

REGULATION OF CAPSID SIZES OF LARGE TAILED BACTERIOPHAGES

by

Jianfei Hua

B.S., University of Science and Technology of China, 2004

Submitted to the Graduate Faculty of
Art and Science in partial fulfillment
of the requirements for the degree of
Master of Science

University of Pittsburgh

2010

UNIVERSITY OF PITTSBURGH

School of Art and Science

This dissertation was presented

by

Jianfei Hua

It was defended on

November 20, 2009

and approved by

Michael Cascio, PhD, Department of Molecular Genetics and Biochemistry

Peijun Zhang, PhD, Department of Structural Biology

Dissertation Advisors: James Conway, PhD, Department of Structural Biology, &

Roger Hendrix, PhD, Department of Biological Sciences

Copyright © by Jianfei Hua

2010

REGULATION OF CAPSID SIZES OF LARGE TAILED BACTERIOPHAGES

Jianfei Hua, M.S.

University of Pittsburgh, 2010

Many bacteriophages and eukaryotic viruses, which share little sequence similarities, have icosahedral protein capsids containing their genetic materials. Generally, an icosahedral viral capsid is assembly of 12 pentamers and a certain number of hexmers of the major capsid protein, following Caspar and Klug's quasi-equivalence rule. The arrangement of these pentamers and hexmers is characterized by the triangulation (T) number. Questions arise whether viruses have evolved from a few common ancestors, and how the assembly of the icosahedral capsids has been regulated to achieve a defined capsid size and geometry. I present studies of the capsids of several large icosahedral bacteriophages, which broaden our understanding of the regulation of viral capsid assembly.

Bacteriophage SPO1 may share common ancestry with herpesvirus, according to the similarities in their T numbers and in the asymmetric molecules slightly off the local three-fold symmetry positions on the outer surface of both capsids. However, the cryo-EM structure of the SPO1 capsid assembled from the uncleaved major capsid protein show that, unlike the herpesvirus asymmetric molecule, the SPO1 asymmetric protein may not be required for the initial procapsid assembly, suggesting that the two asymmetric molecules may have different origins. Phage P1 is excellent for studying size determination in viral capsid since it produces virions of three sizes. The cryo-EM structures of the three capsids and internal capsid proteins identified suggests a

control mechanism for P1 capsids, in which the DarA protein functions as a semi-scaffolding protein to assist the assembly of the P1 big capsid. Jumbo phages have been rarely studied. The structural studies on four jumbo phages showed their T numbers. N3, PAU and 121Q are the first T = 19, 25 and 28 viral capsids found. These results suggest that T-numbers larger than 16 may generally be allowed.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	0xkk
CHAPTER 1 BACKGROUND.....	1
1.1 INTRODUCTION	1
1.2 PHAGE BIOLOGY	2
1.2.1 The nature of bacteriophages	2
1.2.2 The infection process of tailed phages	4
1.3 THE CAPSID OF TAILED PHAGES.....	5
1.3.1 The icosahedral structure and the triangulation number	5
1.3.2 Capsid proteins	6
1.3.3 Capsid assembly.....	8
1.3.4 The evolution of virus capsids	8
1.4 CRYO-EM.....	10
CHAPTER 2 SPO1	14
2.1 INTRODUCTION	14
2.2 MATERIALS AND METHODS	17
2.2.1 Strains and vectors	17
2.2.2 Media and Buffers	18
2.2.3 Methods	19

2.2.3.1	Phage manipulations.....	19
2.2.3.2	SPO1 phage and capsid purification	20
2.2.3.3	DNA and protein techniques.....	21
2.2.3.4	Isolation of SPO1 amber mutant 3.2am1.....	24
2.2.3.5	Cryo electron microscopy and capsid reconstruction	24
2.3	RESULTS	25
2.3.1	Empty capsids from wild type SPO1 lysate	25
2.3.2	Cloning and expression of SPO1 capsid genes.....	27
2.3.3	Capsids isolated from 3.2am1 lysate	28
2.4	SUMMARY AND DISCUSSION	34
CHAPTER 3 PHAGE P1		37
3.1	INTRODUCTION	37
3.2	MATERIALS AND METHODS	39
3.2.1	Strains.....	39
3.2.2	Media and buffers.....	40
3.2.3	P1 phage manipulation.....	40
3.3	RESULTS AND DISCUSSION	42
3.3.1	P1 morphology	42
3.3.2	P1 proteins.....	45
3.4	SUMMARY AND DISCUSSION	48
CHAPTER 4 JUMBO PHAGES		50
4.1	INTRODUCTION	50
4.2	MATERIAL AND METHODS	51

4.3	RESULTS	52
4.4	SUMMARY AND DISCUSSION.....	55
	FINAL SUMMARY.....	57
	BIBLIOGRAPHY	59

LIST OF TABLES

Table 3.1 A summary of the amount of data used and final resolutions in P1 page reconstruction.....	45
Table 4.1 A summary of the amount of data used and final resolutions in jumbo page reconstruction.....	53

LIST OF FIGURES

Figure 1. 1 Schematics of three types of tailed phages.....	3
Figure 1. 2 Construction of icosahedrons by a hexagonal lattice with a defined T-number.	7
Figure 1. 3 A generalized phage assembly	9
Figure 1. 4 cryo-EM structures of a few virus capsids	12
Figure 1. 5 The basic procedure of 3D reconstruction.....	13
Figure2. 1 A negative-stain EM picture of bacteriophage SPO1.....	15
Figure2. 2 Comparison of SPO1 and Herpes Simplex Virus Type-1 (HSV-1) capsid structures	16
Figure2. 3 SDS polyacrylamide gel of samples purified from wild type SPO1 lysate.....	26
Figure2. 4 Cryo-EM structures of the empty particle (17.5 Å) and SPO1 phage (20 Å)	28
Figure2. 5 Expression of SPO1 capsid genes in BL21(DE3)	29
Figure2. 6 Sucrose gradient pattern of 3.2am1 lysate and polyacrylamide gel of sucrose gradient bands.	31
Figure2. 7 Proteins from peaks isolated by ion-exchange chromatography of the B2 and B3 mixture	32
Figure2. 8 Cryo-EM structures of capsids in F1 (17.5Å) and SPO1 phage (20 Å).....	33
Figure 3. 1 A negative stain EM image of P1 phages with capsids of different sizes.....	39
Figure 3. 2 Three P1 types isolated from CsCl gradient.....	44
Figure 3. 3 Cryo-EM structures of P1b, P1s and P1m.....	44

Figure 3. 4 Identification of P1 proteins by Edman degradation and mass spectry.....	46
Figure 4. 1 Jumbo phage structures	54

ACKNOWLEDGEMENTS

I am heartily thankful to my advisors, Dr. James Conway and Dr. Roger Hendrix, who have supported me throughout my entire study with their encouragement, patience and guidance. Their curiosity to science, enthusiasm to work and kindness to people will influence me beyond my life in Pittsburgh. They have been continuously making efforts for the best of mine. I simply could not wish for better advisors.

I am deeply grateful to Dr. Robert Duda and Dr. James Conway for teaching me most of the experimental techniques I used. The discussions with Dr. Robert Duda were very pleasant, and extremely helpful to my wet experiments. I feel very lucky to have such a mentor and friend.

I would like to thank my committee members, Dr. Michael Cascio and Dr. Peijun Zhang. I appreciate their time and efforts put in my thesis defence and my comprehensive exam.

Roger's group is always warm and joyful. I want to thank all the current members and the previous members I met, for their help and friendship. Brian Firek has been very helpful in technical supports and on my thesis correction.

Finally, I am indebted to my parents and my girlfriend, Joanne Wu, for their love and understanding all the time.

CHAPTER 1 BACKGROUND

1.1 INTRODUCTION

Bacteriophages, or phages, are “bacteria eaters”, viruses that infect Eubacteria and Archaea. Phages were discovered independently by Frederick W. Twort (1915) [1] and Felix d’Herelle (1917) [2]. Phage research in the 1960s contributed major discoveries in biological sciences, such as proving DNA as the genetic material, deciphering the genetic code, and the discoveries of transduction and messenger RNA [3]. In the 1970s, explorations of phage enzymes and plasmids provided the understanding and the tools that made it possible to develop genetic engineering [3]. However, it is not until relatively recently that the abundance and ubiquity of phages was explored. Phages are one of the most abundant living entities in the biosphere. The estimated range of total amount of phages is 10^{30} to 10^{32} [3]. There are up to 6×10^6 phages in one milliliter of surface sea water [4]. Phages are also one of the most widely distributed, observed in every natural ecosystem where bacteria survive [3]. Nowadays, phage research has entered a new golden age. Researchers have been paying increasing attention to phages for their abundant distribution and significant functions in nature, and their potential applications in biomedical science and industry [3]. With a number of cutting-edge techniques available, phages are being studied from chemistry to structures and from genetics to ecology.

1.2 PHAGE BIOLOGY

1.2.1 The nature of bacteriophages

Phages are parasites of bacteria. They lack their own translation and energy-generating systems and typically have less than a full complement of the functions needed for transcription and DNA replication, and they therefore must rely on the host systems to carry out their life cycles. Generally, a phage particle (virion) is a large molecular complex with its genome (DNA or RNA) enclosed by a protein capsid and in some cases a lipoprotein coat. The genome of one of the smallest phage, MS2, is 3,569 nucleotides long ssRNA contained in a capsid with a diameter of 300Å and encoding 4 proteins: a protease, a capsid protein, an RNA replicase, and a lysis protein [5]. The genome of the largest phage found, phage G, has ~500,000 base pairs of dsDNA enclosed by a capsid of over 1,500 Å in diameter [6]. Though the genome size of G is close to that of the smallest prokaryotes, G still lacks genes encoding essential molecular machineries such as ribosomes.

The morphology of phages is extremely heterogeneous. The virions can be tailed, polyhedral, filamentous or pleomorphic. So far, more than 5000 bacteriophages have been observed by electron microscopy, and 95% of them are tailed phages [7]. A tailed phage typically consists of an icosahedral protein capsid containing its linear dsDNA genome and a tail attached to the capsid. There are three families of tailed phages characterized by tail morphology: *Siphoviridae* (long, flexible tails), *Myoviridae* (long, double-layered, contractile tails) and *Podoviridae* (short tails) (Figure 1.1) [8]. Tailed phages are the most abundant and wide-spread phage group, and infect a wide range of hosts, including gram-negative bacteria, gram-positive bacteria,

cyanobacteria and Archaea. Thus, a number of tailed phages have been intensively studied as model organisms, including the seven “T” or Type phages chosen by Delbruck (1944) such as T4 and T7.

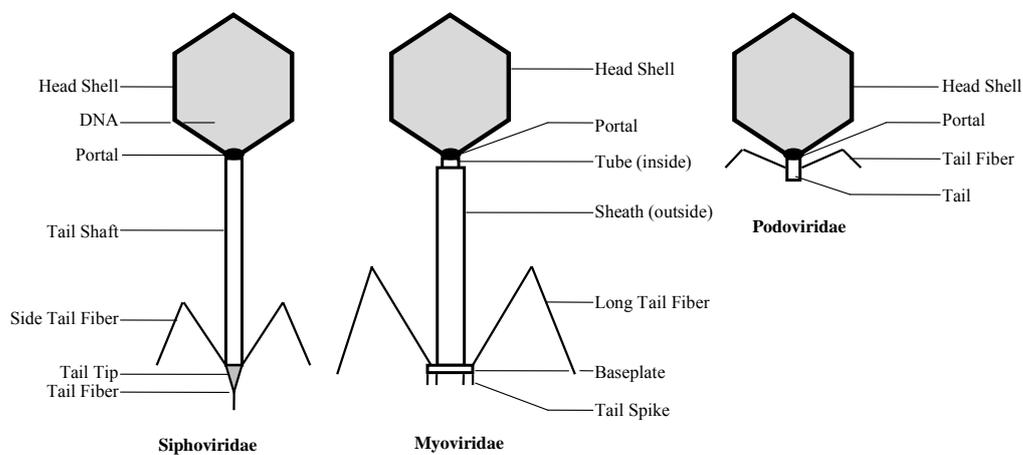


Figure 1. 1 Schematics of three types of tailed phages: Siphoviridae, Myoviridae and Podoviridae. This figure is redrawn according to [8].

There are two types of phage lifestyles: the lytic cycle and the lysogenic cycle, and thus two groups of phages: virulent and temperate. Virulent phages can only live a lytic lifestyle. A virulent phage infects its host by absorption onto the host cell surface and injection of its genome into the host cell. The host metabolism is then controlled by the phage to make a large amount of phage components to assemble into new phages. Minutes or hours later, the host cell lyses to release the new phages. When temperate phages infect host cells, they may choose a lytic cycle

as mentioned above, or a lysogenic cycle. In the latter situation, the phage genome enters an inactive DNA-only state known as a prophage, either by integrating into the host genome or by forming a plasmid. A prophage replicates and segregates into daughter cells in parallel with the host chromosome with the result that all the daughter cells are also lysogenic. During the lysogenic cycle, the transcription of host-lethal genes in the prophage is under tight control, while some host-beneficial genes may be expressed to protect the host from environmental pressures, antibiotics or other phages, or provide toxins to the bacterial host that may promote bacterial and phage growth [3], such as the Shiga toxins from λ prophages [9].

1.2.2 The infection process of tailed phages

Infection begins when phages adsorb to the surface of host bacteria. Tail fibers or spikes are specialized structures for adsorption. However, the receptors on host cell surfaces are highly diversified [3]. For example, the receptors in gram-negative bacteria can be proteins, oligosaccharides or lipopolysaccharides [3]. Many phages need a high concentration of receptors present on the host surface for adsorption. T4, for instance, needs to bind to at least 3 receptors with its long tail fibers to initiate the irreversible adsorption process, during which the base plate rearranges to deploy short tail fibers that irreversibly bind to another receptor. Irreversible adsorption to the host leads to transfer of the phage genome into the host cell. The tail tip has enzymatic activity to penetrate both the peptidoglycan layer and the phospholipid bilayer. The positioning of the tail on the host surface also dismantles the structure blocking the exit of phage DNA. As a result, the phage DNA is transferred from the capsid through the channel within the tail, with a rate up to 4000 base-pairs per second [10]. The mechanisms of DNA delivery are

highly diversified among the different phages; most of them rely on host energy sources such as ATP hydrolysis or transmembrane potential [8].

When the genome of a typical virulent phage enters a host cell, a number of strong phage promoters are quickly recognized by the host RNA polymerase, resulting in the transcription of several immediate early genes. The products of these early genes, many of which are lethal to the host cell, protect the phage genome and alter host components and pathways in favor of phage production [3]. The middle genes are then transcribed to replicate the phage genome followed by the late genes to produce phage structural proteins. Nascent phages are assembled in the host cell and released after cell lysis. In the best-studied cases, lysis is triggered by activities of two phage enzymes: holin that generates pores in the phospholipid bilayer and lysin that digests the peptidoglycan layer [3].

1.3 THE CAPSID OF TAILED PHAGES

1.3.1 The icosahedral structure and the triangulation number

In 1956, Crick and Watson proposed that the capsid of viruses was built up with multiple copies of the same molecule because the genome of a virus might not be sufficient to encode enough different proteins to construct the capsid [11]. They also proposed that the chemical environment of each subunit should be identical, so that the structure of the capsid could have certain symmetries, which they proposed would be one of the cubic symmetries. Later, icosahedral symmetry, which has 5-fold, 3-fold and 2-fold symmetry axes, and helical symmetry, was

proven to be favored by virus capsids. An icosahedron is an isometric structure that comprises 12 pentagonal vertices and 20 equilateral triangular faces, and contains 60 asymmetric units, which are individual protein subunits in the simplest capsids as envisioned by Crick and Watson. In 1962, Caspar and Klug described how the viral icosahedral capsid could be composed of a number of chemically identical subunits larger than 60 (Fig 1.2) [12]. In a flat hexagonal lattice, if some hexagons are properly chosen and replaced by pentagons, the lattice will be able to curve to construct an icosahedral shell [13]. In a virus capsid with 60 subunits, all subunits have identical environment and bonding. However, in a larger capsid with more than 60 chemically identical subunits, each of the 60 asymmetric units is formed by more than one subunit, which are quasi-equivalent because they are chemically identical but their environments are slightly different. The number of subunits in an asymmetric unit, the triangulation (T)-number, is used to characterize the geometric figure of a virus capsid. The T-number can be more strictly defined as follows:

$$T = h^2 + hk + k^2$$

where h and k are non-negative integers, and are the coordinates of a pentagon adjacent to the origin in the hexagonal lattice. An icosahedral capsid following the quasi-equivalence rule has $60 \cdot T$ copies of subunits in T kinds of different environments.

