Copyright

by

Emilio Reyes-Aldrete

2014

The Thesis Committee for Emilio Reyes-Aldrete certifies that this is the approved version of the following dissertation:

The encapsidation protein from *Lactococcus lactis* asccφ28 as a novel system for the study of molecular motors

Committee:

Marc C. Morais, PhD, Mentor

James C. Lee, PhD, Chair

Andres F. Oberhauser, PhD

Michael B. Sherman, PhD

Dean, Graduate School

The encapsidation protein from *Lactococcus lactis* asccφ28 as a novel system for the study of molecular motors

by

Emilio Reyes-Aldrete, B.S.

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas Medical Branch

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science, Biochemistry and Molecular Biology

The University of Texas Medical Branch

December, 2014

Acknowledgements

I would like to thank all the members of my thesis committee for their infinite patience and advice in performing experiments. My interest in molecular motors started during one of Dr. Oberhauser's lectures on single-molecule experiments carried out by Kinosita on the F_1-ATPase. Observing the rotary motion I had only read about in undergrad (with my own eyes!) is still one of the most amazing experiences I've had in my life. Thank you, Dr. Oberhauser. It was also my sincere pleasure to work with Dr. Lee, he is the toughest teacher one can have, but also the best example of what I consider a real scientist; I hope I could have lived up to your expectations. To Dr. Sherman, thanks for attempting to make a decent microscopist out of me, you certainly had you work cut out for you. I also owe special thanks to Dr. Cecile Bussetta as well as Dr. Michal R. Szymanski for their help with carrying out the SAXS (C. Bussetta) and AUC (M. Szymanski) experiments presented in this document. I am also grateful to Dr. Y. Whitney Yin, for sharing some of her time at the synchrotron with us. Finally, two persons have my deepest gratitude: First, Dr Kyung (Kay) Choi, for sharing her synchrotron time, lab space and equipment with me; and second, Dr. Marc C. Morais, the kindest mentor I could have hoped for.

The encapsidation protein from *Lactococcus lactis* asccφ28 as a novel system for the study of molecular motors

Publication No. _____

Emilio Reyes-Aldrete, M.S.

The University of Texas Medical Branch, 2014

Supervisor: Marc C. Morais

Motor proteins, molecules capable of converting the cell's free energy into mechanical work, are prevalent in life. However, despite their pervasiveness, our understanding of these molecular machines remains inadequate. Some of the simplest examples of molecular motors are found in many bacteriophages and some eukaryotic viruses. For example, many phages package their genomes into preformed empty protein shells, or procapsids, in a highly thermodynamically-unfavorable process. This process is thus driven by encapsidation proteins within large macromolecular motor complexes with energy derived from the hydrolysis of ATP. Indeed, *Bacillus subtilis* bacteriophage φ29 has served for many years as a model system for the study of phage motor complexes for its relative structural simplicity and ease of its *in vitro* packaging assay. However, despite its many advantages, the motor protein of φ29, the encapsidation protein (EP), has proven difficult to work with when isolated from the rest of the complex due to its low solubility. In fact, the low solubility of the φ29 EP has hampered crystallization trials, as well as other biochemical and biophysical characterization assays. We thus opted to take a fresh new approach by working on a closely related ortholog which could help us overcome the

φ29 EP's low solubility. A BLAST search was conducted for φ29 EP orthologs which yielded a number of potential candidates, but only one stood out with a favorable score in XtalPred, a crystallizability predicting algorithm. Hence, a recombinant form of the gene product from open reading frame 11 of the *Lactococcus lactis* phage asccφ28, the putative encapsidation protein, has been expressed in *Escherichia coli* and purified to homogeneity (55.6% sequence similarity, Smith-Waterman score). Size-exclusion chromatography suggests that the recombinant protein forms a single oligomeric species in solution, while analytical ultracentrifugation and small-angle x-ray studies reveal a decameric stoichiometry. Surprisingly, preliminary crystallographic, as well as cryo-electron tomography (cryo-ET) and cryo-electron microscopy (cryo-EM) studies, suggest the decameric assembly is in fact composed of a dimer of pentamers, akin to the widely investigated and closest related EP from phage φ29, which forms pentameric rings when attached to viral proheads. The dimer of pentamers assembly of the asccφ28 encapsidation protein is a highly soluble, very stable, and active ATPase. Thus, this recombinant form of the putative encapsidation protein from the *Lactococcus lactis* phage asccφ28 shows great promise as a novel model system for *in vitro* mechanistic studies of this isolated motor protein component of phage dsDNA-packaging motors.

TABLE OF CONTENTS

List of Tables

List of Figures

List of Abbreviations

UTMB	University of Texas Medical Branch
GSBS	Graduate School of Biomedical Science
TDC	Thesis and Dissertation Coordinator
EP	Encapsidation Protein
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded DNA
RNA	Ribonucleic acid
pRNA	Packaging RNA
ATP	Adenosine triphosphate
GTP	Guanosine triphosphate
EMSA	Electrophoretic Mobility Shift Assay
NADH	Nicotinamide adenine dinucleotide (reduced)
NAD	Nicotinamide adenine dinucleotide (oxidized)
PEP	Phosphoenol pyruvate
TKM	Tris, potassium chloride, magnesium sulfate buffer
Cryo-EM	Cryo-electron microscopy
Cryo-ET	Cryo-electron tomography

Chapter 1 Introduction

Controlled motion is a hallmark of life. At the top of the animal food chain, for example, a ballerina earns her keep by matching with motion each note in Sibelius' *Valse Triste*; her controlled movements achieved by carefully rehearsed muscle contractions. In turn, at the microscopic level, each muscle contraction is the combined result of activity by individual molecules of a motor protein. In this instance, the molecule is skeletal muscle myosin, the first motor protein to be identified (Alberts 2002). Indeed, motor proteins such as myosin are at the heart of many diverse events of directed movement in life, most of which occur beyond the reach of sight of the naked eye; examples include vesicle and organelle transport, chromosome segregation, and viral genome encapsidation. Yet, despite the numerous specimens of motor proteins found to date and the extensive biophysical characterization studies carried out on a many number of them (e.g. *in vitro* assembly of their components, single-molecule force generation measurements and high-resolution crystallographic structure calculations) the precise mode by which chemical energy is harvested to induce mechanical work still eludes us.

Richard Feynman often liked to remark that unless you could build *it*, you truly didn't understand how *it* worked. Guided by this simple idea, several *in vitro* assays have been assembled to test these motors in action. For instance, an arrangement of membrane coated organelles and actin filaments proved to Ron Vale that kinesin was the motor protein responsible for organelle trafficking in squid neuron cells (Vale, Reese et al. 1985). However, while many of these experiments have indeed helped us in identifying the motor proteins at the center of these molecular machines, as well as the conditions required for these to function, they still don't tell us how these motor proteins really work. With this goal in mind, many have sought high resolution structures of a diverse number of motor proteins, and many successes have been recorded in the literature,

including crystallographic structures of the aforementioned myosin and kinesin (Rayment, Rypniewski et al. 1993). Remarkably, a high degree of structural similarity has been found for motor proteins with vastly different biological roles, such as muscle contraction or organelle transport, as in the case between myosin and kinesin (Kull, Sablin et al. 1996), or between proton pumping across mitochondrial membranes and genome packaging in viruses, as in the case for the structures of the F_1-ATPase and the bacteriophage T4 genome-encapsidation protein (Sun, Kondabagil et al. 2007).

Our lab seeks to elucidate the workings of genome-packaging motors in bacteriophages, both through structural characterization of their macromolecular components as well as by the development of *in vitro* packaging assays. Previously, our lab and collaborators conducted cryo-electron microscopy (cryo-EM) studies of the phage φ29 DNA-packaging motor complex which revealed the molecular envelopes of each of its constituent macromolecules as well as their relative positions with respect to each other, albeit at low-resolution (Morais, Koti et al. 2008). Bacteriophage φ29 is a small, short-tailed, 19.8kbp dsDNA virus that infects Bacillus subtilis (Anderson, Hickman et al. 1966). Like many other dsDNA phages, the assembly pathway of φ29 involves the pre-assembly of a prohead or procapsid and subsequent genome encapsidation to near-crystalline densities (Earnshaw and Casjens 1980). The process of genome encapsidation is powered by a motor complex, which in phage φ29, is believed to be comprised of a dodecameric assembly of a portal protein, and pentameric rings of both a virally-encoded packaging RNA (pRNA) and encapsidation protein (EP) (Tao, Olson et al. 1998; Simpson, Tao et al. 2000; Morais, Tao et al. 2001). Of these components of the φ29 motor, atomic-resolution structures exist for the portal protein (Simpson, Tao et al. 2000) and for a fragment of the pRNA corresponding to the oligomerization domain (Ding, Lu et al. 2011). Thus, many of our efforts in the last few years have focused on obtaining an x-ray crystallographic structure of the EP, the motor protein that drives the genome packaging process, all to no avail.

Nevertheless, a wealth of knowledge has been accumulated over the past few decades on the φ29 system, including the development of a simple and efficient *in vitro* genome-packaging assay which permitted direct measurement of the force generated by the motor complex during viral genome encapsidation (Smith, Tans et al. 2001). However, despite the many advances made, some questions still remain which cannot be settled by packaging assays or high resolution structures of the individual components alone. For example, some debate still remains over the exact role the pRNA plays; while some view the nucleic acid component simply as an accessory molecule mediating the interaction between the encapsidation and portal proteins, others hypothesize it may be an integral part in the process of force generation itself (Casjens 2011). Unfortunately, the best way to test if the φ29 EP is truly capable of generating force on its own, and thus to clearly distinguish what role each component plays in the macromolecular machine, is to study the individual components in isolation, one at the time. But while some successes have been reported in biophysical studies of isolated portal protein and pRNA preparations, the φ29 EP has proven thus far difficult to work with due to its poor solubility. Indeed, previous studies have shown the φ29 EP as a challenging construct to isolate and purify in large quantities, and whenever significant amounts of soluble protein were obtained, for example by addition of acetone in the buffered solution, the protein construct exhibited barely detectable activity (Huang and Guo 2003). In fact, this was our personal experience when conducting crystallization trials of the wild-type φ29 EP and several other truncated constructs.

In order to find a more amenable subject of study that could teach us more about the φ29 EP, a BLAST search was conducted for closely related orthologs. This search revealed only eight closely related constructs (25-77% identity) both belonging to dsDNA phages and with an "ATP-binding" gene ontology classification (Altschul, Madden et al. 1997). We then employed the XtalPred server (Slabinski, Jaroszewski et al. 2007) in order to narrow the number of targets selected for crystallization trials.

