ABSTRACT

Title of Document: SCAFFOLDING-MEDIATED VIRUS ASSEMBLY: VISUALIZATION AND CHARACTERIZATION OF BACTERIOPHAGE T7 SCAFFOLDING PROTEIN

Charles Stewart Smith, Doctor of Philosophy, 2009

Directed By: Professor Alasdair C. Steven
Laboratory of Structural Biology Research,
National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health
Professor Dorothy Beckett
Chemistry and Biochemistry

In bacteriophage T7 as well as many other dsDNA phages and some animal viruses, scaffolding protein is essential for the accurate formation of the metastable precursor particle (prohead) and thus it is a vital aspect of the viral infection cycle. When purified by anion exchange and gel filtration chromatography, the T7 scaffolding protein (gp9) exists as an extended monomer in solution. These monomers of gp9 associate with other scaffolding monomers to form the extended filaments in T7 procapsids visualized by electron tomography. These filaments interact with the negatively charged inner surface of the procapsid at unique sites, probably via the extended positively charged C-terminus of gp9. Scaffolding protein, via these interactions, facilitates isometric capsid assembly by helping to define the proper

curvature of the viral capsid. Consequently, scaffolding-mediated viral assembly does not require a rigid structural network, nor does it require a stoichiometric amount of gp9 to be present in each procapsid. The observed flexibility of the scaffolding network, association of scaffolding filaments with the core connector complex, as well as the variable copy number of gp9 per particle suggests an assembly mechanism where capsid formation is nucleated around the core/connector complex. In such a mechanism there is more than one path to a single end – production of the correctly formed prohead particles that are required for T7 phage maturation.

SCAFFOLDING MEDIATED VIRUS ASSEMBLY: VISUALIZATION AND CHARACTERIZATION OF BACTERIOPHAGE T7 SCAFFOLDING PROTEIN

By

Charles Stewart Smith

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2009

Advisory Committee:
Professor Dorothy Beckett, Chair
Professor Alasdair C. Steven, Co-Chair
Professor George H. Lorimer
Professor Roger W. Hendrix
Professor Stephen M. Mount

Dedication

For Katie . . .

I wouldn't have made it through science or life in general without your unconditional love and support.

For Mario Edward Cerritelli . . .

You have been there for me since the beginning of my scientific career as my mentor, my *counselor*, and my friend.

. . . and for Sofia

Acknowledgements

There are many people to thank for not only the completion of this work, but for my arrival at this juncture in my career . . .

- I would first like to thank my advisor Alasdair Steven for taking a chance on me as his first graduate student and for seeing this process through to completion.

- I want to thank all the members of the LSBR over the past seven years. You guys taught me everything I know about being successful in the lab . . .

- In particular, I would like to thank Giovanni and Dennis for the many late nights and long hours spent working through the tomographic data with me. Tomography in the lab would not have been possible without Dennis who collected some of my original data sets and taught me to do the same. Then it was Giovanni's patience and teaching that took that data from being interesting pictures to defendable results.

- I would also like to thank Bernard, Norman, and Ulrich who suffered more of my questions than most.

- I'd also like to thank Greg, Giovanni, and Isai for always letting me break their concentration to share a good laugh.

- I especially want to thank Santi who was my mentor for much of the work on the purified gp9, especially the crystallography. He was (and continues to be) a constant source of encouragement.

- I would like to thank the members of my committee (especially Roger Hendrix who traveled a great distance for this defense) who gave of their time to carefully read and thoughtfully critique this document.

- I want to thank my wife and daughter for still loving me even after all of the nights that they never saw me come home.
- My original scientific curiosity I owe to my Mom, and Dad. Thanks for shepherding me along and encouraging me to do what I love.
- I'm grateful for my friends who graciously endured countless episodes of pre-doctoral despair.
- I would also like to thank Stephen Butler for patiently enduring my efforts to complete this document. Stephen kept me focused on writing and received only grumbling in response.

Again, I want to thank Mario who (despite having plenty of other responsibilities) was always available, helpful, and enthusiastic.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

°C	degrees Celsius
[X]	concentration of X
Å	Angstroms
Amp	ampicillin
ATP	adenosine triphosphate
BME	β-mercaptoethanol
C-terminal	carboxy terminal
cryo-EM	cryo-electron microscopy
cryo-ET	cryo-electron tomography
c(s)	sedimentation coefficient distributions
CCD	charge coupled device
DNA	deoxyribonucleic acid
DSC	differential scanning calorimetry
dsDNA	double stranded DNA
DNase	deoxyribonuclease
DTT	dithiothreitol
E. coli	Escherichia coli
EDTA	ethylenediaminetetraacetic acid
eV	electron volt
f/fo	frictional ratio

gp	gene product
His	histidine
HK97	Hong Kong 97
hoc	highly antigenic outer capsid protien
HSV-1	Herpes Simplex Virus type 1
IPTG	isopropyl-β-D-thiogalactopyranoside
Kan	Kanamycin
kb	kilobase pairs
kDa	kilodaltons
krpm	kilorevolutions per minute
LB	Luria-Bertani broth
mL	milliliter
mM	millimolar
MW	molecular weight
Ni-NTA	nickel-nitrilotriacetic acid
N-terminal	amino terminal
nm	nanometers
NMR	nuclear magnetic resonance
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
PMSF	phenylmethanesulphonylfluoride or phenylmethylsulphonyl fluoride

PVDF	polyvinylidene Fluoride
rmsd	root mean square deviation
RNase	ribonuclease
$s_{20,w}$	sedimentation coefficient
SDS	sodium dodecyl sulfate
soc	small outer capsid protein
SPA	single particle analysis
STEM	scanning transmission electron microscopy
T number	triangulation number
TCEP	tris(2-carboxyethyl)phosphine
TEM	transmission electron microscope
TMV	tobacco mosaic virus
Tris	Tris (hydroxymethyl) aminomethane
Trp	tryptophan
μM	micromolar
μm	micrometer
v/v	volume to volume
v	partial specific volume
η	viscosity

Chapter 1: Introduction and Research Overview

1.1 Viruses as Macromolecular Machines

Large multi-protein complexes commonly referred to as macromolecular machines carry out almost every major process in a cell [1]. Programmed protein degradation is controlled by proteosomes, multi-enzyme complexes like pyruvate dehydrogenase catalyze complex reactions [2], and intron removal is carried out by the formidable spliceosome (a complex of over 150 individual proteins) [3]. Cellular interlopers like viruses have been characterized as complex nanomachines with literally hundreds of moving parts. In this regard, viruses have offered structural biologists an opportunity to study the general assembly and functional principles of macromolecular machines in great detail. The viral assembly process has been studied in hopes of better understanding macromolecular assembly principles as well as for finding weaknesses in pathogenic viruses that can be exploited by antiviral therapies.

Viruses cannot produce progeny through division or budding, but instead they must coerce the host cell into producing multiple copies of the viral genome, structural, and functional proteins. The structural proteins of the virus must assemble inside the cell to provide the vessel for conveying the viral genome to the next host. Because the viral life cycle seems to increase in complexity

proportional to the host the virus utilizes, it has proven quite difficult to directly study many of the viruses whose hosts are members of the kingdom eukarya (humans included) [4].

Fortunately many viruses share a great deal of structural similarities regardless of the kingdom of life they invade [5]. This is one reason that double stranded DNA (dsDNA) bacteriophages are an attractive model system for the study of related plant and animal viruses as well as the biosynthesis of macromolecular complexes in general. Further, many dsDNA bacteriophages are ideally suited to the lab due to their tractable genetics and short replication cycles (as little as 13 minutes from infection to lysis for bacteriophage T7). The biosynthesis pathways of many dsDNA phages such as Hong Kong 97 (HK97) [6], T4 [7], T7 [4], and P22 [8] have been studied in great detail and it has been noted that they assemble via a series of common steps with system specific variations at certain points of this generic assembly pathway.

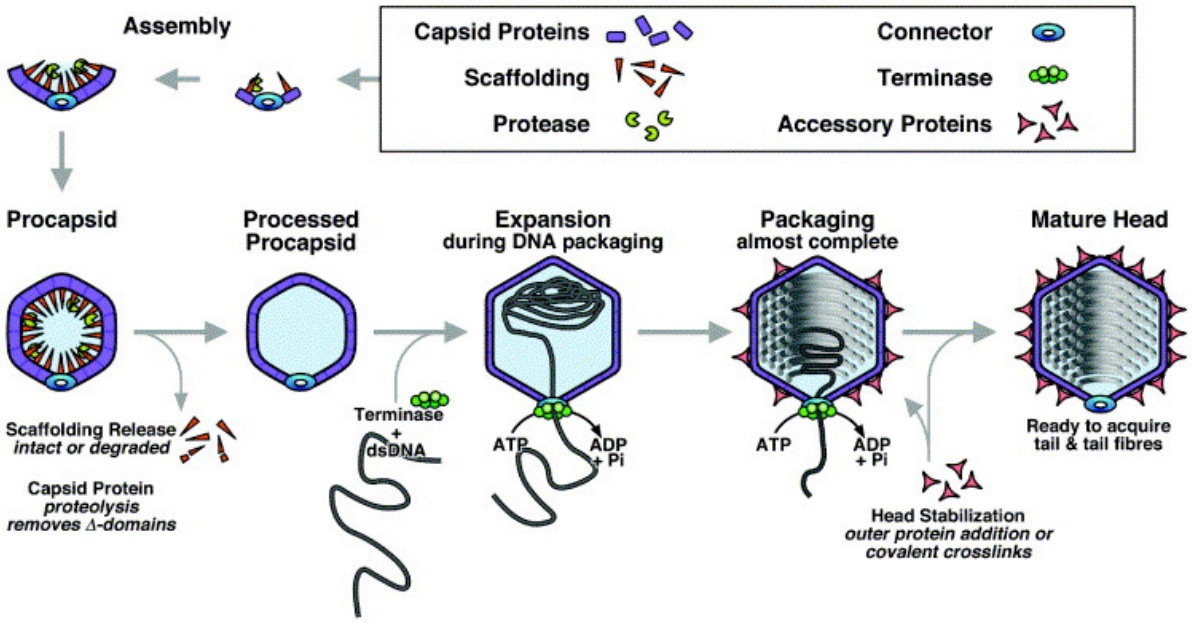
1.2 Double Stranded DNA Phage Assembly

For this work, the discussion of dsDNA phages will be limited to non-enveloped phages that have an icosahedral capsid. It is difficult to make a one-size-fits all scheme to describe phage assembly as each system has its own twist on the common theme. However, a generic assembly pathway for dsDNA phages proceeds through the following steps: At the late stages of infection, phage structural proteins are expressed in the host cell. These

structural proteins include capsid proteins, a scaffolding protein (see later), and a portal or connector protein, which are assembled into an icosahedral precursor particle called the procapsid. The procapsid particle serves as a receptacle into which the dsDNA genome will be packaged. Concomitant with DNA packaging, the procapsid undergoes a process commonly termed expansion in which the internal volume of the virion is increased on the order of 20% and the particle becomes more angular and rigid. Following this expansion, the capsid can be further stabilized by the addition of accessory proteins or by chemical cross-linking. For some dsDNA phages, this particle is then further outfitted with tail and tail fiber proteins that complete the now infectious virion. This progeny phage along with the other newly synthesized infectious phage exit the host upon lysis to encounter new hosts, thus repeating the process. A graphical representation of the generic pathway for dsDNA capsid assembly is provided in Figure 1 (taken from Steven et. al [9]).

While the pathway described above does cover many of the central themes of dsDNA assembly, it is not possible to construct an assembly pathway for which there are no exceptions. The lambdoid phage Hong Kong 97 appeared to lack a scaffolding protein all together; however later it was realized that the HK97 capsid protein gp5 has an N-terminal domain, which is largely alpha-helical [10]and has been shown to serve a scaffolding function during assembly [6]. Interestingly, this domain is cleaved prior to maturation hinting at its role as an assembly factor that is not present in the final virion [11]. At the

Figure 1 Generic assembly pathway for the heads of spherical tailed dsDNA bacteriophage. (taken from Steven *et al*., 2005)


Assembly
Capsid Proteins
Scaffolding
Protease
Connector
Terminase
Accessory Proteins
Procapsid
Processed Procapsid
Expansion during DNA packaging
Packaging almost complete
Mature Head
Scaffolding Release
intact or degraded
Capsid Protein
proteolysis
removes Δ-domains
Terminase
+
dsDNA
ATP
ADP + Pi
ATP
ADP + Pi
Head Stabilization
outer protein addition or
covalent crosslinks
Ready to acquire
tail & tail fibres

other end of the spectrum, the T-even phage T4 possesses two essential scaffolding proteins, which along with a collection of other proteins, make up a large scaffolding complex called the T4 scaffolding core [12].

In addition to inclusion of the portal protein (which has been found in all phage studied thus far), some phage synthesize and package additioxnal internal proteins that form ringed complexes at the portal vertex. In P22, these proteins are referred to as the pilot/injection proteins (gp7, gp16, and gp20), which exit the capsid ahead of the DNA during host infection [13-15]. In T7, an even more substantial complex of proteins known as the internal core (see figure 3 later) resides at the portal vertex in the interior of the procapsid [16]. Mutational studies of T7 phage lacking the core proteins suggest that they play an important role in enhancing the efficiency of assembly [17]. These proteins are also reported to function in the injection of DNA into the host cell upon infection [18].

Other differences exist among dsDNA phage assembly pathways such as the inclusion or exclusion of a maturational protease. Some phage such as HK97 and T4 initiate maturation via the activities of a prepackaged protease whereas T7 and P22 control the initiation of expansion without such an enzyme. In addition, some phage employ additional capsid stabilization measures such as covalent cross-linking of the HK97 capsid proteins upon maturation [19], or the

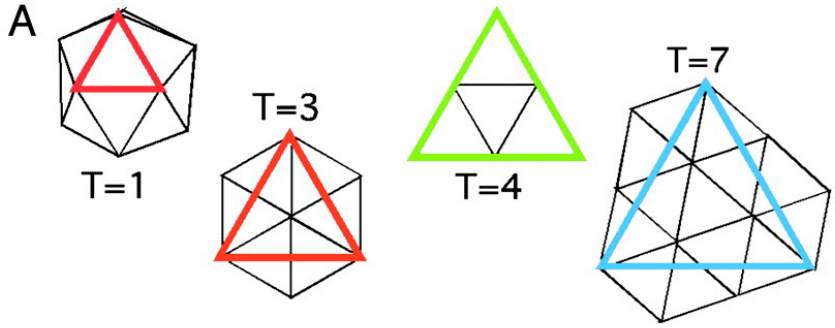
binding of accessory protein clamps between adjacent capsid protein subunit in phages T4 (hoc and soc) [20] and lambda (gpD) [21].

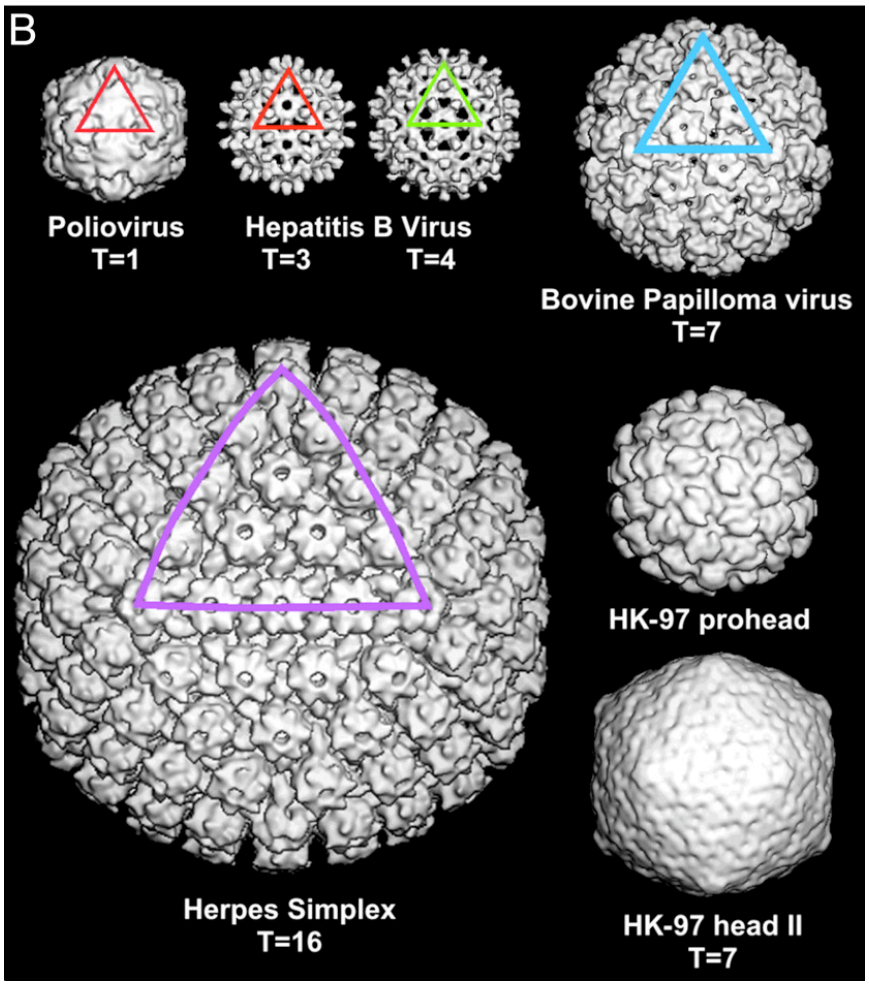
1.3 Icosahedral Geometry and Quasi-equivalence

The capsids of spherical dsDNA phage form regular icosahedral structures [22]. A regular icosahedron is a polyhedron with 20 equilateral triangular faces arranged such that there are 12 five-fold vertices and 30 edges (Figure 2). Viral capsids adopt this geometry because these highly symmetric structures can be built from multiple copies of a single building block (i.e. a capsid protein subunit) [22]. The icosahedron has a 60-fold symmetry, and therefore a virus can only build a strictly regular icosahedron from 60 equivalent subunits. Most capsid proteins are roughly 200 to 400 amino acids in length, and a capsid made of only 60 copies of such a protein would not be able to contain even the more modest sized dsDNA phage genomes. Rather than carry an additional gene for a second, third or fourth type of capsid protein to enlarge their shells, virus capsids adhere to the more forgiving quasi-equivalent geometry first detailed by Caspar and Klug [23].

Viral capsid proteins form icosahedral capsids from more than sixty copies of the same protein by utilizing a single protein that can adopt multiple conformations in the assembled capsid lattice [24]. This means that each copy of capsid protein is capable of forming hexamers and pentamers. The number

Figure 2 Icosahedral symmetry and quasi-equivalence in viral capsids. Panel A shows *T* = 1, T=3, T=4, and T=7 icosahedrons. This panel shows how larger tiles can be cut out of hexagonal lattice, resulting in faces where smaller triangles are arranged around local quasi-six-folds in the manner of a hexameric capsomer. (*B*) A series of cryoEM reconstructions of virus capsids with different *T* numbers. A triangular face is outlined on selected images. (taken from [25])


A
T=1
T=3
T=4
T=7


B
Poliovirus
T=1
Hepatitis B Virus
T=3
T=4
Bovine Papilloma virus
T=7
HK-97 prohead
Herpes Simplex
T=16
HK-97 head II
T=7

of capsid protein molecules used to form each asymmetric unit of the icosahedron (Figure 2), is also called the T number [23]. For example, a capsid with a T number of 4 has four capsid protein molecules in its asymmetric unit (Figure 2). The larger the T number, the larger the size of the viral capsid, and the larger the number of unique local conformations a single capsid protein must adopt in the assembled capsid in order to achieve a complete/closed geometric structure.

By employing the modular assembly strategy described above, viruses can produce a sufficiently large capsid structure from the genetic information contained on a relatively small segment of its genome. In order for such an ambitious construction process to be successful, the capsid protein building blocks must not only fit together correctly, but the forces that direct assembly must either be intrinsic to the starting material or come from an external chaperone. Just as molecular chaperones oversee the proper folding of complex protein structures, a class of molecules termed scaffolding proteins has been shown to behave as quaternary structure chaperones that guide the assembly of large multi-protein complexes like viral capsids [26].

1.4 Procapsid Assembly

In the case of spherical dsDNA bacteriophage, the capsid assembly process proceeds through an intermediate complex called the prohead or procapsid. A

fixed number of capsid proteins (several hundred copies in most cases) assemble along with a chaperone molecule called the scaffolding protein. The capsid and scaffolding protein self-assemble and during the process incorporate -- at a single vertex -- a dodecameric ring of protein referred to as the "connector" (for its role as an attachment point of the viral tail) or the "portal" (a term that describes its function as an opening through which the viral genome both enters and exits the capsid). Structurally, procapsids are generally more spherical and have thicker walls relative to the mature phage capsid structure.

1.5 Procapsid Maturation

As mentioned earlier, assembly of the viral procapsid is followed by a series of large-scale conformational changes called maturation [9]. During this process, the capsid expands, becomes more angular, and its walls become thinner and more rigid. These significant changes are a result of rearrangements of the capsid protein molecules within the icosahedral lattice that can be driven by the combined actions of a maturational protease and the incoming dsDNA genome [27]. The repulsive force of the packaged DNA exert pressure on the capsid lattice that results in rearrangement of the prohead capsomers [28]. The skewed hexamers become more symmetric, and the mature capsid is stabilized. In phages like P22 and T7, the process of phage maturation is not triggered by a maturational protease. Instead, the

scaffolding protein exits the procapsid intact concomitant with DNA packaging and maturation [4]. The most detailed study of maturation is in the HK97 system where several intermediate complexes have been structurally characterized [29]. These structural studies of maturation are also greatly enhanced by the availability of a high-resolution structure for the major capsid protein gp5 of HK97 [30].

1.6 Scaffolding Proteins

Scaffolding protein molecules are also required in substantial numbers ("hundreds of molecules" [31]) for the formation and stabilization of the procapsid. These proteins are required for the capsid assembly process to proceed to completion, and remain as part of the procapsid particle. During maturation the scaffolding proteins are released. The idea of a protein functioning as a scaffold in capsid assembly arose from early studies of phage P22 [31]and T4 [32]. In these studies, it was shown that a particular class of protein molecules was found in the procapsids, but not in the final mature phage particle. Further, it was shown that mutation of the genes for these particular proteins almost ablated capsid assembly entirely. Thus, the scaffolding protein term was adopted.

Scaffolding protein molecules have been proposed to determine the structure of large complexes by inhibiting improper assembly of the structural subunits themselves [33]. This idea has been furthered by detailed in-vitro assembly

studies of viral capsids in which the scaffolding to capsid protein ratios can be controlled to induce formation of a range of aberrant capsid related structures [34-37]. Scaffolding protein molecules themselves are generally extended and alpha helical. To date the most complete atomic structure for a viral scaffolding protein is that of the 11 kDa gp7 of the *Bacillus subtilus* phage phi29. The structure of this scaffolding protein molecule has been solved to high resolution by X-ray crystallography [38]. The protein was crystallized as an extended alpha-helical coiled-coil dimer (a coiled coil is a structural motif in which adjacent alpha helices come together to form an extended multi-helix bundle). In this same study, the purified protein was shown to behave in solution as a dimer and thus it was proposed that this scaffolding protein oligomerization could help to lower the barrier for capsid nucleation by increasing the effective concentration of associated capsid protein molecules.

A 65 residue segment of the carboxy terminus of the scaffolding protein from the *Salmonella* phage P22 has been solved to atomic resolution by NMR [39]. This C-terminal segment is similar to the N-terminus of the phi29 scaffolding protein [38]. Although the complete structure of the ~300 residue P22 scaffolding protein is not available, its functional and biophysical properties have been studied in great detail [33]. The P22 scaffolding protein behaves as a dimer in solution, and can also form tetrameric species [40]. These studies have been combined with Small-Angle X-Ray Scattering Studies [41] to show that the molecule has an extended shape in solution. It has also been shown that the

C-terminal region of the protein (residues 292-303) is necessary and sufficient for capsid protein binding and capsid assembly [42]. In both the P22 and phi29 phage systems, scaffolding protein dimerization has been proposed to play a role in nucleating the assembly of the procapsid [40]. Much of this information regarding the capsid assembly functions of the P22 scaffolding protein has been derived from the robust in-vitro assembly system developed for phage P22 in which procapsid like particles can be assembled from in-vitro mixing of purified capsid and scaffolding proteins [37, 41, 43, 44].

1.7 Cryoelectron Microscopy Studies of Phage Procapsids

The procapsids of many spherical dsDNA viruses have been studied by cryoelectron microscopy (cryoEM) coupled with computer image processing. This technique, often called single particle analysis (SPA), actually relies on imaging many thousands of identical complexes oriented randomly with respect to the viewing axis [45]. Once the images are collected and digitized, the data are used to calculate a three dimensional reconstruction by assigning orientations to each particle and averaging the particles that represent identical views of the complex. For a more thorough discussion of this technique see [46]. This approach is well suited for the study of large protein complexes like procapsids and has yielded subnanometer resolution reconstructions of procapsids from multiple phage systems [47-49]. CryoEM reconstructions like these, combined with atomic structures for capsid protein monomers have

created pseudo-atomic models of viral capsids that have led to a thorough understanding of the interactions between capsid protein molecules that govern self assembly of these viral particles [50].

Unfortunately, despite the availability and ease of purification of scaffolding containing procapsid particles for several phage systems, the organization and structure of scaffolding protein networks within the assembled procapsid is unclear. Unlike the well-ordered icosahedral capsid protein lattice, the scaffolding protein network (if such a thing exists) does not appear to rigidly obey icosahedral symmetry in the procapsid. Density attributable to scaffolding protein is almost entirely absent in icosahedral reconstructions of phage procapsids produced from SPA of cryoEM data [38, 47]. Even in the P22 system where difference imaging was used to compare scaffolding containing particles with procapsid particles that lacked scaffolding protein, it was only possible to visualize faint hints of density that could be related to the scaffolding protein [51].

