

**Biophysical and structural characterisation of the
bacteriophage HK97 DNA packaging system**

Kar Ho Herman Fung

Doctor of Philosophy

University of York

Biology

September 2017

Abstract

DNA packaging is a key step in the assembly of dsDNA viruses such as tailed bacteriophages and herpes viruses, whereby empty capsids are filled with a copy of the viral genome. The task is mediated by a 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. *Pac* phages use a signal to mark only the beginning of their genome. Protein-primed replication in ϕ 29-like phages produces unit-length genome copies for packaging. For *cos* phages, the structures of assemblies that initiate, perform and terminate packaging are unknown. The structures of individual terminase proteins are also unknown.

To further elucidate the mechanisms of DNA packaging in *cos* phages, a new packaging motor was assembled *in vitro* based on *Escherichia coli* bacteriophage HK97. A minimal motor could be formed using proheads and large terminase. Structures determined by X-ray crystallography reveal that the HK97 small terminase is a nonameric assembly with circularly displayed helix-turn-helix motifs, supporting a model where *cos* DNA wraps around the protein; the presence of unique structural elements in the large terminase suggests a subtle difference in mechanism for the regulation of ATPase and endonuclease activities during packaging. An additional lysine, important for catalysis, is present in the ATPase active site. This lysine potentially substitutes for the arginine finger of *pac* and ϕ 29 homologues. Cryo-EM reconstructions of the active motor and counting of large terminase subunits tagged with GFP in a stalled motor by stepwise photobleaching show that the HK97 motor contains five large terminase subunits. The conserved architecture suggests that *cos*, *pac* and ϕ 29-like motors share a common mechanism for DNA translocation, despite their different initiation and termination behaviours.

Table of contents

Abstract	2
Table of contents	3
List of tables	9
List of figures	10
List of accompanying material.....	14
Acknowledgements	15
Declaration	16
1. Introduction	17
1.1 Viruses.....	17
1.2 Bacteriophages	18
1.3 Phage assembly	20
1.3.1 Making heads and tails.....	20
1.3.2 Functional classification of phages: cos, pac and ϕ 29-like.....	22
1.4 DNA replication	23
1.5 Prohead assembly and capsid maturation	24
1.5.1 Prohead assembly.....	24
1.5.2 Capsid maturation	25
1.6 Packaging initiation.....	28
1.7 DNA packaging.....	33
1.8 Packaging termination.....	38

1.9 Assembly of other viruses	39
1.10 Bacteriophage HK97 as a model system for the study of DNA packaging	40
2. <i>In vitro</i> reconstitution of the HK97 DNA packaging motor	42
2.1 Introduction	42
2.2 Materials and methods	42
2.2.1 Phage production and purification	42
2.2.2 Phage DNA extraction	43
2.2.3 Cloning of the HK97 small terminase, large terminase and the cos region	44
2.2.4 Overproduction and purification of HK97 small and large terminase proteins ..	44
2.2.5 Production of HK97 proheads.....	45
2.2.6 Production of HK97 GP74	46
2.2.7 DNase protection assays	47
2.3 Overproduction of HK97 small terminase and large terminase.....	47
2.4 Isolation of empty HK97 proheads	49
2.5 Reconstitution of the DNA packaging motor.....	51
2.6 Biochemical characterisation of the DNA packaging motor	53
2.7 The role of small terminase in packaging termination.....	56
2.8 An endonuclease co-factor for packaging initiation and termination	57
2.9 Discussion	59
3. Structure and function of the HK97 small terminase.....	61
3.1 Introduction	61

3.2 Materials and methods	61
3.2.1 Protein and DNA production	61
3.2.2 Size exclusion chromatography with multi-angle laser-light scattering analysis	62
3.2.3 Native mass spectrometry	62
3.2.4 Crystallisation and structure determination	63
3.2.5 Surface plasmon resonance	63
3.2.6 Nuclear magnetic resonance	64
3.2.7 Electrophoretic mobility shift assay	65
3.3 Oligomeric state of the HK97 small terminase	66
3.4 Crystal structure of the HK97 small terminase	67
3.4.1 Crystallisation and structure determination	67
3.4.2 The HK97 small terminase crystal structure	73
3.5 Biophysical analysis of the <i>Bacillus</i> phage SF6 small terminase N-terminal domain	76
3.5.1 The DNA-binding properties of the N-terminal domain by surface plasmon resonance	76
3.5.2 Identification of the DNA binding surface by NMR	78
3.6 Interactions of the HK97 small terminase with DNA	85
3.6.1 The DNA-binding properties of the HK97 small terminase	85
3.6.2 Preparation of a 95-bp DNA substrate for co-crystallisation studies	86
3.7 Discussion	87
4. Structural and biophysical analysis of the HK97 packaging motor	91

4.1 Introduction	91
4.2 Materials and Methods	91
4.2.1 Protein and DNA production	91
4.2.2 Electron microscopy.....	91
4.2.3 Subunit counting by total internal reflection fluorescence microscopy.....	92
4.3 Negative-stain and cryo-EM imaging of an active HK97 DNA packaging motor ...	95
4.4 Cryo-EM reconstructions of the HK97 prohead and DNA packaging motor.....	96
4.2.1 Cryo-EM reconstructions of the HK97 prohead	96
4.2.2 Cryo-EM reconstructions of the HK97 DNA packaging motor	99
4.5 Measurement of subunit stoichiometry for the HK97 motor by total internal reflection fluorescence microscopy.....	102
4.5.1 Photobleaching assay development.....	102
4.5.2 Routines for photobleaching data analysis.....	104
4.5.3 Photobleaching data analysis	105
4.6 Discussion	108
5. Structure and function of the HK97 large terminase	110
5.1 Introduction.....	110
5.2 Materials and methods	110
5.2.1 Protein production.....	110
5.2.2 Crystallisation and structure determination	111
5.2.3 Size exclusion chromatography with multi-angle laser-light scattering analysis	111

5.2.4 Sedimentation velocity analytical ultracentrifugation	112
5.2.5 Measurement of ATPase activity	112
5.2.6 Far-UV circular dichroism spectropolarimetry	113
5.3 Crystal structure of the HK97 large terminase	113
5.3.1 Crystallisation	113
5.3.2 Structure determination	115
5.3.3 The HK97 large terminase crystal structure	118
5.4 Biochemical and biophysical characterisation of the HK97 large terminase	124
5.5 Molecular basis of HK97 large terminase ATPase activity	125
5.6 Model of the HK97 DNA packaging motor	129
5.7 Discussion	130
6. Conclusions and future perspectives	132
Appendix A. DNA and protein sequences	135
Appendix B. Theory	139
B.1 X-ray crystallography	139
B.1.1 Diffraction theory	139
B.1.2 Anomalous scattering	140
1.11.3 Substructure solution	142
1.11.4 Phasing and hand determination	143
1.11.5 Density modification	145
B.2 Nuclear magnetic resonance spectroscopy	147

B.3 Electron microscopy	149
B.4 Total internal reflection fluorescence microscopy	152
Abbreviations	154
References	156

List of tables

Table 3.1. Crystallographic data collection and refinement statistics.....	69
Table 4.1. Event statistics.....	106
Table 5.1. Crystallographic data collection and refinement statistics.....	116
Table A.1. DNA oligonucleotides.....	135

List of figures

Figure 1.1. Viruses	17
Figure 1.2. The structural genes of phage λ	20
Figure 1.3. The assembly process of tailed <i>cos</i> and <i>pac</i> bacteriophages	21
Figure 1.4. Prohead assembly	24
Figure 1.5. Icosahedral capsids	25
Figure 1.6. The HK97 capsid	27
Figure 1.7. Packaging initiation in <i>cos</i> phage λ	29
Figure 1.8. Packaging initiation in <i>pac</i> phages	30
Figure 1.9. Crystal structures of <i>pac</i> large terminase proteins and their nuclease domains	32
Figure 1.10. The packaging ATPases of phage ϕ 29 and <i>pac</i> phages T4, Sf6 and P74-26 .	35
Figure 1.11. ATPase fold and reaction mechanism	36
Figure 1.12. Packaging outcomes for <i>cos</i> and <i>pac</i> viruses	38
Figure 1.13. Virus assembly mechanisms.....	40
Figure 1.14. The structural genes of phage HK97	41
Figure 2.1. Purification of HK97 small terminase	48
Figure 2.2. Purification of HK97 large terminase.....	49
Figure 2.3. Production of empty HK97 proheads	50
Figure 2.4. Reconstitution of the HK97 DNA packaging motor	52
Figure 2.5. Packaging in the absence of <i>cos</i>	53
Figure 2.6. The effects of salts and pH on <i>in vitro</i> packaging	54

Figure 2.7. The effects of spermidine on <i>in vitro</i> packaging	55
Figure 2.8. The speed of the HK97 DNA packaging motor	55
Figure 2.9. Packaging as a function of ATP	56
Figure 2.10. DNA packaging and processing in the presence of GP74	58
Figure 3.1 Molecular weight measurements of the HK97 small terminase	66
Figure 3.2. Crystallisation and diffraction experiment	67
Figure 3.3. Experimental phasing and density modification.....	70
Figure 3.4. Secondary structure profile and ribbon representation of a HK97 small terminase monomer	71
Figure 3.5. Asymmetric unit of the crystal structure	72
Figure 3.6. Cyclic symmetry in the HK97 small terminase	73
Figure 3.7. Comparison of small terminase subunits.....	74
Figure 3.8. Electrostatic surface potential of the HK97 small terminase oligomer	75
Figure 3.9. Surface plasmon resonance analysis of the DNA-binding properties of the SF6 small terminase N-terminal domain	77
Figure 3.10. Backbone assignment of the SF6 small terminase N-terminal domain.....	79
Figure 3.11. Amide assignment indicated on the ^1H - ^{15}N HSQC spectrum	80
Figure 3.12. Secondary structure indicators.....	81
Figure 3.13. ^1H - ^{15}N HSQC titration series of 20 bp dsDNA into the SF6 N-terminal domain.....	82
Figure 3.14. ^1H - ^{15}N HSQC titration of a 14-nt DNA hairpin into the SF6 N-terminal domain.....	83
Figure 3.15. Chemical shift perturbation in response to the DNA hairpin	84

Figure 3.16. Mapping of chemical shift perturbation	85
Figure 3.17. Electrophoretic mobility shift assay	86
Figure 3.18. Production of a 95-bp <i>cos</i> segment for crystallisation with small terminase .	87
Figure 3.19. The HK97 <i>cos</i> region.....	89
Figure 4.1. Negative-stain and cryo-images of the HK97 motor.....	95
Figure 4.2. Reconstruction of the HK97 prohead	97
Figure 4.3. Flexible fitting of the prohead structure	98
Figure 4.4. Refinement protocol for the HK97 motor	100
Figure 4.5. Asymmetric reconstruction of the HK97 motor	101
Figure 4.6. Design of the TIRFM photobleaching assay	102
Figure 4.7. Alignment of image channels using TetraSpeck microspheres.....	103
Figure 4.8. GFP fluorescence of a protein spot co-localising with DNA under constant illumination	104
Figure 4.9. GFP photobleaching step counts	106
Figure 4.10. Binomial distribution of step counts for a motor containing five large terminase subunits.....	107
Figure 5.1 Enzyme-coupled ATPase assay.....	112
Figure 5.2. Optimisation of HK97 large terminase crystals.....	114
Figure 5.3. Experimental phasing and model building	117
Figure 5.4. Secondary structure profile and ribbon representation of HK97 large terminase	119
Figure 5.5. Crystal structure of the HK97 large terminase	120
Figure 5.6. Large terminase nuclease domains	121

Figure 5.7. Large terminase active sites.....	123
Figure 5.8. Molecular weight measurements of HK97 large terminase	124
Figure 5.9. Regulation of ATPase activity.....	125
Figure 5.10. Critical residues of the HK97 large terminase ATPase domain	127
Figure 5.11. Far-UV circular dichroism spectra of wild-type HK97 large terminase and mutants	128
Figure 5.12. Multiple sequence alignment of large terminase Walker A motifs	128
Figure 5.13. Model of the HK97 DNA packaging motor	129
Figure B.1. Breaking Friedel's law	142
Figure B.2. Harker constructions for anomalous diffraction experiments.....	144
Figure B.3. Probabilistic SAD Harker construction.....	145

List of accompanying material

Custom MATLAB scripts (archived on CD attached to the back of this thesis)

Acknowledgements

I would like to thank my family for their support; Prof Fred Antson and Dr Christoph Baumann for their supervision and faith in this undertaking; Mrs Maria Chechik and Dr Huw Jenkins, for inspiration and support; Prof Maggie Smith and Prof Mark Leake for their feedback over the years in TAP meetings; my CIDCATS cohort (Hayley, Jason, Mike and Eva), Alexandra, Łukasz, Becky, Imogen, Claire, Mui, Marc, Eleni, Bas, Paul B, Amina, Adam, Tamara, Jen, Sophie, Dan W, Rosie, and many more past and present colleagues in the York Structural Biology Laboratory and Department of Biology for their company, support and encouragement. I would also like to express my gratitude to Drs Shelley Grimes, Paul Jardine, Robert Duda, James Conway and Alexis Huet, and Prof Roger Hendrix for their patience and mentorship, and to the members of their labs for their hospitality. I would like to acknowledge the Wellcome Trust for their financial support and training opportunities provided through the CIDCATS programme. Finally, I would like to dedicate this thesis to the memory of Dr Shelley Grimes and Prof Roger Hendrix.

Declaration

I declare that this thesis is a presentation of original work and composed by myself. This work has not previously been presented for an award at this, or any other, University. Work described in Chapter 3.5 was previously published in *Nucleic Acids Research*. Contributions by collaborators are explicitly stated in the text and as follows. Bacteriophage and bacterial cultures were provided by Dr Robert Duda and Prof Roger Hendrix, University of Pittsburgh. Procapsids were produced jointly with Dr Robert Duda and Mr Joshua Maurer, University of Pittsburgh. The HK97 motor was reconstituted jointly with Dr Shelley Grimes and Dr Paul Jardine, University of Minnesota. Final EM data for reconstruction were collected by Dr Alexis Huet and Dr James Conway, University of Pittsburgh. Plasmid for GP74 overproduction was a gift from Prof Alan Davidson, University of Toronto. Native mass spectrometry data were collected by Dr Joseph Gault, University of Oxford. NMR data were collected and analysed jointly with Dr Pedro Aguiar and Prof Jennifer Potts. Crystallographic data were collected jointly with Dr Johan Turkenburg and Mr Sam Hart. Finally, references are provided for all sources cited.

1. Introduction

1.1 Viruses

Viruses are small infectious agents which carry their genetic material in protein capsids. Some viruses are surrounded by lipids and contain or are decorated with additional proteins (Figure 1.1). Viruses have no translation machinery of their own. They can self-replicate only by infecting a living host and commandeering its resources. They are incapable of metabolism. Nor can they maintain homeostasis.

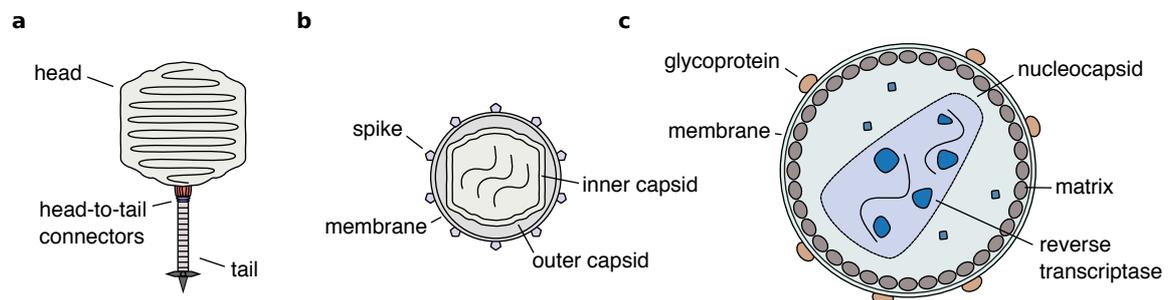


Figure 1.1. Viruses. Schematic of (a) a tailed bacteriophage with a dsDNA genome, (b) phage $\phi 6$ with a segmented dsRNA genome, (c) human immunodeficiency virus with two copies of its ssRNA genome.

Viruses are drivers of evolution. They cause disease and mortality, but they also mediate horizontal gene transfer. To humans, viruses are a mixed blessing. They can be pathogens, or confer pathogenicity and drug resistance to would-be pathogens. They harm livestock and reduce economic productivity. Yet, viruses have been instrumental to our understanding of life. They have led to discoveries such as the chemical basis of genes (1), the nature of the genetic code (2), and methods for producing and manipulating recombinant DNA (3-5). They have become a tool for transduction of genetic material into cells (6,7). Indirectly, they have also given rise to CRISPR technology (8), which has become a popular method for genome editing (9).

The origins of viruses are unknown. However, three hypotheses have been proposed: the escape hypothesis, the regressive hypothesis and the virus-first hypothesis.

Similarities between retroviruses and retrotransposons have been cited as evidence for the escape hypothesis, whereby cellular genetic material gained the ability to leave and enter cells. Retrotransposons that encode a reverse transcriptase and integrase can be reverse-

transcribed once transcribed and inserted elsewhere in the genome. The acquisition of structural proteins that assist with egress and entry might have created the first virus.

With the discovery of nucleocytoplasmic large dsDNA viruses, which have large genomes of different complexities, the regressive hypothesis is gaining popularity. The view is that viruses were once free-living parasites, which became increasingly dependent on their host so much so that they lost their ability to replicate independently. Evidence for this hypothesis includes genes that have no homologues in their host genomes, and the presence of a DNA-directed RNA polymerase for transcription (10), tRNAs and other components for protein translation in some viruses (11).

Proponents of the virus-first hypothesis argue that virus-like self-replicating units existed before the last living universal common ancestor (12). In a confined hydrothermal environment rich in iron-sulfur clusters and organic precursors, future prokaryotic genomes would acquire the ability to synthesis membranes while selfish elements employing a parasitic strategy for self-replication would become viruses.

Lateral gene transfer between viruses with a common host and accretion of genes from the host are common occurrences. An emerging view is that viruses have multiple origins, with lineages escaping from or integrating permanently into the cell at different points in evolutionary history (13,14). Though whatever their origins, viruses have been and will continue to be a significant part of life on our planet.

1.2 Bacteriophages

Much of modern molecular biology began with tailed double-stranded DNA bacterial viruses, or dsDNA bacteriophages. These viruses were discovered by English scientist Frederick Twort in 1915 (15) and French-Canadian scientist Felix d'Herelle in 1917 (16). At the time, viruses were already known as infectious agents smaller than bacteria that would pass through a Chamberland filter. Setting out to grow viruses on artificial media, Twort noted a glassy transformation of *Micrococcus* colonies growing on agar. He found that the transformation coincided with a disappearance of bacteria, and that it was caused by a filterable agent, which could be propagated through further transformations. He concluded that the agent could kill bacteria and in the process multiply itself. While travelling to spread *Coccobacillus* bacteria against locusts, d'Herelle also noticed clear

spots in his cultures, again caused by a filterable agent. On return to France, he isolated a similar agent from the stools of a patient recovering from dysentery. The agent could kill *Shigella* in liquid culture, be serially transmitted but would not grow on artificial media. D'Herelle recognised that this was a virus and in his descriptions coined the name bacteriophage.