Surprisingly, the server returned a classification of "optimal" for only one of the eight pre-selected BLAST hits, belonging to the putative EP from the Lactococcus lactis phage asccφ28 with a 55.6% Smith-Waterman sequence similarity score (Smith and Waterman 1981). Phage asccφ28 was recently identified as an infectious agent of dairy fermentation strains of *Lactococcus lactis* (Kotsonis, Powell et al. 2008). Morphological characterization and genomic analysis carried out in this same study offered evidence that the newly discovered phage asccφ28 is genetically and functionally comparable to phage φ29 of *Bacillus subtilis* and phage Cp-1 of *Streptococcus pneumoniae*. Readily recognizable similarities to the φ29 phage include an equally sized genome (19-20Kbp), a prolate icosahedral capsid and a short non-contractile tail.

Unlike phage asccφ28, the genome-packaging and assembly processes of phage φ29 have been studied thoroughly; yet, even the stoichiometry of all components in the motor assembly is still a matter of some debate. As mentioned before, our lab and collaborators have put forward a model of the motor assembled with a dodecameric portal protein and pentameric rings of the pRNA and EP (Morais, Koti et al. 2008). However, previous solution studies have suggested that the pRNA forms hexameric rings in solution via inter-molecular base-pairing (Guo, Zhang et al. 1998; Zhang, Lemieux et al. 1998). This result has also been supported by cryo-EM studies (Ibarra, Caston et al. 2000) and single-molecule microscopy experiments with fluorescently labeled pRNAs (Shu, Zhang et al. 2007). As for the stoichiometry of the EP itself, all of the previously cited studies, with the exception of our report, have either failed to address the issue directly or have made the assumption that the stoichiometry of the EP should be the same as that of the pRNA in order to avoid a symmetry mismatch. Perhaps the strongest arguments for a hexameric assembly of the EP come from many previous comparative genomic studies which emphasize the functional and structural similarities of the φ29 EP to other P-loop NTPases, most notably by Burroughs et al. (Burroughs, Iyer et al. 2007). The φ29 EP has been classified within the additional strand conserved E (glutamate)

(ASCE) division of the P-loop NTPase fold; which houses many hexameric assembly examples in its helicase (Patel and Picha 2000; Delagoutte and von Hippel 2002) and RecA superfamilies (Yu and Egelman 1997). More specifically, the EP from phage φ29 is believed to belong in the HerA/FtsK superfamily, which partially derives its name from a well-known hexameric complex involved in chromosome segregation during bacterial cell division (Allemand, Maier et al. 2011).

Further uncertainty about the exact relationship between the stoichiometry of the components of genome-packaging motors in phages and their function, is in great part due to the existence of detailed and very satisfying mechanistic models for both pentameric and hexameric ring assemblies (Lisal and Tuma 2005; Sun, Kondabagil et al. 2008; Yu, Moffitt et al. 2010). These mechanisms are based in great part on the only available crystal structures of EPs from phages: The hexameric P4, from a family of dsRNA bacteriophages of Cystoviridae (Kainov, Tuma et al. 2006), and the large terminase from the dsDNA phage T4 (Sun, Kondabagil et al. 2007; Sun, Kondabagil et al. 2008). Unlike P4, the T4 terminase did not crystallize as a pentameric ring assembly but rather, the stoichiometry was assessed from crystal structure fittings into cryo-EM reconstructions of partially assembled phage particles (Sun, Kondabagil et al. 2008). However, as mentioned before, the simplicity of the question about the stoichiometry of the components of any genome encapsidation motor complex is overshadowed by the typical technical difficulties of studying its isolated component *in vitro*. The isolated φ29 EP, for example, has low solubility and barely detectable ATPase activity in solution which complicates stoichiometry measurements and mechanistic studies (Grimes and Anderson 1990). Although phage φ29 has been the subject of many successful single-molecule experiments utilizing in vitro assembled proheads (Chemla, Aathavan et al. 2005; Hugel, Michaelis et al. 2007; Moffitt, Chemla et al. 2009), detailed studies of the isolated components of its motor complex have proven challenging. As mentioned before, acetone and polyethylene glycol were successfully recently used to obtain a more

soluble form of the enzyme which was then found to be slightly more active, though still monomeric (Huang and Guo 2003; Huang and Guo 2003; Lee, Zhang et al. 2008). Nonetheless, development of packaging assays that can be easily investigated and manipulated *in vitro* remains one of the many challenges to overcome towards a better understanding of the working mechanisms of molecular motors. Here we report that the recombinant form of the putative EP from phage asccφ28 is a highly soluble, very stable decameric assembly, and active ATPase which may offer an excellent model system for in vitro studies of this isolated component of phage dsDNA-packaging motors.

Chapter 2 Methods

CLONING

A codon-optimized, synthetic gene coding the previously published sequence for the putative EP of the phage asccφ28 (Kotsonis, Powell et al. 2008) was obtained in a pJExpress401 vector from DNA 2.0. The obtained plasmid was replicated within the provided host strain overnight at 37˚C and isolated the next morning with the Qiagen QIAprep Spin Mini Prep kit (Cat. No. 27104). The plasmid was digested with the NdeI and XhoI restriction enzymes (NEB Cat. No. R0111S and R0146S) on sites introduced during the gene synthesis process. The digested insert was isolated from the plasmid/insert mixture electrophoretically in a 1% Agarose-TBE gel and ligated with the NEB Quick Ligation kit (Cat. No. 2200) into a pET-30a(+) vector (Novagen) that had been previously digested with the same restriction enzymes. The ligated construct was then used to chemically transform E. coli cells, Rosetta strain (EMD Biosciences), as detailed by User Protocol TB009 Rev. F 0104 from Novagen. Plasmid was again isolated from transformed cells as described above and sequenced at the UTMB Molecular Genomics Core Facility using T7 promoter and termination primers. The sequenced DNA fragment was in perfect agreement, when translated, with that of the phage asccφ28 putative EP as reported in the Swiss Institute for Bioinformatics database and verified via their *blastp* service. However, the final synthetic sequence codes for eight additional residues at the c-terminal end: leucine and glutamic acid, one each, corresponding to the XhoI restriction site, and a 6xHis-tag introduced from cloning into the pET-30a(+) for ease in purification.

EXPRESSION AND PURIFICATION

E. coli Rosetta cells transformed with the synthetic construct were grown in LB media at 37˚C until optical density at 600nm reached 0.6 (1 cm path). Maximum yield of soluble protein was obtained by induction with isopropyl β-d-1-thiogalactopyranoside (IPTG) at a final concentration of 0.1mM and expression overnight at 18˚C. Cells were harvested the next morning by centrifugation with a Beckman J2-21 centrifuge equipped with a JA-14 fixed angle rotor spinning at 5,000 rpm (~3,000xg) for 30min at 4˚C. Collected cell pellets were then either stored at -20˚C or immediately resuspended in a lysis/extraction buffer containing 50mM Na_2HPO_4, 1M NaCl, 5mM β-mercaptoethanol (β-ME) pH 8.1; typical volume of lysis/extraction buffer used was 25mL for a pellet harvested from 1L of cell culture. Resuspended cells are kept at 4˚C on ice and lysed by sonication with the Microson XL200 Ultrasonic Cell Disruptor from Misonix. Sonication regime involves thirty seconds pulse treatments with one minute rest intervals until mixture appears completely homogenous (typically after ten pulse treatments). Soluble protein fraction is then separated from cell lysate homogenate by centrifugation in the Beckman J2-21 centrifuge equipped with a JA-20 fixed angle rotor spinning at 10,000 rpm (~12,000xg) for 30min at 4˚C. The supernatant containing the soluble 6xHis-tag protein fraction is then incubated for thirty minutes at 4˚C with Talon resin (BD Biosciences) that has been previously equilibrated in lysis/extraction buffer as described above (typically employing one 1mL of resin for 25mL of supernatant). After incubation period, protein is recovered from talon resin by placement in a 1cm x 10cm glass chromatography column and eluted in 1mL fractions with a linear gradient of at least ten column/resin volumes with increasing imidazole; this gradient is generated by gradual mixing of the starting lysis/extraction buffer and an elution buffer (starting buffer with 150mM imidazole). Late eluting fractions (those eluting in the last third of the imidazole gradient) contained ~95% pure protein as judged by eye in Sodium Dodecyl Sulfate

Polyacrylamide Gel Electrophoresis (fig 1, (Laemmli 1970)). Further purification of early, dirtier fractions was accomplished by size-exclusion chromatography with a High Load 16/60 Superdex 200 in the AKTA FPLC system by GE.

ANALYTICAL ULTRACENTRIFUGATION STUDIES

Purified protein product was dialyzed into a buffer containing 20mM Tris-HCl adjusted to pH 8.1 at 20°C, 400mM NaCl and 5mM β-ME and subsequently centrifuged at 15000rpm for 20min at 4°C. Concentration was determined spectrophotometrically, with the extinction coefficient ε280 = 42400$cm^{-1}.M^{-1}$ (monomer) obtained using an approach based on Edeldoch's method (Edelhoch 1967; Gill and von Hippel 1989). Analytical ultracentrifugation experiments were performed with an Optima XL-A analytical ultracentrifuge (Beckman Inc., Palo Alto, CA), equipped with absorbance optics and An60Ti rotor; all experiments were carried out at 20°C. Absorbance data were collected by scanning the sample cells at wavelength intervals of 0.001cm in the step mode with 5 averages per step. Each experiment was conducted at two or three rotor speeds, starting with the lowest and finishing with the highest rotor speed. The sedimentation was considered to be at equilibrium when consecutive scans, separated by intervals of 8 hours, did not indicate any change.