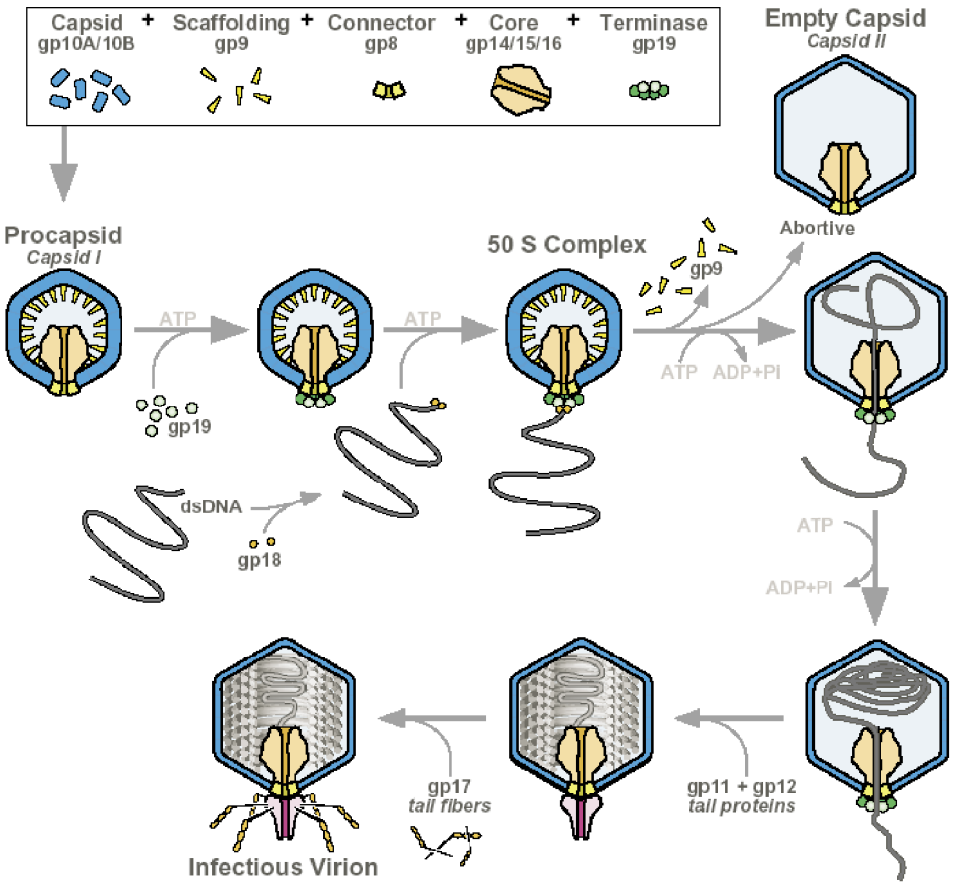
The first hints of internal scaffolding protein networks were only recently seen in the asymmetric reconstructions of phage procapsids [38, 52]. Asymmetric reconstructions do not leverage the high degree of symmetry of the icosahedral particles under study, but instead treat the entire particle as an asymmetric unit so that averaging does not ablate non-icosahedral features. These types of reconstructions require more (usually 60 times the number)

particle images than traditional icosahedral reconstructions to achieve comparable resolutions to traditional icosahedral reconstructions. Only with advances in automated data collection, computation power, and automated processing routines have such methods become practical. In asymmetric reconstructions of the prolate phi29 prohead, concentric shells of density were proposed to be a network of scaffolding protein molecules arranged within the procapsid in a mesh pattern [38]. In contrast, an asymmetric reconstruction of the isometric T7 procapsid showed a faint, disordered protein density that surrounded the connector-core [52]. This density was proposed to be the scaffolding protein network, but detailed interpretation of the network organization was not given and likely not possible from this data. Asymmetric reconstructions in both cases represented an improvement over traditional icosahedral SPA at visualizing scaffolding protein density, but neither offered a concrete picture of the scaffolding network and its interaction with the capsid shell.

Procapsid structure has been studied in detail for many phages, but relatively little information about the scaffolding protein in these procapsids has arisen from such studies. Gel quantitation and mutagenesis studies indicate that scaffolding is necessary for the formation of and present in procapsid particles despite its invisibility in EM reconstructions.

Figure 3 Schematic representation of bacteriophage T7 assembly.

Bacteriophage T7 assembly pathway. Precursor particles assemble from pools of the proteins indicated in the upper box. The DNA enters the particles through the connector in concert with ATP hydrolysis and is packaged in concentric rings. Concurrent with packaging of DNA, the capsid expands and the scaffolding protein is lost. Once packaging is complete, the tail proteins (gp11 and gp12) attach at the connector vertex and six tail fibers are attached to form an infectious virion. Proteins that make up the mature phage are indicated on the schematic. Note that gp9 is not present in the mature phage. Each structure is made of a fixed number of copies of the indicated proteins (copy numbers are not indicated here). Taken from[4].


Capsid
gp10A/10B
+
Scaffolding
gp9
+
Connector
gp8
+
Core
gp14/15/16
+
Terminase
gp19
Empty Capsid
Capsid II
Procapsid
Capsid I
ATP
gp19
ATP
50 S Complex
gp9
ATP
ADP+PI
Abortive
dsDNA
gp18
ATP
ADP+PI
gp17
tail fibers
gp11 + gp12
tail proteins
Infectious Virion

1.8 Bacteriophage T7

In this work, scaffolding mediated virus assembly was studied in the context of bacteriophage T7 (Figure 3). Bacteriophage T7 is a lytic phage with a short, non-contractile tail, and a member of the family podoviridae. T7 is an attractive system for laboratory study not only because of its short replication time, but also because its 40 kb dsDNA genome is well characterized and a number of mutants have been previously described.

1.9 T7 Procapsid Assembly

Bacteriophage T7 has an icosahedral head into which is packaged its double-stranded DNA genome of about 40 kbp [53]. The head or capsid contains 415 copies of the capsid protein gene product 10 (gp10) with one five-fold vertex containing a dodecameric ring called the portal or connector (gp8). Inside of the capsid, at the portal vertex, is a group of proteins called the core (gp14, gp15, and gp16). These core, connector and capsid proteins, along with a scaffolding protein (gp9) make up a meta-stable precursor particle called the procapsid [4]. This particle was seen in wild type infections and was further characterized through electron microscopy studies of a mutant phage lacking DNA polymerase (gp5) and thus unable to package DNA [17]. Infections of wild type *E. coli* with this mutant T7 (termed 5amber28 for the amber mutation found in gene 5 that effects the 28th codon) yielded only unexpanded phage particles lacking DNA and a tail. While these particles had been seen

in small numbers in wild type infections, they were the exclusive isolate of infections with this 5amber28 phage.

1.10 T7 DNA Packaging and Maturation

In the next stage of assembly, the packaging sequence of the DNA is recognized by gp18 or the small subunit of the packaging motor as the large subunit of the same motor assembles in the presence of ATP at the connector vertex [4]. Then, concurrent with DNA packaging, the procapsid undergoes a maturation process known as expansion during which gp9 is lost, the capsid walls become thinner and the capsid as a whole adopts a more angular shape [4]. The loss of gp9 is evidenced by the fact that it appears in SDS PAGE gels stained by coomassie in isolated procapsid particles and is not present in mature phage, or phage particles that have been expanded in vitro [34]. These changes in the capsid due to expansion increase the strength and inner volume of the capsid so that it can accommodate the incoming DNA [54].

While the purpose of expansion is understood, the mechanism by which it occurs in T7 is not. Even in wild type infections, not all expanded particles contain packaged DNA [17]. A population of particles known as abortive heads exist in wild type infections that do not contain DNA, a tail or tail fibers, but are intact, thin walled, rigid and contain all of the protein components of the procapsid with the exception of gp9 [55]. As it has been characterized thus far, the abortive head appears to be a dead end product of the assembly pathway [4].

In other dsDNA bacteriophage, expansion is triggered at the N-terminus by the removal of an N-terminal segment of the capsid protein by a maturational protease [11, 56]. These cleavage controlled expansion events can be reproduced in vitro and visualized by cryoEM [57]. In the case of T7, expansion occurs in the absence of a protease making scaffolding protein gp9 an attractive model for understanding non-proteolytic viral capsid expansion.

DNA packaging reactions are temporally proximal to maturation events in dsDNA phage. These events have been well studied in phage T3, a close relative of T7 [58]. In the case of a successful T3 packaging reaction, ATP is hydrolyzed and 1.8 bases of the dsDNA genome/per hydrolysis event are translocated from the exterior to the interior of the capsid [58]. ATP hydrolysis continues until the entire genome is packaged. In T7 the terminase complex completes the packaging process by cleaving the concatemeric DNA strand [54]. The complete genome is packaged into the expanded T7 capsid and takes on a concentric shell arrangement [54].

The concentric ring DNA motif was observed in the electron microscope when studying a mutant deficient in genes eleven and twelve (the gene products of these genes are the two components of the phage tail) [53]. This tailless phage oriented preferentially in positions that made it possible to visualize the packaged DNA in specific 2D projections [54]. SDS PAGE analysis of these particles compared to wild type particles along with

observations from electron microscopy showed that the tail was composed of gp11 and gp12 and that it's attachment followed the completion of DNA packaging and preceded the attachment of the tail fibers, (gp17) [59]. Attachment of the six tail fibers that serve as host recognition factors completes the assembly of the infectious virion [4]. This process is represented visually in Figure 3. For a more detailed look at T7 biology, see reviews by Dunn and Studier [53] and Steven and Trus [60].

1.11 T7 and Herpes Simplex Virus type 1

The assembly pathways of dsDNA bacteriophages, such as T7, have been studied in great detail in search of common themes that link bacteriophage to the often more complex systems of animal and plant viruses. For example, it has been shown that there are some key similarities to dsDNA bacteriophage found in Herpes Simplex Virus 1 (HSV1) [9]. HSV1 is an enveloped virus that consists of a capsid surrounded by a lipid bi-layer. Like T7, HSV1 assembles via the formation of a prohead-like particle that is stabilized by a scaffolding protein prior to DNA packaging [61]. It has a double-stranded DNA genome and it also contains a portal protein UL6 that is analogous to the connector protein (gp8) in T7 [62]. In both T7 and HSV-1 the connector or portal protein is the site of DNA entry into the capsid. Understanding the capsid architecture and the mechanism of its assembly in T7 can provide new insights into some of the key features of HSV-1, and other animal viruses.

1.12 Previous Structural and Assembly Studies of T7 Procapsids

Previous biochemical studies have shown that the T7 scaffolding protein can be used to catalyze prohead assembly in vitro [34]. In the absence of scaffolding protein, capsid protein (gp10) alone does not assemble into a completed procapsid, but rather yields a myriad of aberrant structures and aggregates [17, 34]. It was also shown by in vitro assembly experiments that a molar excess of scaffolding protein relative to capsid protein can form procapsid-like particles termed 9.10 heads [34]. These particles lack cores and connectors but their capsid morphology matches that of wild type T7 procapsids. Icosahedral reconstructions of T7 9.10 heads revealed projections of density on the internal surface of the procapsid shell [4]. These densities were not seen on the internal surface of reconstructions of the mature capsid. It was proposed that these densities, which were only seen beneath hexameric capsomers, were the product of scaffolding protein contacts with the capsid shell [4].

A recent structural study has characterized the procapsids of T7 by means of cryoelectron microscopy and single particle analysis (SPA) without the imposition of icosahedral symmetry [63]. In this study many details about the nonicosahedral core-connector complex of the T7 procapsids were visually confirmed. However, even without the imposed constraints of icosahedral symmetry, a detailed interpretation of the density newly attributed to

scaffolding protein was limited [63]. These limitations were due in large part to the reliance of SPA on averaging over a large number of particles, which likely contain variable amounts of scaffolding protein [34]. This variation of scaffolding protein amount in T7 will be addressed further in the results and discussion sections of this work.

1.13 Project Overview

The advent of techniques like electron tomography [64] that are most powerful when applied of pleomorphic or polymorphic structures made it possible to apply a novel visualization technique to the question of scaffolding protein organization in the T7 procapsid. Further, the scaffolding protein itself had been previously shown to be available in great abundance via expression of the cloned gene in *E. coli*. This wealth of protein made the biochemical and biophysical characterization of purified scaffolding protein a logical complement to the study of scaffolding protein in the procapsid.

Chapter 2: Methods and Experimental Procedures

2.1 Clones of scaffolding protein

The wild type gene 9 that was used for expression was in plasmid pAR441 [65] in which the sequence for gene 9 including the natural promoter was cloned into pBR322 at the *Bam*HI site in silent (reversed) orientation under the control of its natural phi9 promoter. The sequence of gene 9 was amplified in full length and truncated (residues 1-256) forms using PCR to create the gp9 hexahistadine fusion constructs. The oligos for amplification were:
ATCATATGATGGCTGAATCTAATGCAG (Nde I fwd),
ATCTCGAGGAAGTTCGAATCGATTAC (Xho I rev)
ATCTTCGAGTCAGAAGTTCGAATCGATTAC (Xho I rev Stop)
The full-length (9NHis) and truncated (9_256NHis) gene 9 constructs with the N-terminal HisTags were created by cloning the appropriate gp9 amplification product into pET-28a (Novagen) at the NcoI and XhoI restriction sites. The complementary C-terminal HisTag constructs (9CHis and 9_256CHis) were created by cloning into the same sites in pET-21b (Novagen).

2.2 Expression of the gene 9 clones

All proteins were expressed in BL21 DE3 pLysS (Novagen). For the pET-28 constructs, cultures were grown in LB with 50 ug/mL of kanamycin sulfate. All other cultures were grown in LB with 100 ug/mL of ampicillin. For induction, one liter cultures were incubated at 37^{o}C shaking at 250 rpm in two liter flasks until the optical density reached 0.7 at 600 nm. Cultures were induced with 400 µM IPTG.

Cells were harvested 2 hours after induction by centrifugation. Induction was tracked by SDS PAGE by removing aliquots of culture at 0, 0.5, 1, and 2 hour time points following addition of IPTG.

2.3 Purification of wild type gp9

Cell pellets were resuspended in 50 mL of lysis buffer (20 mM Tris pH 7.4, 3mM TCEP, 4% v/v glycerol, and 2mM EDTA) plus PMSF, leupeptin, and pepstatin. Cells were lysed by freeze/thaw and treated with RNase and DNase. Cell debris were removed by centrifugation and filtration using a 22 μm filter. Clarified lysate was loaded onto a HiTrap Q HP 5mL anion exchange column (GE) equilibrated with binding buffer (lysis buffer + 50mM NH_4Cl). The column was washed with 5 column volumes of binding buffer and bound protein was eluted with a 100 mL linear gradient of 50 mM to 1M NH_4Cl. The peak fractions were pooled in a total of 5mL and equilibrated in binding buffer (including protease inhibitors) using a NAP-25 desalting column (GE). This protein fraction was then loaded onto a 15 ml CHT Hydroxyapatite column (Biorad). Bound protein was eluted using a linear gradient of 0 to 0.5 M sodium phosphate (pH 7.4). The protein eluted at 80 mM sodium phosphate and was pooled and concentrated to 2 ml. This concentrated fraction was loaded onto a Superdex 75 HiLoad 16/60 gel filtration column (GE) equilibrated with binding buffer (lysis buffer + 50mM NH_4Cl). Gp9 eluted at 54 mL and was concentrated to 20.5 mg/ml. In later purifications, gp9 was divided into 1 mL aliquots to which 10% glycerol was added and frozen prior to gel filtration. When needed these frozen aliquots were thawed and purified by gel filtration to remove glycerol and ensure that the protein sample was fresh and non-degraded.

2.4 *Purification of HisTagged gp9*

Following induction, harvested cells were resuspended in 50 mL of HisTag lysis buffer (20 mM Tris pH 7.4, 0.5M NaCl, 3mM TCEP, 4% v/v glycerol, and 2mM EDTA) plus PMSF, leupeptin, and pepstatin. Cells were lysed by freeze/thaw and treated with RNase and DNase. Cell debris was removed by centrifugation and filtration using a 22 µm filter. Clarified lysate was loaded onto a HisTrap HP 5mL nickel sepharose column (GE) equilibrated with HisTag binding buffer (HisTag lysis buffer + 20mM imidazole). The loaded column was washed with 5 column volumes of HisTag binding buffer and bound protein was eluted with a 150 mM imidazole. The peak fractions were pooled in a total of 5mL and equilibrated in binding buffer (including protease inhibitors) using a NAP-25 desalting column (GE). This protein fraction was then loaded onto a Mono Q HR 1.7 mL anion exchange column (GE). Bound protein was eluted using a linear gradient of 50 mM to 1M NH_4Cl. The protein eluted at 0.4 M was pooled and concentrated to 2 ml. This concentrated fraction was loaded onto a Superdex 75 HiLoad 16/60 gel filtration column (GE) equilibrated with binding buffer. Gp9 eluted at 54 mL and was concentrated to 20.5 mg/ml. In later purifications, gp9 was divided into 1 mL aliquots to which 10% glycerol was added and frozen prior to gel filtration. When needed these frozen aliquots were thawed and purified by gel filtration to remove glycerol and ensure that the protein sample was fresh and non-degraded.

2.5 *Analytical Ultracentrifugation*

Sedimentation velocity experiments were conducted at 20.0°C on a Beckman Optima XL-I analytical ultracentrifuge. Samples of full-length and truncated gp9 were prepared in 100 mM NH4Cl, 20 mM Tris.HCl (pH 7.5), 4% (v/v) glycerol, 2 mM EDTA and 3 mM TCEP. Proteins

were loaded at nominal 0.5, 1.0 and 2.0 mg/ml. For each sample, 400 μL was loaded in double sector, Epon charcoal centerpiece cells and centrifuged at 50 krpm in duplicate. 200 single absorbance measurement scans were acquired at 3.35 minute intervals using a radial spacing of 0.003 cm. SEDFIT 11.3 [66] was used to analyze data in terms of a continuous c(s) distribution, or a single discrete species to obtain a sedimentation coefficient and molecular mass. Solution densities ρ were measured at 20.00°C on a Mettler Toledo DE51 density meter. Viscosities η and protein partial specific volumes ν were calculated and corrected for glycerol content using the program SEDNTERP 1.09 [67]. Sedimentation coefficients were corrected to $s_{20,w}$. The c(s) analyses were carried out using an s-value range of 0.05 to 4.0 with a linear resolution of 100 and a confidence level (F-ratio) of 0.68. Both the c(s) and single species analyses returned similar root mean square deviation (rmsd) values for the best fits, ranging from 0.0051 to 0.0098 absorbance units.

2.6 Differential Scanning Calorimetry

The VP-DSC calorimeter was employed (Microcal, Northampton, MA) for calorimetric measurements. Samples were dialyzed overnight against phosphate buffered saline pH 7.4 (PBS) and this dialysis buffer was used as the reference for all scans. Protein samples and buffer reference solutions were thoroughly degassed and carefully loaded into the cells to avoid artifacts caused by gas bubbles. Protein concentrations used in the DSC studies varied between 2.95 and 73.7 μM. To achieve baseline stability, at least ten reference scans with buffer in both cells were carried out before each sample run. The standard DSC experiment consisted of a heating scan at 1.5 degrees/min followed by a second scan to probe the reversibility of the protein melting event. The difference in the heat capacity between the initial and final states was modeled by a

sigmoidal baseline derived from a manual fitting to each state performed as described [68]. The data were analyzed via nonlinear least squares fitting procedures of the ORIGIN 7.0 software.

2.7 ***Trypsin Digests***

Protease digests were carried out in PBS (pH 7.4) buffer. For each digest, 75 μg of purified gp9 was incubated with 0.25 μg of trypsin in a 20 μl reaction (final volume). Reactions were incubated at 37°C for 30 minutes. Samples from each reaction were taken at 5 minute intervals. Protease digestions were quenched by boiling prior to loading onto 10% SDS BisTris Novex gels (Invitrogen). Gels were either stained with Simply Blue (Invitrogen) or protein fragments were transferred to PVDF membranes for mass spectrometry or N-terminal sequencing respectively. An additional set of identical protease reactions were quenched by addition of HCl prior to analysis by HPLC and N-terminal sequencing.

2.8 Mass Spectrometry

Following limited trypsin digestion of purified wild type gp9, fragments were separated by SDS PAGE on 12%, Bis-Tris NuPAGE gels (Invitrogen). The stained bands were excised, washed from salts and dried under vacuum. These bands were then rehydrated with trypsin solution and digested overnight at 37°C. Peptides were extracted and salts were removed on a reverse phase micro-column (Zip-tip, Millipore); and eluted directly on the mass spectrometer sample plate.

Peptides were analyzed on an Applied Biosystems Voyager DE-Pro

time-of-flight mass spectrometer. Trypsin autolysis peaks were used for internal mass calibration of raw spectra. Protein identification was performed on the Mascot Search program, (http://www.matrixscience.com/search_form_select.html licensed and locally run as http://biospec.nih.gov/) against the SProt/Tremble database.

2.9 N-terminal Sequencing

Amino acid analysis was used to confirm trypsin cleavage sites. The SDS gels with the gp9 fragments from limited trypsin proteolysis were blotted on a polyvinylidene difluoride (PVDF) membrane. The proteins were stained with Coomassie brilliant blue R-250, and the stained bands were excised and sequenced for several Edman degradation cycles. The resulting N-terminal sequences identified the trypsin cleavage sites. Peptide fragments that were too small to be resolved by SDS PAGE were separated by HPLC on a 2.1 x 250-mm C18 column (Phenomenex, Torrance, CA). Eluted peptides were harvested, covalently attached to a solid support, and subjected to Edman degradation cycles as before.

2.10 Protein Crystallization

Initial crystallization screening was performed using Hampton Crystal Screen HT with hanging-drop vapor-diffusion trays set by a mosquito robot (TTP LabTech - Cambridge, MA) at 20°C. Drops containing 200 nl of gp9 at an initial concentration of 10.5 mg/ml in crystallization buffer (20 mM Tris pH 7.4, 50 mM NH_4Cl, 3mM TCEP, 4% v/v glycerol, 2mM EDTA) and 200 nl of reservoir solution were equilibrated against 50 µl of reservoir solution. The protein crystallized in multiple

conditions all containing magnesium and PEG 400. Optimized crystals were obtained in 10 days from 0.1 M Hepes-Na (pH 7.5), 180 mM $MgCl_2$, and 24% v/v PEG 400. The crystals were diamond shaped and reached a size of ~50 μm in diameter.

2.11 X-ray diffraction and indexing of diffraction patterns

Prior to X-ray diffraction, crystals were washed in a cryoprotectant solution [36% PEG 400, 0.1 M Hepes-Na (pH 7.5), 180 mM $MgCl_2$], mounted in nylon loops of 0.1 mm in diameter and flash-frozen in liquid nitrogen. Initial diffraction data was collected in house on a Rigaku Raxis IV image plate mounted on a Rigaku RUH3R rotating anode with multilayer focusing optics for CuKα at 100ºK. A higher-resolution dataset to 4.2 Å was collected at ID-22 SER-CAT beamline in the Advanced Photon Source (APS, Argonne National Laboratory, Argonne, IL) on a Mar225 CCD detector at 100ºK. Diffraction data was processed and indexed with program HKL-2000 [69]. Crystals were looped and set at 400 mm from the detector. Data was collected to completion in 1degree rotational increments with 8secs per exposure at 50% beam attenuation.

2.12 Secondary structure prediction

The gp9 amino acid sequence was submitted as a single sequence to JPRED [70] - http://jura.ebi.ac.uk:8008/jpred) a consensus server based on the neural network JNET [71]. A multiple sequence alignment was constructed using the gp9 sequence as the query using three iterations of the PSI-BLAST algorithm. Redundant sequences and gaps that appeared in the query sequence were removed prior to running the prediction algorithms. The multiple sequence alignment is then scored

using the Position-specific scoring matrix [72] and hidden Markov model profiles [73]. A consensus prediction of secondary structure based on a number of different prediction methods including JNET [71] was generated using a three state prediction (helix, beta strand or random coil). JPRED reports an average prediction accuracy per protein of 76.4% with a standard deviation of 8.4% for all predictions [71].

2.13 Detection of heptad repeats

The gp9 amino acid sequence was submitted to the COILS server (http://www.ch.embnet.org/software/COILS_form.html) as a user defined sequence. The scoring profiles for MTK and MTIDK scoring matrices were used on successive runs as the type of coiled-coil interactions for gp9 were not known. Because the gp9 sequence is highly charged, the weighting option was activated to avoid false coiled-coil probabilities that were known to occur in the absence of heptad periodicity in such charged sequences. The sequence was scanned using window sizes of 14, 21, and 28 residues.

2.14 Wild type T7 prohead sample preparation

Wild type proheads were purified from *E. coli* BL21 lysates of cells infected with T7 phage that have an amber mutation in gene 5 (DNA polymerase). Briefly, cell lysates of infected *E. coli* BL21 are adjusted to 0.5M NaCl, and DNase I and RNase treated two hours post-infection. Following this treatment, cell debris is removed by low speed centrifugation, and the clarified lysate is then mixed with PEG 8000 at 10% w/v. The PEG precipitate is spun down at high speed and resuspended in 20mM Tris-HCl pH 8.0 with 10 mM $MgCl_2$ and 2 mM EDTA. The particles are then purified by isopynic banding on a Cesium

Chloride gradient. Cesium Chloride is removed by dialysis against 50 mM Tris-HCl, 10 mM $MgCl_2$, 50 mM NaCl prior to sample vitrification.

2.15 Cryo-electron Microscopy

4.5 uL of sample at roughly 2 mg/ml protein concentration were mixed with a suspension of 10-nm BSA-gold particles (Aurion, Wageningen, the Netherlands) and applied to grids covered with lacey carbon films prepared as described [74]. Sample grids were blotted by hand using Whatman Number 1 filter paper, and plunge-frozen in liquid ethane at -180 C using a KF-80 vitrification system (Leica, Bannockburn, IL.). The vitrified specimens were imaged at 38,500x magnification with a Technai 12 microscope (FEI, Eindhoven, The Netherlands) operating at an accelerating voltage of 120 keV and fitted with a Gatan (Warrendale, PA) model 626 liquid nitrogen cooled cryo-holder . The microscope was also equipped with an energy filter (Gatan GIF 2002) that was operated in the zero-energy-loss mode with a slit width of 20 eV. Single-axis tilt series at 1° steps were recorded at a nominal 4 µm defocus on a 2,048 × 2,048-pixel CCD camera (Gatan), covering ranges of approximately −70° to +70°. Data were collected using SerialEM for automated tilting, tracking, focusing, and image acquisition [75]. The cumulative electron dose was ~ 70 electrons/$Å^2$ per tilt series.

2.16 Image Alignment and Reconstruction

The projection images in each tilt series were aligned by using a combination of cross correlation and fiducial gold markers using IMOD [76]. The aligned images were then used to calculate a tomogram using the weighted back-projection method also in IMOD. The majority of the image analyses were performed on binned tomograms with a voxel size

of 7.8 Å^3. The tomograms were then subjected to 30 iterations of non-linear anisotropic denoising in Bsoft [77] with a λ of 4.8 to reduce noise while preserving the structural information of the particles. The resolution of each tomogram was assessed by the NLOO-2D method as implemented in ELECTRA [78].

2.17 Image Classification, Averaging, and Visualization

Alignment and classification procedures were performed with routines from Bsoft [77], modified as needed and wrapped into Python scripts. Routines from IMOD [76] were used for the selection of particles from tomograms, whereas extraction, alignment and averaging of sub-tomograms were performed using routines from Bsoft. Specifically, 3dmod from the IMOD package[76] was used to select the particles in the tomograms. Sub-tomograms were then extracted with bpick, aligned with bfind and averaged together with badd [77]. All of the procedures with the exception of the selection were automated with the aid of python scripts. These scripts provided the user with a single interface that controlled the workflow for the analysis, seamlessly interfacing the different routines and thereby minimizing the need for user intervention. Figures were prepared using output from the UCSF Chimera package [79] along with images from 3dmod of the IMOD collection of programs [76]. Segmentation, filament length analysis, and path tracings were carried out using the Volume Path Tracer and Distance tools in Chimera.

2.18 Scaffolding Contact Point Analysis

The goal of the analysis was to assess the location of additional density below the shell of the capsid, possibly related to scaffold proteins, that were not necessarily present at all the equivalent positions in the

icosahedrally symmetric procapsid. Individual wild type particles were extracted from full tomograms in cubic volumes of 84x84x84 voxels (7.8 Å/voxel). The particles were then aligned against a reference map as described in the previous section. After alignment, a cylindrical mask was applied to all the particles around the connector-core vertex, in order to exclude the density from this complex from the analysis. Next, particles were normalized and a separate contact point analysis was performed on each one. All the voxels in one iscosahedral asymmetric unit were explored and compared to the equivalent ones in the remaining 59 asymmetric units. For each location, only the maximum density was stored and applied uniformly to all the asymmetric units. For this procedure, we used a variant of the program bsym in the Bsoft package [80], which applies given symmetries to a particle.

The results obtained from the analysis of each particle were finally integrated by averaging, in order to decrease the noise in the final output. The results were visualized by using the density values to color the surface of a modeled icosahedron, with the tools 'Icosahedral Surface' and 'Surface Color' in Chimera [79].

