction method (*166*) then filtered using the smooth (3x3 kernel) and median (size 3) options in IMOD. Select tomograms were annotated using Amira v2019.2 (ThermoFisher Scientific).

Scanning Electron Microscopy

SEM Preparation and Imaging

GV particles were imaged using a JEOL JSM-7500F scanning electron microscope. Prior to imaging, virus particles were desiccated using an EM CPD300 critical point dryer, fixed with glutaraldehyde onto poly-L-Lysine treated SEM slides, and sputter coated with a ~2.7nm layer of iridium using a Q150T Turbo Pumped Coater. Particles were imaged between 8,500 X and 85,000 X nominal magnification.

Differential Mass Spectrometry

Sample Preparation

SMBV and TV particles were dialyzed against 20 mM sodium phosphate buffer, adjusted to pH 2, for 2 hours, as described above. An aliquot of each virus was left undialyzed as a control (Material Applied, MA). Following dialysis, proteins that had been released from the viral particles were separated from the virions via centrifugation in a microcentrifuge at 8,000 x g for 15 minutes. Visible viral pellets were resuspended in the same volume as the supernatant using 20 mM sodium phosphate buffer, adjusted to pH 7.0. Two technical replicates were created for each sample. An aliquot of each sample was used for SDS-PAGE.

Each sample was TCA precipitated and submitted for LC/MS/MS analysis to the MSU Proteomics Core. Prior to submission, samples were run on a 15% polyacrylamide gel at a voltage of 200 V for 45 minutes. TV and SMBV gel bands visible by Coomassie blue stain were excised and submitted for MS analysis as well.

Proteolytic Digestion

TCA precipitated pellets were re-suspended in 270uL of 100mM ammonium bicarbonate supplemented with 10% trifluoroethanol. Samples were reduced and alkylated by adding TCEP and Iodoacetamide at 10mM and 40mM, respectively and incubating for 5min at 45C with shaking at 1400 rpm in an Eppendorf ThermoMixer. Trypsin, in 100mM ammonium bicarbonate, was added at a 1:100 ratio (wt/wt) and the mixture was incubated at 37C overnight. Final volume of each digest was ~300uL. After digestion, the samples were acidified to 2% TFA and subjected to C18 solid phase clean up using StageTips (*236*) to remove salts.

LC/MS/MS and Data Analysis

An injection of 5uL was automatically made using a Thermo EASYnLC 1200 onto a Thermo Acclaim PepMap RSLC 0.075mm x 20mm C18 trapping column and washed for ~5min with buffer A. Bound peptides were then eluted over 95min with a gradient of 8%B to 42%B in 84min, ramping to 100%B at 85min and held at 100%B for the duration of the run (Buffer A = 99.9% Water/0.1% Formic Acid, Buffer B = 80% Acetonitrile/0.1% Formic Acid/19.9% Water) at a constant flow rate of 300nl/min. Column temperature was maintained at a constant temperature of 50 °C using and integrated column oven (PRSO-V2, Sonation GmbH, Biberach, Germany). Eluted peptides were sprayed into a ThermoScientific Q-Exactive HF-X mass spectrometer using a FlexSpray spray ion source. Survey scans were taken in the Orbi trap (60,000 resolution, determined at m/z 200) and the top ten ions in each survey scan are then subjected to automatic higher energy collision induced dissociation (HCD) with fragment spectra acquired at 7,500 resolution. The resulting MS/MS spectra are converted to peak lists using MaxQuant v1.6.0.1 (*237*) and searched using the Andromeda (*238*) algorithm against a protein database containing sequences from SMBV or TV and *Acanthamoeba castellanii* (each downloaded from NCBI, www.ncbi.nlm.nih.gov). Common laboratory contaminants were included in the Andromeda search. Protein and peptide FDR for all searches were set to 1%.

Mass Spectrometry Data Synthesis

The percentage of the total LFQ signal each protein was responsible for in each sample was calculated by dividing the individual protein LFQ signal by the total LFQ signal for the sample, excluding contaminates. Proteins that are released from the viral particles are expected to make up a higher percentage of the supernatant sample than the whole virion (MA), so the

ratios of these two percentages were calculated (Table 3.3, Table S3.4). Proteins with a supernatant:MA ratio > 1 were selected for further analysis.

Classification/Functional Annotation of Proteins Identified via MS

TV and SMBV proteins released at low pH were classified via their predicted functions and domains. Primary functional annotation had been carried out previously for both TV (*3*) and SMBV (*95*). Additional functional prediction, as well as homology prediction between the two viruses, was carried out through the use of the NCBI BLAST database (NCBI) as well as the HHBLITS server (*224*) and the InterPro database (*225*). Domain prediction was carried out by searching the InterPro database and utilizing the PSIPRED server (*226*) with the DISOPRED3 (*227*) functionality activated.

Quantification and Statistical Analysis

Mass Spectrometry Analysis

LFQ intensities for both SMBV and TV spectra were detected in triplicate. For each virus, the initial run did not produce high quality data so these intensities were disregarded. LFQ intensities from the remaining two runs were averaged together to produce the reported intensity (Table 3.3).

Data and Software Availability

Three-dimensional tomograms have been deposited to the Electron Microscopy Database (EMDB) under the ID codes EMD-20747 (pH 2, Movie S4), EMD-20748 (100 °C, Movie S5), EMD-20746 (pH 2 + 100 °C, Movie S6), and EMD-20745 (pH 2 + 100 °C, Movie S9).

SUPPLEMENTARY MATERIALS

Supplementary Video 2: Untreated SMBV Bubblegram Imaging. Bubblegram image series of a native SMBV particle demonstrating the buildup of radiation damage over time. A clear star-shaped radiation damage pattern is observed around the 11:00 position on the particle. Each frame represents a two second exposure (14 e-/Å^2). Total exposure time = 24 seconds (~140 e-/Å^2). Related to Figure 3.1.

Supplementary Video 3: Untreated SMBV Tomogram. Slice-by-slice view of a tomogram of a native SMBV particle. Related to Figure 3.2B-C.

Supplementary Video 4: Low pH-Treated SMBV Tomogram. Slice-by-slice view of a tomogram of a pH 2-treated SMBV particle. Note the opening in the stargate vertex as well as the sac exiting the capsid. Related to Figure 3.2F-G.

Supplementary Video 5: Tomogram of SMBV Incubated at High Temperature. Slice-by-slice view of a tomogram from an SMBV particle incubated at 100 °C for 6 hours. Note the fully open stargate vertex, the exodus of the nucleocapsid, and the apparent tethers between the capsid and the nucleocapsid. Related to Figure 3.2J-K.

Supplementary Video 6: Tilt Series of High Temperature Incubated SMBV. Tilt series of an SMBV particle incubated at 100 °C. Tilts were acquired every 2 degrees ranging from +/- 50 degrees. Related to Figure 3.2J-K

Supplementary Video 7: Low pH and High Temperature-Treated SMBV Tomogram. Slice-by-slice view of a tomogram of an SMBV particle treated with both low pH and high temperature. Tomogram segmentation was carried out using Amira v2019.2. Colors represent the following: Red- Outer Capsid Layer, Orange- Inner Capsid Layer, Blue- Starfish Seal Complex, and Yellow- Lipid. Note the flexibility of the innermost capsid layer and the residual density within the capsid interior. Related to Figure 3.2N-O.

Supplementary Video 8: Low pH and High Temperature Treated SMBV Tilt Series. Tilt series of an SMBV particle treated with both pH 2 and 100 °C. Tilts were acquired every 2° ranging from +/- 50°degrees. Related to Figure 3.2N-O.

Supplementary Video 9: Low pH and High Temperature-Treated SMBV Tomogram. Slice-by-slice view of a tomogram of five SMBV particles treated with both low pH and high temperature. These particles all have open stargate vertices, and one is oriented in a top-down view, providing additional structural information about the SMBV particle.

Supplementary Video 10: Low pH and High Temperature-Treated SMBV Tilt Series. Tilt series of an SMBV particle treated with both pH 2 and 100 °C. Tilts were acquired every 2° ranging from +/- 50°degrees. Five distinct SMBV particles are visible within this tilt series.

CHAPTER 4

CONCLUSIONS AND FUTURE DIRECTIONS

SIGNIFICANCE

Newly discovered (*79*), giant viruses (GV) represent an understudied and ever-expanding segment of virology. These viruses are ubiquitous (*4*) and possess incredible capsid stability, allowing some GV to survive for millennia frozen in ice (*62, 63*). While this capsid stability aids GV in surviving extreme environments (*3, 4, 63, 204*), it also represents an energetic barrier that the viruses must overcome during infection. Many GV contain a unique capsid vertex that opens to facilitate genome release (*83, 87*). To prevent premature genome loss, these vertices are sealed prior to infection by one of two structurally conserved seal complexes; inter-capsid cork-like seals or external starfish-shaped seal complexes.

This structural conservation provides an opportunity to use model systems to study the GV lifecycle, especially the GV genome release process. Towards this end, we used Samba virus (SMBV), a Brazilian GV, as a model system to generate novel data on the GV genome release process. This work establishes a procedure for visualizing SMBV for structural biology studies and demonstrates that SMBV does, in fact, contain a starfish-shaped external seal complex. This work also characterizes the molecular forces that stabilize the SMBV starfish seal complex and demonstrates that these molecular interactions are conserved across *Mimiviridae*. By manipulating these forces, we were able to generate the first *in vitro* system for studying GV genome release and were able to identify proteins released from both SMBV and Tupanvirus virions during the initiation of infection. In answering these questions, my Thesis has generated some of the first molecular information surrounding the GV genome release process and even identified proteins that are critical for the early stages of GV infection. The key findings from each chapter are highlighted below.

SUMMARY

Chapter 2: Microscopic Characterization of the Brazilian Giant Samba Virus

The work presented in Chapter 2 of this Thesis established a procedure for imaging SMBV particles at Michigan State University (MSU) and used this procedure to answer outstanding structural and biological SMBV questions. A summary of this procedure and the questions answered in this chapter can be found below.

How Does One Visualize a Biological Entity as Large as SMBV?

To answer questions on the biology and structure of SMBV, we established a procedure to visualize these particles through various microscopic techniques (*96*). These techniques included cryo-electron microscopy (cryo-EM), cryo-electron tomography (cryo-ET), scanning electron microscopy (SEM), and fluorescence microscopy. SMBV particles are incredibly large (~850 nm capsid diameters) and this size proves challenging for transmission electron microscopy (TEM). As the size of the specimen increases, the percentage of transmitted electrons decreases, lowering the signal-to-noise ratio of the micrographs (*68*). The JEOL 2200-FS at MSU is equipped with an omega energy filter (JEOL USA, Inc., Peabody, MA) as well as a DE-20 direct electron detector (Direct Electron, LP, San Diego, CA). Using these two pieces of equipment we were able to visualize SMBV particles through TEM and generate structural data on this GV.

Does SMBV Utilize a Stargate Vertex to Facilitate Genome Release?

Previous structural studies of SMBV were performed using negatively stained thin section TEM (*95*). This technique is prone to structural artifacts (*68*). Using the imaging procedure described above, we were able to visualize SMBV particles in a near-native state. Through cryo-EM and cryo-ET we were able to confirm the presence of an SMBV stargate vertex (through the presence of spontaneously opened SMBV particles) and its corresponding starfish-shaped seal complex (through cryo-ET and free-floating seal complexes).

How Structurally Similar are SMBV and APMV?

We also compared the structural features of SMBV with mimivirus (APMV). Through this comparison we identified structural differences between the two viruses, with the largest difference being the increased heterogeneity of the SMBV particles. Using SEM and fluorescence imaging, we were able to observe the ultrastructural behavior of these viruses. SMBV particles formed an ordered hexagonal lattice when observed through both SEM and fluorescence, whereas APMV particles appeared to form disordered aggregates. As we were using a fiberless APMV variant (M4 (*64*)) and were not able to remove the fibers from the SMBV particles, any differences in ultrastructure may simply be due to the presence/absence of the external fiber layer.

Chapter 3: Boiling Acid Mimics Intracellular Giant Virus Genome Release

The work presented in Chapter 3 of this Thesis investigated the mechanism and molecular constituents of the GV genome release process. First, the molecular forces responsible for maintaining SMBV starfish seal complex stability were determined. Next these forces were disrupted in other icosahedral GV, demonstrating that the forces responsible for starfish seal disruption are conserved across *Mimiviridae*. By modulating these molecular forces, we were able to mimic four distinct stages of the GV genome release process and determine the fate of the external seal complex during GV genome release. Finally, we were able to identify and compare the proteins released from the SMBV and Tupanvirus soda lake (TV) capsids at the initiation of infection. All told, the work presented in this chapter represents the first *in vitro* GV genome release system and provides novel molecular data concerning the forces and proteins governing this process. A summary of these procedures and the individual questions answered in this chapter can be found below.

What Molecular Forces Promote SMBV Starfish Seal Complex Stability?

To determine which molecular forces are responsible for SMBV starfish seal complex stability and subsequent stargate vertex opening, we treated SMBV particles with conditions that have been shown previously to disrupt the capsids of other viruses. These conditions included a pH range, a temperature range, urea (9M), guanidinium hydrochloride (6M), the organic solvents DMSO and chloroform, Triton X-100, DNase I, and an enzyme cocktail of bromelain, proteinase K, and lysozyme. The majority of these conditions had no effect on SMBV particles, however low pH and high temperature did result in dramatic increases in the percentage of open SMBV particles. These conditions disrupt electrostatic interactions (low pH) and increase the entropy in

the system (high temperature), suggesting that these forces play a role in the stability of the SMBV starfish seal complex.

Are These Molecular Forces Conserved Across Mimiviridae?

To determine the conservation of the seal complex-stabilizing forces amongst *Mimiviridae* we treated APMV, Antarctica virus, and TV with low pH and high temperature. Under these conditions, the GV particles all demonstrated open capsids, suggesting that the forces stabilizing the external seal complex are conserved across *Mimiviridae*.

Which Stages of the GV Genome Release Process can be Mimicked In Vitro*?*

By modulating disruption of the stabilizing forces mentioned above we were able to generate four distinct stages of the SMBV genome release process. By not disrupting either force, electrostatic interactions or system entropy, we were able to generate SMBV particles in a **Pre-Release** (native) state. By disrupting the electrostatic interactions (low pH treatment) we were able to disrupt the starfish seal complex but not completely open the stargate vertex. In this state, SMBV particles had small cracks in their capsids through which the extra membrane sac and any free-floating proteins were able to escape the capsid, mimicking the **Initiation of Infection**. By increasing the entropy in the system (high temperature incubation) we were able to again disrupt the starfish seal complex. Unlike low pH, these SMBV particles had completely opened their stargate vertices, facilitating release of the nucleocapsid and mimicking the **Nucleocapsid Release** stage of SMBV infection. By both increasing entropy and disrupting

electrostatic interactions we were able to generate fully empty particles. These particles had completely released their nucleocapsid and mimicked the **Completion of Genome Release**.

What is the Fate of the External Seal Complex?

Although it was known that the external seal complex must be disrupted to facilitate GV genome release, the ultimate fate of this complex was unknown. This complex could leave the capsid *en masse*, evidenced by the presence of free-floating starfish seal complexes in both APMV (*151*) and SMBV (*96*), or it could unzip and remain attached to the capsid. To determine the fate of the SMBV starfish seal complex we imaged each of the *in vitro* genome release stages (described above) through SEM. Low pH and high temperature (individually) treated samples did not provide definitive evidence for the fate of the seal complex. SMBV particles that had completed genome release, however, demonstrated that the external seal complex remains attached to the capsid following genome release. APMV and Antarctica virus seal complexes also remained attached to the capsid whereas the TV seal complex did not. These data suggest that the fate of the external seal complex may be conserved amongst lineage A *Mimiviridae* but not necessarily amongst *Mimiviridae* as a whole.

Which Proteins are Released From SMBV and TV Capsids at the Initiation of Infection?

With the ability to mimic distinct stages of the GV genome release process we have developed a system capable of determining the proteins that are released from the GV capsid at each stage. As a proof of principle we used mass spectrometry to determine the proteins released from both SMBV and TV virions at the Initiation of Infection (low pH treatment). Through this

analysis we identified 86 SMBV proteins and 56 TV released proteins. These proteins run the gamut of functional roles including membrane fusion, virion protection inside of the phagosome, regulation of host processes, and catabolism. We then compared the proteins released by the two viruses through sequential as well as predicted functional and structural homology. Only 17 of the released proteins shared predicted homology. This work provides some of the only data available concerning the proteins released from GV capsids during infection.

CONCLUSIONS

Through modulation of two molecular forces, entropy and electrostatic interactions, we were able to mimic four distinct stages of the GV genome release process *in vitro*; **Pre-Release**, **Initiation of Infection**, **Nucleocapsid Release**, and **Completion**. **Pre-Release** particles represent GV virions that have yet to encounter a host cell or virions that have been phagocytosed but have not yet initiated seal complex disruption. *In vivo*, the proteins or forces that trigger seal complex disruption remain unknown, but this process typically occurs 1-3 hours post infection (*92*). Particles trapped at the **Initiation of Infection** (low pH treated particles) represent GV particles that have been phagocytosed by the host cell and have triggered genome release. Virions in the **Nucleocapsid Release** stage (high temperature incubation) represent GV particles that have opened their stargate vertices, facilitating nucleocapsid exodus. Inside of the phagosome the nucleocapsid would be fusing with the phagosomal membrane at this stage of infection, releasing the genome into the cytoplasm. Finally, particles that have reached the **Completion** of the genome release process (low pH and high temperature treatment) mimic empty particles that remain in the phagosome following nucleocapsid fusion. The fate of these particles *in vivo* has not been studied.

Through the work presented here, combined with previous studies on other GV, we are able to propose a model for the GV genome release process (Figure 3.8). Following phagocytosis, GV particles unzip their external seal complexes, facilitating release of the extra membrane sac along with ~50-90 proteins, depending on the virus. We hypothesize that some of these proteins help protect the nucleocapsid from the harsh conditions of the host phagosome. Following seal complex disruption, the stargate vertex opens and the nucleocapsid fuses with the phagosomal membrane. We also hypothesize that some of the proteins released at the initiation

of infection may play a role in this fusion event as many of these proteins contain predicted transmembrane domains. As the nucleocapsid fuses with the phagosomal membrane the genome is released into the host cytoplasm where a viral factory forms and GV progeny are produced.

FUTURE DIRECTIONS

Future directions to this work include determining the specific functions of the proteins released during the initiation of infection, determining the proteins released from other GV (*e.g.* APMV or Antarctica virus), and determining the forces/proteins that are responsible for seal complex disruption *in vivo*.

Between 56 (TV) and 86 (SMBV) proteins are released during the initiation of GV infection. Many (30-40) of these proteins have no known function and are annotated as hypothetical proteins (GenBank KF959826.2). Of the remaining proteins, many have annotations based solely on predicted functional domains. Regardless of the annotation state, the role of these proteins in the GV genome release process must be assessed. We are able to speculate at the role that some of the proteins play in this process. For example, TV releases a Cu-Zn superoxide dismutase (AUL78503.1) that we hypothesize protects the genome from the high levels of heavy metals and ROS that the host uses to degrade proteins inside its phagosome (*216*). Additionally, now that the released proteins have been identified, these specific proteins can be cloned and expressed for additional structural/functional analyses.

Due to sample limitations, and the lack of an Antarctica virus genome sequence, only the proteins released from SMBV and TV virions were analyzed and identified via mass spectrometry. Initial SDS-PAGE experiments suggest that APMV and Antarctica virus also release discrete proteins following low pH treatment. Using the protocol established for SMBV and TV, identification of these proteins should be possible. This identification would provide an opportunity to investigate the conservation of early infection proteins amongst *Mimiviridae*.

Boiling GV particles in acid does not represent a physiological condition in any known GV host. However, the molecular forces identified in this study likely play a role in disrupting

GV seal complexes *in vivo*. It is likely that inside of the phagosome there are host proteins or other, unknown forces that initiate seal disruption. Isolation of amoebal phagosomes, treatment of GV with the contents, and identification of the components that are critical for genome release is a promising method for identifying a GV host receptor.

Overall, the work presented in this Thesis represents a monumental leap in the study of GV. We established the first system for mimicking GV genome release *in vitro*, a system that has been long sought after in the field. Using this system, we were able to characterize the molecular forces responsible for GV seal complex stability, determine the fate of the *Mimiviridae* external seal complexes post-disruption, and identify proteins that are released by GV at the initiation of infection. This *in vitro* genome release system can be used across *Mimiviridae* and potentially even in non-icosahedral GV. This work acts as a basis for future GV genome release studies and may even lead to the discovery of a GV host receptor protein.

APPENDICES

APPENDIX A

SAMBA VIRUS AND TUPANVIRUS SODA LAKE MASS SPECTROMETRY

Analysis and synthesis of this data is presented in the following manuscript:

Schrad, J.R., Abrahão, J.S., Cortines, J.R., Parent, K.N. 2019. Boiling Acid Mimics Intracellular Giant Virus Genome Release. *Cell* (in revision, preprint available through bioRxiv doi: https://doi.org/10.1101/777854).

