

**Cryo-EM analysis of DNA packaging by the HK97
bacteriophage**

Dorothy Hawkins

Doctor of Philosophy

University of York,

Chemistry,

March 2022

Abstract

DNA packaging is a key step in the assembly of dsDNA viruses, such as tailed bacteriophages and evolutionarily related herpes viruses, whereby empty capsids are filled with a copy of the viral genome. The task is mediated by an ATP powered DNA packaging motor, made of terminase proteins, interacting with the portal vertex of a capsid. *Cos* phages use a defined signal, recognised by the terminase machinery, to mark the beginning and end of their genome in newly replicated, concatemeric DNA.

This thesis focuses on cryo-EM analysis of the DNA packaging machine from the *Escherichia coli cos* bacteriophage HK97. A monomeric crystal structure of large terminase (PDB 6Z6D) is available, but accurate structural information for assemblies that initiate, process, and terminate packaging remained unknown. The motor was assembled *in vitro*, using proheads and terminase proteins purified separately, and DNA protection assays indicated that low concentrations of ATP could stall the packaging motor at a *cos* site of the substrate DNA. This behaviour permitted particles with an assembled packaging motor to be selected using a novel magnetic bead assay, that were suitable for structural studies.

Cryo-EM reconstructions elucidated a high-resolution structure of prohead II and the portal protein. An *in situ* asymmetric reconstruction of the portal protein was also achieved, showing unique contacts between the portal and surrounding capsomeres. Reconstruction of the motor displayed a ring of 5 large terminase monomers with DNA in the central channel. The motor is tilted relative to the portal axis, with only a single C-terminal domain of the terminase making clear contact. The N-terminal ATPase domains, constitute a second ring below, displaying variable degrees of extension away from the C-terminal domains. This is indicative of a novel packaging mechanism, where contraction between the 2 large terminase domains appears to drive translocation, with terminase-portal interactions varying throughout the ATP hydrolysis cycle.

Table of Contents

<i>Abstract</i>	2
<i>Table of Contents</i>	3
<i>List of Figures</i>	8
<i>List of Tables</i>	13
<i>List of Accompanying Material</i>	14
<i>Acknowledgements</i>	15
<i>Declaration</i>	16
<i>Project timeline</i>	17
<i>Chapter 1: Introduction</i>	18
<i>Virus introduction</i>	18
<i>Bacteriophage introduction</i>	21
<i>Bacteriophage infection</i>	23
<i>Bacteriophage assembly</i>	23
<i>Bacteriophage structure</i>	26
<i>HK97 fold and capsid expansion</i>	27
<i>Genome packaging strategies</i>	30
<i>Portal proteins</i>	33
<i>Small terminases</i>	36
<i>HK97 small terminase</i>	38
<i>ASCE ATPases</i>	39
<i>Translocation kinetics of the large terminase packaging motor</i>	40
<i>Models for DNA translocation by large terminase</i>	42

<i>Structure of the HK97 large terminase</i>	47
<i>Transmission Electron microscopy</i>	49
<i>Sample preparation for TEM</i>	50
<i>Transmission Electron Microscope anatomy</i>	50
<i>Image formation in the microscope</i>	53
<i>Building a 3D model</i>	58
<i>Estimating resolution</i>	60
<i>Chapter 2: Stalling the HK97 DNA packaging motor for structural investigation</i>	62
<i>Introduction</i>	62
<i>Methods</i>	67
<i>Phage amplification</i>	67
<i>Plaque assays</i>	67
<i>Capsid purification</i>	68
<i>Packaging assays</i>	69
<i>Negative stain electron microscopy</i>	69
<i>Magnetic bead packaging assays</i>	69
<i>Results</i>	70
<i>Optimisation of capsid production protocol to avoid expansion</i>	70
<i>Stalling the DNA packaging motor at a cos site</i>	73
<i>Restarting the stalled motor</i>	74
<i>Design of a magnetic bead assay to specifically isolate stalled packaging complexes</i>	75
<i>Optimisation of assay conditions</i>	78
<i>Collecting an initial cryo-EM data set</i>	79
<i>Introducing nucleosides and analogues to the assay to prevent motor detachment</i>	82
<i>Discussion</i>	84

<i>Optimisation of the purification of HK97 prohead II</i>	84
<i>Stalling of the HK97 DNA packaging motor</i>	84
<i>Development of a magnetic bead assay for selective purification of stalled packaging complexes</i>	86
<i>Conclusion</i>	86
<i>Chapter 3: Increasing the concentration of the stalled HK97 packaging sample</i>	88
<i>Introduction</i>	88
<i>Methods</i>	88
<i>Cryo electron microscopy screening</i>	89
<i>Data collection parameters</i>	89
<i>Concentration of the biotinylated sample using streptavidin coated affinity grids</i>	90
<i>Overexpression and purification of small terminase</i>	91
<i>Results</i>	92
<i>Cryo-EM analysis of an initial data set</i>	92
<i>Processing the cryo-EM data</i>	97
<i>Concentration of the final sample</i>	97
<i>Concentration using a PET concentrator</i>	98
<i>Concentration by PEG precipitation</i>	99
<i>Concentration by evaporation</i>	99
<i>Concentration by ammonium sulfate precipitation</i>	99
<i>Adjustments to the magnetic bead purification assay</i>	101
<i>Increase in ionic strength of magnetic bead assay buffers</i>	101
<i>Streptavidin coated affinity grids for concentration</i>	104
<i>Removing the biotin selection purification step</i>	106
<i>Discussion</i>	108

<i>Conclusion</i>	109
<i>Chapter 4: Structure of the prohead and portal protein</i>	110
<i>Introduction</i>	110
<i>Methods</i>	111
<i>Data collection parameters</i>	111
<i>Data processing pipeline</i>	114
<i>Results</i>	119
<i>Icosahedral reconstruction of the prohead</i>	120
<i>Reconstruction of the in situ portal protein structure from prohead II</i>	121
<i>Asymmetric reconstruction of the stalled packaging complex</i>	124
<i>Local reconstruction of the unique portal vertex</i>	124
<i>Discussion</i>	126
<i>Structure of the prohead</i>	126
<i>Structure of the in situ prohead portal</i>	127
<i>Structure of the unique portal-containing capsid vertex</i>	131
<i>Conclusion</i>	132
<i>Chapter 5: The structure of the motor</i>	134
<i>Introduction</i>	134
<i>Methods</i>	135
<i>Expression and purification of the K92 large terminase mutant</i>	135
<i>Crystallisation screens</i>	136
<i>Results</i>	136
<i>Processing strategies for the motor</i>	136
<i>1.Masking for the motor</i>	137
<i>2. Re-extraction of motor density</i>	138

2a) C5 symmetry expansion of symmetrical classes _____	138
2b) C5 symmetry expansion and re-extraction _____	140
2c) Isolating the CTD pentameric ring _____	141
3. Classification of expanded icosahedral particles by motor density _____	142
4) C12 symmetry expansion of the portal _____	144
5) Comparison of C12 and C5 angles for C1 reconstruction _____	145
6) Focussed classification and subtraction from a single data set _____	146
6a) Focussed classification and subtraction from a single data set for the portal/motor complex _____	149
Discussion _____	151
Fitting a monomeric large terminase X-ray structure into oligomeric assembly _____	151
Potential modes of DNA translocation _____	151
T4 like packaging _____	151
Φ 29 like packaging _____	153
FtsK like packagingr _____	155
Examination of angular distribution _____	157
Alignment between the prohead and motor _____	158
Absence of small terminase _____	160
Probing a potential glutamate switch _____	160
Conclusion _____	163
Accompanying Material _____	165
Glossary _____	167
Bibliography _____	169

List of Figures

<i>Figure 0.1: Timeline of cryo-EM data collections</i>	17
<i>Figure 1.1: The virus life cycle</i>	19
<i>Figure 1.2: Schematic of major ds DNA phage families</i>	21
<i>Figure 1.3: Expansion states of phage HK97</i>	23
<i>Figure 1.4: Bacteriophage assembly pathway</i>	24
<i>Figure 1.5: Depiction of Caspar Klug theory of the icosahedral phage architecture</i>	25
<i>Figure 1.6: Ribbon depiction of the HK97 fold</i>	26
<i>Figure 1.7: HK97 capsid expansion pathway</i>	28
<i>Figure 1.8: Differential ds DNA packaging strategies between bacteriophage</i>	31
<i>Figure 1.9: Ribbon depiction of the T4 portal structure 3JA7</i>	32
<i>Figure 1.10: Ribbon diagram of small terminase nonomers from the PDB</i>	35
<i>Figure 1.11: Ribbon diagram of HK97 small terminase nonomer</i>	38
<i>Figure 1.12: Depiction of the dwell burst cycle of DNA translocation by the Φ29 packaging motor</i>	40
<i>Figure 1.13: Depiction of Φ29 DNA translocation mechanism.</i>	42
<i>Figure 1.14: Structural models for T4 and Φ29 ds DNA packaging motors</i>	44
<i>Figure 1.15: Structure of the HK97 large terminase monomer</i>	46
<i>Figure 1.16: Structure of the HK97 large terminase nuclease domain determined by X-ray crystallography</i>	46
<i>Figure 1.17: Electron microscope anatomy where the red dotted line represents the central axis</i>	51

<i>Figure 1.18: Possible interactions between incoming electrons and the biological sample in an electron microscope</i>	52
<i>Figure 1.19: Vector diagram for phase contrast</i>	54
<i>Figure 1.20: Depiction of the Contrast Transfer Function</i>	56
<i>Figure 1.21: The Fourier projection slice theorem</i>	58
<i>Figure 1.22: An example of an FSC curve</i>	59
<i>Figure 2.1 DNA protection assay for HK97 DNA packaging</i>	63
<i>Figure 2.2: Current model of HK97 complete packaging machinery</i>	65
<i>Figure 2.3: Negative stain electron micrographs tracking HK97 prohead II purification from the amC2 amber mutant phage expressed in E. coli strain 594</i>	70
<i>Figure 2.4: Negative stain electron micrographs tracking HK97 prohead II final purification step</i>	71
<i>Figure 2.5: DNA protection assays mapping stalling of the HK97 packaging motor at decreasing concentrations of ATP. Protected DNA with, and without, an engineered cos site is shown by agarose gel electrophoresis.</i>	73
<i>Figure 2.6: DNA protection assay displaying the restarting of stalled HK97 packaging motors.</i>	74
<i>Figure 2.7: Schematic of magnetic bead assay to isolate stalled HK97 packaging complexes</i>	76
<i>Figure 2.8: Optimisation of parameters for a magnetic bead purification assay</i>	78
<i>Figure 2.9: Electron micrographs and models of HK97 prohead and portal, from an initial cryo-EM data collection</i>	80
<i>Figure 2.10: Optimised production of stalled HK97 packaging complexes</i>	82
<i>Figure 2.11: Potential mechanisms of stalling at the cos site by HK97 small terminase</i>	84

<i>Figure 3.1: Cryo-EM screening micrographs for stalled HK97 packaging complexes isolated by a magnetic bead assay</i>	93
<i>Figure 3.2: Distinct 3D classes of the HK97 packaging motor.</i>	95
<i>Figure 3.3: Monitoring of sample concentration techniques by negative stain electron micrographs</i>	99
<i>Figure 3.4: Purification of stalled HK97 packaging complexes by magnetic bead assay at 250 mM KGlu</i>	101
<i>Figure 3.5: Investigation of HK97 DNA packaging at variable concentrations of KGlu</i>	102
<i>Figure 3.6: Use of streptavidin affinity grids for concentration of stalled HK97 packaging complexes</i>	104
<i>Figure 3.7: Negative stain electron micrographs of optimised stalled HK97 packaging complexes</i>	106
<i>Figure 4.1: Cryo-EM screening micrograph for stalled HK97 packaging complexes.</i>	110
<i>Figure 4.2: Schematic depicting early data processing steps and combining of data sets</i>	115
<i>Figure 4.3: symmetry expansion and focused classification for identification of the portal/motor class</i>	117
<i>Figure 4.4: Icosahedral reconstruction of the HK97 prohead II to 3.06 Å resolution</i>	119
<i>Figure 4.5: C12 reconstruction of the HK97 in situ portal protein structure at 3.30 Å resolution</i>	120
<i>Figure 4.6: Local resolution for the map of the in situ portal protein</i>	122
<i>Figure 4.7: Asymmetric refinement of the HK97 full packaging complex</i>	123
<i>Figure 4.8: Local resolution of the asymmetric refinement of unique portal vertex</i>	125
<i>Figure 4.9: 2D classes from HK97 packaging complex sample</i>	126

<i>Figure 4.10: Comparison of portal protein structures</i>	127
<i>Figure 4.11: HK97 portal architecture in prohead II with and without an associated motor</i>	129
<i>Figure 4.12: Local resolution of the asymmetric refinement of unique portal vertex without angular restraints</i>	131
<i>Figure 5.1: Processing scheme 1 - masking the motor from the portal/mask 3D class.</i>	136
<i>Figure 5.2: Processing scheme 2 - re-extraction of the motor from the portal/mask 3D class</i>	137
<i>Figure 5.2a): Processing scheme 2a - symmetry expansion of classes showing C5 and reclassification</i>	139
<i>Figure 5.2b): Processing scheme 2b - symmetry expansion of classes showing C5, re-extraction to include portal signal, and reclassification in C12</i>	140
<i>Figure 5.2c): Processing scheme 2c - symmetry expansion of classes showing C5 symmetry, masked to include only CTD signal, and reclassification in C5</i>	141
<i>Figure 5.3): Processing scheme 3 -Re-extraction of icosahedral expanded particles in the motor locus followed by 3D classification</i>	142
<i>Figure 5.4): Processing scheme 3 - C12 symmetry expansion of portal refinement, re-extraction to include both motor and portal, and only motor, followed by downstream classification</i>	143
<i>Figure 5.5): Processing scheme 5 - C12 and C5 angles from portal and prohead refinement respectively compared for C1 angles. 3D classification with only local angular searches.</i>	145
<i>Figure 5.6): Processing scheme 6 - particle subtraction of the motor on eBIC data set</i>	147
<i>Figure 5.6a): Particle subtraction of the portal/motor complex from the eBIC data set</i>	149
<i>Figure 5.7: Pseudo-atomic structure of the large terminase</i>	151
<i>Figure 5.8: Schematic of cracked helix to planar model of DNA translocation</i>	153

<i>Figure 5.9: Proposed mechanism for DNA translocation by HK97 large terminase.</i>	_____	154
<i>Figure 5.10: The FtsK DNA translocation mechanism.</i>	_____	155
<i>Figure 5.11: Angular distribution for motor map 5.6 depicted from the side and from outside the prohead</i>	_____	157
<i>Figure 5.12: Alignment between the prohead and motor</i>	_____	158
<i>Figure 5.13: Alignment of HK97 large terminase ATPase domain with the large terminase of P74-26</i>	_____	160
<i>Figure 5.14: AlphaFold prediction of HK97 large terminase K92Q</i>	_____	161
<i>Figure 5.15: Motor and portal class following icosahedral symmetry expansion and focussed classification</i>	_____	162

List of Tables

<i>Table 2.1: data collection parameters for initial cryo-EM data collection of stalled HK97 packaging complexes</i>	79
<i>Table 3.1: data collection parameters for initial cryo-EM data collection of stalled HK97 packaging complexes</i>	89
<i>Table 3.2: Sample preparation of stalled HK97 packaging complexes for visualisation by transmission electron microscopy on streptavidin coated affinity grids</i>	90
<i>Table 4.1: Data collection parameters for three cryo-EM data collections of stalled HK97 packaging complexes</i>	112

List of Accompanying Material

Script provided by Dr Pavol Bardy for assymmetric reconstruction _____ 164

Acknowledgements

I dedicate this thesis to my oldest and youngest friends, Lady Hilda Cain and Daphne Bray Hawkins.

Firstly, I'd like to thank my supervisor Prof Fred Antson for supporting me throughout the project, always encouraging my ideas and providing exciting research opportunities. I'm also extremely grateful to Dr. Oliver Bayfield for day-to-day mentoring, support, and friendship. The whole Antson group have been exceptionally friendly and helpful- in particular, thanks to Dr. Pavol Bardy and Dr. Huw Jenkins for guidance with data processing, and to Dr. Sandra Grieve and especially to Maria Chechik for advice on the project and practical help in the lab. Dr Herman Fung also kindly introduced me to the project and set me up with skills and protocols. I'm grateful to the department of YSBL, for encompassing such a welcoming and collaborative community and providing excellent technical support. Thank you also to Mr Nick Burton who guided my industry placement, to my TAP member Dr. Jon Aguire, and to my UKRI-BBSRC iCASE studentship for funding my research and supporting my outreach projects.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. Cryo-EM data at the Astbury Centre for Structural Molecular Biology in Leeds, was collected with Dr. Emma Hesketh, Dr. Rebecca Thompson and Dr. Dan Maskell. At eBIC, cryo-EM data collections were facilitated by Dr Kyle Morris and Dr James Gilchrist. Cryo-EM data collection and screening at the University of York was performed with help from Dr. Jamie Blaza. All data collections were aided by Dr. Johan Tukenburg and Mr Sam Hart who coordinated data transfer and screening. Packaging assay protocols were established by Dr. Herman Fung in collaboration with Dr. Shelley Grimes and Dr. Paul Jardine, University of Minnesota. Original bacterial cultures and bacteriophage were provided by Dr. Robert Duda and Mr Joshua Maurer, University of Pittsburgh. All sources are acknowledged as references or where others contributed to work it is stated within the text.

Project timeline

The major aim of this project was to determine the structure of the HK97 packaging machinery by cryo-EM. Figure 0.1 summarises the structural work done prior to the project start (1), and the progress through multiple cryo-EM data collections. The structural data presented in Chapters 4 and 5 was from 3 combined data sets.

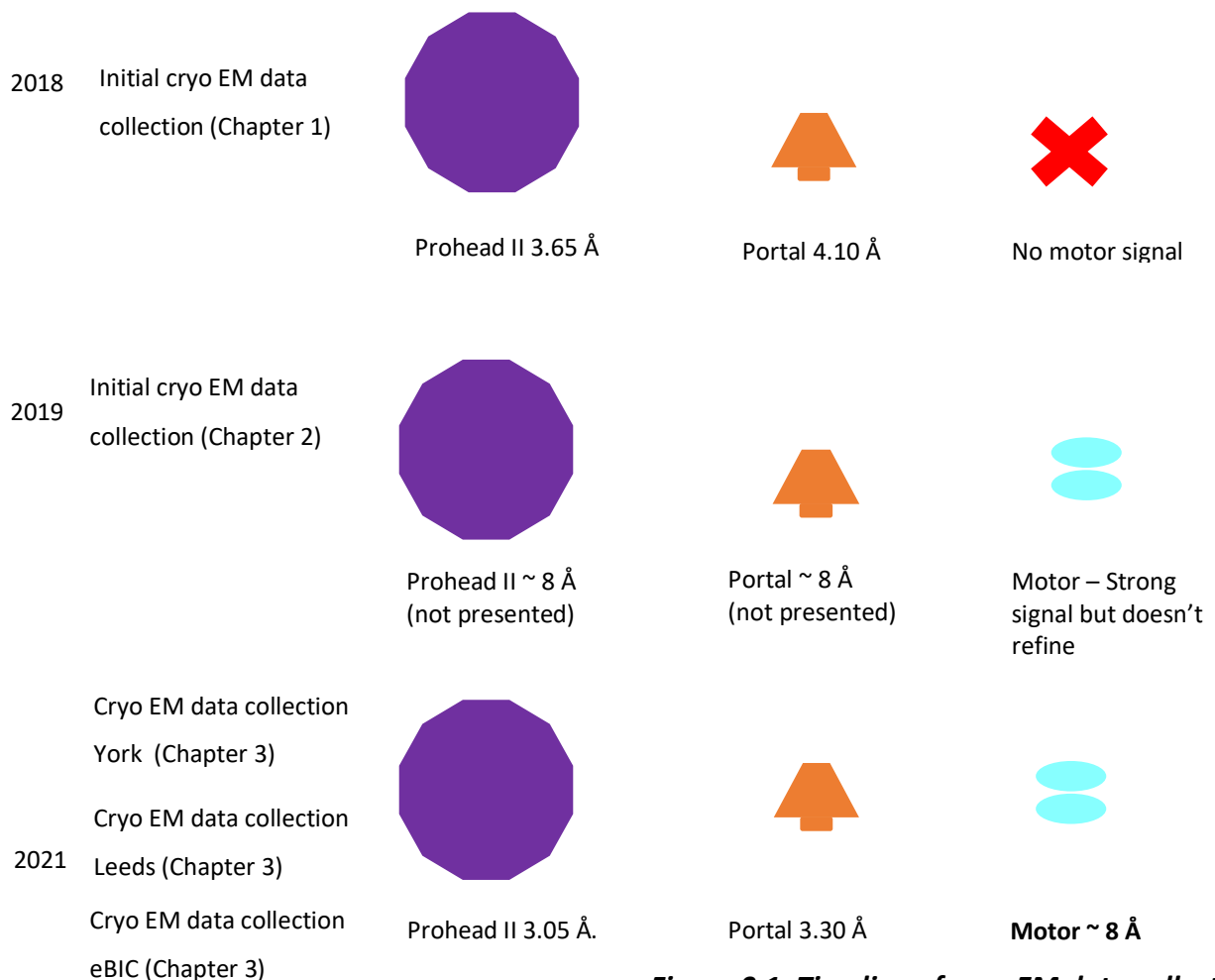
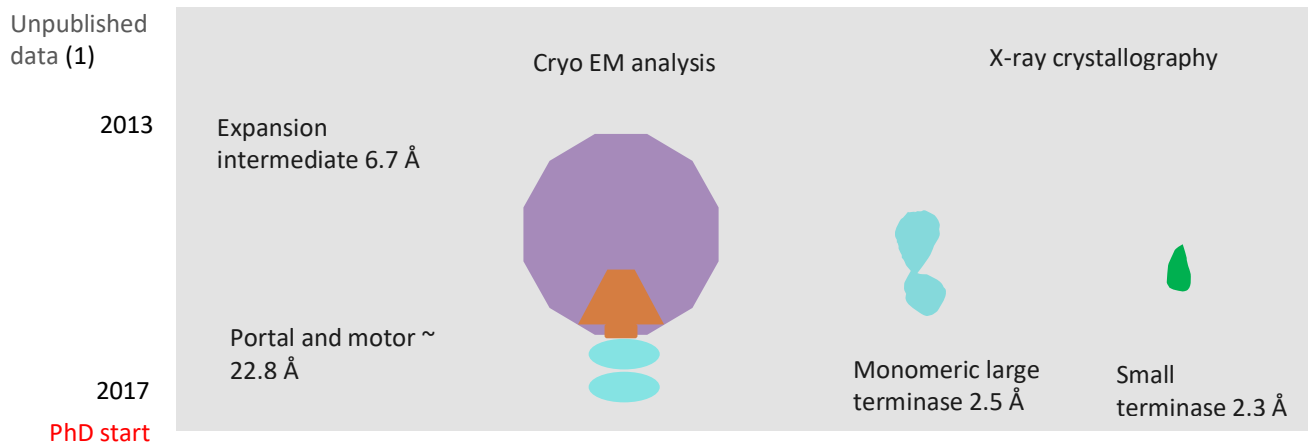


Figure 0.1: Timeline of cryo-EM data collections

Chapter 1: Introduction

Virus introduction

Viruses are microscopic parasitic agents. The number of viruses in the ecosystem of our planet is estimated to be in the range of 10^{31} , with 10^{23} infection events taking place every second within the oceans (2–4). Almost every known organism has associated viruses: including animals, fungi, archaea, and bacteria. In common with other parasites, the relationship with the host spans from cooperativity to selfishness (5). Whilst lytic viruses destroy the host cell, they are also critical in driving genetic evolution (6).

In their simplest form, viruses represent genetic material protected by a protein coat, or capsid. Some viruses are further shielded by an envelope of lipids, containing virally encoded glycoproteins, which facilitates entry into certain host cells. Viral variation is prolific: the largest known *Mimiviruses* exploit genomes of greater than 1000 kb (7), whilst *Pithovirus sibericum* displays a capsid diameter of 1500 nm (8). This is significantly larger than some bacteria. In contrast, *MS2*, which infects *E. coli*, encodes only 4 proteins, and viral genome sizes amongst *Circoviridae* can be as small as 1 kb, with a capsid diameter of just 20nm (9–11).

The virus life cycle begins with attachment to the host cell, usually mediated by interaction of a viral protein with a receptor. It must then penetrate the cell *via* endocytosis or direct fusion. On entry, the virus sheds its protein shell to release the viral genome. The host replication machinery is subsequently hijacked to replicate the genome, which is in turn translated into viral proteins. New viruses are assembled and released. Exit from the cell may be mediated by lysis of the host cell, or additionally for enveloped viruses, *via* budding (12,13) (Figure 1.1).

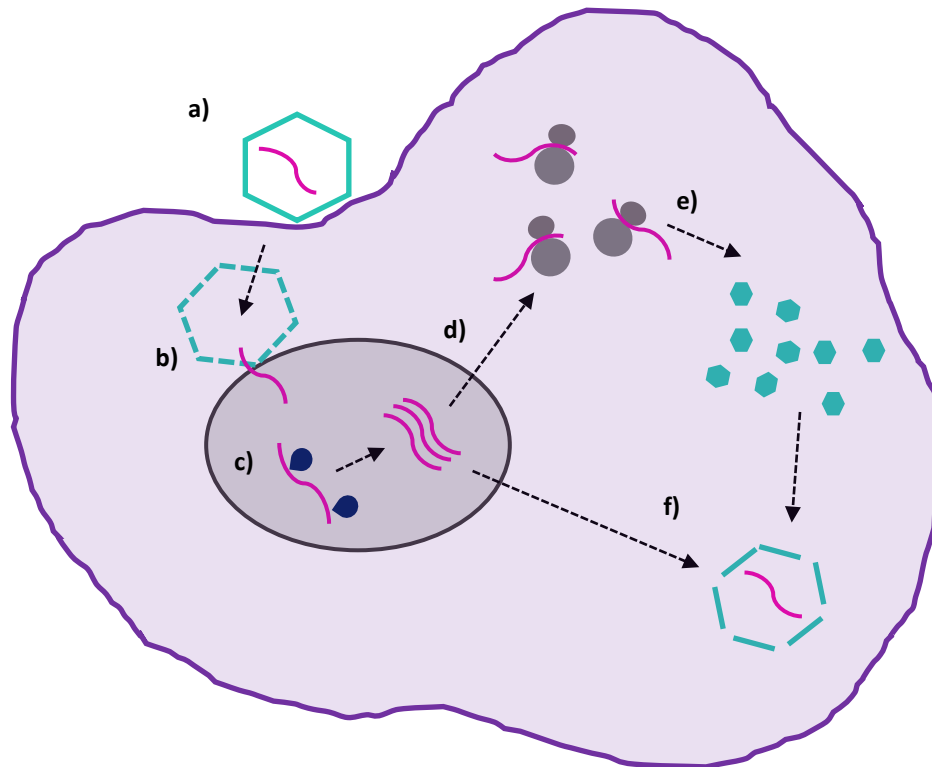


Figure 1.1: The virus life cycle a) the virus attaches to the host cell b) on cell entry the viral genome is released c)d) the host cell machinery is hijacked to produce copies of the viral genome and mRNA for transcription e) host ribosomes make viral proteins f) new virions are assembled for release

Viral genetic material can be DNA or RNA, single-stranded or double-stranded. Double stranded DNA viruses may utilise host or viral DNA and RNA polymerases, for genome replication and mRNA production respectively. These viruses include *Poxviruses* and *Herpesviruses* as well as many phages (which infect bacteria). RNA dependant RNA viruses, like *Flaviviridae*, instead utilise single-stranded mRNA genomes, which can be translated directly. Meanwhile, negative-sense RNA viruses such as *Paramyxoviridae*, encode single-stranded RNA complementary to mRNA. RNA genomes may also be double stranded, or indeed ambisense, as for the *Arenavirida* family, where genomes represent mosaics of positive and negative stranded RNA (14–16). Another unusual class of RNA viruses are the retroviruses, *Retroviridae* (17). These contain 2 copies of single stranded RNA, where replication proceeds through a DNA intermediate. This requires virally encoded reverse transcriptase. In a mirrored system, the *Hepadnaviridae* family encode DNA, for which an RNA intermediate is required for replication (18).

Broadly, genomes must encode three types of proteins. Firstly, viruses must be able to replicate their genomes and produce mRNA to produce other viral proteins. For RNA viruses, RNA polymerase or replicase are therefore essential (19). DNA viruses also tend to encode additional DNA polymerase or replication initiation proteins. Next, the structural proteins are required for the assembly of new viruses, known as virions outside of a host. The final set of proteins, involved in disrupting the host defence mechanism, is less ubiquitous (18).

There are three major hypotheses as to the origin of viruses. The first theory, known as “virus first”, describes early viruses as capable of independent replication (20). This is supported by unique and highly conserved viral genome motifs, known as virus hallmark genes, which are not present in any cellular life forms. This suggests viruses may predate cellular life, emerging from a primordial pool of genetic elements. Extensive mixing would allow for the evolution of divergent genetic information utilised by viruses (21). The second, ‘reduction’ hypothesis instead depicts viruses as a ‘streamlined’ evolution of unicellular organisms. Here, many cellular elements would be lost but genetic material retained (22). Giant viruses such as *Mimiviruses* and *Megaviruses* share many features with parasitic bacteria, and thus support this theory (23–25). Finally, the escape hypothesis links viruses to transferable genetic elements such as plasmids. In theory, this element could have become engulfed by a vesicle during asymmetrical cell division. This theory provides rationale for the embedding of modern day viral genetic material into host genomes (26,27).

Bacteriophage introduction

Bacteriophage were first recognised as agents which replicate and kill bacteria in the early twentieth century 1915 by William Twort, and in 1917 by Felix d'Herelle (28,29). Subsequent studies with bacterial plaques allowed a life cycle of infection, latency, and lysis to be established (30–32). From the 1930s, bacteriophage were studied using electron microscopy to reveal polyhedral heads and tubular tails (33,34). During this time, several *E. coli* bacteriophage acted as tools for much molecular biology advancement, enabling ground breaking research. This included deciphering the genetic code, the discovery of messenger RNA, and understanding of restriction enzymes (35–38).

Bacteriophage are wildly abundant and diverse. Indeed, 100 million phage species are predicted to inhabit the planet and each bacterium is expected to host approximately 10 unique phage (39–42). Roughly 96% of studied bacteriophage belong to the *Caudovirales* family: the tailed bacteriophage. This family is further classified by tail structure including *Myoviridae* which show long contractile tails, *Siphoviridae* which use long non-contractile tails, and *Podoviridae* which display short tails (Figure 1.2)(43–45). The remaining phage may instead display a “cubic,” filamentous, or pleomorphic shape (46) while several phage have internal constituents such as lipids in envelopes. Multiple novel morphologies are also newly being discovered for archaeal phages (47–50) .

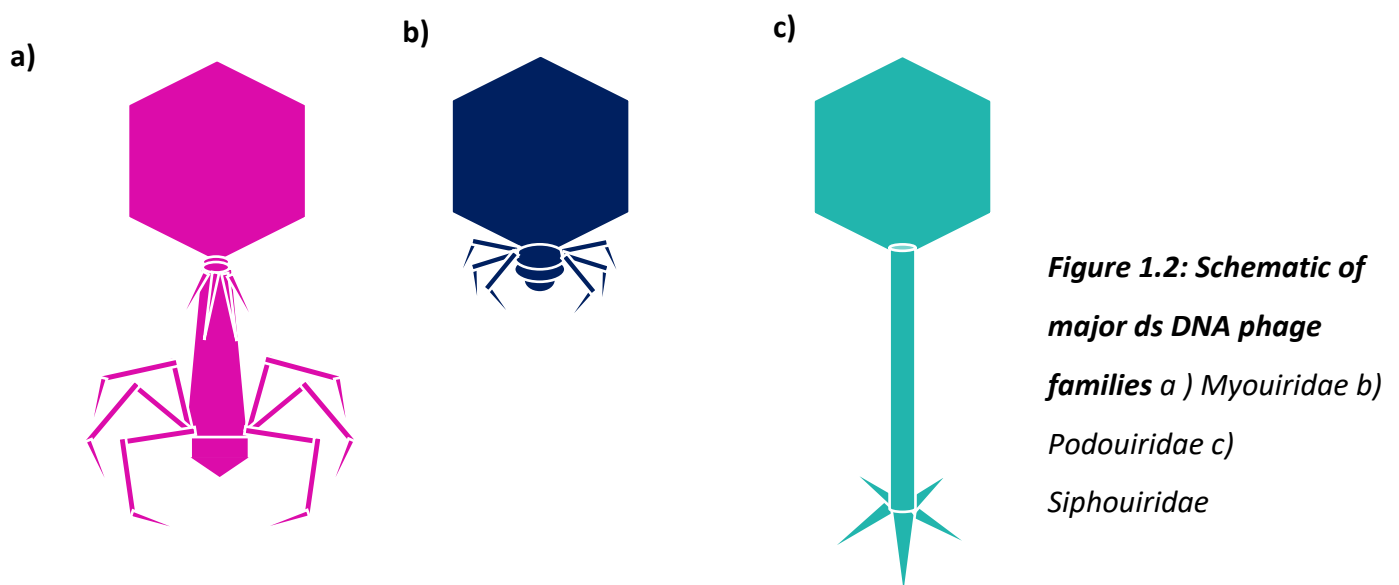


Figure 1.2: Schematic of major ds DNA phage families a) Myoviridae b) Podoviridae c) Siphoviridae

Bacteriophage infection

The *Caudovirales* life cycle begins as a virion, external to the host. Infection is mediated by the tail, which is responsible for both locating and binding to the host receptor. This confers specificity to the phage (51,52). Receptor binding proteins may reside in tail spikes or baseplates, and irreversibly bind to a specific protein receptor on the host membrane surface. This instigates a major irreversible conformational change which is relayed through the tail. The highly compact viral genome is now ejected through the bacterium membrane via the tail (53), which may fuse directly with the membrane, or instead interact with host cell membrane channels. For both gram-positive and gram-negative bacteriophage, some tail substructures also harbour peptidoglycan degrading activity (54,55).

Bacteriophage show divergent genomes with most utilising double-stranded DNA (56), but a minority adopting single-stranded DNA, single stranded RNA and more rarely still double stranded RNA. The genome length can vary between approximately 1 kb and 1 Mb (57,58) and phage may employ lytic or lysogenic life cycles. Lytic phage, such as the tailed phages T1- T7 (59), rapidly infect and kill their hosts. For instance, T4 encodes its own DNA polymerase for genome replication, as well as transcription activator proteins which bind to viral genome promoters to regulate transcription by the host RNA polymerase. Promoters are activated in distinct early, middle and late phases to regulate viral protein expression with the phage life cycle (60,61). New virions may then be assembled and released from the cell by lysis.

In contrast the genetic material of lysogenic phage remains dormant within the host, either integrated into the host genome (known as a prophage), or as a separate plasmid (62). Integrated genetic material may then be replicated as part of the bacterial genome whilst extrachromosomal viral genomes must include genes for plasmid inheritance to be adopted by host progeny. For temperate phage, such as λ , lysogenic phage may switch to lytic life cycles by the process of induction. This is thought to be a survival strategy employed by the phage when the host bacterium is at risk of death (63–65). Following induction, λ phage employ the host DNA polymerase for rolling replication of its circular genome and modulate the host RNA polymerase by transcriptional anti-termination. Two separate phage encoded

proteins bind to RNAP to prevent termination of transcription in the early and late phases of transcription (61,66). In this way temperate phage may exploit the benefits of both infection strategies.

Bacteriophage assembly

Viral assembly is remarkably well conserved among the double stranded DNA bacteriophages. Early electron micrographs show the insertion of DNA into preformed protein proheads, which consist of numerous capsid protein monomers, and often additional scaffolding proteins. These assist coordination for self-assembly. However, several phage, including HK97 instead utilise the N terminal portion of the major capsid protein for this role. In order to digest scaffolding proteins after assembly, many phage encode proteases. Digested scaffolding proteins may then be released through pores in the prohead. Unusually, both Φ 29 and P22 release scaffold proteins intact (67)(68).

DNA enters the prohead through a dodecameric portal protein at a unique pentameric vertex. This represents the site for both prohead assembly and later for tail attachment. DNA packaging triggers the maturation into expanded, DNA-filled capsids which show thinner shells, more pronounced vertices and a larger diameter despite displaying the same number of viral coat proteins (58,69,70) . Figure 1.3 depicts a comparison of empty HK97 proheads and mature phage visualised by negative stain electron microscopy.

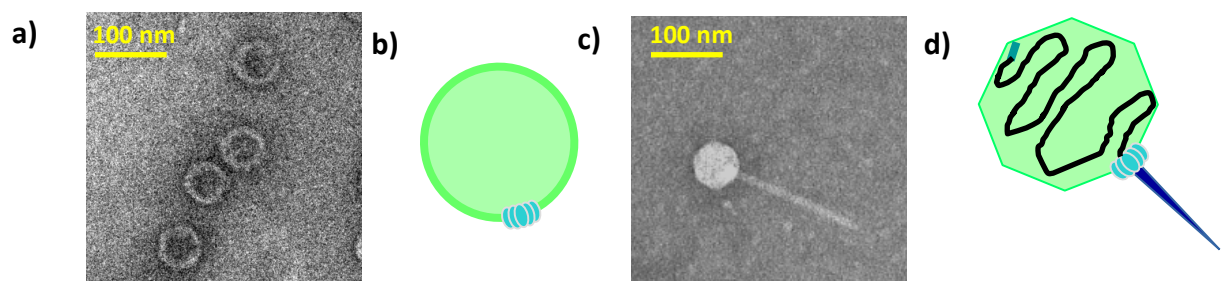


Figure 1.3: Expansion states of phage HK97 a) negatively stained electron micrograph of HK97 proheads b) diagram of prohead depicting capsid proteins (green) and portal protein (blue). c) negatively stained electron micrograph of HK97 virion d) diagram of virion depicting tail (dark blue), portal protein (light blue), capsid (green) and DNA (black).

The DNA packaging machinery of double stranded (ds) DNA phage is known as the terminase complex. This recruits viral DNA to the immature prohead where it docks on to the portal (71). A ring of five large terminase subunits represent the DNA motor, which translocates the genome into the prohead in an ATP dependant manner (72–76). Small terminase subunits have been hypothesised to form a second ring of variable subunit numbers beneath . After packaging is completed, the terminase cleaves the DNA, dissociates, and the complex is replaced by neck proteins to prevent the release of genetic material. This is followed by attachment of a pre-assembled tail or tail proteins (77–79)(Figure 1.4). Now the phage can be released from the cell, via lysis, as a virion.

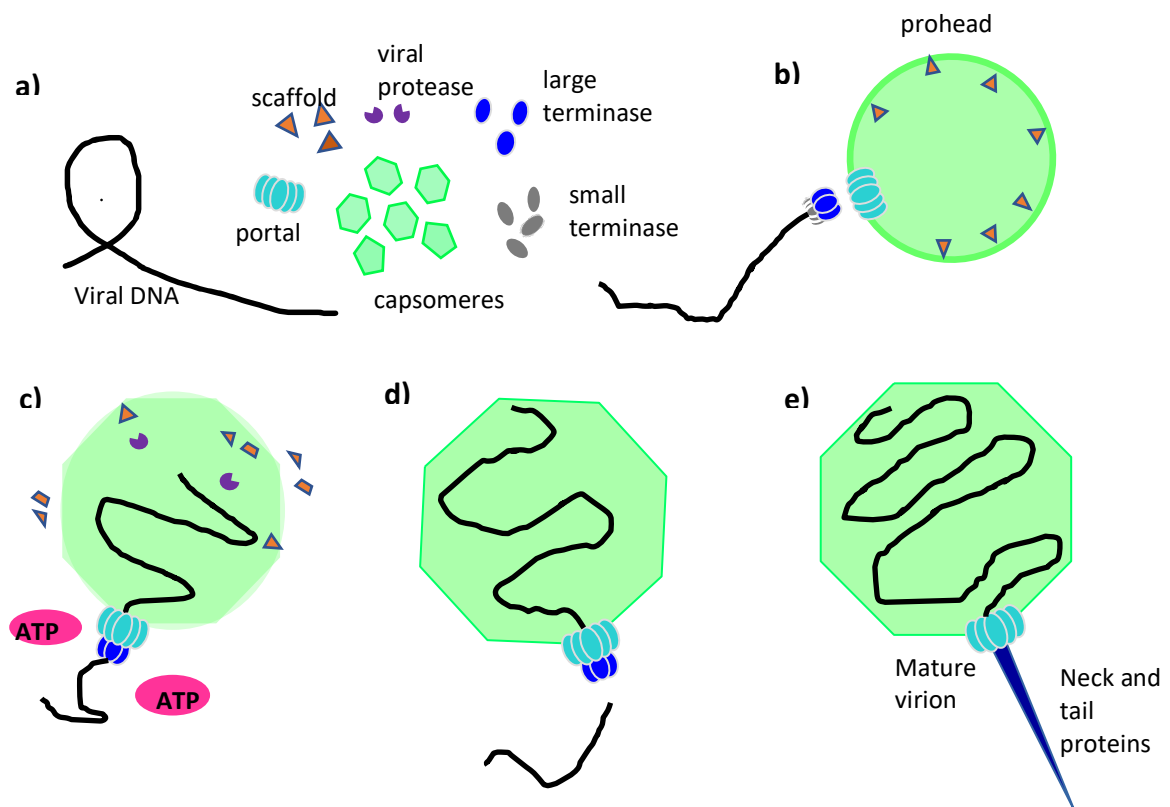


Figure 1.4: Bacteriophage assembly pathway a) viral DNA is replicated by the host and translated into viral proteins b) proheads self-assemble around portal and scaffold proteins, while small terminase recruits large terminase to the viral genome c) large terminase powers DNA packaging through the portal, and scaffold proteins are digested and released d) the prohead expands into a mature capsid and DNA is cleaved after one genome length has been packaged e) the terminase complex dissociates and is replaced by neck and tail proteins to form a mature virion

Bacteriophage structure

Approximately 75% of studied tailed have icosahedral structures (56). This structural arrangement, coined by Caspar and Klug (the CK theory), describes 20 triangles which form the icosahedron faces. 5-fold rotational symmetry is exhibited at each vertex with two-fold rotational symmetry along each edge. The triangle may be divided into three equivalent parts, so that the centre of each triangle marks a 3-fold symmetrical axis (80).

Here, the triangulation number, T , is equivalent to the number of major capsid protomers in an asymmetric unit. The total number of capsid proteins in an icosahedral capsid is equal to T multiplied by the number of asymmetric units, 60. Subtracting one pentamer from the unique portal-containing vertex, the number of capsid proteins in the capsid becomes 5 less. T numbers in viruses studied range from 1 to 52 (81), with the capsids of many well-studied phages including λ , HK97, P22, Sf6 and SPP1 have a T number of 7, giving a total of 415 subunits (82–85) (Figure 1.4). A minority of phage instead adopt a prolate head, where hexamers are used to elongate the capsid along one fivefold axis. Here the T number refers to the “capping” icosahedron, and the Q number is for the extension. Now the total number of subunits equates to $(30(t+q) - 5)$. This assembly is adopted by bacteriophages $\Phi 29$ ($T = 3$ and $Q = 5$) and $T4$ ($T = 13$ and $Q = 21$)(86–88).

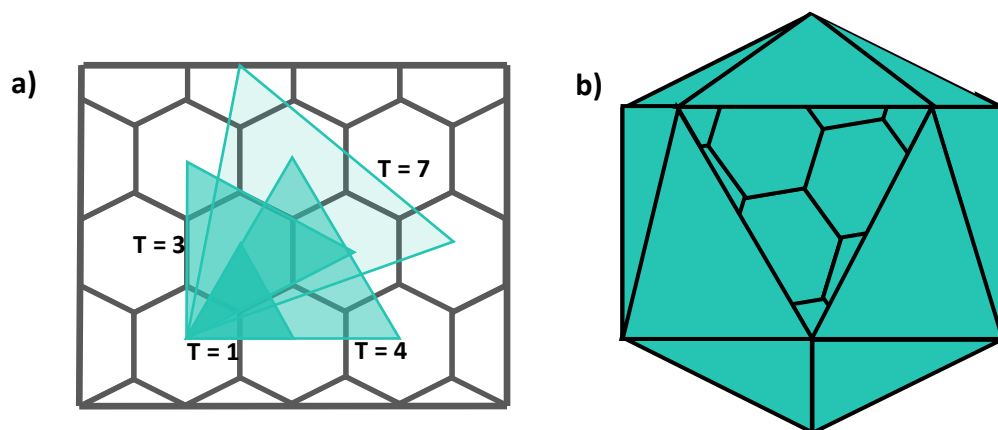


Figure 1.5: Depiction of Caspar Klug theory of the icosahedral phage architecture a) Triangular number depiction for icosahedral phage b) $T = 7$ icosahedral capsid

HK97 fold and capsid expansion

The structure of the bacteriophage HK97 icosahedral capsid was one of the earliest capsid proteins to be determined at high resolution (85). Thus, the fold of the major capsid protein gp5 was fold coined the 'HK97' fold. Subsequently this protein fold was found to be shared not only by all other tailed bacteriophages, but also by evolutionarily related eukaryotic viruses such as *Herpesviridae* (81) and eukaryotic encapsulins. The major capsid protein of HK97 assembles into hexameric and pentameric capsomeres, that can further produce the T=7 icosahedral structure. In the structure of this protein the A domain exists at the capsomere axis and encapsulates a 4-stranded β sheet with 2 helices, while the P domain, at the capsomeres periphery, has a three stranded antiparallel beta sheet against a long helix (89). The full fold, including an additional N terminal arm and Extended loop, is depicted in Figure 1.6 in the conformation present in the mature *HK97* head.

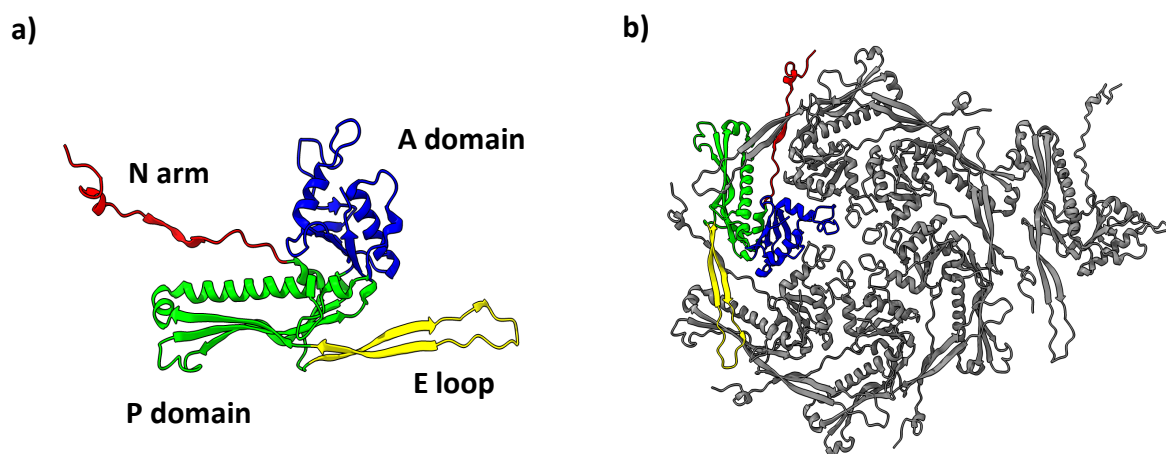


Figure 1.6: Ribbon depiction of the HK97 fold a) individual b) within one asymmetric unit of the HK97 mature head from PDB structure 2FT1

HK97 employs a unique prohead maturation strategy, which unlike other phage does not require the presence of the portal for capsid assembly or utilise a separate scaffolding protein (89). Instead, the major capsid protein, gp5, is cleaved by protease gp4 from 42 kDa to 31 kDa, and a network of covalent bonds interlinking neighbouring subunits is established (91). Transition from prohead I to prohead II requires cleavage, and release of this delta domain of gp5. This delta domain shares many characteristics with traditional viral scaffold proteins: it is essential for capsid assembly, resides within the procapsid and is expelled after cleavage. This domain also has a high alpha helical content and predicted capacity to form parallel coiled-coils. This domain is thought to prevent premature expansion (92).

Proheads exhibit a “skewed” hexamer, consisting of two trimer blocks which appear to have been dislocated by approximately 25 Å (91). Once prohead I is cleaved, Prohead II morphs through a series of three expansion intermediates before assuming a fixed mature structure (93). This structure shows a 60 Å radial expansion, relative to the prohead, and non-skewed hexamers with approximate 6-fold symmetry (94). Expansion from prohead II can be stimulated *in vitro*, by acidification to pH 4, transitioning through expansion intermediate EI-I, EI-II, EI-III, and EI-IV, representing an 86% increase in internal volume. Neutralisation of pH then drives the transition to mature Head I, and crosslinking between neighbouring subunits stabilises the mature capsid Head II (93). Crosslinking occurs spontaneously and is facilitated by conformational changes during expansion. Glutamate 363 is thought to catalyse amide bond formation between lysine 169 and asparagine 356, forming the isopeptide bond between separate subunits at the local 3-fold axes of the capsid (95). Valine 163 is also essential and contributes to a hydrophobic pocket surrounding these three amino acids (96).

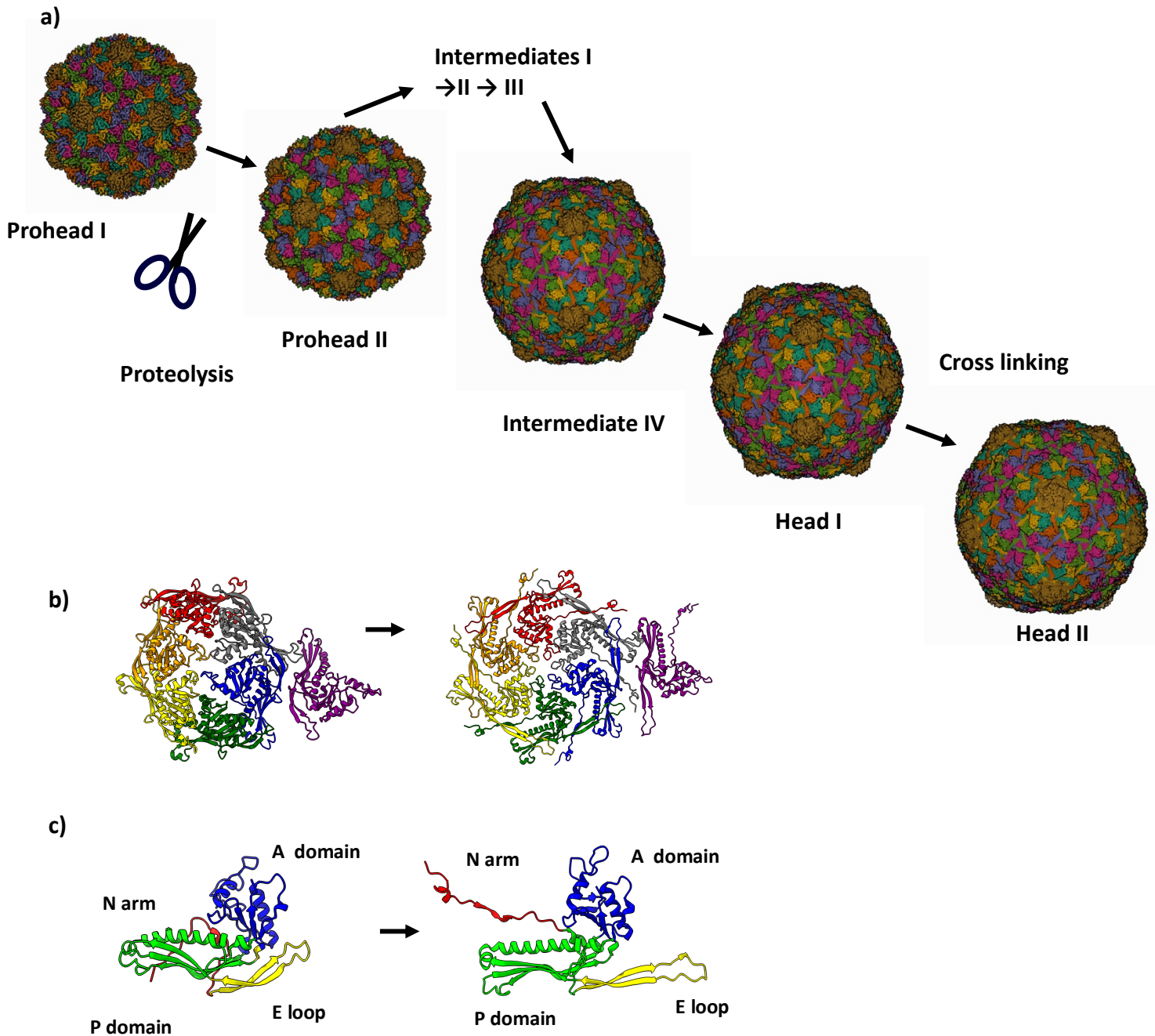


Figure 1.7:HK97 capsid expansion pathway a) Expansion of the HK97 immature prohead to mature head through PDB structures 3QPR, 3E8K, 2FTE, 2F53 and 2FT1 b) Rearrangement of capsomeres from prohead II to head II c) Adaptation of the HK97 fold from prohead II to head II

Structures have been solved for each intermediate within the expansion process (Figure 1. 7 a). Structural analysis suggests that expansion is driven largely by rigid-body rotations and translations of the subunits. This is supported by H/2H exchange which shows that all expansion intermediates maintain intercapsomere integrity at all local and icosahedral 3-

fold axes. Refolding within subunits is confined to two small regions: the N-terminal arm and the E-loop (Figure 1.7 b,c). Scanning calorimetry has shown that while the prohead structure represent an energy minimum, expansion is driven irreversibly to completion by proteolysis and crosslinking which act as 'locks' (97) .

Genome packaging strategies

Although broad mechanisms of packaging are conserved among the dsDNA bacteriophages further classifications can be highlighted based on processing of the viral DNA. Most dsDNA bacteriophage DNA is replicated as long genome multimers, known as concatemers (98). Packaging occurs in a series of unidirectional packaging events along the concatemer, where cleavage of the DNA substrate at termination, produces a fresh dsDNA end for a subsequent initiation event. Packaging series typically consist of 2 to 5 full genome packaging events but may elongate dependant on infection conditions (99).