2.19 Scanning Transmission Electron Microscopy (STEM)

Freshly prepared T7 proheads were prepared as described above and checked by negative stain TEM prior to shipping. Samples were prepared as freeze dried specimens according to the methods of Brookhaven National Laboratories STEM facility (www.biology.bnl.gov/stem/stem.html). Briefly, titanium grids were first coated with a thick, holey (5–10 μm) carbon film that served as a support for a continuous layer of thin carbon (~3 nm thick) that has been applied in separate preparation and floated onto the surface of water. 0.2

µg of tobacco mosaic virus (TMV) in solution was added to the sample drop and incubated on the grid for 1 min and then washed with multiple applications of water. Then 2 µl of the specimen was injected into the drop and incubated for an additional minute. Water is washed over the grid several more times following which, the grid was blotted between the two pieces of filter paper (leaving a ~1 µm layer of solution on the grid). The grid is then immediately plunge frozen in liquid nitrogen slush after which it is freeze-dried overnight in an ion-pumped freeze dryer by gradual warming to -80 °C. The grid is then transferred under vacuum to the microscope. Data was collected by scanning with a 0.25 nm probe with a dwell time of 30 µs/pixel at -150 °C on a custom-built STEM at 40 kV.

2.20 Analysis of STEM Images

Dark-field micrographs of freeze-dried sample (512 x 512 pixels) were recorded and rasterized at 1.0 or 2.0 nm per pixel. The images were visualized and analyzed using PCMass [81] software package provided by the Brookhaven STEM Resource Center. Square boxes of 50 nm were used for the measurements of proheads and 9.10 heads while rectangular boxes of 22nm long were used to measure the mass of the TMV control. Background signal was removed using circular and rectangular masks applied within PCMass prior to data collection.

Chapter 3: Results

3.1 Purification and Characterization of gp9

3.1.1 Scaffolding protein purification

Wild type T7 scaffolding protein gp9 and four variants were constructed and expressed in *E. coli* BL21 DE3 plys S (Table 1). The wild type gp9 was purified using a procedure modified from Cerritelli et. al [34]. *E. coli* BL21 DE3 plysS cells containing gp9 plasmid pAR9-441 were grown in a 1 L culture of LB media and gene 9 expression was induced with IPTG. This induction was monitored over a two hour period by SDS PAGE during which time gp9 expression levels reached their maximum (Figure 4). Gp9 bound to a Q-Sepharose anion exchange column and was eluted on a linear gradient in approximately 380 mM NH_4Cl (Figure 5). The peak fractions from this column were pooled and further purified using hydroxyl-apatite and size exclusion chromatography to extremely high purity based on SDS PAGE (see gel insert for Figure 19). The purified gp9 was concentrated to 20.5 mg/ml using a centrifugal filter device.

3.1.1.1 Gp9 carboxy-terminal histidine tags bind NiSepharose

The full-length and truncated gp9 C-terminal His-Tag constructs were purified with the benefit of affinity chromatography. Following the

Table 1

Protein	Gp9 residues	Molecular Weight kDa (calculated)*	Terminal His-Tag
Wild type gp9	1 to 307	33.9	None
9_CHis	1 to 307	35.0	Carboxy
9_NHis	1 to 307	35.1	Amino
9_256CHis	1 to 256	28.3	Carboxy
9_256NHis	1 to 256	28.4	Amino

*All calculated molecular weights include first methionine residue and all residues associated with the His-Tag and linkers supplied by expression vector.

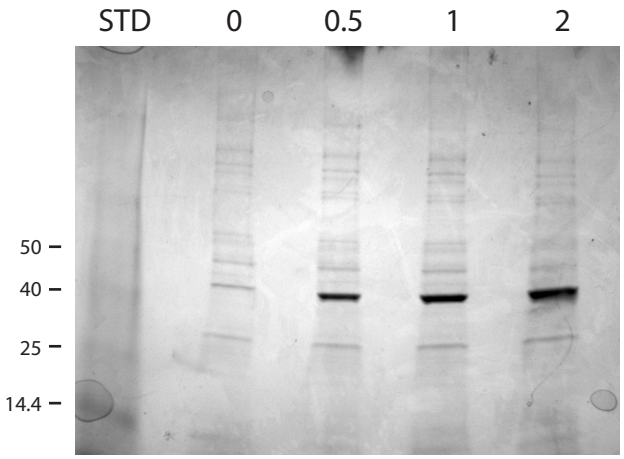

STD
0
0.5
1
2
50
40
25
14.4

Figure 4 Induction of pAR9-441. 20 uL aliquots of a one liter culture of *E. coli* BL21 DE3 carrying pAR9-441 that codes for gp 9 were mixed with SDS PAGE sample buffer and loaded onto a 4-20% SDS polyacrylamide gel at time zero and indicated time points (in hours) following induction with 400 uM IPTG. Samples were boiled for 30 seconds prior to loading and gel was stained with Coomassie R-250. Bench Mark Prestained Standard from Invitrogen was loaded as molecular weight marker.

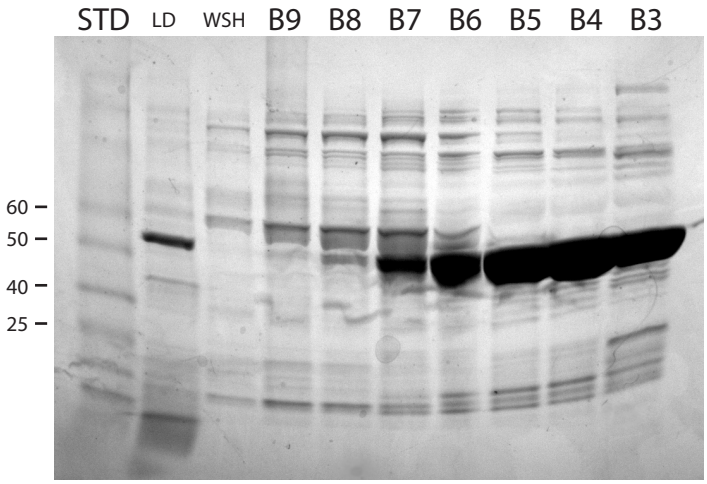

STD
LD
WSH
B9
B8
B7
B6
B5
B4
B3
60
50
40
25

Figure 5 Anion Exchange profile of wt gp9. 50 mL of a cell extract from *E. coli* BL21 DE3 carrying pAR9-441 induction were loaded onto a HiTrap Q HP anion exchange column at a flow rate of 2mL/min. Bound proteins were eluted with a linear gradient of NH_4Cl over 50 mL. 10 uL samples of the Load (LD), wash (WSH), and indicated fractions (B9 through B3) were boiled for 30 seconds prior to loading onto the gel. Gel was stained with Coomassie R-250. Mark 12 Standard from Invitrogen was loaded as molecular weight marker.

same procedure as with wild type gp9 induction, BL21 DE3 plysS cells containing plasmids 9_CHis and 9_256CHis were cultured and induced with Isopropyl β-D-1-thiogalactopyranoside (IPTG) to produce gp9 with C-terminal HisTags in full-length and truncated form respectively. The cell extracts from one liter cultures were passed separately over affinity columns containing Ni Sepharose. Both the 9_CHis and 9_256CHis forms of gp9 bound to the affinity column and were eluted with 150 mM imidazole (Figure 6). The peak fractions for each form of the C-terminally HisTagged gp9 were pooled and further purified by application to a high resolution anion exchange column followed by size exclusion chromatography to greater than 95% purity based on SDS PAGE. The purified proteins were concentrated to 21.5 mg/ml for 9_CHis, and 23.1 mg/ml for 9_256CHis.

3.1.1.2 Amino-terminal histidine tags decrease gp9 expression levels and do not bind NiSepharose

The two complementary N-terminally His-tagged constructs (9_NHis and 9_256NHis) were created, but purification was problematic. Expression levels for the full-length gp9 with the N-terminal His-tag (9_NHis) were substantially lower than the levels seen with either wild type gp9 or 9_CHis (Figure 7). Further, when 9_NHis was passed over a NiNTA column under native conditions, none of the protein bound to the column (Figure 8). Sequencing of the plasmid p9_NHis confirmed that the construct was correct. To test for the presence of

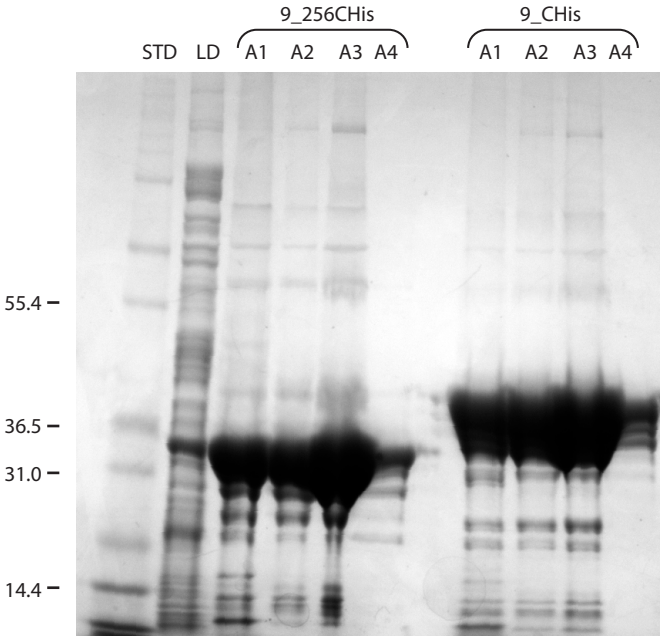

9_256CHis
9_CHis
STD
LD
A1
A2
A3
A4
A1
A2
A3
A4
55.4
36.5
31.0
14.4

Figure 6 Nickel Sepharose column profile of C-terminal HisTag gp9 constructs. 100 mL of a cell extract from induced *E. coli* BL21 DE3 carrying p9_256CHis were loaded onto a HisTrap HP 5 ml affinity column (GE) at a flow rate of 2.5 mL/min. Bound proteins were eluted with 250mM imidizole. 10 μL of the Load fraction (LD) and indicated fractions (9_256CHis A1-A4) plus 4 μl of SDS PAGE sample buffer were boiled for 30 seconds prior to loading onto the SDS polyacrylamide gel. Fractions from a separate run of 9_CHis extract under identical conditions is also shown (9_CHis A1-A4). Gel was stained with Coomassie R-250. Bench Mark Standard from Invitrogen was loaded as molecular weight marker in the first lane.

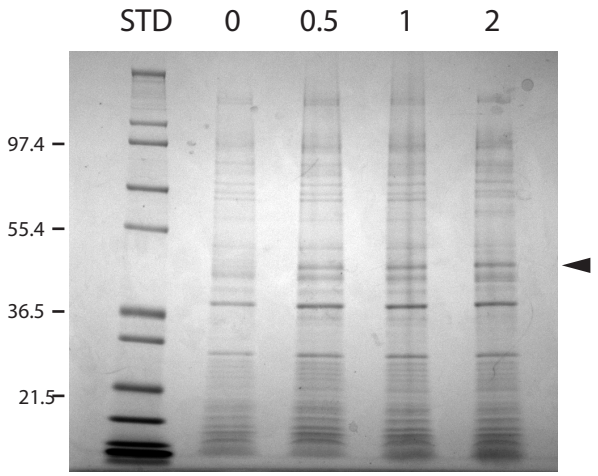

STD
0
0.5
1
2
97.4
55.4
36.5
21.5

Figure 7 Induction of p9_NHis. 20 μl aliquots of a one liter culture of *E. coli* BL21 DE3 carrying p9_NHis that codes for gp 9 with an amino terminal HisTag were combined with 4 μl of SDS PAGE sample buffer and loaded onto a 4-20% SDS polyacrylamide gel at time zero and indicated time points (in hours) following induction with 400 μM IPTG. Samples were boiled for 30 seconds prior to loading and gel was stained with Coomassie R-250. The arrow indicates the migration position of the HisTagged gp9. Molecular weight standard from Invitrogen was loaded as a marker in the first lane.

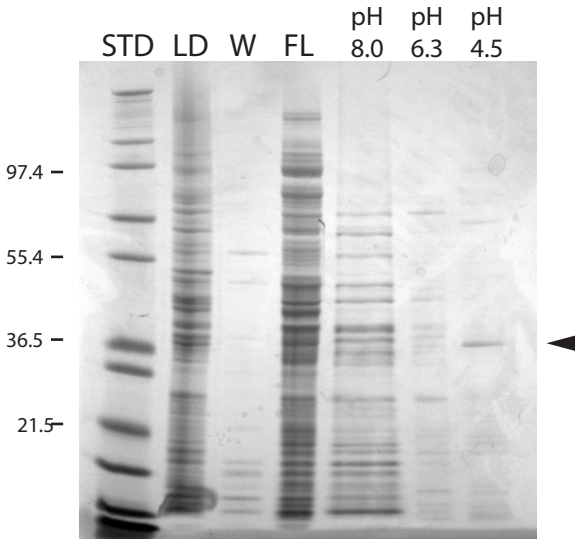

STD
LD
W
FL
pH 8.0
pH 6.3
pH 4.5
97.4
55.4
36.5
21.5

Figure 8 Nickel Sepharose column profile of denatured N-terminal HisTag gp9 construct. 20 mL of a cell extract from induced *E. coli* BL21 DE3 carrying p9_NHis was denatured in 8 M Urea and loaded onto a HisTrap HP 5 ml affinity column (GE) using a 25cc syringe. Bound proteins were eluted with a pH step gradient at pH 8.0, pH 6.3, and pH 4.5 (5 column volumes each). 10 µl samples of the Load (LD), Wash (W), Flow Through (FL), and indicated pH fractions were boiled for 30 seconds prior to loading onto the SDS polyacrylamide gel. Gel was stained with Coomassie R-250. Arrow indicates the gp9 HisTagged fusion that was eluted at pH 4.5.

the HisTag in the translated gp9 fusion, 9_NHis was denatured in 8M Urea and passed over the affinity column. This denatured 9_NHis was found to bind to the Ni Sepharose column confirming the presence of the HisTag in the expressed protein construct (Figure 8). Only in the denatured state was the N-terminal HisTag able to bind the column suggesting that this affinity tag was not exposed in the native form.

Attempts were also made to purify the truncated gp9 (residues 1-256) with an N-terminal HisTag (9_256NHis). Even several hours after induction, expression of this truncated form of gp9 with the N-terminal HisTag was not detectable by SDS PAGE. There was no detectable binding when cell extracts of 9_256NHis expression were passed over a NiNTA column.

3.1.2.1 Gel filtration indicates that gp9 is an extended monomer or compact dimer

Purified wild type gp9, 9_CHis, and 9_256CHis were run over a Superdex 75 HiLoad 16/60 column (GE Biosciences). All forms of gp9 eluted as a single, sharp peak with no traces of high molecular weight gp9 multimers or aggregates in the sample (Figure 9), and no significant protein loss during purification. Wild type gp9 was eluted from the column at 52.5 ml corresponding to a protein with a molecular weight between 44 and 67 kDa

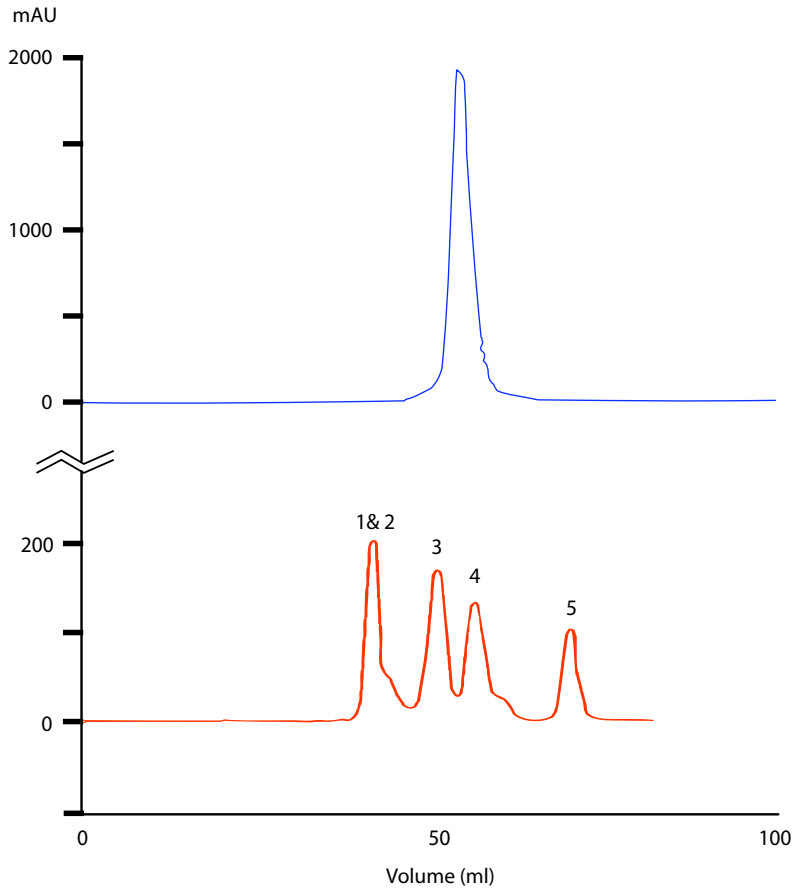

mAU
2000
1000
0
200
0
1& 2
3
4
5
0
50
100
Volume (ml)

Figure 9 Gel filtration column chromatogram of purified full length gp9. 5 ml peak fractions from hydroxyl-apatite column were pooled, concentrated to 2ml, and loaded onto a Superdex 75 HiLoad 16/60 size exclusion column (GE) at 1.0 ml/min. Gp9 eluted at 54 ml as a single sharp peak (top). Molecular weight calibration standards [Ferritin (1), IgG (2), Albumin (3), Ovalbumin (4), and Myoglobin (5)] (GE) were run on the same column and their elution profile is shown (bottom). According to the standards (bottom), gp9 elution corresponds to molecular weight between 67 kDa (3) and 43 kDa (4).

based on calibration standards (Figure 9). This elution profile indicates either a compact dimer or a highly extended monomer based on a calculated molecular weight of 33.9 kDa from the gp9 sequence. The elution profile of 9_CHis was indistinguishable from that of the wild type gp9. The truncated 9_256CHis also migrated through the column as a single peak and eluted at 54.5 ml (a slightly larger elution volume than the wild type and 9_CHis, suggesting a smaller molecule).

3.1.3 Analytical Ultracentrifugation confirms gp9 is an elongated monomer in solution

In order to unambiguously determine the oligomeric status of purified gp9 in solution, a series of sedimentation velocity experiments were carried out by my collaborator Dr. Rodolfo Ghirlando of the LMB in NIDDK at various loading concentrations on full-length gp9 as well as the C-terminally truncated construct 9_256CHis. Data analysis in terms of a continuous c(s) distribution returned excellent fits showing the presence of a monodisperse single species for the full length protein (Figure 10). The weight average sedimentation coefficient for the full length gp9 was not dependent on sample concentration. Analysis in terms of a single discrete species (Figure 10) returned an average sedimentation coefficient of 2.26 ± 0.01 S and a molecular mass of 32.6 ± 0.6 kDa (n = 0.93 ± 0.02). The c(s) analysis returns a frictional ratio f/f_o of 1.75.

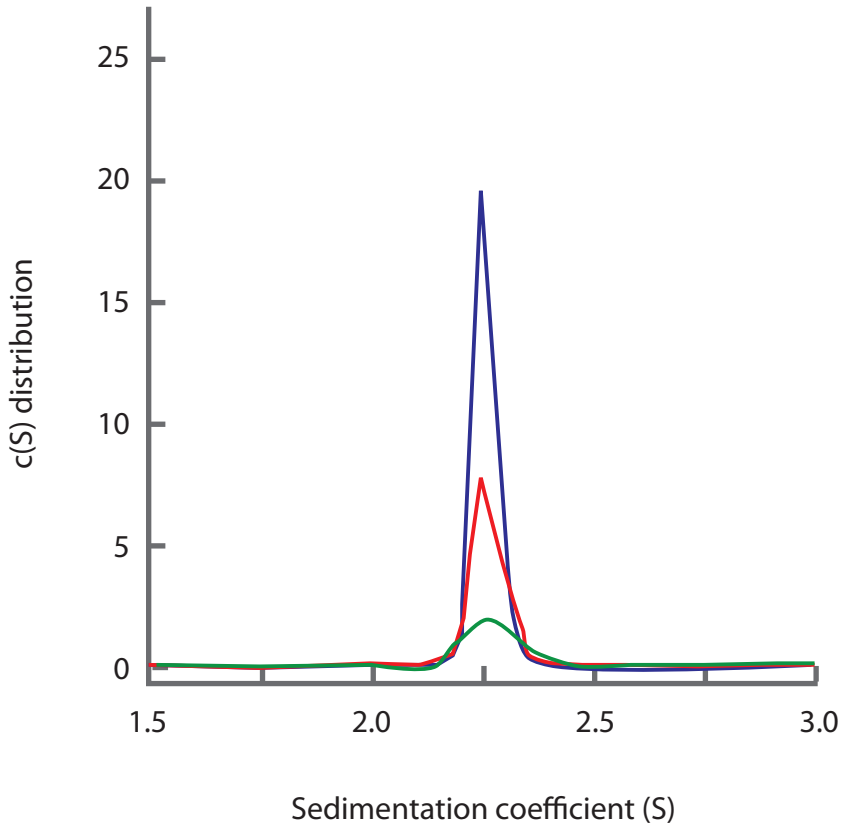
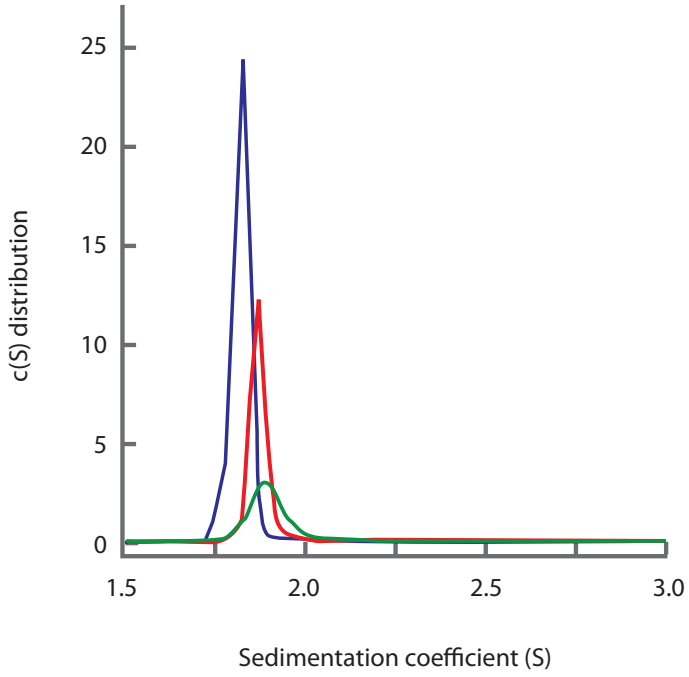

25
20
15
10
5
0
c(S) distribution
1.5
2.0
2.5
3.0
Sedimentation coefficient (S)

Figure 10 Sedimentation velocity experiments on full length *gp9*. The sedimentation coefficient distributions, c(*s*), based on sedimentation velocity data collected at 50 krpm, 280 nm and 20.0ºC are plotted for full length *gp9*. The loading concentrations of 0.5 mg/ml (green), 1 mg/ml (red) and 2 mg/ml (blue) were analyzed. For each load, only a single oligomeric species of full length gp9 was detected.

a


25
20
15
10
5
0
c(S) distribution
1.5
2.0
2.5
3.0
Sedimentation coefficient (S)

b

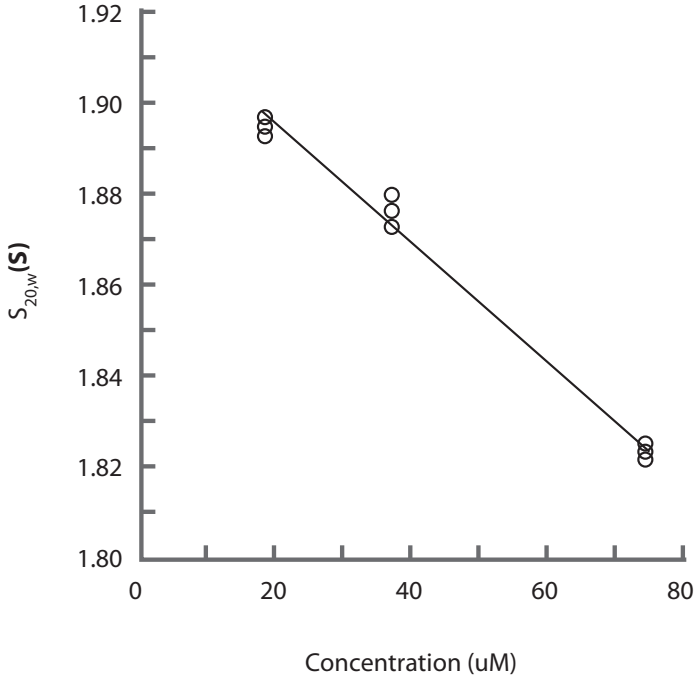

1.92
1.90
1.88
1.86
1.84
1.82
1.80
$S_{20,w}$ (S)
0
20
40
60
80
Concentration (uM)

Figure 11 Sedimentation velocity experiments on truncated *gp9* (residues 1-256). Plot of c(*s*) distributions based on sedimentation velocity data collected at 50 krpm and 20.0ºC for the C-terminally truncated form of *gp9* (a). Loading concentrations of 0.5 mg/ml (green), 1 mg/ml (red) and 2 mg/ml (blue) were analyzed. Dependence of the sedimentation coefficients obtained from the single discrete species and c(*s*) analyses on the loading concentration are shown in (b). The black line in (b) indicates the best linear fit to the data.

molecular mass of 27.1 ± 0.3 kDa (n = 0.96 ± 0.01) was obtained. The experimental frictional ratio for the lowest loading concentration of 9_256CHis gp9 was f/f_o of 1.87.

In order to derive an approximate shape of the molecule, the program SEDNTERP (http://www.rasmb.bbri.org/) developed by Hayes, Laue, and Philo was used to calculate an axial ratio for gp9. Axial ratios were calculated for both full length and truncated forms of gp9 as a prolate ellipsoid, an oblate ellipsoid, and a cylinder. A summary of these ratios is found in Table 2.

3.1.4 Differential scanning calorimetric analysis reveals broad gp9 melting event

Purified gp9 was subjected to analysis by differential scanning calorimetry to probe for temperature dependent protein unfolding events that would suggest the presence of a molecule that is natively folded. Gp9, loaded at a concentration of 2.5 mg/ml, was scanned across a temperature range of 15 to 120 degrees Celsius (Figure 12). The protein underwent a melting event starting at ~35^{o}C and persisting to 85^{o}C. This melting persisted on consecutive scans of the sample that was re-cooled to 10 ^{o}C prior to the scan (Figure 13). The melting event was readily detectable at protein concentrations of 2.5, 0.5, and 0.1 mg/mL. The melting curve was fit to a non-two state model in which two separate melting events with Tm's of 49.4 C and 62.2 C overlap to produce the broad peak seen in

Table 2

Calculated Axial Ratios

	Full length gp9	Truncated gp9
Prolate ellipsoid	9.917	10.76
Oblate ellipsoid	12.20	13.38
Cylinder	7.18	7.88

Axial ratios were calculated from velocity sedimentation data using SEDNTERP (http://www.rasmb.bbri.org/) for both the full-length gp9 (9CHis) and the truncated gp9 (9_256CHis).