From the 1930s to the 1950s, Max Delbrück, Alfred Hershey, Salvador Luria, and peers worked to define the life cycle of bacteriophages. They did so by devising quantitative methods to measure phage and bacterial numbers, by electron microscopy and isotope labelling. The development of single-step growth experiments (17), where unadsorbed phage was chased away so that progeny from infected cells was produced synchronously, allowed the latent period of infections to be measured. Application of Poisson statistics gave the average burst size. The rate of phage adsorption was related to the amount of unabsorbed phage over time by a differential equation. Later, Luria and Thomas Anderson documented by electron microscopy the head and tail structures of phages (18), and, together with Delbrück, captured images of phage adsorbing to cells and the release of more particles on cell lysis (19). This provided evidence that phages self-replicate inside cells. Work by André Lwoff, Evelyne Ritz and colleagues in 1950 showed that a phage could lie dormant and be propagated as genetic material, or prophage, in a lysogenic bacterium (20). A prophage could be induced by UV irradiation, after which new phage particles were assembled and the cell would lyse. In 1952, Hershey and Martha Chase infected cells with phage containing ^{35}S -labelled protein or ^{32}P -labelled DNA (1). Using a blender and centrifugation to separate adsorbed and unadsorbed phage particles from cells, respectively, they found that phage DNA remained with the cells while phage protein did not, demonstrating that phage DNA was the carrier of genetic information. Parallel studies using radioactive tracers established that recombination and duplication of phage DNA occurred inside the cell (21,22). With the discovery of the structure of DNA (23) and ensuing experiments (24), the molecular basis of heredity became clear.

1.3 Phage assembly

1.3.1 Making heads and tails

Many studies followed in the 1960s and the 1970s to elucidate the mechanisms by which phage DNA directs the viral life cycle. For the purpose of this thesis, studies on the dsDNA phage assembly process will be discussed. Studying the morphology of T4 conditional lethal mutants by electron microscopy, Richard Epstein and Edward Kellenberger recognised that the assembly of a tailed phage particle is sequential, that different gene products interact in a specific sequence of events to form the full, intact phage particle (25). The discovery of petit lambda ($p\lambda$), hollow λ head particles lacking in gpD auxiliary or decoration protein (26), led to the hypothesis that empty precursor heads were produced first during phage assembly and then filled with DNA. This hypothesis was confirmed for phage λ through the use of suppressible nonsense mutants (27).

Suppressible nonsense mutants were constructed by UV treatment of phage λ before and after adsorption to suppressor-containing bacteria (28). Most suppressors in bacteria are tRNA molecules that can target a stop codon. Mutants were grouped into classes, which corresponded to genes, by *in vitro* complementation assays, where lysate of suppressor-free bacteria infected with a mutant was mixed with lysate of bacteria infected with another mutant and then tested for infectivity. Further complementation assays using induced lysogens, and protein radiolabelling and SDS-PAGE revealed that the formation of empty precursor heads, or proheads, required gene products gpE, gpB, gpC and gpNu3 (Figure 1.2), and the host groEL/groES chaperonin system (29-31).



Figure 1.2. The structural genes of phage λ . Genes responsible for prohead assembly, packaging and tail assembly are coloured red, blue and orange, respectively. Genes W and F_{II} encode the head-to-tail connectors. Gene product gpF_I increases the efficiency of packaging.

Addition of linear DNA, gpNu1, gpA and ATP would convert proheads into unstable DNA-filled heads, implying that gpNu1 and gpA were essential for DNA filling or packaging (32,33). The unstable filled heads could then be stabilised by addition of gpD. The tails of phage λ appeared to assemble independently of heads (34). In fact, the genes for head and tail assembly mapped to different clusters in the genome (35). Furthermore, filled heads could be combined with *in vitro* tails to form infectious particles (34), with gene products gpW and gpF_{II} serving as connectors between the head and the tail (36). Later, it would become apparent that the assembly process of phage λ consists of the following steps: DNA replication, prohead assembly, packaging initiation, DNA translocation, packaging termination, and tail attachment (Figure 1.3).

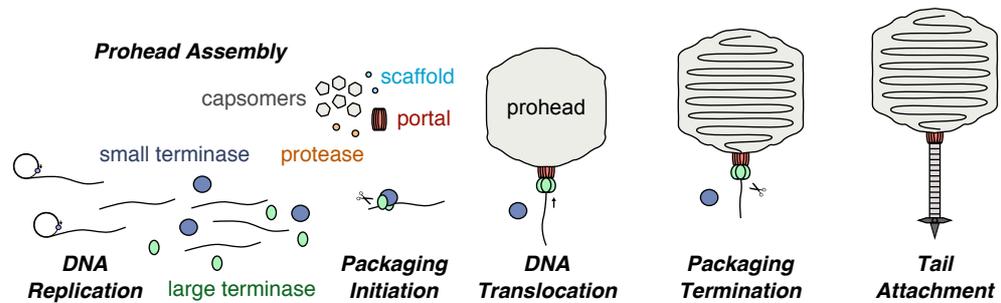


Figure 1.3. The assembly process of tailed *cos* and *pac* bacteriophages. Empty proheads form when capsid protein and scaffolding protein nucleate around a portal and polymerise. New viral DNA, which is mostly concatemeric, is recognised by terminase proteins. The DNA is processed to initiate packaging. A packaging motor then forms, whereby DNA is translocated into the prohead. A second processing event terminates packaging. The free DNA end becomes the starting point of the next packaging event on a new prohead. An independently formed tail is attached to the filled head via head-to-tail connector proteins. The expression of lysis genes causes the host cell to lyse, liberating the mature phage.

1.3.2 Functional classification of phages: *cos*, *pac* and ϕ 29-like

Studies on other tailed bacteriophages found that the assembly of phages follows the same basic scheme as phage λ . However, the structure of the packaged DNA can differ from phage to phage. Mature DNA isolated from phage λ would always end in the same sequence, but mature DNA from phage P22 could end in different sequences (37). DNA from phage ϕ 29 would have ends that are consistent in sequence but fused covalently to protein (38). A new classification system was thus created: a phage may be a *cos* phage, a *pac* phage, or a ϕ 29-like phage, depending on the structure of its mature DNA.

Cos viruses contain mature genomes that have consistent, cohesive ends, or *cos* ends. Examples of *cos* viruses are phages λ , N15 and HK97. *Pac* viruses contain genomes that have variable ends. More than 100% of the genome is contained within each virus particle. DNA ends would differ from particle to particle but for each packaged molecule the two ends would overlap in sequence, that is, they are terminally redundant. Examples of *pac* viruses are phages T4, SPP1, SF6, P74-26, P22 and Sf6. Finally, ϕ 29-like viruses contain genomes with consistent ends but ones that are fused to protein. The different DNA ends in ϕ 29-like viruses is the result of a different DNA replication mechanism (38,39). The different ends in *cos* and *pac* viruses are due to the replication of DNA as concatemers and differences in packaging termination mechanisms. Despite what is known already, structural understanding of the *cos* phage assembly process is limited. This work seeks to close a gap in knowledge by focusing on the assembly process of *cos* phage HK97. However, before introducing phage HK97 in detail, an overview of what is known about each step of the dsDNA phage assembly process will first be provided.

1.4 DNA replication

The *cos* ends of mature phage λ DNA has 5' overhangs which are complementary (40-43). On cell entry, the ends anneal and the DNA circularises under the action of host ligase. Expression of delayed early genes triggers five to six rounds of bidirectional θ replication, where replication forks travel in opposite directions around the DNA circle (44). Later in infection, the majority of DNA is produced by rolling-circle σ replication. The switch to σ replication can occur through Red-mediated recombination or through a unidirectional θ phase, with the 3' end of the leading strand eventually displacing its 5' end from template DNA (45). In combination with lagging strand synthesis, σ replication in phage λ generates concatemeric dsDNA.

The assembly of replication complexes at the origin of λ DNA replication, *ori λ* , requires DnaB helicase activity (46). Protein gpO bound to *ori λ* recruits gpP in complex with DnaB. Chaperones DnaJ, DnaK and GrpE are then required to displace gpP to release the inhibition on DnaB. Helicase activity by DnaB enables the recruitment of primase DnaG and the DNA polymerase III holoenzyme. Finally, DnaA-dependent transactivation of *ori λ* at the *pR* promoter signals for replication to begin (45). The trigger for σ replication is unknown but appears to be dependent on DnaA.

Pac genomes such as that of P22 circularise via homologous recombination upon cell entry (37). The DNA is duplicated via bidirectional θ replication and rolling-circle replication. Similar replication complexes to those that form for λ DNA replication feature during this process. Other *pac* genomes such as that of phage T4 do not circularise. Bidirectional replication initiates from RNA-DNA hybrids that form at some origins. Y-shaped, branched intermediates are observed because progression of the leftward fork from the origin is delayed (47). Recombination-dependent replication dominates later in infection. Repeated strand invasion by previously synthesised DNA and ensuing elongation produces concatemeric DNA (48).

For ϕ 29-like viruses, single unit-length genome copies are produced during replication due to a protein-primed DNA replication mechanism (38,39). This gives rise to DNA whose ends are always fused covalently to protein.

1.5 Prohead assembly and capsid maturation

1.5.1 Prohead assembly

Studies based on petit λ showed that the λ prohead consists of gpE, proteolytically processed gpB, and fragments of gpC and gpE fused together, and that gpNu3 is required for assembly but subsequently lost (31,49). It is now understood that gpE is the major capsid protein; gpB is the portal protein; gpNu3 is the scaffolding protein (50). First, gpB and gpNu3 associate to form a nucleation centre for gpE (51). Then, gpE, with the help of additional gpNu3, polymerises into an icosahedron; gpB is processed by gpC, the capsid maturation protease (52). Crosslinks form between gpC and some copies of gpE; the fusion is cleaved, and the products remain bound to the prohead (53,54). Finally, gpNu3 leaves and DNA packaging begins (Figure 1.4). A similar reaction involving portal, scaffolding and capsid protein is observed for other tailed phages, including phage P22 (55).

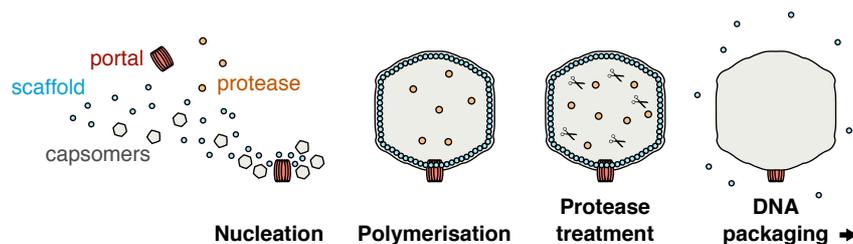


Figure 1.4. Prohead assembly. Capsid protein nucleates around a dodecamer of portal protein in the presence of scaffolding protein and polymerises into an immature prohead. Proteolysis takes place under the action of capsid maturation protease, and the scaffolding leaves either before or during DNA packaging.

Most viruses have icosahedral symmetry or helical symmetry. Some are conical (56), and some are pleomorphic (57). In the late 1950s, Crick and Watson reasoned that the viral genome does not encode enough proteins to form a capsid if only one copy of each protein were made (58,59). Based on the X-ray diffraction patterns of tobacco mosaic virus (60) and turnip yellow mosaic virus (61), they proposed that viral capsids are made of symmetrically arranged, identical subunits. Caspar and Klug noted in 1962 that the most efficient way to enclose a sphere while maintaining an identical environment for each subunit is to arrange subunits in pentamers in the shape of an icosahedron (62). Hexamers can be inserted among pentamers to form larger capsids, with minor distortion to the protein-protein interface (Figure 1.5a). Many viruses have prolate heads, for example, phages ϕ 29 (63) and T4 (64). Prolate heads consist of two end-caps, each with five equilateral triangular facets, and a mid-section containing ten triangular facets (Figure

1.5a). The scaffolding protein seems key for prolate head formation; deletion of the scaffolding can lead to isometric heads (65).

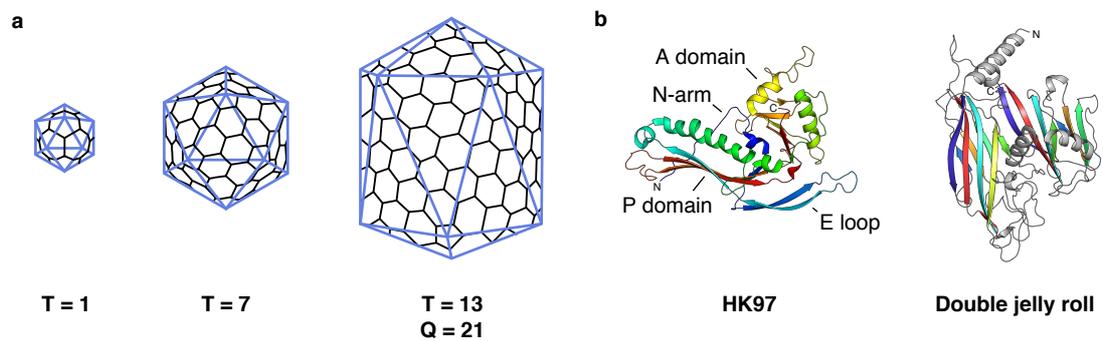


Figure 1.5. Icosahedral capsids. (a) An icosahedron has six 5-fold, ten 3-fold and fifteen 2-fold rotation axes. The quasi-equivalence principle describes how proteins can assemble into larger capsids but maintain similar chemical environments. Capsids are described by their triangulation number, $T = h^2 + hk + k^2$, where h and k are the number of hexamers separating pentamers in orthogonal directions. For prolate capsids, the Q number describes the number of asymmetric units composing a triangular facet in the mid-section. (b) The HK97 fold (PDB code 3E8K) and the double jelly-roll fold (PDB 1CJD). The double jelly-roll fold consists of two single jelly rolls fused together. The β -strands of each jelly-roll are coloured in order from blue to red.

Tailed bacteriophages generally encode capsid proteins of the HK97 fold (66). Eukaryotic and archaeal viruses encode capsid proteins of the HK97 or jelly-roll fold (Figure 1.5b) (67). The portal of tailed phages occupies a special 5-fold vertex of the icosahedral capsid. A ring consisting of twelve copies of the portal protein provides a tunnel through which DNA can enter and exit. The portal of phages with prolate heads is located at the tip of one of the two end caps. In all systems studied so far, the portal has twelve subunits. This creates a symmetry mismatch, as capsid proteins surrounding the portal are related by 5-fold symmetry. The symmetry mismatch has led to questions over the stoichiometry of the packaging assembly, and whether the portal rotates with respect to capsid during packaging (68).

1.5.2 Capsid maturation

Proheads are known to expand during DNA packaging, doubling in volume in some cases to accommodate the viral genome (66). The expansion is also referred to as a maturation process. Unlike DNA packaging, capsid maturation is well documented for *cos* phages in structural terms. The capsid of phage HK97 was, in fact, the first dsDNA phage capsid for which crystal structures were determined; hence, the naming of the fold (69,70).

The HK97 capsid is a $T = 7$ icosahedron. Folding of the HK97 capsid protein gp5 requires the host groEL/groES chaperonin system and produces a mixture of pentamers and hexamers (71). These pentamers and hexamers are interconvertible *in vitro* and can assemble spontaneously into capsids (72). Unique to the HK97 system, the scaffolding protein is fused to the capsid protein as an N-terminal delta domain. Polymerisation of capsid protein with the delta domain produces the capsid conformation known as Prohead I (Figure 1.6a). Removal of the delta domain by the capsid maturation protease gp4 produces Prohead II, which is believed to be the state prior to DNA packaging (73-75).

During packaging, negative charges on the DNA and growing pressure within the capsid are thought to trigger a conformational change in the capsid, causing it to expand. Crystal structures and EM reconstructions of empty, portal-less proheads, heads, and intermediates later reveal the conformations sampled as the capsid expands (Figure 1.6a) (66,70,76,77). The N-arm of the capsid protein transitions from a rigid state in Prohead I, directed towards the delta domain in the interior of the capsid; then to a more mobile state; and finally to a rigid state again, resting between adjacent capsid subunits. Accompanying this change, subunits undergo a twisting motion about the three β -strands of the P domain while the long spine helix un-kinks (Figure 1.6b).

While most phages encode a cementing or auxiliary protein (64,78), which interacts with the 3-fold vertices of capsids to stabilise their expanded form, HK97 does not encode an auxiliary protein. Instead, crosslinks form in the mature head between residues K169 and N356 of neighbouring subunits, catalysed by residue E363 (70,79). This creates a protein 'chainmail' (Head II, Figure 1.6c), which locks pentamers and hexamers together to stabilise the capsid (80).

A more detailed structural investigation into the HK97 DNA packaging assembly could reveal the molecular triggers for capsid expansion and subsequently crosslink formation. This forms another motivation for the work described in this thesis.

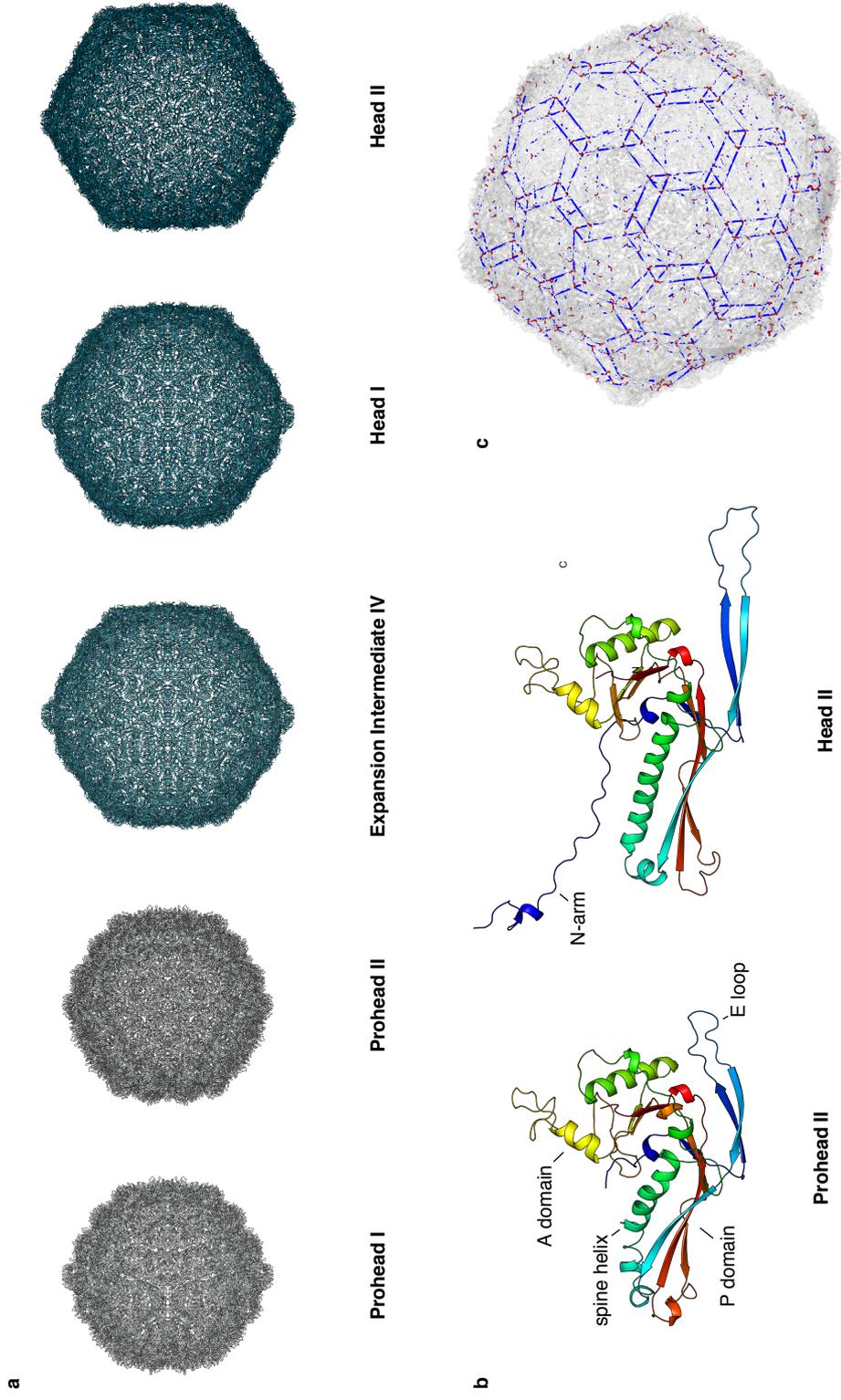


Figure 1.6. The HK97 capsid. (a) The maturation pathway of the HK97 capsid, captured by X-ray crystallography. Capsid protein first polymerises into Prohead I (PDB code 3QPR). Removal of the delta domain generates Prohead II (3E8K). Treatment with acid triggers expansion, resulting in Expansion Intermediate IV (2FRP). On reversion to neutral pH, crosslinking occurs, giving Head I (1OHG). Head I (2FS3), the state prior to crosslinking, is captured via a K169Y mutation. (b) Conformational changes in the capsid protein. (c) Protein chainmail formed by crosslinks (red). Blue lines connect residues K169 and N356 of the same subunit.