Table A.1

Protein	*Accession ID*	*Intensities*					
		MA 2	**MA 3**	**Pellet 2**	**Pellet 3**	**Supe 2**	**Supe 3**
Myosin heavy chain*	AAA27709.1	0	0	0	0	0	0
Cytochrome c oxidase subunit 1+2*	AOS85694.1	3212000	2059100	0	1269600	0	0
ATP synthase subunit 9*	AOS85732.1	21864000	16361000	917830	9821000	240990	1428700
STP synthase subunit alpha*	AOS85698.1	5401500	6380000	1195200	2242200	2301600	690380
ORFB (mitochondrion)*	AOS85720.1	12593000	5453900	1527700	6609100	0	0
ubiquitin-like protein Ublp94.4*	AAQ16627.1	7167400	0	0	1717700	0	10181000
iron-superoxide dismutase*	AAT91955.1	3639700	2755600	0	951880	0	0
Rpl7A, partial*	AAY21190.1	12065000	13193000	844350	3491600	2099700	8704700
translocase of outer mitochondrial membrane 40*	ADZ24223.1	5914100	4237000	0	6318400	0	0
histidyl tRNA synthetase *	DAA64396.1	0	0	332640000	0	95086000	0
hypothetical protein	AHJ39842.2	0	0	1704700	834350	2712000	0
virion-associated membrane protein	AHJ39877.1	1576000	1651200	328170	2451700	0	718000
hypothetical protein	AHJ39903.2	2570400	966080	0	962470	0	0
hypothetical protein	AHJ39934.1	3968500	3959900	0	0	0	0
amine oxidase	AHJ39955.1	69398000	81042000	4577000	55055000	983650	3703400
DNA-dependent RNA polymerase subunit RPB9	AHJ39967.2	16054000	9044000	1922700	19595000	242060	0
DNA-directed RNA polymerase subunit 6	AHJ39968.1	0	0	0	0	0	0
DNA topoisomerase 1	AHJ39974.1	216510	0	0	0	0	0
DNA-directed RNA polymerase subunit 5	AHJ39981.1	0	0	0	928590	0	0
hypothetical protein	AHJ39982.1	0	0	0	0	0	0

Table A.1 SMBV Mass Spectrometry Intensities. Raw intensity values for SMBV proteins identified through mass spectrometry. MA = Material Applied, untreated SMBV pellet. Pellet = pH 2.0-treated SMBV pellet. Supe = pH 2.0-treated SMBV supernatant. Pellets and supernatants were separated via centrifugation (as described in Chapter 3).
**Acanthamoeba castellanii* protein

Table A1 (cont'd)

Protein	*Accession ID*	*Intensities*					
		MA 2	**MA 3**	**Pellet 2**	**Pellet 3**	**Supe 2**	**Supe 3**
DNA-directed RNA polymerase subunit 2	AHJ39988.2	6584900	375740	0	2417900	0	203380
DNA-directed RNA polymerase subunit 2	AHJ39990.2	1154000	0	0	0	0	0
DNA-directed RNA polymerase subunit 2	AHJ39991.1	1141500	0	0	0	0	0
ubiquitin-conjugating enzyme e2	AHJ39993.2	312760000	266210000	47065000	232820000	13941000	24068000
hypothetical protein	AHJ39994.1	3364700	3483900	0	0	0	0
WD repeat-containing protein	AHJ40002.1	79358000	52835000	1499700	33007000	599630	2621800
b-type lectin protein	AHJ40019.2	12494000	7015100	987070	17392000	529860	732680
ubiquitin carboxyl-terminal hydrolase	AHJ40023.2	1518500	4327600	0	1335100	0	0
kinesin-like protein	AHJ40024.1	39265000	10107000	692840	79292000	0	870120
serine threonine-protein kinase	AHJ40028.1	0	0	0	0	0	0
protein phosphatase 2c	AHJ40032.1	2476300	744390	0	0	0	0
hypothetical protein	AHJ40033.1	0	0	0	0	0	0
hypothetical protein	AHJ40034.2	0	0	0	0	0	0
formamidopyrimidine-DNA glycosylase	AHJ40038.1	0	0	0	0	0	0
myristoylated membrane protein	AHJ40045.1	0	0	0	0	0	0
hypothetical protein	AHJ40049.1	0	453160	0	0	0	0
hypothetical protein	AHJ40051.1	51550000	21799000	2059300	41160000	214650	2998600
poly(A) polymerase catalytic subunit	AHJ40056.1	0	0	0	0	0	0
hypothetical protein	AHJ40060.1	43338000	44461000	0	34190000	570680	0
hypothetical protein	AHJ40061.1	9954600	5712000	299570	9999100	0	0
transcription termination factor	AHJ40063.1	2112400	989540	0	934960	0	0
hypothetical protein	AHJ40065.1	0	0	0	0	0	0
DNA-directed RNA polymerase II subunit N	AHJ40068.2	1829200	3053900	152310	2726200	0	470310

Table A1 (cont'd)

Protein	***Accession ID***	***Intensities***					
		MA 2	**MA 3**	**Pellet 2**	**Pellet 3**	**Supe 2**	**Supe 3**
thioredoxin domain-containing protein	AHJ40071.1	1108900000	731130000	24146000	792400000	4971200	39646000
ATP-dependent RNA helicase	AHJ40072.2	0	0	0	0	0	0
glycosyltransferase	AHJ40078.2	2110300	1273700	0	2179400	0	0
NTPase	AHJ40081.1	0	0	0	0	0	0
mRNA-capping enzyme	AHJ40083.1	9126300	2869900	208160	2780200	0	84213
putative FtsJ-like methyltransferase	AHJ40084.1	1613600	0	0	1823200	0	0
hypothetical protein	AHJ40087.2	359940000	116230000	9336900	478130000	0	9079800
low complexity protein	AHJ40093.1	554440000	334130000	82096000	397910000	6422600	30274000
putative serine/threonine-protein kinase	AHJ40094.2	3343800	1876400	0	714060	0	0
core protein	AHJ40101.1	378370000	122500000	10086000	292410000	3500800	7650200
hypothetical protein	AHJ40107.2	28839000	21834000	1372700	19291000	217500	1828500
short-chain type dehydrogenase/reductase	AHJ40109.2	0	0	0	0	0	0
capsid protein 1	AHJ40114.2	282400000	143220000	1464500	242040000	4367500	3930800
hypothetical protein	AHJ40128.1	735290000	256460000	3764800	583950000	2920000	10994000
thioredoxin domain-containing protein	AHJ40129.2	285260000	217370000	5550500	199480000	3031600	52726000
hypothetical protein	AHJ40133.1	5838900	3547100	0	4956400	0	0
hypothetical protein	AHJ40139.1	23026000	11459000	2610200	54831000	0	3024500
hypothetical protein	AHJ40142.1	14736000	6618800	0	22282000	0	0
hypothetical protein	AHJ40144.1	26750000	7653500	8539200	31157000	969550	513490
Zn-finger domain-containing protein	AHJ40157.1	3777400	0	0	4539000	0	0
hypothetical protein	AHJ40160.2	125060000	48999000	5763600	87345000	1154100	4978100
hypothetical protein	AHJ40162.1	503870000	542860000	31724000	281370000	26795000	46650000
zinc-type alcohol dehydrogenase-like protein	AHJ40166.2	0	2476900	0	0	0	0
hypothetical protein	AHJ40169.1	388820000	222890000	34520000	327170000	11608000	26910000
DNA-directed RNA polymerase subunit 1	AHJ40170.1	3928500	4459700	1010300	3473900	0	0

Table A1 (cont'd)

Protein	*Accession ID*	*Intensities*					
		MA 2	**MA 3**	**Pellet 2**	**Pellet 3**	**Supe 2**	**Supe 3**
DNA-directed RNA polymerase subunit 1	AHJ40172.1	9643200	6115500	1839800	10449000	2053200	0
hypothetical protein	AHJ40180.1	8058600	2905100	2150400	24521000	338320	674010
hypothetical protein	AHJ40183.2	5598400	1168400	348210	4225900	0	303830
alpha beta hydrolase/esterase/lipase	AHJ40190.1	53691000	8623500	5755000	15275000	0	125120
5-3 exonuclease 20	AHJ40191.2	0	0	0	0	0	0
VVI8 helicase	AHJ40204.1	0	0	0	9927200	0	0
hypothetical protein	AHJ40207.1	2450600	0	0	1690300	0	0
hypothetical protein	AHJ40211.1	0	0	0	0	0	0
hypothetical protein	AHJ40212.2	643690	0	0	0	0	0
hypothetical protein	AHJ40213.2	390490000	265400000	15777000	256640000	4481000	15700000
hypothetical protein	AHJ40220.1	72760000	19080000	9769100	71881000	374970	0
hypothetical protein	AHJ40228.1	2571900	2446100	0	798090	0	0
hypothetical protein	AHJ40230.1	102360000	108450000	20994000	157560000	9938800	11677000
hypothetical protein	AHJ40231.1	0	0	0	0	0	0
hypothetical protein	AHJ40232.2	14689000	7787000	0	6560800	0	0
thiol oxidoreductase e10r	AHJ40236.1	1888300	0	0	1312600	0	0
hypothetical protein	AHJ40243.1	22679000	20486000	433570	18570000	289770	795620
structural PPIase-like protein	AHJ40244.1	0	0	0	0	0	0
mannose-6P isomerase	AHJ40247.1	27475000	630650	3887600	25056000	1093200	3610700
dual specificity S/Y phosphatase	AHJ40253.2	2036800	261980	283450	1971300	0	0
hypothetical protein	AHJ40254.1	31550000	14699000	1776800	27185000	418410	671410
hypothetical protein	AHJ40269.1	32449000	2223400	286590	23463000	0	493370
hypothetical protein	AHJ40271.2	82263000	83890000	10689000	277960000	1653300	7477100
Tat pathway signal sequence domain protein	AHJ40276.1	24858000	10298000	6755400	102680000	2197900	6195000
collagen triple helix repeat containing protein	AHJ40289.2	35364000	32035000	1838600	17627000	458880	1458000
collagen-like protein 7	AHJ40290.2	60650000	41532000	2795100	10319000	1098300	2992400
hypothetical protein	AHJ40316.2	95751000	60680000	920320	20327000	496620	3556700
hypothetical protein	AHJ40318.2	16146000	24142000	859160	42916000	541970	1447900
hypothetical protein	AHJ40319.1	1618500	600140	0	3330500	0	483310

Table A1 (cont'd)

Protein	*Accession ID*	*Intensities*					
		MA 2	**MA 3**	**Pellet 2**	**Pellet 3**	**Supe 2**	**Supe 3**
hypothetical protein	AHJ40320.2	0	0	0	0	0	0
serine protease inhibitor	AHJ40325.2	0	0	0	0	0	0
hypothetical protein	AHJ40326.2	41712000	22223000	14019000	52853000	5196000	3740500
low complexity protein	AHJ40329.1	3772700000	2763700000	1867300000	3292700000	356300000	458050000
hypothetical protein	AHJ40330.1	0	0	0	1184900	0	0
hypothetical protein	AHJ40333.1	369100000	255110000	19375000	877720000	4760400	60539000
chemotaxis protein	AHJ40337.1	3284900	0	0	8822200	0	0
hypothetical protein	AHJ40339.1	30857000	29317000	1426800	12077000	0	1143600
hypothetical protein	AHJ40340.1	1828800	1447300	448800	0	0	1287300
ubiquitin thioesterase	AHJ40341.2	159400000	151150000	8345800	58336000	2432700	5258000
hypothetical protein	AHJ40367.2	1018800	0	0	1230700	0	0
lanosterol 14-alpha-demethylase	AHJ40393.1	99928000	89711000	3424500	38026000	1097000	5245500
glucose-methanol-choline oxidoreductase	AHJ40412.1	22485000	20854000	0	4776800	0	0
hypothetical protein	AHJ40423.1	24887000	26891000	3128100	27548000	1682100	2222200
transcription factor jumonji domain-containing protein	AHJ40444.2	0	0	0	0	0	0
endonuclease exonuclease phosphatase	AHJ40450.1	44209000	42975000	1860200	36268000	0	2968300
hypothetical protein	AMK61731.1	0	0	0	0	0	0
putative lipoxygenase	AMK61740.1	26518000	20749000	207000	10637000	0	0
collagen triple helix repeat containing protein	AMK61745.1	299410000	193730000	1462800	77859000	458960	3046800
GMC-type oxidoreductase	AMK61775.1	24260000	16030000	348990	9170200	99067	609320
choline dehydrogenase-like protein	AMK61776.1	9988900000	7866900000	394030000	5548900000	89947000	556400000
hypothetical protein	AMK61784.1	0	0	0	1442400	0	0
hypothetical protein	AMK61785.1	1249900	0	0	2377200	0	0
DNA topoisomerase 1b	AMK61799.1	0	0	0	0	0	0
probable glutaredoxin	AMK61800.1	15955000	13399000	889360	12214000	398560	2539300
collagen triple helix repeat containing protein	AMK61820.1	834400	652920	0	0	0	0
hypothetical protein	AMK61829.1	317520000	134730000	32595000	382600000	1146200	75513000

Table A1 (cont'd)

Protein	*Accession ID*	*Intensities*					
		MA 2	**MA 3**	**Pellet 2**	**Pellet 3**	**Supe 2**	**Supe 3**
hypothetical protein	AMK61837.1	22654000	21271000	0	5600800	0	0
hypothetical protein	AMK61849.1	15880000	6316800	276690	14814000	0	553970
hypothetical protein	AMK61851.1	1913300	0	0	1935200	0	0
DNA polymerase family x protein	AMK61854.1	1174600	0	0	1844100	0	0
hypothetical protein	AMK61856.1	361310000	304750000	20695000	176900000	0	5407400
early transcription factor large subunit	AMK61857.1	2949100	0	0	3093100	0	0
hypothetical protein	AMK61858.1	45410000	45374000	1395800	31288000	966600	0
regulator of chromosome condensation	AMK61866.1	18280000	14979000	0	9812200	0	0
thiol protease	AMK61869.1	0	0	0	0	177860	0
hypothetical protein	AMK61870.1	5371900	0	0	3627800	0	385600
hypothetical protein	AMK61889.1	1649500	1321500	0	0	0	0
hypothetical protein	AMK61891.1	24776000	10398000	15009000	18500000	577450	0
hypothetical protein	AMK61892.1	45196000	35263000	0	36223000	0	3271500
hypothetical protein	AMK61902.1	199050000	273620000	31582000	92049000	6369500	13627000
anaerobic nitric oxide reductase transcription regulator NorR	AMK61903.1	104210000	55554000	6701500	111960000	745740	5549100
hypothetical protein	AMK61908.1	11334000	761000	0	11220000	0	0
ankyrin repeat protein	AMK61918.1	0	0	0	4551300	0	0
PAN domain-containing protein	AMK61919.1	2115500	1254300	0	2191900	0	0
hypothetical protein	AMK61920.1	1066500000	904620000	39827000	462620000	5646300	110080000
hypothetical protein	AMK61929.1	5010300	471940	200180	7861300	0	0
hypothetical protein	AMK61934.1	0	0	0	0	0	0
hypothetical protein	AMK61942.1	627130000	327650000	2974600	876770000	1120200	7328400
ATP-dependent RNA helicase	AMK61946.1	0	0	0	0	0	0
N-acetyltransferase	AMK61955.1	38112000	10739000	1074500	31942000	0	2132100
hypothetical protein	AMK61957.1	722840	658120	0	793710	0	0
prolyl 4-hydroxylase	AMK61959.1	19679000	5930500	376560	12788000	864120	0
hypothetical protein	AMK61967.1	6465100	2847800	0	7899000	0	575150

Table A1 (cont'd)

Protein	***Accession ID***	***Intensities***					
		MA 2	**MA 3**	**Pellet 2**	**Pellet 3**	**Supe 2**	**Supe 3**
proline rich protein	AMK61968.1	298520000	213770000	7602800	219080000	12090000	35779000
stomatin family protein	AMK61970.1	5419700	0	0	4986900	0	0
hypothetical protein	AMK61977.1	23388000	28574000	1345900	15654000	0	1585700
hypothetical protein	AMK61986.1	0	1360700	0	827060	0	0
NHL repeat-containing protein	AMK61987.1	7630600	8172600	441770	23102000	0	519480
hypothetical protein	AMK61989.1	2096300	369380	0	2268000	0	0
serine threonine-protein kinase	AMK61995.1	0	703940	0	1230400	0	0
hypothetical protein	AMK62013.1	44883000	23693000	427100	62168000	549250	22665000
bifunctional metalloprotease Ub-protein ligase	AMK62014.1	3288100	0	200500	7631700	0	0
hypothetical protein	AMK62016.1	2877800	3458800	0	0	0	0
hypothetical protein	AMK62033.1	5414300	4400800	1372100	6376100	6697700	2277400
hypothetical protein	AMK62059.1	3054700000	2910500000	125850000	1755500000	31168000	136560000
hypothetical protein	AMK62082.1	0	0	0	0	0	0
outer membrane lipoprotein	AMK62087.1	1214500	554800	0	1986800	0	0
choline dehydrogenase-like protein	AMK62096.1	607350000	782590000	18357000	311710000	3747400	15797000
Actin-1*	CAA23399.1	111750000	86549000	25453000	67526000	36029000	21639000
Ubiquitin-60S ribosomal protein L4*	CAA53293.1	78853000	64131000	2678800	28781000	3390900	6702300

Table A.2

Protein	Accession ID	LFQ (Label Free Quantification) Intensities					
		MA 2	MA 3	Pellet 2	Pellet 3	Supe 2	Supe 3
Myosin heavy chain*	AAA27709.1	0	0	0	0	0	0
Cytochrome c oxidase subunit 1+2*	AOS85694.1	0	0	0	1269600	0	0
ATP synthase subunit 9*	AOS85732.1	0	0	0	0	0	1428700
STP synthase subunit alpha*	AOS85698.1	0	0	0	0	0	0
ORFB (mitochondrion)*	AOS85720.1	0	0	0	6609100	0	0
ubiquitin-like protein Ublp94.4*	AAQ16627.1	0	0	0	0	0	9091200
iron-superoxide dismutase*	AAT91955.1	0	0	0	951880	0	0
Rpl7A, partial*	AAY21190.1	11344000	11190000	1768600	4781400	2748400	8565600
translocase of outer mitochondrial membrane 40*	ADZ24223.1	6652400	4772400	0	5044600	0	0
histidyl tRNA synthetase *	DAA64396.1	0	0	0	0	0	0
hypothetical protein	AHJ39842.2	0	0	1630100	0	2786600	0
virion-associated membrane protein	AHJ39877.1	0	0	0	0	0	0
hypothetical protein	AHJ39903.2	0	0	0	962470	0	0
hypothetical protein	AHJ39934.1	0	3959900	0	0	0	0
amine oxidase	AHJ39955.1	75612000	74241000	3801500	54949000	894650	4171700
DNA-dependent RNA polymerase subunit RPB9	AHJ39967.2	12022000	17092000	8582300	16396000	0	0
DNA-directed RNA polymerase subunit 6	AHJ39968.1	0	0	0	0	0	0
DNA topoisomerase 1	AHJ39974.1	0	0	0	0	0	0
DNA-directed RNA polymerase subunit 5	AHJ39981.1	0	0	0	0	0	0

Table A.2 SMBV Mass Spectrometry LFQ Intensities. LFQ (Label Free Quantification) intensity values for SMBV proteins identified through mass spectrometry. MA = Material Applied, untreated SMBV pellet. Pellet = pH 2.0-treated SMBV pellet. Supe = pH 2.0-treated SMBV supernatant. Pellets and supernatants were separated via
centrifugation (as described in Chapter 3).
**Acanthamoeba castellanii* protein

Table A2 (cont'd)

Protein	Accession ID	LFQ (Label Free Quantification) Intensities					
		MA 2	MA 3	Pellet 2	Pellet 3	Supe 2	Supe 3
hypothetical protein	AHJ39982.1	0	0	0	0	0	0
DNA-directed RNA polymerase subunit 2	AHJ39988.2	10843000	0	0	5251200	0	0
DNA-directed RNA polymerase subunit 2	AHJ39990.2	0	0	0	0	0	0
DNA-directed RNA polymerase subunit 2	AHJ39991.1	0	0	0	0	0	0
ubiquitin-conjugating enzyme e2	AHJ39993.2	357250000	269180000	46762000	227400000	17983000	30072000
hypothetical protein	AHJ39994.1	0	0	0	0	0	0
WD repeat-containing protein	AHJ40002.1	112620000	65406000	1951900	55600000	0	2499600
b-type lectin protein	AHJ40019.2	4702700	3246600	587320	18379000	0	576290
ubiquitin carboxyl-terminal hydrolase	AHJ40023.2	0	3043300	0	0	0	0
kinesin-like protein	AHJ40024.1	34002000	6730600	0	67500000	0	1053800
serine threonine-protein kinase	AHJ40028.1	0	0	0	0	0	0
protein phosphatase 2c	AHJ40032.1	0	0	0	0	0	0
hypothetical protein	AHJ40033.1	0	0	0	0	0	0
hypothetical protein	AHJ40034.2	0	0	0	0	0	0
formamidopyrimidine-DNA glycosylase	AHJ40038.1	0	0	0	0	0	0
myristoylated membrane protein	AHJ40045.1	0	0	0	0	0	0
hypothetical protein	AHJ40049.1	0	0	0	0	0	0
hypothetical protein	AHJ40051.1	58187000	26292000	2141200	46870000	0	0
poly(A) polymerase catalytic subunit	AHJ40056.1	0	0	0	0	0	0
hypothetical protein	AHJ40060.1	19463000	18204000	0	14341000	0	0
hypothetical protein	AHJ40061.1	9679000	4022700	0	6428700	0	0
transcription termination factor	AHJ40063.1	0	899920	0	1269500	0	0
hypothetical protein	AHJ40065.1	0	0	0	0	0	0

Table A2 (cont'd)

Protein	Accession ID	LFQ (Label Free Quantification) Intensities					
		MA 2	MA 3	Pellet 2	Pellet 3	Supe 2	Supe 3
DNA-directed RNA polymerase II subunit N	AHJ40068.2	0	0	0	0	0	470310
thioredoxin domain-containing protein	AHJ40071.1	595120000	215760000	32453000	318880000	5719700	31261000
ATP-dependent RNA helicase	AHJ40072.2	0	0	0	0	0	0
glycosyltransferase	AHJ40078.2	0	0	0	0	0	0
NTPase	AHJ40081.1	0	0	0	0	0	0
mRNA-capping enzyme	AHJ40083.1	15659000	3937200	0	3228400	0	0
putative FtsJ-like methyltransferase	AHJ40084.1	0	0	0	0	0	0
hypothetical protein	AHJ40087.2	238060000	62856000	19079000	317140000	0	12710000
low complexity protein	AHJ40093.1	523920000	236490000	46603000	312900000	7443700	29451000
putative serine/threonine-protein kinase	AHJ40094.2	0	1876400	0	0	0	0
core protein	AHJ40101.1	411430000	123000000	16113000	316980000	4869100	11068000
hypothetical protein	AHJ40107.2	31001000	8617300	0	18920000	0	0
short-chain type dehydrogenase/reductase	AHJ40109.2	0	0	0	0	0	0
capsid protein 1	AHJ40114.2	517550000	240590000	3125800	343750000	0	3807700
hypothetical protein	AHJ40128.1	606170000	194430000	2258800	438010000	1379900	11254000
thioredoxin domain-containing protein	AHJ40129.2	234660000	163430000	9553400	127460000	0	46373000
hypothetical protein	AHJ40133.1	0	3439400	0	5001500	0	0
hypothetical protein	AHJ40139.1	29482000	8504500	2975800	48815000	0	3827100
hypothetical protein	AHJ40142.1	15549000	0	0	18198000	0	0
hypothetical protein	AHJ40144.1	22782000	8709800	12466000	34743000	0	0
Zn-finger domain-containing protein	AHJ40157.1	0	0	0	0	0	0
hypothetical protein	AHJ40160.2	133670000	62202000	13190000	104340000	0	9020800
hypothetical protein	AHJ40162.1	491130000	414460000	30336000	187060000	19957000	28277000
zinc-type alcohol dehydrogenase-like protein	AHJ40166.2	0	0	0	0	0	0
hypothetical protein	AHJ40169.1	317830000	181660000	32449000	302100000	8263800	19007000

Table A2 (cont'd)

Protein	Accession ID	LFQ (Label Free Quantification) Intensities					
		MA 2	MA 3	Pellet 2	Pellet 3	Supe 2	Supe 3
DNA-directed RNA polymerase subunit 1	AHJ40170.1	4650800	0	0	0	0	0
DNA-directed RNA polymerase subunit 1	AHJ40172.1	14594000	9396700	0	13010000	0	0
hypothetical protein	AHJ40180.1	0	0	0	24550000	0	0
hypothetical protein	AHJ40183.2	8780600	1080900	0	4363900	0	352740
alpha beta hydrolase/esterase/lipase	AHJ40190.1	46583000	28863000	0	27380000	0	0
5-3 exonuclease 20	AHJ40191.2	0	0	0	0	0	0
VVI8 helicase	AHJ40204.1	0	0	0	0	0	0
hypothetical protein	AHJ40207.1	2605600	0	0	2386700	0	0
hypothetical protein	AHJ40211.1	0	0	0	0	0	0
hypothetical protein	AHJ40212.2	0	0	0	0	0	0
hypothetical protein	AHJ40213.2	345720000	257310000	24751000	190960000	10305000	20410000
hypothetical protein	AHJ40220.1	81363000	22334000	10684000	54057000	0	0
hypothetical protein	AHJ40228.1	0	0	0	798090	0	0
hypothetical protein	AHJ40230.1	90896000	90141000	19415000	120300000	6928300	8555200
hypothetical protein	AHJ40231.1	0	0	0	0	0	0
hypothetical protein	AHJ40232.2	12091000	9970200	0	6976200	0	0
thiol oxidoreductase e10r	AHJ40236.1	0	0	0	0	0	0
hypothetical protein	AHJ40243.1	22146000	19962000	0	18874000	0	0
structural PPIase-like protein	AHJ40244.1	0	0	0	0	0	0
mannose-6P isomerase	AHJ40247.1	28235000	2444600	11579000	28412000	762710	2704500
dual specificity S/Y phosphatase	AHJ40253.2	0	0	0	0	0	0
hypothetical protein	AHJ40254.1	18279000	15795000	2120500	24096000	0	0
hypothetical protein	AHJ40269.1	33371000	0	0	25558000	0	0
hypothetical protein	AHJ40271.2	105800000	101520000	11142000	303540000	0	9816600
Tat pathway signal sequence domain protein	AHJ40276.1	31817000	14112000	7033200	80732000	4465600	9259500
collagen triple helix repeat containing protein	AHJ40289.2	29889000	33683000	0	21453000	0	0
collagen-like protein 7	AHJ40290.2	70054000	24055000	0	8189500	3463600	10830000
hypothetical protein	AHJ40316.2	99910000	61928000	0	20679000	0	2951800

