Structural Studies on Bacteriophage Portal Proteins

Weisha Luan

A Thesis submitted for the degree of Doctor of Philosophy

CHEMISTRY

The University of York

Department of Chemistry

July 2013

Abstract

In tailed bacteriophages and evolutionarily related herpes viruses, the portal protein is a central component of the DNA packaging molecular motor, which translocates viral genomic DNA into a preformed procapsid. The motor is the most powerful molecular machine discovered in nature, generating forces reaching ~50 pN and translocating DNA with a speed of several hundred bp/sec using ATP as an energy source. The oligomeric portal protein ring is situated at a unique vertex of the procapsid forming a conduit for DNA entry and exit. Although the three-dimensional structure has already been determined for portal proteins from bacteriophages P22, SPP1 and phi29, several important questions about the role of individual protein segments in DNA translocation and their interaction with other components of the motor remain unanswered. Structural information on portal proteins from other bacteriophages, like T4 for which a wealth of biochemical information is already available, will help to answer at least some of these questions.

The portal protein of bacteriophage SPP1 (gp6) can form circular oligomers containing 12 or 13 subunits. It is found as a 12-subunit oligomer when incorporated into the viral capsid and as a 13-subunit assembly in its isolated form. The X-ray structure of the SPP1 portal protein is available only for the isolated 13-subunit assembly of the N365K mutant form. Because this mutation results in a reduction in the length of packaged DNA, determining the structure of the wild type portal protein would shed light on the mechanism of DNA translocation. Elucidation of the mechanism of DNA packaging depends also on the availability of accurate structural information on the SPP1 portal protein in its 12-subunit biologically active state. Such structural knowledge would be particularly useful in future, for designing a stable molecular machine that can function *in vitro*.

In this thesis, experiments were designed to promote the formation of the dodecameric gp6: viz fusing gp6 with TRAP protein that forms a stable circular dodecamer as well as the co-expression of gp6 with the SPP1 scaffolding protein gp11. The protein targets were cloned, expressed and purified, and the oligomeric state of gp6 was characterised by a combination of biochemical, biophysical and structural approaches. The structure of the wild type gp6 was solved at 2.8 Å

resolution, revealing a 13-fold symmetrical molecule. The protein's fold is the same as for the N365K mutant, with most significant conformational differences observed in the tunnel loop and in segments of the clip and crown domains. Comparison with the structure of N365K mutant reveals significant differences in subunit-subunit interactions formed by tunnel loops, including different hydrogen bonding and van der Waals interactions. It is likely that these differences account for the different amount of packaged DNA, indicating involvement of tunnel loops in DNA packaging.

The portal protein of bacteriophage CNPH82, cn3, was also successfully cloned, expressed and purified. Promising crystallisation conditions have been identified that yield crystals diffracting to 4.2 Å. Further optimisation should lead to determination of the X-ray structure of this protein in not too distant future. Self-rotation function calculations and SEC-MALLS analysis indicate that the cn3 protein forms 13-subunit assemblies, in common with the SPP1 portal protein. Foundation work has been carried out for the bacteriophage T4 portal protein, aimed at identifying suitable production and purification conditions. In addition, the full-length bacteriophage SPP1 scaffolding protein gp11 has been cloned, purified and crystallised. Degradation was observed in the full length gp11 protein and therefore a series of truncations were designed, cloned and purified aiming to improve the stability. Further studies on limited proteolysis of the full-length gp11 should lead to a stable gp11 tuncation that will form crystals with better diffraction.

List of Contents

Figures

Tables

Abbreviations

AEBSF	4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride
APS	Ammonium Persulphate
ATP	Adenosine-5'-triphosphate
bp	Base pairs
CCP4	Collaborative Computational Project, number 4
DNA	Deoxyribonucleic acid (ssDNA for single-stranded, dsDNA for double-stranded)
DNase	Deoxyribonuclease
dNTP	The deoxyribonucleotides dATP, dTTP, dCTP and dGTP
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic Acid
EM	Electron Microscopy
ESRF	European Synchrotron Radiation Facility
FT	Flow Through
GST	Glutathione-S-Transferase
HEPES	2-[4-(2-hydroxyethyl) piperazin-1-yl] ethanesulfonic acid
kb	kilobases
IPTG	Isopropyl-β-D-thiogalacto-pyranoside
LB	Luria–Bertani / Lysogeny broth
LS	Light Scattering
MALDI-MS	Matrix Assisted Laser Desorption Ionisation-Mass Spectrometry
MES	2-(N-morpholino) ethanesulfonic acid
MPD	2-methylpentane-2,4-diol
NMR	Nuclear Magnetic Resonance
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEG	Polyethylene Glycol
RMSD	Root Mean Square Deviation
RNA	Ribonucleic acid (ssRNA for single-stranded, dsRNA for double-stranded)
RI	Refractive Index

SDS PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC-MALLS	Size Exclusion Chromatography coupled with Multi-Angle Laser Light Scattering
TEMED	N,N,N′,N′-Tetramethylethylenediamine
TerL	Large Terminase
TerS	Small Terminase
TRAP	Tryptophan RNA-Binding Attenuation Protein
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
TRAP	Tryptophan RNA-Binding Attenuation Protein
WT	Wild Type

Acknowledgements

I would like to express my special thanks to a large group of people, because this doctoral thesis would never have been possible without help and support from them.

Firstly, I would like to thank my supervisor Professor Fred Antson, who gave me the opportunity to work on this exciting project. He has always been happy to provide invaluable guidance and support on both an academic and a personal level throughout the four years of my PhD studies.

I would like to express my deepest gratitude to my Independent Panel Member, the late Professor Guy Dodson for his inspiration, encouragement, and suggestions in his many personal communications and in my TAP meetings.

I am truly grateful for people who have made valuable contributions towards the work described in this thesis. Particularly, a huge thank you to my co-supervisor, Callum Smits, for his encouragement and guidance to help me get started in the laboratory and throughout the project. Special thanks to Maria Chechik for all the instruction and help with molecular cloning and protein purification. Thanks to Yuriy Chaban and Elena Orlova for the contribution on EM studies, Huw Jenkins for help in solving the structure of gp6, Jochen Fesseler for the work on cn3, Oliver Bayfield and Sherry Chan for the work on the portal-TRAP fusion proteins, Carina Büttner for help in data processing of cn3, Elena Blagova and Fiona Whelan for helpful discussions on protein purification and crystallisation, Chao-sheng Chen for discussions and suggestions on the study of TRAP, Andrew Leech for assistance on SEC-MALLS, Johan Turkenburg and Sam Hart for collecting diffraction data, Simon Grist and Sally Lewis for running the laboratory, University of York Overseas Scholarship and Wild Fund for financial support.

For proof reading the thesis, a big thanks goes to Sandra Greive who went through the entire thesis, gave me very helpful suggestions and corrected my mistakes. I am also very grateful for the help from Lowri Williams, Dan Peters and Oliver Bayfield to carefully check some chapters and correct my mistakes, as well as for critical comments and useful feedback from Fred Antson.

I would also like to thank Maria Chechik, Callum Smits, Fiona Whelan, Elena Blagova, Glyn Hemsworth, Sandra Greive, Huw Jenkins, Lowri Williams, Dan Peters, Juan Loredo Varela, Chao-sheng Chen, Yuan He, Annika Frank, James Tunaley, Christopher Watson and everyone in YSBL for keeping me company and for promoting a welcoming academic and social environment.

Last but not least, I am grateful for the support, encouragement and understanding from my family and my best friend Feng Tan. This thesis is dedicated to them.

Declaration

The results discussed here are my own original work, and have not been presented elsewhere, except for the following aspects of the studies:

1. Part of the work on fusing the SPP1 gp6 with the natural *B. stearothermophilus* 11-subunit TRAP was performed by Sherry Chan (BSc Biochemistry, York, 2012) under my supervision, and was presented earlier in her final year project report entitled "How viruses fill their capsids with DNA; structure and function of a DNA-translocating molecular motor ".

2. Part of the work on the crystallisation of the portal-TRAP fusion proteins and cloning of the full-length gp11 was performed by Oliver Bayfield (2011) under my supervision, and was presented earlier in his project summary report entitled" Molecular machinery: characterisation of the SPP1 *B. subtilis* bacteriophage portal protein".

3. Cloning and purification of the portal protein from bacteriophage CNPH82 was performed by Jochen M. Fesseler (MSc Biochemistry, University of Bayreuth, 2010) in his project report entitled "Portal protein and procapsid formation of the bacteriophage CNPH82". A paper on the crystallisation and crystallographic analysis of this protein was published in 2012 entitled "Recombinant portal protein from *Staphylococcus epidermidis* bacteriophage CNPH82 is a 13-subunit oligomer".

1 Introduction

1.1 The Bacteriophages

Bacteriophages are viruses that infect and replicate within bacteria, often causing bacterial lysis following their replication. Bacteriophages were independently discovered by Frederick Twort and Félix d'Hérelle in the 1910s (Twort, 1925). However, due to their small size, bacteriophages were not visualised until the advent of the electron microscope in the 1940s (Luria et al, 1943). As the most abundant organisms in the biosphere, bacteriophages display striking diversity in their morphological properties - shape (spherical, helical, rod, polyhedral, etc.), size and auxiliary structures such as tails and envelopes (Orlova, 2009). Notably, the genetic material in the majority of bacteriophages is double-stranded DNA (dsDNA), but bacteriophages containing single-stranded DNA (ssDNA), double-stranded RNA (dsRNA) and single-stranded RNA (ssRNA) have also been discovered. The morphology and nucleic acid properties provide the basis for the classification of bacteriophages, and nineteen families have now been defined (Calendar, 2006).

Bacteriophages are obligate intracellular parasites of bacteria. Viral protein production and genome replication are fulfilled by bacterial machinery. These viruses are categorised as lytic bacteriophages or temperate bacteriophages based on the ability to cause lytic or lysogenic infection of the host cell (Bertani, 1953). Lytic bacteriophages can merely replicate throughout the lytic life cycle and cause lysis of the host bacteria, whereas temperate bacteriophages can also integrate their genetic materials into the bacteria's genome as a non-infectious “prophage”.

During lytic infection by DNA bacteriophages, viral DNA replaces the bacterial genomic DNA as the template for both replication and transcription in host cells. Viral proteins, such as the coat proteins and tail proteins, are synthesised using host cell ribosomes, tRNAs and amino acids. The process of DNA replication, protein synthesis, and viral assembly is sequentially coordinated in the lytic life cycle (Figure 1-1). The first step of infection is adsorption, during which a phage recognises a specific receptor on the surface of the bacterial cell and adheres to the site, often by means of the tail fibres (Rakhuba et al, 2010). Following adsorption,

viral DNA is injected into the host cell. The synthesis of "early proteins" is triggered immediately after viral DNA penetration, to produce virus-specific DNA polymerases, which exclusively replicate phage DNA. "Late proteins" are synthesised after the replication of the bacteriophage genome, subsequently followed by the assembly of infectious virion particles. Mature virions are released from host cells with the assistance of phage-encoded lytic enzymes, such as lysozyme, to break down the peptidoglycan in bacterial cell walls. It takes about 25-35 minutes for a T-phage to complete a life cycle, during which some 50 to 200 phage particles may be produced from every infected bacterium (Calendar, 2006).

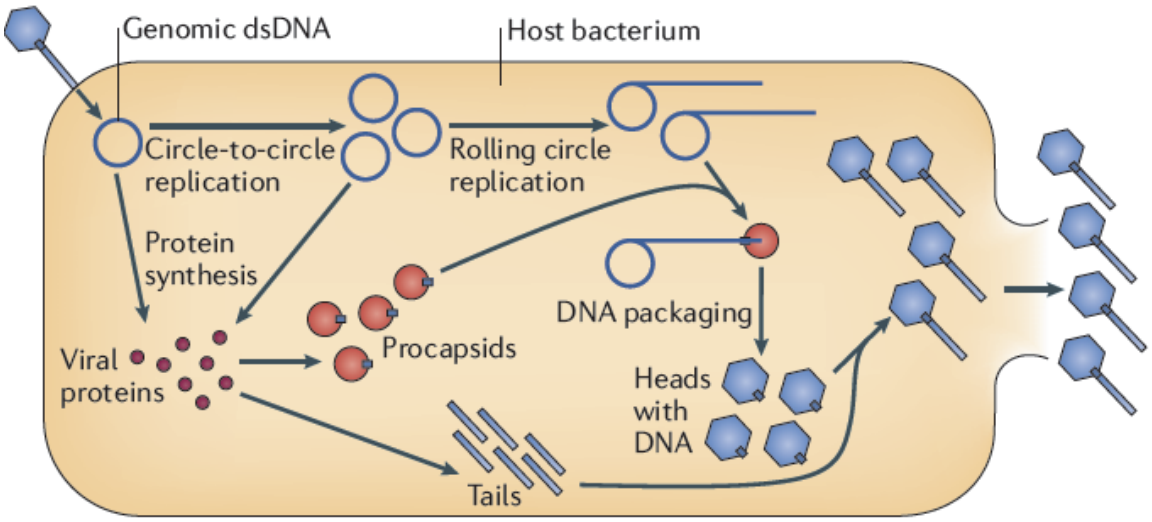


Figure 1-1 The major steps in the lytic life cycle of a typical tailed bacteriophage.
This figure was adapted from (Casjens, 2011).

The potential for using bacteriophages to treat bacterial infections was recognised early on, with phages being introduced as antibacterial agents in the 1920s (Summers, 2001). The development of phage therapy was stalled because of the discovery and wide application of antibiotics. However, with the recent emergence of multi-antibiotic resistant bacteria, phage therapy is now considered a promising tool to treat bacterial infections that cannot be cured by conventional antibiotics (Thiel, 2004). Phage therapy has been successfully applied in the agricultural, food-processing and fishery industries (Inal, 2003). As the understanding of bacteriophage biology advances, the potential of phage therapy as a means of treating or preventing human diseases is likely to become more evident.

1.2 The Virus Assembly Process

Research on the viral assembly process progressed rapidly with the isolation of conditional lethal mutants and the characterisation of the structures that accumulate in mutant-infected cells (Casjens & King, 1975). The viral assembly process, which proceeds along an ordered morphogenetic pathway, has been extensively studied. A more detailed understanding of the viral assembly process could enable the design of new antiviral agents, and lead to the development of new routes for designing nanoparticles (Harvey et al, 2009).

This thesis is concerned with proteins involved in the assembly of tailed dsDNA bacteriophages. It is a tightly regulated and strictly ordered process, which often involves shell formation, DNA encapsidation and tail attachment. Bacteriophage heads and tails assemble independently and join together to form mature virions following the translocation of DNA into heads. The empty protein shell, termed the "procapsid", can form spontaneously providing the capsid proteins are present at sufficient concentrations. In most cases, the formation of the procapsid is facilitated by other structural proteins, such as the auxiliary scaffolding proteins and the oligomeric portal protein. A DNA translocation molecular motor, located in one vertex of the procapsid, is responsible for driving the viral genome into the procapsid. The assembly process is often completed by the attachment of tail proteins. The morphogenic pathway of bacteriophage P22 is shown here (Figure 1-2) as a typical example of tailed dsDNA bacteriophages (Teschke & Parent, 2010).

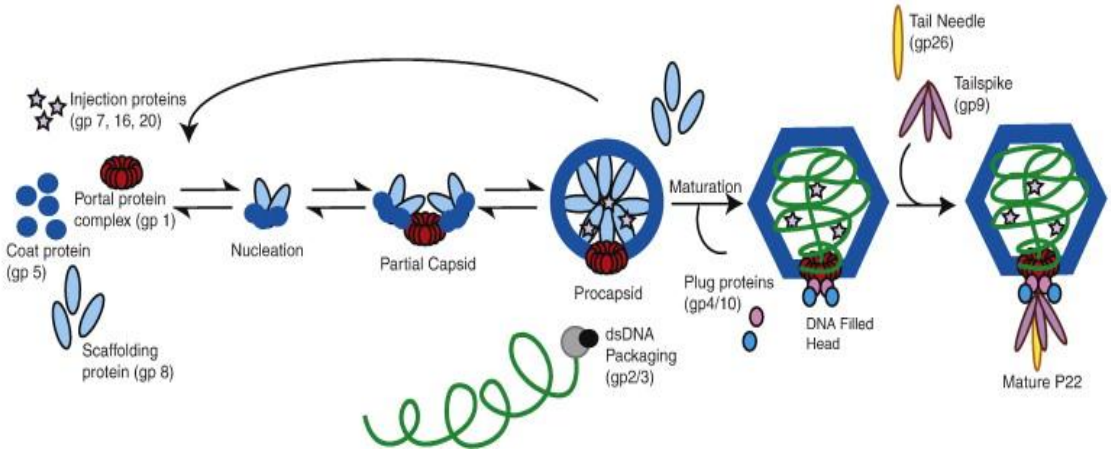


Figure 1-2 The assembly pathway of bacteriophage P22.
This figure was adapted from (Teschke & Parent, 2010).

1.3 The DNA Packaging Molecular Motor

The double-stranded DNA bacteriophages and herpes viruses adopt common machinery – a DNA translocating molecular motor - to encapsidate the genomic DNA into a preformed procapsid protein shell. The packaging motor is powered by ATP hydrolysis, converting chemical energy into mechanical force to pump DNA into the procapsid against large internal pressure (Casjens, 2011).

The DNA translocation process is initiated when the small terminase recognises and specifically binds to a recognition site on the concatemeric phage DNA (the *pac* site for phages P22, SPP1, T4 and T7 and the *cos* site for the λ-like and P2/P4-like phages) (Rao & Feiss, 2008). The small terminase also plays a role in recruiting the large terminase (Buttner et al, 2011). The large terminase displays endonuclease activity and makes an initial cut in the phage DNA at a specific site close to the recognition site. Following the initial cut, a small terminase - large terminase - DNA complex is formed and then docked to the portal protein, which is positioned at one vertex of the procapsid. Powered by ATP hydrolysis, the large terminase drives translocation of the viral DNA, via the portal protein, into the pre-formed procapsid. Once the capsid is filled, the large terminase is triggered to make a second cut in the DNA, thereby releasing the packaged DNA from the remainder of the concatemeric DNA (Rao & Feiss, 2008). Precisely one genome-length of DNA is packaged into phages in which both cuts are made in a sequence specific manner (for example bacteriophage λ). For phages in which the second cut is carried out in a sequence independent manner, slightly more than one genome length (102-110%) of phage DNA is normally packaged.

The properties of the DNA packaging molecular motor can be investigated by single-molecule experiments, involving the use of optical tweezers (Figure 1-3). In this technique, a single active procapsid-motor complex is tethered to a microsphere coated with an anti-capsid protein antibody, and the distal end of the biotinylated DNA is tethered to another microsphere coated with streptavidin. Both microspheres are captured in optical traps (the narrowest point of a focused laser beam), and packaging is initiated by moving the packaging motor bead into close contact with the fixed DNA bead. The dynamics of the DNA packaging

molecular motor such as translocation speed, force and step size can then be studied by manipulating the microspheres (Rao & Feiss, 2008; Smith et al, 2001).

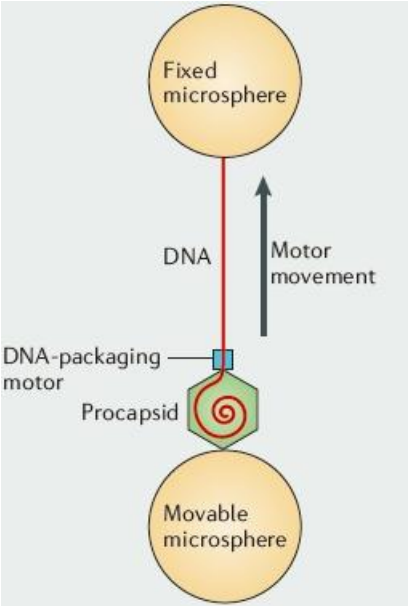


Figure 1-3 Single-molecule analysis of DNA packaging using optical tweezers.
This figure was adapted from (Casjens, 2011).

The bacteriophage DNA packaging molecular motor is one of the most powerful biological motors reported. The dsDNA molecule is as compact as crystalline DNA in most mature capsids (Earnshaw & Casjens, 1980). In order to package genomic DNA, the motor has to work against a high internal pressure that builds up inside the capsid with the filling of DNA, and overcome the extremely unfavourable energetic environment resulted from extensive DNA bending and charge repulsion of the compacted phosphate backbone. The force generated by the motors of phi29, λ, and T4 is approximately 60 pN (Smith et al, 2001).

An average translocation step size of two base pairs of DNA per molecule of ATP hydrolysed was deduced for phages phi29 and T3. The velocity of DNA translocation varies in different phages and appears to be correlated with the phage's genome size. The highest rate recorded to date was approximately 2000 bp/sec in the case of bacteriophage T4 (Fuller et al, 2007). As packaging proceeds and internal pressure increases, the translocation rate decreases and eventually falls to zero when approximately 100% of the genome is packaged (Smith et al, 2001).

The mechanism underlying the packaging of viral genomic DNA by the DNA packaging molecular motor is a fascinating research topic. A better understanding of the packaging motor will provide insights into the assembly of tailed bacteriophages, and also shed light on the potential applications of this nanomotor in nanotechnology and gene therapy. The structural basis of the motor components has been the subject of intensive research over the past two decades, and recent findings are reviewed in the following sections.

1.4 Components of DNA Packaging Motors

In general, a fully functional packaging motor relies on the co-operation of three essential components: the portal protein, the small terminase and the large terminase (Figure 1-4) (Casjens, 2011). One exception is the packing motor of phage phi29, which lacks the small terminase and instead requires a small 174-nt packaging RNA (pRNA), which docks the large terminase to the portal protein (Guo, 2002).

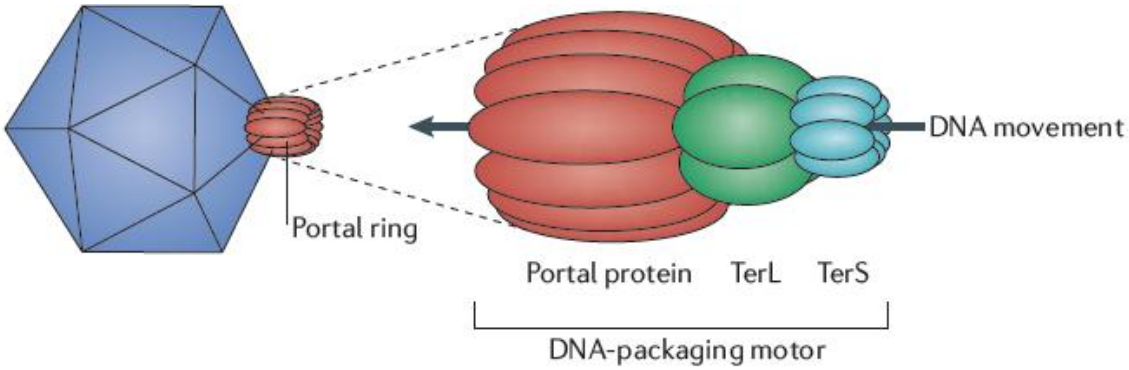


Figure 1-4 Schematics of the bacteriophage DNA packaging molecular.
TerL: Large terminase, TerS: Small terminase. The portal protein is shown as a dodecameric ring embedded in the 5-fold symmetrical vertex of an icosahedral bacteriophage capsid. The portal protein, large terminase and small terminase form the dsDNA translocating motor. The large and small terminase proteins are shown as rings made up of five and eight subunits, respectively, but their exact oligomeric states may differ in different bacteriophages. All known functional portal proteins are dodecamers (Rixon, 2008). The position of the small terminase is speculative. The arrow indicates the position and direction of dsDNA movement during translocation. This figure was adapted from (Casjens, 2011).

1.4.1 The Central Component - The Portal Protein

The portal protein assembly is located at a unique vertex of the procapsid shell, and forms a passage through which DNA can be translocated. As an essential component of the DNA packaging motor, the portal protein plays several indispensable roles in the assembly of viral particles. Not only can the portal protein initiate procapsid assembly and control the size of the procapsid, but it is also a central component of the DNA translocating molecular motor, a headful sensor for DNA packaging and a connector for tail attachment (Bazinet & King, 1988; Casjens et al, 1992; Droege & Tavares, 2005; Tavares et al, 1992).

The portal proteins from different bacteriophage species are highly diverged homologues, and there is no detectable similarity in amino acid sequences. In spite of the lack of homology between the amino acid sequences, a close resemblance in the overall architecture of bacteriophage portal proteins was revealed by X-ray structural and cryo-electron microscopy studies (Guasch et al, 2002; Lebedev et al, 2007; Leiman et al, 2004; Olia et al, 2011; Orlova et al, 2003; Simpson et al, 2000b). Portal proteins have a cone-shaped structure containing a central channel, with the wider end situated inside the capsid and the narrower end protruding out of the capsid. The channel is lined by 12 α-helices that expose several acidic residues which allow efficient passage of negatively charged DNA through the tunnel (Figure 1-5A). A single subunit of portal proteins can be subdivided into four regions, namely the clip, stem, wing and crown (Lebedev et al, 2007). The core domain of the portal protein from different phages demonstrated a strikingly similar fold, especially in the region spanning from helix α3 to helix α6 (Figure 1-5B), which is observed in three-dimensional structures of all portal proteins.

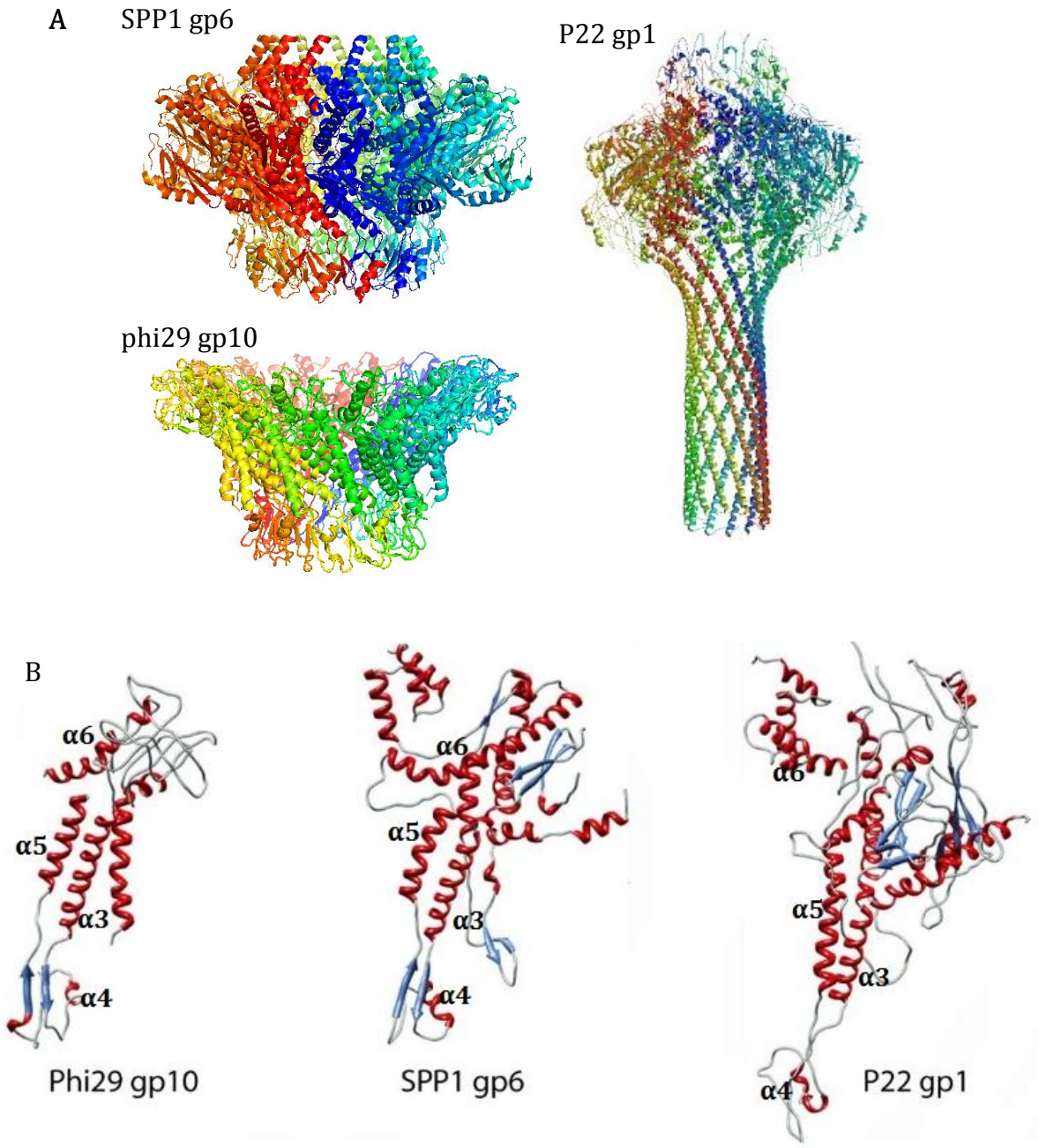


Figure 1-5 Conservation of the overall architecture and core domain of portal proteins

(A) Crystal structures of the portal proteins from bacteriophages phi29 gp10 (PDB: 1FOU), SPP1 gp6 (PDB: 2JES) and P22 gp1 (PDB: 3LJ4) (Lebedev et al, 2007; Olia et al, 2011; Simpson et al, 2000a); (B) The core domains of the portal proteins show that the four α-helices, helices α3 to α6, are observed in all three structures. Cyan arrows correspond to β-strands, red ribbons correspond to α-helices and loop regions are coloured in grey. Variability is localised in the wing regions, which differ in both conformation and size (Veesler & Cambillau, 2011) .

The molecular mass of different portal proteins varies greatly, with the smallest one found in phage phi29 (~35 kDa per subunit) and the largest one found in phage P22 (~83 kDa per subunit). In phage P22, the portal protein has a 200 Å long C-terminal extension, which forms an α-helical barrel and extends into the cavity of the capsid (Olia et al, 2011). In comparison with phi29, the extra molecular weight of the SPP1 portal gp6 (57.3 kDa per subunit) resides in the considerably larger wing and longer C-terminal region (Lebedev et al, 2007; Orlova et al, 2003). Given the large differences in size, the length and spatial arrangement of the helices in the core domain is surprisingly comparable. This conserved core segment accounts for 46% of the phi29 portal protein sequence and is likely to be an ancient structural unit with an important role in genome packaging (Lebedev et al, 2007). The conservation of the core architecture of portal proteins also suggests bacteriophages employ a similar mechanism to translocate genomic DNA into procapsids.

It is believed that the portal proteins exist exclusively as dodecamers in their biologically active state (Rao & Feiss, 2008). However, following ectopic expression and purification, the oligomeric state of the portal proteins from bacteriophages T7, T3, SPP1 and P22 has been found to range from an 11-mer to a 14-mer *in vitro*, and heterogeneous populations have also been reported (Cingolani et al, 2002; Dube et al, 1993; Kocsis et al, 1995; Trus et al, 2004a; Valpuesta et al, 2000). Portal proteins retain a similar overall architecture in different oligomeric states with conformational changes occurring in specific segments such as helix α6 of the crown region and tunnel loops (Lebedev et al, 2007). The factors determining the stoichiometry of the portal protein rings have not yet been identified conclusively.

1.4.2 The Large Terminase - the ATPase and Endonuclease of the DNA Packaging Molecular Motor

The genome of most dsDNA viruses is replicated as a concatemer without free ends, and a linear chromosome must be generated for packaging. The ter protein, from phage λ, was the first protein found to be required for termini formation (Mousset & Thomas, 1969). The term “terminase” remains in use because of the role of the terminases play in generating new DNA termini from viral genomic

concatemers. The nuclease activity actually only resides in the C-terminal domain of the large terminase (Bhattacharyya & Rao, 1993; Kuebler & Rao, 1998; Rentas & Rao, 2003), whereas the N-terminal domain, with its ATPase activity, fuels the motor to pump DNA into the procapsid. Interestingly, co-operation of the ATPase domain and the nuclease domain is thought to be crucial for the endonuclease activity. Evidence has shown that the nuclease activity of the C-terminal nuclease domain is completely lost, or significantly decreased, in the absence of the N-terminal domain (Alam et al, 2008; Alam & Rao, 2008; Smits et al, 2009).

2.1.1.1 The Nuclease Domain of the Large Terminase

During the packaging process, two events require nuclease activity, namely the initiation of DNA packaging and the "headful" termination cuts. The initial cleavage of the concatemer is commonly sequence specific, and occurs near the recognition site of the small terminase, after the large terminase has been recruited to the small terminase:DNA complex (Gual et al, 2000). The packaging of genomic DNA starts from the end generated by this cut. The second cleavage is triggered when the virus particle is fully packaged. For λ-like phages the second cut is also sequence specific, and precisely one genome-length of DNA is packaged accordingly. For phages such as SPP1, T4 and P22, the second cut is sequence-independent and triggered by a headful-sensing mechanism when the capsid has been filled with DNA (Rao & Feiss, 2008). In this case, slightly more than one genome-length of DNA (102 -110%) is packaged (Smits et al, 2009).

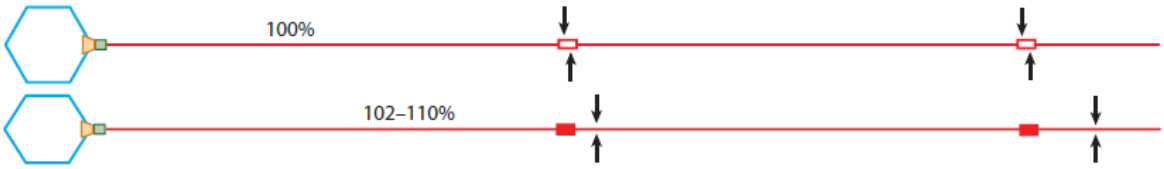


Figure 1-6 The headful packaging of genome DNA by tailed dsDNA phages.
Top: one genome-length of DNA is packaged by λ-like phages with a sequence specific second cut; Bottom: "headful filling" packaging with a sequence-independent second cut as in phages SPP1, T4 and P22, etc. This figure was adapted from (Rao & Feiss, 2008).

Three-dimensional crystal structures of the nuclease domain of the large terminase have been determined for bacteriophages P22, SPP1, T4 and T4-like

RB49 (Alam et al, 2008; Smits et al, 2009; Sun et al, 2008). Despite displaying low sequence homology, these proteins all share a similar RNase H fold, comprising a central β-sheet sandwiched between two clusters of α-helices, and conserved catalytic acidic residues (Smits et al, 2009). Notably, a structure-based alignment revealed that the large terminases possess several distinct, conserved features, for example extension of the central β-sheet and a C-terminal β-hairpin (β8 and β9) (Smits et al, 2009). These additional structural elements are not shared with other members of the RNase H family, indicating their importance for the function of the terminase.

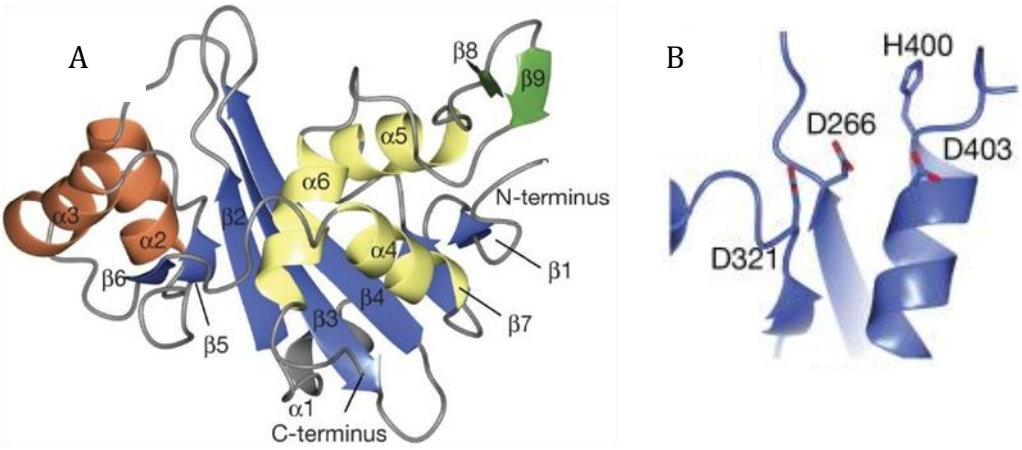


Figure 1-7 The structure of the nuclease domain of the SPP1 large terminase.
The central β-sheet (blue) is surrounded by a three-helix bundle on the concave face (yellow) and two helices on the convex face (orange). The C-terminal β-hairpin (β8 and β9) is shown in green, as shown in (A). The three conserved acidic residues are shown in (B). This figure was adapted from (Smits et al, 2009).

RNase H family members catalyse the hydrolysis of RNA-DNA hybrids by means of a two metal ion-dependent mechanism (Tadokoro & Kanaya, 2009). During catalysis, it has been proposed that one metal ion activates a nucleophilic hydroxyl group and the other metal ion stabilises the transition state. Conserved acidic residues co-ordinate the two metal ions and are required for catalysis. The large terminases of some bacteriophages, such as bacteriophage λ and RB49, require magnesium ions (Mg^{2+}) for catalysis (Sun et al, 2008; Tomka & Catalano, 1993). In contrast, the large terminases of the bacteriophage SPP1 and the evolutionarily related HSV-1 depend on manganese ions (Mn^{2+}) (Cornilleau et al, 2013; Nadal et al, 2010). Sequence-based and structure-based alignments of large terminase

proteins revealed three highly conserved acidic residues within the catalytic site. Mutagenesis analyses showed that these three acidic residues are essential for the nuclease activity of the large terminases, (Alam et al, 2008; Smits et al, 2009; Sun et al, 2008).

On the basis of the structures, several mechanisms have been proposed to explain the headful nuclease activity of the large terminases that occurs following DNA packaging. Due to the conformational changes that occur in the portal protein upon maturation (Figure 1-8), it was proposed that expansion of portal protein might potentiate the dissociation of the large terminase from the portal protein and thus trigger DNA cleavage (Alam et al, 2008; Xiang et al, 2006). After defining the unique fold of the C-terminal β-hairpin in the large terminase, a mechanism was proposed based on the regulation by the β-hairpin. The β-hairpin is flexible and can therefore adopt different conformations which can obstruct or promote interactions between large terminase with substrate DNA molecule (Smits et al, 2009).

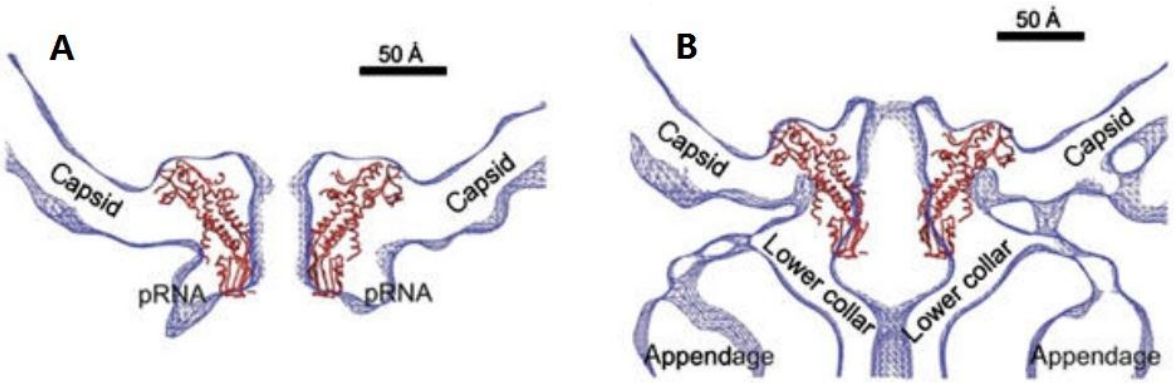


Figure 1-8 Conformational changes of the portal protein upon phage maturation.
(A) Model of the portal protein crystal structure (C_α trace in red) fitted into the cryo-EM density of the five-fold averaged reconstruction of the procapsid, prior to DNA packaging; (B) The cryoEM density of the asymmetric reconstruction of mature viral particles, following removal of the packaged DNA. The narrow end of the portal protein would have to increase its radius in order to fit into the density (dark blue) of the mature viral particle. This figure was adapted from (Xiang et al, 2006).

2.1.1.2 The ATPase domain of the large terminase

Viral genomic DNA is packaged into the procapsid to near-crystalline density, using the energy released from ATP hydrolysis (Sun et al, 2007). The ATPase activity resides in the N-terminal domain of the large terminase proteins (Kanamaru et al, 2004; Rao & Mitchell, 2001). The N-terminal ATPase catalytic centre is strictly conserved in terms of sequence signatures and secondary structure motifs among different bacteriophages, which contains the adenine-binding motif, catalytic carboxylate, the walker A motif and the walker B motif (Feiss & Rao, 2012; Goetzinger & Rao, 2003; Walker et al, 1982).

The atomic structure of the N-terminal domain of the T4 large terminase gp17 with the D255E/E256D mutations (Figure 1-9) revealed a flat structure consisting of two spatially separate subdomains (Sun et al, 2007). Subdomain I contains the nucleotide-binding domain (NBD, Figure 1-9) with the characteristic Rossmann-fold motif - a conserved β-sheet core with six parallel β-strands (Rao & Feiss, 2008; Rossmann et al, 1974; Sun et al, 2007). The apo- and ADP/ATP-bound structures are very similar. The only difference is the conformation of the adenine binding loop, which may reflect the state of the ATP binding pocket (Sun et al, 2007). Subdomain II contains residues 1–58 from the N-terminus and residues 314–360 from the C-terminus.

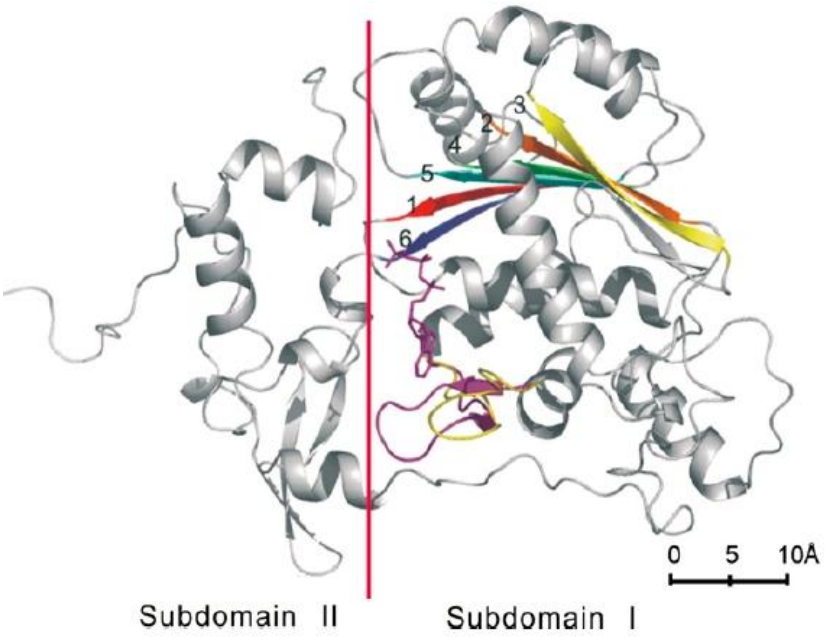


Figure 1-9 The structure of the N-terminal domain of the T4 large terminase.

A ribbon diagram shows the structure of the N-terminal domain of the large terminase gp17 N360-ED mutant from the bacteriophage T4. The Rossmann-fold motif in Subdomain I follows a 324516 topology and the β strands are coloured in sequence with red, orange, yellow, green, cyan, and blue. The conformational change in the adenine binding loop is shown between the ATP-bound state (purple) and the apo state (yellow). This figure was adapted from (Sun et al, 2007).

1.4.3 The Small Terminase

The small terminase is the third component of the DNA packaging motor and is responsible for recognising the viral DNA for the initiation of packaging (Camacho et al, 2003). The small terminase binds to specific sequences in the viral DNA (either *cos* or *pac* sites) from where packaging is initiated (Lin et al, 1997). The role of the small terminase in the packaging process is not yet fully understood. It may actively participate in the regulation of the ATPase and nuclease activities of the large terminase during DNA translocation (Buttner et al, 2011). However, small terminases are not necessarily required for packaging pre-cut DNAs in most *in vitro* packaging systems (T4, T3, and λ) (Hamada et al, 1986; Rao & Black, 1988; Rubinchik et al, 1995).