1.6 Packaging initiation

The process of packaging initiation on proheads differs among phages. $\phi 29$ DNA is replicated as unit-length, protein-fused genome copies. Thus, to initiate packaging, DNA-gp3 simply needs to be recruited to the prohead. This involves supercoiling of the DNA by the packaging ATPase gp16 (81). Prohead RNA or pRNA molecules associate with the special vertex of the prohead to assist with recruitment of the ATPase to the prohead. The pRNA contributes not only to initiation but also to the selection of DNA substrates for packaging (82,83).

The initiation of packaging for *cos* and *pac* viruses is more complex, as the substrate is concatemeric. *In vitro* complementation assays using mutants of phage λ showed that gpA is essential for the generation of *cos* ends as well as filling of proheads (33,84). The minimal region required for gpA activity is referred to as the *cos* site. Mapping studies using a λ mutant, where an additional *cos* site is inserted asymmetrically into the genome, showed that λ DNA is not entirely processed into monomers before packaging (85). Instead, most processing occurs during packaging. Further complementation assays using DNA substrates with intact *cos* sites to the left and right of a genome copy, according to the genetic map, showed that λ DNA is packaged from left to right from one *cos* site to the next (86). Therefore, the *cos* site marks the beginning and end of a genome copy in *cos* viruses.

A holoenzyme later isolated from induced lysogens suggested that *cos* site processing, also known as maturation, and packaging requires gpA, which contains an N-terminal ATPase and C-terminal nuclease domain, as well as a second gene product, gpNu1 (87). The gpNu1 protein was purified and found to have specificity for three direct repeats downstream of the cleavage site (88-90). The region spanning the repeats was later named *cosB* (Figure 1.2). An integration host factor (IHF) consensus recognition sequence is also present within *cosB*. Addition of IHF increases the affinity and specificity of gpNu1 for *cosB* DNA *in vitro* (91,92). Analytical ultracentrifugation data suggest that gpNu1 dimers would associate with one molecule of gpA each and then tetramerise subsequently to form the holoenzyme (93,94). Thus, the holoenzyme, or terminase, would contain eight small gpNu1 terminase subunits and four large gpA terminase subunits.

The N-terminal region of gpNu1 is a winged helix-turn-helix domain (Figure 1.7a) (95). Both wing and helix-turn-helix motif are implicated in contacting DNA. The current model of packaging initiation in phage λ is that *cos* DNA is bent around the circularly arranged N-terminal domains of gpNu1 to be presented to gpA as a substrate for DNA cleavage; and IHF binding assists with the bending of the DNA (Figure 1.7b). The genomes of other *cos* phages lack an IHF binding site in the vicinity of the cleavage site. In these cases, IHF is not necessary for DNA maturation or packaging (96).

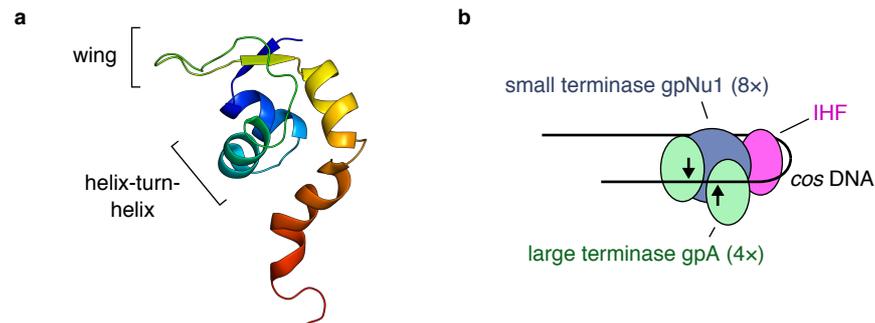


Figure 1.7. Packaging initiation in *cos* phage λ . (a) Solution NMR structure of the λ small terminase gpNu1 N-terminal domain, coloured from blue to red, N- to C-terminus. Both wing and helix-turn-helix elements are indicated. (b) Current model of packaging initiation in phage λ . DNA is bent by host IHF and wraps around the λ maturation complex with eight small terminase gpNu1 and four large terminase gpA subunits. Cleavage occurs upstream of the gpNu1 binding sites, indicated by black arrows.

Other than nuclease activity, the C-terminal nuclease domain of λ gpA has also shown ATPase and helicase activities (97), which stimulate *cos* end generation and separation, respectively. The domain contains a Walker A sequence, likely a nucleotide-binding site, and a basic leucine zipper region implicated in the dimerisation of the protein on DNA. These features are present in some but not all *cos* viruses outside of the lambdoid family. No further structural information is available on *cos* terminase proteins. Thus, the molecular detail of packaging initiation in *cos* viruses remains unclear.

Pac viruses encode similar small terminase and large terminase proteins. A *pac* site marks the beginning of the first genome copy packaged from concatemeric DNA, but only the beginning (98-100). Cleavage at the *pac* site is also not highly precise; cleavage can occur over a range of five to twelve nucleotides around the primary cleavage site (101,102). Crystal structures and native mass spectrometry data show that *pac* small terminase proteins form rings of eight to eleven subunits (103-105). Most *pac* small terminase proteins have N-terminal helix-turn-helix (HTH) regions, which are circularly displayed

(Figure 1.8a) (103,104). In a DNase footprinting assay using phage SPP1 small terminase, repeating patterns of protection and sensitivity were observed over a region of ~100 bp in the *pac* site of the SPP1 genome (106). This has led to a model where *pac* DNA, which is AT-rich, wraps around the small terminase protein oligomer for packaging initiation (Figure 1.8b). For phages with small terminase proteins that lack an N-terminal helix-turn-helix, for example, phage 44RR, the C-terminus is implicated in DNA recognition instead (107).

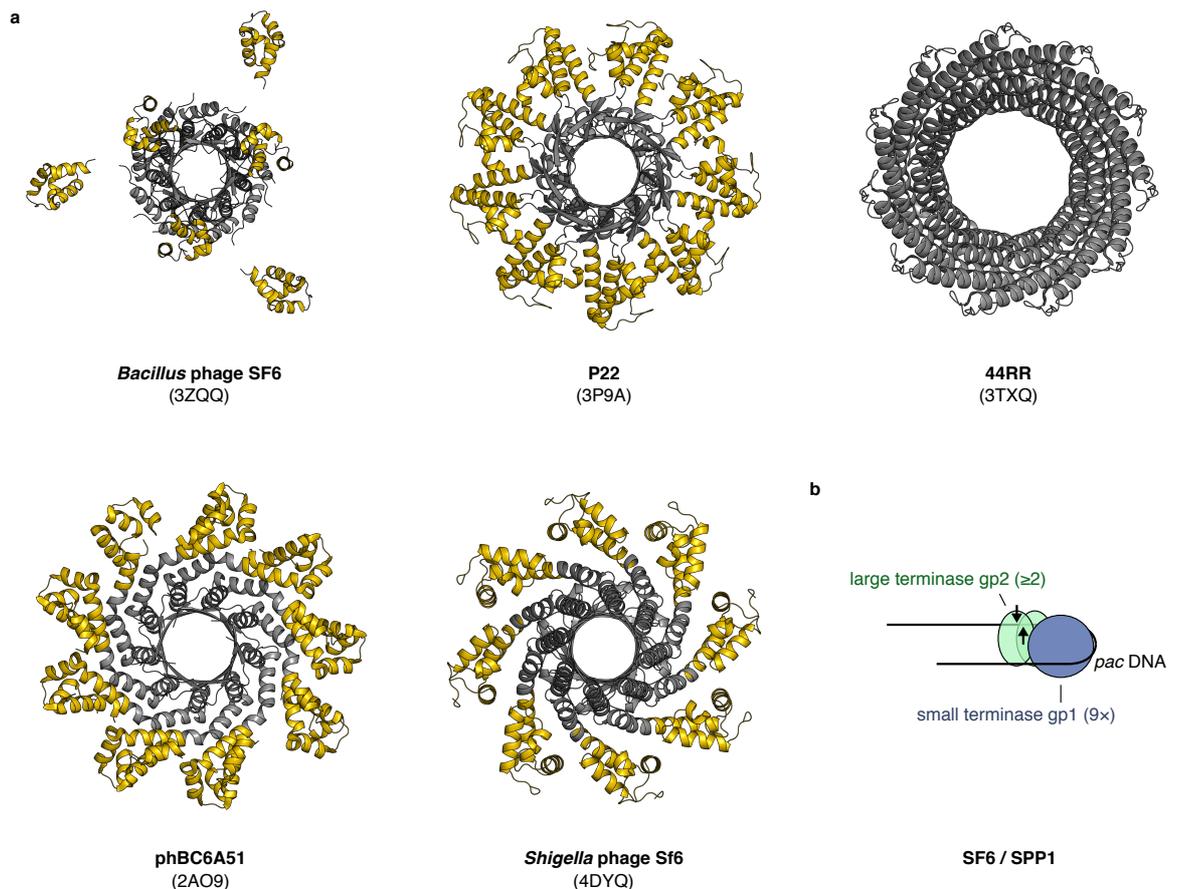


Figure 1.8. Packaging initiation in *pac* phages. (a) Crystal structures of *pac* small terminase proteins. PDB codes are listed in brackets. All but the small terminase of phage 44RR contain N-terminal helix-turn-helix structures (coloured in yellow). Three of nine N-terminal domains of the SF6 small terminase are disordered in the crystal structure. (b) Current model of initiation in *Bacillus* phages SF6 or SPP1. DNA wraps around nonameric small terminase. At least two large terminase subunits are recruited for the cleavage event. Cleavage occurs downstream of the small terminase binding sites, indicated by black arrows.

For *pac* phages, large terminase appears to be recruited through its N-terminal end to the C-terminus of small terminase. On recruitment, the C-terminal domain of large terminase becomes activated and cleaves the DNA. At least two large terminase molecules are thought to be present in the maturation assembly. Crystal structures of *pac* large terminase proteins suggest that DNA cleavage occurs via a two-metal mechanism similar to that of RuvC-like proteins in the RNase H superfamily (Figure 1.9) (108,109). A β -hairpin unique to terminase proteins is implicated in regulating this nuclease activity. Once the DNA is cleaved, DNA along with large terminase is recruited to the portal vertex of a prohead for packaging. The conformational intermediates involved in this process are unknown. It is also unclear whether small terminase remains bound to the complex after initiation.

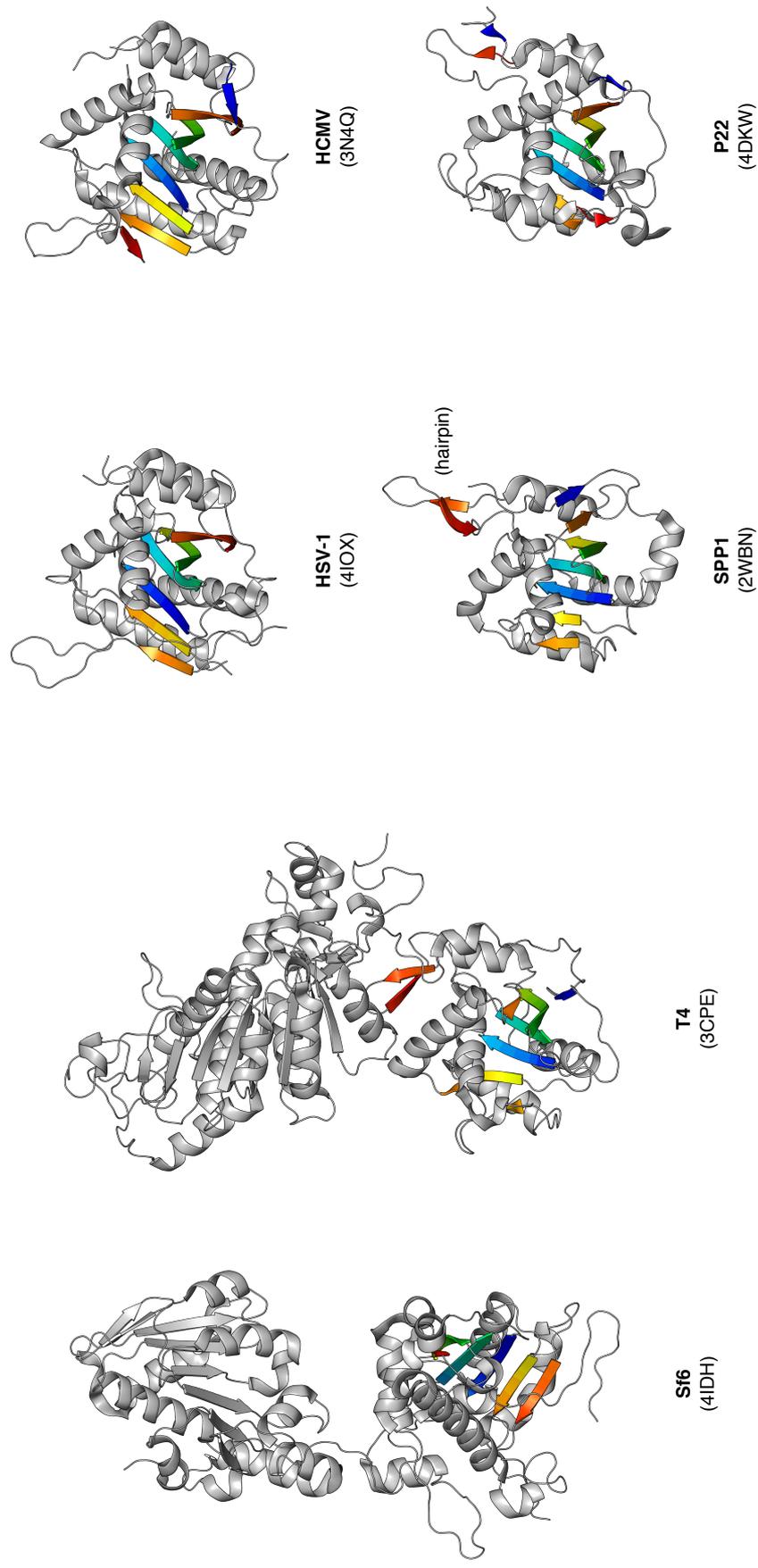


Figure 1.9. Crystal structures of *pac* large terminase proteins and their nuclease domains. PDB codes are listed in brackets. β -strands in the nuclease domain are coloured from blue to red, N- to C-terminus. The β -hairpin (orange/red), or sometimes extended loop, implicated in regulating nuclease activity is labelled.

1.7 DNA packaging

In *cos* and *pac* viruses, a complex forms between the prohead, large terminase and DNA to package the DNA. In ϕ 29-like viruses, the complex is formed between the prohead and pRNA, the packaging ATPase and DNA. Till the late 1980s, this complex, known as the packaging motor, had only been investigated through complementation assays, where purified components or lysates of infected cells or induced lysogens were mixed together and the assembly of phage particles tested for using plaque assays. The development of an *in vitro* packaging assay (110), using purified components in combination with DNase treatment and agarose gel electrophoresis to analyse the packaged material, provided the means to understand the motor more thoroughly.

Single-molecule biophysical studies became feasible with the identification of conditions that favour motor assembly (111-113). Optical tweezers experiments showed that the ϕ 29 motor can package DNA at speeds up to 165 bp/s (114); the λ motor can package up to ~600 bp/s (115); and the T4 motor can package up to 2000 bp/s (116). The speed appears to correlate with genome length: 19.3 kb for ϕ 29, 48.5 kb for λ , and 171 kb for T4. All three motors can work against forces greater than 50 pN, applied to the end of the DNA. Thus, viral DNA packaging motors represent some of the more powerful motors found in nature. For comparison, *E. coli* FtsK DNA translocase moves at speeds up to 5000 bp/s and has a stall force of 60 pN (117). RecBCD moves at 500 bp/s and has a stall force of 8 pN (118). *E. coli* RNA polymerase elongates at speeds up to 10 nt/s and stalls at 23 pN of applied force (119,120). Chicken myosin V moves 200 nm/s and stalls at 2.5 pN(121). It should be noted though that stall force does not always correlate with the ability of the motor to clear a roadblock (122).

Asymmetric cryo-EM reconstruction of the ϕ 29 motor showed that five gp16 ATPase molecules and five pRNA molecules associate with the portal vertex of the prohead (123,124). There is again a symmetry mismatch, but the assembly is stabilised by pRNA bridging between the capsid and the ATPase pentamer. Reconstructions of the *pac* T4 motor revealed a similar architecture, with five large terminase molecules associating with the portal of the prohead (125). The full-length structure of the T4 large terminase (126) was initially docked with the N-terminal domain contacting the portal. It is now known through Förster resonance energy transfer studies that the C-terminal nuclease domain of

T4 large terminase interacts with the portal (127). Similarly, the C-terminus of λ large terminase is found to be important for interaction with the portal of proheads (128).

Despite the symmetry mismatch between the capsid and portal, it was found by fluorescent labelling of the portal that the portal does not rotate with respect to the capsid during DNA translocation (129). Rather, DNA rotates with respect to the ATPase, portal and capsid as it rolls from one ATPase subunit to the next during translocation (130). Single-nucleotide-resolution optical tweezers experiments showed that the ϕ 29 motor translocates in bursts of four 2.5-bp steps followed by a relatively long dwell phase (131). By challenging the system with ATP γ S, ADP and ADP-vanadate, it was found that ATP hydrolysis and phosphate release occur during the burst phase, in tandem with each working stroke, and ADP release and ATP binding occur during the dwell phase (132,133). ATP binding did not appear cooperative, and the dependence of the dwell duration on ADP concentration was well described by a competitive-inhibition model, indicating that ADP is released from one subunit and ATP binds before ADP is released from the next subunit. Based on the distribution of burst sizes prior to ATP γ S-induced pausing, it was concluded that all five subunits hydrolyse ATP though only four provide a working stroke in every cycle. Hydrolysis by the non-translocating subunit either leads or terminates the burst phase. So far, it has been difficult to test this model in another system, because DNA translocation is too fast in the λ and T4 motors for individual steps to be resolved. However, mutants have now been identified in the T4 system that will slow down packaging (134). This should enable studies of motor coordination for *pac* systems.

The ATPases of viral packaging motors have a unique fold (Figures 1.10, 1.11a). While they are members of the Additional Strand Catalytic E (ASCE) superfamily (135), they differ from AAA+, RecA and FtsK/HerA proteins in the secondary structure elements that they contain in addition to the ASCE core, which has the Rossmann fold. Unlike AAA+, RecA and FtsK/HerA proteins, they form hexamers instead of pentamers. ATP hydrolysis involves the coordination of a magnesium ion by the Walker A serine or threonine, stabilisation of the magnesium via a water-mediated interaction with the aspartate adjacent to the catalytic glutamate, coordination of the β and γ phosphates of ATP by the magnesium and Walker A lysine, abstraction of a proton from a water molecule by the catalytic glutamate, followed by nucleophilic attack at the γ phosphorous (Figure 1.11b).