Replication of the phage Φ 29 DNA unusually produces unit genome length copies, where each 5' end is covalently bound to viral protein gp3 (100). This complex is then packaged by the motor, which shows large terminase linked to the portal via a unique RNA hexamer (101,102). The prohead RNA (pRNA) is necessary for both packaging initiation and substrate selectivity (103,104)

Pac viruses include P22, P1, SPP1 and T4, and packaging is initiated from a '*pac*' (packaging) site within the genome (105–108). For SPP1, two separate small terminase oligomers are thought to bind DNA at sequences *pac* L (100 base pairs) and *pac* R (30 base pairs)(109). This loops DNA between binding sites, exposing a third sequence, *pac* C, to cleavage by large terminase. *Pac* virus termination occurs via a 'head full' mechanism (110), whereby pressure changes within the capsid are relayed to the terminase. This instigates the ATPase to nuclease activity switch, required for DNA cleavage. In practice, a small amount of redundant DNA is always packaged: SPP1 virions harbour approximately 103% of the genome length (111).

Cos viruses, such as λ , HK97 and P2, instead package genomic DNA units (112,113). Upon bacterial infection, the linear genome is injected into the cell, where complimentary sticky ends are sealed by host ligases. This forms a circular viral genome, which serves as a template for rolling circle replication. The site of adhesion is known as the *cos* site. This sequence is recognised by small terminase, which recruits large terminase to instigate a staggered, sequence-specific cut, producing complimentary overhangs with either a 5' or 3' extension (114).

For λ phage, the core element within the 200 base pair *cos* sequence is *cos* N, where large terminase produces nicks separated by 12 base pair overhangs (115–117). Downstream element *cos* B is also required for efficient initiation of packaging (116,118,119). Packaging termination occurs at the next *cos* site, where large terminase once again cleaves the *cos* site. Efficient termination is reliant on upstream element *cos* Q (120). Termination is also proposed to encompass a headful element, as the efficiency of cleavage is highly dependent on the length of DNA packaged (121). Figure 1.8 summarises these three major packaging strategies.

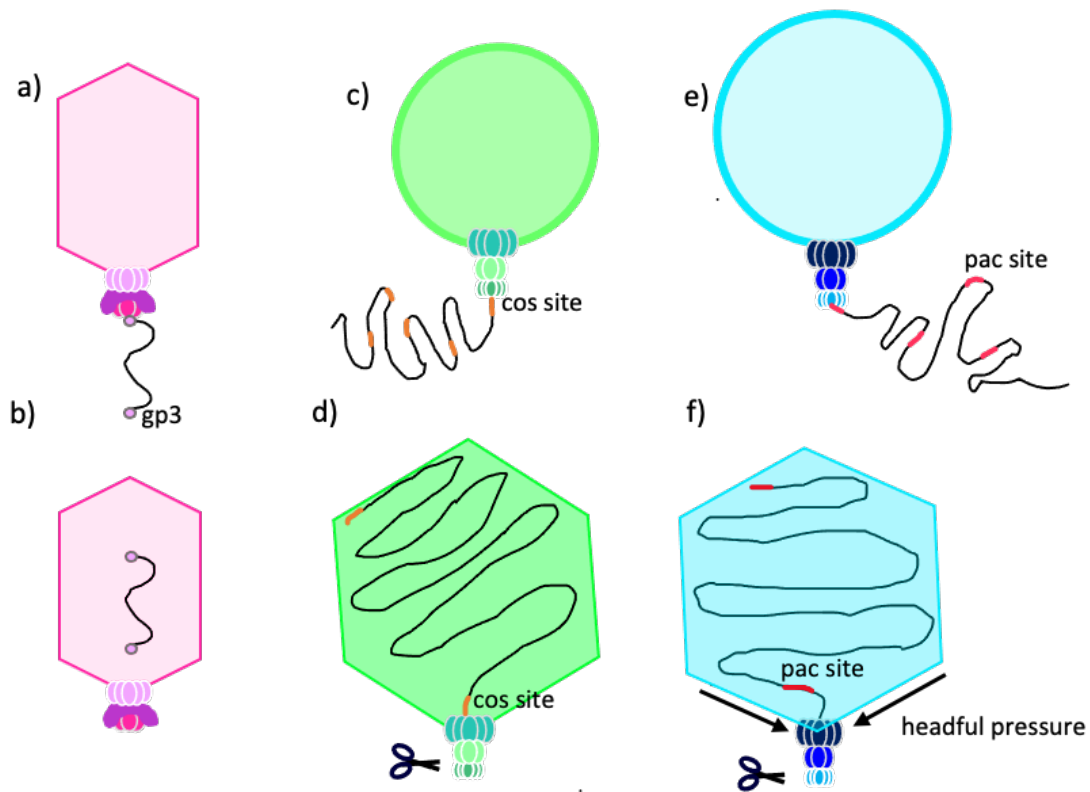


Figure 1.8: Different ds DNA packaging strategies between bacteriophage a) b) Φ 29 packages single unit genome lengths capped by gp3 c) d) cos bacteriophage package from one cos site to the next e) f) pac bacteriophage package from a pac site to just downstream of the next cos site

Portal proteins

The portal protein exists at the unique prohead vertex and acts as a door in to the prohead. Portal proteins show highly divergent sequences and molecular weights. However, in the capsid they are always assembled as a dodecameric ring (122–124), with a negatively charged central lumen for the smooth passage of DNA. The monomeric portal protein structures share homology: containing a clip, stem, wing, and crown domains with a similar fold. The wing domain shows the widest structural variation, and coordinates contact with the capsid proteins. Meanwhile, the stem lines the internal channel, and the crown interacts with packaged DNA. The clip domain makes contact with the packaging machinery during assembly, and later the adaptor proteins for tail attachment. An additional barrel domain is found above the crown within the portal proteins of many P22-like *podoviruses* (125). Portal incorporation is critical for native capsid formation in phages T4, SPP1, and Φ 29 (126,127). Figure 1.9 shows the archetypal T4 portal structure (128).

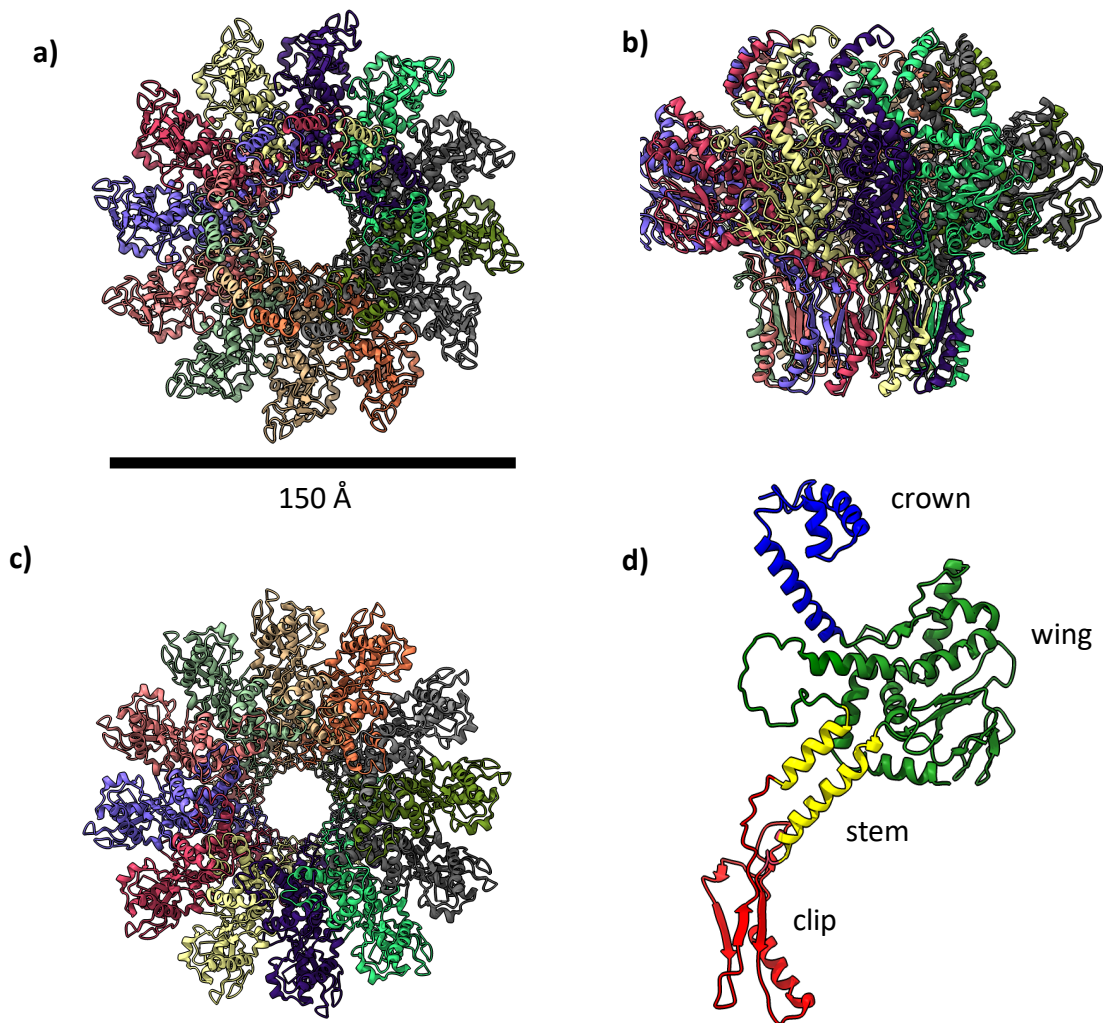


Figure 1.9: Ribbon depiction of the T4 portal structure 3JA7 **a)** viewed from outside the capsid **b)** viewed from the side **c)** viewed from inside the capsid **d)** monomer showing conserved portal fold

Portal proteins exhibit remarkable plasticity, which is proposed to facilitate symmetry mismatches throughout phage assembly. For instance, a recent structure of the PaP3 portal displays a “corkscrew” architecture: employing helical, rather than rotational quaternary structure (129). Meanwhile, the structure of the T4 unique portal-containing vertex displays minimal structural deviation within coat proteins, relative to those surrounding other pentameric vertices. However, dramatic structural variation within the portal N-terminus permits similar interactions with the capsid, in spite of the different environment each subunit experiences (130). *In situ* structures of portals in Φ 29, P23-45 and P68 show poorly

resolved N-terminal whiskers (131–133). This suggests this region does not conform to 12-fold symmetry and potentially morphing of the N terminus to interact with coat proteins is ubiquitous. Alternatively, this region may be disordered for some phage. For *pac* phage P22, the mature phage portal shows twelve-fold symmetry and negligible affinity for pentameric large terminase. Meanwhile, in the crystal structure of preformed portal rings, the clip domain appears to exhibit an asymmetric quasi-fivefold symmetry which shows a strong binding affinity for large terminase (128). While this could be a result of the non-native environment of the protein, it undoubtedly demonstrates the plasticity of the portal. These preformed portals also show a tendency to be incorporated into proheads over mature p22 capsids indicating biological relevance (134).

Packaging termination is thought to be instigated through high internal pressure which increases as DNA fills the capsid. This pressure change is relayed to the large terminase through the portal protein. Indeed, mature phage P22 shows DNA spooled round the portal, which modelling indicates is incompatible with the procapsid portal form (128). The role of the portal as a DNA pressure sensor is supported by the over-packing phenotype resulting from several discrete mutations within the portal core of P22. Typically, the portal exhibits an increase in density and reduction in volume with the progression of DNA packaging. However, biophysical studies indicate that each overpackaging mutant alters these normal volume fluctuations and causes the portal to remain in a more procapsid like form. Thus, the structural change within the portal is not propagated and packaging is not terminated (135). Mutations within λ phage portal protein core also curtail termination (136).

This role in signal transduction is supported by divergent portal structures between cognate procapsids and mature heads. For instance, in the mature head of T7, the portal is compressed outward from the capsid centre, reducing the total length by 12 Å and exposing the clip domain. This is facilitated by a rigid body 25° rotation of the wing domain (137). Meanwhile, Φ 29 shows a 16 Å reduction of the diameter of the clip domain, and a comparable expansion at the top of the wing domain (133).

The portal protein also plays a key role in preventing DNA leakage. P22 mature head portal shows a dramatic decrease in the channel diameter from the procapsid form, from 40 Å to

25. Å(128,138,139). This facilitated by a 10° increase in the angle of tunnel helices towards the central axis. A similar phenomenon is apparent for Thermophage P23-45, where the *in situ* prohead structure displays a constricted channel relative to the crystal structure. Here, channel loop conformations are inverted, reducing the diameter to just 14 Å and subsequently altering the nature of the channel interface from hydrophilic to hydrophobic. This is accompanied by an upward movement of the crown domain (140). In addition to these broad structural rearrangements, portal proteins also display mechanisms to prevent DNA slippage during packaging. Pairs of charged residues within the clip region (K200 and K209 for Φ29 and K331 with K342 for SPP1 (141) are thought to act as a DNA clamp. Meanwhile, single particle studies on T4 indicate that the portal prevents DNA release during motor slipping or stalling (142).

Small terminases

Most dsDNA phage employ small terminase proteins for recognition of the viral genome (143–146). The structure of numerous small terminases have been determined to reveal a central oligomerisation domain joining the N-terminal DNA-binding domain and C-terminal large terminase binding domain. Figure 1.10 shows 3 examples.

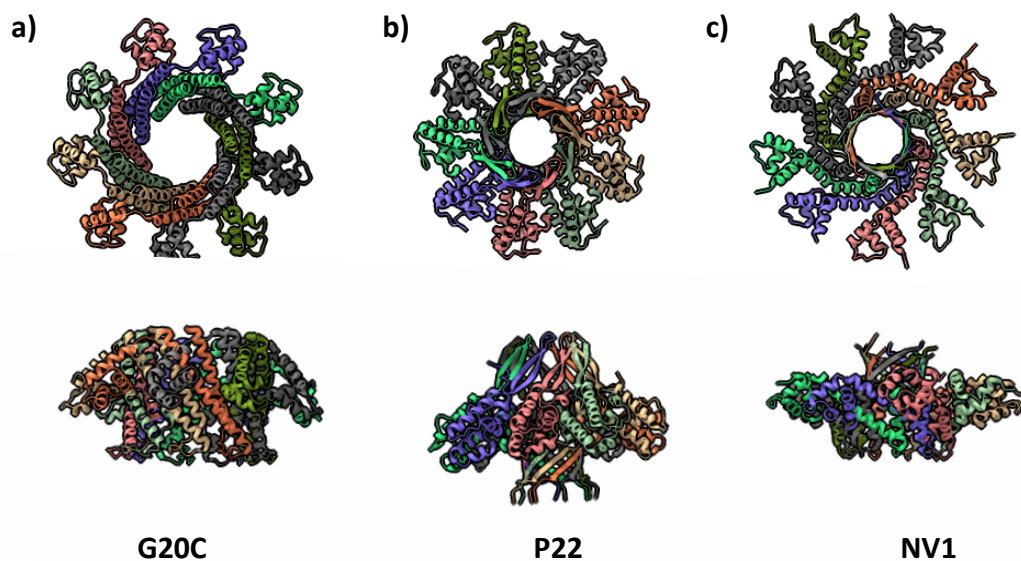


Figure 1.10: Ribbon diagram of small terminase nonomers from the PDB a) 4XVN from phage G20C (151) b) 3P9A from phage P22 (149) c) 7JOQ from phage PaP3 (256)

Small terminases have been shown to stimulate large terminase packaging activity (147). For instance, binding of homologous T4 small terminase (gp16) to large terminase (gp17) is proposed to stimulate a conformational change which repositions residues within the catalytic pocket (147,148). Numerous sites of interaction between the two proteins have been identified, suggesting that multiple weak interactions could facilitate the rapid multimeric assembly and disassembly required for packaging. Meanwhile, P22 small terminase must bind to its cognate viral DNA in order to stimulate large terminase. This depicts an elegant way of discriminating against wasteful packaging of host DNA (149). As a transducing phage however, P22 occasionally packages host DNA which shares a homologous sequence to the native phage pac site (150). Interestingly, P74-26 small terminase is shown to stimulate the ATPase activity of large terminase, while simultaneously inhibiting the endonuclease activity (151): presumably preventing the motor from premature termination.

The nature with which small terminase oligomers interact with viral DNA throughout packaging is currently unclear. For instance, a single small terminase oligomer may remain bound throughout packaging, acting as an extension of the DNA channel below large terminase. However, interaction may also be transient: confined to packaging initiation and termination. Here, small terminases could position large terminase at the correct sequence, before dissociating from the complex, and without participating in the ATP-dependent packaging step. DNA may also feed through the small terminase central channel during packaging, or instead wrap around the outside of the oligomeric ring, and numerous oligomers may also be required. SPR and EPR data for Sf6 small terminase favour the notion that DNA wraps around small terminase, as binding is shown to be weak but cooperative (152). Indeed, this small terminase can bind the genome of related phage SPP1 and so specificity is thus thought to be determined more by DNA shape (i.e. intrinsically bent) rather than sequence. This is supported by the structure of the T4 small terminase oligomer (153) and nonameric structure of P74-6 which each show DNA binding HTH motifs arranged radially (151). However, it has been argued that such bending of DNA is energetically unfavourable. Alternatively, a high-resolution structure of the P22 small terminase oligomeric ring features a sufficiently wide lumen for hydrated B-DNA (154).

HK97 small terminase

A structure of the HK97 small terminase has been solved by X ray crystallography to 2.3 Å resolution (Figure 1.11 a). The oligomeric state of the was determined by SEC-MALS to be nonameric. A 9-subunit oligomer could then be modelled by rotation around the crystallographic 3-fold axis to display long helices and an N terminal HTH motif which should form a positive band round the centre of the ring (1).

More recently a cryo-EM structure of the nonameric small terminase in complex with DNA has been determined (Figure 1.11b) (155). The essential DNA binding site was identified by electromagnetic shift assays (EMSA), and DNA density is apparent within the nonameric tunnel. However, the DNA backbone is both tilted and kinked, and appears shifted towards one side of the channel. Here it interacts with positively charge residues, including Lys100, Arg107 and Arg114, from 3 or 4 of the 9 subunits. The orientation of DNA could be resolved to reveal the *cos* cleavage site at the top of the protein (which interacts with large terminase) and downstream nucleotides feed into the small terminase tunnel. DNA also appears to interact with K4, R5, R7 in the first helix of chain A, seemingly creating a second DNA binding site (155).

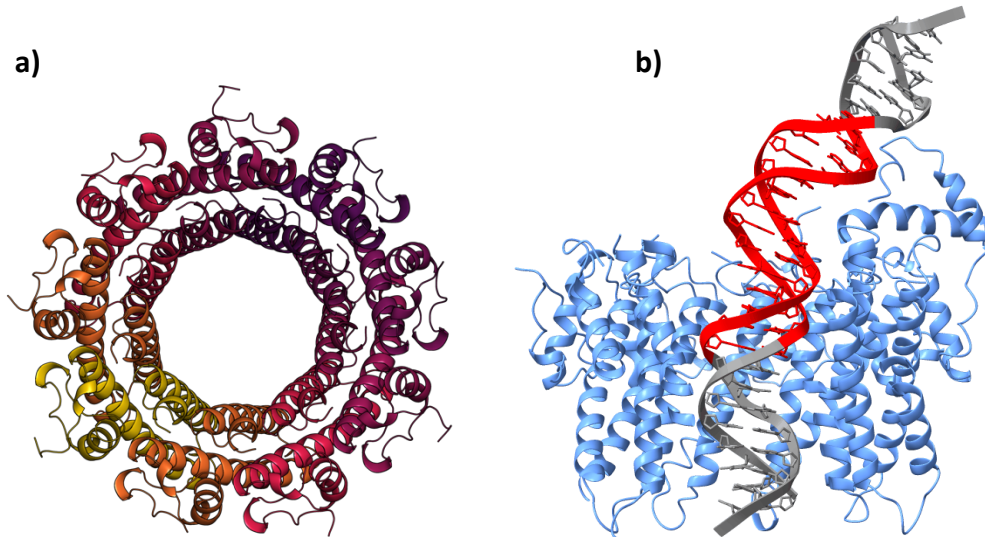


Figure 1.11: Ribbon diagram of HK97 small terminase nonomer determined by a) X-ray crystallography (PDB model 6Z6E (1)) b) cryo-EM in complex with DNA . The *cos* cleavage site is coloured red.

ASCE ATPases

Translocation of DNA into the capsid is powered by the large terminase, that contains an N-terminal ATPase domain (156). This motor is thought to reach forces of up to 100 pN (157–159) and thus represents the most powerful biological machine studied. The terminase functions akin to other ring-shaped, oligomeric translocases: utilising ATP hydrolysis cycles to translocate biological monomers through a central pore (161).

Terminases belong to the additional strand, conserved glutamate (ASCE) subset of P-loop NTPases, which employ conserved Walker A and B motifs for ATP binding and hydrolysis (162–164). The ASCE family utilise a second conserved acidic residue within the Walker B domain, and a β -strand inserted between the Walker A and B domains (165,166). Beta and gamma phosphates of ATP are coordinated by the Walker A domain, while the Walker B motifs coordinates Mg^{2+} (167). This facilitates the conserved catalytic carboxylate to activate water for ATP hydrolysis, which in turn instigates conformational change to translocate the substrate (168,169). Each hydrolysis event instigates a subsequent ATP

hydrolysis event in the neighbouring subunit (72,170). Structural and mechanistic studies of large terminase proteins from different viruses have led to several alternative models of DNA translocation and subunit coordination. However, structural information on complete packaging machines comprising the capsid, large terminase and DNA, are limited and the extent to which mechanisms are conserved in different dsDNA viruses is unknown.

Translocation kinetics of the large terminase packaging motor

Much mechanistic understanding has been drawn from single particle optical tweezer studies on the Φ 29 DNA packaging motor. Here, a DNA substrate was tethered to a micro bead held in place by a laser beam (171–173). When a small external force is applied, as when the DNA is packaged, the force can be measured by the laser beam which applies an equal force in the opposite direction. Broadly, Φ 29 packaging was suggested to proceed through two alternating modes: the dwell phase and burst phase. During the dwell phase ATP binds cooperatively to each subunit around the ATPase ring. During the burst phase, DNA translocation into the prohead occurs in 4 subsequent 2.5 bp steps corresponding to 4 ATP hydrolysis events (170). As packaging of the genome approaches completion, pausing and slipping of the motor occurs more frequently and the burst phase decreases to approximately 2.3bp translocation per hydrolysis event, whilst dwells lengthen. Concomitantly, the rotation of the DNA increases from a rotational density of 1.5 to 5. This compensates for the rotation lost by the additional base pair, which likely repositions the phosphate backbone to make contact with the unique subunit and restart the dwell-burst cycle (Figure 1.12)(174).

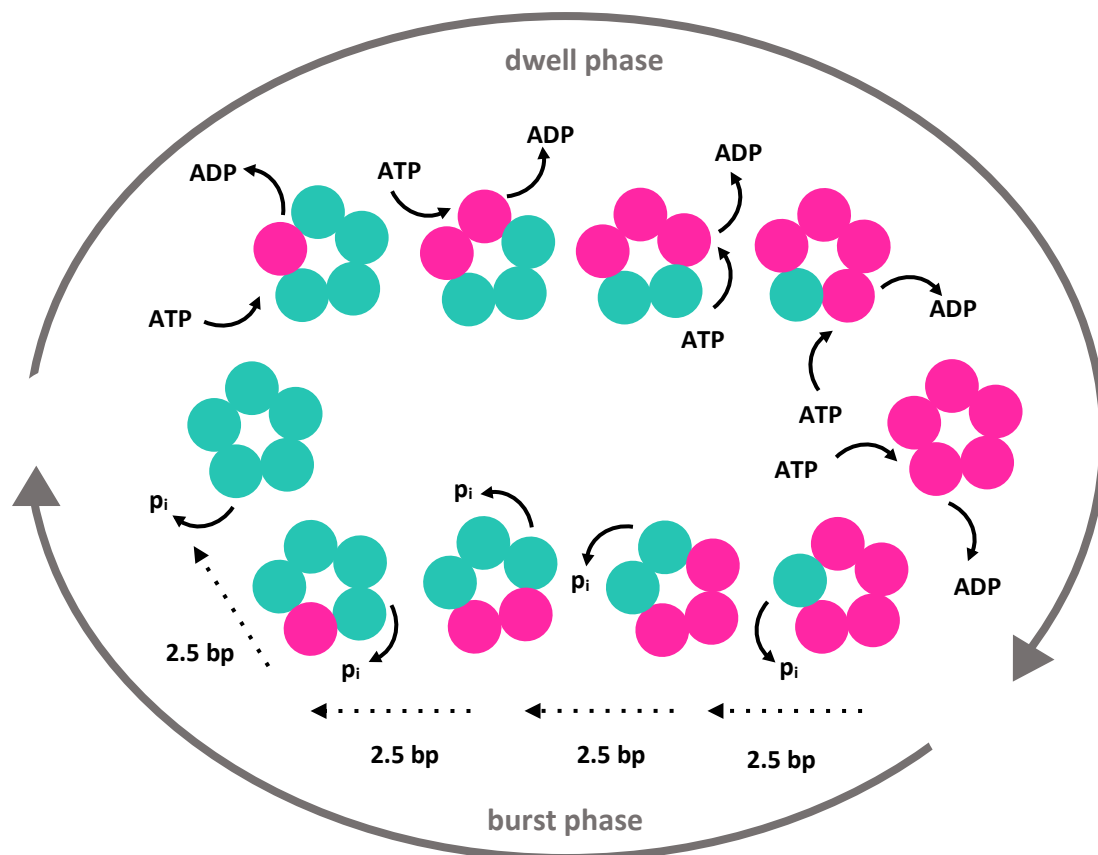


Figure 1.12: Depiction of the dwell burst cycle of DNA translocation by the $\Phi 29$ packaging motor. In the burst phase 5 ATP hydrolysis events produce just 4 steps of DNA translocation, each moving 2.5 pb into the capsid. This is followed by the dwell phase where ADP is replaced by ATP, and DNA remains stationary.

Single molecule studies have also been used to probe T4 and λ DNA packaging machines. For each phage, the rate of translocation, reaching an average of 700 bp/s for T4 (142), is proportional to the size of the genome. This means the total genome packaging time is roughly 2-3 minutes for each. Slowing of packaging, as internal pressure increased towards packaging completion, was also ubiquitous. However, for λ an early pressure peak and drop was also present after approximately 30 % of the genome had been packaged. This pressure is considered responsible for prohead expansion, where the mature capsid displays a twofold increase in volume, relieving some internal pressure (159). For T4, packaging appears especially versatile, with packaging into a mature capsid occurring with equal velocity to the immature prohead (175). Single molecule fluorescence studies have also

revealed that multiple DNA substrates may be packaged into a single prohead in sequential packaging events (176).

Models for DNA translocation by large terminase

A recent high resolution cryo-EM reconstruction of the intact $\Phi 29$ packaging motor, comprising the capsid, pRNA, ATPase and DNA, shows the five large terminase ATPase domains, in a “cracked” helical conformation, stabilised by ATP γ S (177). This contrasts previously determined planar structures of the $\Phi 29$ ATPase (178,179), as well as an ADP bound planar structure of the highly related phage ascc- $\phi 28$ large terminase (180). The necessary transition required between these two states has been modelled by molecular dynamics simulations and inspired a translocation model in fitting with the burst dwell cycle (177,180) (Figure 1.12). In the ATP-bound cracked helical conformation, all subunits tightly grip the DNA. When Subunit 1 hydrolyses ATP it releases DNA, and Subunit 2 is moved up into the S1 plane, moving approximately 2.5 bp of DNA into the prohead. Subunit 1 residue K105 is now poised to trigger ATP hydrolysis in Subunit 2, which in turn loses grip of the DNA and brings Subunit 3 into the Subunit 1 plane. Thus, four ATP hydrolysis steps occur, resulting in four 2.5bp steps of DNA translocation. When all five subunits are present in a planar ring, the pentamer has twice the buried surface area between subunits as the cracked helical state, and thus represents a more stable state. Subunit 5 must then hydrolyse ATP to release the DNA substrate and prime S1 for nucleotide exchange and the beginning of the dwell phase. Here, ADP-ATP exchange again occurs sequentially, allowing each subunit in turn to move down the helix and make contact with the DNA (Figure 1.13). DNA binding is energetically favourable, compensating for the less stable protein conformation, with decreased interaction interfaces between subunits. ATP hydrolysis is thus required to climb out of the energy minima in the transition to the planar state once again.

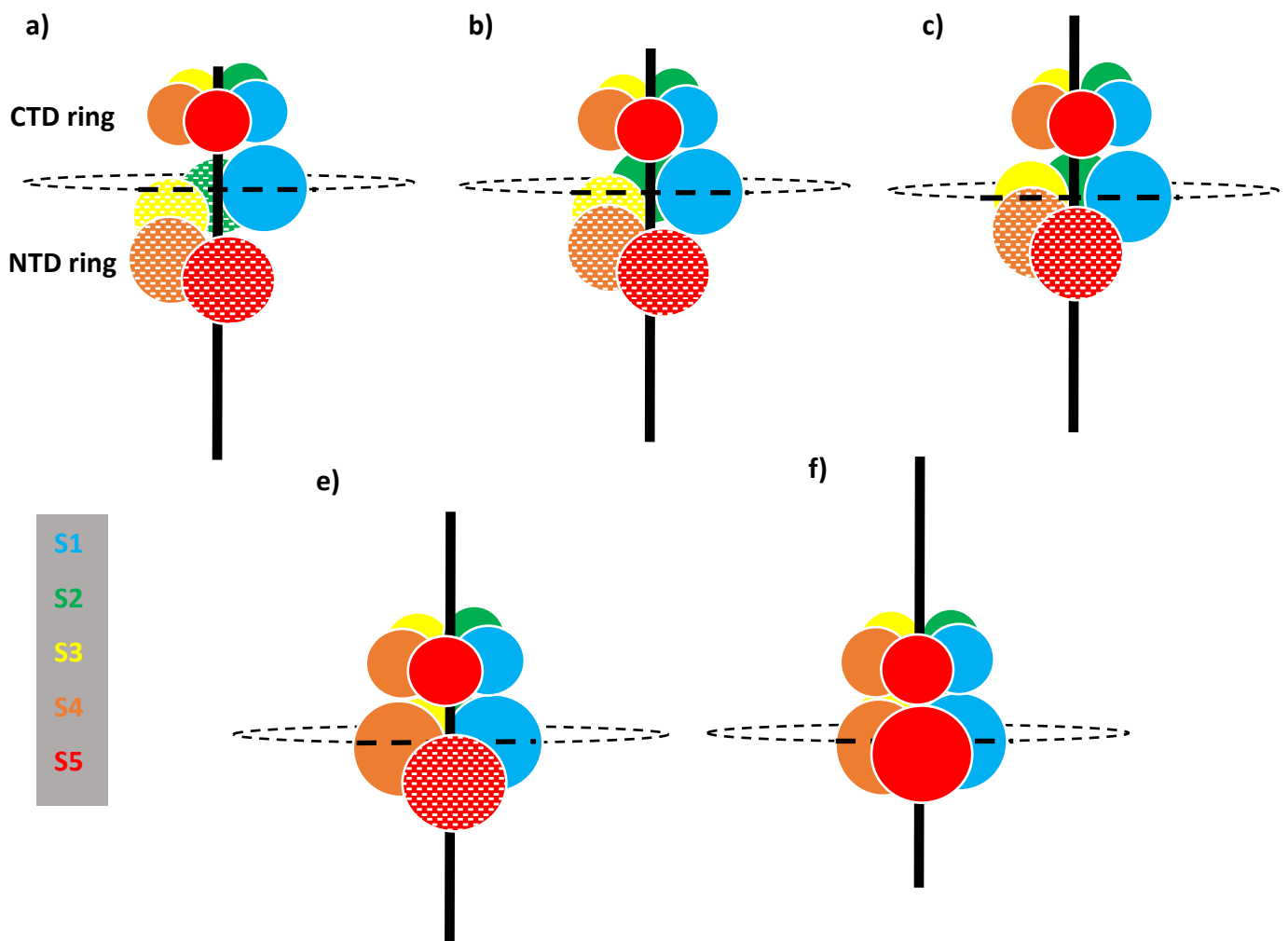


Figure 1.13: Depiction of $\Phi 29$ DNA translocation mechanism. The top ring represents the C terminus domain and bottom ring shows the N terminus domain a) the cracked helix ATP formation b) S1 hydrolyses ATP and S2 is brought into the S1 plane moving 2.5 bp of DNA into the prohead c)d)e) subsequent hydrolysis and translocation events occur for subunits S2 S3 and S4 f) S5 hydrolysis does not produce DNA translocation but resets the motor for sequential ADP-ATP exchange resetting of the cracked helix formation depicted in 1.

Movement between the ATPase and nuclease domains is coordinated by a linker, folded into a 3-helix bundle, which also makes contact with the adjacent subunit. Variations in twist and pitch of the helices allow subunits to adopt the different orientations required for translocation. The ATPase active site exists at the subunit periphery, where ATP appears to be sandwiched between R146 and K105 from the neighbouring (trans acting) subunit. Whilst all five subunits appear to contact DNA, K56 in S2 - S5 is positioned to the track the 5' - 3'

DNA strand. Meanwhile in S1, K56 is closer to the 3'-5' strand, perhaps helping to distinguish a unique role for this subunit in initiating each burst cycle.

For the bacteriophage T4, a crystal structure of the apo form of the large terminase and fully assembled packaging motor solved by cryo-EM are also available (181). The monomer depicts a "tense", more compact state relative to a "relaxed" extended state seen in the low resolution cryo-EM reconstruction. It has been proposed that the ATPase domain lies directly below the portal, with the nuclease beneath, in a contrasting orientation to $\Phi 29$. The nuclease domain is thought to bind DNA in the relaxed state, moving it towards the ATPase domain. This was proposed to induce a conformational change which flips an arginine finger from the same subunit (cis) into the active site, catalysing ATP hydrolysis. In turn the nuclease domain rotates aligning charge pairs producing a 2 bp movement of DNA via electrostatic attraction. This represents the tense state. Release of ATP hydrolysis products triggers the relaxation of the large terminase and positions DNA ready to bind the next subunit, so that at any point just one of the five subunits adopts the tense state (153). The portal binding site has been ascribed to a helix turn helix motif, within the ATP hydrolysis-controlling subdomain II. Interestingly, inhibiting this portal- HTH interaction blocks DNA translocation (182). The cryo-EM structure of $\Phi 29$ and T4 packaging motors are presented in Figure 1.14.

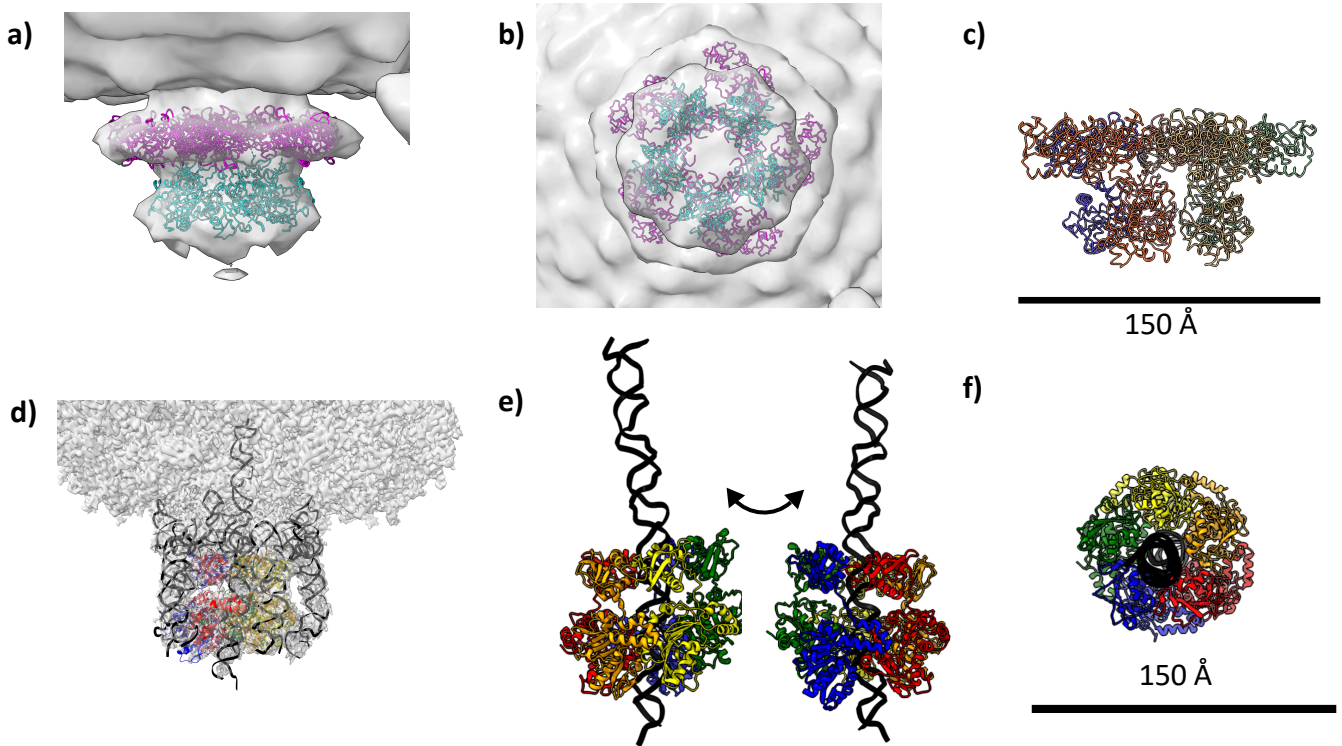


Figure 1.14: Structural models for T4 and Φ 29 ds DNA packaging motors a)b) Crystal structure of T4 large terminase (ribbon model 3EZK) fitted into C5 symmetric low resolution EM reconstruction of the motor (EMD-1573) c) 3EZK alone d) Ribbon depiction of the Φ 29 packaging motor determined by cryo-EM (7JQQ), within the map EMD-6560 e) 7JQQ viewed from the side f) 7JQQ viewed from below

Crystal structures for monomeric forms of large terminases from many other viruses have been determined, in apo, ADP-bound, and ATP analogue-bound conformations. Structural comparisons and molecular dynamic simulations have informed additional models for conserved and unique features of DNA translocation. T4, P74-26 and Sf6 each show the conserved catalytic glutamate within the Walker B domain, and an additional conserved arginine in the Walker A motif. Molecular dynamics indicate that this arginine may act as a ‘toggle’, switching coordination from the active site, to a glutamate within the ATPase lid (or linker domain) on substrate release (183). This is thought to coordinate the dramatic subunit rearrangement seen between ATP analogue-bound and apo structures: an approximate 13° rotation occurs between the ATPase and lid subdomains of P74-26 large terminase. Studies on P74-26 large terminase mutants indicated a trans acting arginine finger (R139) as critical for coordination of the gamma phosphate of the neighbouring substrate.

For Sf6 terminase, a putative nucleotide binding site spans the nuclease and lid domains. A crystal structure with bound ATP γ S shows critical Arginine residue R24 coordinating the gamma phosphate. On ATP hydrolysis and Pi release, R24 instead interacts with and E187 in the lid domain, as shown in the ADP bound structure. Molecular dynamics simulate a 0.5 Å movement of the whole lid domain. This conformational change is proposed to be propagated to the bound DNA, which is subsequently 'pushed' into the prohead (184) .

MD simulations have also indicated a glutamate switch residue as critical to the transition between the ATP bound DNA tight binding state, and the ADP bound state. In each of the four large terminases investigated, Sf6, ϕ 29, and ascc- ϕ 28 and P74-26, a polar or charged switch residue appears to "fix" the catalytic glutamate to point away from the active site rendering it inactive. For each terminase molecular dynamics simulations have been used to calculate two-dimensional energy landscapes of χ 1, χ 2 angles in the catalytic glutamate on ATP binding. In the ATP bound state each glutamate favours the 'active' state. P74-26 is unique in this group in having a polar switch residue (glutamine) rather than charged – this is reflected by a weaker preference for the inactive state which is only metastable in the ADP bound state. MD simulations are supported by crystal structures of terminase mutants: when charged ϕ 29 switch residue Arg53 is mutated to uncharged, the catalytic glutamate favours the active pose in the absence of ATP. Meanwhile, when polar glutamine is mutated to charged Lysine in P74-26 glutamate 150 adopts the inactive pose. This indicates that the active site glutamate conformation is entirely dependant on the switch residue. The switch residue in turn relays structural rearrangement at the DNA binding site to promote binding when in the presence of ATP (185).

Structure of the HK97 large terminase

The structure of the monomeric form of the large terminase from bacteriophage HK97 has been determined by crystallography to a resolution of approximately 2.5 Å (PDB code 6Z6D)(1). The overall fold is classical – showing the ATPase domain to consist of an ATPase core of 8 beta sheet with 6 parallel and 2 antiparallel strands, and a three helix lid. This is adjoined to the nuclease domain by a short linker. The Walker A domain resides between the first and second beta sheets while the Walker B residues are present at the end of Beta 6. There is a potential adenosine binding pocket between Q33 and hydrophobic residues F239 and M27. The structure and active site are depicted in Figure 1.15.

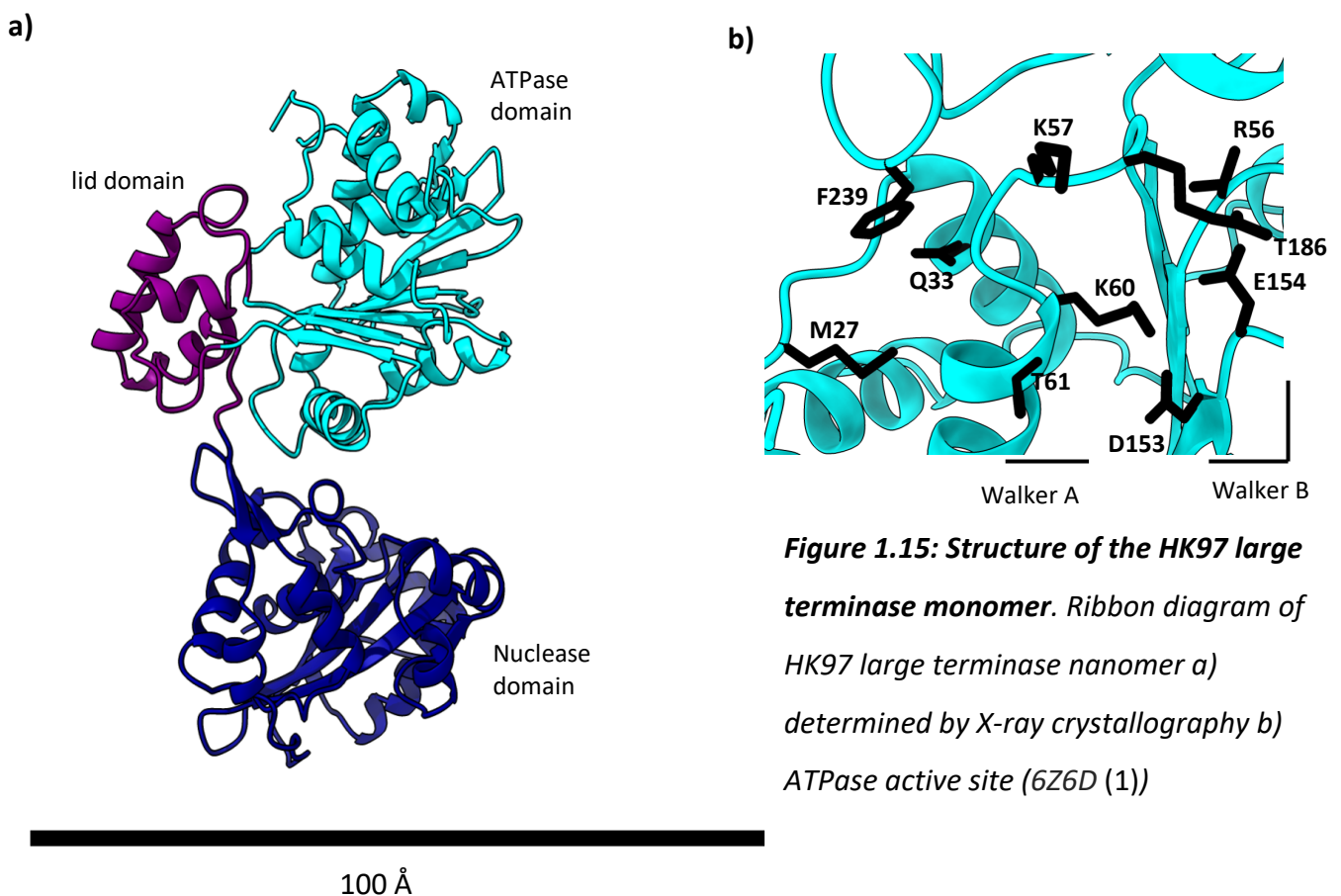


Figure 1.15: Structure of the HK97 large terminase monomer. Ribbon diagram of HK97 large terminase monomer a) determined by X-ray crystallography b) ATPase active site (6Z6D (1))

Mutation studies in combination with ATPase and packaging assays have been used to determine key functional residues (1). E154 adopts the role of the catalytic Walker B glutamate, whilst essential residue K60 acts as the conserved Walker A lysine. Mutation of R56 also produces serious disruption of ATPase activity and packaging. This residue is equivalent to R162 proposed in T4 as a cis-acting arginine finger, whilst in P74-26 the

equivalent arginine is thought to act as the arginine toggle which coordinates conformational change. Mutations of numerous potential arginines did not reveal any likely trans acting arginine finger candidates. In the absence of this donated stimulus, the Walker A residue K57, which is also essential for ATPase activity and packaging, is proposed to provide the additional charge required for catalysis. This lysine is conserved in a number of other dsDNA viruses including T5, herpes simplex and cytomegalovirus.

Structure of the HK97 large terminase nuclease domain

The C-terminal domain of large terminases strongly resemble each other (181,186–190), as well as well as other nucleases including RNases H, topoisomerases, integrases, and DNA/RNA polymerases. Each adopts a 2-metal catalysis approach to cleave the oligonucleotide backbone. However, there is some discrepancy as to the exact mechanisms employed. A structure of Sf6 large terminase suggests that the 2 metals may be brought unusually close (2.42 Å) during catalysis providing a highly positive charge and allowing selectivity of magnesium over calcium (186). However, the structure of G20C nuclease domain indicates a Ruv-c type mechanism (188).

Regulation of the nuclease domain is critical during packaging to prevent early termination. Stimulation is thought to be modulated by the N-terminal domain. Indeed, in the absence of the ATPase domain, the isolated P22 endonuclease domain shows reduced or obliterated activity. An additional extended β -sheet and an auxiliary β -hairpin are seen in terminase nuclease families relative to their RNaseH type nucleases. The hairpin appears to partially block the active site and thus movement could cause rearrangement of the active site access to DNA (189). Such conformational changes may be instigated through linker domain. The signal for cleavage is likely relayed by conformational changes through the portal, as discussed.

The structure of HK97 nuclease domain (Figure 1.16) shows some divergence from this general structure, exhibiting an $\alpha\alpha\beta$ insertion after the anti-parallel section of the β -sheet (1). The active site is formed by residues D302, D388 and D471 found next to the conserved

nuclease hairpin, with K463 coordinating a water molecule. An equivalent lysine in Sf6 is thought to act as a switch for metal binding.

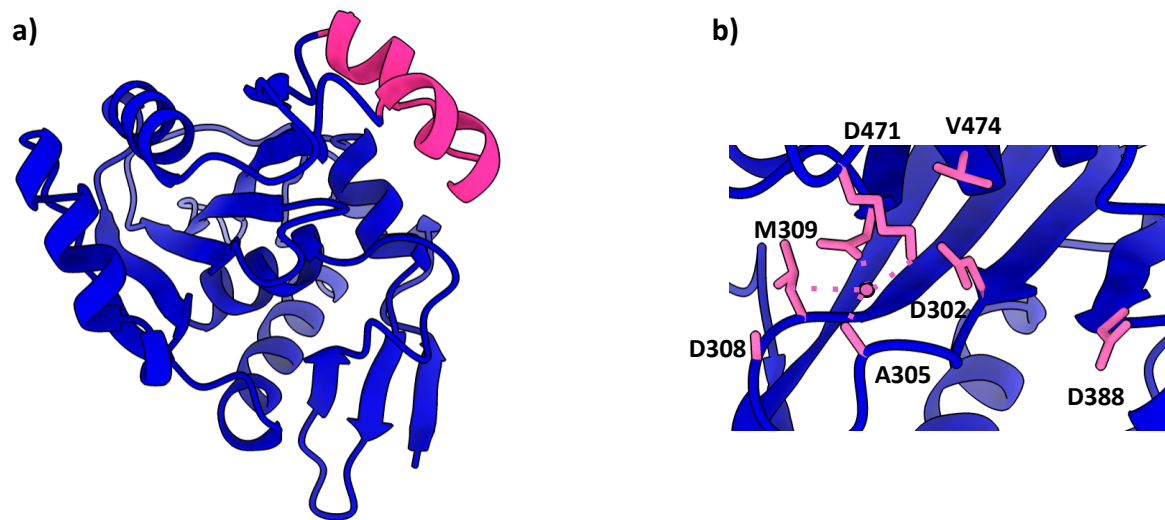


Figure 1.16: Structure of the HK97 large terminase nuclease domain determined by X-ray crystallography (6Z6D (1)). Ribbon model is shown with the additional helix relative to other phage terminase highlighted in pink b) nuclease active site with individual residues labelled in pink

Transmission Electron microscopy

Electron waves hold great potential for probing biological samples, as they have a very small wavelength (2 pm) and can be focussed by electromagnetic lenses. The main limitation of electron waves is their inability to penetrate thick samples, and the inevitable radiation damage to biological samples. Electron doses must therefore be very low, reducing detectable signal, and sample preparation techniques must be employed to protect the sample from damage (191).

Sample preparation for TEM

For negative stain electron microscopy, a heavy metal salt, typically uranyl acetate at 1% w/v concentration, is applied to the sample before blotting the excess (192). The stain dries around the sample forming a cast, creating a negative image. Contrast arises from the differential electron scattering of electrons due to superior electrostatic potential of heavy atoms over the light atoms (C,N,O) present in the biological sample (193,194). This technique quickly enables broad, low resolution, features to be identified, including symmetry, homogeneity, and overall shape, and reconstructions can be resolved to approximately 20 Å (195,196).

For higher resolutions, performing microscopy at cryogenic temperatures helps to significantly reduce sample damage caused by ionising radiation (197). Samples are flash frozen in liquid ethane, using an automated plunger, and maintained at -180 °C in liquid nitrogen (198). The speed of freezing produces vitreous ice, and prevents ice crystal formation, whilst allowing samples to remain in a frozen hydrated state which retains the native conformations of proteins (199). Ideally the ice thickness should match the particle dimensions as closely as possible, to maximise contrast and reduce multiple scattering events (200).

Transmission Electron Microscope anatomy

An electron beam is generated by an electron gun: essentially a high current passing through a wire where electrons with enough energy can 'escape'. The most common electron gun is a field electron gun (FEG): whereby escaped electrons are 'pulled' out of a fine tip by an electromagnetic field. The beam of electrons then passes through a series of accelerators within the microscope column, each at a sequentially higher voltage, so when the beam hits the sample at 0 V, each electron has been accelerated by the full potential of the microscope (typically 200 or 300 kV for modern instruments). The entire column must

be kept in a strict vacuum, typically (10^{-7} to 10^{-8} Torr (201)) as electrons will interact with any molecules present.

Electrons are focussed by three sets of lenses which allows for spatial and temporal cohesion of electrons. This ensures electrons reaching the image plate are aligned. Each lens system consists of electromagnets, parallel to the electron path. The field force produced, causes electrons to be focussed along the optical axis with the same energy. Deflectors at the top of the lens represent additional electromagnets, this time perpendicular to the path, which first deflect electrons toward the centre of the column. Finally, the aperture at the base of the lens means that only electrons close to the optical axis are able to pass, removing non convergent radiation. The condenser lenses control the intensity and centering of the beam, ensuring parallel illumination. The objective lens then controls the initial image formation and quality. This can be modified by stage adjustments including the tilt and Z height. Finally, the projector lens system is responsible for cumulative magnification up to a million-fold amplification.

At the image plane, electrons produce an image using a detector. Traditional electron detectors are either photographic film or charge coupled detectors (CCDs), whereby electrons are converted to photons by a scintillator, before being focussed through a fibre optic bundle where the photon is converted to charge. However, huge advances within the field have been produced by the emergence of direct detectors which directly convert electrons to charge. These can take images several hundred times a second, recording micrographs over a series of frames (movies) that can subsequently be aligned. Movement induced by the electron beam or in the stage can be corrected post-image acquisition, known as 'drift correction'. In addition, the dose of electrons is varied throughout the movie. Earlier frames experience the most "drift" but also the lowest electron dose and so contain the highest resolution information. Later frames accumulate a higher electron dose, producing a better signal to noise ratio but are subject to greater sample damage. Averaging of frames can be adjusted based on these factors, in a process known as dose weighting, to improve the final signal (202–205). This also allows the total dose to be raised from approximately $20 \text{ e}^-/\text{\AA}^2$ to $50 \text{ e}^-/\text{\AA}^2$ (206). Critically, direct detectors also eliminate the secondary scattering of photons which occurs in CCDs, increasing sensitivity and resolution

significantly. This allows for the detection of individual electron events. The combined effect of improved microscopes, electron sources, and detectors has allowed for smaller samples to be imaged, down to 40 kDa (207). A summary of the microscope anatomy is described in Figure 1.17.