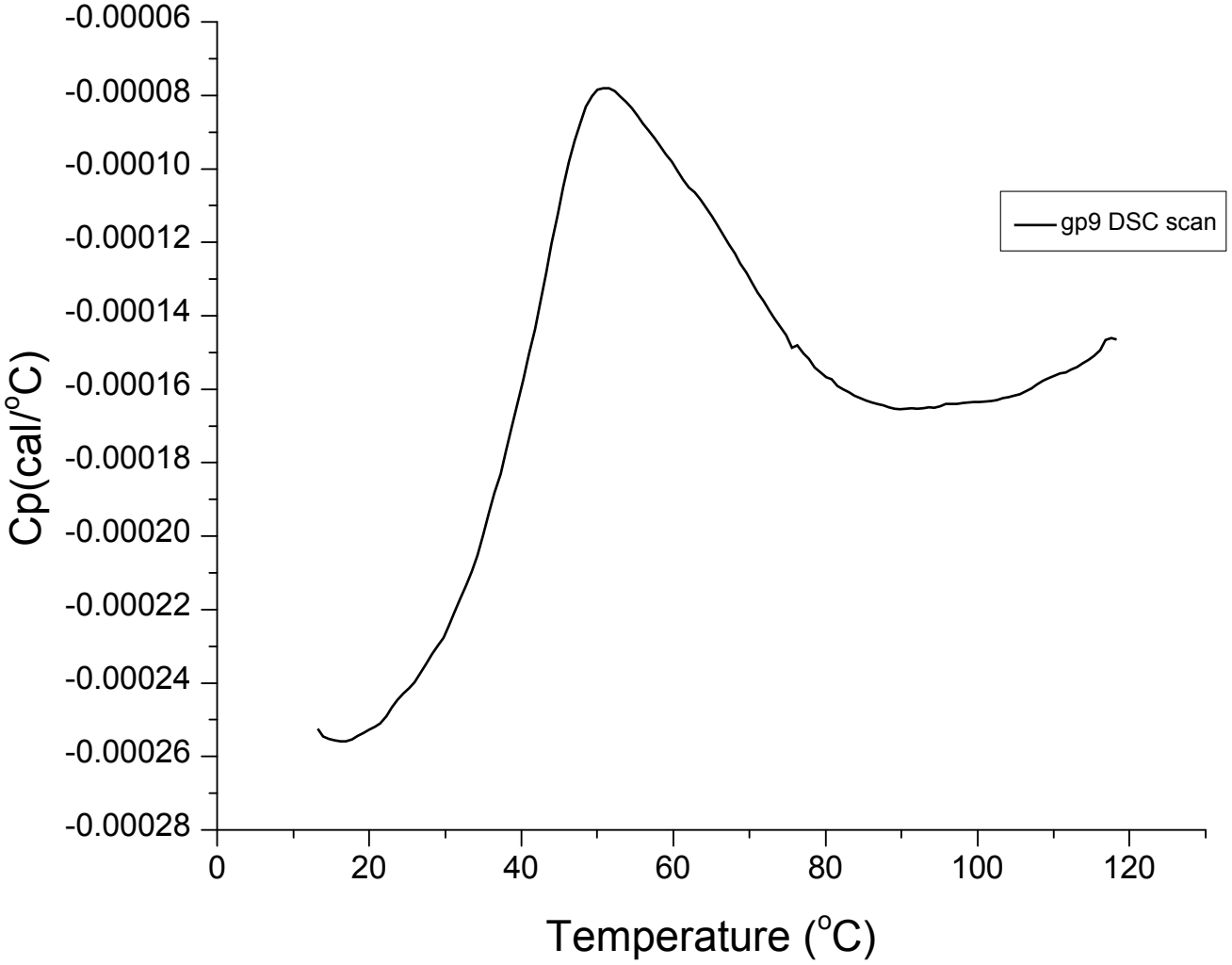

-0.00006
-0.00008
-0.00010
-0.00012
-0.00014
-0.00016
-0.00018
-0.00020
-0.00022
-0.00024
-0.00026
-0.00028
Cp(cal/°C)
0
20
40
60
80
100
120
Temperature (°C)
gp9 DSC scan

Figure 12 Buffer normalized differential scanning calorimetric trace of wild type gp9. Purified gp9 in 1X PBS was diluted to 2 mg/ml and scanned from 15^{o} C to 120^{o} C. The resulting trace is shown in black.

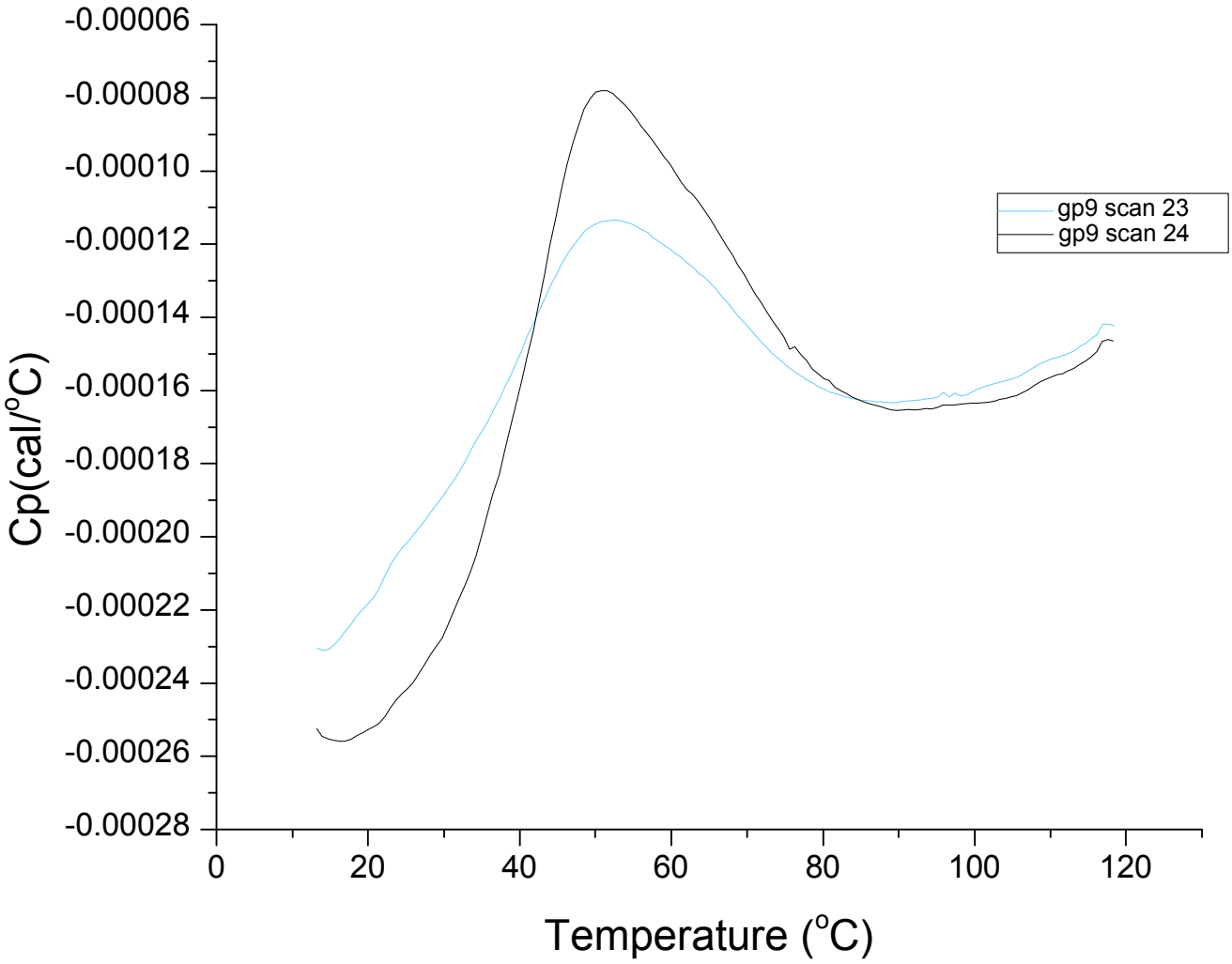

-0.00006
-0.00008
-0.00010
-0.00012
-0.00014
-0.00016
-0.00018
-0.00020
-0.00022
-0.00024
-0.00026
-0.00028
Cp(cal/°C)
0
20
40
60
80
100
120
Temperature (°C)
gp9 scan 23
gp9 scan 24

Figure 13 Consecutive buffer normalized differential scanning calorimetric traces of wild type gp9. Purified gp9 in 1X PBS was diluted to 2 mg/ml and scanned from 15^{o} C to 120^{o} C (black). The same sample was then cooled and rescanned over the same temperature range (blue).

the scans (Figure 14). It is possible that these two melting events correspond to the denaturation of two regions or domains of the protein. The total change in enthalpy for this transition was calculated by integrating the area under the two melting events and found to be 1.46 x 10^3 kcal/mol of gp9.

3.1.5 Secondary structure prediction indicates gp9 is primarily alpha-helical and contains heptad repeats

A variety of secondary structure prediction algorithms (see Methods) were applied to the amino acid sequence of gp9 and their predictions were in excellent agreement regarding the presence of particular secondary structural elements. A consensus from these predictions indicated that 46 % of the residues in the gp9 sequence form alpha-helical motifs, and 47 % form random coil (Figure 15). Only seven percent of the 306 residue sequence is predicted to contain beta elements. The N-terminal region (residues 1 through 78) contains all of the predicted beta elements and is separated from the largely alpha-helical central region by 32 residues of predicted random coil. The N-terminal region contains a large number of acidic residues, which contribute the proteins calculated acidic pI of 4.2. In the central region of gp9 (residue 110 to residue 245), roughly 85 percent of the sequence was predicted to be alpha-helical. The helices in this region are separated by short stretches of coil from 2 to 6 residues in length. The longest helices in this

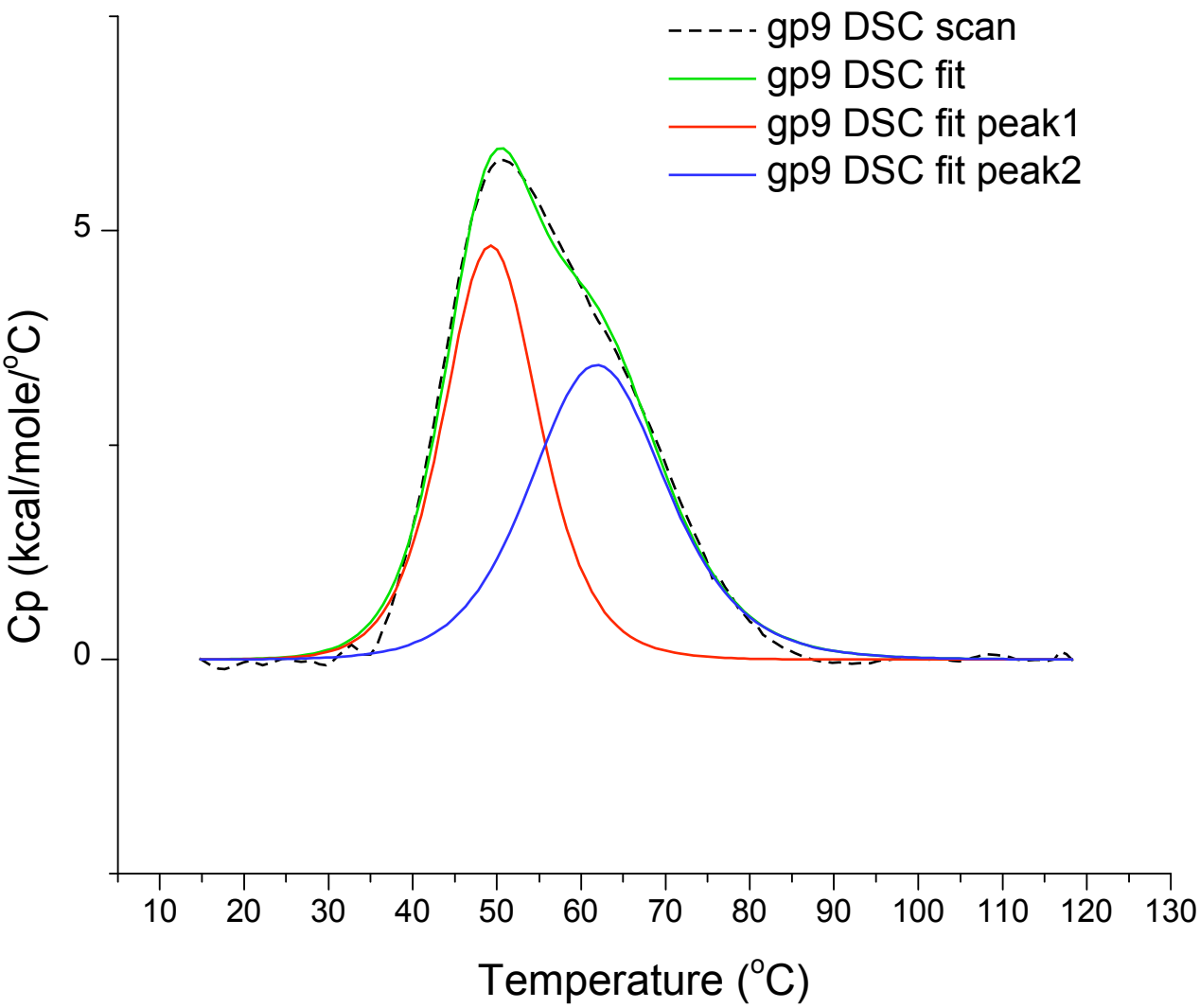

gp9 DSC scan
gp9 DSC fit
gp9 DSC fit peak1
gp9 DSC fit peak2
Cp (kcal/mole/°C)
5
0
10
20
30
40
50
60
70
80
90
100
110
120
130
Temperature (°C)

Figure 14 Curve fitting of differential scanning calorimetric trace of wild type gp9. Purified gp9 scanned from 15^{o} C to 120^{o} C (dashed line) was fitted to a model curve for a non-two state melting event (green) using Origin for MicroCal (GE). The two component peaks that make up the melting event (red and blue) have Tm's of 49^{o} C and 62^{o} C respectively.

```
                : 1---------11--------21--------31--------41--------51--------61--------71--------81--------91--------101-------
gp9seq          : MAESNADVYASFGVNSAVMSGGSVEEHEQNMLALDVAARDGDDAIELASDEVETERDLYDNSDPFGQEDDEGRIQVRIGDGFEPTDVDTGEEGVEGTEGSEEFTSLGETP

pred            : -------EEEEE-----EEE----HHHHHH--------------EEE-------------------------EEEEEE----------------------EE-------
```

```
                : 111-------121-------131-------141-------151-------161-------171-------181-------191-------201-------211-------
gp9seq          : EELVAASEQLGEHEEGFQEMINIAAERGMSVETIEAIQREYEENEELSAESYAKLAEIGYTKAFIDSYIRGQEALVEQYVNSVIEYAGGRERFDALYNHLETHNPEAAQS

pred            : HHHHHHHHHHHHHHHHHHHHHHHHHH----HHHHHHHHHHHH-----HHHHHHHHH----HHHHHHHHHHHHHHHHHHHHHHHH----HHHHHHHHH---HHHHH
```

```
                : 221-------231-------241-------251-------261-------271-------281-------291-------301----
gp9seq          : LDNALTNRDLATVKAIINLAGESRAKAFGRKPTRSVTNRAIPAKPQATKHEGFADRSEMIKAMSDPRYRTDANYRRQVEQKVIDSNF

pred            : HHHHHH---HHHHHHHHHHHHHH----------------------------HHHHHHHHH---------HHHHHHHHHHHH---
```

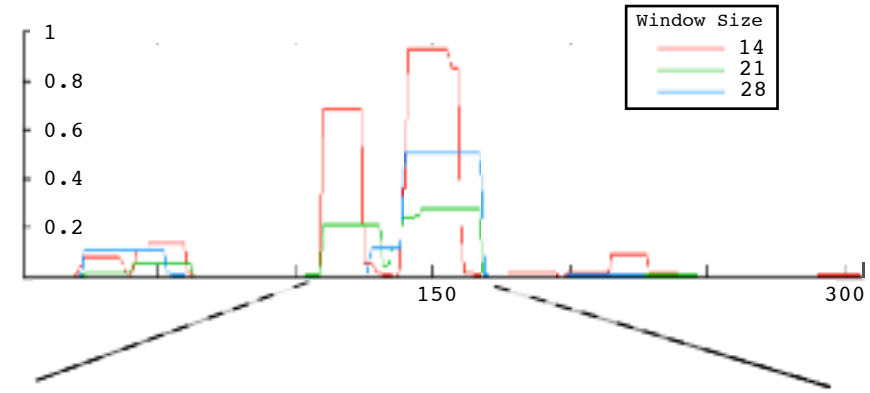
Figure 15 Secondary structure prediction profile for wild type gp9. The gp9 sequence was subjected to analysis by multiple secondary structure prediction algorithms (See Methods). The consensus prediction from these routines is displayed below the sequence. Predicted residue conformations are indicated as follows: Helix - H (red), Beta element - E (yellow), and random coil - dashed line (black).

region were predicted to be 25 residues each. The roughly 60 remaining C-terminal residues were predicted to form irregular secondary structure with the exception of two short helical stretches. These helices together span residues 276 to 303 and are separated by nine amino acids of predicted random coil (Figure 15).

Heptad repeats are the characteristic sequence motif for a protein that has a propensity for the formation of alpha-helical coiled-coil [82]. Evaluation of the gp9 amino acid sequence by the COILS algorithm [83] detected two types of heptad repeats (Figure 16). The first sets of repeats were detected in residues 110 to 124 and 142 to 156 with a 14 residue window size. The second repeat detected was a four heptad motif (residues 141 to 169). The predicted heptad repeat motifs are found in the region of gp9 predicted to be largely alpha-helical (see earlier).

3.1.6 Trypsin proteolysis coupled with mass spectrometry and N-terminal sequencing further defines gp9 domains

In an attempt to support the presence of secondary structure elements predicted for gp9 as well as search for protease stable domains for crystallization, the purified T7 scaffolding protein was subjected to proteolysis by trypsin. Treatment with trypsin results in a series of defined fragments that were resolved by SDS PAGE (Figure 17). The fragments that were too small

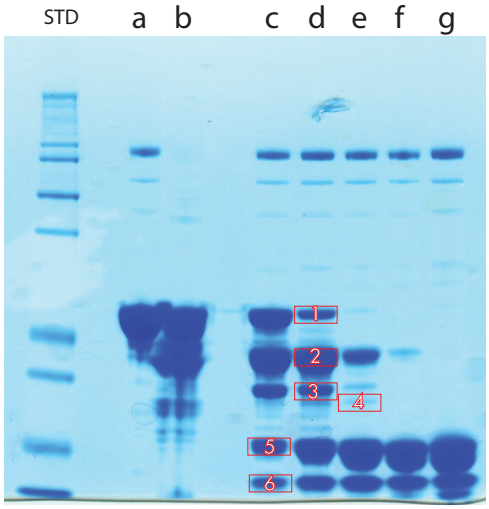


110– EELVAASEQLGEHEEGFQEMINIAAERGMSVETIEAIQREYEENEELSAESYAKLAEIG–170
fgabcdefgabcde fgabcdefgabcdefg

110– EELVAASEQLGEHEEGFQEMINIAAERGMSVETIEAIQREYEENEELSAESYAKLAEIG–170
efgabcdefgabcdefgabcdefgabcd

Figure 16 Detection of heptad repeats by COILS algorithm in wild type gp9. The gp9 sequence was analyzed by the COILS algorithm (Lupas et. al 1997). The sequence was scanned with window sizes of 14, 21, and 28 residues. Top: The probability scores are plotted for each window size in red (14), green (21), and blue (28). Bottom: Regions of sequence scoring above 0.5 are shown in detail with the predicted heptads highlighted in red (14) and blue (28). The predicted heptad position for each amino acid is indicated by the multicolored a-g below each highlighted residue.

A



1 Full length gp9
2 gp9 1 - 259 (28.2 kDa)
3 gp9 78 - 307 (28.2 kDa)
4 gp9 78 - 259 (19.5 kDa)
5 gp9 138 - 259 (13.6 kDa)
6 gp9 1 - 77 (8.2 kDa)

7 gp9 255 - 307 (6.1 kDa)
8 gp9 260 - 307 (5.6 kDa)
9 gp9 245 - 281 (4.1 kDa)
10 gp9 247 - 281 (3.9 kDa)

B

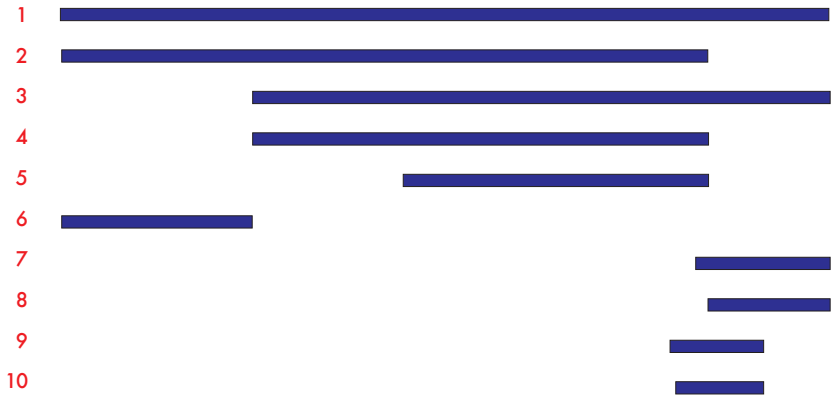
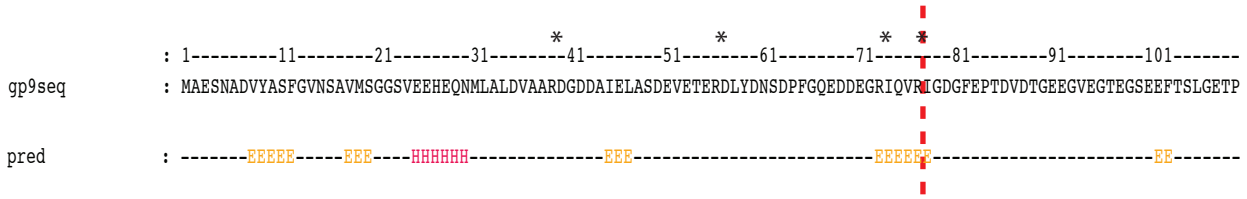


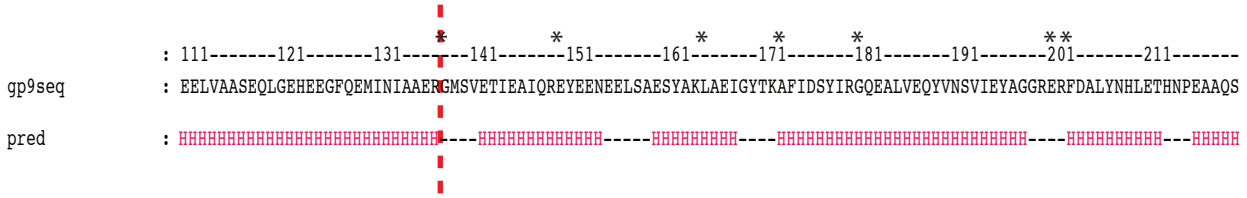
Figure 17 Analysis of protease digestion profile of wild type gp9. Roughly 100 μg of gp9 was subjected to digestion by less than 1 μg of trypsin in a 20 μl reaction. The reactions were carried out over a period of 20 minutes and samples were removed and immediately boiled at 1, 5, 10, 15, and 20 minutes (A: lanes c-g respectively). Gp9 stored for periods of one week (a), and one month (b) at 4° C were also loaded onto the 10% Bis-Tris polyacrylamide gel (A). The bands indicated by the red boxes as well as smaller digestion products that were resolved using HPLC were identified by a combination of mass spectrometry and N-terminal sequencing (see A red numbers 1-6 and 7-10 respectively). The residue numbers and molecular weights for each digestion product (A) are indicated as well as the size and position of each fragment in the gp9 sequence (B).

to be resolved via SDS PAGE were separated using HPLC reverse phase chromatography.

The isolated fragments from polyacrylamide gels and HPLC were analyzed by mass spectrometry. In addition, trypsin digests were separated on SDS polyacrylimide gels and transferred to PVDF membranes and sequenced by N-terminal sequencing methods. Together with the mass spectra these data were used to determine the molecular weight and sequence of each trypsin fragment (Figure 17). These experiments were carried out by my collaborator Dr. Lyuben Marekov of the LSBR in NIAMS.

These fragments were then used to identify the trypsin sites that were cleaved during the proteolysis time course (Figure 18). The gp9 sequence contains 26 trypsin sites spanning the majority of the sequence. Thirteen trypsin sites were found in the C-terminal 67 amino acid residues (Figure 18). The protease digestion profile shows that seven of the 26 available trypsin sites are accessible for protease cleavage. The majority of the cleavage sites mapped to areas of the peptide predicted to be random coil or extended loops (see earlier), thereby supporting the sequence-based secondary structure predictions for those regions. Only two of the 13 trypsin sites found in the first 240 residues were cleaved, while 5 of the 13 sites in the remaining C-terminal 67 residues were cut by the protease (Figure 18).


: 1---------11--------21--------31--------41--------51--------61--------71--------81--------91--------101-------
gp9seq : MAESNADVYASFGVNSAVMSGGSVEEHEQNMLALDVAARDGDDAIELASDEVETERDLYDNSDPFGQEDDEGRIQVRIGDGFEPTDVDTGEEGVEGTEGSEEFTSLGETP
pred : -------EEEEE-----EEE----HHHHHH--------------EEE-------------------------EEEEEE----------------------EE-------


: 111-------121-------131-------141-------151-------161-------171-------181-------191-------201-------211-------
gp9seq : EELVAASEQLGEHEEGFQEMINIAAERGMSVETIEAIQREYEENEELSAESYAKLAEIGYTKAFIDSYIRGQEALVEQYVNSVIEYAGGRERFDALYNHLETHNPEAAQS
pred : HHHHHHHHHHHHHHHHHHHHHHHHHH----HHHHHHHHHHHH-----HHHHHHHHH----HHHHHHHHHHHHHHHHHHHHHHHH----HHHHHHHHH---HHHHH

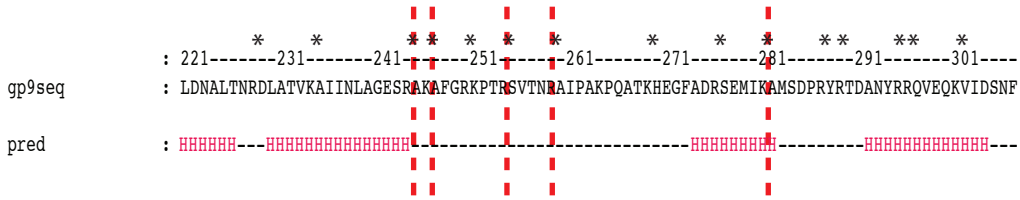

: 221-------231-------241-------251-------261-------271-------281-------291-------301----
gp9seq : LDNALTNRDLATVKAIINLAGESRAKAFGRKPTRSVTNRAIPAKPQATKHEGFADRSEMIKAMSDPRYRTDANYRRQVEQKVIDSNF
pred : HHHHHH---HHHHHHHHHHHHHH----------------------HHHHHHHHH---------HHHHHHHHHHHH---

Figure 18 Protease cleavage results and secondary structure prediction for wild type gp9. The gp9 sequence is plotted with trypsin sites indicated by * and cleaved trypsin sites indicated by a dashed red line based on the results of the limited proteolysis experiments. The gp9 secondary structure prediction results are shown for reference with the resulting protease digest pattern. Predicted residue conformations are indicated as follows: Helix - H (red), Beta element - E (yellow), and random coil - dashed line (black).

Protease-sensitive sites at R137 and K280 were both cleaved by trypsin. These sites fell in regions of predicted coil between two helices (Figure 16). Based on the secondary structure predictions and the protease accessibility to these regions

of the sequence, it is likely that these regions are involved in a helix-turn-helix (HTH) motifs. The significance of these observations will be discussed later (see Discussion).