Table A2 (cont'd)

Protein	Accession ID	LFQ (Label Free Quantification) Intensities					
		MA 2	MA 3	Pellet 2	Pellet 3	Supe 2	Supe 3
hypothetical protein	AHJ40318.2	28276000	36381000	900770	78912000	0	1294500
hypothetical protein	AHJ40319.1	0	0	0	3927500	0	0
hypothetical protein	AHJ40320.2	0	0	0	0	0	0
serine protease inhibitor	AHJ40325.2	0	0	0	0	0	0
hypothetical protein	AHJ40326.2	90722000	58762000	18367000	98564000	7248300	8820200
low complexity protein	AHJ40329.1	5556400000	4482100000	1995100000	6377200000	440580000	495790000
hypothetical protein	AHJ40330.1	0	0	0	0	0	0
hypothetical protein	AHJ40333.1	290130000	190610000	21428000	504590000	12180000	52945000
chemotaxis protein	AHJ40337.1	0	0	0	9012500	0	0
hypothetical protein	AHJ40339.1	35795000	32106000	1566300	11850000	0	1312200
hypothetical protein	AHJ40340.1	1950300	0	0	0	0	0
ubiquitin thioesterase	AHJ40341.2	183720000	120620000	11386000	61438000	3929000	7124500
hypothetical protein	AHJ40367.2	0	0	0	0	0	0
lanosterol 14-alpha-demethylase	AHJ40393.1	24878000	105490000	3089700	59016000	1809900	3224900
glucose-methanol-choline oxidoreductase	AHJ40412.1	21962000	20458000	0	5695200	0	0
hypothetical protein	AHJ40423.1	30584000	24545000	3371100	25998000	2483300	5724000
transcription factor jumonji domain-containing protein	AHJ40444.2	0	0	0	0	0	0
endonuclease exonuclease phosphatase	AHJ40450.1	0	0	0	0	0	2276300
hypothetical protein	AMK61731.1	0	0	0	0	0	0
putative lipoxygenase	AMK61740.1	25894000	20523000	0	12304000	0	0
collagen triple helix repeat containing protein	AMK61745.1	629900000	382950000	2513900	70645000	0	5199600
GMC-type oxidoreductase	AMK61775.1	16690000	13274000	2679600	17166000	0	0
choline dehydrogenase-like protein	AMK61776.1	8730000000	8111500000	322160000	4321600000	83357000	293410000
hypothetical protein	AMK61784.1	0	0	0	0	0	0
hypothetical protein	AMK61785.1	0	0	0	0	0	0
DNA topoisomerase 1b	AMK61799.1	0	0	0	0	0	0
probable glutaredoxin	AMK61800.1	12745000	13517000	0	8715600	0	3356300

Table A2 (cont'd)

Protein	Accession ID	LFQ (Label Free Quantification) Intensities					
		MA 2	MA 3	Pellet 2	Pellet 3	Supe 2	Supe 3
collagen triple helix repeat containing protein	AMK61820.1	0	0	0	0	0	0
hypothetical protein	AMK61829.1	598380000	80661000	26222000	257310000	10148000	37068000
hypothetical protein	AMK61837.1	17051000	16797000	0	5811000	0	0
hypothetical protein	AMK61849.1	15760000	6649100	0	14676000	0	0
hypothetical protein	AMK61851.1	0	0	0	0	0	0
DNA polymerase family x protein	AMK61854.1	0	0	0	0	0	0
hypothetical protein	AMK61856.1	303580000	243810000	22254000	140110000	0	5586000
early transcription factor large subunit	AMK61857.1	0	0	0	7001500	0	0
hypothetical protein	AMK61858.1	35475000	0	1721100	23451000	0	0
regulator of chromosome condensation	AMK61866.1	18853000	14613000	0	10726000	0	0
thiol protease	AMK61869.1	0	0	0	0	0	0
hypothetical protein	AMK61870.1	0	0	0	0	0	0
hypothetical protein	AMK61889.1	0	0	0	0	0	0
hypothetical protein	AMK61891.1	19571000	0	24610000	15185000	0	0
hypothetical protein	AMK61892.1	26505000	11917000	0	34061000	0	0
hypothetical protein	AMK61902.1	192540000	160350000	23706000	121130000	0	10031000
anaerobic nitric oxide reductase transcription regulator NorR	AMK61903.1	98709000	51820000	6790200	91968000	1081000	4806600
hypothetical protein	AMK61908.1	16995000	0	0	6629600	0	0
ankyrin repeat protein	AMK61918.1	0	0	0	7354000	0	0
PAN domain-containing protein	AMK61919.1	0	0	0	0	0	0
hypothetical protein	AMK61920.1	1245300000	663870000	38871000	459540000	7417900	58282000
hypothetical protein	AMK61929.1	8036800	0	0	7315500	0	0
hypothetical protein	AMK61934.1	0	0	0	0	0	0
hypothetical protein	AMK61942.1	412940000	126840000	4356000	104340000	1295700	3135000
ATP-dependent RNA helicase	AMK61946.1	0	0	0	0	0	0
N-acetyltransferase	AMK61955.1	46944000	0	0	38786000	0	0

Table A2 (cont'd)

Protein	Accession ID	LFQ (Label Free Quantification) Intensities					
		MA 2	MA 3	Pellet 2	Pellet 3	Supe 2	Supe 3
hypothetical protein	AMK61957.1	0	0	0	0	0	0
prolyl 4-hydroxylase	AMK61959.1	17194000	13232000	0	13126000	0	0
hypothetical protein	AMK61967.1	5512500	3566900	0	11082000	0	0
proline rich protein	AMK61968.1	295320000	228780000	10183000	214540000	8007500	35537000
stomatin family protein	AMK61970.1	0	0	0	4986900	0	0
hypothetical protein	AMK61977.1	13516000	17588000	4080700	35079000	0	0
hypothetical protein	AMK61986.1	0	0	0	0	0	0
NHL repeat-containing protein	AMK61987.1	6185000	10932000	0	21044000	0	0
hypothetical protein	AMK61989.1	0	0	0	0	0	0
serine threonine-protein kinase	AMK61995.1	0	0	0	0	0	0
hypothetical protein	AMK62013.1	45647000	24085000	0	50388000	0	19943000
bifunctional metalloprotease ubiquitin-protein ligase	AMK62014.1	5900800	0	0	7820900	0	0
hypothetical protein	AMK62016.1	0	3458800	0	0	0	0
hypothetical protein	AMK62033.1	0	0	0	0	0	0
hypothetical protein	AMK62059.1	2755400000	2558200000	162630000	1494900000	34452000	124910000
hypothetical protein	AMK62082.1	0	0	0	0	0	0
outer membrane lipoprotein	AMK62087.1	0	0	0	0	0	0
choline dehydrogenase-like protein	AMK62096.1	567030000	559750000	22405000	238710000	5269100	25666000
Actin-1*	CAA23399.1	89967000	68741000	36087000	57990000	38353000	31388000
Ubiquitin-60S ribosomal protein L4*	CAA53293.1	72618000	65974000	3379300	21626000	4465000	8933700

Table A.3

Protein	Accession ID	# of Peptides						Unique Peptides						Sequence Coverage (%)					
		MA		Pellet		Supe		MA		Pellet		Supe		MA		Pellet		Supe	
		2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3
Myosin heavy chain*	AAA27709.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cytochrome c oxidase subunit 1+2*	AOS85694.1	1	1	0	1	0	0	1	1	0	1	0	0	1	1	0	1	0	0
ATP synthase subunit 9*	AOS85732.1	1	1	1	1	1	1	1	1	1	1	1	1	12.7	12.7	12.7	12.7	12.7	12.7
STP synthase subunit alpha*	AOS85698.1	2	2	1	1	1	1	2	2	1	1	1	1	4.6	4.6	2.1	2.5	2.1	3.3
ORFB (mitochondrion)*	AOS85720.1	1	1	1	1	0	0	1	1	1	1	0	0	7	7	7	7	0	0
ubiquitin-like protein Ublp94.4*	AAQ16627.1	2	2	1	2	1	2	1	1	0	1	0	1	2.6	2.6	1.1	2.6	1.1	2.6
iron-superoxide dismutase*	AAT91955.1	1	1	0	1	0	0	1	1	0	1	0	0	5.1	5.1	0	5.1	0	0
Rpl7A, partial*	AAY21190.1	2	2	2	2	2	2	2	2	2	2	2	2	4	4	4	4	4	4
translocase of outer mitochondrial membrane 40*	ADZ24223.1	2	2	0	2	0	0	2	2	0	2	0	0	5	5	0	5	0	0
histidyl tRNA synthetase *	DAA64396.1	0	0	1	0	1	0	0	0	1	0	1	0	0	0	2.2	0	2.2	0
hypothetical protein	AHJ39842.2	0	1	2	1	2	0	0	1	2	1	2	0	0	1.5	3.5	2	3.5	0
virion-associated membrane protein	AHJ39877.1	1	1	1	1	0	1	1	1	1	1	0	1	3.4	3.4	3.4	3.4	0	3.4
hypothetical protein	AHJ39903.2	1	1	0	1	0	0	1	1	0	1	0	0	4.8	4.8	0	4.8	0	0
hypothetical protein	AHJ39934.1	1	1	0	1	0	0	1	1	0	1	0	0	6.3	6.3	0	6.3	0	0
amine oxidase	AHJ39955.1	7	7	5	7	2	4	7	7	5	7	2	4	17.3	17.3	12.2	17.3	7.5	9.4

Table A.3 SMBV Peptide Counts and Sequence Coverage. Number of peptides identified via mass spectrometry for the SMBV proteins. The number of peptide counts identified in each sample, as well as the percentage of the protein sequence that is covered by these counts, is reported.
**Acanthamoeba castellanii* protein

Table A3 (cont'd)

Protein	Accession ID	# of Peptides						Unique Peptides						Sequence Coverage (%)					
		MA		Pellet		Supe		MA		Pellet		Supe		MA		Pellet		Supe	
		2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3
DNA-dependent RNA polymerase subunit RPB9	AHJ39967.2	5	3	1	6	1	0	5	3	1	6	1	0	32.3	12.5	5.2	32.3	5.2	0
DNA-directed RNA polymerase subunit 6	AHJ39968.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DNA topoisomerase 1	AHJ39974.1	1	0	0	0	0	0	1	0	0	0	0	0	1.3	0	0	0	0	0
DNA-directed RNA polymerase subunit 5	AHJ39981.1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	7.3	0	0
hypothetical protein	AHJ39982.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DNA-directed RNA polymerase subunit 2	AHJ39988.2	4	1	0	3	0	1	4	1	0	3	0	1	12.6	2.9	0	12	0	2.9
DNA-directed RNA polymerase subunit 2	AHJ39990.2	1	0	0	0	0	0	1	0	0	0	0	0	5.6	0	0	0	0	0
DNA-directed RNA polymerase subunit 2	AHJ39991.1	1	0	0	0	0	0	1	0	0	0	0	0	1.9	0	0	0	0	0
ubiquitin-conjugating enzyme e2	AHJ39993.2	6	6	4	5	4	6	6	6	4	5	4	6	18.3	18.3	17.5	15.3	14.9	18.3
hypothetical protein	AHJ39994.1	2	2	0	0	0	0	2	2	0	0	0	0	48.3	48.3	0	0	0	0
WD repeat-containing protein	AHJ40002.1	5	6	2	5	1	2	5	6	2	5	1	2	5.7	6.6	2.7	5.1	1.2	2.7
b-type lectin protein	AHJ40019.2	1	1	1	1	1	1	1	1	1	1	1	1	8.3	8.3	8.3	8.3	8.3	8.3
ubiquitin carboxyl-terminal hydrolase	AHJ40023.2	1	3	0	1	0	0	1	3	0	1	0	0	2.7	7.9	0	2.7	0	0
kinesin-like protein	AHJ40024.1	13	5	1	16	0	2	13	5	1	16	0	2	6	2.3	0.3	7.1	0	0.6
serine threonine-protein kinase	AHJ40028.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table A3 (cont'd)

Protein	Accession ID	# of Peptides						Unique Peptides						Sequence Coverage (%)					
		MA		Pellet		Supe		MA		Pellet		Supe		MA		Pellet		Supe	
		2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3
protein phosphatase 2c	AHJ40032.1	1	1	0	0	0	0	1	1	0	0	0	0	4.1	4.1	0	0	0	0
hypothetical protein	AHJ40033.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
hypothetical protein	AHJ40034.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
formamidopyrimidine-DNA glycosylase	AHJ40038.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
myristoylated membrane protein	AHJ40045.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
hypothetical protein	AHJ40049.1	0	1	0	0	0	0	0	1	0	0	0	0	0	4.5	0	0	0	0
hypothetical protein	AHJ40051.1	3	1	1	2	1	1	3	1	1	2	1	1	20	7.5	7.5	20	3.4	7.5
poly(A) polymerase catalytic subunit	AHJ40056.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
hypothetical protein	AHJ40060.1	1	1	0	1	1	0	1	1	0	1	1	0	12.8	12.8	0	12.8	12.8	0
hypothetical protein	AHJ40061.1	4	4	1	3	0	0	4	4	1	3	0	0	8	7.4	2.1	7.4	0	0
transcription termination factor	AHJ40063.1	1	2	0	2	0	0	1	2	0	2	0	0	1.3	3.1	0	3.1	0	0
hypothetical protein	AHJ40065.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DNA-directed RNA polymerase II subunit N	AHJ40068.2	1	1	1	1	0	1	1	1	1	1	0	1	17.3	17.3	17.3	17.3	0	17.3
thioredoxin domain-containing protein	AHJ40071.1	15	17	10	18	6	6	15	17	10	18	6	6	40.5	39.6	27.5	41	22.5	23.7
ATP-dependent RNA helicase	AHJ40072.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
glycosyltransferase	AHJ40078.2	1	1	0	1	0	0	1	1	0	1	0	0	5.2	5.2	0	5.2	0	0
NTPase	AHJ40081.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
mRNA-capping enzyme	AHJ40083.1	3	3	1	3	0	1	3	3	1	3	0	1	2.4	2.9	0.6	3.2	0	0.6
putative FtsJ-like methyltransferase	AHJ40084.1	1	0	0	1	0	0	1	0	0	1	0	0	3.7	0	0	3.7	0	0
hypothetical protein	AHJ40087.2	7	3	3	9	0	3	7	3	3	9	0	3	40.6	22.8	8.5	45.1	0	8.5
low complexity protein	AHJ40093.1	12	11	8	13	2	8	12	11	8	13	2	8	19	19	13.6	21.6	7.7	13.6

Table A3 (cont'd)

Protein	Accession ID	# of Peptides						Unique Peptides						Sequence Coverage (%)					
		MA		Pellet		Supe		MA		Pellet		Supe		MA		Pellet		Supe	
		2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3
putative serine/threonine-protein kinase	AHJ40094.2	1	2	0	1	0	0	1	2	0	1	0	0	2.4	4.9	0	2.4	0	0
core protein	AHJ40101.1	22	15	5	22	3	6	22	15	5	22	3	6	36.2	31.2	8.2	41.3	6.7	11.5
hypothetical protein	AHJ40107.2	3	2	1	2	1	1	3	2	1	2	1	1	8	5.7	3.3	5.7	3.3	3.3
short-chain type dehydrogenase/reductase	AHJ40109.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
capsid protein 1	AHJ40114.2	14	14	2	19	2	5	14	14	2	19	2	5	30.3	28.1	4.2	31.1	4.2	13.2
hypothetical protein	AHJ40128.1	42	35	6	39	4	7	42	35	6	39	4	7	35.8	32.3	4.4	32.8	6.2	7.9
thioredoxin domain-containing protein	AHJ40129.2	8	8	3	8	1	6	8	8	3	8	1	6	47.1	47.1	24	55.4	8.3	31.4
hypothetical protein	AHJ40133.1	1	2	0	2	0	0	1	2	0	2	0	0	6.3	12.2	0	12.2	0	0
hypothetical protein	AHJ40139.1	8	4	7	14	0	5	8	4	7	14	0	5	9.1	5.1	7.8	13.8	0	5.7
hypothetical protein	AHJ40142.1	2	1	0	2	0	0	2	1	0	2	0	0	9	4.8	0	9	0	0
hypothetical protein	AHJ40144.1	2	2	3	4	1	1	2	2	3	4	1	1	7.2	11.9	10.8	18	4.6	4.6
Zn-finger domain-containing protein	AHJ40157.1	1	0	0	1	0	0	1	0	0	1	0	0	9.4	0	0	10	0	0
hypothetical protein	AHJ40160.2	6	7	4	7	1	4	6	7	4	7	1	4	34.5	43.1	23.9	43.1	10.2	33.5
hypothetical protein	AHJ40162.1	6	7	3	6	2	3	6	7	3	6	2	3	32.4	32.4	22.3	32.4	22.3	22.3
zinc-type alcohol dehydrogenase-like protein	AHJ40166.2	0	1	0	0	0	0	0	1	0	0	0	0	0	7.4	0	0	0	0
hypothetical protein	AHJ40169.1	3	3	2	3	2	2	3	3	2	3	2	2	32.5	32.5	32.5	32.5	32.5	32.5
DNA-directed RNA polymerase subunit 1	AHJ40170.1	2	1	1	1	0	0	2	1	1	1	0	0	8	3.7	3.7	3.7	0	0
DNA-directed RNA polymerase subunit 1	AHJ40172.1	3	3	1	4	1	0	3	3	1	4	1	0	3.2	2.6	0.8	3.8	0.8	0
hypothetical protein	AHJ40180.1	1	1	1	2	1	1	1	1	1	2	1	1	7.1	8.3	7.1	15.5	7.1	7.1
hypothetical protein	AHJ40183.2	3	2	1	2	0	2	3	2	1	2	0	2	3.5	2	1.4	2	0	2.6
alpha beta hydrolase/esterase/lipase	AHJ40190.1	2	2	1	4	0	1	2	2	1	4	0	1	13.9	13.9	3.2	14.7	0	2.6

Table A3 (cont'd)

Protein	Accession ID	# of Peptides						Unique Peptides						Sequence Coverage (%)					
		MA		Pellet		Supe		MA		Pellet		Supe		MA		Pellet		Supe	
		2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3
5-3 exonuclease 20	AHJ40191.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VVI8 helicase	AHJ40204.1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	2.3	0	0
hypothetical protein	AHJ40207.1	2	0	0	2	0	0	2	0	0	2	0	0	10.3	0	0	7	0	0
hypothetical protein	AHJ40211.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
hypothetical protein	AHJ40212.2	1	0	0	0	0	0	1	0	0	0	0	0	3.7	0	0	0	0	0
hypothetical protein	AHJ40213.2	16	14	8	16	5	7	16	14	8	16	5	7	40.9	39.3	20	36.1	9.9	9.9
hypothetical protein	AHJ40220.1	7	3	3	8	1	0	7	3	3	8	1	0	23.4	11.2	6.2	24.1	2.8	0
hypothetical protein	AHJ40228.1	1	1	0	1	0	0	1	1	0	1	0	0	5.6	5.6	0	5.6	0	0
hypothetical protein	AHJ40230.1	3	4	3	4	3	3	3	4	3	4	3	3	34.1	62.6	34.1	62.6	34.1	31.9
hypothetical protein	AHJ40231.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
hypothetical protein	AHJ40232.2	3	2	0	2	0	0	3	2	0	2	0	0	6.8	6	0	4.5	0	0
thiol oxidoreductase e10r	AHJ40236.1	1	0	0	1	0	0	1	0	0	1	0	0	3.1	0	0	3.1	0	0
hypothetical protein	AHJ40243.1	2	2	1	2	1	1	2	2	1	2	1	1	12.6	12.6	12.6	12.6	12.6	12.6
structural PPIase-like protein	AHJ40244.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
mannose-6P isomerase	AHJ40247.1	6	2	2	5	3	2	6	2	2	5	3	2	43.8	13.8	13.8	50	25	18.8
dual specificity S/Y phosphatase	AHJ40253.2	1	1	1	1	0	0	1	1	1	1	0	0	4.7	4.7	4.7	4.7	0	0
hypothetical protein	AHJ40254.1	2	2	2	2	1	1	2	2	2	2	1	1	4.1	4.1	4.1	4.1	4.1	4.1
hypothetical protein	AHJ40269.1	4	1	1	4	0	1	4	1	1	4	0	1	9.3	3.9	3.9	9.3	0	5.4
hypothetical protein	AHJ40271.2	3	3	3	4	1	2	3	3	3	4	1	2	19.8	19.8	19.8	23.4	4	13.2
Tat pathway signal sequence domain protein	AHJ40276.1	4	4	4	7	2	3	4	4	4	7	2	3	22.5	22.5	22.5	30	12.8	18.5
collagen triple helix repeat containing protein	AHJ40289.2	2	2	1	2	1	1	2	2	1	2	1	1	3	3	1.2	3	1.2	1.2
collagen-like protein 7	AHJ40290.2	4	2	1	3	2	2	4	2	1	3	2	2	2.5	1.9	0.5	2.5	1.1	1.1
hypothetical protein	AHJ40316.2	2	2	1	1	1	1	2	2	1	1	1	1	16.2	16.2	9	9	9	9
hypothetical protein	AHJ40318.2	4	3	2	4	1	2	4	3	2	4	1	2	14.7	9.6	5.1	11.8	2.1	5.1
hypothetical protein	AHJ40319.1	1	1	0	2	0	1	1	1	0	2	0	1	3.3	3.3	0	12	0	3.3

Table A3 (cont'd)

Protein	Accession ID	# of Peptides						Unique Peptides						Sequence Coverage (%)					
		MA		Pellet		Supe		MA		Pellet		Supe		MA		Pellet		Supe	
		2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3
hypothetical protein	AHJ40320.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
serine protease inhibitor	AHJ40325.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
hypothetical protein	AHJ40326.2	3	2	2	3	2	2	3	2	2	3	2	2	21.2	10.8	10.8	17.2	10.8	10.8
low complexity protein	AHJ40329.1	18	16	15	18	13	14	18	16	15	18	13	14	31.6	30.7	25.9	31.6	24.6	24.6
hypothetical protein	AHJ40330.1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	4	0	0
hypothetical protein	AHJ40333.1	5	4	4	5	2	5	5	4	4	5	2	5	33.5	29.4	33.5	33.5	11.9	33.5
chemotaxis protein	AHJ40337.1	1	0	0	2	0	0	1	0	0	2	0	0	7	0	0	12.6	0	0
hypothetical protein	AHJ40339.1	3	4	2	4	0	2	3	4	2	4	0	2	16.1	17	10.6	17	0	10.6
hypothetical protein	AHJ40340.1	2	1	1	0	0	1	2	1	1	0	0	1	8.5	5.8	8.9	0	0	3.1
ubiquitin thioesterase	AHJ40341.2	9	9	3	6	2	3	9	9	3	6	2	3	31.7	31.7	10.9	23.9	10.9	14.1
hypothetical protein	AHJ40367.2	1	0	0	1	0	0	1	0	0	1	0	0	5.9	0	0	5.9	0	0
lanosterol 14-alpha-demethylase	AHJ40393.1	2	2	2	2	2	2	2	2	2	2	2	2	2.7	2.7	2.7	2.7	2.7	2.7
glucose-methanol-choline oxidoreductase	AHJ40412.1	4	4	0	3	0	0	4	4	0	3	0	0	15	15	0	13.2	0	0
hypothetical protein	AHJ40423.1	4	4	3	4	2	2	4	4	3	4	2	2	31	48	30	31	10	10
transcription factor jumonji domain-containing protein	AHJ40444.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
endonuclease exonuclease phosphatase	AHJ40450.1	1	1	1	1	0	1	1	1	1	1	0	1	1.7	1.7	1.7	1.7	0	1.7
hypothetical protein	AMK61731.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
putative lipoxygenase	AMK61740.1	4	4	1	3	0	0	4	4	1	3	0	0	8.7	8.7	1.6	5.7	0	0
collagen triple helix repeat containing protein	AMK61745.1	4	5	2	5	1	2	4	5	2	5	1	2	4.7	5.9	2.4	6.9	0.8	2.4
GMC-type oxidoreductase	AMK61775.1	3	3	2	3	1	1	3	3	2	3	1	1	9.8	9.8	5.3	9.8	3	3
choline dehydrogenase-like protein	AMK61776.1	51	49	30	50	19	32	51	49	30	50	19	32	58.1	58.3	42	56.1	29.5	43.2