Biochemical and mutational analyses revealed that the small terminase consists of three domains, namely an N-terminal DNA-binding domain, a central oligomerisation domain, and a C-terminal large terminase-binding domain (Al-Zahrani et al, 2009). Over the past decade, considerable knowledge has been accumulated regarding the structure of the small terminases. The structures of the small terminase from several bacteriophages have been determined, including the cryo-EM and crystal structures of the small terminase from phage P22 (Nemecek et al, 2008; Roy et al, 2012), the NMR structure of the DNA-binding domain of the small terminase gpNu1 from phage λ (de Beer et al, 2002), the full-length crystal structures of the phage Sf6 (Podoviridae family) small terminase gp1 and the phage SF6 (Siphoviridae family) small terminase G1P (Buttner et al, 2011; Zhao et al, 2010) and the crystal structure of the central domain of the small terminase from a T4 family phage, phage 44RR, in two different oligomeric states (Sun et al, 2012).

Although there is no significant overall sequence homology, the 3-D structures of the small terminases revealed several common structural motifs required for the DNA packaging process (Rao & Feiss, 2008). A helix-turn-helix (HTH) motif located at the N-terminus is often associated with DNA binding (Brennan & Matthews, 1989). The central domain, consisting of two long antiparallel α-helices, plays a key role in oligomerisation (Kondabagil & Rao, 2006). The C-terminal domain is composed mainly of β-strands, which form an intersubunit β-Barrel. The crystal structure of the full-length small terminase from phage SF6 is shown in Figure 1-10.

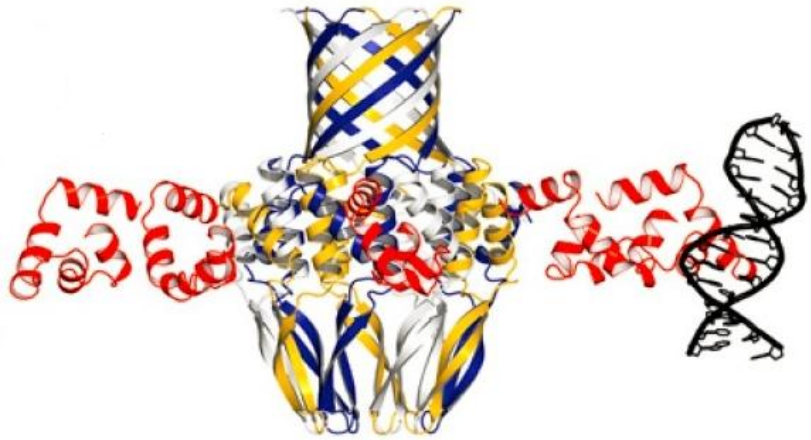


Figure 1-10 A ribbon diagram of the full-length SF6 small terminase.
For clarity, DNA binding domains (red ribbons) are shown only for five subunits. A 10-bp dsDNA is fitted with the helix-turn-helix motif of a single subunit. The C-terminal β-barrel and the main body of the oligomerisation core domain are shown with individual subunits in alternating colours (white, yellow, blue). This figure was adapted from (Buttner et al, 2011).

The oligomeric state of the small terminase varies from eightfold to twelve-fold, with full-length Sf6 small terminase dispaying eight-fold symmetry, the SF6 small terminase displaying nine-fold or ten-fold symmetry and the T4 family phage 44RR small terminase displaying eleven-fold or twelve-fold symmetry (Sun et al, 2012). Interestingly, the full-length SF6 small terminase exclusively forms nine-subunit assemblies, whereas protein constructs missing the C-terminal β-barrel form both nine-subunit and ten-subunit assemblies, indicating the importance of the C-terminus for defining the oligomeric state (Buttner et al, 2011).

1.5 Aims of the Project

As described above, the portal protein assembly is the major component of the DNA packaging molecular motor, which translocates the viral genomic DNA into a preformed procapsid protein shell fuelled by the energy of ATP hydrolysis. The oligomeric portal protein ring is situated at a unique vertex of the procapsid shell to form the conduit for DNA entry and exit. The SPP1 portal protein is a circular homo-oligomer with an intrinsic 12-fold symmetry incorporated into the viral capsid, but forms a 13-mer assembly in its isolated form. Although the three-dimensional structure of the isolated SPP1 portal protein is available, it is not for the wild-type form but for N365K mutant. This mutation results in reduction of the length of packaged DNA (Tavares et al, 1992). During my PhD studies, my work was mainly focused on the structural studies of the portal proteins from double-stranded DNA bacteriophages SPP1, T4 and CNPH82.

My major aim was to produce gp6 in its biologically relevant dodecameric state for *in vitro* studies, to gain three-dimensional structural information on the wild-type portal protein and to understand structural reasons that result in the reduced length of packaged DNA in the case of capsid containing N365K portal protein. In order to promote the formation of the dodecameric gp6, two strategies involving protein engineering of fusion proteins and co-expression of gp6 with gp11, were designed and investigated using a combination of biochemical, biophysical and structural approaches.

Finally, the portal proteins from bacteriophages CNPH82 and T4, cn3 and gp20 respectively, were also studied to gain more general information about the assembly of portal proteins from other phages. The goal of this research was to explore methods of producing soluble proteins and to obtain diffracting crystals for structural studies.

2 Materials and Methods

2.1 Introduction

In this chapter, the general experimental procedures involved in the projects such as molecular cloning, protein production, purification and characterisation, protein crystallisation as well as macromolecular X-ray crystallography are summarised. The application of relevant techniques will be detailed in the following chapters.

2.2 Molecular Cloning

As a fundamental laboratory technique, molecular cloning facilitates the production of large quantities of protein, which would otherwise naturally be only found in small amounts, for use in structural and functional studies. It is also used to engineer protein molecules by allowing the introduction of mutations, truncations and tags etc. to the target protein. The process involves the insertion of a foreign DNA fragment into a vector capable of replicating autonomously in a host cell (usually *Escherichia coli*). Multiple copies of the inserted DNA will then be produced with the host cell growth.

2.2.1 Insert Preparation

2.2.1.1 Polymerase Chain Reaction (PCR)

The polymerase chain reaction (PCR) has revolutionized molecular biology research since its invention by Kary Mullis in 1983 (Saiki et al, 1985). Nowadays, it has become a routine technique to amplify a large number of copies of a specific DNA fragment *in vitro* from trace amounts of the template material. PCR has a variety of applications in biological and medical research: gene isolation, DNA/genome sequencing, detection, diagnosis of genetic diseases, etc.

Three distinct steps controlled by temperature are involved in PCR reactions – denaturation, annealing and elongation. Prior to the beginning of a PCR reaction, there is normally an extended denaturation step at ~95 °C for 2-5 minutes to ensure efficient denaturation of the template DNA. Double-stranded DNA template is first heated typically to 95 °C to separate its complementary single strands

during denaturation process. The reaction is then rapidly cooled to allow for primer-template hybridization, usually between 40 °C and 72 °C dependant on the melting temperature (*Tm*) of the oligonucleotide primers. Template single strands are too long and complex to reanneal within the same time frame. At last, DNA will be synthesised by a thermostable DNA polymerase at its optimal working temperature (e.g., 72 °C for Taq DNA polymerase). 20-40 cycles of the three PCR steps should be performed to obtain sufficient DNA products for a specific application purpose. Finally, there is an extended 72 °C incubation step to ensure full-length synthesis of PCR products.

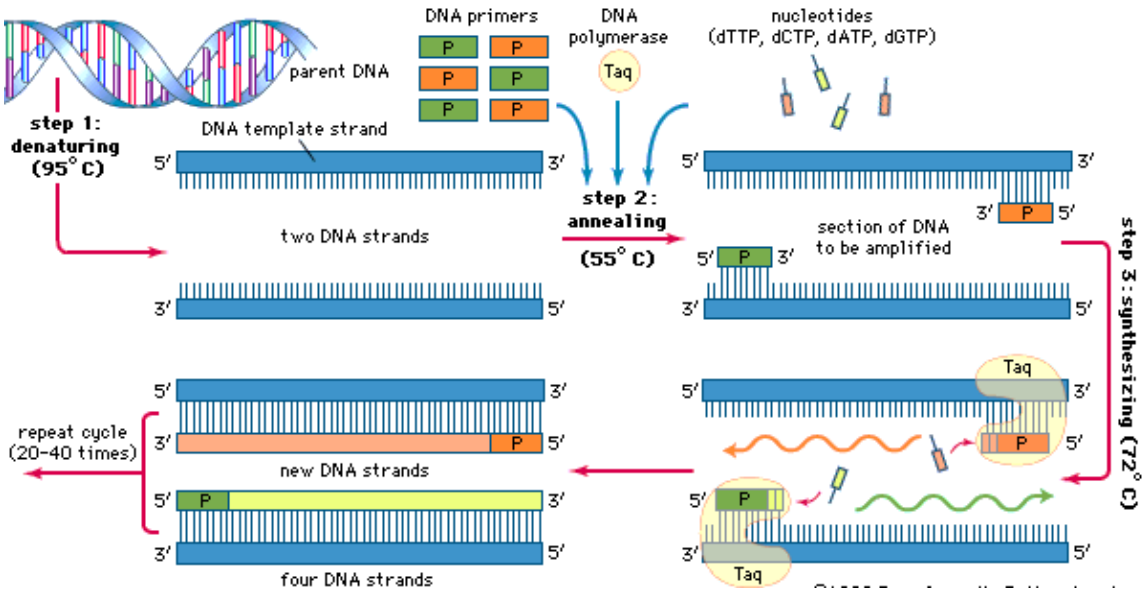


Figure 2-1 Schematic diagram of the PCR process.
The three steps involved in PCR reactions – denaturation, annealing and elongation are illustrated. This figure was adapted from British Encyclopaedia.

Short synthetic oligonucleotides are required in a PCR reaction, because DNA polymerases can only add new nucleotides to the 3' end of an existing strand. Primers are designed to be reverse complementary sequences of the template DNA, and play a crucial role in guiding a precise DNA synthesis process. Optimal primer design is the most critical parameter for a successful PCR. Poorly designed primers can result in PCR reaction failure due to nonspecific amplification and/or primer-dimer formation even if all other parameters properly optimised (Apte & Daniel, 2009). Several factors should be taken into consideration for a successful PCR primer design, such as length, G-C content and Tm. It is crucial that the

forward primer and reverse primer have similar melting temperature, since annealing ought to occur simultaneously for both primers. Meanwhile, formation of secondary structures by intermolecular or intramolecular interactions should be avoided, because the presence of primer secondary structures can lead to poor or no yield of the PCR product.

2.2.1.2 Agarose Gel Electrophoresis

Agarose gel electrophoresis is an easy and effective way to separate, identify, and purify DNA fragments. DNA fragments are separated according to their sizes and visualised under a UV transilluminator after staining with a fluorescent dye such as ethidium bromide or SYBR® Safe stain. When an electrical field is conducted, DNA, negatively charged at neutral pH due to its phosphate backbone, will migrate toward the anode. The migration rate of a linear DNA fragment through an agarose gel is inversely proportional to the $\log_{10}$ of its molecular weight. The migration rate can also be affected by the agarose concentration, the conformation of the DNA, the applied current, base composition and temperature. DNA fragments with sizes varying from 50 base pairs to several mega base pairs can be separated with appropriate percentage of agarose gel (from 0.75% to 2.0%) and electrophoresis condition. A DNA ladder normally runs alongside samples, as a reference to estimate the size of samples.

A solution of agarose at an appropriate concentration (w/v) for separating DNA fragments with particular sizes was prepared in 1x TAE electrophoresis buffer containing 40mM Tris pH 8.0, 20mM acetic acid, and 1mM EDTA . Agarose slurry was heated in a microwave oven until fully dissolved. SYBR® Safe DNA gel stain (10,000x concentrate in DMSO) was then added to agarose gel solution to a 1x final concentration after the gel solution cooled. DNA samples were mixed with blue or orange dye before loading, and a suitable DNA ladder (e.g., 2-Log DNA Ladder, New England Biolabs) was also loaded alongside samples. Agarose gels were usually run at 100-120 V for approximately 60 minutes. Gel pictures were taken on the Syngene Bio Imager.

2.2.1.3 Quantitation of DNA Concentration

DNA concentration can be quantified by optical density measurement, because nucleic acids maximally absorb UV light at 260 nm wavelength.

Spectrophotometric analysis provides an easy way to assess the concentration of DNA in solution, especially with a NanoDrop Spectrophotometer - a microvolume spectrophotometer that utilises a patented sample retention system and thus eliminates the need for cuvettes and capillaries. 2 µL of sample is pipetted directly onto the pedestal when the arm is open. With a closed arm, a sample column is formed between two optical surfaces. Then, the pedestal moves to automatically adjust for an optimal path length in vertical direction (0.05 mm - 1 mm). The concentration range of dsDNA varying between 2 ng/µL and 3700 ng/µL can be assessed by the measurement of absorbance at 260 nm wavelength using NanoDrop Spectrophotometer.

2.2.2 Vector Preparation

2.2.2.1 Vector

Vectors are DNA molecules carrying exogenous DNA insert into host cells for the purpose of DNA cloning. There are three common features for all cloning vectors : (i) a replication origin for independent replication; (ii) a multiple cloning site (also called the polylinker region) with a number of unique restriction endonuclease cleavage sites where exogenous DNA fragments can be inserted; (iii) a selectable marker, usually antibiotic resistance genes or enzyme genes absent in the host cell, to distinguish positively transformed cells. Four major types of vectors were categorised based on the capacity to accommodate insert DNA into plasmids, viral vectors, cosmids, and artificial chromosomes with insert size of 10, 20, 45 and several hundred kilobases (kb), respectively.

2.2.2.2 Plasmid Vector

Plasmids are circular double-stranded DNA molecules, which can be found in almost all bacteria and yeast, and in some fungi, protozoa, plants and animals. They are separated from the chromosomal DNA as extrachromosomal self-replicating DNA molecules, and range in size from 1 to over 1000 kb. Plasmids from *E. coli* are commonly used for molecular cloning, as many have been artificially constructed for the optimal usage as vectors - the length dramatically reduced for easy manipulation with little more than the above-mentioned three essential elements. The pET vector system were originally designed by Studier and colleagues in 1986 (Studier & Moffatt, 1986) and further developed at Novagen. It

is the most widely used vector system for gene cloning and expression in *E. coli,* with different selectable markers and different characteristics. Transcription of the insert DNA is controlled by the strong T7 RNA promoter, which is specifically recognised by bacteriophage T7 RNA polymerase. T7 RNA polymerase is not naturally existed in *E. coli* cells, and therefore host *E. coli* cells are genetically engineered to incorporate the T7 RNA polymerase gene in the host chromosome under the control of the *lac* promoter. Once Lactose or lactose analogue IPTG (isopropyl-beta-D-thiogalactopyranoside) triggers the production of T7 RNA polymerase, transcription and translation of the insert DNA will be induced subsequently (Garrett & Grisham, 2010).

2.2.2.3 pBlueScript II Phagemid System and Blue/White Screen

The pBluescript II phagemids - plasmids with a phage origin and 21 restriction enzyme recognition sites in the polylinker region, are designed for cloning, sequencing, *in vitro* mutagenesis, and so on. The polylinker region is located within the lacZ' gene, which encodes the N-terminal fragment of β-galactosidase also known as the α-peptide. Host *E. coli* cells with *lacZ* deletion mutant can only produce inactive mutant β-galactosidase, so a α-peptide produced by pBluescript vector is required to rescue the function of β-galactosidase and cleave colourless X-gal into blue coloured 5-bromo-4-chloro-indoxyl. Thus, blue coloured colonies indicate the existence of uninterrupted vectors with no insert DNA.

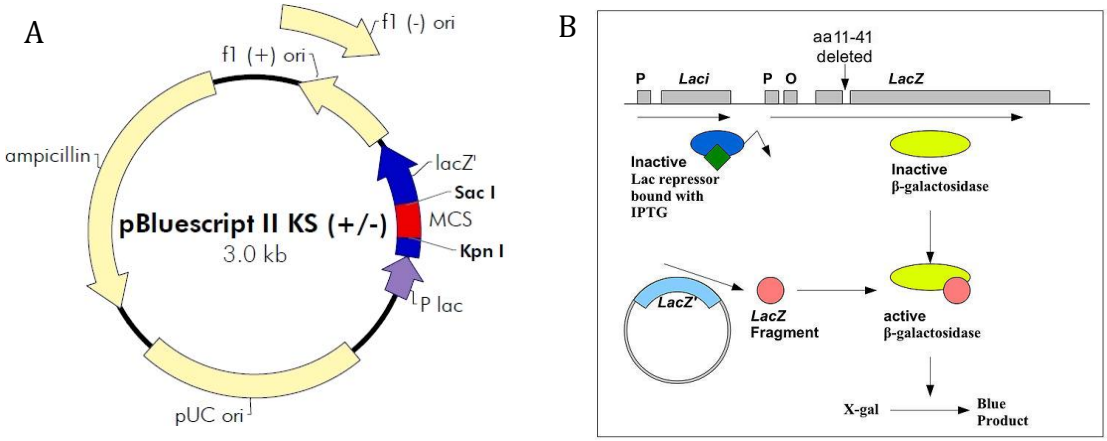


Figure 2-2 Schematic representation of blue white screen
(A) pBluescript II KS (+/-) vector, and (B) blue white screen. This figure was adapted from Stratagene.

2.2.3 Ligation Dependent Cloning

A restriction enzyme digestion step and a ligation step are required for the classical ligation dependent cloning approach. The vector and insert DNA should contain the same endonuclease recognition site in order to be recognised and cut by the specific restriction endonucleases. Complementary sticky ends or blunt ends will be generated after a digestion step. The digested insert and linearised vector are joined together with covalent bonds by DNA ligase, most commonly T4 DNA ligase. Dephosphorylation of linearised vectors to remove the 5' phosphate group and prevent self-ligation of vectors is a crucial step for a successful ligation reaction.

Double digestion using two different restriction endonucleases is usually performed to generate non-complementary sticky ends, which will inhibit self-ligation of vectors and ensure target DNA fragments are incorporated into vectors in the desired orientation. In the digestion reaction, an appropriate amount of DNA, restriction endonucleases, the particular buffer for the specific restriction endonucleases and sometimes BSA protein solution are mixed and incubated at the optimal temperature for enzyme activity, usually 37 °C, for about one hour. Composition of a typical restriction digestion reaction is shown in Table 2-1. Most restriction enzymes can be inactivated by heating at 65 °C for 20 minutes. An extra dephosphorylation of vector is performed by incubation with 1 µL Antarctic phosphatase (5 U/µL) and 1X phosphatase buffer at 37 °C for 20 minutes and subsequent heat inactivation at 65 °C for 10 minutes.

Table 2-1 Composition of a typical double digestion reaction

Component	Amount
Insert/ Vector	~1 µg
Restriction endonuclease I (20 U/ µL,NEB)	5 µL
Restriction endonuclease II (20 U/ µL,NEB)	5 µL
Restriction endonuclease buffer (10X,NEB)	5 µL
Milli-Q water	Up to 50 µL

For a ligation reaction to take place, three main components are required: (1) DNA fragments with compatible ends; (2) A buffer containing 0.25 mM - 1 mM ATP to provide the necessary energy for the reaction; and (3) T4 DNA ligase. Composition

of a typical restriction digestion reaction is shown in Table 2-2, and the ideal ratios for ligating insert to vector with sticky ends ranges between 1:1 and 3:1. 5 µL of ligation product is transformed into 100 µL of *E. coli* competent cells after overnight ligation at room temperature or 16 °C.

Table 2-2 Composition of a typical ligation reaction

Component	Amount
Double digested vector	100 ng
Double digested insert	X ng*
T4 DNA ligase (400 U/ µL, NEB)	1 µL
T4 DNA ligase buffer (10X,NEB)	2 µL
Milli-Q water	Up to 20 µL

*the amount (ng) of the double digested insert added in a typical ligation reaction is normally calculated by the formula:

$$Insert\ Mass\ in\ ng = 6 \times \left[\frac{Insert\ Length\ in\ bp}{Vector\ Length\ in\ bp}\right] \times Vector\ Mass\ in\ ng$$

2.2.4 Transformation

Transformation is the process of introducing foreign DNA into bacterial strains. With successful transformation, the bacteria cells will inherit specific characteristics owing to the acquisition of foreign DNA. The delivery of foreign DNA into competent cells can be achieved via two main approaches - chemical transformation and electroporation. Only chemical transformation is employed in the projects reported here. During chemical transformation, bacterial cells are heat-shocked in a water bath, which opens pores on the cell membrane to allow for the entry of plasmid DNA into cells.

5 µL of the ligation product was added into 100 µL of freshly thawed competent *E. coli* cells, mixed gently and incubated on ice for 30 minutes. Exactly one minute heat shock was carried out at 42 °C, followed by a further incubation on ice for 2 minutes. 200 µL LB medium was added and the mixture was incubated in a shaker at 37 °C for one hour. The transformed cells were plated out on a LB-agar plate with appropriate antibiotics. After overnight incubation at 37 °C, colonies growing on the plate should carry foreign DNA, however further verification is necessary to

confirm the successful transformation. Meanwhile, appropriate positive and negative controls were treated identically for results interpretation.

For general cloning, *E. coli* strains DH5α and NovaBlue were used, because they are competent for plasmid maintenance with high transformation efficiency and high plasmid yield. For the purpose of protein production, BL21 (DE3 lys) and B834 (DE3 lys) strains were used, because the *lacI* gene, *lacUV5* promoter and T7 RNA polymerase gene are incorporated in the chromosome of these strains to make them suitable for hosting the pET vector system. Additionally, the lack of *lon* protease and *ompT* outer membrane protease in these strains may prevent protein degradation during purification process.

2.2.5 Colony PCR

The appearance of numerous antibiotic resistant *E. coli* colonies after transformation can be a good indicator of a successful cloning reaction. However, further verifications by colony PCR or/and restriction analysis always prove to be necessary to confirm the existence of foreign DNA insert, especially when a significant number of colonies are also shown on the control plate.

Colony PCR is a fast way to screen numerous colonies, during which the PCR reaction will only work on condition that the insert DNA is present using vector flanking primer pairs. PCR premix solution was prepared according to the number of colonies for screening (Table 2-3). Some random colonies was picked using a sterile yellow tip, and transferred into the reaction mixture by pipetting up and down several times and streaked onto a fresh replicate agar plate. If bands relating to the PCR products with the expected lengths were observed on an agarose gel, the corresponding colonies growing on a replicate agar plate were cultured.

Table 2-3 Components to set up colony PCR premix

Components	Final concentration	1 x pre-mix
Primer 1 (2 μM)	0.2 μM	2.5 μL
Primer 2 (2 μM)	0.2 μM	2.5 μL
10 X PCR buffer	1X	2.5 μL
dNTPs (25 mM each)	0.2 mM	0.2 μL
Taq Polymerase (5 U/μL)	0.05 U	0.25 μL
dH_2O		17.05 μL

2.3 Expression of Recombinant Protein

The production of recombinant protein requires an expression system involving a combination of a plasmid and a host (Sorensen & Mortensen, 2005). In microbial expression systems, *Escherichia coli* is a widely used host for recombinant expression of heterologous proteins (Baneyx, 1999). In this study, DE3 lysogenic strains of BL21 and B834 were used to host pET vector system for the over-expression of target proteins controlled by T7 promoter.

2.3.1 Small Scale Protein Expression

A small-scale protein expression experiment is normally performed before a large-scale cell culture and purification, to optimise conditions for the expression and solubilisation of the target protein. The key factors involved for optimal protein expression and solubility are as follows: induction condition, cell strains, growth medium, incubation temperature and lysis buffer (Sivashanmugam et al, 2009).

For small-scale expression trials, two vials of 5 mL LB media with appropriate antibiotics were inoculated with starter cultures in a ratio of 1:100 (v/v) at 37 °C until an OD_{600} of between 0.6 and 0.8 was reached. 1 mL of cell culture was harvested by centrifugation and the cell pellet was kept at – 20 °C, before addition of 1 mM IPTG to the rest of the cell culture. Overnight incubation at 16 °C or three to five hours incubation at 37 °C was carried out for the over-expression of the target protein. Four aliquots of 1 mL cell culture were harvested by centrifugation

at 13000 g for one minute after OD_{600} measurement. Cells were re-suspended in 200 µL lysis buffer and lysed by sonication (10 s, 3 cycles, 50% power). Cell debris was pelleted by centrifugation at 13000 g for 10 minutes. 15 µL of lysate containing total proteins and supernatant containing soluble proteins were analysed by SDS-PAGE to evaluate the expression, yield and solubility of the target protein in different cell strains, induction temperatures and lysis buffers. The expression conditions were optimised based upon the pilot experimental results to maximise the production of target proteins. High-throughput protein expression screenings and buffer condition screenings can be performed to identify the best buffer condition for solubilisation of the target protein.

The auto-induction strategy was developed by F. W. Studier based upon the ability of lactose to induce protein expression in *E. coli* when cells reach saturation (Grabski et al, 2005; Studier, 2005). In the growth medium ZY-P-5052 for auto-induction, glucose is used as the early carbon source and prevents induction by lactose, while glycerol is the late carbon source. With the depletion of glucose, lactose can enter cells to turn on the synthesis of the T7 RNA polymerase. Cell cultures are inoculated and incubated for 20-24 hours at 37 °C with shaking at 300 rpm to reach a typical cell density of approximately OD_{600} 5-6. The yields of proteins produced by the auto-induction method are typically several fold higher than the conventional way involving IPTG induction (Studier, 2005).

2.3.2 Large Scale Protein Expression

Once optimum conditions have been identified for the over-expression of a specific recombinant protein, the volume of cell culture can be scaled up to produce sufficient proteins for purification and further structural or functional studies. Overnight cell culture of a single colony was used to inoculate a large volume of LB medium, usually 1 litre to 2 litres, in a ratio of 1:100. The over-expression of the target protein was conducted under the optimum conditions determined by the expression test. Cells were harvested by centrifugation at 5000 rpm for 30 minutes at 4 °C (JLA-8.100 rotor, Beckman Coulter Avanti J-HC). After decanting the supernatant, cell pellets were re-suspended with lysis buffer (5 mL buffer per gram cell) for purification, or stored at -80 °C.

Table 2-4 Components of auto-induction medium for protein expression

Stock solutions	
100 mL 50× 5052	25 g glycerol 2.5 g glucose 10 g α-lactose 73 mL water autoclave
100 mL 20× NPS	6.6 g $(NH_4)_2SO_4$ 13.6 g KH_2PO_4 14.2 g Na_2HPO_4 90 mL water autoclave N.B. pH of 20-fold dilution in water should be ~6.75
1 L ZY	10 g N-Z-amine AS (or tryptone) 5 g Yeast extract 925 mL water autoclave
100 mL 1000× metals	Autoclave individual solutions before mixing
	50 mL 0.1M $FeCl_3$-$6H_2O$ (in 0.1M HCl) 2 mL 1M $CaCl_2$ 1 mL 1M $MnCl_2$-$4H_2O$ 1 mL 1M $ZnSO_4$-$7H_2O$ 1 mL 0.2M $CoCl_2$-$6H_2O$ 2 mL 0.1M $CuCl_2$-$2H_2O$ 1 mL 0.2M NiCl2-$6H_2O$ 2 mL 0.1M Na_2MoO_4-$5H_2O$ 2 mL 0.1M Na_2SeO_3-$5H_2O$ 2 mL 0.1M H_3BO_3 36 mL water
To make 1L of auto-induction medium with kanamycin	
1M $MgSO_4$ (sterile)	1 mL
1000× metals	1 mL
50× 5052	200 mL
Mix well	
20× NPS	50 mL
30 mg/mL kanamycin	1 mL
ZY	To 1 L

2.4 Protein Purification

Obtaining pure protein is a prerequisite for structural studies and biochemical assays. Isolation of a specific protein from a crude mixture requires individual purification strategies dependent upon the properties of the desired protein, such as size, physico-chemical properties, binding affinity and biological activity. The key to successful and efficient protein purification relies in the application of appropriate purification methods and procedures specific to an individual protein.

2.4.1 Cell Lysis and Protein Extraction

Cell lysis is the first step for protein extraction and purification, because cellular contents are separated from the extracellular environment by a plasma membrane. The plasma membrane of bacterial cell is also surrounded by a rigid peptidoglycan cell wall.

Physical disruption using a sonicator is a common way to extract cellular contents. Before sonication, cell pellets were re-suspended with 5 mL/g freshly prepared lysis buffer. Additive components in the lysis buffer included lysozyme and protease inhibitors such as serine protease inhibitor 4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF) and aspartyl proteases inhibitor pepstatin. Lysozymes can help to damage bacterial cell walls, and protease inhibitors are required to prevent protein degradation by unregulated endogenous protease. Since overheating may cause the formation of aggregates (Stathopulos et al, 2004), it is crucial to keep the sample on ice and have an interval between each sonication period. Four to five rounds of 20-second sonication pulses were performed using a Soniprep 150 sonicator (MSE) with two minutes interval between each sonication pulse. After sonication, the lysate was cleared by centrifugation to precipitate cell debris, and the supernatant was syringe filtrated through a 0.45 μm filter to remove any remaining aggregates.

2.4.2 Affinity Chromatography

Affinity chromatography is a technique used as a capture step with high selectivity, high resolution, and usually high binding capacity based on the ligand specificity of target proteins. Affinity purification requires a biospecific ligand covalently

attached to a chromatography matrix. The ligand can specifically bind to the target protein, and the binding between them should be reversible to allow for the elution of the target protein in an active form. Target proteins are commonly engineered as fusion proteins to express the binding partner for affinity interaction with a specific ligand, such as poly-histidine for Ni^{2+} affinity chromatography and glutathione-S-transferase for GSH-affinity chromatography.

Ni^{2+} based immobilized metal ion affinity chromatography (IMAC) is a purification method used to separate protein with poly-histidine tag based on the tendency of histidine to form complexes with divalent metal cations around neutral pH (Porath et al, 1975). A polyhistidine-tag is an amino acid motif of at least five histidine (His) residues, often at the N- or C-termini of proteins and can be cleavable or non-cleavable.

In this study, the HiTrap Chelating HP and HisTrap HP columns (GE Healthcare) were used for Ni^{2+} affinity chromatography. The metal ions (Ni^{2+}) have been coupled to the chelating agarose matrix of prepacked columns. Charged columns were first equilibrated with buffer A containing 20mM imidazole before loading cell lysate. A low concentration of imidazole (typically 20 mM) was included in buffer A to inhibit non-specific binding of contaminant proteins to the column. Separations were performed under the control of Äkta liquid chromatography equipment (GE Healthcare). After binding of the target protein to the column, around 10 column volumes of buffer A was washed through the column to eliminate any non-specifically bound impurity. The target protein was then eluted with a linear concentration gradient of imidazole by the addition of Buffer B, which normally contained the same components as buffer A except increasing the imidazole concentration to 500 mM. The elution fractions were collected and analysed by SDS-PAGE. The verified fractions were then pooled together for the following purification steps. Cleavable His-tag can then be removed by protease treatment and a second Ni^{2+} affinity chromatography

2.4.3 Ion Exchange Chromatography

Purification using ion exchange chromatography takes advantage of the interaction between charged solute molecules and immobilized ion exchange

groups. The ionic groups on the stationary phase surface electrostatically attract oppositely charged solute molecules, and molecules with the same charge will flow through. Anion exchange chromatography is usually performed to purify a negatively charged protein - dissolved in a buffer with a pH greater than its isoelectric point. Proteins with net negative charge interact with resin and are retained on column under low ionic strength condition, while positively charged molecules flow through the column. A wash step is then performed to remove any weakly or non-specifically bound proteins whilst the tightly bound proteins will be eluted by high ionic strength buffer.

In this study, the low ionic loading buffer (referred to as buffer A) and the high ionic elution buffer (referred to as buffer B) were prepared to contain 50mM NaCl and 1M NaCl, respectively. The pH values of both buffers were at least one unit greater than the pI value of the target protein predicted on the basis of its amino acids composition. MonoQ columns (GE Healthcare) were used for anion exchange chromatography, and separations were performed under the control of Äkta liquid chromatography equipment (GE Healthcare). The MonoQ column was pre-equilibrated with the buffer A before loading sample and a linear gradient concentration of NaCl was maintained by the Äkta to elute the target protein, after washing out any weakly or non-specifically bound proteins with 10 column volumes of buffer A. The elution fractions were collected and analysed by SDS-PAGE, and verified fractions were then pooled together for further purification.

2.4.4 Ammonium Sulfate Precipitation

Ammonium sulfate precipitation is a method used to purify proteins by altering their solubility in the solution, known as salting out. The solubility of proteins varies with the ionic strength of the solution. Salting in refers to the enhanced solubility of the protein with increasing salt concentration (i.e. increasing ionic strength), when the salt concentration is low. However, the solubility of the protein begins to decrease until a sufficiently high ionic strength is reached, at which point the protein completely precipitates out of solution. Salting out can be a very effective technique for separating proteins, because the solubility of proteins differs greatly at high ionic strength.

Ammonium sulfate is a commonly used salt for the purpose of purifying proteins by salting out. In this study, a certain amount of solid ammonium sulfate was added into protein solutions according to Table 2-5 to achieve a desired concentration of ammonium sulfate. The ammonium sulfate concentration that precipitated the maximum proportion of desired protein was identified by SDS-PAGE. Precipitated protein was recovered by dissolving precipitates in fresh buffer.

Table 2-5 Ammonium sulfate precipitation table

Initial concentration of ammonium sulfate	Percentage saturation at 0°																
	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
	Solid ammonium sulfate (grams) to be added to 1 liter of solution																
0	106	134	164	194	226	258	291	326	361	398	436	476	516	559	603	650	697
5	79	108	137	166	197	229	262	296	331	368	405	444	484	526	570	615	662
10	53	81	109	139	169	200	233	266	301	337	374	412	452	493	536	581	627
15	26	54	82	111	141	172	204	237	271	306	343	381	420	460	503	547	592
20	0	27	55	83	113	143	175	207	241	276	312	349	387	427	469	512	557
25		0	27	56	84	115	146	179	211	245	280	317	355	395	436	478	522
30			0	28	56	86	117	148	181	214	249	285	323	362	402	445	488
35				0	28	57	87	118	151	184	218	254	291	329	369	410	453
40					0	29	58	89	120	153	187	222	258	296	335	376	418
45						0	29	59	90	123	156	190	226	263	302	342	383
50							0	30	60	92	125	159	194	230	268	308	348
55								0	30	61	93	127	161	197	235	273	313
60									0	31	62	95	129	164	201	239	279
65										0	31	63	97	132	168	205	244
70											0	32	65	99	134	171	209
75												0	32	66	101	137	174
80													0	33	67	103	139
85														0	34	68	105
90															0	34	70
95																0	35
100																	0

This table was Adapted from (Dawson et al, 1986).

2.4.5 Size Exclusion Chromatography

Size Exclusion Chromatography (SEC), also called gel filtration chromatography, was used in the work described in this thesis as a polishing step to achieve the required level of purity by separating molecules on the basis of their size and shape (hydrodynamic radius). When molecules with different sizes pass through a gel filtration medium packed in a column, smaller molecules experience a more complex pathway to exit than larger molecules, causing differences in retention time. Both molecular weight and three-dimensional shape contribute to the degree of retention. Large molecules (e.g., aggregates) are eluted in or just after the void volume (Vo), which is equivalent to approximately 30% of the total column volume for a well-packed column. Small molecules (e.g., salts) are usually eluted just before one total column volume (Vt), since they have full access to the medium. Proteins are detected by monitoring UV absorbance at A_{280} nm, and salts are

detected by monitoring the conductivity of the buffer. There are many types of media available, each with different fractionation ranges, such as Sephadex, Sephacryl, Superose. In conclusion, gel filtration is a useful technique in protein purification for final polishing, desalting and buffer exchange.

A gel filtration column was chosen according to the size of the target protein, and the column was equilibrated with one column volume of elution buffer before loading the sample. The sample was concentrated to the volume that was less than 2% of the column volume and precipitated aggregates were removed by centrifugation or syringe filtration. Separations were performed under the control of Äkta liquid chromatography equipment (GE Healthcare) with one column volume of elution buffer. The elution fractions containing protein of interest were collected and analysed by SDS-PAGE.

2.5 Protein Characterisation

2.5.1 Denaturing Polyacrylamide Gel Electrophoresis

Denaturing polyacrylamide gel electrophoresis (SDS-PAGE) is used for separating proteins merely according to their size (no other physical feature) ranging from 5 to 2,000 kDa. The denaturing detergent SDS (sodium dodecyl sulphate) is used to convert all proteins into the same linear shape with only primary structures, and impart all proteins with large net negative charges. Polyacrylamide is a polymer of acrylamide monomers, and the pore size is controlled by the concentration. The polymerisation of acrylamide is driven by free radicals formed by the addition of ammonium persulfate (APS) and promoted by the addition of N,N,N',N'-Tetramethylethylenediamine (TEMED). Typically, resolving gels can be made in 6%, 8%, 10%, 12% or 15%, and the percentage chosen depends on the size of the protein to be identified in the sample - the smaller the protein's size is, the higher percentage the resolving gel should be. Stacking gel (3%) is poured on top of the resolving gel to concentrate the protein sample into very narrow bands prior to separation on the resolving gel, and a gel comb is then inserted to form wells and define lanes. When an electrical field is applied across the gel, SDS coated proteins will migrate towards the anode.

In this study, stacking gel (3%) was prepared with Tris-HCl buffer pH 6.8; resolving gel (12% or 15%) was prepared with Tris-HCl buffer pH 8.8 according to the recipe in Table 2-6. The samples were prepared by mixing with loading dye and heating to 95 °C for 5 minutes to ensure complete denaturation of proteins. A low range molecular weight marker (Bio-Rad) with bands at 97.4, 66.2, 45, 31, 21.5 and 14.4 kDa was applied to run along with samples for the estimation of sample's molecular weight. After loading samples, gels were usually run at 200V until the tracking dye reached the bottom of the gel. Gels were stained with coomassie brilliant blue, and then destained in 5% propan-2-ol and 7% acetic acid before visualising protein bands. Approximately 1-10 μg of protein was loaded for a clear band.

Table 2-6 SDS-PAGE components

Stacking gel (3%)
0.13 M Tris-HCl (pH 6.8)
0.1% SDS
3% acrylamide
0.04% bis-acrylamide
For polymerisation:
0.025% APS
Approx. 10 µL TEMED per 10 mL gel mixture
Resolving gel (example, 15% polyacrylamide)
0.38 M Tris-HCl (pH 8.8)
0.1% SDS
15% acrylamide
0.2% bis-acrylamide
For polymerisation:
0.025% APS
Approx. 10 µL TEMED per 10 mL gel mixture
Running buffer
25 mM Tris
200 mM Glycine (pH 8.8)
Loading dye (4×)
60 mM Tris-HCl (pH 6.8)
10% glycerol
2% SDS
0.02% Bromophenol blue
5% β-mercaptoethanol (for reducing conditions)

2.5.2 Size Exclusion Chromatography Multi-Angle Laser Light Scattering (SEC-MALLS)

Size Exclusion Chromatography coupled with Multi-Angle Laser Light Scattering (SEC-MALLS) is a characterisation method to estimate the molecular mass and as well as determine the oligomeric state of proteins (Mogridge, 2004). Fractionation is performed by a HPLC system (Shimadzu) on a gel filtration column. Fractions are then delivered to three detection systems: 1) a UV/Visible light spectrometer to monitor the protein elution; 2) a Refractive Index (RI) monitor to enable measurement of the protein's concentration; and 3) a Light Scattering (LS) detection system to record light scattering data. The amount of light scattered by proteins in solution is directly proportional to the product of the weight-averaged molar mass and the macromolecular solute concentration, i.e., LS $\sim M_w \cdot c$ (Wyatt, 1993).

60 µL of protein sample at a concentration of 0.5 mg/mL was applied to a pre-equilibrated gel filtration column BioSep-SEC-s3000 column (Phenomenex). Size-exclusion chromatography was carried out on a Shimadzu HPLC system comprising LC-20AD pump, SIL-20A Autosampler and SPD20A UV/Vis detector. The elution of the target protein was monitored at 280 nm by the SPD20A UV/Vis detector. The light-scattering data were recorded by a Dawn HELEOS-II 18-angle light scattering detector and the concentration of the eluted protein was measured by an in-line Optilab rEX refractive index monitor (Wyatt Technology). Data were analysed with the ASTRA V software package. Molecular mass was calculated based on Zimm's formalism of the Rayleigh-Debye-Gans light scattering model for dilute polymer solutions using 0.183 mL/g as the refractive index increment (dn/dc) value.

2.5.3 Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry (MALDI-MS)

Mass spectrometry is an effective analytical technique tool routinely used for measuring the molecular mass of a protein sample (Gross & Strupat, 1998; Strupat, 2005). Matrix-assisted laser desorption/ionization (MALDI) was developed as a soft ionization technique to produce gas-phase ions of proteins by laser irradiation (Hillenkamp & Karas, 1990; Karas et al, 1987). It provides an accurate and

sensitive method for measuring the molecular weights of large, non-volatile and labile molecules like proteins by mass spectrometry (Li et al, 1994).

Usually time-of-flight mass spectrometry is used, in which the mass-to-charge ratio of the ion can be determined by the time measurement. As little as one pmol of sample is required to be detected. The experiments were performed by Simon Grist.

2.5.4 Negative Staining Electron Microscopy

Electron microscopy is a useful visualisation tool to examine the shape and stoichiometry of a large protein assembly. The negative staining technique uses heavy metal salts (e.g. uranyl acetate) to enhance the contrast between the background and the image of protein assembly.

Protein samples were diluted to a certain concentration (~ 0.05 mg/mL) for the single-particle imaging. A 2 μL aliquot of the sample was applied to glow-discharged, continuous-carbon and formvar covered copper grids with a 300-square mesh. The dried sample was washed with MiliQ water and then stained with 2% uranyl acetate for two minutes. The grids were air dried before being inspected on a 200 kV FEI F20 electron microscope at a magnification of 100, 000x. Imaging was performed in a low electron dose mode. Approximately 10 micrographs were recorded and 1200 protein particles were selected for image analysis. The experiments were performed with Yuriy Chaban and Elena Orlova at Birkbeck College, University of London.

2.6 Crystallisation

2.6.1 Theory of Crystallisation

High-quality crystals are required to obtain the three-dimensional structure of proteins by single crystal diffraction method. However, producing diffracting crystals can be very challenging and sometimes a bottleneck in structure determination (Durbin & Feher, 1996). Proteins normally need to be pure with homogeneity of 95% at high concentration (> 5 mg/mL) to form crystals. The crystallisation of molecules is a phase transition phenomenon driven by the

minimisation of free energy (Weber, 1991). When the concentration of protein is higher than its solubility limitation and it reaches the supersaturation state, the transformation from liquid phase to solid crystalline phase occurs. A phase diagram (Figure 2-3) is used to illustrate the process of crystallisation, during which the stable state as liquid, crystalline or precipitate is formed under a variety of crystallisation parameters such as the concentration of protein, precipitant(s), additive(s) and so on (Chayen, 2004).

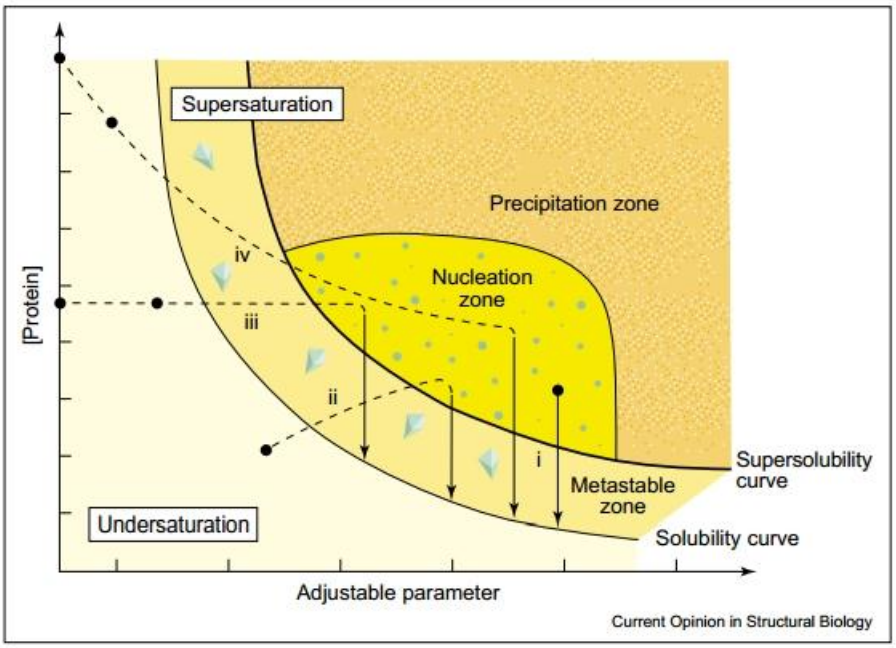


Figure 2-3 Schematic illustration of a typical protein crystallisation phase diagram. *This figure was adapted from (Chayen, 2004).*

Four zones representing different degrees of supersaturation are shown on the diagram: proteins will precipitate in the zone of "high supersaturation"; spontaneous nucleation will take place in the zone of "moderate supersaturation"; the growth of well-ordered crystals will occur under suitable conditions in the metastable zone of "lower supersaturation"; and in the zone of "undersaturation", proteins are fully dissolved and will never crystallise. Ideally, the concentration of protein solution will drop with the formation of nuclei, which will lead the system into the metastable zone to promote the growth of single crystals.