3.1.7 Gp9 Protein Crystals indexed as F 2 3 cubic (icosahedral) space group

The wild type T7 scaffolding protein was purified to 20.5 mg/ml in the presence of 100 mM NH4Cl, and screened using automated high throughput methods. Initial 200 nl vapor diffusion drops gave small crystals in conditions containing Mg salt and polyethylene glycol 400. These conditions were optimized in larger 2 μl hanging drops where diamond shaped crystals appeared after five days at pH 7.5 in the presence of Magnesium with 24 % PEG 400 as the precipitant. These crystals grew to a size of ~50 μm ten days after they appeared and then stopped growing (Figure 19B and C). The crystals were transferred to 36 % PEG 400 for cryoprotection and frozen in liquid Nitrogen.

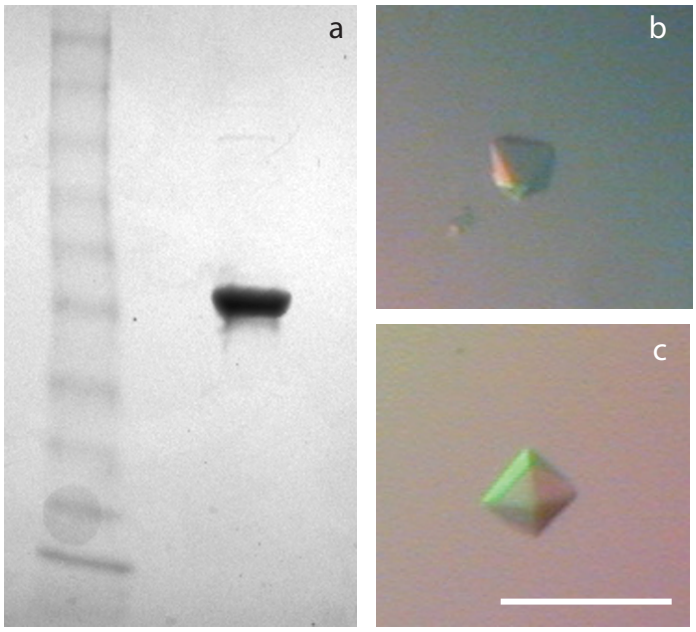

a
b
c

Figure 19 Crystallization of T7 scaffolding protein, gp9. a) Roughly 10 ugs of gp9 loaded onto SDS PAGE to assay purity prior to crystallization. b and c) Crystals of gp9 grown at 20^{o} C in the presence of 180 mM $MgCl_2$ and 24% PEG 400. In each case the largest dimension was ~50 μm in length.

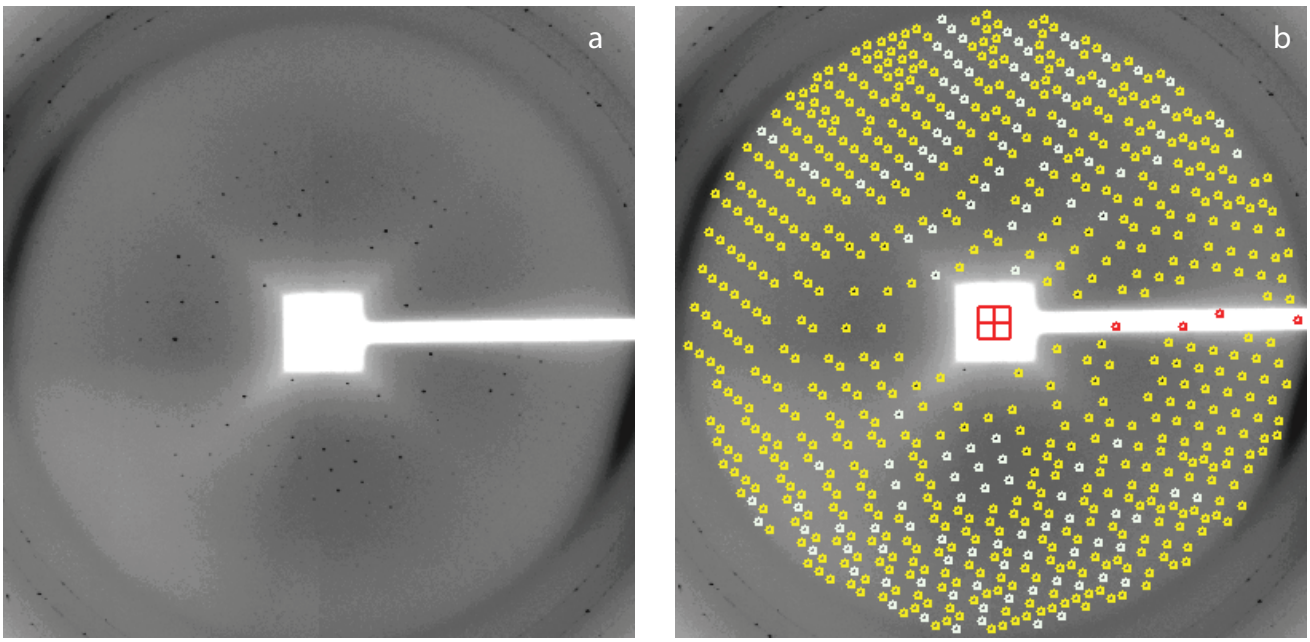

a
b

Figure 20 Indexing of T7 scaffolding protein, gp9 X-ray diffraction patterns. a) Diffraction image of a crystal of gp9. b) Indexing of the diffraction pattern shown in (a). The white circles indicate diffraction spots detected in this reflection. The yellow circles indicate spots that are predicted to appear in an "ideal" reflection based on the predicted space group. Red circles indicate predicted spots that are in the area of the beam stop. The crystal space group was found to be F 2 3 cubic based on the results of the indexing.

The frozen crystals were tested for diffraction on an in-house X-ray generator and found to diffract to 6.2 Å. The crystals were then shipped to the Advanced Photon Source (APS) where they yielded 4.3 Å and 4.6 Å data sets (Figure 20A). The APS 4.3 Å resolution data set was indexed (Figure 20B) and revealed that the crystals belong to the cubic space group F 2 3 (unit-cell parameters a = 178.1, b = 178.1 c = 178.1 Å). Considerations of the packing density suggests the presence of one subunit in the asymmetric unit, which would correspond to a Matthews coefficient (VM) of 3.47 $Å^3$ Da-1 and a solvent content of 64% [84]. This packing geometry is icosahedral and the monomeric asymmetric unit is consistent with the protein's behavior in solution as shown by the analytical ultracentrifugation and gel filtration.

3.2 Characterization of scaffolding protein (gp9) in prohead particles

3.2.1 Cryoelectron Tomography of T7 prohead particles

Grids of frozen hydrated wild type (wt) proheads (t7 5amber28) were scanned at low magnification for areas of ideal ice thickness and particle distribution, and upon locating suitable areas on the grid, tilt series were collected. Collection of tilt images was attempted over a range of -70^o to $+ 70^o$ in 1 degree increments for three tomograms. In some cases it was not possible to

Table 3

Tilt Series [a]	Tilt Range (^{o})	No. of particles	Resolution (nm) [b]
1	-69 to 67	59	4.8
2	-65 to 69	56	4.4
3	-64 to 70	91	4.3
4	-63 to 67	3	5.1
5	-66 to 65	4	4.6
6	-68 to 70	7	4.3
7	-64 to 68	4	4.3

[a] All tilt series were collected with a nominal defocus setting of 4 micrometers with a CTF first zero of 3.7 nm.
[b] In-plane resolution of the tomograms in nm was measured over three particles selected from each tomogram by the NLOO2D criterion with a threshold of 0.3.

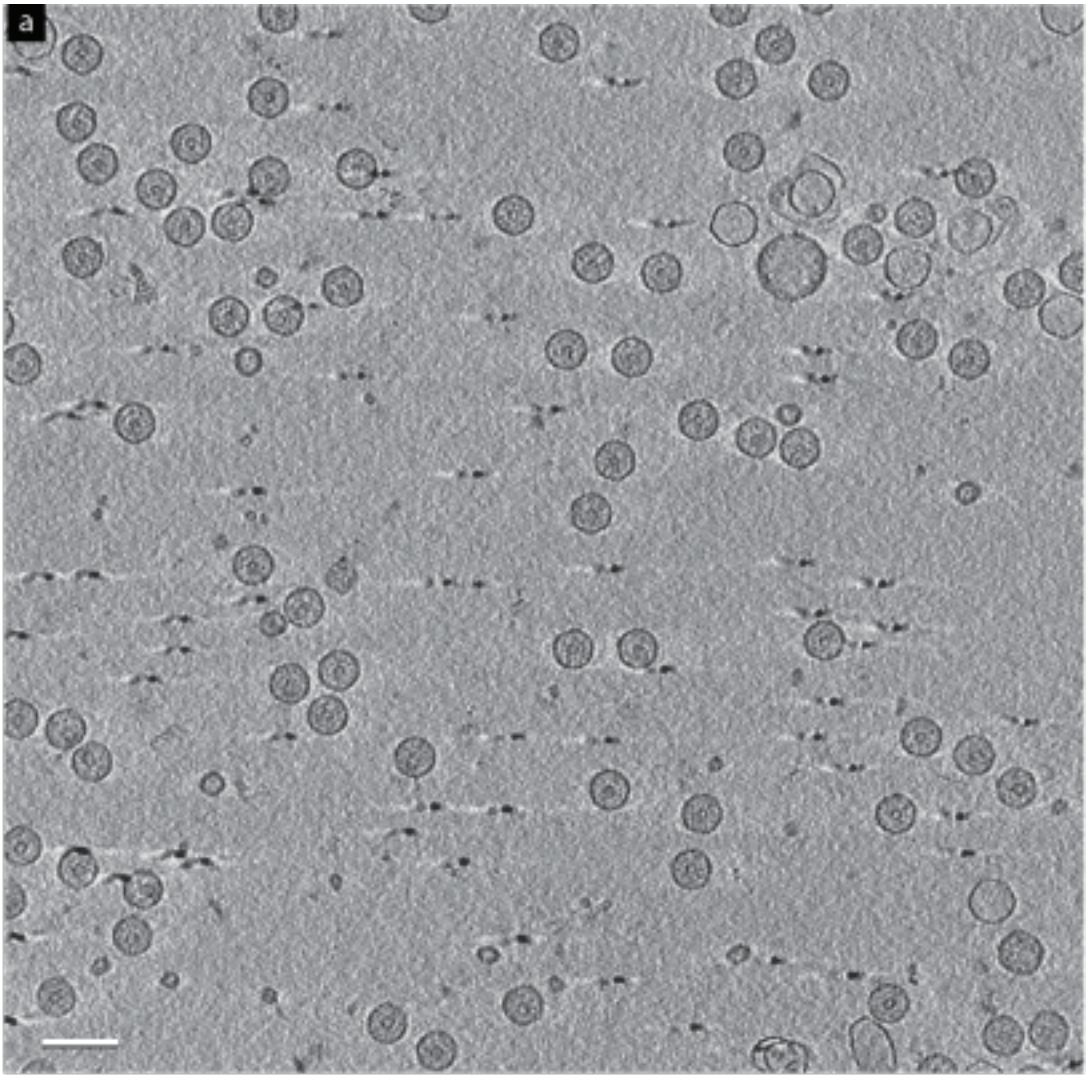
Tomograms 1-3 were collected on preparations of wt proheads that were purified by PEG precipitation and density gradient centrifugation. Tomograms 4-7 were collected on cell lysates, and as such have lower particle numbers than the tomograms collected on purified prohead preparations.

collect high angle tilt images, or in other cases those images were collected but found to contain little useful information and were discarded. The total electron dose for each tomogram was roughly 70 e-/ Å 2. The quality of each data set was assessed by the NLOO-2D method [78]. A summary of the angular ranges, resolution, and other relevant parameters for each tomogram are given in Table 3. Four sets of additional tilt series data were collected on cell lysates of *E. coli* following 5am28 infection for reasons that will be discussed later. Because these particles were not subjected to purification and concentration procedures, the number of particles per tomogram was much lower. The parameters for these tomograms can also be found in Table 3.

3.2.1.1 Tomograms of T7 wt proheads

From eight tomograms, 224 wt prohead particles were extracted and closely examined. A central section of a denoised tomographic reconstruction is shown in Figure 21A. The majority of the particles appear rounded with a central ring of density clearly visible inside of the capsid. This view of the prohead is commonly encountered when vitrified T7 particles are imaged by cryoEM [16, 54]. The cause is thought to be a hydrophobic portion of the prohead that preferentially migrates to the air-water interface [16] (Figure 21B).

In the tomograms, prohead particles were found to be roughly spherical with a diameter of 50nm (Figure 21). The prohead shells


a


b

Figure 21 a) Central section through a denoised cryotomogram of a field of wild type T7 prohead particles. There are also what appear to be broken and expanded T7 capsids visible in the field of view. The bar represents 100 nm. b) Schematic representation of T7 proheads oriented in an ice section.

measured 5 nm thick on average and displayed the characteristic corrugated surface of T7 procapsids previously studied [4, 47, 52]. The internal connector-core complex, a hallmark of the T7 system, was visible in the prohead particles in the tomograms. This complex is composed of the dodecameric gp8 connector and the core proteins gp14, gp15, and gp16. It is positioned at a unique five-fold vertex of the procapsid, the point of attachment for the DNA packaging machinery as well as the point of tail attachment in the mature phage [16].

3.2.1.2 Internal protein features of wild type proheads

Wild type T7 procapsids oriented in ice were grouped by eye into two primary views. The first view, into which the majority of the particles were positioned, presented the central channel of the connector-core complex parallel to the line-of-sight. Particles presenting this view were likely subject to the preferential orientation phenomenon mentioned above and found at both air water interfaces of the vitreous preparation (Figure 21 inset). These particles were referred to as "top views" and were easy to spot because of the appearance of a central ring of density attributed to the core that was visible in the central sections of the prohead (Figure 22). The second orientation, which was not as prominent in the data set, was referred to as a "side view" because it presented a view of the particle in which the connector-core

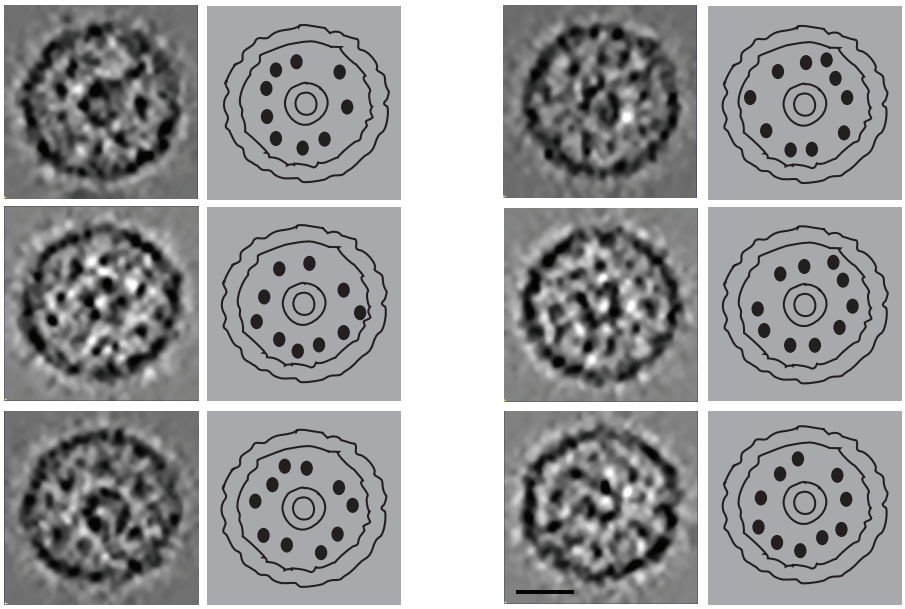


Figure 22 A gallery of "top views" of denoised wild type prohead particles extracted from a full tomogram. Each set of panels contains a 0.78 nm thick central slice of a prohead particle (left) with a schematic representation of that same particle (right). The prohead particles were extracted from a tomogram and then denoised by non-linear anisotropic diffusion. The connector-core channel is parallel to the viewing axis and surrounded by a roughly circular pattern of densites. The capsid, connector-core, and punctate internal densities have been indexed for each particle (right). The bar represents 20 nms.

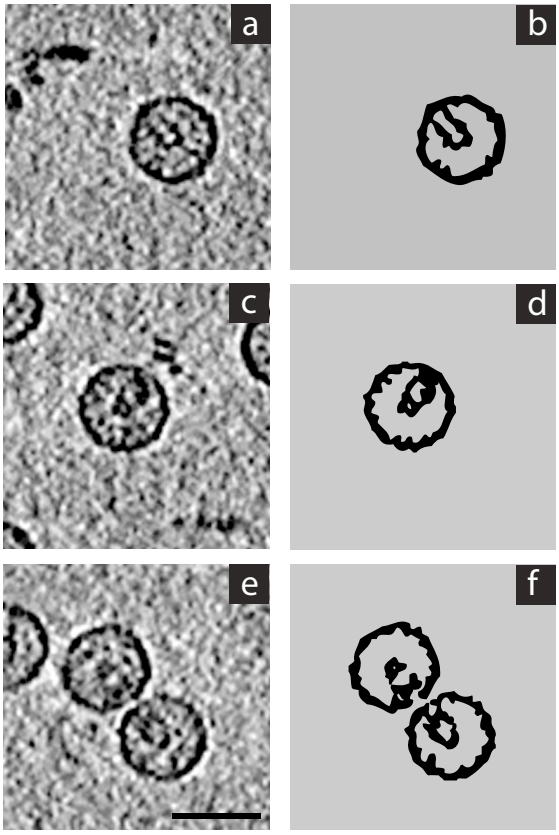

a
b
c
d
e
f

Figure 23 A gallery of "side views" of denoised wild type prohead particles extracted from a full tomogram. The central connector-core channel is roughly prependicular to the viewing axis in these 0.78 nm thick sections of the prohead particle (a, c, and e). The particles were extracted from a tomogram and then denoised by non-linear anisotropic diffusion. Schematic representation of each particle side view showing capsid and connector cores (b, d, and f). The particles in panel (e) form a "prohead dimer" likely through interactions between their connector-core complexes. The bar represents 50 nms.

channel was roughly perpendicular to the viewing axis (Figure 23). These side views were most distinct in particles whose connectors seemed to be interacting with connectors of adjacent particles (Figure 23). This association was also noted in previous studies of phage T7 by cryoEM [54]. There were also particles that could not be classified as top or side views by eye. These particles were likely oriented such that the connector-core axis was neither parallel nor perpendicular to the line of sight.

3.2.1.3 Analysis of punctate density patterns suggests an organized internal protein network

In addition to the connector-core complex, a circular pattern of discrete densities persisted at a radius of 15 nm from the center of the particle (Figure 22). This radius was just larger than that of the connector-core complex, but smaller than that of the capsid. In previous studies of single particle averages of T7 procapsids [52], a diffuse density seen at this radius had been attributed to scaffolding protein. In central sections of prohead top views extracted from tomograms, these patterns of densities were commonly made up of as many as 10 discrete spots with density values equivalent to those measured in the capsid protein region (Figure 22). The angular

spacing between adjacent spots was measured for the data set. The average angular spacing was found to be 34.6 degrees with a SD of 2.8. This would allow for 10.4 filaments in a full 360 degrees. When these punctate patterns were compared across the entire data set, it was apparent that the patterns were roughly similar.

The average of 93 particles from a single tomogram, aligned to a synthetic reference model, shows the icosahedral shell and connector core- complex at a single 5 fold vertex (Figure 24) surrounded by a diffuse density layer. This faint pattern of density present in the region of the procapsid corresponding to the spot patterns seen in individual particles is similar to what was observed when the particles were analyzed by single particle methods [52]. The strong signal for the capsid protein as well as the connector-core complex in the serial sections of the tomographic average indicates that particles were well aligned with respect to the capsid and connector-core. A comparison of the tomographic average and the single particle reconstruction [52] is shown in Figure 25. The two reconstructions are in good agreement with one another with respect to the capsid shell, connector core complex, and the diffuse putative scaffolding density.

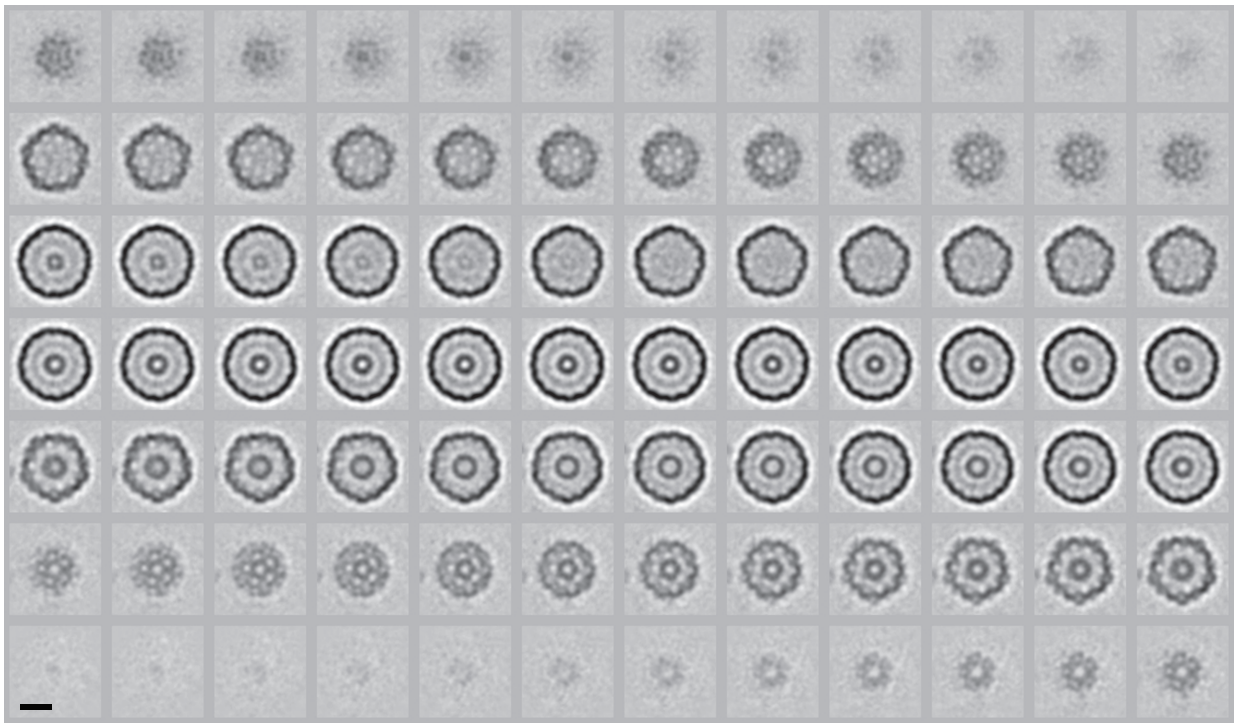


Figure 24 Average of wild type T7 prohead particles from tomography. Particles were extracted as sub-volumes from a reconstructed tomogram and aligned to an asymmetric reference using correlation analysis with no symmetry constraints imposed. The resulting average of 93 aligned particles is shown as a montage of 7.8 Å slices. The bar represents 200 Å.

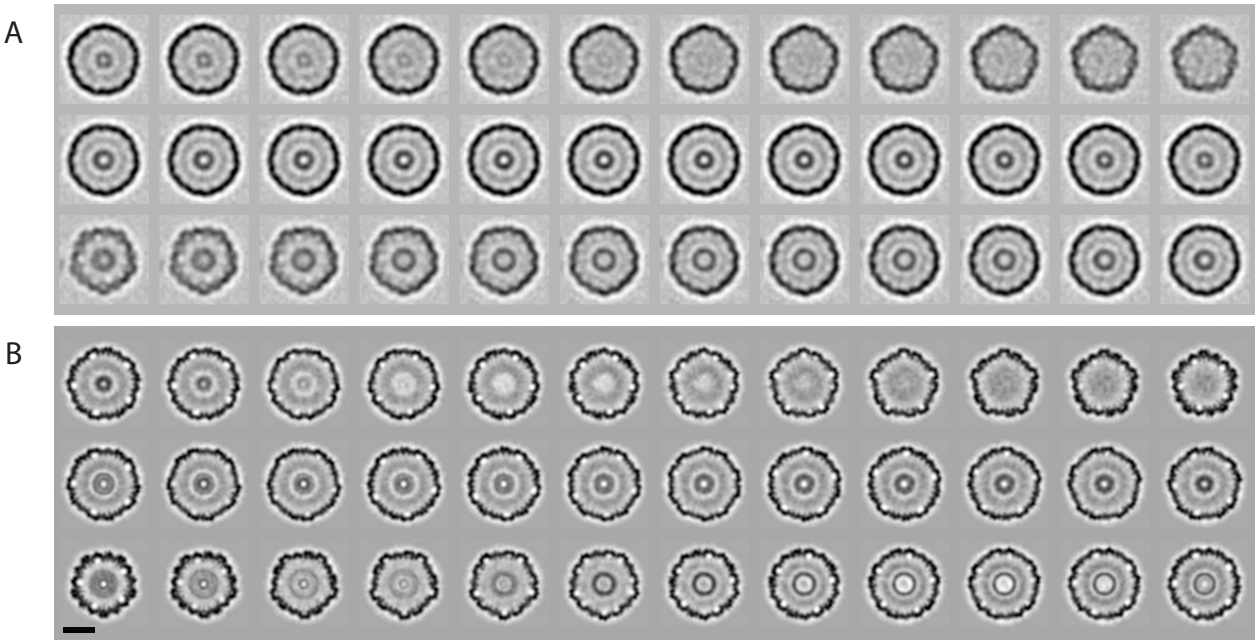

A
B

Figure 25 Comparison of average from tomography to a single particle reconstruction of wild type T7 prohead particles. A: The central 36 sections from the average of 93 aligned tomographic subvolumes as a montage of 0.78 nm slices. B: The corresponding sections of a single particle reconstruction of wild type prohead particles (Agirrezabala et. al 2005) filtered to 4 nm resolution. The bar represents 20 nm.

3.2.1.5 Extended filaments visualized in wild type proheads

Upon closer examination, the punctate densities seen in central sections were found to be cross sections of extended filaments. These densities persist through 10 nm of serial sections in the tomographic reconstructions (Figure 26). The filament densities can also be seen in individual particle side views. Because all proteins found in the T7 prohead with the exception of gp9 can be assigned to the capsid and connector-core [16, 47], the extended filaments are the only possible densities that can be immediately assigned to the scaffolding protein (gp9). Additional data presented in this work also supports this interpretation (see later).

The scaffolding filaments were manually traced and analyzed using Chimera's Volume Path tracer tool as described in Methods. Filaments were observed making connections with points on the inner surface of the capsid as well as connections with points on the connector core. A total of 2081 filaments were isolated from the data set of 224 particles. These filaments lengths were measured based on the cumulative distance of the marker models used for their isolation. A representative set of scaffolding filament traces are shown in Figure 27. The lengths of the filaments were found to be variable, and ranged

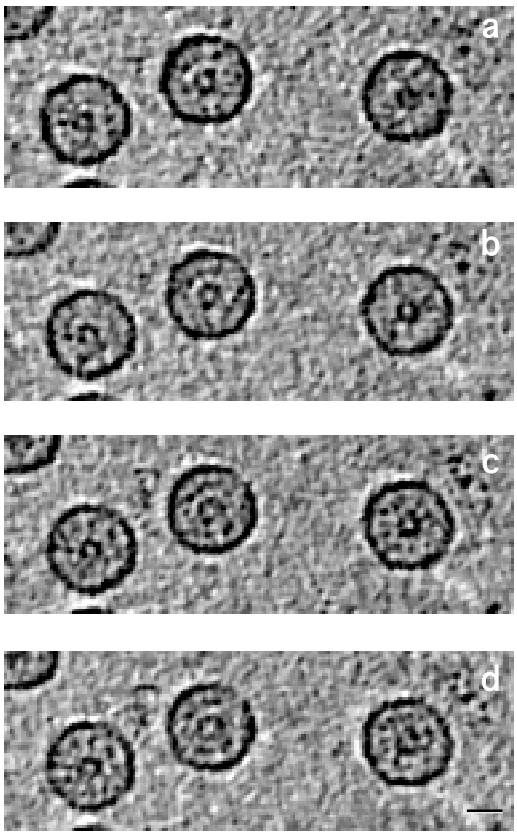

a
b
c
d

Figure 26 Serial sections through a cryo-tomogram of wild type T7 prohead particles. The images are 0.78 nm sections of a reconstructed tomogram that has been subjected to non-linear anisotropic denoising. The bar represents 20 nm.

from 6 to 30 nm with an average of 15nm $\pm$ 7nm (S.D.). The diameter was roughly uniform and measured 3nm on average. The average total filament length per particle was 140 nm. Based on the total number of particles that were analyzed, there were an average of 9.4 filaments per particle. These numbers are subject to the missing wedge effect, which is an intrinsic limitation of electron tomography. This "missing wedge" refers to the wedge-shaped region in Fourier space that contains no data due to the less than 180 degree angular range that was sampled [64]. In the case of the wt prohead particles, the average tomogram was sampled over 120 degrees of a possible 180. This means that it is possible that the filaments sampled do not represent the entire number of filaments from each particle, however they would represent a majority of filaments that are present.