1.3.2 Capsid proteins

The major capsid protein (MCP) is the main block that builds the icosahedral capsid shell. The MCP is the only absolutely required virus protein for capsid assembly of the T= 7 phage HK97 [14]. Each HK97 capsid has 12 pentons and 60 hexons formed by 420 copies of the MCP [15]. The size of an icosahedral capsid is largely determined by the number of MCP copies, since the

molecular weight of tailed-phage MCPs is roughly 50 kDa. All tailed phage virions have a single penton replaced by a dodecameric hollow protein complex with 12-fold rotational symmetry called portal or connector [16], because phage DNA passes through it during DNA transfer and tails attach to capsids by connecting to it [17]. The portal also initiates capsid assembly in phage T4 and probably all others [18].

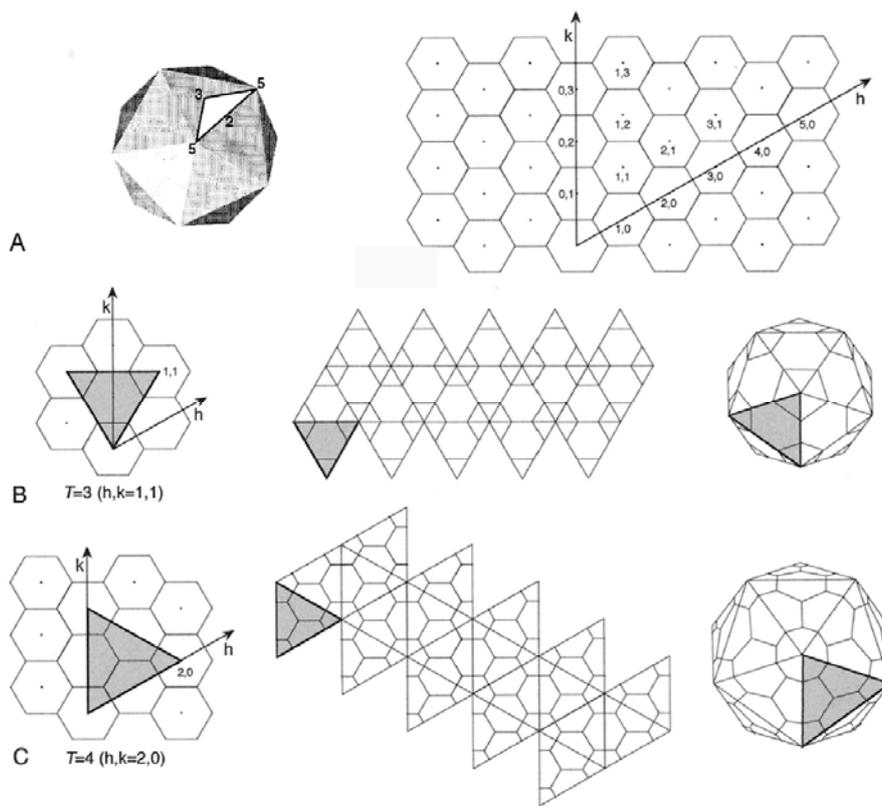


Figure 1. 2 Construction of icosahedrons by a hexagonal lattice with a defined T-number. (A). Left: an icosahedron with symmetry axes shown by numbers and an asymmetric subunit outlined; Right: a flat array of hexagons. By converting some hexagons to pentagons, curvature is introduced at the pentagons. An icosahedral shell with a specified T-number ($T=3$ in B and $T=4$ in C) is formed by appropriate replacement of 12 hexagons with 12 pentagons. This figure is adapted from [13] with permission.

In addition to the MCP, some phages have accessory capsid shell proteins called decoration proteins that bind to the outer surface of the mature capsid [8]. Decoration proteins stabilize the mature capsid, but are dispensable at least in some growth conditions [19]. Phages like T4 and P1 also have proteins packed inside the capsid and injected with their genomes during infection. The T4 internal proteins may assist capsid assembly [20] or phage infection [21].

1.3.3 Capsid assembly

Tailed-phage assembly follows a well-defined pathway in which each step does not proceed until the previous step completes [8]. This manner prevents the production of dead-end structures with missing components. Despite the variations in structural proteins and details in individual assembly, all tailed phages share notable similarities in the overall pathway [8, 22]. Figure 1.3 illustrates a generalized tailed phage assembly. The major capsid protein, portal protein, internal scaffolding protein and protease assemble into a round particle called a procapsid [8]. The N-terminal end of the major capsid protein is then cleaved by the protease, and the scaffolding protein exits the procapsid directly or after being cleaved by the protease [8]. Phage DNA is then pumped into the capsid, coupled with the expansion and maturation of the capsid. The tail is either attached to the completed capsid as a preassembled tail in long-tailed phages, or built on the portal by sequential addition of tail proteins [8].

1.3.4 The evolution of virus capsids

Because all viruses characterized to date lack ribosomes, the classification of living organisms based on 16S rRNA does not work for viruses. The great diversity of virus genome sequences

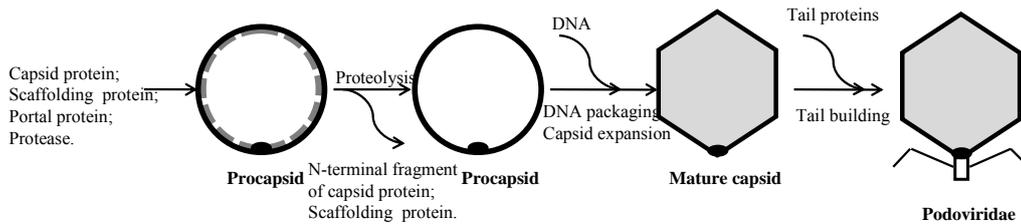


Figure 1. 3 A generalized phage assembly. Redrawn according to [8].

also limits the possibilities for evolution study by sequence comparisons alone [23]. Therefore, viruses were once thought to have multi-origins which depend on the evolutionary types of their hosts [24]. However, recent evidence from the structure of virus capsid proteins has suggested that viruses might have developed from a few common ancestors despite their differences in host classification [25]. The most remarkable structure-based viral evolutionary connection comes from a dsDNA phage PRD1 and the human Adenovirus. Their shared signature is the trimeric MCP characterized by a double eight-stranded antiparallel β -barrel structure [26, 27]. The trimerization of the MCPs forms hexagonal capsomer motifs which build their icosahedral virus shells together with a separate penton protein [28]. In addition to PRD1 and Adenovirus, this distinguishable viral β -barrel fold is shared by most viruses for which atomic resolution MCP

structures have been solved including plant viruses TBSV and TYMV, animal viruses SV40 and Rhinovirus, and bacteriophage ϕ X174 [23]. Another structural lineage of virus capsids is observed in the human Herpesvirus and the large group of tailed phages. Comparison of cryo-EM capsid structures with the HK97 crystal structure reveals that MCPs of these viruses most likely adopt the HK97 gp5 fold [29-31].

1.4 CRYO-EM

Nano-particles like bacteriophages are beyond the capacity of the light microscope. On the other hand, phage particles are too big for Nuclear Magnetic Resonance (NMR) and, with a few exceptions, for x-ray crystallography, techniques that are capable of delivering molecular structures at atomic resolution. Electron microscopy (EM) has been the primary tool for phage classification and structure study since Ruska (1943) pioneered the application of EM in phage research. In the past two decades, the use of EM has been expanded rapidly for phage structure study because of two important approaches in EM techniques in biological research: the use of Cryo-EM that makes it possible to preserve structure details of fragile biological samples [32], and the development of efficient algorithms that succeed in producing high-resolution 3-dimensional (3D) structures from cryo-EM images [33-35]. These two developments allow 3D structure analyses for a wide range of viruses (Figure 1.4). A cryo-EM virus structure at 20 to 30Å is able to visualize capsid features including the surface shapes of capsid proteins and symmetries of the capsid. Moreover, the integration of cryo-EM structures of molecular complexes with NMR and x-ray diffraction subunit models at atomic resolution leads to the ‘divide and conquer’ strategy which has become increasingly popular in structural biology [32].

With state-of-art cryo electron microscopes and powerful computing, the latest record of the resolution for virus capsids by cryo-EM has exceeded 4Å, a resolution where secondary structure can be recognized and side chains of bulky amino acid residues can be assigned [36-38].

For cryo-EM, only a small amount of concentrated virus sample (for example, 5 µl at 3 mg/ml) is needed. The sample is applied to an EM grid with holey carbon support. The grid is then blotted by filter paper to remove most of sample so that only a very thin layer of sample is maintained in the holes of the carbon support. The blotted grid is quickly dipped into cryogen to vitrify the water, avoiding crystallization and preserving the structure of the protein sample. During the whole cryo-EM experiment, the grid is kept frozen below -160°C to avoid formation of cubic phase ice that damages the capsid structure [13, 39, 40]. Cryo-electron microscopy is performed under an electron dose as low as about 10 to 20 e⁻/Å² to avoid damaging the structure of the biological particles which are sensitive to radiation effects [13]. Areas with concentrated but separated particles embedded in a thin layer of ice are imaged with a magnification of about × 20,000 to × 50,000, depending on camera types. The frozen biological sample without any stain is intrinsically low in contrast, and a phase contrast image at focus has no contrast. Therefore, Cryo-EM micrographs are typically taken underfocused in a range of 1 µm to 4 µm to enhance specimen contrast.

Cryo-EM images of phage capsids are essentially two dimensional (2D) projections of the 3D particle. These images are interpreted as different 2D views of identical structures. 3D reconstruction consists of three basic steps [13, 33] (Figure 1.5). Firstly, the initial

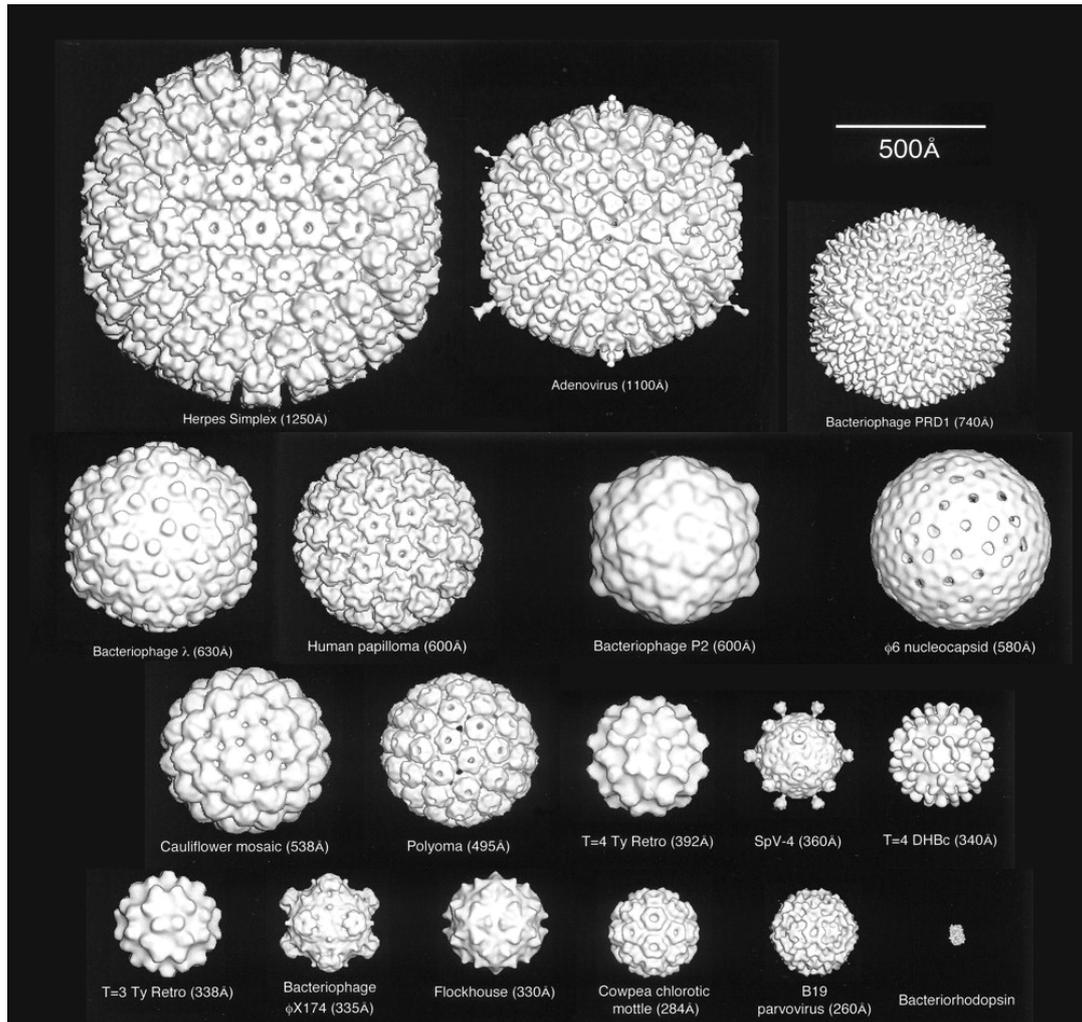


Figure 1. 4 cryo-EM structures of a few virus capsids. Adapted from [13] with permission.

orientation and center parameters of the particles are determined by comparing these particles with a model, either from an existing related structure or from reconstruction of particles with randomly assigned orientations. Secondly, a new model is produced based on these particles and the particle parameters are refined by comparing the particles with the new model. Thirdly, a final 3D structure is calculated from data with optimized parameters. The quality of a 3D structure is commonly estimated by comparing two independent reconstructions from two

subsets of data. A Fourier Shell Correlation score is calculated at all spatial frequencies: a score close to 1 means that the subset structures at this resolution level agree with each other, and FSC = 0.5 is considered as a conservative estimation of the resolution limit.

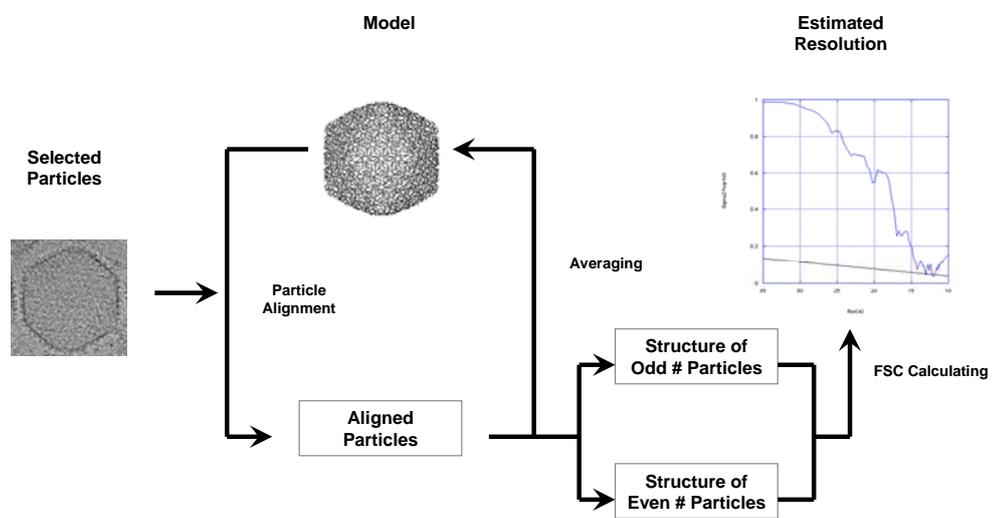


Figure 1. 5 The basic procedure of 3D reconstruction.

CHAPTER 2 SPO1

2.1 INTRODUCTION

SPO1 is a large virulent Myovirus of *Bacillus subtilis*, first isolated by S. Okubo from soil in Japan [41]. Like several other large virulent phages such as SP8, SP5C, SP82, 2C, Φ e and H1, the SPO1 virion has a low GC content genome with hydroxymethyluracil in place of thymine [42, 43]. SPO1 has been the most intensively studied phage among this group, mainly because it is the first example of programmed gene transcription cascades controlled by a set of sigma factors [44]. The complete SPO1 genome is 132,562 bp long, with a 13,185 bp terminal redundancy [45]. The virion has a 1,400Å long contractile tail and a T=16 icosahedral capsid with a maximum diameter of 1,080Å [46-48] (Figure 2.1). The mature SPO1 capsid consists of 955 copies of major capsid protein gp6.1 with about 2kDa removed from the N-terminal end [45, 49].