2.6.2 Initial Screening

A crystallisation experiment is usually started with a flexible sparse matrix initial screen for the identification of initial crystallisation conditions. The roles of pH, precipitant, additives, and temperature are examined in the initial screen whilst precipitants are used to decrease the solubility of the protein. Polymers such as PEG, salts, and organic solvents are the most popular precipitants in protein crystallisation experiments. Since pH can greatly affect a protein's solubility due to ionisation of charged residues at the surface, a suitable buffer must be present during crystallisation (e.g. MES, HEPES, Tris). Most proteins display a strong solubility dependence on temperature, and crystallisation experiments are usually carried out at 4 °C or 25 °C, because extreme temperature tends to cause denaturation of proteins.

To enhance the possibility of crystal formation, initial crystallisation screens are usually designed based on a rational combination of chemical conditions. There are large numbers of commercial screens available, such as INDEX, PACT, JCSG, and MPD (D'Arcy et al, 2003). After setting up initial screening trays by a pipetting robot, plates should be inspected under a microscope at regular time intervals (e.g., after 1 day, 2 days, 5 days, and 10 days), and the conditions that generate initial crystallisation hits will be further optimised for better quality crystals.

2.6.3 Crystallisation Optimisation

Once an initial condition is identified through initial screenings, optimisation of the condition is usually necessary to grow a well-diffracting crystal. The optimisation process takes the initial conditions with promising hits and then varies the concentration of the protein, precipitant, salt and additives in a systematic manner. The additives could be biologically relevant cofactors, substrates and so on. The optimisation trays can be set up either manually or using a Mosquito robot.

3 Engineering Fusion Constructs between the SPP1 Portal and TRAP Proteins

3.1 Introduction

3.1.1 The Bacteriophage SPP1 Portal Protein – Gp6

Bacteriophage SPP1, a lytic phage that infects *Bacillus subtilis*, was first described by Riva et al. (Riva et al, 1968). SPP1 belongs to the *Siphoviridae* family and comprises an isometric icosahedral capsid attached to a long (177 nm), flexible, non-contractile tail (Riva et al, 1968). The tail fibre at the tip of the tail is responsible for host cell attachment (Alonso et al, 2006).

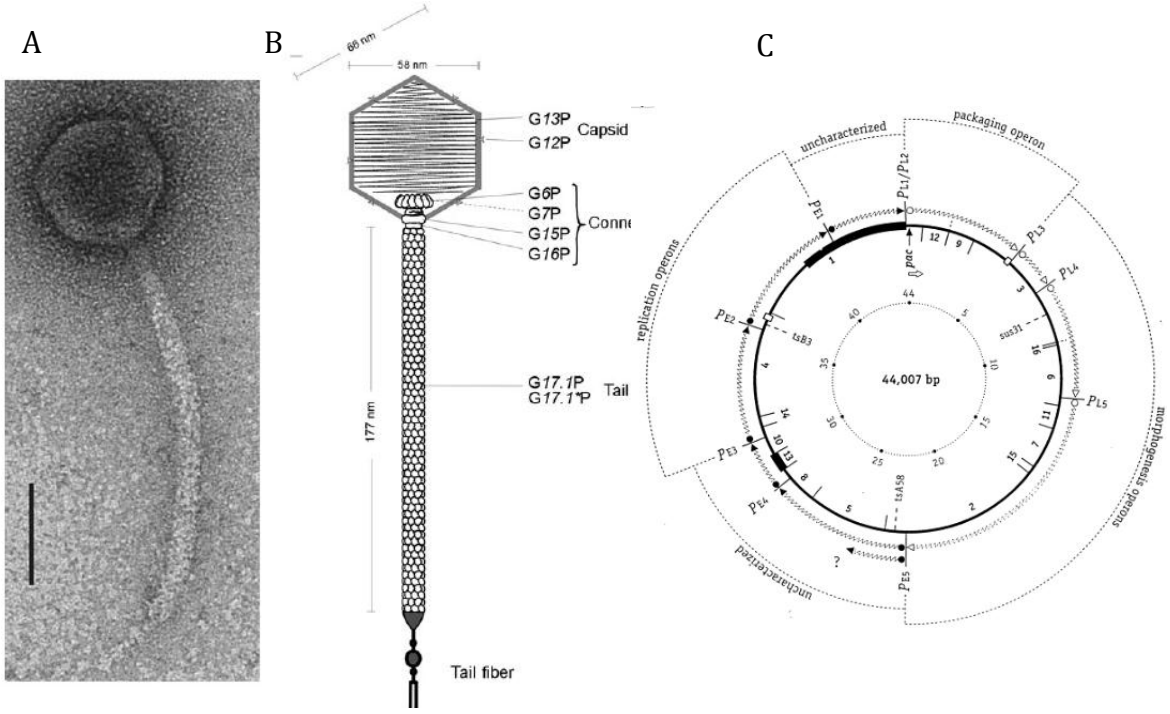


Figure 3-1 Structure and genetic map of bacteriophage SPP1.
(A): An SPP1 virion visualised by negative staining electron microscopy. The bar represents 50 nm. (B): A diagram illustrates the structural organisation of the mature SPP1 virion. (C): Physical and genetic map of bacteriophage SPP1. This figure was adapted from (Alonso et al, 2006).

In the mature SPP1 phage, the viral genome, a linear double-stranded DNA molecule of ~45.9 kb, is enclosed within the viral capsid (Droege & Tavares, 2005). The complete nucleotide sequence of the SPP1 genome was reported in 1997, with

a size of 44007 bp and a base composition of 43.7% dG + dC (Alonso et al, 1997). The complementary DNA strands can be separated into a purine-rich heavy strand and a light strand (Riva et al, 1968). The DNA circularises upon its infection of the Gram-positive bacterium *Bacillus subtilis*, and only the heavy strand is transcribed during the asymmetric transcription process (Alonso et al, 1997; Riva, 1969). Functionally related gene modules were identified, which comprise clusters of genes encoding proteins with complimentary functions. Approximately 47% of the genome encodes proteins involved in phage assembly and DNA packaging (Alonso et al, 1997). Bacteriophage SPP1 is used as a model system for viral capsid assembly, because it is one of the most highly characterised phages whose biochemistry and structure have been studied in considerable detail.

The assembly of the SPP1 capsid follows a pathway common among tailed bacteriophages and herpes viruses, shown in Figure 3-2 (White et al, 2012). Firstly, a spherically shaped protein shell "procapsid" is formed by many copies of the major capsid protein gp13, 100–180 copies of internal scaffolding proteins gp11, one 12-subunit portal protein gp6, and several copies of the minor protein gp7 (Becker et al, 1997; Droge et al, 2000). The scaffolding protein gp11 is often required for a functional procapsid to ensure the polymerisation of gp13 into a shell with the correct geometry (T=7). Stable complexes of portal protein gp6 and minor protein gp7 were detected both *in vivo* and *in vitro*. *In vitro* studies showed gp7 was not a structural necessity for the formation of biologically active procapsids, although the biological activity would be reduced 5 to 10-fold in the absence of gp7 (Droge et al, 2000).

A complex consisting of the viral terminases and the viral DNA docks on to the portal protein to assemble the DNA translocation molecular motor. The DNA packaging motor is composed of the portal protein gp6, the small terminase gp1 and the large terminase gp2, which pump the viral genome into the procapsid with ~4% terminal redundancy controlled by a headful packaging mechanism. The packaging of DNA is accompanied by the release of the internal scaffolding protein and the expansion of the procapsid size from ~55 nm to ~61 nm in diameter (White et al, 2012). Upon the termination of DNA encapsidation, head completion proteins (gp15 and gp16) sequentially attach to the portal vertex to form a head-to-tail connector following the detachment of gp1 and gp2 from the portal.

Lastly, an independently assembled long non-contractile tail attaches to the portal protein to yield an infective virion.

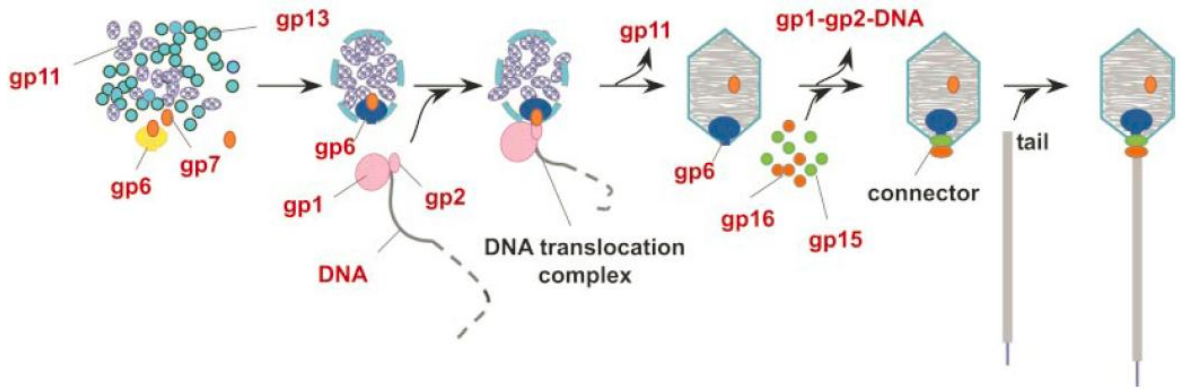


Figure 3-2 Morphogenesis process of bacteriophage SPP1. *This figure was adapted from (Orlova et al, 2003).*

The portal protein of bacteriophage SPP1 is encoded by the gene *gp6*. Each subunit of the portal protein consists of 503 amino acids with a molecular weight of 57 kDa. Although the amino acid sequence of gp6 shows no significant similarity to portal proteins from other bacteriophages, gp6 plays the same role in the morphogenesis process to initiate the formation of procapsid, whilst providing a site for the attachment of the DNA packaging motor to the procapsid and the tail to the filled capsid (Alonso et al, 2006; Droge et al, 2000; Tavares et al, 1992).

The SPP1 portal protein gp6 is a circular homo-oligomer situated at a single vertex of the icosahedral capsid. The three-dimensional structure of gp6 was first determined using electron microscopy at liquid-helium temperatures and angular reconstitution, which revealed a 13-fold symmetric oligomer with three distinct regions: stem, wing and a fringe of small 'tentacles' (Orlova et al, 1999). The structures of the connector - a complex of gp6 and the head completion proteins gp15 and gp16, and the isolated gp6 alone were then determined at 10 Å and 9 Å resolution respectively by cryo-electron microscopy and single particle analysis (Orlova et al, 2003). Statistical analysis revealed the intrinsic 12-fold symmetry within gp6 from the connector in comparison to the 13-mer of isolated gp6. In spite of the difference in oligomeric state, similar overall structural organisations were observed from gp6, with the exception of some conformational changes at the bottom of the stem region. The base of the stem moved outwards to interact with and accommodate gp15 (Figure 3-3).

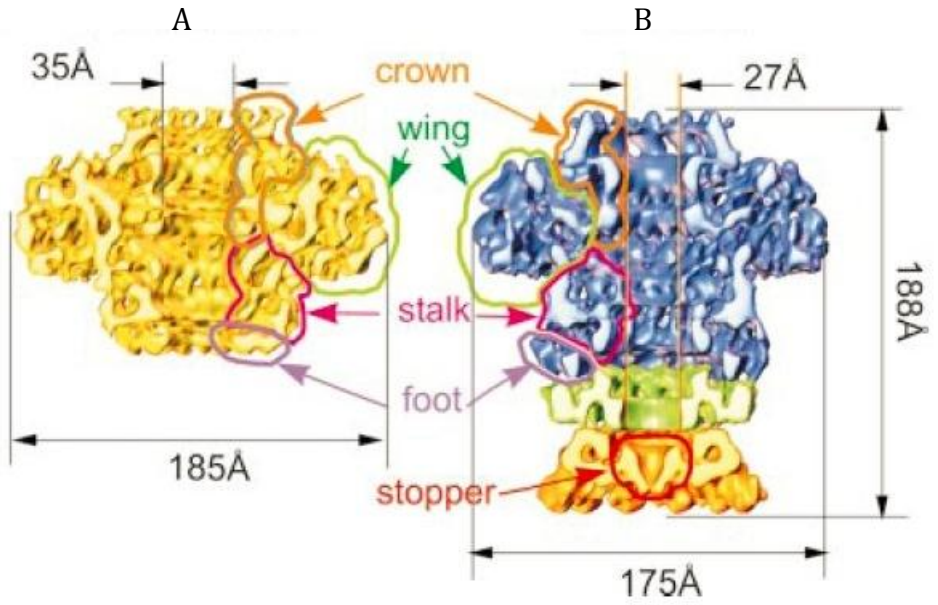


Figure 3-3 Cut-away views of the SPP1 connector and isolated portal protein by cryo-EM.
The isolated portal protein is shown in yellow (A), and the portal protein in the connector is shown in blue, gp15 in green and gp16 in orange (B). The domains of gp6 are outlined: the crown in orange, the wing in green, the stem stalk in magenta, the stem foot in purple and the gp16 stopper in red. This figure was adapted from (Orlova et al, 2003).

The X-ray structure of the 13-subunit assembly of isolated gp6 at 3.4 Å resolution provided more information to help elucidate the structural basis for DNA translocation (Lebedev et al, 2007). A single gp6 subunit can be divided into four regions, namely the clip, stem, wing and crown (Figure 3-4). The basal clip region has an α/β fold and is exposed to the viral head exterior. The stem is composed of two helices, α3 and α5, tilting relative to the central tunnel axis by ~50° and ~30°, respectively, and connects the clip to the wing. The wing region is largely made up of α-helices, with helix α6 being 40-residue long and spanning from the central tunnel to the peripheral rim. There is a 45° kink at the distal end of helix α6, which is stabilised by interaction with the C-terminus of helix α5. A 15-residue loop, referred to as the "tunnel loop", protrudes into the tunnel and connects helix α5 and α6. These tunnel loops constitute the most constrained part of the portal channel with a diameter of 27.7 Å in a 13-mer assembly. Meanwhile, the crown is composed of three α-helices and there are 40 additional residues at the C-terminus that are disordered.

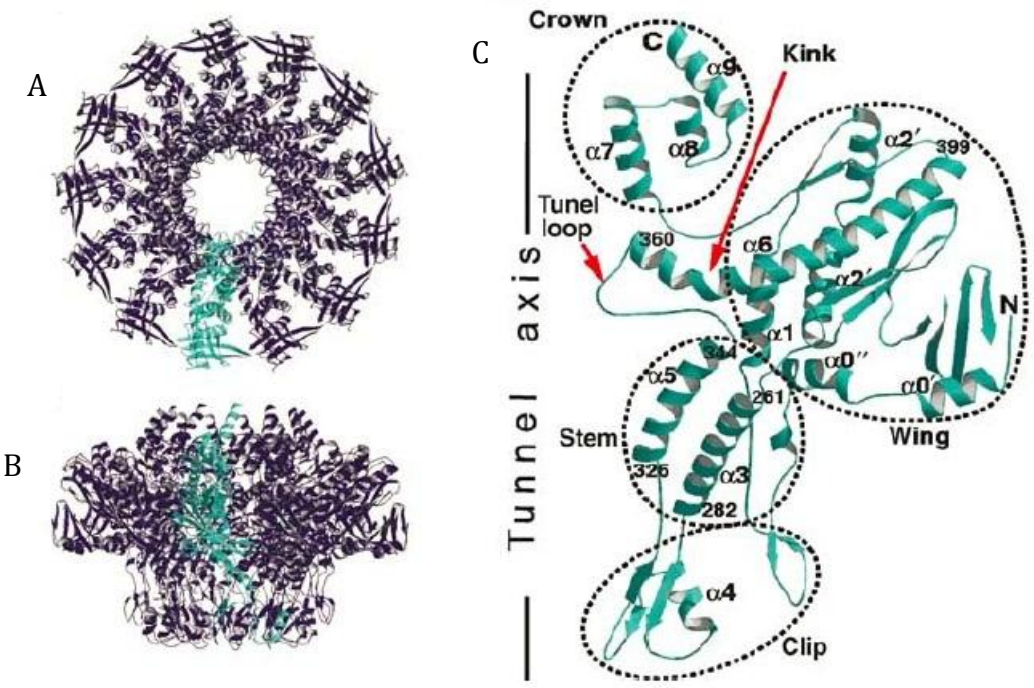


Figure 3-4 X-ray structure of the SPP1 portal protein 13-subunit assembly.
(A, B) Ribbon diagrams of the portal protein along and perpendicular to the 13-fold axis. (C) Ribbon diagram of a single portal protein subunit to illustrate the four major domains: crown, wing, stem and clip with the tunnel axis in vertical orientation. This figure was adapted from (Lebedev et al, 2007).

Because no crystallographic structure is available for the dodecameric SPP1 portal assembly, the pseudo-atomic structure of the 12-mer gp6 was generated by fitting single subunits taken from the X-ray structure of the 13-mer into the EM map of the connector (10 Å resolution). The accuracy of such a procedure was assessed by fitting subunits as rigid bodies into the EM map of the isolated 13-mer of gp6 (9 Å resolution), and the reconstructed structure proved to be in excellent agreement with the X-ray structure. The overall architecture of gp6 subunits in the two different oligomeric states was very similar. However, conformational changes in some specific segments occurred when switching from the 13-mer to the 12-mer: (i) the N-terminal end of α6 moved ~3 Å towards helix α5 (Figure 3-5); (ii) an outer loop of the clip region moved outwards; and (iii) the tunnel loops had more conformational variability. Accompanying the oligomeric state transition from 13-mer to 12-mer, the diameter of the tunnel loop belt was significantly decreased from 27.7 Å to 18.1 Å due to subunits reorientation - the decrease of

approximately 10 Å appeared to be much greater than a simple scaling down by a factor of 12/13 (Figure 3-5).

The SPP1 portal protein is consistently found to be a dodecameric assembly incorporated into procapsids (Lurz et al, 2001), but presented as 13-mer in the isolated form (Dube et al, 1993). How is gp6 capable of assembling into different homo-oligomeric rings? A possible explanation lies in the flexible inter-subunit interactions (Lebedev et al, 2007). The direct main-chain/main-chain inter-subunit hydrogen bonds are only within the clip, which is the most stable region. The rest of the structure has relatively few contacts between neighbouring subunits, with only 4.3 hydrogen bonds observed per 1000 $Å^2$ (12 per 2800 $Å^2$), half the expected number of direct inter-subunit hydrogen bonds (Lebedev et al, 2007). The determinant factor for the stoichiometric transition of SPP1 portal protein from 12-mer to 13-mer remains unknown, and there is no method to produce an isolated gp6 dodecamer for *in vitro* studies yet.

It has been found that the 13-subuint assembly of purified gp6 is dependent on the ionic environment and forms a monodispersed population of 13-mers in the hundred millimolar range of univalent salt (≥ 250mM NaCl), or in the presence of a millimolar range of bivalent cations (≥ 5mM $MgCl_2$). In the absence of bivalent cations and at univalent salt concentrations below 250 mM, there will be an association-dissociation equilibrium mainly between the monomer and the 13-mer with a minor population of intermediate oligomers (Jekow et al, 1999).

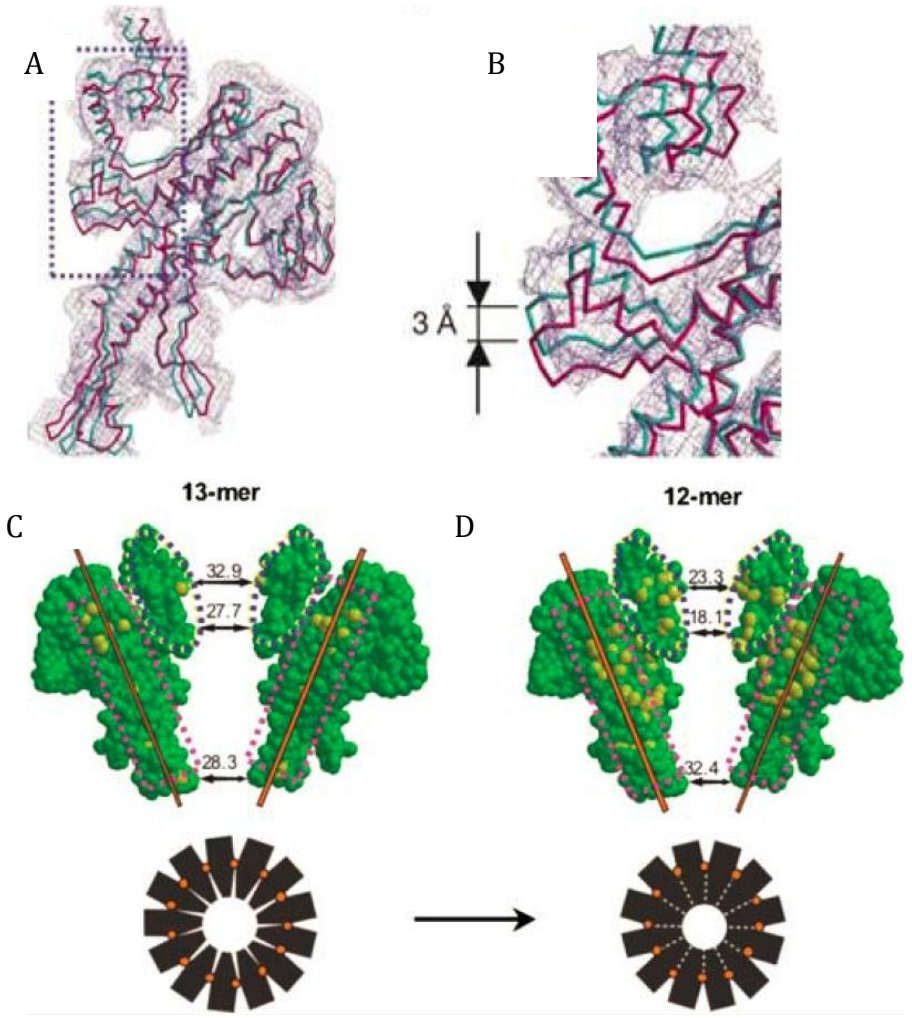


Figure 3-5 Structural changes upon 13/12-mer transition of gp6.
(A, B) The superposition of single subunits from the gp6 12-mer (coloured in magenta) and 13-mer (coloured in cyan) showed the N-terminal end of α6 moved ~3 Å towards helix α5; (C,D) Two diametrically arranged subunits from the 13-mer and 12-mer are shown for the clip, tunnel loops and crown areas with the van der Waals size of the tunnel diameter. The simple packing model schematized in black shows how the diameter of 12-mer tunnel is considerably reduced compared to the 13-mer tunnel by this rocking motion. This figure was adapted from (Lebedev et al, 2007).

3.1.2 Engineering of the Portal-TRAP Fusion Proteins

Protein engineering has proved to be pivotal in developing a fundamental molecular understanding of natural protein construction, and in the adaption of proteins to generate novel proteins. The coding sequence responsible for folding, structure, and function of the protein may be tractably manipulated during the protein engineering process. The approaches and technologies that have enabled

protein engineering include: simple site directed mutagenesis, advanced computational design, development of new and useful biocatalysts, integration of functional biological parts with fabricated devices, and construction of next generation biopharmaceuticals.

In the case of the SPP1 portal protein, previous attempts to obtain 12-subunit SPP1 portal protein *in vitro* by means of truncation and site directed mutagenesis have been unsuccessful. In this study, SPP1 gp6 is fused at its C-terminus with the TRAP protein (Figure 3-6) in order to promote the formation of portal protein oligomers containing 12 subunits. The reason that TRAP is chosen to construct the fusion proteins with gp6 is that TRAP forms an extremely stable dodecameric ring with much stronger subunit-subunit interactions than gp6. It is hoped that the oligomeric state of the fusion proteins could be dominated by TRAP as shown in the model (Figure 3-6), when the fusion proteins are over-expressed in *E. coli* cells.

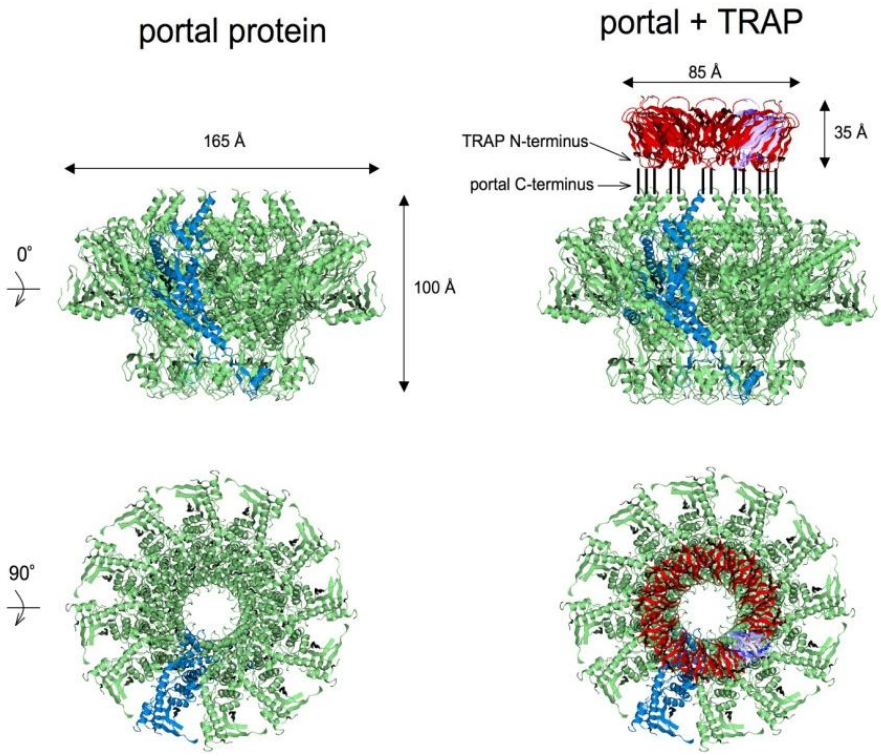


Figure 3-6 Model of the portal-TRAP fusion protein assembly. *SPP1 gp6 is fused at its C-terminus with the TRAP protein.*

3.1.3 The Tryptophan RNA-Binding Attenuation Protein

The tryptophan RNA-binding attenuation protein (TRAP), encoded by gene *mtrB*, is a ring-shaped oligomer involved in the regulation of tryptophan biosynthesis in Bacilli (Gollnick et al, 2005; Szigeti et al, 2004). TRAP regulates the *trp* operon by a transcription attenuation mechanism. When tryptophan is in excess, TRAP is activated and binds to the triplet repeats of the *trp* operon, favouring the formation of an intrinsic terminator hairpin to stop RNA synthesis (Figure 3-7B). When tryptophan is limiting, TRAP is not activated, allowing the formation of the alternative antiterminator hairpin and the transcription of the trp synthesis and transport genes (Figure 3-7A).

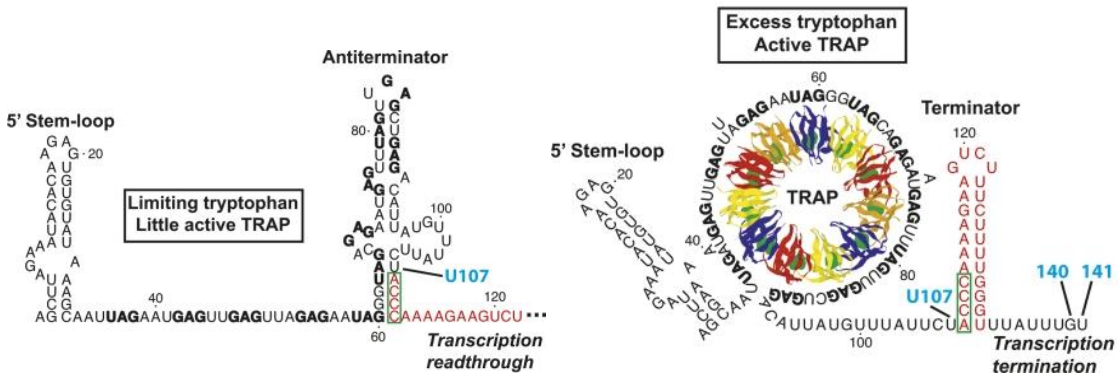


Figure 3-7 Model of the *B. subtilis* trpEDCFBA operon transcription attenuation mechanism.
(A) When tryptophan is limiting, TRAP is not activated, and antiterminator structure is formed to allow transcriptional readthrough. (B) When tryptophan is in excess, TRAP becomes activated by tryptophan and binds to the triplet repeats. The formation of an intrinsic terminator hairpin stops RNA synthesis of the downstream trp synthesis and transport genes. This figure was adapted from (McGraw et al, 2009).

TRAP from *Bacillus halodurans,* an alkaliphilic bacterium can grow at pH higher than 9.5, is 76-amino acid long and shares ~71% sequence identity with *Bacillus subtilis* TRAP. The *Bacillus stearothermophilus* mtrB gene encodes a 74-amino acid long polypeptide with 77% sequence identity compared to TRAP from *B. subtilis*. Three conserved residues - Lys37, Lys56 and Arg58 in *B. subtilis* TRAP - were identified to be involved in RNA binding (Yang et al, 1997).

X-ray structures of *B. subtilis*, *B. stearothermophilus* and *B. halodurans* TRAP revealed circular homo-oligomers with essentially identical architecture (Antson et al, 1995; Chen et al, 2011a; Chen et al, 1999). Like the bacteriophage portal protein, TRAP forms oligomers composed of multiple subunits. While *B. subtilis* TRAP and *B. stearothermophilus* TRAP exist as 11-mers, *B. halodurans* TRAP is a natural 12-subunit assembly. Removal of the five C-terminal residues in *B. stearothermophilus* TRAP or introducing a point mutation S72N in *B. subtilis* TRAP can induce the formation of stable 12-subunit oligomers (Bayfield et al, 2012; Chen et al, 2011a). The β-strands are the fundamental secondary structural elements in the TRAP monomer, which is composed of two β-sheets lying face-to-face. These β-sheets form extended inter-subunit β-sheets through main chain-main chain interactions to reinforce the oligomeric structure.

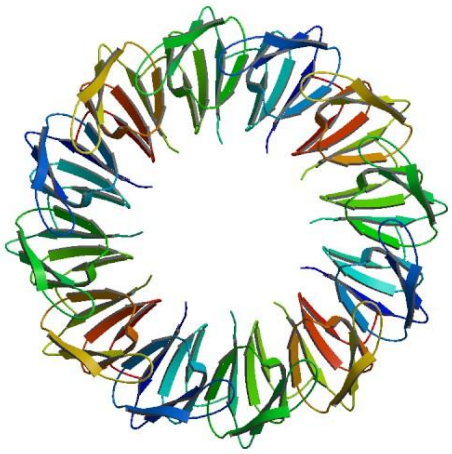


Figure 3-8 Structure of *B. halodurans* TRAP (3ZZL).
Ribbon diagram viewed along the 12-fold axis. This figure was adapted from (Chen et al, 2011a).

The stability of several wild type and mutant TRAP proteins in the presence of L-tryptophan were examined by dye-based scanning fluorimetry and CD spectroscopy (Figure 3-9). Among the wild type TRAPs, *B. stearothermophilus* TRAP is found to be the most stable, whilst *B. subtilis* TRAP is the least stable one. Significantly, the thermal stability of the mutant 12-subunit assemblies is increased compared to the wild-type 11-mers.

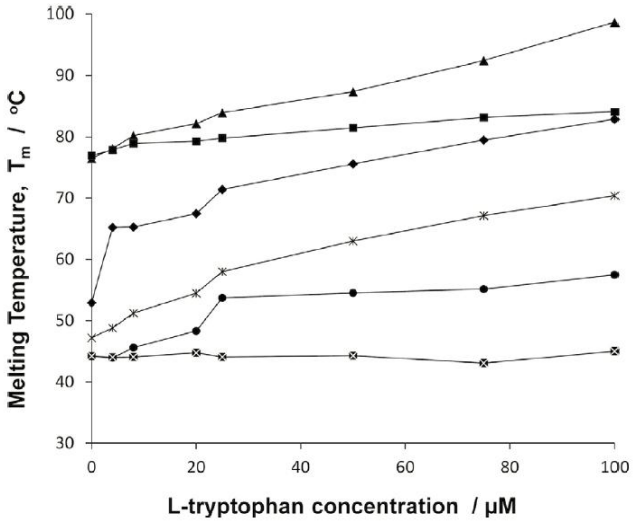


Figure 3-9 Melting temperatures of different TRAP oligomers assessed by dye-based scanning fluorimetry.

The dependence of melting temperature as a function of L-tryptophan concentration for six TRAP oligomers is show. From the top: B. stearothermophilus E71stop 12-mer (triangle), B. stearothermophilus wild type 11-mer (square), B. halodurans wild type 12-mer (diamond), B. subtilis K71stop 12-mer (open cross), B. subtilis S72N 12-mer (circle), B. subtilis wild type 11-mer (crossed square). This figure was adapted from (Bayfield et al, 2012).

3.2 Materials and Methods

3.2.1 Cloning of Fusion Protein Constructs with N-terminal His-tag

SPP1 genomic DNA was used as a template for PCR amplification of gp6 coding fragments; and plasmid DNA with *B. halodurans mtrB* gene in pET28a or plasmid DNA with *B. stearothermophilus mtrB* gene in pET9a was the template for TRAP encoding gene amplification. Primers were designed to introduce specific restriction recognition sites to the termini of PCR products: *NheI* and *BamHI* restriction sites to gp6 fragments for fusions with *B. halodurans* TRAP, *BamHI* and *XhoI* sites to *B. halodurans mtrB*; *BamHI* and *NotI* restriction sites to gp6 fragments for fusions with *B. stearothermophilus* TRAP, and *NotI and XhoI* restriction sites to *B. stearothermophilus mtrB.* A summary of the primers used in this study is listed in Table 3-1. The primers were synthesised by Eurofins MWG Operon, Ebersberg, Germany (purification level: salt free). For gp6 and *B. halodurans* TRAP fusion constructs, double digested PCR products of gp6 and *B. halodurans mtrB* as well as *NheI* and *XhoI* double digested pET28a (Novagen) were ligated together in one vial using T4 DNA ligase (NEB, UK) at room temperature for two hours. For the ligation of gp6 and *B. stearothermophilus* TRAP fusion constructs, double digested PCR products of gp6 and *B. stearothermophilus mtrB* as well as *BamHI* and *XhoI* double digested pCDFDuet-1 vector (Novagen) were ligated together in one vial as above. The correct inserts were verified by DNA sequencing (GATC Biotech) and alignment analysis (Clustal W).

3.2.2 Expression and Purification of Portal-TRAP Fusion Proteins

The plasmids coding for gp6-*B. halodurans* TRAP fusion constructs were transformed into *Escherichia coli* expression strain BL21(DE3) cells, and the plasmids coding for gp6-*B. stearothermophilus* TRAP fusion constructs were transformed into *Escherichia coli* expression strain B834 cells. A single colony from a Luria–Bertani (LB) agar plate containing the appropriate antibiotics (30 μg/mL kanamycin for pET28a vector and 50 μg/mL streptomycin for pCDFDuet-1 vector) was picked to inoculate a small overnight culture at 37 °C. 5 ml aliquots of the overnight culture were used to inoculate a 500 mL culture the following day. The 500 mL Culture were grown at 37 °C until the mid-log phase (OD600 of 0.6–0.8) was reached, at which point protein expression was induced by the addition

of isopropyl-d-1-thiolgalactopyranoside (IPTG) to a final concentration of 1 mM. Lastly, overnight incubation of the culture at 16 °C and 180 r.p.m was carried out for the over-expression of the fusion proteins. Cells were recovered by centrifugation in SORRALL® RC 5B plus for 20 min at 5000 g, 4 °C. Supernatant was removed and cell pellet was stored at -20 °C.

Table 3-1 Primers used for cloning of portal-TRAP fusion protein constructs

Primer	Sequence(5'-3')
Gp6-F 1	CCCCTAGCTAGCGCAGAACCGGACACAACCATG
Gp6-R1	CGCGGATCCTTGGTTCATTTCCTCTTCTATGCG
Gp6-R2	CGCGGATCCGCCCTGCATTTCAGCGTATTGG
Gp6-R3	CGCGGATCCGCCCTCATCGTCGAGTAGGTTGC
Gp6-R5	CGCGGATCCTGTAACGCCTTGTACCTGACTC
TRAP-F1	CGCGGATCCTCATCAAACTTTTTTGTCATAAAAGCAAAGG
TRAP-F2	CGCGGATCCTCAGGTTCTTCGTCAAACTTTTTTGTCATAAAAGCAAAGG
TRAP-F3	CGCGGATCCTCAGGTTCTTCAGGATCATCGTCAAACTTTTTTGTCATAAAAGCAAAGG
TRAP-F4	CGCGGATCCTCAGGTTCTTCAGGATCATCGATGAACGTGGGGGATAACTCAAAC
TRAP-F5	CGCGGATCCTCAAACTTTTTTGTCATAAAAGCAAAGG
TRAP-R1	GGGCCGCTCGAGTTATTCATCTTTCTCCGTATCTAGCG
GP6_I425Q_F	GACAGCGAGCAGGTTCAGAGT
GP6_I425Q_R	ACTCTGAACCTGCTCGCTGTC
GP6N27EcoR	GGAATTCGGCGGAACCGGATACCACCATGATTCAGAAACTG
GP6R479Not1	ATAGTTTAGCGGCCGCAGCAGCGCCTTCATCATCCAGCAGGTTGCCCT
TRAP_stea_F1	ATAAGAATGCGGCCGCAGCTAGCGACTTTGTTGTCATT
TRAP_stea_F2	ATAAGAATGCGGCCGCAGCTGCTAGCGACTTTGTTGTCATT
TRAP_stea_R1	CCGCTCGAGTTACTTTTTCCCTTCCGA

The cell pellet was re-suspended in nickel affinity chromatography binding buffer supplemented with a combination of 100 µg/mL lysozyme and protease inhibitors (1 mM AEBSF and 0.7 µg/mL pepstatin). For all the gp6-*B. halodurans* TRAP fusion proteins, the binding buffer was 50 mM imidazole pH 8.0, 250 mM NaCl, and 10 mM $MgCl_2$. Cells were disrupted by sonication with a large probe in a glass beaker using short pulses of 30 seconds, with a two-minute resting time in between

pulses. The sonication was carried out on ice, with caution to minimise thermal damage to the protein extract. The cell debris was removed by centrifugation at 38758 g for 30 minutes at 4 °C using a Sorvall SS34 rotor. The supernatant was collected and then cleared with a 0.45 mm filter (Millipore). The filtrate was digested overnight by a combination of RNase A and RNase T1 to remove RNA contamination at 4 °C. It was then loaded onto a 5 mL HiTrap column (GE Healthcare) and equilibrated with binding buffer for nickel affinity chromatography purification.

The nickel-bound protein was eluted with an increasing proportion of elution buffer containing 500 mM imidazole. For all the gp6-*B. halodurans* TRAP fusion proteins, the elution buffer was 500 mM imidazole pH 8.0, 250 mM NaCl, 10 mM $MgCl_2$ and 10% glycerol. Fractions containing the portal-TRAP fusion proteins were pooled together, and concentrated using Vivascience 30kDa molecular weight cut-off concentrators. The concentrated sample (less the 10 mg/mL) was applied to a Superose 6 size-exclusion chromatography column as a final polishing step to remove any high-molecular-weight aggregates and obtain fusion proteins with the appropriate molecular weight. The fusion proteins were eluted with GF buffer (50 mM imidazole pH 8.0, 250 mM NaCl, and 10 mM $MgCl_2$), concentrated to at least 10 mg/mL and stored at -80 °C for crystallisation.

3.2.3 Size-Exclusion Chromatography coupled with Multi-Angle Laser Light Scattering (SEC–MALLS)

The molecular mass was determined by size-exclusion chromatography coupled with multi-angle laser light scattering (SEC–MALLS). The protein sample (60 µL, 0.5 mg/mL) was applied on a BioSep SEC-s3000 gel filtration column (Phenomenex) equilibrated with buffer containing 50 mM imidazole pH 8.0, 250 mM NaCl, and 10 mM $MgCl_2$. Size-exclusion chromatography was carried out on a Shimadzu HPLC system and the elution was monitored at 280 nm by an SPD20A UV/Vis detector. Light-scattering data were recorded by a Dawn HELEOS-II 18-angle light-scattering detector and the concentration of the eluting protein was measured by an in-line Optilab rEX refractive-index monitor (Wyatt Technology). Data were analysed with the ASTRA V software package (Wyatt Technology). Molecular mass was calculated using Zimm's formalism of the Rayleigh–Debye–

Gans light-scattering model for dilute polymer solutions and a refractive-index increment (dn/dc) of 0.183 mL/g was used for the protein molecular mass estimation.

3.2.4 Crystallisation

The sitting-drop vapour diffusion method was used for the initial crystallisation screening. Drops containing 150 nL protein solution and 150 nL reservoir solution were dispensed by a Mosquito Nanolitre Pipetting robot (TTP Lab-tech) in 96-well plates, and equilibrated against 60 µL of reservoir solution. To obtain initial crystal hits, crystal trays with several commercial screens such as Index, PACT, Clear Strategy Screens I and II, Morpheus and MPD were set up. Conditions in which small crystals grew were optimised in 24-well hanging-drop plates with manual pipetting. Crystals were tested using a Rigaku RU-H3R X-ray generator with rotating anode, equipped with Osmic multilayer optics and a MAR345 (MarResearch) imaging-plate detector.

3.2.5 Negative Staining Electron Microscopy

The purified fusion proteins and the wild-type SPP1 portal protein were diluted to ~ 0.05 mg/mL for the single-particle imaging. A 2 µL aliquot of the sample was applied to glow-discharged, continuous-carbon and formvar covered copper grids with a 300-square mesh. The dried sample was washed with MiliQ water and then stained with 2% uranyl acetate for two minutes. The grids were air-dried before being inspected on a 200 kV FEI F20 electron microscope at a magnification of 100, 000x. Imaging was performed in a low electron dose mode. 10 micrographs were recorded and approximately 1200 protein particles were selected for image analysis. The experiments were performed with Yuriy Chaban and Elena Orlova at Birkbeck College, University of London.

3.2.6 Dissociation-Reassociation Experiments of the Fusion Protein GP6(27-466)-GSSGSS-TRAP(7-76) (YM92)

To probe if re-association could generate 12-subunit oligomers, a dissociation-re-association experiment was designed. This was based on the observation that the assembly of portal protein oligomer is dependent on the ionic

environment, forming a 13-subunit assembly at and above 250 mM NaCl or at and above 5 mM $MgCl_2$. Subunit contacts in TRAP were deemed to be strong enough to preserve the 12-subunit oligomeric state of TRAP, which can then lead to the assembly of a 12-mer gp6-TRAP fusion construct after gp6 segment was dissociated at low salt concentration followed by re-association with the addition of $MgCl_2$. The fusion protein GP6(27-466)-GSSGSS-TRAP(7-76) was partially dissociated by overnight dialysis from the original purification buffer to low salt buffer containing 20 mM HEPES pH7.7, 150 mM NaCl, 1 mM L-Tryptophan at 4 °C. The re-association was achieved by adding a 2 M $MgCl_2$ solution to the final concentration of 10 mM. The re-associated assembly was inspected by negative staining electron microscopy to determine the subunit number.