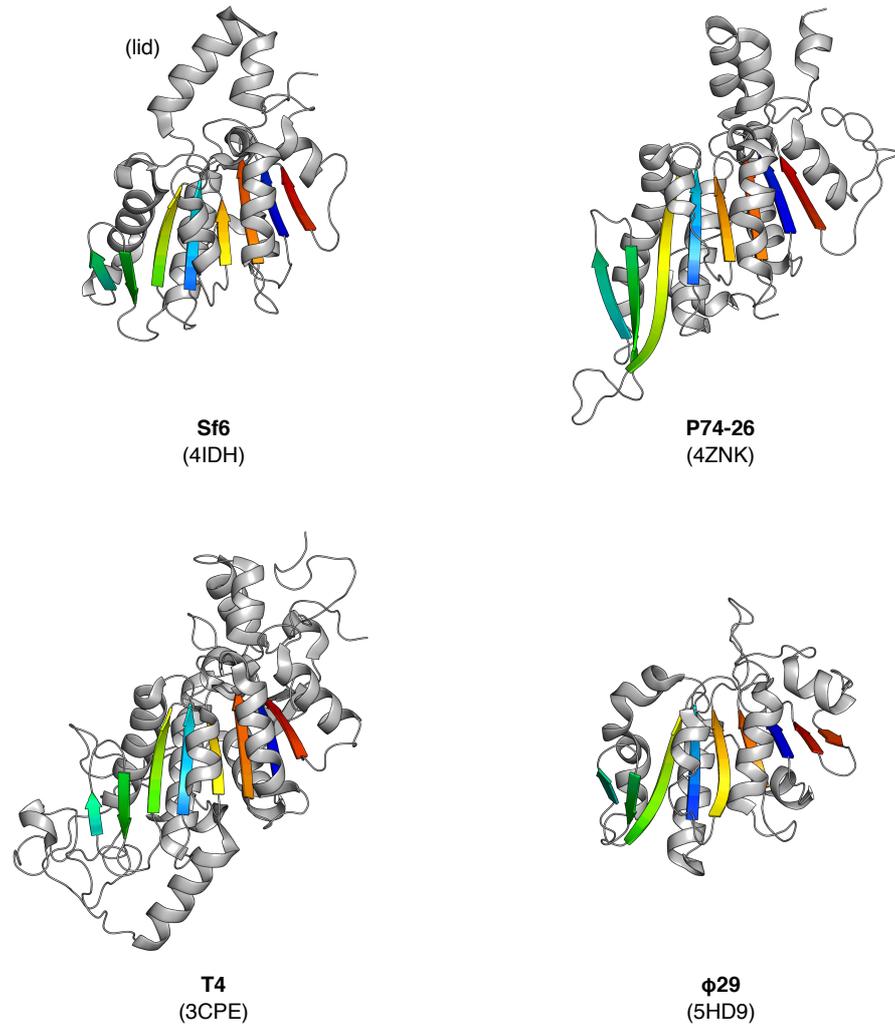


Figure 1.10. The packaging ATPases of phage $\phi 29$ and *pac* phages T4, Sf6 and P74-26. PDB codes are listed in brackets. β -strands in the nuclease domain are coloured from blue to red, N- to C-terminus. The C-terminal helical bundle, or lid, of the ATPases is indicated. The structures determined are of monomeric protein. The ATPase functions as a pentamer in the packaging motor.

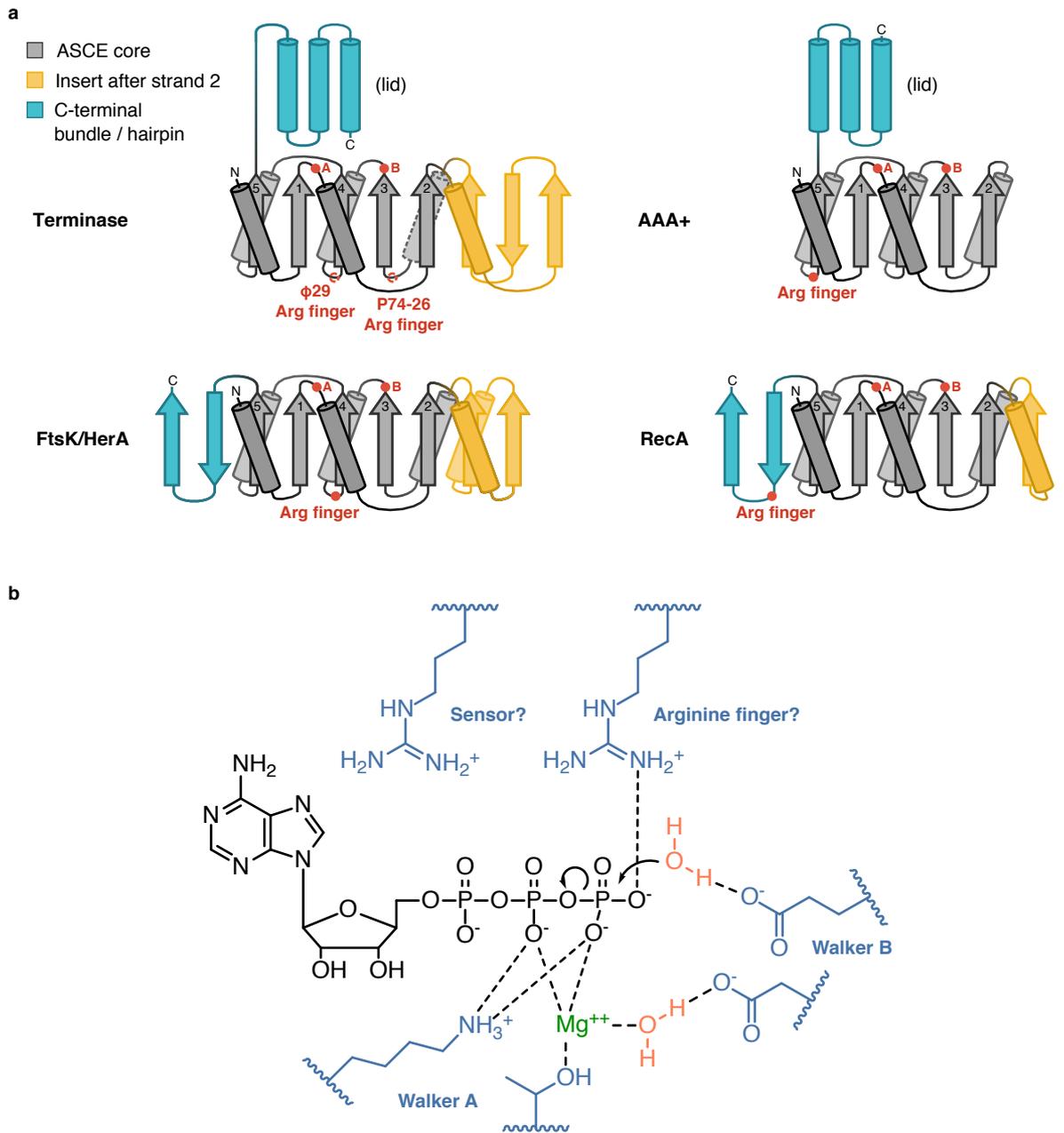


Figure 1.11. ATPase fold and reaction mechanism. (a) The ATPase of viral DNA packaging motors has a distinct fold within the ASCE superfamily. It has a mixture of inserts which feature separately in AAA+, FtsK/HerA and RecA proteins. Arginine fingers have been identified in two different loops within the ASCE core of terminase ATPases, indicated by red dotted circles. Red circles mark the location of the Walker A and B sequences. The C-terminal bundle of terminase and AAA+ ATPases is often referred to as a lid. (b) Putative mechanism for ATP hydrolysis. ATP is coordinated by the Walker A lysine and magnesium ion, which is in turn coordinated by the Walker A threonine/serine and an aspartate residue from the Walker B sequence. The catalytic glutamate abstracts a proton from a water molecule, which then attacks the γ phosphorus of ATP.

There is an ongoing debate over how the leaving phosphate group is stabilised during hydrolysis. In T4, conformational changes during oligomerisation are thought to reposition a conserved arginine residue in the P loop so that it would coordinate the leaving phosphate (126). In ϕ 29 and P74-26, the conserved arginine is said to be a sensor II-like residue that mediates conformational changes during ATP turnover. Instead, *trans*-acting arginine fingers have been identified which would donate a positive charge to the active site of the adjacent subunit to catalyse hydrolysis (124,136). As with most ATPases, the conformational changes that accompany ATP hydrolysis are harnessed to produce work. For viral packaging ATPases, the relative movement between the ASCE core and the C-terminal lid is believed to drive DNA translocation.

1.8 Packaging termination

In the later stages of packaging, smaller steps are taken during the burst phase and long-lived pauses become more frequent during the dwell phase, likely due to the growing pressure inside the capsid as it fills (130). Because the DNA substrate is often still concatemeric in *cos* and *pac* viruses, cleavage of the DNA is required to terminate packaging. Cleavage in *cos* viruses occurs precisely at the next *cos* site. As a result, the same *cos* ends are generated in every copy of the genome packaged (Figure 1.12a). Termination in λ is dependent on the presence of a *cosQ* region on the right end of the genome (137) and the presence of an I2 element downstream of the cleavage site (138). It is unclear how the ATPase is deactivated and how the nuclease is activated. However, it was found that cleavage at artificial internal *cos* sites was more likely when the prohead was more full (85,139). This indicates that there is likely a small headful element to the termination process of *cos* viruses.

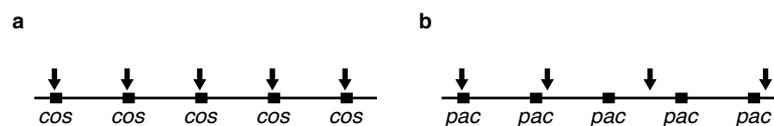


Figure 1.12. Packaging outcomes for *cos* and *pac* viruses. (a) Unit-length copies with identical ends are packaged due to sequence-specific termination at the *cos* site. (b) More than a full length of a genome is packaged in a *pac* virus because of the headful termination mechanism. As a result, DNA is packaged in a moving window. Each packaged species will have ends that overlap in sequence.

Termination in *pac* viruses is entirely headful. Cleavage of the DNA is independent of the downstream *pac* site, but dependent on the structural changes that occur when the capsid becomes full. The mechanism by which this signal is transmitted to the large terminase is unknown, but it is thought to involve changes in portal conformation due to expansion of the capsid or due to pressure from the packaged DNA itself (140,141). Cleavage can occur over thousands of base pairs after the downstream *pac* site. As a result, a typical phage particle will contain more than 100% of the genome. The liberated end becomes the start point of the next copy of DNA packaged. The *pac* site is no longer relevant for subsequent packaging events. Due to the headful mechanism, the ends of each packaged species will overlap in sequence and are thus terminally redundant (Figure 1.12b).

Studies using filled λ heads and tails showed that the right end of the genome is sensitive to nuclease treatment unless a tail is added on (142). For tail addition, the terminase must

leave and head-to-tail connectors must interact with the portal. The portal of the mature head is thought to have a lower affinity for terminase and higher affinity for the head-to-tail connectors (143). Finally, the expression of holins and lysins to create holes in the membrane and degrade the peptidoglycan cell wall, respectively, allows assembled phage particles to escape from the host.

1.9 Assembly of other viruses

The assembly of tailed dsDNA bacteriophages is a highly orchestrated process, but the assembly process of other viruses is no less complex. For completeness of discussion, a summary of assembly strategies used by other viruses is provided here.

Among dsDNA viruses, herpes viruses use a similar strategy to that of tailed phages to package their DNA. A packaging motor translocates DNA into empty pre-formed capsids in the nucleus (Figure 1.13a) (144). Instead of associating with a tail, filled capsids bud into the perinuclear space and fuse with the outer nuclear membrane to be released into the cytoplasm (145). Tegument protein associates with the capsids; the capsids pass through the *trans*-Golgi network, becoming enveloped and acquiring glycoprotein in the process. Finally, the assembled viruses exit the host via exocytosis. Tailed bacteriophages belong to the order Caudovirales, which contain families *Siphoviridae*, *Myoviridae* and *Podoviridae* differing in tail morphology. *Tectiviridae* is a bacteriophage family with no assigned order. An internal membrane surrounds the genome, which has protein-fused ends, inside the viral capsid. The capsid protein of phage PRD1 has the jelly-roll fold (146). A hexameric FtsK-like ATPase assembly associates with the dodecameric portal at the special vertex to form the DNA packaging motor (147). The portal, in turn, associates with a transmembrane protein complex, which offers passage through the internal membrane (Figure 1.13b). Some dsDNA viruses do not use a packaging motor. For example, polyomaviruses make use of host histones to condense their DNA along with coat protein into a nucleocapsid (148,149). This process resembles that in ssRNA viruses where coat protein condenses with RNA to form the nucleocapsid.

Many ssDNA viruses, for example, parvoviruses, produce empty procapsids. A packaging motor then fills these procapsids with ssDNA. However, the packaging motor is a helicase rather than a translocase (150,151). One strand is fed into the procapsid as DNA is unwound. Phage ϕ X174 uses a different strategy. DNA from rolling circle replication is

fed directly into the prohead (Figure 1.13c) (152,153). ϕ X174 has another unusual feature: the prohead shrinks, rather than expands, during packaging (154).

dsRNA viruses in the *Cystoviridae* family such as phage ϕ 6 and in the *Reoviridae* family such as rotavirus have segmented genomes. Plus-strand RNA segments are packaged into the procapsid by a hexameric NTPase (155). RNA-directed RNA polymerase contained within the procapsid will synthesise negative-strand RNA from the plus-strand RNA, converting the segments into dsRNA (Figure 1.13d). *Reoviridae* capsids carry an additional enzyme for the addition of 5' caps to newly synthesised viral mRNA during an infection (156). Genome encapsidation in ssRNA viruses occurs through interaction of coat protein or core protein with hairpin packaging signals in the genome. RNA and protein condense together to form the nucleocapsid (Figure 1.13e). This can occur in the cytoplasm (157) or at the cell membrane (158). The process of genome encapsidation through packaging signals is reviewed in (159).

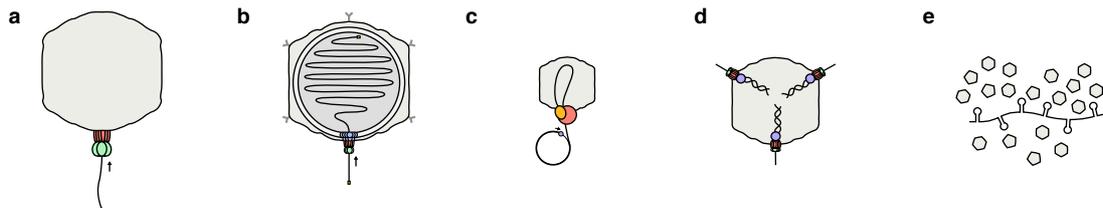


Figure 1.13. Virus assembly mechanisms. (a) Tailed bacteriophages and herpes viruses encode a terminase-based dsDNA packaging motor. (b) The *Tectiviridae* family encodes a FtsK-like dsDNA packaging motor. (c) Packaging of ssDNA is coupled to rolling circle replication in phage ϕ X174. (d) RNA-directed RNA polymerase synthesises dsRNA from plus-strand ssRNA entering the capsid due to helicase activity in *Cystoviridae* and *Reoviridae*. (e) Capsid and accessory protein condense together with the RNA genome to form the nucleocapsid in ssRNA viruses.

1.10 Bacteriophage HK97 as a model system for the study of DNA packaging

Current structural understanding of DNA packaging in *cos* viruses is limited. The only terminase structure available is that of the N-terminal domain of λ small terminase gpNu1 (Figure 1.7) (95). The structure and stoichiometry of the complete motor are unknown. By studying the structural, biophysical and biochemical properties of an as yet uncharacterised system, new insights may be gained into the mechanisms of packaging in *cos* viruses.

Tailed bacteriophage HK97 is a *cos* virus. Originally isolated from pig dung in Hong Kong (160), it has an isometric icosahedral head and a long, flexible tail. The genome of HK97 has a similar gene order to that of λ (Figure 1.2). The two share 32–95% amino acid sequence identity in: their tail fibre proteins, regulatory proteins cI, cII and cIII, and proteins associated with DNA replication, integration and cell lysis (161). Perhaps because of the similarities in their tail fibres, both phages adsorb to the LamB receptor of *E. coli* for infection (160,162). Sequence similarity in the structural proteins between HK97 and λ is poor. However, the identity of the small and large terminase proteins can be inferred based on gene order and active site motifs (Figure 1.14). It is expected that the HK97 maturation complex for packaging initiation will be arranged differently to that of λ , because the HK97 *cos* ends contain 12-nt 3' overhangs rather than 10-nt 5' overhangs (161).

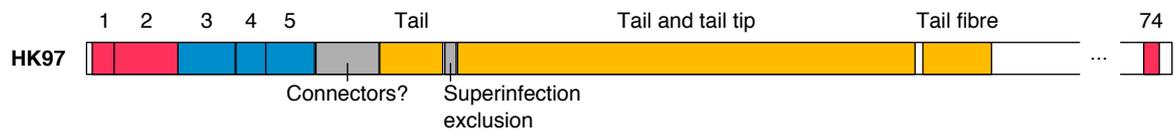


Figure 1.14. The structural genes of phage HK97. Genes responsible for prohead assembly, packaging and tail assembly are coloured red, blue and orange, respectively. Genes 1 to 5 of HK97 encode the small terminase, large terminase, portal protein, capsid maturation protease and capsid protein, respectively. The scaffolding is fused to the capsid protein. No auxiliary protein is encoded. A moron element is found among the structural genes. This was later shown to be a superinfection exclusion protein (163). Gene 74 encodes an endonuclease cofactor which has been implicated in stimulating *cos* DNA cleavage during packaging (164).

The wealth of knowledge on the genetics, biology and biochemistry of phage HK97 makes it an attractive system for studying DNA packaging. Therefore, the primary objectives of this work were to reconstitute a functional DNA packaging motor *in vitro* based on phage HK97 (Chapter 2); to study the structure and function of the small terminase and large terminase (Chapters 3 and 5); and to determine the stoichiometry and structure of the HK97 DNA packaging motor (Chapter 4), with an aim to elucidate the mechanisms of packaging initiation, DNA translocation and packaging termination. The theory underlying the structural biology and biophysical methods important to this work is reviewed in Appendix B.

2. *In vitro* reconstitution of the HK97 DNA packaging motor

2.1 Introduction

Phage HK97 is a *cos* tailed bacteriophage. It has consistent ends with 12-nt 3' overhangs. The structural genes of phage HK97 are located on the left end of the genome, by reference to the λ genome (Figure 1.10). This chapter details the assembly of a functional HK97 DNA packaging motor using purified components. The roles of small terminase, large terminase and GP74 in initiation, translocation and termination were explored. The biochemical knowledge gained provided a foundation for ensuing biophysical and structural studies.

2.2 Materials and methods

2.2.1 Phage production and purification

To produce phage HK97 particles on a large scale, clear plaque mutants were used to ensure a lytic cycle. First, plaques were generated. An overnight culture of *E. coli* LE392 cells (165) (*supE44*, *supF58*) was prepared in LB medium which had been supplemented with 0.4% (w/v) maltose (Fisher Scientific). Maltose was added to ensure expression of the LamB receptor, a recognition site for the phage. In the absence of maltose, MalT, which drives *lamB* expression, is inhibited by MalK binding. Maltose causes MalK to become part of the MalFGK₂ transporter complex, thereby relieving its inhibition on MalT (166). Serial dilutions of the phage stock were prepared in TMG (10 mM Tris-Cl, pH 7.5, 10 mM MgSO₄, 0.01% (w/v) gelatin). Volumes of 50, 100, 200 μ L were taken from the diluted stocks and incubated at 37 °C with 150 μ L of cells from an overnight culture for 20 min. Then, 2 mL molten top agar (0.7% (w/v) agar in LB) was added to the mixture, mixed well, and the entire volume transferred onto warm solid bottom agar (1.5% (w/v) agar in LB). Once the top agar had set, the plates were incubated upside down overnight at 37 °C.

One to three plaques were picked the following day with a glass Pasteur pipette and deposited into 500 μ L TMG for incubation at 37 °C for 1 h. An equal volume of overnight culture, as above, was added and the preparation was incubated for another 20 min. 500 mL LB in a 2.5 L flask was inoculated with this mixture and the flask incubated with shaking at 37 °C for 6–8 h. As lysis was sometimes not apparent, 50 μ L chloroform was

added to a small aliquot of cells to test for the presence of phage-infected cells (167). To harvest, 2.5 mL chloroform was added per 500 mL culture to destabilise and lyse all phage-infected cells. MgSO₄ and DNase I was added to a final concentration of 1 mM and 0.1 µg/mL, respectively, while the lysate was left to cool to room temperature, that is, 20–22 °C. Cell debris was removed by centrifugation at 6000 × g for 10 min at 4 °C.