ACTIVITY ASSAYS

Levels of ATPase activity for the putative EP were determined by a coupled assay with the enzymes pyruvate kinase and lactate dehydrogenase. The reaction scheme involves uptake by pyruvate kinase of the adenosine diphosphate generated by the EP, and transfer of the phosphate group from phosphoenolpyruvate (PEP) to regenerate adenosine triphosphate and pyruvate. The released pyruvate is then taken up by the

enzyme lactate dehydrogenase, along with reduced nicotinamide adenine dinucleotide (NADH), to generate lactate and oxidized nicotinamide adenine dinucleotide (NAD). In this manner, it is possible to monitor the progression of the reaction by exploiting the differences in absorbance at 340nm between the NADH and NAD species. The reaction mixture contains excess amounts of pyruvate kinase and lactate dehydrogenase enzymes of rabbit muscle origin, PEP and NADH. The exact amounts were 2mM PEP, 0.3mM NADH, 10u/mL of pyruvate kinase and 15u/mL lactate dehydrogenase. Reaction was carried out in a 50mM Tris, 72mM KCl, 7.2mM $MgSO_4$ (TKM) buffer as first described by (Bucher and Pfleiderer 1955); all reagents were purchased from Sigma Aldrich. Data was analyzed as the simplest case described by the Michaelis-Menten equation where the steady-state rate of ATP hydrolysis, v, is:

$$v/([\varphi 28\ EP])=(k_{cat}\ [NTP])/(K_m+[NTP])$$

SOLUTION SMALL-ANGLE X-RAY SCATTERING STUDIES

Solution small-angle scattering studies were carried out at the Lawrence Berkeley National Laboratory in Berkeley, CA. (ALS beamline 12.3.1, SIBYLS). Scattering intensities I(q) for protein and buffer solutions were measured as a function of scattering vector q ($q = 4\pi\sin\theta/\lambda$, where λ is the X-ray wavelength and 2θ is the scattering angle). Experiments were executed at 20°C and the distance from sample to detector was 1.5m, which yielded a q range from $0.01 Å^{-1}$ to $0.32 Å^{-1}$; λ was 1.0Å. The strategy employed for data collection was first reported in Hura et al. (Hura, Menon et al. 2009). In summary, SAXS data was gathered for two buffer samples and three different protein concentrations (1.4, 2.0, and 2.7mg/ml). Samples were subjected to three X-ray exposures to gauge radiation damage: one short (1s) exposure, followed by a long (10s) exposure, and a final short (1s) exposure. The program ogreNew available at the SIBYLS beamline was used to subtract buffer scattering contributions from sample scattering data.

Futher data analysis was completed using the program package PRIMUS from the ATSAS suite 2.3 (Konarev, Volkov et al. 2003). Guinier and kratky plots were employed to check for sample aggregation and folded state, respectively. The radius of gyration R_G and forward scattering intensity I(0) were determined using the Guinier approximation: $I(q) \approx I(0) \exp(-q^2R_G^2/3)$, with the limits $qR_G < 1.3$. In addition, these parameters were also calculated from the pair-distance distribution function P(R), which was in turn computed from full scattering patterns via indirect Fourier inversion of the scattering intensity I(q) utilizing the program GNOM (Svergun 1992). The parameter D_{max} or maximum particle diameter was also approximated from the P(R). The hydrated volume V_P of the particle was determined according to the Porod equation: $V_P = 2\pi 2I(0)/Q$, where I(0) is the extrapolated scattering intensity at zero angle and Q is the Porod invariant (Porod 1982; Mertens and Svergun 2010). Finally, the molecular mass of a globular protein can be approximated from the value of its hydrated volume. It has been found empirically that the hydrated volume in nm^3 should be between 1.5 to 2 times the molecular mass in kDa (Mertens and Svergun 2010).

CRYSTAL PREPARATION AND CRYSTALLOGRAPHIC ANALYSIS

Crystals were grown by vapor diffusion in 8–20 days at 17ºC using a droplet containing 5 mg/ml EP protein, 400mM NaCl, 50mM $NaPO_4$, pH 8.1, surrounded by petroleum jelly (HUMCO) and equilibrated against 100mM KPO_4, pH 7.4, 10% isopropanol and ~1M sodium citrate as precipitant. Crystals were transferred into paraffin oil for stabilization and improvement of diffraction resolution. These crystals were highly reproducible and diffracted up to a maximum resolution of 4.5Å at the APS beamline of the Argonne national laboratory, and 5.5Å at our home source (Rigaku Ultimate Home Lab® equipped with FR-E++DW Superbright x-ray generator and RAXIS-IV^{++} dual 30cm imaging plate). The X-ray data were processed and scaled using HKL3000

(Otwinowski and Minor 1997). Crystals belong to the tetragonal lattice system, with unit cell dimensions of a=b=107.5Å and c=349.1Å.

CRYO-ELECTRON TOMOGRAPHY AND SINGLE-PARTICLE CRYO-MICROSCOPY STUDIES

For cryo-ET studies, purified asccφ28 EP at a concentration of ~0.2mg/mL was combined with a suspension of 15-nm gold beads that serve as fiducial markers for the subsequent alignment of images in a tilt series. Sample mixtures of protein and gold beads (3.5 µl) were applied onto 200-mesh holey EM grids (R2/2 Quantifoil; Micro Tools GmbH, Jena, Germany). Filter paper was then used for blotting the grids before submersion into liquid ethane. This process yielded grids with EP embedded in a thin layer of vitreous water spanning across the holes in the carbon supporting film. Grids were then moved into the microscope cryoholder and kept at a constant temperature near −176°C. Grid imaging was done with a JEOL 2200FS microscope (200 keV) equipped with an Omega energy filter at a magnification of ×10,000. Cryo-ET tilt series were imaged and data frames recorded on a slow-scan 4K charge-coupled device (CCD) camera (UltraScan 895; GATAN, Inc.); dosage was ~1 $e^-/Å^2$/frame.

Data for single-axis tilt series was collected in 2° steps over a ±68° tilt range. Full 4K CCD frames were recorded (typically 80 to 89 frames for each tilt series). Data collection was carried out using the program SerialEM (Mastronarde 2005) which performs automatic tilting, tracking, focusing, and image acquisition under low-dose conditions. The program IMOD (Mastronarde 1997) was then used to process the tilt series and reconstruct tomograms. All images in the tilt series were aligned with the gold beads serving as fiducial markers, and the final tomograms were computed by a weighted back projection algorithm. Final computed tomograms measured 4K by 4K by 600 voxels. Individual EP particles were boxed from the full tomograms applying the program TRIMVOL included in the IMOD suite.

For single-particle experiments, phage asccφ28 EP at a concentration of ~0.2mg/mL was frozen on a thin layer of ice resting over holey carbon film grids (C-flat; Protochips, Raleigh, NC or R2x2 Quantifoil; Micro Tools GmbH, Jena, Germany). This was accomplished by applying 3.5µl of protein suspension to the holey film grids, blotting with filter paper, and submerging grids into liquid ethane cooled in a liquid nitrogen bath. Frozen grids were then stored under liquid nitrogen while waiting to be transferred to a cryo-specimen holder, also kept under liquid nitrogen, before being loaded into a Polara G2 electron microscope equipped with a field emission gun (FEG) operating at 300 keV. All grids were maintained at near-liquid nitrogen temperature (−172 to −180°C) during imaging.

Protein assemblies were imaged at a ×60,000 nominal magnification with a 4,000- by 4,000-pixel slow-scan charge-coupled device (CCD) camera (Tietz). Images were acquired with a ~20-electron/$Å^2$ dose; the CCD pixel size corresponded to 1.49Å on the specimen scale. We used a 1- to 3.5-µm defocus range for imaging. Overall, 12,213 individual particles images from 20 micrographs were processed, with 9,431 particles used in the final reconstruction. Boxing of individual protein particle images from micrographs was done using the BOXER program from the EMAN suite (Ludtke, Baldwin et al. 1999). Further steps in the reconstruction process such as CTF correction, generation of an initial search model, and particle classification and reconstruction, were all done with programs available as part of the EMAN suite.

Chapter 3 Results

LACTOCOCCUS LACTIS ASCCΦ28 EP FORMS A STABLE HOMO-DECAMER IN SOLUTION

A recombinant form of the open reading frame 11 from the Lactococcus lactis phage asccφ28, the putative EP, was successfully expressed in E. coli Rosetta cells. A codon optimized DNA sequence coding for the putative EP was synthesized by DNA 2.0 in a pJExpress401 vector and subsequently cloned into a PET30a(+) vector (Novagen). Transfer into the PET30a(+) added a 6x-His coding sequence at the c-terminal end which facilitated purification of the protein product. The greatest amount of soluble protein product was obtained after induction with 1mM IPTG and expression at 18˚C overnight. A first purification step with Talon resin (BD Biosciences) in a 50mM Sodium Phosphate and 1M NaCl buffer yielded late-eluting fractions which were over 95% pure as judged by SDS-PAGE (Fig 1). However, in order to obtain the highest yield of mostly homogenous sample, a second purification step was carried out by size-exclusion chromatography in an AKTA system equipped with a High Load 16/60 Superdex 200 (GE). The chromatographic profile obtained during this second purification step suggested that the protein exists as a large assembly. Comparison to a typical elution profile of standard proteins in the same chromatographic medium and flow rate pointed towards an apparent molecular weight for the putative EP between that of IgG (158 KDa) and Ferritin (440 KDa) (Fig 2).

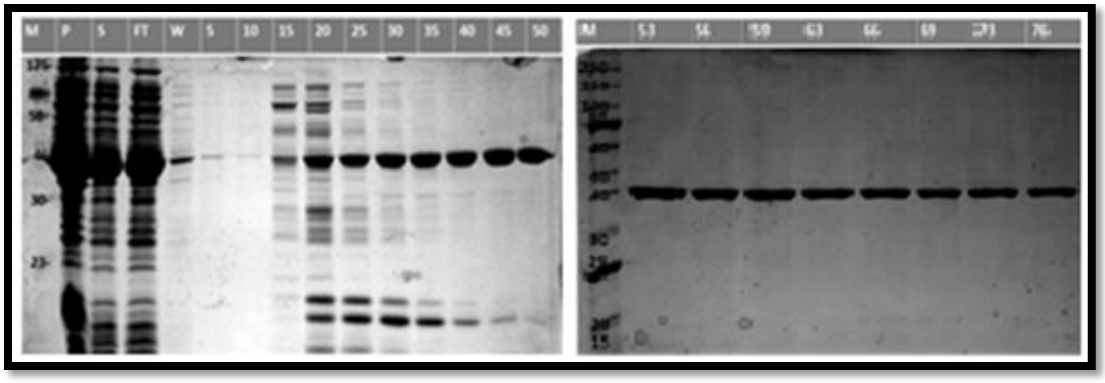


Figure 1. SDS-PAGE illustrating typical yield and purity of asccφ28 EP product obtained after a single purification step with a metal-affinity resin.
Fractions obtained during the first purification steps were analyzed by SDS-PAGE. Cell lysate obtained after sonication was spun at 12000xg to separate soluble (S lane in left gel above) from non-soluble fractions (P lane on left gel). The soluble fraction was then incubated with Talon resin for 30min, allowed to settle inside a 1cmx10cm glass column by allowing the majority of supernatant to slowly flow through (FT lane on left gel), washed with 5 column volumes of the lysis/extraction described above (W lane left gel) and finally eluted from the resin with the same buffer with an increasing linear gradient of 0-150mM Imidazole at a rate of 0.5mL/min. Every fifth of the early eluting (low imidazole concentration) fractions is shown on the left gel above (5-50). The right gel shows every third fraction of late eluting (high imidazole concentration) fractions (53-76). Although the late eluting fractions showed only one and highly pure protein species, fractions 53 and above were combined and concentrated and subjected to a second-purification step by size-exclusion chromatography.