3.3 Results

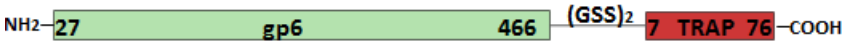
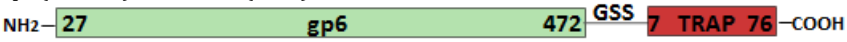
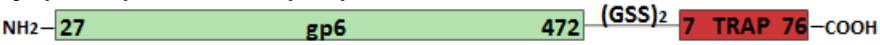
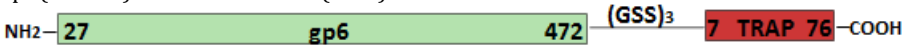
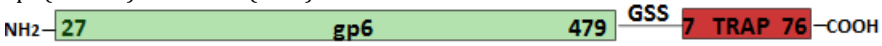
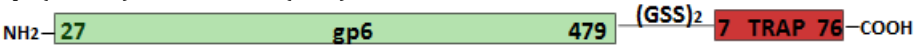
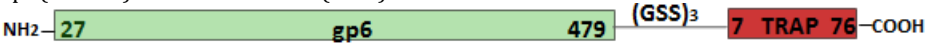
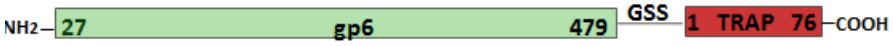
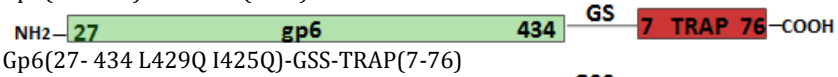
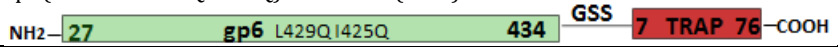
3.3.1 Design of the Portal-TRAP Fusion Proteins

SPP1 portal protein and TRAP were fused together using a short peptide linker between the two proteins. The linkers were composed of flexible residues such as glycine and serine so that the adjacent proteins are free to move relative to one another. The linkers were designed to have an appropriate length, to impose enough assembling power from TRAP on the portal protein. We reasoned that if the linker was too long, the assemblies would not direct each other. However, if the linker was too short, the correct folding of each protein could be affected.

I. Protein constructs (YM91-YM100) were designed to fuse the truncated SPP1 portal protein Δ27-466, Δ27-472, Δ27-479 with a natural 12-subunit TRAP from *B. halodurans* using linkers GSS, $(GSS)_2$ or $(GSS)_3$.

II. In order to reduce the difference between the tunnel diameters of the portal and TRAP, two fusion constructs were designed with a longer truncation to the C-terminus of portal protein. The tunnel diameter of gp6 Δ27-434 is the same as in TRAP, at approximately 30 Å. Shorter linkers (GSS and GS) were chosen to strengthen the contact and reduce flexibility between the gp6 and TRAP domains of the fusion protein.

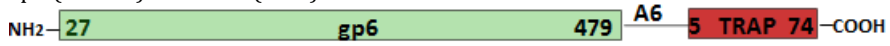
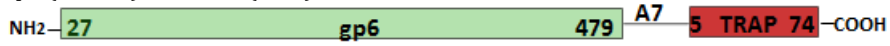
Table 3-2 Constructs of gp6-*B. halodurans* TRAP fusion proteins

Plasmid	Fusion protein constructs
pYM91	GP6(27-466)-GSS-TRAP(7-76)
	
pYM92	GP6(27-466)-GSSGSS-TRAP(7-76)
	
pYM93	GP6(27-466)-GSSGSSGSS-TRAP(7-76)
	
pYM94	Gp6(27-472)-GSS-TRAP(7-76)
	
pYM95	Gp6(27-472)-GSSGSS-TRAP(7-76)
	
pYM96	Gp6(27-472)-GSSGSSGSS-TRAP(7-76)
	
pYM97	Gp6(27-479)-GSS-TRAP(7-76)
	
pYM98	Gp6(27-479)-GSSGSS-TRAP(7-76)
	
pYM99	Gp6(27-479)-GSSGSSGSS-TRAP(7-76)
	
pYM100	Gp6(27-479)-GSS-TRAP(1-76)
	
pYM147	Gp6(27- 434)-GS-TRAP(7-76)
	
pYM148*	Gp6(27- 434 L429Q I425Q)-GSS-TRAP(7-76)
	

**Double mutations (L429Q and I425Q) are introduced to improve solubility and stability*

III. Based on the thermal stability analysis, the oligomeric state of *B. stearothermophilus* TRAP is the most stable among all TRAP species. Two constructs were designed to fuse the SPP1 portal protein (residues 27-479) with a natural 11-subunit TRAP from *B. stearothermophilus* (residues 5-74) using polyalanine linkers (6 x Ala and 7 x Ala).

Table 3-3 Constructs of gp6-*B. Stearothermophilus* TRAP fusion proteins

Plasmid	Fusion protein constructs
pYM198	Gp6 (27-479)-A6-TRAP (5-74)
	
pYM199	Gp6 (27-479)-A7-TRAP (5-74)
	

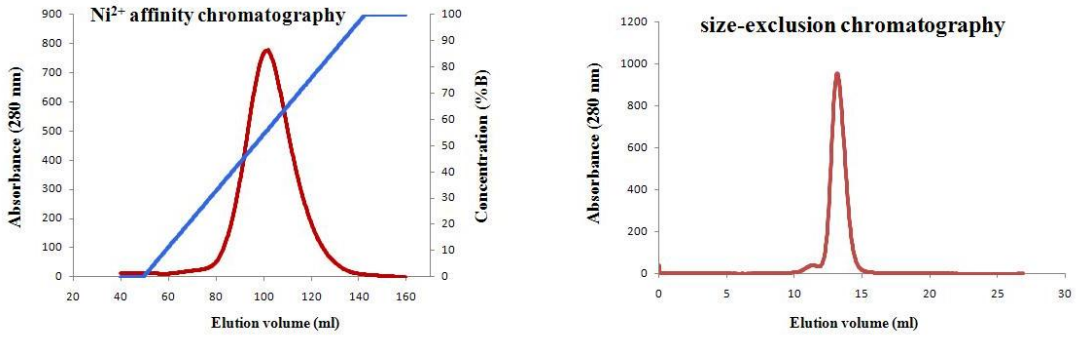
3.3.2 Cloning, Expression and Purification of the Portal-TRAP Fusion Proteins

3.3.2.1 Fusion proteins of truncated SPP1 gp6 with natural 12-subunit *Bacillus halodurans* TRAP

The DNA segments encoding the truncated portal gp6-*B. halodurans* TRAP fusion proteins were cloned into the restriction sites *BamHI* and *XhoI* of vector pET28a for the expression of N-terminal hexahistidine-tagged protein.

The fusion proteins were successfully overexpressed in *E. coli* B834 as a soluble form. Pure and homogenous proteins were obtained after Ni^{2+} affinity chromatography and size-exclusion chromatography in 50 mM imidazole pH 8.0, 250 mM NaCl, and 10 mM $MgCl_2$. Overnight digestion was performed to remove/eliminate RNA contamination during the purification process using a combination of RNase A and RNase T1. The eluted proteins from the major peak of the size-exclusion chromatography corresponded to the oligomer of portal-TRAP fusion proteins (Figure 3-10).

A



B

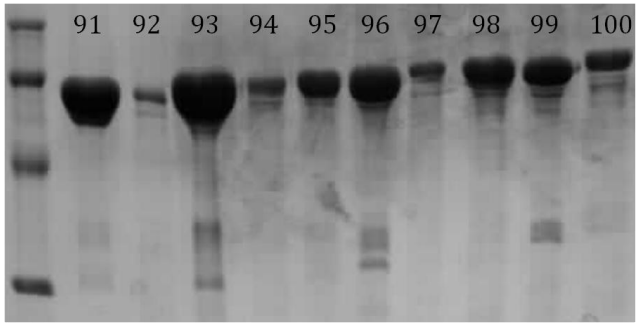


Figure 3-10 Purification of the gp6-*B. halodurans* TRAP fusion protein.

(A) Typical elution profiles of the two-step purification process containing Ni^{2+} affinity chromatography and size-exclusion chromatography. The elution profiles of

the protein YM92 duirng the purification process is shown here as an exmaple. (B) Characterisation of purified fusion proteins YM91-YM100 on a 12% SDS–PAGE gel.

3.3.2.2 Fusion proteins of SPP1 gp6 Δ27-479 with natural *B. stearothermophilus* 11-subunit TRAP Δ5-74

The DNA segments encoding the protein product gp6 Δ27-479 and *B. stearothermophilus* TRAP Δ5-74 were ligated and cloned into the restriction sites *BamHI* and *XhoI* of vector pCDFDuet-1 for the expression of fusion proteins with an N-terminal hexahistidine tag.

The fusion proteins were successfully over-expressed in *Escherichia coli* strain B834 at 16 °C. However, the majority of the protein was insoluble (Figure 3-11A). To improve the solubility, a buffer screen was carried out to investigate the effect of pH, salt type, salt concentration or additive on the solubility of the target protein. According to SDS-PAGE results (Figure 3-11B), the overexpressed protein was found to show preference for 20 mM Tris, pH 6.5, 500 mM NaCl and 5% glycerol with regard to solubility.

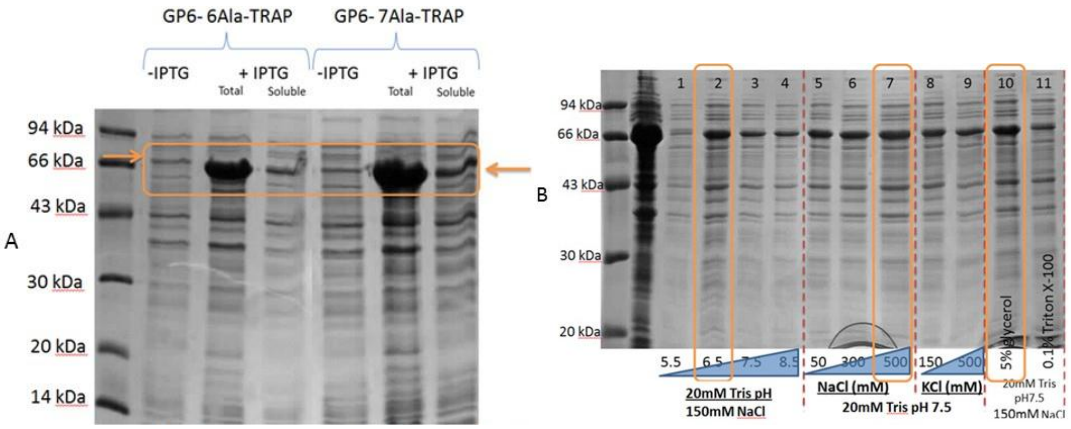


Figure 3-11 Expression and solubility test of the gp6-*B. stearothermophilus* TRAP fusion protein.
(A) "-IPTG" indicates expression cell lysate obtained before induction of the fusion proteins; "+IPTG" indicates the sample after induction of protein expression. Intense bands with the molecular weight of 66 kDa, highlighted, from the 'Total' lanes of both constructs were detected. However, only a thin band was present in the soluble sample. (B) Solubility test of crude cell extract in different buffers. Lane T contained total protein, while the rest contained only the soluble proteins. Lanes 1-4: solubility comparison with buffers of different pH, the most intense band at 66 kDa was shown

with 20 mM Tris pH 6.5. Lanes 5-9: the effects of salt type and concentration on the protein solubility were compared, and protein showed preference towards NaCl, but was not sensitive to the concentration of NaCl. Lanes 10 and 11 tested two additives, 5% glycerol and 0.1% triton X-100, an intense band was observed in lane 10, with 5% glycerol.

Nickel affinity chromatography was then performed using the buffer identified from the solubility screen as promoting the highest solubility (20 mM Tris, pH 6.5, 500 mM NaCl and 5% glycerol) with the addition of 10 mM $MgCl_2$ and 0.1 mM L-tryptophan. The addition of 10 mM $MgCl_2$ could facilitate the oligomerisation of gp6 (Jekow et al, 1999), and 0.1 mM L-tryptophan may bind and stabilise TRAP (McElroy et al, 2002). According to Figure 3-12, the fusion protein was eluted when the concentration of imidazole in the buffer was approximately 250 mM and 300 mM (40%~50% buffer B). However, the solubility and yield of the protein was insufficient for further characterisation and crystallisation.

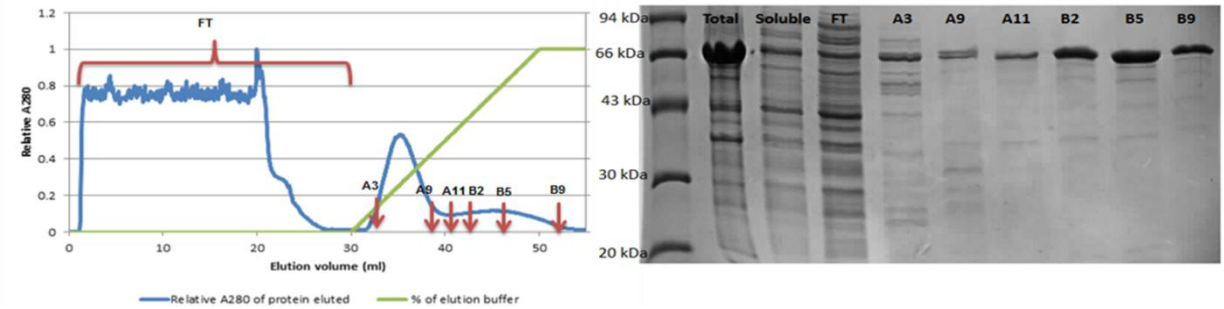


Figure 3-12 Nickel affinity chromatography of gp6-*B. stearothermophilus* TRAP fusion protein.
20 mM Tris, 500 mM NaCl, 5% glycerol, 10 mM MgCl2, 0.1 mM L-tryptophan, 10 mM imidazole, pH 6.5 was used as binding buffer. For the elution of bound his-tagged protein YM198, Gp6 (27-479)-A6-TRAP (5-74) , elution buffer with an imidazole concentration of 500 mM was used as Buffer B. Fractions indicated by arrows were resolved in SDS PAGE. The comparison of total and soluble samples of crude cell extract suspended in binding buffer showed the low solubility of the protein from a large-scale expression. FT: flow-through from the nickel column.

To further improve the stability, a buffer screen at pH 8.5 was carried out, because the theoretical isoelectric point of the fusion proteins is approximately 4.9. The

effect of $MgCl_2$ on the solubility was also investigated. According to Figure 3-13, the fusion protein was shown to tolerate low NaCl concentrations, and was most soluble at 100 mM NaCl. By comparing lanes 2 and 5 of both gel images, the addition of 10 mM $MgCl_2$ was shown to reduce the protein solubility. By comparing lane 3 of both gel images, the fusion protein was also shown to be more soluble in buffer with a pH of 8.5 than of 6.5. As a result, a buffer at pH 8.5 with no $MgCl_2$ was chosen for the following experiment.

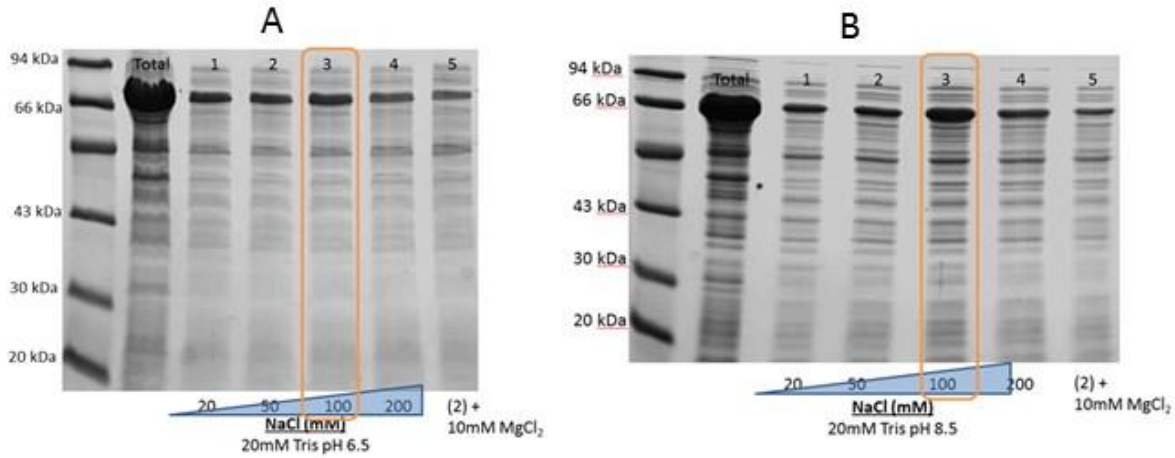
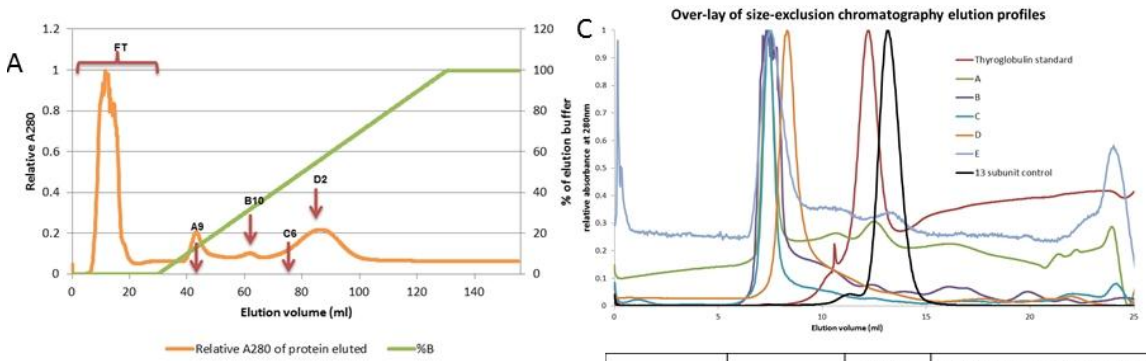


Figure 3-13 Solubility tests of gp6-*B. stearothermophilus* TRAP fusion protein. *The target protein is protein construct YM198: Gp6 (27-479)-A6-TRAP (5-74). Comparing the solubility dependency on NaCl concentration and effects of 10mM $MgCl_2$ in pH 6.5 (left) and 8.5 (right). Lane 5 of both gels has the same buffer composition as lane 2 but with 10mM $MgCl_2$ added.*

To purify the fusion protein, nickel affinity chromatography followed by size-exclusion chromatography was performed. The N-terminal histidine tagged protein (6xHis-tag) could bind to a nickel column and eluted at 300 mM imidazole (Figure 3-14A, B). Purification by size-exclusion chromatography generated aggregates with either 100 mM NaCl or 500 mM NaCl in the purification buffer, impeding purification. Decreasing the sample's concentration to ~1 mg/mL didn't help to resolve this issue (Figure 3-14C). Anion exchange chromatography was performed using a 1mL HiTrap Q HP column (GE Healthcare). The initial binding buffer was 20 mM Tris pH 8.5, 0.1 mM L-tryptophan, and varied with different NaCl concentrations of 0, 20 and 50 mM (data not shown). This likewise proved unsuccessful in improving solubility or yield.



	Construct	Sample conc. (mg/ml)	Elution buffer/ 0.1mM L- tryptophan, pH 8.5	
			Buffer	NaCl (mM)
A[1]	GP6 – 6Ala – TRAP	0.37	20mM Tris	100
B[1]	GP6 – 7Ala – TRAP	10.6		
C		1.6		
D		1.1	50mM imidazole	
E		0.31	20mM Tris	500

Figure 3-14 Purification of gp6- *B. stearothermophilus* TRAP fusion proteins.
(A-B) Nickel affinity chromatography purification of YM199: binding buffer was 20 mM Tris pH 8.5, 500 mM NaCl, 0.1 mM L-tryptophan, 20 mM imidazole, and imidazole concentration was raised to 500mM in elution buffer. Fractions indicated by arrows were resolved in SDS PAGE (B), FT: flow-through from the nickel column. (C) Gel filtration with different buffer conditions and sample concentrations. Protein concentration listed were final sample concentration loaded onto the gel filtration column, determined using the Bradford assay. Buffer conditions were varied with regard to NaCl concentration and buffer type. Chromatogram of 13-subunit control obtained from a previously characterised gp6-B. halodurans TRAP fusion protein.

3.3.3 Characterisation of Portal-TRAP Fusion Proteins

3.3.3.1 Molecular weight determination by MALDI-MS

For the fusion proteins of SPP1 gp6 truncations fused with natural *B. halodurans* 12-subunit TRAP, the purified protein was firstly analysed by MALDI-MS for rapid identification. The molecular weight of the fusion subunit detected by MALDI-MS was shown to be consistent with the theoretical molecular weight (an example shown in Figure 3-15). Also, the MALDI data excluded the possibility of contamination by Chaperone GroEL during the process of purification, because there was no peak corresponding to 57356.9 Da - the molecular weight of the GroEL subunit.

3.3.3.2 Protein Identification by Trypsin Digestion and Mass Spectrometry

The gp6-*B. stearothermophilus* TRAP fusion proteins have a theoretical molecular weight of approximately 62 kDa. However, the estimated molecular weight of the over-expressed fusion protein based on the migration of protein band on a 12% SDS PAGE was significantly larger than expected. SDS PAGE bands of fractions from nickel affinity chromatography purification containing reasonably pure target protein were submitted for identification by trypsin digestion and mass spectrometry. Sequence alignment confirmed the purified proteins as gp6-TRAP fusions.

3.3.3.3 Oligomeric State Analysis by size-exclusion chromatography coupled with multi-angle laser light scattering (SEC-MALLS)

The theoretical molecular mass of each gp6-*B. halodurans* TRAP fusion protein was calculated by ProtParam based on the amino acids composition. During SEC-MALLS analysis, a single monodispersed peak was observed for samples YM91, YM92, YM93, YM96, and YM97, with a mean molecular weight of the eluted species close to that expected for 13 subunits per oligomer (Figure 3-16, Table 3-4). In contrast, fusion protein YM94, YM95, YM98, YM99, and YM100 showed polydispersity, giving broad peaks in gel filtration elution profiles, probably due to the sample heterogeneity and potential aggregation. The molecular weight of the peak containing target proteins indicated 13 subunits per oligomer for YM95 and YM98 constructs and 12 subunits per oligomer for YM94, YM99, and YM100 fusions (Figure 3-16, Table 3-4).

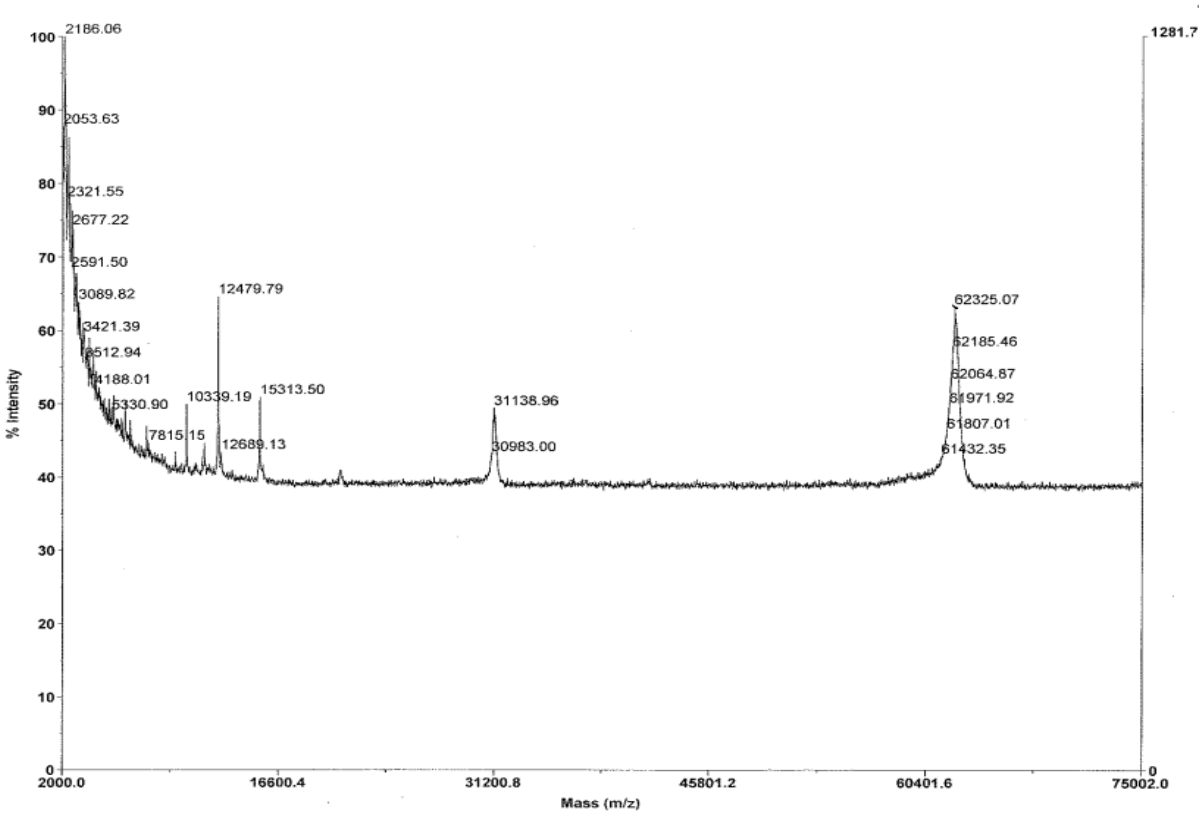


Figure 3-15 The molecular weight of the purified YM92 subunit characterised by MALDI-MS.

Only one major peak was detected corresponding to the molecular weight of 62325.07 Da, very close to theoretical molecular weight of 62399.3 Da.

Table 3-4 Estimated molecular weight of the fusion proteins by SEC-MALLS

Protein	Length (aa)	Mw (subunit, Da)	SEC-MALLS data (kDa)	Subunits number per oligomer
YM91	536	61239.2	837.5 ± 6	13.7
M92	539	61470.4	808.3 ± 8	13.1
YM93	542	61701.6	805.2 ± 6	13.1
YM94*	542	61919	766.0 ± 5	12.4
YM95	545	62150.2	779.7 ± 5	12.5
YM96	548	62381.4	795.4 ± 6	12.7
YM97	549	62675.7	809.1 ± 6	12.9
YM98	552	62906.9	815.5 ± 90	12.9
YM99*	555	63138.2	757.3 ± 4	12.0
YM100*	555	63306.4	768.9 ± 4	12.2

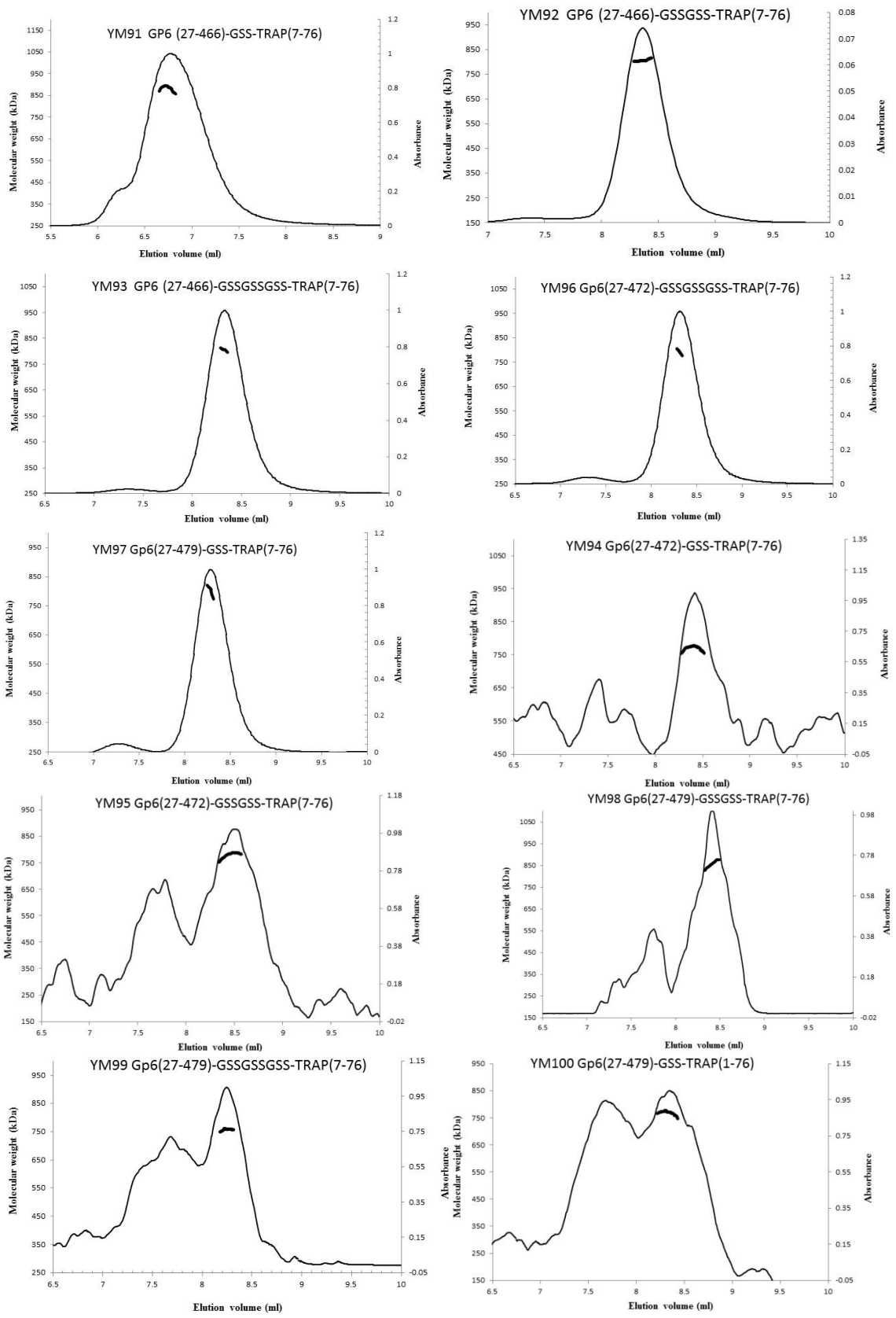


Figure 3-16 Characterisation of the oligomeric state by SEC–MALLS.
The thin line corresponds to the absorbance monitored at 280 nm. The thick line shows the molecular weight calculated for the eluted species. Fusion protein YM91, YM92, YM93, YM96, and YM97 showed single monodispersed peaks from gel

filtration elution profiles, while the fusion proteins YM94, YM95, YM98, YM99, and YM100 showed polydispersity giving broad peaks in gel filtration elution profiles.

3.3.3.4 Oligomeric State Characterisation by Negative Staining Electron Microscopy

Negative staining electron microscopy was employed as an effective and accurate method in characterising the oligomeric state of the fusion assemblies. Ring-like particles were distributed throughout the electron micrographs, with a detectable extra ring of TRAP above the gp6 ring in the fused proteins compared to the wild-type gp6. The preliminary analysis of eigen images performed in Prof. Orlova's laboratory (Birkbeck College, London) clearly showed the 13-fold symmetry for fusion proteins containing truncated gp6 Δ27-466, Δ27-472 and Δ27-479 (Figure 3-17). Furthermore, no oligomeric state change was observed from the 13-mer YM92 fusion protein (containing gp6 Δ27-466) after the dissociation-re-association procedure. For the fusion constructs containing gp6 Δ27-434, the preliminary results indicated mixtures of 13-mer and 14-mer assemblies (data not shown).

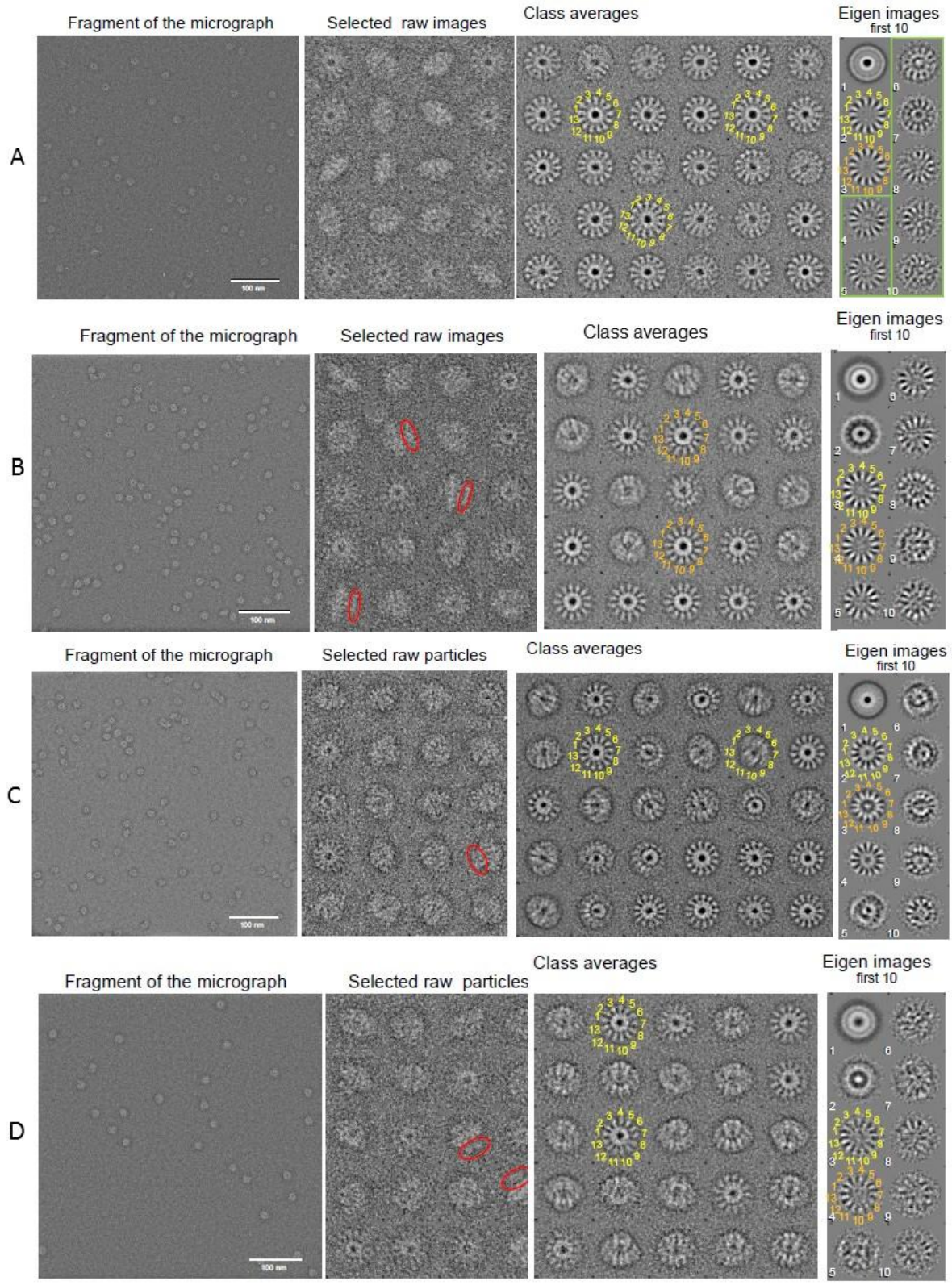


Figure 3-17 Electron microscopy of negative stained fusion proteins of gp6-*B. halodurans* TRAP.

Fragments of EM micrographs, selected raw particles, class averages and eigen images were shown for each sample: A- control wild type gp6, B - YM92, C - YM97

and D - YM96. Red circles indicate an extra ring corresponding to TRAP protein, and the bar represents 100 nm.

3.3.3.5 Crystallisation

Among all TRAP fusion constructs formed with Δ27-466, Δ27-472 and Δ27-479 gp6 (YM91-YM100), the construct YM96, with truncated 26 N-terminal residues, 31 C-terminal residues and $(GSS)_3$ linker, formed crystals most readily. The 10 mg/mL protein solution was in 50 mM imidazole pH 8.0, 250 mM NaCl, and 10 mM $MgCl_2$. YM96 formed ~50 μm cubic crystals with the reservoir containing 0.2 M NaCl, 40% MPD. A number of different plate-like crystals were obtained with reservoirs containing either 0.2 M ammonium acetate, 0.1 M BIS-TRIS pH 5.5, 45% MPD, or 0.2 M ammonium acetate, 0.1 M BIS-TRIS pH 6.5, 45% MPD (Figure 3-18, Index™ screen, Hampton Research). Crystals grown with 0.2 M $MgCl_2$ and 40% MPD in the reservoir diffracted to ~16 Å and crystals grown with 0.2 M ammonium acetate, 0.1 M BIS-TRIS pH 5.5, 45% MPD in the reservoir diffracted to ~17 Å. Optimisation of the conditions giving initial crystal hits did not generate crystals with enhanced diffraction. There were no crystals obtained for fusion proteins containing gp6 Δ27-434 that appeared to be mixtures of 13-mers and 14-mers according to preliminary EM analysis (data not shown).

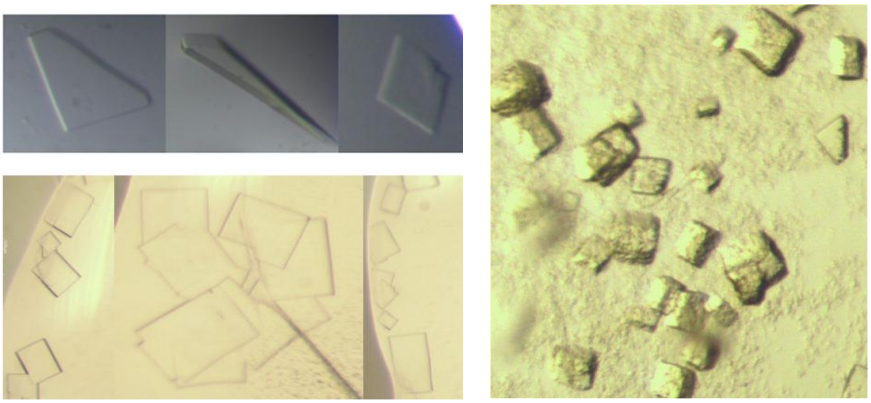


Figure 3-18 Crystals of portal-TRAP fusion protein YM96.

3.4 Discussion

The fusion proteins of the SPP1 portal protein with *B. halodurans* TRAP can be successfully overexpressed in *E. coli* BL21 cells with an N-terminal histidine tag. These fusion proteins can be purified by Nickel affinity chromatography followed by size-exclusion chromatography, in which the major peak was shown to

correspond to the stable oligomer. According to the SEC-MALLS measurements, the molecular weights of most purified fusion proteins are closer to the theoretical molecular weight of a 13-subunit oligomer, with the exception of three polydispersed protein samples that appeared to contain 12-mer assemblies. The accuracy of the SEC-MALLS technique however does not allow for a decisive conclusion about the oligomeric state of proteins containing more than ~10 subunits. This is due to a typical systematic error of 5% during the molecular weight determination (Folta-Stogniew & Williams, 1999).

By electron microscopy studies, ring-like particles are distributed throughout the electron micrograph. Compared to the wild-type gp6, an extra ring corresponding to the TRAP assembly is detectable above the gp6 ring in the fusion proteins. The preliminary analysis of eigen images clearly demonstrated the 13-fold symmetry of fusion proteins with gp6 truncations Δ27-466, Δ27-472 and Δ27-479 (Figure 3-17). However, whether the TRAP protein contains 12 or 13 subunits in these assemblies is unknown.

For the fusion proteins with gp6 Δ27-434, the preliminary results indicated mixtures of 13-mer and 14-mer assemblies (data not shown). According to previous studies, the removal of the whole crown domain (α7- α9) would result in 14-mer assembly of gp6 (personal communication with Paulo Tavares). Since gp6 Δ27-434 keeps only the α7 helix in the crown domain, the removal of α8 and α9 helices might explain the co-existence of the 13-mer and 14-mer assemblies in solution. The heterogeneous oligomeric state of these proteins is the likely reason for these proteins not forming crystals.

The oligomeric state of the gp6 fusion construct YM92 did not change after the dissociation re-association procedure, with the protein still found in the 13-subunit oligomerisation state. The dissociation-re-association experiment was aimed to determine which of the two oligomeric proteins (gp6 or TRAP) defines the oligomeric state of the fused construct. The result suggests that the oligomeric state of the fusion proteins is determined by gp6, whose size is significantly larger than TRAP (50.7 kDa per subunit versus 8.3 kDa in the case of TRAP), although its subunit-subunit interactions are much weaker than TRAP.

The study of the TRAP oligomer stability from different *Bacilli* showed that *B. stearothermphilus* wild type and E71Stop TRAP form the most stable oligomers. Hence, these are most likely suitable for further fusion proteins studies with the portal protein. However, several fusion proteins constructs with the *B. stearothermphilus* TRAP, generated in this study, proved to be insoluble.

4 Co-expression of the SPP1 Portal Protein with the Scaffolding Protein

4.1 Introduction

Viral assembly is a multistep process involving a specific order of protein-protein and protein-nucleic acid interactions. For the assembly of biologically active SPP1 procapsids, three essential structural components are required - the major capsid protein gp13, the scaffolding protein gp11 and the portal protein gp6. The geometrically correct polymerisation of gp13 is directed by the scaffolding protein (Becker et al, 1997), while the formation of regularly sized procapsids is controlled by the portal protein, which influences the copolymerisation of gp11 and gp13 (Droge et al, 2000). Procapsid-like structures formed when the scaffolding protein and major capsid protein encoding genes 11 and 13 were co-expressed in a heterologous host. However, there was no interaction observed between the two proteins, when they were produced alone and then mixed *in vitro* (Becker et al, 1997; Droge et al, 2000). Similarly, stable interactions between the portal protein gp6 and scaffolding protein gp11 could be detected only when the proteins were coproduced (Droge et al, 2000). The interaction between the portal and scaffolding proteins was also observed in the cryo-EM studies of phage P22 (Chen et al, 2011b).

The SPP1 portal protein is a cyclical dodecamer in the virion (Lurz et al, 2001), while the protein produced *in vitro* is a 13-subunit assembly (Lebedev et al, 2007). It was proposed that the interaction of the portal protein with other procapsid proteins, scaffolding protein or major capsid protein, would impose an increased bend between gp6 subunits and maintain a 12-mer assembly (Lurz et al, 2001). In this study, the SPP1 portal protein and scaffolding protein were co-expressed in the heterologous host *E. coli,* with the aim of promoting the assembly of a 12-subunit portal protein.

4.2 Methods and Materials

4.2.1 Cloning Strategy for Co-Expression of Gp6 and Gp11

The plasmid pYM70 encoding gp6 Δ27-472 was first transformed into *E. coli* B834 cells. The transformant was treated with calcium chloride to make *E. coli* B834 competent cells carrying plasmid pYM70. The plasmid pYM184 encoding full-length gp11 was transformed into the *E. coli* B834 competent cells with plasmid pYM70.

4.2.2 Purification of the Gp6 and Gp11

A single colony carrying plasmids pYM70 and pYM184 was picked from a Luria–Bertani (LB) agar plate containing 30 μg/mL kanamycin and 50 μg/mL chloramphenicol to inoculate a small overnight culture at 37 °C. 5 mL aliquots of the overnight culture were used to inoculate 500 mL cultures the following day. The 500 mL cultures were grown at 37 °C until the mid-log phase (OD600 of 0.6–0.8) was reached, at which point protein expression was induced by the addition of isopropyl-d-1-thiolgalactopyranoside (IPTG) to a final concentration of 1 mM. Lastly, overnight incubation of the culture at 16 °C and 180r.p.m was carried out for the over-expression of proteins. Cells were recovered by centrifugation at 5000 g (SORVALL® RC 5B Plus rotor) at 4 °C for 20 minutes. Supernatant was removed and cell pellet was stored at -20 °C.

The cell pellet was re-suspended in nickel affinity chromatography binding buffer, 20 mM Tris pH 7.5, 150 mM NaCl, 10 mM $MgCl_2$, 50 mM Imidazole, supplemented with a combination of 100 μg/mL lysozyme and protease inhibitors (1 mM AEBSF and 0.7 μg/mL pepstatin). Cells were disrupted by sonication with large probe in glass beaker using short pulses of 30 seconds, with a 2 minutes resting time in between pulses. The sonication was carried out on ice, with caution to minimise thermal damage to protein extract. The cell debris was removed by centrifugation at 38758 g for 30 minutes at 4 °C using a Sorvall SS34 rotor. The supernatant was collected and then cleared with a 0.45 mm filter (Millipore). The filtrate loaded onto a 5 mL HiTrap column (GE Healthcare) equilibrated with binding buffer for nickel affinity chromatography purification. The bound protein was eluted with an increasing proportion of elution buffer containing 20 mM Tris pH 7.5, 150 mM

NaCl, 10 mM $MgCl_2$, and 500 mM imidazole. Fractions containing both gp6 and gp11 were pooled together, and concentrated using Vivascience 30 kDa molecular weight cut-off concentrators. The concentrated sample (less the 10 mg/mL) was applied to a HiPrep 16/60 Sephacryl S-500 HR column (GE Healthcare) for size exclusion chromatography. The proteins were eluted with 20 mM Tris pH 7.5, 150 mM NaCl, 10 mM $MgCl_2$, and concentrated to at least 10 mg/mL and stored at -80 °C before crystallisation.