3.2.1.6 Scaffolding protein filaments can make specific contacts with the connector-core complex

In wt prohead subtomograms, the internal connector-core extends roughly ~27 nm from one five-fold vertex of the capsid shell into the particle interior. In the averaged reconstruction of extracted and aligned wild type proheads, the connector-core ranged from 18 nm at its widest, to 11 nm in the thinnest region near its apex (Figure 24). In

individual subtomograms of prohead particles, filamentous projections emanate from the top of the connector-core and make contact with the internal surface of the capsid. These types of filament connections are represented by the blue traces in Figure 27. Scaffolding filaments are also seen contacting the base of the connector-core, the region composed of dodecameric connector (red traces in Figure 27).

3.2.2 Electron Tomographic Analysis of 9 10 heads

To further substantiate that the extended filaments seen in the wild type proheads were gp9, particles composed exclusively of scaffolding protein (gp9) and capsid protein (gp10) called 9-10 heads were prepared as described [34] and vitrified for analysis by cryo-electron tomography. Prior to vitrification, the sample quality was assayed by SDS PAGE to confirm the presence of gp9 in the preparation; by native agarose gel to confirm that the 9-10 heads were unexpanded gp9 containing particles [34]; and by negative stain TEM to visually confirm that the particles were intact.

Tilt series were collected in the same manner described earlier for wild type proheads. For this sample two tilt series were acquired by tilting the specimen over a range of -70 to + 70 in 1 degree increments at nominal defocus values between 4 to 6 ums. The angular ranges, electron dose, and other parameters for both tomograms are given in Table 4.

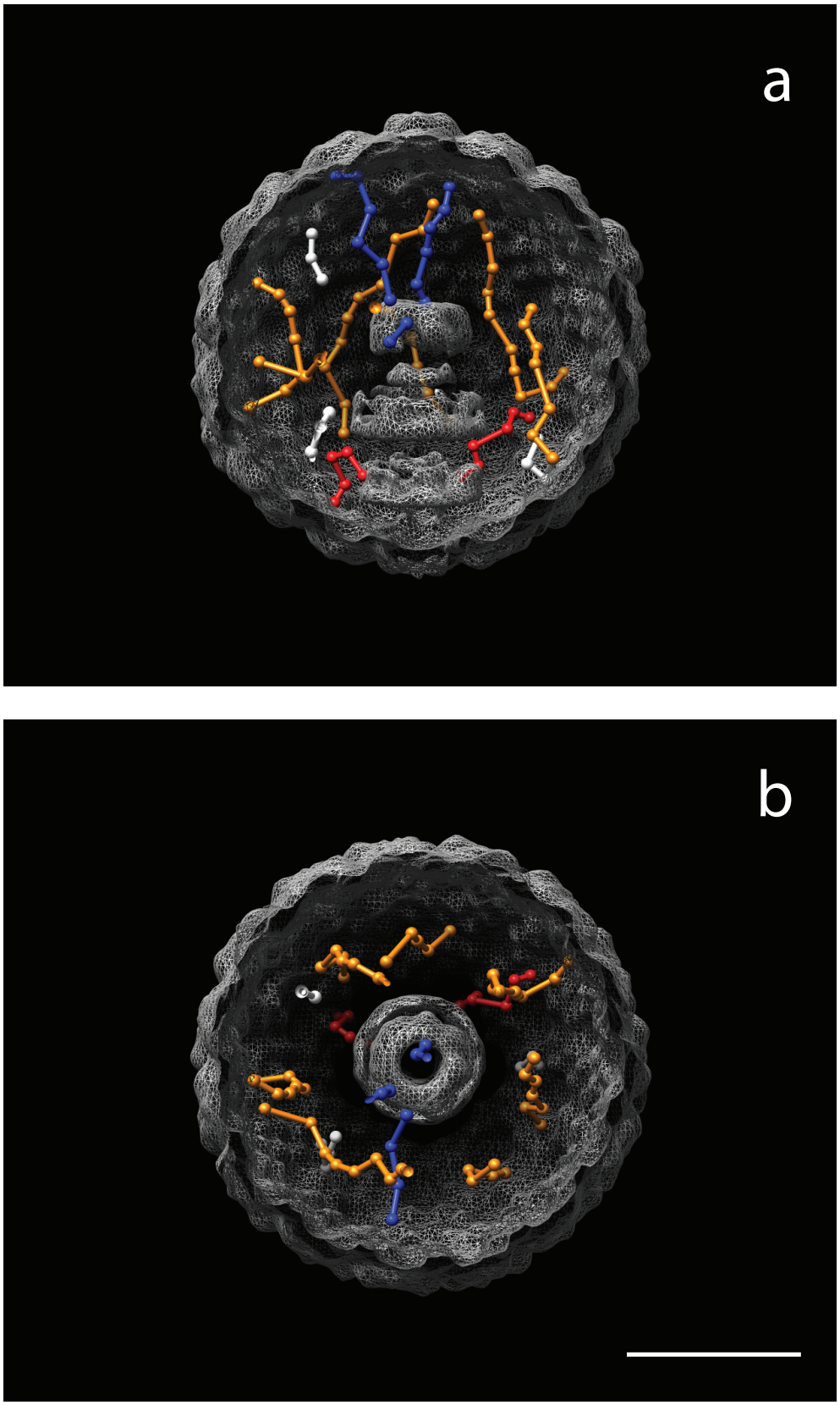

a
b

Figure 27 Three-dimensional tracings of extended scaffolding densities seen in individual tomographic reconstructions of particles that were aligned to a reference map of the procapsid(shown here as gray mesh). The front of the mesh capsid has been cut away to show the interior of the particle. a and b: The particle and traces are viewed (a) perpendicular to or (b) along the connector-core axis. Filaments can run parallel to the connector core axis and connect two points of the capsid inner surface (orange). Other filaments connect the capsid to the connector core (red and blue). Some small filaments are seen making only one contact with the inner surface of the capsid (white). The traces shown are representative of the types of filaments observed in the larger data set. The figure was created in Chimera and the filament tracings were done using the Volume Path Tracer tool in the same program. Bar is 20 nm.

Table 4

Tilt Series	Defocus [a] (μm)	Tilt Range (^{o})	No. of particles	Resolution [b] (nm)
1	-4 (3.7)	-67 to 69	88	4.3
2	-6 (4.5)	-60 to 68	34	4.4

[a] The nominal defocus setting is given in micrometers with a CTF first zero in parenthesis.
[b] In-plane resolution of the tomograms was measured over three particles selected from each tomogram by the NLOO2D criterion with a threshold of 0.3.

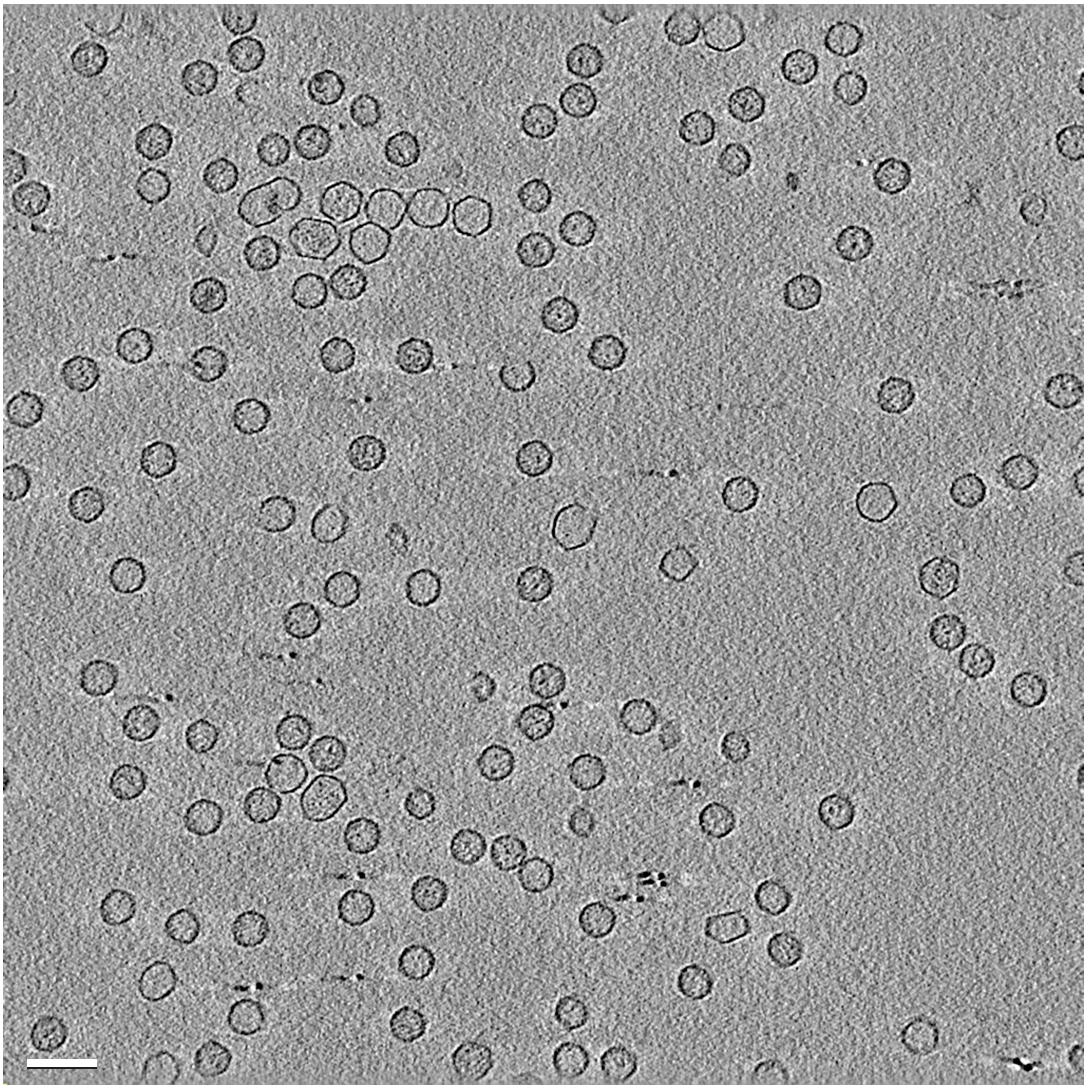


Figure 28 Central section through a denoised cryotomogram of a field of T7 9.10 head particles. The bar represents 100 nm.

Two tilt series were reconstructed using IMOD and a total of 122 particles were extracted from the resulting tomograms (Figure 28). 9.10 heads do not have an internal connector core complex and thus there is not a major feature to visually orient the particles to a specific vertex. Further, without a connector core at one vertex, the particles were free to randomly orient independent of the air-water interface of the vitreous specimen [16].

3.2.2.1 9.10 heads contain scaffolding protein (gp9) filaments

Despite this presence of an increased number of views, it was still possible to find 9 10 heads that displayed punctate densities similar to those seen in wild type proheads (Figure 29). The arrangement of these densities, however, was even more variable in appearance than patterns seen in the wild type proheads. This was likely due to the absence of the connector core, which in wild type procapsids occupies a substantial portion of the internal volume of the procapsid and may limit the number of possible arrangements for scaffolding.

3.2.2.2 A small subset of 9.10 heads contain trapped capsid assembly subunits

A small number of particles seen in the 9.10 head preparations contained densities inside the capsid that were roughly planar with the same thickness as the walls of the capsid (Figure 30). Upon closer examination, these densities appeared to be pieces of assembled capsid that were trapped in the interior of the particles based on the similarity

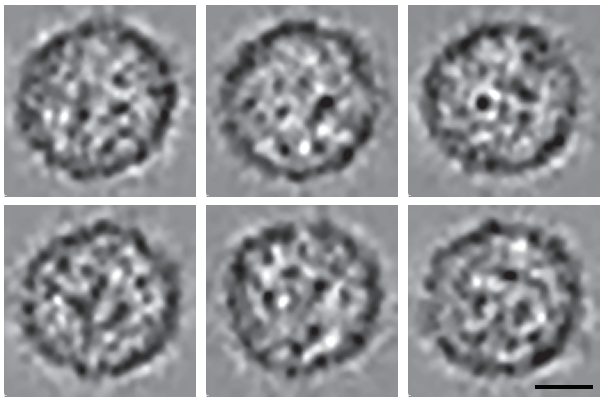


Figure 29 A gallery of denoised 9.10 head particles extracted from a full tomogram. The particles contain punctate densities similar to those seen in tomograms of wild type proheads. The particles were extracted from a tomogram and then denoised by non-linear anisotropic diffusion. The bar represents 20 nm.

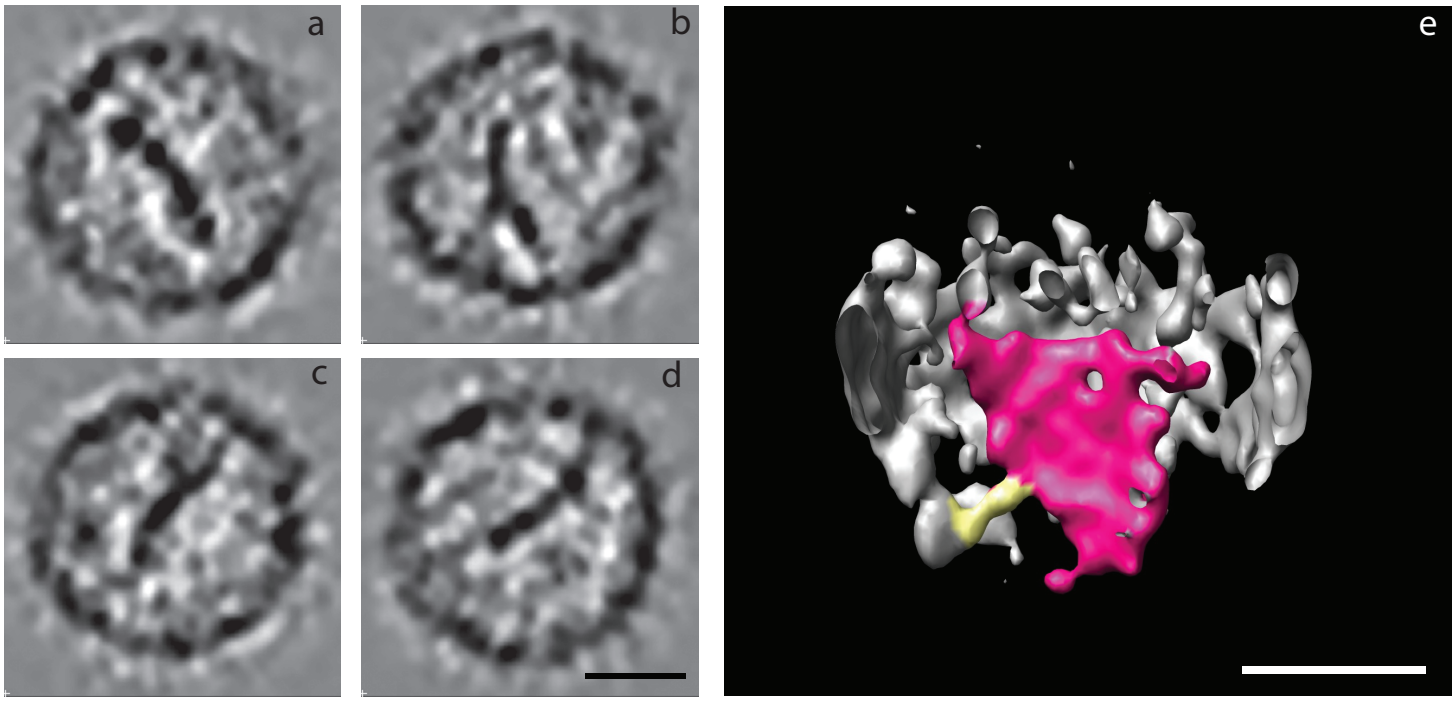

a
b
c
d
e

Figure 30 Tomograms of 9.10 head particles that contain plates of density that resemble the capsid shell in thickness and curvature. The particles contain large sheets of densities that appear to be trapped assemblies of capsid shell proteins. The particles were extracted from a tomogram and then denoised by non-linear anisotropic diffusion (a-d). Panel f is a surface rendering of one particle containing a large sheet of putative capsid protein. The internal density has been colored to highlight the capsid protein (pink) and a possible scaffolding filament (yellow). The bar represents 20 nm.

be linked to extended scaffolding filaments (Figure 30E). This phenomenon was not observed in tomograms of wild type proheads. 9.10 heads do not have an internal connector-core complex, and thus have additional internal volume to accommodate these putative trapped capsid assembly subunits.

3.2.3 Contact Point Analysis identifies specific regions of scaffolding interaction with the procapsid shell

Contact sites on the inner surface of the gp10 shell were more rigorously defined with respect to the asymmetric unit of the icosahedral capsid by a statistical procedure applied to the tomographic data. This procedure uses a modified averaging technique that better represents the signal from low occupancy regions of the larger structure. The analysis was performed for each particle in the data set and the average of these analyses were used to define regions of contact between the capsid shell and the scaffolding protein filaments. This technique revealed high instances of scaffolding connections at two locations on the internal surface of the capsid. One region of interaction was at the three-fold symmetry axis of the capsid. This axis is the most planar symmetry axis and a point of predicted high stability of the capsid protein subunits [47]. The second point for which there appears to be a high probability of interaction is at sites on hexamers adjacent to the five fold axes

of the procapsid shell. In both cases the interaction point seemed to be related to the "nubbins" of density that were previously noted protruding from the capsid's inner surface [4]. The contact points are indicated on the icosahedral surface by the colored zones (Figure 31).

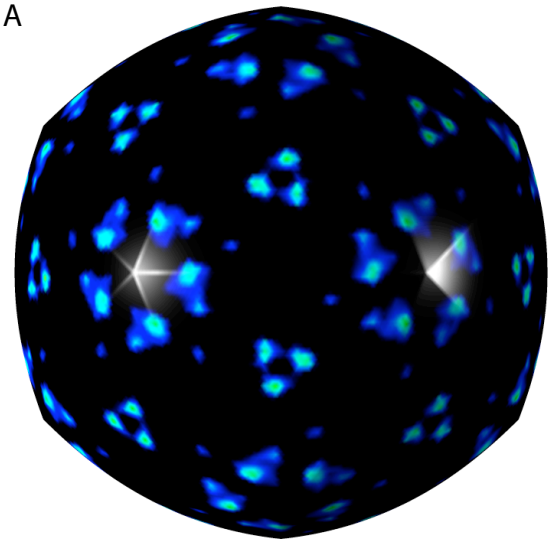
3.3 STEM Analysis of T7 procapsids, 9.10 heads, and mature phage

Analysis of the scaffolding protein filaments suggested that the number of filaments can vary from one procapsid to the next. We used Scanning Transmission Electron Microscopy (STEM) on wild type procapsids and 9.10 heads. Because the amount of capsid protein, the connector, as well as the core components is known [16, 60, 63], any additional mass would relate to scaffolding protein.

3.3.1 STEM data reveals broad mass distribution in wild type proheads

For the STEM mass analysis of wild type proheads, 973 prohead particles were selected from STEM micrographs based on criteria of intact shells and expected diameter (Figure 32). Expanded capsids, aberrant, and broken capsids (less than 5 % of the sample) were ignored. The average particle mass was 22.6 MDa. The mass of the capsid and connector-core were subtracted based on known molecular weights and copy number, and the remaining mass was converted to scaffolding protein copy number using the gp9 sequence

A



B

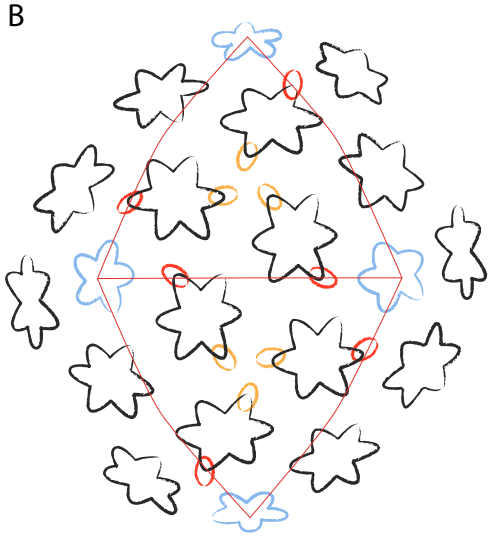


Figure 31 Statistical analysis of the tomographic data to locate scaffolding contacts with the capsid shell. A: Particles from the tomogram were masked to remove the core/connector complex and aligned to a symmetric reference particle. The sixty asymmetric units from an individual particle were compared to one another to find the maximum density value at each point. All particle results were then averaged to reduce noise. The resulting average asymmetric unit was propagated as an icosahedron and the result is displayed as an icosahedral slice. The surface of the icosahedron is colored based on the density values that cross it at each point. Density values have been colored such that light blue represents the highest density and the decreasing values are colored with darker shades of blue and finally black. B: The contact points of scaffolding filaments with the capsomers on two triangular faces of the icosahedron are represented by two sets of colored ellipses. Red ellipses indicate contact with a subunit adjacent to a five-fold vertex, and orange indicate contact with a subunit at the three-fold symmetry axis. Hexamers of capsid protein are shown in black and pentamers in blue. Two triangular faces of the capsid are outlined in red.

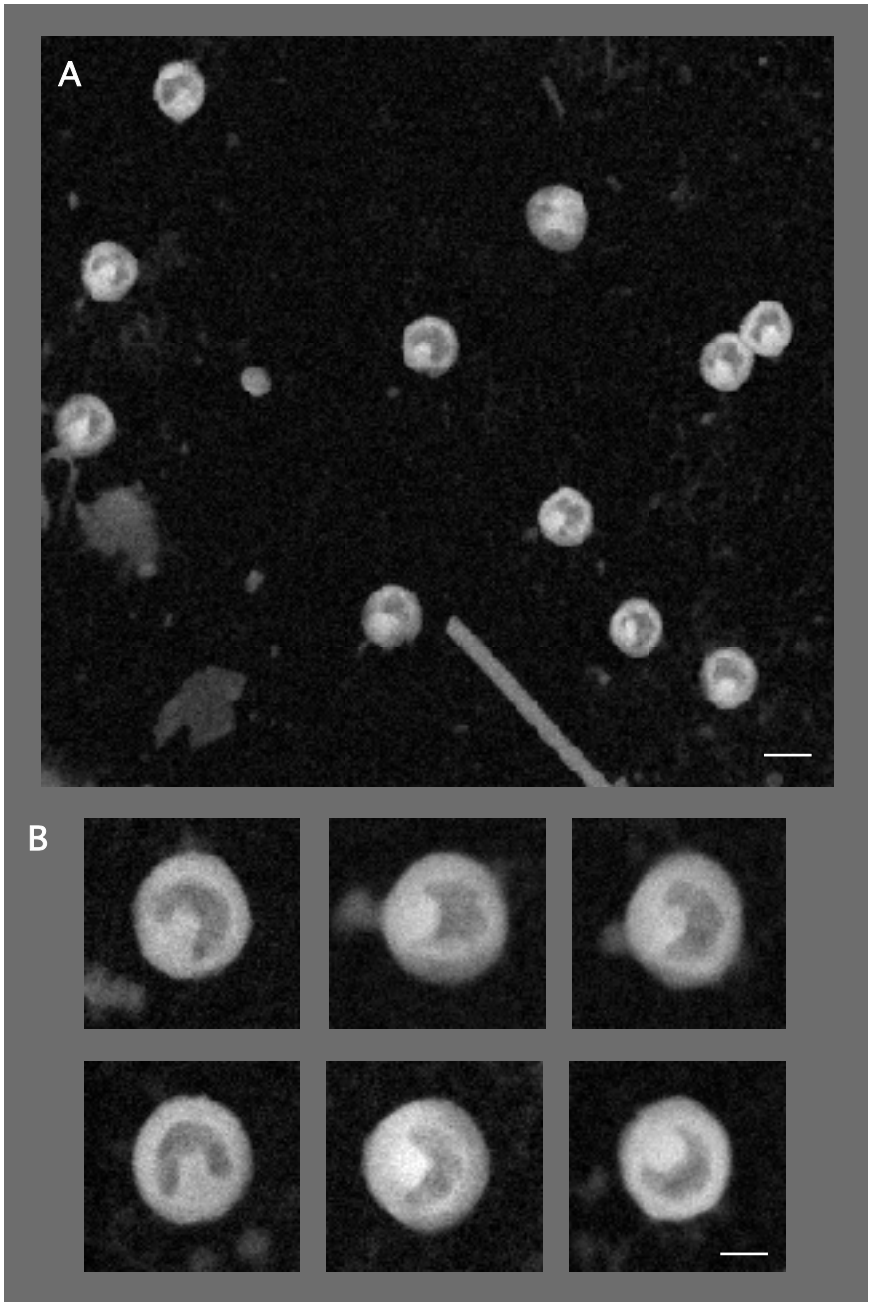
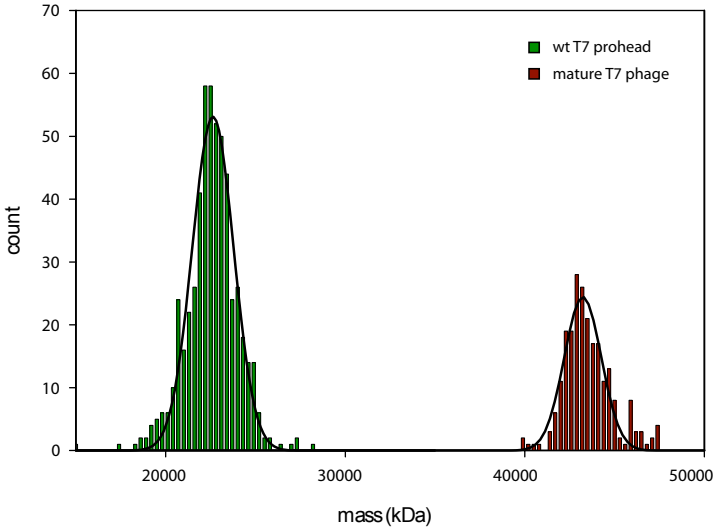

A
B

Figure 32 Scanning Transmission Electron Microscopy (STEM) of freeze dried wild type prohead particles. A: Representative STEM image of a field of unstained wild type prohead particles. The bar represents 50 nm. B: Gallery of individual particles showing thick capsid shell and density corresponding to a connector-core complex. The bar represents 25 nm.

calculated molecular weight of 33.8 kDa per monomer. Based on this analysis, the average number of copies of scaffolding protein per wild type prohead particle was 142.3. Particle mass measurements have a range of 8.9 MDa. The prohead masses were plotted as a histogram and fit by a Gaussian equation (Figure 33A). In order to rule out sample preparation or some other T7 protein as a source of variation, a control sample of mature T7 phage (particles that contain a full complement of capsid protein and connector-core proteins, but lack gp9) was imaged by STEM. A total of 228 mature phage particle masses were recorded and plotted as percent standard deviation so that the histograms of proheads and mature phage could be easily compared. Because the number of component molecules are understood to be the same from one mature T7 particle to the next [60], the variation of 2.3% from the mean seen in this sample is attributed to STEM measurement error. This error value is consistent with 2.2% the observed error for masses measured from the internal calibration standard TMV (not shown).