SPO1 shares the triangulation number T=16 with the herpesviruses. An evolutionary connection of the two capsids was proposed based on a careful comparison of their cryo-EM structures [48] (Figure 2.2). The local three-fold symmetry positions of the SPO1 capsid outer surface have extra density at the symmetry centers, and spikes near the centers protrude out slightly off the local symmetry axis [48]. Since the extra density and the spikes are as strong as the rest of the

capsid and are absent from the local three-fold symmetries near the pentons, this density is reasonably considered to be from a minor capsid protein rather than a part of gp6.1. In the herpesvirus HSV-1 capsid, there are ‘triplex’ molecules sitting in the equivalent positions that also break the local 3-fold symmetry [50] (Figure 2.2). The triplex is also present at sites of local three-fold symmetry adjacent to pentons, but the binding of triplex to these locations is weaker than to the other three-fold sites [51].

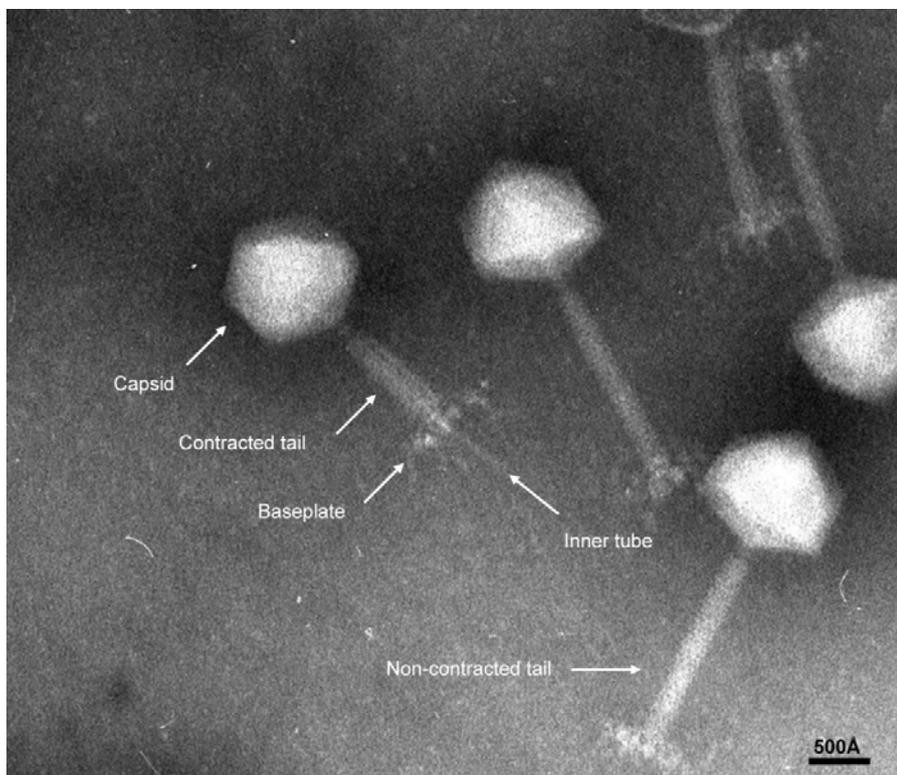


Figure2. 1 A negative-stain EM picture of bacteriophage SPO1. Image courtesy of Dr Robert Duda.

Minor capsid proteins are thought to play two kinds of role in capsid assembly. GpD in phage λ is a ‘decoration’ protein that binds to the mature capsid and stabilizes it but is not required for the

initial assembly of the capsid shell [48]. In contrast, the triplex in HSV-1 is essential to capsid assembly because it mediates the interaction between MCPs in the procapsid [51]. The similarities between SPO1 and HSV-1 capsid structures suggest that the minor capsid protein in SPO1 may be required for procapsid formation as is the HSV-1 triplex, and thus the two viruses would share a similar capsid assembly mechanism. However, it will not be clear if SPO1 minor capsid protein acts as HSV-1 triplex until the SPO1 procapsid is isolated and the presence of minor capsid protein in the procapsid is examined.

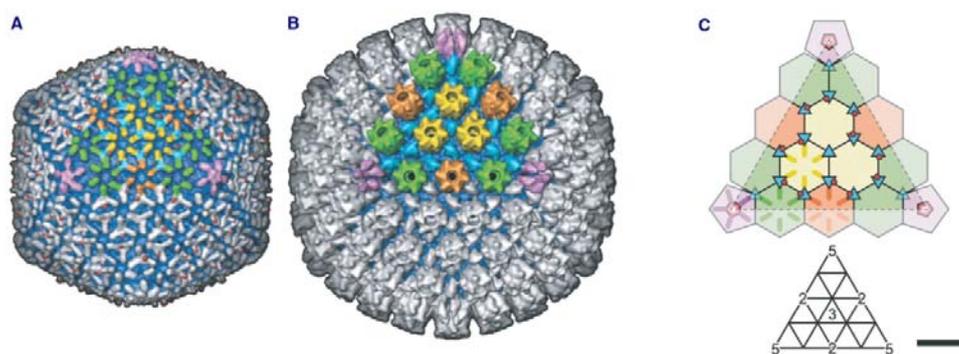


Figure 2. Comparison of SPO1 and Herpes Simplex Virus Type-1 (HSV-1) capsid structures. A) SPO1 capsid structure. Red: asymmetric spikes; light blue: extra densities at local and global 3-fold symmetries. B) HSV-1 capsid structure. Light blue: triplex. C) Schematic of symmetries and asymmetric spikes (red) in one facet of the SPO1 capsid. Adapted from [48] with permission.

As discussed in Chapter 1, comparisons of the structures of pseudo-T=25 phage PRD1 and pseudo-T=25 Adenovirus demonstrate the evolutionary connection between the two viruses. Therefore, comparisons of the functions of the minor capsid proteins of phage SPO1 and herpesvirus in capsid assembly may expand our evidence for common ancestry between tailed phages and animal viruses. In this project, I identified the minor capsid protein of SPO1 and

studied the structure of the SPO1 procapsid with particular focus on the location and function of the minor capsid protein.

2.2 MATERIALS AND METHODS

2.2.1 Strains and vectors

Strains

(Source: Hendrix laboratory collection except as noted)

1) Wild type phage SPO1

Wild type phage SPO1 was used in lysate preparation.

2) Wild type *Bacillus subtilis*

Wild type *Bacillus subtilis* was used as a non-suppressor host.

3) 168M

168M *Bacillus subtilis trpC2* [52] was used as a non-suppressor host in 3.2am1 screening.

(Source: *Bacillus* Genetic Stock Center, The Ohio State University, Columbus, Ohio)

4) CB313

CB313 *Bacillus subtilis sup 3* suppresses amber and ochre mutations and inserts lysine at

UAG and UAA. It was used as a suppressor host in 3.2am1 screening. (Source: Charles

Stewart, Rice University, Houston, Texas)

5) DH10b

DH10b *Escherichia coli F endA1 recA1 galE15 galK16 nupG rpsL ΔlacX74 Φ80lacZΔM15*

araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) λ⁻ was used in cloning.

6) BL21(DE3)

BL21(DE3) *Escherichia coli* *F⁻ ompT gal dcm lon hsdS_B(r_B⁻ m_B⁻) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])* was used in SPO1 capsid protein expression.

Vectors

1) pT7-5

pT7-5 (ampicillin resistance; T7 promoter) was used in cloning and protein expression.

2) pBluescript (II) SK(+)

pBluescript (II) SK(+) (ampicillin resistance; lac promoter) was used in cloning.

3) pLysS

pLysS has Chloramphenicol resistance; pLysS carries the gene encoding T7 lysozyme which lowers the background expression of genes controlled by the T7 promoter. pMK4 [53]

pMK4 is a shuttle vector for *Bacillus subtilis* (chloramphenicol resistance) and *Escherichia coli* (chloramphenicol and ampicillin resistance).

2.2.2 Media and Buffers

1) Luria Broth (LB)

1% (w/v) tryptone (Difco), 0.5% (w/v) yeast extract (Difco) and 0.5% (w/v) NaCl in ddH₂O, autoclaved for 25min. Salts, antibiotics and extra nutrients were added before using.

2) LB agar

LB plus 1.5% (w/v) agar, autoclaved for 25min.

3) Soft agar

1% (w/v) tryptone, 0.5% (w/v) NaCl and 0.7% (w/v) agar in ddH₂O, autoclaved for 25min.

4) Transformation media for DH10b

Autoclaved LB plus 2.5 mM KCl, 10 mM MgCl₂ and 0.4% (w/v) glucose.

5) Transformation media and buffer for 168M and CB313

a) 10x T-Base:

150 mM ammonium sulfate, 800 mM potassium hydrogen phosphate (K₂HPO₄), 440 mM potassium dihydrogen phosphate (KH₂PO₄), 35 mM sodium citrate.

b) SpC (minimal culture media):

100 mL 1x T-Base with 1.0 ml 50% Glucose, 1.5 mL 1.2% MgSO₄, 2.0 mL 10% yeast extract and 2.5 mL 1% casamino acids.

c) SpII (starvation media):

100 mL 1x T-Base with 1.0 mL 50% Glucose, 7.0 mL 1.2% MgSO₄, 1.0 mL 10% yeast extract, 1.0 mL 1% casamino acids and 0.5 mL 100 mM CaCl₂.

6) SPO1 phage dilution buffer

20 mM Tris-HCl pH 7.5, 10 mM MgSO₄ and 40 mM NaCl in ddH₂O, 0.2 µm filtered.

2.2.3 Methods

2.2.3.1 Phage manipulations

1) Preparation of liquid lysate

Fresh stationary *Bacillus subtilis* culture (10-hour-old overnight culture or 1/20 overnight culture incubated at 37°C on a 250 rpm shaker for 3 hours) was diluted 500-fold with LB

and grown at 37°C on a 250 rpm shaker. The cell density was monitored by reading the absorption at 550 nm in a spectrometer. After ~2.5 hours, the 550 nm measurement reached 0.1, which indicates the cell density as about 1×10^8 cells/mL. SPO1 phage stock was added to the culture with a multiplicity of infection (m.o.i) = 0.1 for wild type phage and m.o.i. = 2 for amber phage. In ~1.5 hours, the lysis completed and the culture became clear.

2) Phage titering

10 µl of SPO1 phage dilution was mixed with 100 µl of fresh stationary *Bacillus subtilis* culture. 3 ml of warm melted soft agar was added and poured on LB agar plates. After the soft agar was hardened, the plates were incubated at 37°C overnight.

2.2.3.2 SPO1 phage and capsid purification

1) Phage particle purification and concentration

1L of SPO1 liquid lysate was centrifuged in a JA10 rotor at 9k rpm, 4°C for 15 minutes. The supernatant was carefully collected and the pellet was discarded. 0.5 M NaCl and 15% (w/v) PEG 8000 was added to the supernatant while it was vigorously agitated. After the PEG 8000 was completely dissolved, the mixture was incubated at 4°C overnight. Second day, the mixture was centrifuged in a JA10 rotor at 9k rpm, 4°C for 15 minutes. The pellet was resuspended in 50 mL SPO1 buffer. The resuspended particles were centrifuged in a JLA16.250 rotor at 9k rpm, 4°C for 10 minutes. The supernatant was carefully collected, diluted, and centrifuged in a Ti45 rotor at 35k rpm, 4°C for 1.5 hours to concentrate the particles and remove the retained PEG. The pellet was then resuspended in 5 mL SPO1 buffer at 4°C overnight. Third day, the insoluble pellet was removed by minicentrifugation.

2) Procapsid and empty capsid purification

5% - 45% sucrose gradient in SPO1 buffer was prepared by GradientMaster. The concentrated particles were run through the gradient in a SW41 rotor at 35k rpm, 20°C for 30 minutes. The procapsids and empty capsids, identified by light scattering, were located in the middle of the gradients. The sucrose in the procapsid sample was removed by dialysis against 2L SPO1 dilution buffer at 4°C overnight, and the sample was further purified by BioCAD Sprint ion-exchange chromatography.

3) Purification of phage by continuous CsCl gradient

CsCl salt was added to the concentrated lysate to a density of 1.5 g/mL. The mixture was sealed in a plastic tube and centrifuged in a Ti 70.1 rotor at 18°C 38k rpm for 16 hours. The phage band in the middle of the gradient was collected by syringes and the CsCl in the sample was removed by dialysis.

2.2.3.3 DNA and protein techniques

1) Plasmid extraction

Plasmids were extracted and purified with a QIAGEN miniprep kit. For plasmid extraction from *Bacillus subtilis*, 1 mg/mL chicken egg white lysozyme was added after the cell pellet was completely resuspended in buffer P1. The reaction was incubated on the bench for 30 min to digest the peptidoglycan layer.

2) Phage DNA preparation

SPO1 DNA was extracted from phages purified by CsCl gradient by phenol extraction [54].

3) Plasmid construction

a) Genes 6.1 and 3.3 cloning

SPO1 phage DNA was digested with TaqI (NEB) and electrophoresed through a 0.8%

agrose gel. The 3kb fragment containing genes *6.1* and *3.3* was extracted from gel band, blunted by Klenow (NEB). The fragment was subcloned into pBluescript (II) SK(+) vector and then transferred into the expressing vector pT7-5. The plasmid is marked as pT7-5-SPO12G.

b) Gene *29.2* cloning

Gene *29.2* was amplified by PCR reaction using two primers

gtgtccgagctcGATATTATGCCTTTCCTGTGATCTGA carrying a SacI site and

tcgaccgatccAGGATACGAAGTATCCTCTTTCCTGC carrying a BamHI site, and SPO1

phage DNA as the template. (The sequences in lowercase are the restriction sites introduced

by the primers.) The fragment was cloned into a pT7-5 vector. After the sequence was

verified by sequencing, the *29.2* was cloned into pT7-5-SPO12G. The plasmid containing

6.1, *3.3* and *29.2* is marked as pT7-5-SPO13G.

c) Gene *3.1* and *3.2* cloning

Genes *3.1* and *3.2* was amplified by PCR reaction using two primers

tatagatccGCATGCACGTACAGTGTATGC carrying BamHI and

ACTTTTGAAGGCGCTCAGTAGTGC carrying NheI, and SPO1 phage DNA as the

template. The fragment was cloned into a pT7-5 vector, and verified by sequencing. The

fragment was cloned into pT7-5-SPO12G and pT7-5-SPO13G, marked as pT7-5-SPO14G

and pT7-5-SPO15G.

d) Gene *3.2am* cloning

An AAG to UAG mutation in gene *3.2* in plasmid was introduced to the DNA as described

in [55]. Two fragments were amplified by PCR reaction using phage DNA as the template

and two pairs of primers: atatgaattcTCTGTCATCAACCGTTATGTTA and

CTCAGCAAATGGATAACCCA; TAGTGCGTAGTCACTGAAAAA and
atatggatccGCTGAGACTTAACAATACCTTG. The first and fourth primers were treated
with phosphatase before the PCR reaction. The two fragments were ligated and the joint
fragment was amplified by PCR reaction using the ligation product as the template and the
first and fourth primers. The 1.5 kb fragment was subcloned into pT7-5 vector, and cloned
into the shuttle vector pMK4 after the sequence was verified by sequencing. An AAG to
UAG mutation in gene 3.2 was introduced by the first three nucleotides of the third primer.

4) Phage protein preparation for SDS-PAGE

Phage stock was minicentrifuged at 14k rpm at room temperature for 30 minutes (this step
was skipped if the protein concentration was higher than 5mg/ml). Phages were resuspended
in 10 mM DTT and 10 mM EDTA was added. The mixture was incubated in a 70°C water
bath for 2 minutes, and vortexed every minute. This turned the sample to jelly. The sample
was then incubated at 30°C mixed with 10 mM MgSO₄ and 0.05 mg/mL DNase I, until the
viscosity of the sample had decreased appreciably.

5) Restriction digestion, ligation and transformation

Restriction digestion and ligation were performed according to the manuals of specific
enzymes used. Transformation in DH10b was conducted by electroporation of 1 µl ligation
mixed with 40 µl competent cells by a BioRAD GenePulser. Competent *Bacillus subtilis*
was prepared by 1) dilution of the overnight culture 10-fold with SpC medium and
incubation of the culture at 37°C 250 rpm for 3 hours; and 2) dilution of the culture by
adding an equal volume of SpII medium and incubation at 37°C 250rpm for 2 hours.