4.2.3 His Tag Pull-Down Assay

400 µg of individually purified His6-gp11 were immobilized by 400 µL nickel chelate beads and incubated with 87.5 µg of individually purified untagged gp6 Δ27-472 in 20 mM Tris pH 7.5, 150 mM NaCl, 10 mM $MgCl_2$ (total volume: 465 µL) for one hour at room temperature. The beads were washed three times with binding buffer to remove unbound proteins. Samples of total proteins, unbound proteins and bound proteins were analysed by 12% SDS-PAGE. As a control, the untagged gp6 Δ27-472 alone and the untagged gp11 alone were also incubated with 200 µL nickel chelate beads in buffer 20 mM Tris pH 7.5, 150 mM NaCl, 10 mM $MgCl_2$ for one hour at room temperature.

4.2.4 Size-Exclusion Chromatography coupled with Multi-Angle Laser Light Scattering (SEC–MALLS)

The molecular mass was determined by size-exclusion chromatography coupled with multi-angle laser light scattering (SEC–MALLS). The protein sample, 60 µL of fraction A12 from size chromatography using a Superose 6 HR 10/30 column containing both gp6 Δ27-472 and gp11 at a concentration of 0.5 mg/mL, was applied to a BioSep™ SEC-s3000 gel filtration column (Phenomenex) equilibrated with buffer containing 20 mM Tris pH 7.5, 150 mM NaCl, 10 mM $MgCl_2$. As a control, the separately purified gp6 Δ27-472 was analysed under the same conditions. Size-exclusion chromatography was carried out on a Shimadzu HPLC system and the elution was monitored at 280 nm by an SPD20A UV/Vis detector. Light-scattering data were recorded on a Dawn HELEOS-II 18-angle light-scattering detector and the concentration of the eluting protein was measured by an in-line Optilab rEX refractive-index monitor (Wyatt Technology). Data were analysed with the ASTRA V software package (Wyatt Technology).

Molecular mass was calculated using Zimm's formalism of the Rayleigh–Debye–Gans light-scattering model for dilute polymer solutions and a refractive-index increment (dn/dc) of 0.183 mL/g was used for the protein molecular mass estimation.

4.2.5 Negative Staining Electron Microscopy

The fraction A12 from size exclusion chromatography using a Superose 6 HR 10/30 column containing both gp6 Δ27-472 and gp11 was diluted to a concentration suitable for single particle observation (~ 0.05 mg/mL). A 2 µL aliquot of the sample was applied to glow-discharged, continuous-carbon and formvar covered copper grids with a 300-square mesh. The dried sample was washed with MiliQ water and then stained with 2% uranyl acetate for two minutes. The grids were air-dried before inspection on a 200 kV FEI F20 electron microscope at a magnification of 100, 000x. Imaging was performed in a low electron dose mode. 10 micrographs were recorded and approximately 1200 protein particles were selected for image analysis. These experiments were performed by Elena Orlova at Birkbeck College, University of London.

4.2.6 Crystallisation and Data Collection

The sitting-drop vapour diffusion method was used for initial crystallisation screening. Drops containing 150 nL protein solution and 150 nL reservoir solution were dispensed by a Mosquito Nanolitre Pipetting robot (TTP Lab-tech) in 96-well plates, and equilibrated against 60 µL of reservoir solution. To obtain initial crystal hits, crystal trays with the MPD screen were set up. Conditions in which small crystals grew were optimised in 24-well hanging-drop plates by manual pipetting. Crystals were tested in-house using a Rigaku RU-H3R rotating-anode generator equipped with Osmic multilayer optics and a MAR Research MAR345 imaging-plate detector. Native data were collected from a single crystal on the Diamond I04-1 beamline at the synchrotron to 2.8 Å resolution at 0.92 Å wavelength. Cryosolutions were not added for crystal freezing as 40% MPD was already present in the crystallisation condition. The data were processed using XDS.

4.2.7 Structure Determination and Refinement

All crystallographic calculations were performed using the CCP4 program package (Winn et al, 2011b). The X-ray structure of gp6 Δ27-472 was solved by molecular replacement using PHASER (Mccoy et al, 2007) with a single subunit of gp6 Δ29-466 as the search model. The search model was generated by the combination of two structures - residues 29 to 340 (including the disordered wing region) from a 2.9 Å resolution structure of 14-mer (unpublished data available in the group), and residues 341 to 466 from the 3.4 Å resolution structure of 13-mer.

Refinement was carried out by REFMAC (Murshudov et al, 1997). The model was built using COOT (Emsley & Cowtan, 2004), and further corrected using maps calculated with maximum likelihood-weighted coefficients $2|F_o| - |F_c|$ and $|F_o| - |F_c|$. All figures were generated using PyMOL (Schrodinger, 2010).

4.3 Results

4.3.1 Cloning, Co-expression and Purification of Portal Protein and Scaffolding Protein

The gene encoding full-length gp11 was cloned into the restriction sites *NdeI* and *XhoI* of vector pET28a to form plasmid pYM184 for the expression of N-terminal hexahistidine-tagged protein, and the plasmid pYM70 was successfully cloned with the gene encoding gp6 Δ27-472 ligated into the *NcoI* and *BamHI* sites of vector pACYCDuet to express the untagged gp6 Δ27-472. The plasmids pYM70 and pYM184 with compatible replicon and antibiotic resistance were transformed into the same competent *E. coli* cell for the co-expression of the His-tagged full-length scaffolding protein gp11 and the untagged truncated portal protein gp6 Δ27-472.

Gp6 Δ27-472 and full-length gp11 were both over-expressed in soluble form in the same *E. coli* B834 cells at 16 °C (Figure 4-1). The purification was performed by nickel affinity chromatography followed by size exclusion chromatography. The fractions containing both gp6 and gp11 were eluted at an imidazole concentration between 162 mM and 255 mM (Figure 4-2A). During size exclusion chromatography, the elution profile on a Superose 6 HR 10/30 column showed peaks corresponding to aggregate, fractions containing both gp6 and gp11

(A11-B2), and fractions containing excess gp11 (Figure 4-2B). However, the elution profile on a HiPrep 16/60 Sephacryl S-500 HR column corresponded to an overlapping elution of the two separate proteins, but no fraction containing both proteins was eluted before the gp6 fractions (Figure 4-2C).

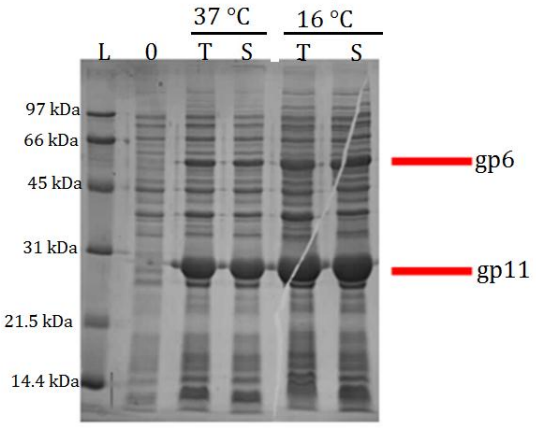
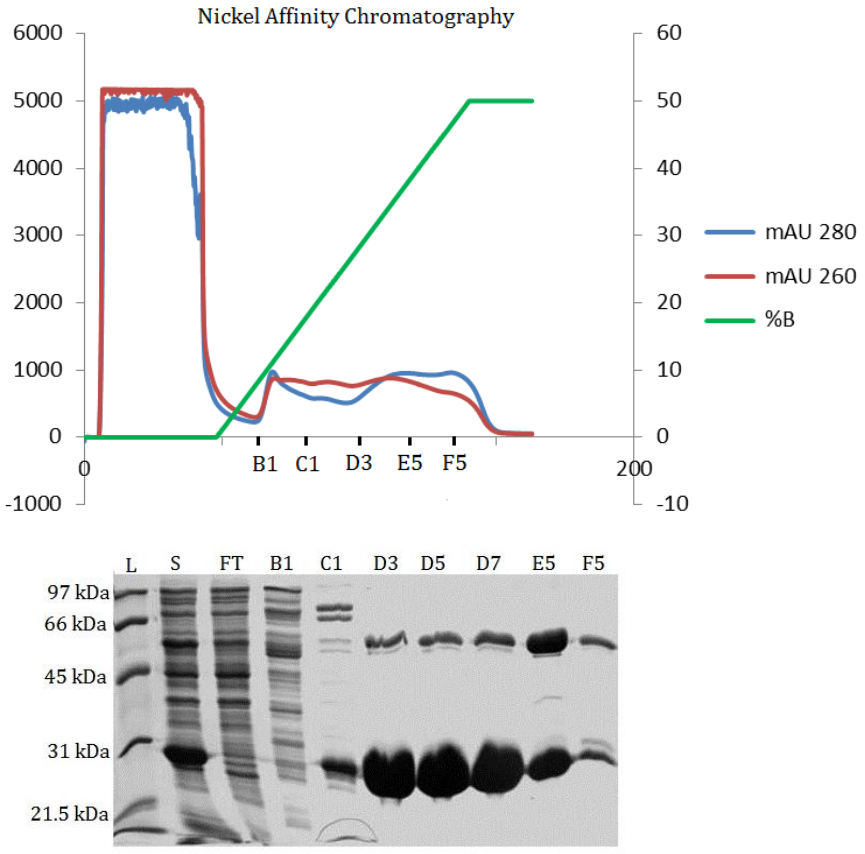
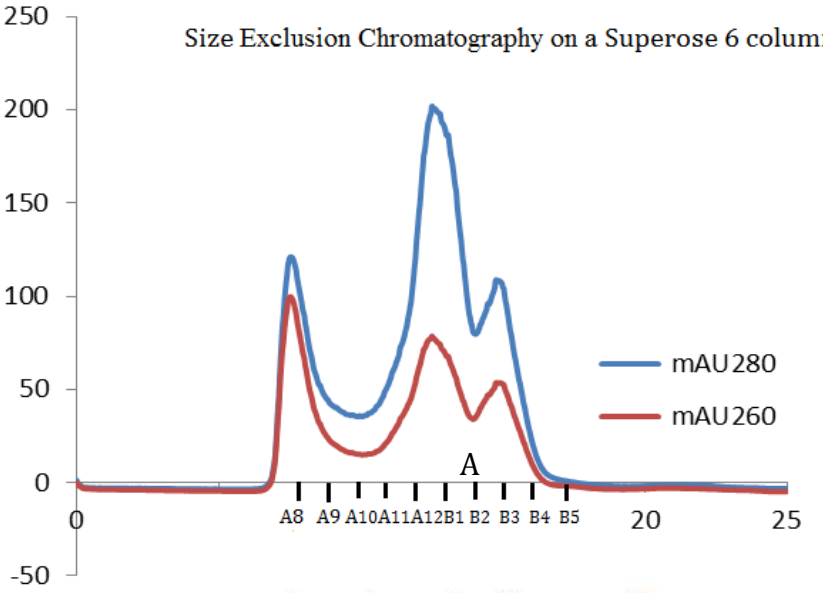


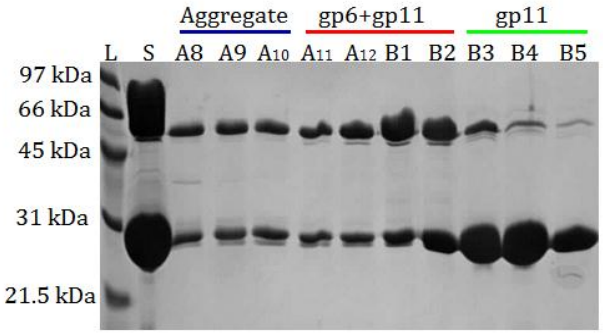
Figure 4-1 Co-expression of gp6 Δ27-472 and full-length gp11.
The co-expression of the two proteins was carried out in E. coli B834 cells induced at 16 °C.

A

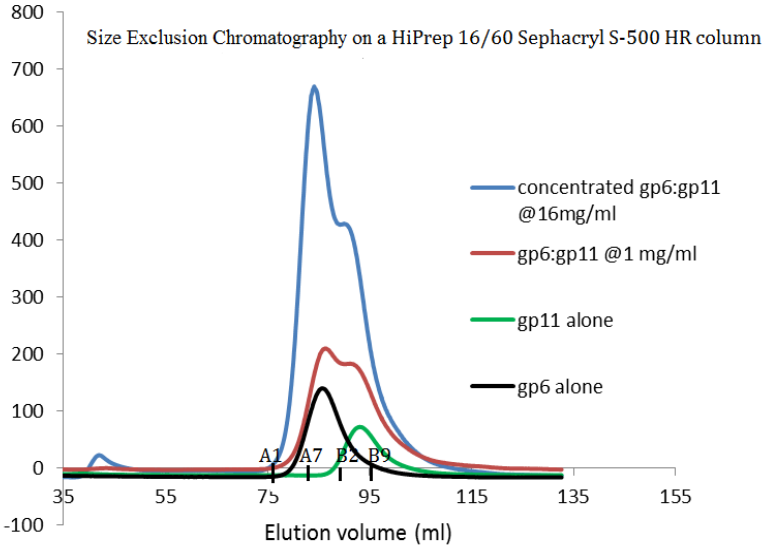


B


Size Exclusion Chromatography on a Superose 6 column
250
200
150
100
50
0
-50
0
20
25
mAU280
mAU260
A
A8 A9 A10 A11 A12 B1 B2 B3 B4 B5


Aggregate
gp6+gp11
gp11
L S A8 A9 A10 A11 A12 B1 B2 B3 B4 B5
97 kDa
66 kDa
45 kDa
31 kDa
21.5 kDa

C


Size Exclusion Chromatography on a HiPrep 16/60 Sephacryl S-500 HR column
800
700
600
500
400
300
200
100
0
-100
35
55
75
95
115
135
155
Elution volume (ml)
concentrated gp6:gp11 @16mg/ml
gp6:gp11 @1 mg/ml
gp11 alone
gp6 alone
A1 A7 B2 B9

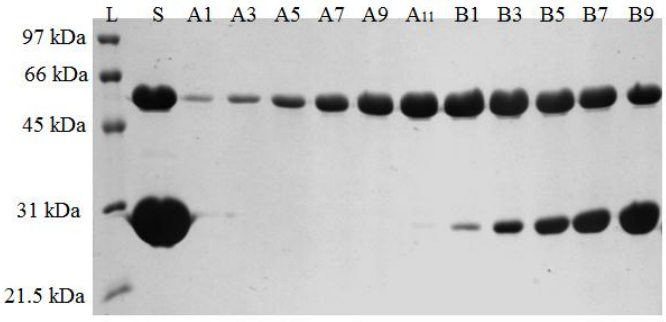

L S A1 A3 A5 A7 A9 A11 B1 B3 B5 B7 B9
97 kDa
66 kDa
45 kDa
31 kDa
21.5 kDa

Figure 4-2 Purification of the co-expressed gp6 Δ27-472 and gp11.
(A) Purification by nickel affinity chromatography; (B) Purification by size exclusion chromatography on a Superose 6 HR 10/30 column, the sample loaded on the column was the pool of fractions containing both gp6 and gp11 from nickel affinity chromatography (Fractions D3-F5), and the running buffer was 20 mM Tris pH 7.5, 150 mM NaCl, 10 mM $MgCl_2$; (C) Purification by size exclusion chromatography on a HiPrep 16/60 Sephacryl S-500 HR column, the sample loaded on the column was the pool of fractions containing both gp6 and gp11 from nickel affinity chromatography (Fractions D3-F5), and the running buffer was 20 mM Tris pH 7.5, 150 mM NaCl, 10 mM $MgCl_2$. The elution profile was shown on the top panel. The same protein samples at 16 mg/mL (coloured in blue) and 1 mg/mL (coloured in red) were separately run on the column to compare the effect of the sample's concentration on the elution profile. Samples containing gp6 alone (coloured in black) and gp11 alone (coloured in green) were run as controls. As shown on the bottom panel, the elution fractions from the run of the concentrated sample at 16 mg/mL were analysed by 12% SDS-PAG. S: the sample for purification containing both gp6 and gp11 from nickel affinity chromatography (Fraction D3-F5). A1-B9: the purification fractions as illustrated on the elution profile on the top panel.

4.3.2 Probing If a His-Tag Pull-Down Assay Could Be Used to Detect Interaction between the Portal Protein and the Scaffolding Protein

Because the full-length gp11 was expressed with an N-terminal histidine affinity tag and the gp6 Δ27-472 didn't contain a histidine tag, the initial plan was to detect the interaction between gp6 and gp11 by His-tag pull down assays. Unexpectedly, preliminary results showed that the untagged gp6 Δ27-472 and the untagged gp11 were both able to bind to nickel-chelate beads in the absence of the polyhistidine tag (Figure 4-3). Consequently, His tag pull down assays cannot be employed to detect interactions between the two proteins.

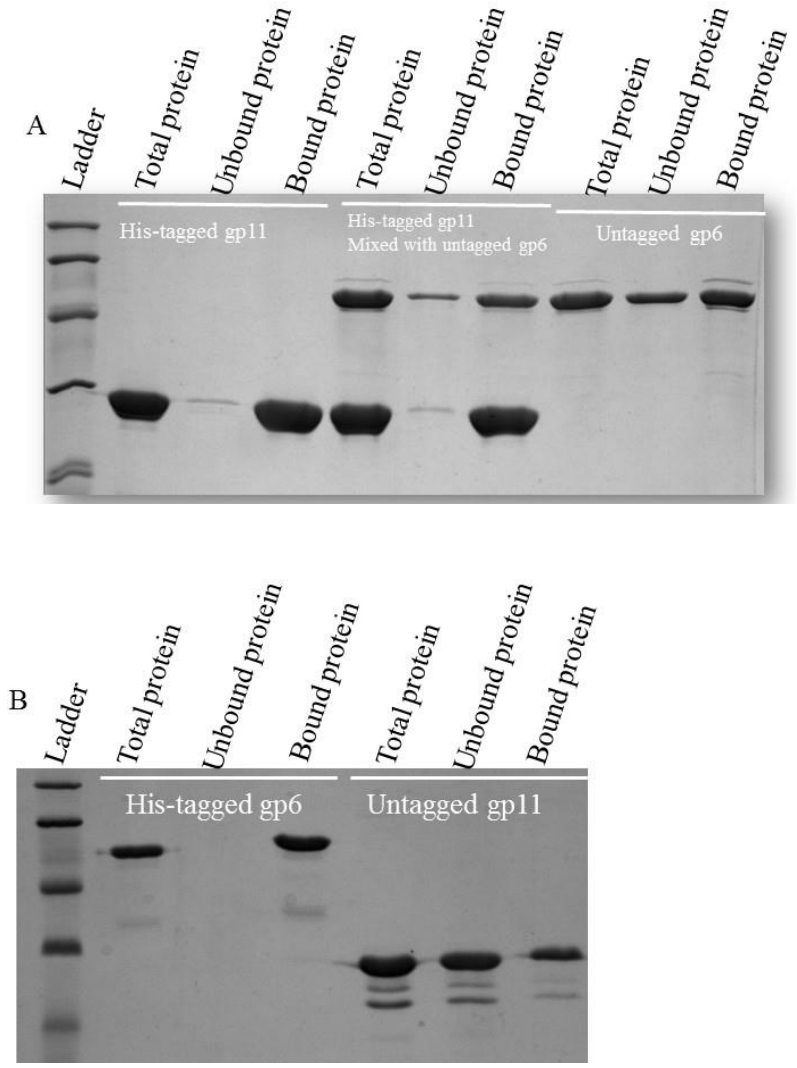


Figure 4-3 His-tag pull down assays of gp6 Δ27-472 and gp11.
Untagged gp6 Δ27-472 and untagged gp11 were able to bind to nickel chelate beads. (A) Pull down assay with samples: his-tagged gp11, untagged gp6, and the mixture of His-tagged gp11 with untagged gp6; (B) Pull down assay with samples his-tagged gp6 and untagged gp11. Total proteins, unbound proteins and bound proteins were analysed by 12% SDS-PAGE.

4.3.3 Oligomeric State Analysis by SEC-MALLS

The recombinant protein construct YM70 consists of 446-amino acid long untagged portal protein gp6 Δ27-472 and two residues introduced by the vector - Met and Gly. The theoretical molecular mass calculated by ProtParam is 51615.6 Da.

In order to determine the oligomeric state of the portal protein after co-expression with the scaffolding protein, the fraction A12 from size-exclusion chromatography on a Superose 6 HR 10/30 column containing both gp6 and gp11 was investigated by SEC-MALLS on a BioSep™ SEC-s3000 gel filtration column (Phenomenex) equilibrated with 20 mM Tris (pH 7.5), 150 mM NaCl, and 10 mM $MgCl_2$.

The elution profile of the size-exclusion chromatography on a BioSep™ SEC-s3000 gel filtration column showed separate peaks corresponding to gp6 alone and gp11 alone respectively, but there was no peak corresponding to the complex of the two proteins. SEC-MALLS data showed the mean molecular weight of the peak corresponding to gp6 Δ27-472 alone as 629.2 kDa, or ~12.2 subunits per oligomer in solution (Figure 4-4). Meanwhile, the separately produced gp6 Δ27-472 was also subject to SEC-MALLS characterisation under the same experimental conditions as a control. The molecular weight of gp6 Δ27-472 control was determined to be 679.3 kDa, or ~13.2 mer subunits per oligomer, which suggested gp6 as a 13mer in solution.

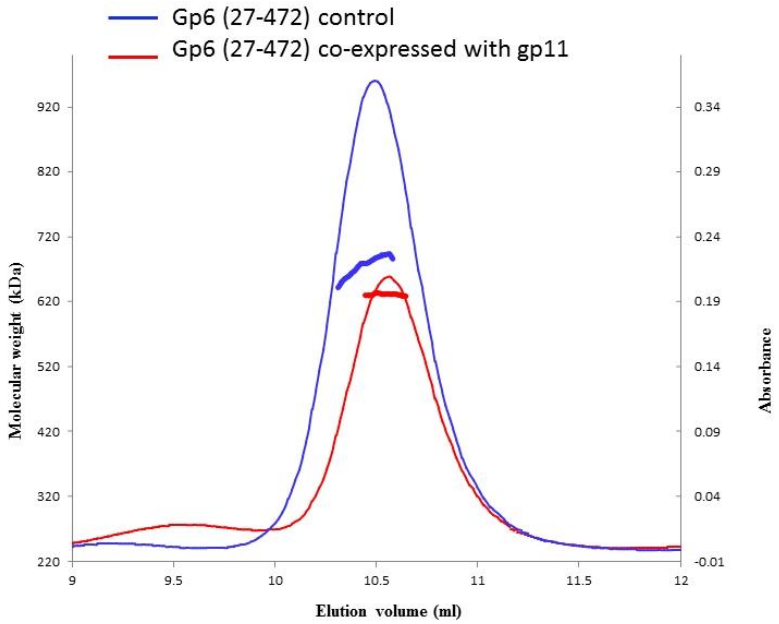


Figure 4-4 Characterisation of the oligomeric state of the co-expressed gp6 by SEC–MALLS.
The thin line corresponds to the absorbance monitored at 280 nm. The thick line shows the molecular weight calculated for the eluted species. Separately produced gp6 was applied as a control and coloured in blue, whereas the elution of co-expressed gp6 was coloured in red.

4.3.4 Determination of the Oligomeric State by Negative Staining Electron Microscopy

The fraction A12 from size exclusion chromatography on a Superose 6 HR 10/30 column containing both gp6 Δ27-472 and gp11 was also investigated by negative staining electron microscopy to characterise the oligomeric state of gp6, and the interaction between gp6 and gp11. Gp6 ring-like particles were observed and the preliminary analysis of eigen images clearly showed 13-fold symmetry of gp6 (Figure 4-5). Interestingly, there seemed to be several gp11 molecules attached to the gp6 complex as well.

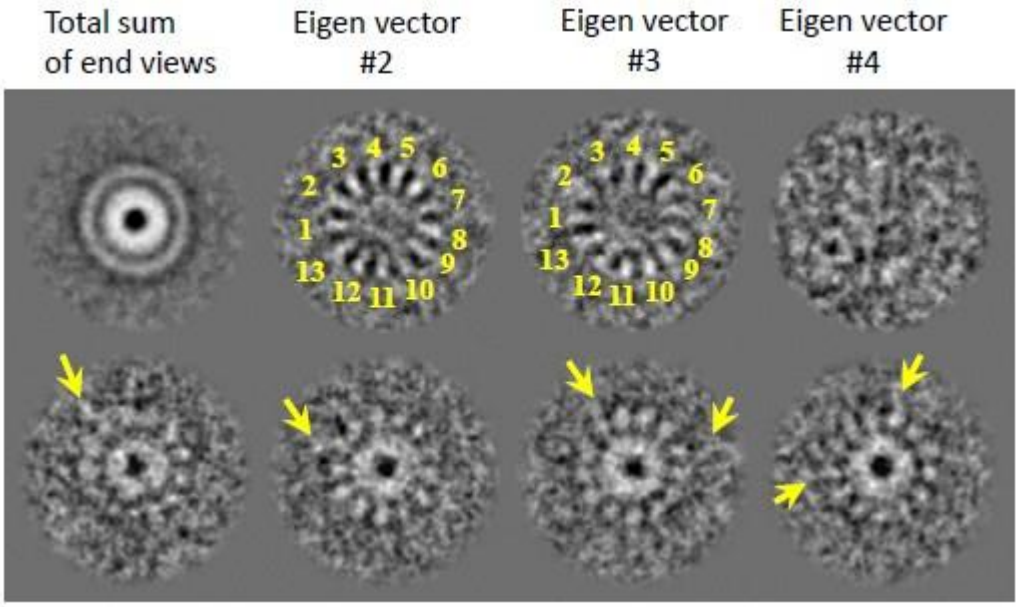


Figure 4-5 Negative staining electron microscopy analysis of the co-expressed gp6 and gp11.
The bottom row represented averages of end views (not symmetrised), and arrows indicated possible positions of gp11. The sample subject to observation was the fraction from gel filtration purification containing the mixtures of gp6 Δ27-472 and gp11.

4.3.5 Crystallisation of Gp6 Δ27-472

The protein sample used for crystallisation was prepared after co-expression with gp11, nickel affinity chromatography and size exclusion chromatography on a HiPrep 16/60 Sephacryl S-500 HR column. The fractions containing gp6 alone were collected and concentrated to 19 mg/mL. Here, gp6 Δ27-472 exhibited a very

high crystallisation propensity under MPD screen conditions. The best diffracting crystals were produced under the conditions 0.2 M Lithium acetate, 40% MPD (condition B12 of MPD screen) and 0.1 M Tris pH 8.0, 40% MPD (condition F5 of MPD screen). Crystals from both conditions were tested in-house, with diffraction extending to approximately 3.7 Å resolution. The best crystals were stored and shipped to the Diamond Light Source for collection of a complete data set. Sets of complete X-ray data were collected from three different crystals. The best data set extended to a resolution of 2.8 Å (beamline I04-1, Diamond Light Source, Figure 4-6).

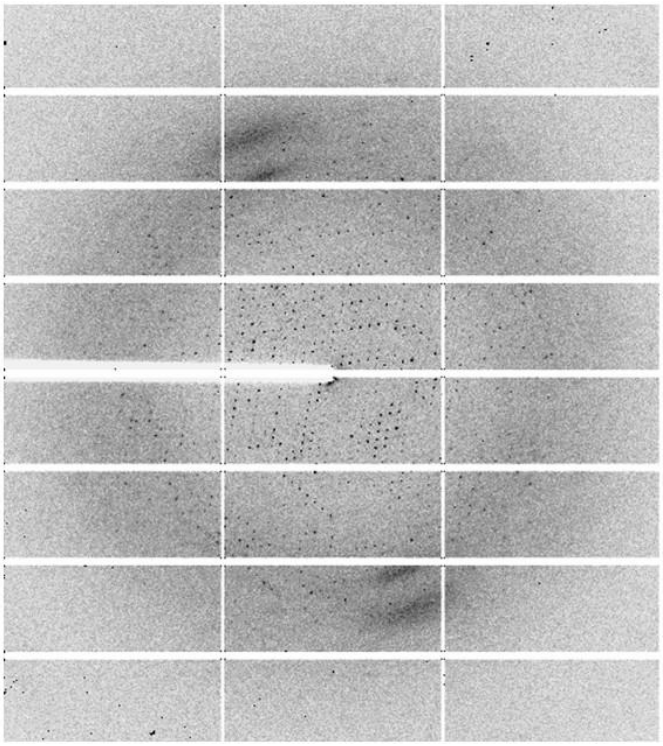


Figure 4-6 Diffraction image.
Resolution at the edges of the plate is 1.88 Å.

4.3.6 Structure Determination

The X-ray dataset from the best crystal was indexed and integrated using XDS and the space group was determined to be C2 with unit cell dimensions of a=224.8 Å, b=176.8 Å, c=208.1 Å, β=112.45°. The data were further processed using the CCP4 program AIMLESS. Statistics of the merged data are shown in Table 4-1.

The crystallographic calculations were performed using the CCP4 program package (Winn et al, 2011b). The structure of gp6 Δ27-472 was solved by molecular replacement using PHASER (Mccoy et al, 2007), with a single subunit of gp6 Δ29-466 as the search model. The model was refined at 2.8 Å to the final Rfree of 26.6% (Rfactor=20.7%) (Table 4-1). The electron density of the wild type gp6 is well defined for most amino acids, including Asn365 (Figure 4-7 shows electron density at 2 σ coutouring level). This is the residue which was substituted by lysine in the previously determined structure of gp6 (Lebedev et al, 2007).

Like N365K mutant gp6, the isolated wild type protein forms 13-subunit assembly (Figure 4-8). The superposition of Cα traces for the subunits of the wild type and the N365K mutant gp6 reveals identical folds (Figure 4-9). Most significant conformational differences are found in the tunnel loop segment, a loop in the clip, and in the crown domain (Figure 4-9). In addition, the segment from residue 170 to 238, which was poorly defined in the structure of N365K gp6, is well defined in the current model.

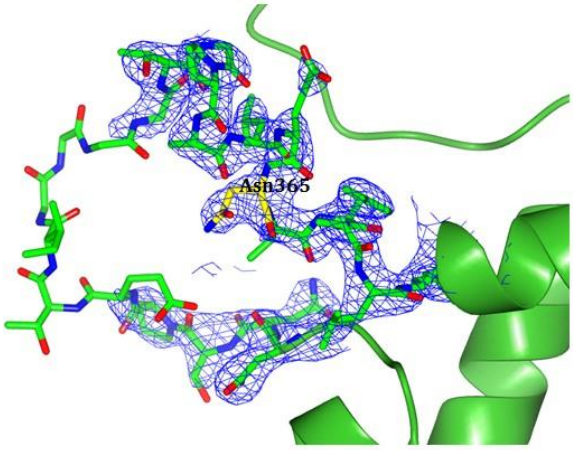


Figure 4-7 Electron density of the residue Asn365 in the WT gp6 shown on $2|F_o| - |F_c|$ map.
The coutouring level is 2 σ.

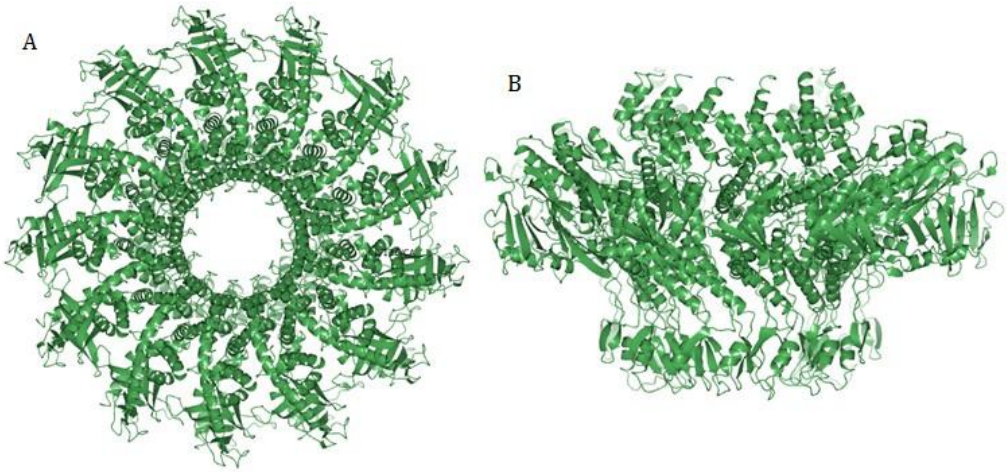


Figure 4-8 Structure of the wild-type gp6.

(A) Ribbon diagram representing the top view of the wild-type gp6; (B) Ribbon diagram representing a side view of the wild-type gp6.

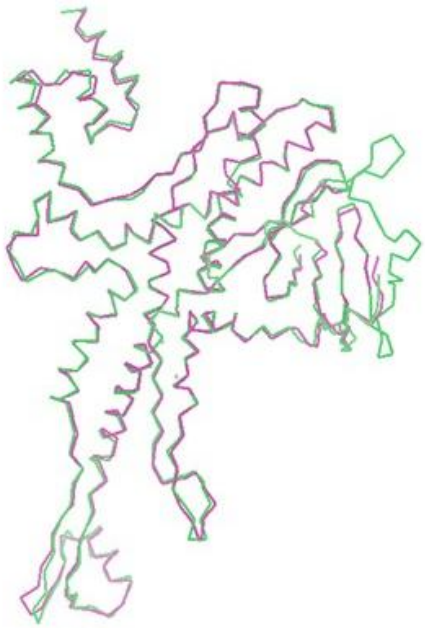


Figure 4-9 Superposition of Cα traces for the subunit of the WT gp6 and the N365K mutant gp6.

The wild type gp6 is coloured in green, and the N365K mutant gp6 structure is coloured in magenta.

In both wild type and mutant gp6, residue 365 is involved in subunit-subunit interactions. In the wild type gp6, Asn356 participates in the formation of two hydrogen bonding interactions with Asn349 of adjacent subunit: one hydrogen bond is formed between the side chain amine hydrogen of Asn356 and the backbone carbonyl oxygen of Asn349, and the other hydrogen bond is formed

between the side chain carbonyl oxygen of Asn356 and the side chain amine hydrogen of Asn349 (Figure 4-10A). In contrast, in the N365K gp6, Lys365 forms only one hydrogen bonding interaction, between its side chain amine hydrogen and the backbone carbonyl oxygen of Gly357 from an adjacent subunit (Figure 4-10B). The side chain of this lysine residue also makes van der waals interactions with Asn349 of this tunnel loop, being directed towards its centre by the hydrogen bonding interactions. These inter-subunit interactions could play an important role in maintaining the functional structure of the 15-residue tunnel loops (345-359), which constitute the most constricted part of the tunnel and would form close contacts with DNA during its translocation (Lebedev et al, 2007). The flexibility of tunnel loops is considered to be crucial for the packaging of DNA into procapsids. It is noteworthy that the residues involved in the hydrogen bonding contacts with Asn365/Lys356 are different - Gly357 in the mutant gp6 and Asn349 in the wild type gp6. Both residues are located within the tunnel loop, however, occupy opposite positions - Gly357 is situated at the top , close to the C-terminus of the tunnel loop, while Asn349 is at the bottom in the N-terminal part of the loop. Compared to the wild type gp6, the inter-subunit hydrogen bonding interaction between Gly357 and Lys365 and van der Waals interactions in the mutant gp6 could restrict tunnel loop's conformation reducing its flexibility. This suggestion is supported by comparison of the tunnel loop's electron density which reveals it is better defined in the case of the N365K mutant gp6. The reduced flexibility of the tunnel loops could account for the different behaviour of capsids containing the mutant gp6: previous studies showed that the N365K substitution resulted in reduction of the length of the encapsidated DNA, although the DNA packaging process was not impaired (Tavares et al, 1992).

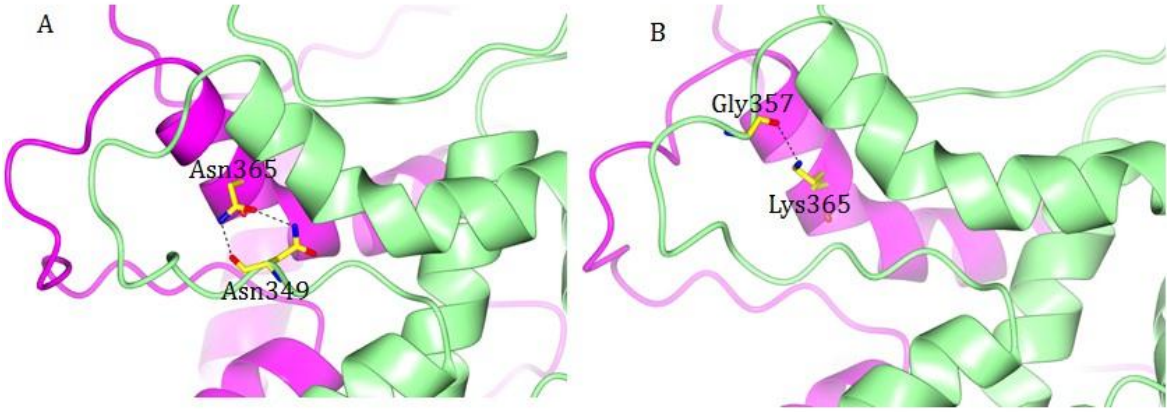


Figure 4-10 Comparison of the hydrogen-bonding interactions formed by Asn365 in the WT gp6 and Lys365 in the mutant gp6.
(A) Two hydrogen bonds are formed between Asn365 and Asn349 in the wild type gp6, (B) one hydrogen bond is formed between Lys365 and Gly357 in the mutant gp6. All strong hydrogen bonding interactions are shown as dotted lines, and the adjacent subunits are coloured in green and magenta, respectively.

Table 4-1 Crystallographic statistics

Data collection statistics	
X-ray source	I04-1, Diamond
Wavelength, Å	0.92
Temperature, K	100
Space group	C2
Unit cell parameters, Å	a=224.8, b=176.8, c=208.1
	α=90°, β=112.5°, γ=90°
Resolution Range, Å	48.30 – 2.80 (2.85-2.80)
No. of unique reflections	179891(8876)
Rmerge[a], %	10.1 (107.5)
Completeness, %	98.9 (98.8)
Redundancy	3.5 (3.5)
Average I/ **σ**(I)	9.4 (1.1)
Refinement statistics	
No. reflections	178069
R_{work}/ R_{free},%	20.7/26.6
Number of water molecules	328
R.m.s deviation from ideal bond length, Å	0.0046
R.m.s deviation from ideal bond angles, °	0.8391
Ramachandran Plot, %	
In preferred regions	94.72
In allowed regions	4.72
Outliers	0.56

[a]Rmerge = $\sum_{hkl} \sum_i |I_i (h,k,l) - \langle I(h,k,l) \rangle| \; / \; \sum_{hkl} \sum_i I_i (h,k,l)$, where $I(h,k,l)$ is the intensity of reflection, $\langle I(h,k,l) \rangle$ is the average value of the intensity, the sum $\sum_{hkl}$ is over all measured reflections and the sum $\sum_i$ is over i measurements of a reflection. Values in parentheses are for the highest-resolution shell.

4.4 Discussion

In this study, the full-length SPP1 scaffolding protein gp11 was co-expressed with the portal protein gp6 Δ27-472 in *E. coli* cells with the aim to induce the formation of a 12-subunit gp6 assembly in the presence of the capsid morphogenesis protein gp11. The two proteins were successfully co-expressed in the same cell. Purification of the two proteins was carried out by nickel affinity chromatography and size exclusion chromatography. Fractions containing both proteins were eluted from size exclusion chromatography with a Superose 6 HR 10/30 column, whereas fractions eluted from a HiPrep 16/60 Sephacryl S-500 HR column corresponded to separate proteins. His-tag pull-down assays revealed the untagged gp6 Δ27-472 was able to bind to nickel chelate beads. Therefore, the binding of untagged gp6 to a HiTrap affinity column was unlikely to be caused by the interaction with his-tagged gp11. Moreover, fractions containing both proteins that were eluted from size exclusion chromatography are more likely due to an overlapping elution of the two separate proteins instead of the gp6-gp11 complex.

The SEC-MALLS data on the size-exclusion chromatography fraction containing both gp6 and gp11 showed two peaks corresponding to individual proteins. Calculated oligomeric state of the peak corresponding to gp6 was 12.2 subunits, suggesting a 12-mer assembly. A control experiment with separately produced gp6 had a calculated oligomeric state of 13.2 subunits, indicating a 13-mer assembly. However, the oligomeric state of the gp6 co-expressed with gp11 was shown to be 13 subunits by negative staining electron microscopy and by crystallographic analysis. These observations demonstrate that the SEC-MALLS data could be misleading in determining the oligomeric state of proteins containing more than ~10 subunits. This is due to the limitation on the accuracy of a single SEC/LS experiment, which is ±5% for protein standards ranging from 12 to 475 kDa (Folta-Stogniew & Williams, 1999). Compared to SEC-MALLS, analysis of oligomers by negative staining electron microscopy can provide more reliable data for characterisation of large assemblies like gp6.

The 2.8 Å resolution structure of the wild type gp6 Δ27-472 shows that the architecture of the 13-fold symmetrical molecule is the same as observed

previously in the 3.4 Å resolution structure of the N365K mutant protein (PDB entry: 2JES). The structure reveals differences in interaction between tunnel loops of adjacent subunits, notably differences in hydrogen bonding interactions, which are likely to be important for normal functioning of the motor. Residues from 170 to 238, poorly defined in the structure of the mutant gp6, are resolved in the current structure.

Research described in this thesis used truncated gp6 protein lacking 31 residues at the C-terminus. Such truncation has led to better diffracting crystals of the portal protein. The C-terminal segment is negatively charged due to the presence of stretches of Asp and Glu residues. Since the C-terminus is exposed towards the interior of the capsid, this negatively charged segment could be important for the interaction with gp11. Future work should be directed towards co-expressing gp6 containing a complete C-terminus with the full-length gp11 and characterising the oligomeric state of gp6 by negative staining electron microscopy. Crystallographic analysis would be carried out if the natural 12-mer assembly of portal protein, observed in mature viral particles, could be obtained.

5 The SPP1 Scaffolding Protein

5.1 Introduction

5.1.1 Scaffolding Proteins from dsDNA Bacteriophages

The procapsid is a transient precursor structure composed of the portal protein, coat proteins and scaffolding proteins present in double-stranded DNA bacteriophages and herpes viruses. Some viruses require both internal scaffolding proteins and external scaffolding proteins to accomplish procapsid assembly, and others only require internal scaffolding proteins. Scaffolding proteins play a catalytic role in the polymerisation of coat proteins with correct geometry. In several viral systems, a mixture of purified scaffolding and coat proteins can yield virus-like particles *in vitro* (Cerritelli & Studier, 1996; Dokland et al, 2002; Fu et al, 2007; Lee & Guo, 1995; Newcomb et al, 1999; Prevelige et al, 1988; Wang et al, 2000). Absence of scaffolding proteins will result in an aberrant procapsid structure and thus a non-infectious viral particle. Scaffolding proteins act as chaperones to prevent improper interactions and ensure the fidelity of the assembly process (Morais et al, 2004), but are subsequently removed during DNA encapsidation and are absent from the mature virion (Chang et al, 2008; Ziegelhoffer et al, 1992).

One of the most well characterised scaffolding proteins is from *Salmonella typhimurium* Bacteriophage P22, a member of the *Podoviridae* family. P22 scaffolding protein is found to initiate shell assembly and regulate the polymerisation of coat subunits into icosahedral procapsids *in vitro* (Prevelige et al, 1988). There are approximately 250 molecules of internal scaffolding proteins located in the interior of the procapsid. Truncation mutants designed to define the regions of scaffolding protein responsible for the different aspects of its function (Weigele et al, 2005) revealed that amino acids 1-20 are nonessential. The scaffolding protein interacts with the coat protein via the C-terminal coat protein-binding domain (residues 238-303), with residues 280-294 constituting the minimal coat protein-binding site. The observations indicated that the N-terminal 57 residues are important for sensing the scaffolding protein release

signal to allow the scaffolding protein to exit the procapsid. Several residues between 229 and 238 are required for portal protein recruitment.