The masses of the wild-type proheads were found to vary by 5.1% about the mean value. This suggests that 2.8% of the variation in prohead masses is due to variation in the number of scaffolding protein molecules. The normalized distributions of the wild type prohead masses and mature phage masses are presented for visual comparison in Figure 33B. The difference in the width of the two peaks is interpreted as the variation in mass from prohead to prohead due to varying amounts of scaffolding protein.

A


wt T7 prohead
mature T7 phage
count
mass (kDa)
0
10
20
30
40
50
60
70
20000
30000
40000
50000

B

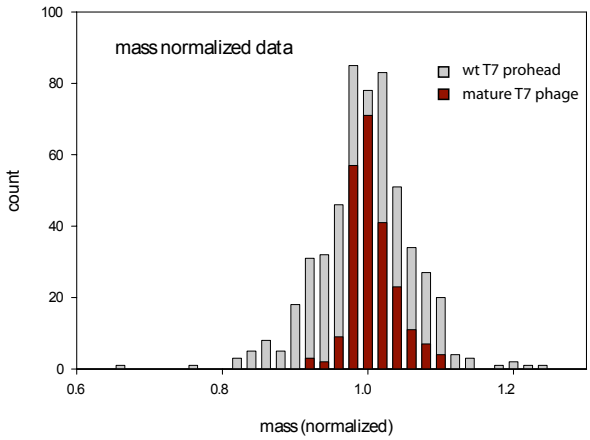

mass normalized data
wt T7 prohead
mature T7 phage
count
mass (normalized)
0
20
40
60
80
100
0.6
0.8
1.0
1.2

Figure 33 Scanning Transmission Electron Microscopy (STEM) mass analysis of T7 particles. A: Histograms of STEM mass data for wild type proheads (green) and wild type T7 phage (red). The histograms have been fit with a simple Gaussian curve as described previously [85]. The mean mass measurement for each data set is shown. B: Comparison of the spread of masses for particles containing scaffolding protein (wild type proheads in grey) and those lacking scaffolding protein (mature T7 phage in red). The plots and data analysis were done using Sigma Plot (SPSS Inc).

3.3.2 9.10 heads show similar mass distribution to that of wild type proheads

The 9.10 heads were also analyzed by STEM to determine their mass distribution and gp9 copy numbers. As with wild type proheads, 9.10 heads were selected for analysis based on the criteria of having intact capsids and having diameters consistent with those observed in previous structural studies. The masses of 321 unstained 9.10 head particles were recorded and normalized to the TMV internal control. These masses were converted to gp9 copy number and plotted in the histogram shown in Figure 34. The particles had a mean mass of 20.7 MDa corresponding to 149.1 copies of gp9 on average. The histogram of normalized measurements shows a broad distribution similar to that of the wild type proheads, indicating the same variation in scaffolding protein was also occurring with the 9.10 heads.

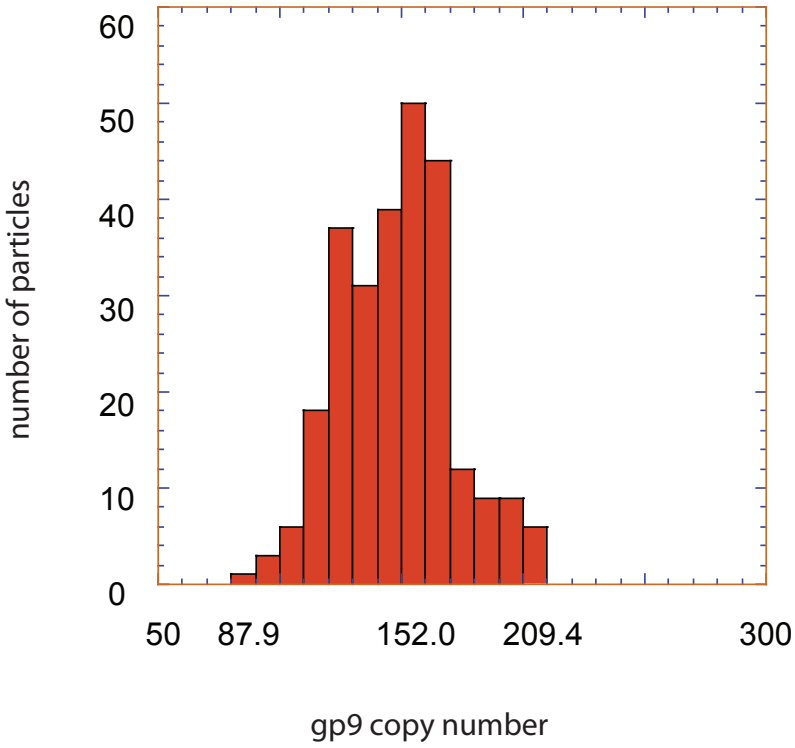

60
50
40
30
20
10
0
number of particles
50
87.9
152.0
209.4
300
gp9 copy number

Figure 34 Distribution of gp9 copy number for T7 9.10 heads as determined by STEM analysis. 270 particles were analyzed by scanning transmission electron microscopy. The masses of these particles were normalized and converted to gp9 copy number by subtracting the mass of 420 copies of capsid protein and dividing by the molecular weight of gp9 (33.9 kDa). The mean copy number for the data set was 149.5 gp9 molecules per 9.10 head. The plots and data analysis were done using KaleidaGraph (Synergy).

Chapter 4: Discussion

4.1 Gp9 in Solution

The T7 scaffolding protein (gp9) was expressed and purified and found to be a monomer in solution. Gp9 elutes on a high performance gel filtration column as a single, sharp peak. Although this peak could correspond to a protein with a molecular weight that indicates a compact gp9 dimer, a highly extended gp9 monomer, or some equilibrium between a population of dimers and monomers, analytical ultracentrifugation sedimentation velocity spins at three loading concentrations provided confirmation of the presence of a single discrete species of extended monomer. It is possible that the attempted sedimentation equilibrium studies would have unambiguously confirmed the monomeric state of gp9. The monomeric nature of T7 gp9 is different from that of other scaffolding proteins that have been studied, such as phi29 gp7 [38], P22 gp8 [40], and SPP1 gp11 [86]. These scaffolding proteins from other phage systems have all been reported to form oligomers (primarily dimers) in solution in the absence of other phage or host proteins. The results from this work indicate that purified T7 scaffolding protein does not oligomerize on its own. Perhaps in the presence of other phage proteins, such as capsid protein or other assembly factors, the scaffolding protein oligomerizes as part of the assembly process.

In addition to being an extended monomer in solution the T7 scaffolding protein gp9 has proven to be exceedingly soluble (this work and [34]). Despite these observations, the peptide does not appear to be natively unfolded. Purified gp9, when subjected to differential scanning calorimetric (DSC) analysis, underwent a broad melting transition (Figure 12). This melting transition was readily detectable at multiple protein concentrations, and there were no exothermic events detected in the scan (an indicator of protein aggregation). Upon repeated scans of the same gp9 sample (Figure 13) the melting event persisted and its amplitude was only marginally diminished. These results indicate that the reaction was reversible, and once the sample was cooled, the protein reverted to its original conformation. These findings are supported by previous accounts of gp9's ability to remain soluble in boiled cell extracts [34]. The reversible nature of gp9 folding is consistent with an extended molecule with little or no complex globular structure. If this were the case, it could easily refold during the time between calorimetric scans. In contrast, similar calorimetric studies of the major capsid proteins of different phages (those proteins equivalent to gp10) have shown that these head proteins are unable to refold once they have been denatured [68, 87, 88]. Gp9, on the other hand, is very stable and can independently refold. Perhaps it is this property that allows the scaffolding protein to serve a chaperone function for the unstable [34], compact and globular [47] capsid protein gp10.

The DSC melting profile for purified gp9 could be modeled by a non-two state model in which overlapping endotherms occur at T_ms of 49.4 ^oC and 62.2 ^oC. Because purified gp9 is not oligomeric in solution, the melting event seen in the calorimetry scans was attributed to the denaturation of at least two regions within the gp9 monomer. The two component events of the broad endotherm could involve the separate unfolding of two roughly equivalent domains, or perhaps one large domain and two smaller domains. Based on the total change in enthalpy for the event, ~136 residues were involved in the transition (assuming 10.5 kcal/mol of amino acid [89]). Further structural and biophysical characterizations including circular dichroism, NMR, and additional calorimetric studies would be needed to determine the exact nature and number of residues involved in this transition.

A variety of secondary structure prediction algorithms (see Methods) were applied to the gp9 sequence and their predictions were in excellent agreement regarding the presence and location of particular secondary structural elements (Figure 15). These predictions suggest that greater than one half of the gp9 sequence forms alpha-helical stretches separated by short segments of random coil. Less than five percent of the 307 residue sequence is predicted to contain beta elements. On the whole, the T7 scaffolding protein gp9 is predicted to be primarily alpha-helical, which is also the case for the phi29 [38] scaffolding protein gp7, and the P22 [90] scaffolding protein gp8. These

proteins perform similar functions in their respective phage systems despite an obvious sequence homology.

The secondary structure predictions were in agreement with a series of proteolysis assays that were performed on purified gp9 (Figure 18). The protease sites in the sequence predicted to be involved in the formation of secondary structural elements remained uncut while sites falling in regions of predicted random coil, were cleaved. Taken together, the analytical ultracentrifugation, DSC, structural prediction and protease digestion experiments suggest that gp9 is divided into three putative domains separated by stretches of flexible linker sequence. These three domains are:

1) The N-terminal domain

The N-terminal portion of the protein (residues 1 - 75) contains all of the 16 residues predicted to adopt beta structure. The fact that all protease sites in this region remained uncleaved, coupled with the high occurrence of hydrophobic residues (greater than 40% of the amino acids in this region), suggests a globular domain. This property could explain why hexahistadine tags appended to the N-terminus reduced expression levels of full length gp9 by > 80 percent (Figure 7), and appending the same type of HisTag to the extended C-terminal region did not decrease gp9 expression levels relative to those of the wild type protein. This would also imply that the N-terminus of the protein

is buried within a folded core, and not accessible to the solvent, an interpretation that is further supported by the fact that only denatured 9NHis was able to bind the NiNTA column. The limited proteolysis of N-terminal domain, the high percentage of hydrophobic residues, the observed binding behavior of HisTags appended to this region, and its sensitivity to modification suggest that this domain is likely globular and perhaps more compact than other regions of the protein.

2) The C-terminal domain

The high instance of protease sensitive sites seen in the C-terminal domain (251 to 307) along with the predicted irregular secondary structure profiles suggests that this domain is tethered to the rest of the protein by a solvent-exposed flexible linker region. The flexibility of this domain with respect to the core of the protein was further supported by the fact that its removal did not diminish the molecule's elongated profile as evidenced by the velocity sedimentation studies (see Table 2). The C-terminal domain contains two predicted helices with an arrangement of helices and coils similar to what has been observed in the scaffolding protein of phage P22 (another example of a dsDNA phage that forms a T=7 capsid) [90]. In the P22 system, this domain was essential for capsid binding functions of the scaffolding protein [90].

3) The central alpha-helical core domain

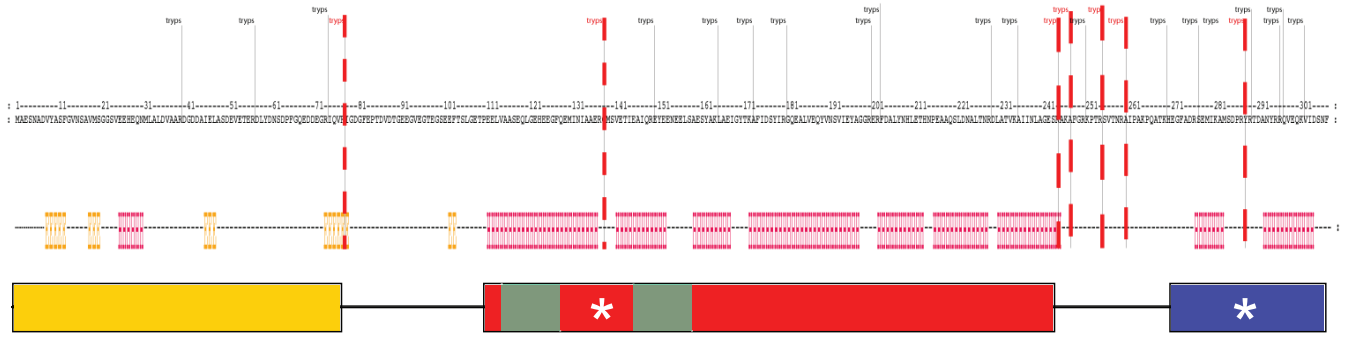
Eighty-five percent of the sequence in the middle of gp9 (from residue 110 to residue 242) was predicted to be alpha-helical. The axial ratios derived from the velocity sedimentation data and the calculated partial specific volume of the protein based on sequence were used to suggest that the gp9 monomer is a prolate molecule whose longest dimension is roughly 100 Å or 10 nm. The axial ratio for the truncated form is slightly higher than that of the full-length molecule suggesting that removal of 51 C-terminal residues did not diminish the protein's extended profile (Table 2). Therefore it is likely that this alpha-helical core domain is primarily responsible for the molecules extended profile in solution. The length of this domain, based on a single continuous helix for the predicted helical content, would be approximately 150 Å or 15 nm. However, the helical motifs in this domain are separated by short stretches of random coil based on the sequence prediction. A single protease site (R137) in this region is cleaved, possibly indicating a "turn" that occurs between two adjacent helices. Heptad repeats were also found in this region (Figure 16) indicating the possibility of intra-molecular interactions (see below). Were gp9 to fold back upon itself in this region, its predicted helical profile would be in good agreement with the prolate ellipsoid or cylinder shape derived from the analytical ultracentrifugation.

Evaluation by the COILS algorithm indicates gp9 contains heptad repeats in the alpha-helical core domain (Figure 16). Heptad repeats are a sequence motif that consists of a repeating pattern of hydrophobic and polar amino acids, and the periodicity of these amino acid side chains dictates the helical interaction face between multiple α helices. Heptad repeats are indicators of a protein's predisposition to form alpha-helical coiled-coils [82]. The helices in coiled-coil interactions can either be anti-parallel or parallel, and can involve either intra- or inter-molecular interactions. Given the monomeric state of gp9 in solution, anti-parallel α helices could be involved in formation of intra-molecular alpha-helical coiled-coil [91]. Protease accessibility data along with the profile derived from the velocity sedimentation data suggests that the polypeptide folds back on itself via a flexible region between two of the heptad repeats (Figure 35).

Model for gp9 in Solution

Based on the results from this work, as well as previous studies of gp9 and other viral scaffolding proteins, a model for gp9 in solution is presented. The results of the analytical ultracentrifugation, secondary structure prediction, the discrete protease digestion profiles, as well as the DSC, can be interpreted as arising from a protein with three distinct domains separated by stretches of random coil sequence (Figure 35A). The N-terminal domain is likely compact

A



B

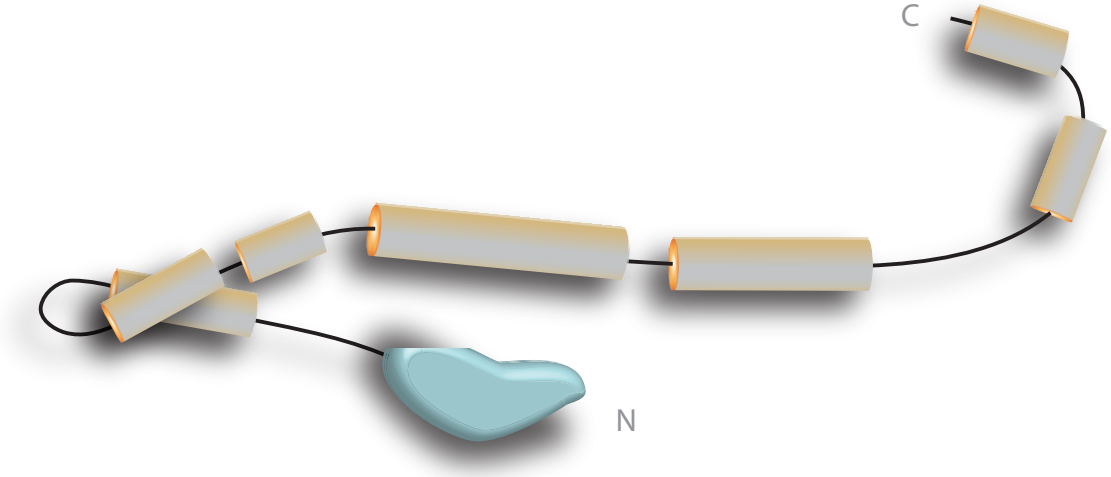


Figure 35 Model for gp9 monomer in solution. **A)** The gp9 sequence is displayed along with the results of the secondary structure prediction and the proteolysis assays. A schematic for putative domains is given below the structure prediction. The domains are: globular N-terminus (yellow), alpha-helical core (red), and C-terminal helix turn helix (blue). The protease sites that were cleaved within a domain are represented by * (white), and the regions of heptad repeats detected by the COILS algorithm (Lupas et. al 1997) are shown as green highlights in the alpha-helical core domain. **B)** A cartoon model for gp9 depicts a compact globular N-terminus (cyan) attached to a highly extended, primarily alpha-helical molecule that folds back onto itself via an intra-molecular coiled-coil.

(possibly globular) as it resists proteolysis, has a high hydrophobic content, and is sensitive to modification (based on the expression profiles for the N-terminally modified clones of gp9). The majority of the C-terminal residues were predicted to adopt irregular secondary structure with the exception of two short helical stretches, which could form a helix-turn-helix motif similar to the capsid-binding regions seen in the scaffolding proteins of phi29 [38] and P22 [90]. The alpha-helical core domain of T7 scaffolding protein contains heptad repeats that were detected by the COILS algorithm [83]. The only cleaved trypsin protease site found in this region falls in a stretch of predicted coil between two sets of heptad repeats (Figure 35A). Based on the position of this protease accessible region and the fact that gp9 was found to be a monomer in solution, I propose that gp9 forms an intra-molecular coiled-coil structure in the monomeric state as depicted in Figure 35B. In this model, the entire molecule would exist as an extended monomer that folds back onto itself by formation of an anti-parallel, intra-molecular coiled-coil. These types of self-interactions could explain why purified gp9 does not appear to oligomerize via a coiled-coil interactions in solution. However, the formation of higher-ordered oligomers of gp9 is possible (see later) and could be triggered by assembly-related interactions with the T7 capsid protein (gp10). The predicted α helices containing the heptad repeats of gp9 could conceivably switch from intra- to inter-molecular coiled coils when T7 capsid protein is present. This type of switching between parallel and anti-parallel coiled coil in synthetic peptides has been demonstrated and characterized

previously [92]. Since conformational changes in capsid protein have been shown to be involved in the maturation stage of the virus life cycle [9], conformational changes in the T7 scaffolding protein may play an important role in the capsid assembly process as well.

The alpha-helical topology mixed with coiled elements seems to be a common feature of scaffolding proteins in dsDNA phage [38, 90] as well as in scaffolding proteins of animal viruses like VP22a of HSV-1 [93]. These primarily alpha-helical motifs are also seen in the delta domain of HK97 gp5 [10, 11], which is a domain of the capsid protein that acts as a covalently fused scaffolding protein during assembly [9]. A previous study by Agirrezabala et. al [52] has proposed a model for the scaffolding protein in which it forms an inter-molecular alpha-helical coiled-coil dimer. This model is based exclusively on sequence-based prediction and structural modeling methods that were guided by the phi29 scaffolding protein dimer. I did not find any evidence that such a dimeric species would arise from T7 scaffolding protein.

4.2 Crystallization of gp9

Preliminary crystallization trials of gp9 yielded small crystals that were slow growing and difficult to optimize, but in due course diffraction quality crystals were obtained (Figure 19). Unfortunately, the diffraction data (4.3 Å) were not of high enough resolution to obtain a density map suitable for

unambiguous structure determination, however some important details have emerged.

Indexing of the diffraction patterns from T7 scaffolding protein crystals indicated an uncommon packing geometry (Figure 20). The molecules in the crystal were arranged in an F 2 3 space group (icosahedral), which has only been reported for 27 out of 53,660 structures deposited in the Protein Data Bank (www.pdg.org -[94]). Based on Matthew's coefficient calculations and gel analysis (not shown), the crystals were found to contain one monomer of gp9 per asymmetric unit with 48 asymmetric units in a single unit cell. This means that the gp9 molecules in these crystals are making two distinct types of intermolecular contacts (2-fold and 3-fold symmetry contacts). These crystal contacts alone are not sufficient to imply biologically significant gp9 interactions. However, it is worth noting that a molecule responsible for directing the assembly of an icosahedral capsid is capable of adopting this unique space group geometry in crystalline form.

The primary role of gp9 is to interact with capsid protein (gp10) to facilitate capsid assembly [17, 34]. Unlike scaffolding protein, the T7 capsid protein (gp10) is difficult to purify because it rapidly forms insoluble aggregates *in vitro* [95]. In general phage capsid protein folding is not reversible as has been shown by DSC studies in P22 [87], HK97 [68], and T4 [96]. In T7 infection, scaffolding protein expression is triggered in the cell prior to capsid protein

expression [53] and available scaffolding protein could interact with newly synthesized capsid protein in a manner similar to what has been observed in *in vitro* assays of mixtures of the purified native proteins [34]. The highly soluble, reversible and monomeric gp9 could be thought to balance the irreversible and highly reactive gp10 during capsid assembly.

4.3 Gp9 in Proheads and 9.10 heads

The T7 prohead and related procapsid particles have been studied in detail by cryo-electron microscopy (cryoEM) and single particle analysis (SPA) [4, 47, 52]. Single particle methods provide three-dimensional structural information by averaging images of many identical complexes, and are well suited to the study of rigid regular complexes like virus capsids. In the case of T7, recent cryoEM studies using SPA have resulted in pseudoatomic map of the T7 procapsid [47], as well as a lower resolution but more comprehensive picture of the non-icosahedral core and connector components within the procapsid [52]. Unfortunately, these studies have only provided limited information about the organization of scaffolding protein in the procapsid. This is often the case for complexes that are pleomorphic such as influenza virions [97] or highly polymorphic capsids such as those of retroviruses [98]. Cryo-electron tomography (cryoET), on the other hand, can provide three-dimensional structural information about a frozen hydrated complex regardless of its regularity. CryoET involves imaging the same specimen at multiple angles

and combining those images into a three-dimensional reconstruction. A comprehensive review of the techniques and applications of cryo-electron tomography can be found [64].

When examining T7 proheads by cryo-electron tomography (cryoET), scaffolding protein (gp9) was found to form extended filaments of variable length (6 to 30 nm) and uniform diameter (3nm) in the capsid interior. These filaments were also seen in wild type prohead particles that were imaged by cryoET immediately after cell lysis eliminating the possibility of the filaments arising during the purification process.

Gp9 makes up 20 percent of the protein composition of the procapsid, the core 10 percent, and the capsid shell the remaining 70 percent, according to STEM mass measurements and SDS gel quantitation (this work and [34, 53, 60]). High-resolution cryo-electron microscopy (cryoEM) using single particle analysis (SPA) has confirmed that the capsid protein alone is sufficient to account for all the density present in the capsid shell of reconstructed proheads [47]. This rules out the possibility of scaffolding molecules being integrated into the capsid shell. The protein components of the connector and core have also been convincingly assigned in previous studies [16, 63]. Previously reported asymmetric SPA reconstructions [52], as well as the tomographic average of the T7 prohead (Figure 24) are in good agreement and, show that the capsid and core are clearly visible and defined. Therefore, the only extra densities inside

the prohead that are attributable to scaffolding protein are the filamentous structures, which were present in tomograms of purified wild type prohead particles. All evidence from this work along with findings from previous studies, point to gp9 as the source of these filaments.

In central sections of the procapsids from tomography, as well as in cryo images of procapsids projected down the central axis of the connector-core [16] scaffolding filament cross-sections formed distinct circular patterns. The $\sim 35^{o}$ angular spacing of these filament cross-sections dictates that 10 filaments would cross plane of the prohead seen in the central sections of the tomographic volumes. This suggests that the scaffolding filaments are related to the five-fold symmetric capsid lattice. This seems likely as many of the filaments in individual tomograms made contact with the internal surface of the capsid and extend into the procapsid space between the connector-core and the capsid shell (see below). It may be gp9's interaction with gp10 during the course of assembly that leads to the formation of the extended scaffolding filaments seen inside the procapsid.

To better understand the scaffolding protein's role in assembly, the specific points of contact that the scaffolding makes within the completed procapsid shell were determined. This was made possible by the use of a novel averaging technique (see Methods) on our tomographic data (Figure 31A). By analyzing the individual particle sub-tomograms and then averaging the

results of the entire analysis, it was possible to localize scaffolding filament contact points that would otherwise not have been visible by standard averaging methods. The contact points were localized to the three fold axes of the triangular faces as well as to points on the inner surface immediately adjacent to the pentameric vertices (Figure 31B). Previous high resolution cryoEM and molecular modeling studies [47] have indicated that the threefold symmetry axis position in the T7 procapsid lattice is a point of relatively high stability with respect to the local gp10:gp10 interactions. It is also the interface between three hexamers of gp10 that make up the triangular face of the T=7 icosahedral capsid. This site could be an important contact point for capsid and scaffolding protein subunits during capsid assembly (see later).

Model for gp9 as Filaments

The STEM and tomography data suggest that the filaments of gp9 are composed of multiple copies of scaffolding protein. The tomography data indicates that there are an average of 9.4 filaments per particle, which is in good agreement with the total number of scaffolding filaments (10) derived from the angular spacing of the circular filament patterns of the prohead. This data also give an average of 140 nm of filament length per particle. Taken together with the copy number results from the STEM analysis (an average of 142 gp9 molecules per prohead), this relates to one molecule of gp9 per ~1 nm of filament length. Based on the extended profiles modeled from the sedimentation velocity data (10 nm wide ellipsoid), full-length gp9 would

have a significant overlap with itself in order to satisfy the 1 nm of filament length per molecule packing arrangement. Further experiments would be required to form a more detailed picture of the organization of the gp9 molecule within the filament.

The results from the contact point analysis of the filaments seen in assembled procapsids suggest that extended gp9 molecules are bound to the inner capsid surface at the icosahedral three-fold symmetry axis. When bound to the capsid at these three-fold sites, multiple molecules of gp9 are in close proximity. The heptad repeat containing alpha-helical core domains of these adjacent scaffolding protein molecules could oligomerize via an inter-molecular coiled-coil to form the extended filaments seen by tomography. The 3nm diameter of scaffolding filaments observed in tomograms would accommodate such an oligomeric bundle (perhaps trimeric) formed by adjacent scaffolding molecules at the three-fold symmetry axis of the assembled capsid shell. This would require the allosteric conformational switching of coiled-coils from intra-molecular to inter-molecular, a mechanism that has been proposed for many biological signaling processes including viral glyco-protein membrane fusion [99]. It is possible that gp9 utilizes such a switch creating a means by which gp9 can go from monomer to oligomer during prohead assembly. This switch is likely dependent on the presence of capsid protein as such an oligomerization does not occur during the course of DSC experiments.