Phage particles were partially purified by PEG precipitation. PEG 8000 (Amresco, biotechnology grade) and NaCl were added to 10% (w/v) and 0.5 M NaCl final concentrations, respectively. The mixture was left overnight at 4 °C and precipitate collected the following day by centrifugation at 7000 × g for 20 min at 4 °C. The pellet was resuspended in TMG with stirring over 10 min on ice and undissolved material removed by centrifugation at 6000 × g for 10 min. To prepare a concentrated stock, particles were pelleted by ultracentrifugation at 20,000 rpm for 1.5 h at 4 °C in a Type 45 Ti Rotor (Beckman Coulter) and resuspended overnight in TMG. Final titres were measured using a plaque assay.

2.2.2 Phage DNA extraction

Before DNA extraction, wild-type HK97 particles were further purified by isopycnic centrifugation using a CsCl gradient. 0.7 g CsCl was added per mL of phage suspension for a final density of ~1.4 g/mL. Following ultracentrifugation in heat-sealed tubes for 16 h at 38,000 rpm and 18 °C in a Type 70 Ti Rotor (Beckman Coulter), purified particles were extracted with a needle and syringe, and then dialysed against TMG in SnakeSkin dialysis tubing, 10K MWCO (Thermo Fisher), at 4 °C. Equal volumes of phage (~1 × 10¹² pfu/mL) and phenol, equilibrated with 10 mM Tris-Cl, pH 8.0, 1 mM EDTA (Sigma, BioReagent grade), were mixed together at room temperature. Aqueous and organic phases were separated by centrifugation at 13,000 × g for 10 min. The top aqueous phase was extracted and subjected to further rounds of phenol treatment until clear. 10 mM Tris, pH 8.5, was added to the organic fractions and undisturbed interfaces for ‘back-extraction’ of DNA. The final aqueous fractions were pooled; 0.5 volume of chloroform was added to remove the dissolved phenol, followed by centrifugation to remove the organic phase. Finally, DNA was precipitated with 0.1 volume of 3 M NaOAc, pH 5.2, and 2.5 volumes of ethanol overnight in an SS34 tube at –20 °C. Precipitated DNA was collected by centrifugation at 20,000 × g for 15 min. The pellet was washed with 70% (v/v) ethanol and finally dried at room temperature. The pellet was dissolved overnight with gentle rocking at 4 °C in TE

(10 mM Tris, pH 8.0, 1 mM EDTA). Excessive pipetting or shaking was avoided and only 1000 μ L tips were used throughout the process to avoid shearing. The quality of the DNA was confirmed by agarose gel electrophoresis. DNA was stored in aliquots at 4 °C. To separate annealed *cos* ends and generate single genome units, the DNA was heated to 65 °C for 10 min and then cooled on ice before use.

2.2.3 Cloning of the HK97 small terminase, large terminase and the *cos* region

The coding sequences of small and large terminase proteins (residues 2–161 and 1–526, respectively) were amplified from CsCl-purified wild-type HK97 phage particles and cloned into a pET22a-based vector using the ligation-independent In-Fusion Cloning system (Clontech) to generate N-terminal His-SUMO fusions. The HK97 *cos* region was recovered by extracting a 784-bp segment (–312..+472) around the *cos* cleavage site by overlap-extension PCR using wild-type phage particles, and cloned using *Bam*HI and *Eco*RI sites into a pUC18 plasmid. Q5 High-Fidelity DNA Polymerase (New England Biolabs) were used for all PCR reactions. Primers, listed in Appendix A, were obtained from Eurofins Genomics. Plasmids were propagated using chemically competent *E. coli* Stellar cells (Takara) and purified using the silica-based NucleoSpin Miniprep kit or the anion exchange chromatography-based NucleoBond Xtra Midi kit (Macherey-Nagel).

2.2.4 Overproduction and purification of HK97 small and large terminase proteins

Proteins were overproduced using transformed *E. coli* BL21(DE3) pLysS cells in LB medium containing 30 μ g/mL kanamycin and 33 μ g/mL chloramphenicol. Expression was induced at OD₆₀₀ \approx 0.8 with 0.4 mM IPTG at 37 °C for 4 h for small terminase and at 16 °C for 18 h for large terminase. Cells were harvested by centrifugation and lysed by sonication 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. The soluble fraction after centrifugation was applied to a HisTrap FF Ni Sepharose column (GE Healthcare) and the protein eluted with a 20–500 mM imidazole gradient on an ÄKTA Fast Protein Liquid Chromatography system. 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.

Small terminase was further purified by cation exchange chromatography on a Mono S 10/100 column (Amersham), eluting with a 200–1000 mM NaCl gradient in 20 mM Tris-

Cl, pH 8.0, and size exclusion chromatography on a Superdex 200 16/600 column (GE Healthcare) in 20 mM Tris-Cl, 300 mM NaCl, pH 7.5. 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. The proteins were concentrated using Vivaspin centrifugal concentrators (Sartorius) with MWCOs of 3000 and 30,000, respectively. Protein concentration was measured by absorbance at A_{280} using a NanoDrop spectrophotometer (Thermo Scientific) based on theoretical extinction coefficients of 12,490 and 52,940 $M^{-1}\cdot\text{cm}^{-1}$ for small and large terminase proteins, respectively. Purity was assessed based on the A_{260}/A_{280} ratio and by SDS-PAGE and Coomassie staining, with Mark12 Unstained Standard (Invitrogen) and PageRuler Unstained Protein Ladder (Thermo Scientific) as molecular weight markers. Small and large terminase proteins were resolved on 10% (w/v) and 18% (w/v) polyacrylamide gels (37.5:1 acrylamide:bisacrylamide), respectively.

2.2.5 Production of HK97 proheads

Proheads were produced by infection of *E. coli* 594 cells (34) (*supF58*⁻) with the HK97 amber mutant amC2 (74,168), propagated as before using LE392 cells. The amber mutation is found within the C fragment of the HK97 genome treated with *EcoRI*, and has been shown to produce empty proheads. 500 mL LB medium was inoculated in 2.5 L flasks with an overnight culture of 594 cells, grown in LB with 0.4% (w/v) maltose, to a starting OD_{600} of 0.05. Cells were infected at $OD_{600} \approx 0.4$ with 20 infectious phage particles per cell, assuming a cell density of 8×10^8 cells/mL at an OD_{600} value of 1. Cells would lyse ~90 min after infection, releasing proheads into the supernatant. Chloroform was not added to promote cell lysis to avoid triggering prohead expansion (168).

Proheads in the supernatant were precipitated by addition of NaCl and PEG 8000 to 0.5 M and 15% (w/v) final concentrations, respectively, followed by 1 h incubation on ice. Precipitate was collected by centrifugation at $10,000 \times g$ for 15 min at 4 °C. Unless specified, all subsequent steps were carried out on ice where possible, and centrifugation, at 4 °C. The pellet was resuspended in 20 mM Tris-Cl, pH 7.5, 40 mM NaCl, with stirring. Undissolved material was removed by centrifugation at $12,000 \times g$ for 15 min. To reduce the amount of PEG and low molecular weight species in solution, the dissolved proheads were further diluted for ultracentrifugation at 35,000 rpm for 2 h in a Type 45 Ti Rotor

(Beckman Coulter) followed by resuspension overnight at 4 °C in the same buffer. Undissolved material was again removed by centrifugation at $12,000 \times g$ for 15 min.

10–30% (v/v) glycerol gradients in 20 mM Tris-Cl, pH 7.5, 100 mM NaCl, were prepared in Ultra-Clear 38.5 mL tubes (Beckman Coulter) using a Gradient Master (BioComp). 3.7 mL dissolved proheads were layered on top and transferred into buckets for ultracentrifugation in a SW28 Ti rotor (Beckman Coulter) at 27,000 rpm for 2 h. Proheads, forming a band in the centre of the tube, were extracted with a needle and syringe, and dialysed against 20 mM Tris-Cl, pH 7.5, 100 mM NaCl.

Proheads were then applied to a POROS HQ 20 μm anion exchange column (Applied Biosystems) on a BioCad Perfusion Sprint Chromatography Workstation and eluted with a 100–1000 mM NaCl gradient. Fractions were pooled and diluted for ultracentrifugation at 35,000 rpm for 2 h in a Type 45 Ti Rotor. The pellet was dissolved overnight in 20 mM Tris-Cl, pH 7.5, 100 mM NaCl and stored at 4 °C. Protein concentration was measured by absorbance at A_{280} based on a theoretical extinction coefficients of $14\,204\,019\text{ M}^{-1}\cdot\text{cm}^{-1}$. Proheads were precipitated by incubation with 10% (w/v) trichloroacetic acid (TCA) and 0.04% (w/v) sodium deoxycholate (DOC) on ice, washed with 3 volumes of acetone and resuspended in SDS-containing gel running buffer. Proteins were resolved on low-crosslinking 12% (w/v) polyacrylamide gels (33.5:0.3 acrylamide:bisacrylamide) (169).

2.2.6 Production of HK97 GP74

A p11-based vector (Structural Genomics Consortium) encoding a His-GP74 fusion, cleavable with TEV protease, was a gift from Dr Karen Maxwell, University of Toronto. The GP74 protein has NCBI Reference Sequence number NP_037756.1. The protein was overproduced and purified using a similar protocol as described by Kala et al. (164). The protein was overproduced in BL21(DE3) pLysS cells grown in LB medium containing 50 μM ZnSO_4 , 100 $\mu\text{g}/\text{mL}$ ampicillin and 33 $\mu\text{g}/\text{mL}$ chloramphenicol overnight at 16 °C. Cells were lysed in 20 mM Tris-Cl, 0.5 M NaCl, 10 μM ZnSO_4 , 10% (v/v) glycerol, 20 mM imidazole, 0.05% (v/v) β -mercaptoethanol, pH 7.5, with 100 μM AEBSF, 1 μM leupeptin, 1 μM pepstatin A and 10 $\mu\text{g}/\text{mL}$ RNase A. The soluble fraction was applied to a HisTrap FF Ni Sepharose column and the protein eluted with a 20–500 mM imidazole gradient. Eluate was dialysed against 20 mM Tris-Cl, 50 mM NaCl, 0.05% (v/v) β -mercaptoethanol, pH 7.5, at 4 °C overnight, with 1:100 (w/w) TEV protease, applied to an anion exchange

MonoQ column, eluted using a 50–1000 mM NaCl gradient, and further purified using a Superdex 75 16/600 size exclusion column. The protein was concentrated using a Vivaspin centrifugal concentrator, MWCO 3000. Concentration was estimated based on a theoretical extinction coefficient of $17,990 \text{ M}^{-1}\cdot\text{cm}^{-1}$.

2.2.7 DNase protection assays

Linear plasmid DNA was prepared using FastDigest restriction endonucleases (Thermo Scientific). Unless specified, DNA was linearised using *ScaI* and the assay carried out in 20 μL reactions in packaging buffer (20 mM Tris-Cl, pH 7.5, 10 mM MgSO_4 , 30 mM potassium glutamate, 0.1 mM β -mercaptoethanol). Assays with mature phage HK97 genomic DNA were performed in packaging buffer supplemented with 4 mM spermidine. To reconstitute the packaging motor, 0.5 μg DNA (2 nM, 13.5 nM and 15 nM for HK97 genomic DNA, linear pUC18 and cos-containing pUC18 DNA, respectively) was mixed with proheads in a 1:2 molar ratio, and then large and small terminase were added to 1 μM and 2 μM final protein monomer concentrations, followed by incubation for 5 min at room temperature. Large and small terminase were diluted from concentration stocks beforehand into 20 mM Tris-Cl, pH 7.5, 10 mM MgSO_4 , 300 mM potassium glutamate, 1 mM β -mercaptoethanol, and mixed to produce a $10 \times$ pre-mix. ATP (Sigma-Aldrich), adjusted to pH 7.0 with NaOH, was added to 1 mM concentration to initiate packaging. After 30 min at room temperature, unpackaged DNA was digested by incubation with 1 $\mu\text{g}/\text{mL}$ DNase I (Roche) for 10 min. DNase was subsequently inactivated and packaged DNA was liberated from capsids by incubation with 25 mM EDTA, pH 8.0, 500 $\mu\text{g}/\text{mL}$ Proteinase K (Roche) at 65 $^\circ\text{C}$. Finally, the DNA was analysed by electrophoresis on a 0.8% (w/v) agarose gel in $1 \times$ TBE and visualised using pre-included ethidium bromide at 0.2 $\mu\text{g}/\text{mL}$ concentration and a UV gel documentation system. GP74 was added to the reactions at 1 μM concentration. 1 kb Plus DNA Ladder (Invitrogen) and 2-Log DNA Ladder (New England Biolabs) were used for size reference.

2.3 Overproduction of HK97 small terminase and large terminase

To generate full-length recombinant small terminase and large terminase, the native coding sequence of each protein was cloned for expression in *E. coli* as a His-SUMO-fusion construct. This design offers three advantages: the presence of SUMO enhances solubility;

an N-terminal hexahistidine tag enables purification by affinity chromatography; lastly, removal of SUMO with SUMO protease produces protein with a native N-terminus.

Both proteins were produced at levels detectable by SDS-PAGE and were soluble in buffers containing 1 M NaCl (Figure 2.1a, 2.2a). Following nickel-affinity chromatography, the NaCl concentration could be reduced to 200 mM without precipitation of the protein. The proteins were treated with SUMO protease and then further purified by ion exchange chromatography, according to their theoretical pI, and size exclusion chromatography (Figure 2.1b, 2.1c, 2.2a).

Cleavage of the His-SUMO large terminase fusion was never complete, likely due to an inaccessible N-terminus. In attempt to improve cleavage efficiency, a range of additives were screened. Greater efficiencies were achieved in the presence of EDTA, RNaseA, glycerol and urea, but the reactions still did not go to completion. Nevertheless, the cleaved protein could be separated from uncleaved protein, because it would elute earlier during anion exchange chromatography and later during size exclusion chromatography.

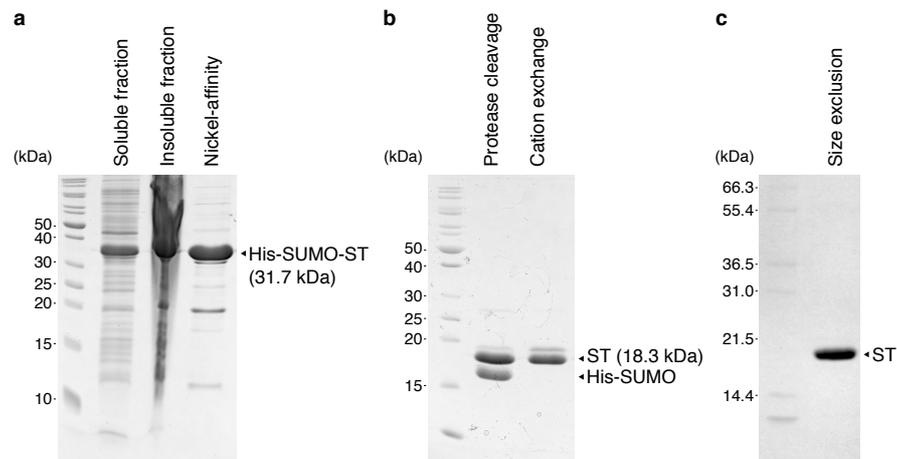


Figure 2.1. Purification of HK97 small terminase. (a) Small terminase overexpressed as a His-SUMO fusion (His-SUMO-ST) was solubilised in 20 mM Tris-Cl, 1 M NaCl, 10% (v/v) glycerol, 20 mM imidazole, 0.05% (v/v) β -mercaptoethanol, pH 8.0, purified by nickel-affinity chromatography. Samples were analysed by SDS-PAGE. (b) Cleavage by SUMO protease and purification by cation exchange chromatography followed by size exclusion chromatography in 20 mM Tris-Cl, 300 mM NaCl, pH 7.5, yielded > 90% pure protein.

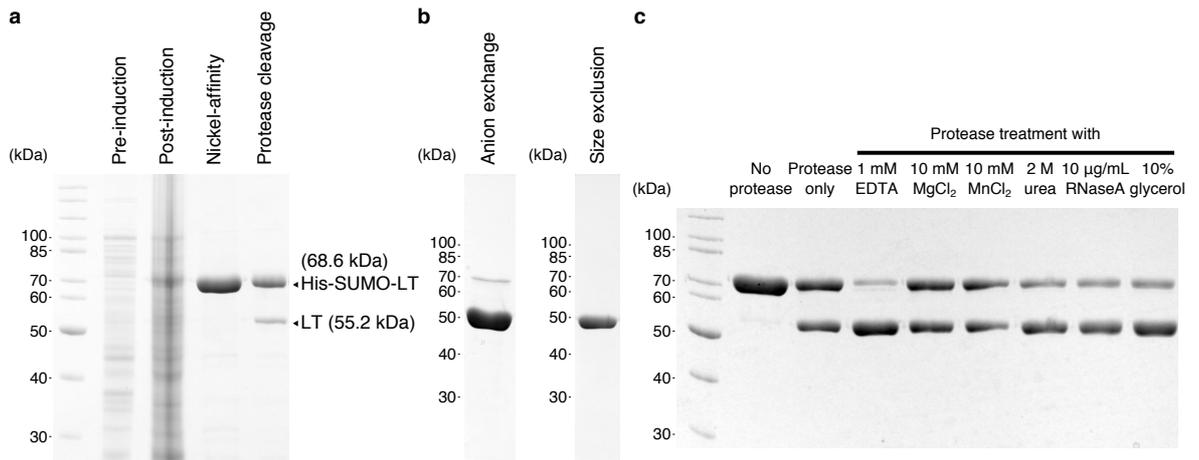


Figure 2.2. Purification of HK97 large terminase. (a) Large terminase overproduced as a His-SUMO fusion (His-SUMO-LT) was solubilised in 20 mM Tris-Cl, 1 M NaCl, 10% (v/v) glycerol, 20 mM imidazole, 0.05% (v/v) β -mercaptoethanol, pH 8.0, and purified by nickel-affinity chromatography, treated with SUMO protease and further purified by (b) anion exchange chromatography and size exclusion chromatography in 20 mM Tris-Cl, 300 mM NaCl, 2 mM DTT, pH 7.5. (c) Screening of additives for SUMO protease cleavage.

2.4 Isolation of empty HK97 proheads

Empty proheads can be overproduced in *E. coli* using a plasmid, pVP0, which contains the native bacteriophage sequence spanning the portal protein, the capsid protein and capsid maturation protease coding regions under control of a T7 promoter. While recombinant expression simplifies the production process, it gives rise to a sub-population of non-portal-containing proheads, formed from self-assembling capsid protomers. This can be recognised by a lower ratio of portal protein in SDS-PAGE analysis and slower capsid expansion kinetics (R. Duda, personal communication). So far, it has not been possible to separate non-portal-containing proheads from the 13.3-MDa portal-containing proheads, as they differ in mass by only 0.4 MDa and have similar solubilities and charge properties. The presence of non-portal-containing proheads renders a prohead preparation less active in DNA packaging experiments and is non-ideal for structural and biophysical studies.

A more reliable method to generate empty portal-containing proheads is to infect *E. coli* cells with an amber bacteriophage mutant. Protein production is likely better tuned during infection, resulting in higher levels of portal incorporation. Amber mutants defective in DNA packaging have been isolated from a hydroxylamine genetic screen (168). The amC2 mutant contains an amber mutation in the C fragment of the HK97 genome treated with *EcoRI*. The mutant would propagate normally in *supF⁺* cells, where the amber codon is

sometimes translated due to expression of a tyrosine amber suppressor tRNA. In *supF*⁻ cells infected with the mutant, proheads are produced instead. Infected *supF*⁻ cells would lyse, releasing proheads into the supernatant. The proheads could then be purified by PEG precipitation, followed by rate-zonal density-gradient centrifugation, anion exchange chromatography, and finally sedimentation by ultracentrifugation (Figure 2.3).