Structural characteristics of the C-terminus from P22 scaffolding protein have been studied using circular dichroism, and nuclear magnetic resonance (NMR) (Parker et al, 1997; Sun et al, 2000). Stable secondary structure elements are present in the 163-amino acid carboxyl-terminal fragment (Parker et al, 1997), and a helix-loop-helix motif stabilized by a hydrophobic core is defined by NMR in the solution structure of the coat protein-binding domain (residues 238-303, Figure 5-1) (Sun et al, 2000). The association of scaffolding and coat proteins is mediated mainly by ionic interactions. Residues R293 and K296 are particularly important for coat protein binding (Cortines et al, 2011; Padilla-Meier et al, 2012). There is limited knowledge of the structure of the P22 scaffolding protein beyond the NMR structure of the extreme C-terminus coat protein-binding domain. It is suggested that the full-length protein structure is predominantly composed of two α-helical elongated domains connected by unstructured regions (Tuma et al, 1998). A conformational change upon assembly occurs, because the N-terminus and the C-terminus are proximate in solution, however, the N-terminus is no longer accessible to the C-terminus when assembled into procapsids (Padilla-Meier & Teschke, 2011).

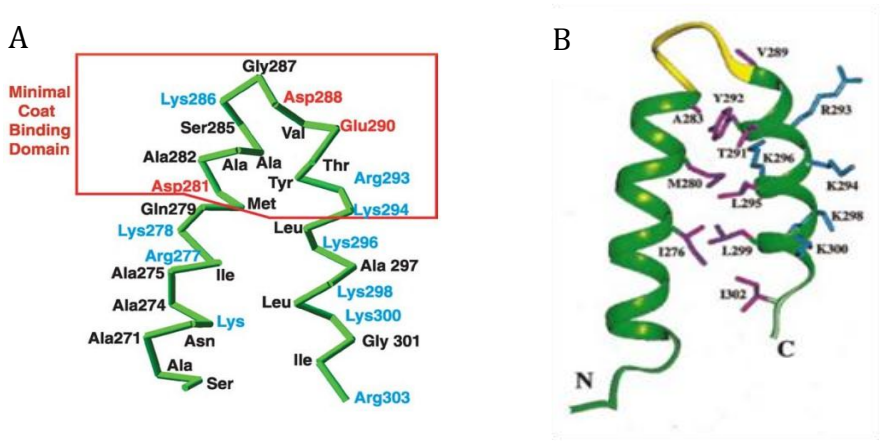


Figure 5-1 The structure of the coat protein-binding domain of the P22 scaffolding protein.
(A) The minimal coat binding domain is highlighted; (B) The hydrophobic residues involved in the hydrophobic core are illustrated in purple and basic residues in blue.

The structure was determined by NMR, and the figure was adapted from (Sun et al, 2000).

The reconstructed P22 procapsid at 8.7Å resolution by cryo-EM (Chen et al, 2011b) showed a set of scaffolding subunits interacting with both the portal and coat subunits (Figure 5-2). The "loop" in the helix-loop-helix motif of the scaffolding protein is in close contact with the wing domain of the portal protein indicating an extensive interaction, and the number of scaffolding protein interacting with the portal is suggested to be 10 (Figure 5-2).

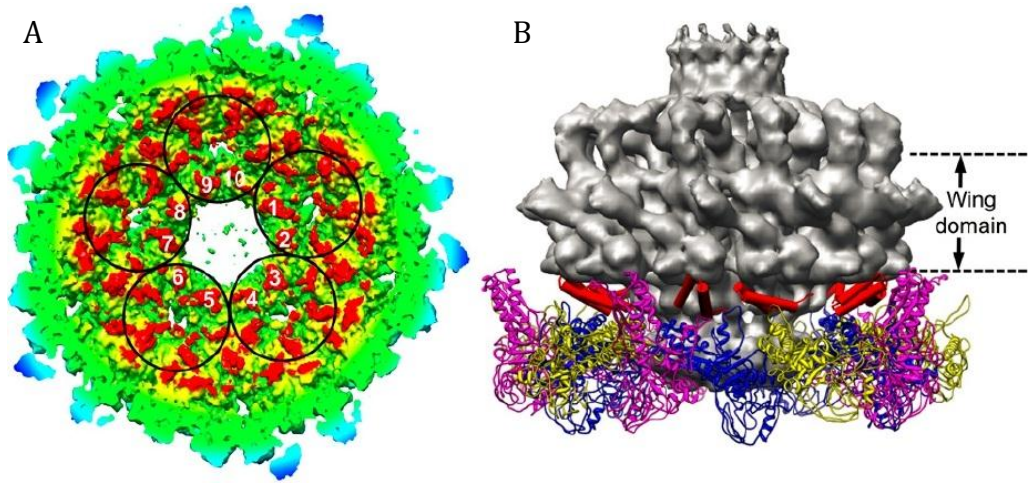


Figure 5-2 Cryo-EM reconstruction of the P22 procapsid.
(A) The C termini of 10 scaffolding proteins are labelled from 1 to 10, from five hexamers (circled) surrounding the portal vertex. (B) Side view of the interaction model among the 10 scaffolding protein C termini (red cylinders), coat proteins (ribbons), and the 12-fold averaged portal density (grey). The figure was adapted from (Chen et al, 2011b).

Scaffolding proteins have been successfully purified from other bacteriophages such as bacteriophage λ (Ziegelhoffer et al, 1992), SPP1 (Poh et al, 2008), and phi29. Biophysical data indicate that scaffolding proteins possess an unusually elongated shape (Ziegelhoffer et al, 1992). Currently, the only available x-ray structures of scaffolding proteins are from phi29, from both before and after prohead assembly (Figure 5-3). The structures revealed that phi29 scaffolding protein is a homodimer that resembles an arrow with a four helix bundle composing the arrowhead and a coiled coil forming the tail (Morais et al, 2003).

The α-helical structure of the scaffolding protein is shown to be a conserved characteristic, which assists the association of the major capsid protein (Morais et al, 2003; Sun et al, 2000).

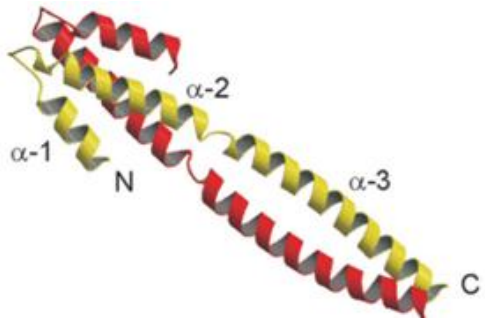


Figure 5-3 Ribbon diagram of the structure of the phi29 scaffolding protein.
The figure was adapted from (Morais et al, 2003).

5.1.2 The SPP1 Scaffolding Protein

The SPP1 scaffolding protein is gp11, with the full-length protein being 214 amino acids in size and having a subunit molecular mass of 23.5 kDa. A study on head morphogenesis genes discovered the importance of gp11 in the formation of normal proheads by the major capsid protein gp13 (Becker et al, 1997). Gp11 directs the polymerisation of major capsid protein subunits to form the required correct geometry for the icosahedral procapsid structure (Becker et al, 1997). Gp11 only interacts with gp13 when they are co-produced where they can yield procapsid-like structures, whilst no interaction was detected when synthesised separately and mixed *in vitro* (Droge et al, 2000). When gp11 is co-expressed with gp6 and gp13, biologically active procapsids competent for DNA packaging *in vitro* can be formed. The stable interaction between gp6 and the two major procapsid proteins gp13 and gp11 was detected only when the three proteins were co-produced (Droge et al, 2000).

As with internal scaffolding proteins from other phages, the SPP1 scaffolding protein is also predicted to be an α-helix rich molecule with a very elongated shape. The organisation of purified oligomeric gp11-His6 was characterised to be a tetramer in solution. The MALDI mass spectra of gp11-His6 cross-linked with 0.25%

glutaraldehyde implied the tetramer is a dimer formed from gp11-His6 dimers (Poh et al, 2008). There is very low amino acid sequence similarity shown among the scaffolding proteins from the bacteriophage SPP1, P22 and phi29.

```
SPP1            ----------------------------------------------------MSLK-
phi29           ----------------------------------------------------MPLKP
P22             MEPTTEIQATEDLTLSGDHAAASADSLVVDNANDNAGQEEGFEIVLKDDETAPKQDPAKN
                                                                    . *

SPP1            --------EQLGEELYGQVLAKLGEGAKLVDIS----------DGSFIPKEKFDAVNSEK
phi29           EEH-----EDILNKLLDPELAQSERTEALQQLRVN--------YGSFVSE----------
P22             AEFARRRIERKRQRELEQQMEAVKRGELPESLRVNPDLPPQPDINAYLSEEGLAKYDYDN
                        *   :      :          .:            .::::.:

SPP1            KS------------LEQQLTDRDQQLQELSTKATGHDELSAKIADLQKANEEAKQAFE--
phi29           ----------------------------------------YNDLTKSH----EKLA--
P22             SRALAAFNAANTEWLMKAQDARSNAVAEQGRKTQEFTQQSAQYVEAARKHYDAAEKLNIP
                                                         :  . :    : :

SPP1            --AEKQQLKYEHALETALRDSGAKNPKAVKALLDTESIKLDGDKLLGFEDQIKALKEQED
phi29           --AEKDDLIV--------------------------SNSKLFRQIGLTDKQEEDHKKAD
P22             DYQEKEDAFMQLVPPAVGADIMRLFPEKSAALMYHLGANPEKARQLLAMDGQSALIELTR
                   **::                                : .  . :   *  .   :

SPP1            YLFKGT-EPNGGVQGTPPPGKGADLGGLPTKKNPFKQGPDFNLTEQGILFRENPELAKKL
phi29           ISETITIE------------------------------------------------DL
P22             LSERLTLKPRGKQISSAPPADQPITGDVSAANK-----DAIRKQMDAAASKGDVETYRKL
                     * :                                                .*

SPP1            QAEAQ---
phi29           EAK-----
P22             KAKLKGIR
                :*:
```

Figure 5-4 Multiple sequence alignment of scaffolding proteins from the bacteriophage SPP1, P22 and phi29. *The alignment was performed using ClustalW.*

5.2 Methods and Materials

5.2.1 Cloning of Gp11 Constructs

The SPP1 genomic DNA was served as the template for the PCR reaction. The primers were designed to introduce specific endonuclease restriction sites and amplify the required DNA fragments (Table 5-3) and synthesised by Eurofins MWG Operon, Ebersberg, Germany (purification level: salt free). Amplification of the required ORF by PCR was performed in a 50 μL reaction vial with the components listed in Table 5-1 under thermocycling conditions shown in Table 5-2. The insert was cloned into vectors pET28a, pCDFDuet-1, pGEX6P-3, Lic (-), Lic (+), and pET22b for different purposes. The correct inserts were verified by DNA sequencing (GATC Biotech) and alignment analysis (Clustal W).

Table 5-1 PCR reaction components

Component	Amount in 50 µL reaction	Final concentration
F Primer (2 µM)	12.5 µL	0.5 µM
R Primer (2 µM)	12.5 µL	0.5 µM
5X Phusion HF Buffer	10 µL	1X
dNTPs (25 mM each)	0.4 µL	0.2 mM
Phusion® DNA Polymerase	0.5 µL	1.0 unit
Template DNA	0.5 µL	< 250 ng
dH_2O	Up to 50 µL	

Table 5-2 Thermocycling conditions for a routine PCR

Step	Temperature	Time
Initial Denaturation	98 °C	2 minutes
30 Cycles	98 °C 55 °C	30 seconds 30 seconds
Final Extension	72 °C	5 minutes
Hold	4 °C	

Table 5-3 Primers for the cloning of gp11 constructs

Primer	Sequence (5'-3')
GP11-F1	GGG AAT TCC ATA TGA TGA GTT TGA AAG AGC AG
GP11-R1	GAG CCG CTC GAG TTA CTG TGC TTC TGC TTG TAG
GP11_F2_BamH1	CGG GAT CCA TGA GTT TGA AAG AGC AGT TAG
GP11-R2	GAG GAG AAG GCG CGT TAC TGT GCT TCT GCT TGT AGC
GP11-F3	CAC CAC CAC CAC CCG AAA GAA AAG TTT GAC GCT GTT
GP11-F3_3c	CCA GGG ACC AGC ACC GAA AGA AAA GTT TGA CGC TGT T
GP11-F4	CAC CAC CAC CAC TTT GAC GCT GTT AAC AGT GAG AAG
GP11-F4_3c	CCA GGG ACC AGC ATT TGA CGC TGT TAA CAG TGA GAA G
GP11-F5	CAC CAC CAC CAC GGC CCG GAT TTC AAC CTC A
GP11-F5_3C	CCA GGG ACC AGC AGG CCC GGA TTT CAA CCT CA
GP11-R3	GAG GAG AAG GCG CGT TAA CCT TTA AAG AGA TAG TCC TC
Gp11_infusion_F1	AAG GAG ATA TAC ATA TGA TGA GTT TGA AAG AGC AGT TAG G
Gp11_infusion_R1	GTG CGG CCG CAA GCT TTT ATT TAA AGA GAT AGT CCT CTT GT
Gp11_infusion_F2	ATC ACC ACC ACC ACA TGA GTT TGA AAG AGC AGT TAG G
Gp11_infusion_R2	TGA GGA GAA GGC GCG TTA TTT AAA GAG ATA GTC CTC TTG T

5.2.2 Expression and Purification of Gp11

The plasmids were transformed into *Escherichia coli* expression strain Rosetta2 cells. The plasmid pYM209 encoding the N-terminal GST tagged gp11 and the plasmids (pYM210, pYM212, pYM222, pYM223, and pYM224) encoding the truncated gp11 were transformed into *E. coli* expression strain B834. A single colony from a Luria–Bertani (LB) agar plate containing the appropriate antibiotics for the different vectors (e.g. 30 μg/mL kanamycin for pET28a vector and 50 μg/mL streptomycin for pCDFDuet-1 vector) was picked to inoculate a small overnight culture at 37 °C. 5 mL aliquots of the overnight culture were used to inoculate a 500 mL culture the following day. The 500 mL Culture were grown at 37 °C until the mid-log growth phase (OD600 of 0.6–0.8) was reached, at which point protein expression was induced by the addition of isopropyl-d-1-thiolgalactopyranoside (IPTG) to a final concentration of 1 mM. Lastly, overnight incubation of the culture at 16 °C and 180 r.p.m was carried out for the over-expression of proteins. Cells were recovered by centrifugation in SORRALL® RC 5B plus for 20 minutes at 5000 g, 4 °C. Supernatant was removed and the cell pellet was stored at -20 °C.

The cell pellet was resuspended in nickel affinity chromatography binding buffer, 20 mM Tris pH 7.5, 150 mM NaCl, 50 mM Imidazole, supplemented with a combination of 100 μg/mL lysozyme and protease inhibitors (1 mM AEBSF and 0.7 μg/mL pepstatin). Cells were disrupted by sonication with a large probe in glass beaker using short pulses of 30 seconds, with a two-minute resting time between pulses. The sonication was carried out on ice, with caution to minimise thermal damage to the protein extract. The cell debris was removed by centrifugation at 38758 g for 30 minutes at 4 °C using a Sorvall SS34 rotor. The supernatant was collected and then cleared with a 0.45 mm filter (Millipore). The filtrate loaded onto a 5 mL HiTrap column (GE Healthcare) equilibrated with binding buffer for nickel affinity chromatography purification.

The bound protein was eluted with an increasing proportion of elution buffer containing 20 mM Tris pH 7.5, 150 mM NaCl, and 500 mM imidazole. Fractions containing gp11 were pooled together. To obtain untagged gp11, a thrombin digestion step was performed to remove the histidine tag. Instead of a second

nickel affinity chromatography, MonoQ high-resolution ion-exchange chromatography was carried out to remove thrombin and the cleaved histidine tag, because the untagged gp11 can also bind to nickel. To ensure the complete removal of the histidine tag, an excess amount of thrombin (Amersham Biosciences) was used - 10 cleavage units of thrombin solution per mg for overnight digestion at 4 °C. The cleavage of the tag was confirmed by MALDI-MS. Once digestion was complete, the buffer was exchanged by extensive dialysis against the low ionic loading buffer 20 mM Tris pH 7.5, 50 mM NaCl for ion exchange chromatography.

Fractions containing gp11 from nickel affinity chromatography and buffer exchanged against the Ion exchange chromatography loading buffer were loaded onto an 8 ml column of Mono Q 10/100 GL (GE Healthcare) equilibrated with loading buffer 20 mM Tris pH 7.5, 50 mM NaCl, and washed with 10 column volumes of the loading buffer to wash out unbound proteins. The bound protein was eluted with an increasing proportion of elution buffer containing 20 mM Tris pH 7.5, 1 M NaCl. Fractions containing gp11 were pooled together, and concentrated using Vivascience 30 kDa molecular weight cut-off concentrators.

The concentrated sample (less the 10 mg/mL) was applied to a HiPrep 16/60 Sephacryl S-500 HR column (GE Healthcare) for size exclusion chromatography as the final polishing step to remove any high-molecular-weight aggregates and obtain the fusion proteins with the appropriate molecular weight. The proteins were eluted with GF buffer (20mM Tris pH 7.5, 150 mM NaCl), concentrated to at least 10 mg/mL and stored at -80 °C for crystallisation.

5.2.3 Size-Exclusion Chromatography coupled with Multi-Angle Laser Light Scattering (SEC–MALLS)

The molecular mass was determined by size-exclusion chromatography coupled with multi-angle laser light scattering (SEC–MALLS). The protein sample (60 µL) with a concentration of 0.5 mg/mL was applied on a BioSep SEC-s3000 gel filtration column (Phenomenex) equilibrated with buffer containing 20mM Tris pH 7.5, 150 mM NaCl. Size-exclusion chromatography was carried out on a Shimadzu HPLC system and the elution was monitored at 280 nm by an SPD20A UV/Vis

detector. Light-scattering data were recorded by a Dawn HELEOS-II 18-angle light-scattering detector and the concentration of the eluting protein was measured by an in-line Optilab rEX refractive-index monitor (Wyatt Technology). Data were analysed with the ASTRA V software package (Wyatt Technology). Molecular mass was calculated using Zimm's formalism of the Rayleigh–Debye–Gans light-scattering model for dilute polymer solutions and a refractive-index increment (dn/dc) of 0.183 mL/g was used for the protein molecular mass estimation.

5.2.4 Crystallisation

The sitting-drop vapour diffusion method was used for the initial crystallisation screen. Drops containing 150/300 nL protein solution and 150 nL reservoir solution were dispensed by a Mosquito Nanolitre Pipetting robot (TTP Lab-tech) in 96-well plates, and equilibrated against 54 µL of reservoir solution at 20 °C or 4 °C. The search for suitable crystallisation conditions was performed using several commercial screens such as Index, PACT, Clear Strategy Screens I and II, Morpheus, JCSG, Ammonium sulphate and MPD. Conditions from Morpheus™ screen (Molecular Dimensions) growing small crystals were optimised in 24-well hanging-drop plates with manual pipetting. Crystals were tested using a Rigaku RU-H3R rotating-anode X-ray generator equipped with Osmic multilayer optics and a MAR Research MAR345 imaging-plate detector.

5.2.5 Limited Proteolysis

Chymotrypsin concentration used in proteolysis studies was 50 ng/µL, and the protein concentration used was 2 mg/mL. The total volume of the reaction was 10 µL, and 20mM Tris pH 7.5, 150mM NaCl was the buffer for the reaction. Untagged full-length gp11 (after tag removal from YM184) and chymotrypsin were mixed in ratios of 1:200, 1:100 and 1:50, and incubated for one hour at room temperature. The digestion of gp11 by chymotrypsin was monitored on 15% SDS – PAGE. The reaction was quenched by the addition of PMSF to the final concentration of 2 mM before the sample was sent for analysis by electrospray ionisation mass spectrometry.

Table 5-4 Conditions for chymotrypsin digestion of full length gp11

Reaction	YM184 (2 mg/ml)	Chymotrypsin (50 ng/µL)	Buffer	Protease: protein ratio	Time
1	5 µL	1 µL	4 µL	1:200	1hr
2	5 µL	2 µL	3 µL	1:100	1hr
3	5 µL	4 µL	1 µL	1:50	1hr
	YM184 (2 mg/ml)	Chymotrypsin (50 ng/µL)	Buffer	Protease: protein ratio	Time
4	5 µL	1 µL	4 µL	1:200	45 min
5	5 µL	2 µL	3 µL	1:100	45 min
6	5 µL	4 µL	1 µL	1:50	45 min
7- Control	5 µL	0 µL	5 µL	NA	NA

5.3 Results

5.3.1 Cloning, Expression and Purification of the Full-Length Gp11

The gene encoding full-length gp11 was cloned into the restriction sites *NdeI* and *XhoI* of vector pET28a to form plasmid pYM184 for the expression of N-terminal hexahistidine-tagged protein, and cloned into the restriction sites *NdeI* and *XhoI* of vector pCDFDuet-1 to form plasmid pYM185 for the expression of untagged protein. Also, the full length gp11 encoding DNA sequence was cloned and ligated into the *BamHI* and *XhoI* sites of vector pGEX6P-3 to form plasmid pYM209 for the expression of the protein with an N-terminal glutathione S-transferase tag.

The full-length gp11 with an N-terminal hexahistidine tag or N-terminal glutathione S-transferase tag were both successfully overexpressed in *E. coli* in soluble form. Pure and soluble His-tagged gp11 protein was obtained by a three-step purification procedure consisting of nickel affinity chromatography, MonoQ high-resolution ion-exchange chromatography, and size exclusion chromatography. The yield of the His-tagged gp11 from a 500 mL culture was up to ~75 mg with a concentration of 30 mg/mL in 20 mM Tris pH 7.5, 50 mM NaCl. An extra thrombin digestion step with an excess amount of thrombin and overnight digestion was used to completely remove the histidine tag. GST-tagged gp11 was purified by batch method with Glutathione Sepharose™ 4B, and then polished by gel filtration using a Superose® 6 HR 10/30 column. After the two-step purification the protein was not pure enough for further biochemical analysis, with several impurities still present (Figure 5-5A).

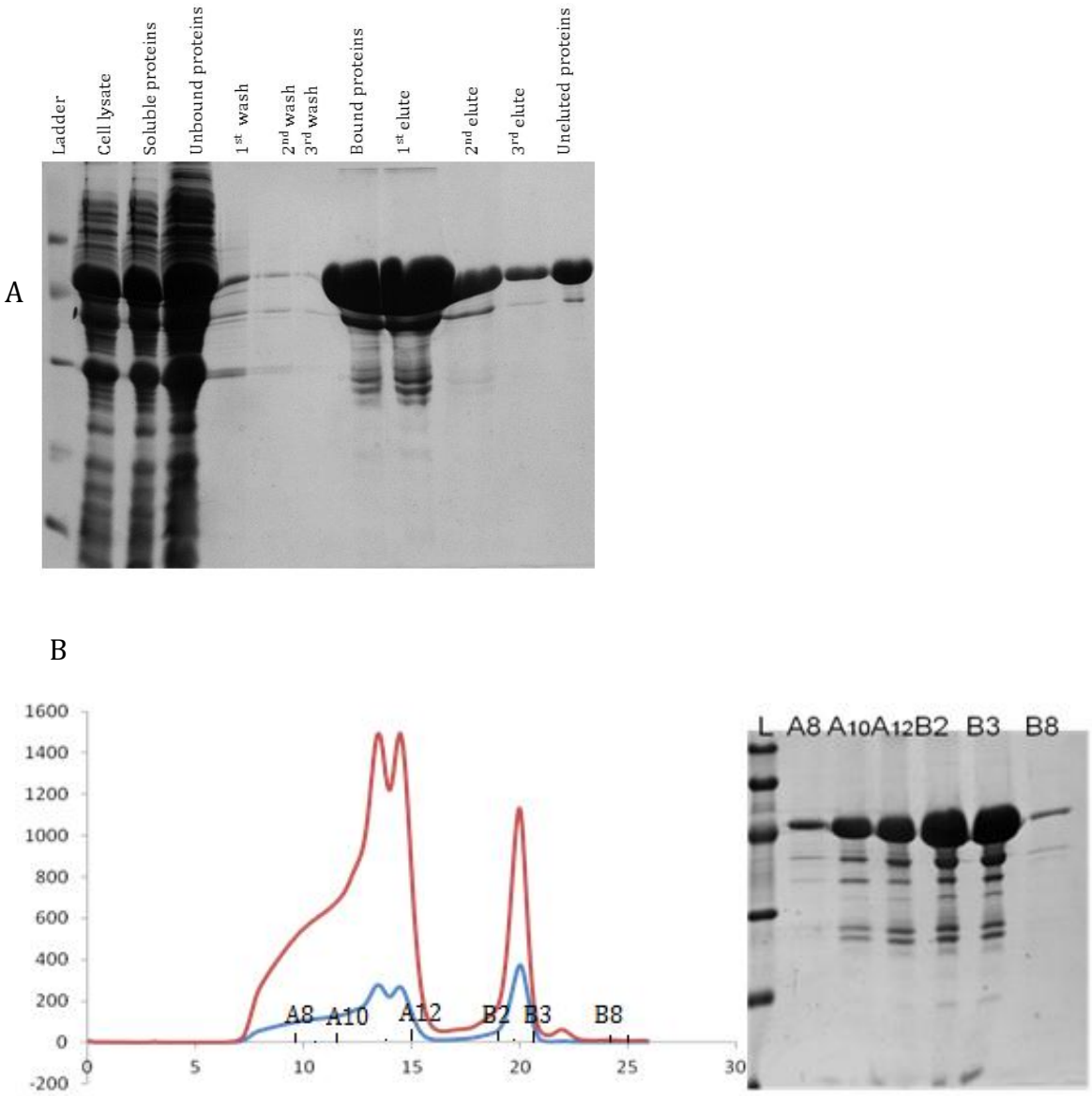


Figure 5-5 Purification of the GST-tagged full-length gp11.

(A) The purification procedure of the GST-tagged gp11 by batch method using Glutathione Sepharose 4B beads was monitored on a 15% SDS PAGE gel, and the protein samples of cell lysate, soluble proteins, unbound proteins, bound proteins and eluted proteins were shown respectively; (B) The purification of the GST-tagged gp11 by gel filtration using a Superose 6 HR 10/30. The elution profile was shown on the left panel with the red line indicating A260 abosorbance and the blue line indicating A280 absorbance. The corresponding fractions A8-B8 were analysed by 15% SDS PAGE gel as shown on the right panel.

5.3.2 Characterisation of the Full Length Gp11 by MALDI-MS and SEC-MALLS

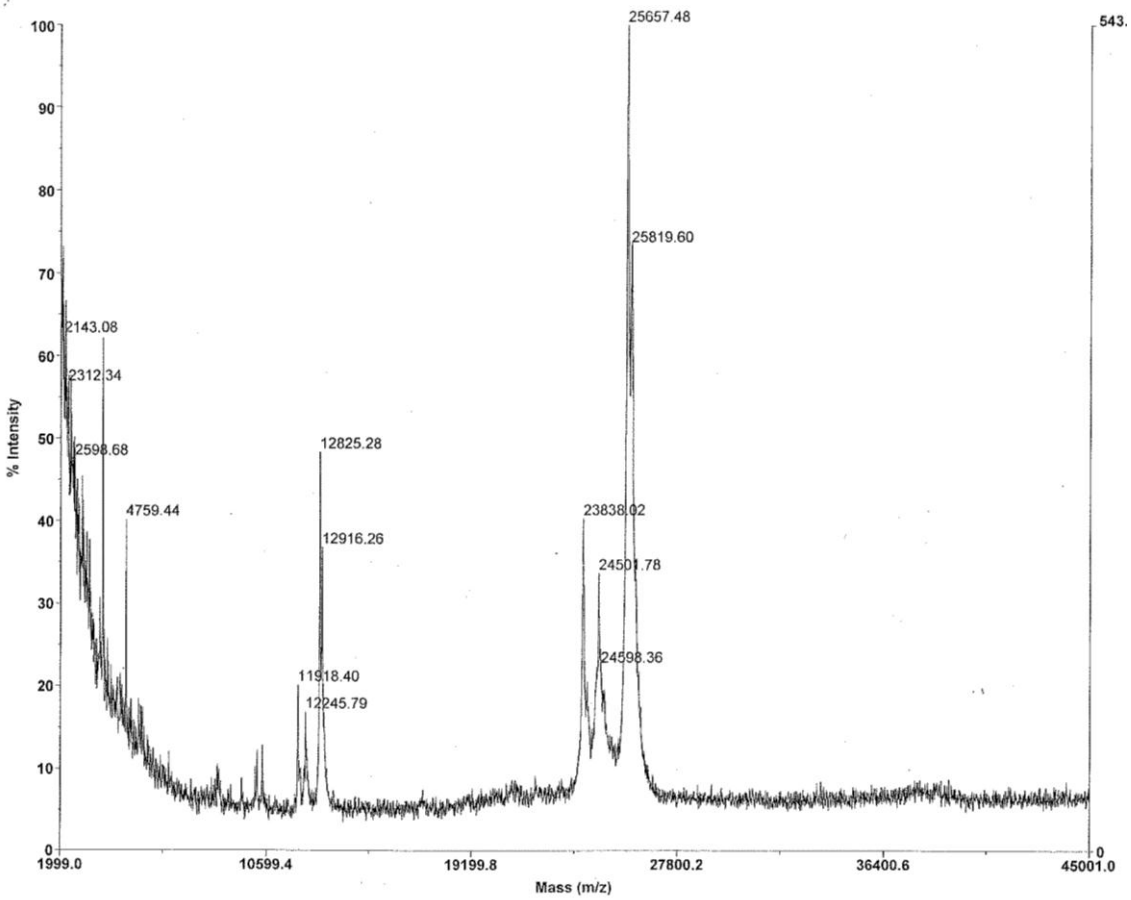
The hexahistidine-tagged gp11 consists of 234 amino acids including the 16-amino acid long thrombin cleavable histidine tag (HHHHHHSSGLVPRGSH) with a theoretical molecular mass of 25626.6 Da. The purified protein was analysed by MALDI-MS for the rapid identification and characterisation of the subunit molecular weight. The molecular weight detected by MALDI-MS was 25657.48 Da, which was in good agreement with the theoretical molecular weight (Figure 5-6 A).

A single monodispersed peak was observed during the size-exclusion chromatography, although the elution profile was broad, possibly due to the elongated shape of the scaffolding protein. SEC-MALLS data showed the mean molecular weight of the eluted species as 53.5 kDa, or ~2.1 subunits per oligomer, suggesting that gp11 is a dimer in solution (Figure 5-6 B).

5.3.3 Gradual Degradation and Limited Proteolysis of Full-Length Gp11

The full-length gp11 degraded gradually at 4 °C with a series of protein fragments detected on 15% SDS PAGE gel (Figure 5-7 A). A limited proteolysis experiment was performed to probe the stable truncated gp11. Gp11 was fully digested by chymotrypsin after 30 minutes at room temperature. The protein fragment after chymotrypsin digestion was subject to electrospray ionisation mass spectrometry, and the major peak detected corresponded to a molecular mass of 17.48 kDa (Figure 5-7 B). The potential digestion sites of gp11 by chymotrypsin were predicted using the ExPASy PeptideCutter tool (Wilkins et al, 1999). Chymotrypsin preferentially cleaves peptide amide bonds where the carboxyl side of the amide bond (the P1 position) is a tyrosine, tryptophan, or phenylalanine. According to the online tool and the molecular mass of the digested fragment determined by ESI-MS, three possible sets of digestion sites were predicted 40-199, 64-214, and 1-153.

A



B

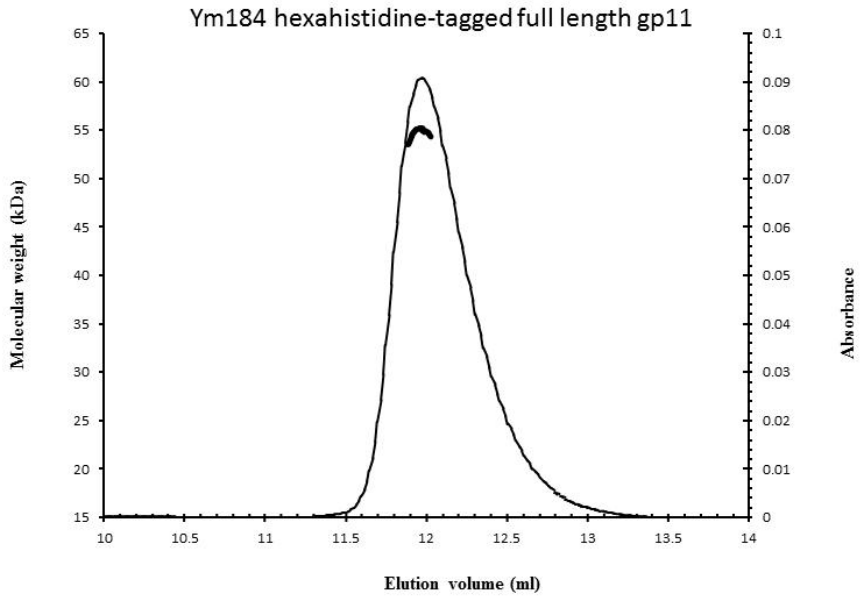
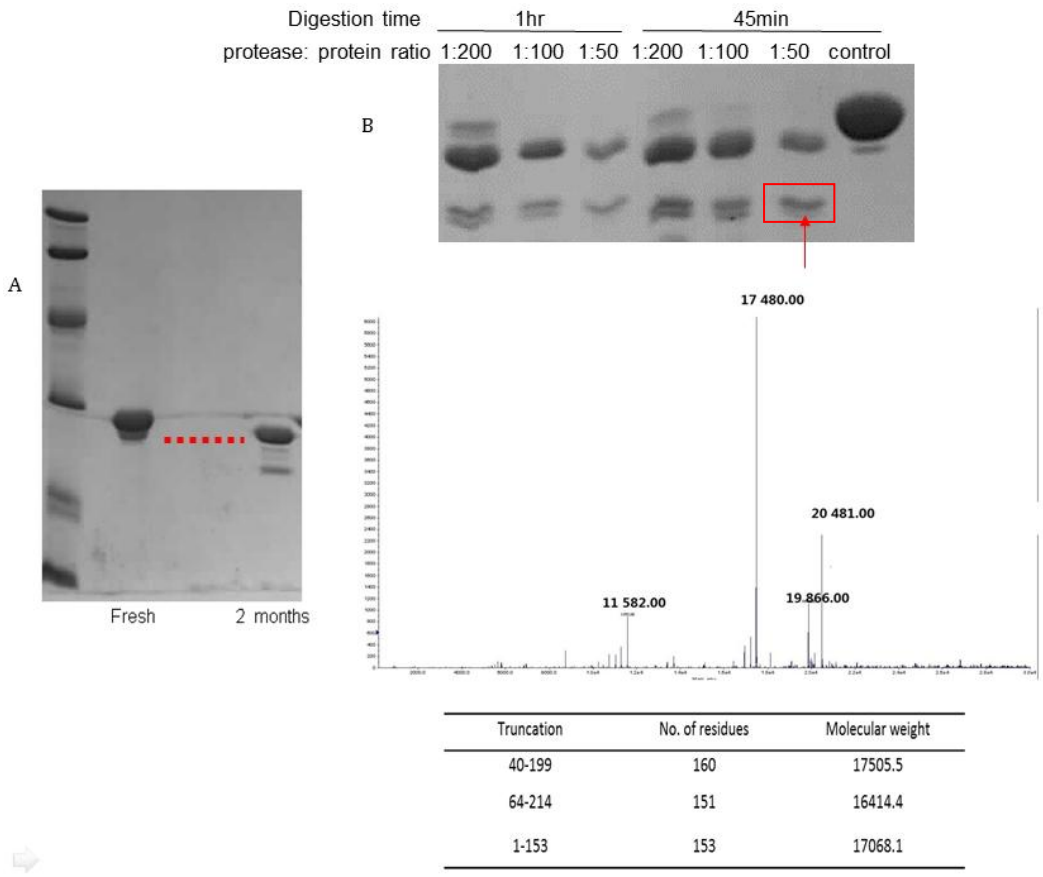


Figure 5-6 Analysis of His-tagged full-length gp11 by MALDI-MS (A) and SEC-MALLS (B).



Truncation	No. of residues	Molecular weight
40-199	160	17505.5
64-214	151	16414.4
1-153	153	17068.1

Figure 5-7 Limited proteolysis analysis of full-length gp11.
(A) Gradual degradation of the full-length gp11 was shown on a 15% SDS-PAGE gel; (B) The full length gp11 was digested with different protein: chymotrypsin ratios and digestion times, and the samples after digestion were analysed on a 15% SDS-PAGE gel shown on the top panel. The sample after digestion was also analysed by ESI-MS shown on the bottom panel, and the ESI-MS spectrum showed the major peak with a molecular weight of 17.48 kDa corresponding to the digested gp11 fragment that was highlighted on the top panel.

5.3.4 Design, Cloning, Expression and Purification of the Truncated Gp11

Truncated constructs were designed based on the information from chymotrypsin digestion and disorder prediction, which suggested the C terminus starting from amino acid 155 was extremely disordered (Figure 5-8, Table 5-5). The corresponding DNA sequences were cloned and ligated into the vectors Lic (-), Lic (+), and pET22b to express the protein with: a non-cleavable N-terminal his tag, a cleavable N-terminal his tag and without any tag, respectively.

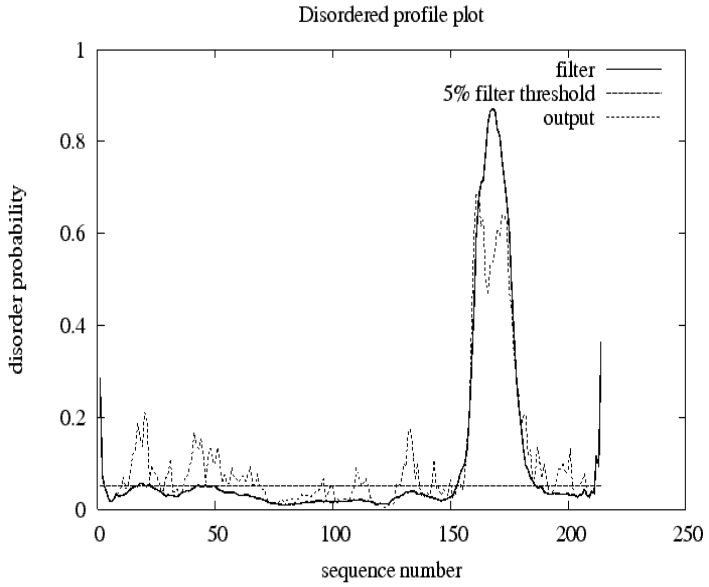


Figure 5-8 Disorder in the C- terminal segment of gp11 starting from amino acid 155.
The prediction was performed by PrDOS (Ishida & Kinoshita, 2007).

Table 5-5 List of truncated constructs of gp11

plasmid	Constructs	Vector	Tag
pYM210	39-214	Lic(-)	N-His6
pYM212	39-155	Lic(-)	N-His6
pYM222	35-214	Lic(+)	N-His6
pYM223	39-155	Lic(+)	N-His6
pYM224	1-154	pET22b	No tag

All the truncated gp11 proteins were produced in soluble form. Purification of the His-tagged protein was carried out by nickel affinity chromatography, MonoQ high-resolution ion-exchange chromatography, and lastly size exclusion chromatography. For the untagged protein, a Q FF column was used for the first step. Although it was possible to purify the truncated proteins, they proved to be less soluble and stable in comparison with the full-length gp11 and precipitated at protein concentrations higher than 10 mg/mL.

5.3.5 Crystallisation of the Full-Length Gp11

Among all the gp11 construct variants, the full-length gp11 exhibited higher crystallisation propensity than the constructs with truncated N- and C-termini. The hexahistidine-tagged gp11 at a concentration of 10 mg/mL in 20 mM Tris pH 7.5, 50 mM NaCl was crystallised under reservoir conditions of 10% PEG 4000, 20% glycerol, 0.03 M of each divalent cation ($CaCl_2$ and $MgCl_2$), 0.1 M bicine/Trizma base pH 8.5 (condition A11 of Morpheus™ screen, Figure 5-9A). Optimisation based on this condition with variation of the PEG concentration generated more crystals of larger size - 50µm×45µm×35µm - that were checked for diffraction using the in-house Rigaku X-ray generator with Mar345 detector (Figure 5-9B). However, there was no diffraction observed from the optimised crystal. When the histidine tag was cleaved, the full length gp11 at a concentration of 14 mg/mL formed crystals under conditions of 10% PEG 4000, 20% glycerol, 0.03 M of each divalent cation ($CaCl_2$ and $MgCl_2$), 0.1 M bicine/Trizma base pH 8.5 (A11 of Morpheus™ screen, Figure 5-10 A), 10% PEG 8000, 20% ethylene glycol, 0.02 M of each alcohol, 0.1 M MOPS/HEPES-Na pH 7.5 (D6 of Morpheus™ screen, Figure 5-10 B), and 10% PEG 8000, 20% ethylene glycol, 0.03 M of each ethylene glycol, 0.1 M MOPS/HEPES-Na pH 7.5 (E6 of Morpheus™ screen, Figure 5-10 C). Further optimisation of the untagged gp11 based on these conditions also failed to produce better diffracting crystals.

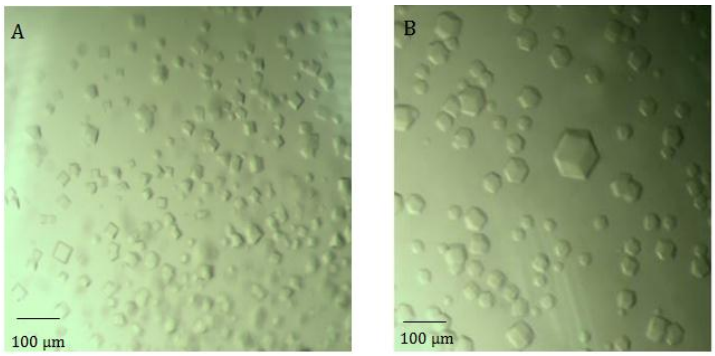


Figure 5-9 Crystals of his-tagged full-length gp11.
Crystals obtained using MorpheusTM screen condition A11 before (A) and after (B) optimisation.

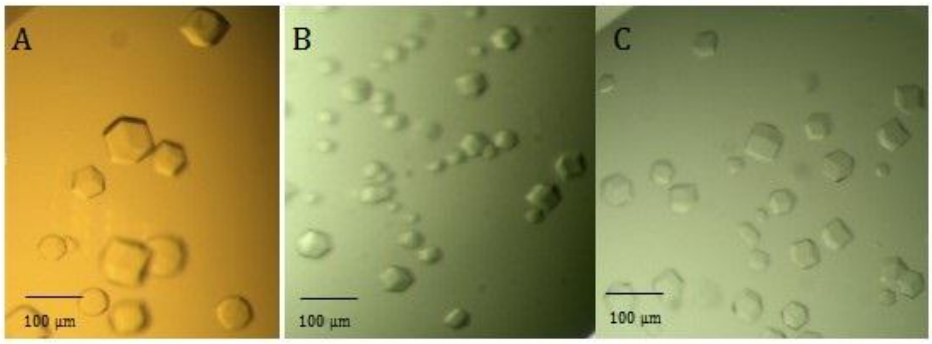


Figure 5-10 Crystals of untagged full-length gp11.
Initial hits from Morpheus™ screen conditions A11 (A), D6 (B) and E6 (C).

5.4 Discussion

The full-length scaffolding protein gp11 from bacteriophage SPP1, with and without an N-terminal histidine tag, was successfully cloned, purified and crystallised. The GST-tagged version of the full-length gp11 was also cloned, but purification was not satisfactory, with considerable amounts of impurities presented after a two-step purification by batch method with Glutathione Sepharose™ 4B beads followed by size exclusion chromatography. The oligomeric state of the histidine tagged scaffolding protein in solution was determined to be dimeric. This result contradicts previous observations on the SPP1 gp11 that showed that a C-terminally histidine tagged protein formed tetramers in solution: a dimer of dimers was observed by sedimentation equilibrium centrifugation and MALDI mass spectrometry of cross-linked protein samples (Poh et al, 2008).