Transformation in *Bacillus subtilis* was performed by incubating 500 mL competent cell and
50 ng plasmid at 37°C 250rpm for 30 minutes.

6) Protein expression

Genes cloned into pT7-5 plasmid were expressed in BL21(DE3)plysS cells with the same method used for HK97 gp5 expression [14].

2.2.3.4 Isolation of SPO1 amber mutant 3.2am1

In order to produce procapsid, an SPO1 amber mutant in protease gene 3.2 was isolated by in vivo recombination which is reported at a rate of about 10^{-4} [56]. A piece of ~1.5 kb DNA which contains gene 3.2 in the middle was chosen for mutagenesis. An AAG to UAG mutation in gene 3.2 was introduced to the DNA as described in [55]. The DNA was subcloned into a pT7-5 vector, and cloned into a shuttle vector pMK4. The pMK4 plasmid was then transformed into 168M cells. Recombination occurred when wild type SPO1 infected 168M cells carrying the plasmid, and a small number of SPO1 mutants were released along with the wild type majority. The lysate was titered to produce scattered plaques. Of more than 14,000 plaques tested, one produces tiny plaques on suppressor strain CB313 plate but does not infect non-suppressor strain 168M. The amber mutation in gene 3.2 was confirmed by sequencing.

2.2.3.5 Cryo electron microscopy and capsid reconstruction

Cryo-EM experiments were performed on an FEI Tecnai F20 microscope. Micrographs were recorded by a CCD camera or a film camera at a magnification of 50k for the CCD camera or 30k for the film camera. The films were digitalized by a Nikon Coolscan 9000 scanner at a digitalization step of 6.35 μm , corresponding to 2.12 \AA /pixel at the sample. Particles of high quality were picked by the X3D program [57]. The defocus of each micrograph was determined

by SUMPS and CTFZERO (James Conway, University of Pittsburgh), or by the Bshow software in the BSOFTE package [58, 59]. Capsid reconstruction was done with AUTO3DEM [60].

2.3 RESULTS

If the function of the minor capsid protein in SPO1 capsid resembles that of the HSV-1 triplex, this protein is required for procapsid assembly and should be found in the procapsid. To study the role of the minor capsid protein in SPO1 capsid assembly, three methods were tried to isolate SPO1 procapsids.

2.3.1 Empty capsids from wild type SPO1 lysate

Empty particles were first isolated from wild type SPO1 lysate. These particles were concentrated by PEG precipitation and purified by sucrose gradient and ion-exchange chromatography, and identified by negative-staining EM to be empty of DNA. The capsid only has two proteins visible in SDS polyacrylamide gel (Figure 2.3). Both of them are also found in the phage protein sample in the same gel but not in the baseplate and tail samples, which confirms that they are capsid proteins. The light band was identified by Edman degradation as gp29.2, which has a predicted mass of 19.8 kDa, compatible with its mobility on the SDS gel.

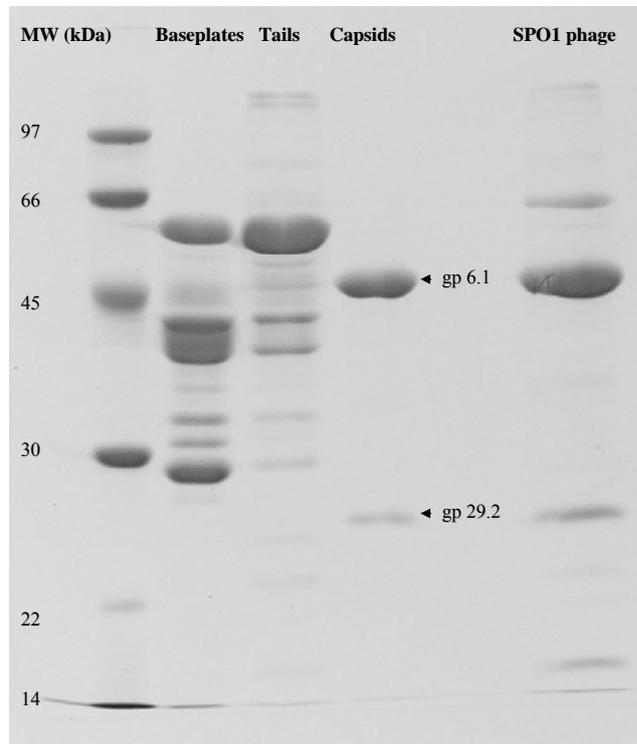


Figure 2. 3 SDS polyacrylamide gel of samples purified from wild type SPO1 lysate. The capsid sample is purified by sucrose gradient and ion-exchange chromatography. The two capsid proteins identified by Edman degradation are cleaved gp6.1 and gp29.2. The baseplate and tail samples are dialysed sucrose gradient bands. The phage sample is purified by CsCl gradient.

The 51 kDa heavy band gave the N-terminal sequence of cleaved gp6.1 (MCP). The portal protein gp3.1 is not visible on the gel. Cryo-electron microscopy of the particles was performed by Professor James Conway and a 17.5Å structure was calculated using 790 particles out of 1240 picked from 25 CCD images. Comparison of this structure with a 20Å phage capsid structure shows that the outer surface features are very similar (Figure 2.4). Notably, the structure of the isolated particle has asymmetric density distributed exactly as the phage structure, so the gp29.2 must be the minor capsid protein. A typical procapsid is more spherical compared with the

mature capsid, and is composed of uncleaved major capsid protein. Therefore, the empty particle is not procapsid, but expanded mature capsid that has lost its DNA or did not package DNA.

2.3.2 Cloning and expression of SPO1 capsid genes

Encouraged by the highly efficient procapsid assembly from HK97 gp5 expression in *Escherichia coli* [14], I cloned and expressed a number of SPO1 capsid genes to produce procapsids. There is a cluster of capsid genes including 3.1 (portal), 3.2 (protease), 3.3 (scaffolding protein) and 6.1 (major capsid protein) in the SPO1 genome. 29.2 is far away from this cluster (Figure 2.5A). Four combinations of capsid genes were cloned into pT7-5 plasmids and expressed in BL21(DE3)-plysS cells, including:

- 1) 2G: 3.3 and 6.1 (Scaffolding protein and MCP)
- 2) 3G: 3.3, 6.1 and 29.2 (Scaffolding protein, MCP and minor capsid protein)
- 3) 4G: 3.1, 3.2, 3.3 and 6.1 (Portal, protease, scaffolding protein and MCP)
- 4) 5G: 3.1, 3.2, 3.3, 6.1 and 29.2 (Portal, protease, scaffolding protein, MCP and minor capsid protein)

Figure 2.5B shows SDS-PAGE of ³⁵S-methionine labeled expressions of the four combinations. In the 2G and 3G expressions, gp6.1 and gp3.3 are expressed at a considerable level. The 4G and 5G expressions are much lower. Gp29.2 is not visible because it does not have any methionine. The expression products were purified as described in [14], and examined with negative-stain EM. Long tube-like polyheads were observed in the 2G and 3G products (Figure 2.5C), and no procapsid-like particles were found in the 4G and 5G products. The 2G and 3G results suggest that gp6.1 was correctly folded and formed hexamers, but failed to assemble into icosahedral

capsids in BL21(DE3) cells. The accumulation of gp6.1 led to the formation of polyheads. Other proteins may be required for SPO1 procapsid assembly, such as portal or protease. However, gp3.1 and gp3.2 may be toxic to the cell, which would lead to the poor expressions of 4G and 5G.

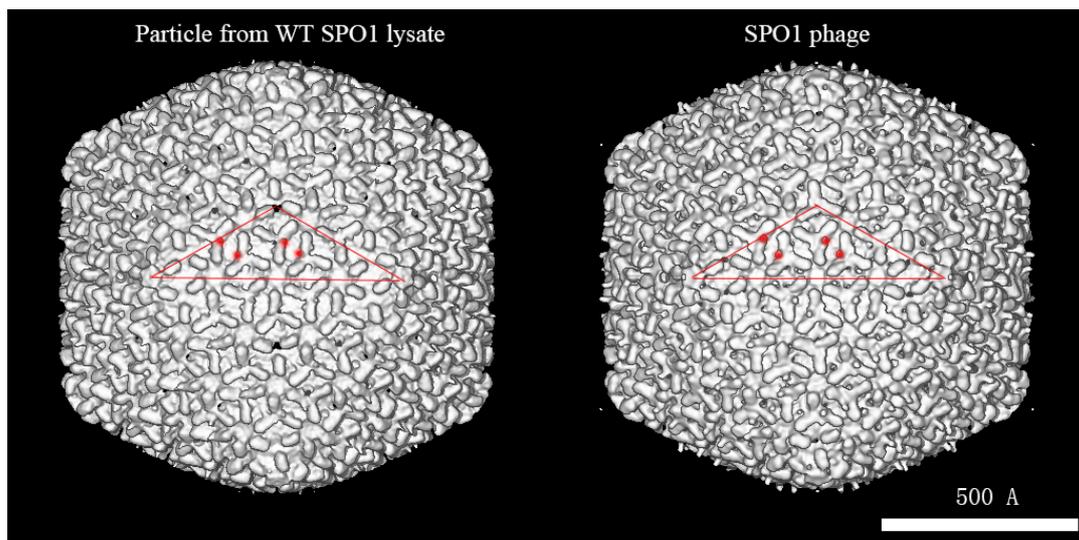


Figure 2.4 Cryo-EM structures of the empty particle (17.5 Å) and SPO1 phage (20 Å). The surface features of the two structures are very similar. Both of them have the asymmetric densities off the local 3-fold. An asymmetric unit is indicated by the red triangle.

2.3.3 Capsids isolated from 3.2am1 lysate

A mutant 3.2am1 with an amber mutation within the protease gene 3.2 was isolated. 3.2am1 forms tiny plaques on soft agar plates of the suppressor strain CB313 (<0.5mm in diameter in 0.3% soft agar plate for 3.2am1, and ~2mm for wild type). The reversion rate of 3.2am1 is high. In all three individual 3.2am1 plate stocks which were prepared from single plaques, about 0.1% of phages were able to form plaques on the non-suppressing host 168M. Nevertheless, these plate

stocks were still used to produce procapsids by one time infection to 168M cells at an m.o.i of 2.0, and particles were purified by sucrose gradient and ion-exchange chromatography.

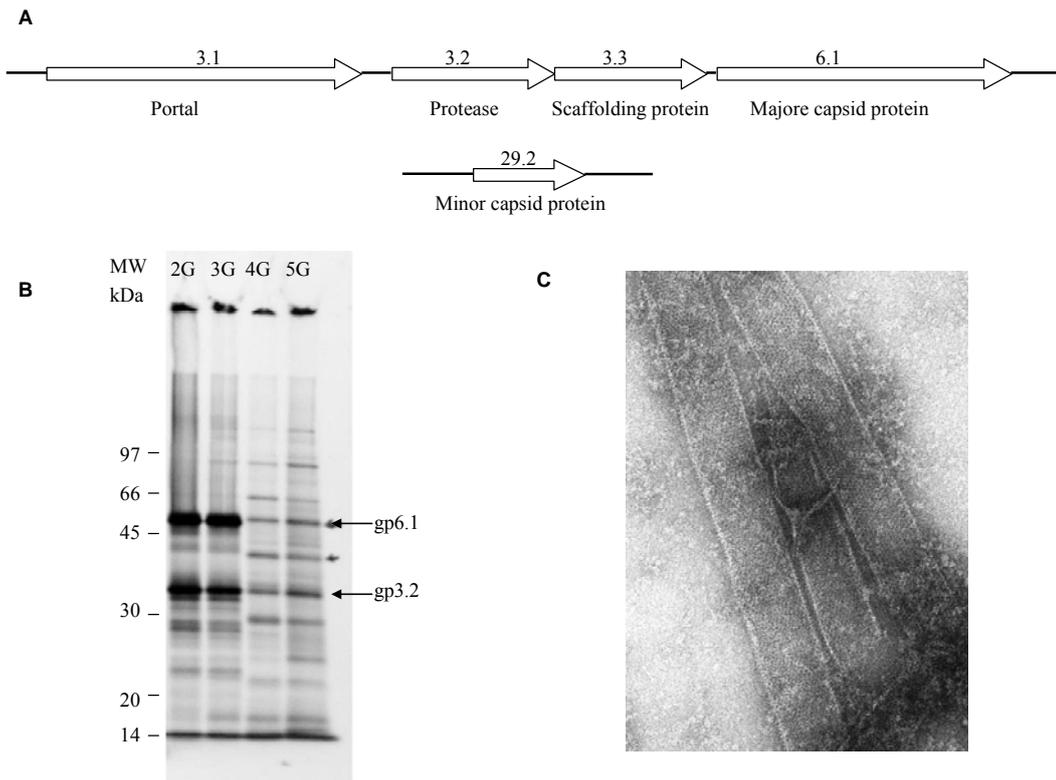


Figure 2.5 Expression of SPO1 capsid genes in BL21(DE3). A) Positions and gene numbers of portal, protease, scaffolding protein, major capsid protein and minor capsid protein in SPO1 genome. B) ³⁵S-Methionine labeled expressions of 4 combinations of capsid genes. The 2G and 3G expressions are high and the 4G and 5G expressions are very low. Gp6.1 and gp3.2 are marked. C) Negative-stain EM image of 2G expression product.

The concentrated lysate was fractionated on a 5% - 45% sucrose velocity gradient and generated four bands (Figure 2.6A). Three bands B1-3 were collected, dialysed and electrophoresed through an SDS polyacrylamide gel. B1 is the tail band conformed by negative stain EM (Figure 2.6B).

B2 and B3 ran closely with each other at the middle where empty capsids are expected to be found (Figure 2.6A). The two bands contain the same capsid proteins including gp6.1 as the major band, some cleaved gp6.1, gp3.3 and gp29.2 (Figure 2.6B). This result suggests that the protease activity is effectively suppressed, and both B2 and B3 may contain procapsids with largely uncleaved MCP, scaffolding protein and the minor capsid protein. A small amount of cleaved gp6.1 is expected because the phage stock has 0.1% WT SPO1 which produces expanded capsids as described in 2.3.1. B3 capsids should have more scaffolding protein than B2 capsids because of the higher gp3.3/gp6.1 ratio in B2 showing in the gel. This result explains why B3 runs faster in the sucrose gradient: B3 capsids contain more gp3.3 and are heavier than B2 capsids. It also indicates that the isolated capsid may slowly lose gp3.3 during purification. B2 and B3 were collected together for further purification because: 1) they are very close and difficult to separate; 2) they may be better separated in the next step of purification; 3) B3 particles may slowly turn into B2 particles by losing gp3.3.

The mixture of B2 and B3 was further purified by Biocad SPRINT ion-exchange chromatography. Figure 2.7 shows the chromatograph and the SDS-PAGE of the four samples fractionated. All of the four fractions contain uncleaved gp6.1 as their major protein. Unlike the F2 lane, the F1 lane does not have a visible gp29.2 band. In addition, F1 is not contaminated by WT expanded capsids because the F1 lane does not have a cleaved gp6.1 band. The protein concentrations of F3 and F4 are very low.

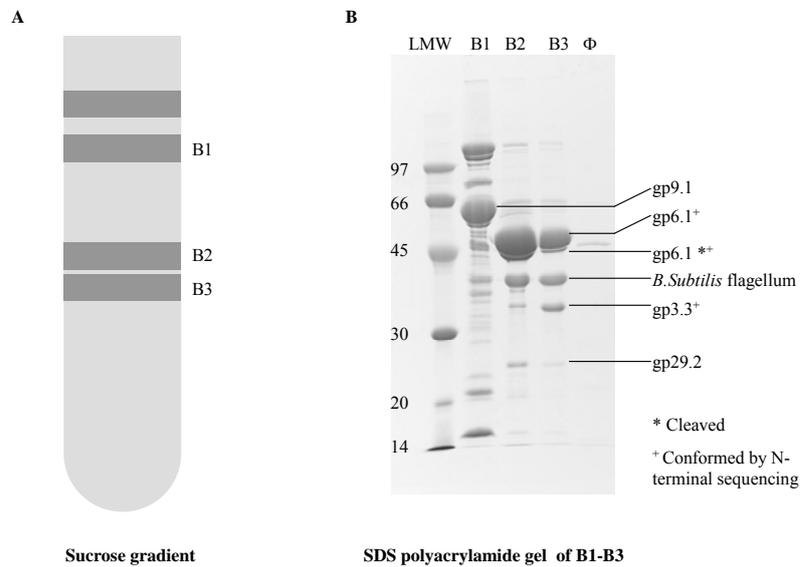


Figure 2.6 Sucrose gradient pattern of 3.2am1 lysate and polyacrylamide gel of sucrose gradient bands. A) Pattern of 3.2am1 lysate running through 5% - 45% sucrose gradient. The top band should be baseplate according to 2.3.1. B) B1-3 were electrophoresed in a SDS polyacrylamide gel. B1 has a heavy band of sheath protein gp9.1 and other tail protein bands. Both B2 and B3 have a large amount of gp6.1 band, and smaller amounts of gp6.1* (the cleaved form of gp6.1), gp3.3 and gp29.2 bands. B3 has a higher gp3.3/gp6.1 mass ratio than B2.