Table A3 (cont'd)

Protein	Accession ID	# of Peptides						Unique Peptides						Sequence Coverage (%)					
		MA		Pellet		Supe		MA		Pellet		Supe		MA		Pellet		Supe	
		2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3
hypothetical protein	AMK61784.1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	3.7	0	0
hypothetical protein	AMK61785.1	1	0	0	1	0	0	1	0	0	1	0	0	5.3	0	0	5.3	0	0
DNA topoisomerase 1b	AMK61799.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
probable glutaredoxin	AMK61800.1	2	3	1	3	1	2	2	3	1	3	1	2	29.4	42.2	7.8	47.1	7.8	29.4
collagen triple helix repeat containing protein	AMK61820.1	1	1	0	0	0	0	1	1	0	0	0	0	1.6	1.6	0	0	0	0
hypothetical protein	AMK61829.1	13	8	4	12	2	4	13	8	4	12	2	4	48.4	33.9	19.4	45.5	12.5	19.4
hypothetical protein	AMK61837.1	5	5	0	4	0	0	5	5	0	4	0	0	25.5	25.5	0	21.2	0	0
hypothetical protein	AMK61849.1	3	2	1	3	0	1	3	2	1	3	0	1	14.8	14.2	5.3	14.8	0	8.9
hypothetical protein	AMK61851.1	1	0	0	2	0	0	1	0	0	2	0	0	4.3	0	0	8.6	0	0
DNA polymerase family x protein	AMK61854.1	1	0	0	1	0	0	1	0	0	1	0	0	2.8	0	0	2.8	0	0
hypothetical protein	AMK61856.1	16	15	9	12	0	6	16	15	9	12	0	6	27.5	25.2	22	27	0	12.8
early transcription factor large subunit	AMK61857.1	2	0	0	3	0	0	2	0	0	3	0	0	1.3	0	0	2.6	0	0
hypothetical protein	AMK61858.1	2	1	2	2	1	0	2	1	2	2	1	0	21.2	21.2	21.2	21.2	21.2	0
regulator of chromosome condensation	AMK61866.1	5	5	0	3	0	0	5	5	0	3	0	0	5.3	5.3	0	3.7	0	0
thiol protease	AMK61869.1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	4.7	0
hypothetical protein	AMK61870.1	2	0	0	2	0	1	2	0	0	2	0	1	4.5	0	0	4.5	0	2
hypothetical protein	AMK61889.1	1	1	0	0	0	0	1	1	0	0	0	0	1.8	1.8	0	0	0	0
hypothetical protein	AMK61891.1	4	2	2	2	1	0	4	2	2	2	1	0	8.4	5.1	5.3	4.8	3.4	0
hypothetical protein	AMK61892.1	6	3	0	6	0	1	6	3	0	6	0	1	9.2	3.3	0	8	0	1.5
hypothetical protein	AMK61902.1	7	9	4	6	1	3	7	9	4	6	1	3	19.7	21.5	8.4	19.7	2.4	10
anaerobic nitric oxide reductase transcription regulator NorR	AMK61903.1	5	6	4	6	2	4	5	6	4	6	2	4	15.7	18.8	15.7	18.8	7.2	15.4
hypothetical protein	AMK61908.1	4	1	0	4	0	0	4	1	0	4	0	0	3.5	0.6	0	3.5	0	0

Table A3 (cont'd)

Protein	Accession ID	# of Peptides						Unique Peptides						Sequence Coverage (%)					
		MA		Pellet		Supe		MA		Pellet		Supe		MA		Pellet		Supe	
		2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3
ankyrin repeat protein	AMK61918.1	0	0	0	4	0	0	0	0	0	4	0	0	0	0	0	3	0	0
PAN domain-containing protein	AMK61919.1	1	1	0	1	0	0	1	1	0	1	0	0	7.2	7.2	0	7.2	0	0
hypothetical protein	AMK61920.1	20	17	12	17	3	12	20	17	12	17	3	12	42.8	39	31.7	40.8	11.1	29.2
hypothetical protein	AMK61929.1	2	1	1	3	0	0	2	1	1	3	0	0	3.8	1.2	1.2	6.9	0	0
hypothetical protein	AMK61934.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
hypothetical protein	AMK61942.1	7	7	2	7	2	2	7	7	2	7	2	2	61.9	61.9	20	61.9	20	20
ATP-dependent RNA helicase	AMK61946.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N-acetyltransferase	AMK61955.1	3	1	1	4	0	1	3	1	1	4	0	1	6.1	3.9	3.9	10.1	0	3.9
hypothetical protein	AMK61957.1	1	1	0	1	0	0	1	1	0	1	0	0	4.6	4.6	0	4.6	0	0
prolyl 4-hydroxylase	AMK61959.1	4	2	1	4	1	0	4	2	1	4	1	0	27.9	14	6.8	27.9	6.8	0
hypothetical protein	AMK61967.1	2	2	0	2	0	1	2	2	0	2	0	1	8.3	8.3	0	8.3	0	2.2
proline rich protein	AMK61968.1	10	10	6	10	2	6	10	10	6	10	2	6	29.8	35.2	20.5	34.9	8.1	21.4
stomatin family protein	AMK61970.1	1	0	0	1	0	0	1	0	0	1	0	0	4.3	0	0	4.3	0	0
hypothetical protein	AMK61977.1	3	4	2	4	1	1	3	4	2	4	1	1	8.1	11.2	6.5	11.2	3.1	3.1
hypothetical protein	AMK61986.1	0	1	0	1	0	0	0	1	0	1	0	0	0	7.7	0	7.7	0	0
NHL repeat-containing protein	AMK61987.1	2	2	1	3	0	1	2	2	1	3	0	1	8	6.5	2	13.8	0	2
hypothetical protein	AMK61989.1	1	1	0	2	0	0	1	1	0	2	0	0	2.4	1.3	0	3.7	0	0

Table A3 (cont'd)

Protein	Accession ID	# of Peptides						Unique Peptides						Sequence Coverage (%)					
		MA		Pellet		Supe		MA		Pellet		Supe		MA		Pellet		Supe	
		2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3
serine threonine-protein kinase	AMK61995.1	0	1	0	1	0	0	0	1	0	1	0	0	0	1.2	0	1.2	0	0
hypothetical protein	AMK62013.1	3	3	1	4	1	4	3	3	1	4	1	4	14.6	14.6	2	19.8	2	19.8
bifunctional metalloprotease ubiquitin-protein ligase	AMK62014.1	3	0	1	2	0	0	3	0	1	2	0	0	5.4	0	2.1	4	0	0
hypothetical protein	AMK62016.1	1	1	0	0	0	0	1	1	0	0	0	0	2.4	2.4	0	0	0	0
hypothetical protein	AMK62033.1	1	1	1	1	1	1	1	1	1	1	1	1	5	5	5	5	5	5
hypothetical protein	AMK62059.1	32	29	14	26	14	17	32	29	14	26	14	17	62.1	60	24.2	61	28.2	34.2
hypothetical protein	AMK62082.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
outer membrane lipoprotein	AMK62087.1	1	1	0	1	0	0	1	1	0	1	0	0	11	11	0	11	0	0
choline dehydrogenase-like protein	AMK62096.1	22	24	7	23	4	8	22	24	7	23	4	8	47.6	51.6	12.7	47.9	9	13.9
Actin-1*	CAA23399.1	7	8	10	9	8	8	7	8	10	9	8	8	28.6	30.7	28.1	32.6	23.3	23.3
Ubiquitin-60S ribosomal protein L4*	CAA53293.1	6	5	3	5	3	3	5	4	2	4	2	2	37.5	37.5	30.5	37.5	21.1	30.5

Table A.4

Protein	Accession ID	Intensities					
		MA 2	MA 3	Pellet 2	Pellet 3	Supe 2	Supe 3
profilin, chain A*	AAA27710.1	4856300	7395200	0	0	0	7747200
Cytochrome c oxidase subunit 1+2*	AAD11820.1	0	11275000	0	4641300	0	0
ATP synthase subunit 9*	AOS85732.1	123140000	141950000	153820000	79427000	0	3651000
ATP synthase subunit alpha*	AOS85698.1	0	0	0	0	1683500	2970500
ubiquitin-like protein Ublp94.4*	AAQ16627.1	976250	0	0	0	0	12544000
iron-superoxide dismutase*	AAT91955.1	17808000	26383000	6859600	0	0	0
lactate dehydrogenase-like protein, partial*	ABD46577.1	0	0	0	22453000	0	0
encystation-mediating serine proteinase*	ABY63398.1	2653100	0	0	0	272310	7892700
protein kinase C8, partial (plastid)*	AFD36237.1	66638000	273980000	100480000	185340000	0	2186300
fork head domain-containing protein	AUL77470.1	0	14520000	0	12438000	0	0
putative ORFan	AUL77479.1	0	50530000	76543000	0	0	527760
hypothetical protein	AUL77482.1	2531400	11504000	3626600	10685000	0	0
hypothetical protein	AUL77486.1	2072200	9630000	5762000	6179900	0	0
hypothetical protein	AUL77492.1	5400500	44880000	21936000	36988000	0	0
mg749 protein	AUL77517.1	29725000	68069000	25018000	37440000	0	458180
tyrosine-protein phosphatase	AUL77518.1	26459000	37955000	39493000	2995700	0	0
putative ORFan	AUL77532.1	21671000	80035000	80598000	49804000	0	1825500
alkylated dna repair protein alkb-like 8 isoform x1	AUL77540.1	5757300	11865000	5166700	10103000	0	0

Table A.4 TV Mass Spectrometry Intensities. Raw intensity values for TV proteins identified through mass spectrometry. MA = Material Applied, untreated TV pellet. Pellet = pH 2.0-treated SMBV pellet. Supe = pH 2.0-treated SMBV supernatant. Pellets and supernatants were separated via centrifugation (as described in Chapter 3).
**Acanthamoeba castellanii* protein

Table A4 (cont'd)

Protein	Accession ID	Intensities					
		MA 2	MA 3	Pellet 2	Pellet 3	Supe 2	Supe 3
mannose-6P isomerase	AUL77553.1	33439000	53031000	0	31798000	0	9435100
hypothetical protein	AUL77579.1	0	0	0	0	0	28260000
putative oxireductase	AUL77599.1	138660000	785350000	127210000	395010000	0	23783000
hypothetical protein	AUL77600.1	410410000	636220000	979840000	410600000	0	6117300
ubiquitin-conjugating enzyme e2	AUL77610.1	39729000	78424000	53799000	31031000	0	0
hypothetical protein	AUL77622.1	0	8970700	0	4683600	0	0
hypothetical protein	AUL77647.1	60808000	96904000	106500000	59432000	0	0
structural ppiase-like protein	AUL77649.1	25021000	75609000	39659000	60820000	0	554700
thiol oxidoreductase E10R	AUL77655.1	18597000	23951000	38161000	16037000	0	0
mg709 protein	AUL77661.1	341940000	567980000	128340000	338190000	0	28425000
hypothetical protein	AUL77666.1	18498000	91009000	128610000	49944000	0	0
putative ORFan	AUL77678.1	0	158850000	0	114660000	0	0
putative N-acetyl transferase	AUL77680.1	10939000	79666000	28348000	24611000	0	0
putative helicase	AUL77687.1	11852000	0	67527000	51032000	0	0
hypothetical protein	AUL77688.1	3357400000	8669400000	4378900000	5221100000	0	108520000
hypothetical protein	AUL77694.1	1300300000	5436000000	3090900000	5864200000	0	53301000
glycosyl hydrolase family 18	AUL77711.1	100190000	262810000	139970000	218410000	0	0
hypothetical protein	AUL77718.1	100580000	132000000	41449000	23224000	0	302720000
putative DNA repair protein	AUL77721.1	0	5793300	0	0	0	0
hypothetical protein	AUL77723.1	296630000	1780500000	345140000	391770000	0	49149000
putative ORFan	AUL77729.1	446240000	1190300000	748750000	1205200000	0	5841000
hypothetical protein	AUL77752.1	231580000	795650000	231610000	810340000	0	257360
hypothetical protein	AUL77753.1	11035000	91770000	0	0	0	0
putative ATP-dependent RNA helicase	AUL77758.1	20569000	16542000	35122000	6934500	0	0
atp-dependent rna helicase	AUL77773.1	3463400	15814000	9939500	13845000	0	0

Table A4 (cont'd)

Protein	Accession ID	Intensities					
		MA 2	MA 3	Pellet 2	Pellet 3	Supe 2	Supe 3
putative 5-3exonuclease20	AUL77795.1	5213000	72637000	65979000	52018000	0	0
phosphoesterase-like protein	AUL77796.1	34413000	151240000	60068000	142420000	0	0
hypothetical protein	AUL77813.1	17956000	22038000	51189000	0	0	0
hypothetical protein	AUL77820.1	42331000	137390000	54980000	64850000	0	772640
kinesin-like protein	AUL77838.1	119540000	223770000	237400000	113880000	0	2284800
putative protein phosphatase 2c	AUL77859.1	52239000	99238000	110180000	153230000	0	0
hypothetical protein	AUL77863.1	15879000	78878000	16491000	33822000	0	0
hypothetical protein	AUL77885.1	1934400	10142000	0	12303000	0	0
DNA polymerase family X	AUL77886.1	10802000	26075000	21653000	45634000	0	0
hypothetical protein	AUL77896.1	497220000	1573500000	855800000	1294400000	0	11279000
putative early transcription factor	AUL77899.1	90552000	601940000	436280000	699420000	0	1118500
hypothetical protein	AUL77902.1	0	0	1669700	0	0	459040
hypothetical protein	AUL77903.1	58601000	122400000	100460000	128690000	0	0
mg574 protein	AUL77905.1	25836000	5122200	0	77484000	0	0
hypothetical protein	AUL77907.1	438820000	1345100000	710120000	1325600000	0	123470000
mg18 protein	AUL77928.1	4355100	36953000	5016600	32427000	0	0
polyA polymerase catalitic subunit	AUL77929.1	74431000	303550000	211290000	288450000	0	1699100
hypothetical protein	AUL77930.1	32989000	75088000	0	5047700	0	16063000
hypothetical protein	AUL77933.1	74077000	301630000	196220000	230750000	0	1434200
hypothetical protein	AUL77936.1	147810000	489110000	600700000	564660000	0	7544500
SNF2 family helicase	AUL77941.1	49791000	102680000	194360000	190270000	0	0
hypothetical protein	AUL77944.1	201530000	429160000	398190000	436820000	0	2599300
putative ORFan	AUL77950.1	3735800	21304000	6773900	14684000	0	0
hypothetical protein	AUL77952.1	18030000	62004000	57425000	20014000	0	0
hypothetical protein	AUL77961.1	40261000	162910000	93368000	90850000	0	888140
hypothetical protein	AUL77962.1	0	3555100	1855300	7051800	0	0
thioredoxin domain-containing protein	AUL77963.1	1066500000	3754300000	812610000	643590000	0	127530000

Table A4 (cont'd)

Protein	Accession ID	Intensities					
		MA 2	MA 3	Pellet 2	Pellet 3	Supe 2	Supe 3
putative ATP-dependent RNA helicase	AUL77999.1	0	0	0	0	0	0
NUDIX hydrolase	AUL78015.1	0	0	0	0	0	0
dna-directed rna polymerase subunit	AUL78016.1	56851000	109530000	118070000	119340000	0	30346000
hypothetical protein	AUL78017.1	1017200	0	1808100	1912700	0	0
hypothetical protein	AUL78019.1	3140900	0	0	0	0	4160400
NTPase	AUL78021.1	137380000	290240000	264070000	229350000	0	384930
putative leucine-rich repeat protein	AUL78028.1	1998900	0	17411000	0	0	9316400
mRNA capping enzyme	AUL78031.1	124070000	440020000	166770000	274810000	0	3233400
FtsJ-like methyl transferase	AUL78032.1	48292000	111180000	126520000	172240000	0	0
ubiquitin domain-containing protein	AUL78040.1	146950000	437070000	177510000	261250000	1382700	296500000
hypothetical protein	AUL78045.1	31936000	182840000	57321000	90332000	0	0
hypothetical protein	AUL78046.1	25821000	96461000	53680000	106700000	0	0
putative heat shock 70 kDa protein	AUL78049.1	2481400	5898700	3255600	4039400	0	0
hypothetical protein	AUL78055.1	76018000	266780000	66570000	94848000	0	73763000
hypothetical protein	AUL78059.1	37757000	94968000	61138000	101410000	0	0
hypothetical protein	AUL78061.1	71692000	70018000	0	2067700	0	0
serine threonine-protein kinase	AUL78063.1	3301600	20822000	13900000	7647400	0	0
hypothetical protein	AUL78067.1	209280000	590160000	597930000	491110000	0	5218600
hypothetical protein	AUL78068.1	29108000	60289000	71290000	70558000	0	0
hypothetical protein	AUL78073.1	2619500	22967000	15533000	14415000	0	0
hypothetical protein	AUL78080.1	1967400	12470000	0	0	0	0
major core protein	AUL78082.1	4896000000	12109000000	10970000000	8145100000	0	137450000
putative ORFan	AUL78086.1	0	0	19724000	29704000	0	0
putative ORFan	AUL78088.1	114970000	420510000	12266000	18228000	0	669070000
hypothetical protein	AUL78091.1	21042000	117280000	9505300	40451000	0	22989000
hypothetical protein	AUL78092.1	0	19352000	0	7726800	0	20472000

Table A4 (cont'd)

Protein	Accession ID	Intensities					
		MA 2	MA 3	Pellet 2	Pellet 3	Supe 2	Supe 3
hypothetical protein	AUL78093.1	28435000	422590000	109490000	530880000	0	0
catalase HPII	AUL78097.1	446710000	1291900000	550700000	868170000	0	164100000
putative ORFan	AUL78106.1	2717200	6588500	0	0	0	0
hypothetical protein	AUL78108.1	19202000	78964000	21592000	51112000	0	3387600
dna topoisomerase 1b	AUL78109.1	35127000	124400000	60832000	61543000	0	0
cyclopropane fatty acyl phospholipid synthase	AUL78111.1	0	0	0	6971200	0	0
putative ORFan	AUL78114.1	0	0	9064500	15247000	0	0
extracellular ligand-binding receptor	AUL78119.1	26201000	19153000	33064000	25186000	0	0
putative ORFan	AUL78120.1	47102000	112650000	77054000	48897000	0	753130
glyoxalase	AUL78134.1	73251000	181240000	97410000	167450000	0	3765400
hypothetical protein	AUL78135.1	1750500000	4389800000	2730600000	2743400000	0	30439000
putative ORFan	AUL78142.1	29055000	42625000	11587000	52676000	0	0
hypothetical protein	AUL78143.1	4502700000	13832000000	7293800000	10519000000	294710	164970000
capsid protein 1	AUL78147.1	42249000000	91156000000	82274000000	95948000000	1905700	1246000000
hypothetical protein	AUL78155.1	71697000	281320000	159690000	190150000	0	0
putative ORFan	AUL78156.1	0	0	0	0	0	2480000
hypothetical protein	AUL78183.1	4602000	13170000	9437600	0	0	868410
hypothetical protein	AUL78191.1	7405500000	18328000000	8364900000	14584000000	413810	1025500000
thioredoxin domain-containing protein	AUL78192.1	55573000	173490000	245750000	51906000	0	0
hypothetical protein	AUL78198.1	38576000	85669000	34861000	41345000	0	0
putative ORFan	AUL78206.1	426010000	1263000000	894700000	493060000	0	0
putative pore coat assembly factor	AUL78211.1	218070000	691340000	217150000	385900000	0	23168000
hypothetical protein	AUL78214.1	1162400000	2238600000	1242600000	1566700000	0	49255000
hypothetical protein	AUL78219.1	1244500000	4415400000	2576900000	4252700000	0	44491000
hypothetical protein	AUL78232.1	252850000	501620000	106690000	83346000	0	5882200
dna directed rna polymerase subunit	AUL78244.1	13555000	64257000	70546000	58898000	0	0
hypothetical protein	AUL78246.1	163640000	364390000	454750000	417860000	0	2661600
putative ORFan	AUL78253.1	23178000	31506000	9830600	0	0	22418000
hypothetical protein	AUL78254.1	32601000	83010000	9041500	34586000	0	0

Table A4 (cont'd)

Protein	Accession ID	Intensities					
		MA 2	MA 3	Pellet 2	Pellet 3	Supe 2	Supe 3
arylsulfatase	AUL78269.1	204920000	917850000	380390000	650180000	0	10308000
peptidase inhibitor I9	AUL78271.1	0	0	0	0	0	0
putative ankyrin repeat protein	AUL78278.1	32040000	66143000	87084000	51558000	0	0
hypothetical protein	AUL78280.1	35968000	60459000	68818000	71427000	0	1386100
hypothetical protein	AUL78287.1	1509100000	4901400000	1116600000	1765600000	198080	162490000
hypothetical protein	AUL78288.1	487750000	852350000	374280000	259260000	0	44479000
hypothetical protein	AUL78292.1	340870000	647230000	319960000	575960000	0	1018400
hypothetical protein	AUL78295.1	0	5460200	0	0	0	0
hypothetical protein	AUL78301.1	6388200	54825000	45489000	76866000	0	0
dna-directed rna polymerase subunit 1	AUL78302.1	348540000	1002800000	951830000	1126900000	298080	25094000
hypothetical protein	AUL78318.1	2599800	21965000	5615900	44539000	0	0
hypothetical protein	AUL78319.1	76174000	122610000	147710000	88921000	0	0
peptidase inhibitor I9	AUL78329.1	0	37750000	0	0	0	0
peptidase inhibitor I9	AUL78330.1	56407000	161020000	78856000	56653000	0	1192300
hypothetical protein	AUL78347.1	103560000	171870000	7731100	22738000	0	33895000
hypothetical protein	AUL78348.1	11111000	49461000	0	0	0	4622800
hypothetical protein	AUL78354.1	28838000	34692000	6304300	8094200	0	64830000
intein-containing DNA-directed RNA polymerase subunit 2	AUL78361.1	42598000	146810000	137500000	181170000	0	0
intein-containing dna-directed rna polymerase subunit 2	AUL78362.1	121460000	426080000	272410000	456980000	0	2227900
DNA-directed RNA polymerase subunit 6	AUL78368.1	24897000	47055000	22495000	13676000	0	20558000
putative fibril associated protein	AUL78400.1	3468000000	8883800000	6162500000	7596500000	0	69631000
putative major capsid protein	AUL78403.1	2172400	16225000	8033100	15315000	0	0
putative neurocan core protein	AUL78405.1	3735700	20126000	15574000	9303900	0	224820
hypothetical protein	AUL78410.1	34402000	0	9197700	6603400	0	0
putative ORFan	AUL78423.1	0	11590000	0	15241000	0	0

Table A4 (cont'd)