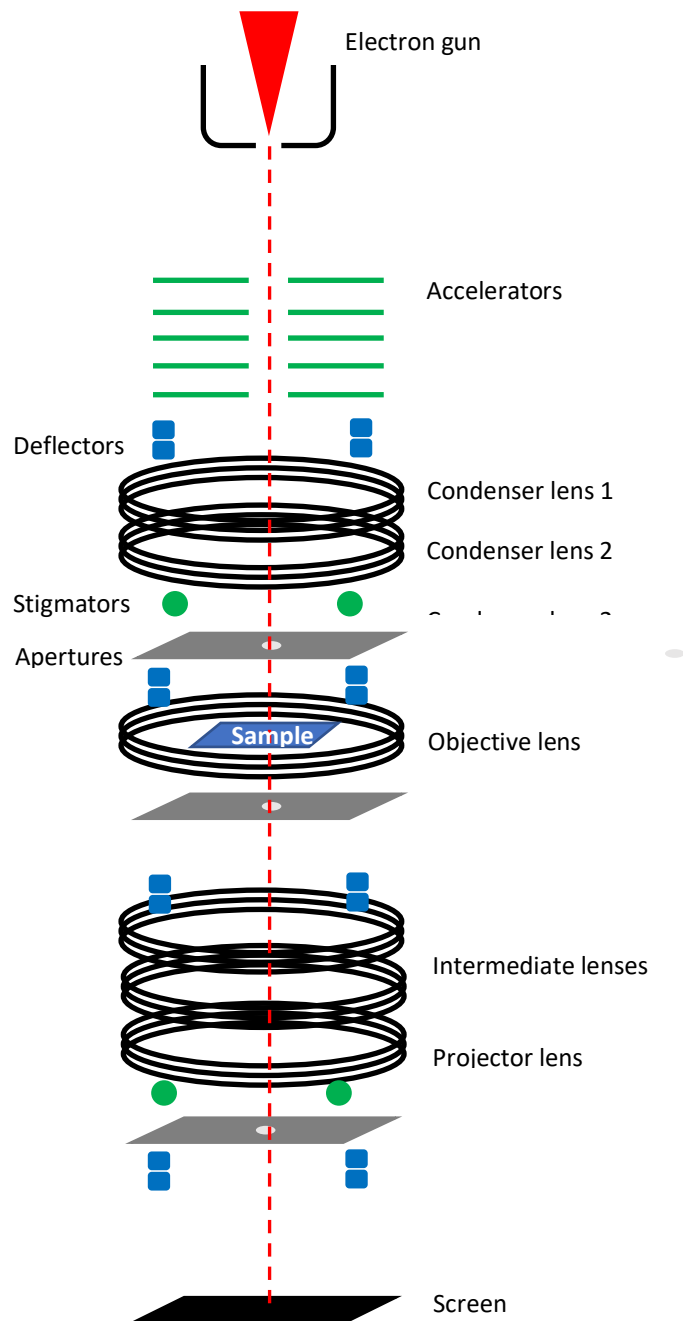


Figure 1.17: Electron microscope anatomy where the red dotted line represents the central

Image formation in the microscope

Electrons may interact with the sample in several different ways. For small molecules, such as those found in biological samples, most electrons will pass through undeterred, with some experiencing deflection. Since negligible radiation is absorbed, the intensity and energy of the incident and scattered waves are equal. This results in a lack of contrast within the image.

Electrons may be scattered elastically (maintaining energy) or inelastically (losing energy to the sample). This lost energy may result in ionisation; chemical bond rearrangement; x-ray emissions; or secondary electron release; each of which equates to sample damage.

Elastically scattered electrons are instead deflected from the electromagnetic field of the nucleus, causing them to take a different path length through the specimen, and emerge out of phase with the incident beam (208)(Figure 1.18).

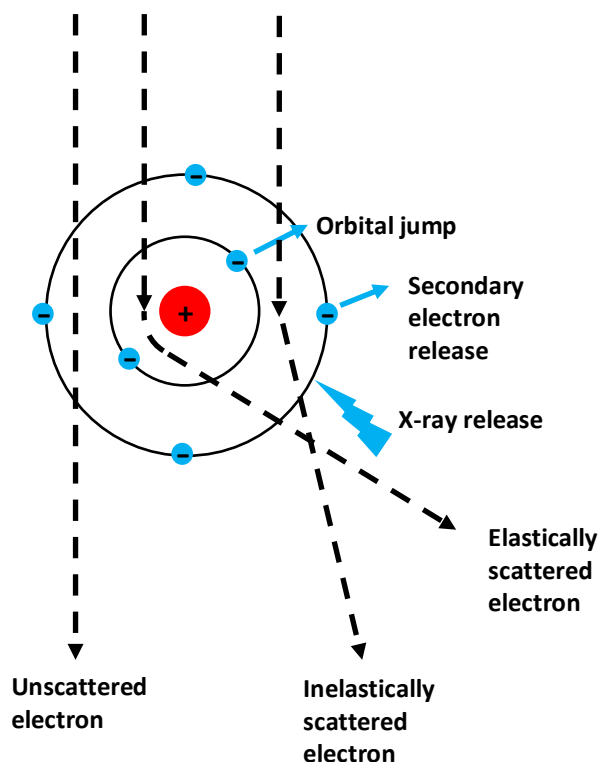


Figure 1.18: Possible interactions between incoming electrons and the biological sample in an electron microscope. Adapted from (271)

The scattering angle of an elastically scattered electron is proportional to electron potential of the atom (proportional to atomic number). This allows the transmitted electron wave to be described as

$$\Psi_{\text{sam}}(\vec{r}) = \Psi_0 \exp(i \sigma \phi_{\text{pr}}(\vec{r}))$$

Ψ_{sam}	Transmitted wave
Ψ_0	Incident wave
σ	$m_e \lambda / (2 \pi \hbar^2)$
\hbar	$h / 2\pi$ where h is Planck's constant
$\phi_{\text{pr}}(\vec{r})$	$\int_{-t/2}^{+t/2} \phi(\vec{r}, z)$
z	Specimen potential in z axis i.e along microscope optical axis
\vec{r}	Vector in image plane
t	Sample thickness

Atoms within biological samples (C, O, H and N) are relatively small, and therefore only weakly scatter electrons. This allows the exponential term within this equation to be approximated to zero, and the transmitted wave to be as modelled as:

$$\Psi_{\text{sam}}(\vec{r}) \approx \Psi_0 (1 + i \sigma \phi_{\text{pr}}(\vec{r}))$$

Here, the first term represents the unscattered wave and the second, imaginary component, describes the scattered wave. Interference between the scattered and unscattered wave is known as the phase contrast. The intensity of the image is modelled by the equation:

$$I(\vec{r}) = \Psi_{\text{sam}}(\vec{r}) \Psi_{\text{sam}}^*(\vec{r}) \approx 1 + (\sigma \phi_{\text{pr}}(\vec{r}))^2$$

When the phase of the scattered component is shifted by 90° , the equation changes to:

$$I(\vec{r}) = \Psi_{\text{sam}}(\vec{r}) \Psi_{\text{sam}}^*(\vec{r}) \approx 1 - 2 \sigma \phi_{\text{pr}}(\vec{r})$$

Now, the invisible phase shift is converted to a visible increase in amplitude (seen as contrast and depicted in Figure 1.19).

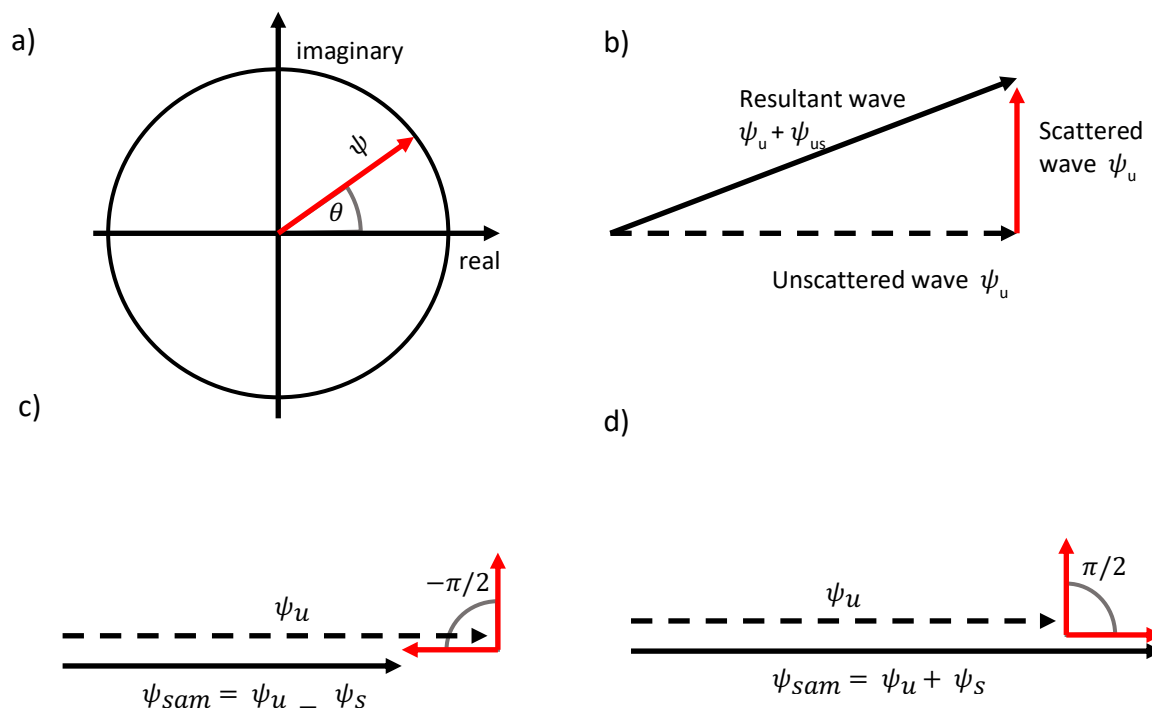


Figure 1.19: Vector diagram for phase contrast a) Complex plane representation of wave vector ψ with phase θ b) vector representation of the scattered wave, unscattered wave and resultant wave where amplitudes of unscattered and resultant waves are similar c) $-\pi/2$ phase shift produces negative contrast d) phase shift of $+\pi/2$ produces positive phase contrast. Adapted from(271)

In practice this can be achieved by introducing a change in spherical aberration and increasing defocus. Spherical aberration describes the increase in lens induced refraction experienced by electron waves at the periphery of the beam, relative to the centre. Meanwhile, defocus is measured as the distance from the sample at which the beam converges (the focal point).

This phenomenon can be explained mathematically by the point spread function: the distortion experienced by a single dot within the sample, to produce the equivalent dot in the image. Distortions are introduced by a number of factors within the microscope

including lens aberration, sample radiation damage and electron coherence as well as general temporal/ spatial instability throughout imaging. These can be described in Fourier (diffraction) space by the equation:

$$F \{PSF (R \rightarrow)\} = CTF (R \rightarrow). E (R \rightarrow)$$

Where the Contrast Transfer function is described by $\exp(i\gamma)$.

$$(R \rightarrow) = -2\pi (\lambda \Delta R^{-2}) - \frac{1}{4} C_s \lambda^3 R^{-4}$$

PSF	Point spread function
E(R)	Envelope function
CTF (R)	contrast transfer function of the microscope = $\exp(i\gamma)$
γ	phase shift caused by aberrations
C_s	coefficient of spherical aberration
R	spatial frequency
Δ	Defocus
λ	Wavelength of electron beam

The contrast transfer function describes the extent to which the phase shift is converted to amplitude at different frequencies and depicts a 2D oscillating wave in the “reciprocal” Fourier space (Figure 1.20). Information is therefore lost at certain spatial frequencies, but much can be recovered by utilising a range of defocus values which allows for CTF correction of images. At high frequencies, corresponding to high resolution information, the CTF declines, dependent on defocus and other imaging parameters.

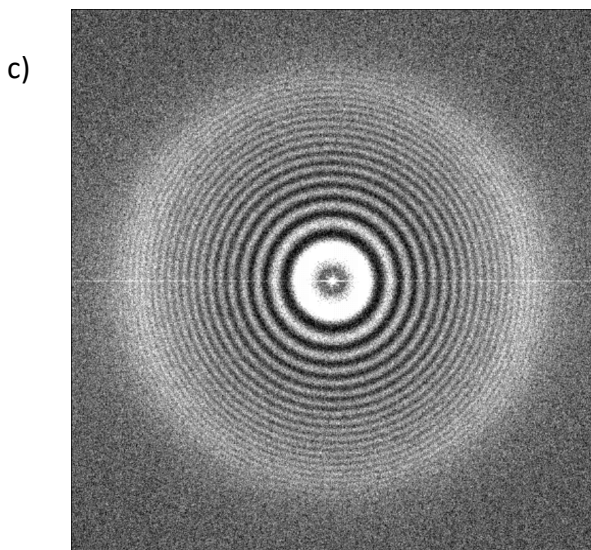
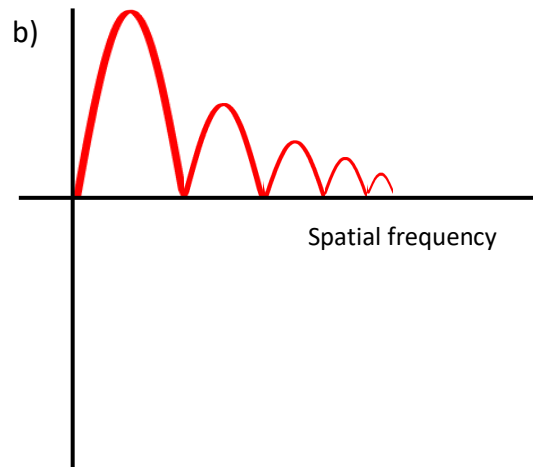
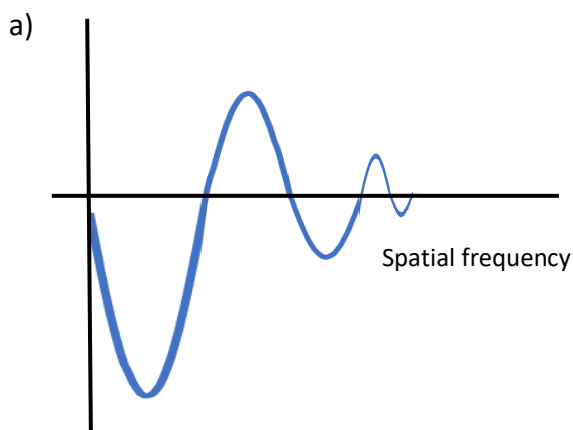


Figure 1.20: Depiction of the Contrast Transfer Function. a) CTF curve at a single defocus b) CTF curve after phase correction c) Example CTF of an electron micrograph (power spectrum)

Contrast within images may also be enhanced by the use of a phase plate, which shifts the path of scattered electrons relative to unscattered electrons by 90° (209). Energy filters are also employed to remove inelastically scattered electrons, which have lower energy than scattered electrons, and therefore are deflected along different paths (210,211). A physical slit can therefore be used to restrict which electrons contribute to image formation. Such energy filters may exist within the column, or post column, and typically allow electrons with energies in the range of approximately 20 eV to pass through (212).

Building a 3D model

The low dose rate required for cryo-EM inevitably causes a low signal to noise ratio (SNR), producing images with poorly resolved features. This poses challenges for both the selection of particles and their alignment to reconstruct a 3-dimensional model. Automated particle picking by RELION utilises a Laplacian-of-Gaussian filter to select for areas of rapid intensity change, or instead manual picking of particles can be used to provide a reference image (213). Alternatively, machine learning algorithms, such as that employed by TOPAZ software, use neural network to train for the selection of 'positive' templates (214).

The Fourier projection-slice theorem describes how a Fourier transform of a 2D projection image of a 3D object, shares identical phases and amplitudes as a central slice of the 3D Fourier transform of the object in question (Figure 1.21). Using cross correlation to identify lines of intersection between 2D transforms of particles, a 3D Fourier transform can therefore be built. In turn, a reverse Fourier transform then produces a real space model (215). The process of iterative alignment allows for angular assignments to be progressively corrected and improved. 2D projection images (i.e. particles) are cross correlated with a 3D reference in Fourier space in order to estimate angular assignments. These new angular assignments are now used to build the new reference map for the next refinement iteration. Iterative refinement is also used for 2D classification, allowing particles in distinct orientations to be clustered and enhancing the SNR (216).

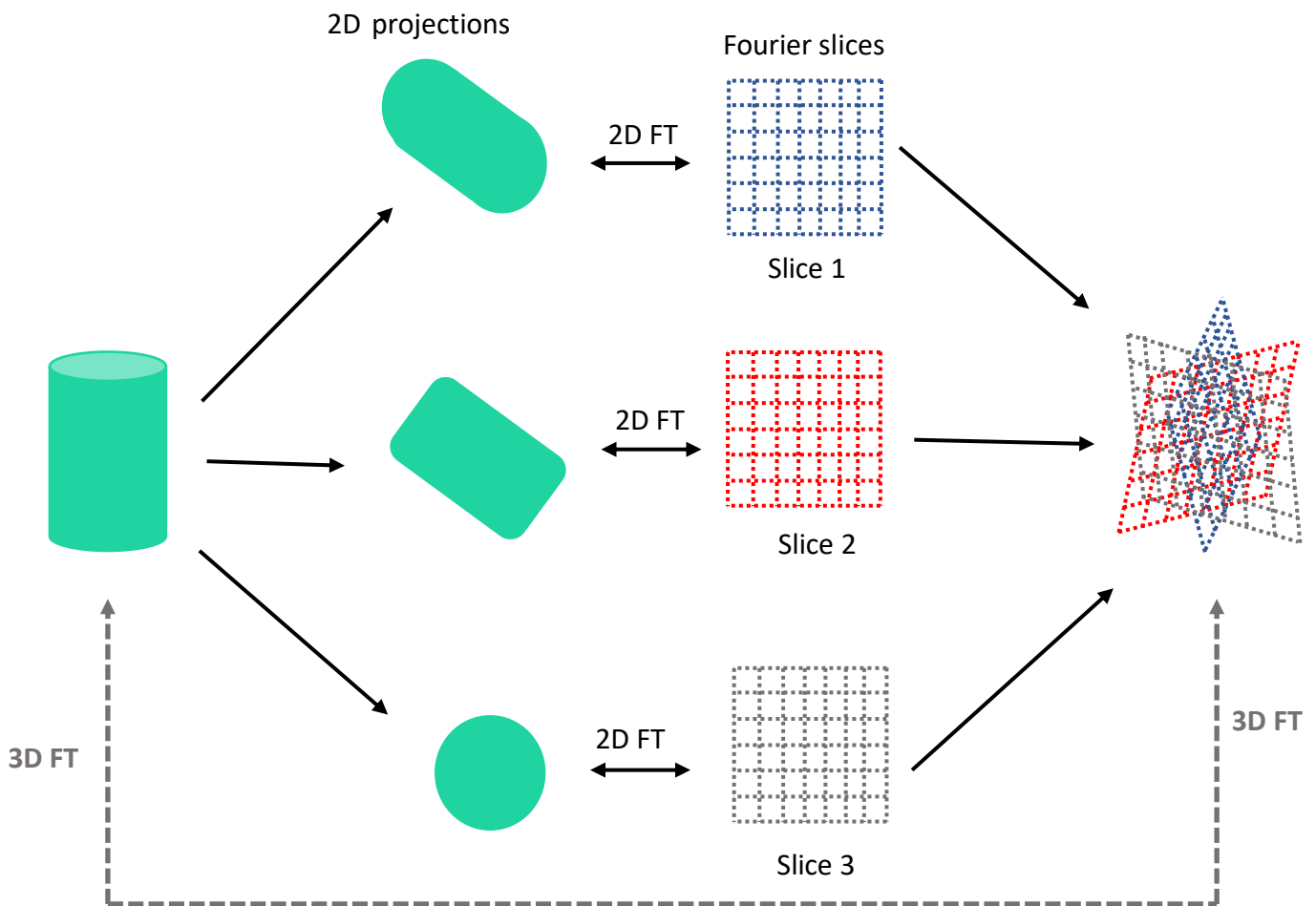


Figure 1.21: The Fourier projection slice theorem adapted from (215)

Unlike previously used maximum likelihood approaches, whereby particles are assigned the most likely angles during refinement, RELION uses Bayesian statistical approaches. This means the angular assignment are weighted based on likelihood, and thus preserve information on the range of possible angular distributions. Such an approach limits the risk of overfitting: where noise particles are randomly aligned to a starting model, and thus preserve these arbitrary assignments in later iterations reinforcing the starting model (217). Indeed, random noise can be aligned to reinforce any input model as demonstrated by the famous 'Einstein from noise' example (218). Where a previously determined model is used as a reference, model bias is also reduced.

Estimating resolution

The resolution of cryo-EM derived reconstructions corresponds to the highest frequency data within the map which can be reliably distinguished from noise (216). Resolution is typically measured by dividing the data into two and reconstructing each set into separate 'half maps'. Strict independence of these two reconstructions from both the initial model and each other is critical for the calculation as characterised by the gold standard approach (219). The half maps may now be subject to cross correlation in Fourier space.

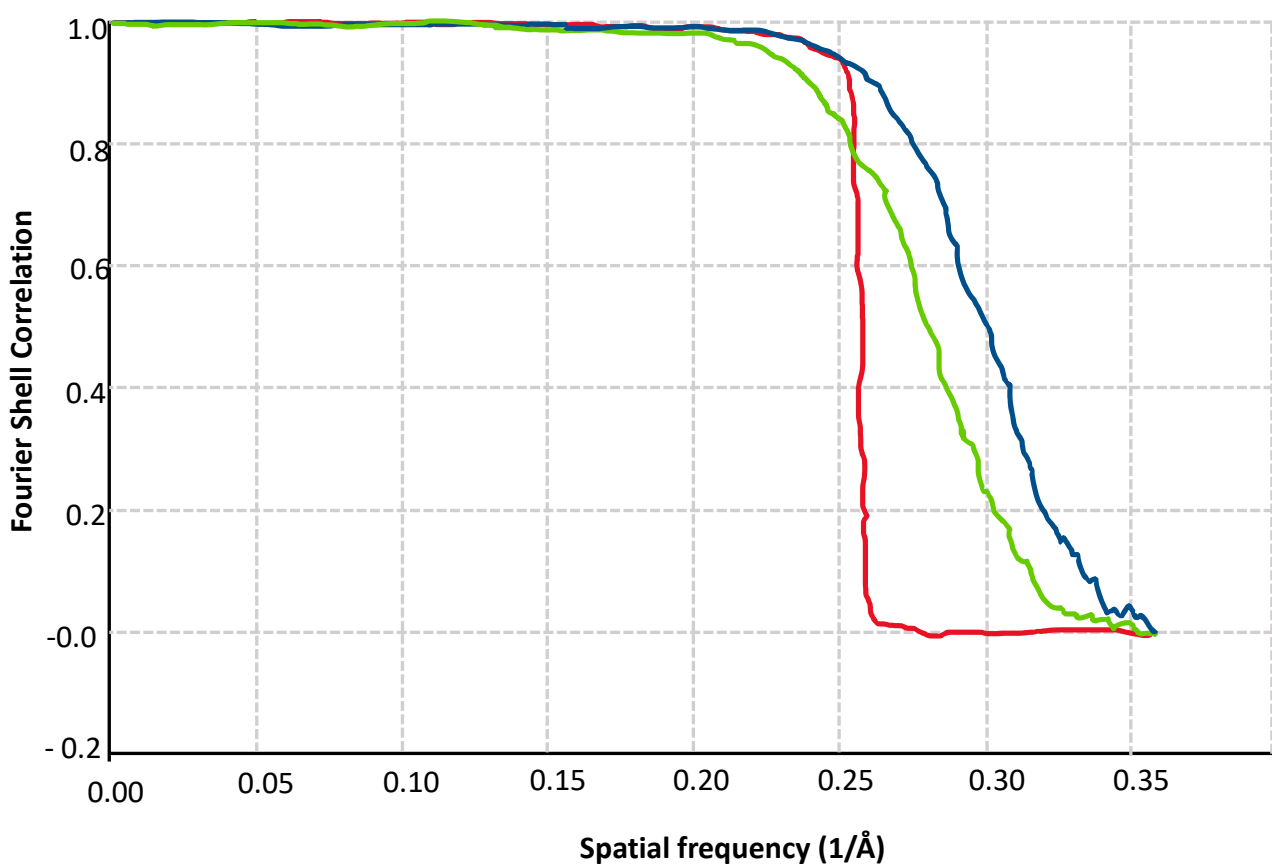


Figure 1.22: An example of an FSC curve where the *red* curve shows phase randomise masked maps, the *green* curve shows unmasked maps, and the *blue* curve shows masked maps

An example plot of correlation against spatial frequency is depicted in Figure 1.22. Where the coefficient is equal to 1 the two maps show perfect correlation, whilst a coefficient of 0 indicative of pure noise. Resolution can be identified by identifying the spatial frequency at a chosen threshold of correlation: a matter of hot debate. Most recently the value of 0.143

has been adopted in order to simulate values comparable to X ray crystallography data of a similar quality (220). However, theoretical calculations have indicated that this “fixed threshold” estimation is an unreliable criterion for resolution assessment (221). An alternative technique for estimation of high-resolution maps is known as ‘map versus model’. Here the map is used to build an atomic model, which is in turn used as a template for a new noise free map. The quality of the original map may then be assessed by comparison (222). FSC calculations may also be made by first dividing the maps into sections, allowing local resolution to be estimated (223).

Chapter 2: Stalling the HK97 DNA packaging motor for structural investigation

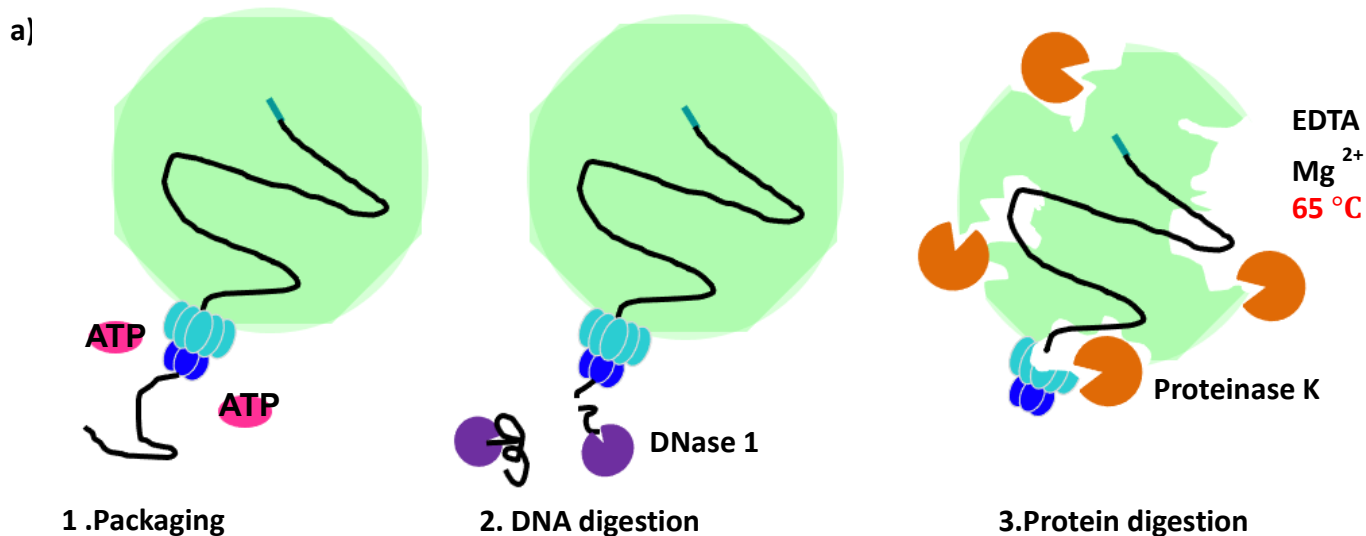
Introduction

To truly understand the DNA packaging machinery of HK97 and related viruses, a cryo-EM reconstruction of the active motor would be invaluable. Whilst a wealth of structural data comprising individual packaging proteins has emerged, the field is only starting to move towards analysing complete packaging systems of dsDNA viruses. The recent structure of the bacteriophage Φ 29 DNA packaging machinery has inspired a novel model for DNA translocation: whereby the large terminase ATPase domains transition from a cracked helix to planar ring (177). Meanwhile, structural data on the bacteriophage T4 suggests oscillation of the large terminase monomers between a compact and an extended conformational state (181). This contraction is thought to push DNA into the prohead.

However, each of these systems varies dramatically from HK97. For instance, neither T4 nor Φ 29 utilise a small terminase for packaging initiation, and for Φ 29 a unique pRNA links the portal to large terminase (104). Meanwhile termination of packaging in T4 occurs in a headful manner, while Φ 29 requires the protein gp3 (100). A high-resolution structure of the active HK97 packaging motor would therefore shed light on a number of unanswered questions for *cos* viruses: How does the virus overcome the symmetry mismatch between the 12-subunit portal and pentameric large terminase against enormous internal pressure? Does small terminase remain bound throughout packaging? What is the mechanism of coordination around the large terminase ring and how are the chemical events of ATP hydrolysis coupled to mechanical translocation of DNA?

Packaging occurs only transiently in a virus life cycle. For single particle analysis by cryo-EM, numerous packaging events must be captured simultaneously, it is therefore critical to initiate packaging *in vitro*, under controlled conditions. For HK97 much work has been done

to establish a highly effective *in vitro* packaging assay (1). Proheads, large terminase and small terminase produced by recombinant expression, are primed with a double stranded DNA substrate, and packaging is instigated by the addition of ATP. After 30 minutes, the mixture is incubated with DNase I to digest any unpackaged DNA. DNase I is in turn inhibited by EDTA, and subsequent digestion with proteinase K allows the packaged DNA to be visualised on an agarose gel (1). Figure 2.1 shows a schematic representation of the assay and reproduction of a typical packaging assay at 10 mM ATP. The input reaction is incubated without DNase and thus shows that the DNA substrate is kept intact during packaging. The control reaction is not incubated with ATP and so indicates that all packaging is strictly ATP dependant, and the DNase is active



b)

2log Input Reaction Control

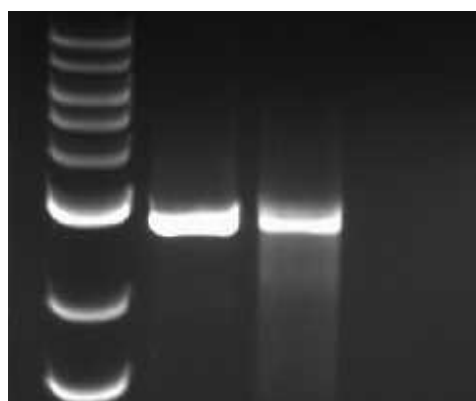


Figure 2.1 DNA protection assay for HK97 DNA packaging a) Schematic of a DNA protection assay. Packaging ensues for 30 minutes, and excess DNA is degraded by DNase. After 15 minutes DNase is inhibited by EDTA and protected DNA is released by proteinase K. b) Agarose gel showing protected DNA after HK97 packaging reaction at 10 mM ATP.

The purification protocol for prohead II has been established using the *amC2* HK97 mutant, deficient in the terminase gene (224), as described in methods. For early packaging reactions, large terminase and small terminase were expressed and purified by Maria Chechik. DNA substrates were synthesised using a PCR reaction to linearise a pUC18 plasmid with a 'cos site' introduced using restriction enzymes. The cos site represented -312 to +471 of the DNA cleavage site, within the HK97 viral genome (1).

Previous work varying packaging conditions has also elucidated several features of packaging. Firstly, small terminase was found to have a stimulating effect on packaging but was not essential. The rate of packaging was also determined at approximately 600bp/s. This speed is comparable to λ phage, which concurrently has a comparable genome length of 48.5 kb relative to 39.7 kb for HK97 (112,114,159). Packaging appeared more tolerant to potassium glutamate relative to sodium chloride (225)(1).

Perhaps most interestingly, at an ATP concentration of 50 μ M, the motor appeared to stall at the *cos* site (1). This was only apparent when small terminase was included in the packaging assay. DNA was not cleaved, but partial cleavage could be instigated by the addition of the HNH endonuclease gp74, produced *in vitro*. This was deduced by the presence of two discrete bands in the input lane, corresponding to the length between each free DNA end to the *cos* site. These bands were also present in the reaction lane as protected DNA, but it is unclear if cleavage occurred prior to packaging at initiation, or on encountering the *cos* site. Overnight incubation of large terminase, small terminase, DNA and proheads in the presence and absence of ATP indicate that gp74 does stimulate specific cleavage at the *cos* site as well as limited non-specific cleavage (1).

Finally, a packaging solution was applied directly to a Quantifoil R 2/2 grid for analysis by cryo-EM. Proheads are partially expanded, divided between two intermediate classes with diameters of 600 Å and 604 Å. These are resolved to 6.7 Å and 8.9 Å respectively. An asymmetric reconstruction of the packaging phage shows signal for large terminase, DNA, and portal, and is resolved to 22.8 Å (1). Figure 2.2 b) depicts this asymmetric reconstruction and Figure 2.2 a) shows a negative stain electron micrograph recreating the conditions used in this data set. Several areas for potential improvement are highlighted: the presence of tails, expansion of some proheads and unsynchronised packaging.

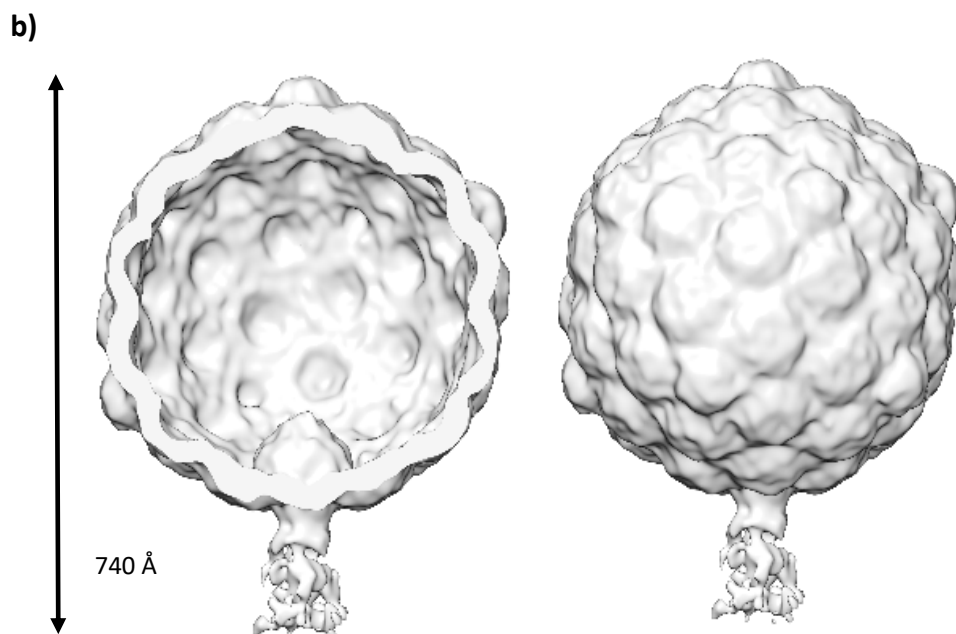
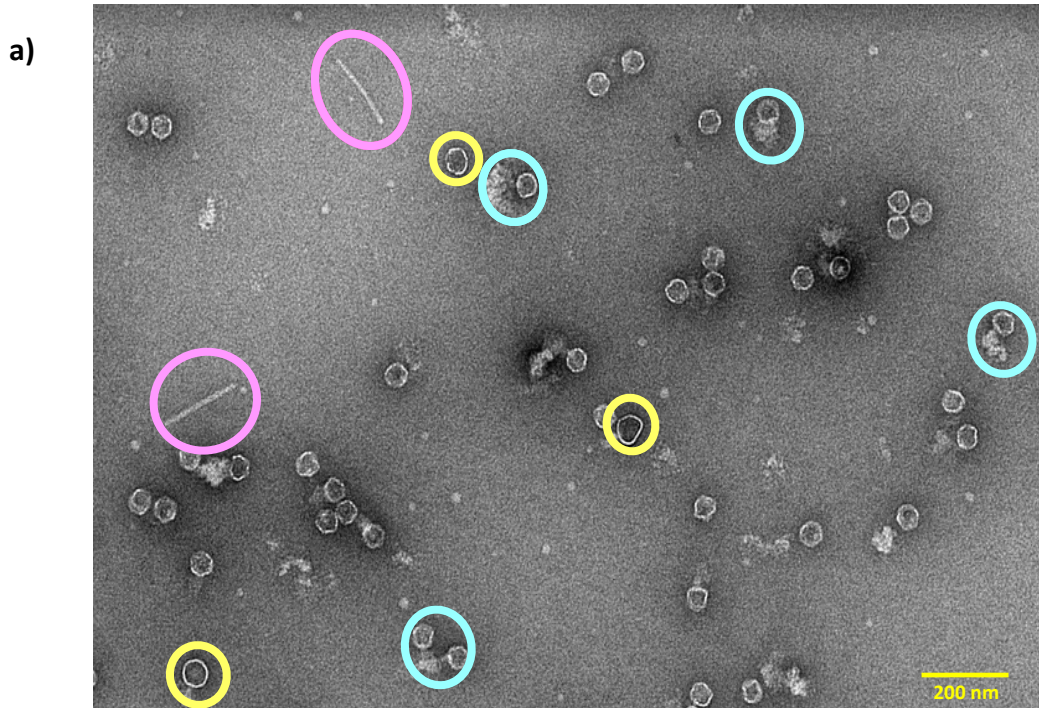


Figure 2.2: Current model of HK97 complete packaging machinery a) micrographs of HK97 packaging DNA at 10 mM ATP, with samples prepared using Uranyl Acetate. Contaminating tails are circled in pink, expanded proheads in yellow and unsynchronised packaging events in blue b) A cryo-EM asymmetric reconstruction of the HK97 packaging complex at a resolution of 22.8 Å (1).

In this chapter the optimisation of sample preparation for further investigation by cryo-EM is described. Firstly, the prohead purification is modified to inhibit prohead expansion. Stalling of packaging is optimised in order to 'fix' motors mid-packaging. Lastly, a novel technique using magnetic beads is designed, for the specific selection and isolation of stalled packaging complexes.

Methods

Phage amplification

Glycerol stocks of LE392 *E. coli* (226) were streaked out on an LB-agar plate overnight at 37°C. A single colony was then picked to prepare an overnight culture in LB and 0.4% maltose for inoculation of volumes of 500 ml LB, each in 2.5 L flasks. Cultures were infected with stock mutant phage amC2 (224) at an OD₆₀₀ of 0.1 and incubated at 37 °C for a further 2 hours at an RPM of 200. Subsequent steps were carried out at 4 °C or on ice where possible. Chloroform was added to a concentration of 25 mM and stirred for 20 minutes to encourage bacterial lysis. Remaining unlysed cells were then harvested by centrifugation at 6000 G and the supernatant was collected for PEG precipitation. This proceeded through incubation with 0.5 M NaCl and 2.5% (w/v) PEG 8000 overnight. Precipitate was collected by centrifugation at 7000 G for 20 minutes and the pellet redissolved in 25 ml TMG (10 mM Tris-Cl, pH 7.5, 10 mM MgSO₄, 0.01% (w/v) gelatin) per litre. The inclusion of gelatin in buffers is thought to improve long term viability of phage (227). Undissolved debris was then removed by centrifugation at 6000 xg. Finally, phage were pelleted by ultracentrifugation using a T45 rotor at 20,000 xg for 2 hours and the resultant pellet again redissolved in TMG.

Plaque assays

Virus titres were estimated using plaque assays. LE392 cells were grown overnight from a single colony with additional 4 % (w/v) maltose. This is thought to enhance expression of the HK97 receptor. Serial dilutions of stock phage were prepared in TMG buffer and 25 µL of each mixed with 150 µL of LE392 cells. This infection was then incubated at 37 °C for 20 minutes with a rpm of 200. 3 mL of molten 0.75 % LB-agar was next added to each mixture

and vortexed briefly before plating onto a bed of 1.5% LB-agar. Plates were incubated overnight at 37 °C. The number of plaques grown for each dilution was used to estimate the stock titre.

Capsid purification

Procapsids were produced by infection of the amber suppressor *E. coli* cell line 594 (*supF58*) (228) with the amplified amber amC2 stock. An overnight culture of 594 cells, grown with 0.4% (w/v) maltose, was used to inoculate 500 ml volumes of LB in 2.5 L flasks at 37 °C. From an initial OD₆₀₀ of 0.05, cultures were infected with phage mutant at OD₆₀₀ 0.4 at a ratio of 10 viral particles per cell. Bacteria had largely lysed after 90 minutes and remaining whole cells were removed by centrifugation at 4000 xg for 30 minutes. Following steps were all performed at 4 °C or on ice where possible.

The supernatant was subject to PEG precipitation using 15% (w/v) PEG 8000 and 0.5M NaCl overnight. The precipitate was pelleted by centrifugation at 10,000 G for 15 minutes and redissolved in 20 mM Tris pH 7.5, 100 mM NaCl before a second round of centrifugation at 12000 G for 15 minutes to remove undissolved material. Proheads were then pelleted by ultracentrifugation using a T45 rotor at 35,000 xg and redissolved overnight in the same buffer.

The viral particles were next subject to analytical ultracentrifugation using a glycerol gradient of 10%–30% and created in Ultra-Clear (Beckman Coulter™) 38.5 mL tubes using a Gradient Master (BioComp). The dissolved pellet was layered on top of the tubes which were then spun at 27,000 G for 2 hours in a SW28 Ti rotor. Proheads formed a visible band in the centre of the tube which was extracted using a needle and dialysed again 20 mM Tris pH 7.5, 100 mM NaCl.

Lastly, proheads were applied to a POROS HQ 20 µm anion exchange column and eluted in a 100–1000 mM NaCl gradient. Selected fractions were pooled and pelleted again by ultracentrifugation at 35,000 xg for 2 hours in a Type 45 Ti Rotor. The pellet was redissolved overnight into a concentration of approximately 500 nM.

Packaging assays

Packaging assays were performed using the protocol developed by Dr Herman Fung (1). 20 μ l reactions were performed in a buffer of 20mM Tris pH 7.5, 10 mM $MgCl_2$, and 50 mM Kglu. Proheads were present at 30 nM in a 2:1 molar ratio to DNA (linearised pUC18 (229)) and large terminase and small terminase were added to 1 μ M and 2 μ M concentration respectively. Packaging was initiated by the addition of ATP to 10 mM and allowed to proceed for thirty minutes at room temperature. The reaction was then incubated with DNase I for 10 minutes to digest unpackaged DNA before degradation was terminated by the addition of EDTA. Proteinase K produced in house was next added to the reaction mixture for 30 minutes incubation at room temperature followed by 30 minutes incubation at 60 °C. Thus, packaged proheads were digested and DNA was released for visualisation on a 0.8% agarose gel run at 72V for 2 hours.

Negative stain electron microscopy

For each grid, 7 μ l of sample was blotted onto carbon-coated copper grids and washed three times with MQ water before staining with 1% uranyl acetate. Grids were imaged using a Tecnai 12 G2 microscope operating at 120kV using a ISS MegaView III CCD camera.

Magnetic bead packaging assays

The volume of Sera-Mag Magnetic Streptavidin Coated particles beads used was equal to half the total reaction volume (a packaging assay featuring a biotinylated DNA substrate). Beads were washed with two successive 500 μ l volumes of TM buffer prior to assays. Next packaging was initiated as for standard assays, with an additional 5 minutes to ensure complete DNA binding. Critically, ATP was added to a final concentration of 75 μ M. After 30 mins incubation at room temperature, beads were separated magnetically and any unbound material removed. Beads were washed with two successive volumes of elution buffer (20 mM Tris pH 7.5, 10 mM Mg^{2+} , 50 mM KGlu) before re suspending into a chosen third volume of elution buffer and 0.3 μ l of the restriction enzyme *ScaI*. Incubation preceded for 1 hour at 37 °C before removing beads magnetically and releasing cleaved complexes.

Results

Optimisation of capsid production protocol to avoid expansion

In vivo, HK97 genome packaging is initiated when DNA bound small terminase recruits large terminase. Large terminase in turn docks onto the portal protein, at the unique vertex of prohead II, and packaging can ensue. The prohead passes through a series of expansion intermediates before adopting the final stable state Head II. In the available cryo-EM data set, two expansion intermediates were identified with diameters of 600 Å and 604 Å (1). These may be less packaging competent than prohead II, and may contain conformational differences within the portal structure, as changes in capsid structure are expected to be relayed to the portal protein during expansion (128). Thus, it was considered important to prevent proheads from expansion during the purification, in order to mimic *in vivo* packaging as closely as possible.

The purification of prohead II from *amC2* mutant phage, deficient in terminase gene, has been previously established as described in methods. Purification steps were tracked using negative stain electron microscopy to identify a locus where expansion could occur. In the penultimate stage of purification, where procapsids are applied to a POROS ion exchange column, the elution profile showed two peaks at ~300 mM and ~400 mM NaCl. The first peak contained predominantly expanded capsids and tails, whereas the second contained more compact and rounded proheads (Figure 2.3).

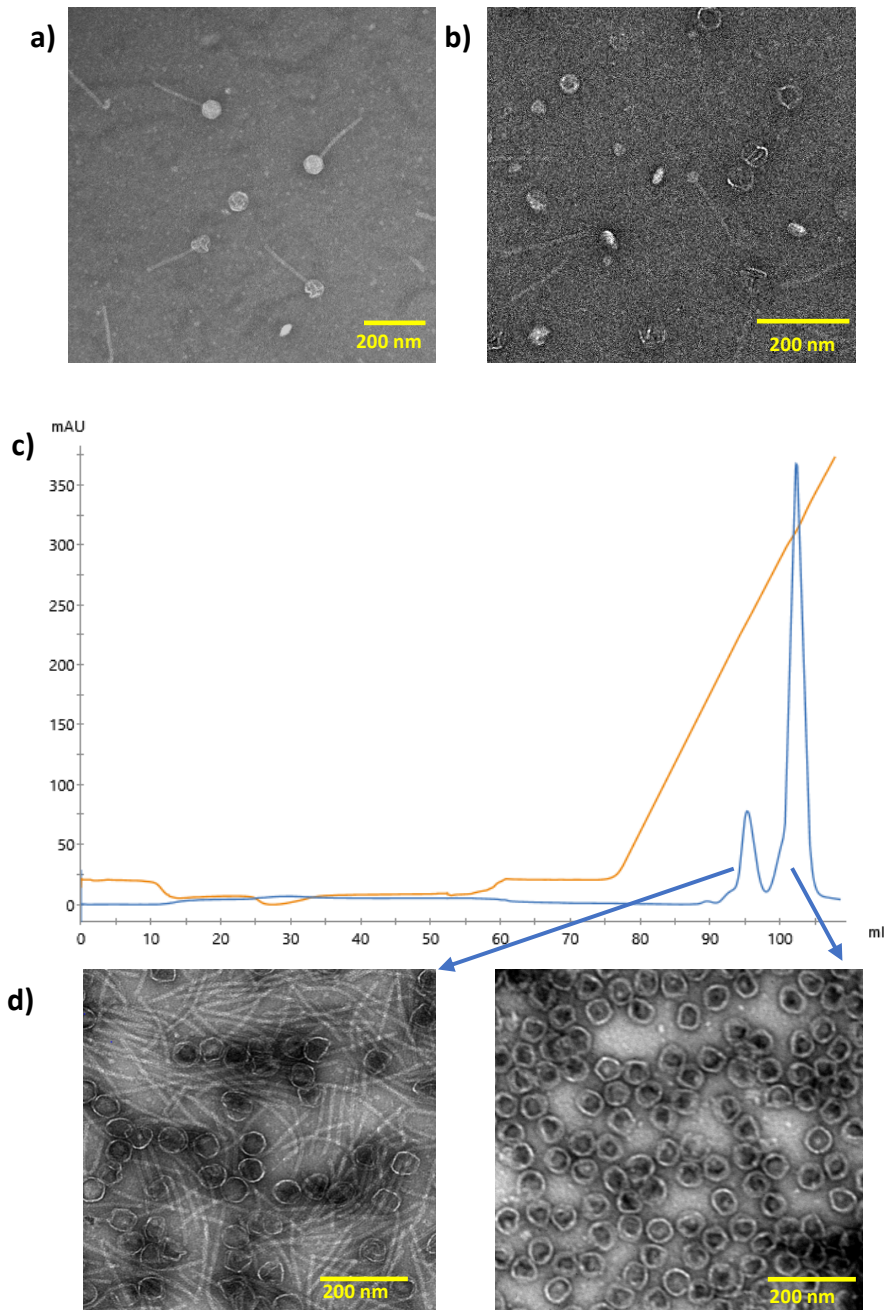


Figure 2.3: Negative stain electron micrographs tracking HK97 prohead II purification from the *amC2* amber mutant phage expressed in *E. coli* strain 594. a) wild type phage b) after analytical ultracentrifugation in a glycerol gradient c) chromatogram from application to a POROS column d) peak 1 elution e) peak 2 elution

For packaging reactions these compact proheads required dilution, in order to lower salt concentration, followed by concentration. Initial attempts to dilute proheads, and subsequently concentrate by ultracentrifugation resulted in expansion of proheads (negative stain electron micrograph depicted in Figure 2.4 a). Concentration was instead trialled with a Amicon[®] cellulose concentrator, allowing for step wise dilution into a buffer with lower ionic strength. Unfortunately, this technique produced damaged and expanded capsids (Figure 2.4 b). Assessment by A₂₈₀ readings on a nanodrop also indicated that the ultimate concentration had not significantly increased. Together, these results suggested proheads may be adhering to the membrane surface within the concentrator.

The final technique employed was to pellet the prohead peak, directly in the high salt buffer. It remained important to use additional buffer to ensure the tube for ultracentrifugation was adequately filled. The pellet was then resuspended into a buffer of the desired ionic strength. This method proved successful in preventing expansion (Figure 2.4 c) and the resultant proheads were used in all subsequent packaging assays.

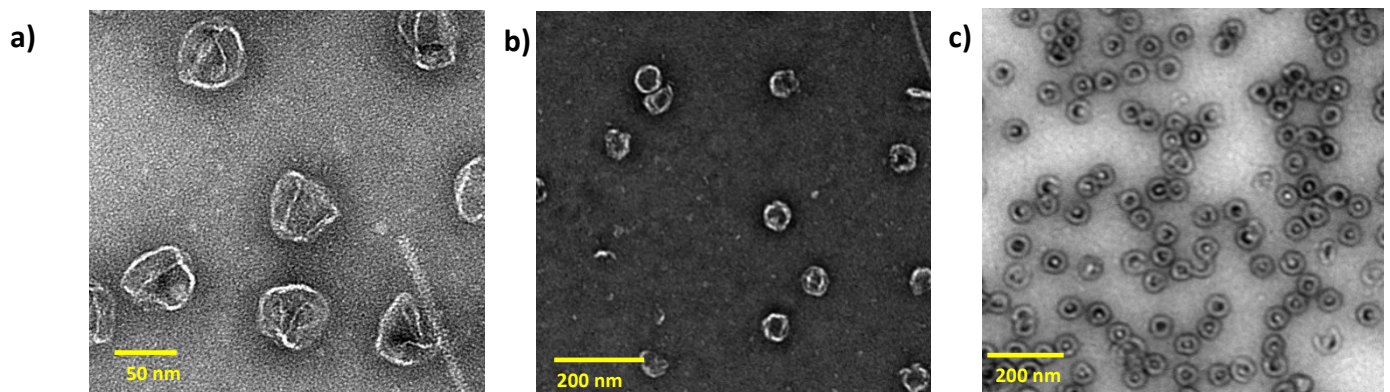


Figure 2.4: Negative stain electron micrographs tracking HK97 prohead II final purification step a) dilution into a low salt buffer and pelleting b) concentration in a 300 kDa commercial concentrator c) dilution into high salt buffer, pelleting and resuspension into low salt buffer

Stalling the DNA packaging motor at a *cos* site

The potential to stall the HK97 motor in action was considered a highly valuable approach for future cryo-EM analysis of the complete motor assembly. By introducing a *cos* site into the DNA substrate of a packaging reaction, the motor appears 'fixed' in place at a low concentration of ATP (1). This should alleviate the possibility of the motor detaching, after packaging the full substrate. Stalling should also force each motor into synchronisation: with each terminase oligomer poised at the same position, with an identical length of DNA within the prohead. This should produce a homogeneous sample.

After establishing packaging of linearised pUC18 with newly purified proheads, multiple conditions within the packaging assay were varied in order to optimise stalling. Firstly, the ATP concentration was varied from 50 μ M to 10 mM shown in Figure 2.5. Stalling at the *cos* site occurs partially at 100 μ M, with some additional full-length packaging, and occurs completely at 75 μ M ATP. This is shown by the emergence of a lower molecular weight band corresponding to the length of DNA from the 5' end to the 'cos'. The concentration of large terminase and small terminase were varied from 50 nM to 6 μ M, and 0.1 μ M to 12 μ M respectively. Optimal stalling occurred with 3 μ M large terminase and 6 μ M small terminase corresponding to 100 and 200 molecules per prohead. These conditions were carried forward for all future stalling assays.

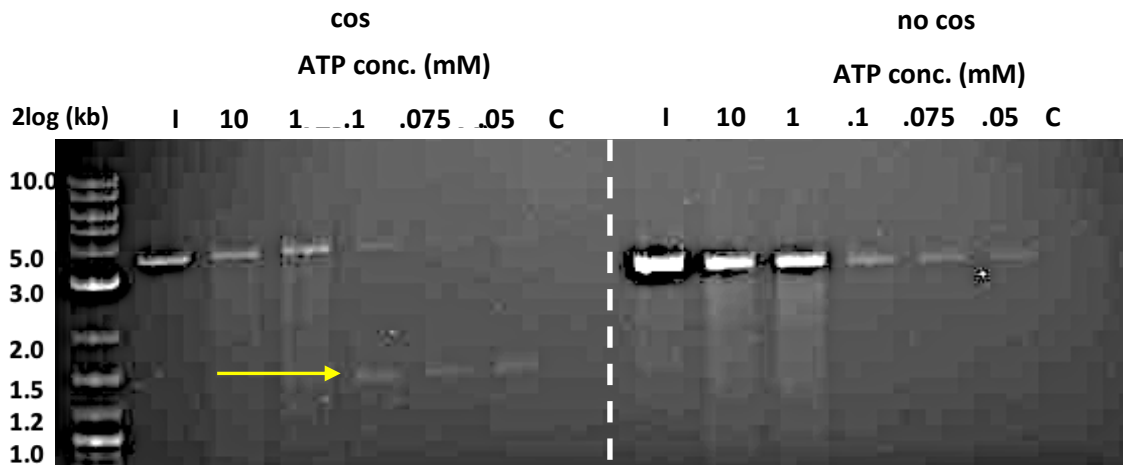


Figure 2.5: DNA protection assays mapping stalling of the HK97 packaging motor at decreasing concentrations of ATP. Protected DNA with, and without, an engineered “cos” site is shown by agarose gel electrophoresis. ‘I’ represents the input lane without DNase and ‘C’ the control without ATP.

Restarting the stalled motor

To test that motors are stalled in a biologically relevant and functional state, an experiment was designed to ‘restart’ stalled motors. A standard packaging assay was performed at 75 μ M ATP, before dividing the solution into two equal volumes. Additional ATP was then incubated with one sample, raising the final concentration to 10 mM. After 5 minutes, protected DNA was examined on an agarose gel. While the control solution shows a band at 1.3kb, corresponding to stalling, for the solution with additional ATP a smearing of DNA bands upwards towards the full DNA length is visible. This suggests that stalled motors can restart and are therefore still functional.