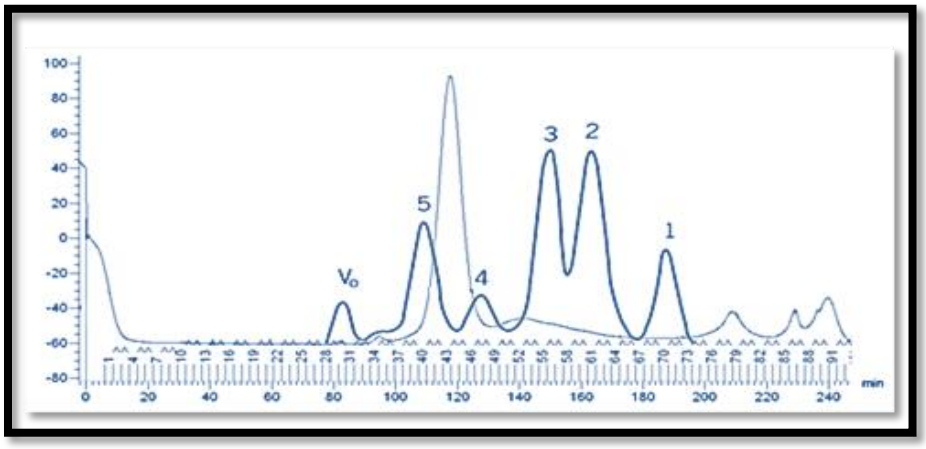


Figure 2. The asccφ28 EP elutes in a size-exclusion chromatography column with an apparent molecular weight between 158 and 440 kDa
Typical elution profile in a size-exclusion chromatography GE Superdex 200 (16/60) column for the combined late-eluting fractions shown in Figure 1 (light blue). Dark blue overlay shows the typical expected elution profile for standard proteins under the same flow rate conditions: 1. Myoglobin, Mr 17 000; 2. Ovalbumin, Mr 43 000; 3. Albumin, Mr 67 000; 4. IgG, Mr 158 000; 5. Ferritin, Mr 440 000.

In order to better assess the molecular weight of the assembly observed in size-exclusion chromatographic studies, we carried out analytical ultracentrifugation (AUC) measurements. Sedimentation equilibrium experiments were performed for different protein concentrations of samples purified as described above. Figure 3 below shows AUC data collected at a protein concentration of 0.353μM, the figure shows protein absorbance values as a function of radial distance, at three-different speeds. The smooth curves overlaying the data are simulations using the best fit parameters resulting from a global NLLS fit to above equation with M_i and b as fitting parameters. For the n-component system, the total concentration at radial position r, c_r, is defined by:

$$c_r = \sum_{i=1}^{n} c_{b\,i} \exp\left[\frac{(1 - \bar{v}_i \rho)\omega^2 M_i (r^2 - r_b^2)}{2RT}\right] + b$$

where c_{bi}, v_i, and M_i are the concentration at the bottom of the cell, partial specific volume, and molecular mass of the "I" component, respectively, ρ is the density of the solution, ω is the angular velocity, and b is the base-line error term. Partial specific volume of the φ28 EP protein (v_i) was calculated from the amino acid composition of the protein according to Lee and Timasheff (Lee and Timasheff 1974) and was 0.738 mL/g. Molecular mass for the purified protein sample was found to be 443 472 ± 8 000 Da (Figure 3).

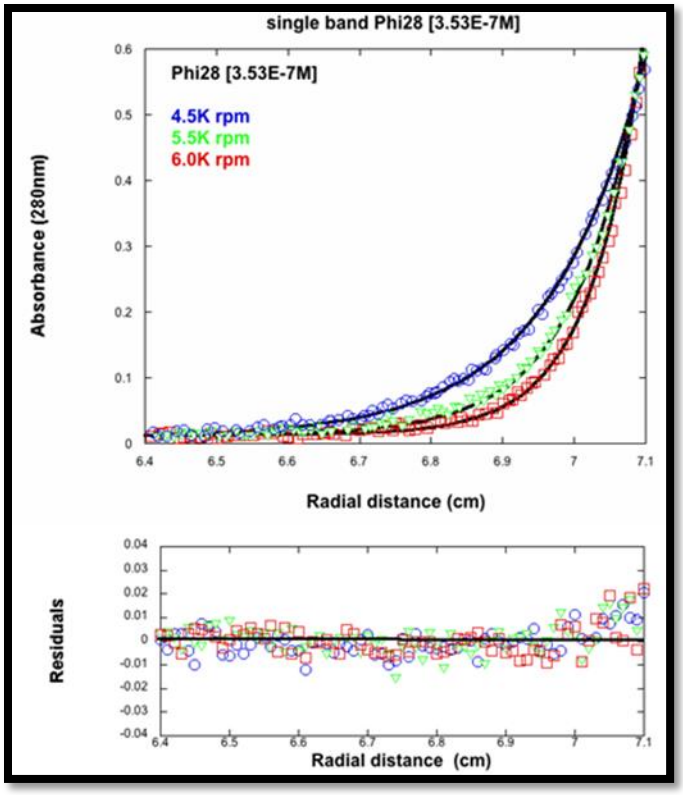


Figure 3. Sedimentation equilibrium data suggests a decameric assembly for the asccφ28 EP.
Protein absorbance data was collected at three different speeds: 4.5K RPM (blue circles), 5.5K RPM (green circles) and 6K RPM (red squares). Dark overlaying lines are simulations using the best fitting parameters resulting from a global NLLS to the equation above, with M_i and b as fitting parameters. M_i value found in this way was 443,472 ± 8000Da consistent with a decameric assembly of the encapsidation protein (monomer molecular mass is 44,585 Da).

With the aim to further characterize the size and shape of the assemblies in solution, we conducted solution x-ray scattering experiments in the same Tris/NaCl buffer used in AUC experiments (Figure 4). Since no radiation damage was detected over a long (10 sec) exposure time, this data was used for the analysis. Three different concentrations ranging from 1.4 to 2.7 mg/mL were selected to ensure scattering data for the sample did not show any variations at different protein concentrations. No protein concentration dependence was detected as data for these three different concentrations superimposed well. The radius of gyration (R_G) and the maximal dimension (D_{max}) are

two parameters addressing the size of the particle in solution that can be determined from the SAXS data. Two independent methods (the Guinier approximation and the pair-distribution function P(R)) were used to calculate the R_G value for each protein concentration. Both methods provided R_G values in agreement to each other with ≈ 55 ± 2 Å calculated from the Guinier approximation and ≈ 53.7 ± 1 Å estimated from the pair-distance distribution function P(R) (Table 1). The maximal dimension of the particle, D_{max}, which is the value where the pair distribution function equals zero and x≠0, was found to be 155 ± 5 Å for the three protein concentrations. The hydrated particle volume (V_P) was estimated between 745.6 and 778.6 nm^3, corresponding to average molecular masses between 426.1 and 444.9 kDa and thus to a decameric assembly in solution (Table 1).

Table 1: Summary of SAXS data.

Protein concentration	R_G (Guinier) (Å)	D_{max} (P(R)) (Å)	R_G (real space, P(R)) (Å)	Hydrated volume (Porod) (nm^3)	Estimated molecular mass (kDa)	Multimeric state
1.4 mg/mL	54.8	155	53.8	745.6	426.1	9.5
2.0 mg/mL	55.1	155	53.7	778.6	444.9	9.9
2.7 mg/mL	55.2	155	53.6	760.6	434.6	9.7

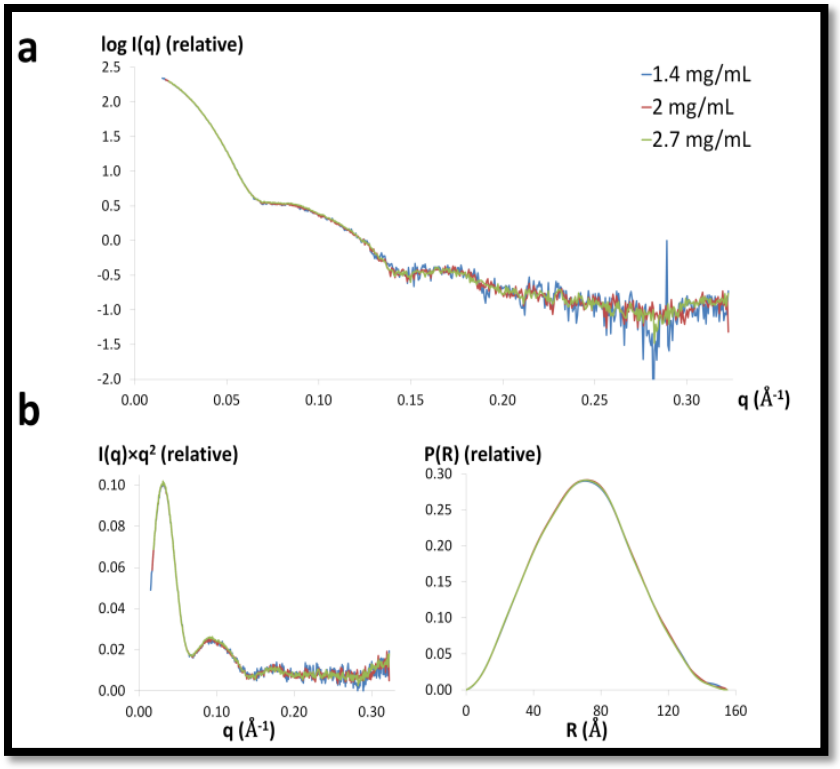


Figure 4: Small-angle x-ray scattering data indicates the asccφ28 EP is a well-folded ring assembly with a maximum dimension of 155±5Å.
Scattering intensities were recorded as a function of the scattering vector for three different protein concentrations: 1.4mg/mL (blue profile), 2mg/mL (red profile) and 2.7mg/mL (green profile) (figure 2a). An additional representation of the data is the Kratky plot (figure 2b, left) in which the assembly clearly presents a profile typical of well-folded protein. The pair distribution function (figure 2b, right) reveals an assembly with a maximum dimension of 155±5Å.