Protein	Accession ID	Intensities					
		MA 2	MA 3	Pellet 2	Pellet 3	Supe 2	Supe 3
putative ORFan	AUL78440.1	4748700	8470300	0	5572500	0	0
hypothetical protein	AUL78464.1	153580000	826200000	150340000	282150000	0	77726000
hypothetical protein	AUL78466.1	60583000	119930000	75913000	86573000	0	13399000
hypothetical protein	AUL78468.1	26663000	186930000	46668000	47794000	0	113590000
hypothetical protein	AUL78479.1	2255400	12748000	7458500	7546600	0	0
hypothetical protein	AUL78481.1	59567000	315340000	143130000	207030000	0	911210
putative ORFan	AUL78496.1	0	12393000	5712400	20821000	0	0
hypothetical protein	AUL78500.1	0	25977000	0	11704000	0	0
putative ORFan	AUL78514.1	69753000	172350000	134650000	103120000	0	4012200
putative ORFan	AUL78530.1	2956000	8733600	29777000	56010000	0	0
hypothetical protein	AUL78545.1	8510400	64681000	7532900	36438000	0	1184600
hypothetical protein	AUL78577.1	26029000	152210000	138110000	109980000	0	0
putative lipocalin	AUL78583.1	0	8074700	0	7057200	0	0
putative ORFan	AUL78586.1	5629500	6734600	0	0	0	0
Tlr 6Fp protein	AUL78587.1	0	0	0	0	0	0
hypothetical protein	AUL78601.1	278000000	173740000	177110000	59774000	0	3956300
putative protein kinase	AUL78629.1	354210000	596610000	552290000	351300000	0	4158800
Ig family protein	AUL78630.1	214960000	431130000	25082000	0	970820	582990000
hypothetical protein	AUL78631.1	0	0	0	0	0	7640700
putative ORFan	AUL78635.1	226040000	548070000	67937000	235990000	0	179600
chemotaxis	AUL78637.1	33618000	100590000	76714000	96843000	0	0
hypothetical protein	AUL78639.1	0	9642500	0	0	0	0
hypothetical protein	AUL78659.1	85367000	221890000	99267000	91058000	0	0
hypothetical protein	AUL78681.1	188610000	385990000	46931000	58019000	0	158930000
putative chemotaxis protein ched	AUL78687.1	14323000	19436000	23552000	15460000	0	0
bifunctional metalloprotease ubiquitin-protein ligase	AUL78691.1	14238000	88267000	61101000	101390000	0	778310
putative ORFan	AUL78708.1	8064200	8879800	0	0	0	0
hypothetical protein	AUL78717.1	0	9137100	0	7813300	0	0
putative CfxQ-like protein	AUL78719.1	121330000	650480000	202930000	604050000	0	949640
glutaredoxin	AUL78724.1	56591000	74204000	22607000	43547000	0	86378000

Table A4 (cont'd)

Protein	Accession ID	Intensities					
		MA 2	MA 3	Pellet 2	Pellet 3	Supe 2	Supe 3
hypothetical protein	AUL78731.1	20902000	32026000	0	6332100	0	0
hypothetical protein	AUL78736.1	2079200	9682000	5140300	8150600	0	0
DNA-dependent RNA polymerase subunit Rpb9	AUL78739.1	46185000	167160000	91369000	106680000	0	32980000
p87	AUL78740.1	0	0	64046000	10833000	0	0
DNA-directed RNA polymerase subunit 6	AUL78741.1	21885000	84760000	69932000	89636000	0	0
Actin-1*	CAA23399.1	8465600	41232000	127580000	23642000	14033000	33837000
Ubiquitin-60S ribosomal protein L40*	CAA53293.1	0	0	0	0	0	13388000
Myosin-2 heavy chain*	CAA68663.1	436370000	982390000	717240000	955280000	0	0

Table A.5

Protein	Accession ID	LFQ (Label Free Quantification)					
		MA 2	MA 3	Pellet 2	Pellet 3	Supe 2	Supe 3
profilin, chain A*	AAA27710.1	0	0	0	0	0	7747200
Cytochrome c oxidase subunit 1+2*	AAD11820.1	0	0	0	4641300	0	0
ATP synthase subunit 9*	AOS85732.1	0	0	0	0	0	3651000
ATP synthase subunit alpha*	AOS85698.1	0	0	0	0	0	2970500
ubiquitin-like protein Ublp94.4*	AAQ16627.1	0	0	0	0	0	10730000
iron-superoxide dismutase*	AAT91955.1	0	0	6859600	0	0	0
lactate dehydrogenase-like protein, partial*	ABD46577.1	0	0	0	22453000	0	0
encystation-mediating serine proteinase*	ABY63398.1	0	0	0	0	0	8207300
protein kinase C8, partial (plastid)*	AFD36237.1	0	0	0	0	0	2186300
fork head domain-containing protein	AUL77470.1	0	0	0	12438000	0	0
putative ORFan	AUL77479.1	0	0	0	0	0	527760
hypothetical protein	AUL77482.1	0	0	0	10685000	0	0
hypothetical protein	AUL77486.1	0	0	0	6179900	0	0
hypothetical protein	AUL77492.1	7559700	47633000	16965000	29280000	0	0
mg749 protein	AUL77517.1	29870000	83320000	24939000	33731000	0	0
tyrosine-protein phosphatase	AUL77518.1	25879000	0	43925000	0	0	0

Table A.5 TV Mass Spectrometry LFQ Intensities. LFQ intensity values for TV proteins identified through mass spectrometry. MA = Material Applied, untreated TV pellet. Pellet = pH 2.0-treated SMBV pellet. Supe = pH 2.0-treated SMBV supernatant. Pellets and supernatants were separated via centrifugation (as described in Chapter 3).
**Acanthamoeba castellanii* protein

Table A5 (cont'd)

Protein	Accession ID	LFQ (Label Free Quantification)					
		MA 2	MA 3	Pellet 2	Pellet 3	Supe 2	Supe 3
putative ORFan	AUL77532.1	22501000	57015000	79021000	42656000	0	0
alkylated dna repair protein alkb-like 8 isoform x1	AUL77540.1	0	12094000	0	10981000	0	0
mannose-6P isomerase	AUL77553.1	0	43777000	0	0	0	0
hypothetical protein	AUL77579.1	0	0	0	0	0	28260000
putative oxireductase	AUL77599.1	142990000	561600000	199550000	304440000	0	0
hypothetical protein	AUL77600.1	282290000	665460000	666150000	336850000	0	2974800
ubiquitin-conjugating enzyme e2	AUL77610.1	0	97859000	65442000	0	0	0
hypothetical protein	AUL77622.1	0	0	0	0	0	0
hypothetical protein	AUL77647.1	41990000	87024000	77610000	79148000	0	0
structural ppiase-like protein	AUL77649.1	25303000	66457000	34325000	47781000	0	0
thiol oxidoreductase E10R	AUL77655.1	15724000	19669000	41529000	13346000	0	0
mg709 protein	AUL77661.1	303940000	641320000	150040000	306440000	0	33717000
hypothetical protein	AUL77666.1	27450000	52828000	165470000	50455000	0	0
putative ORFan	AUL77678.1	0	0	0	0	0	0
putative N-acetyl transferase	AUL77680.1	0	91491000	0	0	0	0
putative helicase	AUL77687.1	0	0	0	51032000	0	0
hypothetical protein	AUL77688.1	2526300000	7596100000	3406100000	5235100000	0	62601000
hypothetical protein	AUL77694.1	1119100000	5026200000	1957100000	4852800000	0	82865000
glycosyl hydrolase family 18	AUL77711.1	96169000	253900000	131740000	215160000	0	0
hypothetical protein	AUL77718.1	0	102350000	0	0	0	0
putative DNA repair protein	AUL77721.1	0	5793300	0	0	0	0
hypothetical protein	AUL77723.1	285770000	811390000	431250000	430110000	0	105460000
putative ORFan	AUL77729.1	547160000	967650000	574990000	1250100000	0	8613000
hypothetical protein	AUL77752.1	388770000	798780000	493520000	1111600000	0	0

Table A5 (cont'd)

Protein	Accession ID	LFQ (Label Free Quantification)					
		MA 2	MA 3	Pellet 2	Pellet 3	Supe 2	Supe 3
hypothetical protein	AUL77753.1	0	124200000	0	0	0	0
putative ATP-dependent RNA helicase	AUL77758.1	0	0	0	0	0	0
atp-dependent rna helicase	AUL77773.1	0	0	0	13845000	0	0
putative 5-3exonuclease20	AUL77795.1	0	76504000	70419000	51185000	0	0
phosphoesterase-like protein	AUL77796.1	37900000	235490000	71749000	134510000	0	0
hypothetical protein	AUL77813.1	0	0	51189000	0	0	0
hypothetical protein	AUL77820.1	59758000	0	63181000	62186000	0	0
kinesin-like protein	AUL77838.1	104980000	254530000	281540000	76777000	0	4227400
putative protein phosphatase 2c	AUL77859.1	47814000	155690000	98462000	141980000	0	0
hypothetical protein	AUL77863.1	15011000	66059000	18261000	46555000	0	0
hypothetical protein	AUL77885.1	0	0	0	12303000	0	0
DNA polymerase family X	AUL77886.1	0	37499000	0	44684000	0	0
hypothetical protein	AUL77896.1	0	0	0	0	0	0
putative early transcription factor	AUL77899.1	204430000	812900000	498800000	628620000	0	0
hypothetical protein	AUL77902.1	0	0	0	0	0	459040
hypothetical protein	AUL77903.1	59253000	121320000	124240000	105950000	0	0
mg574 protein	AUL77905.1	0	0	0	0	0	0
hypothetical protein	AUL77907.1	495210000	1189300000	707420000	1029300000	0	97830000
mg18 protein	AUL77928.1	0	35640000	0	32751000	0	0
polyA polymerase catalitic subunit	AUL77929.1	60644000	355590000	135710000	315060000	0	0
hypothetical protein	AUL77930.1	38619000	70551000	0	0	0	18721000
hypothetical protein	AUL77933.1	85813000	295130000	200810000	228140000	0	0
hypothetical protein	AUL77936.1	121740000	536700000	473740000	489710000	0	17981000
SNF2 family helicase	AUL77941.1	59980000	120720000	212000000	251270000	0	0
hypothetical protein	AUL77944.1	0	0	184090000	114400000	0	0

Table A5 (cont'd)

Protein	Accession ID	LFQ (Label Free Quantification)					
		MA 2	MA 3	Pellet 2	Pellet 3	Supe 2	Supe 3
putative ORFan	AUL77950.1	0	0	0	14684000	0	0
hypothetical protein	AUL77952.1	0	37551000	48076000	22709000	0	0
hypothetical protein	AUL77961.1	59389000	160590000	116910000	108280000	0	0
hypothetical protein	AUL77962.1	0	0	0	7051800	0	0
thioredoxin domain-containing protein	AUL77963.1	1259600000	3410200000	876940000	574250000	0	125890000
putative ATP-dependent RNA helicase	AUL77999.1	0	0	0	0	0	0
NUDIX hydrolase	AUL78015.1	0	0	0	0	0	0
dna-directed rna polymerase subunit	AUL78016.1	58025000	110820000	105340000	130170000	0	38870000
hypothetical protein	AUL78017.1	0	0	0	0	0	0
hypothetical protein	AUL78019.1	0	0	0	0	0	4160400
NTPase	AUL78021.1	139190000	310720000	257040000	299270000	0	0
putative leucine-rich repeat protein	AUL78028.1	0	0	0	0	0	9316400
mRNA capping enzyme	AUL78031.1	94948000	327810000	225770000	375160000	0	0
FtsJ-like methyl transferase	AUL78032.1	48347000	128190000	135230000	155750000	0	0
ubiquitin domain-containing protein	AUL78040.1	153420000	502600000	141990000	203770000	4522800	283660000
hypothetical protein	AUL78045.1	23714000	149890000	48915000	127930000	0	0
hypothetical protein	AUL78046.1	0	0	0	106700000	0	0
putative heat shock 70 kDa protein	AUL78049.1	0	0	0	0	0	0
hypothetical protein	AUL78055.1	83201000	255620000	79059000	94013000	0	80317000
hypothetical protein	AUL78059.1	0	0	0	0	0	0
hypothetical protein	AUL78061.1	47142000	28383000	0	0	0	0
serine threonine-protein kinase	AUL78063.1	0	0	0	7647400	0	0
hypothetical protein	AUL78067.1	231300000	629320000	567030000	516350000	0	8247900
hypothetical protein	AUL78068.1	30858000	61226000	69192000	71925000	0	0
hypothetical protein	AUL78073.1	0	17523000	18023000	17205000	0	0

Table A5 (cont'd)

Protein	Accession ID	LFQ (Label Free Quantification)					
		MA 2	MA 3	Pellet 2	Pellet 3	Supe 2	Supe 3
hypothetical protein	AUL78080.1	0	12470000	0	0	0	0
major core protein	AUL78082.1	3900300000	12314000000	10696000000	8085800000	0	154520000
putative ORFan	AUL78086.1	0	0	19724000	0	0	0
putative ORFan	AUL78088.1	106970000	451510000	0	125550000	0	366430000
hypothetical protein	AUL78091.1	0	0	0	0	0	22989000
hypothetical protein	AUL78092.1	0	0	0	0	0	0
hypothetical protein	AUL78093.1	69634000	450280000	103590000	274400000	0	0
catalase HPII	AUL78097.1	473730000	1451100000	625780000	874290000	0	57211000
putative ORFan	AUL78106.1	0	0	0	0	0	0
hypothetical protein	AUL78108.1	0	0	0	0	0	3387600
dna topoisomerase 1b	AUL78109.1	33597000	114850000	48242000	89146000	0	0
cyclopropane fatty acyl phospholipid synthase	AUL78111.1	0	0	0	6971200	0	0
putative ORFan	AUL78114.1	0	0	0	15247000	0	0
extracellular ligand-binding receptor	AUL78119.1	0	0	0	25186000	0	0
putative ORFan	AUL78120.1	0	0	0	0	0	753130
glyoxalase	AUL78134.1	0	0	0	0	0	0
hypothetical protein	AUL78135.1	1737500000	4218900000	2923400000	2703500000	0	33949000
putative ORFan	AUL78142.1	0	0	0	52676000	0	0
hypothetical protein	AUL78143.1	3828100000	14023000000	8499100000	8208400000	0	109700000
capsid protein 1	AUL78147.1	33575000000	82361000000	79996000000	86487000000	1174200	1096500000
hypothetical protein	AUL78155.1	44240000	248050000	140470000	140960000	0	0
putative ORFan	AUL78156.1	0	0	0	0	0	2480000
hypothetical protein	AUL78183.1	0	0	9437600	0	0	0
hypothetical protein	AUL78191.1	6267800000	18980000000	7810800000	14762000000	0	983720000
thioredoxin domain-containing protein	AUL78192.1	79886000	209780000	239360000	0	0	0
hypothetical protein	AUL78198.1	39809000	90031000	0	43937000	0	0
putative ORFan	AUL78206.1	277770000	952870000	0	0	0	0

Table A5 (cont'd)

Protein	Accession ID	LFQ (Label Free Quantification)					
		MA 2	MA 3	Pellet 2	Pellet 3	Supe 2	Supe 3
putative pore coat assembly factor	AUL78211.1	223780000	708090000	303760000	392150000	0	33713000
hypothetical protein	AUL78214.1	962490000	2395100000	1355300000	2187300000	0	29663000
hypothetical protein	AUL78219.1	1161000000	4322100000	2726100000	4254600000	0	43820000
hypothetical protein	AUL78232.1	201270000	361040000	44954000	125340000	0	3685500
dna directed rna polymerase subunit	AUL78244.1	15251000	51318000	99323000	42839000	0	0
hypothetical protein	AUL78246.1	153580000	415560000	489210000	484250000	0	0
putative ORFan	AUL78253.1	0	0	0	0	0	22418000
hypothetical protein	AUL78254.1	0	0	0	34586000	0	0
arylsulfatase	AUL78269.1	276740000	819830000	674880000	547190000	0	14338000
peptidase inhibitor I9	AUL78271.1	0	0	0	0	0	0
putative ankyrin repeat protein	AUL78278.1	17921000	40008000	67041000	44264000	0	0
hypothetical protein	AUL78280.1	38436000	50207000	73606000	59663000	0	0
hypothetical protein	AUL78287.1	1279600000	4586800000	1105100000	1942200000	0	52709000
hypothetical protein	AUL78288.1	531570000	1019000000	438730000	282790000	0	75975000
hypothetical protein	AUL78292.1	478630000	962160000	101770000	287030000	0	0
hypothetical protein	AUL78295.1	0	0	0	0	0	0
hypothetical protein	AUL78301.1	0	62361000	59558000	101610000	0	0
dna-directed rna polymerase subunit 1	AUL78302.1	172350000	1169500000	476820000	1350700000	0	20654000
hypothetical protein	AUL78318.1	0	0	0	44539000	0	0
hypothetical protein	AUL78319.1	72902000	134510000	149520000	112050000	0	0
peptidase inhibitor I9	AUL78329.1	0	0	0	0	0	0
peptidase inhibitor I9	AUL78330.1	0	0	0	0	0	1192300
hypothetical protein	AUL78347.1	0	0	0	0	0	33895000
hypothetical protein	AUL78348.1	6546500	13116000	0	0	0	4899700
hypothetical protein	AUL78354.1	42106000	0	0	0	0	57517000

Table A5 (cont'd)

Protein	Accession ID	LFQ (Label Free Quantification)					
		MA 2	MA 3	Pellet 2	Pellet 3	Supe 2	Supe 3
intein-containing DNA-directed RNA polymerase subunit 2	AUL78361.1	35569000	167750000	158640000	184920000	0	0
intein-containing dna-directed rna polymerase subunit 2	AUL78362.1	123160000	458380000	265510000	435270000	0	3039500
DNA-directed RNA polymerase subunit 6	AUL78368.1	61122000	51978000	0	14752000	0	10963000
putative fibril associated protein	AUL78400.1	4061300000	7791300000	7252300000	8333300000	0	91195000
putative major capsid protein	AUL78403.1	0	15421000	0	15004000	0	0
putative neurocan core protein	AUL78405.1	0	0	0	0	0	224820
hypothetical protein	AUL78410.1	0	0	0	6603400	0	0
putative ORFan	AUL78423.1	0	0	0	15241000	0	0
putative ORFan	AUL78440.1	0	0	0	5572500	0	0
hypothetical protein	AUL78464.1	0	792810000	150010000	285700000	0	86069000
hypothetical protein	AUL78466.1	62827000	133580000	106400000	42358000	0	8900300
hypothetical protein	AUL78468.1	49180000	177010000	39800000	70266000	0	97986000
hypothetical protein	AUL78479.1	0	0	0	7546600	0	0
hypothetical protein	AUL78481.1	51089000	322970000	117130000	232420000	0	1246200
putative ORFan	AUL78496.1	0	13607000	0	20865000	0	0
hypothetical protein	AUL78500.1	0	0	0	11704000	0	0
putative ORFan	AUL78514.1	0	0	0	0	0	4012200
putative ORFan	AUL78530.1	0	0	0	0	0	0
hypothetical protein	AUL78545.1	0	0	0	33230000	0	0
hypothetical protein	AUL78577.1	0	144940000	129500000	108170000	0	0
putative lipocalin	AUL78583.1	0	0	0	7057200	0	0
putative ORFan	AUL78586.1	0	0	0	0	0	0
Tlr 6Fp protein	AUL78587.1	0	0	0	0	0	0
hypothetical protein	AUL78601.1	196380000	168190000	201330000	148480000	0	0

Table A5 (cont'd)

Protein	Accession ID	LFQ (Label Free Quantification)					
		MA 2	MA 3	Pellet 2	Pellet 3	Supe 2	Supe 3
putative protein kinase	AUL78629.1	389350000	646020000	577820000	371190000	0	11539000
Ig family protein	AUL78630.1	244440000	0	0	0	0	476700000
hypothetical protein	AUL78631.1	0	0	0	0	0	7640700
putative ORFan	AUL78635.1	180320000	468830000	109910000	387480000	0	0
chemotaxis	AUL78637.1	36240000	119120000	90418000	96896000	0	0
hypothetical protein	AUL78639.1	0	0	0	0	0	0
hypothetical protein	AUL78659.1	0	262470000	0	0	0	0
hypothetical protein	AUL78681.1	96179000	409910000	37016000	51203000	0	106410000
putative chemotaxis protein ched	AUL78687.1	13177000	0	18416000	0	0	0
bifunctional metalloprotease ubiquitin-protein ligase	AUL78691.1	21510000	99824000	50320000	77730000	0	0
putative ORFan	AUL78708.1	0	0	0	0	0	0
hypothetical protein	AUL78717.1	0	0	0	7813300	0	0
putative CfxQ-like protein	AUL78719.1	0	476800000	329480000	592800000	0	0
glutaredoxin	AUL78724.1	43362000	108360000	21599000	0	0	56970000
hypothetical protein	AUL78731.1	21578000	38161000	0	0	0	0
hypothetical protein	AUL78736.1	0	0	0	8150600	0	0
DNA-dependent RNA polymerase subunit Rpb9	AUL78739.1	85272000	195640000	169410000	112720000	0	0
p87	AUL78740.1	0	0	0	10833000	0	0
DNA-directed RNA polymerase subunit 6	AUL78741.1	0	0	82948000	85312000	0	0
Actin-1*	CAA23399.1	42317000	35244000	108310000	28011000	19116000	40395000
Ubiquitin-60S ribosomal protein L40*	CAA53293.1	0	0	0	0	0	13388000
Myosin-2 heavy chain*	CAA68663.1	436370000	0	0	0	0	0

Table A.6

Protein	Accession ID	# of Peptides						Unique Peptides						Sequence Coverage (%)					
		MA		Pellet		Supe		MA		Pellet		Supe		MA		Pellet		Supe	
		2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3
profilin, chain A*	AAA27710.1	1	1	0	0	0	1	1	1	0	0	0	1	12	12	0	0	0	12
Cytochrome c oxidase subunit 1+2*	AAD11820.1	0	1	0	1	0	0	0	1	0	1	0	0	0	2.3	0	2.3	0	0
ATP synthase subunit 9*	AOS85732.1	1	1	1	1	0	1	1	1	1	1	0	1	12.7	12.7	12.7	12.7	0	12.7
ATP synthase subunit alpha*	AOS85698.1	0	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	2.1	2.1
ubiquitin-like protein Ublp94.4*	AAQ16627.1	2	1	1	1	1	2	1	0	0	0	0	1	2.6	1.1	1.1	1.1	1.1	2.6
iron-superoxide dismutase*	AAT91955.1	1	1	1	0	0	0	1	1	1	0	0	0	5.1	5.1	5.1	0	0	0
lactate dehydrogenase-like protein, partial*	ABD46577.1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	16.9	0	0
encystation-mediating serine proteinase*	ABY63398.1	1	0	0	0	1	2	1	0	0	0	1	2	1.9	0	0	0	1.9	4.4
protein kinase C8, partial (plastid)*	AFD36237.1	1	1	1	1	0	1	1	1	1	1	0	1	2.1	2.1	2.1	2.1	0	2.1
fork head domain-containing protein	AUL77470.1	0	1	1	1	0	0	0	1	1	1	0	0	0	4.3	4.3	4.3	0	0
putative ORFan	AUL77479.1	0	1	1	0	0	1	0	1	1	0	0	1	0	17.4	17.4	0	0	17.4
hypothetical protein	AUL77482.1	1	1	1	1	0	0	1	1	1	1	0	0	1.6	1.6	1.6	1.6	0	0
hypothetical protein	AUL77486.1	1	1	1	1	0	0	1	1	1	1	0	0	2.9	2.9	2.9	2.9	0	0
hypothetical protein	AUL77492.1	3	3	3	3	0	0	3	3	3	3	0	0	5.8	5.8	5.8	5.8	0	0
mg749 protein	AUL77517.1	1	1	1	1	0	1	1	1	1	1	0	1	5.3	5.3	5.3	5.3	0	5.3
tyrosine-protein phosphatase	AUL77518.1	3	1	3	2	0	0	3	1	3	2	0	0	11	3.7	11	4.1	0	0

Table A.6 TV Peptide Counts and Sequence Coverage. Number of peptides identified via mass spectrometry for the TV proteins. The number of peptide counts identified in each sample, as well as the percentage of the protein sequence that is covered by these counts, is reported.
**Acanthamoeba castellanii* protein

Table A6 (cont'd)