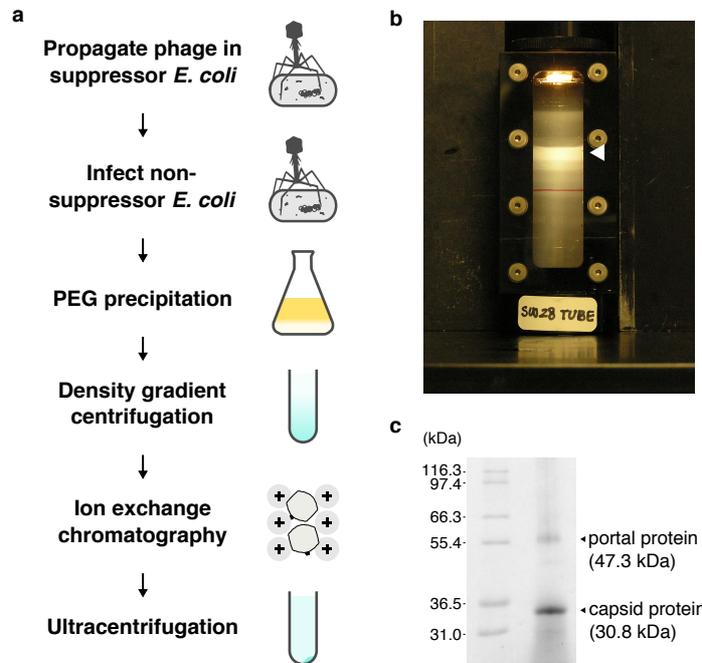


Figure 2.3. Production of empty HK97 proheads. (a) Protocol for prohead production. (b) Prohead band in 10–30% (v/v) glycerol gradient in 20 mM Tris-Cl, pH 7.5, 100 mM NaCl, after ultracentrifugation for 2 h in a SW28 Ti rotor at 27,000 rpm. White arrowhead indicates the prohead fraction. (c) Purified proheads TCA-precipitated for SDS-PAGE analysis. TCA precipitation prevents isopeptide bond formation between capsid proteins during incubation in SDS.

2.5 Reconstitution of the DNA packaging motor

To reconstitute the DNA packaging motor, substrate DNA was incubated with empty proheads in a 1:2 molar ratio and at least 5- and 10-fold excess of large terminase and small terminase, respectively, in the presence of magnesium and ATP. DNA packaging activity was measured using a DNase protection assay (Figure 2.4a), where packaged DNA is protected inside the capsid and unpackaged DNA is degraded (110). After treatment with DNase, EDTA and proteinase K are added to disable the DNase and to release the packaged DNA from capsids for agarose gel electrophoresis. For comparison, an ‘input’ sample is included. This ‘input’ sample is the same packaging reaction but not DNase-treated. It allows the detection of cleavage events that might have occurred during packaging. To ensure that DNase treatment is effective in removing unpackaged DNA, ATP is omitted from the packaging reaction and the reaction is treated with DNase. This reaction is annotated as ‘negative’ in Figure 2.4 and subsequent figures.

The assay shows that the reconstituted motor was active and could package mature HK97 genomic DNA into proheads (Figure 2.4b). The motor could also package linearised plasmid DNA containing a 784-bp insert from the HK97 *cos* region, which spans the genome boundary as seen in a concatemer from rolling-circle replication (Figure 2.4c). However, the pattern of protection was unexpected. The absence of smaller fragments in the ‘input’ indicates that there has been no cleavage of the DNA during packaging. This suggests that packaging initiated from a DNA end rather than the embedded *cos* region. As a result, the entire substrate could be packaged. The presence of smaller protected fragments, whose lengths coincide approximately with the distances between the ends of the substrate and the *cos* cleavage site, suggests that the motor could start from either end of the substrate but sometimes stalled at the *cos* region.

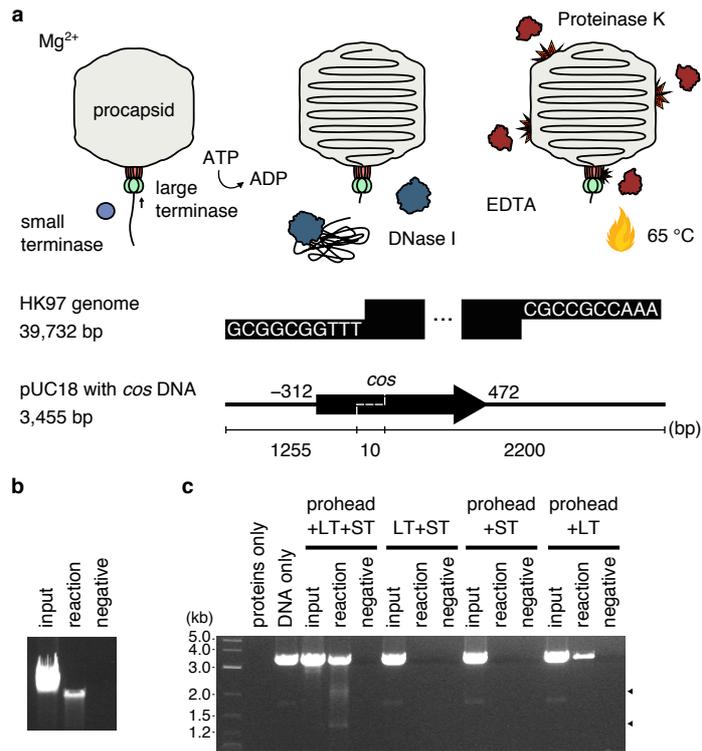


Figure 2.4. Reconstitution of the HK97 DNA packaging motor. (a) DNA (0.5 μ g, 2–15 nM) was incubated with proheads in a 1:2 molar ratio, 1 μ M large and 2 μ M small terminase monomers, magnesium and ATP. Unpackaged DNA was removed by DNase treatment. Packaged DNA was subsequently released by protease treatment and heating, DNase deactivated with EDTA, and the DNA analysed by agarose gel electrophoresis. Mature phage HK97 DNA and *ScaI*-linearised pUC18 DNA with a 784-bp insert from around the HK97 *cos* cleavage site were used as DNA substrates for packaging. (b) Packaging of mature phage DNA. The input reaction received ATP but was not treated with DNase. The negative reaction had no ATP but was treated with DNase. (c) Packaging of linear, *cos*-containing pUC18 DNA. Proheads, large terminase (LT) and small terminase (ST) were removed individually from different reactions.

The lack of protection when proheads were omitted confirms that DNA is packaged inside proheads when protected. Though small terminase had a stimulatory effect, it was not essential for packaging activity *in vitro*. Likewise, removal of the *cos* region, the binding site of small terminase for packaging initiation, did not affect activity (Figure 2.5). This is in line with the idea that the motor was initiating from either end of the substrate. Furthermore, the motor did not distinguish between blunt DNA ends and sticky ends, which have 4-nt 5' or 3' overhangs, or 10-nt 3' overhangs as found in the mature HK97 genome. This property of the motor indicates an uncoupling of DNA translocation from packaging initiation in the *in vitro* system, but has enabled further characterisation of the translocation machinery.

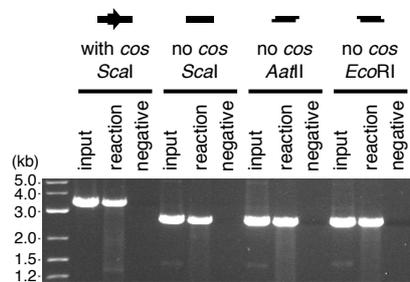


Figure 2.5. Packaging in the absence of *cos*. Linear pUC18 backbone DNA, without the *cos* insert, was used as the substrate. The DNA was linearised using *ScaI*, *AatII* and *EcoRI* to generate blunt ends, and sticky ends with 3' and 5' overhangs, respectively. DNA was analysed by agarose gel electrophoresis.

2.6 Biochemical characterisation of the DNA packaging motor

DNA translocation in viral packaging motors is coupled to ATP hydrolysis in the large terminase subunit. As such, the dependence of the motor on divalent metal ions was investigated. To remove all traces of divalent metal ions before the assay, individual components were incubated with Chelex 100 resin, which contains cation-chelating iminodiacetate functional groups. The motor was reconstituted in the presence of magnesium, manganese, or zinc, or a combination thereof, and finally supplemented with magnesium before DNase treatment. Both magnesium and manganese supported DNA packaging (Figure 2.6a). Zinc did not support packaging and even inhibited activity when magnesium was present. These results were not unexpected because ATPase activity is generally sensitive to the geometry and charge distribution of the metal-ATP complex within the active site.

Further analysis shows that the motor is active across pH range 6–9 (Figure 2.6b). The intracellular pH of *E. coli* is maintained between 7.4 and 7.8, given an external pH range of 5.5–9 (170). Hence, the motor is relatively robust against acidity changes inside the cell. The HK97 motor tolerates up to 300 mM potassium glutamate in solution but only up to 150 mM sodium chloride (Figure 2.6c). This differential sensitivity is reminiscent of other nucleotide-dependent, DNA-acting machines such as the λ packaging motor (171) and DNA polymerase III (172). It has been noted that proteins are more stable in the presence of potassium *in vitro* (173) and that glutamate preserves protein-DNA interactions at higher salt concentrations (174). For phage HK97, this apparent preference may represent an adaptation to the intracellular environment of *E. coli* during log phase, when potassium and glutamate ions are in great abundance (175,176) and resources are available for virus assembly.

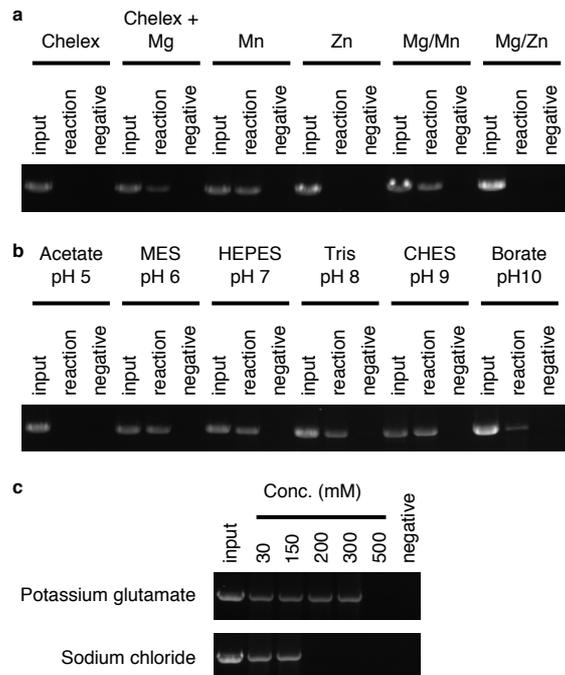


Figure 2.6. The effects of salts and pH on *in vitro* packaging. (a) Divalent metal ions were sequestered by incubation with Chelex resin before motor assembly after which metal was added again. (b) Reactions were carried out in buffers at different pH values. (c) Reactions were carried out at different monovalent salt concentrations, ethanol-precipitated and re-dissolved in an identical gel running buffer for electrophoresis. *ScaI*-linearised, *cos*-containing pUC18 DNA was used for all reactions.

Polyamines such as spermidine and putrescine have been found in mature bacteriophages where they serve as counterions to the densely packed negatively charged DNA (177). *In vitro*, polyamines are known to facilitate packaging of the lambda genome by the lambda motor (33). Assay with the *in vitro*-assembled HK97 motor shows that a polyamine such as spermidine is essential for packaging of the HK97 genome (Figure 2.7a). For activity, a spermidine concentration between 2 and 6 mM was optimal, while higher concentrations led to precipitation of the DNA. By DNase treatment at different time points after addition of ATP and comparison of the protected fragment to a DNA ladder, the speed of packaging at 4 mM spermidine concentration was approximately 600 bp/s (Figure 2.8). Polyamines appear to be dispensable for the packaging of short DNA substrates (Figure 2.7b). These substrates would not fill the capsid and were therefore unlikely to be packed as tightly as the genome. Divalent and monovalent cations in solution probably sufficed in neutralising charges on the shorter DNA.

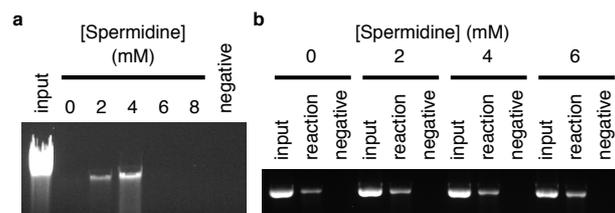


Figure 2.7. The effects of spermidine on *in vitro* packaging. Packaging of (a) mature phage HK97 DNA and (b) linear, *cos*-containing pUC18 DNA.

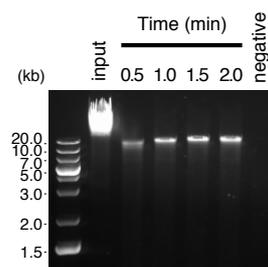


Figure 2.8. The speed of the HK97 DNA packaging motor. Reactions containing mature phage HK97 DNA were treated with DNase at different time points following the addition of ATP.

2.8 An endonuclease co-factor for packaging initiation and termination

Part of the packaging initiation and termination processes is the cleavage of *cos* signals in a DNA concatemer. Phage HK97 encode a HNH endonuclease-like protein, GP74, which has been shown (164). *In vitro*, the protein stimulates cleavage of *cos* DNA by large terminase and small terminase. Empty proheads are produced *in vivo* when the protein is mutated.

Addition of recombinant GP74 during reconstitution of the motor on a *cos*-containing substrate gave rise to smaller fragments in the 'input'. The sizes of these fragments indicate that there had been cleavage at the embedded *cos* region (Figure 2.10a). This cleavage led to a more discrete pattern of DNase protection, particularly in the 2.2-kb region, where non-uniform fragments were observed previously. However, it is unclear whether cleavage took place before or after DNA translocation and from where packaging initiated. Packaging might have initiated from the internal *cos* site, with cleavage, and proceeded in the forward direction. This would produce a ~2.2-kb fragment. The reduction of non-uniform fragments in the 2.2-kb region in the presence of GP74 supports this interpretation. Packaging might also have initiated from a DNA end. Stalling or cleavage after stalling when the *cos* site was encountered in the forward direction would produce a ~1.3 kb fragment. Cleavage and translocation might also be uncoupled. The DNA might be cut first, released, and then fragments were individually packaged.

2.9 Discussion

The reconstitution of a functional packaging motor for *in vitro* characterisation requires the ability to produce large and small terminase proteins, and empty proheads. Using a T7 expression system, recombinant full-length large and small terminase proteins were produced and purified (Figures 2.1, 2.2). Proheads were produced by infection of suppressor-free *E. coli* cells with an amber phage mutant (Figure 2.3).

In the presence of magnesium, ATP and a free DNA end, a minimal motor could be assembled using proheads and large terminase protein only (Figure 2.4). Neither small terminase nor the *cos* site was required for the packaging of free linear DNA *in vitro* (Figures 2.4, 2.5). However, small terminase did improve the efficiency of packaging, suggesting that it does have a role, likely during packaging initiation. As with phage T4, the question is raised whether small terminase participates only during initiation or remains a part of the motor during translocation. Under ATP-limited conditions, the motor would stall upon encountering an internal *cos* site in the forward direction (Figure 2.9). The stalling was dependent on small terminase. As such, small terminase bound to the *cos* site likely acted as an obstacle for the packaging motor. This indicates a role for small terminase in termination. However, whether the stalling was due to the same small terminase molecule that initiated packaging and remained bound to the motor, or a different small terminase molecule interacting with the downstream *cos* site is unclear. The lack of DNA cleavage despite stalling suggested that the capsid was not filled enough to trigger a conformational rearrangement to enable cleavage, as was the case for some λ mutants (139), or that a co-factor was required for endonuclease activity, for example, GP74.

GP74 is an HNH endonuclease. Its coding sequence is located at the right end of the mature genome, which in the context of closed circular or concatemeric DNA places it next to the small terminase coding sequence. The protein has been suggested to aid in the cleavage of *cos* DNA (164). Packaging and cleavage assays with GP74 showed that the protein led to greater processing efficiency at the *cos* site but at the same time gave rise to a low level of non-*cos*-specific cuts (Figure 2.10). At physiological ATP concentrations *in vitro*, the DNA substrate used might be protected from non-specific cuts because the DNA was packaged faster than it could be cleaved. *In vivo*, host proteins might decorate the viral genome, conferring some protection against non-specific DNA cleavage. Given its non-

specific activity and the proximity of its gene to lysin and holin genes, GP74 perhaps also facilitates cell lysis during the lytic cycle of phage HK97.

Further biochemical characterisation showed that manganese would also support packaging, that the motor was sensitive to sodium chloride but tolerant of potassium glutamate (Figure 2.6), and that the motor translocates DNA at approximately 600 bp/s (Figure 2.8). The speed of packaging is comparable to that of the λ system *in vitro*, again correlating with genome length. Knowledge of the biochemical properties of the motor and the ability to uncouple DNA translocation from initiation by supplying a DNA substrate with free ends have facilitated further characterisation of the motor as discussed in the following chapters.

3. Structure and function of the HK97 small terminase

3.1 Introduction

The small terminase proteins of *cos* and *pac* viruses recognise and recruit large terminase to the *cos* and *pac* sites of newly replicated viral DNA, respectively, to initiate packaging. Biochemical data in the previous chapter suggest that the small terminase is also important for termination in a *cos* system. In phage λ , the maturation complex for initiation consists of eight small terminase and four large terminase subunits (93). The λ small terminase has N-terminal winged helix-turn-helix domains that are thought to dimerise for *cos* DNA binding (95). In contrast, *pac* small terminase proteins often contain an odd number of subunits. Their wingless N-terminal helix-turn-helix (HTH) domains do not dimerise.

Despite the difference in quaternary arrangement in the small terminase protein, DNA, whether *cos* or *pac*, is thought to encircle the small terminase during packaging initiation. It is unclear whether the HK97 small terminase contains the HTH elements that would allow DNA to encircle it. Furthermore, the full-length structure of a *cos* small terminase has never been determined. The arrangement of HTH elements in a *cos* protein is also unclear. Therefore, the structure of the HK97 small terminase was determined by crystallography. The protein surprisingly has wingless HTH elements. Because the DNA-binding properties of a *pac* HTH element have not been investigated directly, the N-terminal domain of *pac* phage SF6 was characterised using SPR and NMR. Finally, the cognate DNA sequence within the genome for HK97 small terminase was searched for in preparation for future structural and biophysical studies.

3.2 Materials and methods

3.2.1 Protein and DNA production

The HK97 small terminase was produced as described in the previous chapter. The SF6 N-terminal domain was overproduced as a N-terminal His-tagged fusion in *E. coli* and purified by nickel-affinity, cation exchange and size exclusion chromatography as described in (178). For isotope labelling, transformed cells were grown in M9 minimal medium supplemented with vitamins and trace metals (179), 30 $\mu\text{g/mL}$ kanamycin, and 0.5 g/L $^{15}\text{NH}_4\text{Cl}$ and 3 g/L $\text{U-}^{13}\text{C}$ -glucose as required. Protein expression was induced by

addition of IPTG to a final concentration of 1 mM and shaking for 18 h at 16 °C. DNA for electrophoretic mobility shift assays was amplified from a vector containing HK97 *cos* DNA, as described previously, using Q5 DNA polymerase (New England Biolabs), with a 10 s extension time, 15 s annealing time, and annealing temperatures at 63 °C, 58 °C and 63 °C for fragments I, II and III, respectively. The amplified DNA was purified using the GeneJET PCR Purification Kit (Thermo Fisher) and quantified based on absorbance at 260 nm and theoretical extinction coefficient calculated from the sequence according to Cavaluzzi and Borer (180). For co-crystallisation, unmodified DNA oligonucleotides (0.6 mM) were phosphorylated using 2 U/μL T4 polynucleotide kinase (New England Biolabs) in T4 DNA ligase buffer supplemented with 10 mM ATP, and then annealed by heating with the complementary oligonucleotide to 95 °C for 5 min and cooling slowly in the heat block. The double-stranded pieces were ligated in a one-pot overnight reaction at 16 °C with fresh 10 mM ATP, 0.8 mM DTT and 30 U/μL T4 DNA ligase at a final concentration of 80 μM. The DNA was purified by size exchange chromatography using a Superose 6 10/300 column in 20 mM HEPES, 100 mM potassium glutamate, pH 7.0, followed by anion exchange chromatography on a MonoQ 5/50 column with a 0.4–1 M potassium glutamate gradient. Eluate was concentrated and buffer-exchanged using a Vivaspinn centrifugal concentrator, MWCO 3000 (Sartorius). Purity of the DNA was assessed by electrophoresis on a 12% (w/v) polyacrylamide gel (37.5:1 acrylamide:bisacrylamide) alongside a GeneRuler Ultra Low Range DNA ladder (Thermo Fisher).