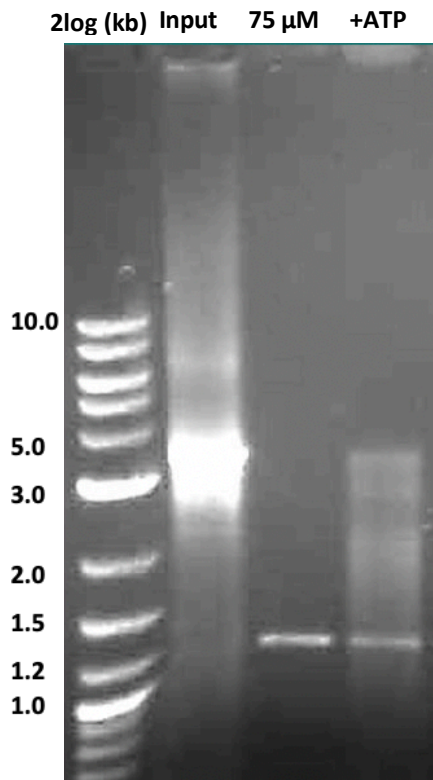


Figure 2.6: DNA protection assay displaying the restarting of stalled HK97 packaging motors. Agarose gel electrophoresis shows protected DNA after a HK97 packaging reaction at an ATP concentration of 75 μ M (stalling), and a subsequent increase to 10 mM (restarting).

Design of a magnetic bead assay to specifically isolate stalled packaging complexes

Stalling of packaging produces a weak band of protected DNA indicative of incomplete stalling among the entire prohead population. It was thus considered important to introduce a further purification step, to select specifically for stalled packaging complexes over empty proheads. Complete incorporation of the motor is especially desirable due to the large size of the prohead at approximately 14,000 kDa, which limits the number of potential particles per micrograph, at typical magnifications used for high-resolution single-particle analysis, even at optimum concentration. Here, design of a magnetic bead assay is presented which elegantly discriminates between empty, fully packaged, and stalled HK97 particles.

Firstly, a DNA substrate was synthesised using a biotinylated 3' primer to linearise pUC18 with a *cos* insert. A restriction site for *ScaI* was engineered just upstream of the biotin label. Using the low ATP conditions established, packaging in the 5' to 3' direction should stall at

the *cos* site, leaving a free biotinylated end. Packaging complexes could then be mixed with streptavidin coated magnetic beads, to anchor the free biotinylated DNA end.

One advantage of the design is that proheads which have packaged the full length of DNA would not bind, as the biotin label would not be accessible (packaging appeared unaffected by the presence of a biotin label in an alternative position indicting the ability to package through the marker). Equally, any proheads without motors would remain untethered. Flow through could then be removed by adhering beads to a magnet, and wash steps introduced by resuspending beads in buffer. Stalled packaging complexes could then be released by a restriction enzyme, into a desired volume, giving the option to concentrate the sample. In theory this should produce a homogenous sample of packaging complexes each stalled in the same locus. The assay design is summarised in Figure 2.7.

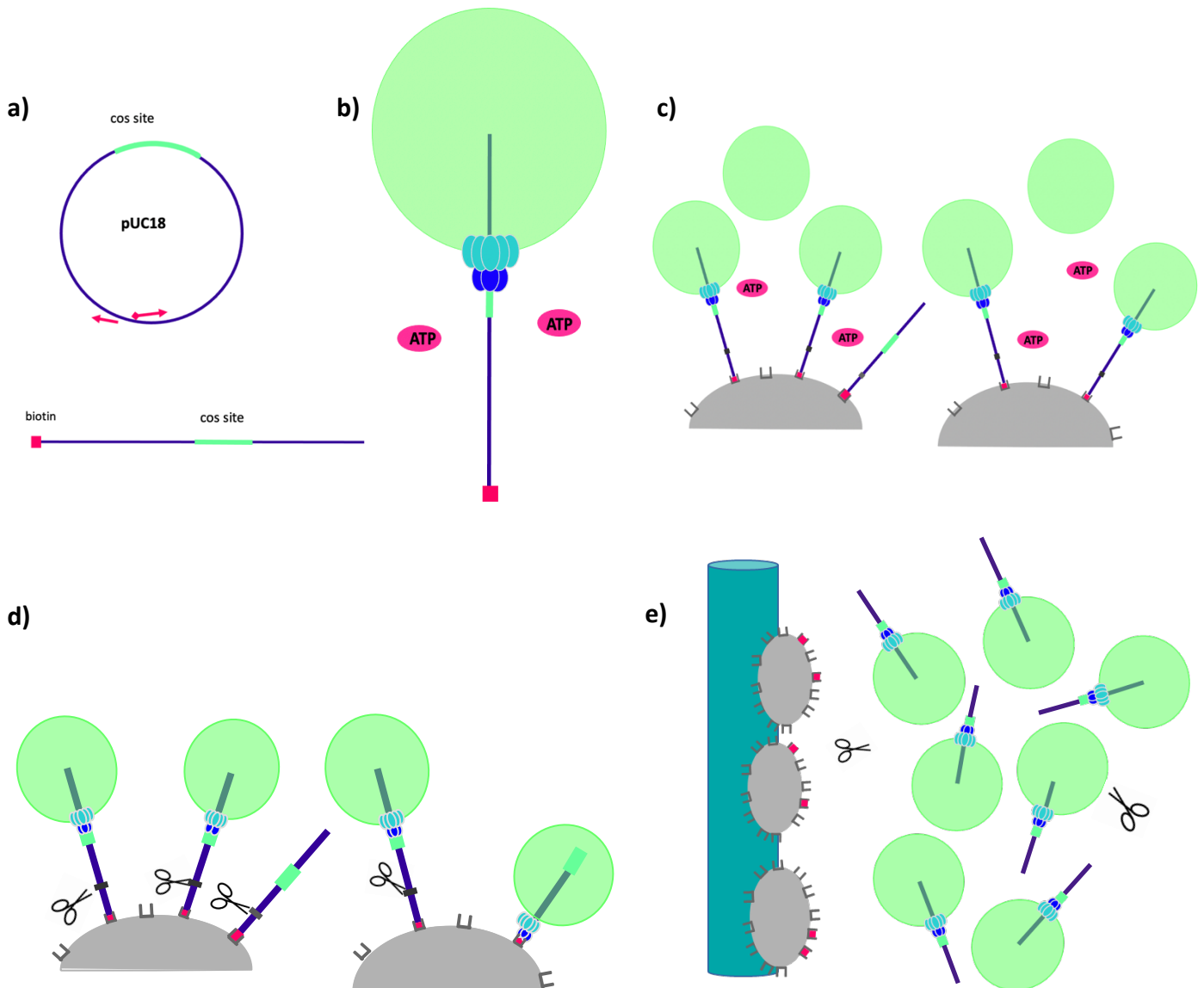


Figure 2.7: Schematic of magnetic bead assay to isolate stalled HK97 packaging complexes.
a) Biotinylated linear DNA containing cos site is synthesised by PCR b) Packaging of the DNA is instigated and stalled at the cos site using 75 μ M ATP c) Stalled complexes are bound to beads and contaminants and incomplete packaging complexes are washed away d) DNase digests all DNA e) packaging complexes are released and separated from beads

Optimisation of assay conditions

For the assay to be used as an effective purification step, multiple parameters required optimisation. The binding capacity of beads was first established, to allow for the minimum volume of beads while ensuring 100% anchoring of packaging complexes. This was determined by incubation of 10 μl beads with the biotinylated DNA substrate in the packaging buffer for 10 minutes. Unbound material was then removed and visualised on an agarose gel (Figure 2.8 a). The maximum DNA concentration which does not produce a visible band, thus represents maximum binding capacity. This was determined as 250 ng DNA per 5 μl of beads, and subsequent sample preparation used 40 μl reaction volumes with 1 μg DNA and 20 μl beads.

Cleavage of the bead-bound DNA to release the complexes from beads was also identified as a possible locus for sample loss. Complete cleavage is required for all complexes to be released, but incubation at 37 °C also needs to be minimised to prevent destabilisation of the sample. It was deemed unfavourable to flood the sample with restriction enzyme, as this would remain present as contamination in the final sample. 250 ng DNA bound to 5 μL beads was therefore incubated with 0.25 U of HF Scal over a time scale of 0 to 380 minutes. DNA which remained bound to beads was removed by a magnet and the resultant concentration of cleaved DNA examined on an agarose gel. Total cleavage occurs at after approximately 1 hour at 37 °C shown in (Figure 2.8 b).

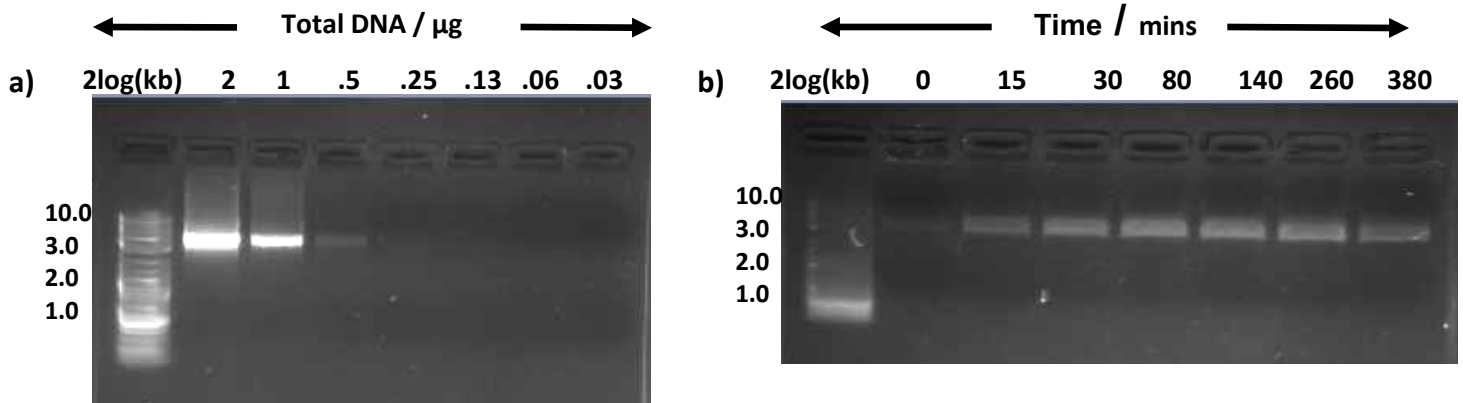


Figure 2.8: Optimisation of parameters for a magnetic bead purification assay. a) Agarose gel showing unbound DNA after incubation with streptavidin coated magnetic beads b) Agarose gel showing DNA released from streptavidin coated magnetic beads after incubation at 37 °C with restriction enzyme *ScaI* from 0 to 380 minutes

While restriction enzyme cleavage is highly specific, the long incubation time it requires was likely destabilising to complex preservation. Subsequent assays instead utilised DNase I in place of *ScaI*. This allowed incubation to be reduced to 10 minutes at room temperature, and degradation of any excess DNA present.

Collecting an initial cryo-EM data set

Using the conditions optimised, the released packaging complexes were examined by Transmission Electron Microscopy, initially by negative staining using the T12 instrument and subsequently under frozen hydrated conditions. Data was collected at Leeds using a 300kV Titan Krios instrument and processing was performed using RELION (219) as described in methods, Chapter 4. Data collection parameters are summarised in Table 2.1. Approximately 40,000 particles gave rise to a 3.6 Å resolution map of prohead II by applying averaging using icosahedral symmetry and a 4.1 Å resolution reconstruction of the portal protein using the C12 rotational symmetry. Figure 2.9 a) shows screening micrographs, with prohead and portal reconstructions depicted in Figure 2.9 b) and c) respectively.

Table 2.1: data collection parameters for initial cryo-EM data collection of stalled HK97 packaging complexes

Location	Leeds (Astbury Centre)
Grid type	Ultrathin Carbon Film on a Lacey Carbon Support Film
Hole size (μM)	N/A (lacey carbon)
Session	48 hr
Microscope	Krios 1
Detector (mode)	Falcon III (integrating mode)
Accelerating Voltage (keV)	300
Pixel size (\AA)	1.065
Nominal Mag	75,000 x
Total dose ($\text{e}\text{\AA}^{-2}$)	86
Number of fractions	79
C2 Aperture (μM)	70
Objective Aperture (μM)	100
Defocus range (μM)	-0.7 and -2.5

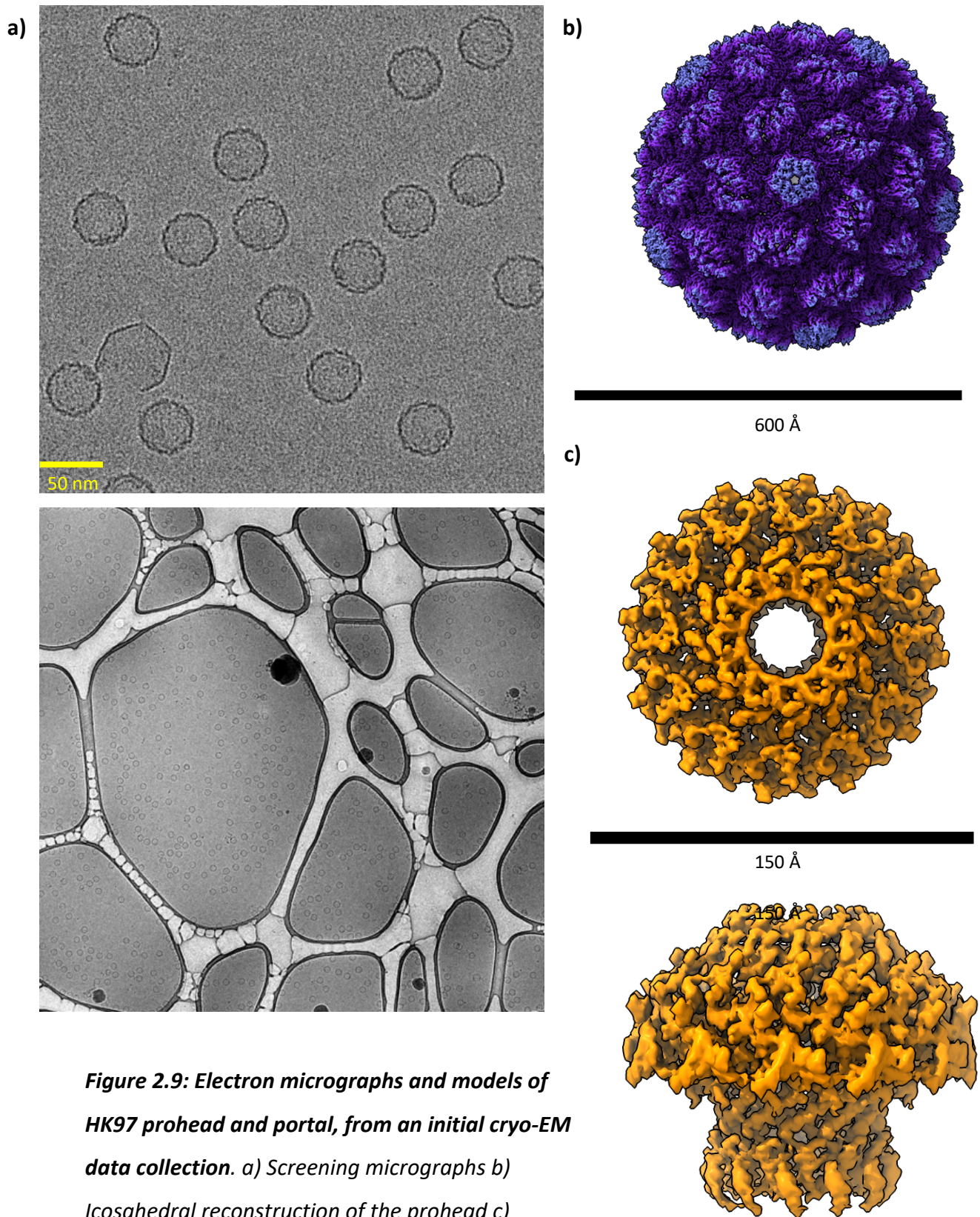


Figure 2.9: Electron micrographs and models of HK97 prohead and portal, from an initial cryo-EM data collection. a) Screening micrographs b) Icosahedral reconstruction of the prohead c) reconstruction of the isolated portal protein by applying C12 symmetry.

Despite extensive processing, there was no detectable signal for the motor. This was arguably caused by the absence of nucleoside within the wash buffers of the assay, causing large terminase to release ATP and lose grip of the DNA. In subsequent sample preparations the nucleoside was therefore added to both wash and elution buffers.

Introducing nucleosides and analogues to the assay to prevent motor detachment

The first nucleoside analogue trialled was the ATP analogue ATP γ S. This was used to stabilise the packaging motor of both Φ 29 and T4 (177,181), and likewise in other cyclic translocase ATPases such as P97 and FtsK (230,231). Protection of DNA in each step of the assay was monitored by DNA protection assays and visualised on an agarose gel. When ATP γ S is added to a sample of stalled packaging complexes, to a final concentration of 2.5 mM, DNA remains protected (Figure 2.10 a). After mixing with streptavidin coated beads, only a very faint band can be seen in the flow through, indicating most complexes are successfully bound. There is no band corresponding to full length DNA suggesting all DNA is bound to the beads. Meanwhile, following several wash steps with ATP γ S buffer, the eluted sample shows a reasonably strong band (Figure 2.10 a). This indicates effective retention and release of the stalled complexes and preservation of their stalled status.

To verify these conclusions, samples containing the stalled complex prior to bead-binding, unbound material, and eluted complex were each visualised using negative stain electron microscopy (Figure 2.10 b-d). The stalled sample before the bead-binding step showed a good concentration but was heterogeneous with contaminating tails. The unbound sample consisted of proheads that often appeared damaged or expanded. More interestingly, the eluted sample clearly showed proheads with density corresponding to a motor, and a small amount of DNA inside. Motors appear present on nearly all proheads and the sample did not appear to have orientation bias.

In light of the relatively large quantities of ATP γ S required for wash and elution buffers, alternative nucleosides and analogues were trialled for their stalling potential. ADP and AMPpPi as well as ATP at constant concentration of 75 μ M were each trialled. Surprisingly,

maintaining ATP concentration at 75 μM seemed to preserve stalled motors with near equivalent efficacy as the addition of ATP γS .

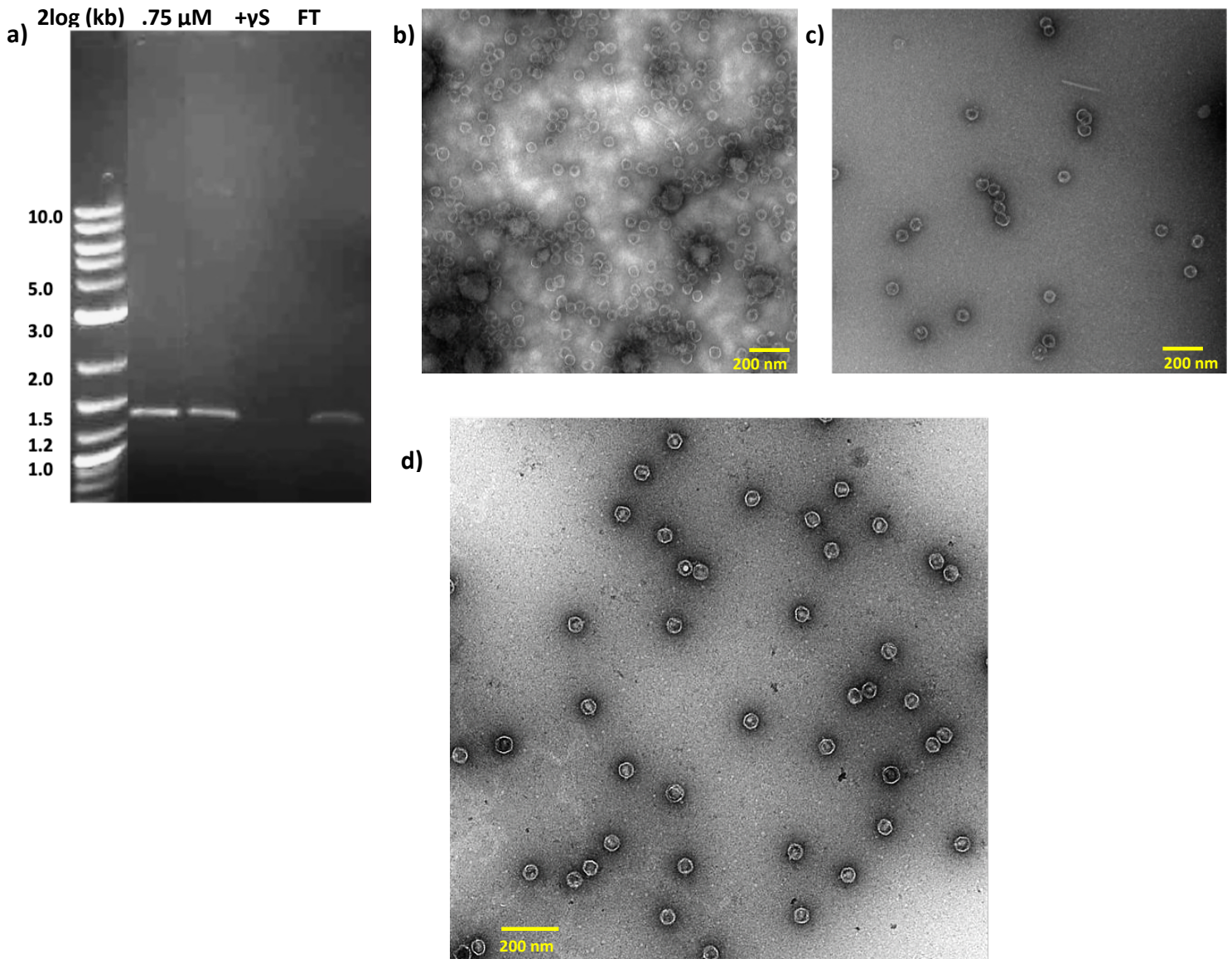


Figure 2.10: Optimised production of stalled HK97 packaging complexes. a) Agarose gel showing protected DNA after a HK97 packaging assay at 75 μM ATP, unbound material on mixing with streptavidin coated magnetic beads and eluted material by DNase. Negative stain micrographs also depicting b) stalled packaging complexes pre binding c) unbound material d) eluted material

Discussion

Optimisation of the purification of HK97 prohead II

The purification protocol of prohead II was firstly altered to avoid partial expansion and produce uniform proheads in the native state for DNA packaging initiation. This required dilution into a high ionic strength buffer following elution from a POROS HQ/20 column. Ultracentrifugation then produced a prohead pellet which could be resuspended into the buffer of choice. Thus, proheads were no longer subject to a sudden decrease in ionic strength, which may have instigated expansion. Alterations in buffer conditions (specifically pH) have previously been shown to drive the expansion pathway of HK97 prohead II (224).

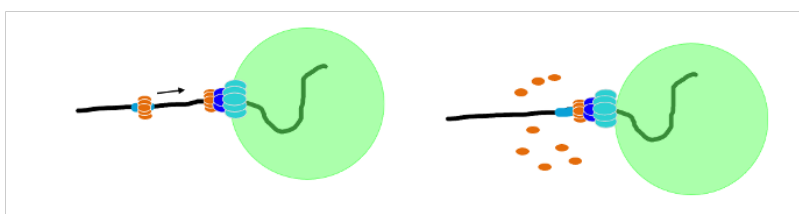
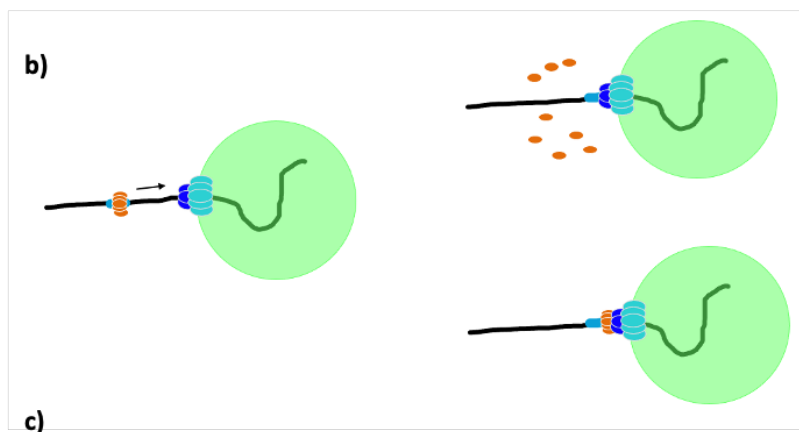
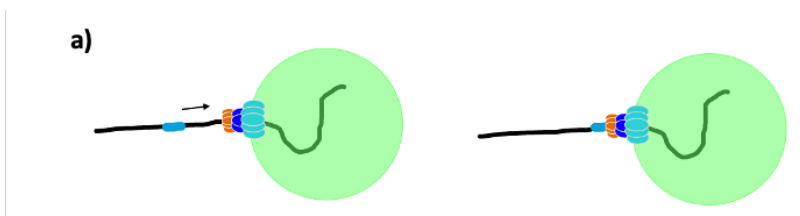
Stalling of the HK97 DNA packaging motor

Next conditions for stalling the packaging motor complex at the *cos* site of a DNA substrate were explored. The ideal ATP concentration was pinpointed at 75 μM which is roughly half the estimated ATP concentration within the *E. coli* cytosol (232). Large terminase and small terminase were present at 3 μM and 6 μM respectively, corresponding to molar ratios of 100:1 and 200:1 which is significantly higher than their predicted oligomeric states of 5 and 9. Both small terminase and an integral *cos* site were essential for stalling.

Stalling of the HK97 motor at a *cos* site likely represent an equilibrium between the forward packaging force provided by large terminase and the binding energy between small terminase and the *cos* site. At higher concentrations of ATP, large terminase provides substantial power to package past the *cos* site. The introduction of additional ATP, after stalling, also appears to cause motors to restart (Figure 2.6). This is demonstrated by a smearing of higher molecular weight bands of protected DNA (indicating that longer sections of DNA have been packaged) but an absence of shorter lengths of DNA (indicating new packaging events have not started). This suggests that motors remain in an active conformation.

It is unclear if small terminase remains bound to large terminase throughout packaging. Three possible models can explain stalling: firstly, oligomeric small terminase creates an

extension of the DNA channel within large terminase, and DNA is fed through the double ring from the free DNA end. On reaching the *cos* site, the small terminase DNA binding affinity increases, causing the whole motor to stall. In the second model small terminase is already present at the *cos* site and provides a roadblock to the incoming large terminase. On reaching the small terminase bound *cos* site, large terminase may interact with small terminase, or displace it. However, in this situation it is unclear how small terminase could stimulate large terminase processivity (147,148). The third option represents a combination of models 1 and 2, with one small terminase oligomer bound to large terminase throughout packaging, and a second already bound to the downstream *cos* site. These hypotheses are summarised in Figure 2.11.






 portal
  large terminase
  small terminase

Figure 2.11: Potential mechanisms of stalling at the *cos* site by HK97 small terminase. a) small terminase remains bound to large terminase throughout packaging b) small terminase is bound to the downstream *cos* site throughout c) separate small terminases are bound to large terminase and the downstream *cos* site

In vitro, two other elements are likely to contribute to cleavage, as opposed to stalling, at the *cos* site. HNH like endonuclease gp74 has been shown to interact with large terminase

and to stimulate cleavage at the *cos* site even in the absence of ATP (233) (1). As for *pac* viruses, (107,111,189,234) 'headful' pressure is suspected to contribute, in part, to phage packaging termination. Pressure changes within the prohead are thought to be relayed to large terminase via structural changes in the portal. The length of DNA used in packaging assays (3.3 kb) is only ~12 % of the HK97 total genome length of 39.7 kb, so there is unlikely significant internal pressure within the system at the time of stalling. This could result in the endonuclease domain of the large terminase being partially engaged, but not fully stimulated. However, this hypothesis is challenged by the success in restarting motor packaging without DNA cleavage.

Development of a magnetic bead assay for selective purification of stalled packaging complexes

The ability to stall the motor gave rise to the development of a magnetic bead-based assay, to specifically select for stalled packaging complexes. Unfortunately, analysis of the eluted sample by cryo-EM indicated that the motor was not retained by the prohead. By maintaining the ATP concentration throughout the assay, or flooding the stalled sample with ATP γ S, negative stain electron microscopy indicated that the packaging machinery could be kept intact. It was concluded that the motor must be present in the ATP bound state in order to retain grip of the DNA, as predicted by the glutamate switch hypothesis: large terminase subunits are expected to cycle between an ATP bound state with high affinity for DNA, and an ADP bound state with a low affinity for DNA(177,185).

Conclusion

Stalling of the HK97 DNA packaging motor is dependent on the interaction between small terminase and the *cos* site, which causes large terminase to stop packaging at ATP concentrations close to 75 μ M. DNA protection assays suggest that the DNA is not cleaved, as occurs during packaging termination *in vivo*, and that motors can be restarted upon

addition of ATP. However, it is unclear whether small terminase remains bound to the motor on stalling.

An assay was designed for the selective purification of stalled packaging complexes. Here, biotinylated DNA, emerging from stalled complexes, is bound to streptavidin coated beads. Incomplete complexes are then removed, and the beads are subject to further wash steps to remove excess proteins. Isolated stalled complexes are released by DNase and beads are removed using a magnet. Initial screening by negative stain electron microscopy and cryo-EM revealed the need for a constant nucleoside analogue during wash steps and elution. When excess ATP γ S was present throughout, or a constant ATP concentration of 75 μ M, analysis by negative stain electron microscopy revealed visibly stalled motors. The sample was homogenous and suitable for further analysis by cryo-EM.

Chapter 3: Increasing the concentration of the stalled HK97 packaging sample

Introduction

Based on the conditions established, described in Chapter 2, a cryo-EM data set of stalled HK97 packaging complexes were collected. Particles in micrographs were well spaced, resembled unexpanded procapsids and showed nearly 100% incorporation of the motor. This indicated success of the magnetic bead assay designed to select for stalled packaging motors. The data set was processed using RELION (219) as discussed in methods Chapter 4. The prohead and motor could be clearly resolved. However, the total number of particles (close to 10,000) was insufficient to allow a reconstruction of the motor to be derived.

There are several intrinsic reasons limiting the resolution of the motor. Firstly, the large dimensions of the prohead mean that only a limited number of particles can be present on a single micrograph before overlapping. For a reconstruction of the prohead, the low particle number may be compensated for by the high-order symmetry (icosahedral). However, for an asymmetric feature such as the motor, many more particles will be needed to obtain a reasonable resolution.

Secondly, stalling the motor using low ATP concentrations, produces a faint DNA band relative to the strong band of protected DNA witnessed using high concentrations of ATP. From this we can conclude that much DNA remains unpackaged and correspondingly many proheads are not constituents of active packaging complexes. Therefore, the total concentration of stalled packaging particles, released from the magnetic bead assay will be far lower than the input concentrations. In addition, not all of the stalled complexes will be retrieved by the assay: proheads show a tendency to adhere to surfaces and may well be lodged onto the beads even after the addition of DNase.

In this chapter, concentration of the sample is explored and optimised. Many techniques are rendered unsuitable by the fragility of the sample, and requirements of the assay.

Ultimately it is necessary to remove the biotin selection to achieve the concentration and number of particles suitable for cryo-EM analysis. Thus, efforts are diverted to enhancing stalling efficiency, rather than specifically selecting stalled complexes.

Methods

Cryo electron microscopy screening

Grids for cryo-EM were prepared using the Vitrobot IV instrument. Initial samples used 3 μ l of sample blotted onto Quantifoil 1.2/1.3 400 mesh copper grids before blotting and plunge freezing in liquid ethane. The starting parameters used were blot time 3 s, blot force 0, 10 °C and 100 % humidity. Subsequent optimisation steps are summarised in results. Grids were imaged using Tecnai 12 G2 microscope using a ISS MegaView III CCD camera at 120 kV a.

Data collection parameters

Data collection was performed at eBIC using a Titan Krios 2 microscope with a Gatan K3 detector. The collection proceeded for 48 hours and was controlled remotely from York. A large pixel size was chosen in order to maximise the number of particles and the defocus range was relatively close to focus owing to the large particle size, Table 3.1 below:

Table 3.1: data collection parameters for initial cryo-EM data collection of stalled HK97 packaging complexes

Location	eBIC
Grid type	Ultrathin Carbon Film on a Lacey Carbon Support Film Carbon Support Film
Hole size (μM)	N/A (lacey)
Session	48 hr
Microscope	Krios II
Detector (mode)	Gatan K3 (counting super-resolution)
Accelerating Voltage (keV)	300
Pixel size (\AA)	1.065
Nominal Mag	75,000 x
Total dose ($\text{e}\text{\AA}^{-2}$)	50
Number of fractions	45
C2 Aperture (μM)	70
Objective Aperture (μM)	100
Defocus range (μM)	-0.3 and -1.5

Concentration of the biotinylated sample using streptavidin coated affinity grids

Streptavidin coated C-SMART grids were purchased from Dune Sciences. A hydration chamber was created in a glass petri dish with a base of damp filter paper. Droplets of blotting buffer, sample, wash buffer and stain were placed on a parafilm wrapped slide and the grid floated on top. Grids were blotted with filter paper between droplets. See Table 3.2 for details.

Table 3.2: Sample preparation of stalled HK97 packaging complexes for visualisation by transmission electron microscopy on streptavidin coated affinity grids

	Composition	Volume (μ l)	Incubation time (mins)
Blotting buffer	20 mM Tris pH 7.5, 10 mM MgCl ₂ , 50 mM KGlu, 0.5% w/v Tween-20	10	30
Sample	Stalled packaging complexes diluted in wash buffer	10	15
Wash buffer	20 mM Tris pH 7.5, 10 mM MgCl ₂ , 50 mM KGlu, 5 mM ATP γ S	10 x 3	5 x 3
Stain	1% w/v Uranyl acetate	10	3

Overexpression and purification of small terminase

Fresh small terminase was expressed and purified for the assays and screens described in this chapter.

The small terminase gene (amino acids residues 2–161) was provided in pET22a-based vector with an N terminal His and SUMO tag (235). This was transformed into *E. coli* BL21(DE3) pLysS cells. Bacteria were grown in LB medium at 37°C with 30 μ g/mL kanamycin and 33 μ g/mL chloramphenicol, and induction occurred at OD₆₀₀ \approx 0.6 for 4 hours. Cells were harvested at 4 °C, and suspended in 50 mM Tris pH7.5, 1M NaCl, 5% v/v glycerol, 30 mM Imidazole with additional protease inhibitor, DNase and RNase stimulated by 1mM MgCl₂. The resuspension was next sonicated in 3 second blasts, at 75% power for a total time of 20 minutes.

Lysate was spun at 20,000 xg for 40 minutes. The supernatant was then applied to a 5 ml His-trap column and a washed with 3M salt for 5 column volumes to remove bound nucleic acid, as judged by a reduction in 260/280 absorbance ratio. Protein was eluted in 1M

imidazole in a step elution and diluted two-fold in 25mM Tris pH 7 to reduce the salt concentration to 500 mM. The elution was then incubated overnight at 4 °C with in a 1:100 w/v ratio with Ulp1 protease stimulated by 2 mM DTT.

Cleaved protein was applied to a 5 ml Hi-Trap heparin column and eluted in a salt gradient of 25mM Tris pH 7.5, 200 mM NaCl to 1M NaCl. The remaining uncleaved protein eluted separately from the main peak allowing it to be removed from the sample. Cleaved protein was then diluted into 25mM Tris pH 7 to a salt concentration of 200 mM. This was applied to an 8 ml Mono S column, where elution into 25 mM Tris pH 7 with 1M NaCl produced 2 closely separated peaks. Each was concentrated in a 10 kDa Amicon® cellulose concentrator and purified by gel filtration.

Results

Cryo-EM analysis of an initial data set

The sample preparation parameters optimised in Chapter 2 are summarised in Table 3.1, Although the sample showed a reasonable concentration in negative stain micrographs, the concentration for cryo-EM is typically ~10 times lower (236). Therefore, several different grid types were selected with the aim to enhance concentration. Firstly, gold grids were chosen as they can minimise the unwanted specimen movement (237). Grids with a thin (~2nm) layer of amorphous carbon over lacey carbon were selected as the particles tended to show an affinity for the carbon support. Although background signal from the carbon can be problematic for achieving very high-resolution reconstructions, obtaining an adequate number of particles was deemed more important for the motor reconstruction (238). Poly-lysine coated carbon support grids were also trialled, as such grids are designed to adhere negative particles to the surface (239). Finally, graphene oxide coated grids were selected as they are also thought to encourage adherence of proteins in dilute solutions, whilst producing less background signal (240). An extremely weak blot force and long blot time were selected during plunging, in an attempt to minimise the number of particles removed by blotting.

Table 3.3: Vitrification conditions for cryo electron microscopy screening of stalled HK97 packaging complexes

Sample volume (ul)	Grid type	Glow discharge charge (mA)	Glow discharge time (s)	Blot force	Blot time (s)
6	Gold	20	180	-10	4
6	Lacey	20	30	-10	4
6	Lacey + poly-lysine	20	30	-10	4
6	Graphene coated	20	30	-10	4

Screening of grids prepared for cryo-EM is depicted in Figure 3.1. Clearly the concentration is low for each of the grid type. Poly-lysine grids showed some areas of high protein density, but this appeared to represent broken and aggregated capsids. Lacey carbon and gold grids were selected as the most promising, and 2 fresh grids of each were glow discharged as before. For each grid type, 3 μ l of both the dilute and concentrated samples was applied using the Vitrobot IV, with blot force -10 and blot time 1s. Each condition was also repeated with three sequential sample applications in attempt to enhance particle adhesion. The initial two applications were performed by hand with the last application using the Vitrobot IV as described (241). The grid selected for data collection was applied three times onto a Lacey Carbon grid.

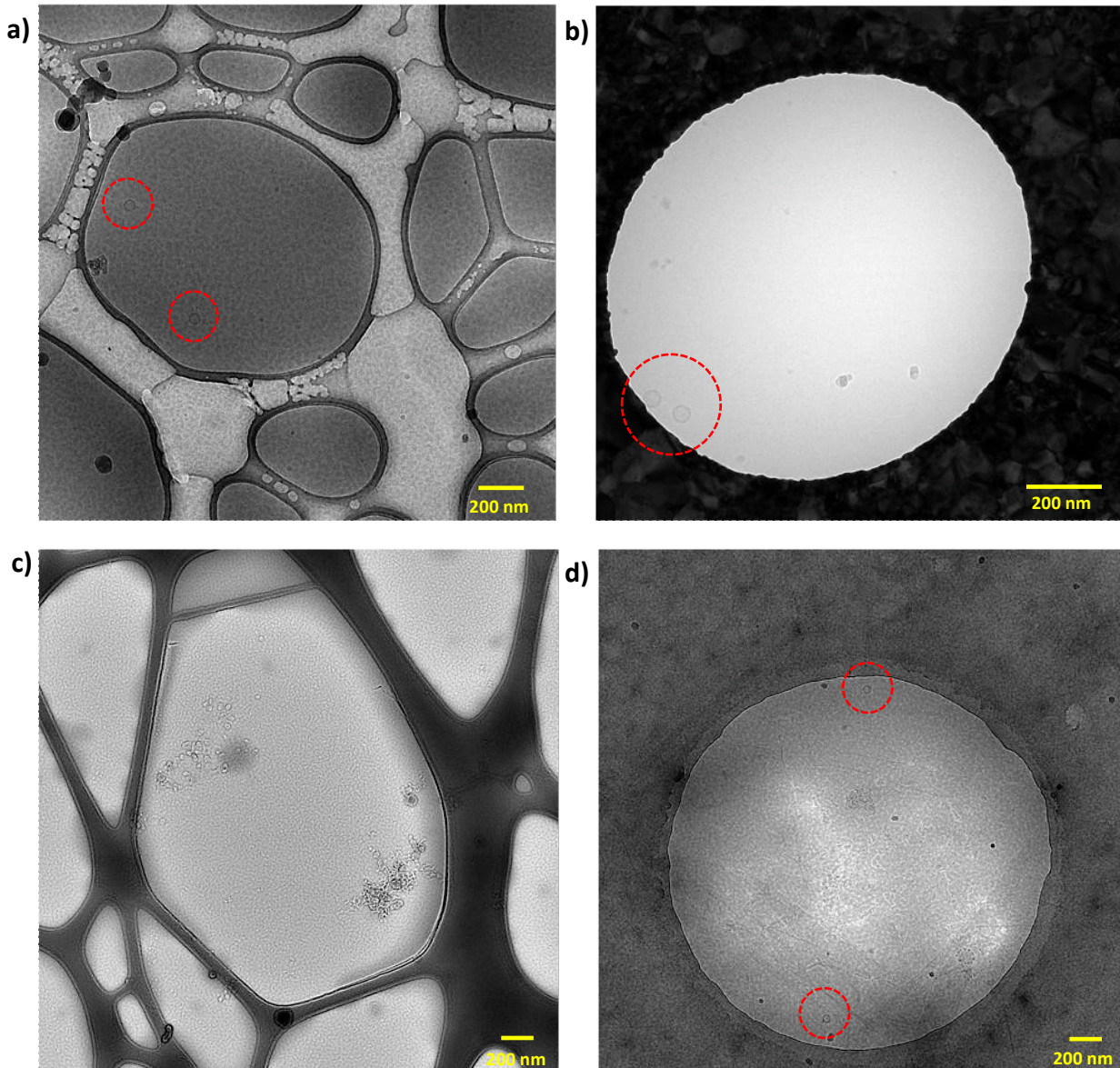


Figure 3.1: Cryo-EM screening micrographs for stalled HK97 packaging complexes isolated by a magnetic bead assay. Obvious particles are circled in red. a) Lacey carbon grid, Ultrathin Carbon Film on a Lacey Carbon Support Film, 400 Mesh Cu b) Gold grid UltrAuFoil R2/2 200 c) poly lysine coated lacey carbon grid, Ultrathin Carbon Film on a Lacey Carbon Support Film, 400 Mesh Cu grid. d) graphine oxide coated Gold grid UltrAuFoil R2/2 200

Processing the cryo-EM data

Data were processed using the RELION software (219), as described in methods in Chapter 4. Particles were picked using RELION autopick, using a low pick threshold to ensure no particles were missed. Any expanded capsids, ice or carbon could then be removed manually due to the low particle numbers. Maps were obtained for both the prohead and portal, using I3 and C12 symmetric averaging, respectively. These structures will be discussed in later chapters following the collection and processing of subsequent data sets, which gave rise to higher resolution reconstructions.

Encouragingly, the density below the portal where the motor is situated appeared strong. Selective masking for the motor region, produced a comparable number of particles as for the portal region alone. This indicated that a majority of portals had retained bound motors. However, for 3D classification, masking for both the motor alone, and motor and portal complex, could not reveal much detail on the motor structure. It appeared that the motor was present in multiple conformations with classification into 2 classes producing one largely 'planar' conformation compared to a second visibly 'twisted' class (Figure 3.2).

However, improving the resulting resolution of the reconstruction was unfeasible based on the low number of total particles.

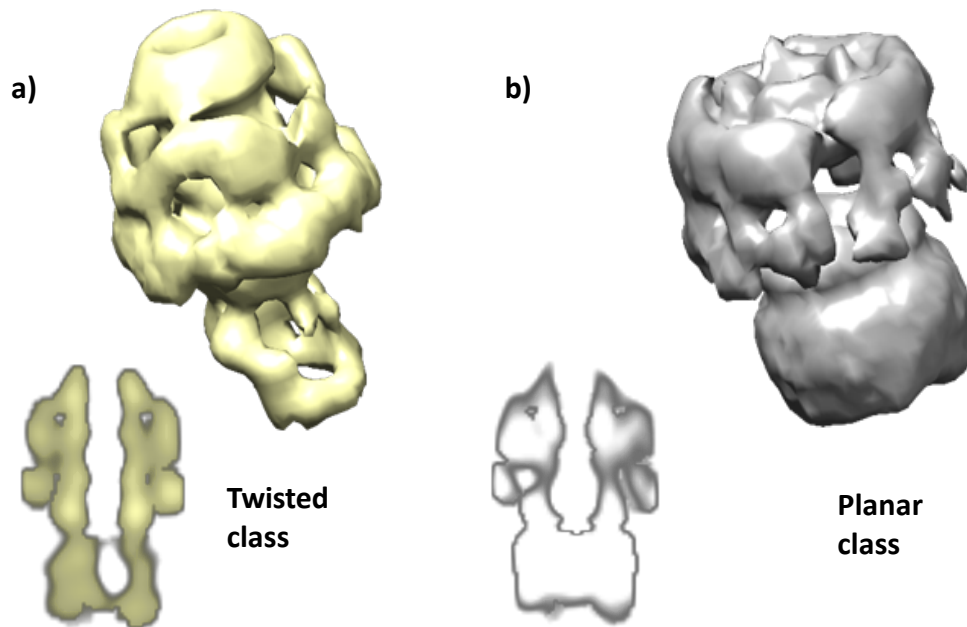


Figure 3.2: Distinct 3D classes of the HK97 packaging motor. a) a twisted and b) planar conformations.

Concentration of the final sample

In light of the processed data, efforts were diverted to increase the number of particles, which was deemed to be the major factor limiting the resolution. A fresh sample of the small terminase protein was produced for these optimisations, as described in methods. In theory, concentrating the sample should be possible simply by eluting from the beads into a smaller volume. However, in practice no significant increase was observed by negative stain electron microscopy when the elution volume was reduced from 25 μL to 12.5 μL .

Presumably, using this smaller volume (which was smaller than the initial bead volume) was unfeasible as the elution process required full resuspension of beads. Therefore, the possibility of scaling up, simply by using a larger reaction volume and larger number of beads would also require a larger elution volume and produce a sample of the same concentration.

Instead, numerous techniques were employed to reduce the volume of the final sample post-assay. However, concentration of such a delicate sample created new challenges. Firstly, the integrity of the packaging complexes is likely time sensitive: the bead assay alone takes a minimum of 40 minutes and any additional time allowed for the sample to incubate is undesirable. Secondly, substantial alteration to buffer composition may cause dissociation of the motor or prohead expansion.

Concentration using a PET concentrator

The first obvious technique employed to concentrate the sample was the use of a Vivaspin® PES membrane centrifugal concentrator with a MWCO of 300 kDa. A standard bead assay was performed, using a reaction volume of 300 µl and an elution volume of 150 µl. The concentrator was equilibrated in elution buffer at 4 ° C, and the elution applied. This was spun at a low speed (1000 xg) in 2-minute intervals with intermittent mixing. These conditions were selected to minimise the harsh forces applied to the complexes, and so keep them intact. One possible advantage of this technique is that DNase I (MW 31 kDa) should pass straight through the filter, in addition to any excess large or small terminase. Once the volume had been reduced 10-fold to 15 µl, 5 µl was applied to a grid and examined by negative stain electron microscopy (Figure 3.3 c).

Whilst some motors appeared intact, the concentration does not appear to be significantly increased (Figure 3.3 c). This is likely a result of proheads adhering to the filter paper within the concentrator. This has been previously demonstrated when optimising the production and purification of proheads (Chapter 2 Figure 2.4 b)

Concentration by PEG precipitation

An attempt to draw water out of the sample by osmosis, thereby increasing the concentration of packaging complexes, was performed by adding PEG 8000 to a dialysis buffer (242). The buffer conditions within the sample should remain constant, as small molecules can pass through a cellulose acetate dialysis membrane freely, whilst PEG molecules and proteins cannot, and so remain separate. Once again, a standard magnetic

bead assay was performed, with a reaction volume of 300 μl and an elution volume of 150 μl . The elution was divided into 4 equal volumes, and each placed into a 1ml Costar™ Spin-X™ Centrifuge Tube Filter. These were suspended in small beakers of identical buffer, one with additional 12.5% w/v PEG 8000, one with 25 % w/v PEG 8000 (the saturation concentration), and the final beaker filled with dry PEG 8000. Samples were incubated on ice for 1 hour. The only condition which resulted in a substantial reduction in volume was use of dry PEG 8000. Here, the volume dropped to approximately 10 μl , and this was examined by negative stain electron microscopy (Figure 3.3 a). Clearly the sample is aggregated and unusable presumably as result of the sudden osmotic shock.

Concentration by evaporation

A gentler way of decreasing the buffer volume was attempted using evaporation. An 150 μl sample elution of packaging complexes was subject to gentle airflow, using a nitrogen tap. After approximately 30 minutes incubation on ice, the initial sample volume had reduced 15-fold to 10 μl . 5 μl was examined by negative stain electron microscopy (Figure 3.3 b).

This technique proved highly effective, as shown by the large number of proheads. However, whilst some viruses appear to retain intact motors, a majority do not. There is also a strong level of protein background contamination, presumably consisting of concentrated DNase I and detached large and small terminase. The initial ionic strength of 50 mM K₂Glu would have increased to 500 mM, following evaporative concentration, whilst the MgSO₄ concentration would have increased to approximately 100 mM. These condition alterations are not only unsuitable for protein-protein and protein-nucleic acid interactions but also can impact on vitrification and image quality in cryo-EM. In addition, evaporation may increase the number of particles at the air-water interface, by decreasing the drop volume. This could also contribute to destabilising the sample (243).

Concentration by ammonium sulfate precipitation

Ammonium sulfate, $(\text{NH}_4)_2\text{SO}_4$, can induce precipitation of soluble proteins at high concentrations. This well-established technique for the purification of protein samples is also known as 'salting out' (244). Following a 300 μl magnetic bead assay, the eluted 150 μl HK97 packaging complexes were added to a ~ 4.2 M solution of $(\text{NH}_4)_2\text{SO}_4$, pH 7, in a 1:1 volume ratio. The mixture was incubated on ice for 30 minutes. Precipitation was then collected by centrifugation at 2000 xg. The pellet was resuspended in 10 μl of sample buffer and spun again at 15,000 rpm, for a further 10 minutes, to remove remaining precipitation. This volume reduction should produce a 10-fold increase in the protein sample concentration. The supernatant was visualised using negative stain electron microscopy, which showed a highly concentrated sample (Figure 3.3 d). However, in parallel to the evaporation condition, very few motors were visibly attached to proheads. This may be a result of the extra incubation period, or the harsh forces applied during the high-speed centrifugation.

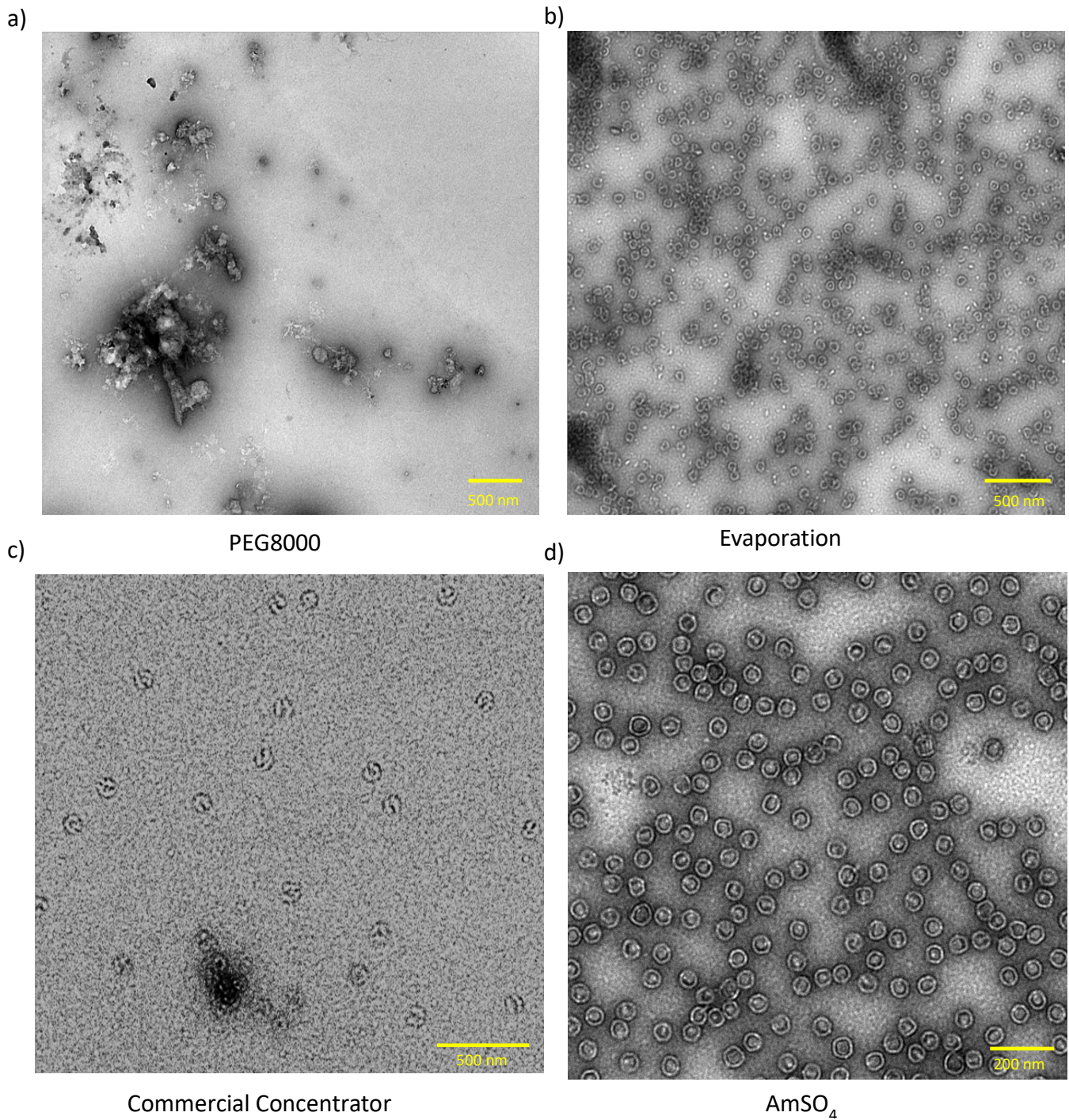


Figure 3.3: Monitoring of sample concentration techniques by negative stain electron micrographs. Samples were prepared by different techniques to attempt to concentrate stalled packaging complexes following a magnetic bead assay using a) PEG precipitation b) evaporation c) a commercial 300 kDa concentrator and d) Ammonium Sulphate precipitation

Adjustments to the magnetic bead purification assay

Unfortunately, efforts to concentrate the magnetic bead elution, intact, proved broadly unsuccessful. Adaptations of the assay protocol was thus considered as an alternative to achieve a higher concentration of particles. Whilst the magnetic bead assay proves highly effective at isolating stalled packaging complexes, there was an approximate 10-fold reduction in concentration observed from the protein input. Efforts were made to reduce this sample loss and to find alternative methods to isolate stalled packaging complexes.