The full-length gp11 protein was observed to degrade gradually and therefore a series of truncations were designed aiming to improve stability based on the

results of limited proteolysis experiments and disorder prediction analysis. The truncated proteins were also successfully cloned and purified. Surprisingly the full-length gp11 exhibited higher crystallisation propensity than the N- and C-terminally truncated constructs, which were initially expected to feature less degradation and be more suitable for production of diffracting crystals. The N- and C-terminally truncated proteins showed less solubility and stability compared with the full-length gp11 during the purification process, and formed amorphous precipitates in the majority of drops during crystallisation, with no crystal formation.

6 Recombinant Portal Protein from *Staphylococcus epidermidis* Bacteriophage CNPH82 Is A 13-Subunit Oligomer

6.1 Introduction

CNPH82 is a bacteriophage infecting the opportunistic pathogen *Staphylococcus epidermidis. S. epidermidis* is normally a human skin commensal bacterium but turns into a very common nosocomial infection pathogen in immunocompromised patients with implanted medical devices (Otto, 2009; Ziebuhr et al, 2006). The therapeutic challenge of treating the *S. epidermidis* infections originates from its rapid development of antibiotic resistance and formation of biofilms (Otto, 2009). Upcoming multiresistency to several *S. epidermidis* strains was connected to horizontal gene transfer, which is commonly mediated by bacteriophages. One of those phages is CNPH82, a member of the *Siphoviridae* family and the *Caudovirales* order.

Transmission electron microscopy micrographs showed that CNPH82 contains an isometric head and noncontractile tail (Figure 6-1) (Daniel et al, 2007). The complete genome of CNPH82 has been sequenced (Daniel et al, 2007). However, unlike the well characterised double-stranded DNA bacteriophages such as T4, T7, P22 and SPP1, no X-ray structural information has yet been deduced for proteins of this essential pathogen related phage.

The portal protein serves as a major component of the ATP-dependent genome translocation molecular motor in tailed bacteriophages and herpes viruses (Casjens, 2011). As an essential requirement during viral morphogenesis process, the portal protein plays indispensable roles in several aspects: it initiates pro-capsid assembly, and is a central component of the DNA translocation molecular motor, headful sensor and connector assembly (Rao & Feiss, 2008). The portal proteins from different tailed bacteriophages and herpes simplex viruses vary dramatically in both amino acid sequence and molecular mass, but share a common characteristics: cyclical *homo-oligomers* arranged radially with a turbine-like shape and a central channel for DNA passage (Orlova et al, 1999; Rao & Feiss, 2008). In the functional mature virion or in the isolated connector bound

to tail factors, portal proteins were consistently presented as 12-subunit assemblies (Olia et al, 2011; Orlova et al, 2003; Simpson et al, 2000b). Nevertheless, the oligomeric state of portal proteins from some viruses, like SPP1 and herpes virus, could change to 13 when heterologously expressed in *E. coli*, possibly due to conformational rearrangements (Cardone et al, 2007; Lebedev et al, 2007; Orlova et al, 2003; Trus et al, 2004b). Each subunit of SPP1 portal protein consists of four regions – termed the clip, stem, wing and crown (Lebedev et al, 2007). CNPH82 portal protein shares 32% amino acid sequence identity with the SPP1 portal protein. Moreover, high sequence similarity between other head morphogenesis proteins such as the major capsid and scaffolding proteins of CNPH82 and SPP1, imply similar morphogenesis processes.

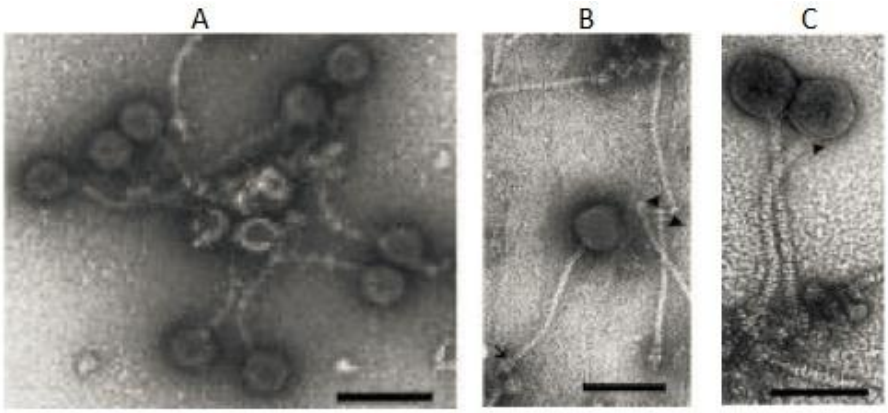


Figure 6-1 Ultrastructures of the phage CNPH82.
The structure was observed by transmission electron microscopy, and the figure was adapted from (Daniel et al, 2007).

6.2 Materials and Methods

6.2.1 Cloning, Expression and Purification of CNPH82 Portal Protein

The partial gene encoding truncated portal protein cn3 (E25-Q456) was amplified by PCR and ligated into the *NheI*/*HindIII* sites of vector pET28a (Novagen). Sequencing and alignment was performed to confirm the correct insert. The portal protein cn3 with a cleavable N-terminal hexahistidine tag was overexpressed in *Escherichia coli* strain B834 cells. Cells were grown in Luria–Bertani medium with 30 μg/ml kanamycin at 310 K to the mid-log phase (OD600 around 0.6-0.8). The portal protein expression was induced by the addition of 1mM IPTG carried out for

20 h at 289 K. The cell pellet was lysed using a cell disruptor (Constant Cell Disruption Systems) at 4° C with a pressure of 25 kpsi in lysis buffer containing 20 mM Tris (pH 7.5), 500 mM NaCl, 10 mM $MgCl_2$, 10 mM Imidazole, 100 µg/ml lysozyme, 1 mM AEBSF, 0.7 µg/ml pepstatin. Nickel Affinity chromatography was performed on a 5 mL HiTrapTM chelating HP column (GE Healthcare) and the protein sample was further purified on a Superose 6 size-exclusion column (GE Life Sciences). Purity was assigned by denaturing polyacrylamide gel-electrophoresis. The molecular mass of the purified sample was confirmed by matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS).

6.2.2 Molecular weight determination by SEC-MALLS

The molecular mass of cn3 (E25-Q456) was determined by Size Exclusion Chromatography coupled with Multi-Angle Laser Light Scattering (SEC-MALLS). The protein sample (60 µl) with a concentration of 0.5 mg/ml was applied on a BioSepTM SEC-s3000 Gel Filtration column (Phenomenex) equilibrated with buffer containing 20 mM Tris (pH 7.5), 500 mM NaCl, 10 mM $MgCl_2$. Size-exclusion chromatography was carried out on a Shimadzu HPLC system and the elution was monitored at 280 nm by a SPD20A UV/Vis detector. Light-scattering data were recorded by a Dawn HELEOS-II 18-angle light scattering detector and the concentration of the eluting protein was measured by an in-line Optilab rEX refractive index monitor (Wyatt Technology). Data were analysed with ASTRA V software package. Molecular mass was calculated using Zimm's formalism of the Rayleigh-Debye-Gans light scattering model for dilute polymer solutions and a refractive-index increment (dn/dc) of 0.183 ml g^{-1} was used for the protein molecular mass estimation.

6.2.3 Crystallisation

The protein cn3 (E25-Q456) was crystallized at 293 K by the sitting drop vapour diffusion method using 15 mg/ml protein solution in 20 mM Tris (pH 7.5), 500 mM NaCl, 10 mM $MgCl_2$. Drops containing 300 nl cn3 solution and 150 nl reservoir solution were dispensed by a Mosquito Nanolitre Pipetting robot (TTP Lab-tech) and equilibrated against 60ul of reservoir solution. To overcome the hurdle of high salt concentration in the protein solution, 500mM NaCl was added into the

reservoir solution after the screen was set up. The best crystal was obtained with the reservoir containing 0.2 M ammonium acetate and 40 % (v/v) MPD.

6.2.4 X-ray Data Collection and Processing

X-ray data were collected from a single crystal at the ESRF beam line ID14-4 at a wavelength of 0.9393Å with the crystal-to-detector distance of 652.7 mm. Data were collected at 100 K using an oscillation range of 0.5° per image with a total crystal rotation of 180°. Diffraction images were indexed and integrated using HKL-2000 (Otwinowski & Minor, 1997) and were further analysed with CCP4 program package (Winn et al, 2011). The self rotation function was calculated using MOLREP (Vagin & Teplyakov, 2010), in the resolution range 5 – 10 Å with the radius of integration sphere of 87 Å. To solve the structure by molecular replacement, BALBES (Long et al, 2008), MOLREP (Vagin & Teplyakov, 2010) and Phaser (Mccoy et al, 2007) were tried and SPP1 portal protein gp6 was used as a search model (PDB access code 2JES).

6.3 Results

6.3.1 Cloning, Expression and Purification

The portal protein was cloned and overexpressed in *E. coli* B834 cells. Homogeneous protein was obtained after Ni-affinity and size-exclusion chromatography. The protein was concentrated to ~35 mg ml^{-1} in solution containing 20 mM Tris (pH 7.5), 500 mM NaCl, 10 mM $MgCl_2$.

6.3.2 Oligomeric State of CNPH82 Portal Protein Cn3

The truncated CNPH82 portal protein cn3 (E25-Q456) consists of 432 amino acids with a theoretical molecular mass of 53.074 kDa. The molecular weight of the purified protein measured by matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) is 53.094 kDa, in good agreement with the theoretical value. A single monodispersed peak was observed during the size-exclusion chromatography of cn3. SEC-MALLS showed the mean molecular weight of the eluted species of 685.9 kDa, or ~12.9 subunits per oligomer, suggesting that cn3 contains 13-subunits per oligomer in solution (Figure 6-2).

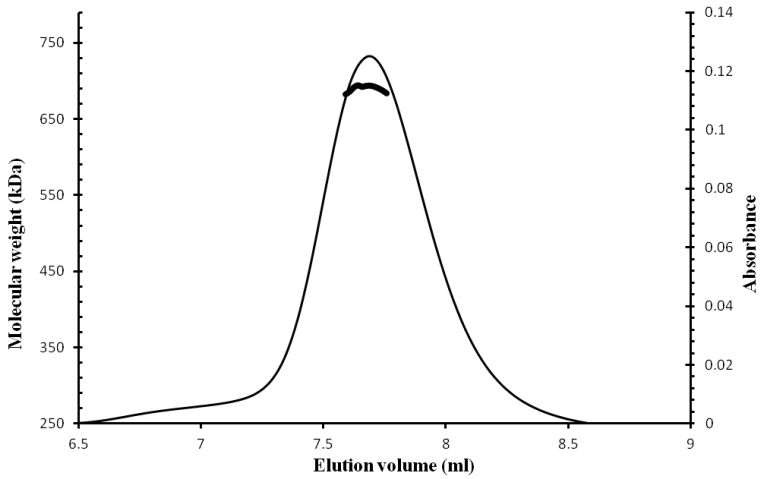


Figure 6-2 Characterization of cn3 oligomeric state by SEC–MALLS.
The thin line corresponds to the absorbance monitored at 280 nm. The thick line shows the molecular weight calculated for the eluted species.

6.3.3 Crystallisation

Several hits appeared in the initial MPD crystallisation screen (Hampton) with the best diffracting crystals growing from 40% MPD containing either 0.2 M ammonium nitrate or 0.2 M ammonium acetate. Both conditions were optimised. A complete native data set to a resolution of 4.2 Å was collected at ESRF using a crystal grown from 0.2 M ammonium acetate and 40 % (v/v) MPD (Figure 6-3).

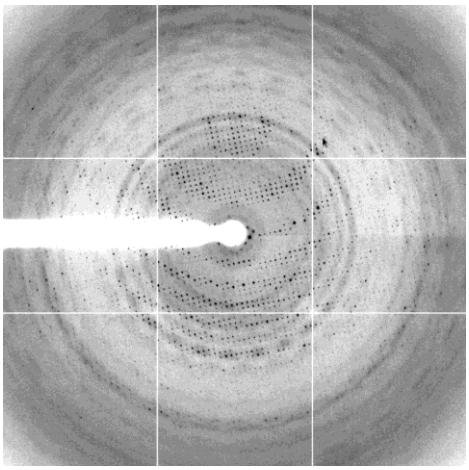


Figure 6-3 Diffraction image.
Resolution at the edge of the plate is 3.9 Å.

6.3.4 Crystallographic Analysis

The crystal belongs to the space group $C222_1$, with a = 252.4 Å, b = 367.0 Å, c = 175.5 Å (Table 6-1). The self-rotation function R (Φ,Ψ,Κ) (Crowther, 1972) was calculated to deduce the internal symmetry of the CNPH82 portal protein. The 13-fold symmetry was identified from peaks appearing in κ sections 360°/13 and κ = 180° (Figure 6-4). Peaks in the κ = 180° section were spaced from each other by 27.7° (Figure 6-4A). Although the sequence identity between cn3 and the portal protein, gp6, of SPP1 is as high as 32%, attempts to solve the structure by molecular replacement proved unsuccessful.

Table 6-1 X-ray data statistics

X-ray source	ID14-4, ESRF
Wavelength, Å	0.9393
Temperature, K	100
Space group	$C222_1$
Unit cell parameters, Å	a=252.4, b=367.0, c=175.5
Resolution Range, Å	100 - 4.2 (4.35 – 4.20)
No. of unique reflections	54776 (4970)
Rmerge[a], %	12.6 (65.4)
Completeness, %	98.1 (90.1)
Redundancy	3.6 (2.9)
Average I/sigma(I)	8.2 (1.4)

Values in parentheses are for the highest resolution shell.

[a] $R_{merge} = \sum_{hkl}\sum_i |I_i(h) - \langle I(h)\rangle| / \sum_{hkl}\sum_i I_i(h)$, where $I(h)$ is intensity of reflection h, $\langle I(h)\rangle$ is average value of intensity, the sum $\sum_{hkl}$ is over all measured reflections and the sum $\sum_i$ is over i measurements of a reflection.

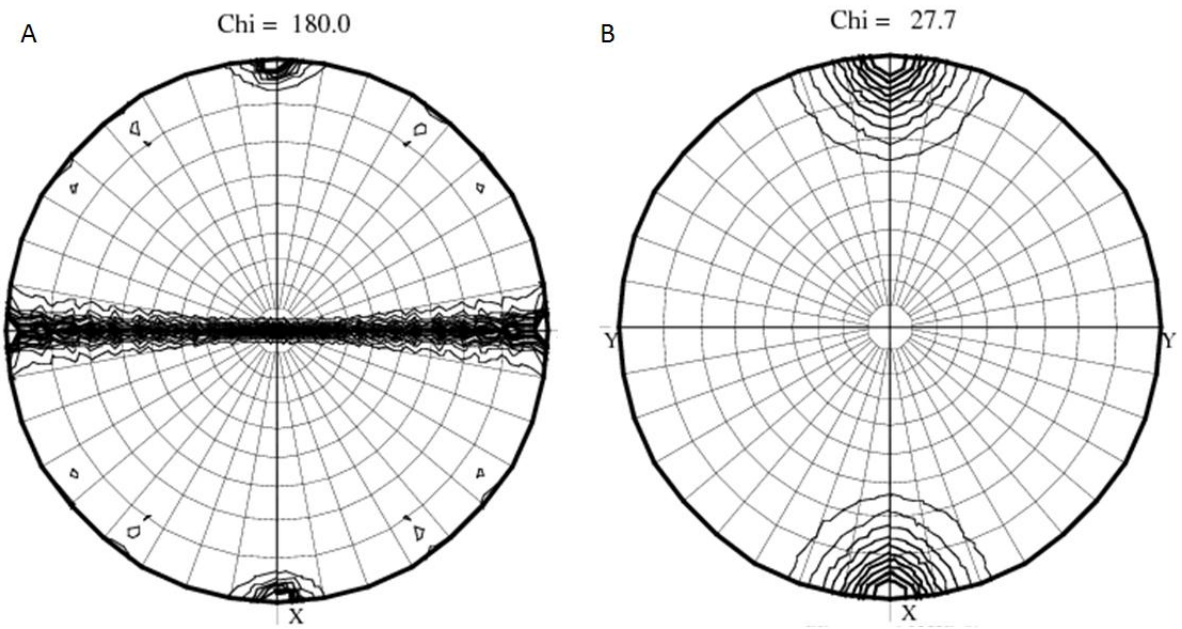


Figure 6-4 Stereographic projections =180° (A) and κ = 27.7° (B) of the self-rotation function.

6.4 Conclusions

In conclusion, the truncated portal protein cn3 of bacteriophage CNPH82 was successfully purified and crystallised. The X-ray data set collected from a native crystal was to the resolution of 4.2 Å. The oligomeric state was characterised to be 13 mer by SEC-MALLS and crystallographic analysis. Elucidating the structure of the portal protein will provide insights into the phage assembly, in particular the mechanism of viral DNA encapsidation.

7 The bacteriophage T4 portal protein – Gp20

7.1 Introduction

7.1.1 Bacteriophage T4

Bacteriophage T4 is a large tailed double-stranded DNA (dsDNA) virus that infects *Escherichia coli.* It belongs to *Myoviridae* family, and is the the most thoroughly investigated representative of the T-even phages (Calendar, 2006). The mature T4 virion consists of a 1200 Å long, 860 Å wide prolate icosahedral head with a 172 kb dsDNA chromosome; a 1000 Å long, 210 Å diameter cocylindrical contractile tail, terminated with a 460 Å diameter baseplate; and six 1450 Å long fibers attached to the baseplate (Leiman et al, 2003). The assemblies of head, tail, and fibers are independent processes, and these components will join together to form a mature virion.

The T4 DNA translocation machinery is one of the fastest and most powerful packaging motors reported to date, with a packaging velocity of 2,000 bp/s (Fuller et al, 2007). It is composed of three components: the dodecameric portal gp20 (61 kDa), the pentameric large terminase gp17 (70 kDa) and the 11- or 12-meric small terminase gp16 (18 kDa). With the ubiquitous distribution of T4-type bacteriophage, the tailed DNA phage T4 has served as an excellent model for the elucidation of the mechanisms underlying head assembly of T-even phages as well as general large icosahedral viruses (Rao & Black, 2010).

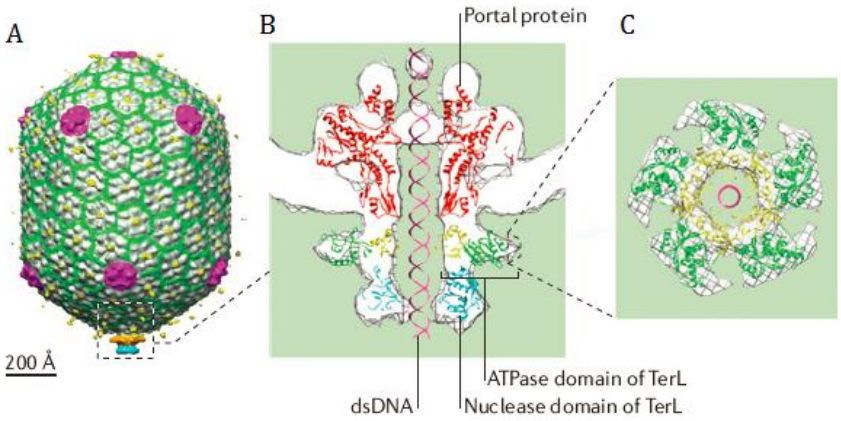


Figure 7-1 The DNA-packaging motor of phage T4.

(A): A three-dimensional cryo-electron microscopy reconstruction of the phage T4 procapsid with bound subunits of the large terminase subunit (TerL). (B): A magnification of the portal vertex. The white areas represent electron density, in which ribbon diagrams for TerL and the portal protein are modeled. (C): A view along the central channel of the indicated ring of electron density. This figure was adapted from (Casjens, 2011).

7.1.2 Studies of T4 Portal Protein Gp20

Bacteriophage T4 portal protein is the product of gene 20, encoding a 61 kDa polypeptide of 524 amino acids. The T4 portal gp20 is well documented to perform multiple functions: initiation of prohead assembly, initiation of core/scaffolding assembly, terminase interaction in DNA packaging, the headful packaging gauge, prohead expansion and stabilization, and head to tail connection.

Purification and crystallisation studies of gp20 began nearly two decades ago, but limited progress has been made due to its hydrophobic properties (Driedonks et al, 1981). Gp20 was first purified from proheads in the presence of certain amount of detergents, such as 0.5% of Nonidet P-40 and 8M urea (Driedonks et al, 1981). Then, fusion proteins of gp20 with Green Fluorescent Protein (GFP), still purified in the presence of 8M urea, were constructed for *in vitro* functional study (Baumann et al, 2006). The gp20 structure was finally observed by negative staining electron microscopy with the majority of particles displaying a ring-like appearance (Driedonks et al, 1981) (Figure 7-2). As expected, the overall structure of gp20 shares common characteristics with homologues from other double-stranded bacteriophages as a dodecamer with a central DNA channel (Driedonks, 1981). As observed by cryo electron microscopy, the narrower end of gp20 protrudes out of the capsid, and the wider end is inside the capsid (Leiman et al, 2004). However, a high-resolution atomic structure of gp20 remains unavailable, because high quality crystals cannot be obtained in conditions containing such high concentrations of urea.

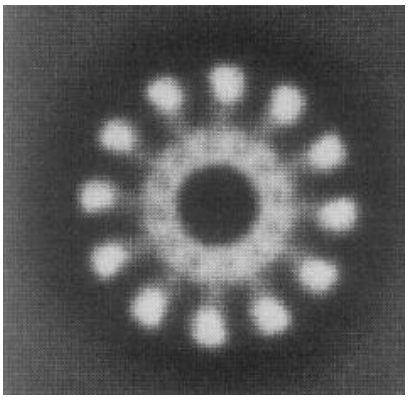


Figure 7-2 Negative-staining EM analysis of the T4 portal protein.
The portal protein from bacteriophage T4 was shown to be a 12-subunit assembly, and the figure was adapted from (Driedonks et al, 1981).

As a central component of the DNA translocation machinery, gp20 forms a channel with a diameter of ~35 Å, through which the DNA is translocated into the procapsid. However, the exact role of gp20 in viral DNA translocation is not yet clear. A recent study using direct binding assays, mutagenesis, and structural analyses, confirmed the specific interaction between gp20 and the large terminase gp17, and identified the interaction sites: part of the protruding αβ domain in gp20 and a conserved helix-turn-helix (HLH) in subdomain II of gp17. Therefore, a molecular lever mechanism coupling ATP hydrolysis to DNA movement has been proposed for the T4 DNA translocation motor (Hegde et al, 2012).

7.1.3 Aim of the Project

According to previous studies on the T4 portal protein gp20, a soluble protein can only be purified in the presence of harsh denaturing agents such as a high concentration of urea (Driedonks et al, 1981) which presents a big challenge for obtaining high quality crystals for X-ray structural studies. In this project, different cloning strategies and purification methods were carried out in order to produce soluble portal protein suitable for structural analysis.

7.2 Materials and Methods

7.2.1 Cloning of the Gp20 Constructs

The plasmid pAF20 with the full-length gene 20 in the expression vector pET14b was used as the template for the PCR reaction. The primers (Table 7-2) were designed to introduce *NheI* and *XhoI* restriction sites within the PCR product of gp20 DNA. Amplification of required ORF by PCR was performed in a 50 μL reaction vial (Table 7-1). The cycling conditions were as follows: initial denaturation at 98 °C for 30 s, denaturation of the template at 98 °C for 10 s, primers annealing at 55 °C for 20 s, extension at 72 °C for 1 min and final extension at 72 °C for 5 min. The three steps – denaturation, annealing and extension were repeated 29 cycles. The insert was then cloned into vector pET28a to express gp20 with an N-terminal cleavable His tag, or vector pGEX 6P-3 to express the protein with a cleavable N-terminal GST tag. The correct inserts were verified by DNA sequencing (GATC Biotech) and alignment analysis (Clustal W).

Table 7-1 Components of PCR reaction (items were added in this order)

Total reaction volume	Amount/reaction (ul)	Final concentration
dH2O	13.6	
5 × Phusion® HF buffer	10	1×
dNTPs (25 mM each)	0.4	0.2 mM each
Primer 1 (2 μM)	12.5	0.5 μM
Primer 2 (2 μM)	12.5	0.5 μM
Template DNA	0.5	
Phusion® DNA Polymerase(2 U/μl)	0.5	0.02 U/ μl

Table 7-2 Primers for cloning of gp20 constructs

Primer	Sequence(5'- 3')
Gp20-F1	CCCCTAGCTAGCATGAAATTTAATGTATTAAGTTTGTTTG
Gp20-R1	GGGCCGCTCGAGTTAAAAATCCTCTTGTTCTTGG
Gp20-R2	GGGCCGCTCGAGTTAAGTCATCTGCAAAATGTC
Gp20-R3	GGGCCGCTCGAGTTATTCTATTTCTTCATCAGTCAT
GP20-R4	GGGCCGCTCGAGTTACTCTTTAGACTCTTCTTCAA
Gp20-F2	CGCGGATCCGAAAAAGAAGATCTTGTTTCC

7.2.2 Expression and Purification of the T4 Portal Protein Gp20

7.2.2.1 Low Temperature (12 °C) Expression of Gp20

The plasmid pYM145 was transformed into *Escherichia coli* ArcticExpress™ (DE3)RIL cells (Stratagene). A single colony from a Luria–Bertani (LB) agar plate containing 30 mg/μL kanamycin was picked to inoculate a small overnight culture at 37 °C. At this stage, 20 mg/μL gentamycin was also added into the LB medium for the selection of the chaperonin-encoding plasmid. 5 mL aliquots of the overnight culture were used to inoculate a 500 mL culture the following day. The 500 mL culture was grown at 37 °C until the mid-log phase (OD600 of 0.6–0.8) was reached, at which point protein expression was induced by the addition of IPTG to a final concentration of 1 mM. Lastly, overnight incubation of the culture at 12 °C and 180 r.p.m was carried out for the over-expression of gp20. Cells were recovered by centrifugation in SORRALL® RC 5B plus for 20 minutes at 5000g at 4 °C. Supernatant was removed and cell pellet was stored at -20 °C.

7.2.2.2 Autoinduction

The plasmid pYM145 was transformed into *Escherichia coli* B834 cells (Stratagene). A single colony from a Luria–Bertani (LB) agar plate containing 30 μg/mL kanamycin was picked to inoculate in LB medium overnight at 37 °C to make a starting culture. 1L of autoinduction medium with 30 μg/mL kanamycin was prepared. Cell cultures were inoculated (1:100) and grown overnight with shaking at 180 r.p.m at 16 °C to produce gp20. Cells were recovered by

centrifugation in SORRALL® RC 5B plus for 20 minutes at 5000 g, 4 °C. Supernatant was removed and cell pellet was stored at -20 °C.

7.2.2.3 Nickel affinity chromatography

The cell pellet was re-suspended in nickel affinity chromatography binding buffer containing 20mM Tris pH 8.0, 250 mM NaCl, 10 mM $MgCl_2$, and 50 mM imidazole supplemented with a combination of 100 µg/mL lysozyme and protease inhibitors (1 mM AEBSF and 0.7 µg/mL pepstatin). Cells were disrupted by sonication with a large probe in a glass beaker using short pulses of 30 seconds, with a 2 minutes resting time in between pulses. The sonication was carried out on ice, with caution to minimise thermal damage to protein extract. The cell debris was removed by centrifugation at 38758 g for 30 minutes at 4 °C using a Sorvall SS34 rotor. The supernatant was collected and then cleared with a 0.45 mm filter (Millipore). The filtrate loaded onto a 5 mL HiTrap column (GE Healthcare) equilibrated with binding buffer for nickel affinity chromatography purification. The bound protein was eluted with an increasing proportion of elution buffer containing 20 mM Tris pH 8.0, 250 mM NaCl, 10 mM $MgCl_2$, and 500 mM imidazole. The samples were analysed by 12% SDS-PAGE.

7.2.2.4 Ammonium sulphate purification of gp20 (M1-T494)

Appropriate amount of solid ammonium sulphate, as indicated in the Table 2-5, was added into the protein solution to precipitate gp20 at room temperature in the presence of different concentrations of ammonium sulphate using the following step-wise increase of its concentration: 20%, 25%, 30%, 35% and 40%. Fresh buffer 20 mM HEPES pH 7.0, 50 mM NaCl and 10% glycerol was subsequently used to dissolve the precipitated gp20.

7.2.2.5 Batch Purification of the GST Fused Gp20 Δ24-494

The protein was produced from 1 litre of *E. coli* B834 cells grown in LB medium. A standard induction protocol entails shifting log-phase cultures (A600 = ~0.6) from 37 °C to 16 °C and adding IPTG to a final concentration of 1 mM. After an overnight of constant vigorous shaking, bacterial cells were recovered by centrifugation at 5000 g (SORVALL® RC 5B Plus rotor) at 4 °C for 20 minutes and stored at -80 °C until purification.

A solubility screen was performed with conditions shown in Table 7-3 before attempting the affinity purification, and the solubility of the protein sample was visualised by 12% SDS-PAGE. The cell pellets were lysed by sonication in 20 mL lysis Buffer 20 mM Tris pH 8.5, 150 mM NaCl, and the insoluble cell debris was removed by centrifugation at 38758 g for 30 minutes at 4 °C using a Sorvall SS34 rotor. To immobilise the GST fused gp20 Δ24-494 on Glutathione Sepharose, the clarified lysate was incubated with Glutathione-Sepharose 4B (AP Biotech) on a tumbler at 4 °C for 4 hours (or one hour at room temperature) in 20 mM Tris pH 8.5, 150 mM NaCl. The resin was washed four times before elution of the GST-gp20 Δ24-494 with 50 mM reduced glutathione added in 20 mM Tris pH 8.5, 150 mM NaCl. The samples were analysed by 10% SDS-PAGE.

Table 7-3 Buffer conditions for solubility test

	Buffer / 20mM Tris	Salt (mM)	Additives
1	pH 5.5	150 mM NaCl	-
2	pH 6.5		-
3	pH 7.5		-
4	pH 8.5		-
5	pH 7.5	50 mM NaCl	-
6		300 mM NaCl	-
7		500 mM NaCl	-
8		150 mM KCl	-
9		500 mM KCl	-
10		150 mM NaCl	5% glycerol
11			0.1% Triton X-100

7.3 Results

7.3.1 Cloning, Expression and Purification of the Full-Length T4 Portal Protein Gp20

The DNA segment encoding the full-length gp20 was cloned into the restriction sites *NheI* and *XhoI* of vector pET28a (forming construct pYM145) for the expression of a recombinant protein containing an N-terminal hexahistidine tag with a thrombin cleavage site.

Results of the initial small-scale expression test showed temperature was the key factor affecting the solubility of the target protein. The low temperature expression in *E. coli* ArcticExpress™ (DE3)RIL cells at 12 °C and autoinduction in *E. coli* B834 cells at 16 °C both significantly improved the solubility (Figure 7-3). However, the solubility of gp20 from large-scale ArcticExpress cell culture was significantly reduced in comparison with the solubility shown from the small-scale expression test. In addition, a considerable amount of cold-adapted chaperones Cpn60/Cpn10 were co-purified with gp20 (Figure 7-3).

The purification of full-length gp20 produced by autoinduction at 16 °C in *E. coli* B834 cells was performed by nickel affinity chromatography and size exclusion chromatography (Figure 7-4). The molecular mass of the protein sample after this two-step purification was determined by MALDI-MS. The major peak from the MALDI mass spectrum, with an approximate mass of 74 kDa (Figure 7-5), corresponded to the *E. coli* chaperone protein DnaK (Hsp70). However, there was no peak with the molecular mass of 63.1 kDa corresponding to the gp20 subunit with the N-terminal hexahistidine tag.

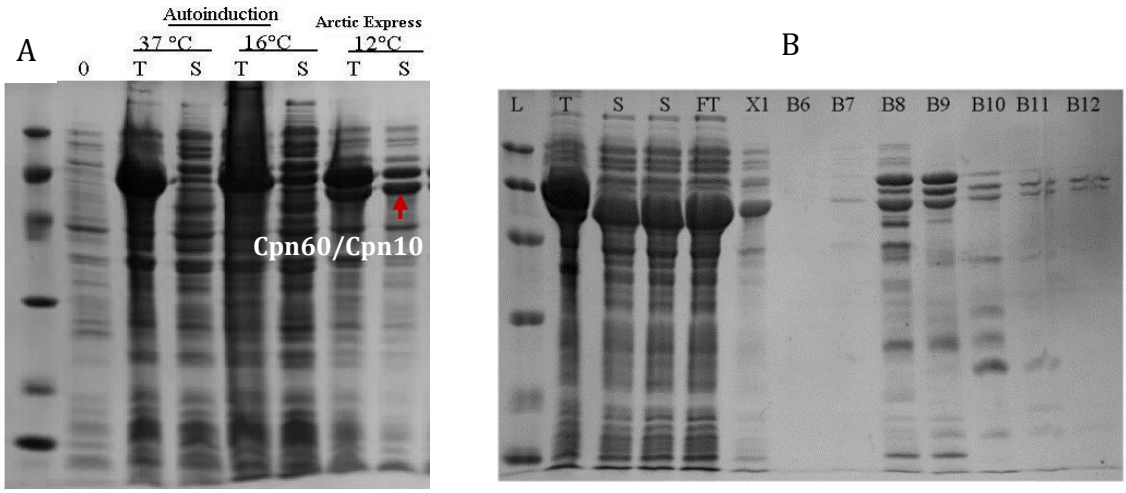


Figure 7-3 Purification of the full-length gp20.
(A) Small-scale expression test of the full-length gp20. A considerable amount of gp20 was soluble when produced in E. coli ArcticExpress (DE3)RIL cells at 12 °C and in E. coli B834 cells by autoinduction at 16 °C. The red arrow indicated the co-expression of the cold-adapted chaperones Cpn60/Cpn10 in E. coli ArcticExpress cells at 12 °C. (B) Nickel affinity purification of the full-length gp20 over-expressed in ArcticExpress cells at 12 °C. The fractions corresponding to elution peaks were resolved on a 12% SDS-PAGE gel L: low molecular weight ladder; T: the total protein lysate; S: soluble proteins after centrifugation; FT: flow-through;X1-B12: fractions from nickel affinity chromatography.

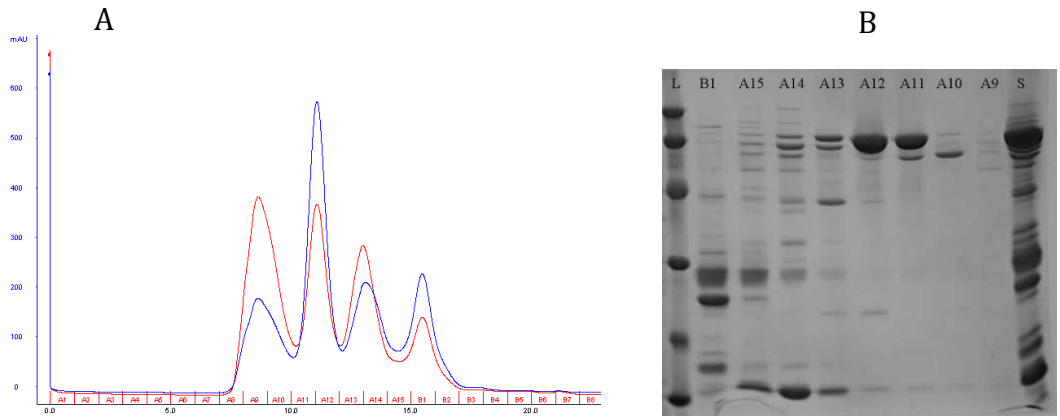


Figure 7-4 Purification of the full-length gp20 produced by autoinduction at 16°C.
(A) Chromatogram profile of size exclusion chromatography of the full-length gp20 on a Superdex 200 column. Absorption at 280 nm was shown in blue, 260 nm in red. (B)The fractions corresponding to elution peaks from size exclusion chromatography

were analysed on a 12% SDS-PAGE gel. L: low molecular weight marker (BioRad); S: loaded sample; A9-B1: fractions from size exclusion chromatography.

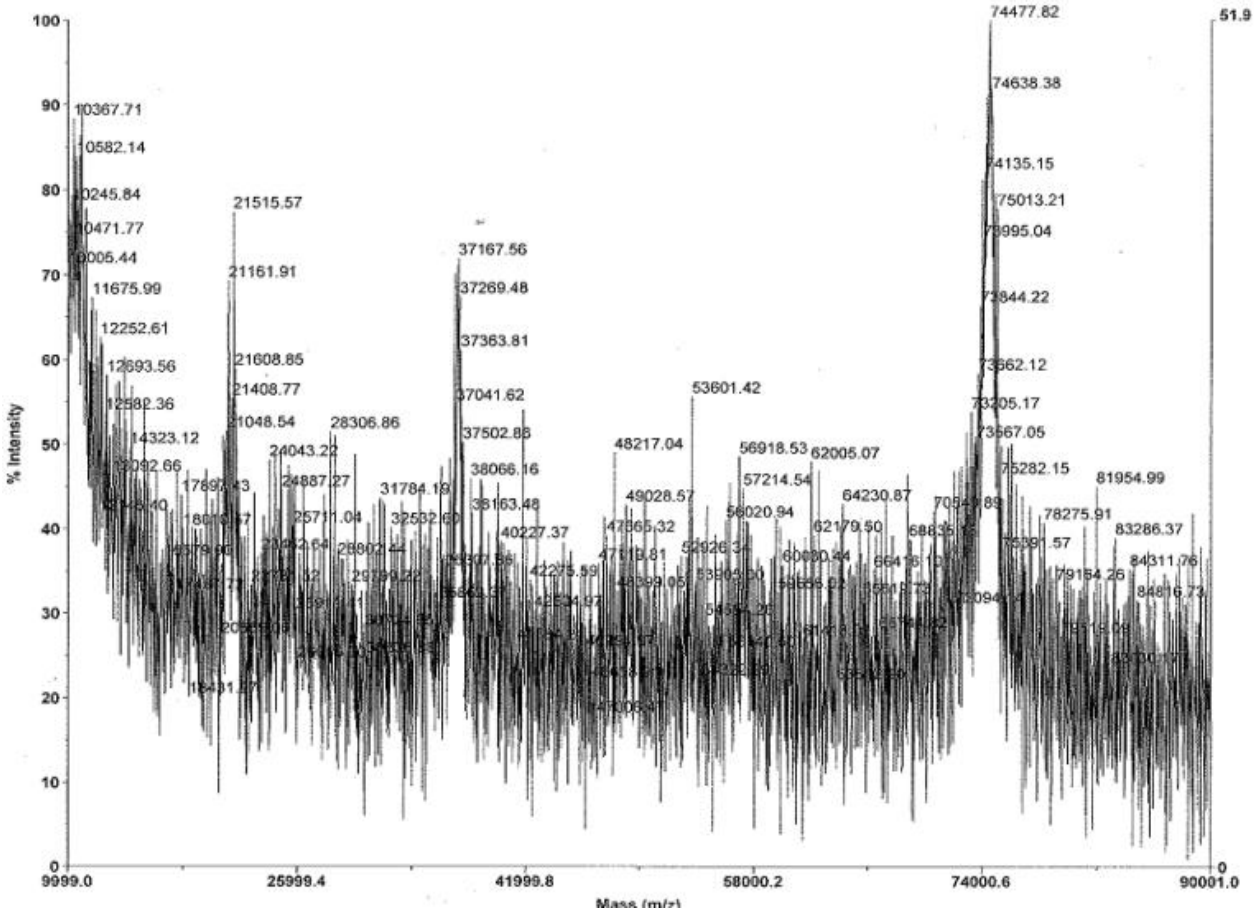


Figure 7-5 MALDI-MS characterisation of the sample after purification of the full-length gp20.
The protein was purified after the two-step purification. The major peak in the MALDI mass spectrum, with an approximate mass of 74 kDa, corresponds to the E. coli chaperone protein DnaK (Hsp70).

7.3.2 Cloning, Expression and Purification of the C-terminally Truncated Gp20

In order to design constructs with enhanced solubility, a secondary structure analysis by Jpred (Figure 7-6) and disorder prediction by PrDOS (Figure 7-7) were carried out to locate potentially disordered regions in gp20. The last 30 amino acids at the C-terminus appear to be disordered. Accordingly, a C-terminally truncated construct gp20 Δ1-494 was designed.

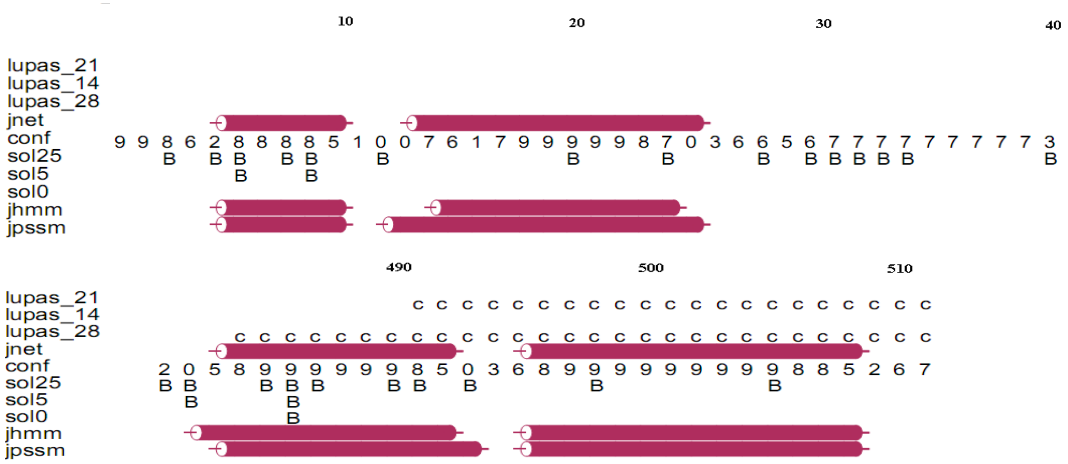


Figure 7-6 Secondary structure prediction of gp20.
Only ~40 amino acid N-terminal and ~40 amino acid C-terminal segments are shown. The prediction was performed by Jpred.

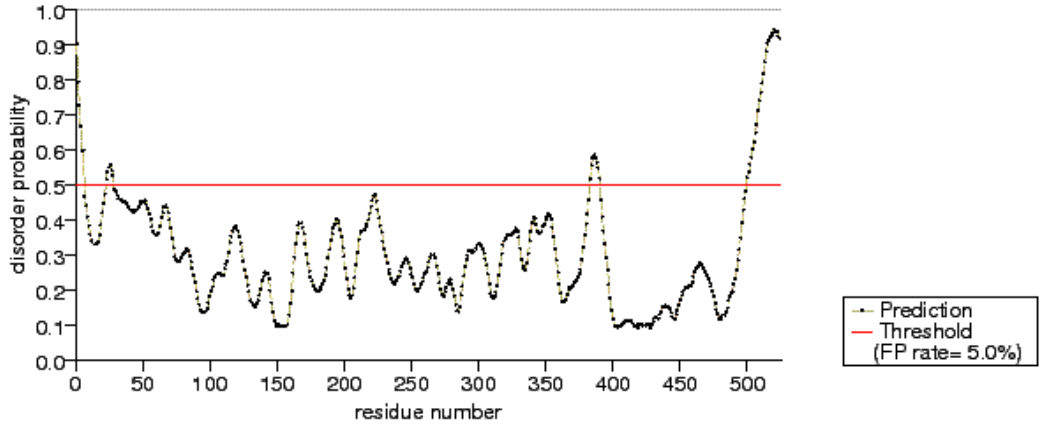


Figure 7-7 Disorder prediction of gp20. *The prediction was performed by PrDOS.*

The DNA segment encoding the C-terminally truncated gp20 Δ1-494 was cloned into the restriction sites *NheI* and *XhoI* of vector pET28a (forming construct pYM149) for the expression of a recombinant protein containing an N-terminal hexahistidine tag with a thrombin cleavage site.

The protein was over-expressed in *E. coli* B834 cells by autoinduction at 16 °C, and a solubility screen with different buffers was performed to optimise the solubility of the protein (Figure 7-8). According to SDS-PAGE, the over-expressed protein was soluble in 20 mM HEPES pH 7.0, 50 mM NaCl and 10% glycerol. The same SDS-PAGE gel was stained with His-tag in-gel stain for visualisation of the

His-tagged proteins, which confirmed the over-expressed protein as the target protein gp20.