3.2.2 Size exclusion chromatography with multi-angle laser-light scattering analysis

The HK97 small terminase (100 μL 2.7 mg/mL) was applied to a Superdex 200 10/300 GL column (GE Healthcare) in 20 mM HEPES, 300 mM KCl, 1 mM DTT, pH 7.0. Eluate was analysed using an on-line Dawn HELEOS-II light scattering detector and Optilab rEX refractometer (Wyatt). Data were processed based on a refractive index increment of 0.183 mL/g and scattering intensities at seven different angles in ASTRA 5.3.4 (Wyatt).

3.2.3 Native mass spectrometry

Small terminase was buffer-exchanged into 200 mM ammonium acetate, pH 7.5, using Micro Bio-Spin polyacrylamide P-6 columns (Bio-Rad). Ammonium acetate is a volatile and, hence, non-adduct-forming buffer solution. Protein was delivered in gold-plated capillary needles for analysis on a custom built Q-ToF2 mass spectrometer (Waters) with

capillary voltages at 1–2 kV and cone voltages up to 100 V. All mass spectrometry data were acquired and processed by Dr Joseph Gault at the University of Oxford.

3.2.4 Crystallisation and structure determination

Crystallisation conditions were screened by sitting drop vapour diffusion using commercially available screens (Hampton Research, Qiagen, Molecular Dimensions). Reservoir solution (54 μ L) was transferred using a Hydra robot (Matrix Technologies) into MRC 2-well plates. Protein and reservoir solution were mixed 1:1 in 300 nL drops using a Mosquito robot (TTP Labtech). Final crystals of the HK97 small terminase (22 mg/mL in 20 mM HEPES, pH 7.0, 300 mM potassium glutamate) were obtained in 2 μ L hanging drops against 500 μ L 0.1 M succinic acid, pH 7.0, 9% (w/v) PEG 3350 in 24-well cell culture plates (Greiner Bio-One) at 20 °C. Glass coverslips were siliconised using AquaSil (Thermo Scientific) and sealed over wells using vacuum grease (Dow Corning). Crystals were soaked in 10 mM HEPES pH 7.0, 50 mM succinic acid pH 7.0, 18% PEG 3350, 15 % (v/v) glycerol, with 250 mM or 500 mM KI, for 5 min. Data were collected at 100 K with exposure times of 0.04 s in 0.1° oscillations over 200° at Diamond Light Source beamlines I02 and I03, integrated and scaled with XDS (181). Phases were determined by substructure solution in SHELXD (182), phase calculation in Phaser (183) and density modification in Parrot (184). A partial model built using Buccaneer (185) was refined against data collected from a crystal soaked with 250 mM KI, which diffracted best. The model was completed and refined using Coot (186) and REFMAC (187), respectively. Molecular graphics were generated using PyMol (Schrödinger). Electrostatic potentials were calculated using APBS (188,189) under the SWANSON force field (190). Buried surface areas were calculated using PISA (191). For co-crystallisation, protein and DNA were incubated for 30 min at room temperature at 24.2 and 26.6 μ M final concentrations, respectively, before mixing with reservoir solution in a 1:1 ratio for screening by sitting drop vapour diffusion.

3.2.5 Surface plasmon resonance

DNA was immobilised on a low-density streptavidin SAD500L chip (Xantec) by injection of a 5'-biotinylated oligonucleotide (100 nM) followed by the complementary strand (10 μ M) in 20 mM HEPES, 150 mM KCl, pH 7.0, 0.05% (v/v) Tween 20 in a Biacore T200 instrument (GE Healthcare). Protein (0.049–1500 μ M) in the same buffer was injected in

triplicates at 30 $\mu\text{L}/\text{min}$ over the chip surface, the surface washed with 1 M KCl before each injection. Responses in an empty upstream flow cell were subtracted from responses in the test cell to correct for bulk refractive index changes. Data were analysed under the Tsodikov-modified McGhee–von Hippel model (192,193):

$$\frac{v}{L} = K_a (1 - nv) \left(\frac{1 - nv}{1 - (n-1)v} \right)^{n-1} \left(\frac{N - n + 1}{N} \right)$$

and transformed as follows:

$$v = \frac{M_{\text{DNA}}}{NM_{\text{protein}}} \cdot \frac{R_{\text{eq}}}{R_L}$$

Least squares fitting was performed in R (194). The residual sum of squares in v were minimised for each integer value of n using the L-BFGS-B method (195).

3.2.6 Nuclear magnetic resonance

NMR spectra were recorded at 600 μM protein concentration at 25 $^{\circ}\text{C}$ in 20 mM MES, 150 mM KCl, pH 6.0, 10% (v/v) D_2O , on a Bruker AVANCE II 700 MHz spectrometer with a triple-resonance probe. For spectral assignments, HNCO (196), HN(CA)CO (197), CBCANH (198) and CBCA(CO)NH (199) spectra of uniformly ^{13}C , ^{15}N -labelled protein were acquired with 50% overall non-uniform sampling (NUS) in the indirect dimensions. ^{13}C -decoupled ^{15}N - ^1H TOCSY-HSQC (200) and ^{15}N - ^1H NOESY-HSQC spectra (201) were also acquired, with mixing times of 60 ms and 160 ms, respectively, and 40% overall NUS in the indirect dimensions. Data were reconstructed and processed using mddNMR v2.4 (202) and NMRPipe (203), respectively, and referenced to 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS; 0.1 M). Spectra assignments were carried out in CcpNmr Analysis v2.4.1 and deposited under BMRB entry 25876. For chemical shift perturbation studies, DNA and protein were dialysed against the same buffer. ^1H - ^{15}N HSQC spectra (204) were acquired as protein (300 μM) was titrated against DNA from a 7.5 mM stock. Combined chemical shift difference magnitudes were given by

$$\text{combined } \Delta\delta = \sqrt{\Delta\delta_{\text{H}}^2 + (\alpha \cdot \Delta\delta_{\text{N}})^2}$$

where $\alpha = 0.2$ for glycine and $\alpha = 0.14$ for all other residues (205).

3.2.7 Electrophoretic mobility shift assay

Composite gels containing 3% (w/v) polyacrylamide (100:1 acrylamide/bisacrylamide) and 0.5% (w/v) agarose in running buffer (25 mM Tris-Cl, 192 mM glycine, pH 8.3, 100 mM Na₂SO₄, 10 mM MgSO₄) were prepared by mixing the components together at 45 °C and leaving to cool in empty mini 1.5-mm Novex gel cassettes. DNA (50 nM) was incubated with protein (0.08–10 μM) in running buffer at room temperature for 30 min. Electrophoresis was performed in a XCell SureLock Mini-Cell (Thermo Fisher) at 10 mA (~20 V), 4 °C, for 18 h. DNA was visualised by staining in 0.4 μg/mL ethidium bromide in running buffer.

3.3 Oligomeric state of the HK97 small terminase

The oligomeric state of small terminase is important for understanding how the viral genome is recognised and how higher-order assemblies involving small terminase and large terminase are organised during packaging initiation and termination. The elution profile of the HK97 small terminase during size exclusion chromatography suggests that the protein is an oligomer in solution. In order to determine the oligomeric state, the molecular weight of the protein was measured using two complementary methods.

During size exclusion chromatography with multi-angle laser-light scattering (SEC-MALLS) analysis, the majority of protein eluted in one monodisperse peak with a weight-average molecular weight of 160.5 kDa (Figure 3.1a). A smaller, later peak was observed but with little light scattering and no detectable traces of protein by SDS-PAGE and Coomassie staining. Similarly, by native electrospray ionisation mass spectrometry, only one population of protein was observed, measuring 165.020 ± 0.0045 kDa in mass (Figure 3.1b). Given a monomeric mass of 18.3 kDa, the observed masses corresponded to 8.8 and 9.0 subunits, respectively.

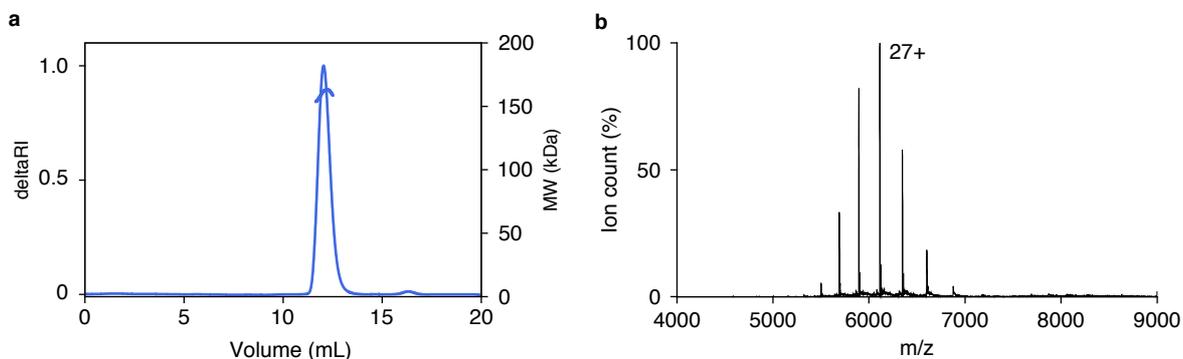


Figure 3.1 Molecular weight measurements of the HK97 small terminase. (a) Size-exclusion chromatography and multi-angle laser-light scattering analysis using a Superdex 200 10/300 column in 20 mM HEPES, 300 mM KCl, 1 mM DTT, pH 7.0. (b) Native mass spectrum of the small terminase in 200 mM ammonium acetate, pH 7.5. A charge envelope centred around the $[M+27H]^{27+}$ ion was observed.

3.4 Crystal structure of the HK97 small terminase

3.4.1 Crystallisation and structure determination

The oligomeric state of the HK97 small terminase could also be deduced *in crystallo*. Crystals of the protein were first obtained in a sparse matrix screen by sitting drop vapour diffusion against 0.1 M succinic acid, pH 7.0, 15% PEG 3350. Further optimisation with larger drop volumes and hanging drop vapour diffusion yielded larger, better diffracting crystals (Figure 3.2a), which were used in the final diffraction experiments (Figure 3.2b). All crystals indexed in space group $R3$ in the hexagonal setting.

Taking the Fourier transform of observed intensities to generate a Patterson map and calculating a self rotation function (206), nine peaks could be seen in the $\kappa = 180^\circ$ section, including an origin peak and two peaks of equal height due to crystallographic symmetry (Figure 3.2c). The appearance of six additional peaks suggests a 3-fold rotational non-crystallographic (NCS) symmetry in the asymmetric unit.

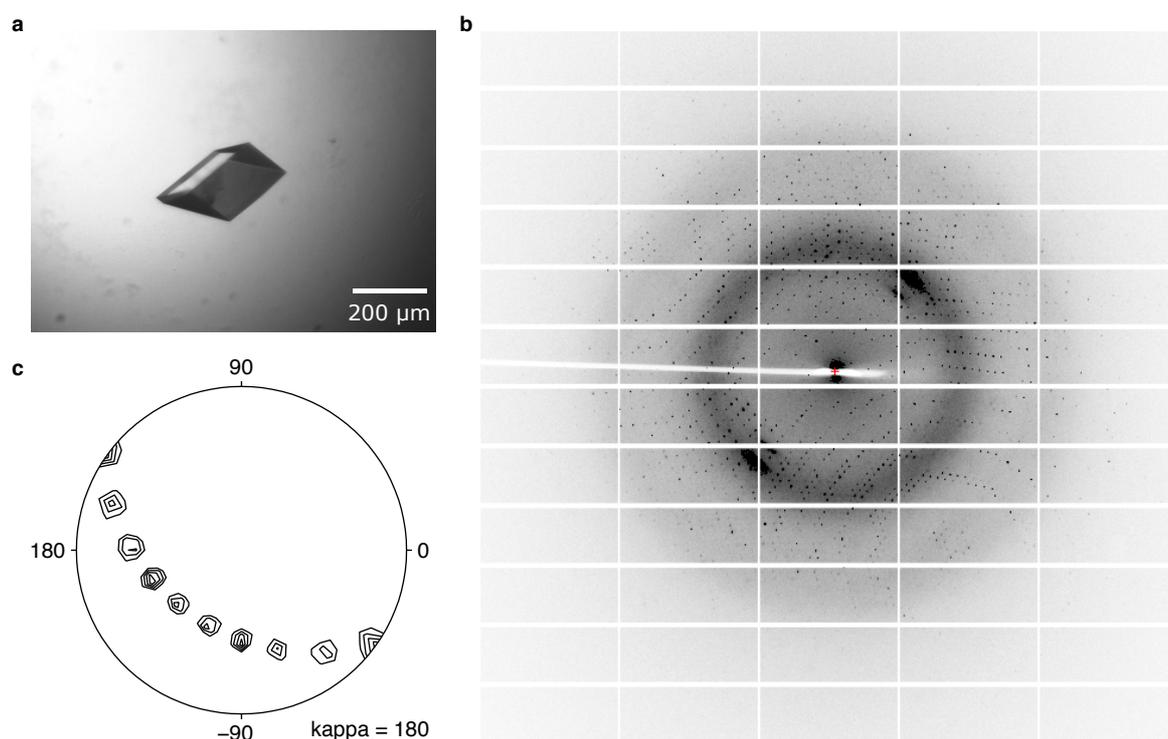


Figure 3.2. Crystallisation and diffraction experiment. (a) Crystal of the HK97 small terminase obtained by hanging drop vapour diffusion against 0.1 M succinic acid, pH 7.0, 9 % PEG 3350. (b) Diffraction image recorded on a Pilatus 6MF detector at Diamond Light Source. Resolution corresponded to 1.4 Å at the edges of the detector. (c) Section of the self-rotation function in polar coordinates, $\kappa = 180^\circ$, calculated using POLARREN (207).

Molecular replacement using available small terminase structures did not yield a clear solution. Solutions from Phaser all had low translation function Z-scores and discontinuous densities outside the search model. Thus, phases were determined experimentally. A halide soak was performed as described by Dauter et al. (208). Cracks appeared when crystals were soaked in cryoprotectant containing 1 M KI with loss of diffraction. Crystals in 500 mM or 250 mM KI remained intact and diffracted well in-house. Data were subsequently collected from these crystals using synchrotron radiation (Table 3.1).

The absorption edges of iodine cannot be accessed easily at most synchrotron beamlines. However, single-wavelength anomalous diffraction (SAD) phasing is possible because there is still significant anomalous scattering at the longer wavelengths accessible. High energy remote data were collected at 1.6984 Å wavelength, where the theoretical imaginary contribution is 7.9 e⁻ (Figure 3.3a). For high multiplicity and minimum radiation damage, data were collected in a helical fashion using a highly attenuated beam and short exposure times of 0.04 s in 0.1° oscillations over 200°.

Data were truncated based on CC_{1/2} (Figure 3.3b). Good correlation in anomalous differences between half sets (Figure 3.3c) and anomalous signal (Figure 3.3d) were observed at least to 2.3 Å resolution. Substructure solution using SHELXD (Figure 3.3e,f) identified 16 iodide sites with at least 0.25 occupancy. The substructure was refined and phases were calculated using Phaser. Density modification using Parrot led to identification of the correct hand (Figure 3.3g,h), with NCS averaging contributing greatly to the improvement in phases. A partial structure built using Buccaneer refined to R_{work} and R_{free} values of 30.5% and 33.0%, respectively.

Higher-resolution data were collected from the 250 mM KI crystal at a shorter wavelength away from the absorption edge to reduce the absorbed dose and thereby radiation damage (Figure 3.3a). Phases were calculated using the partial structure obtained in the SAD experiment. Manual model building in Coot and reciprocal-space refinement using REFMAC produced a final structure with R_{work} and R_{free} values of 16.7% and 18.1%, respectively (Figures 3.4, 3.5).

Table 3.1. Crystallographic data collection and refinement statistics.

	500 mM KI soak	250 mM KI soak
Data collection		
Wavelength (Å)	1.6984	0.9763
Space group	<i>R</i> 3	<i>R</i> 3
Cell dimensions		
<i>a</i> , <i>c</i> (Å)	123.3, 79.4	123.3, 79.0
Resolution (Å)	44.31–1.79 (1.83–1.79)	44.23–1.40 (1.43–1.40)
<i>R</i> _{merge}	0.080 (0.539)	0.049 (0.707)
<i>I</i> / $\sigma(I)$	15.2 (1.4)	10.3 (1.1)
CC _{1/2} (%)	99.6 (51.5)	99.7 (52.5)
CC _{anom} (%)	69.5 (23.2)	21.1 (34.9)
Completeness (%)	95.3 (47.3)	98.8 (88.6)
Multiplicity	4.4 (1.2)	2.7 (2.0)
Wilson B (Å ²)	19.0	18.6
No. unique reflections	40412 (1184)	86871 (3870)
Refinement		
Resolution (Å)		44.23–1.40
No. reflections		
Working		82402
Free		4469
<i>R</i> _{work} / <i>R</i> _{free}		0.159/0.170
No. atoms		
Protein		2474
Water		338
Ligand		9
B factors (Å ²)		
Protein		21.9
Water		33.0
Ligand		30.9
R.m.s. deviations		
Bond lengths (Å)		0.011
Bond angles (°)		1.4
Ramachandran		
Favoured (%)		100.0
Outlier (%)		0.0

† Values in parentheses correspond to the highest resolution shell.