Increase in ionic strength of magnetic bead assay buffers

An increase in ionic strength within the wash/elution buffer was trialled as a means to inhibit non-specific adhesion of proteins to beads and thus increase recovery. Figure 3.4 a) shows a negative stain electron micrograph of packaging complexes eluted at increased ionic strength (250 mM KGlu). The number of particles eluted from the streptavidin beads is greatly increased, relative to 50 mM KGlu buffers, and appears a suitable concentration for cryo-EM. This is supported by DNA protection assays for samples eluted from the beads at 50 mM KGlu compared to 250 mM KGlu. Figure 3.4b shows significantly more DNA is protected within proheads, when a higher ionic strength buffer is used for washing and elution.

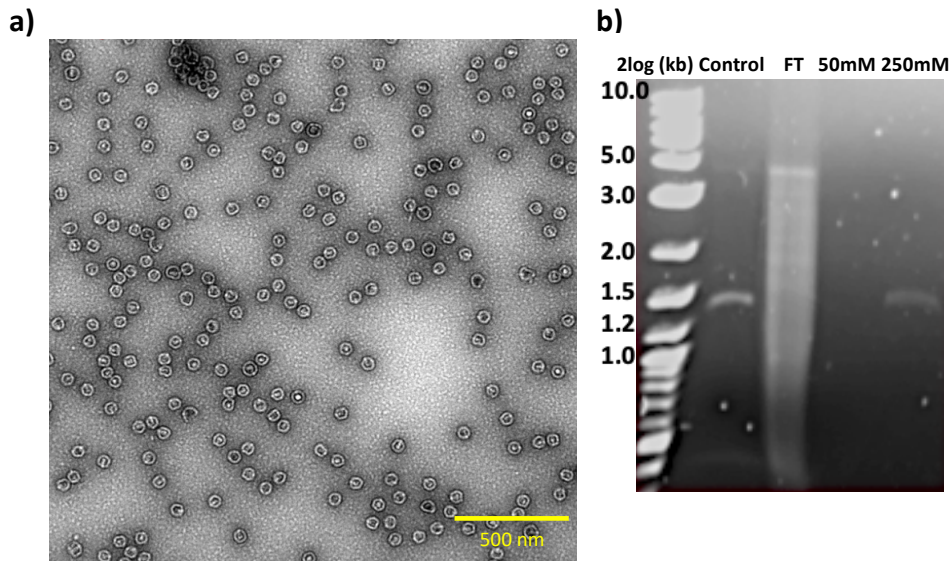


Figure 3.4: Purification of stalled HK97 packaging complexes by magnetic bead assay at 250 mM K_{Glu} a) negative stain electron micrographs depicting a sample of HK97 stalled packaging complexes following a magnetic bead assay at constant K_{Glu} concentration of 250 mM b) Agarose gel electrophoresis showing protected DNA after a HK97 packaging assay at 75 μ M ATP, and eluted material from streptavidin coated magnetic beads by DNase at 50 mM and 250 mM K_{Glu}

Whilst these results are encouraging, it is of note that very few motors appear visibly attached to the proheads on micrographs. Further packaging assays were performed to probe motor retention. Figure 3.5 a) indicates that at higher ionic strength, packaging at high concentrations of ATP actually appeared more efficient, shown by a stronger band of protected full-length DNA. Meanwhile, at low concentrations of ATP, stalling of the motor was apparently unaffected by ionic strength.

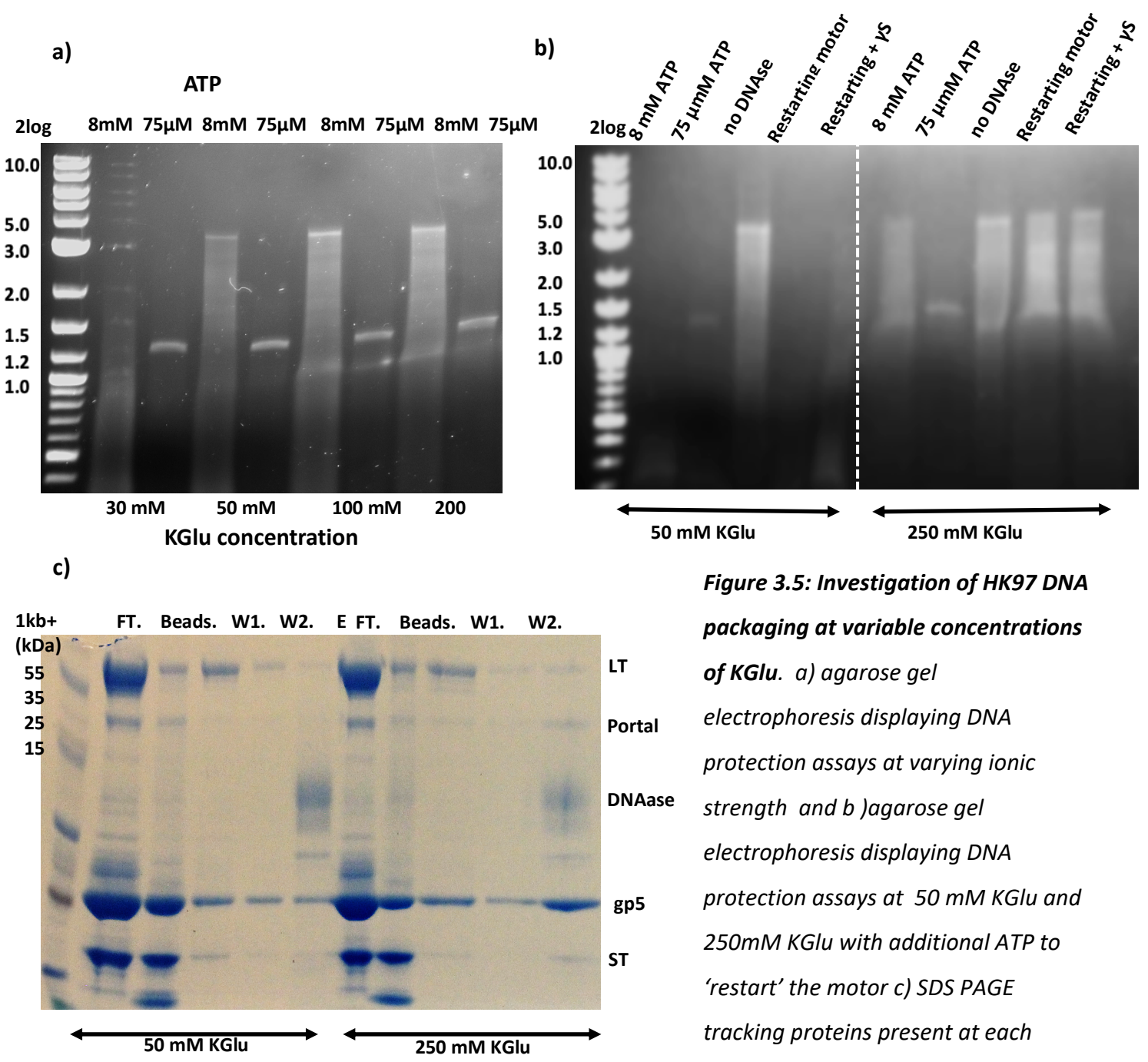


Figure 3.5: Investigation of HK97 DNA packaging at variable concentrations of KGlU. a) agarose gel electrophoresis displaying DNA protection assays at varying ionic strength and b) agarose gel electrophoresis displaying DNA protection assays at 50 mM KGlU and 250mM KGlU with additional ATP to 'restart' the motor c) SDS PAGE tracking proteins present at each stage of streptavidin magnetic bead assay performed with 50 mM KGlU and 250 mM KGlU

An alternative explanation for the lack of motor retention is activation of the endonuclease activity of large terminase. This would cause cleavage of DNA at the *cos* site, and *in vivo*

leads to subsequent detachment of the motor. To determine this, DNA protection assays were performed at 50 mM K₂Glu and 250 mM K₂Glu, in the absence of DNase, to see if full length DNA remained intact (Figure 3.5 b). A faint band at the *cos* site at high salt suggests cleavage is a possibility, though incomplete. To probe further, additional ATP was added after stalling in attempt to restart motors. The smear of protected DNA between the *cos* site and full-length DNA at high salt indicates that at least some motors can package through the *cos* site when additional ATP is added, and thus remain active and engaged with the portal.

Next, magnetic bead assays performed with a constant K₂Glu concentration of 50 mM and 250 mM were tracked by SDS-PAGE. Here we can see an obvious increase in the quantity of the major capsid protein gp5 eluted at increased ionic strength (Figure 3.5 c). Accordingly, a band corresponding to the portal protein is also visible. A faint band is present for large terminase at both ionic strengths, but the increase is less discernible. This is in accordance with the negative stain images at high ionic strength which show decidedly more proheads, where only a small proportion carry a visible motor. Small terminase does appear more prominent at a high ionic strength but this may, at least partially, represent DNA bound small terminase without the associated large terminase and motor.

Together these findings indicated that increasing ionic strength does result in a dramatic increase on the concentration of the sample eluted from the magnetic bead assay. However, this does not translate to an increase in particles with attached motors. Preservation of buffer conditions was therefore considered important.

Streptavidin coated affinity grids for concentration

Streptavidin-coated grids were selected as a potentially ideal alternative for increasing the concentration of a biotinylated sample (245). Here, the grid is effectively acting in place of the streptavidin beads, with biotin-streptavidin binding fixing packaging complexes onto the grid *via* the DNA (Figure 3.6 a). Grids are not glow discharged, so that unbound proteins can be washed away easily, thus removing empty proheads, excess terminase, and motors which have packaged the full length of DNA and biotin label. The particle concentration should not significantly vary between negative stain and cryo-EM samples. Furthermore, the

long flexible DNA linker (approximately 2.2 kb) between the biotinylated end and *cos* site should provide ample freedom of movement for the packaging complex and should not contribute to orientation bias (Figure 3.6 b).

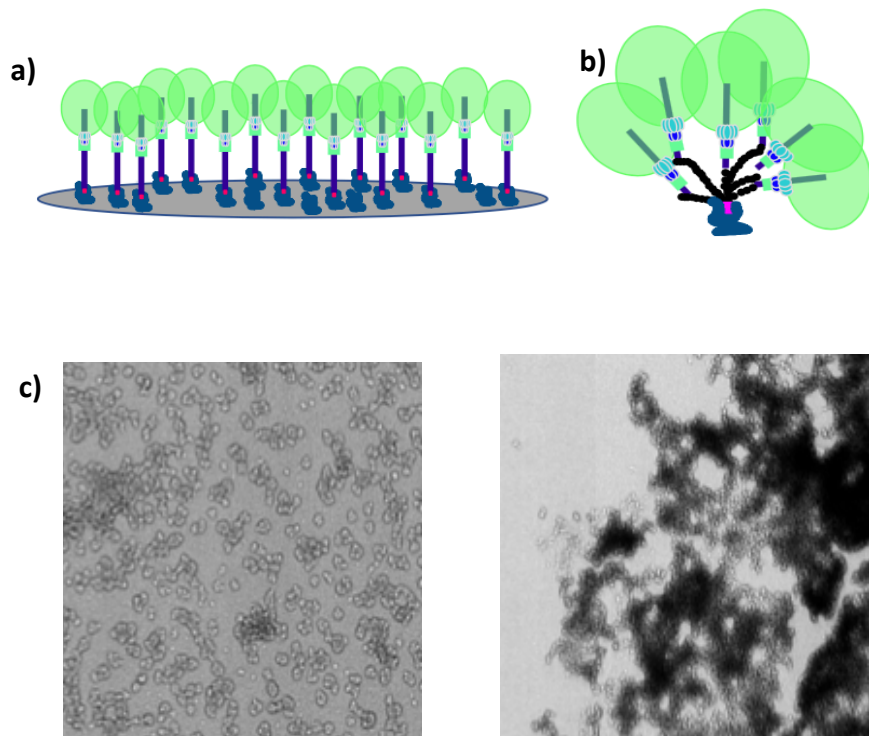


Figure 3.6: Use of streptavidin affinity grids for concentration of stalled HK97 packaging complexes. a) Concentration of stalled HK97 packaging complexes onto an affinity grid via biotin-streptavidin binding b) flexibility provided by DNA linker c) negative stain electron micrographs of stalled HK97 packaging complexes on a streptavidin grid

Dilutions in standard packaging buffer were trialled to a maximum dilution of 10^{-5} , which was predicted to be the lowest concentration detectable by negative stain electron microscopy, assuming all particles present bound. However, in each case proheads were clustered into large aggregates and appeared deformed (Figure 3.6 c,d). Meanwhile, for more dilute samples swathes of the grid appeared empty.

In an attempt to limit aggregation, additional biotinylated DNA was added to the sample as a competitive binder (246). Each streptavidin binds four biotin molecules cooperatively

(247) and so binding of multiple complexes to one streptavidin may crowd particles. Excess biotinylated DNA could therefore help to space out complexes on the grid. Biotinylated DNA was added in 10-fold and 100-fold excess of the number of biotinylated complexes in the 10^{-4} diluted sample. However, the grid appearance was unchanged, and particles remained aggregated and distorted.

Removing the biotin selection purification step

In a final attempt to increase the number of motor particles, stalled packaging complexes were applied directly to grids without a biotin selection step. This greatly increased the concentration of particles, as predicted, but did not select between empty proheads, stalled packaging complexes, and those which had packaged the full length of DNA. Furthermore, an excess of large terminase, small terminase and DNA remained present in the sample. The inherent limitation on the number of motor particles per micrograph, based on the large size of the associated proheads, means retaining a high proportion of motors is essential. A number of protocol modifications were trialled in attempt to enhance motor retention.

Firstly, packaging conditions were varied to establish the shortest time which allowed for stalling. This should limit the time for dissociation between motor and prohead. Assays were stopped by the addition of ATPγs, followed by DNase, at 0s, 10s, 1 min and 3 min and 10 min (Figure 3.7 d). After 1 minute there is a significant band of protected DNA and by 3 minutes the band shows equal intensity as for 10 minutes. Subsequent assays were incubated with ATP for 2 minutes which appeared to increase the proportion of motor bound proheads significantly (Figure 3.7 b).

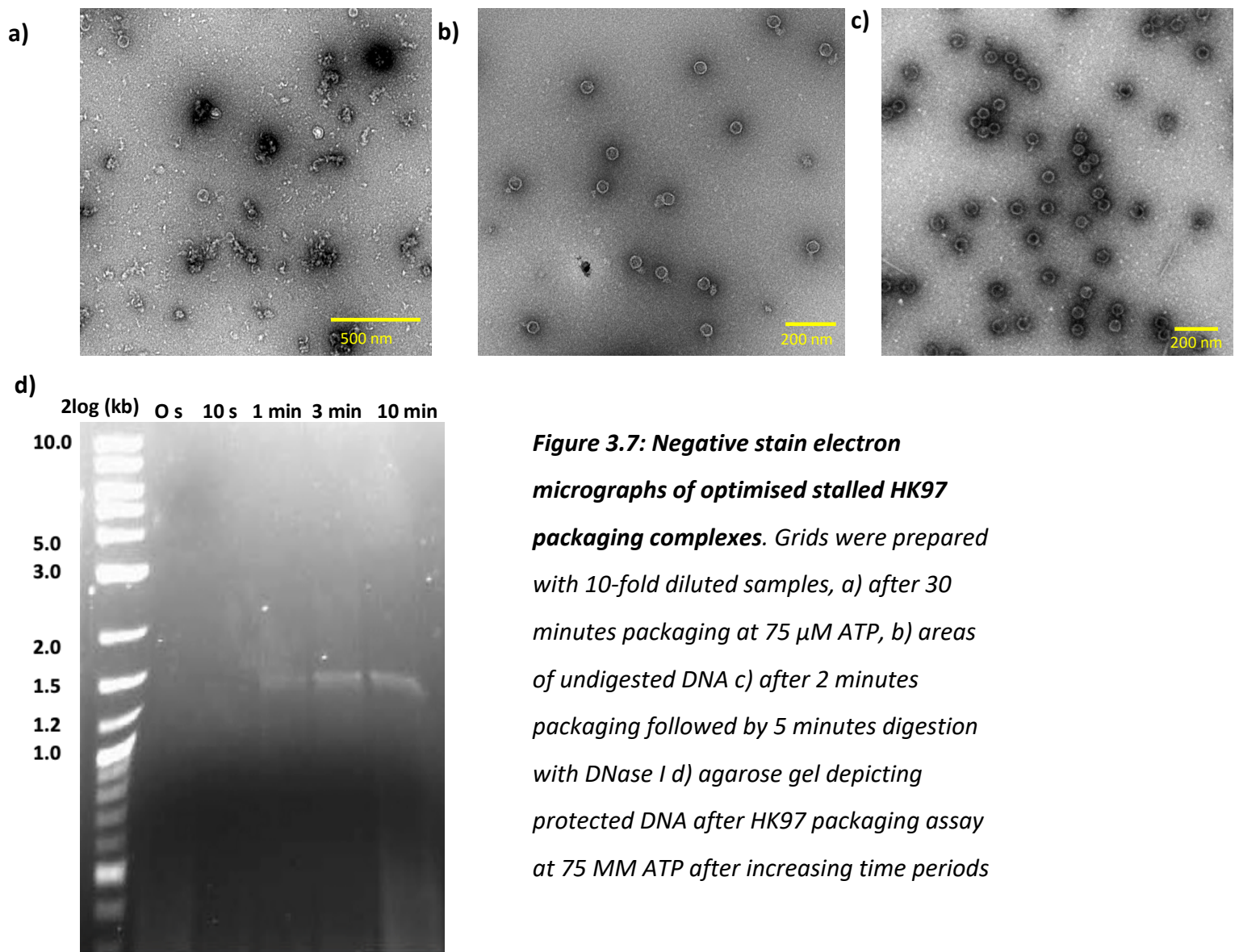


Figure 3.7: Negative stain electron micrographs of optimised stalled HK97 packaging complexes. Grids were prepared with 10-fold diluted samples, a) after 30 minutes packaging at 75 μM ATP, b) areas of undigested DNA c) after 2 minutes packaging followed by 5 minutes digestion with DNase I d) agarose gel depicting protected DNA after HK97 packaging assay at 75 μM ATP after increasing time periods

To reduce the potential background from excess protein concentrations, a packaging assay was also performed to identify minimum terminase concentrations required for stalling motors. This was established as 5 μM small terminase and 2.5 μM large terminase per 30 nM proheads. The addition of DNase was also trialled, both to digest excess DNA from the

sample, and potentially break up groups of motors which show a tendency to cluster together around the DNA substrate. Whilst it seems possible that this may destabilise motors, this is not apparent in negative stain micrographs (Figure 3.7c). In addition, DNase was used to elute package complexes in the previous data set, which showed almost complete motor retention.

Discussion

In this chapter, multiple techniques were employed to improve the yield and concentration of stalled HK97 packaging motors. Analysis by cryo-EM suggested that despite close to 100% retention of the packaging machinery, improving the resolution of the motor assembly structure was broadly unsuccessful. The low number of particles was clearly a limiting factor for the high-resolution reconstruction, in addition to intrinsic heterogeneity and flexibility. Thus, efforts were diverted to increasing the concentration of motor assemblies for further data collection.

Although ultimately the magnetic bead assay did not result in a sufficient improvement in sample concentration for structural studies, it represented an important step in protocol development and understanding packaging and stalling. Analysis of this initial data also affirmed its effectiveness as a technique. As such, it may hold potential for the isolation of other DNA processing enzymes, or indeed other protein-nucleic acid complexes. In theory, any protein which is known to bind a specific nucleic acid sequence with a weak affinity could be separated from excess unbound protein. For unstable or time-sensitive complexes this holds a clear advantage to alternative methods such as ultracentrifugation or analytical size exclusion chromatography which may take longer and expose the sample to destabilising forces. In addition, the magnetic bead assay requires only very small volumes. The technique could also be adapted to isolate any DNA binding proteins from a protein mixture.

With respect to HK97 packaging complexes specifically, investigation of the sample preparation and its stability at higher ionic strength raises several important questions

about the HK97 DNA packaging mechanism. A higher K₂Glu concentration appeared to accelerate packaging, without inhibiting stalling, and produced high yields of protected DNA from the bead assay. However, motor assemblies were only scarcely visible on negative stain electron micrographs, indicating that despite stalling, the large terminase-portal interaction was not preserved. The presence of DNA within proheads indicated that the portal likely acts as a one-way valve, forbidding the exit of DNA in the absence of the motor, as happens at the termination of packaging (248). This notion is corroborated by packaging assays at high concentrations of ATP, where a thick band of protected DNA is visible on an agarose gel, but only a few proheads with attached motors are visible by negative stain EM analysis. Combined with a symmetry mismatch that likely exists between the portal and large terminase, this may suggest there is an electrostatic element to binding, which could be disrupted by salt. Within an *E. coli* bacterium, the ionic strength ranges from 30-300 mM K⁺ with roughly 10 mM NaCl (249). Therefore, within this range the motor can be expected to be able to both attach and detach freely during phage replication.

Limiting the packaging time from 30 minutes to 2 minutes significantly improved motor retention. This corresponds reasonably well to the established packaging speed of 600bp/s, which estimate just 6 seconds to package from the 5' end of the DNA to the *cos* site. Intriguingly, although the band of protected DNA visualised by agarose gel is far weaker than for the full length of packaged DNA, the negative stain micrographs suggests that nearly all proheads retain a packaging motor. The discrepancy in oligonucleotide length can explain some of variation between DNA signal for stalled and fully packaged complexes, with DNA within stalled complexes at just 1.3 kb, compared to full length DNA at 3.4 kb. Multiple DNA molecules may also be packaged by a single motor at high concentrations of ATP. However, proheads are always present in a close to 1:1 ratio with DNA, limiting the opportunity for sequential packaging events.

Conclusion

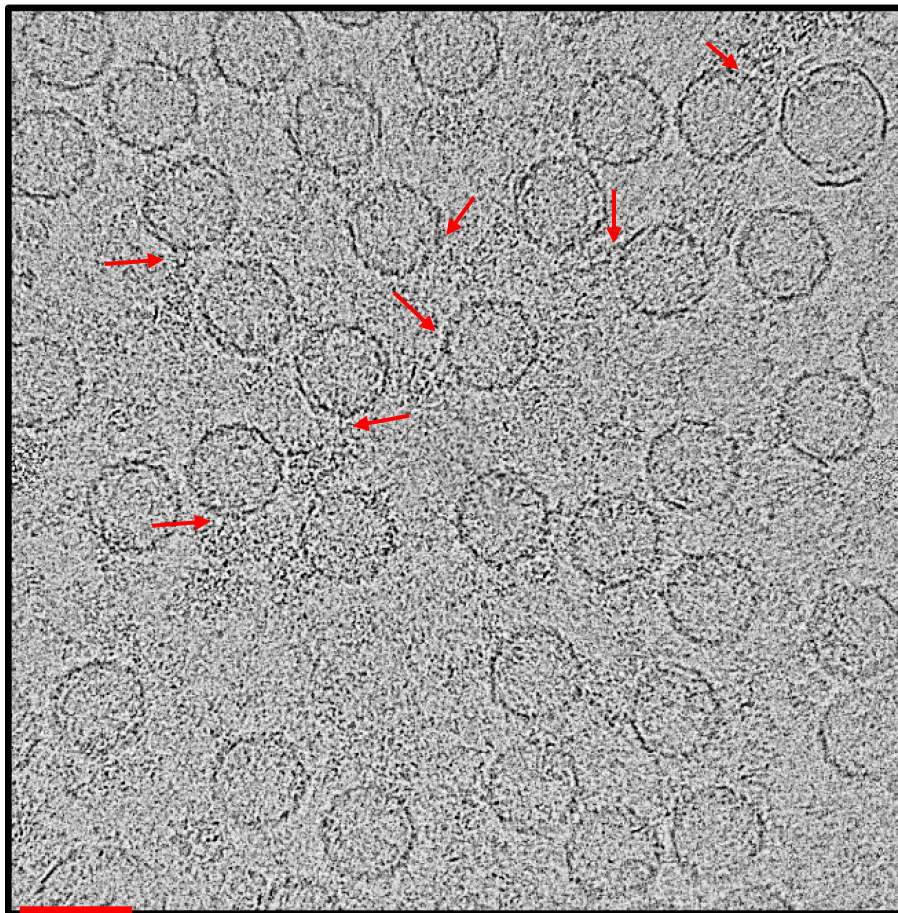
Efforts to concentrate the sample for further structural analysis, initially focused on the elution from the magnetic bead assay. Evaporation, PEG precipitation, AmSO_4 precipitation, and Vivaspin[®] concentrators were each trialled. Whilst these techniques were variously effective, they proved too harsh to keep the complexes intact. Increasing ionic strength of buffer solutions greatly improved the yield of particles recovered from the magnetic bead purification step and did not hinder stalling of the motor or protection of DNA. However, motors did not remain associated after stalling. Streptavidin-coated affinity grids were also trialled in attempt to anchor complexes to the grid by a biotin label. This has potential to harness 100 % of stalled complexes and still allow unbound contaminants to be washed away. Unfortunately, the resultant sample was aggregated and unusable. Ultimately the most effective method for increasing particle density proved to be removing the biotin selection step. The packaging time was reduced to 2 minutes, and DNase was added to reduce background signal from unpacked DNA. This produced a high concentration of stalled packaging complexes at a high concentration, suitable for cryo-EM.

Using this adapted protocol, the concentration of is dramatically increased. The inclusion of a minority of empty proheads, is thus deemed tolerable, as it should be possible to remove these particles as a feature of data processing.

Chapter 4: Structure of the prohead and portal protein

Introduction

After the extensive optimisation presented in Chapter 3, the sample preparation was deemed ready for cryo-EM grid preparation and screening, performed at York using the 200kV Glacios instrument. Imaging revealed proheads which were highly concentrated and predominantly unexpanded. More promisingly still, DNA was clearly visible entering the unique portal vertex of many particles, as shown in Figure 4.1. This sample represents a high concentration of HK97 packaging complexes, largely incorporating an active motor, stalled at the *cos* site, and stabilised by ATPγs.



50 nm

Figure 4.1: Cryo-EM screening micrograph for stalled HK97 packaging complexes. Sample prepared using R 3.5/1 grids with 2 nm Ultrathin Carbon film. Red arrows highlight obvious DNA packaging centres.

The grids were subject to three separate 72 hour data collections at York, Leeds, eBIC, in order to obtain the maximum number of particles. Such an extensive collection was deemed necessary due to the inherent properties of the sample: a large structure with an adjoined smaller, asymmetric and flexible feature. Particles from each data set were ultimately picked using TOPAZ software (214), and pixel sizes were scaled before combining for downstream processing using RELION (250), as described in methods. A total particle number of approximately 330,000 was a promising starting point for image processing.

High-resolution reconstructions of the prohead and portal are described in the following chapter, where each holds realistic potential for model building. The structure of the phage HK97 portal protein has not been determined previously, but a comparison with an available structure from a related *Corynebacterium diphtheriae* phage with sequence identity of 65%, as well the portal within an empty prohead, presented in chapter 2, showed strong correlation. Finally, an asymmetric reconstruction of the unique portal-containing capsid vertex was generated, allowing an examination of the portal-prohead interactions.

Methods

Data collection parameters

The grids used were R 3.5/1 Quantifoil, with 2 nm Ultrathin Carbon, 200 mesh copper. These grids were glow discharged for 60 seconds at 15 mA in a PELCO easiGlow™. They share the advantage of a thin carbon coating with previously used Lacey carbon grids. In addition, regular holes allow for automated data collection and extinguishes the risk of swathes of unusable areas produced by the lacey carbon pattern. Such large holes (300 µm), also allows for up to 8 shots per hole maximising the number of micrographs collected before moving the stage. Two samples were prepped immediately prior to vitrifying using the Vitribot IV, with blot force -5 and blot time 2 seconds. A final grid was prepped with three consecutive applications of sample as before.

On screening, all three grids exhibited a majority of unexpanded proheads with clear density for DNA entering the unique portal vertex. Most importantly, the concentration for each was good. In fact, grid 3 (where multiple sample applications had been applied) actually displayed a number of overlapping particles. All grids were therefore saved for data collection with the aim to collect as much data as possible. Grid 1 was subject to a 72-hour data collection at York on the Glacios instrument, then a further 72 hours at eBIC on a Krios. Data was collected for grid 3 at the Astbury centre at Leeds also for 72 hours using a Krios instrument. The data collection parameters are summarised in Table 4.1.

Table 4.1: Data collection parameters for three cryo-EM data collections of stalled HK97 packaging complexes

Location	York (Dodson building)	Leeds (Astbury Centre)	eBIC
Grid type	R 3.5/1 with Ultrathin Carbon	R 3.5/1 with Ultrathin Carbon	R 3.5/1 with Ultrathin Carbon
Hole size (μM)	3.5 μM	3.5 μM	3.5 μM
Hole spacing (μM)	4.5	4.5	4.5
Session	72 hr	72hr	72hr
Microscope	Glacios	Krios 2	Krios 2
Detector (Mode)	Falcon 4 (counting)	Falcon 4 (counting)	K3 (counting super-resolution)
Accelerating Voltage (kV)	200	300	300
Calibrated pixel size (\AA)	1.20	1.40	1.34
Nominal Mag.	49,000x	59,000x	64,000x
Total dose ($\text{e}\text{\AA}^{-2}$)	50	40.4	40
Number of fractions	66	53	40
C2 Aperture (μm)	70	70	70
Objective Aperture (μm)	100	100	100
Nominal defocus range (μm)	-1 to -3	-0.5 to -2.5	-0.5 to -2.5

The following section describes the processing protocol for these three data sets in RELION (219). For the data set described in chapters 1 and 2, `relion_autopick` was used in place of Topaz.

Data processing pipeline

During data collection motion is introduced within the sample by the electron beam. Micrographs consisting of multiple frames therefore must first be adjusted to account for this movement. Motion correction was performed using MotionCor2 (251) within the RELION pipeline. This algorithm models motion as local deformation, which varies smoothly throughout exposure. Micrographs are split into patches to determine motion, then calculated shifts remapped using this function.

The motion corrected micrographs were then subject to CTF estimation using CTFFIND 4. This program models the microscope's Contrast Transfer Function (CTF) in the power spectrum of the image. The astigmatism of each micrograph is also estimated (252). Recording micrographs over a range of defocus values allows for the CTF zeros to be compensated for, such that all spatial frequencies are represented in the dataset.

Approximately 1000 particles for each data set were picked manually. For early data sets these particles were extracted and used for 2D classification into 5 classes. The best classes were then selected as templates for `relion_autopick`. However, despite optimising picking threshold parameters, a substantial number of particles were left unpicked, while particles corresponding to ice or edges of holes were selected.

Optimised picking parameters were achieved using Topaz, as an external RELION job (214). This approach involved denoising micrographs using the Topaz-Denoise deep learning algorithm, which preserves structural features while reducing background signal. This was not necessary for the York data set as the higher defocus range used produced higher contrast images with a stronger signal-to-noise ratio. Topaz training and autopicking was

then performed on denoised micrographs. The topaz algorithm uses convolutional neural networks to train between “positive” and unlabelled templates. The parameters chosen were designed to “over-pick” rather than miss potential particles, and duplicates were then removed in RELION by selecting a minimum 300 Å distance threshold between particle centres. Coordinates were then applied to original motion corrected micrographs, extracted, and classified into 100 2D classes. The best classes were selected for icosahedral 3D refinement.

For each data set the particles were extracted to yield a pixel size of approximately 1.4 Å and a box size of 600 Å. A *de novo* model was then created, whereby a Stochastic Gradient Descent (SGD) algorithm is employed to create a low-resolution model from 2D classes, without an initial 3D reference. This was used the starting model used for icosahedral refinement. The `relion_refine` protocol automatically estimates when the refinement has converged, using gold-standard method to process images and to calculate the Fourier Shell Correlation (FSC) for resolution estimation. The FSC measures correlation between the two half-maps as a function of spatial frequency. The resolution is estimated from the FSC curve using a fixed-threshold of 0.143. The angular accuracy is also estimated throughout iterations.

To assess if the calibrated pixel sizes of each 3 datasets were properly scaled, a script `determine_relative_pixel_size`, was used (250). Leeds and eBIC icosahedrally-averaged capsid reconstructions were scaled against that obtained from the York dataset, which was chosen as the reference, which gave pixel sizes of 1.340 Å and 1.373 Å, respectively. Micrograph star files obtaining CTF parameters, and the icosahedral refinement `run_data.star` files, were subsequently altered to represent the new relative pixel sizes. Maps with scaled pixel sizes were then checked visually in chimera, and a volume subtraction between prohead pairs at threshold 0.02 was used to generate difference maps. In each case remaining signal was negligible, demonstrating successful scaling. The optics group for Leeds data was labelled group 2 and eBIC data labelled group 3. This order was maintained throughout downstream processing for continuity. The three `run_data.star` files were joined for using JoinStar in RELION and refined again with icosahedral symmetry and local angular searches, using the York refinement as a starting model and a template for a

tight mask. Data sets were also subject to a second round of icosahedral refinement separately, in case scaling was not entirely successful. The processing strategy to the point of data set merging is summarised in Figure 4.2.

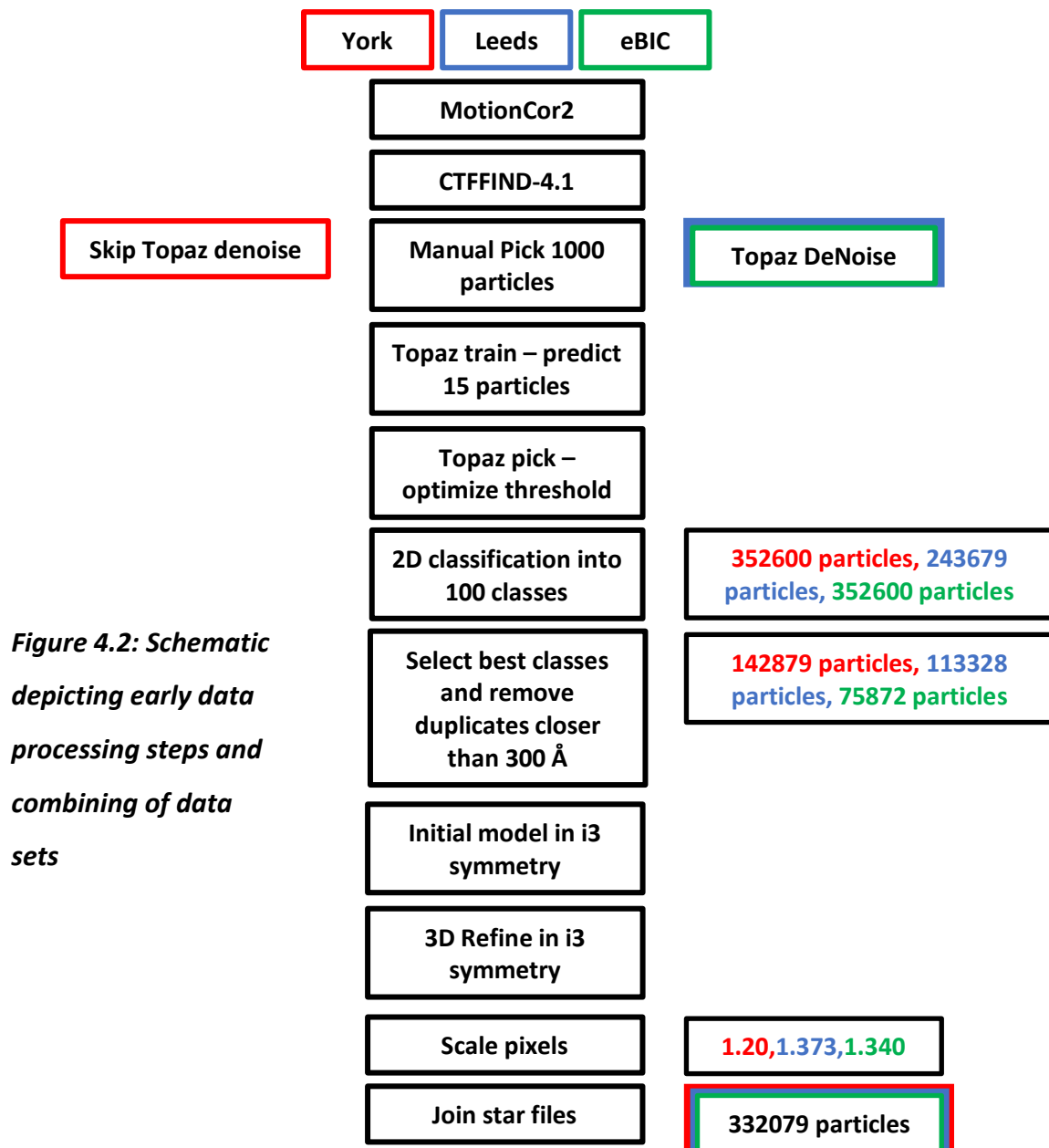


Figure 4.2: Schematic depicting early data processing steps and combining of data sets

From the processing of the three datasets, the highest resolution prohead map was obtained using the eBIC-only data set (3.1 Å compared to 3.7 Å for all three combined). This refinement was subject to postprocessing and subsequently ctf refinement. Postprocessing allows for sharpening of the map, whilst masking signal outside of the mask to provide a

resolution estimate (253). For ctf refinement, any symmetric or asymmetric aberrations or anisotropic magnification within the data set are identified in order to model the CTF parameters more accurately, whilst refining the defocus value for each individual particle. Two successive rounds of ctf refinement were performed, the first estimating anisotropic magnification, and the second instead fitting per particle defocus and objective astigmatism. The output file was then used in 3D refinement once more and post processed.

During icosahedral refinement signal from asymmetric features of the virus are averaged out evenly over the map. Thus, in order to model the packaging machinery, the icosahedral 3D reconstruction of the combined data sets was first subject to RELION symmetry expansion (in symmetry point group I3), which generates a 60-fold increased set of particles.

Particles were then reextracted into smaller boxes, utilising a shift along the Z axis to centre particles on the capsid vertices, producing 60 subparticles per capsid particle. Subparticles were subject to 3D classification into 10 classes to separate subparticles containing portal and motor signal from those with capsid only. Initially a cylindrical mask, created using `relion_helix_toolbox`, was used to mask out contributing signal from the capsid, centred on the expected position for the portal and motor. One clear portal-containing class emerged, which was used as a template for a tighter mask, which was applied to the original subparticle set for further classification. Figure 4.3b) shows the resultant portal-motor class and Figure 4.3 a) shows the convergence behaviour during 3D classification, showing little change in class distribution metric for the portal protein after only ~12 iterations. Flattening of the data line indicates convergence, with the final number of particles only just lower than the predicted fraction of ~0.083 (5/60) indicated by the red line.

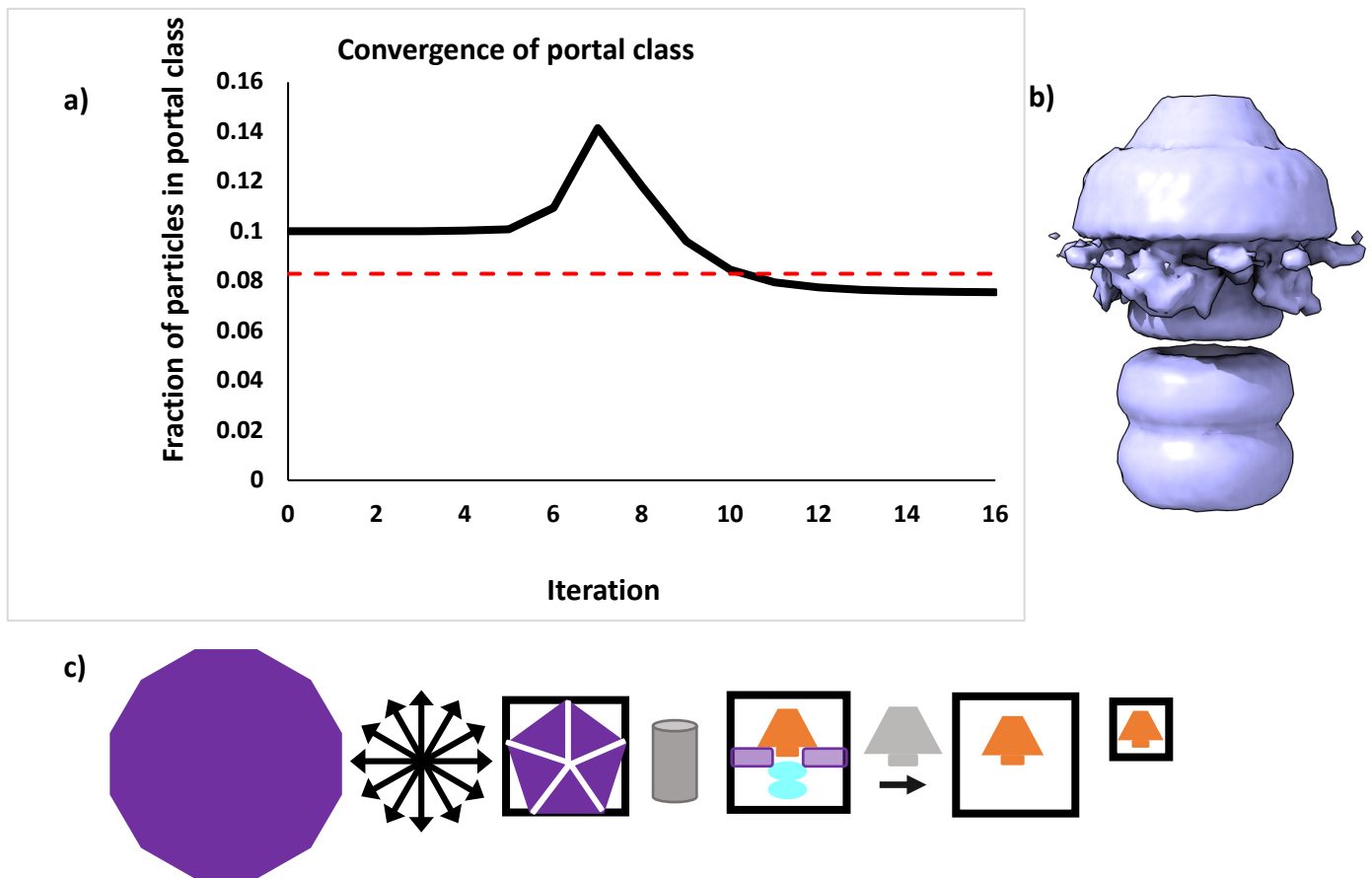


Figure 4.3: symmetry expansion and focused classification for identification of the portal/motor class a) Graph showing convergence of the portal/motor class following symmetry expansion and focussed classification b) portal/motor 3D class c) schematic depicting processing scheme for C12 refinement of the portal protein

To reconstruct the portal protein structure, particles from the portal-motor class were selected and subjected to 3D classification into 5 classes using a portal mask and imposing C12 symmetrical averaging. This produced 4 good classes out of 5, where each showed clear symmetry and secondary structure. The class displaying the most defined secondary structure was selected, and any remaining particle duplicates from symmetry expansion removed and re-extracted into a smaller box encompassing only the portal. These particles were refined using C12 symmetry (Figure 4.5). The 3D map was then subject to postprocessing and local resolution estimation to calculate resolution within different local sections of the map. This showed marked improvement in the clip region which exhibits lower resolution than the rest of the portal protein.

Results

Icosahedral reconstruction of the prohead

The prohead map shows clear icosahedral symmetry and well resolved density for the major capsid protein protomers that comprise the capsid shell (Figure 4.4). It measures a diameter of 550 Å and shows dislocated trimers, within the hexamers, in common with previous Structures of prohead II (254). Fitting the current model of the asymmetric unit (one complete hexamer with a single subunit of the adjacent pentamer) from the PDB model 3E8K (254), demonstrates a very good fit. (Figure 4.4 d). The resolution of this map at 3.06 Å shows a marked improvement of the previously available structure determined at 3.65 Å, allowing for a significantly more accurate model to be built.

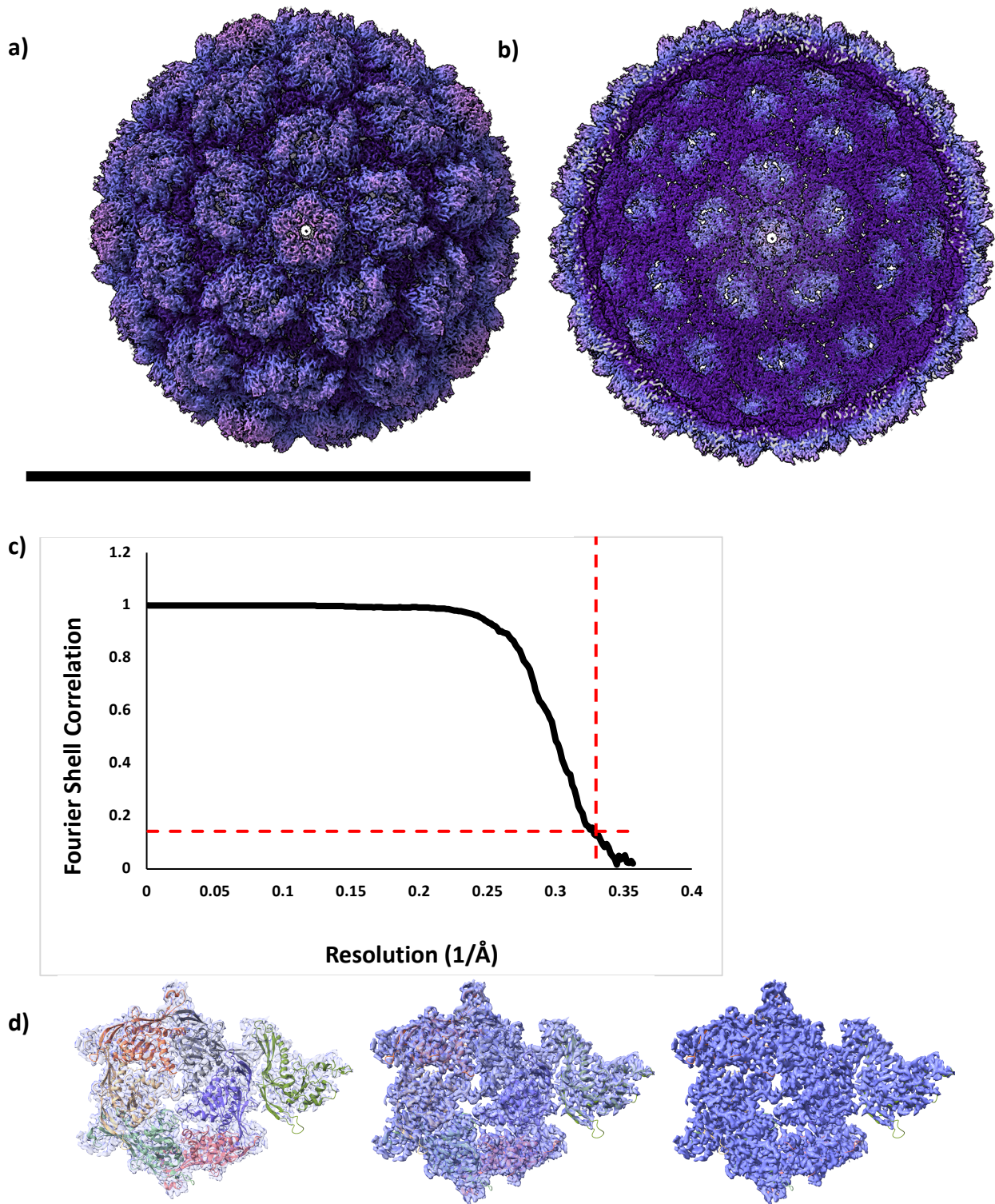
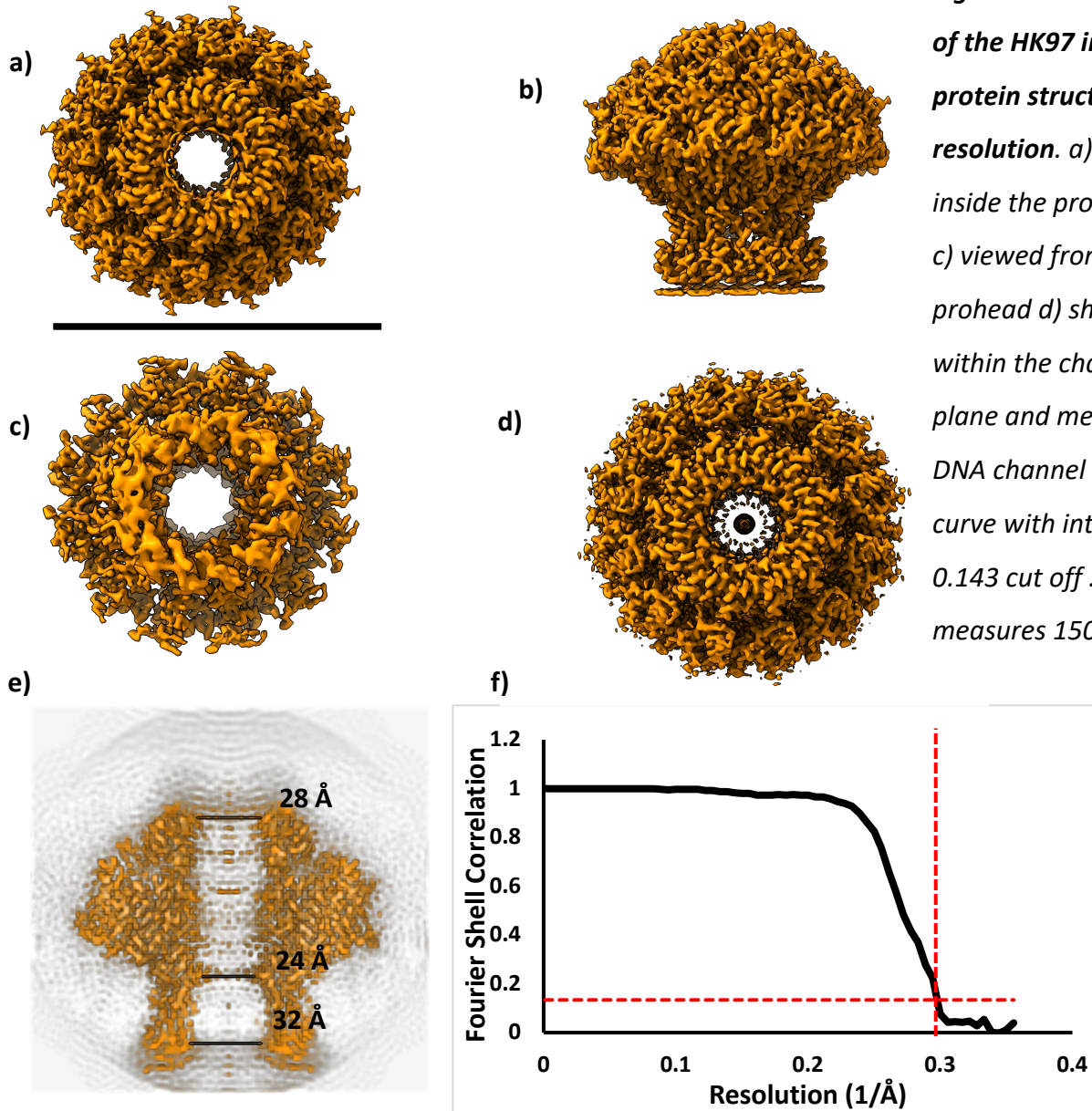


Figure 4.4: Icosahedral reconstruction of the HK97 prohead II to 3.06 Å resolution a) viewed along the five-fold axis b) cross sectional view c) FSC curve with interception at the 0.143 cut off d) Crystal structure 3E8K fit into the asymmetric unit of the prohead II map with decreasing transparency. Scale bar measures 600 Å.

Reconstruction of the *in situ* portal protein structure from prohead II

Figure 4.5: C12 reconstruction of the HK97 *in situ* portal protein structure at 3.30 Å resolution. a) viewed from inside the prohead b) side view c) viewed from outside the prohead d) showing DNA within the channel e) X axis plane and measurement of the DNA channel diameter f) FSC curve with interception at the 0.143 cut off. Scale bar in (a) measures 150 Å.



The portal protein reconstruction shows clear 12-fold symmetry with an overall “mushroom” shaped architecture. Each of the clip, stem, and crown domains are reasonably elongated creating extended DNA channel, with the wing domain spanning outwards. This map provides novel structural information on the phage HK97 portal and the high-resolution

will greatly aid future model building. At high thresholds, DNA is visible inside the portal channel (Figure 4.5 d), although it is less well defined than DNA observed later in the motor channel. The width of the central channel through the portal density is variable, measuring approximately 28 Å at the crown domain, 32 Å in the clip domain and 24 Å in the tightest part of the stem. This is sufficient to allow hydrated B-form DNA (which has a diameter of ~23 Å) into the prohead (255).

Figure 4.6 displays the local resolution map (i.e. locally low-pass filtered) of the portal protein. Clearly the wing domain is particularly well resolved with lower resolution observed for the crown, and especially the clip domain. Flexibility within each of these domains has been indicated in studies of other viruses, in particular during DNA packaging. For the P23-45 bacteriophage, significant differences within the portal crown are seen between structure determined *in situ* by cryo-EM and by crystallography (140). Morphing of the clip domain to adopt a pseudo-C5 symmetry has been observed for a prohead-derived portal protein in crystal (128). There is a growing understanding of a plasticity in the portal protein, exemplified in a recent “corkscrew” conformation of phage PaP3 (256). This conformational plasticity may also facilitate signal transduction, allowing internal pressure within the prohead to stimulate the large terminase nuclease domain (128,135,136). This is essential for headful packaging termination.