Cryo-EM visualization of the F1 sample was performed by Professor Conway, and from these images I calculated a cryo-EM structure. For unknown reasons, both F1 and F2 samples tend to form a mono layer on the cryo-EM grid of closely packed empty capsids which preferentially offer the 3-fold symmetry view. The packing is so close that distortions in the shell of some capsids are visible to the eye, which causes variations in individual capsids and may harm the resolution of the reconstruction.

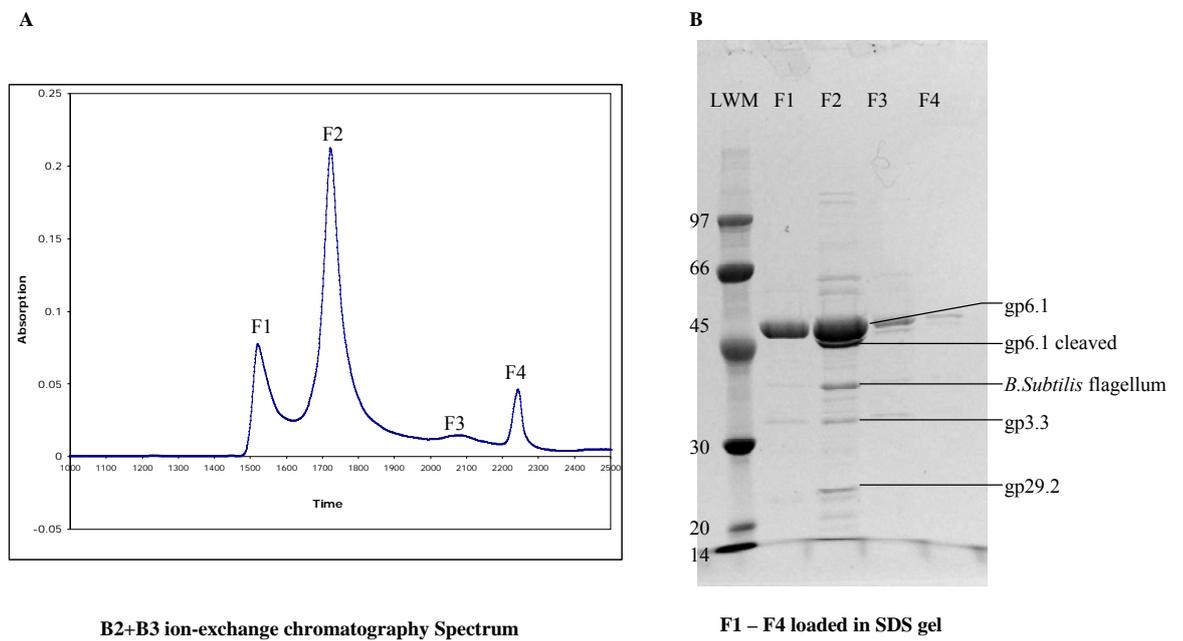


Figure 2. 7 Proteins from peaks isolated by ion-exchange chromatography of the B2 and B3 mixture. A) Spectrum and fractioned peaks of the ion-exchange chromatography. B) SDS polyacrylamide gel of F1-4. All of the four Peaks have the uncleaved gp6.1 as their major protein. F1 is the only one that does not have a visible cleaved gp6.1 band.

In the reconstruction, 5013 particles were selected from a total number of 6333 which were picked from 17 scanned film micrographs. The 17.5Å cryo-EM structure of capsids in F1 is shown in Figure 2.6 left. The F1 capsid has an overall icosahedral structure similar to the phage. However, the asymmetric spike (highlighted with red dots in one asymmetric unit in Figure 2.8 right) located near the local 3-fold symmetries of SPO1 phage structure is missing in the F1 capsid structure. Furthermore, comparison of the centers of the local 3-fold symmetries (one example is circled with red) of the two structures shows that another piece of density is missing

in the left structure. The data here strongly suggest that the missing asymmetric spikes and central density in the local 3-fold symmetries are a result of lacking gp29.2. A question arises from this conclusion: why is there the extra density at the center of the global 3-fold sites in the phage structure, but no asymmetric spike? An explanation is that the icosahedral symmetries including the global 3-fold symmetries were fully utilized during capsid reconstruction so the asymmetric spike was averaged out. Other minor differences in the surface of the two structures are observed, but at the current resolution it is difficult to tell if they are caused by the missing gp29.2.

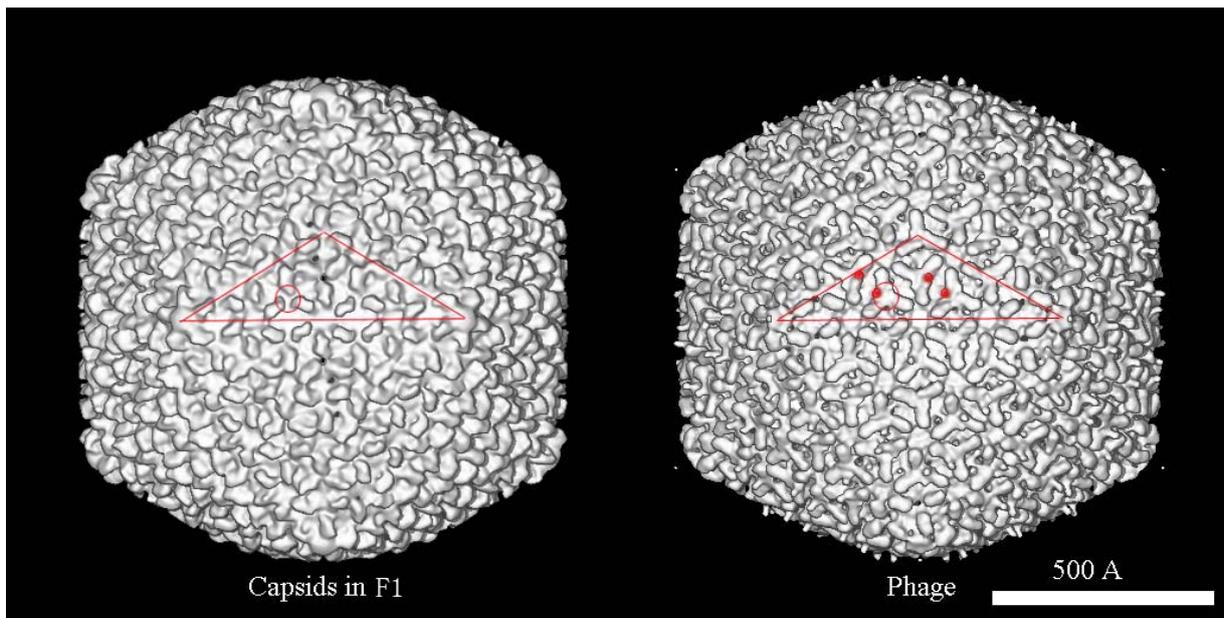


Figure2. 8 Cryo-EM structures of capsids in F1 (17.5Å) and SPO1 phage (20 Å). The overall structures of the two structures are similar. However, the structure of the capsid F1 does not have the spikes off the local three-fold symmetry axes, and the density at the centers of local three-fold symmetry as seen on the phage (circled by red).

2.4 SUMMARY AND DISCUSSION

In the SPO1 project, I identified the minor capsid protein gp29.2 by SDS-PAGE analysis of the empty SPO1 capsid. Combinations of capsid genes were expressed in BL21(DE3), but failed to assemble into procapsids either because of wrong assembly into tubes or low expression levels.

I isolated an amber mutant 3.2am1 in the protease gene 3.2, hoping that capsid particles produced by SPO1 without protease function will stay at the procapsid stage. Two types of capsid containing uncleaved gp6.1 with or without the minor protein gp29.2 were isolated. Both of them contain gp3.3 scaffold and may lose it during purification. A cryo-EM structure was obtained for the F1 capsid, which lacks gp29.2. The surface of the F1 capsid structure resembles the empty SPO1 mature capsid. The new structure does not have the asymmetric spikes off local 3-fold symmetry axes and the extra density at the centers of local and global 3-fold symmetries, presumably a result of the lack of gp29.2, demonstrating that gp29.3 may not be essential for capsid assembly.

Current knowledge on phage capsid assembly suggests that the procapsid undergoes major conformational changes (expansion) during DNA packaging. A procapsid is usually more spherical than its expanded counterpart. For instance the HK97 procapsid I cryo-EM structure looks smaller and less angular than the structures of the mature capsid and expanded procapsid I induced by 5 M urea [61]. The shell of HK97 procapsid I is apparently thicker and protein mass is more unevenly distributed [61]. The expanded and unexpanded T4 polyheads exhibit large differences in the surface lattice images, suggesting that there are significant conformational changes during capsid maturation [62]. However, the overall structure of the SPO1 F1 capsid

structure of SPO1 is similar to the mature capsid structure, suggesting that it has been expanded without proteolysis. Although the structure of the F2 capsid has not been solved, it is probably expanded too, because the F2 capsid ran in a very similar position in the chromatographic profile as the empty capsid did, which implies that they have similar surface properties. The 2kDa N-terminus of gp6.1 and the gp3.3 should not affect the interaction with the ion-exchange column because they are most likely inside in the capsid.

Our original hypothesis is that if the minor capsid protein is essential to procapsid assembly, then the minor capsid protein should be found in procapsids. The F1 and F2 particles that unexpectedly expanded without proteolysis open two possibilities: 1) gp29.2 is required for procapsid assembly but dispensable for expanded/mature capsids. The gp29.2 binds to the procapsid, but expansion lowers its binding affinity and gp29.2 slowly dissociates from expanded capsids; 2) gp29.2 is not required for procapsid assembly and binds to the expanded capsid. The observation that empty particles isolated from WT lysate do not lose gp29.2 (data not shown), which suggests that binding affinity of gp29.2 is mature-capsid-oriented, would favor the latter possibility. However, neither can be excluded until an amber mutant in gene 29.2 is isolated or an unexpanded SPO1 procapsid (if it is different from the expanded capsid, similar to other phages) is purified and studied. The condition of high concentrations of sucrose in gradient purification may cause abnormal conformational changes in HK97 procapsid preparation (Brian Firek, personal communication), so other SPO1 procapsid separation methods are worth trying for SPO1. Examination by EM of fresh 3.2am1 lysate may show if there are unexpanded capsids before purification.

Another combination of SPO1 capsid genes, *3.1*, *3.3*, *6.1* and *29.2* may be worth expressing. The portal protein is required for procapsid assembly of some phages. The failure of the previous combinations could be because gp3.1 (portal) is required for capsid assembly and gp3.2 (protease) is toxic to BL21(DE3). In addition, other phage proteins or the host *Bacillus subtilis* environment may be required for the correct assembly of SPO1 procapsids.

CHAPTER 3 PHAGE P1

3.1 INTRODUCTION

P1 is a model temperate phage with a wide host range that includes *Escherichia coli* and several other enteric bacteria, and was first isolated by Bertani in 1951. After infection, the choice of entering a lytic or lysogenic path is controlled by the translation and activity of a repressor, C1, sensing a number of environmental factors [63, 64]. The P1 prophage is maintained as a circular plasmid with one copy per bacterium. Because of its capacity for carrying large pieces of DNA from hosts, P1 has been used as a generalized transducing phage in *Escherichia coli*. The complete P1 genome was reported to be 94 kb long, encoding at least 117 genes [63].

P1 has the unusual ability to produce virions, in normal infections, with identical tails but icosahedral capsids of different sizes [65, 66] (Figure 3.1). The three capsids have face-to-face diameters of 86nm (P1b), 65nm (P1s) and 47nm (P1m) estimated from negative stained EM images [65]. Their T-numbers were estimated to be 16, 9 and 4 respectively [65]. Although all three virions are able to inject their DNA into a host cell, only the largest virion P1b has enough internal volume to carry the complete genome and is the viable virion [65]. In a lysate of wild type P1, P1b is always the majority, and P1m is the rarest. The proportions of P1s and P1m may depend on host strains [65]. P1 mutants with an 8.86kb dispensable region between the invertible

C-segment and the resident *IS1* element deleted exhibit the Vad- (viral architecture determinant) phenotype which produces a majority of P1s [67, 68]. SDS-PAGE analysis of P1 wild type and Vad- mutants shows that P1s has the same proteins as P1b, and that the vad- virions lack at least 4 proteins present in wild type virions, including DarA, which is encoded within the 8.86kb region, and another protein, DarB [68]. Both DarA and DarB (“dar” is an acronym for “defense against restriction”) are known as dispensable proteins that protect phage genomes against host restriction systems [69, 70].

SDS-PAGE of P1 virions also shows two groups of unusual ladders of protein bands with 1-2 kDa spacing [68]. One group of ladder is related to capsid proteins and the other to tail proteins [68]. SDS-PAGE of a close relative of P1, phage P7, shows a very similar protein pattern to P1 but no ladders [68]. The ladders may be related to cysteine modification because P1 phage proteins treated with reducing agents show a reduced number and intensity of ladders [68].

The ubiquity and abundance of the bacteriophage world means that the competition among different phages is intense. Therefore, it is very important for phages to efficiently produce viable virions. In terms of capsid assembly, the major capsid protein should be selected by evolution to be able to precisely assemble into a normal geometry. As discussed in 1.3.2, the HK97 MCP alone can precisely assemble into a T=7 capsid [14]. However, bacteriophage P1 is a rare example that produces three capsids of different T-numbers under normal laboratory growth conditions. Study of the structures and proteins of the three P1 phages may provide insight into the mechanism that controls the size of viral capsids. This chapter discusses the capsid structures

of three P1 phages, the identification of capsid proteins that may participate in size-control in P1 capsid assembly, and the attempts to identify the P1 ladders by mass spectrometry.

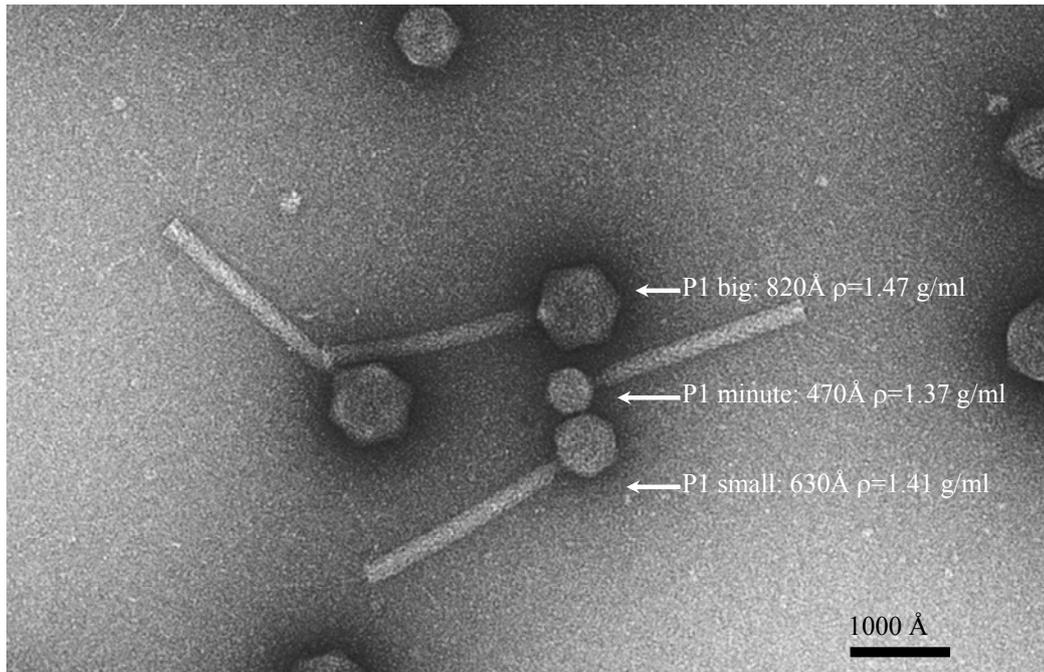


Figure 3. 1 A negative stain EM image of P1 phages with capsids of different sizes. Image courtesy of James Conway.