Protein	Accession ID	# of Peptides						Unique Peptides						Sequence Coverage (%)					
		MA		Pellet		Supe		MA		Pellet		Supe		MA		Pellet		Supe	
		2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3
putative ORFan	AUL77532.1	2	2	2	2	0	1	2	2	2	2	0	1	3.9	3.9	3.9	3.9	0	2.4
alkylated dna repair protein alkb-like 8 isoform x1	AUL77532.1	1	2	1	2	0	0	1	2	1	2	0	0	3.9	6.9	3.9	6.9	0	0
mannose-6P isomerase	AUL77553.1	1	2	0	1	0	1	1	2	0	1	0	1	7.5	14.5	0	7.5	0	7.5
hypothetical protein	AUL77579.1	0	0	0	0	0	3	0	0	0	0	0	3	0	0	0	0	0	8.8
putative oxireductase	AUL77599.1	4	7	3	7	0	2	4	7	3	7	0	2	6.9	14.1	5.9	14	0	2.2
hypothetical protein	AUL77600.1	2	2	3	2	0	2	2	2	3	2	0	2	5.4	5.4	7.6	5.4	0	5.4
ubiquitin-conjugating enzyme e2	AUL77610.1	1	1	1	1	0	0	1	1	1	1	0	0	4.5	4.5	4.5	4.5	0	0
hypothetical protein	AUL77622.1	0	1	0	1	0	0	0	1	0	1	0	0	0	2.9	0	2.9	0	0
hypothetical protein	AUL77647.1	5	4	4	4	0	0	5	4	4	4	0	0	7.8	7.2	5.8	7.2	0	0
structural ppiase-like protein	AUL77649.1	3	4	4	4	0	1	3	4	4	4	0	1	13.1	13.5	13.5	13.5	0	3.7
thiol oxidoreductase E10R	AUL77655.1	2	2	2	2	0	0	2	2	2	2	0	0	4.9	4.9	4.9	4.9	0	0
mg709 protein	AUL77661.1	2	3	2	3	0	2	2	3	2	3	0	2	14.2	17.6	14.2	17.6	0	14.2
hypothetical protein	AUL77666.1	2	4	3	3	0	0	2	4	3	3	0	0	15	49.1	15	42.5	0	0
putative ORFan	AUL77678.1	0	1	0	1	0	0	0	1	0	1	0	0	0	21.5	0	21.5	0	0
putative N-acetyl transferase	AUL77680.1	1	1	1	1	0	0	1	1	1	1	0	0	3.5	3.5	3.5	3.5	0	0
putative helicase	AUL77687.1	1	0	2	2	0	0	1	0	2	2	0	0	1.3	0	1.9	1.9	0	0
hypothetical protein	AUL77688.1	4	5	4	4	0	3	4	5	4	4	0	3	17.3	20.4	17.3	17.3	0	15.7

Table A6 (cont'd)

Protein	Accession ID	# of Peptides						Unique Peptides						Sequence Coverage (%)					
		MA		Pellet		Supe		MA		Pellet		Supe		MA		Pellet		Supe	
		2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3
hypothetical protein	AUL77694.1	5	7	6	6	0	3	5	7	6	6	0	3	10.6	17.6	11.5	11.5	0	9.7
glycosyl hydrolase family 18	AUL77711.1	2	2	2	2	0	0	2	2	2	2	0	0	3.5	3.5	3.5	3.5	0	0
hypothetical protein	AUL77718.1	1	2	1	1	0	1	1	2	1	1	0	1	1.1	2.4	1.1	1.1	0	1.1
putative DNA repair protein	AUL77721.1	0	1	0	0	0	0	0	1	0	0	0	0	0	0.9	0	0	0	0
hypothetical protein	AUL77723.1	6	7	6	5	0	6	6	7	6	5	0	6	11	16.9	13	13	0	11
putative ORFan	AUL77729.1	6	7	5	7	0	2	6	7	5	7	0	2	12	12	12	12	0	5.4
hypothetical protein	AUL77752.1	6	8	7	8	0	1	6	8	7	8	0	1	27.2	29	32.1	29	0	9.3
hypothetical protein	AUL77753.1	1	2	0	0	0	0	1	2	0	0	0	0	8.4	8.4	0	0	0	0
putative ATP-dependent RNA helicase	AUL77758.1	1	1	1	1	0	0	1	1	1	1	0	0	1.3	1.3	1.3	2.2	0	0
atp-dependent rna helicase	AUL77773.1	1	1	1	1	0	0	1	1	1	1	0	0	1.6	1.6	1.6	1.6	0	0
putative 5-3exonuclease20	AUL77795.1	1	4	3	4	0	0	1	4	3	4	0	0	1.7	7	5.7	7	0	0
phosphoesterase-like protein	AUL77796.1	4	4	3	3	0	0	4	4	3	3	0	0	6.1	6.1	4.5	4.5	0	0
hypothetical protein	AUL77813.1	1	1	1	0	0	0	1	1	1	0	0	0	6.4	6.4	6.4	0	0	0
hypothetical protein	AUL77820.1	2	1	2	2	0	1	2	1	2	2	0	1	2.2	0.9	2.2	2.2	0	0.9
kinesin-like protein	AUL77838.1	6	6	9	5	0	2	6	6	9	5	0	2	2.8	2.8	3.9	2.4	0	0.9
putative protein phosphatase 2c	AUL77859.1	3	3	3	3	0	0	3	3	3	3	0	0	8.6	8.6	8.6	8.6	0	0
hypothetical protein	AUL77863.1	2	2	2	2	0	0	2	2	2	2	0	0	10.4	10.4	10.4	10.4	0	0

Table A6 (cont'd)

Protein	Accession ID	# of Peptides						Unique Peptides						Sequence Coverage (%)					
		MA		Pellet		Supe		MA		Pellet		Supe		MA		Pellet		Supe	
		2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3
hypothetical protein	AUL77885.1	1	1	0	1	0	0	1	1	0	1	0	0	3.2	3.2	0	3.2	0	0
DNA polymerase family X	AUL77886.1	1	3	1	3	0	0	1	3	1	3	0	0	5.2	9.1	5.2	9.4	0	0
hypothetical protein	AUL77896.1	1	1	1	1	0	1	1	1	1	1	0	1	1.9	1.9	1.9	1.9	0	1.9
putative early transcription factor	AUL77899.1	8	11	14	15	0	1	8	11	14	15	0	1	4.9	7.9	10.9	11.5	0	0.3
hypothetical protein	AUL77902.1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	2.9	0	0	2.9
hypothetical protein	AUL77903.1	3	3	3	3	0	0	3	3	3	3	0	0	5.3	5.3	5.3	5.3	0	0
mg574 protein	AUL77905.1	1	2	1	1	0	0	1	2	1	1	0	0	7.1	12.1	7.1	7.1	0	0
hypothetical protein	AUL77907.1	9	9	9	9	0	7	9	9	9	9	0	7	44.2	44.2	44.2	44.2	0	42.6
mg18 protein	AUL77928.1	1	2	2	2	0	0	1	2	2	2	0	0	2.9	7.6	7.6	7.6	0	0
polyA polymerase catalitic subunit	AUL77929.1	6	6	10	8	0	2	6	6	10	8	0	2	7.8	11.5	17	14.9	0	4.3
hypothetical protein	AUL77930.1	2	3	0	1	0	2	2	3	0	1	0	2	1.7	2.9	0	1.7	0	1.7
hypothetical protein	AUL77933.1	5	4	5	5	0	1	5	4	5	5	0	1	5	4.2	4.8	4.8	0	1.2
hypothetical protein	AUL77936.1	6	5	5	5	0	3	6	5	5	5	0	3	10.4	8.6	8.6	8.6	0	6.9
SNF2 family helicase	AUL77941.1	4	6	5	7	0	0	4	6	5	7	0	0	8.1	11.1	9.3	12.6	0	0
hypothetical protein	AUL77944.1	2	2	4	3	0	2	2	2	4	3	0	2	12.3	11.7	11.7	11.7	0	12.3
putative ORFan	AUL77950.1	1	1	1	1	0	0	1	1	1	1	0	0	9.3	9.3	9.3	9.3	0	0
hypothetical protein	AUL77952.1	1	4	2	2	0	0	1	4	2	2	0	0	2.2	8.2	4	4	0	0
hypothetical protein	AUL77961.1	1	4	1	3	0	1	1	4	1	3	0	1	2.5	8.8	2.5	8.8	0	2.5

Table A6 (cont'd)

Protein	Accession ID	# of Peptides						Unique Peptides						Sequence Coverage (%)					
		MA		Pellet		Supe		MA		Pellet		Supe		MA		Pellet		Supe	
		2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3
hypothetical protein	AUL77962.1	0	1	1	1	0	0	0	1	1	1	0	0	0	7.7	7.7	7.7	0	0
thioredoxin domain-containing protein	AUL77963.1	6	8	5	7	0	5	6	8	5	7	0	5	17.9	19.8	14.2	17.9	0	11.8
putative ATP-dependent RNA helicase	AUL77999.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NUDIX hydrolase	AUL78015.1	0	1	0	0	0	0	0	1	0	0	0	0	0	3	0	0	0	0
dna-directed rna polymerase subunit	AUL78016.1	3	3	3	3	0	2	3	3	3	3	0	2	15.1	15.1	15.1	15.1	0	11.1
hypothetical protein	AUL78017.1	1	0	1	1	0	0	1	0	1	1	0	0	2.1	0	2.1	3.9	0	0
hypothetical protein	AUL78019.1	1	0	0	0	0	1	1	0	0	0	0	1	1.9	0	0	0	0	1.9
NTPase	AUL78021.1	9	11	10	8	0	1	9	11	10	8	0	1	9.5	12.4	10.2	7	0	0.8
putative leucine-rich repeat protein	AUL78028.1	1	0	1	0	0	1	1	0	1	0	0	1	4.2	0	2.7	0	0	4.2
mRNA capping enzyme	AUL78031.1	5	6	6	7	0	1	5	6	6	7	0	1	4.9	6.6	5.4	7	0	0.9
FtsJ-like methyl transferase	AUL78032.1	5	5	5	5	0	0	5	5	5	5	0	0	16	16	16	16	0	0
ubiquitin domain-containing protein	AUL78040.1	5	5	4	5	1	6	4	4	3	4	0	5	45	45	45	52.5	11.2	52.5
hypothetical protein	AUL78045.1	3	3	3	2	0	0	3	3	3	2	0	0	4.2	4.2	4.2	2.2	0	0
hypothetical protein	AUL78046.1	1	1	2	1	0	0	1	1	2	1	0	0	1.4	1.4	2.3	1.4	0	0
putative heat shock 70 kDa protein	AUL78049.1	1	1	1	1	0	0	1	1	1	1	0	0	3.9	3.9	5.4	3.9	0	0
hypothetical protein	AUL78055.1	2	4	2	4	0	3	2	4	2	4	0	3	14.8	20.2	14.8	20.2	0	14.8

Table A6 (cont'd)

Protein	Accession ID	# of Peptides						Unique Peptides						Sequence Coverage (%)					
		MA		Pellet		Supe		MA		Pellet		Supe		MA		Pellet		Supe	
		2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3
hypothetical protein	AUL78059.1	2	2	2	2	0	0	2	2	2	2	0	0	4.3	4.3	4.3	4.3	0	0
hypothetical protein	AUL78061.1	3	3	0	1	0	0	3	3	0	1	0	0	5.3	5.3	0	1.8	0	0
serine threonine-protein kinase	AUL78063.1	1	1	1	1	0	0	1	1	1	1	0	0	2.9	2.9	2.9	2.9	0	0
hypothetical protein	AUL78067.1	6	8	6	6	0	2	6	8	6	6	0	2	6.6	8.3	6.3	6.7	0	2.8
hypothetical protein	AUL78068.1	5	5	6	6	0	0	5	5	6	6	0	0	11	11	12.6	12.6	0	0
hypothetical protein	AUL78073.1	1	3	2	2	0	0	1	3	2	2	0	0	4.1	8.9	8.6	8.6	0	0
hypothetical protein	AUL78080.1	1	1	0	0	0	0	1	1	0	0	0	0	6.4	6.4	0	0	0	0
major core protein	AUL78082.1	28	31	34	31	0	18	28	31	34	31	0	18	37	45.5	49.6	46.6	0	29
putative ORFan	AUL78086.1	0	0	2	1	0	0	0	0	2	1	0	0	0	0	11.3	3.6	0	0
putative ORFan	AUL78088.1	6	6	1	2	0	6	6	6	1	2	0	6	38	33.3	12.5	12	0	38
hypothetical protein	AUL78091.1	1	1	1	1	0	1	1	1	1	1	0	1	4.1	4.1	4.1	4.1	0	4.1
hypothetical protein	AUL78092.1	1	1	1	2	0	1	1	1	1	2	0	1	3.4	3.4	3.4	3.4	0	3.4
hypothetical protein	AUL78093.1	4	4	4	4	0	0	4	4	4	4	0	0	9.5	9.5	9.5	9.5	0	0
catalase HPII	AUL78097.1	6	10	7	9	0	5	6	10	7	9	0	5	9.2	13.8	10.4	15.1	0	6.6
putative ORFan	AUL78106.1	1	1	0	0	0	0	1	1	0	0	0	0	5.1	5.1	0	0	0	0
hypothetical protein	AUL78108.1	1	1	1	1	0	1	1	1	1	1	0	1	7.4	7.4	7.4	7.4	0	7.4
dna topoisomerase 1b	AUL78109.1	6	9	6	6	0	0	6	9	6	6	0	0	11.5	16.5	13.5	11	0	0

Table A6 (cont'd)

Protein	Accession ID	# of Peptides						Unique Peptides						Sequence Coverage (%)					
		MA		Pellet		Supe		MA		Pellet		Supe		MA		Pellet		Supe	
		2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3
cyclopropane fatty acyl phospholipid synthase	AUL78111.1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	2.6	0	0
putative ORFan	AUL78114.1	0	0	1	1	0	0	0	0	1	1	0	0	0	0	6	6	0	0
extracellular ligand-binding receptor	AUL78119.1	1	1	1	1	0	0	1	1	1	1	0	0	3.7	3.7	3.7	3.7	0	0
putative ORFan	AUL78120.1	1	1	1	1	0	1	1	1	1	1	0	1	6.4	6.4	6.4	6.4	0	6.4
glyoxalase	AUL78134.1	1	1	1	1	0	1	1	1	1	1	0	1	5.9	5.9	5.9	5.9	0	4.9
hypothetical protein	AUL78135.1	9	8	8	10	0	5	9	8	8	10	0	5	25.2	24.6	22	27.8	0	12.5
putative ORFan	AUL78142.1	1	1	1	1	0	0	1	1	1	1	0	0	14.3	14.3	14.3	14.3	0	0
hypothetical protein	AUL78143.1	6	7	7	7	1	3	6	7	7	7	1	3	19.5	23.2	23.2	24.7	5.6	13.9
capsid protein 1	AUL78147.1	39	40	44	43	3	20	39	40	44	43	3	20	58.1	58.3	60.2	63.6	4.1	38.4
hypothetical protein	AUL78155.1	2	2	2	2	0	0	2	2	2	2	0	0	28.1	28.1	28.1	28.1	0	0
putative ORFan	AUL78156.1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1.2
hypothetical protein	AUL78183.1	1	1	2	0	0	1	1	1	2	0	0	1	0.3	0.3	1.7	0	0	0.3
hypothetical protein	AUL78191.1	25	29	23	28	1	17	25	29	23	28	1	17	17	18.1	14.9	18.1	0.7	9.8
thioredoxin domain-containing protein	AUL78192.1	1	2	2	1	0	0	1	2	2	1	0	0	14.1	14.1	14.1	14.1	0	0
hypothetical protein	AUL78198.1	2	2	1	2	0	0	2	2	1	2	0	0	4.4	4.4	3.9	4.4	0	0
putative ORFan	AUL78206.1	3	4	1	1	0	0	3	4	1	1	0	0	7.5	7.5	3.2	3.2	0	0
putative pore coat assembly factor	AUL78211.1	8	9	6	8	0	5	8	9	6	8	0	5	25.2	26.6	21.5	22.1	0	20.6
hypothetical protein	AUL78214.1	7	7	6	8	0	3	7	7	6	8	0	3	31.8	31.8	31.8	43	0	17.5

Table A6 (cont'd)

Protein	Accession ID	# of Peptides						Unique Peptides						Sequence Coverage (%)					
		MA		Pellet		Supe		MA		Pellet		Supe		MA		Pellet		Supe	
		2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3
hypothetical protein	AUL78219.1	7	7	7	8	0	3	7	7	7	8	0	3	30.4	30.4	30.4	32.5	0	15.8
hypothetical protein	AUL78232.1	3	3	2	3	0	2	3	3	2	3	0	2	17.3	17.3	13.4	17.3	0	13
dna directed rna polymerase subunit	AUL78244.1	2	2	2	2	0	0	2	2	2	2	0	0	7.5	6.4	7.5	7.5	0	0
hypothetical protein	AUL78246.1	9	10	12	9	0	1	9	10	12	9	0	1	5.6	6	7.4	5.6	0	0.6
putative ORFan	AUL78253.1	1	1	1	0	0	1	1	1	1	0	0	1	12.3	12.3	12.3	0	0	12.3
hypothetical protein	AUL78254.1	2	2	1	2	0	0	2	2	1	2	0	0	9.8	9.8	3.6	9.8	0	0
arylsulfatase	AUL78269.1	5	8	4	8	0	2	5	8	4	8	0	2	17.3	19.4	15.7	19.4	0	9.8
peptidase inhibitor I9	AUL78271.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
putative ankyrin repeat protein	AUL78278.1	5	3	5	3	0	0	5	3	5	3	0	0	2.9	1.9	2.9	1.9	0	0
hypothetical protein	AUL78280.1	2	2	2	2	0	1	2	2	2	2	0	1	10.1	10.1	10.1	10.1	0	10.1
hypothetical protein	AUL78287.1	11	13	9	11	1	4	11	13	9	11	1	4	7.4	7.6	7.2	7.6	1.4	5
hypothetical protein	AUL78288.1	3	5	3	3	0	3	3	5	3	3	0	3	17.4	26	17.4	17.4	0	17.4
hypothetical protein	AUL78292.1	2	2	3	2	0	2	2	2	3	2	0	2	13.1	13.1	22	13.1	0	13.1
hypothetical protein	AUL78295.1	0	1	0	0	0	0	0	1	0	0	0	0	0	19.1	0	0	0	0
hypothetical protein	AUL78301.1	1	2	2	2	0	0	1	2	2	2	0	0	12.9	32.9	32.9	32.9	0	0
dna-directed rna polymerase subunit	AUL78302.1	22	23	29	25	1	5	22	23	29	25	1	5	19	18.7	23.3	18.8	0.7	4.8
hypothetical protein	AUL78318.1	1	2	1	2	0	0	1	2	1	2	0	0	2.3	11.7	2.3	11.7	0	0

Table A6 (cont'd)

Protein	Accession ID	# of Peptides						Unique Peptides						Sequence Coverage (%)					
		MA		Pellet		Supe		MA		Pellet		Supe		MA		Pellet		Supe	
		2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3
hypothetical protein	AUL78319.1	4	4	4	3	0	0	4	4	4	3	0	0	6.5	6.5	6.5	6.5	0	0
peptidase inhibitor I9	AUL78329.1	0	1	0	0	0	0	0	1	0	0	0	0	0	5.5	0	0	0	0
peptidase inhibitor I9	AUL78330.1	1	1	1	1	0	1	1	1	1	1	0	1	1.4	1.4	1.4	1.4	0	1.4
hypothetical protein	AUL78347.1	1	1	1	1	0	1	1	1	1	1	0	1	1.7	1.7	1.7	1.7	0	1.7
hypothetical protein	AUL78348.1	2	3	0	0	0	3	2	3	0	0	0	3	10.1	10.8	0	0	0	10.8
hypothetical protein	AUL78354.1	2	1	1	1	0	4	2	1	1	1	0	4	13	7	7	7	0	15.8
intein-containing DNA-directed RNA polymerase subunit 2	AUL78361.1	6	7	7	5	0	0	6	7	7	5	0	0	8.9	10	9.9	7.7	0	0
intein-containing dna-directed rna polymerase subunit 2	AUL78362.1	6	7	7	7	0	2	6	7	7	7	0	2	7.8	7.8	7.8	7.8	0	3.6
DNA-directed RNA polymerase subunit 6	AUL78368.1	2	2	2	2	0	5	2	2	2	2	0	5	13.7	13.7	13.7	12.2	0	24.4
putative fibril associated protein	AUL78400.1	12	15	15	13	0	6	12	15	15	13	0	6	29.6	35.6	35.6	30.8	0	14.6
putative major capsid protein	AUL78403.1	1	2	1	2	0	0	1	2	1	2	0	0	2.4	4.7	2.4	4.7	0	0
putative neurocan core protein	AUL78405.1	1	1	1	1	0	1	1	1	1	1	0	1	5.1	5.1	5.1	5.1	0	5.1
hypothetical protein	AUL78410.1	1	0	1	1	0	0	1	0	1	1	0	0	2.2	0	2.2	2.2	0	0
putative ORFan	AUL78423.1	0	1	0	1	0	0	0	1	0	1	0	0	0	12.6	0	12.6	0	0
putative ORFan	AUL78440.1	1	1	1	1	0	0	1	1	1	1	0	0	5.3	5.3	5.3	5.3	0	0

Table A6 (cont'd)

Protein	Accession ID	# of Peptides						Unique Peptides						Sequence Coverage (%)					
		MA		Pellet		Supe		MA		Pellet		Supe		MA		Pellet		Supe	
		2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3
hypothetical protein	AUL78464.1	2	2	2	2	0	2	2	2	2	2	0	2	14.7	14.7	14.7	14.7	0	14.7
hypothetical protein	AUL78466.1	2	4	2	3	0	3	2	4	2	3	0	3	12.4	27.2	12.4	15.4	0	15.4
hypothetical protein	AUL78468.1	2	3	3	2	0	3	2	3	3	2	0	3	5.7	12.3	12.3	5.7	0	12.3
hypothetical protein	AUL78479.1	1	1	1	1	0	0	1	1	1	1	0	0	4	4	4	4	0	0
hypothetical protein	AUL78481.1	3	2	3	3	0	1	3	2	3	3	0	1	5.9	3.3	5.9	5.9	0	1.5
putative ORFan	AUL78496.1	0	1	1	1	0	0	0	1	1	1	0	0	0	8.9	8.9	8.9	0	0
hypothetical protein	AUL78500.1	0	1	0	1	0	0	0	1	0	1	0	0	0	9.2	0	9.2	0	0
putative ORFan	AUL78514.1	1	1	1	1	0	1	1	1	1	1	0	1	14.9	14.9	14.9	14.9	0	14.9
putative ORFan	AUL78530.1	1	1	1	1	0	0	1	1	1	1	0	0	3.3	3	3	3.3	0	0
hypothetical protein	AUL78545.1	1	1	1	2	0	1	1	1	1	2	0	1	3.5	3.5	3.5	5.8	0	3.5
hypothetical protein	AUL78577.1	1	2	3	2	0	0	1	2	3	2	0	0	1.7	8	8	8	0	0
putative lipocalin	AUL78583.1	0	1	0	1	0	0	0	1	0	1	0	0	0	3.9	0	3.9	0	0
putative ORFan	AUL78586.1	1	1	1	0	0	0	1	1	1	0	0	0	2	2	2	0	0	0
Tlr 6Fp protein	AUL78587.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
hypothetical protein	AUL78601.1	2	4	2	3	0	1	2	4	2	3	0	1	11.6	14.4	11.6	7.7	0	5
putative protein kinase	AUL78629.1	3	3	3	3	0	2	3	3	3	3	0	2	4.1	4.1	4.1	4.1	0	2.3
Ig family protein	AUL78630.1	2	1	1	0	1	3	2	1	1	0	1	3	4.6	1.5	1.5	0	3.1	5.9
hypothetical protein	AUL78631.1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	7.1
putative ORFan	AUL78635.1	2	2	2	1	0	1	2	2	2	1	0	1	8.1	8.1	8.1	4.3	0	3.8
chemotaxis	AUL78637.1	3	4	3	4	0	0	3	4	3	4	0	0	7.1	9.3	7.1	9.3	0	0
hypothetical protein	AUL78639.1	0	1	0	0	0	0	0	1	0	0	0	0	0	2.1	0	0	0	0

Table A6 (cont'd)

Protein	Accession ID	# of Peptides						Unique Peptides						Sequence Coverage (%)					
		MA		Pellet		Supe		MA		Pellet		Supe		MA		Pellet		Supe	
		2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3
hypothetical protein	AUL78659.1	1	1	1	1	0	0	1	1	1	1	0	0	8.2	8.2	8.2	8.2	0	0
hypothetical protein	AUL78681.1	2	2	2	2	0	2	2	2	2	2	0	2	2	2	2	2	0	2
putative chemotaxis protein ched	AUL78687.1	2	1	2	1	0	0	2	1	2	1	0	0	8.6	3.7	8.6	3.7	0	0
bifunctional metalloprotease ubiquitin-protein ligase	AUL78691.1	2	3	3	3	0	1	2	3	3	3	0	1	3	4.3	4.3	4.3	0	1.1
putative ORFan	AUL78708.1	1	1	0	0	0	0	1	1	0	0	0	0	3	3	0	0	0	0
hypothetical protein	AUL78717.1	0	1	0	1	0	0	0	1	0	1	0	0	0	1.9	0	1.9	0	0
putative CfxQ-like protein	AUL78719.1	2	2	2	2	0	1	2	2	2	2	0	1	6.7	6.7	6.7	6.7	0	2.1
glutaredoxin	AUL78724.1	2	2	2	1	0	2	2	2	2	1	0	2	24.8	24.8	24.8	14.9	0	24.8
hypothetical protein	AUL78731.1	2	2	0	1	0	0	2	2	0	1	0	0	15.4	15.4	0	9.5	0	0
hypothetical protein	AUL78736.1	1	1	1	1	0	0	1	1	1	1	0	0	0.7	0.7	0.7	0.7	0	0
DNA-dependent RNA polymerase subunit Rpb9	AUL78739.1	2	4	2	3	0	2	2	4	2	3	0	2	13.9	25.3	13.9	19.6	0	12.4
p87	AUL78740.1	0	0	1	1	0	0	0	0	1	1	0	0	0	0	2.8	2.8	0	0
DNA-directed RNA polymerase subunit 6	AUL78741.1	1	1	2	2	0	0	1	1	2	2	0	0	3.8	3.8	6.4	6.4	0	0
Actin-1*	CAA23399.1	3	4	6	2	5	5	3	4	6	2	5	5	7.5	13.6	19.5	7.8	14.7	15.8
Ubiquitin-60S ribosomal protein L40*	CAA53293.1	1	1	1	1	1	2	0	0	0	0	0	1	7	7	7	7	7	21.1
Myosin-2 heavy chain*	CAA68663.1	2	1	1	1	0	0	2	1	1	1	0	0	2.9	1.6	1.6	1.6	0	0

APPENDIX B

SUPPLEMENTARY VIDEOS

These Supplementary Materials were originally published in *Viruses* and as a preprint at bioRxiv. This work is reused here under the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/). Links to the supplemental movies can be found in the Movie Legends.