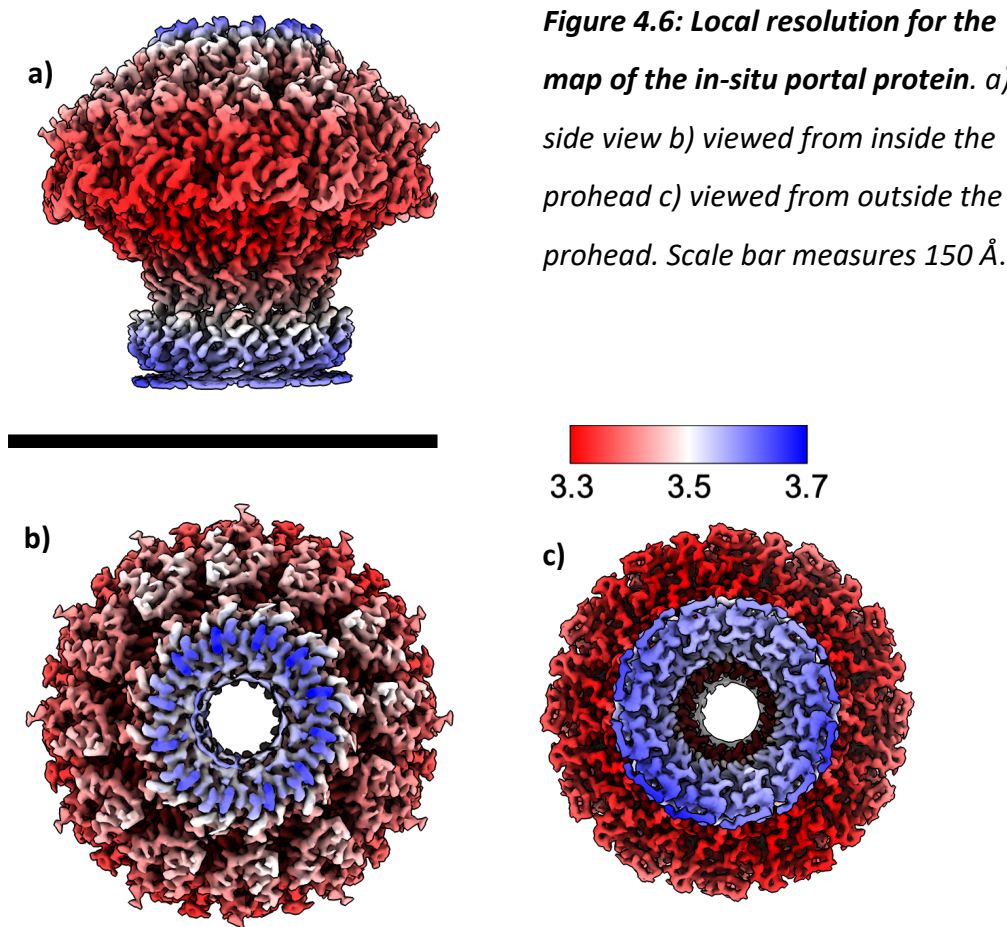


Figure 4.6: Local resolution for the map of the in-situ portal protein. a) side view b) viewed from inside the prohead c) viewed from outside the prohead. Scale bar measures 150 Å.

To improve the quality of the map at the clip domain, attempts were made to mask out all portal signal bar the clip domain, and 3D classification and refinements were explored in C12, C5 and C1 symmetry. Whilst the best resolution was achieved in C12 symmetry, none of these reached a comparable resolution or angular accuracy to the full portal model. Despite masking it may be that the signal from the remaining portal is driving this symmetry. Another possibility is that the conformation of the clip domain changes throughout packaging, as a feature of changing interaction with large terminase. For instance, the portal may interact with between one and all five subunits at different stages within the translocation cycle. Lastly, the portal structure likely represents a mixture of portal with, and without association with large terminase. These two structures are likely impossible to classify out, due to broad structural similarity, but may well exhibit significant differences within the clip domain.

Asymmetric reconstruction of the stalled packaging complex

In order to examine the overall architecture of the packaging complexes the portal-motor class following icosahedral expansion and re-extraction was reextracted with a reverse shift along the Z axis to include the whole prohead. A subset of these particles were reconstructed without alignment and the resultant map was used as a template for 3D refinement. Refinement was performed in C1 and is depicted in Figure 4.7. Here we can see the composition of the particles clearly: an icosahedral prohead, with a portal protein at one vertex protruding deep into the capsid. The strong signal seen for the motor will be discussed in Chapter 5.

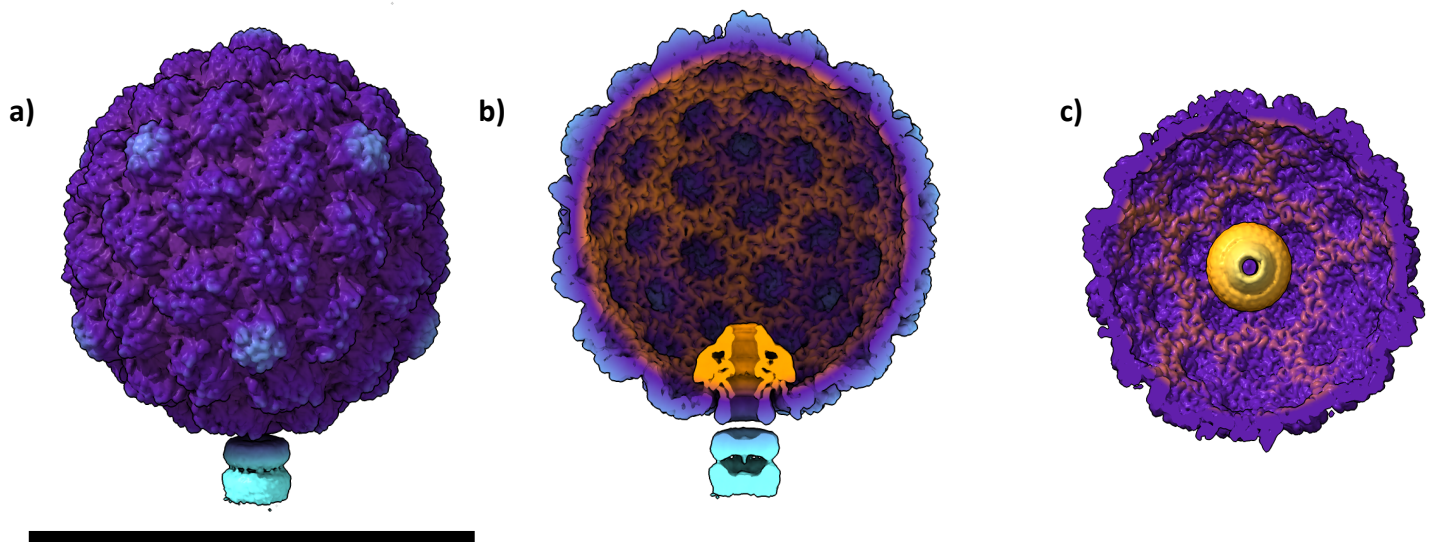


Figure 4.7: Asymmetric refinement of the HK97 full packaging complex a) side view b) cross section along the portal axis c) cross section perpendicular to the portal axis . Scale bar measures 600 Å.

Local reconstruction of the unique portal vertex

At the unique portal vertex, there is 5-fold redundancy in terms of possible prohead orientations with respect to the portal. Therefore, even performing an asymmetric reconstruction without symmetrical averaging, gives five possible orientations for each image to match the reference. In turn, the regions of asymmetry will not be visible in the

final reconstruction. In order to examine the portal- prohead points of contact, asymmetric refinement was facilitated by the use of a script provided by Dr Pavol Barty (257) .

Particles from the portal/motor class after icosahedral expansion were selected and duplicates removed. Reconstruction of the portal region only, in C12, was then performed using a tight mask producing a high resolution map with good angular accuracy. The same process was applied to the surrounding prohead density, which was instead refined with C5 symmetry. The run_data.star files from each job were then fed into the script which compares angles from the C12 and C5 refinements, plus angles in multiples 30° and 72 ° respectively. This effectively produces all the possible input angles. Where the angle from the 2 data sets match, this must represent the angle where the portal is aligned with respect to the prohead. The resultant star file was then 3D classified into a single class, with angular refinement, using very limited local searches of 1.8 ° and 1.8° degrees. This was used as a starting model and mask template for asymmetric 3D refinement, post processing and local resolution calculation shown in Figure 4.8 a).

The resultant map exhibits a resolution of approximately 4.2 Å, with lower resolution at the crown and clip domain of the portal as seen before (Figure 4.8a). Only a single data set was utilised for the script, (eBIC), so it's possible that the resolution could be further improved with adding particles from other datasets. It is possible to see unique contacts between each capsomere and the portal at different depths into the prohead (Figure 4.8 b,c). Importantly, there is also no suggestion that the resolution is lower at the portal-prohead interface. Deviations from C12 symmetry, within the portal, are suggestive of an accurate C1 refinement rather than any symmetry averaging from either the portal or prohead signal.

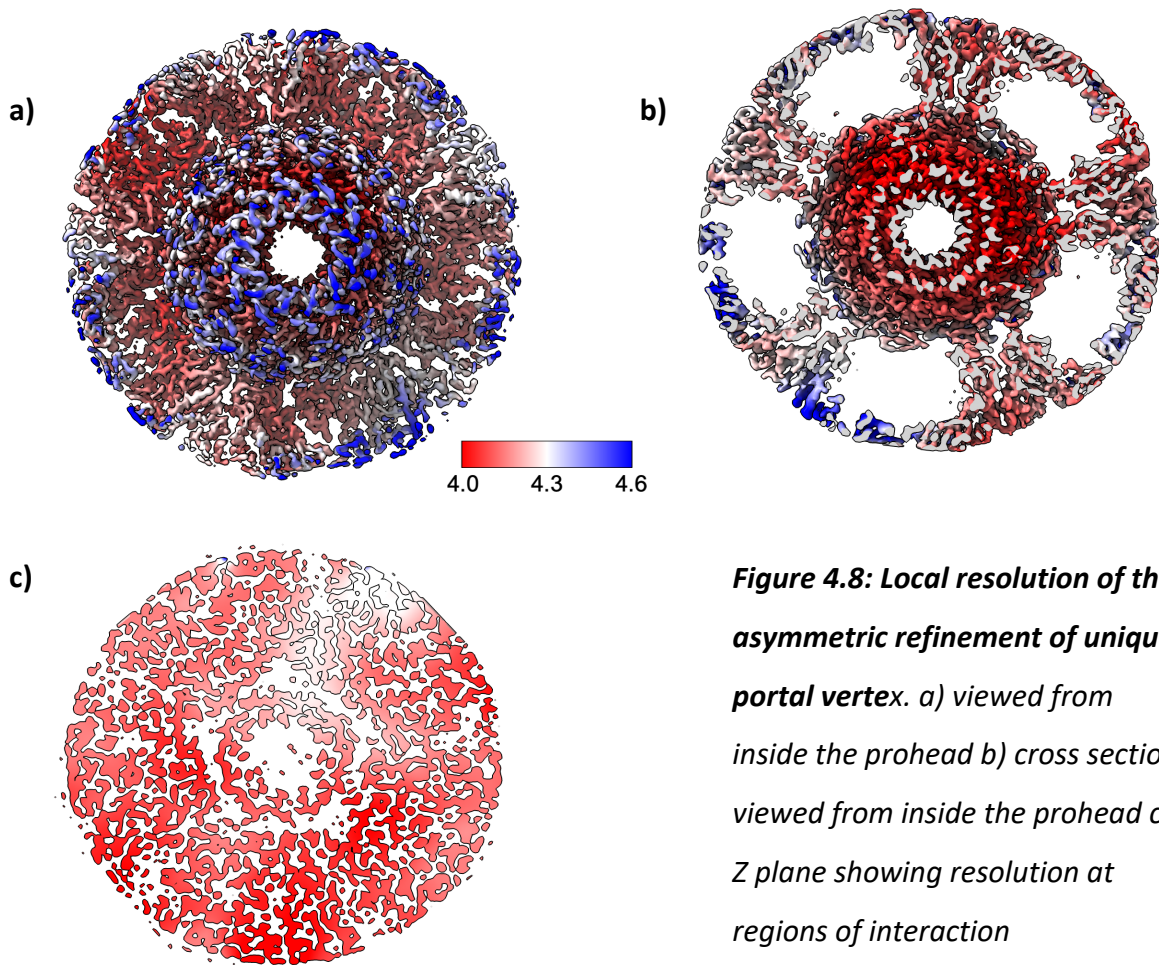


Figure 4.8: Local resolution of the asymmetric refinement of unique portal vertex. a) viewed from inside the prohead b) cross section viewed from inside the prohead c) Z plane showing resolution at regions of interaction

Discussion

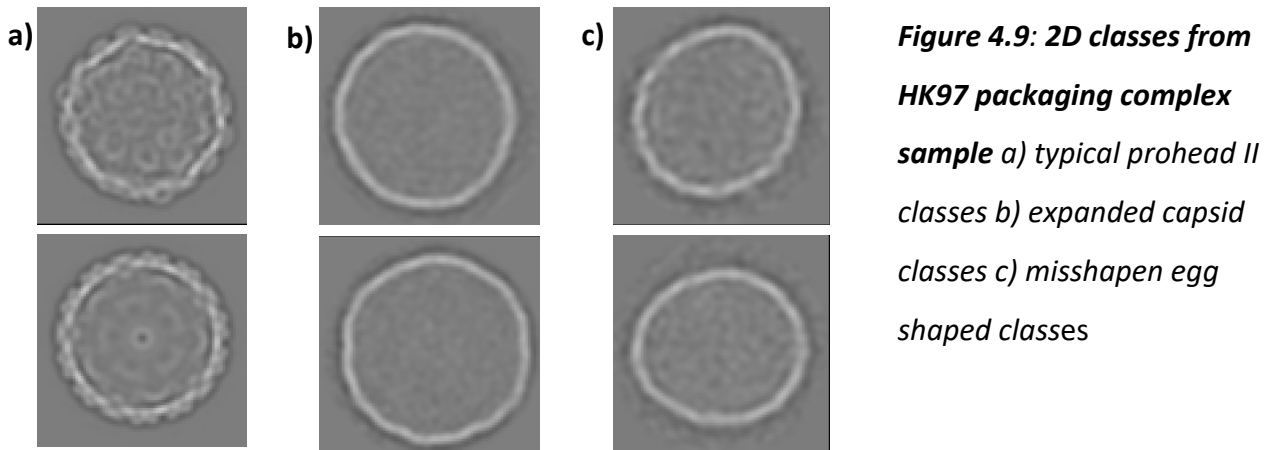
Structure of the prohead

Although every effort was made to inhibit expansion of proheads, 2D classification revealed several potential expansion intermediates. Figure 4.9 a) shows two typical prohead classes, with a compact spherical shape and a thick shell. The portal vertex is visible in one.

Meanwhile Figure 4.9 b) shows classes encompassing more angular, expanded capsids.

Finally, Figure 4.9 c) displays 2 classes with an egg like shape, where one end of the prohead

is distinctly wider than the other. A similar class was present for each data set. Whilst these particles may be deformed from purification, they could also depict an intermediate stage of expansion. However, these particles were present in sufficiently low numbers that they did not warrant further optimisation.



Structure of the *in situ* prohead portal

Isolating the unique portal vertex is challenging due to the dominating signal from the prohead, which displays strict icosahedral symmetry and has a total mass of approximately 14,000 kDa, relative to only ~ 567 kDa for the portal and ~ 276 kDa for the large terminase pentamer. Any asymmetric features are therefore averaged out during the refinement. To locate portal vertices, the icosahedral refinement was subject to I3 symmetry expansion followed by re-extraction and focussed classification. This produced one class with clear portal and motor signal.

The portal map shares the predicted symmetry, overall architecture and domain composition as structures of portal protein from other dsDNA phage. Figure 4.10 a,b) shows comparisons with $\Phi 29$, T4 and a HK97 like phage which has 65 % sequence identity with the HK97 portal protein (258). The structure of this HK97-like phage was determined by crystallography to a resolution of 2.9 Å, with the model comprising amino acids 43–310 (PDB entry 3KDR, unpublished). The HK97 map clearly shows density corresponding to an extended crown region relative to this related portal protein as well as portal proteins of T4 and $\Phi 29$. The crown region is implicated in DNA spooling (259), so this crown extension may

well influence DNA packaging and arrangement within the mature virion. A prediction of the portal monomer structure was also calculated with AlphaFold, which displays reasonable alignment with the portal model of the HK97-like phage (Figure 4.10 c). This model prediction fits the map reasonably well and would make a good starting point for model building and structure refinement.

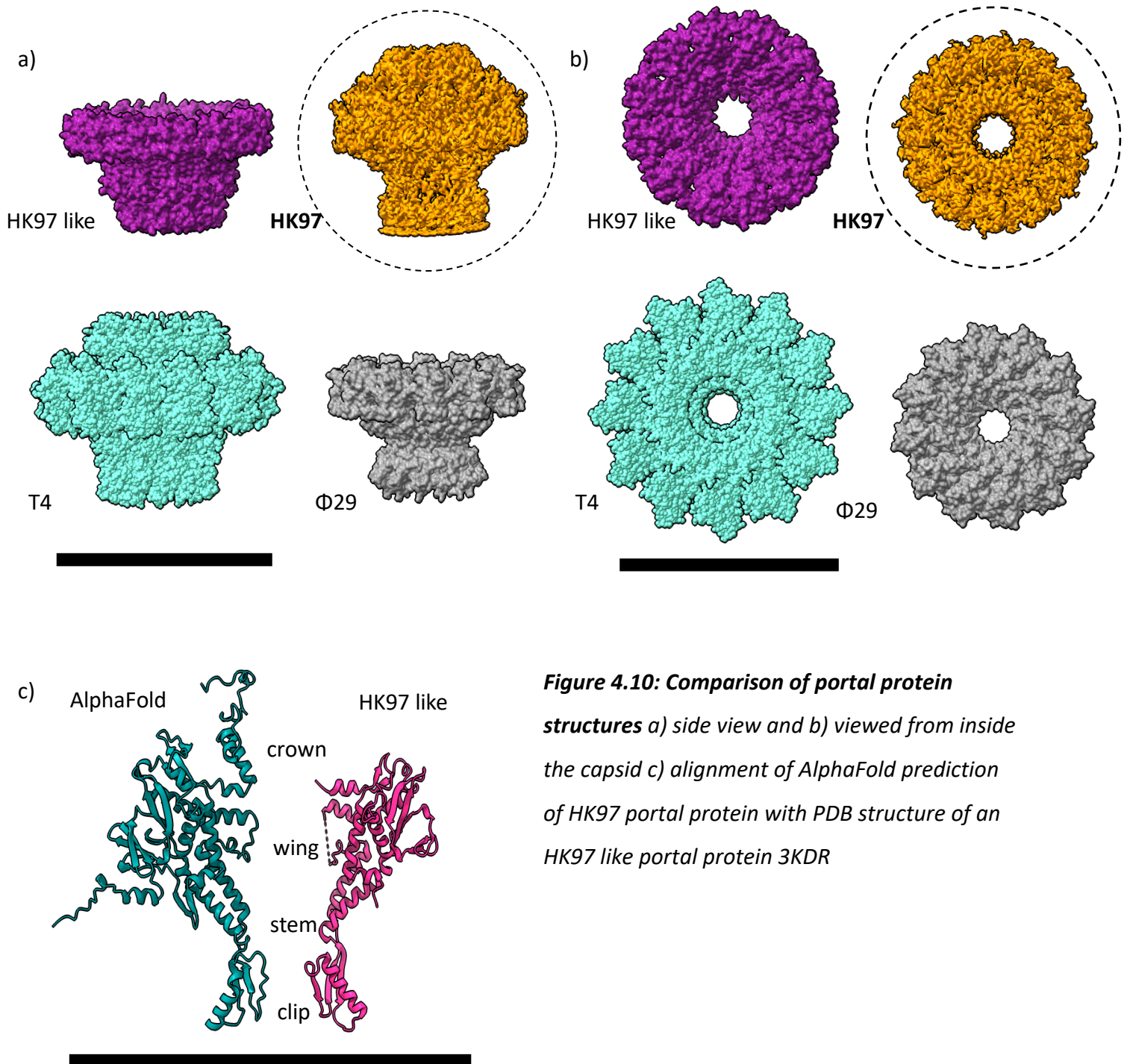


Figure 4.10: Comparison of portal protein structures a) side view and b) viewed from inside the capsid c) alignment of AlphaFold prediction of HK97 portal protein with PDB structure of an HK97 like portal protein 3KDR

Another interesting comparison of the portal reconstruction can be made with the portal map from a previous data set (Chapter 2), where the portal was present within prohead II but without an associated motor. Although the global resolution of this map is lower, the quality of the map corresponding to the clip domain is relatively high, with individual α -helices well-resolved (Figure 4.11 a-e). This suggests that the poor resolution seen in the clip domain may well be due to potential morphing away from C12 symmetry to interact with large terminase. The AlphaFold predicted model of the portal monomer fits both maps reasonably well, but more closely aligns within the clip region in this motor-free map (Figure 4.11 f,g). This supports the notion that the clip domain undergoes conformational rearrangements during packaging.

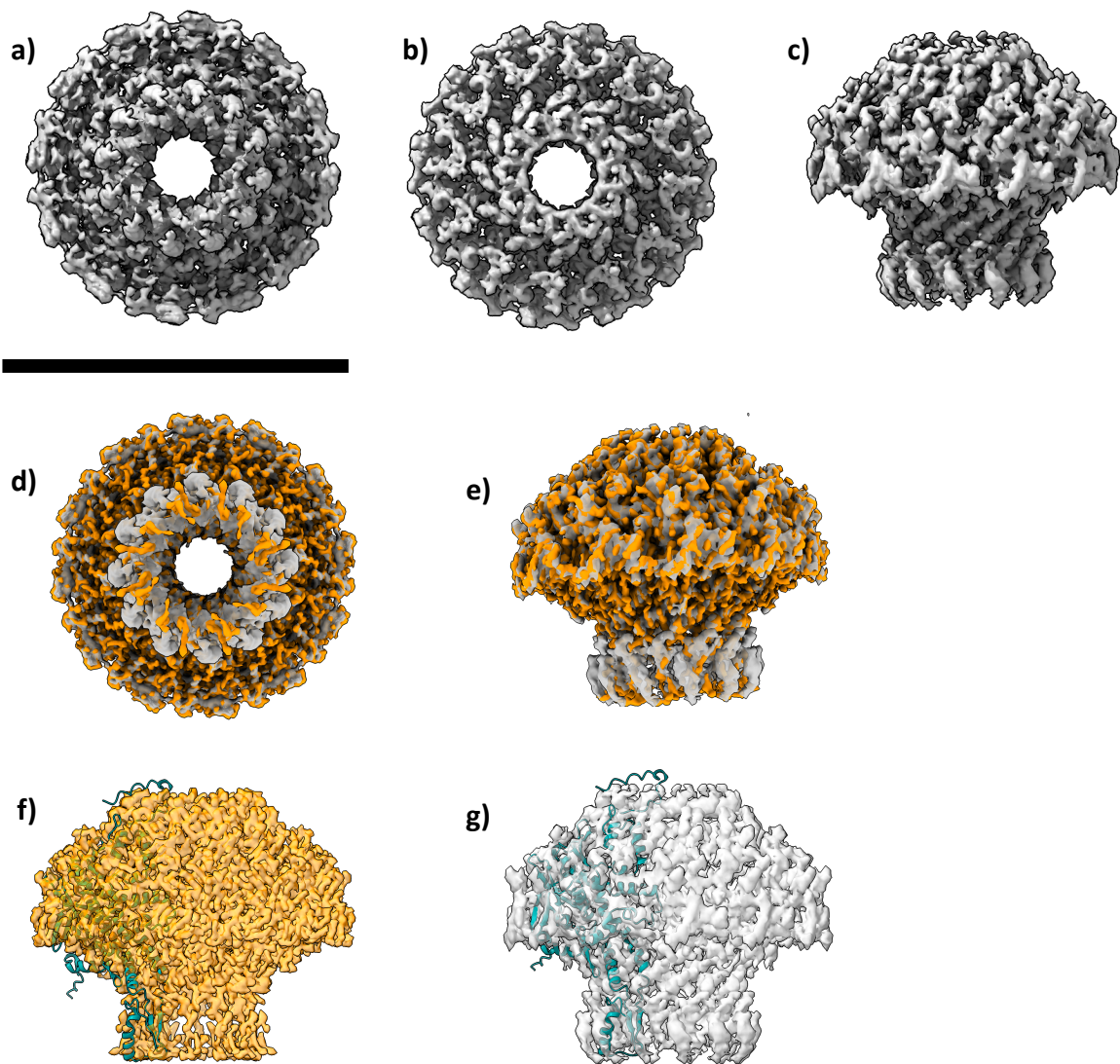


Figure 4.11: HK97 portal architecture in prohead II with and without an associated motor. C12 refinement of HK97 in situ portal in empty prohead particles to a resolution of 4.1 Å a) viewed from inside the prohead b) viewed from beneath the prohead c) side view. Superimposition of portal maps with and without attached motor d) viewed from beneath the prohead e) side view. Alpha fold prediction of portal monomer fit into f) portal map with motor and g) portal map without motor

A structure of the portal protein within the mature HK97 capsid would give valuable insight into the structural deviation of the portal protein throughout phage maturation. For instance, we may predict that the portal pore diameter would be dramatically reduced. This occurs within the mature capsids of Φ 29 and P22, allowing the channel to act as a valve

preventing DNA leakage (128,138,139) (133). Within the mature phage the portal must also bind to the connector protein, which acts as a base for the tail (260,261). The structure of the HK97 connector protein has been determined by crystallography at 2.1 Å, in an oligomeric state comprising 13 circularly arranged subunits (3JVO (262)). Assuming this structure represents the native conformation, association with the connector would thus result in an additional symmetry mismatch, also requiring conformational change within the portal clip domain.

Structure of the unique portal-containing capsid vertex

At the portal vertex of phage T4, portal plasticity allows for each capsomere to interact with a similar environment with little deviation seen within surrounding capsomeres (130). Deviation from C12 symmetry is also seen within C1 reconstruction of the *in-situ* portal protein presented here. To validate the asymmetric nature of the portal and the prohead alignment, 3D refinement was performed using the identical input particles, starting model and mask, but without the predetermined angular limits. Refinement was instead subject to global searches. The resultant map exhibits a far lower resolution (Figure 4.12 a,b) particularly for the portal, where C12 symmetry is not apparent. In the cross section shown, we can see blurring of subunits into a ring. The prohead signal is also far stronger on one side of the portal protein.

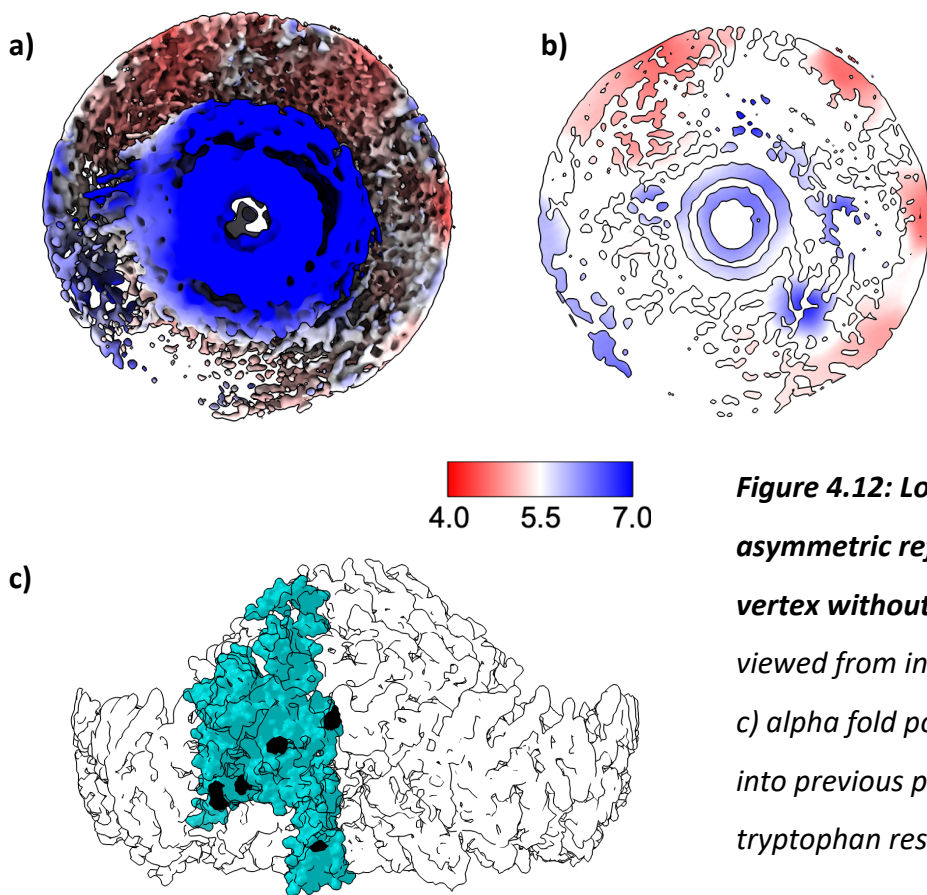


Figure 4.12: Local resolution of the asymmetric refinement of unique portal vertex without angular restraints a) viewed from inside the prohead b) z plane c) alpha fold portal monomer model fit into previous portal vertex map, tryptophan residues are marked black

For some dsDNA viruses, the portal–prohead interaction is assumed to be mediated largely by hydrophobic contacts (263,264). In conjunction with the portal clip morphing this may help to compensate for the symmetry mismatch. Indeed, many portals show a band of tryptophan residues across the wing domain which is thought to help position the portal in place. AlphaFold prediction of the HK97 portal model structure indicates that tryptophan residues indeed span the wing domain, and approximately align to contact with prohead, as shown in Figure 4.12 c).

Conclusion

In this chapter an icosahedral reconstruction of HK97 prohead II, and the *in-situ* portal protein is presented. The prohead map, derived at 3.05 Å resolution, matches the available structural information well, but exhibits a significantly higher resolution, and as such should

allow for a more accurate model to be built. This expansion intermediate is of particular interest as it represents the beginning of the packaging process. The portal protein was also reconstructed with C12 symmetry to approximately 3.3 Å resolution. This represents new structural data for phage HK97, an important model system in the study of dsDNA virus genome packaging. Relative to the portal structure presented in empty proheads in Chapter 2, the clip domain is poorly resolved, suggesting plasticity in this domain may facilitate binding to large terminase.

Finally, an accurate asymmetric reconstruction of the portal vertex was also determined, to examine the interface between the portal and surrounding capsomers in the prohead. This was achieved by identifying C1 angles using a comparison of C5 (vertex) and C12 (portal) refinements. The resultant map, derived at 4.2 Å resolution, shows unique features at different portal-capsid points of contact, indicating success of the asymmetric reconstruction. Together, these three high resolution reconstructions allow for an accurate model of the HK97 prohead and *in situ* portal protein to be built

Chapter 5: The structure of the motor

Introduction

Following successful reconstructions of the prohead and the portal protein, processing efforts were focussed on the motor. The high resolution of both maps thus far was encouraging, indicating that the data is of good quality, that separate data sets have been successfully combined, and that there is an adequate number of particles. Furthermore, the portal-motor class extracted after icosahedral expansion showed very strong signal for the motor.

However, there are also several major processing challenges intrinsic to the system. Firstly, the dominating prohead signal is a likely source of misalignment of motor particle. For the portal, which largely shows strict C12 symmetry, this problem was successfully resolved with re-extraction and masking. However, for an asymmetric and highly flexible feature such as the motor, the overriding signal may complicate processing. Furthermore, there is inherent symmetry mismatch between the portal and large terminase proteins (which may be independent of the symmetry mismatch between capsid and portal), and in turn, potential symmetry mismatch between large terminase and small terminase proteins. These three proteins form oligomers composed of 12, 5 and 9 subunits, respectively (1). In addition, the motor may well be tilted away from the central axis of the portal, as seen for $\Phi 29$ (177).

The large terminase is also expected to display flexibility, as movement between the C- and N-terminal domains is thought to drive translocation cycle (177,180) (184)(153)DNA translocation by the $\Phi 29$ motor is predicted to proceed through 5 conformational states linked with chemical events of ATP hydrolysis. This mechanism was derived on the basis of the structure of the *in situ* $\Phi 29$ motor, 'fixed' with ATP γ S and structures in ADP bound

planar states (76,177,180). The HK97 sample presented here was also flooded with ATP γ S in an attempt to confine the motor to a single state.

Finally, based on sample preparation, a substantial proportion of proheads are predicted to be empty. These will encompass a portal but no motor. Such particles may be difficult to identify and exclude, due to the overriding density from the portal and prohead. Empty particles included in refinements are likely to contribute to noise.

Methods

Expression and purification of the K92Q large terminase mutant

Primers were designed for site directed mutagenesis using the Quick Change mutagenesis system (265). The gene for large terminase (residues 1- 1–526) was provided in a pET22a-based vector with an N terminal SUMO-his tag. A successful clone was transformed into *E. coli* BL21(DE3) pLysS cells and grown in LB medium containing 30 μ g/mL kanamycin and 33 μ g/mL chloramphenicol at 37° C .At an OD₆₀₀ of 0.6 the temperature was reduced to 16 °C for 18 hours.

Cells were harvested at 4 °C, in 20 mM Tris-Cl, 1 M NaCl, 10% (v/v) glycerol, 20 mM imidazole, 0.05% (v/v) β -mercaptoethanol, pH 8.0, with 100 μ M AEBSF, 1 μ M leupeptin, 1 μ M pepstatin A and 10 μ g/mL RNaseA and DNase stimulated by 1mM MgCl₂. The resuspension was next sonicated in 3 second / 10 second blasts, at 75% power for a total time of 20 minutes. Lysate was spun at 20 kg for 40 minutes and the supernatant applied to a His-trap before eluting the protein in a gradient of 20 -500 mM imidazole. The eluate was dialysed against 20 mM Tris-Cl, 200 mM NaCl, 1 mM dithiothreitol (DTT), pH 8.0, at 4 °C overnight, with 1:100 (w/w) SUMO protease, which was produced in-house.

Large terminase was further purified by anion exchange chromatography on a Mono Q 10/100 column, eluting with a 200–1000 mM NaCl gradient in 20 mM Tris-Cl, 1 mM DTT, pH 8.0, and size exclusion chromatography on a Superdex 75 16/600 column in 20 mM Tris-Cl, 300 mM NaCl, 2 mM DTT, pH 7.5

Crystallisation screens

Screening of suitable crystallisation conditions was performed with protein concentrated to 20 mg/ml in 20 mM Tris pH 8, 300 mM NaCl, 10 % w/v glycerol, 1 mM DTT, using commercial screens Morpheus and Index. 150 µl drops were mixed with the reservoir solution in a 1:1 volume ratio. Plates were prepared using a mosquito robot and stored at 18 °C.

Results

Processing strategies for the motor

Multiple strategies, utilising localised reconstruction approaches, were employed to isolate density for the DNA packaging motor and produce a high-resolution reconstruction. These are presented in results sections 1-6. In each case, particles were initially down-sampled substantially to pixel sizes in the range of 3-5 Å. It was hoped that this would speed up processing and facilitate the alignment of broad features. From an initial 3D classification without angular searches, particles from classes with defined features were selected. These were subject to a second round of classification, into a single class using local refinement (angular searches of 1.8°, and translational searches of 5 pixels). The resultant map was used as a template for a mask and starting model for 3D refinement. Refinement also proceeded with local angular searches (1.8° and 1.8 °) as larger angular searches produced featureless and low-resolution maps. Figures 5.1 to 5.6 depict the resultant 3D refinements. This allows for a comparison of models, and the establishment of common structural features as well as points of variation. When discussing the generated maps, the “top” ring of density, directly beneath the clip domain of the portal is predicted to be the large terminase C-terminal domain (CTD), and “bottom” ring the N-terminal domain (NTD) in accordance with data on other viruses (177,266).

1.Masking for the motor

Following identification of the portal-motor class from icosahedral symmetry expansion, the region corresponding to the motor density was segmented in chimera/segger. This was used as a template for a mask. Particles were then subject to 3D classification without angular refinement into a range of class numbers (approximately 5-100) using this new mask to reduce portal and prohead signal. This process was also repeated after removing remaining duplicates closer than 100 Å together using the RELION select tool. The classes with the most defined features were combined and reextracted into a smaller box around the motor before refinement (Figure 5.1 a).

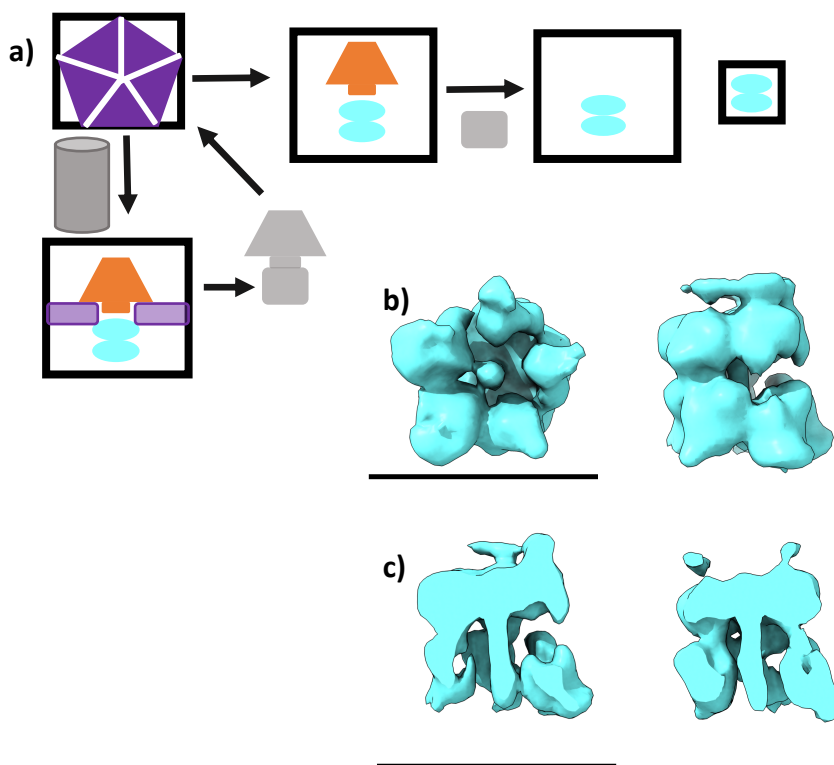


Figure 5.1: Processing scheme 1 - masking the motor from the portal/mask 3D class a) schematic b) resultant 3D refinement viewed from outside the prohead and from the side c) cross section views on X and Y planes. Scale bar measures 150 Å.

This map clearly shows five subunits, with the assumed CTD tilted away from the portal clip in a flat plane, seemingly contacting the portal via 1–2 large terminase subunits (Figure 5.1 b). The NTD is then tilted in the opposite direction creating a shallow “Z” shape within the

cross section (Figure 5.1 c). This causes the subunits in contact with clip to show separated domains, perhaps corresponding to an extended linker region. Meanwhile, the two opposite subunits, which do not appear to contact the portal, show a far more compact state, more closely resembling the HK97 large terminase crystal structure. DNA is clearly visible in the cross-section views, and it appears to form closer contact with 2 of the 5 subunits of the NTD.

2. Re-extraction of motor density

Here, the portal-motor class particles were reextracted with a shift in the z plane (the axis running along the portal's central tunnel). This was designed to eliminate signal from the portal and prohead more successfully than masking alone, allowing the motor signal to dominate alignments. A subset of particles was reconstructed using `reliion_reconstruct` to build a template for a starting model and mask. Again, particles were classified into a range of class numbers (between 10 and 100), removing duplicates before classification (Figure 5.2 a).

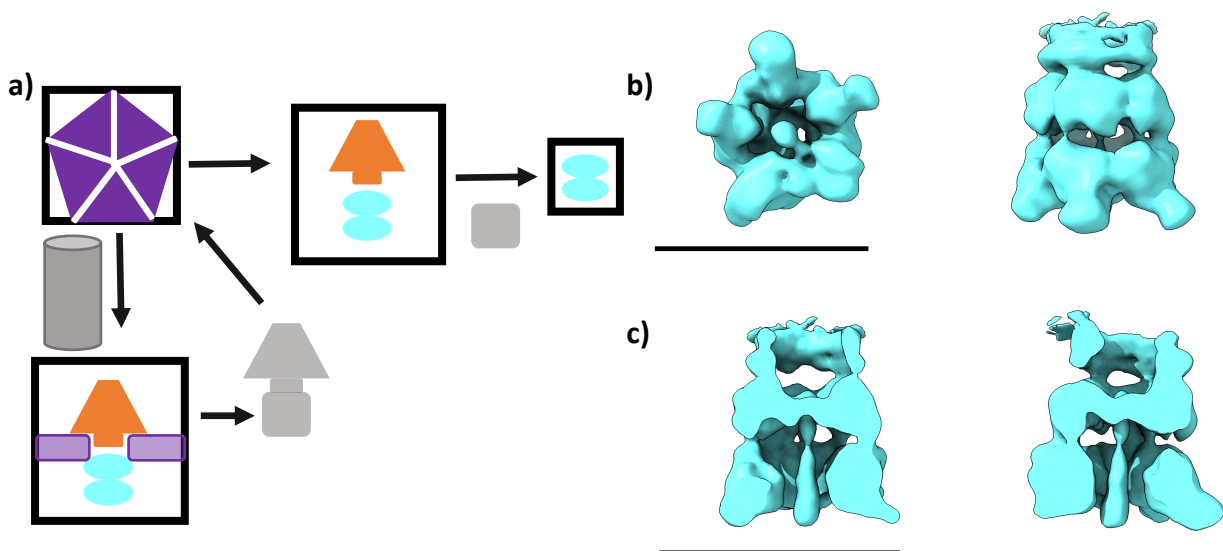


Figure 5.2: Processing scheme 2 - re-extraction of the motor from the portal/mask 3D class
 a) schematic b) resultant 3D refinement viewed from outside the prohead and from the side
 c) cross section views on X and Y planes. Scale bar measures 150 Å.

This map appears similar in architecture to that depicted using method 1, but with less pronounced asymmetry (Figure 5.2 b). The same Z shape can be seen in the cross-section

and contact with the portal appears significantly closer on one side (Figure 5.2 c). This, once again, corresponds to the subunit which is most extended, with no visible density between domains. Tilting occurs in the y plane with a cross section in the x plane appearing far more symmetrical. An interesting feature is the outward extension of density at the bottom of the NTD. Masking this region out did not significantly alter the reconstruction, highlighting these features as the densest within the map. It is unclear if this represents a true extension of the NTD, which seems unlikely based on the crystal structure, or corresponds to emerging DNA or small terminase.

2a) C5 symmetry expansion of symmetrical classes

Whilst the motor is unlikely to adopt strict C5 symmetry, it is highly possible that a subset of motors will be present in a pseudo C5 planar state. This is seen in several motor structures with complete ADP bound occupancy (76). Several classes from method 2 (re-extraction of motor density from the portal-motor class and classification) showed near C5 symmetry. Particles from these classes were pooled and classified using C5 symmetrical averaging and local searches, generating a single class. This class was then subject to C5 symmetry expansion and reclassified once again, using local searches, into 5 classes. After removing duplicates, the best class was refined using C1 symmetry (Figure 5.2a: a).

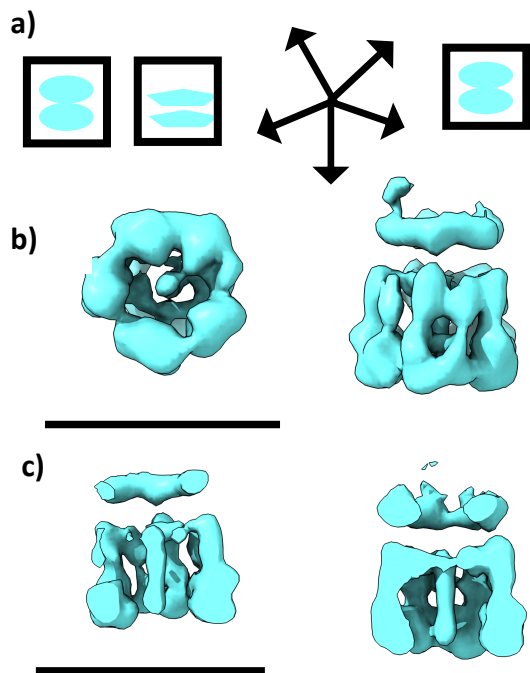


Figure 5.2a): Processing scheme 2a - symmetry expansion of classes showing C5 and reclassification a) schematic b) resultant 3D refinement viewed from outside the prohead and from the side c) cross section views on X and Y planes. Scale bar measures 150 Å.

Whilst the broad architecture of the motor resembles that from previous reconstructions, there is no visible contact between large terminase and portal clip (Figure 5.2 b). This is likely due to the averaging out of signal, as a result of classifying particles with C5 symmetrical averaging. This averaging may have also obscured the off-axis tilt observed in the NTD and CTD in other methods above, effectively aligning the portal, CTD and NTD on the same axis, in parallel planes. More promisingly, however, individual subunits appear clearly defined, and more closely resemble the crystal structure. Protraction of the NTD outward is also no longer apparent.

2b) C5 symmetry expansion and re-extraction

In a variation of the previous approach, classes displaying apparent C5 symmetry following re-extraction of motor density from the portal-motor class, were selected. These particles were once again classified using C5 symmetrical averaging and local searches, into a single class. This class was subject to C5 symmetry expansion, and resultant particles reextracted with an upward shift to include portal signal. 3D classification into 5 classes was then performed with C12 symmetrical averaging, and a mask made to encompass only the portal

density. The best class was then selected for refinement in C1 (Figure 5.2b: a). By selecting based on both C5 motor symmetry and C12 portal symmetry, two separate measures of validation were used in choosing the final particles.

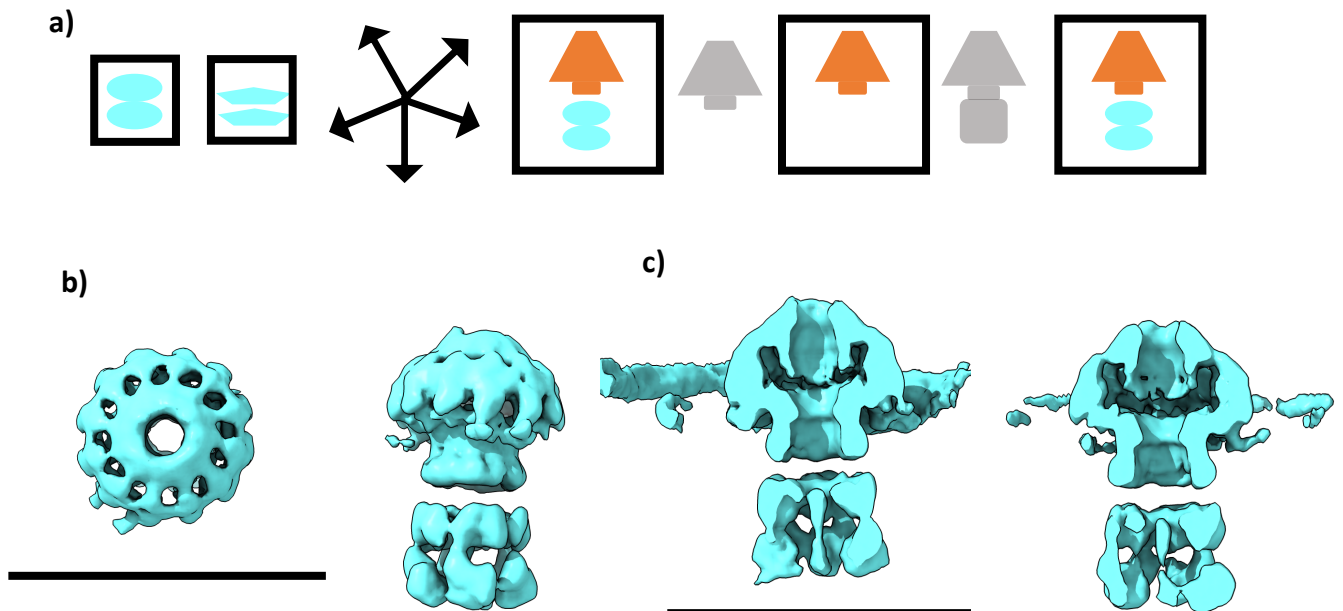


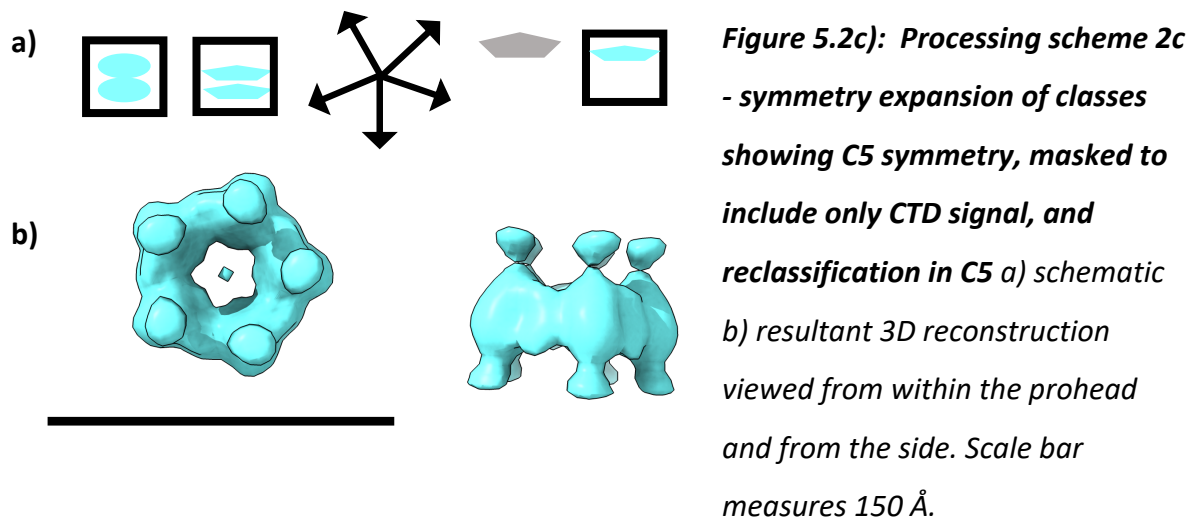
Figure 5.2b): Processing scheme 2b - symmetry expansion of classes showing C5, re-extraction to include portal signal, and reclassification in C12 a) schematic b) resultant 3D refinement seen from top and side views c) cross section views on X and Y planes. Scale bar measures 250 Å. section views on X and Y planes. Scale bar measures 250 Å.

Although only approximately 8500 particles were present in this processing round, the reconstruction showed reasonable angular accuracy (7.65 °) and resolution (10.5 Å) relative to earlier maps. As for Method 2a), individual subunits appear distinct and reasonably well resolved and compact, and DNA is clearly present within the large terminase channel. However, as for Method 2a) there is no contact between the portal and large terminase (Figure 5.2b: b,c) and there is no “Z” shape tilt within the cross section.

2c) Isolating the CTD pentameric ring

For $\Phi 29$, the CTD ring of the large terminase sits above the NTD and displays near C5 planar symmetry (177). Attempts were therefore made to isolate this domain within the HK97 map

and use appropriate symmetry for averaging during refinement. A reasonable 3D class could then act as a starting model for alignment of the likely more variable NTD. Classes displaying near C5 symmetry from Method 2 were again selected as input particles and a mask created to illuminate only the CTD ring. These were subject to C5 classification followed by refinement with C5 symmetry (Figure 5.2c: a).



Although a map of approximately the correct dimensions was obtained (Figure 5.2c: b), the angular accuracy never improved beyond 19 °. This suggests the refinement was effectively unable to align particles, and the map was not a valid starting model for aligning the NTD. It is possible (and indeed apparent in previous 3D classifications and refinements) that despite assuming a planar state the CTD plane tilts at varying angles away from both the portal clip and NTD domains, forming the Z-shape in cross-section. The variation in tilt angle may well complicate attempts to align particles during refinement.

3. Classification of expanded icosahedral particles by motor density

One unexpected feature which emerged from previous maps was the almost complete absence of classes containing only the portal clip domain without large terminase. This seems surprising based on the sample preparation and negative stain micrographs. If empty particles are included in refinements, they will contribute noise and hamper particle alignment. In an attempt to remove empty proheads, the portal-motor class from icosahedral expansion class was used to build a motor only template and mask. This mask was then reapplied to the icosahedral symmetry expanded shifted particle set, in further 3D

classification. By establishing the locus and rough shape of the motor, extracting based on signal in this region only should eliminate some empty particles.

Classification of symmetry expanded particles into 10 classes showed one clear motor class complete with the clip region of portal. This class was subject to two more successive rounds of 3D classification into 5 classes each. Once duplicates were removed from the final selected class, 361,000 particles remained, which was greater than the total number of particles for prohead refinement, at 330,000. This indicates that the technique was not successful at removing empty particles. Further classification produced three strong classes which were combined for refinement (Figure 5.3 a).

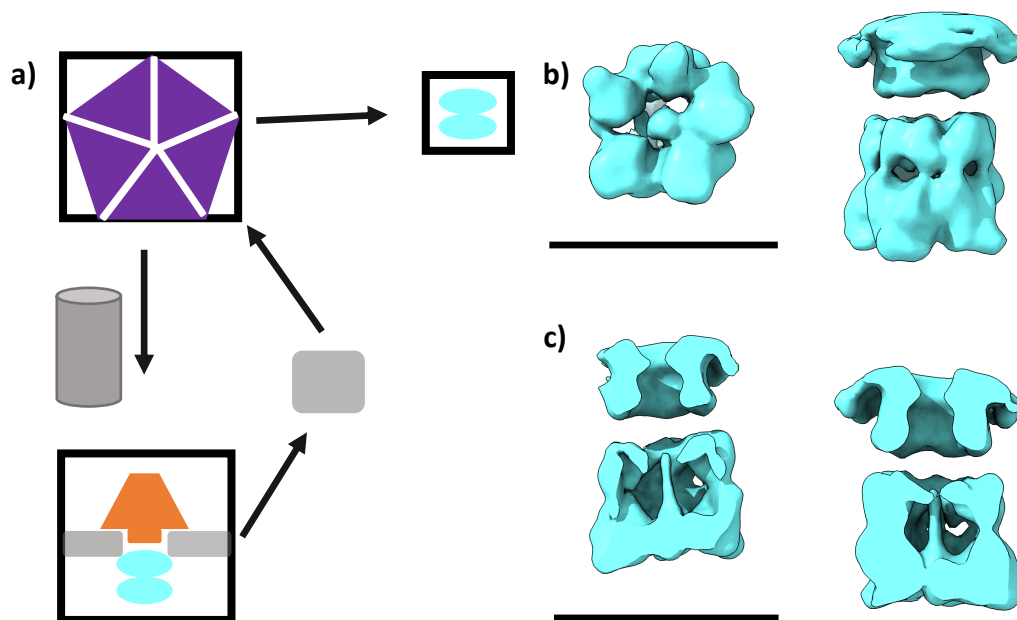


Figure 5.3): Processing scheme 3 -Re-extraction of icosahedral expanded particles in the motor locus followed by 3D classification a) schematic b) resultant 3D refinement viewed from outside the prohead and from the side c) cross section views on X and Y planes. Scale bar measures 150 Å.