3.2 MATERIALS AND METHODS

3.2.1 Strains

1) P1_{vir}

P1_{vir} is a virulent mutant whose lysogenic pathway is suppressed.

2) LE392

LE392 *Escherichia coli* *glnV44 supF58 (lacY1 or ΔlacZY) galK2 galT22 metB1 trpR55 hsdR514(rK-mK+)* was used as a P1 host.

3) C600

C600 *Escherichia coli* *F- tonA21 thi-1 thr-1 leuB6 lacY1 glnV44 rfbC1 fhuA1 λ-* was used as a P1 host.

3.2.2 Media and buffers

1) P1 dilution buffer

10 mM Tris-HCl, pH 7.5, 5 mM MgSO₄ and 2.5 mM CaCl₂ in ddH₂O, 0.2 μm filtered.

2) 8x LB nutrients

8% (w/v) tryptone (Difco), 4% (w/v) yeast extract (Difco) dissolved in ddH₂O, autoclaved for 25 minutes.

3.2.3 P1 phage manipulation

1) Preparation of P1 lysate

Overnight LE392 or C600 culture was diluted 500-fold with LB containing 5 mM MgSO₄, 2.5 mM CaCl₂ and 0.2% glucose. The culture was incubated at 37°C on a 250 rpm shaker. The cell density was monitored by reading the absorption of the culture at 550 nm with a spectrometer. After ~2 hours, the 550 nm measurement reached 0.1. P1_{vir} phage stock was added to the culture with a multiplicity of infection (m.o.i) = 0.2. In ~1.5 hours, the lysis was complete and the culture became clear.

2) Purification of P1b, P1s and P1m

Cell debris in the P1 lysate was removed by centrifugation in a JA 10 rotor at 4°C, 9k rpm for 15 minutes. The supernatant was carefully collected and 1 µg/mL DNase I was added. The supernatant was then incubated at 30°C for 30 minutes. The lysate was concentrated by centrifugation in a Ti45 rotor at 4°C, 35k rpm for 1.75 hours. Phage pellets were resuspended with P1 dilution buffer and further purified in a 1.4 g/mL CsCl gradient, which separates P1s and P1m bands better than the standard 1.5 g/mL CsCl gradient.

3) Purification of P1b empty capsid

Lysate was prepared essentially as described in 3.2.3, except that 50 mL 8x LB nutrients was added immediately after the addition of phage stock. P1b empty capsids were purified from this lysate using the same protocol as SPO1 empty capsids described in 2.2.3.2.

4) Sample preparation for mass spectrometry

MALDI MS samples were prepared using in-gel trypsin digestion as described in [71]. Protein bands were cut from SDS polyacrylamide gel and soaked in CH₃CN. The protein was reduced by 10 mM DTT in 50 mM NH₄HCO₃, and then alkylated by 50 mM iodoacetamide in 50 mM NH₄HCO₃. The water in the gel pieces was substituted by CH₃CN by sequentially soaking in mixtures of 50 mM NH₄HCO₃ buffer and CH₃CN with increasing CH₃CN ratios. The gel pieces were dried using a speed vac and then soaked in 0.02 mg/ml trypsin in 50 mM NH₄HCO₃ at 4°C for 1 hour. The excess liquid was removed, and the gel pieces were covered by 50 mM NH₄HCO₃ buffer. The reaction was incubated at 37°C for 20 hours. The supernatant was collected, and the peptides remaining in the gel were extracted by water and CH₃CN, sequentially. The supernatants were merged and dried with a speed vac.

5) Alkylation of P1 empty capsid and P1 whole phage proteins by iodoacetamide (IIA) 20 μ L saturated urea at 37°C was added to 20 μ L P1b empty capsid or 20 μ L P1b phage protein sample, and incubated at 55°C for two hours. 40 μ L 0.1M IIA in 8 M urea 0.1 M Tris-HCl pH 8.0 was added to each reaction. The reaction was incubated in the dark at room temperature for 1 hour. 8 μ L β -mercaptoethanol, 20 μ L 30% SDS and 8 μ L 50% glycerol were added to 20 μ L of each reaction. The mixtures were then incubated at 100°C for 2.5 minutes.

3.3 RESULTS AND DISCUSSION

3.3.1 P1 morphology

With different DNA/protein mass ratios, the three types of P1 virion have different densities. Thus, they can be separated by a CsCl equilibrium gradient. The concentrated P1 *vir* lysate was separated into 5 visible bands in a CsCl gradient (Figure 3.2). The bands were examined by negative stain EM. The contents and densities of these bands are (from top to bottom):

- a Empty phages, lipid, P1m & P1s, $\rho = 1.368$ g/mL.
- b P1b, P1s, empty phages & lipid, $\rho = 1.382$ g/mL.
- c P1s, P1b & P1m, $\rho = 1.413$ g/mL.
- d P1s, P1b, P1m & lipid, $\rho = 1.425$ g/mL.
- e P1b & P1s, $\rho = 1.438$ g/mL.

Bands that have the highest concentrations of P1b, P1s and P1m are bands e, c, and a, respectively. The three bands were examined by cryo-EM by Professor James Conway and Dr. Dalaver Anjum. 3D structures of P1b (27Å), P1s (15Å) and P1m (27Å) were reconstructed (Figure 3.3) based on the data list in Table 3.1. P1b is a T = 13 capsid with a maximum diameter (vertex to vertex) of 930Å. P1s is a T = 7 capsid with a maximum diameter of 710Å. P1m is a T = 4 capsid with maximum diameter of 540Å. Despite the minor differences in surface details which may be a result of structures at different resolutions, the outer surface view of pentons and hexons of the three capsids are similar. Also, there is no extra density showing in any region where pentons and hexons join in any capsid. These results suggest that the exterior surface of the three capsids have the same proteins. Therefore, proteins that control the size of P1 capsid, such as DarA or DarB, must be inside the capsid.

Table 3.1 A summary of the amount of data used and final resolutions in P1 phage reconstruction.

Phage	Micrographs scanned	Particles selected	Particles in final map	Resolution (Å)
P1b	27* 8	7	141	27
P1s	23	1780 21	62 15	
P1m	15 9	8	130	27

* CCD images.

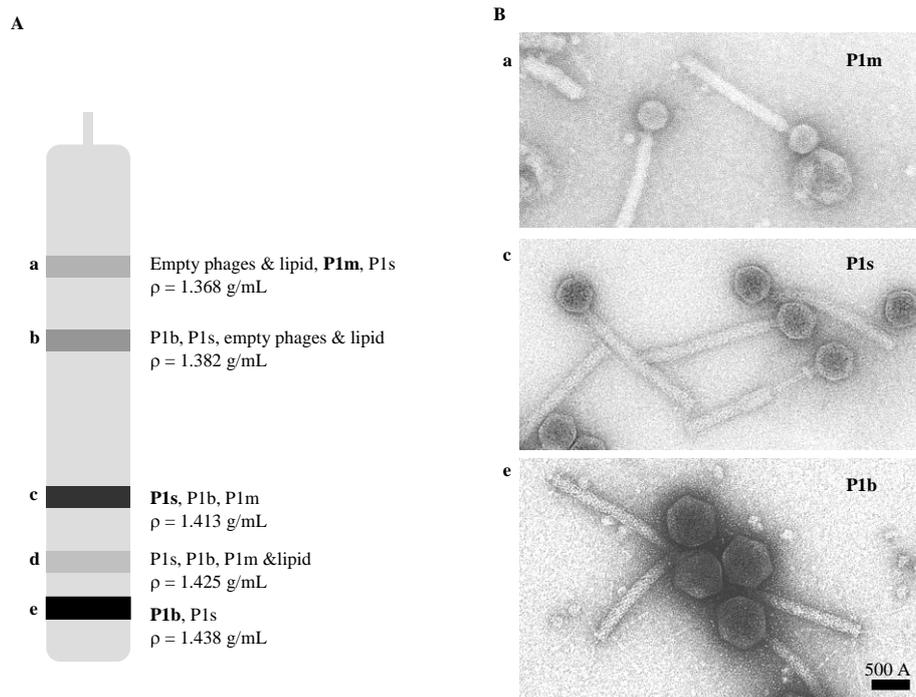


Figure 3. 2 Three P1 types isolated from CsCl gradient. A) A schematic diagram of 1.4 g/mL CsCl gradient banding of P1 vir lysate. The contents and density of each band are indicated. B) Negative stain images of bands a, c and e showing P1m, P1s and P1b.

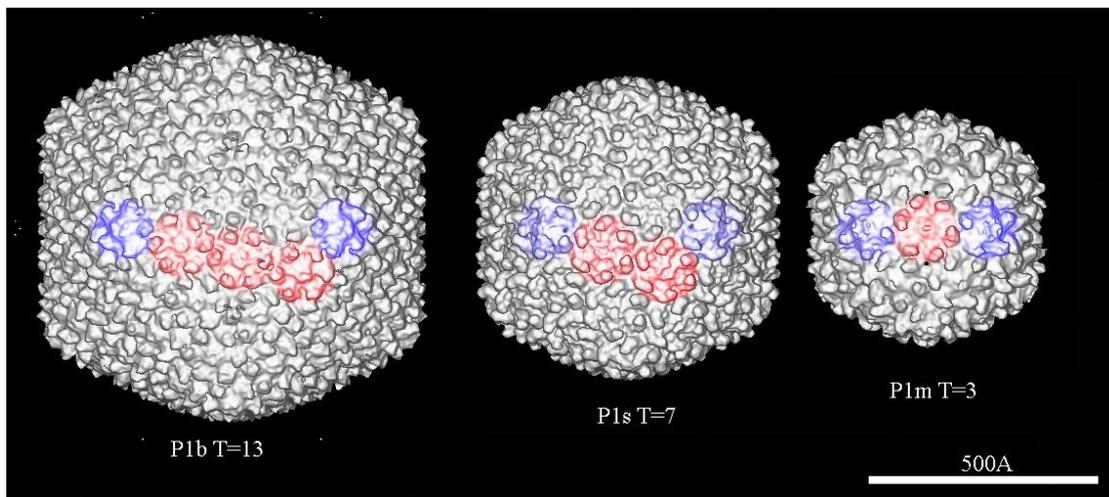


Figure 3. 3 Cryo-EM structures of P1b, P1s and P1m. Two adjacent pentons in each capsid are colored in blue and the hexons along the path between the two pentons were colored in red.

3.3.2 P1 proteins

To find the proteins that may be involved in the size determination in P1 capsid assembly, a handful of capsid proteins were identified by Edman degradation and mass spectrometry (Figure 3.4). SDS-PAGE of whole P1b phage and empty capsid were compared: ideally, any protein present at both lanes is a capsid protein; proteins only present at the whole phage lane are tail proteins or capsid proteins lost in empty capsids. Four proteins were identified to be capsid proteins, beside the MCP gp23 and its modified or cleaved forms. The copy numbers of these four proteins in each capsid were estimated based on their relative amounts in the gel compared with that of gp23, and that an empty P1b capsid should have 775 copies of gp23 (The T=13 P1b capsid is an icosahedral capsid with one penton replaced by the portal vertex, so there are $60 \times 13 - 5 = 775$ copies of gp23 in one capsid.). There are 120 copies of 60 kDa cleaved DarA, 120 copies of 20 kDa cleaved gp21, 50 copies of 76 kDa cleaved DdrB and 6 copies of 220 kDa DarB in an empty P1b capsid. Nothing is known about the functions of DdrB and gp21 (they are reported as baseplate/tail tube protein and protease in [63]). The roles of DarA and DarB during P1 capsid assembly are also not clear. However, since gp21 and DarA are present in large amounts and their cleaved forms remain in the empty capsid that has lost DNA, it is possible that they bind to the inner surface and serve as scaffolding proteins in the procapsid. A model of P1 capsid size control is discussed in section 3.4.

Edman degradation of the ladder region between the gp23 main band and DdrB indicated that the ladder bands have the same N-terminal ends as the cleaved gp23, so the ladder proteins are modified gp23. The gp23 sequence has 6 Cys residues and the cleaved gp23 has 4 Cys residues, which is unusually high compared to the MCPs of other phages such as HK97 (0 Cys) or T4 (1

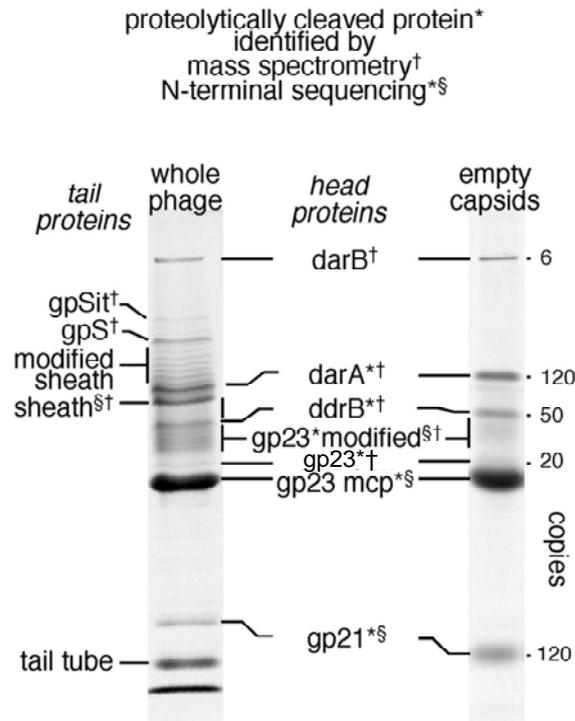


Figure 3. 4 Identification of P1 proteins by Edman degradation and mass spectry. Image courtesy of Dr. Robert Duda, University of Pittsburgh.

Cys). P1 phage proteins and empty capsid proteins were treated with dithiothreitol (DTT) and iodoacetamide (IIA) to test if the ladder is related to Cys modification (Figure 3.5). The capsid protein sample treated with 100mM DTT does not show any visible difference in the ladder region. However, both the P1b phage protein and the P1b capsid protein alkylated by iodoacetamide (IIA) show reduced intensity in the ladder region. This result suggests that the ladder may result from a type of Cys modification different from disulfide bonding. The observation that there are four evenly spaced bands in the gp23 ladder in the whole phage lane in

Figure 3.4 is consistent with the conclusion that the ladder is gp23 with 1 to 4 Cys residues modified by the same mechanism.

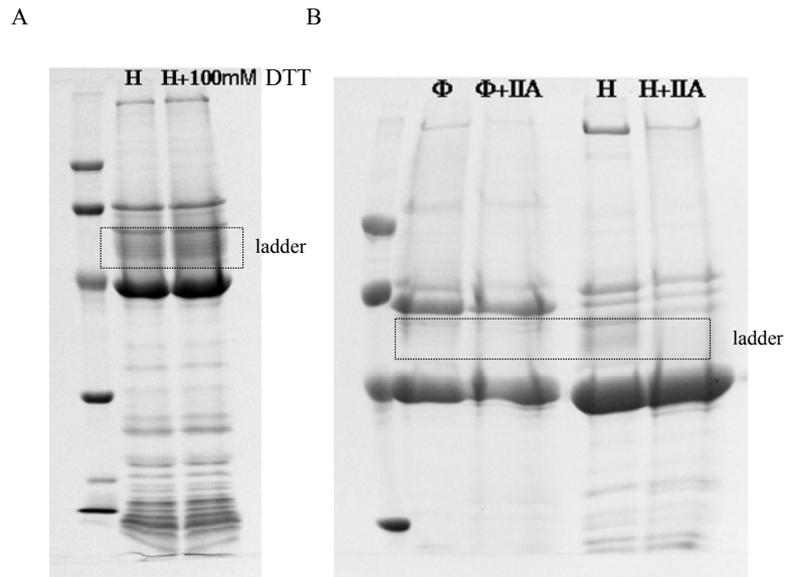


Figure 3.5 Cys modification of phage protein (Φ) and empty capsids (H) by 100 mM DTT and IIA. A) No change in ladder with DTT treatment. B) Reduced intensity of the ladder with IIA treatment.