Scaffolding protein filaments also make specific points of contact with the internal connector-core complex. This is the roughly cylindrical hetero-oligomeric complex that is required for (among other things) efficient assembly in T7 (see earlier). Scaffolding filaments make contact at the regions that correspond to the gp8 connector ring, as well as at the gp16 tetramer at the apex of the core. Many of these filaments join the connector-core complex to the internal surface of the capsid. It is likely that these filaments are responsible for the contacts that are seen around the five fold vertices. The notion of scaffolding interacting with the connector/portal protein is not unique to the T7 system. These interactions in T4 [100], HK97 [6], phi29 [101], SPP1 [102] and HSV1 [103] have been implicated as means of nucleating capsid assembly at a five-fold vertex.

4.4 Proposed Mechanism

In light of the results from this work and taking into account the observations from related research on T7 and other viral systems, a possible mechanism for T7 procapsid assembly is offered.

Nucleation and Assembly-

In the cell, extended scaffolding monomers could interact with capsid protein to activate both molecules for assembly. Such gp9:gp10 interactions were reported previously between the purified proteins in

T7 in vitro assembly experiments [34]. These interactions may prompt the switch of gp9 from a monomer to an oligomer and in the process stabilize a complex of capsid and scaffolding protein to create an assembly center. These hetero-oligomeric assemblies could be further organized into a procapsid nucleation complex via scaffolding protein's interactions with connector-core. These proposed interactions between the scaffolding protein and the connector-core complex would explain the significant number of scaffolding filaments seen in the tomograms that were associated with both the capsid protein and the connector-core (Figure 27). As further support, the number of aberrant structures seen in T7 infections of mutants that lack any of the connector core proteins [17] and *in vitro* assembly reactions of the capsid and scaffolding protein [34] is greatly increased relative to wild type T7 infections.

This type of connector-based nucleation has been proposed elsewhere in the literature for T7 [52] and other phage systems such as HSV-1 [104], phi29 [101], and HK97 [6].

Additional scaffolding bound capsid protein molecules could interact with those capsid protein molecules already associated with the connector core. These interactions could be stabilized by the set of scaffolding contact sites seen at the three-fold symmetry axis of the triangular faces (Figure 27). The coming together of scaffolding

proteins at this three-fold site in the capsid could trigger a switch of scaffolding protein molecules that form the filaments seen in the tomograms. These filaments could form as a means for stabilization of the assembly of the newly polymerized competent building blocks. This theoretical intermediate may resemble the bowl-shaped partial capsid assemblies observed in P22 [37] and HSV-1 [103] systems.

The scaffolding contact points that guide assembly through the steps following nucleation are located at the centers of the triangular faces of the icosahedral shell and points near adjacent pentameric vertices. Each scaffolding connection can be mirrored by a similar contact point on the opposite pole of the capsid, thus providing a means for establishing the proper handedness and curvature without the need for filling in all available scaffolding contact points. This redundant assembly scheme is attractive as it takes into account the flexible nature of the observed scaffolding filaments, the contact points on the capsid shell, as well as the variable copy number of scaffolding protein molecules derived from the STEM mass measurements.

Several variations of assembly mechanisms exist for other dsDNA phages (i.e. HK97, P22, and Lambda) that form T=7 capsids, however none of them are

able to explain all aspects of the Bacteriophage T7 system. For instance, in the Hong Kong 97 (HK97) system, gp5 capsid protein molecules are assembled into hexamers and pentamers for prohead assembly [105]. HK97 capsid protein has a scaffolding protein like domain that acts as an assembly chaperone and is then cleaved prior to maturation, whereas T7 has a separate scaffolding and capsid proteins. In P22, a mechanism based on the assembly of scaffolding activated monomers of capsid protein driven by oligomeric scaffolding protein is proposed [106]. However, P22 lacks the large core complex of T7 and thus appears to rely on a more robust set of scaffolding:capsid protein interactions than those seen in T7.

The connector-core nucleated mechanism is not the only path to T7 procapsid shell lattice assembly, however it is the most efficient mechanism and the only assembly pathway that can lead to infectious virions in this system. The T7 connector-core is necessary for subsequent assembly steps such as tail attachment, and it has been proposed to play a role in ensuring that packaged DNA does not escape the virion prior to infection [52] and has been proposed to play a role in the DNA injection itself [18]. The mechanism described here where highly soluble scaffolding protein monomers undergo a conformational change to form filaments details a means for providing the proper curvature during assembly and ensures only a single connector-core is incorporated. The connector-core nucleated assembly, as proposed here with "scaffolding extensions" connecting "triangular face building blocks", provides a rationale

for the data obtained from this study as well as past observations [4, 17, 34, 47, 52, 60]. It also provides a logical explanation for the different forms of gp9 that were observed, in addition to providing a way for gp9 to rapidly exit the procapsid by reverting back to an extended monomer during DNA packaging and expansion.

Bibliography

1. Alberts, B., The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* **1998,** *92* (3), 291-4.

2. Perham, R. N., Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions. *Annu Rev Biochem* **2000,** *69*, 961-1004.

3. Jurica, M. S., Detailed close-ups and the big picture of spliceosomes. *Curr Opin Struct Biol* **2008,** *18* (3), 315-20.

4. Cerritelli, M. E.; Conway, J. F.; Cheng, N.; Trus, B. L.; Steven, A. C., Molecular mechanisms in bacteriophage T7 procapsid assembly, maturation, and DNA containment. *Adv Protein Chem* **2003,** *64*, 301-23.

5. Bamford, D. H.; Grimes, J. M.; Stuart, D. I., What does structure tell us about virus evolution? *Curr Opin Struct Biol* **2005,** *15* (6), 655-63.

6. Hendrix, R. W.; Duda, R. L., Bacteriophage HK97 head assembly: a protein ballet. *Adv Virus Res* **1998,** *50*, 235-88.

7. Leiman, P. G.; Kanamaru, S.; Mesyanzhinov, V. V.; Arisaka, F.; Rossmann, M. G., Structure and morphogenesis of bacteriophage T4. *Cell Mol Life Sci* **2003,** *60* (11), 2356-70.

8. Prevelige, P. E., Jr., Bacteriophage P22. In *The Bacteriophages*, Calendar, R., Ed. Oxford University Press: New York, 2006; pp 457-468.

9. Steven, A. C.; Heymann, J. B.; Cheng, N.; Trus, B. L.; Conway, J. F., Virus maturation: dynamics and mechanism of a stabilizing structural transition that leads to infectivity. *Curr Opin Struct Biol* **2005,** *15* (2), 227-36.

10. Nemecek, D.; Overman, S. A.; Hendrix, R. W.; Thomas, G. J., Jr., Unfolding Thermodynamics of the Delta-Domain in the Prohead I Subunit of Phage HK97: Determination by Factor Analysis of Raman Spectra. *J Mol Biol* **2008**.

11. Conway, J. F.; Duda, R. L.; Cheng, N.; Hendrix, R. W.; Steven, A. C., Proteolytic and conformational control of virus capsid maturation: the bacteriophage HK97 system. *J Mol Biol* **1995,** *253* (1), 86-99.

12. Showe, M. K.; Black, L. W., Assembly core of bacteriophage T4: an intermediate in head formation. *Nat New Biol* **1973,** *242* (116), 70-5.

13. Botstein, D.; Waddell, C. H.; King, J., Mechanism of head assembly and DNA encapsulation in Salmonella phage p22. I. Genes, proteins, structures and DNA maturation. *J Mol Biol* **1973,** *80* (4), 669-95.

14. Hoffman, B.; Levine, M., Bacteriophage P22 virion protein which performs an essential early function. II. Characterization of the gene 16 function. *J Virol* **1975,** *16* (6), 1547-59.

15. Israel, V., E proteins of bacteriophage P22. I. Identification and ejection from wild-type and defective particles. *J Virol* **1977,** *23* (1), 91-7.

16. Cerritelli, M. E.; Trus, B. L.; Smith, C. S.; Cheng, N.; Conway, J. F.; Steven, A. C., A second symmetry mismatch at the portal vertex of bacteriophage T7: 8-fold symmetry in the procapsid core. *J Mol Biol* **2003,** *327* (1), 1-6.

17. Roeder, G. S.; Sadowski, P. D., Bacteriophage T7 morphogenesis: phage-related particles in cells infected with wild-type and mutant T7 phage. *Virology* **1977,** *76* (1), 263-85.

18. Molineux, I. J., No syringes please, ejection of phage T7 DNA from the virion is enzyme driven. *Mol Microbiol* **2001,** *40* (1), 1-8.

19. Duda, R. L., Protein chainmail: catenated protein in viral capsids. *Cell* **1998,** *94* (1), 55-60.

20. Baschong, W.; Baschong-Prescianotto, C.; Engel, A.; Kellenberger, E.; Lustig, A.; Reichelt, R.; Zulauf, M.; Aebi, U., Mass analysis of bacteriophage T4 proheads and mature heads by scanning transmission electron microscopy and hydrodynamic measurements. *J Struct Biol* **1991,** *106* (2), 93-101.

21. Yang, F.; Forrer, P.; Dauter, Z.; Conway, J. F.; Cheng, N.; Cerritelli, M. E.; Steven, A. C.; Pluckthun, A.; Wlodawer, A., Novel fold and capsid-binding properties of the lambda-phage display platform protein gpD. *Nat Struct Biol* **2000,** *7* (3), 230-7.

22. Crick, F. H.; Watson, J. D., Structure of small viruses. *Nature* **1956,** *4506* (0028-0836 (Print)), 473-475.

23. Caspar, D. L.; Klug, A., Physical principles in the construction of regular viruses. *Cold Spring Harbor symposia on quantitative biology* **1962,** *27* (0091-7451 (Print)), 1-24.

24. Johnson, J. E.; Speir, J. A., Quasi-equivalent viruses: a paradigm for protein assemblies. *J Mol Biol* **1997,** *269* (5), 665-75.

25. Zlotnick, A., Viruses and the physics of soft condensed matter. *Proc Natl Acad Sci U S A* **2004,** *101* (44), 15549-50.

26. Dokland, T., Scaffolding proteins and their role in viral assembly. *Cell Mol Life Sci* **1999,** *56* (7-8), 580-603.

27. Duda, R. L.; Hempel, J.; Michel, H.; Shabanowitz, J.; Hunt, D.; Hendrix, R. W., Structural transitions during bacteriophage HK97 head assembly. *J Mol Biol* **1995,** *247* (4), 618-35.

28. Smith, D. E.; Tans, S. J.; Smith, S. B.; Grimes, S.; Anderson, D. L.; Bustamante, C., The bacteriophage straight phi29 portal motor can package DNA against a large internal force. *Nature* **2001,** *413* (6857), 748-52.

29. Gan, L.; Speir, J. A.; Conway, J. F.; Lander, G.; Cheng, N.; Firek, B. A.; Hendrix, R. W.; Duda, R. L.; Liljas, L.; Johnson, J. E., Capsid conformational sampling in HK97 maturation visualized by X-ray crystallography and cryo-EM. *Structure* **2006,** *14* (11), 1655-65.

30. Wikoff, W. R.; Liljas, L.; Duda, R. L.; Tsuruta, H.; Hendrix, R. W.; Johnson, J. E., Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* **2000,** *289* (5487), 2129-33.

31. King, J.; Botstein, D.; Casjens, S.; Earnshaw, W.; Harrison, S.; Lenk, E., Structure and assembly of the capsid of bacteriophage P22. *Philos Trans R Soc Lond B Biol Sci* **1976,** *276* (943), 37-49.

32. van Driel, R.; Couture, E., Assembly of the scaffolding core of bacteriophage T4 preheads. *J Mol Biol* **1978,** *123* (4), 713-9.

33. Fane, B. A.; Prevelige, P. E., Jr., Mechanism of scaffolding-assisted viral assembly. *Adv Protein Chem* **2003,** *64*, 259-99.

34. Cerritelli, M. E.; Studier, F. W., Assembly of T7 capsids from independently expressed and purified head protein and scaffolding protein. *J Mol Biol* **1996,** *258* (2), 286-98.

35. Choi, K. H.; Morais, M. C.; Anderson, D. L.; Rossmann, M. G., Determinants of bacteriophage phi29 head morphology. *Structure* **2006,** *14* (11), 1723-7.

36. Greene, B.; King, J., Scaffolding mutants identifying domains required for P22 procapsid assembly and maturation. *Virology* **1996,** *225* (1), 82-96.

37. Parent, K. N.; Zlotnick, A.; Teschke, C. M., Quantitative analysis of multi-component spherical virus assembly: scaffolding protein contributes to the global stability of phage P22 procapsids. *J Mol Biol* **2006,** *359* (4), 1097-106.

38. Morais, M. C.; Kanamaru, S.; Badasso, M. O.; Koti, J. S.; Owen, B. A.; McMurray, C. T.; Anderson, D. L.; Rossmann, M. G., Bacteriophage phi29 scaffolding protein gp7 before and after prohead assembly. *Nat Struct Biol* **2003,** *10* (7), 572-6.

39. Sun, Y.; Parker, M. H.; Weigele, P.; Casjens, S.; Prevelige, P. E., Jr.; Krishna, N. R., Structure of the coat protein-binding domain of the scaffolding protein from a double-stranded DNA virus. *J Mol Biol* **2000,** *297* (5), 1195-202.

40. Parker, M. H.; Stafford, W. F., 3rd; Prevelige, P. E., Jr., Bacteriophage P22 scaffolding protein forms oligomers in solution. *J Mol Biol* **1997,** *268* (3), 655-65.

41. Tuma, R.; Tsuruta, H.; French, K. H.; Prevelige, P. E., Detection of intermediates and kinetic control during assembly of bacteriophage P22 procapsid. *J Mol Biol* **2008,** *381* (5), 1395-406.

42. Parker, M. H.; Casjens, S.; Prevelige, P. E., Jr., Functional domains of bacteriophage P22 scaffolding protein. *J Mol Biol* **1998,** *281* (1), 69-79.

43. Fuller, M. T.; King, J., Assembly in vitro of bacteriophage P22 procapsids from purified coat and scaffolding subunits. *J Mol Biol* **1982,** *156* (3), 633-65.

44. Prevelige, P. E., Jr.; Thomas, D.; King, J., Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells. *Biophys J* **1993,** *64* (3), 824-35.

45. Baumeister, W.; Steven, A. C., Macromolecular electron microscopy in the era of structural genomics. *Trends Biochem Sci* **2000,** *25* (12), 624-31.

46. Frank, J., *Three-dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State*. 2 ed.; Oxford University Press: 2006; p 410.

47. Agirrezabala, X.; Velazquez-Muriel, J. A.; Gomez-Puertas, P.; Scheres, S. H.; Carazo, J. M.; Carrascosa, J. L., Quasi-atomic model of bacteriophage t7 procapsid shell: insights into the structure and evolution of a basic fold. *Structure* **2007,** *15* (4), 461-72.

48. Jiang, W.; Li, Z.; Zhang, Z.; Baker, M. L.; Prevelige, P. E., Jr.; Chiu, W., Coat protein fold and maturation transition of bacteriophage P22 seen at subnanometer resolutions. *Nat Struct Biol* **2003,** *10* (2), 131-5.

49. Morais, M. C.; Choi, K. H.; Koti, J. S.; Chipman, P. R.; Anderson, D. L.; Rossmann, M. G., Conservation of the capsid structure in tailed dsDNA bacteriophages: the pseudoatomic structure of phi29. *Mol Cell* **2005,** *18* (2), 149-59.

50. Wikoff, W. R.; Conway, J. F.; Tang, J.; Lee, K. K.; Gan, L.; Cheng, N.; Duda, R. L.; Hendrix, R. W.; Steven, A. C.; Johnson, J. E., Time-resolved molecular dynamics of bacteriophage HK97 capsid maturation interpreted by electron cryo-microscopy and X-ray crystallography. *J Struct Biol* **2006,** *153* (3), 300-6.

51. Thuman-Commike, P. A.; Greene, B.; Malinski, J. A.; Burbea, M.; McGough, A.; Chiu, W.; Prevelige, P. E., Jr., Mechanism of scaffolding-directed virus assembly suggested by comparison of scaffolding-containing and scaffolding-lacking P22 procapsids. *Biophys J* **1999,** *76* (6), 3267-77.

52. Agirrezabala, X.; Martin-Benito, J.; Caston, J. R.; Miranda, R.; Valpuesta, J. M.; Carrascosa, J. L., Maturation of phage T7 involves structural modification of both shell and inner core components. *EMBO J* **2005,** *24* (21), 3820-9.

53. Dunn, J. J.; Studier, F. W., Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J Mol Biol* **1983,** *166* (4), 477-535.

54. Cerritelli, M. E.; Cheng, N.; Rosenberg, A. H.; McPherson, C. E.; Booy, F. P.; Steven, A. C., Encapsidated conformation of bacteriophage T7 DNA. *Cell* **1997,** *91* (2), 271-80.

55. Masker, W. E.; Serwer, P., DNA packaging in vitro by an isolated bacteriophage T7 procapsid. *J Virol* **1982,** *43* (3), 1138-42.

56. Showe, M. K.; Isobe, E.; Onorato, L., Bacteriophage T4 prehead proteinase. II. Its cleavage from the product of gene 21 and regulation in phage-infected cells. *J Mol Biol* **1976,** *107* (1), 55-69.

57.Conway, J. F.; Wikoff, W. R.; Cheng, N.; Duda, R. L.; Hendrix, R. W.; Johnson, J. E.; Steven, A. C., Virus maturation involving large subunit rotations and local refolding. *Science* **2001,** *292* (5517), 744-8.

58. Morita, M.; Tasaka, M.; Fujisawa, H., DNA packaging ATPase of bacteriophage T3. *Virology* **1993,** *193* (2), 748-52.

59. Serwer, P., Internal proteins of bacteriophage T7. *J Mol Biol* **1976,** *107* (3), 271-91.

60. Steven, A. C.; Trus, B. L., The Structure of Bacteriophage T7. In *Electron Microscopy of Proteins*, Harris, J. R. H. R., Ed. Academic Press: London, 1986; Vol. 5, pp 1-35.

61. Newcomb, W. W.; Trus, B. L.; Cheng, N.; Steven, A. C.; Sheaffer, A. K.; Tenney, D. J.; Weller, S. K.; Brown, J. C., Isolation of herpes simplex virus procapsids from cells infected with a protease-deficient mutant virus. *J Virol* **2000,** *74* (4), 1663-73.

62. Cardone, G.; Winkler, D. C.; Trus, B. L.; Cheng, N.; Heuser, J. E.; Newcomb, W. W.; Brown, J. C.; Steven, A. C., Visualization of the herpes simplex virus portal in situ by cryo-electron tomography. *Virology* **2007,** *361* (2), 426-34.

63. Agirrezabala, X.; Martin-Benito, J.; Valle, M.; Gonzalez, J. M.; Valencia, A.; Valpuesta, J. M.; Carrascosa, J. L., Structure of the connector of bacteriophage T7 at 8A resolution: structural homologies of a basic component of a DNA translocating machinery. *J Mol Biol* **2005,** *347* (5), 895-902.

64. Frank, J., *Electron Tomography Methods for Three-Dimensional Visualization of Structures in the Cell*. 2nd ed.; Springer: New York, 2006.

65. Studier, F. W.; Moffatt, B. A., Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **1986,** *189* (1), 113-30.

66. Schuck, P., Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* **2000,** *78* (3), 1606-19.

67. Hayes, D. B.; Laue, T.; Philo, J. *SEDNTERP*, 1.09; Thousand Oaks, CA, 2006.

68. Ross, P. D.; Cheng, N.; Conway, J. F.; Firek, B. A.; Hendrix, R. W.; Duda, R. L.; Steven, A. C., Crosslinking renders bacteriophage HK97 capsid maturation irreversible and effects an essential stabilization. *EMBO J* **2005,** *24* (7), 1352-63.

69. Otwinowski, Z.; Minor, W., Processing of X-ray Diffraction Data Collected in Oscillation Mode. In *Methods in Enzymology*, Academic Press: New York, 1997; Vol. 276, pp 307-326.

70. Cole, C.; Barber, J. D.; Barton, G. J., The Jpred 3 secondary structure prediction server. *Nucleic Acids Res* **2008,** *36* (Web Server issue), W197-201.

71. Cuff, J. A.; Barton, G. J., Application of multiple sequence alignment profiles to improve protein secondary structure prediction. *Proteins* **2000,** *40* (3), 502-11.

72. Jones, D. T., Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* **1999,** *292* (2), 195-202.

73. Eddy, S. R., Profile hidden Markov models. *Bioinformatics* **1998,** *14* (9), 755-63.

74. Cheng, N.; Conway, J. F.; Watts, N. R.; Hainfeld, J. F.; Joshi, V.; Powell, R. D.; Stahl, S. J.; Wingfield, P. E.; Steven, A. C., Tetrairidium, a four-atom cluster, is readily visible as a density label in three-dimensional cryo-EM maps of proteins at 10-25 A resolution. *J Struct Biol* **1999,** *127* (2), 169-76.

75. Mastronarde, D. N., Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* **2005,** *152* (1), 36-51.

76. Kremer, J. R.; Mastronarde, D. N.; McIntosh, J. R., Computer visualization of three-dimensional image data using IMOD. *J Struct Biol* **1996,** *116* (1), 71-6.

77. Heymann, J. B.; Cardone, G.; Winkler, D. C.; Steven, A. C., Computational resources for cryo-electron tomography in Bsoft. *J Struct Biol* **2008,** *161* (3), 232-42.

78. Cardone, G.; Grunewald, K.; Steven, A. C., A resolution criterion for electron tomography based on cross-validation. *J Struct Biol* **2005,** *151* (2), 117-29.

79. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **2004,** *25* (13), 1605-12.

80. Heymann, J. B.; Belnap, D. M., Bsoft: image processing and molecular modeling for electron microscopy. *J Struct Biol* **2007,** *157* (1), 3-18.

81. Wall, J. S.; Simon, M. N., Scanning transmission electron microscopy of DNA-protein complexes. *Methods Mol Biol* **2001,** *148*, 589-601.

82. Lupas, A.; Van Dyke, M.; Stock, J., Predicting coiled coils from protein sequences. *Science* **1991,** *252* (5010), 1162-4.

83.Lupas, A., Predicting coiled-coil regions in proteins. *Curr Opin Struct Biol* **1997,** *7* (3), 388-93.

84.Matthews, B. W., Solvent content of protein crystals. *J Mol Biol* **1968,** *33* (2), 491-7.

85. Sen, A.; Baxa, U.; Simon, M. N.; Wall, J. S.; Sabate, R.; Saupe, S. J.; Steven, A. C., Mass analysis by scanning transmission electron microscopy and electron diffraction validate predictions of stacked beta-solenoid model of HET-s prion fibrils. *J Biol Chem* **2007,** *282* (8), 5545-50.

86. Poh, S. L.; el Khadali, F.; Berrier, C.; Lurz, R.; Melki, R.; Tavares, P., Oligomerization of the SPP1 scaffolding protein. *J Mol Biol* **2008,** *378* (3), 551-64.

87. Galisteo, M. L.; King, J., Conformational transformations in the protein lattice of phage P22 procapsids. *Biophys J* **1993,** *65* (1), 227-35.

88. Steven, A. C.; Greenstone, H. L.; Booy, F. P.; Black, L. W.; Ross, P. D., Conformational changes of a viral capsid protein. Thermodynamic rationale for proteolytic regulation of bacteriophage T4 capsid expansion, co-operativity, and super-stabilization by soc binding. *J Mol Biol* **1992,** *228* (3), 870-84.

89. Privalov, P. L., Thermodynamics of protein folding. *J Chem Thermodynamics* **1997,** *29*, 447-474.

90. Tuma, R.; Parker, M. H.; Weigele, P.; Sampson, L.; Sun, Y.; Krishna, N. R.; Casjens, S.; Thomas, G. J., Jr.; Prevelige, P. E., Jr., A helical coat protein recognition domain of the bacteriophage P22 scaffolding protein. *J Mol Biol* **1998,** *281* (1), 81-94.

91. Lupas, A. N.; Gruber, M., The structure of alpha-helical coiled coils. *Adv Protein Chem* **2005,** *70*, 37-78.

92. Schnarr, N. A.; Kennan, A. J., pH-Switchable strand orientation in peptide assemblies. *Org Lett* **2005,** *7* (3), 395-8.

93. Hong, Z.; Beaudet-Miller, M.; Durkin, J.; Zhang, R.; Kwong, A. D., Identification of a minimal hydrophobic domain in the herpes simplex virus type 1 scaffolding protein which is required for interaction with the major capsid protein. *J Virol* **1996,** *70* (1), 533-40.

94. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res* **2000,** *28* (1), 235-42.

95. Cerritelli, M. E.; Studier, F. W., Purification and characterization of T7 head-tail connectors expressed from the cloned gene. *J Mol Biol* **1996,** *258* (2), 299-307.

96. Ross, P. D.; Black, L. W.; Bisher, M. E.; Steven, A. C., Assembly-dependent conformational changes in a viral capsid protein. Calorimetric comparison of successive conformational states of the gp23 surface lattice of bacteriophage T4. *J Mol Biol* **1985,** *183* (3), 353-64.

97. Harris, A.; Cardone, G.; Winkler, D. C.; Heymann, J. B.; Brecher, M.; White, J. M.; Steven, A. C., Influenza virus pleiomorphy characterized by cryoelectron tomography. *Proc Natl Acad Sci U S A* **2006,** *103* (50), 19123-7.

98. Butan, C.; Winkler, D. C.; Heymann, J. B.; Craven, R. C.; Steven, A. C., RSV capsid polymorphism correlates with polymerization efficiency and envelope glycoprotein content: implications that nucleation controls morphogenesis. *J Mol Biol* **2008,** *376* (4), 1168-81.

99. Deng, Y.; Liu, J.; Zheng, Q.; Yong, W.; Lu, M., Structures and polymorphic interactions of two heptad-repeat regions of the SARS virus S2 protein. *Structure* **2006,** *14* (5), 889-99.

100. Kato, H.; Baschong, C., Isolation of a gp20-complex and its role in in vitro assembly of both prohead and core of bacteriophage T4. *Virology* **1997,** *227* (2), 400-8.

101. Fu, C. Y.; Morais, M. C.; Battisti, A. J.; Rossmann, M. G.; Prevelige, P. E., Jr., Molecular dissection of o29 scaffolding protein function in an in vitro assembly system. *J Mol Biol* **2007,** *366* (4), 1161-73.

102. Droge, A.; Santos, M. A.; Stiege, A. C.; Alonso, J. C.; Lurz, R.; Trautner, T. A.; Tavares, P., Shape and DNA packaging activity of bacteriophage SPP1 procapsid:

protein components and interactions during assembly. *J Mol Biol* **2000,** *296* (1), 117-32.

103. Newcomb, W. W.; Homa, F. L.; Thomsen, D. R.; Booy, F. P.; Trus, B. L.; Steven, A. C.; Spencer, J. V.; Brown, J. C., Assembly of the herpes simplex virus capsid: characterization of intermediates observed during cell-free capsid formation. *J Mol Biol* **1996,** *263* (3), 432-46.

104. Newcomb, W. W.; Thomsen, D. R.; Homa, F. L.; Brown, J. C., Assembly of the herpes simplex virus capsid: identification of soluble scaffold-portal complexes and their role in formation of portal-containing capsids. *J Virol* **2003,** *77* (18), 9862-71.

105. Xie, Z.; Hendrix, R. W., Assembly in vitro of bacteriophage HK97 proheads. *J Mol Biol* **1995,** *253* (1), 74-85.

106. Prevelige, P. E., Jr.; King, J., Assembly of bacteriophage P22: a model for ds-DNA virus assembly. *Prog Med Virol* **1993,** *40*, 206-21.