THE ASCCΦ28 EP IS AN ACTIVE ATPASE IN VITRO

We tested levels of ATPase activity for the putative EP in order to assess whether the synthetic gene indeed coded for an active enzyme product, and whether the large assembly observed was the result of folded monomer units and not a denatured aggregate. ATPase activity for the putative EP was determined by a coupled assay with the enzymes pyruvate kinase and lactate dehydrogenase as described in the materials and methods section. Decrease in absorbance at 340nm as a result of NADH depletion was monitored

to assess the progression of the reaction. Reaction solution contained excess amounts of the coupling enzymes, PEP, and NADH, in a 50mM Tris, pH 7.4, 72mM KCl, 7.2mM $MgSO_4$ (TKM) buffer. The rate of absorbance changes was obtained from the linear slopes of progress curves after steady state rates were observed for periods over five minutes. The values obtained from these slopes were then plotted as a function of substrate concentration (Figure 5) and the resulting curve was modeled by a nonlinear combination of the parameters V_{max} and K_m in the modified Michaelis-Menten equation:

$$v = \frac{V_{max}[NTP]}{K_m + [NTP]}$$

The values for V_{max} and K_m obtained from the nonlinear regression analysis were $7.6x10^{-7}Msec^{-1}$ and 28μM, respectively. Also, from $V_{max}=[asccφ28EP]k_{cat}$, k_{cat} was found to be $0.849sec^{-1}$.

Since all NTPases are somewhat promiscuous in the selection of their substrates; that is, they all are capable of hydrolyzing any of the nucleoside triphosphates, we decided to verify our construct was indeed primarily an ATPase. This is important since the nucleoside triphosphate preference can offer clues as to the protein's primary function. While there are many examples of GTPases within the P-loop fold, a fold common in motor proteins, no GTPase has been found involved in genome encapsidation in phages, or in related motor proteins involved in nucleic acid recombination or unwinding (Leipe, Wolf et al. 2002). The kinetic parameters found for GTP as a substrate were: K_m=344μM, $V_{max}=7.31x10^{-7}Msec^{-1}$ and $k_{cat}=0.817sec^{-1}$. Based on the 10-fold increase in the K_m value that occurred from switching from ATP to GTP, 28 μM to 344 μM, we can be confident that the φ28 EP is indeed primarily an ATPase.

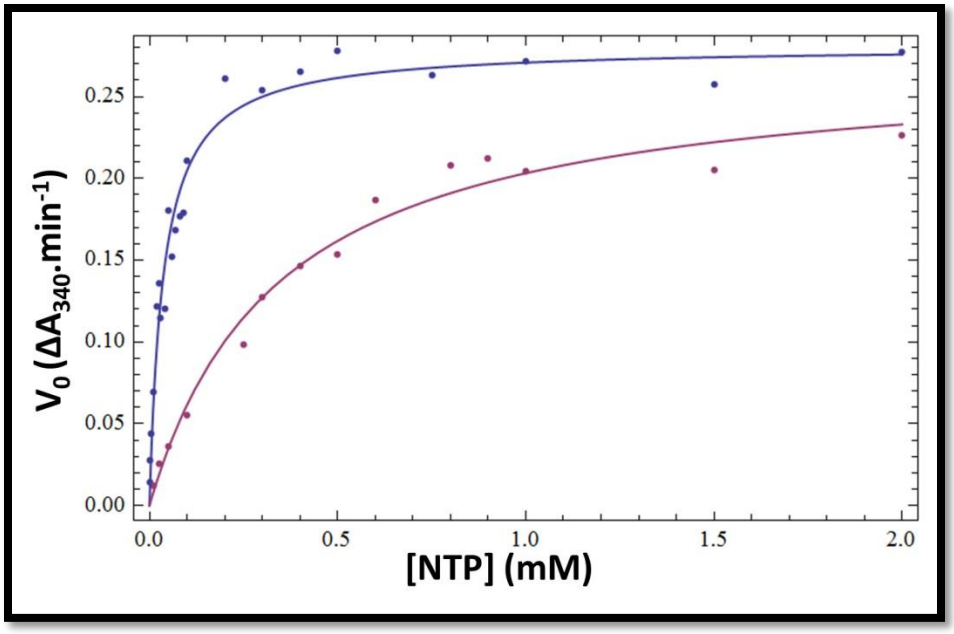


Figure 5. Steady-state kinetics for the asccφ28 encapsidation protein.
ATPase (blue profile) and GTPase (magenta) activity was evaluated by a coupled assay with excess amounts of the enzymes pyruvate kinase and lactate dehydrogenase. Decline in absorbance at 340nm as a result of NADH depletion was monitored to assess the progression of the reaction. The plot above shows the initial velocities calculated from reaction progress curves as a function of substrate concentration. The resulting curve was then modeled by a nonlinear combination of the parameters V_{max} and K_m in a modified Michaelis-Menten equation (see text above). The values obtained for ATP from the nonlinear regression analysis for V_{max} and K_m were $7.53x10^{-7}Msec^{-1}$ and 37μM, respectively. In the same manner, for GTP, K_m=344μM, $V_{max}=7.31x10^{-7}Msec^{-1}$.

X-RAY CRYSTALLOGRAPHIC DATA FOR THE ASCCΦ28 EP SUGGESTS THE PRESENCE OF 5 MONOMERS IN THE ASYMMETRIC UNIT

A great majority of our efforts were spent on the structural characterization of the asccφ28 assembly and the φ29 homolog by x-ray crystallography and cryo-electron microscopy (cryo-EM). In the crystallography front, we had a partial success as we obtained crystals for the asccφ28 EP in 100mM phosphate buffer at pH 7.4, 1M sodium citrate and 10% Isopropanol (Figure 6). These crystals are highly reproducible and typically diffract up to 4.5 Å. We collected several complete native data sets at our home source. However, the best native data set was collected at the APS beamline of the

Argonne national laboratory. Analysis of the diffraction data from the APS beamline indicated that the crystals belong to the tetragonal lattice system, with unit cell dimensions of a=b=107.5Å and c=349.1Å, similar to the cell dimensions determined for the previously collected data sets at our home source. Evaluation of crystallographic symmetry and systematic absences showed that within the tetragonal lattice system, the crystals likely belong to spacegroup $P4_12_12$. Combining cell dimensions and spacegroup information with normalized probabilities for the Matthews coefficient (and hence solvent content) for a representative group of proteins in the PDB, five monomers are predicted to be most likely present in the asymmetric unit. Given that AUC and SAXS data indicated that the protein is a decamer, the most likely interpretation of all available data is that the oligomeric assembly would consist of two pentameric rings related to each other by a two-fold symmetry axis.

However, we have been unsuccessful thus far in attempting to solve the structure from the native data set by molecular replacement using a homology model (described below) or the structure of the N-terminal core of φ29 EP (M. Morais, unpublished data), or by SAD and SIRAS methodology using additional data sets of iodine-derivatized crystals collected at our home source. On the other hand, we have successfully produced selenomethionine labeled protein with no changes detected in its oligomeric state by native gel electrophoresis. The labeled construct yielded crystals in the same conditions as the wild-type isoform and will likely yield anomalous data at a synchrotron radiation source in the near future. This data will allow the calculation of phases needed to determine the structure of the assembly at high resolution.

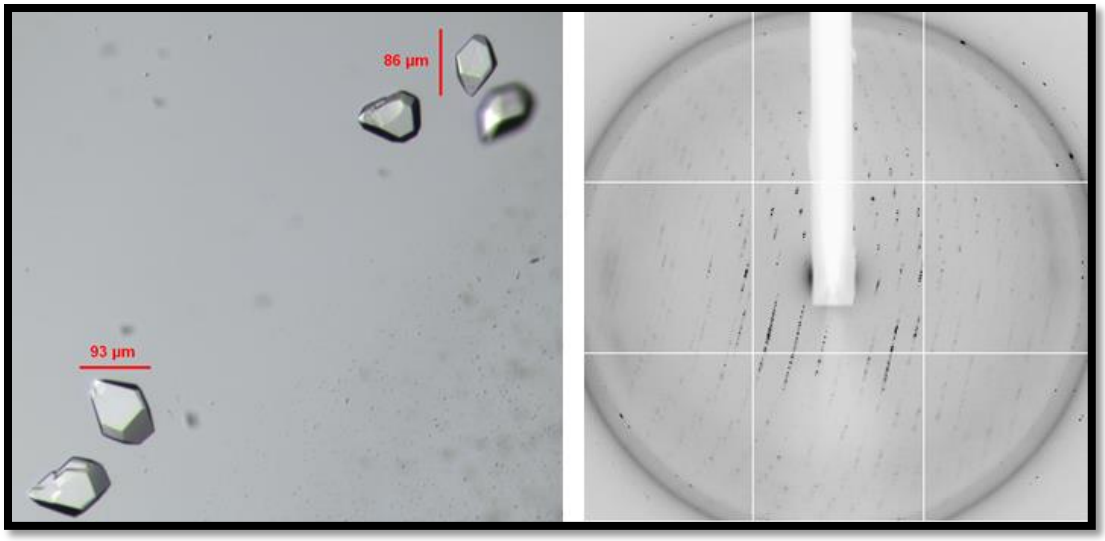


Figure 6. Typical crystal morphology and best diffraction pattern obtained for crystallized asccφ28 EP.
Crystals were obtained for the asccφ28 EP in 100mM phosphate buffer at pH 7.4, 1M sodium citrate and 10% Isopropanol. These crystals are highly reproducible and typically diffract up to a maximum resolution of 4.5 Å. Analysis of diffraction data at the APS beamline indicated that the crystals belong to the tetragonal lattice system, with unit cell dimensions of a=b=107.5Å and c=349.1Å, and likely belongs to spacegroup $P4_12_12$.

THE ASCCΦ28 EP ASSEMBLY IS COMPOSED OF A DIMER OF PENTAMERIC RINGS

We have also performed cryotomography and single-particle electron microscopy studies (Data collected in collaboration with Dr. Michael B. Sherman). These studies confirmed the presence of a large assembly with dimensions of 120±12.5Å by 145±7.5Å, width and height dimensions, as determined from the tomographic reconstruction (Figure 6). Even though the exact symmetry could not be assessed from this reconstruction, the presence of a dihedral symmetry was nevertheless plausible. However, a closer look at figure 7 may raise some questions in the mind of an inquisitive reader. First, the assembly shows a large cavity in its core (7a), the poles being held together by only two small contact points; while this would not be impossible, it is highly unlikely that such an arrangement would yield a stable assembly. Second, the view from above (7b) appears rather square, with very marked edges; this is also unlikely to be found in many

biological assemblies, at least at this scale. Thus, it is more likely that this is an artifact of the reconstruction process and further processing of the data is needed.