The resultant map displays a combination of features from previous of reconstructions (Figure 5.3 b). As for Methods 2a and 2b, subunits are very well resolved and do not display extension outward. However, there is also a clear tilting within the motor which is displayed in the cross section in different planes (Figure 5.3 c). Whilst some subunits appear more

extended than others, there is no complete absence of density joining the two domains as for methods 1 and 2. From beneath, we can clearly see DNA in contact with 2 N terminus subunits (Figure 5.3 b). The estimated resolution is 12.9 Å and angular accuracy at 9.2 °.

4) C12 symmetry expansion of the portal

To examine the interaction between the portal clip domain and large terminase, the C12 reconstruction of the portal was subject to C12 symmetry expansion, to search for possible matching motor orientations that may be obscured by the portal's C12 symmetry. The resultant particles were reextracted with a shift in the z direction, to include the motor density, and classified into a range of class numbers without angular refinement. This process was repeated using particles reextracted to include only the motor region (Figure 5.4). Unfortunately, there was no clear definition in any of the classes. This may be because the portal clip domain does not adopt strict C12 symmetry, as demonstrated in the reconstruction, even if upper regions of the portal (e.g. wing domain) do conform to C12 symmetry. As mentioned in Chapter 4, portal particles may also represent a mixture of those bound to a motor, and unbound portal proteins. The respective clip domains are likely to adopt different conformations, with variability in the extent to which they conform to C12 symmetry.

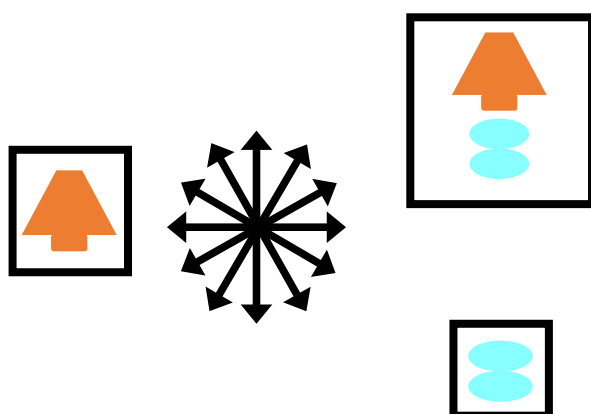


Figure 5.4): Processing scheme 3 - C12 symmetry expansion of portal refinement, re-extraction to include both motor and portal, and only motor, followed by downstream classification

5) Comparison of C12 and C5 angles for C1 reconstruction

A script provided by Dr Pavol Bardy was employed to assign angular constraints (257) for an asymmetric reconstruction of the motor. In Chapter 4, this method was successfully employed to produce a high-resolution asymmetric reconstruction of the *in-situ* portal in contact with the prohead vertex.

Particles from the portal-motor class after icosahedral expansion were selected and duplicates removed. Reconstruction of the portal region only, in C12, was then performed using a tight mask producing a high-resolution map with good angular accuracy. The same process was applied to the surrounding prohead vertex density, which was refined with C5 symmetry. The run_data.star files from each job were then fed into the script and the resultant star file 3D classified into a single class, where angular refinement was limited to very local searches (5 pixels, 1.8 degrees) (Figure 5.5 a). Figure 5.5 b) and c) depict the resultant 3D class. A 3D refinement using limited angular searches did not produce observable features. 3D classification into multiple classes was also trialled in an attempt to divide contributing motor conformations, but did not produce maps with interpretable density, whilst resolution was limited.

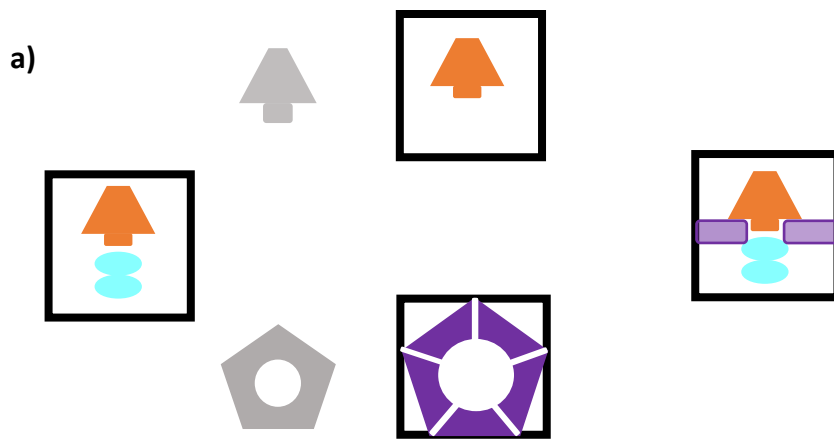
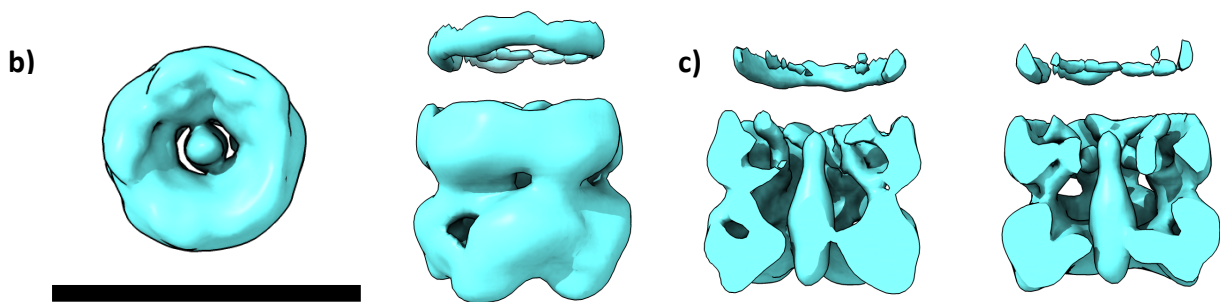


Figure 5.5): Processing scheme 5 - C12 and C5 angles from portal and prohead refinement respectively compared for C1 angles. 3D classification with only local angular searches a) schematic b) resultant 3D refinement viewed from outside the prohead and from the side c) cross section views on X and Y planes. Scale bar measures 150 Å.



This model shows near C5 symmetry, but clearly lacks detail within individual subunits (Figure 5.5 b). However, the cross-section shows clear distinction between domains and strong density for DNA (Figure 5.5 c). There is, no contact between the portal and large terminase. Based on the processing scheme, asymmetric contact should be resolved. It is thus likely that the portal-large terminase contact changes throughout the packaging cycle. This could affect the overall tilt of the motor relative to the prohead, which has potentially produced extra complications for classification and refinement. Alternatively, the portal - large terminase contact may be constant, but fully independent to the portal – prohead mismatch.

6) Focussed classification and subtraction from a single data set

The processing of the eBIC data set alone, produced the highest resolution prohead map, so this data set was selected for solo focussed classification. As for method 2, the portal/motor class particles were re-extracted into a smaller box size and subject to 3D

classification without alignment. The class displaying clear individual subunits was used as a template for a mask, and particles from this class were refined without particle alignment, and without a mask. The resultant run_data.star file was subject to particle subtraction using the newly made mask. Successful particle subtraction was confirmed using relion_reconstruct, which showed much of the prohead density removed. Subtracted particles were subsequently subject to refinement with local angular searches (1.8° and 1.8°). The resolution and angular accuracy of the map were improved to 9.6 Å and 9.2° respectively. Following post-processing, the structure appeared more defined, and the resolution reached 8.8 Å. (Figure 5.6). The overall architecture shows significantly clearer features. Each of the five subunits shows strong density in both domains and there is clear contact between two large terminase monomers and the portal clip.

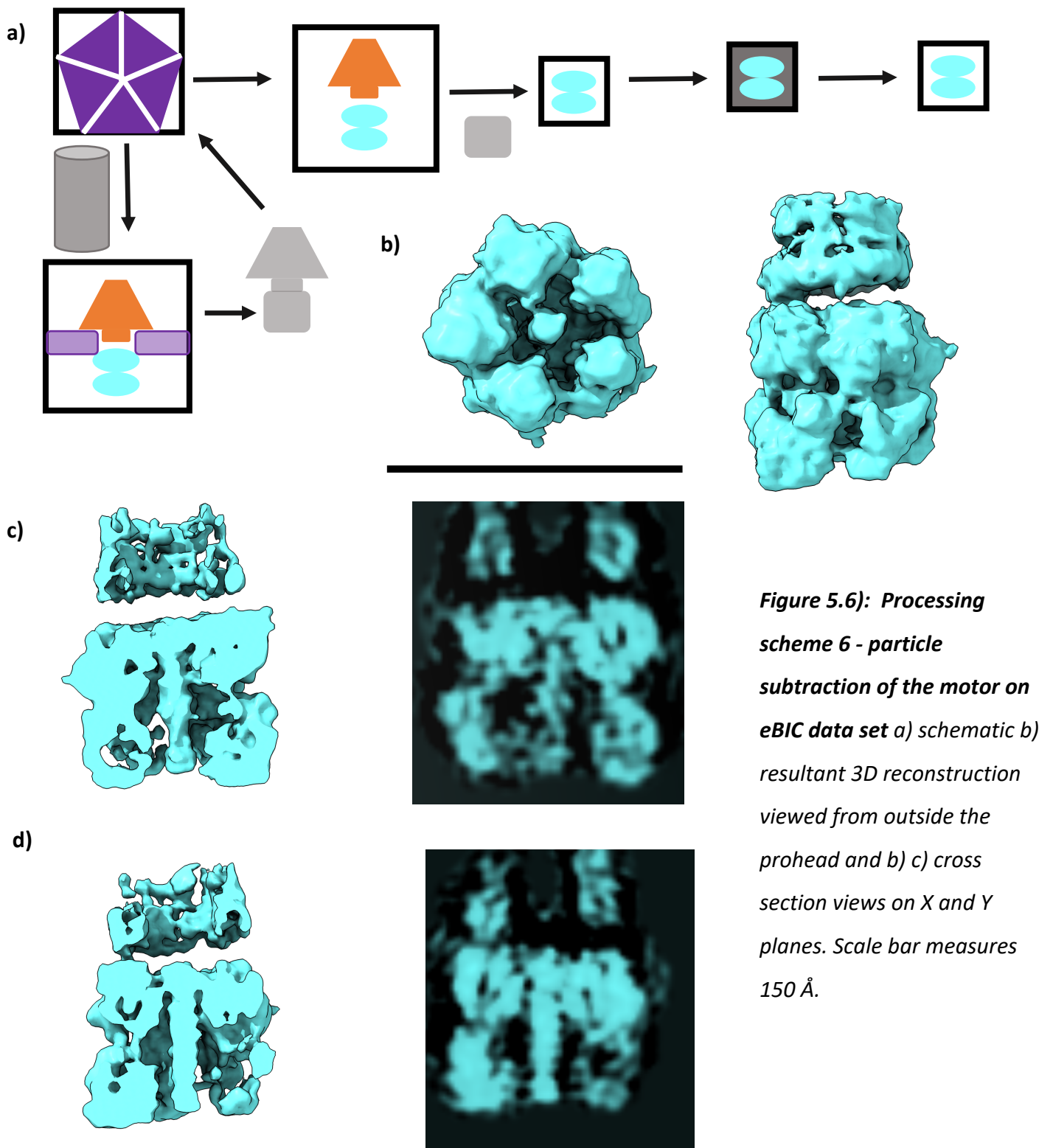


Figure 5.6): Processing scheme 6 - particle subtraction of the motor on eBIC data set a) schematic b) resultant 3D reconstruction viewed from outside the prohead and b) c) cross section views on X and Y planes. Scale bar measures 150 Å.

6a) Focussed classification and subtraction from a single data set for the portal/motor complex

Based on the success of this technique, the run_data.star file from the refinement was used as a template for re-extraction to include portal density. The same protocol of particle subtraction was employed, using a mask to encompass the entire motor/portal complex. Subtracted particles were then refined with limited angular searches of 1.8° and 1.8° . This gave rise to a 7.4 Å reconstruction, with an estimated angular accuracy of 3.7° . Figure 5.6a displays the resultant reconstruction following post processing. The use of broader angular searches during refinement, or iterative refinement, gave rise to higher overall resolution estimates, but reconstructions were dominated by the 12-fold signal of the portal, with a less well resolved motor.

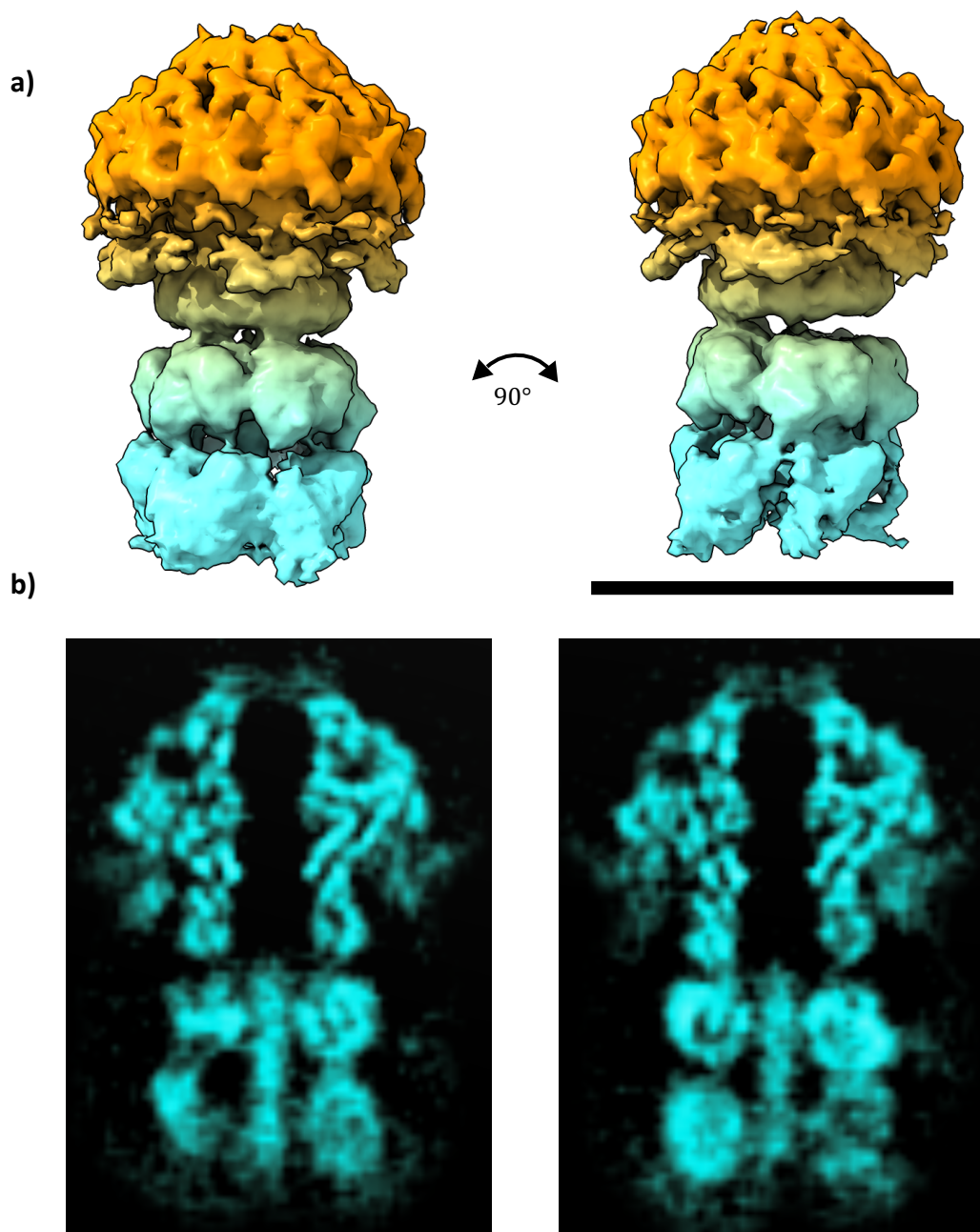


Figure 5.6a): Particle subtraction of the portal/motor complex from the eBIC data set a) resultant 3D reconstruction b) cross section views on X and Y planes. Scale bar measures 150 Å.

Discussion

The reconstructions of the motor and portal/motor complex presented in Figures 5.6 and 5.6 a) respectively show well defined features and clear contacts between proteins and protein subunits. One striking feature is the apparent lack of DNA density within the portal channel. Packaging assays have consistently shown protected DNA within proheads, suggesting that the motor may be trapped in the packaging termination state – having cleaved DNA but remaining bound to the portal. This would produce the blunt DNA end seen at the top of the motor. However, DNA packaging assays performed in the absence of DNAase consistently show complete DNA substrate (for example Figure 3.5 a presented in chapter 2 figure) suggesting the motor is not producing DNA cuts. It has also been suggested that the HNH-endonuclease gp74 is critical for DNA cleavage (1).

A second surprising characteristic is the asymmetrical interaction between the large terminase and the portal clip domain. Contact appears only between one (or possibly 2) large terminase subunits, positioning the entire motor tilted away from the portal axis. The seemingly fragile nature of the interaction is particularly surprising given the enormous internal pressure the motor must work against at late-stage packaging.

Fitting a monomeric large terminase X-ray structure into oligomeric assembly

In order to examine the architecture of the motor in more detail, five subunits of the large terminase, for which crystal structure was already available (6Z6D (1))were fitted into the reconstruction presented in figure 5.6. The C- and N-terminal domains of the crystal structure were divided, to allow for independent rigid-body fitting of each domain, and for varied linker lengths around the ring. Whilst the motor is clearly highly dynamic and flexible, this motion is likely largely confined to inter-subunit and interdomain movement as opposed to broad domain rearrangement. This is supported by the globular nature of the two domains and the ability to crystallise single-domain constructs more readily than full-length constructs (160,185,187–189,267,268). Rigid body fitting the domain structures into the map, in chimera, indicated that the CTD is more compact and fits better into the “top”

ring, directly below the portal, with the NTD below. The apparent cleaved DNA end at the top of the channel supports this topology. Despite the low resolution of the HK97 map, the fitted structure and map show good agreement (Figure 5.7 a,b).

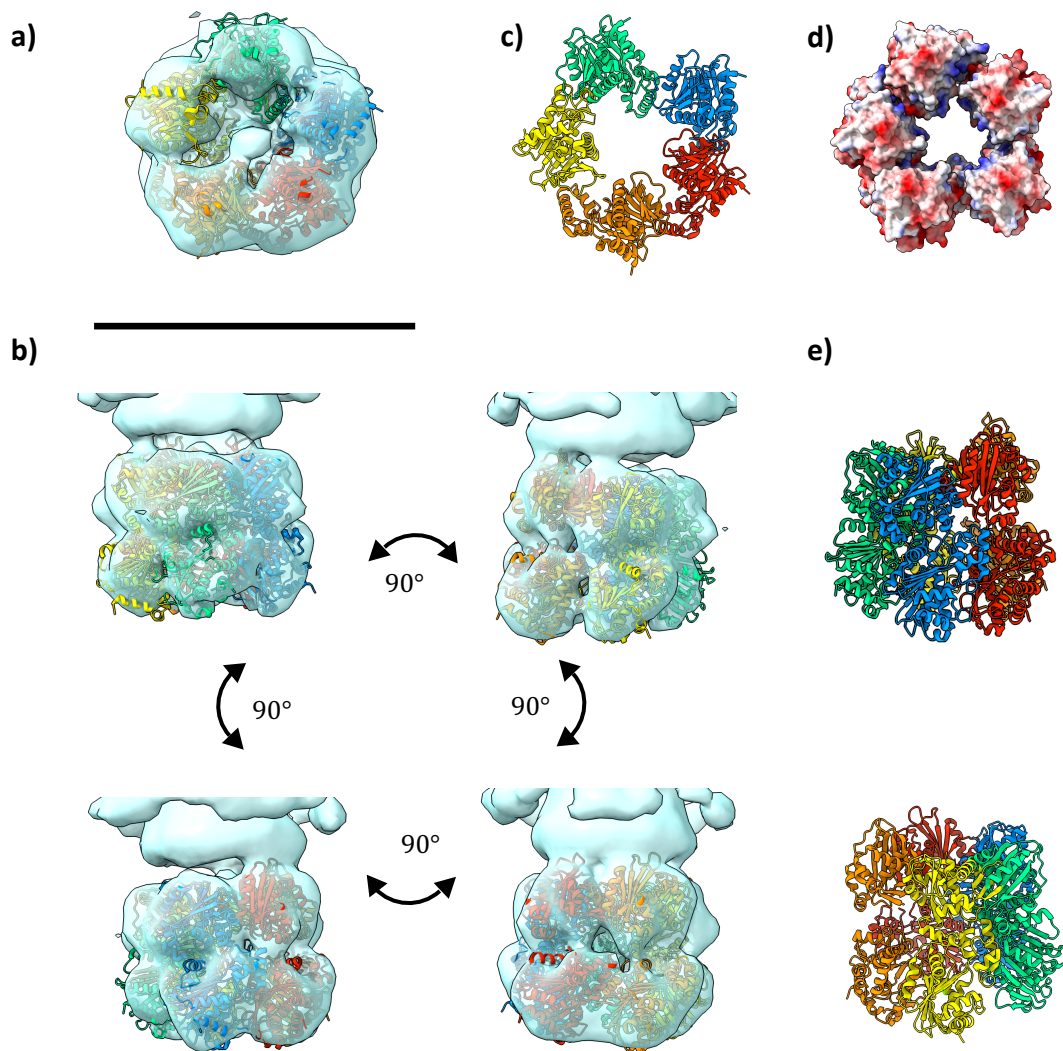


Figure 5.7: Pseudo-atomic structure of the large terminase. Crystal structure of HK97 large terminase fit into motor map 5.6 after dividing N and C terminus domains a) viewed from the prohead b) side views c) NTD ring with CTD removed d) surface electrostatic potential of the NTD ring e) side views of the pentameric atomic model of large terminase

Positively charged residues within the N terminus domain were explored to investigate potential DNA binding residues. Figure 5.7 d) shows the electrostatic surface of the NTD ring, with multiple positively charged residues within the central tunnel marked in blue. These include Arg 90 and Lys 102.

Potential modes of DNA translocation

T4 like packaging

The docked crystal structure indicates variability within the extension of the linker domain between the NTD and CTD. In Figure 5.7 e) the subunits depicted in blue and green appear compact, with a short distance between domains. In contrast, the red and orange subunits are further separated, and the linker between them more extended. Contraction from an extended 'relaxed' ATP bound state to a contracted 'tense' state is proposed as a translocation mechanism for T4. In contrast to HK97, a cryo-EM structure of the active T4 packaging motor models N terminus attached to the portal clip, with the CTD ring below (181). However, this map has a low resolution, and FRET measurements instead indicate that the CTD of T4 contacts the portal clip(266) . For T4, ATP hydrolysis is proposed to occur sequentially around the ring, so that only a single subunit is ever present in the tense contracted state. This does not fit well with the HK97 structure presented here, where subunits appear to span a range of conformations.

Φ29 like packaging

Like HK97, Φ29 large terminase is orientated with the large terminase CTD ring closer to the portal. However, in this case the interaction is mediated by the additional pRNA. A high resolution cryo-EM map indicates that sequential movement of the 5 NTD domains from a cracked helical conformation into a planar ring may drive translocation (177). A summary of the translocation model proposed for Φ29 is described in Figure 5.8.

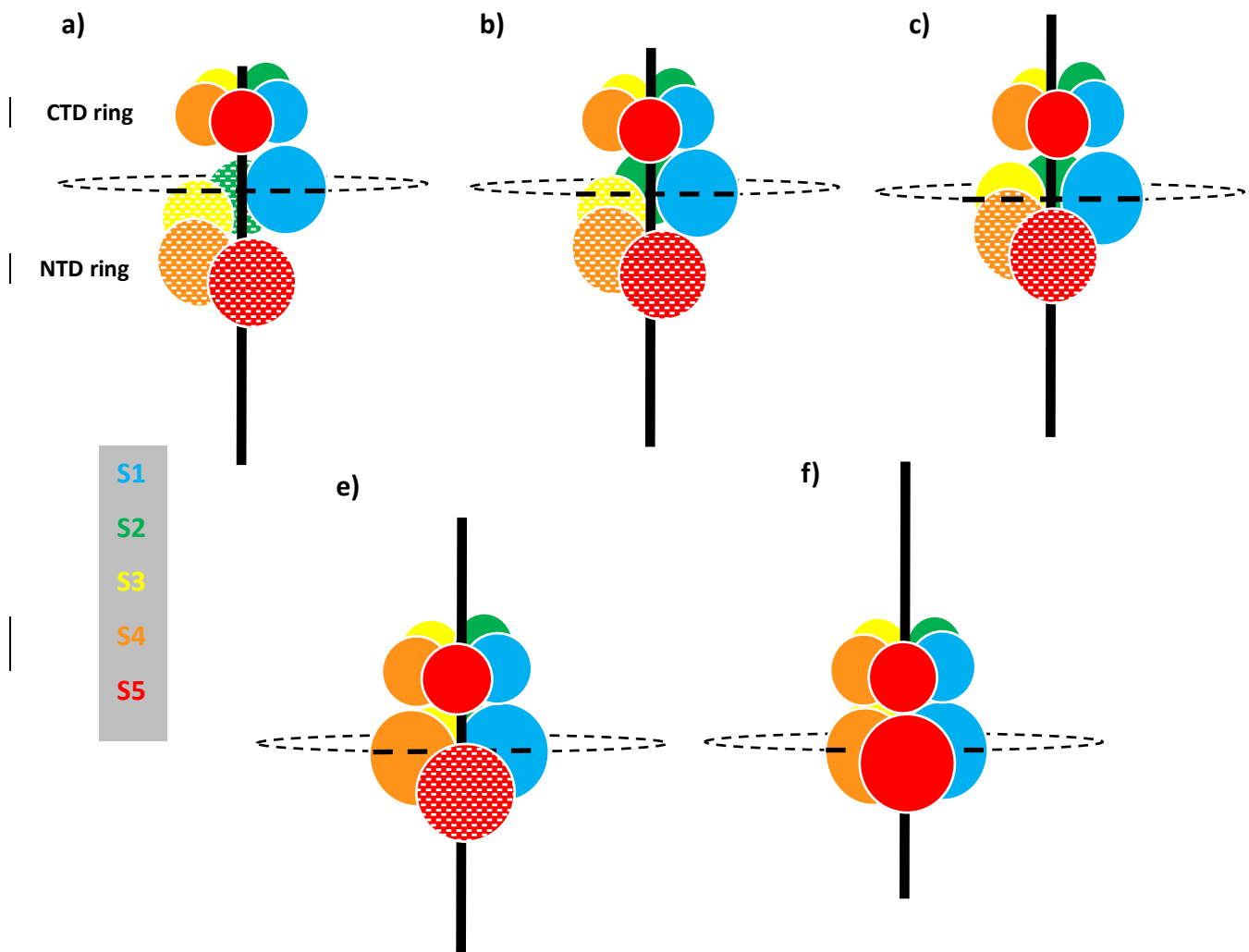


Figure 5.8: Schematic of cracked helix to planar model of DNA translocation. a) ATP bound state with NTD in cracked helix conformation b) c) d) successive NTD subunits move into plane as ATP is hydrolysed e) ADP bound conformation in planar state

For the HK97 motor reconstruction, however, the variation in subunit length does not appear to strictly fit the cracked helix conformation. Instead, the most extended subunits appear adjacent from the most compact, echoed by the “Z” shape architecture in previous asymmetric reconstructions. In addition, based on the inability to resolve the CTD ring it seems unlikely that this component remains rigid and planar in HK97, as for $\Phi 29$.

An alternative packaging model to explain the features of the HK97 reconstruction is presented in Figure 5.9. Here, ATP hydrolysis induces a conformational change which is propagated through the linker, causing the NTD and CTD to contract together, as suggested

for other viruses (177,181,184). This has the double effect of pushing DNA into the prohead and disrupting the CTD -portal interaction. This mechanism is supported by visible interaction between extended large terminase monomers with DNA, in favour of compact large terminase monomers (Figures 5.1 b, 5.3 b and 5.6 a). However, it would not clearly fit the dwell – burst translocation cycle established for Φ 29, as a fully ADP bound pentamer would lose contact with the portal entirely. It would also suggest that the model shown is not in the fully ATP γ S bound state, where we would expect all subunits to be bound to the portal.

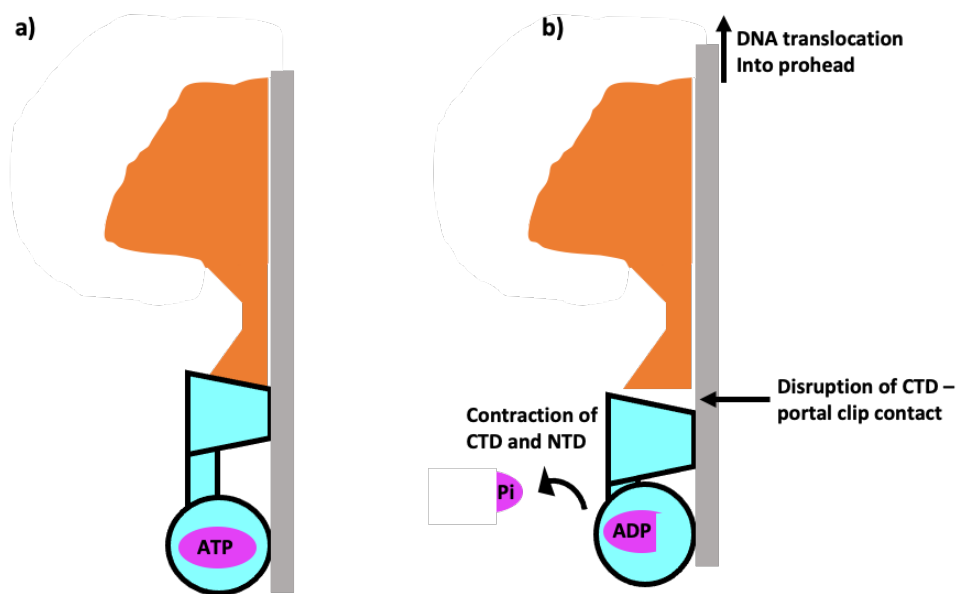


Figure 5.9: Proposed mechanism for DNA translocation by HK97 large terminase a) The CTD contacts the portal clip, the ATP bound NTD is extended away and binds DNA b) on ATP hydrolysis, conformational changes within the link domain cause NTD and CTD to contract. This moves DNA into the prohead and simultaneously breaks the contact between the CTD and portal clip.

FtsK like packaging

Perhaps a more promising comparison can be made to the proposed translocation mechanism for *E. coli* DNA translocase FtsK. For this molecular machine, a sample of the

hexameric motor with DNA and ATP γ S produced a variety of different 3D classes. The best resolved subset shows all 6 subunits adopting a unique conformation, with variable angles of separation between the α and β domains (Figure 5.10 b). 4 subunits engage in DNA binding, forming two strands of DNA binding residues in a spiral around the DNA. Three of these active sites show clear bound ATP γ S while the remaining appears to be caught in a state of ATP γ S hydrolysis. The remaining 2 subunits are ADP bound and do not engage with DNA (230). (fig) ADP bound, apo, and DNA free states have also been resolved: each adopting a broadly planar hexameric states (230,269). These structures largely resemble apo and ADP bound states in terminase motors. However, the proposed translocation mechanism for FtsK never passes through a fully ADP bound state, with the rolling change in conformations around the ring shifting sequentially akin to a spiral escalator. The changes in angles associated with each position causes DNA to be translocated (230).

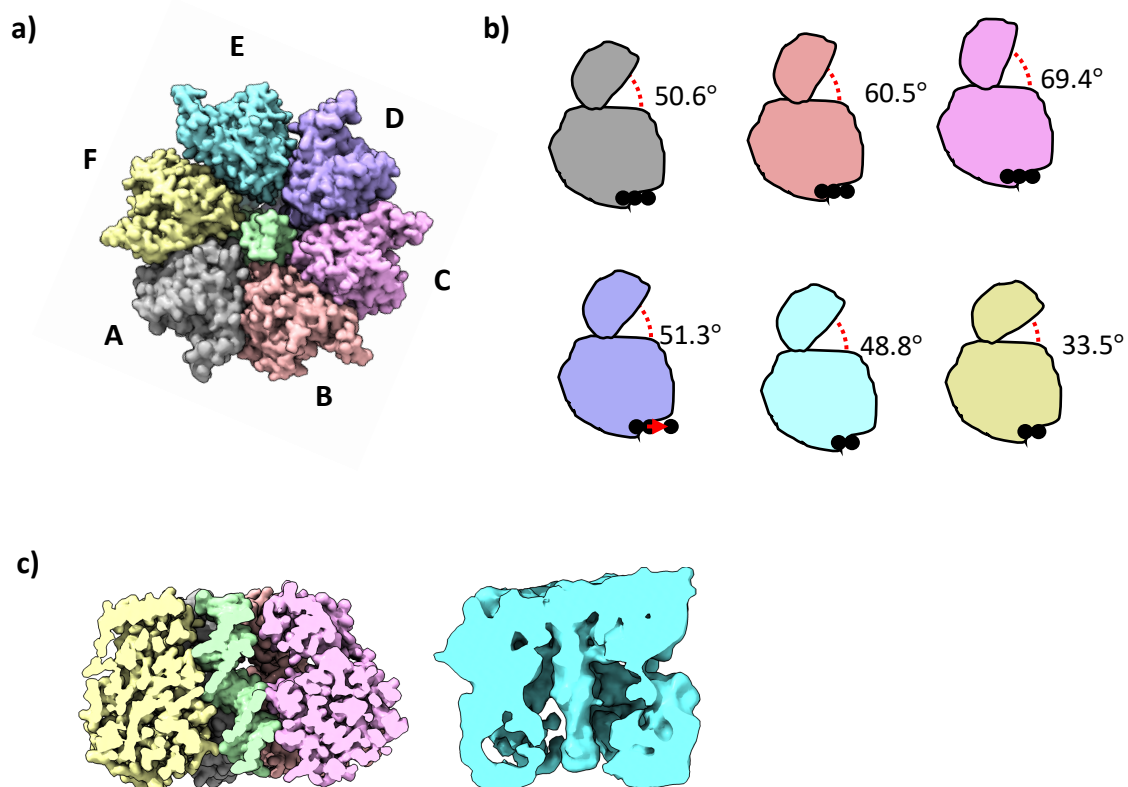


Figure 5.10: The FtsK DNA translocation mechanism: a) Viewed from the β domain
 b) Angular separation between the α and β domains of each FtsK subunit
 c) Cross sectional views of the FtsK motor and the HK97 motor showing comparative asymmetry

This mechanism of translocation fits well with the HK97 motor reconstruction. As for FtsK, each subunit adopts a slightly different orientation. DNA is also engaged with the more extended, assumed ATP γ S, bound subunits within the map. The challenge of this comparison is the necessary adaptation of this 6 subunit mechanism to 5- subunit motor, as well as negotiating the portal-protein interaction. However, the ever-changing conformations around the ring could actually explain the seemingly transient contact with the portal: perhaps subunits only engage in portal contact in one or 2 of the unique positions so that contact is broadly 'averaged out' over any combination of states

Examination of angular distribution

The angular distribution of the same map (Figure 5.6) was examined, to assess whether this was a potential factor limiting resolution and angular assignments (Figure 5.11). Clearly, angular distribution of particles between Euler angles is uneven, with the majority of views from directly beneath the motor, along the fivefold axis, which could limit resolution. However, the orientation bias is not as extreme as in some documented cases, with all angles in the C1 refinement represented. Based on the small number of particles used in the reconstruction relative to the whole data set (17,500 out of 75,900), this does not necessarily indicate orientation bias of the entire picked particle set. Instead, it may be that the most promising 3D classes represent 'head on' motor views, as interference from prohead and portal signal is less influential.

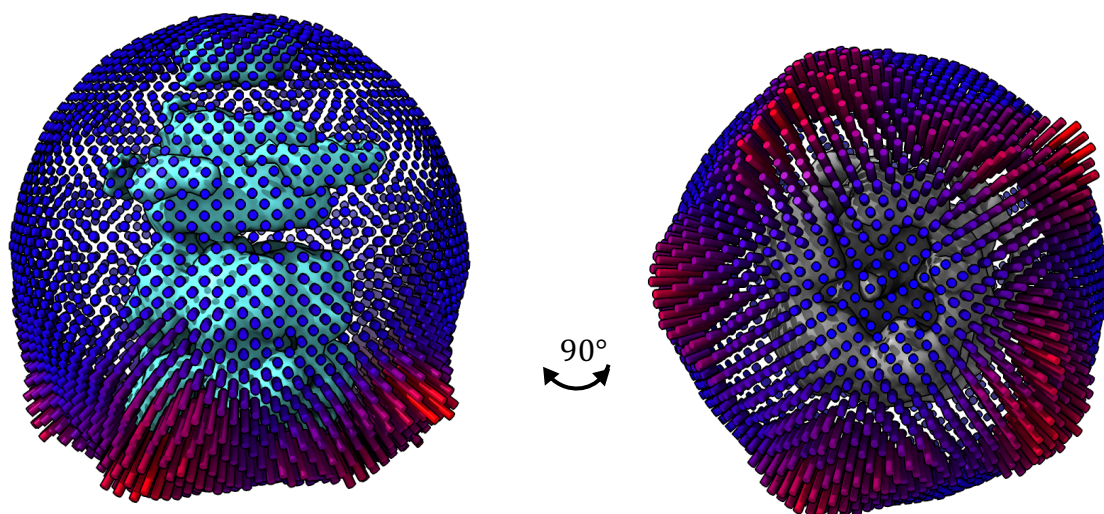


Figure 5.11: Angular distribution for motor map 5.6 depicted from the side and from outside the prohead

Alignment between the prohead and motor

Additional refinements were performed in order to establish if the orientation of the motor with respect to the portal had any dependence on the orientation of the portal with respect

to the capsid vertex. Firstly, the portal and motor model presented in method 2b) was refined again, with C5 symmetry, limited angular searches and no mask (Figure 5.12 b).

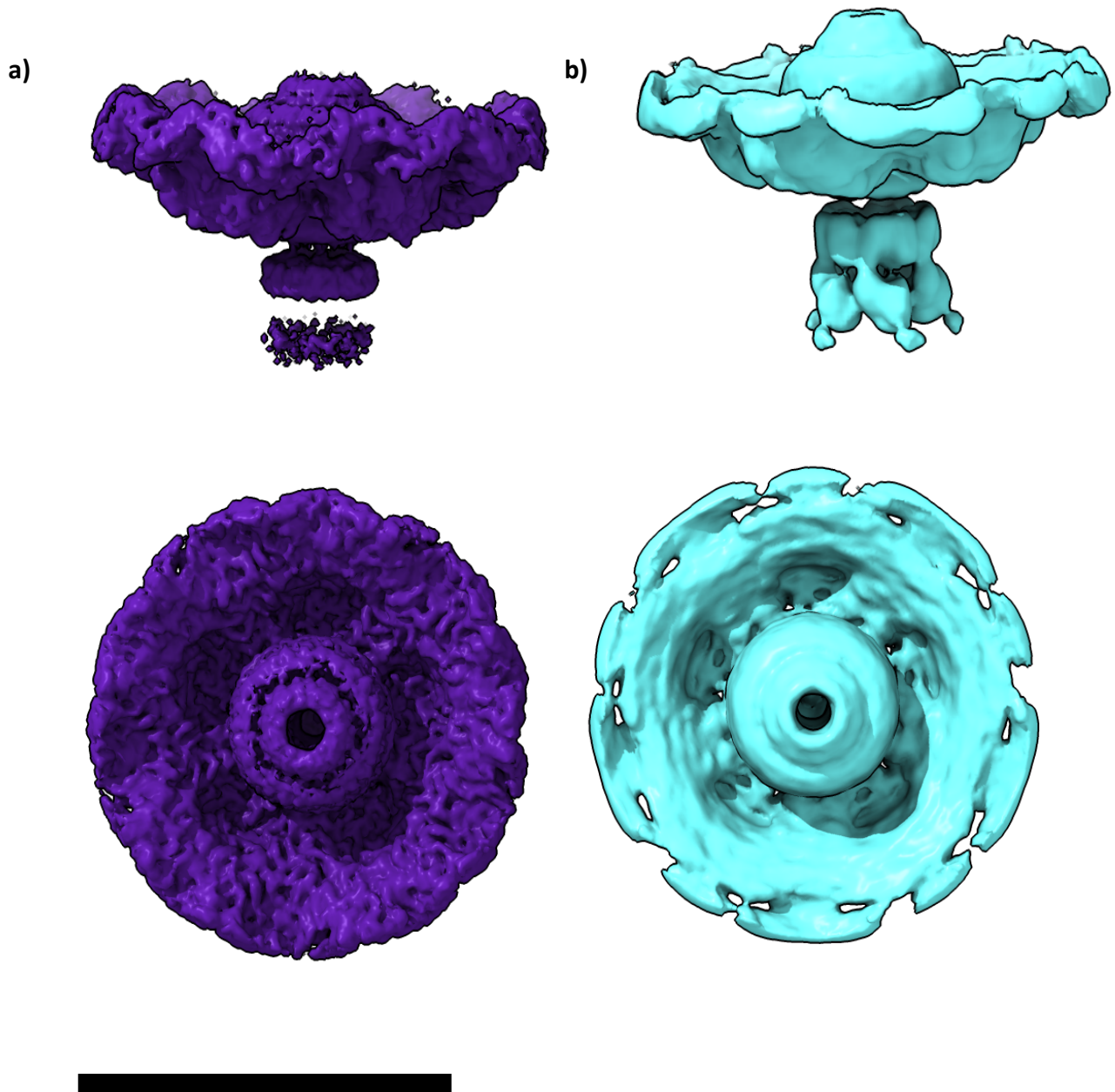


Figure 5.12: Alignment between the prohead and motor. Refinement of *run_data.star* from map 2b) from side and viewed from the prohead a) using wide angular searches and b) using limited angular searches. Scale bar measures 300 Å.

Here, the motor is well defined but the prohead density lacks definition. As a consequence, the global resolution estimate is low, even though the quality of the motor map region is high. The refinement was repeated with identical particles but with a wider angular search,

and masking out of the portal and motor density, so that only the capsid signal was used in alignments. This produced a well resolved capsid vertex but showed only two broad rings for the motor (Figure 5.12 a). Clearly the pentameric nature of large terminase is not simply an artifact of the C5 prohead symmetry, and indeed these the symmetries of the two complexes don't appear to depend strongly on one another. In addition, this observation appears to oppose the idea that morphing of the portal clip domain by the prohead facilitates binding to large terminase.

Absence of small terminase

Although small terminase is essential for stalling there is no obvious density for the oligomer within the motor structures. The nonameric ring of small terminase has a molecular weight of 160 kDa, which should be discernible. This suggests that the mechanism of stalling may be model 2, presented in Chapter 2 (Figure 2.11b), where small terminase bound at the *cos* site is displaced by incoming large terminase. A challenge for this displacement hypothesis, is the ability of small terminase to increase efficiency of packaging by large terminase (1). *In vivo*, small terminase is required for packaging initiation at the *cos* site, so perhaps more DNA is packaged in assays due to a greater number of initiation events, as opposed to enhanced speed.

There still remains a small possibility that signal from so many larger components make it impossible to distinguish small terminase, which could interact with the motor in multiple possible conformations with respect to the bulk of the complex. There is also some unexplained density extending outward from large terminase in several maps (Figure 5.2 b,c) which could be attributed to small terminase. Based on the recent structure of small terminase (Figure 1.10 b, (155)) the *cos* cut site resides approximately 10 bp upstream of the small terminase binding site. This stretch could correspond to the length of DNA passing through the NTD central tunnel, seen in all reconstructions.

Probing a potential glutamate switch

Recent work on several related phages has shed light on a potentially conserved mechanism utilising a conserved glutamate or glutamine to coordinate the transition of individual NTD domains from an active DNA binding state, to an inactive state, on ATP hydrolysis (185). In each case the polar or charged switch residue appears to “fix” the catalytic glutamate to point away from the active site, rendering it inactive. To determine if HK97 terminase could also utilise this glutamate switch mechanism the crystal structure was aligned with multiple terminase ATPase domains. Perfect alignment was seen between the catalytic glutamate residues of HK97 and P74-26 (glutamate 154 and 150). In turn, the glutamine switch residue of P74-26 also overlaid perfectly to glutamine 92 in HK97 (Figure 5.13).

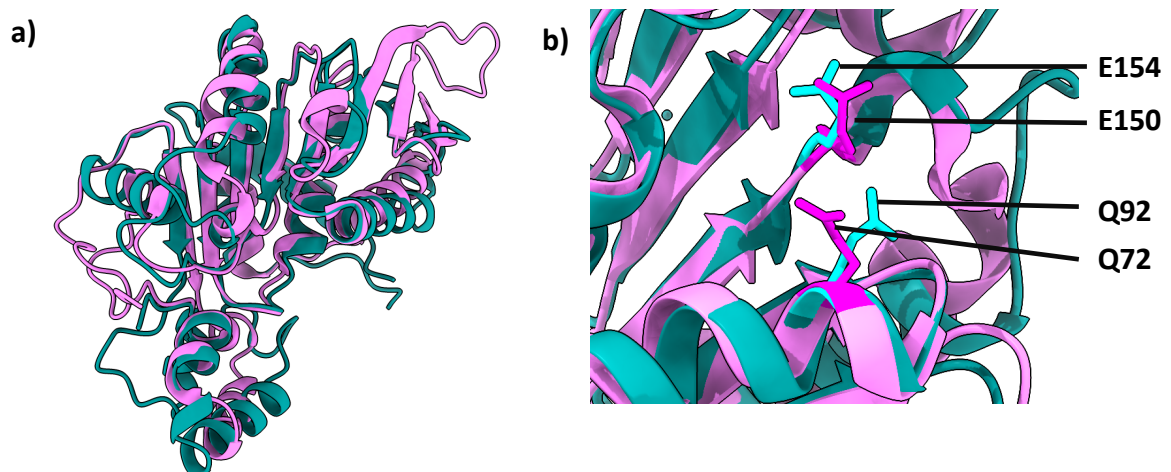


Figure 5.13: Alignment of HK97 large terminase ATPase domain with the large terminase of P74-26 a) full structures with HK97 in cyan and P74-26 in pink and b) alignment of catalytic glutamate residues E154 and E150 and potential glutamine switch residues Q92 and Q72

This suggests that, like P74-26, in the apo state HK97 may adopt a metastable state between the active and inactive positions. An analogous mutant to P74-26, which in HK97 is K92Q, was cloned and purified, but did not yield crystals during crystallisation trials. However, based on AlphaFold modelling, this K92Q mutant shows an approximate 60 ° rotation in the

χ_1 angle of the catalytic glutamate E145, positioning the residue towards the catalytic switch and thus away from the ATP γ -phosphate (Figure 5.13). The distance between the catalytic glutamate E145 and critical catalytic K60 within the walker A motif is increased. This is apparent in comparison to both the crystal structure of large terminase, and an AlphaFold prediction of the wild type domain. The active site is thus likely rendered inactive. The mutant also shows minor rearrangements within the lid domain of the ATPase, which is thought to coordinate movement between the ATPase and nuclease domain (Figure 5.14 c)d) (189,270). Again, this is significant relative to both reference structures. Future work should aim to monitor ATPase activity and DNA packaging ability of the K92Q mutant protein.

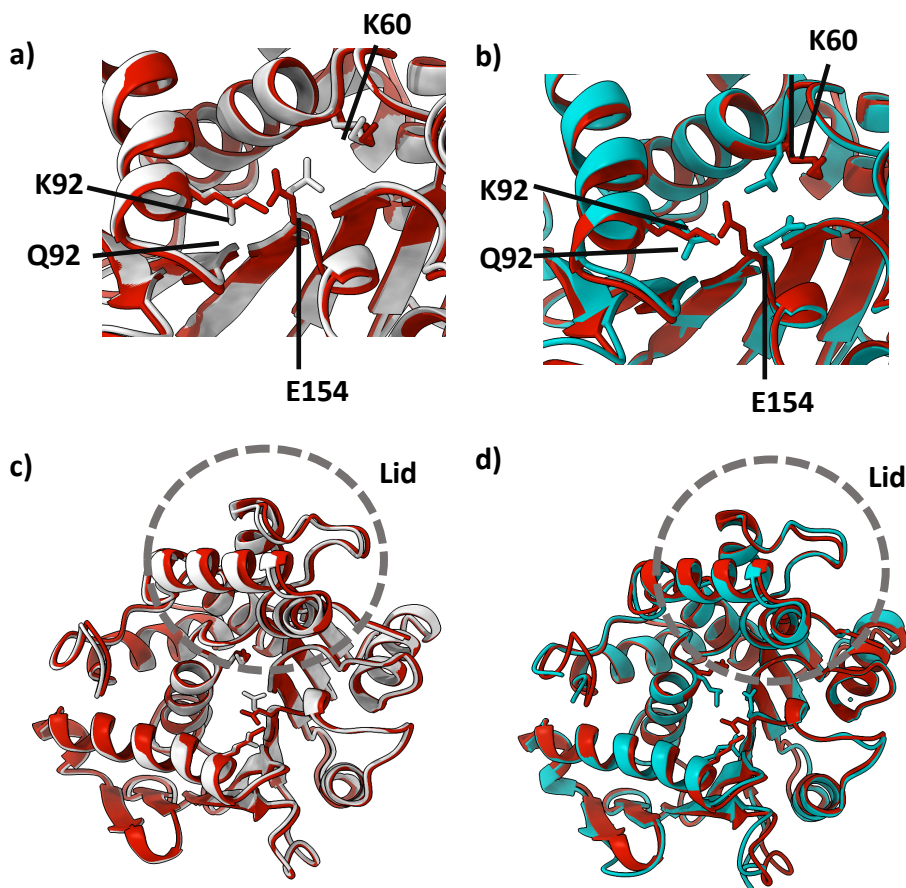


Figure 5.14: AlphaFold prediction of HK97 large terminase K92Q (red) a) Alignment with AlphaFold prediction with wild type prediction active site (white) b) wild type crystal structure active site (cyan) c) wild type prediction whole domain (white) and d) wild type crystal structure whole domain

Conclusion

In the single portal-motor class following icosahedral symmetry expansion and classification, two uniform stacked rings are present below the portal. This conformation is also apparent in the 3D asymmetric reconstruction of the whole packaging complex including the prohead (Figure 5.15). There is no obvious contact between the motor rings, or between the portal and top ring, indicating that this structure represents an average of numerous states.

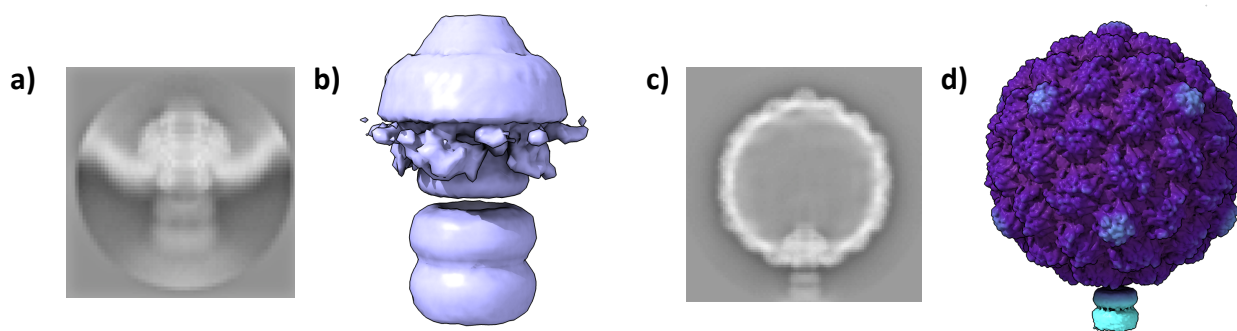


Figure 5.15: Motor and portal class following icosahedral symmetry expansion and focussed classification a) 2D class b) 3D class. Asymmetric reconstruction of complete packaging machinery c) 2D class d) 3D refinement

Multiple processing strategies were employed in attempt to separate out the contributing conformations and refine the motor. In each case, only a relatively small number of particles is present in the final reconstruction (between 8,000 and 40,000) supporting the notion of a mixture of discrete states within the data set.

Certain broad features are present in all refinements: 5 distinct subunits are clearly visible, as indicated by total internal reflection fluorescence microscopy analysis, and analogous to other double stranded DNA phage and herpes motors (72–76). Whilst some 3D classes display near five-fold symmetry, there is no clear planar state, as seen for $\Phi 29$ and ascc- $\phi 28$, in the ADP bound conformations (76). Cross sections in X and Y planes of each reconstruction also suggest broad contact of DNA within the CTD ring, compared to interaction with just one or 2 NTD subunits. This could signify that the motor adopts the

termination state, where the CTD is fully engaged. The abrupt disappearance of DNA density at the entrance to the portal channel also indicates that packaging may be terminated.

When fivefold symmetric averaging is employed within the processing strategy, large terminase subunits more closely resembled the compact crystal structure. The portal clip, CTDs and NTDs tend to reside as flat planes on the same axis, with no portal – large terminase contact. This indicates that portal-motor contact is not symmetric around the ring. Interestingly, the fivefold symmetry of the pentameric motor does not align with the fivefold axis of the prohead which challenges the notion that portal is prepped for motor binding by morphing of the clip domain to match the prohead (181) .

When asymmetric 3D classification is performed around the masked motor density, classes with clear features show strong asymmetry. Large terminase subunits are separated by variable distances, and the cross sections show a “Z” shape: with contact between the portal and CTD on one side only, where subunits are most extended. This is suggestive of a translocation mechanism whereby both contact between the portal and CTD, and extension between the CTD and NTD vary throughout the packaging cycle.

Dividing the crystal structure of the large terminase monomer into individual N- and C-terminal domains allowed them to fit independently into the motor map depicted in figure 5.6, with which they showed good agreement. This model supports the assumed topology, with the CTD ring below the portal. Positive amino acids line the NTD lumen, whilst Q92 is suggested to adopt the role of a glutamine switch: coordinating the transition from an ATP bound DNA binding state to an ADP bound low DNA affinity state. There is no clear density for the small terminase in any of the reconstructions, suggesting it may have been displaced by large terminase on stalling.