Three attempts were made to identify the type of modification. Unfortunately, none succeeded in giving any detail about the modification mechanism due to the very low amount of target proteins and the high background of the sample. In collaboration with Dr Lyuben Marekov at the NIH, specialized approaches to reduce background and increase sensitivity in MS were used, including 1) ProteasMax (Promega V2071) treatment which enhances protein solubility and

promotes protease in-gel digestion; 2) O-methyl isourea which converts C-terminal lysine residues to homoarginine to increase MS sensitivity; 3) HPLC separation of the ladder region from the gel to remove background. In one of these MS experiments, a C-terminal fragment with the fourth Cys residue was identified, suggesting that this Cys residue may not be modified. However, this result does not rule out the possibility that the number and locations of modified Cys vary among different gp23 peptides, which is likely according to the gel.

3.4 SUMMARY AND DISCUSSION

In the study of the P1 capsid, the cryo-EM structures of three P1 capsids with different sizes were determined, 4 capsid proteins were identified and their stoichiometry estimated, and evidence was obtained suggesting that the unusual ladder in P1 protein gel is the result of Cys modification. Comparison of the surface structure of the three capsids suggests that there may be only gp23 on the surface, and thus the protein that controls capsid size during P1 capsid assembly may be inside the capsid. Studies on P1 capsid protein show that the P1 virion has at least 4 internal proteins, and 2 of them, DarA and gp21, are present in large quantities. We propose that DarA functions as a scaffolding protein which assists the ‘correct’ assembly of gp23 into viable P1b virions (Figure 3.6). P1 gp23 has an N-terminal 120 amino acid “delta” domain which may also have some scaffolding function. However, P1 may not have a specialized scaffolding protein to assist gp23 assembly. Instead, the dispensable protein DarA would act as a semi-scaffolding protein that interacts with the delta domain of gp23 to facilitate assembly of the viable P1b capsid. When DarA is knocked out, gp23 can still assemble into P1b capsids, but more gp23 fails to assemble correctly, resulting in increased production of P1s and P1m capsids.

This multi-capsid assembly pathway is quite different from that of HK97 and related phage which produce a single capsid geometry. It may be an indication of the evolutionary history of P1 capsid size growth that could have stepped from a T=4 capsid, through a T=7 intermediate, to the current T=13 size thereby allowing incremental increases in the genome.

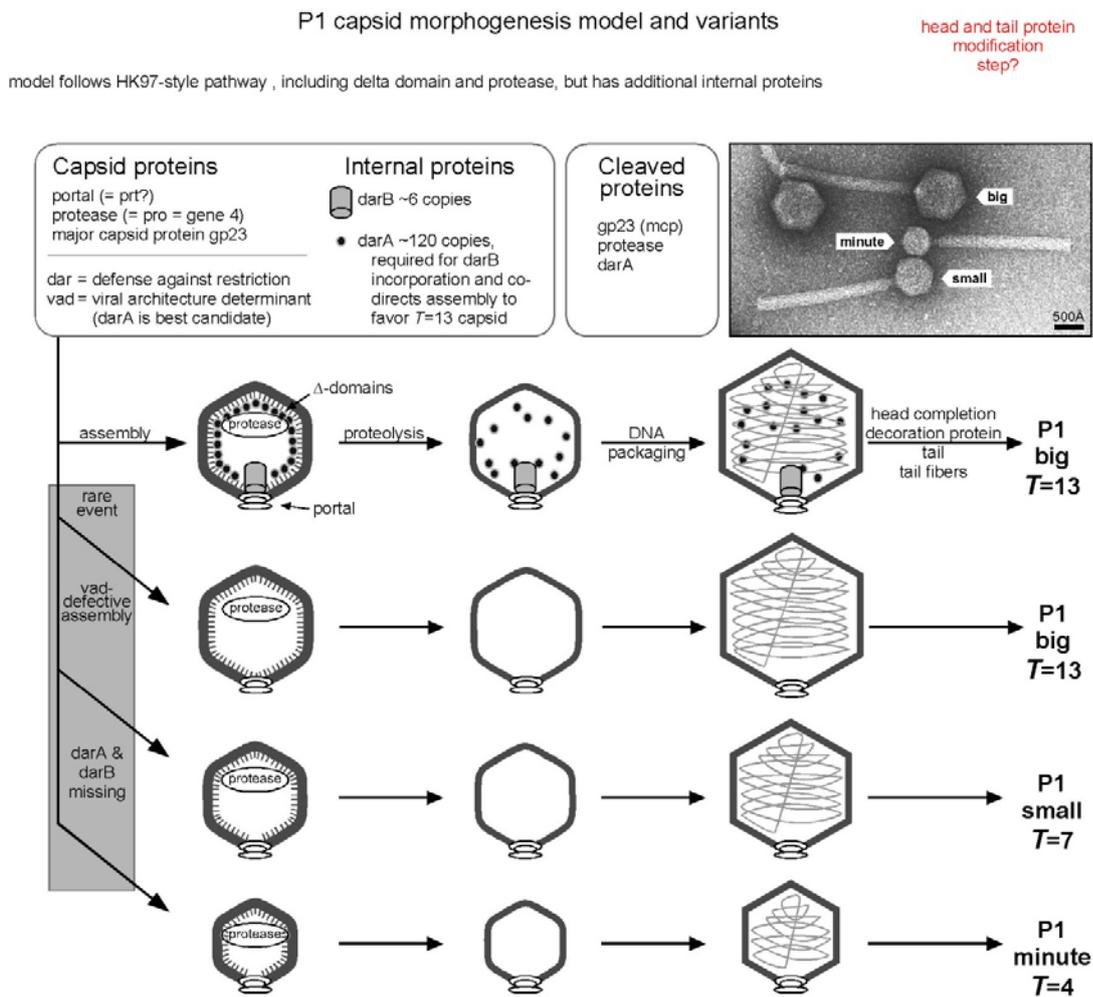


Figure 3.6 A mechanism of P1 capsid size control proposed by Dr. Robert Duda, University of Pittsburgh. P1 major capsid protein has a “delta” domain about 120 amino acid residues. DarA interacts with delta domain, to produce P1b capsids. After proteolysis, the delta domains and a 9 kb fragment of DarA are cleaved. The cleaved DarA remains as an internal protein. When DarA is not expressed, alternative prohead assembly produces a large amount of P1s and P1m.

CHAPTER 4 JUMBO PHAGES

4.1 INTRODUCTION

The structural diversities in the bacteriophage world are far from being fully explored, including the diversity in capsid size. Most intensively studied phages are small or medium sized ones with a triangulation number less than 16. Below is a list of possible triangulation numbers predicted by the Caspar-Klug theory, with the T-numbers that have been found in previous studies of phages shadowed:

1 3 4 7 9 12 13 16 19 21 25 27 28 31 36 37 39 43 48 49 52 57 61 63 64 67 73 ...

Obviously, phages with T-numbers larger than 16 have rarely been discovered, and those sizes indicated above are represented by examples: by the T=27 phage phiKZ [72] and T=52 phage G (James Conway and Roger Hendrix, unpublished data). This observation raises the interesting question whether the “missing” T numbers are simply waiting to be found or whether some of them may be in some sense forbidden.

Phages with genomes larger than 200kbp are designated as jumbo phages [73]. Interest in jumbo phages is increasing, particularly in the genome organization and capsid evolution from small phages to jumbo ones, and the large number of genes with unknown functions in jumbo phage genomes [73]. A few jumbo phages have been sequenced by the Pittsburgh Bacteriophage

Institute. This chapter shows preliminary structural studies on four jumbo phages: a *Sinorhizobium meliloti* phage N3 [74], a *Sphingomonas paucimobilis* phage PAU [75], a *Bacillus subtilis* phage PBS1 [76] and an *Escherichia coli* phage 121Q [77].

4.2 MATERIAL AND METHODS

Strains

1) PAU, PBS1 and 121Q

Either Hendrix Lab or Genome Center stocks, University of Pittsburgh, Department of Biological Sciences, Pittsburgh, PA.

2) N3

Courtesy of Dr. Valerie Oke, University of Pittsburgh, Pittsburgh, PA.

3) RM1021

RM1021 is a *Sinorhizobium meliloti* strain with streptomycin resistant. Courtesy of Dr. Valerie Oke, University of Pittsburgh, Pittsburgh, PA.

Buffers

1) Dilution buffer

10 mM Tris-HCl pH 7.5, 10 mM MgSO₄ in ddH₂O

Methods:

1) Purification of N3 phage

Twenty RM1021 colonies were completely resuspended into 5 mL LB with 250 ug/mL streptomycin and incubated at 30°C on a 250 rpm shaker overnight. The second day, a mixture of 5 mL warm soft agar, 5 mL warm LB, 1×10^4 N3 phages and 300 μ l RM1021 overnight culture supplemented with 5 mM MgSO₄, 2.5 mM CaCl₂ and 250 ug/mL streptomycin was poured onto each D×H 150 mm × 25 mm LB plate. The plates were incubated at 30°C overnight. On the third day, 10 mL dilution buffer was added on top of the soft agar. The plates were incubated at 30°C for two days. On the fifth day, N3 phages were extracted from the soft agar plates, and concentrated by centrifugation in a Ti45 rotor at 4°C and 35k rpm for 1 hour. Phages were resuspended from the pellet and purified by 10% - 45% sucrose gradient. Phages in the sucrose gradient band were concentrated again by centrifugation in a Ti45 rotor at 4°C 35k rpm for 1 hour. N3 phages resuspended from the pellet were then prepared for cryo-EM. Note that N3 phages are very sensitive to osmotic shock and PEG precipitation.

4.3 RESULTS

The cryo-EM structures of four jumbo phages were determined as shown in Figure 4.1, based on the data listed in table 4.1. Cryo-EM imaging was done by Professor James Conway and I performed the reconstructions. N3 is the first T=19 bacteriophage found. The N3 structure (13Å resolution) shows that the capsid has a finely detailed outer surface and decoration proteins

appear on the pentons. PAU is the first T=25 virus that exactly follows the quasi-equivalence rule. (The capsids of Adenovirus and bacteriophage PRD1 are a pseudo-T=25, in which the hexon is made by 3 copies of the MCP and the penton is a different protein.) The PAU structure (26Å) shows ‘arms’ on hexons adjacent to pentons, and extra density above centers of other hexons. PBS1 (24Å) is T=27, the second phage with this triangulation number. PBS1 capsid is similar in size and surface features to the first T=27 phage phiKZ [72]. There are also significant similarities in the genomes of the two phages when the genome sequences were aligned (Robert Duda, University of Pittsburgh, personal communication). 121Q (20Å) is T=28, the first viral capsid with this triangulation number. The characteristics of jumbo phage structures, T numbers, diameters, interior volume and possible decoration proteins are summarized in Table 4.2.

Table 4.1 A summary of the amount of data used and final resolutions in jumbo phage reconstruction.

Phage	Micrographs scanned	Particles selected	Particles in final map	Resolution (Å)
N3	50 1	0014	6484	13
PAU	17	806 76	5 26	
PBS1	19	687 57	5 24	
121Q	50	5238 39	16 20	

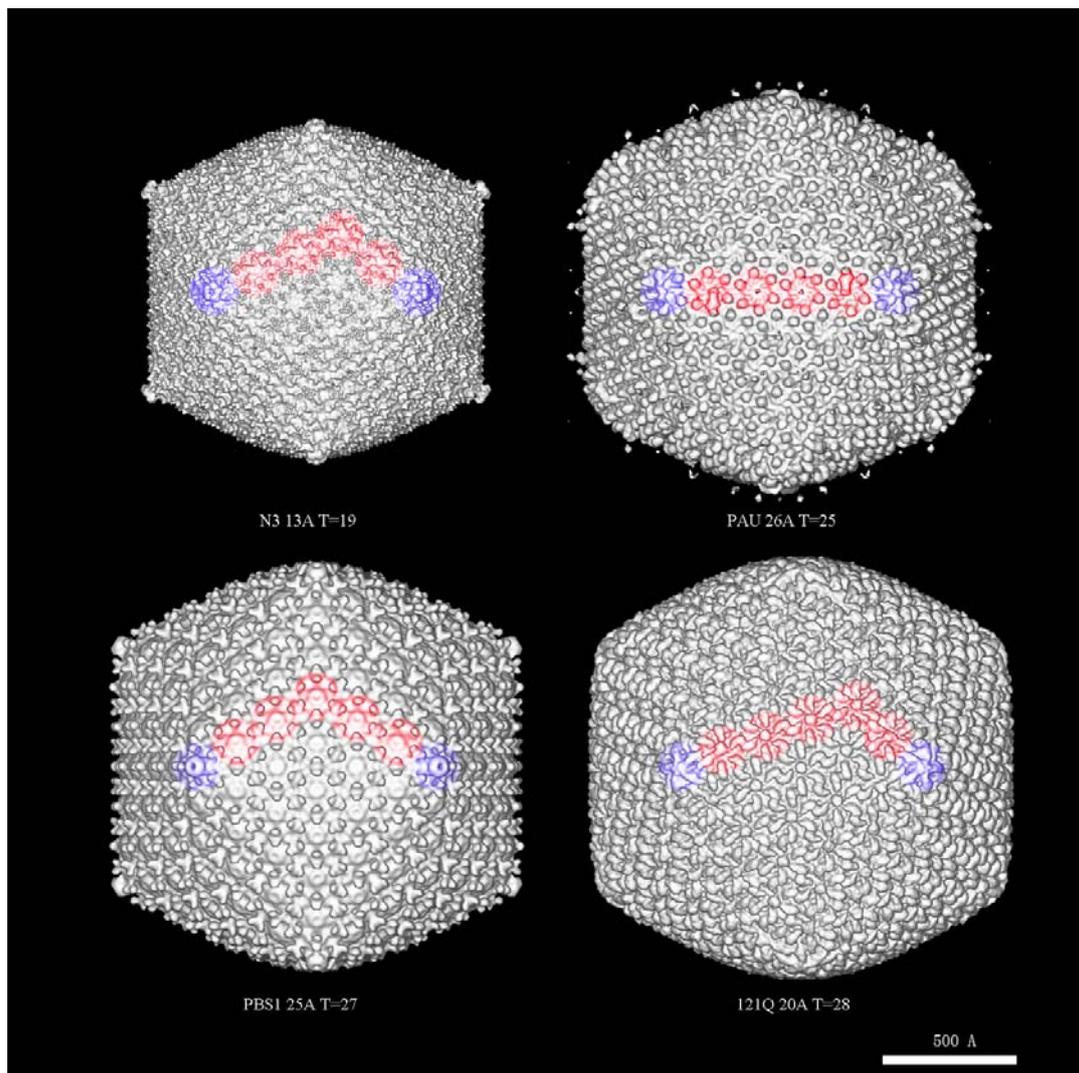


Figure 4. 1 Jumbo phage structures. N3: 13Å, T=19; PAU: 26Å, T=25; PBS1 25Å, T=27; and 121Q: 20Å, T=28. Two adjacent pentons in each capsid are colored in blue and the hexons along the path between the two pentons were colored in red.

4.4 SUMMARY AND DISCUSSION

In this preliminary study of four jumbo phages, 3D cryo-EM structures of these phages were obtained. Three of them have unique T-numbers, suggesting that T-numbers larger than 16 may generally be allowed. Other T-numbers may be found when more jumbo phages are studied. This result is consistent with the headful capsid evolution theory, which suggests that all T-numbers should be possible in principle because a larger virus is a descendant of a virus with a smaller capsid but close in size of capsid.

Table 4.2 Summary of the characteristics of jumbo phage structures: T numbers, diameters, space and possible decoration proteins.

Phage	T	D _{max} (Å)*	D _{in} (Å)**	Volume (Å ³)***	Possible decoration proteins
PBS1	27 (2 nd)	1403.1	305.7	.06x10 ⁸	None. None for phiKZ.
PAU	25 (1 st)	1311.1	222.5	.48x10 ⁸	Arms around pentons; extra density above hexons.
N3	19 (1 st)	1185.1	053.3	.70x10 ⁸	Above pentons.
121Q	28 (1 st)	1408.1	310.7	.12x10 ⁸	Not observed.

* D_{max}: the distance between the exterior surfaces of two pentons along a 5-fold axis.