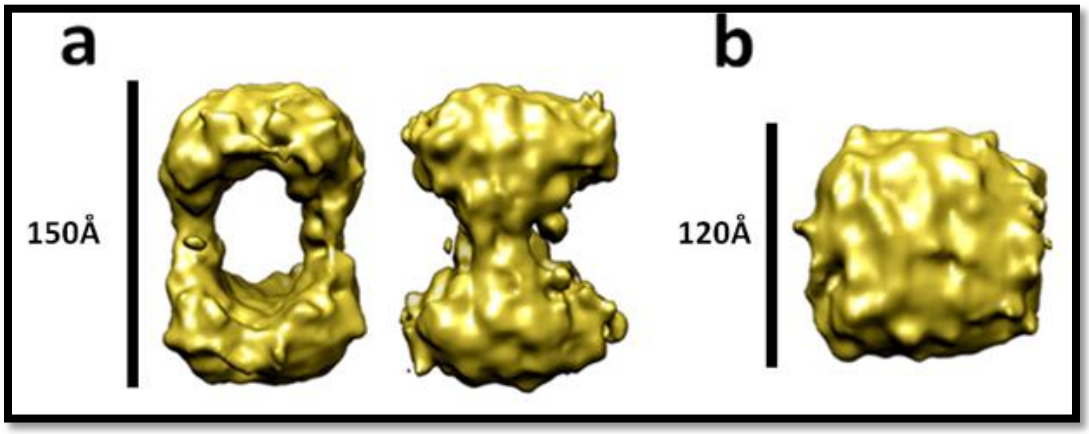


Figure 7. Molecular envelopes for the asccφ28 EP as determined by cryo-ET clearly showing two well separated lobes.
Side views (**a**, 90˚ apart) of the molecular envelope determined from sub-volume averaging of cryo-electron tomography reconstructions clearly showing two well separated lobes linked by fairly elongated regions. Since it is unlikely that each lobe is composed of only one monomer, dihedral symmetry is likely present (reconstruction envelope shown here at contour level 2 σ; also **b** top view).

Cryo-ET differs from single-particle cryo-EM in one fundamental way. In single-particle cryo-EM, information is collected on thousands of single particles that have been frozen in random orientations, and a three-dimensional volume is reconstructed by combining all this information after the orientation of each particle has been determined. In contrast, in cryo-ET, information is collected on single particles that have been frozen in random orientations but a three-dimensional volume is reconstructed by collecting information for each of these particles in different orientations. This is done by changing the point of view by tilting the microscope's stage. Thus, whereas cryo-EM often requires an initial model to proceed with the reconstruction process, cryo-ET has no such requirement. However, microscope stages have a limited range of tilt, leaving certain angles of view with no information available; this is also known as "the missing wedge" problem of cryo-ET. Consequently, once information has been collected on each of the

individual particles at different angles, these so called sub-volumes are then averaged to increase the signal to noise ratio and minimize the effects of the "missing wedge". Sometimes, during the process of averaging, it is possible that alignment of sub-volumes may be dominated by the missing wedge rather than by real structural similarities between particles. This often leads to artifacts such as the one observed here. Nevertheless, although further analysis of the data is required, we are confident that the dihedral symmetry assumption still holds, since the missing wedge effects on the sub-volume averaging process would not alter the larger structural arrangement of the assembly.

Similar dimensions were also calculated from single-particle reconstructions, 105±10Å by 145±5Å; these dimensions are in close agreement with those calculated from x-ray scattering experiments. Thus, a reconstruction carried out with only d1 symmetry applied and using the tomography reconstruction as a starting model yielded a model with easily discernible d5 symmetry (Figure 8).

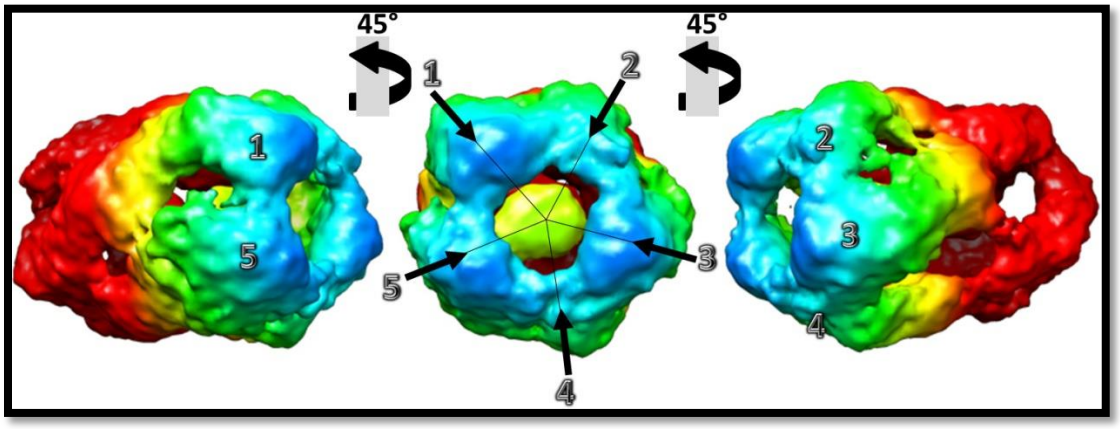


Figure 8. Single-particle cryo-EM reconstruction of the asccφ28 EP, with no symmetry imposed, displaying an assembly of two rings, each with five protrusions. Three views, approximately 45˚ apart, of the EP's molecular envelope as determined from a cryo-EM reconstruction from single-particle studies of the asccφ28 EP (d1 symmetry, contour level of 3σ). In place of the two lobes seen before by cryotomography, two rings are now clearly visible. Also, each of these rings shows five protrusions, with three of them slightly more defined, see top view (middle).

We thus applied the d5 symmetry to an additional reconstruction of the same data set and obtained a molecular envelope for the assembly with a resolution of 10.4Å, as determined by a 0.5 cutoff for the Fourier shell correlation in an even-odds test (Figure 8). The resulting molecular envelope clearly shows two pentameric ring structures facing each other with strong electron densities observed in the core and near the equator of the resulting assembly.

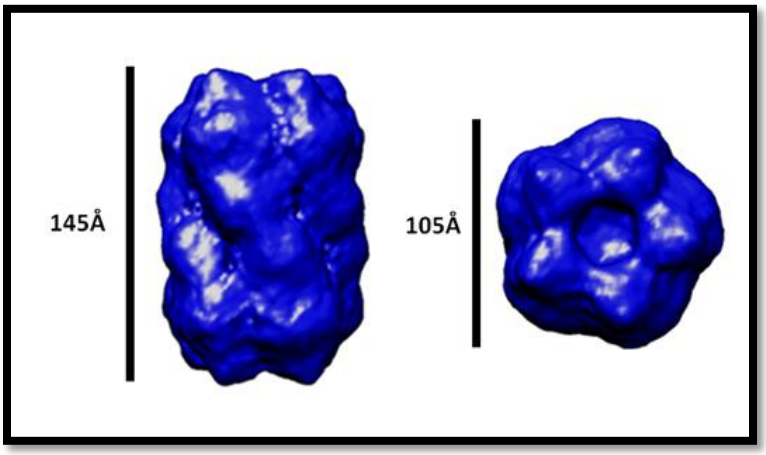


Figure 9. Single-particle cryo-EM reconstructions of the EP with d5 symmetry imposed.
The dimensions of the envelope are 105±10Å by 145±5Å; calculated resolution was 10.4Å (contour level of 3σ).

Chapter 4 Discussion

We began with a quest to understand the φ29 molecular machine encouraged by the wealth of information available on the system and its relative simplicity. As mentioned before, the molecular motor of φ29 is comprised of only three components: a dodecameric assembly of portal protein and pentameric rings of a pRNA and encapsidation protein. With the aim of understanding the role that each component played in the assembly, we attempted to further characterize the EP in isolation, only to corroborate previous studies reporting on the protein's poor solubility. Now, with the finding of a closely related EP belonging to phage asccφ28, we have not only discovered a stable, soluble and relatively active novel construct, we may have also stumbled upon a relatively simpler model system of genome encapsidation in bacteriophages. In fact, it is possible that the stability and activity observed with asccφ28 EP is indicative of its ability to assemble on the portal protein on its own, without mediation by pRNA, as is the case in φ29. This would not be surprising, as φ29 and CP-1 are the only two phage families found to date with an active RNA component in its genome encapsidation machinery (Casjens 2011). Furthermore, analysis of potential transcripts exhibiting secondary elements typical of these pRNA's in noncoding regions of the asccφ28 genome revealed no matches (Kotsonis, Powell et al. 2008). Obviously, corroboration that the asccφ28 molecular machine is indeed simpler and composed of only two components, a portal assembly and the dimeric pentamer of EP, will necessitate the production of partially assembled particles *in vitro* and further characterizations experiments as previously done with φ29 (Grimes and Anderson 1990; Grimes, Jardine et al. 2002).

The stoichiometry discovered for the asccφ28 also raises some interesting questions. As stated before, some debate still exists on whether the φ29 EP functions as a pentameric or hexameric assembly. Moreover, a model has been proposed that incorporates observations from single-molecule and cryo-EM experiments on the φ29

system. This is the most detailed model proposed to date for the φ29 molecular motor and requires a pentameric assembly of EP to fit one of its central assumptions (Moffitt, Chemla et al. 2009). Our studies on the closely related EP from asccφ28 suggest that a pentameric assembly might indeed be the correct stoichiometry present in φ29 and support the above mentioned model. The fact that we have a dimer of pentamers, instead of a single pentameric ring, may also suggest that the extra ring present in asccφ28 has perhaps taken over the role played by the pentameric ring of pRNA present in φ29. While it would be a remarkable adaptation that an assembly of a single type of species, a protein, has taken over the role played by assemblies of two distinct species, a protein and a ribonucleic acid, it would not be entirely unlikely or novel. It could be certainly possible that the φ28 EP is a conformational heterodimer, with each "monomer" being in fact composed of a pentameric ring. In this way, one conformation of the pentameric ring mediates attachment to the portal protein, while the second ring is free to interact with incoming genomic material and initiate packaging. However, our preliminary cryo-ET data appear to show an assembly with dihedral symmetry, which implies that the two pentameric rings face each other. If each of the pentameric rings of EP in asccφ28 is capable of generating force on its own, as has been the assumption at least for EP in the φ29 system, this would imply that the two rings are working against each other. On the other hand, if the dimer of pentamers is actually a single operating unit, this likely indicates that the pRNA and EP in φ29 also work as a unit, unless the φ29 EP has managed to carry out exactly the same function with a significantly smaller assembly. At this point, lacking further data, we cannot completely rule out that the dimer of pentamers is just an artifact of *in vitro* expression, namely, that the biologically functional unit is indeed a pentamer. We also cannot completely rule out the existence of pRNA in φ28. Thus, all this remains exciting conjecture for the time being.