Overall, multiple approaches to localised reconstruction of the DNA packaging motor, have revealed significant insights into both the symmetry and architecture of the motor. Further studies should be directed at improving the resolution, to better understand the translocation mechanism. A structure of the isolated large terminase pentamer in complex with DNA and, potentially small terminase, would also be highly informative.

Accompanying Material

The following script was provided by Dr Pavol Bardy for asymmetric reconstructions of the unique portal vertex (257).

```
#!/usr/bin/env python

c5In=open('input1.star', 'r')
c12In=open('input2.star', 'r')
starOut=open('output.star', 'w')

c5 = [0.0, 120.0, 240.0]
c12 = [0.0, 72.0, 144.0, 216.0, 288.0]

c12Inc = 72.0
c5Inc = 120.0

c5Line = c5In.readlines()
c12Line = c12In.readlines()

count=0
for line in c5Line:

    c5Split = line.split()
    c12Split = c12Line[count].split()
    count=count+1

    if (1 <= len(c5Split) <= 2) and (1 <= len(c12Split) <= 2):

        starOut.write(" ".join(c5Split))
        starOut.write("\n")

        if c5Split[0] == "_rlnAngleRot":
            rotColC5 = c5Split[1][1:]

        if c12Split[0] == "_rlnAngleRot":
            rotColC12 = c12Split[1][1:]

    if (len(c5Split) > 2 and len(c12Split) > 2):

        rotC5=float(c5Split[int(rotColC5)-1])
        rotC12=float(c12Split[int(rotColC12)-1])

        while not ((rotC5 >= 0.0) and (rotC5 < c5Inc)):
```

```

if rotC5 >= c5Inc:
    rotC5=rotC5-c5Inc
else:
    rotC5=rotC5+c5Inc

while not ((rotC12 >= 0.0) and (rotC12 < c12Inc)):
    if rotC12 >= c12Inc:
        rotC12=rotC12-c12Inc
    else:
        rotC12=rotC12+c12Inc

c5Vals=[i+rotC5 for i in c5]
c12Vals=[i+rotC12 for i in c12]
diffs=[]
for i in c5Vals:
    diffs.append(min([(x-i)*(x-i)**0.5 for x in c12Vals]))

resVal=c5Vals[diffs.index(min(diffs))]

c5Split[int(rotColC5)-1] = str(resVal)
starOut.write("\t".join(c5Split))
starOut.write("\n")

starOut.close()

```

Glossary

A ₂₈₀	Absorbance at 280 nm
ASCE	Additional Strand, Conserved Glutamate
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
ATP-γ-S	Adenosine-5'-o-(3-thio-triphosphate), Tetralithium Salt
CTD	C terminal domain
CCD	Charge Coupled Device
Cryo-EM	Cryo Electron Microscopy
EPR	Electron Paramagnetic Resonance
EMSA	Electrophoretic Mobility Shift Assay
EDTA	Ethylenediaminetetraacetic Acid
FEG	Field Electron Gun
FSC	Fourier shell correlation
HTH	Helix Turn Helix
LB	Lysogeny Broth
MWCO	Molecular Weight Cut Off
SEC-MALS	Multi-angle light scattering coupled with size exclusion chromatography
NTD	N Terminal Domain
OD ₆₀₀	Optical Density at 600 nm
PAGE	Polyacrylamide Gel Electrophoresis
PEG	Polyethylene Glycol

PET	Polyethylene Terephthalate
PCR	Polymerase Chain Reaction
Kglu	Potassium Glutamate
PDB	Protein Data Bank
RELION	REgularised Likelihood OptimisatioN
RNA	Ribonucleic Acid
RNAP	RNA polymerase
RPM	Revolutions Per Minute
SDS	Sodium Dodecyl Sulfate
SGD	Stochastic Gradient Descent
SNR	Signal to Noise Ratio
SPR	Surface Plasmon Resonance
TM	Tris Magnesium
TMG	Tris Magnesium Glutamate

References

1. Dr.H. Fung. PhD Thesis University of York. 2017;
2. Hendrix RW, Smith MCM, Burns RN, Ford ME, Hatfull GF. Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. *Proc Natl Acad Sci U S A*. 1999;96(5).
3. Mushegian AR. Are there 10³¹ virus particles on earth, or more, or fewer? *Journal of Bacteriology*. 2020;202(9).
4. Microbiology by numbers. Vol. 9, *Nature Reviews Microbiology*. 2011.
5. Jalasvuori M, Koonin E v. Classification of prokaryotic genetic replicators: Between selfishness and altruism. *Ann N Y Acad Sci*. 2015;1341(1).
6. Koonin E v., Starokadomskyy P. Are viruses alive? The replicator paradigm sheds decisive light on an old but misguided question. *Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences*. 2016;59.
7. Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, et al. The 1.2-megabase genome sequence of Mimivirus. *Science* (1979). 2004;306(5700).
8. Legendre M, Bartoli J, Shmakova L, Jeudy S, Labadie K, Adrait A, et al. Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proc Natl Acad Sci U S A*. 2014;111(11).
9. Mankertz A, Domingo M, Folch JM, Lecann P, Jestin A, Segalés J, et al. Characterisation of PCV-2 isolates from Spain, Germany and France. *Virus Research*. 2000;66(1).
10. Fiers W, Contreras R, Duerinck F, Haegeman G, Iserentant D, Merregaert J, et al. Complete nucleotide sequence of bacteriophage MS2 RNA: Primary and secondary structure of the replicase gene. *Nature*. 1976;260(5551).
11. Finsterbusch T, Mankertz A. Porcine circoviruses-Small but powerful. Vol. 143, *Virus Research*. 2009.
12. Bândea CI. A new theory on the origin and the nature of viruses. *Journal of Theoretical Biology*. 1983;105(4).

13. Durzyńska J, Goździcka-Józefiak A. Viruses and cells intertwined since the dawn of evolution Emerging viruses. Vol. 12, Virology Journal. 2015.
14. Bishop DHL. Ambisense Rna Genomes of Arenaviruses and Phleboviruses. Advances in Virus Research. 1986;31(C).
15. Kolakofsky D. A short biased history of RNA viruses. Vol. 21, RNA. 2015.
16. Li CX, Shi M, Tian JH, Lin XD, Kang YJ, Chen LJ, et al. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. Elife. 2015;2015(4).
17. Poiesz BJ, Ruscetti FW, Gazdar AF, Bunn PA, Minna JD, Gallo RC. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. Proc Natl Acad Sci U S A. 1980;77(12 II).
18. STRAUSS JH, STRAUSS EG. Overview of Viruses and Virus Infection. In: Viruses and Human Disease. 2008.
19. Payne S. Introduction to RNA Viruses. In: Viruses. 2017.
20. Koonin E v., Martin W. On the origin of genomes and cells within inorganic compartments. Trends in Genetics. 2005;21(12).
21. Koonin E v., Dolja V v., Krupovic M. Origins and evolution of viruses of eukaryotes: The ultimate modularity. Vols. 479–480, Virology. 2015.
22. Forterre P. The two ages of the RNA world, and the transition to the DNA world: A story of viruses and cells. In: Biochimie. 2005.
23. Arslan D, Legendre M, Seltzer V, Abergel C, Claverie JM. Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. Proc Natl Acad Sci U S A. 2011;108(42).
24. la Scola B, Audic S, Robert C, Jungang L, de Lamballerie X, Drancourt M, et al. A giant virus in amoebae. Vol. 299, Science. 2003.
25. Chelikani V, Ranjan T, Zade A, Shukla A, Kondabagil K. Genome Segregation and Packaging Machinery in Acanthamoeba polyphaga Mimivirus Is Reminiscent of Bacterial Apparatus. Journal of Virology. 2014;88(11).
26. Forterre P. The origin of viruses and their possible roles in major evolutionary transitions. Vol. 117, Virus Research. 2006.
27. Claverie JM. Viruses take center stage in cellular evolution. Genome Biology. 2006;7(6).

28. Twort FW. AN INVESTIGATION ON THE NATURE OF ULTRA-MICROSCOPIC VIRUSES. The Lancet. 1915;186(4814).
29. D'Herelle F. Sur un microbe invisible antagoniste des bacilles dysenteriques (An invisible microbe that is antagonistic to the dysentery bacillus). Comptes rendus Acad Sciences. 1917;165.
30. LWOFF A, GUTMANN A. Recherches sur un Bacillus megatherium lysogène. Ann Inst Pasteur (Paris). 1950;78(6).
31. KOZLOFF LM, KNOWLTON K, PUTNAM FW, EVANS EA. Biochemical studies of virus reproduction. V. The origin of bacteriophage nitrogen. J Biol Chem. 1951;188(1).
32. Putnam FW, Kozloff LM. On the origin of virus phosphorus. Science (1979). 1948;108(2806).
33. Luria SE, Delbrück M, Anderson TF. Electron Microscope Studies of Bacterial Viruses. Journal of Bacteriology. 1943;46(1).
34. Luria SE, Anderson TF. The Identification and Characterization of Bacteriophages with the Electron Microscope. Proceedings of the National Academy of Sciences. 1942;28(4).
35. Pennazio S, Roggero P. Tobacco Mosaic Virus RNA as Genetic Determinant: Genesis of a Discovery. Rivista di Biologia - Biology Forum. 2000;93(3).
36. Arber W, Dussoix D. Host specificity of DNA produced by Escherichia coli: I. Host controlled modification of bacteriophage λ . Journal of Molecular Biology. 1962;5(1).
37. Brenner S, Jacob F, Meselson M. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. Nature. 1961;190(4776).
38. Gros F, Hiatt H, Gilbert W, Kurland CG, Risebrough RW, Watson JD. Unstable ribonucleic acid revealed by pulse labelling of Escherichia coli. Nature. 1961;190(4776).
39. Suttle CA. Viruses in the sea. Vol. 437, Nature. 2005.
40. Fuhrman JA, Noble RT. Viruses and protists cause similar bacterial mortality in coastal seawater. Limnology and Oceanography. 1995;40(7).
41. Fuhrman JA. Marine viruses and their biogeochemical and ecological effects. Vol. 399, Nature. 1999.
42. Rohwer F. Global phage diversity. Vol. 113, Cell. 2003.

43. Ackermann HW. Tailed bacteriophages: the order caudovirales. Vol. 51, *Advances in virus research*. 1998.
44. Maniloff J, Ackermann HW. Taxonomy of bacterial viruses: Establishment of tailed virus genera and the order Caudovirales. *Archives of Virology*. 1998;143(10).
45. Lefkowitz EJ, Dempsey DM, Hendrickson RC, Orton RJ, Siddell SG, Smith DB. Virus taxonomy: The database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Research*. 2018;46(D1).
46. Ackermann HW. Frequency of morphological phage descriptions in the year 2000. Vol. 146, *Archives of Virology*. 2001.
47. Dellas N, Snyder JC, Bolduc B, Young MJ. Archaeal viruses: Diversity, replication, and structure. *Annual Review of Virology*. 2014;1(1).
48. Krupovic M, Quemin ERJ, Bamford DH, Forterre P, Prangishvili D. Unification of the Globally Distributed Spindle-Shaped Viruses of the Archaea. *Journal of Virology*. 2014;88(4).
49. Erdmann S, Chen B, Huang X, Deng L, Liu C, Shah SA, et al. A novel single-tailed fusiform *Sulfolobus* virus STSV2 infecting model *Sulfolobus* species. *Extremophiles*. 2014;18(1).
50. Attar N. Archaeal biology: Archaeal virus escapology. Vol. 14, *Nature Reviews Microbiology*. 2016.
51. Duckworth DH. History and basic properties of bacterial viruses. In: *Phage Ecology* . 1987.
52. Molineux IJ, Panja D. Popping the cork: Mechanisms of phage genome ejection. Vol. 11, *Nature Reviews Microbiology*. 2013.
53. Letellier L, Boulanger P, Plançon L, Jacquot P, Santamaria M. Main features on tailed phage, host recognition and DNA uptake. Vol. 9, *Frontiers in Bioscience*. 2004.
54. Mikoulinskaia G v., Odinokova I v., Zimin AA, Lysanskaya VY, Feofanov SA, Stepnaya OA. Identification and characterization of the metal ion-dependent l-alanoyl-d-glutamate peptidase encoded by bacteriophage T5. *FEBS Journal*. 2009;276(24).
55. Waseh S, Hanifi-Moghaddam P, Coleman R, Masotti M, Ryan S, Foss M, et al. Orally administered P22 phage tailspike protein reduces *Salmonella* colonization in chickens: Prospects of a novel therapy against bacterial infections. *PLoS ONE*. 2010;5(11).

56. Ackermann HW. 5500 Phages examined in the electron microscope. Vol. 152, Archives of Virology. 2007.
57. Cimolai N. Molecular and Diagnostic Procedures in Mycoplasmology, Volume II: Diagnostic Procedures. Clinical Infectious Diseases. 1996;23(3).
58. Tu AHT, Voelker LRL, Shen X, Dybvig K. Complete nucleotide sequence of the mycoplasma virus P1 genome. Plasmid. 2001;45(2).
59. Abedon ST. The murky origin of snow white and her t-even dwarfs. Vol. 155, Genetics. 2000.
60. Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Rüger W. Bacteriophage T4 Genome. Microbiology and Molecular Biology Reviews. 2003;67(1).
61. Yang H, Ma Y, Wang Y, Yang H, Shen W, Chen X. Transcription regulation mechanisms of bacteriophages: Recent advances and future prospects. Vol. 5, Bioengineered Bugs. 2014.
62. Łobocka MB, Rose DJ, Plunkett G, Rusin M, Samojedny A, Lehnerr H, et al. Genome of bacteriophage P1. Journal of Bacteriology. 2004;186(21).
63. Fry BA. Lambda phage and for the establishment of lysogeny. Journal of General Microbiology. 1959;21(3).
64. Stewart FM, Levin BR. The population biology of bacterial viruses: Why be temperate. Theoretical Population Biology. 1984;26(1).
65. Hobbs Z, Abedon ST. Diversity of phage infection types and associated terminology: the problem with "Lytic or lysogenic." Vol. 363, FEMS microbiology letters. 2016.
66. Court DL, Oppenheim AB, Adhya SL. A new look at bacteriophage λ genetic networks. Vol. 189, Journal of Bacteriology. 2007.
67. Fuller MT, King J. Purification of the coat and scaffolding proteins from procapsids of bacteriophage P22. Virology. 1981;112(2).
68. Chen DH, Baker ML, Hryc CF, DiMaio F, Jakana J, Wu W, et al. Structural basis for scaffolding-mediated assembly and maturation of a dsDNA virus. Proc Natl Acad Sci U S A. 2011;108(4).
69. Murialdo H, Becker A. Head morphogenesis of complex double-stranded deoxyribonucleic acid bacteriophages. Vol. 42, Microbiological Reviews. 1978.
70. Earnshaw WC, Casjens SR. DNA packaging by the double-stranded DNA bacteriophages. Vol. 21, Cell. 1980.

71. Hendrix RW. Bacteriophage DNA packaging: RNA gears in a DNA transport machine. Vol. 94, Cell. 1998.
72. Mao H, Saha M, Reyes-Aldrete E, Sherman MB, Woodson M, Atz R, et al. Structural and Molecular Basis for Coordination in a Viral DNA Packaging Motor. Cell Reports. 2016;14(8).
73. Dai L, Singh D, Lu S, Kottadiel VI, Vafabakhsh R, Mahalingam M, et al. A viral genome packaging ring-ATPase is a flexibly coordinated pentamer. Nature Communications. 2021;12(1).
74. Migliori AD, Keller N, Alam TI, Mahalingam M, Rao VB, Arya G, et al. Evidence for an electrostatic mechanism of force generation by the bacteriophage T4 DNA packaging motor. Nature Communications. 2014;5.
75. Daudén MI, Martín-Benito J, Sánchez-Ferrero JC, Pulido-Cid M, Valpuesta JM, Carrascosa JL. Large terminase conformational change induced by connector binding in bacteriophage T7. Journal of Biological Chemistry. 2013;288(23).
76. Reyes-Aldrete E, Dill EA, Bussetta C, Szymanski MR, Diemer G, Maindola P, et al. Biochemical and biophysical characterization of the dsdna packaging motor from the lactococcus lactis bacteriophage asccphi28. Viruses. 2021;13(1).
77. Rossmann MG, Mesyanzhinov V v., Arisaka F, Leiman PG. The bacteriophage T4 DNA injection machine. Vol. 14, Current Opinion in Structural Biology. 2004.
78. Lander GC, Khayat R, Li R, Prevelige PE, Potter CS, Carragher B, et al. The P22 Tail Machine at Subnanometer Resolution Reveals the Architecture of an Infection Conduit. Structure. 2009;17(6).
79. Camacho A, Jiménez F, Viñuela E, Salas M. Order of assembly of the lower collar and the tail proteins of Bacillus subtilis bacteriophage phi 29. Journal of Virology. 1979;29(2).
80. CASPAR DL, KLUG A. Physical principles in the construction of regular viruses. Cold Spring Harb Symp Quant Biol. 1962;27.
81. Suhanovsky MM, Teschke CM. Nature's favorite building block: Deciphering folding and capsid assembly of proteins with the HK97-fold. Vols. 479–480, Virology. 2015.
82. Lander GC, Evilevitch A, Jeembaeva M, Potter CS, Carragher B, Johnson JE. Bacteriophage Lambda Stabilization by Auxiliary Protein gpD: Timing, Location, and Mechanism of Attachment Determined by Cryo-EM. Structure. 2008;16(9).

83. Parent KN, Gilcrease EB, Casjens SR, Baker TS. Structural evolution of the P22-like phages: Comparison of Sf6 and P22 procapsid and virion architectures. *Virology*. 2012;427(2).
84. White HE, Sherman MB, Brasilès S, Jacquet E, Seavers P, Tavares P, et al. Capsid Structure and Its Stability at the Late Stages of Bacteriophage SPP1 Assembly. *Journal of Virology*. 2012;86(12).
85. Helgstrand C, Wikoff WR, Duda RL, Hendrix RW, Johnson JE, Liljas L. The Refined Structure of a Protein Catenane: The HK97 Bacteriophage Capsid at 3.44 Å Resolution. *Journal of Molecular Biology*. 2003;334(5).
86. Tao Y, Olson NH, Xu W, Anderson DL, Rossmann MG, Baker TS. Assembly of a tailed bacterial virus and its genome release studied in three dimensions. *Cell*. 1998;95(3).
87. Fokine A, Chipman PR, Leiman PG, Mesyanzhinov V v., Rao VB, Rossmann MG. Molecular architecture of the prolate head of bacteriophage T4. *Proc Natl Acad Sci U S A*. 2004;101(16).
88. Morais MC, Choi KH, Koti JS, Chipman PR, Anderson DL, Rossmann MG. Conservation of the capsid structure in tailed dsDNA bacteriophages: The pseudoatomic structure of ϕ 29. *Molecular Cell*. 2005;18(2).
89. Duda RL, Teschke CM. The amazing HK97 fold: versatile results of modest differences. Vol. 36, *Current Opinion in Virology*. 2019.
90. Duda RL, Martincic K, Hendrix RW. Genetic basis of bacteriophage HK97 prohead assembly. *Journal of Molecular Biology*. 1995;247(4).
91. Conway JF, Duda RL, Cheng N, Hendrix RW, Steven AC. Proteolytic and conformational control of virus capsid maturation: The bacteriophage HK97 system. *Journal of Molecular Biology*. 1995;253(1).
92. Oh B, Moyer CL, Hendrix RW, Duda RL. The delta domain of the HK97 major capsid protein is essential for assembly. *Virology*. 2014;456–457(1).
93. Wikoff WR, Conway JF, Tang J, Lee KK, Gan L, Cheng N, et al. Time-resolved molecular dynamics of bacteriophage HK97 capsid maturation interpreted by electron cryo-microscopy and X-ray crystallography. *Journal of Structural Biology*. 2006;153(3).
94. Gan L, Speir JA, Conway JF, Lander G, Cheng N, Firek BA, et al. Capsid Conformational Sampling in HK97 Maturation Visualized by X-Ray Crystallography and Cryo-EM. *Structure*. 2006;14(11).

95. Wikoff WR, Liljas L, Duda RL, Tsuruta H, Hendrix RW, Johnson JE. Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* (1979). 2000;289(5487).
96. Tso DJ, Peebles CL, Maurer JB, Duda RL, Hendrix RW. On the catalytic mechanism of bacteriophage HK97 capsid crosslinking. *Virology*. 2017;506.
97. Ross PD, Conway JF, Cheng N, Dierkes L, Firek BA, Hendrix RW, et al. A Free Energy Cascade with Locks Drives Assembly and Maturation of Bacteriophage HK97 Capsid. *Journal of Molecular Biology*. 2006;364(3).
98. Black LW. DNA packaging in dsDNA bacteriophages. Vol. 43, *Annual Review of Microbiology*. 1989.
99. Adams MB, Hayden M, Casjens S. On the sequential packaging of bacteriophage P22 DNA. *Journal of Virology*. 1983;46(2).
100. Guo P, Peterson C, Anderson D. Prohead and DNA-gp3-dependent ATPase activity of the DNA packaging protein gp16 of bacteriophage ϕ 29. *Journal of Molecular Biology*. 1987;197(2).
101. Reid RJD, Bodley JW, Anderson D. Characterization of the prohead-pRNA interaction of bacteriophage ϕ 29. *Journal of Biological Chemistry*. 1994;269(7).
102. Ding F, Lu C, Zhao W, Rajashankar KR, Anderson DL, Jardine PJ, et al. Structure and assembly of the essential RNA ring component of a viral DNA packaging motor. *Proc Natl Acad Sci U S A*. 2011;108(18).
103. Zhao W, Jardine PJ, Grimes S. An RNA Domain Imparts Specificity and Selectivity to a Viral DNA Packaging Motor. *Journal of Virology*. 2015;89(24).
104. Peixuan G, Erickson S, Anderson D. A small viral RNA is required for in vitro packaging of bacteriophage ϕ 29 DNA. *Science* (1979). 1987;236(4802).
105. Lin H, Black LW. DNA requirements in vivo for phage T4 packaging. *Virology*. 1998;242(1).
106. Sternberg N, Coulby J. Cleavage of the bacteriophage P1 packaging site (*pac*) is regulated by adenine methylation. *Proc Natl Acad Sci U S A*. 1990;87(20).
107. Tavares P, Lurz R, Stiege A, Rückert B, Trautner TA. Sequential headful packaging and fate of the cleaved DNA ends in bacteriophage SPP1. *Journal of Molecular Biology*. 1996;264(5).
108. Tye BK, Huberman JA, Botstein D. Non-random circular permutation of phage P22 DNA. *Journal of Molecular Biology*. 1974;85(4).

109. Chai S, Lurz R, Alonso JC. The small subunit of the terminase enzyme of *Bacillus subtilis* bacteriophage SPP1 forms a specialized nucleoprotein complex with the packaging initiation region. *Journal of Molecular Biology*. 1995;252(4).
110. Jackson EN, Jackson DA, Deans RJ. EcoRI analysis of bacteriophage P22 DNA packaging. *Journal of Molecular Biology*. 1978;118(3).
111. Oliveira L, Tavares P, Alonso JC. Headful DNA packaging: Bacteriophage SPP1 as a model system. Vol. 173, *Virus Research*. 2013.
112. Juhala RJ, Ford ME, Duda RL, Youlton A, Hatfull GF, Hendrix RW. Genomic sequences of bacteriophages HK97 and HK022: Pervasive genetic mosaicism in the lambdoid bacteriophages. *Journal of Molecular Biology*. 2000;299(1).
113. Ziermann R, Calendar R. Characterization of the cos sites of bacteriophages P2 and P4. *Gene*. 1990;96(1).
114. Feiss M, Catalano CE. Bacteriophage Lambda Terminase and the Mechanism of Viral DNA Packaging. In: *Viral Genome Packaging Machines: Genetics, Structure, and Mechanism*. 2007.
115. Hohn B. DNA sequences necessary for packaging of bacteriophage λ DNA. *Proc Natl Acad Sci U S A*. 1983;80(24 I).
116. Feiss M, Widner W, Miller G, Johnson G, Christiansen S. Structure of the bacteriophage lambda cohesive end site: location of the sites of terminase binding (cosB) and nicking (cosN). *Gene*. 1983;24(2–3).
117. Feiss M, Kobayashi I, Widner W. Separate sites for binding and nicking of bacteriophage λ DNA by terminase. *Proc Natl Acad Sci U S A*. 1983;80(4 I).
118. Shinder G, Gold M. The Nul subunit of bacteriophage lambda terminase binds to specific sites in cos DNA. *Journal of Virology*. 1988;62(2).
119. Oh CS, Sippy J, Charbonneau B, Hutchinson JC, Mejia-Romero OE, Barton M, et al. DNA topology and the initiation of virus DNA packaging. *PLoS ONE*. 2016;11(5).
120. Cue D, Feiss M. A site required for termination of packaging of the phage λ chromosome. *Proc Natl Acad Sci U S A*. 1993;90(20).
121. Feiss M, Fisher RA, Crayton MA, Egner C. Packaging of the bacteriophage λ chromosome: Effect of chromosome length. *Virology*. 1977;77(1).
122. Agirrezabala X, Martín-Benito J, Valle M, González JM, Valencia A, Valpuesta JM, et al. Structure of the connector of bacteriophage T7 at 8 Å resolution: Structural

- homologies of a basic component of a DNA translocating machinery. *Journal of Molecular Biology*. 2005;347(5).
123. Valpuesta J, Fujisawa H, Marco S, Carazo J, Carrascosa JL. Three-dimensional structure of T3 connector purified from overexpressing bacteria. *Journal of Molecular Biology*. 1992;224(1).
 124. Doan DNP, Dokland T. The gpQ portal protein of bacteriophage P2 forms dodecameric connectors in crystals. *Journal of Structural Biology*. 2007;157(2).
 125. Tang J, Lander GC, Olia A, Li R, Casjens S, Prevelige P, et al. Peering down the barrel of a bacteriophage portal: The genome packaging and release valve in P22. *Structure*. 2011;19(4).
 126. Dröge A, Santos MA, Stiege AC, Alonso JC, Lurz R, Trautner TA, et al. Shape and DNA packaging activity of bacteriophage SPP1 procapsid: Protein components and interactions during assembly. *Journal of Molecular Biology*. 2000;296(1).
 127. Guo P, Erickso S, Xu W, Olson N, Baker TS, Anderson D. Regulation of the phage ϕ 29 prohead shape and size by the portal vertex. *Virology*. 1991;183(1).
 128. Lokareddy RK, Sankhala RS, Roy A, Afonine P v., Motwani T, Teschke CM, et al. Portal protein functions akin to a DNA-sensor that couples genome-packaging to icosahedral capsid maturation. *Nature Communications*. 2017;8.
 129. David Hou C-F, Swanson NA, Li F, Yang R, Lokareddy RK, Cingolani G. Cryo-EM Structure of a Kinetically Trapped Dodecameric Portal Protein from the *Pseudomonas*-phage PaP3. *Journal of Molecular Biology* [Internet]. 2022 May 15 [cited 2022 Mar 21];434(9):167537. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0022283622001115>
 130. Fang Q, Tang WC, Tao P, Mahalingam M, Fokine A, Rossmann MG, et al. Structural morphing in a symmetry-mismatched viral vertex. *Nature Communications*. 2020;11(1).
 131. Hrebík D, Štveráková D, Škubník K, Füzik T, Pantůček R, Plevka P. Structure and genome ejection mechanism of *Staphylococcus aureus* phage P68. *Science Advances*. 2019;5(10).
 132. Bayfield OW, Klimuk E, Winkler DC, Hesketh EL, Chechik M, Cheng N, et al. Cryo-EM structure and in vitro DNA packaging of a thermophilic virus with supersized T=7 capsids. *Proc Natl Acad Sci U S A*. 2019;116(9).

133. Xu J, Wang D, Gui M, Xiang Y. Structural assembly of the tailed bacteriophage ϕ 29. *Nature Communications*. 2019;10(1).
134. Motwani T, Lokareddy RK, Dunbar CA, Cortines JR, Jarrold MF, Cingolani G, et al. A viral scaffolding protein triggers portal ring oligomerization and incorporation during procapsid assembly. *Science Advances*. 2017;3(7).
135. Bedwell GJ, Prevelige PE. Targeted mutagenesis of the P22 portal protein reveals the mechanism of signal transmission during DNA packaging. *Virology*. 2017;505.
136. Wieczorek DJ, Didion L, Feiss M. Alterations of the portal protein, gpB, of bacteriophage λ suppress mutations in *cosQ*, the site required for termination of DNA packaging. *Genetics*. 2002;161(1).
137. Chen W, Xiao H, Wang X, Song S, Han Z, Li X, et al. Structural changes of a bacteriophage upon DNA packaging and maturation. Vol. 11, *Protein and Cell*. 2020.
138. Dedeo CL, Cingolani G, Teschke CM. Portal Protein: The Orchestrator of Capsid Assembly for the dsDNA Tailed Bacteriophages and Herpesviruses. *Annual Review of Virology*. 2019;6.
139. Olia AS, Prevelige PE, Johnson JE, Cingolani G. Three-dimensional structure of a viral genome-delivery portal vertex. *Nature Structural and Molecular Biology*. 2011;18(5).
140. Bayfield OW, Steven AC, Antson AA. Cryo-EM structure in situ reveals a molecular switch that safeguards virus against genome loss. *Elife*. 2020;9.
141. Chaban Y, Lurz R, Brasilès S, Cornilleau C, Karreman M, Zinn-Justin S, et al. Structural rearrangements in the phage head-to-tail interface during assembly and infection. *Proc Natl Acad Sci U S A*. 2015;112(22).
142. Fuller DN, Raymer DM, Kottadiel VI, Rao VB, Smith DE. Single phage T4 DNA packaging motors exhibit large force generation, high velocity, and dynamic variability. *Proc Natl Acad Sci U S A*. 2007;104(43).
143. Hamada K, Fujisawa H, Minagawa T. A defined in vitro system for packaging of bacteriophage T3 DNA. *Virology*. 1986;151(1).
144. Hamada K, Fujisawa H, Minagawa T. Overproduction and purification of the products of bacteriophage T3 genes 18 and 19, two genes involved in DNA packaging. *Virology*. 1986;151(1).

145. Rao VB, Black LW. 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. *Journal of Molecular Biology*. 1988;200(3).
146. Casjens S, Huang WM, Hayden M, Parr R. Initiation of bacteriophage P22 DNA packaging series. Analysis of a mutant that alters the DNA target specificity of the packaging apparatus. *Journal of Molecular Biology*. 1987;194(3).
147. Al-Zahrani AS, Kondabagil K, Gao S, Kelly N, Ghosh-Kumar M, Rao VB. The small terminase, gp16, of bacteriophage T4 is a regulator of the DNA packaging motor. *Journal of Biological Chemistry*. 2009;284(36).
148. Baumann RG, Black LW. Isolation and characterization of T4 bacteriophage gp17 terminase, a large subunit multimer with enhanced ATPase activity. *Journal of Biological Chemistry*. 2003;278(7).
149. Roy A, Bhardwaj A, Datta P, Lander GC, Cingolani G. Small terminase couples viral DNA binding to genome-packaging ATPase activity. *Structure*. 2012;20(8).
150. Ebel-Tsipis J, Botstein D, Fox MS. Generalized transduction by phage P22 in *Salmonella typhimurium*. I. Molecular origin of transducing DNA. *Journal of Molecular Biology*. 1972;71(2).
151. Hayes JA, Hilbert BJ, Gaubitz C, Stone NP, Kelch BA, Cameron CE. A thermophilic phage uses a small terminase protein with a fixed helix-turn-helix geometry. *Journal of Biological Chemistry*. 2020;295(12).
152. Büttner CR, Chechik M, Ortiz-Lombardía M, Smits C, Ebong IO, Chechik V, et al. Structural basis for DNA recognition and loading into a viral packaging motor. *Proc Natl Acad Sci U S A*. 2012;109(3).
153. Suna S, Gao S, Kondabagil K, Xiang Y, Rossmann MG, Rao VB. Structure and function of the small terminase component of the DNA packaging machine in T4-like bacteriophages. *Proc Natl Acad Sci U S A*. 2012;109(3).
154. Roy A, Bhardwaj A, Cingolani G. Crystallization of the nonameric small terminase subunit of bacteriophage P22. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*. 2011;67(1).
155. Maria Chechik. Manuscript in progress. 2022;
156. Morita M, Tasaka M, Fujisawa H. Dna packaging atpase of bacteriophage t3. *Virology*. 1993;193(2).

157. Chemla YR, Aathavan K, Michaelis J, Grimes S, Jardine PJ, Anderson DL, et al. Mechanism of force generation of a viral DNA packaging motor. *Cell*. 2005;122(5).
158. Rickgauer JP, Fuller DN, Grimes S, Jardine PJ, Anderson DL, Smith DE. Portal motor velocity and internal force resisting viral DNA packaging in bacteriophage ϕ 29. *Biophysical Journal*. 2008;94(1).
159. Fuller DN, Raymer DM, Rickgauer JP, Robertson RM, Catalano CE, Anderson DL, et al. Measurements of Single DNA Molecule Packaging Dynamics in Bacteriophage λ Reveal High Forces, High Motor Processivity, and Capsid Transformations. *Journal of Molecular Biology*. 2007;373(5).
160. Hilbert BJ, Hayes JA, Stone NP, Duffy CM, Sankaran B, Kelch BA. Structure and mechanism of the ATPase that powers viral genome packaging. *Proc Natl Acad Sci U S A*. 2015;112(29).
161. Lyubimov AY, Strycharska M, Berger JM. The nuts and bolts of ring-translocase structure and mechanism. Vol. 21, *Current Opinion in Structural Biology*. 2011.
162. Iyer LM, Leipe DD, Koonin E v., Aravind L. Evolutionary history and higher order classification of AAA+ ATPases. In: *Journal of Structural Biology*. 2004.
163. delToro D, Ortiz D, Ordyan M, Pajak J, Sippy J, Catala A, et al. Functional Dissection of a Viral DNA Packaging Machine's Walker B Motif. *Journal of Molecular Biology*. 2019;431(22).
164. Ogura T, Whiteheart SW, Wilkinson AJ. Conserved arginine residues implicated in ATP hydrolysis, nucleotide-sensing, and inter-subunit interactions in AAA and AAA+ ATPases. In: *Journal of Structural Biology*. 2004.
165. Iyer LM, Makarova KS, Koonin E v., Aravind L. Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: Implications for the origins of chromosome segregation, cell division and viral capsid packaging. Vol. 32, *Nucleic Acids Research*. 2004.
166. Leipe DD, Koonin E v., Aravind L. Evolution and classification of P-loop kinases and related proteins. *Journal of Molecular Biology*. 2003;333(4).
167. Walker JE, Saraste M, Runswick MJ, Gay NJ. 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*. 1982;1(8).

168. Kenniston JA, Baker TA, Fernandez JM, Sauer RT. Linkage between ATP consumption and mechanical unfolding during the protein processing reactions of an AAA+ degradation machine. *Cell*. 2003;114(4).
169. Hanson PI, Whiteheart SW. AAA+ proteins: Have engine, will work. Vol. 6, *Nature Reviews Molecular Cell Biology*. 2005.
170. Moffitt JR, Chemla YR, Aathavan K, Grimes S, Jardine PJ, Anderson DL, et al. Intersubunit coordination in a homomeric ring ATPase. *Nature*. 2009;457(7228).
171. Fuller DN, Gemmen GJ, Rickgauer JP, Dupont A, Millin R, Recouvreur P, et al. A general method for manipulating DNA sequences from any organism with optical tweezers. *Nucleic Acids Research*. 2006;34(2).
172. Moffitt JR, Chemla YR, Smith SB, Bustamante C. Recent advances in optical tweezers. Vol. 77, *Annual Review of Biochemistry*. 2008.
173. Ashkin A, Dziedzic JM, Bjorkholm JE, Chu S. Observation of a single-beam gradient force optical trap for dielectric particles. In: *Optical Angular Momentum*. 2016.
174. Liu S, Chistol G, Hetherington CL, Tafuya S, Aathavan K, Schnitzbauer J, et al. A viral packaging motor varies its DNA rotation and step size to preserve subunit coordination as the capsid fills. *Cell*. 2014;157(3).
175. Zhang Z, Kottadiel VI, Vafabakhsh R, Dai L, Chemla YR, Ha T, et al. A promiscuous DNA packaging machine from bacteriophage T4. *PLoS Biology*. 2011;9(2).
176. Vafabakhsh R, Kondabagil K, Earnest T, Lee KS, Zhang Z, Dai L, et al. Single-molecule packaging initiation in real time by a viral DNA packaging machine from bacteriophage T4. *Proc Natl Acad Sci U S A*. 2014;111(42).
177. Woodson M, Pajak J, Mahler BP, Zhao W, Zhang W, Arya G, et al. A viral genome packaging motor transitions between cyclic and helical symmetry to translocate dsDNA. *Science Advances*. 2021;7(19).
178. Koti JS, Morais MC, Rajagopal R, Owen BAL, McMurray CT, Anderson DL. DNA Packaging Motor Assembly Intermediate of Bacteriophage ϕ 29. *Journal of Molecular Biology*. 2008;381(5).
179. Morais MC, Koti JS, Bowman VD, Reyes-Aldrete E, Anderson DL, Rossmann MG. Defining Molecular and Domain Boundaries in the Bacteriophage ϕ 29 DNA Packaging Motor. *Structure*. 2008;16(8).

180. Pajak J, Dill E, Reyes-Aldrete E, White MA, Kelch BA, Jardine PJ, et al. Atomistic basis of force generation, translocation, and coordination in a viral genome packaging motor. *Nucleic Acids Research*. 2021;49(11).
181. Sun S, Kondabagil K, Draper B, Alam TI, Bowman VD, Zhang Z, et al. The Structure of the Phage T4 DNA Packaging Motor Suggests a Mechanism Dependent on Electrostatic Forces. *Cell*. 2008;135(7).
182. Hegde S, Padilla-Sanchez V, Draper B, Rao VB. Portal-Large Terminase Interactions of the Bacteriophage T4 DNA Packaging Machine Implicate a Molecular Lever Mechanism for Coupling ATPase to DNA Translocation. *Journal of Virology*. 2012;86(8).
183. Ortiz D, DelToro D, Ordyan M, Pajak J, Sippy J, Catala A, et al. Evidence that a catalytic glutamate and an “Arginine Toggle” act in concert to mediate ATP hydrolysis and mechanochemical coupling in a viral DNA packaging motor. *Nucleic Acids Research*. 2019;47(3).
184. Zhao H, Christensen TE, Kamau YN, Tang L. Structures of the phage Sf6 large terminase provide new insights into DNA translocation and cleavage. *Proc Natl Acad Sci U S A*. 2013;110(20).
185. Pajak J, Atz R, Hilbert BJ, Morais MC, Kelch BA, Jardine PJ, et al. Viral packaging ATPases utilize a glutamate switch to couple ATPase activity and DNA translocation. *Proc Natl Acad Sci U S A*. 2021;118(17).
186. Zhao H, Lin Z, Lynn AY, Varnado B, Beutler JA, Murelli RP, et al. Two distinct modes of metal ion binding in the nuclease active site of a viral DNA-packaging terminase: Insight into the two-metal-ion catalytic mechanism. *Nucleic Acids Research*. 2015;43(22).
187. Smits C, Chechik M, Kovalevskiy O v., Shevtsov MB, Foster AW, Alonso JC, et al. Structural basis for the nuclease activity of a bacteriophage large terminase. *EMBO Reports*. 2009;10(6).
188. Xu RG, Jenkins HT, Chechik M, Blagova E v., Lopatina A, Klimuk E, et al. Viral genome packaging terminase cleaves DNA using the canonical RuvC-like two-metal catalysis mechanism. *Nucleic Acids Research*. 2017;45(6).
189. Roy A, Cingolani G. Structure of P22 headful packaging nuclease. *Journal of Biological Chemistry*. 2012;287(33).

190. Nadal M, Mas PJ, Blanco AG, Arnan C, Solà M, Hart DJ, et al. Structure and inhibition of herpesvirus DNA packaging terminase nuclease domain. *Proc Natl Acad Sci U S A*. 2010;107(37).
191. Glaeser RM, Taylor KA. Radiation damage relative to transmission electron microscopy of biological specimens at low temperature: a review. *Journal of Microscopy*. 1978;112(1).
192. Brenner S, Horne RW. A negative staining method for high resolution electron microscopy of viruses. *BBA - Biochimica et Biophysica Acta*. 1959;34(C).
193. Ohi M, Li Y, Cheng Y, Walz T. Negative staining and image classification - Powerful tools in modern electron microscopy. *Biological Procedures Online*. 2004;6(1).
194. de Carlo S, Harris JR. Negative staining and cryo-negative staining of macromolecules and viruses for TEM. Vol. 42, *Micron*. 2011.
195. Scarff CA, Fuller MJG, Thompson RF, Iadanza MG. Variations on negative stain electron microscopy methods: Tools for tackling challenging systems. *Journal of Visualized Experiments*. 2018;2018(132).
196. Burgess SA, Walker ML, Thirumurugan K, Trinick J, Knight PJ. Use of negative stain and single-particle image processing to explore dynamic properties of flexible macromolecules. *Journal of Structural Biology*. 2004;147(3).
197. McDowell AW, Chang J -J, Freeman R, Lepault J, Walter CA, Dubochet J. Electron microscopy of frozen hydrated sections of vitreous ice and vitrified biological samples. *Journal of Microscopy*. 1983;131(1).
198. Tivol WF, Briegel A, Jensen GJ. An improved cryogen for plunge freezing. *Microscopy and Microanalysis*. 2008;14(5).
199. Schultz P. Cryo-electron microscopy of vitrified specimens. *Quarterly Reviews of Biophysics*. 1988;21(2).
200. Sgro GG, Costa TRD. Cryo-EM Grid Preparation of Membrane Protein Samples for Single Particle Analysis. Vol. 5, *Frontiers in Molecular Biosciences*. 2018.
201. Mills DJ. Setting up and operating a cryo-EM laboratory. *Quarterly Reviews of Biophysics*. 2021.
202. Cheng Y, Grigorieff N, Penczek PA, Walz T. A primer to single-particle cryo-electron microscopy. Vol. 161, *Cell*. 2015.

203. Li X, Mooney P, Zheng S, Booth CR, Braunfeld MB, Gubbens S, et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods*. 2013;10(6).
204. Campbell MG, Cheng A, Brilot AF, Moeller A, Lyumkis D, Veesler D, et al. Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. *Structure*. 2012;20(11).
205. Scheres SH w. Beam-induced motion correction for sub-megadalton cryo-EM particles. *Elife*. 2014;3.
206. Zhang Y, Lu PH, Rotunno E, Troiani F, van Schayck JP, Tavabi AH, et al. Single-particle cryo-EM: Alternative schemes to improve dose efficiency. In: *Journal of Synchrotron Radiation*. 2021.
207. Zhang K, Li S, Kappel K, Pintilie G, Su Z, Mou TC, et al. Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 Å resolution. *Nature Communications*. 2019;10(1).
208. Langmore JP, Smith MF. Quantitative energy-filtered electron microscopy of biological molecules in ice. *Ultramicroscopy*. 1992;46(1–4).
209. Danev R, Nagayama K. Optimizing the phase shift and the cut-on periodicity of phase plates for TEM. *Ultramicroscopy*. 2011;111(8).
210. Tsuno K, Kaneyama T, Honda T, Tsuda K, Terauchi M, Tanaka M. Design and testing of Omega mode imaging energy filters at 200 kV. *Journal of Electron Microscopy*. 1997;46(5).
211. Krivanek OL, Friedman SL, Gubbens AJ, Kraus B. An imaging filter for biological applications. *Ultramicroscopy*. 1995;59(1–4).
212. Kim LY, Rice WJ, Eng ET, Kopylov M, Cheng A, Raczkowski AM, et al. Benchmarking cryo-EM single particle analysis workflow. *Frontiers in Molecular Biosciences*. 2018;5(JUN).
213. Zivanov J, Nakane T, Forsberg BO, Kimanius D, Hagen WJH, Lindahl E, et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife*. 2018;7.
214. Bepler T, Morin A, Rapp M, Brasch J, Shapiro L, Noble AJ, et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature Methods*. 2019;16(11).

215. Wang L, Shkolnisky Y, Singer A. A Fourier-based Approach for Iterative 3D Reconstruction from Cryo-EM Images. 2013 Jul 22;
216. Single-particle Cryo-EM of Biological Macromolecules. Single-particle Cryo-EM of Biological Macromolecules. 2021.
217. Scheres SHW. A bayesian view on cryo-EM structure determination. *Journal of Molecular Biology*. 2012;415(2).
218. Henderson R. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. Vol. 110, *Proceedings of the National Academy of Sciences of the United States of America*. 2013.
219. Scheres SHW. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*. 2012;180(3).
220. Rosenthal PB, Henderson R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *Journal of Molecular Biology*. 2003;333(4).
221. van Heel M, Schatz M. Fourier shell correlation threshold criteria. *Journal of Structural Biology*. 2005;151(3).
222. Beckers M, Mann D, Sachse C. Structural interpretation of cryo-EM image reconstructions. *Progress in Biophysics and Molecular Biology*. 2021;160.
223. Rohou A. Fourier shell correlation criteria for local resolution estimation. *bioRxiv*. 2020;
224. Duda RL, Hempel J, Michel H, Shabanowitz J, Hunt D, Hendrix RW. Structural transitions during bacteriophage HK97 head assembly. *Journal of Molecular Biology*. 1995;247(4).
225. Leirimo S, Harrison C, Cayley DS, Record MT, Burgess RR. Replacement of Potassium Chloride by Potassium Glutamate Dramatically Enhances Protein-DNA Interactions in Vitro. *Biochemistry*. 1987;26(8).
226. Yanofsky C, Ito J. Nonsense codons and polarity in the tryptophan operon. *Journal of Molecular Biology*. 1966;21(2).
227. Santos SB, Carvalho CM, Sillankorva S, Nicolau A, Ferreira EC, Azeredo J. The use of antibiotics to improve phage detection and enumeration by the double-layer agar technique. *BMC Microbiology*. 2009;9.
228. Weigle J. Assembly of phage lambda in vitro. *Proc Natl Acad Sci U S A*. 1966;55(6).

229. Norrander J, Kempe T, Messing J. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene*. 1983;26(1).
230. Jean NL, Rutherford TJ, Löwe J. FtsK in motion reveals its mechanism for double-stranded DNA translocation. *Proc Natl Acad Sci U S A*. 2020;117(25).
231. Pan M, Zheng Q, Yu Y, Ai H, Xie Y, Zeng X, et al. Seesaw conformations of Npl4 in the human p97 complex and the inhibitory mechanism of a disulfiram derivative. *Nature Communications*. 2021;12(1).
232. Yaginuma H, Kawai S, Tabata K v., Tomiyama K, Kakizuka A, Komatsuzaki T, et al. Diversity in ATP concentrations in a single bacterial cell population revealed by quantitative single-cell imaging. *Scientific Reports*. 2014;4.
233. Kala S, Cumby N, Sadowski PD, Hyder BZ, Kanelis V, Davidson AR, et al. HNH proteins are a widespread component of phage DNA packaging machines. *Proc Natl Acad Sci U S A*. 2014;111(16).
234. Alam TI, Draper B, Kondabagil K, Rentas FJ, Ghosh-Kumar M, Sun S, et al. The headful packaging nuclease of bacteriophage T4. *Molecular Microbiology*. 2008;69(5).
235. Eschenfeldt WH, Lucy S, Millard CS, Joachimiak A, Mark ID. A family of LIC vectors for high-throughput cloning and purification of proteins. *Methods in Molecular Biology*. 2009;498.
236. Lyumkis D. Challenges and opportunities in cryo-EM single-particle analysis. *Journal of Biological Chemistry*. 2019;294(13).
237. Russo CJ, Passmore LA. Ultrastable gold substrates for electron cryomicroscopy. *Science (1979)*. 2014;346(6215).
238. Drulyte I, Johnson RM, Hesketh EL, Hurdiss DL, Scarff CA, Porav SA, et al. Approaches to altering particle distributions in cryo-electron microscopy sample preparation. *Acta Crystallographica Section D: Structural Biology*. 2018;74(6).
239. Guo F, Jiang W. Single particle cryo-electron microscopy and 3-D reconstruction of viruses. *Methods in Molecular Biology*. 2014;1117.
240. Palovcak E, Wang F, Zheng SQ, Yu Z, Li S, Betegon M, et al. A simple and robust procedure for preparing graphene-oxide cryo-EM grids. *Journal of Structural Biology*. 2018;204(1).
241. Hesketh EL, Saunders K, Fisher C, Potze J, Stanley J, Lomonosoff GP, et al. The 3.3 Å structure of a plant geminivirus using cryo-EM. *Nature Communications*. 2018;9(1).

242. Juckles IRM. Fractionation of proteins and viruses with polyethylene glycol. *BBA - Protein Structure*. 1971;229(3).
243. D'Imprima E, Floris D, Joppe M, Sánchez R, Grininger M, Kühlbrandt W. Protein denaturation at the air-water interface and how to prevent it. *Elife*. 2019;8.
244. Wingfield P. Protein Precipitation Using Ammonium Sulfate. In: *Current Protocols in Protein Science*. 1998.
245. Han BG, Watson Z, Kang H, Pulk A, Downing KH, Cate J, et al. Long shelf-life streptavidin support-films suitable for electron microscopy of biological macromolecules. *Journal of Structural Biology*. 2016;195(2).
246. Zhou T, Teng IT, Olia AS, Cerutti G, Gorman J, Nazzari A, et al. Structure-Based Design with Tag-Based Purification and In-Process Biotinylation Enable Streamlined Development of SARS-CoV-2 Spike Molecular Probes. *Cell Reports*. 2020;33(4).
247. Liu F, Zhang JZH, Mei Y. The origin of the cooperativity in the streptavidin-biotin system: A computational investigation through molecular dynamics simulations. *Scientific Reports*. 2016;6.
248. Fang H, Jing P, Haque F, Guo P. Role of channel lysines and the “push through a one-way valve” mechanism of the viral DNA packaging motor. *Biophysical Journal*. 2012;102(1).
249. Shabala L, Bowman J, Brown J, Ross T, McMeekin T, Shabala S. Ion transport and osmotic adjustment in *Escherichia coli* in response to ionic and non-ionic osmotic. *Environmental Microbiology*. 2009;11(1).
250. Wilkinson ME, Kumar A, Casan al A. Methods for merging data sets in electron cryo-microscopy. *Acta Crystallographica Section D: Structural Biology*. 2019;75.
251. Zheng SQ, Palovcak E, Armache JP, Verba KA, Cheng Y, Agard DA. MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Vol. 14, *Nature Methods*. 2017.
252. Rohou A, Grigorieff N. CTFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology*. 2015;192(2).
253. Scheres SHW. RELION-3.1-tutorial. *Manuals*. 2013;
254. Gertsman I, Gan L, Guttman M, Lee K, Speir JA, Duda RL, et al. An unexpected twist in viral capsid maturation. *Nature*. 2009;458(7238).

255. LEDERER H, MAY RP, KJEMS JK, BAER G, HEUMANN H. Solution structure of a short DNA fragment studied by neutron scattering. *European Journal of Biochemistry*. 1986;161(1).
256. Niazi M, Florio TJ, Yang R, Lokareddy RK, Swanson NA, Gillilan RE, et al. Biophysical analysis of Pseudomonas-phage PaP3 small terminase suggests a mechanism for sequence-specific DNA-binding by lateral interdigitation. *Nucleic Acids Research*. 2020;48(20).
257. Bárdy P, Füzik T, Hrebík D, Pantůček R, Thomas Beatty J, Plevka P. Structure and mechanism of DNA delivery of a gene transfer agent. *Nature Communications*. 2020;11(1).
258. Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*. 1970;48(3).
259. Guo F, Liu Z, Vago F, Ren Y, Wu W, Wright ET, et al. Visualization of uncorrelated, tandem symmetry mismatches in the internal genome packaging apparatus of bacteriophage T7. *Proc Natl Acad Sci U S A*. 2013;110(17).
260. Tavares P, Zinn-Justin S, Orlova E v. Genome gating in tailed bacteriophage capsids. *Advances in Experimental Medicine and Biology*. 2012;726.
261. Orlova E v., Gowen B, Dröge A, Stiege A, Weise F, Lurz R, et al. Structure of a viral DNA gatekeeper at 10 Å resolution by cryo-electron microscopy. *EMBO Journal*. 2003;22(6).
262. Cardarelli L, Lam R, Tuite A, Baker LA, Sadowski PD, Radford DR, et al. The Crystal Structure of Bacteriophage HK97 gp6: Defining a Large Family of Head-Tail Connector Proteins. *Journal of Molecular Biology*. 2010;395(4).
263. Guasch A, Pous J, Ibarra B, Gomis-Rüth FX, Valpuesta JM, Sousa N, et al. Detailed architecture of a DNA translocating machine: The high-resolution structure of the bacteriophage ϕ 29 connector particle. *Journal of Molecular Biology*. 2002;315(4).
264. Yang K, Baines JD. Proline and Tyrosine Residues in Scaffold Proteins of Herpes Simplex Virus 1 Critical to the Interaction with Portal Protein and Its Incorporation into Capsids. *Journal of Virology*. 2009;83(16).
265. Stratagene. QuickChange Site-Directed Mutagenesis Kit. Kit Manual. 2011;

266. Dixit AB, Ray K, Thomas JA, Black LW. The C-terminal domain of the bacteriophage T4 terminase docks on the prohead portal clip region during DNA packaging. *Virology*. 2013;446(1–2).
267. Hilbert BJ, Hayes JA, Stone NP, Xu RG, Kelch BA. The large terminase DNA packaging motor grips DNA with its ATPase domain for cleavage by the flexible nuclease domain. *Nucleic Acids Research*. 2017;45(6).
268. Sun S, Kondabagil K, Gentz PM, Rossmann MG, Rao VB. The Structure of the ATPase that Powers DNA Packaging into Bacteriophage T4 Procapsids. *Molecular Cell*. 2007;25(6).
269. Massey TH, Mercogliano CP, Yates J, Sherratt DJ, Löwe J. Double-Stranded DNA Translocation: Structure and Mechanism of Hexameric FtsK. *Molecular Cell*. 2006;23(4).
270. Kanamaru S, Kondabagil K, Rossmann MG, Rao VB. The functional domains of bacteriophage T4 terminase. *Journal of Biological Chemistry*. 2004;279(39).
271. Orlova E v., Saibil HR. Structural analysis of macromolecular assemblies by electron microscopy. Vol. 111, *Chemical Reviews*. 2011.