** D_{in}: the distance between the interior surfaces of two pentons along a 5-fold axis.

*** The interior volume is estimated according to the following formula:

$$D_{in} = 2r_u;$$

$$r_u = \frac{a}{2} \sqrt{\varphi \sqrt{5}} = \frac{a}{4} \sqrt{10 + 2\sqrt{5}} \approx 0.9510565163 \cdot a;$$

$$V = \frac{5}{12} (3 + \sqrt{5}) a^3 \approx 2.18169499 a^3.$$

As well as discovering 3 phages with capsids having unique T-numbers, this work allows the capsid size and genome size to be correlated. Future work will compare the capsid volume with the size of chromosome packaged, and the unique genomic length of the viral DNA. In the case of PAU, evidence from pulsed-field gel electrophoresis of DNA extracted from virions is consistent with the possibility that the packaged DNA is more than twice as long as the genome sequence, which may indicate that the capsid has recently undergone an expansion in T-number and size. However, the genome has not yet exploited the additional volume by modifying or duplicating genes, or exchanging with host genes (K. Payne and R. Hendrix, unpublished). The work reported here lays the groundwork for future studies on the mechanisms of size determination in viral capsids as well as studies on the relationships among capsid size, genome size and the evolution of the genome.

FINAL SUMMARY

The shared folds of the major capsid proteins in bacteriophages and eukaryotic viruses suggest that viruses have common ancestries [23, 25]. Therefore, the various icosahedral viruses today must be evolved from a few virus prototypes. Yet little is known about the mechanisms how virus capsids have evolved and how the size of virus capsids is determined. I present capsid structures of large bacteriophages, which extend our knowledge about the regulation of capsid sizes.

The asymmetric protein gp29.2 weakly binds to the expanded procapsid of SPO1 3.2am1, and thus gp29.2 may not be essential for the initial assembly of SPO1 procapsid. This result suggests the evolutionary connection between SPO1 and herpesvirus, which is proposed based on the assumption that the asymmetric proteins of the two capsids have conserved function, is probably questionable. However, it is also possible that gp29.2 is released from the procapsid after the unexpected expansion. The SPO1 procapsid is needed to future test of this possibility.

The cryo-EM structures of three different P1 capsids have been solved. The surface features of the three capsids show that no decoration proteins bind to the outer surface, suggesting that the DarA protein which regulates the P1 capsid size is inside the capsid and functions as a semi-

scaffolding protein. This may be tested in future if P1 procapsids or empty capsids of *darA* mutant are isolated.

Preliminary structural studies on four jumbo phages showed their T numbers, and three T numbers are the first found in viral capsids. Therefore, large T numbers may not be restricted for viral capsids. The results provide valuable information for future work on viral capsid evolution, regulation of capsid sizes, and the relationship between the capsid size and the genome size.

BIBLIOGRAPHY

1. Twort, F.W., *An investigation on the nature of ultra-microscopic viruses*. *Lancet*, 1915. **II**: p. 1241.
2. d'Herelle, F., *Sur un microbe invisible antagoniste des bacilles dysentériques*. *Compt Rend Acad Sci*, 1917. **165**: p. 373.
3. Kutter, E., Sulakvelidze, A., *Bacteriophages: biology and applications*. CRC Press: Florida, 2004.
4. Wommack, K.E. and R.R. Colwell, *Virioplankton: viruses in aquatic ecosystems*. *Microbiol Mol Biol Rev*, 2000. **64**(1): p. 69-114.
5. Valegard, K., et al., *The three-dimensional structure of the bacterial virus MS2*. *Nature*, 1990. **345**(6270): p. 36-41.
6. Neitzey, L.M., M.S. Hutson, and G. Holzwarth, *Two-dimensional motion of DNA bands during 120 degrees pulsed-field gel electrophoresis*. *Electrophoresis*, 1993. **14**(4): p. 296-303.
7. Ackermann, H.W., *Frequency of morphological phage descriptions in the year 2000. Brief review*. *Arch Virol*, 2001. **146**(5): p. 843-57.
8. Casjens, S. and R. Hendrix, *dsDNA Phage Assembly*, in *The Bacteriophages*, R. Calendar, Editor. 1988, Plenum Press: New York.
9. Friedman, D.I. and D.L. Court, *Bacteriophage lambda: alive and well and still doing its thing*. *Curr Opin Microbiol*, 2001. **4**(2): p. 201-7.
10. Letellier, L., et al., *Main features on tailed phage, host recognition and DNA uptake*. *Front Biosci*, 2004. **9**: p. 1228-339.
11. Crick, F.H. and J.D. Watson, *Structure of small viruses*. *Nature*, 1956. **177**(4506): p. 473-5.
12. Caspar, D.L. and A. Klug, *Physical principles in the construction of regular viruses*. *Cold Spring Harb Symp Quant Biol*, 1962. **27**: p. 1-24.
13. Baker, T.S., N.H. Olson, and S.D. Fuller, *Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs*. *Microbiol Mol Biol Rev*, 1999. **63**(4): p. 862-922.
14. Duda, R.L., K. Martinic, and R.W. Hendrix, *Genetic basis of bacteriophage HK97 prohead assembly*. *J Mol Biol*, 1995. **247**(4): p. 636-47.

15. Wikoff, W.R., et al., *Topologically linked protein rings in the bacteriophage HK97 capsid*. Science, 2000. **289**(5487): p. 2129-33.
16. Driedonks, R.A., *The quaternary structure of the T4 gene product 20 oligomer*. Prog Clin Biol Res, 1981. **64**: p. 315-23.
17. Bazinet, C. and J. King, *The DNA translocating vertex of dsDNA bacteriophage*. Annu Rev Microbiol, 1985. **39**: p. 109-29.
18. Kato, H. and C. Baschong, *Isolation of a gp20-complex and its role in in vitro assembly of both prohead and core of bacteriophage T4*. Virology, 1997. **227**(2): p. 400-8.
19. Lander, G. C., et al., *Bacteriophage lambda stabilization by auxiliary protein gpD: timing, location, and mechanism of attachment determined by cryo-EM*. Structure, 2008. **16**(9): p. 1399-406.
20. Kurtz, M.B. and S.P. Champe, *Precursors of the T4 internal peptides*. J Virol, 1977. **22**(2): p. 412-9.
21. Bachrach, U. and L. Benchetrit, *Studies on phage internal proteins. 3. Specific binding of T4 internal proteins to T4 DNA*. Virology, 1974. **59**(2): p. 443-54.
22. Casjens, S. and J. King, *Virus assembly*. Annu Rev Biochem, 1975. **44**: p. 555-611.
23. Hendrix, R.W., *Evolution: the long evolutionary reach of viruses*. Curr Biol, 1999. **9**(24): p. R914-7.
24. Fauquet, C.M. and D. Fargette, *International Committee on Taxonomy of Viruses and the 3,142 unassigned species*. Virol J, 2005. **2**: p. 64.
25. Bamford, D.H., J.M. Grimes, and D.I. Stuart, *What does structure tell us about virus evolution?* Curr Opin Struct Biol, 2005. **15**(6): p. 655-63.
26. Benson, S. D., et al., *Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures*. Cell, 1999. **98**(6): p. 825-33.
27. Roberts, M.M., et al., *Three-dimensional structure of the adenovirus major coat protein hexon*. Science, 1986. **232**(4754): p. 1148-51.
28. Krupovic, M. and D.H. Bamford, *Virus evolution: how far does the double beta-barrel viral lineage extend?* Nat Rev Microbiol, 2008. **6**(12): p. 941-8.
29. Baker, M.L., et al., *Common ancestry of herpesviruses and tailed DNA bacteriophages*. J Virol, 2005. **79**(23): p. 14967-70.
30. Fokine, A., et al., *Structural and functional similarities between the capsid proteins of bacteriophages T4 and HK97 point to a common ancestry*. Proc Natl Acad Sci U S A, 2005. **102**(20): p. 7163-8.
31. Zhou, Z.H., et al., *Seeing the herpesvirus capsid at 8.5 Å*. Science, 2000. **288**(5467): p. 877-80.
32. Adrian, M., et al., *Cryo-electron microscopy of viruses*. Nature, 1984. **308**(5954): p. 32-6.
33. Crowther, R.A., *Procedures for three-dimensional reconstruction of spherical viruses by Fourier synthesis from electron micrographs*. Philos Trans R Soc Lond B Biol Sci, 1971. **261**(837): p. 221-30.
34. Crowther, R.A., et al., *Three dimensional reconstructions of spherical viruses by fourier synthesis from electron micrographs*. Nature, 1970. **226**(5244): p. 421-5.
35. Conway, J.F., et al., *Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy*. Nature, 1997. **386**(6620): p. 91-4.
36. Zhang, X., et al., *Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction*. Proc Natl Acad Sci U S A, 2008. **105**(6): p. 1867-72.

37. Jiang, W., et al., *Backbone structure of the infectious epsilon15 virus capsid revealed by electron cryomicroscopy*. Nature, 2008. **451**(7182): p. 1130-4.
38. Yu, X., L. Jin, and Z.H. Zhou, *3.88 Å structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy*. Nature, 2008. **453**(7193): p. 415-9.
39. Lepault, J., F.P. Booy, and J. Dubochet, *Electron microscopy of frozen biological suspensions*. J Microsc, 1983. **129**(Pt 1): p. 89-102.
40. McDowell, A.W., et al., *Electron microscopy of frozen hydrated sections of vitreous ice and vitrified biological samples*. J Microsc, 1983. **131**(Pt 1): p. 1-9.
41. Okubo, S., B. Strauss, and M. Stodolsky, *The Possible Role of Recombination in the Infection of Competent Bacillus Subtilis by Bacteriophage Deoxyribonucleic Acid*. Virology, 1964. **24**: p. 552-62.
42. Hoet, P.P., M. M. Coene, and C.G. Cocito, *Replication cycle of Bacillus subtilis hydroxymethyluracil-containing phages*. Annu Rev Microbiol, 1992. **46**: p. 95-116.
43. Lawrie, J.M., J.S. Downard, and H.R. Whiteley, *Bacillus subtilis bacteriophages SP82, SPO1, and phiε: a comparison of DNAs and of peptides synthesized during infection*. J Virol, 1978. **27**(3): p. 725-37.
44. Losick, R. and J. Pero, *Cascades of Sigma factors*. Cell, 1981. **25**(3): p. 582-4.
45. Stewart, C.R., et al., *The genome of Bacillus subtilis bacteriophage SPO1*. J Mol Biol, 2009. **388**(1): p. 48-70.
46. Parker, M.L. and F.A. Eiserling, *Bacteriophage SPO1 structure and morphogenesis. I. Tail structure and length regulation*. J Virol, 1983. **46**(1): p. 239-49.
47. Parker, M. L., E.J. Ralston, and F.A. Eiserling, *Bacteriophage SPO1 structure and morphogenesis. II. Head structure and DNA size*. J Virol, 1983. **46**(1): p. 250-9.
48. Duda, R.L., et al., *Shared architecture of bacteriophage SPO1 and herpesvirus capsids*. Curr Biol, 2006. **16**(1): p. R11-3.
49. Parker, M.L. and F.A. Eiserling, *Bacteriophage SPO1 structure and morphogenesis. III. SPO1 proteins and synthesis*. J Virol, 1983. **46**(1): p. 260-9.
50. Trus, B.L., et al., *The herpes simplex virus procapsid: structure, conformational changes upon maturation, and roles of the triplex proteins VP19c and VP23 in assembly*. J Mol Biol, 1996. **263**(3): p. 447-62.
51. Newcomb, W.W., et al., *Structure of the herpes simplex virus capsid. Molecular composition of the pentons and the triplexes*. J Mol Biol, 1993. **232**(2): p. 499-511.
52. Zahler, S.A., L. G. Benjamin, B. S. Glatz, P. F. Winter, and B. J. Goldstein, *Genetic mapping of alsA, alsR, thyA, kauA, and citD markers in Bacillus subtilis*, in *Microbiology*, D. Schlessinger and A.S.f. Microbiology, Editors. 1974, American Society for Microbiology: Washington, p. 35-43.
53. Sullivan, M.A., R.E. Yasbin, and F.E. Young, *New shuttle vectors for Bacillus subtilis and Escherichia coli which allow rapid detection of inserted fragments*. Gene, 1984. **29**(1-2): p. 21-6.
54. Rubin, G.M., *The nucleotide sequence of Saccharomyces cerevisiae 5.8 S ribosomal ribonucleic acid*. J Biol Chem, 1973. **248**(11): p. 3860-75.
55. Adereth, Y., et al., *Site-directed mutagenesis using Pfu DNA polymerase and T4 DNA ligase*. Biotechniques, 2005. **38**(6): p. 864, 866, 868.
56. Okubo, S., et al., *The genetics of bacteriophage SPO1*. Biken J, 1972. **15**(2): p. 81-97.

57. Conway, J.F. and A.C. Steven, *Methods for reconstructing density maps of "single" particles from cryoelectron micrographs to subnanometer resolution*. J Struct Biol, 1999. **128**(1): p. 106-18.
58. Heymann, J.B., *Bsoft: image and molecular processing in electron microscopy*. J Struct Biol, 2001. **133**(2-3): p. 156-69.
59. Heymann, J.B. and D.M. Belnap, *Bsoft: image processing and molecular modeling for electron microscopy*. J Struct Biol, 2007. **157**(1): p. 3-18.
60. Yan, X., R.S. Sinkovits, and T.S. Baker, *AUTO3DEM--an automated and high throughput program for image reconstruction of icosahedral particles*. J Struct Biol, 2007. **157**(1): p. 73-82.
61. Ross, P.D., et al., *A free energy cascade with locks drives assembly and maturation of bacteriophage HK97 capsid*. J Mol Biol, 2006. **364**(3): p. 512-25.
62. Kocsis, E., et al., *Multiple conformational states of the bacteriophage T4 capsid surface lattice induced when expansion occurs without prior cleavage*. J Struct Biol, 1997. **118**(1): p. 73-82.
63. Lobocka, M.B., et al., *Genome of bacteriophage P1*. J Bacteriol, 2004. **186**(21): p. 7032-68.
64. Heinrich, J., M. Velleman, and H. Schuster, *The tripartite immunity system of phages P1 and P7*. FEMS Microbiol Rev, 1995. **17**(1-2): p. 121-6.
65. Walker, D.H., Jr. and T.F. Anderson, *Morphological variants of coliphage P1*. J Virol, 1970. **5**(6): p. 765-82.
66. Walker, J.T. and D.H. Walker, Jr., *Coliphage P1 morphogenesis: analysis of mutants by electron microscopy*. J Virol, 1983. **45**(3): p. 1118-39.
67. Iida, S., et al., *Accessory genes in the darA operon of bacteriophage P1 affect antirestriction function, generalized transduction, head morphogenesis, and host cell lysis*. Virology, 1998. **251**(1): p. 49-58.
68. Walker, J.T. and D.H. Walker, Jr., *Structural proteins of coliphage P1*. Prog Clin Biol Res, 1981. **64**: p. 69-77.
69. Iida, S., J. Meyer, and W. Arber, *The insertion element IS1 is a natural constituent of coliphage P1 DNA*. Plasmid, 1978. **1**(3): p. 357-65.
70. Iida, S., et al., *Two DNA antirestriction systems of bacteriophage P1, darA, and darB: characterization of darA- phages*. Virology, 1987. **157**(1): p. 156-66.
71. Rosenfeld, J., et al., *In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis*. Anal Biochem, 1992. **203**(1): p. 173-9.
72. Fokine, A., et al., *A three-dimensional cryo-electron microscopy structure of the bacteriophage phiKZ head*. J Mol Biol, 2005. **352**(1): p. 117-24.
73. Hendrix, R.W., *Jumbo bacteriophages*. Curr Top Microbiol Immunol, 2009. **328**: p. 229-40.
74. Lesley, S.M., *A bacteriophage typing system for Rhizobium meliloti*. Can. J. Microbiol., 1982. **28**: p. 180-9.
75. Ackermann, H.W., et al., *Bacteriophages from Bombyx mori*. Arch Virol, 1994. **137**(1-2): p. 185-90.
76. Eisnerling, F.A., *Bacillus subtilis bacteriophages: structure, intracellular development and conditions of lysogeny*. Ph.D thesis, University of California, Los Angeles., 1963: p. 1-127.

77. Ackermann, H.W. and T.M. Nguyen, *Sewage coliphages studied by electron microscopy*. Appl Environ Microbiol, 1983. **45**(3): p. 1049-59.