In the meantime, since an x-ray crystallographic structure of the assembly has eluded us thus far, we have employed the protein fold recognition algorithm implemented

in PHYRE programs which identified three templates with high confidence for residues 3 up to 220, suggesting the existence of an N-terminal domain or subdomain. Two of these templates belong to RNA helicase families (DEAD/DEAH box and flavivirus DEAD domain families) and the third template is from the phage T4 DNA packaging protein gp17. On the other hand, no satisfactory template was found for the C-terminal end (residues 214 to 366), although all secondary structure predictions indicate a high proportion of α-helical structural content (Figure 9). This N-terminal domain or subdomain templates found by the algorithm has been characterized extensively, and it is known to contain all the structural elements needed for ATP binding and hydrolysis. Additionally, the N-terminal templates have been found to be structurally similar (Rmsd<3.3Å) to many other motor proteins associated with, for example, dsRNA genome packaging (φ12 P4), DNA recombination (RecA) and proton pumping (F_1-ATPASE) (Sun, Kondabagil et al. 2007). Thus, if the fold predictions by this algorithm are correct, this would indicate that 1)the force generation mechanism is likely the same, not just between φ29 and asccφ28, but between these and the distantly related T4 bacteriophage and RNA helicases, and perhaps even to dsRNA packaging motors, recombinases and the F_1-ATPase proton pump; 2)the n-terminal domain is likely the motor core that has been adapted to power different molecular machines for different tasks (i.e genome encapsidation, RNA unwinding, DNA recombination, proton pumping); and 3)the c-terminal domain likely contains the adapting component, for example, in the case of φ28, to bind the portal protein or the phage's genome.

The lack of high resolution structural data and the difficulty in working with large protein complexes are among the many reasons that we still lack a fundamental understanding of how these molecular machines operate. However, the asccφ28 encapsidation protein has thus far been found to be a stable assembly that is also active as an ATPase even when isolated from the rest of the motor complex. This level of activity observed for the asccφ28 encapsidation protein clearly suggests that the assembly may

function independently from other molecular machine components, although its ability to generate a force on its own remains to be tested. It is our belief that this recently characterized assembly will likely yield high resolution structural information in the near future and will serve as a novel system to test current models of genome encapsidation motors in phages.

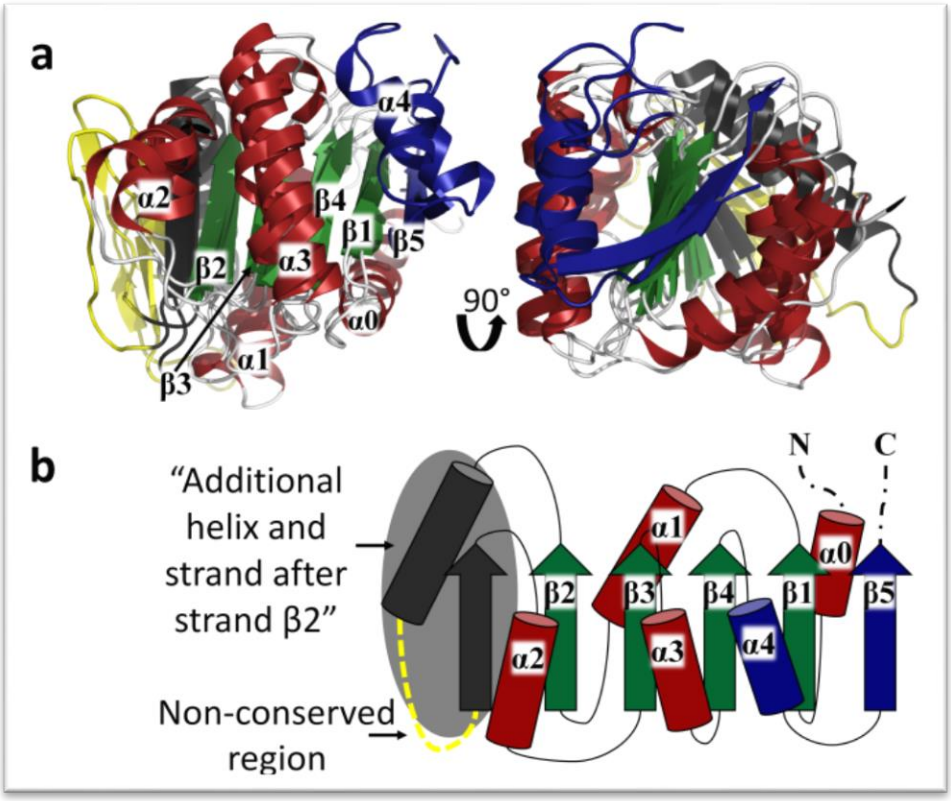


Figure 10. Homology modeling of the asccφ28 encapsidation protein.
(a) Superimposition of the three templates selected by the PHYRE program to model the N-terminal domain of the asccφ28 encapsidation protein. The superimposed structures show a high degree of structural similarity in the ATPase core across multiple families, despite having low primary sequence identity. These structures are colored as follows: green, red and light grey are respectively the β-strands, α-helices and loops of the highly conserved part of the domain core. The black elements correspond to a synapomorphy shared across the different superfamilies represented by the templates and is defined as "additional helix and strand after strand β-2". The yellow region is non-conserved among the templates. The blue α-helix and β-strand represent the missing part in the Flavivirus DEAD helicase family. (b) Common topology of the three templates selected by PHYRE program.

Chapter 5 Future Directions

The fundamental task going forward in the development of φ28 as novel system for the study of motor proteins is to determine the biologically relevant assembly of EP. Here we have reported that the φ28 EP is an active ATPase that assembles into a very stable dimer of pentameric rings. The most convincing evidence that this assembly is indeed the functional unit will likely be acquired only by the development of an in vitro genome-packaging assay, much in the way it was done in the case of φ29 (Grimes, Jardine et al. 2002). In such an assay, an assembly of φ28 procapsids, portal protein, genomic material (or other dsDNA fragments) and EP could be gathered so as to test if genome encapsidation is possible with only these components, and thus corroborate the prediction that φ28 does not require a pRNA to carry out genome packaging. Verification that genome encapsidation events are indeed taking place could be easily done by comparing procapsids before and after the addition of EP or dsDNA in a cryo-EM experiment. This choice of technique would have the added advantage that, in case packaging events are observed, the experiments could be easily fine-tuned with varying amounts of ATP or shorter incubation times before freezing. This would then allow us to capture a significant portion of the particle population arrested in the middle of a packaging event, and thus observe the motor assembly at work. Finally, molecular envelope calculations from cryo-EM reconstructions of EP engaged in genome packaging could be compared to envelopes we have obtained for isolated EP to verify whether the motor protein functions as a pentamer or as a dimer of pentamers.

Alternatively, if the assumption is made that the φ28 EP is indeed a pRNA-independent motor protein, and that the dimer of pentamers is the biologically relevant and functional unit, it is then likely that it may be able to bind DNA and generate a force all on its own. We have attempted preliminary electrophoretic motility shift assays (EMSA) to test if the φ28 EP was capable of binding a short random sequence (~20bp) of

dsDNA (data not shown). Although these experiments failed to show dsDNA-EP complexes, they were by no means exhaustive, mobility assay should be attempted once again with different lengths of dsDNA fragments, different amounts of ATP present in the reaction, different buffer conditions, or with phage genomic DNA. It is also important to bear in mind that EMSA experiments may not be sensitive enough and that many other techniques are available to test protein-DNA interactions, such as capillary electrophoresis, laser-induced fluorescence, or affinity chromatography. It is also possible that any dsDNA that we attempt to use as a substrate of the φ28 EP may need to be labelled, by covalently bonding an accessory protein also coded by the φ28 genome, as this the manner in which phage motors recognize which dsDNA's need to be encapsidated from the myriad of choices available in a host's cell. Nevertheless, if we are able to determine the conditions in which the EP assembly binds dsDNA, we could then begin understanding the force generation mechanism with carefully planned mutations followed by force measurements using atomic force microscopy.

Finally, a lot of the work we have already started has the potential to yield large amounts of useful information. The fold recognition algorithm from PHYRE predicted the presence of N-terminal domain or subdomain which contains all the elements for the ATP binding and catalysis and has been shown to form ring structure in many other structurally related examples. The algorithm also predicted the existence of a second c-terminal domain or subdomain with high α-helical content. A close look at the single-particle cryo-EM volume reconstruction presented here (Fig 9) appears to show the presence of two rings at each of the poles of the reconstructed volume, and "arms" reaching across. Knowing the tendency of the N-terminal ATPase domain to form rings, and the rigid nature of α-helices, which may appear as long-stretching arms in low-resolution density maps, one could assume that the N-terminal domain of the protein lies at the poles and its c-terminus at the equator. That is, it is possible that pentameric rings are primarily formed by N-terminal to N-terminal domain interactions which then

dimerize by interaction of the c-terminal sub-domains on opposite rings. If this is the case, a mutated construct containing only an N-terminal domain may be stable enough and yield single pentameric rings. This construct would provide a new target which may be more suitable for crystallization, yield more information about the quaternary structure of the assembly, and educate us about communication between the two domains or subdomains: Will a construct of the N-terminus alone exhibit similar ATPase activity as the whole construct or is there some sort of feedback from the c-terminal domain needed to stimulate ATP hydrolysis?

Even without significant more planning of new experiments, it is possible that just the incorporation of further data in the single-particle cryo-EM reconstruction process we already began may increase the resolution achieved by this method beyond 10Å. Thus, if subnanometer resolution is indeed achieved for the cryo-EM reconstructions, this will allow for tracing of the protein backbone and reliable fitting of homology models in the molecular envelope to generate a much more refined structural model of the assembly in solution.

References

Alberts, B. (2002). Molecular biology of the cell. New York, Garland Science.

Allemand, J. F., B. Maier, et al. (2011). "Molecular motors for DNA translocation in prokaryotes." Curr Opin Biotechnol.

Altschul, S. F., T. L. Madden, et al. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res **25**(17): 3389-402.

Anderson, D. L., D. D. Hickman, et al. (1966). "Structure of Bacillus subtilis bacteriophage phi 29 and the length of phi 29 deoxyribonucleic acid." J Bacteriol **91**(5): 2081-9.

Bucher, T. and G. Pfleiderer (1955). "Pyruvate Kinase from muscle." Methods is Enzymology **1**(C): 435-440.