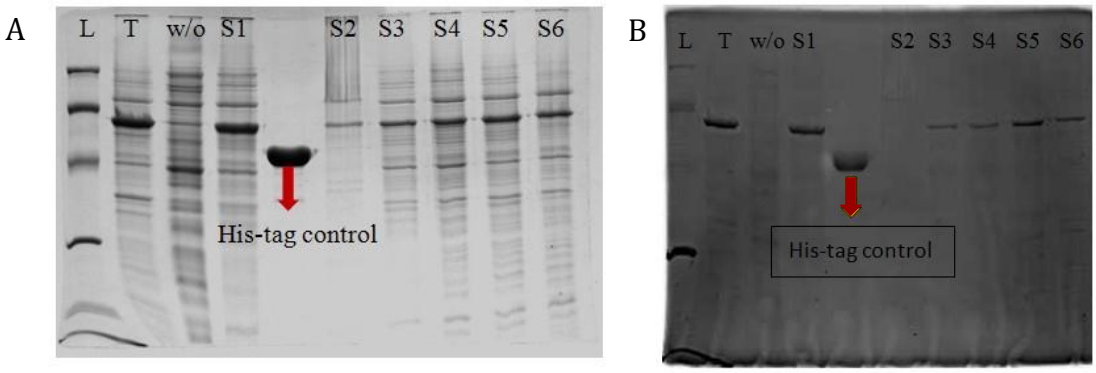


Figure 7-8 Solubility screen of the C-terminally truncated gp20 Δ1-494.
(A) Solubility screen of gp20 Δ1-494 with different buffers produced in E. coli B834 cells by autoinduction at 16°C. L: low molecular weight marker (BioRad); T: the total protein lysate; S: soluble protein after centrifugation. Proteins were resolved on a 12% SDS-PAGE gel; the red arrow indicated the His-tagged control sample. (B) The SDS PAGE gel was stained with His-tag In-gel stain to visualise His-tagged protein.

Nickel affinity chromatography was then performed using the best buffer identified - 20 mM HEPES pH 7.0, 50 mM NaCl and 10% glycerol. However, the protein did not bind to the nickel column, with the majority of the soluble gp20 Δ1-494 going directly to the flow-through (Figure 7-9). The incapability of gp20 Δ1-494 to bind to nickel was also confirmed by a following nickel beads pull down assay.

Other purification techniques without the involvement of histidine tag, such as ammonium sulphate precipitation and anion exchange chromatography were tried with no success. During ammonium sulphate precipitation, most soluble gp20 was precipitated in the presence of 40% ammonium sulphate (Figure 7-9). However, the precipitated protein could not be recovered by centrifugation and dissolved in fresh buffer. During anion exchange chromatography, the soluble gp20 Δ1-494 protein didn't bind to the MonoQ column or HiTrapQ column in the binding buffer 20 mM HEPES pH 7.0, and 50 mM NaCl.

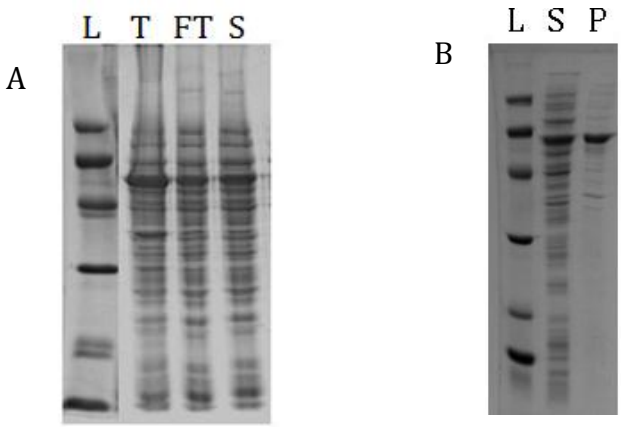


Figure 7-9 Purification of gp20 Δ1-494.
(A)Nickel affinity chromatography showed that most of the soluble protein (S) went to the flow-through (FT). (B), Ammonium sulphate precipitation showed that most of the soluble gp20 Δ1-494 precipitated in the presence of 40% ammonium sulphate. L: low molecular weight marker (BioRad); T: total protein lysate; S: soluble protein after centrifugation; P: precipitated protein. Proteins were resolved on a 12% SDS-PAGE gel.

7.3.3 Cloning, Expression and Purification of GST Fused Gp20 Δ24-494

The DNA fragment encoding the 80.6 kDa GST fused gp20 Δ24-494 with both N-terminal and C-terminal truncations was cloned within the *BamH1-Xho1* sites of the vector pGEX 6P-3, forming construct pYM170, to express the recombinant protein with a cleavable N-terminal GST tag (26 kDa). GST fused gp20 Δ24-494 was over-expressed in *E. coli* B834 cells. A solubility screen was performed to investigate the effect of pH and NaCl concentration on the solubility of the protein (Figure 7-10 A, buffer conditions shown in Table 7-3), during which the best buffer was identified as 20 mM Tris pH 8.5 and 150 mM NaCl.

The GST fused gp20 Δ24-494 was purified by binding to Glutathione-Sepharose beads and then eluted with 50 mM glutathione. The protein was able to bind to Glutathione-Sepharose beads, and 50 mM glutathione successfully competed with GST tag to elute the fusion protein (Figure 7-10 B). Further purification of the eluted proteins using size exclusion chromatography or anion exchange chromatography was not successful. The protein (GST fused gp20 Δ24-494) was

lost during the purification process possibly due to precipitation. PreScission Protease Cleavage was used for the removal of the GST tag, and further concentrating of the cleaved protein induced precipitation. The eluted proteins from Glutathione-Sepharose beads during batch purification were concentrated and checked by MALDI-MS to confirm the existence of the GST fused gp20 Δ24-494. However, there was no signal detected corresponding to the GST fused gp20 Δ24-494, which could be caused by protein precipitation during concentration.

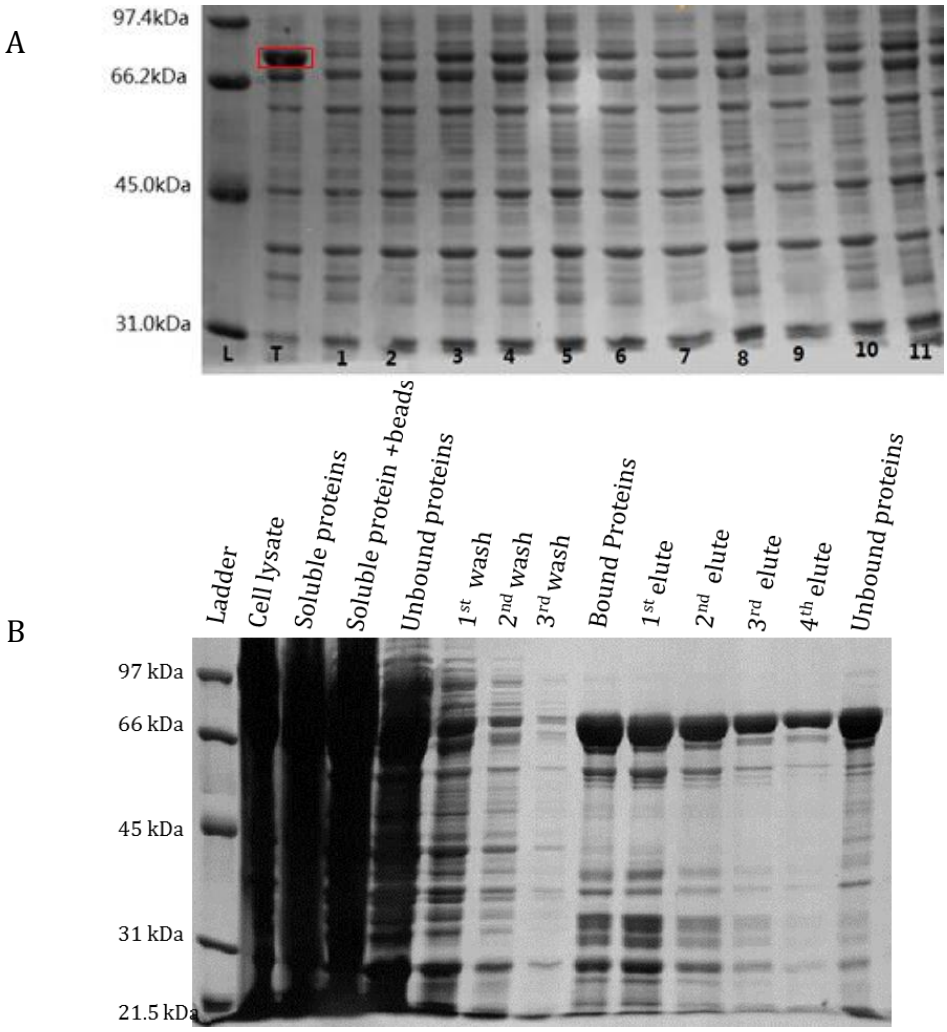


Figure 7-10 Purification of GST fused gp20 Δ24-494.

(A) Solubility screen of the GST fused gp20 Δ24-494 with different buffer conditions. Lane T indicates the total protein lysate, lanes 1 to 11 correspond to the soluble protein in different buffers listed in Table 7-3. (B) Batch purification of the GST fused gp20 Δ24-494 using Glutathione-Sepharose beads. Proteins were resolved on a 12% SDS-PAGE gel.

7.4 Discussion

Solubility is an intrinsic characteristic of a particular protein. The protein solubility can be affected by multiple factors including expression cell strains, the rate of protein synthesis, expression temperature and expression tags. In the case of the T4 portal protein gp20, the expressed protein is insoluble and accumulates in inclusion bodies when expressed at 37 °C. In this study, the solubility of the full-length gp20 was significantly improved when the protein was over-expressed in *E. coli* ArcticExpress RIL cells at 12 °C. The Cpn10 and Cpn60 chaperones were co-expressed to facilitate protein folding at 12 °C. However, the solubility of gp20 in large-scale ArcticExpress cell culture was significantly reduced in comparison to its solubility during small-scale expression tests, with considerable amounts of chaperones co-purified.

Solubility enhancement of the full-length gp20 was also observed when the protein was produced by autoinduction at 16 °C in *E. coli* B834 cells. The protein obtained after a two-step purification using nickel affinity chromatography and size exclusion chromatography turned out to be the *E. coli* chaperone protein DnaK, while soluble gp20 was lost during the purification process.

The C-terminal truncated gp20 Δ1-494, with the C-terminal disordered region removed, showed significant improvement with regard to the solubility. The protein was over-expressed in soluble form in *E. coli* B834 cells by autoinduction at 16 °C. However, this truncated protein could not bind to the nickel column with its N-terminal histidine tag, which might imply the misfolding of the protein. Alternative purification methods such as anion exchange chromatography and ammonium sulphate precipitation have also proved to be unsuccessful for the purification of gp20 Δ1-494.

The gp20 Δ24-494 was fused to a GST tag, which can bind to Glutathione-Sepharose beads and be eluted by 50 mM glutathione. The further purification by size exclusion chromatography or anion chromatography was unsuccessful with protein lost during the purification process. The purity and yield of the GST fused gp20 Δ24-494 after the batch purification was insufficient for further characterisation and crystallisation.

The future work could be directed towards fusing maltose-binding protein (MBP, ~42.5 kDa) as a tag, because MBP could serve as a solubility enhancer and permit one-step purification using amylose resin. It is also worth evaluating a potential use of the available GFP fused T4 portal protein constructs which are soluble in native buffer condition (Baumann et al, 2006).

8 Conclusions

Obtaining the SPP1 portal protein gp6 in its 12-subunit biologically active state *in vitro* is considered to be extremely important for elucidating the mechanism of the DNA packaging molecular motor and for designing a stable molecular machine that can function *in vitro*. Two strategies involving protein engineering of fusion proteins and co-expression of gp6 with gp11, were designed and investigated using a combination of biochemical, biophysical and structural approaches.

Several constructs of the SPP1 portal protein gp6 were fused with *B. halodurans*/*B. stearothermphilus* TRAP. The fusion proteins obtained with the *B. stearothermphilus* TRAP proved to be insoluble. The fusion proteins of gp6 with *B. halodurans* TRAP were successfully overexpressed in *E. coli* BL21 cells and purified, and their oligomeric states were analysed by SEC-MALLS and electron microscopy. The results revealed 13-fold symmetrical assemblies formed by construct containing gp6 truncations Δ27-466, Δ27-472 and Δ27-479, and mixtures of 13-mer and 14-mer assemblies for constructs containing gp6 Δ27-434. The oligomeric state remained 13 when the fusion protein containing gp6 Δ27-466 was treated with the dissociation-re-association procedure. The ring corresponding to the TRAP assembly was detected by electron microscopy; however, the number of subunits contained in the TRAP ring can't be deduced by EM due to the limited resolution. Because the study of the TRAP oligomer stability showed that *B. stearothermphilus* wild type and E71Stop TRAP are the most stable oligomers, future studies on fusion proteins could try to fuse *B. stearothermphilus* wild type and E71Stop TRAP with gp6.

The oligomeric state of the C-terminally truncated gp6, co-expressed in *E. coli* with gp11, was characterised by SEC-MALLS, electron microscopy and crystallographic analysis, and the results showed that gp6 is still a 13-mer. Because the negatively charged C-terminus of gp6 could be important for interaction between gp6 and gp11, future work should be directed towards the co-expression of the full-length gp6 with gp11 and characterising the oligomeric state of gp6 by negative staining electron microscopy. Attempts to produce diffracting crystals of gp11 were unsuccessful; however, further solubility screening of various truncated constructs may prove useful.

The structure of the wild type gp6 Δ27-472 was solved at 2.8 Å resolution, revealing a 13-fold symmetrical molecule. The electron density maps corresponding to Asn365 are very clear. The protein's fold is the same as for the N365K mutant, with most significant conformational differences observed in the tunnel loop and in segments of the clip and crown domains. Comparison with the structure of N365K mutant reveals significant differences in subunit-subunit interactions formed by tunnel loops, including different hydrogen bonds and van der Waals interactions. These differences may account for the different behaviour of capsids containing the mutant portal protein.

The truncated portal protein cn3 of bacteriophage CNPH82 was successfully purified and crystallised. The X-ray data set from a crystal of the wild type protein was collected to the resolution of 4.2 Å. Self-rotation function calculation and SEC-MALLS experiments indicate that cn3 forms 13-subunit assemblies, like the SPP1 portal protein.

The work on T4 portal protein gp20 aimed to produce soluble protein for structural studies. The full-length gp20 could not be purified in a soluble form; while a C-terminal truncated gp20 Δ1-494, as well as GST fused gp20 Δ24-494 showed significant improvements in solubility, production of crystallisation quantities of pure soluble protein was not feasible. Further work could be directed towards exploring the effect of fusing the portal protein with a maltose-binding protein to improve its solubility.

In summary, this work has focused on the study of the SPP1 portal protein gp6. In order to promote the formation of the dodecameric gp6, two strategies involving protein engineering of fusion proteins and co-expression of gp6 with gp11, were designed and investigated. The structure of the wild type gp6 was determined at 2.8 Å resolution, providing structural basis for understand the effect of N365K substitution on the amount of packaged DNA. Finally, cloning, expression, purification and crystallisation studies on the SPP1 scaffolding protein gp11, CNPH82 portal protein cn3, and T4 portal protein gp20 were also performed. In the case of CNPH82 portal protein cn3, promising crystallisation conditions have

been identified. Further optimisation should lead to structure determination in future.

9 References

Al-Zahrani AS, Kondabagil K, Gao S, Kelly N, Ghosh-Kumar M, Rao VB (2009) The small terminase, gp16, of bacteriophage T4 is a regulator of the DNA packaging motor. *J Biol Chem* **284:** 24490-24500

Alam TI, Draper B, Kondabagil K, Rentas FJ, Ghosh-Kumar M, Sun S, Rossmann MG, Rao VB (2008) The headful packaging nuclease of bacteriophage T4. *Mol Microbiol* **69:** 1180-1190

Alam TI, Rao VB (2008) The ATPase domain of the large terminase protein, gp17, from bacteriophage T4 binds DNA: implications to the DNA packaging mechanism. *J Mol Biol* **376:** 1272-1281

Alonso JC, Luder G, Stiege AC, Chai SH, Weise F, Trautner TA (1997) The complete nucleotide sequence and functional organization of Bacillus subtilis bacteriophage SPP1. *Gene* **204:** 201-212

Alonso JC, Tavares P, Lurz R, Trautner TA (2006) Bacteriophage SPP1. *The Bacteriophages (2nd edition)*

Antson AA, Otridge J, Brzozowski AM, Dodson EJ, Dodson GG, Wilson KS, Smith TM, Yang M, Kurecki T, Gollnick P (1995) The structure of trp RNA-binding attenuation protein. *Nature* **374:** 693-700

Apte A, Daniel S (2009) PCR primer design. *Cold Spring Harb Protoc* **2009:** pdb ip65

Baneyx F (1999) Recombinant protein expression in Escherichia coli. *Curr Opin Biotechnol* **10:** 411-421

Baumann RG, Mullaney J, Black LW (2006) Portal fusion protein constraints on function in DNA packaging of bacteriophage T4. *Mol Microbiol* **61:** 16-32

Bayfield OW, Chen CS, Patterson AR, Luan W, Smits C, Gollnick P, Antson AA (2012) Trp RNA-Binding Attenuation Protein: Modifying Symmetry and Stability of a Circular Oligomer. *PLoS One* **7:** e44309

Bazinet C, King J (1988) Initiation of P22 procapsid assembly in vivo. *J Mol Biol* **202:** 77-86

Becker B, de la Fuente N, Gassel M, Gunther D, Tavares P, Lurz R, Trautner TA, Alonso JC (1997) Head morphogenesis genes of the Bacillus subtilis bacteriophage SPP1. *J Mol Biol* **268:** 822-839

Bertani G (1953) Lysogenic Versus Lytic Cycle of Phage Multiplication. *Cold Spring Harb Sym* **18:** 65-70

Bhattacharyya SP, Rao VB (1993) A novel terminase activity associated with the DNA packaging protein gp17 of bacteriophage T4. *Virology* **196:** 34-44

Brennan RG, Matthews BW (1989) The helix-turn-helix DNA binding motif. *J Biol Chem* **264:** 1903-1906

Buttner CR, Chechik M, Ortiz-Lombardia M, Smits C, Ebong IO, Chechik V, Jeschke G, Dykeman E, Benini S, Robinson CV, Alonso JC, Antson AA (2011) Structural basis for DNA recognition and loading into a viral packaging motor. *Proc Natl Acad Sci U S A*

Calendar R (ed) (2006) *The Bacteriophages*: Oxford University press

Camacho AG, Gual A, Lurz R, Tavares P, Alonso JC (2003) Bacillus subtilis bacteriophage SPP1 DNA packaging motor requires terminase and portal proteins. *J Biol Chem* **278:** 23251-23259

Cardone G, Winkler DC, Trus BL, Cheng N, Heuser JE, Newcomb WW, Brown JC, Steven AC (2007) Visualization of the herpes simplex virus portal in situ by cryo-electron tomography. *Virology* **361:** 426-434

Casjens S, King J (1975) Virus assembly. *Annu Rev Biochem* **44:** 555-611

Casjens S, Wyckoff E, Hayden M, Sampson L, Eppler K, Randall S, Moreno ET, Serwer P (1992) Bacteriophage-p22 portal protein is part of the gauge that regulates packing density of intravirion dna. *J Mol Biol* **224:** 1055-1074

Casjens SR (2011) The DNA-packaging nanomotor of tailed bacteriophages. *Nat Rev Microbiol* **9:** 647-657

Cerritelli ME, Studier FW (1996) Assembly of T7 capsids from independently expressed and purified head protein and scaffolding protein. *J Mol Biol* **258:** 286-298

Chang JR, Poliakov A, Prevelige PE, Mobley JA, Dokland T (2008) Incorporation of scaffolding protein gpO in bacteriophages P2 and P4. *Virology* **370:** 352-361

Chayen NE (2004) Turning protein crystallisation from an art into a science. *Curr Opin Struct Biol* **14:** 577-583

Chen CS, Smits C, Dodson GG, Shevtsov MB, Merlino N, Gollnick P, Antson AA (2011a) How to change the oligomeric state of a circular protein assembly: switch from 11-subunit to 12-subunit TRAP suggests a general mechanism. *PLoS One* **6:** e25296

Chen DH, Baker ML, Hryc CF, DiMaio F, Jakana J, Wu WM, Dougherty M, Haase-Pettingell C, Schmid MF, Jiang W, Baker D, King JA, Chiu W (2011b) Structural basis for scaffolding-mediated assembly and maturation of a dsDNA virus. *Proceedings of the National Academy of Sciences of the United States of America* **108:** 1355-1360

Chen X, Antson AA, Yang M, Li P, Baumann C, Dodson EJ, Dodson GG, Gollnick P (1999) Regulatory features of the trp operon and the crystal structure of the trp RNA-binding attenuation protein from Bacillus stearothermophilus. *J Mol Biol* **289:** 1003-1016

Cingolani G, Moore SD, Prevelige PE, Jr., Johnson JE (2002) Preliminary crystallographic analysis of the bacteriophage P22 portal protein. *J Struct Biol* **139:** 46-54

Cornilleau C, Atmane N, Jacquet E, Smits C, Alonso JC, Tavares P, Oliveira L (2013) The nuclease domain of the SPP1 packaging motor coordinates DNA cleavage and encapsidation. *Nucleic Acids Res* **41:** 340-354

Cortines JR, Weigele PR, Gilcrease EB, Casjens SR, Teschke CM (2011) Decoding bacteriophage P22 assembly: identification of two charged residues in scaffolding protein responsible for coat protein interaction. *Virology* **421:** 1-11

D'Arcy A, Mac Sweeney A, Stihle M, Haber A (2003) The advantages of using a modified microbatch method for rapid screening of protein crystallization conditions. *Acta Crystallogr D Biol Crystallogr* **59:** 396-399

Daniel A, Bonnen PE, Fischetti VA (2007) First complete genome sequence of two Staphylococcus epidermidis bacteriophages. *Journal of Bacteriology* **189:** 2086-2100

Dawson RMC, Elliott DC, Elliott WH, Jones KM (1986) *Data for biochemical research*, 3rd ed edn. Oxford: Clarendon Press.

de Beer T, Fang J, Ortega M, Yang Q, Maes L, Duffy C, Berton N, Sippy J, Overduin M, Feiss M, Catalano CE (2002) Insights into specific DNA recognition during the assembly of a viral genome packaging machine. *Mol Cell* **9:** 981-991

Dokland T, Wang S, Lindqvist BH (2002) The structure of P4 procapsids produced by coexpression of capsid and external scaffolding proteins. *Virology* **298:** 224-231

Driedonks RA (1981) The quaternary structure of the T4 gene product 20 oligomer. *Prog Clin Biol Res* **64:** 315-323

Driedonks RA, Engel A, tenHeggeler B, van D (1981) Gene 20 product of bacteriophage T4 its purification and structure. *J Mol Biol* **152:** 641-662

Droege A, Tavares P (2005) *Bacteriophage SPP1 DNA packaging*: Kluwer Academic/Plenum Publ, 233 Spring St, New York, Ny 10013 USA.

Droge A, Santos MA, Stiege AC, Alonso JC, Lurz R, Trautner TA, Tavares P (2000) Shape and DNA packaging activity of bacteriophage SPP1 procapsid: Protein components and interactions during assembly. *J Mol Biol* **296:** 117-132

Dube P, Tavares P, Lurz R, Vanheel M (1993) The Portal Protein of Bacteriophage-Spp1 - a DNA Pump with 13-Fold Symmetry. *Embo J* **12:** 1303-1309

Durbin SD, Feher G (1996) Protein crystallization. *Annu Rev Phys Chem* **47:** 171-204

Earnshaw WC, Casjens SR (1980) DNA packaging by the double-stranded DNA bacteriophages. *Cell* **21:** 319-331

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D* **60:** 2126-2132

Feiss M, Rao VB (2012) The bacteriophage DNA packaging machine. *Adv Exp Med Biol* **726:** 489-509

Folta-Stogniew E, Williams KR (1999) Determination of molecular masses of proteins in solution: Implementation of an HPLC size exclusion chromatography and laser light scattering service in a core laboratory. *J Biomol Tech* **10:** 51-63

Fu CY, Morais MC, Battisti AJ, Rossmann MG, Prevelige PE (2007) Molecular dissection of O29 scaffolding protein function in an in vitro assembly system. *Journal of Molecular Biology* **366:** 1161-1173

Fuller DN, Raymer DM, Kottadiel VI, Rao VB, Smith DE (2007) Single phage T4 DNA packaging motors exhibit large force generation, high velocity, and dynamic variability. *Proc Natl Acad Sci U S A* **104:** 16868-16873

Garrett R, Grisham CM (2010) *Biochemistry*, 4th edn. Belmont, CA: Brooks/Cole, Cengage Learning.

Goetzinger KR, Rao VB (2003) Defining the ATPase center of bacteriophage T4 DNA packaging machine: requirement for a catalytic glutamate residue in the large terminase protein gp17. *J Mol Biol* **331:** 139-154

Gollnick P, Babitzke P, Antson A, Yanofsky C (2005) Complexity in regulation of tryptophan biosynthesis in Bacillus subtilis. *Annu Rev Genet* **39:** 47-68

Grabski A, Mehler M, D. D (2005) The Overnight Express Autoinduction System:

High-density cell growth and protein expression
while you sleep. *Nat Methods* **2:** 233-235

Gross J, Strupat K (1998) Matrix-assisted laser desorption/ionisation mass spectrometry applied to biological macromolecules. *Trac-Trends Anal Chem* **17:** 470-484

Gual A, Camacho AG, Alonso JC (2000) Functional analysis of the terminase large subunit, G2P, of Bacillus subtilis bacteriophage SPP1. *J Biol Chem* **275:** 35311-35319

Guasch A, Pous J, Ibarra B, Gomis-Ruth FX, Valpuesta JM, Sousa N, Carrascosa JL, Coll M (2002) Detailed architecture of a DNA translocating machine: the high-resolution structure of the bacteriophage phi29 connector particle. *J Mol Biol* **315:** 663-676

Guo PX (2002) Structure and function of phi 29 hexameric RNA that drives the viral DNA packaging motor: review. *Prog Nucleic Acid Re* **72:** 415-472

Hamada K, Fujisawa H, Minagawa T (1986) A defined in vitro system for packaging of bacteriophage T3 DNA. *Virology* **151:** 119-123

Harvey SC, Petrov AS, Devkota B, Boz MB (2009) Viral assembly: a molecular modeling perspective. *Phys Chem Chem Phys* **11:** 10553-10564

Hegde S, Padilla-Sanchez V, Draper B, Rao VB (2012) Portal-large terminase interactions of the bacteriophage T4 DNA packaging machine implicate a molecular lever mechanism for coupling ATPase to DNA translocation. *J Virol* **86:** 4046-4057

Hillenkamp F, Karas M (1990) Mass spectrometry of peptides and proteins by matrix-assisted ultraviolet laser desorption/ionization. *Methods Enzymol* **193:** 280-295

Inal JM (2003) Phage therapy: a reappraisal of bacteriophages as antibiotics. *Arch Immunol Ther Exp (Warsz)* **51:** 237-244

Ishida T, Kinoshita K (2007) PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res* **35:** W460-464

Jekow P, Behlke J, Tichelaar W, Lurz R, Regalla M, Hinrichs W, Tavares P (1999) Effect of the ionic environment on the molecular structure of bacteriophage SPP1 portal protein. *European Journal of Biochemistry* **264:** 724-735

Kanamaru S, Kondabagil K, Rossmann MG, Rao VB (2004) The functional domains of bacteriophage t4 terminase. *J Biol Chem* **279:** 40795-40801

Karas M, Bachmann D, Bahr U, Hillenkamp F (1987) Matrix-Assisted Ultraviolet-Laser Desorption of Nonvolatile Compounds. *Int J Mass Spectrom* **78:** 53-68

Kocsis E, Cerritelli ME, Trus BL, Cheng N, Steven AC (1995) Improved methods for determination of rotational symmetries in macromolecules. *Ultramicroscopy* **60:** 219-228

Kondabagil KR, Rao VB (2006) A critical coiled coil motif in the small terminase, gp16, from bacteriophage T4: insights into DNA packaging initiation and assembly of packaging motor. *J Mol Biol* **358:** 67-82

Kuebler D, Rao VB (1998) Functional analysis of the DNA-packaging/terminase protein gp17 from bacteriophage T4. *J Mol Biol* **281:** 803-814

Lebedev AA, Krause MH, Isidro AL, Vagin AA, Orlova EV, Turner J, Dodson EJ, Tavares P, Antson AA (2007) Structural framework for DNA translocation via the viral portal protein. *Embo J* **26:** 1984-1994

Lee CS, Guo P (1995) Sequential interactions of structural proteins in phage phi 29 procapsid assembly. *J Virol* **69:** 5024-5032

Leiman PG, Chipman PR, Kostyuchenko VA, Mesyanzhinov VV, Rossmann MG (2004) Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. *Cell* **118:** 419-429

Leiman PG, Kanamaru S, Mesyanzhinov VV, Arisaka F, Rossmann MG (2003) Structure and morphogenesis of bacteriophage T4. *Cell Mol Life Sci* **60:** 2356-2370

Li Y, Hunter RL, McIver RT, Jr. (1994) High-resolution mass spectrometer for protein chemistry. *Nature* **370:** 393-395

Lin H, Simon MN, Black LW (1997) Purification and characterization of the small subunit of phage T4 terminase, gp16, required for DNA packaging. *J Biol Chem* **272:** 3495-3501

Long F, Vagin AA, Young P, Murshudov GN (2008) BALBES: a molecular-replacement pipeline. *Acta Crystallogr D* **64:** 125-132

Luria SE, Delbruck M, Anderson TF (1943) Electron Microscope Studies of Bacterial Viruses. *J Bacteriol* **46:** 57-77

Lurz R, Orlova EV, Gunther D, Dube P, Droge A, Weise F, van Heel M, Tavares P (2001) Structural organisation of the head-to-tail interface of a bacterial virus. *J Mol Biol* **310:** 1027-1037

Mccoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40:** 658-674

McElroy C, Manfredo A, Wendt A, Gollnick P, Foster M (2002) TROSY-NMR Studies of the 91 kDa TRAP Protein Reveal Allosteric Control of a Gene Regulatory Protein by Ligand-altered Flexibility. *Journal of molecular biology* **323:** 463-473

McGraw AP, Mokdad A, Major F, Bevilacqua PC, Babitzke P (2009) Molecular basis of TRAP-5 ' SL RNA interaction in the Bacillus subtilis trp operon transcription attenuation mechanism. *Rna-a Publication of the Rna Society* **15:** 55-66

Mogridge J (2004) Using light scattering to determine the stoichiometry of protein complexes. *Methods in molecular biology (Clifton, NJ)* **261:** 113-118

Morais MC, Fisher M, Kanamaru S, Przybyla L, Burgner J, Fane BA, Rossmann MG (2004) Conformational switching by the scaffolding protein D directs the assembly of bacteriophage phiX174. *Mol Cell* **15:** 991-997

Morais MC, Kanamaru S, Badasso MO, Koti JS, Owen BAL, McMurray CT, Anderson DL, Rossmann MG (2003) Bacteriophage phi 29 scaffolding protein gp7 before and after prohead assembly. *Nature Structural Biology* **10:** 572-576

Mousset S, Thomas R (1969) Ter, a function which generates the ends of the mature lambda chromosome. *Nature* **221:** 242-244

Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* **53:** 240-255

Nadal M, Mas PJ, Blanco AG, Arnan C, Sola M, Hart DJ, Coll M (2010) Structure and inhibition of herpesvirus DNA packaging terminase nuclease domain. *Proc Natl Acad Sci U S A* **107:** 16078-16083

Nemecek D, Lander GC, Johnson JE, Casjens SR, Thomas GJ, Jr. (2008) Assembly architecture and DNA binding of the bacteriophage P22 terminase small subunit. *J Mol Biol* **383:** 494-501

Newcomb WW, Homa FL, Thomsen DR, Trus BL, Cheng N, Steven A, Booy F, Brown JC (1999) Assembly of the herpes simplex virus procapsid from purified components and identification of small complexes containing the major capsid and scaffolding proteins. *J Virol* **73:** 4239-4250

Olia AS, Prevelige PE, Johnson JE, Cingolani G (2011) Three-dimensional structure of a viral genome-delivery portal vertex. *Nat Struct Mol Biol* **18:** 597-U107

Orlova EV (2009) How viruses infect bacteria? *Embo J* **28:** 797-798

Orlova EV, Dube P, Beckmann E, Zemlin F, Lurz R, Trautner TA, Tavares P, van Heel M (1999) Structure of the 13-fold symmetric portal protein of bacteriophage SPP1. *Nat Struct Biol* **6:** 842-846

Orlova EV, Gowen B, Droge A, Stiege A, Weise F, Lurz R, van Heel M, Tavares P (2003) Structure of a viral DNA gatekeeper at 10 angstrom resolution by cryo-electron microscopy. *Embo J* **22:** 1255-1262

Otto M (2009) Bacterial Sensing of Antimicrobial Peptides. *Contrib Microbiol* **16:** 136-149

Padilla-Meier GP, Gilcrease EB, Weigele PR, Cortines JR, Siegel M, Leavitt JC, Teschke CM, Casjens SR (2012) Unraveling the Role of the C-terminal Helix-Turn-Helix of the Coat-binding Domain of Bacteriophage P22 Scaffolding Protein. *J Biol Chem*

Padilla-Meier GP, Teschke CM (2011) Conformational Changes in Bacteriophage P22 Scaffolding Protein Induced by Interaction with Coat Protein. *Journal of Molecular Biology* **410:** 226-240

Parker MH, Jablonsky M, Casjens S, Sampson L, Krishna NR, Prevelige PE, Jr. (1997) Cloning, purification, and preliminary characterization by circular dichroism and NMR of a carboxyl-terminal domain of the bacteriophage P22 scaffolding protein. *Protein Sci* **6:** 1583-1586

Poh SL, el Khaday F, Berrier C, Lurz R, Melki R, Tavares P (2008) Oligomerization of the SPP1 scaffolding protein. *Journal of Molecular Biology* **378:** 551-564

Porath J, Carlsson J, Olsson I, Belfrage G (1975) Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* **258:** 598-599

Prevelige PE, Jr., Thomas D, King J (1988) Scaffolding protein regulates the polymerization of P22 coat subunits into icosahedral shells in vitro. *J Mol Biol* **202:** 743-757

Rakhuba DV, Kolomiets EI, Dey ES, Novik GI (2010) Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell. *Pol J Microbiol* **59:** 145-155

Rao VB, Black LW (1988) Cloning, overexpression and purification of the terminase proteins gp16 and gp17 of bacteriophage T4. Construction of a defined in-vitro DNA packaging system using purified terminase proteins. *J Mol Biol* **200:** 475-488

Rao VB, Black LW (2010) Structure and assembly of bacteriophage T4 head. *Virol J* **7**

Rao VB, Feiss M (2008) The bacteriophage DNA packaging motor. *Annu Rev Genet* **42:** 647-681

Rao VB, Mitchell MS (2001) The N-terminal ATPase site in the large terminase protein gp17 is critically required for DNA packaging in bacteriophage T4. *J Mol Biol* **314:** 401-411

Rentas FJ, Rao VB (2003) Defining the bacteriophage T4 DNA packaging machine: evidence for a C-terminal DNA cleavage domain in the large terminase/packaging protein gp17. *J Mol Biol* **334:** 37-52

Riva S, Polsinelli M, Falaschi A (1968) A new phage of Bacillus subtilis with infectious DNA having separable strands. *J Mol Biol* **35:** 347-356

Riva SC (1969) Asymmetric transcription of B. subtilis phage SPP1 DNA in vitro. *Biochem Biophys Res Commun* **34:** 824-830

Rixon FJ (2008) A good catch: packaging the virus genome. *Cell Host Microbe* **3:** 120-122

Rossmann MG, Moras D, Olsen KW (1974) Chemical and biological evolution of nucleotide-binding protein. *Nature* **250:** 194-199

Roy A, Bhardwaj A, Datta P, Lander GC, Cingolani G (2012) Small Terminase Couples Viral DNA Binding to Genome-Packaging atpase Activity. *Structure* **20:** 1403-1413

Rubinchik S, Parris W, Gold M (1995) The in-vitro translocase activity of lambda-terminase and its subunits - kinetic and biochemical-analysis. *J Biol Chem* **270:** 20059-20066

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230:** 1350-1354

Schrodinger, LLC. (2010) The PyMOL Molecular Graphics System, Version 1.3r1.

Simpson AA, Tao Y, Leiman PG, Badasso MO, He Y, Jardine PJ, Olson NH, Morais MC, Grimes S, Anderson DL, Baker TS, Rossmann MG (2000a) Structure of the bacteriophage phi29 DNA packaging motor. *Nature* **408:** 745-750

Simpson AA, Tao YZ, Leiman PG, Badasso MO, He YN, Jardine PJ, Olson NH, Morais MC, Grimes S, Anderson DL, Baker TS, Rossmann MG (2000b) Structure of the bacteriophage phi 29 DNA packaging motor. *Nature* **408:** 745-750

Sivashanmugam A, Murray V, Cui C, Zhang Y, Wang J, Li Q (2009) Practical protocols for production of very high yields of recombinant proteins using Escherichia coli. *Protein Sci* **18:** 936-948

Smith DE, Tans SJ, Smith SB, Grimes S, Anderson DL, Bustamante C (2001) The bacteriophage [phis]29 portal motor can package DNA against a large internal force. *Nature* **413:** 748-752

Smits C, Chechik M, Kovalevskiy OV, Shevtsov MB, Foster AW, Alonso JC, Antson AA (2009) Structural basis for the nuclease activity of a bacteriophage large terminase. *EMBO Rep* **10:** 592-598

Sorensen HP, Mortensen KK (2005) Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microb Cell Fact* **4:** 1

Stathopulos PB, Scholz GA, Hwang YM, Rumfeldt JA, Lepock JR, Meiering EM (2004) Sonication of proteins causes formation of aggregates that resemble amyloid. *Protein Sci* **13:** 3017-3027

Strupat K (2005) Molecular weight determination of peptides and proteins by ESI and MALDI. *Methods Enzymol* **405:** 1-36

Studier FW (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41:** 207-234

Studier FW, Moffatt BA (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189:** 113-130

Summers WC (2001) Bacteriophage therapy. *Annu Rev Microbiol* **55:** 437-451

Sun S, Gao S, Kondabagil K, Xiang Y, Rossmann MG, Rao VB (2012) Structure and function of the small terminase component of the DNA packaging machine in T4-like bacteriophages. *Proc Natl Acad Sci U S A* **109:** 817-822

Sun S, Kondabagil K, Draper B, Alam TI, Bowman VD, Zhang Z, Hegde S, Fokine A, Rossmann MG, Rao VB (2008) The structure of the phage T4 DNA packaging motor suggests a mechanism dependent on electrostatic forces. *Cell* **135:** 1251-1262

Sun S, Kondabagil K, Gentz PM, Rossmann MG, Rao VB (2007) The structure of the ATPase that powers DNA packaging into bacteriophage T4 procapsids. *Mol Cell* **25:** 943-949

Sun YH, Parker MH, Weigele P, Casjens S, Prevelige PE, Krishna NR (2000) Structure of the coat protein-binding domain of the scaffolding protein from a double-stranded DNA virus. *Journal of Molecular Biology* **297:** 1195-1202

Szigeti R, Milescu M, Gollnick P (2004) Regulation of the tryptophan biosynthetic genes in Bacillus halodurans: common elements but different strategies than those used by Bacillus subtilis. *J Bacteriol* **186:** 818-828

Tadokoro T, Kanaya S (2009) Ribonuclease H: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes. *Febs J* **276:** 1482-1493

Tavares P, Santos MA, Lurz R, Morelli G, Delencastre H, Trautner TA (1992) Identification of a Gene in Bacillus-Subtilis Bacteriophage Spp1 Determining the Amount of Packaged DNA. *J Mol Biol* **225:** 81-92

Teschke CM, Parent KN (2010) 'Let the phage do the work': Using the phage P22 coat protein structures as a framework to understand its folding and assembly mutants. *Virology* **401:** 119-130

Thiel K (2004) Old dogma, new tricks--21st Century phage therapy. *Nat Biotechnol* **22:** 31-36

Tomka MA, Catalano CE (1993) Physical and Kinetic Characterization of the DNA Packaging Enzyme from Bacteriophage-Lambda. *J Biol Chem* **268:** 3056-3065

Trus BL, Cheng N, Newcomb WW, Homa FL, Brown JC, Steven AC (2004a) Structure and polymorphism of the UL6 portal protein of herpes simplex virus type 1. *J Virol* **78:** 12668-12671

Trus BL, Cheng N, Newcomb WW, Homa FL, Brown JC, Steven AC (2004b) Structure and polymorphism of the UL6 portal protein of herpes simplex virus type 1. *J Virol* **78:** 12668-12671

Tuma R, Parker MH, Weigele P, Sampson L, Sun Y, Krishna NR, Casjens S, Thomas GJ, Jr., Prevelige PE, Jr. (1998) A helical coat protein recognition domain of the bacteriophage P22 scaffolding protein. *J Mol Biol* **281:** 81-94

Twort FW (1925) The discovery of the "bacteriophage.". *Lancet* **1:** 845-845

Vagin A, Teplyakov A (2010) Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* **66:** 22-25

Valpuesta JM, Sousa N, Barthelemy I, Fernandez JJ, Fujisawa H, Ibarra B, Carrascosa JL (2000) Structural analysis of the bacteriophage T3 head-to-tail connector. *J Struct Biol* **131:** 146-155

Veesler D, Cambillau C (2011) A common evolutionary origin for tailed-bacteriophage functional modules and bacterial machineries. *Microbiol Mol Biol Rev* **75:** 423-433, first page of table of contents

Walker JE, Saraste M, Runswick MJ, Gay NJ (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *Embo J* **1:** 945-951

Wang S, Palasingam P, Nokling RH, Lindqvist BH, Dokland T (2000) In vitro assembly of bacteriophage P4 procapsids from purified capsid and scaffolding proteins. *Virology* **275:** 133-144

Weber PC (1991) Physical Principles of Protein Crystallization. *Advances in Protein Chemistry* **41:** 1-36

Weigele PR, Sampson L, Winn-Stapley D, Casjens SR (2005) Molecular genetics of bacteriophage P22 scaffolding protein's functional domains. *Journal of Molecular Biology* **348:** 831-844

White HE, Sherman MB, Brasiles S, Jacquet E, Seavers P, Tavares P, Orlova EV (2012) Capsid structure and its stability at the late stages of bacteriophage SPP1 assembly. *J Virol* **86:** 6768-6777

Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF (1999) Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* **112:** 531-552

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D* **67:** 235-242

Wyatt PJ (1993) Light-scattering and the absolute characterization of macromolecules. *Analytica Chimica Acta* **272:** 1-40

Xiang Y, Morais MC, Battisti AJ, Grimes S, Jardine PJ, Anderson DL, Rossmann MG (2006) Structural changes of bacteriophage phi29 upon DNA packaging and release. *Embo J* **25:** 5229-5239

Yang M, Chen XP, Militello K, Hoffman R, Fernandez B, Baumann C, Gollnick P (1997) Alanine-scanning mutagenesis of Bacillus subtilis trp RNA-binding attenuation protein (TRAP) reveals residues involved in tryptophan binding and RNA binding. *J Mol Biol* **270:** 696-710

Zhao H, Finch CJ, Sequeira RD, Johnson BA, Johnson JE, Casjens SR, Tang L (2010) Crystal structure of the DNA-recognition component of the bacterial virus Sf6 genome-packaging machine. *Proc Natl Acad Sci U S A* **107:** 1971-1976

Ziebuhr W, Hennig S, Eckart M, Kranzler H, Batzilla C, Kozitskaya S (2006) Nosocomial infections by Staphylococcus epidermidis: how a commensal bacterium turns into a pathogen. *Int J Antimicrob Ag* **28:** S14-S20

Ziegelhoffer T, Yau P, Chandrasekhar GN, Kochan J, Georgopoulos C, Murialdo H (1992) The purification and properties of the scaffolding protein of bacteriophage lambda. *J Biol Chem* **267:** 455-461