Schrad, J.R., Young, E.J., Abrahão, J.S., Cortines, J.R., Parent, K.N. 2017. Microscopic Characterization of the Brazilian Giant Samba Virus. *Viruses* doi:10.3390/v9020030.

Schrad, J.R., Abrahão, J.S., Cortines, J.R., Parent, K.N. 2019. Boiling Acid Mimics Intracellular Giant Virus Genome Release. *Cell* (in revision, preprint available through bioRxiv doi: https://doi.org/10.1101/777854).

SUPPLEMENTAL VIDEOS

Supplementary Video 1: Z-slices of a representative SMBV tomogram (central section depicted in Figure 2.3B).

Supplementary Video 2: Untreated SMBV Bubblegram Imaging. Bubblegram image series of a native SMBV particle demonstrating the buildup of radiation damage over time. A clear star-shaped radiation damage pattern is observed around the 11:00 position on the particle. Each frame represents a two second exposure (14 e-/Å^2). Total exposure time = 24 seconds (~140 e-/Å^2). Related to Figure 3.1.

Supplementary Video 3: Untreated SMBV Tomogram. Slice-by-slice view of a tomogram of a native SMBV particle. Related to Figure 3.2B-C.

Supplementary Video 4: Low pH-Treated SMBV Tomogram. Slice-by-slice view of a tomogram of a pH 2-treated SMBV particle. Note the opening in the stargate vertex as well as the sac exiting the capsid. Related to Figure 3.2F-G.

Supplementary Video 5: Tomogram of SMBV Incubated at High Temperature. Slice-by-slice view of a tomogram from an SMBV particle incubated at 100 °C for 6 hours. Note the fully open stargate vertex, the exodus of the nucleocapsid, and the apparent tethers between the capsid and the nucleocapsid. Related to Figure 3.2J-K.

Supplementary Video 6: Tilt Series of High Temperature Incubated SMBV. Tilt series of an SMBV particle incubated at 100 °C. Tilts were acquired every 2 degrees ranging from +/- 50 degrees. Related to Figure 3.2J-K

Supplementary Video 7: Low pH and High Temperature-Treated SMBV Tomogram. Slice-by-slice view of a tomogram of an SMBV particle treated with both low pH and high temperature. Tomogram segmentation was carried out using Amira v2019.2. Colors represent the following: Red- Outer Capsid Layer, Orange- Inner Capsid Layer, Blue- Starfish Seal Complex, and Yellow- Lipid. Note the flexibility of the innermost capsid layer and the residual density within the capsid interior. Related to Figure 3.2N-O.

Supplementary Video 8: Low pH and High Temperature Treated SMBV Tilt Series. Tilt series of an SMBV particle treated with both pH 2 and 100 °C. Tilts were acquired every 2° ranging from +/- 50°degrees. Related to Figure 3.2N-O.

Supplementary Video 9: Low pH and High Temperature-Treated SMBV Tomogram. Slice-by-slice view of a tomogram of five SMBV particles treated with both low pH and high temperature. These particles all have open stargate vertices, and one is oriented in a top-down view, providing additional structural information about the SMBV particle.

Supplementary Video 10: Low pH and High Temperature-Treated SMBV Tilt Series. Tilt series of an SMBV particle treated with both pH 2 and 100 °C. Tilts were acquired every 2° ranging from +/- 50°degrees. Five distinct SMBV particles are visible within this tilt series.

REFERENCES

REFERENCES

1. R. W. Hendrix, Bacteriophage HK97: assembly of the capsid and evolutionary connections. *Advances in virus research* **64**, 1-14 (2005).

2. D. L. Caspar, A. Klug, Physical principles in the construction of regular viruses. *Cold Spring Harbor symposia on quantitative biology* **27**, 1-24 (1962).

3. J. Abrahao *et al.*, Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. *Nature communications* **9**, 749 (2018).

4. A. Andrade *et al.*, Ubiquitous giants: a plethora of giant viruses found in Brazil and Antarctica. *Virology journal* **15**, 22 (2018).

5. E. V. Koonin, M. Krupovic, N. Yutin, Evolution of double-stranded DNA viruses of eukaryotes: from bacteriophages to transposons to giant viruses. *Annals of the New York Academy of Sciences* **1341**, 10-24 (2015).

6. E. V. Koonin, N. Yutin, Origin and evolution of eukaryotic large nucleo-cytoplasmic DNA viruses. *Intervirology* **53**, 284-292 (2010).

7. S. R. Casjens, I. J. Molineux, Short noncontractile tail machines: adsorption and DNA delivery by podoviruses. *Advances in experimental medicine and biology* **726**, 143-179 (2012).

8. A. R. Davidson, L. Cardarelli, L. G. Pell, D. R. Radford, K. L. Maxwell, Long noncontractile tail machines of bacteriophages. *Advances in experimental medicine and biology* **726**, 115-142 (2012).

9. P. G. Leiman, M. M. Shneider, Contractile tail machines of bacteriophages. *Advances in experimental medicine and biology* **726**, 93-114 (2012).

10. D. Prangishvili, The wonderful world of archaeal viruses. *Annual review of microbiology* **67**, 565-585 (2013).

11. E. R. Quemin *et al.*, First insights into the entry process of hyperthermophilic archaeal viruses. *Journal of virology* **87**, 13379-13385 (2013).

12. J. M. Claverie, C. Abergel, Mimivirus and its virophage. *Annual review of genetics* **43**, 49-66 (2009).

13. M. G. Fischer, Giant viruses come of age. *Current opinion in microbiology* **31**, 50-57 (2016).

14. A. G. Kutikhin, A. E. Yuzhalin, E. B. Brusina, Mimiviridae, Marseilleviridae, and virophages as emerging human pathogens causing healthcare-associated infections. *GMS hygiene and infection control* **9**, Doc16 (2014).

15. S. T. Abedon, T. D. Herschler, D. Stopar, Bacteriophage latent-period evolution as a response to resource availability. *Applied and environmental microbiology* **67**, 4233-4241 (2001).

16. A. J. McMichael, Environmental and social influences on emerging infectious diseases: past, present and future. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **359**, 1049-1058 (2004).

17. B. Muhlemann *et al.*, Ancient hepatitis B viruses from the Bronze Age to the Medieval period. *Nature* **557**, 418-423 (2018).

18. Y. A. Helmy, H. El-Adawy, E. M. Abdelwhab, A Comprehensive Review of Common Bacterial, Parasitic and Viral Zoonoses at the Human-Animal Interface in Egypt. *Pathogens* **6**, (2017).

19. S. K. Cohn, Pandemics: waves of disease, waves of hate from the Plague of Athens to A.I.D.S. *Historical journal* **85**, 535-555 (2012).

20. C. Theves, E. Crubezy, P. Biagini, History of Smallpox and Its Spread in Human Populations. *Microbiology spectrum* **4**, (2016).

21. K. F. Shortridge, The 1918 'Spanish' flu: pearls from swine? *Nature medicine* **5**, 384-385 (1999).

22. F. Javed *et al.*, Zika virus: what we need to know? *Journal of basic microbiology* **58**, 3-16 (2018).

23. V. A. Kostyuchenko *et al.*, Structure of the thermally stable Zika virus. *Nature* **533**, 425-428 (2016).

24. S. S. Jadav, A. Kumar, M. J. Ahsan, V. Jayaprakash, Ebola virus: current and future perspectives. *Infectious disorders drug targets* **15**, 20-31 (2015).

25. D. Baltimore, Expression of animal virus genomes. *Bacteriological reviews* **35**, 235-241 (1971).

26. E. V. Koonin, V. V. Dolja, M. Krupovic, Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology* **479-480**, 2-25 (2015).

27. D. A. Kennedy *et al.*, Industry-Wide Surveillance of Marek's Disease Virus on Commercial Poultry Farms. *Avian diseases* **61**, 153-164 (2017).

28. P. Roy, Bluetongue virus structure and assembly. *Current opinion in virology* **24**, 115-123 (2017).

29. S. C. Han, H. C. Guo, S. Q. Sun, Three-dimensional structure of foot-and-mouth disease virus and its biological functions. *Archives of virology* **160**, 1-16 (2015).

30. V. N. Fondong, The Search for Resistance to Cassava Mosaic Geminiviruses: How Much We Have Accomplished, and What Lies Ahead. *Frontiers in plant science* **8**, 408 (2017).

31. C. C. Kao, P. Ni, M. Hema, X. Huang, B. Dragnea, The coat protein leads the way: an update on basic and applied studies with the Brome mosaic virus coat protein. *Molecular plant pathology* **12**, 403-412 (2011).

32. I. Ahmad, F. Ahmed, J. Pichtel, *Microbes and microbial technology : agricultural and environmental applications*. (Springer, New York, 2011), pp. xvi, 516 p.

33. M. B. Marco, V. B. Suarez, A. Quiberoni, S. A. Pujato, Inactivation of Dairy Bacteriophages by Thermal and Chemical Treatments. *Viruses* **11**, (2019).

34. S. A. Pujato, A. Quiberoni, D. J. Mercanti, Bacteriophages on dairy foods. *Journal of applied microbiology* **126**, 14-30 (2019).

35. M. G. Ison, R. T. Hayden, Adenovirus. *Microbiology spectrum* **4**, (2016).

36. D. Z. Rechenchoski, L. C. Faccin-Galhardi, R. E. C. Linhares, C. Nozawa, Herpesvirus: an underestimated virus. *Folia microbiologica* **62**, 151-156 (2017).

37. E. Robilotti, S. Deresinski, B. A. Pinsky, Norovirus. *Clinical microbiology reviews* **28**, 134-164 (2015).

38. L. Zerboni, N. Sen, S. L. Oliver, A. M. Arvin, Molecular mechanisms of varicella zoster virus pathogenesis. *Nature reviews. Microbiology* **12**, 197-210 (2014).

39. S. R. Carding, N. Davis, L. Hoyles, Review article: the human intestinal virome in health and disease. *Alimentary pharmacology & therapeutics* **46**, 800-815 (2017).

40. K. M. Wylie, The Virome of the Human Respiratory Tract. *Clinics in chest medicine* **38**, 11-19 (2017).

41. N. A. Molinari *et al.*, The annual impact of seasonal influenza in the US: measuring disease burden and costs. *Vaccine* **25**, 5086-5096 (2007).

42. P. Gemski, Jr., B. A. Stocker, Transduction by bacteriophage P22 in nonsmooth mutants of Salmonella typhimurium. *Journal of bacteriology* **93**, 1588-1597 (1967).

43. A. Kruger, P. M. Lucchesi, Shiga toxins and stx phages: highly diverse entities. *Microbiology* **161**, 451-462 (2015).

44. M. G. Weinbauer, Ecology of prokaryotic viruses. *FEMS microbiology reviews* **28**, 127-181 (2004).

45. M. G. Weinbauer, F. Rassoulzadegan, Are viruses driving microbial diversification and diversity? *Environmental microbiology* **6**, 1-11 (2004).

46. E. Lara *et al.*, Unveiling the role and life strategies of viruses from the surface to the dark ocean. *Science advances* **3**, e1602565 (2017).

47. S. Haase, A. Sciocco-Cap, V. Romanowski, Baculovirus insecticides in Latin America: historical overview, current status and future perspectives. *Viruses* **7**, 2230-2267 (2015).

48. S. Riedel, Edward Jenner and the history of smallpox and vaccination. *Proceedings* **18**, 21-25 (2005).

49. M. F. Naso, B. Tomkowicz, W. L. Perry, 3rd, W. R. Strohl, Adeno-Associated Virus (AAV) as a Vector for Gene Therapy. *BioDrugs : clinical immunotherapeutics, biopharmaceuticals and gene therapy* **31**, 317-334 (2017).

50. J. L. Santiago-Ortiz, D. V. Schaffer, Adeno-associated virus (AAV) vectors in cancer gene therapy. *Journal of controlled release : official journal of the Controlled Release Society* **240**, 287-301 (2016).

51. R. T. Schooley *et al.*, Development and Use of Personalized Bacteriophage-Based Therapeutic Cocktails To Treat a Patient with a Disseminated Resistant Acinetobacter baumannii Infection. *Antimicrobial agents and chemotherapy* **61**, (2017).

52. N. Chanishvili, Phage therapy--history from Twort and d'Herelle through Soviet experience to current approaches. *Advances in virus research* **83**, 3-40 (2012).

53. B. K. Chan *et al.*, Phage treatment of an aortic graft infected with Pseudomonas aeruginosa. *Evolution, medicine, and public health* **2018**, 60-66 (2018).

54. R. Voelker, FDA Approves Bacteriophage Trial. *Jama* **321**, 638 (2019).

55. A. D. Hershey, M. Chase, Independent functions of viral protein and nucleic acid in growth of bacteriophage. *The Journal of general physiology* **36**, 39-56 (1952).

56. A. Pingoud, G. G. Wilson, W. Wende, Type II restriction endonucleases--a historical perspective and more. *Nucleic acids research* **42**, 7489-7527 (2014).

57. W. Wang *et al.*, Bacteriophage T7 transcription system: an enabling tool in synthetic biology. *Biotechnology advances* **36**, 2129-2137 (2018).

58. A. Pawluk, A. R. Davidson, K. L. Maxwell, Anti-CRISPR: discovery, mechanism and function. *Nature reviews. Microbiology* **16**, 12-17 (2018).

59. F. Jiang, J. A. Doudna, CRISPR-Cas9 Structures and Mechanisms. *Annual review of biophysics* **46**, 505-529 (2017).

60. S. Zhang *et al.*, Neutralization mechanism of a highly potent antibody against Zika virus. *Nature communications* **7**, 13679 (2016).

61. D. V. Rakhuba, E. I. Kolomiets, E. S. Dey, G. I. Novik, Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell. *Polish journal of microbiology* **59**, 145-155 (2010).

62. M. Legendre *et al.*, In-depth study of Mollivirus sibericum, a new 30,000-y-old giant virus infecting Acanthamoeba. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E5327-5335 (2015).

63. M. Legendre *et al.*, Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 4274-4279 (2014).

64. M. Boyer *et al.*, Mimivirus shows dramatic genome reduction after intraamoebal culture. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 10296-10301 (2011).

65. M. Boyer *et al.*, Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 21848-21853 (2009).

66. S. Dhindwal, B. Avila, S. Feng, R. Khayat, Porcine Circovirus 2 Uses a Multitude of Weak Binding Sites To Interact with Heparan Sulfate, and the Interactions Do Not Follow the Symmetry of the Capsid. *Journal of virology* **93**, (2019).

67. S. E. Jacobs, D. M. Lamson, K. St George, T. J. Walsh, Human rhinoviruses. *Clinical microbiology reviews* **26**, 135-162 (2013).

68. T. S. Baker, N. H. Olson, S. D. Fuller, Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. *Microbiology and molecular biology reviews : MMBR* **63**, 862-922, table of contents (1999).

69. L. B. Dustin, B. Bartolini, M. R. Capobianchi, M. Pistello, Hepatitis C virus: life cycle in cells, infection and host response, and analysis of molecular markers influencing the outcome of infection and response to therapy. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **22**, 826-832 (2016).

70. T. Olszak, A. Latka, B. Roszniowski, M. A. Valvano, Z. Drulis-Kawa, Phage Life Cycles Behind Bacterial Biodiversity. *Current medicinal chemistry* **24**, 3987-4001 (2017).

71. I. J. Molineux, D. Panja, Popping the cork: mechanisms of phage genome ejection. *Nature reviews. Microbiology* **11**, 194-204 (2013).

72. J. C. Snyder *et al.*, Insights into a viral lytic pathway from an archaeal virus-host system. *Journal of virology* **87**, 2186-2192 (2013).

73. M. W. Beijerinck, *Ueber ein Contagium vivum fluidum als Ursache der Fleckenkrankheit der Tabaksblätter*. (*Verh. Akad. Wet. Amst.*, 1898).

74. D. Iwanowski, Über die Mosaikkrankheit der Tabakspflanze. *Bull. Acad. Imp. Sci. St. Petersburg N.S. III* **35**, 65-70 (1892).

75. G. Stubbs, S. Warren, K. Holmes, Structure of RNA and RNA binding site in tobacco mosaic virus from 4-A map calculated from X-ray fibre diagrams. *Nature* **267**, 216-221 (1977).

76. C. Xiao *et al.*, Cryo-EM reconstruction of the Cafeteria roenbergensis virus capsid suggests novel assembly pathway for giant viruses. *Scientific reports* **7**, 5484 (2017).

77. Q. Fang *et al.*, Near-atomic structure of a giant virus. *Nature communications* **10**, 388 (2019).

78. X. Yan *et al.*, Structure and assembly of large lipid-containing dsDNA viruses. *Nature structural biology* **7**, 101-103 (2000).

79. B. La Scola *et al.*, A giant virus in amoebae. *Science* **299**, 2033 (2003).

80. P. Colson, X. de Lamballerie, G. Fournous, D. Raoult, Reclassification of giant viruses composing a fourth domain of life in the new order Megavirales. *Intervirology* **55**, 321-332 (2012).

81. V. Sharma, P. Colson, O. Chabrol, P. Pontarotti, D. Raoult, Pithovirus sibericum, a new bona fide member of the "Fourth TRUC" club. *Frontiers in microbiology* **6**, 722 (2015).

82. J. S. Abrahao *et al.*, Acanthamoeba polyphaga mimivirus and other giant viruses: an open field to outstanding discoveries. *Virology journal* **11**, 120 (2014).

83. S. Aherfi, P. Colson, B. La Scola, D. Raoult, Giant Viruses of Amoebas: An Update. *Frontiers in microbiology* **7**, 349 (2016).

84. P. Colson, S. Aherfi, B. La Scola, D. Raoult, The role of giant viruses of amoebas in humans. *Current opinion in microbiology* **31**, 199-208 (2016).

85. F. L. Assis *et al.*, Pan-Genome Analysis of Brazilian Lineage A Amoebal Mimiviruses. *Viruses* **7**, 3483-3499 (2015).

86. N. Philippe *et al.*, Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science* **341**, 281-286 (2013).

87. C. Abergel, M. Legendre, J. M. Claverie, The rapidly expanding universe of giant viruses: Mimivirus, Pandoravirus, Pithovirus and Mollivirus. *FEMS microbiology reviews* **39**, 779-796 (2015).

88. J. S. Abrahao, R. Araujo, P. Colson, B. La Scola, The analysis of translation-related gene set boosts debates around origin and evolution of mimiviruses. *PLoS genetics* **13**, e1006532 (2017).

89. F. Schulz *et al.*, Giant viruses with an expanded complement of translation system components. *Science* **356**, 82-85 (2017).

90. D. Raoult *et al.*, The 1.2-megabase genome sequence of Mimivirus. *Science* **306**, 1344-1350 (2004).

91. C. M. Mizuno *et al.*, Numerous cultivated and uncultivated viruses encode ribosomal proteins. *Nature communications* **10**, 752 (2019).

92. R. A. L. Rodrigues, T. S. Arantes, G. P. Oliveira, L. K. Dos Santos Silva, J. S. Abrahao, The Complex Nature of Tupanviruses. *Advances in virus research* **103**, 135-166 (2019).

93. J. M. Claverie, Viruses take center stage in cellular evolution. *Genome biology* **7**, 110 (2006).

94. T. A. Williams, T. M. Embley, E. Heinz, Informational gene phylogenies do not support a fourth domain of life for nucleocytoplasmic large DNA viruses. *PloS one* **6**, e21080 (2011).

95. R. K. Campos *et al.*, Samba virus: a novel mimivirus from a giant rain forest, the Brazilian Amazon. *Virology journal* **11**, 95 (2014).

96. J. R. Schrad, E. J. Young, J. S. Abrahao, J. R. Cortines, K. N. Parent, Microscopic Characterization of the Brazilian Giant Samba Virus. *Viruses* **9**, (2017).

97. E. Milrot *et al.*, Structural studies demonstrating a bacteriophage-like replication cycle of the eukaryote-infecting Paramecium bursaria chlorella virus-1. *PLoS pathogens* **13**, e1006562 (2017).

98. T. Klose *et al.*, Structure of faustovirus, a large dsDNA virus. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 6206-6211 (2016).

99. D. G. Reteno *et al.*, Faustovirus, an asfarvirus-related new lineage of giant viruses infecting amoebae. *Journal of virology* **89**, 6585-6594 (2015).

100. P. Colson, B. La Scola, D. Raoult, Giant Viruses of Amoebae: A Journey Through Innovative Research and Paradigm Changes. *Annual review of virology* **4**, 61-85 (2017).

101. R. A. L. Rodrigues, S. Mougari, P. Colson, B. La Scola, J. S. Abrahao, "Tupanvirus", a new genus in the family Mimiviridae. *Archives of virology* **164**, 325-331 (2019).

102. N. Brandes, M. Linial, Giant Viruses-Big Surprises. *Viruses* **11**, (2019).

103. K. K. W. To, C. C. Y. Yip, K. Y. Yuen, Rhinovirus - From bench to bedside. *Journal of the Formosan Medical Association = Taiwan yi zhi* **116**, 496-504 (2017).