Burroughs, A. M., L. M. Iyer, et al. (2007). "Comparative genomics and evolutionary trajectories of viral ATP dependent DNA-packaging systems." Genome Dyn **3**: 48-65.

Casjens, S. R. (2011). "The DNA-packaging nanomotor of tailed bacteriophages." Nat Rev Microbiol **9**(9): 647-57.

Chemla, Y. R., K. Aathavan, et al. (2005). "Mechanism of force generation of a viral DNA packaging motor." Cell **122**(5): 683-92.

Delagoutte, E. and P. H. von Hippel (2002). "Helicase mechanisms and the coupling of helicases within macromolecular machines. Part I: Structures and properties of isolated helicases." Q Rev Biophys **35**(4): 431-78.

Ding, F., C. Lu, et al. (2011). "Structure and assembly of the essential RNA ring component of a viral DNA packaging motor." Proc Natl Acad Sci U S A **108**(18): 7357-62.

Earnshaw, W. C. and S. R. Casjens (1980). "DNA packaging by the double-stranded DNA bacteriophages." Cell **21**(2): 319-31.

Edelhoch, H. (1967). "Spectroscopic determination of tryptophan and tyrosine in proteins." Biochemistry **6**(7): 1948-54.

Gill, S. C. and P. H. von Hippel (1989). "Calculation of protein extinction coefficients from amino acid sequence data." Anal Biochem **182**(2): 319-26.

Grimes, S. and D. Anderson (1990). "RNA dependence of the bacteriophage phi 29 DNA packaging ATPase." J Mol Biol **215**(4): 559-66.

Grimes, S., P. J. Jardine, et al. (2002). "Bacteriophage phi 29 DNA packaging." Adv Virus Res **58**: 255-94.

Guo, P., C. Zhang, et al. (1998). "Inter-RNA interaction of phage phi29 pRNA to form a hexameric complex for viral DNA transportation." Mol Cell **2**(1): 149-55.

Huang, L. P. and P. Guo (2003). "Use of acetone to attain highly active and soluble DNA packaging protein Gp16 of Phi29 for ATPase assay." Virology **312**(2): 449-57.

Huang, L. P. and P. Guo (2003). "Use of PEG to acquire highly soluble DNA-packaging enzyme gp16 of bacterial virus phi29 for stoichiometry quantification." J Virol Methods **109**(2): 235-44.

Hugel, T., J. Michaelis, et al. (2007). "Experimental test of connector rotation during DNA packaging into bacteriophage phi29 capsids." PLoS Biol **5**(3): e59.

Hura, G. L., A. L. Menon, et al. (2009). "Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS)." Nature Methods **6**(8): 606-12.

Ibarra, B., J. R. Caston, et al. (2000). "Topology of the components of the DNA packaging machinery in the phage phi29 prohead." J Mol Biol **298**(5): 807-15.

Kainov, D. E., R. Tuma, et al. (2006). "Hexameric molecular motors: P4 packaging ATPase unravels the mechanism." Cell Mol Life Sci **63**(10): 1095-105.

Konarev, P. V., V. V. Volkov, et al. (2003). "PRIMUS: a Windows PC-based system for small-angle scattering data analysis." Journal of Applied Crystallography **36**: 1277-1282.

Kotsonis, S. E., I. B. Powell, et al. (2008). "Characterization and genomic analysis of phage asccphi28, a phage of the family Podoviridae infecting Lactococcus lactis." Appl Environ Microbiol **74**(11): 3453-60.

Kull, F. J., E. P. Sablin, et al. (1996). "Crystal structure of the kinesin motor domain reveals a structural similarity to myosin." Nature **380**(6574): 550-5.

Laemmli, U. K. (1970). "Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4." Nature **227**: 680-685.

Lee, J. C. and S. N. Timasheff (1974). "The calculation of partial specific volumes of proteins in guanidine hydrochloride." Arch Biochem Biophys **165**(1): 268-73.

Lee, T. J., H. Zhang, et al. (2008). "Strand and nucleotide-dependent ATPase activity of gp16 of bacterial virus phi29 DNA packaging motor." Virology **380**(1): 69-74.

Leipe, D. D., Y. I. Wolf, et al. (2002). "Classification and evolution of P-loop GTPases and related ATPases." J Mol Biol **317**(1): 41-72.

Lisal, J. and R. Tuma (2005). "Cooperative mechanism of RNA packaging motor." J Biol Chem **280**(24): 23157-64.

Ludtke, S. J., P. R. Baldwin, et al. (1999). "EMAN: semiautomated software for high-resolution single-particle reconstructions." J Struct Biol **128**(1): 82-97.

Mastronarde, D. N. (1997). "Dual-axis tomography: an approach with alignment methods that preserve resolution." J Struct Biol **120**(3): 343-52.

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

Mertens, H. D. and D. I. Svergun (2010). "Structural characterization of proteins and complexes using small-angle X-ray solution scattering." Journal of Structural Biology **172**(1): 128-41.

Moffitt, J. R., Y. R. Chemla, et al. (2009). "Intersubunit coordination in a homomeric ring ATPase." Nature **457**(7228): 446-50.

Morais, M. C., J. S. Koti, et al. (2008). "Defining molecular and domain boundaries in the bacteriophage phi29 DNA packaging motor." Structure **16**(8): 1267-74.

Morais, M. C., Y. Tao, et al. (2001). "Cryoelectron-microscopy image reconstruction of symmetry mismatches in bacteriophage phi29." J Struct Biol **135**(1): 38-46.

Otwinowski, Z. and W. Minor (1997). Processing of X-ray diffraction data collected in oscillation mode, Elsevier. **276:** 307-326.

Patel, S. S. and K. M. Picha (2000). "Structure and function of hexameric helicases." Annu Rev Biochem **69**: 651-97.

Porod, G. (1982). Chapter 2: General Theory. Small Angle X-ray Scattering. O. Glatter and O. Kratky, Academic Press, London.

Rayment, I., W. R. Rypniewski, et al. (1993). "Three-dimensional structure of myosin subfragment-1: a molecular motor." Science **261**(5117): 50-8.

Shu, D., H. Zhang, et al. (2007). "Counting of six pRNAs of phi29 DNA-packaging motor with customized single-molecule dual-view system." Embo J **26**(2): 527-37.

Simpson, A. A., Y. Tao, et al. (2000). "Structure of the bacteriophage phi29 DNA packaging motor." Nature **408**(6813): 745-50.

Slabinski, L., L. Jaroszewski, et al. (2007). "XtalPred: a web server for prediction of protein crystallizability." Bioinformatics **23**(24): 3403-5.

Smith, D. E., S. J. Tans, et al. (2001). "The bacteriophage straight phi29 portal motor can package DNA against a large internal force." Nature **413**(6857): 748-52.

Smith, T. F. and M. S. Waterman (1981). "Identification of common molecular subsequences." J Mol Biol **147**(1): 195-7.

Sun, S., K. Kondabagil, et al. (2008). "The structure of the phage T4 DNA packaging motor suggests a mechanism dependent on electrostatic forces." Cell **135**(7): 1251-62.

Sun, S., K. Kondabagil, et al. (2007). "The structure of the ATPase that powers DNA packaging into bacteriophage T4 procapsids." Mol Cell **25**(6): 943-9.

Svergun, D. I. (1992). "Determination of the regularization parameter in indirect-transform methods using perceptual criteria." Journal of Applied Crystallography **25**: 495-503.

Svergun, D. I., M. V. Petoukhov, et al. (2001). "Determination of domain structure of proteins from X-ray solution scattering." Biophysical Journal **80**(6): 2946-53.

Tao, Y., N. H. Olson, et al. (1998). "Assembly of a tailed bacterial virus and its genome release studied in three dimensions." Cell **95**(3): 431-7.

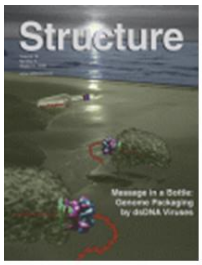
Vale, R. D., T. S. Reese, et al. (1985). "Identification of a novel force-generating protein, kinesin, involved in microtubule-based motility." Cell **42**(1): 39-50.

Yu, J., J. Moffitt, et al. (2010). "Mechanochemistry of a viral DNA packaging motor." J Mol Biol **400**(2): 186-203.

Yu, X. and E. H. Egelman (1997). "The RecA hexamer is a structural homologue of ring helicases." Nat Struct Biol **4**(2): 101-4.

Zhang, F., S. Lemieux, et al. (1998). "Function of hexameric RNA in packaging of bacteriophage phi 29 DNA in vitro." Mol Cell **2**(1): 141-7.

Annex



Structure
Article
Volume 16, Issue 8, 6 August 2008, Pages 1267–1274

Defining Molecular and Domain Boundaries in the Bacteriophage ϕ29 DNA Packaging Motor

Marc C. Morais,[1,2] Jaya S. Koti,[3] Valorie D. Bowman,[2] Emilio Reyes-Aldrete,[1] Dwight L. Anderson,[3,4] and Michael G. Rossmann[2,*]
[1]Department of Biochemistry and Molecular Biology, 301 University Boulevard, University of Texas Medical Branch, Galveston, TX 77555-0647, USA
[2]Department of Biological Sciences, 915 W. State Street, Purdue University, West Lafayette, IN 47907-2054, USA
[3]Department of Diagnostic and Biological Sciences
[4]Department of Microbiology
University of Minnesota, 18-242 Moos Tower, Minneapolis, MN 55455, USA

Abstract

Cryo-electron microscopy (cryo-EM) studies of the bacteriophage phi29 DNA packaging motor have delineated the relative positions and molecular boundaries of the 12-fold symmetric head-tail connector, the 5-fold symmetric prohead RNA (pRNA), the ATPase that provides the energy for packaging, and the procapsid. Reconstructions, assuming 5-fold symmetry, were determined for proheads with 174-base, 120-base, and 71-base pRNA; proheads lacking pRNA; proheads with ATPase bound; and proheads in which the packaging motor was missing the connector. These structures are consistent with pRNA and ATPase forming a pentameric motor component around the unique vertex of proheads. They suggest an assembly pathway for the packaging motor and a mechanism for DNA translocation into empty proheads.

Vita

Emilio Reyes-Aldrete was born in Monterrey, Mexico. He is the son of Emilio Reyes Ortiz and Luisa Aldrete Sanchez. Emilio served five years in the United Stated Navy as a Hospital Corpsman. After his military service, he attended Montana State University in Bozeman, MT where he received a B.S. in chemistry.

Permanent address: 1638 Sunnyside Dr., Harlingen, TX 78552

This dissertation was typed by Emilio Reyes-Aldrete.