104. X. Dai, Z. H. Zhou, Structure of the herpes simplex virus 1 capsid with associated tegument protein complexes. *Science* **360**, (2018).

105. C. Kerepesi, V. Grolmusz, The "Giant Virus Finder" discovers an abundance of giant viruses in the Antarctic dry valleys. *Archives of virology* **162**, 1671-1676 (2017).

106. S. Aherfi *et al.*, Experimental Inoculation in Rats and Mice by the Giant Marseillevirus Leads to Long-Term Detection of Virus. *Frontiers in microbiology* **9**, 463 (2018).

107. M. Khan, B. La Scola, H. Lepidi, D. Raoult, Pneumonia in mice inoculated experimentally with Acanthamoeba polyphaga mimivirus. *Microbial pathogenesis* **42**, 56-61 (2007).

108. E. Ghigo *et al.*, Ameobal pathogen mimivirus infects macrophages through phagocytosis. *PLoS pathogens* **4**, e1000087 (2008).

109. R. A. Rodrigues *et al.*, Mimivirus Fibrils Are Important for Viral Attachment to the Microbial World by a Diverse Glycoside Interaction Repertoire. *Journal of virology* **89**, 11812-11819 (2015).

110. M. Boughalmi *et al.*, First isolation of a giant virus from wild Hirudo medicinalis leech: Mimiviridae isolation in Hirudo medicinalis. *Viruses* **5**, 2920-2930 (2013).

111. K. R. Andrade *et al.*, Oysters as hot spots for mimivirus isolation. *Archives of virology* **160**, 477-482 (2015).

112. S. Aherfi *et al.*, Marseillevirus in lymphoma: a giant in the lymph node. *The Lancet. Infectious diseases* **16**, e225-234 (2016).

113. H. Saadi *et al.*, Shan virus: a new mimivirus isolated from the stool of a Tunisian patient with pneumonia. *Intervirology* **56**, 424-429 (2013).

114. H. Saadi *et al.*, First isolation of Mimivirus in a patient with pneumonia. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **57**, e127-134 (2013).

115. D. Raoult, P. Renesto, P. Brouqui, Laboratory infection of a technician by mimivirus. *Annals of internal medicine* **144**, 702-703 (2006).

116. N. Popgeorgiev, G. Michel, H. Lepidi, D. Raoult, C. Desnues, Marseillevirus adenitis in an 11-month-old child. *Journal of clinical microbiology* **51**, 4102-4105 (2013).

117. N. Shah *et al.*, Exposure to mimivirus collagen promotes arthritis. *Journal of virology* **88**, 838-845 (2014).

118. G. M. Almeida, L. C. Silva, P. Colson, J. S. Abrahao, Mimiviruses and the Human Interferon System: Viral Evasion of Classical Antiviral Activities, But Inhibition By a Novel Interferon-beta Regulated Immunomodulatory Pathway. *Journal of interferon & cytokine research : the official journal of the International Society for Interferon and Cytokine Research* **37**, 1-8 (2017).

119. S. Aherfi, P. Colson, D. Raoult, Marseillevirus in the Pharynx of a Patient with Neurologic Disorders. *Emerging infectious diseases* **22**, 2008-2010 (2016).

120. R. Siddiqui, N. A. Khan, Biology and pathogenesis of Acanthamoeba. *Parasites & vectors* **5**, 6 (2012).

121. C. Abergel, J. M. Claverie, [Pithovirus sibericum: awakening of a giant virus of more than 30,000 years]. *Medecine sciences : M/S* **30**, 329-331 (2014).

122. B. Bean *et al.*, Survival of influenza viruses on environmental surfaces. *The Journal of infectious diseases* **146**, 47-51 (1982).

123. J. A. Muller *et al.*, Inactivation and Environmental Stability of Zika Virus. *Emerging infectious diseases* **22**, 1685-1687 (2016).

124. S. Bousbia *et al.*, Serologic prevalence of amoeba-associated microorganisms in intensive care unit pneumonia patients. *PloS one* **8**, e58111 (2013).

125. D. J. Goetschius, C. R. Parrish, S. Hafenstein, Asymmetry in icosahedral viruses. *Current opinion in virology* **36**, 67-73 (2019).

126. K. N. Parent, J. R. Schrad, G. Cingolani, Breaking Symmetry in Viral Icosahedral Capsids as Seen through the Lenses of X-ray Crystallography and Cryo-Electron Microscopy. *Viruses* **10**, (2018).

127. B. Hu, W. Margolin, I. J. Molineux, J. Liu, Structural remodeling of bacteriophage T4 and host membranes during infection initiation. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E4919-4928 (2015).

128. K. N. Parent *et al.*, OmpA and OmpC are critical host factors for bacteriophage Sf6 entry in Shigella. *Molecular microbiology* **92**, 47-60 (2014).

129. K. N. Parent, E. B. Gilcrease, S. R. Casjens, T. S. Baker, Structural evolution of the P22-like phages: comparison of Sf6 and P22 procapsid and virion architectures. *Virology* **427**, 177-188 (2012).

130. B. Apellaniz, N. Huarte, E. Largo, J. L. Nieva, The three lives of viral fusion peptides. *Chemistry and physics of lipids* **181**, 40-55 (2014).

131. U. Ghosh, L. Xie, D. P. Weliky, Detection of closed influenza virus hemagglutinin fusion peptide structures in membranes by backbone (13)CO- (15)N rotational-echo double-resonance solid-state NMR. *Journal of biomolecular NMR* **55**, 139-146 (2013).

132. S. C. Harrison, Viral membrane fusion. *Virology* **479-480**, 498-507 (2015).

133. A. Borodavka, U. Desselberger, J. T. Patton, Genome packaging in multi-segmented dsRNA viruses: distinct mechanisms with similar outcomes. *Current opinion in virology* **33**, 106-112 (2018).

134. C. P. Long, S. M. McDonald, Rotavirus genome replication: Some assembly required. *PLoS pathogens* **13**, e1006242 (2017).

135. J. J. Carvajal *et al.*, Host Components Contributing to Respiratory Syncytial Virus Pathogenesis. *Frontiers in immunology* **10**, 2152 (2019).

136. T. J. Ruckwardt, K. M. Morabito, B. S. Graham, Immunological Lessons from Respiratory Syncytial Virus Vaccine Development. *Immunity* **51**, 429-442 (2019).

137. N. B. Hubbs, *Hijacking the cell : how bacteriophage Sf6 uses Shigella flexneri outer membrane proteins for infection*. (2017), pp. 1 online resource (xiv, 136 pages).

138. N. B. Porcek, K. N. Parent, Key residues of S. flexneri OmpA mediate infection by bacteriophage Sf6. *Journal of molecular biology* **427**, 1964-1976 (2015).

139. C. Wang, J. Tu, J. Liu, I. J. Molineux, Structural dynamics of bacteriophage P22 infection initiation revealed by cryo-electron tomography. *Nature microbiology* **4**, 1049-1056 (2019).

140. L. Cai, M. Gochin, K. Liu, Biochemistry and biophysics of HIV-1 gp41 - membrane interactions and implications for HIV-1 envelope protein mediated viral-cell fusion and fusion inhibitor design. *Current topics in medicinal chemistry* **11**, 2959-2984 (2011).

141. J. Ning *et al.*, In vitro protease cleavage and computer simulations reveal the HIV-1 capsid maturation pathway. *Nature communications* **7**, 13689 (2016).

142. W. I. Sundquist, H. G. Krausslich, HIV-1 assembly, budding, and maturation. *Cold Spring Harbor perspectives in medicine* **2**, a006924 (2012).

143. J. J. Skehel, D. C. Wiley, Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annual review of biochemistry* **69**, 531-569 (2000).

144. P. Tavares, The Bacteriophage Head-to-Tail Interface. *Sub-cellular biochemistry* **88**, 305-328 (2018).

145. M. McElwee, S. Vijayakrishnan, F. Rixon, D. Bhella, Structure of the herpes simplex virus portal-vertex. *PLoS biology* **16**, e2006191 (2018).

146. F. Piacente *et al.*, The rare sugar N-acetylated viosamine is a major component of Mimivirus fibers. *The Journal of biological chemistry* **292**, 7385-7394 (2017).

147. F. Piacente *et al.*, Giant virus Megavirus chilensis encodes the biosynthetic pathway for uncommon acetamido sugars. *The Journal of biological chemistry* **289**, 24428-24439 (2014).

148. A. J. Rommel, A. J. Hulsmeier, S. Jurt, T. Hennet, Giant mimivirus R707 encodes a glycogenin paralogue polymerizing glucose through alpha- and beta-glycosidic linkages. *The Biochemical journal* **473**, 3451-3462 (2016).

149. J. Andreani *et al.*, Cedratvirus, a Double-Cork Structured Giant Virus, is a Distant Relative of Pithoviruses. *Viruses* **8**, (2016).

150. T. Klose *et al.*, The three-dimensional structure of Mimivirus. *Intervirology* **53**, 268-273 (2010).

151. C. Xiao *et al.*, Structural studies of the giant mimivirus. *PLoS biology* **7**, e92 (2009).

152. C. Xiao *et al.*, Cryo-electron microscopy of the giant Mimivirus. *Journal of molecular biology* **353**, 493-496 (2005).

153. E. Milrot *et al.*, Virus-host interactions: insights from the replication cycle of the large Paramecium bursaria chlorella virus. *Cellular microbiology* **18**, 3-16 (2016).

154. L. K. Dixon, D. A. Chapman, C. L. Netherton, C. Upton, African swine fever virus replication and genomics. *Virus research* **173**, 3-14 (2013).

155. G. Beaud, Vaccinia virus DNA replication: a short review. *Biochimie* **77**, 774-779 (1995).

156. F. P. Dornas *et al.*, Detection of mimivirus genome and neutralizing antibodies in humans from Brazil. *Archives of virology*, (2017).

157. B. La Scola, T. J. Marrie, J. P. Auffray, D. Raoult, Mimivirus in pneumonia patients. *Emerging infectious diseases* **11**, 449-452 (2005).

158. M. G. Fischer, M. J. Allen, W. H. Wilson, C. A. Suttle, Giant virus with a remarkable complement of genes infects marine zooplankton. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 19508-19513 (2010).

159. R. I. Koning, A. J. Koster, Cryo-electron tomography in biology and medicine. *Annals of anatomy = Anatomischer Anzeiger : official organ of the Anatomische Gesellschaft* **191**, 427-445 (2009).

160. P. S. Shen, The 2017 Nobel Prize in Chemistry: cryo-EM comes of age. *Analytical and bioanalytical chemistry* **410**, 2053-2057 (2018).

161. E. Nogales, Cryo-EM. *Curr Biol* **28**, R1127-R1128 (2018).

162. A. R. Faruqi, R. Henderson, Electronic detectors for electron microscopy. *Current opinion in structural biology* **17**, 549-555 (2007).

163. Z. A. Ripstein, J. L. Rubinstein, Processing of Cryo-EM Movie Data. *Methods in enzymology* **579**, 103-124 (2016).

164. S. Q. Zheng *et al.*, MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature methods* **14**, 331-332 (2017).

165. A. Bartesaghi *et al.*, Classification and 3D averaging with missing wedge correction in biological electron tomography. *Journal of structural biology* **162**, 436-450 (2008).

166. D. N. Mastronarde, Dual-axis tomography: an approach with alignment methods that preserve resolution. *Journal of structural biology* **120**, 343-352 (1997).

167. D. N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements. *Journal of structural biology* **152**, 36-51 (2005).

168. W. Wu *et al.*, Localization of the Houdinisome (Ejection Proteins) inside the Bacteriophage P22 Virion by Bubblegram Imaging. *mBio* **7**, (2016).

169. W. Wu *et al.*, Internal Proteins of the Procapsid and Mature Capsids of Herpes Simplex Virus 1 Mapped by Bubblegram Imaging. *Journal of virology* **90**, 5176-5186 (2016).

170. W. Wu, J. A. Thomas, N. Cheng, L. W. Black, A. C. Steven, Bubblegrams reveal the inner body of bacteriophage phiKZ. *Science* **335**, 182 (2012).

171. A. Lwoff, The concept of virus. *Journal of general microbiology* **17**, 239-253 (1957).

172. V. Sharma, P. Colson, P. Pontarotti, D. Raoult, Mimivirus inaugurated in the 21st century the beginning of a reclassification of viruses. *Current opinion in microbiology* **31**, 16-24 (2016).

173. F. P. Dornas *et al.*, A Brazilian Marseillevirus Is the Founding Member of a Lineage in Family Marseilleviridae. *Viruses* **8**, 76 (2016).

174. P. Scheid, C. Balczun, G. A. Schaub, Some secrets are revealed: parasitic keratitis amoebae as vectors of the scarcely described pandoraviruses to humans. *Parasitology research* **113**, 3759-3764 (2014).

175. T. Ekeberg *et al.*, Three-dimensional reconstruction of the giant mimivirus particle with an x-ray free-electron laser. *Physical review letters* **114**, 098102 (2015).

176. N. Zauberman *et al.*, Distinct DNA exit and packaging portals in the virus Acanthamoeba polyphaga mimivirus. *PLoS biology* **6**, e114 (2008).

177. Y. Mutsafi, N. Zauberman, I. Sabanay, A. Minsky, Vaccinia-like cytoplasmic replication of the giant Mimivirus. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 5978-5982 (2010).

178. M. Suzan-Monti, B. La Scola, L. Barrassi, L. Espinosa, D. Raoult, Ultrastructural characterization of the giant volcano-like virus factory of Acanthamoeba polyphaga Mimivirus. *PloS one* **2**, e328 (2007).

179. M. Adrian, J. Dubochet, J. Lepault, A. W. McDowall, Cryo-electron microscopy of viruses. *Nature* **308**, 32-36 (1984).

180. L. J. Reed, Muench, H., A simple method of estimating fifty per cent endpoints. *American Journal of Epidemiology* **27**, 493-497 (1938).

181. Z. Wang *et al.*, An atomic model of brome mosaic virus using direct electron detection and real-space optimization. *Nature communications* **5**, 4808 (2014).

182. J. R. Kremer, D. N. Mastronarde, J. R. McIntosh, Computer visualization of three-dimensional image data using IMOD. *Journal of structural biology* **116**, 71-76 (1996).

183. G. Tang *et al.*, EMAN2: an extensible image processing suite for electron microscopy. *Journal of structural biology* **157**, 38-46 (2007).

184. X. Yan, K. A. Dryden, J. Tang, T. S. Baker, Ab initio random model method facilitates 3D reconstruction of icosahedral particles. *Journal of structural biology* **157**, 211-225 (2007).

185. X. Yan, R. S. Sinkovits, T. S. Baker, AUTO3DEM--an automated and high throughput program for image reconstruction of icosahedral particles. *Journal of structural biology* **157**, 73-82 (2007).

186. S. J. Ludtke, Single-Particle Refinement and Variability Analysis in EMAN2.1. *Methods in enzymology* **579**, 159-189 (2016).

187. Y. G. Kuznetsov *et al.*, Atomic force microscopy investigation of the giant mimivirus. *Virology* **404**, 127-137 (2010).

188. M. Schaffer *et al.*, Cryo-focused Ion Beam Sample Preparation for Imaging Vitreous Cells by Cryo-electron Tomography. *Bio-protocol* **5**, (2015).

189. M. Schaffer *et al.*, Optimized cryo-focused ion beam sample preparation aimed at in situ structural studies of membrane proteins. *Journal of structural biology*, (2016).

190. J. L. Milne, S. Subramaniam, Cryo-electron tomography of bacteria: progress, challenges and future prospects. *Nature reviews. Microbiology* **7**, 666-675 (2009).

191. J. L. Milne *et al.*, Cryo-electron microscopy--a primer for the non-microscopist. *The FEBS journal* **280**, 28-45 (2013).

192. Y. Mutsafi, E. Shimoni, A. Shimon, A. Minsky, Membrane assembly during the infection cycle of the giant Mimivirus. *PLoS pathogens* **9**, e1003367 (2013).

193. M. Barcena *et al.*, Cryo-electron tomography of mouse hepatitis virus: Insights into the structure of the coronavirion. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 582-587 (2009).

194. K. N. Parent *et al.*, Cryo-reconstructions of P22 polyheads suggest that phage assembly is nucleated by trimeric interactions among coat proteins. *Physical biology* **7**, 045004 (2010).

195. J. Y. Khalil *et al.*, High-Throughput Isolation of Giant Viruses in Liquid Medium Using Automated Flow Cytometry and Fluorescence Staining. *Frontiers in microbiology* **7**, 26 (2016).

196. E. Lenk, S. Casjens, J. Weeks, J. King, Intracellular visualization of precursor capsids in phage P22 mutant infected cells. *Virology* **68**, 182-199 (1975).

197. S. Azza, C. Cambillau, D. Raoult, M. Suzan-Monti, Revised Mimivirus major capsid protein sequence reveals intron-containing gene structure and extra domain. *BMC molecular biology* **10**, 39 (2009).

198. P. V. M. Boratto *et al.*, Analyses of the Kroon Virus Major Capsid Gene and Its Transcript Highlight a Distinct Pattern of Gene Evolution and Splicing among Mimiviruses. *Journal of virology* **92**, (2018).

199. F. P. Dornas *et al.*, Mimivirus circulation among wild and domestic mammals, Amazon Region, Brazil. *Emerging infectious diseases* **20**, 469-472 (2014).

200. E. A. Lusi, D. Maloney, F. Caicci, P. Guarascio, Questions on unusual Mimivirus-like structures observed in human cells. *F1000Research* **6**, 262 (2017).

201. R. K. Campos *et al.*, Virucidal activity of chemical biocides against mimivirus, a putative pneumonia agent. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology* **55**, 323-328 (2012).

202. F. P. Dornas *et al.*, Acanthamoeba polyphaga mimivirus stability in environmental and clinical substrates: implications for virus detection and isolation. *PloS one* **9**, e87811 (2014).

203. M. Suomalainen *et al.*, A direct and versatile assay measuring membrane penetration of adenovirus in single cells. *Journal of virology* **87**, 12367-12379 (2013).

204. J. Andreani *et al.*, Orpheovirus IHUMI-LCC2: A New Virus among the Giant Viruses. *Frontiers in microbiology* **8**, 2643 (2017).

205. J. Andreani *et al.*, Pacmanvirus, a New Giant Icosahedral Virus at the Crossroads between Asfarviridae and Faustoviruses. *Journal of virology* **91**, (2017).

206. L. C. F. Silva *et al.*, Microscopic Analysis of the Tupanvirus Cycle in Vermamoeba vermiformis. *Frontiers in microbiology* **10**, 671 (2019).

207. D. J. Cummings, N. L. Couse, G. L. Forrest, Structural defects of T-even bacteriophages. *Advances in virus research* **16**, 1-41 (1970).

208. K. N. Parent *et al.*, P22 coat protein structures reveal a novel mechanism for capsid maturation: stability without auxiliary proteins or chemical crosslinks. *Structure* **18**, 390-401 (2010).

209. C. M. Teschke, A. McGough, P. A. Thuman-Commike, Penton release from P22 heat-expanded capsids suggests importance of stabilizing penton-hexon interactions during capsid maturation. *Biophysical journal* **84**, 2585-2592 (2003).

210. C. M. Teschke, K. N. Parent, 'Let the phage do the work': using the phage P22 coat protein structures as a framework to understand its folding and assembly mutants. *Virology* **401**, 119-130 (2010).

211. J. C. Chan, H. H. Gadebusch, Virucidal properties of dimethyl sulfoxide. *Appl Microbiol* **16**, 1625-1626 (1968).

212. I. M. Cotton, L. S. Lockingen, Inactivation of Bacteriophage by Chloroform and X Irradiation. *Proceedings of the National Academy of Sciences of the United States of America* **50**, 363-367 (1963).

213. D. J. Cummings, V. A. Chapman, S. S. DeLong, Disruption of T-even bacteriophages by dimethyl sulfoxide. *Journal of virology* **2**, 610-620 (1968).

214. J. Griffith, M. Manning, K. Dunn, Filamentous bacteriophage contract into hollow spherical particles upon exposure to a chloroform-water interface. *Cell* **23**, 747-753 (1981).

215. R. S. Flannagan, B. Heit, D. E. Heinrichs, Antimicrobial Mechanisms of Macrophages and the Immune Evasion Strategies of Staphylococcus aureus. *Pathogens* **4**, 826-868 (2015).

216. N. German, D. Doyscher, C. Rensing, Bacterial killing in macrophages and amoeba: do they all use a brass dagger? *Future microbiology* **8**, 1257-1264 (2013).

217. C. A. Lopez, E. P. Skaar, The Impact of Dietary Transition Metals on Host-Bacterial Interactions. *Cell host & microbe* **23**, 737-748 (2018).

218. E. O. Freed, HIV-1 assembly, release and maturation. *Nature reviews. Microbiology* **13**, 484-496 (2015).

219. W. F. Mangel, C. San Martin, Structure, function and dynamics in adenovirus maturation. *Viruses* **6**, 4536-4570 (2014).

220. J. D. Nabarro, Acromegaly. *Clinical endocrinology* **26**, 481-512 (1987).

221. L. Kordyukova, Structural and functional specificity of Influenza virus haemagglutinin and paramyxovirus fusion protein anchoring peptides. *Virus research* **227**, 183-199 (2017).

222. W. W. Newcomb *et al.*, The UL6 gene product forms the portal for entry of DNA into the herpes simplex virus capsid. *Journal of virology* **75**, 10923-10932 (2001).

223. S. M. Huang, S. T. Kuo, H. C. Kuo, S. K. Chang, Assessment of fish iridoviruses using a novel cell line GS-1, derived from the spleen of orange-spotted grouper Epinephelus coioides (Hamilton) and susceptible to ranavirus and megalocytivirus. *The Journal of veterinary medical science* **80**, 1766-1774 (2018).

224. M. Remmert, A. Biegert, A. Hauser, J. Soding, HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature methods* **9**, 173-175 (2011).

225. A. L. Mitchell *et al.*, InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic acids research* **47**, D351-D360 (2019).

226. D. W. A. Buchan, D. T. Jones, The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic acids research* **47**, W402-W407 (2019).

227. D. T. Jones, D. Cozzetto, DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* **31**, 857-863 (2015).

228. C. Camacho *et al.*, BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

229. M. Bastian, Heymann, S., Jacomy, M. , Gephi: An Open Source Software for Exploring and Manipulating Networks. *International AAI Conference on Weblogs and Social Media*, (2009).

230. K. A. Manning, N. Quiles-Puchalt, J. R. Penades, T. Dokland, A novel ejection protein from bacteriophage 80alpha that promotes lytic growth. *Virology* **525**, 237-247 (2018).

231. W. Greene *et al.*, The ubiquitin/proteasome system mediates entry and endosomal trafficking of Kaposi's sarcoma-associated herpesvirus in endothelial cells. *PLoS pathogens* **8**, e1002703 (2012).

232. Y. Mutsafi, Y. Fridmann-Sirkis, E. Milrot, L. Hevroni, A. Minsky, Infection cycles of large DNA viruses: emerging themes and underlying questions. *Virology* **466-467**, 3-14 (2014).

233. G. Oliveira *et al.*, Tupanvirus-infected amoebas are induced to aggregate with uninfected cells promoting viral dissemination. *Scientific reports* **9**, 183 (2019).

234. M. A. Ramakrishnan, Determination of 50% endpoint titer using a simple formula. *World journal of virology* **5**, 85-86 (2016).

235. D. N. Mastronarde, S. R. Held, Automated tilt series alignment and tomographic reconstruction in IMOD. *Journal of structural biology* **197**, 102-113 (2017).

236. J. Rappsilber, M. Mann, Y. Ishihama, Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature protocols* **2**, 1896-1906 (2007).

237. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* **26**, 1367-1372 (2008).

238. J. Cox *et al.*, Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of proteome research* **10**, 1794-1805 (2011).