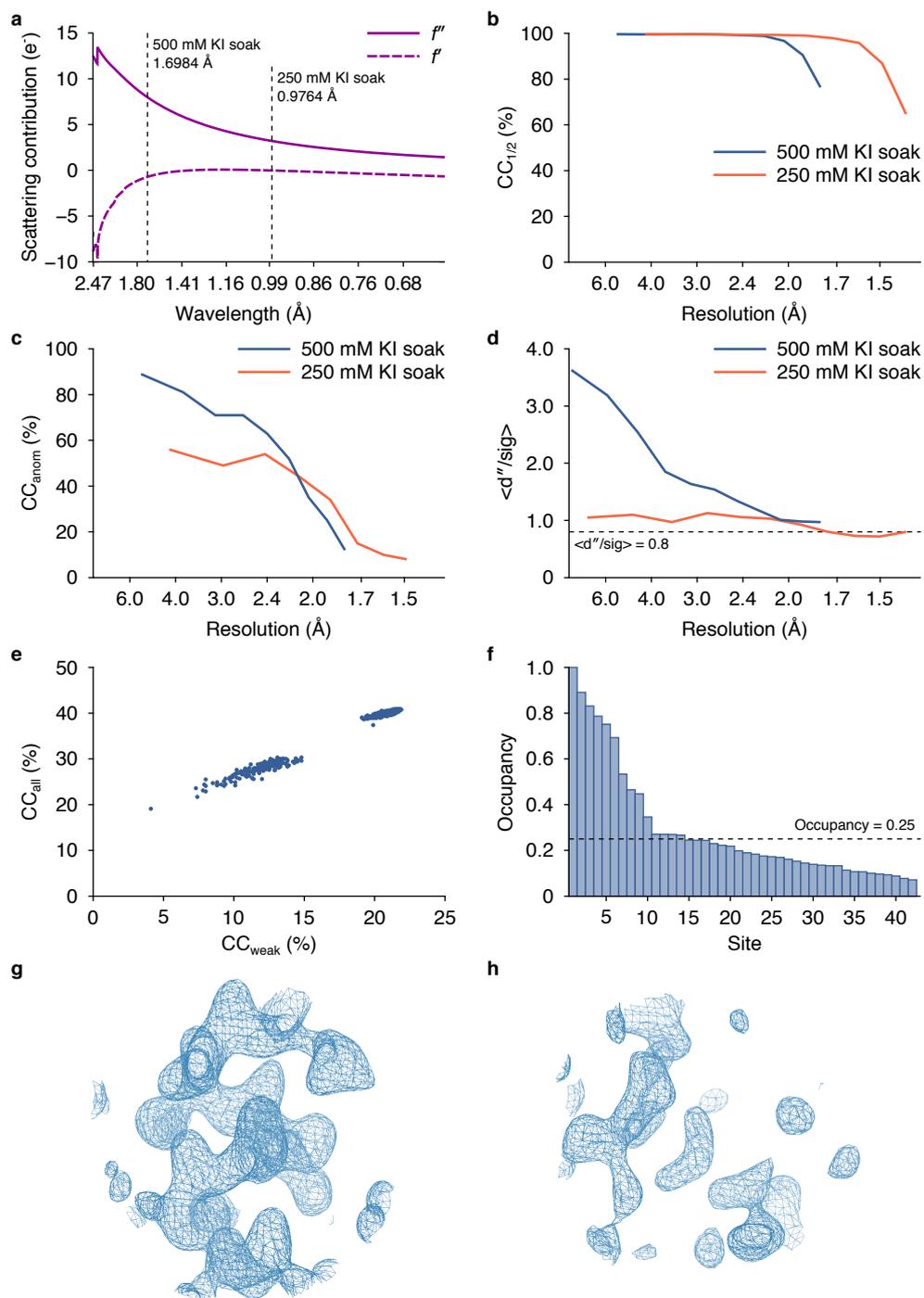


Figure 3.3. Experimental phasing and density modification. (a) Theoretical anomalous scattering contribution of iodine (209,210). Dotted lines indicate the wavelengths at which data were collected for crystals soaked with 500 mM and 250 mM KI, respectively. (b) Correlation coefficient between the average intensities of symmetry-related reflections in two half sets, $CC_{1/2}$, as a function of resolution. (c) Correlation coefficient between the anomalous differences of reflections in two half sets, CC_{anom} . (d) Mean anomalous difference divided by standard deviation, averaged, $\langle d''/sig \rangle$. A value of 0.8 corresponds to random noise. (e) Correlation coefficient between observed and calculated normalized structure factor amplitudes for all reflections, CC_{all} , and reflections not used in substructure solution, CC_{weak} . (f) Occupancy of identified anomalous sites. (g, h) Electron density for the two hands of the solution after density modification.

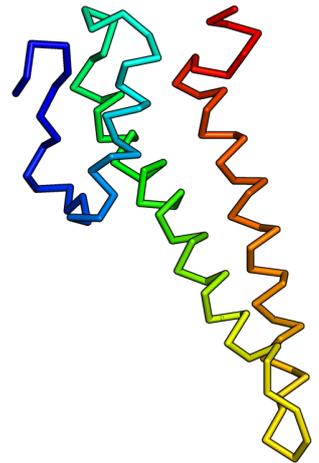
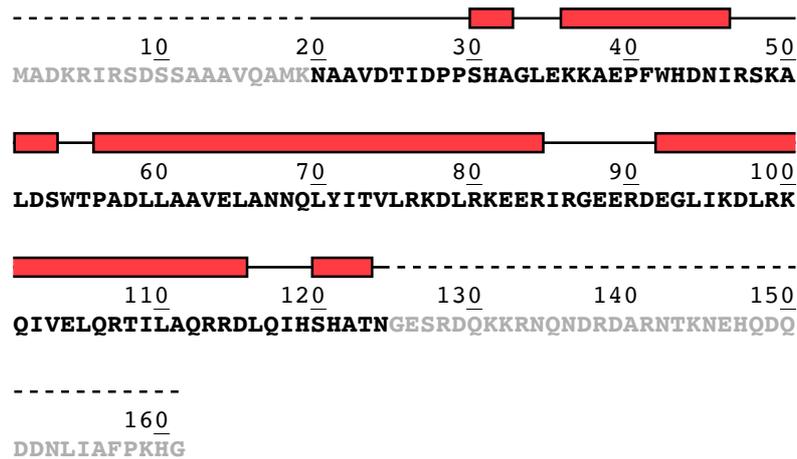


Figure 3.4. Secondary structure profile and ribbon representation of a HK97 small terminase monomer. Helices are indicated by red rectangles, random coil by a solid line, and unresolved residues by a dotted line. The ribbon diagram is coloured blue to red from N- to C-terminus.

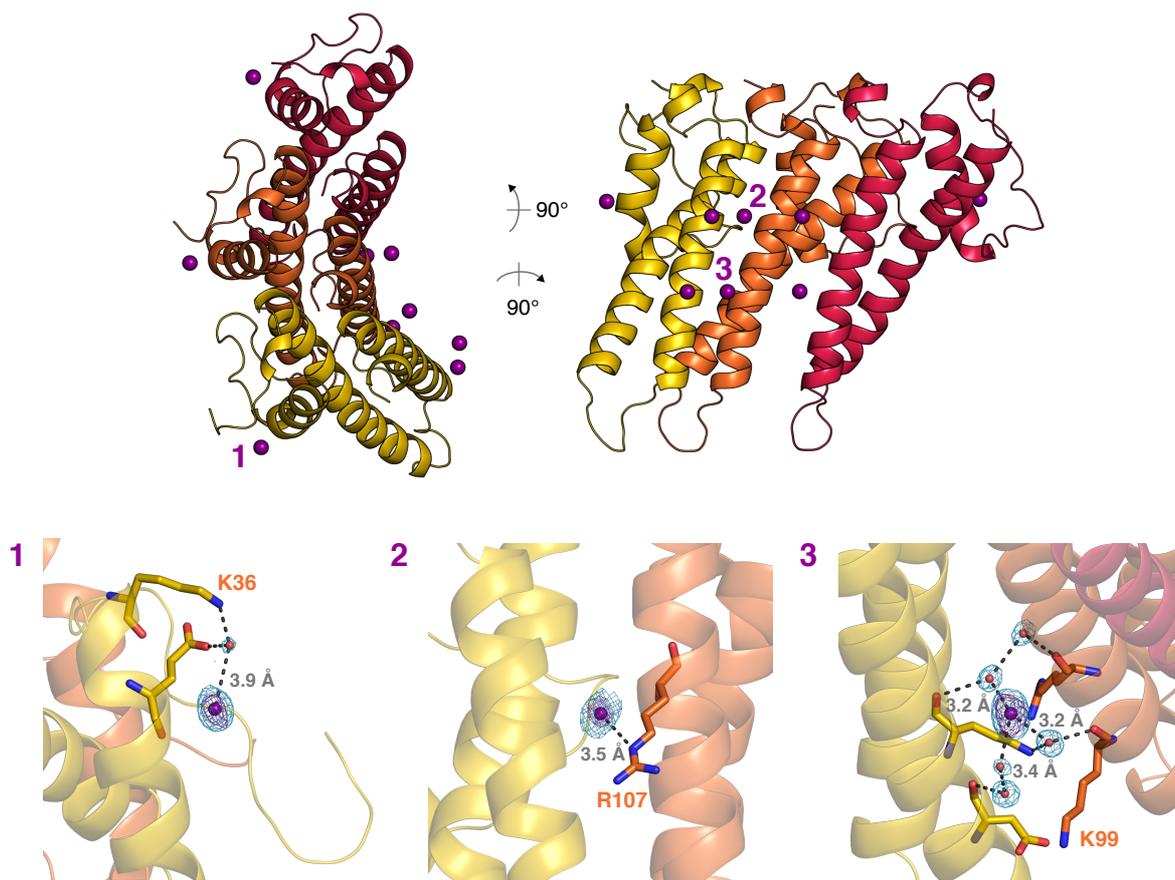


Figure 3.5. Asymmetric unit of the crystal structure. The local environment of each labelled iodide ion (1–3) is shown in an inset. The anomalous difference Fourier map (purple) and $2F_o - F_c$ map (blue), calculated using the final model and data from the 250 mM KI soak, are displayed around all solvent atoms, contoured at 4σ and 1.5σ , respectively, with a carve radius of 1.6 \AA .

Electron densities stronger than that of surrounding water molecules were observed in the solvent region. At 0.9764 \AA wavelength, f'' is 3.9 e^- for iodine. Because the densities also coincided with peaks in the anomalous difference Fourier map, they were attributed to iodide ions. To account for the elongated shape of the densities, the ions were modelled with anisotropic B factors, which led to improvements in R_{free} . The observed interatomic distances agree with published hydrogen bonding distances for iodide ions (211).

3.4.2 The HK97 small terminase crystal structure

Three copies of the small terminase protein are present in the asymmetric unit (Figure 3.5). Of 161 residues, residues 21–124 had well resolved electron density in each copy. A ring of nine subunits is produced on applying crystallographic symmetry (Figure 3.6a). The ring has a outer van der Waals diameter of ~ 85 Å and an inner tunnel ~ 18 Å wide at the narrowest point and ~ 55 Å long. Each interface has a buried solvent-accessible surface area of ~ 1227 Å². The observed quaternary structure is similar to that of small terminase proteins from *pac* viruses, where subunits are arranged circularly with no internal two-fold symmetry (105,107).

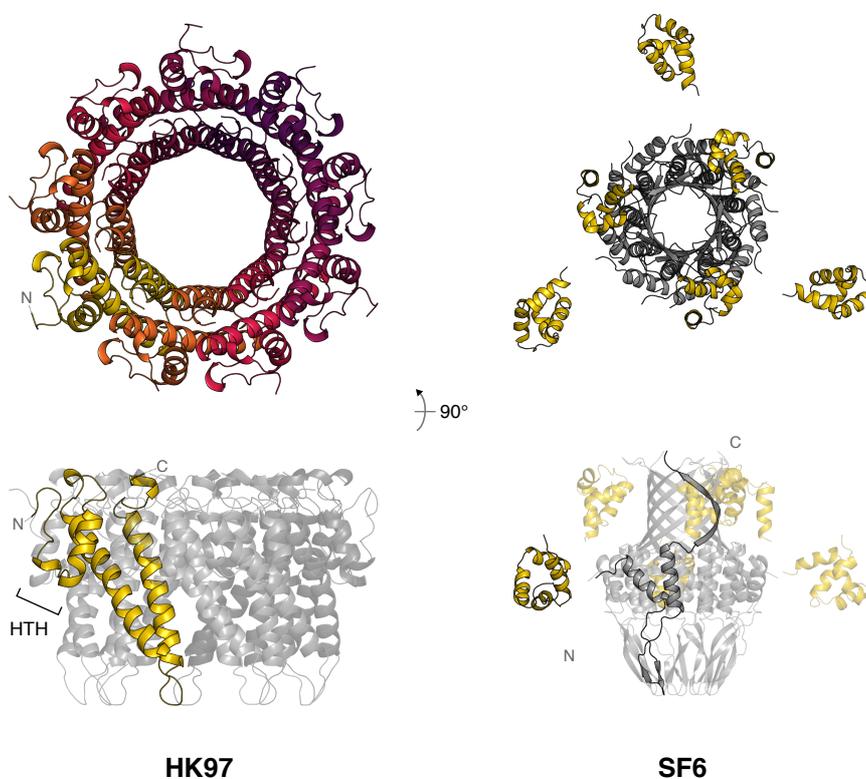


Figure 3.6. Cyclic symmetry in the HK97 small terminase. A 9-subunit oligomer is generated by rotation around the crystallographic 3-fold axis. Similar oligomeric states and rotational symmetry are observed in the small terminase proteins of some *pac* viruses, for example, phage SF6 (PDB code 3ZQQ) as shown. The N-terminal helix-turn-helix (HTH) domains of the SF6 protein are coloured in yellow. Three of nine N-terminal domains are disordered in the SF6 structure. HTH-like elements in the HK97 protein are also indicated.

Each protein chain contains two long helices, which form the oligomerisation interface, and an N-terminal region, which resembles a helix-turn-helix (HTH) motif (Figure 3.7a). This N-terminal region superposes with the winged HTH domain of λ small terminase (95) and the HTH region of *pac* homologues such as the *Bacillus* phage SF6 small terminase

(104), which have been implicated in DNA binding. Positive residues, K36, K37 and R47, occur in conserved positions at the turn and second helix of the motif.

Compared to the N-terminal domain of homologues, the HK97 HTH appears less mobile. The HTH is held in place by interactions between the second helix and the oligomerisation domain: a salt bridge between K49 and D58, a hydrogen bond between N45 and Q69, and hydrophobic contacts among W42, L59 and V63. The first helix of the HK97 HTH is truncated due to two proline residues, P28 and P29, which are strong helix breakers and frequent N-cap residues (212). Instead of a hydrophobic core, which features in homologues, the HTH is held together by direct intramolecular contacts (Figure 3.7b). Contacts include four hydrogen bonds, and a CH- π interaction between the C β atom of P28 and W42. The CH- π interaction has a distance of 3.8 Å, which is comparable to the curated value of 3.7 ± 0.1 Å (mean \pm sd) (213).

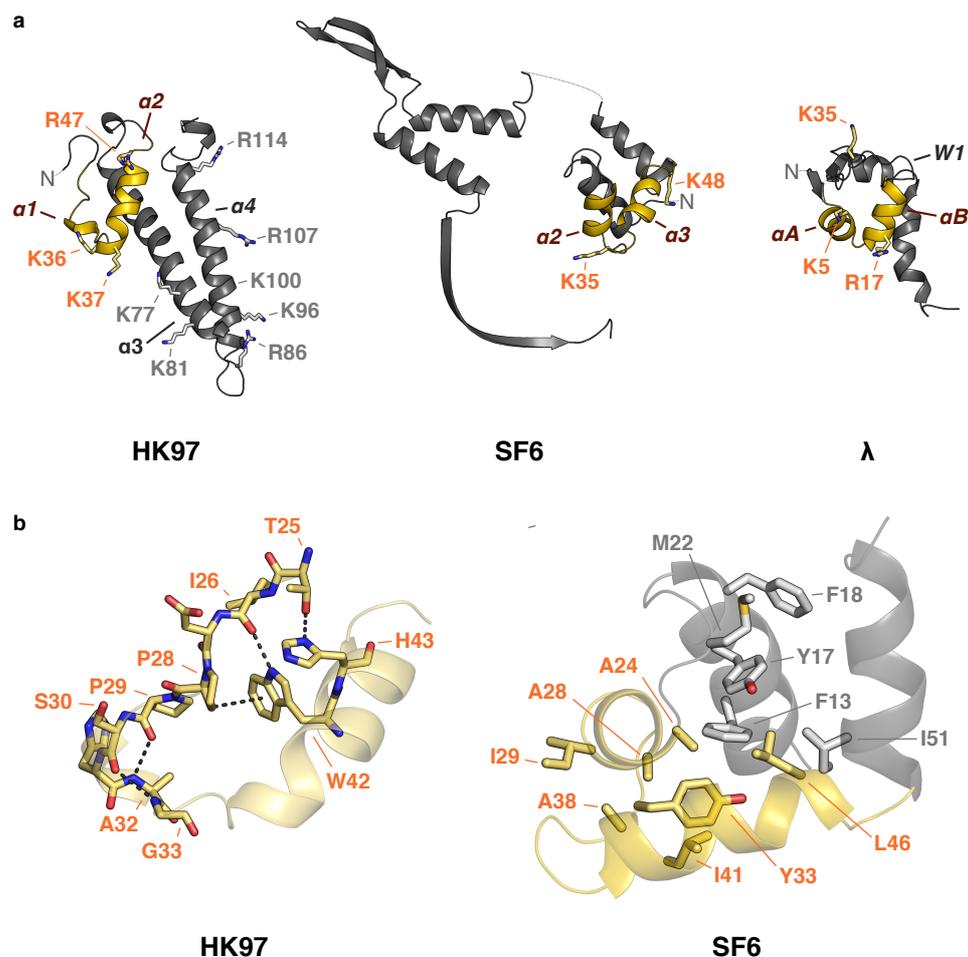


Figure 3.7. Comparison of small terminase subunits. (a) The N-terminal helix-turn-helix regions of the HK97 small terminase, SF6 small terminase (PDB code 3ZQQ) and the λ small terminase (PDB code 1J9A)

are coloured in yellow. Positively charged residues present in all three structures are indicated in orange. (b) Intramolecular contacts in the N-terminal HTH regions of the HK97 and SF6 small terminase proteins.

Together with residues K77 and K81, the conserved positive residues of the HK97 HTH form a positive belt around the small terminase oligomer (Figure 3.8). This identifies a potential DNA binding surface, in line with proposed models where viral DNA is selected (92,106). On the other hand, positive residues line the surface of the inner tunnel (Figure 3.7a). An alternate model of DNA binding where DNA enters through the tunnel is also possible (105). Whatever the case, similarities in structure and symmetry suggest that HK97 and *pac* viruses employ a similar viral DNA recognition mechanism.

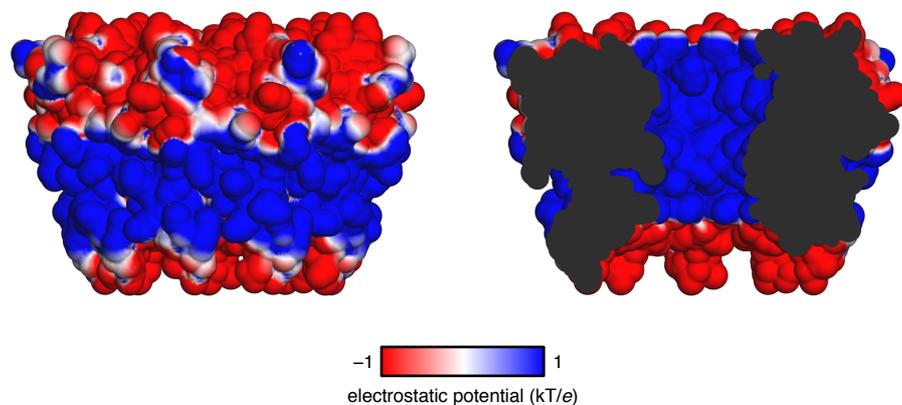


Figure 3.8. Electrostatic surface potential of the HK97 small terminase oligomer. The solvent accessible surface of the oligomer is displayed. Potentials were calculated under the SWANSON force field (190).

3.5 Biophysical analysis of the *Bacillus* phage SF6 small terminase N-terminal domain

3.5.1 The DNA-binding properties of the N-terminal domain by surface plasmon resonance

The structural similarities between *cos* and *pac* small terminase proteins are clear. For a broader understanding of their DNA binding mechanisms, the N-terminal domain of *Bacillus* phage SF6 small terminase was also studied. Despite being implicated in DNA binding, the SF6 HTH has never been shown directly to bind DNA. Hence, the N-terminal domain was produced and analysed by surface plasmon resonance (SPR; Figure 3.9a).

Fast association and dissociation were observed as protein was injected over immobilised DNA (Figure 3.9b). Accompanying work published in *Nucleic Acids Research* showed that the N-terminal domain has no sequence preference throughout the *pac* region of the SF6 genome (178). Rather, it is the circular arrangement of N-terminal domains that enables the selection of viral DNA. Because individual domains bind DNA non-sequence-specifically, the SPR responses at equilibrium were analysed under the Tsodikov-modified McGhee–von Hippel model (192,193), which describes the non-specific binding of ligands to a finite one-dimensional lattice (Figure 3.9c,d). Transformation of variables for analysis under the model is possible because DNA and protein have similar refractive index increments (214) and because the response in SPR is directly proportional to refractive index changes. Against 20 bp and 30 bp DNA, individual N-terminal domains had a footprint of 3 bp and an intrinsic dissociation constant of 0.92–1.85 mM (Figure 3.9c,d).

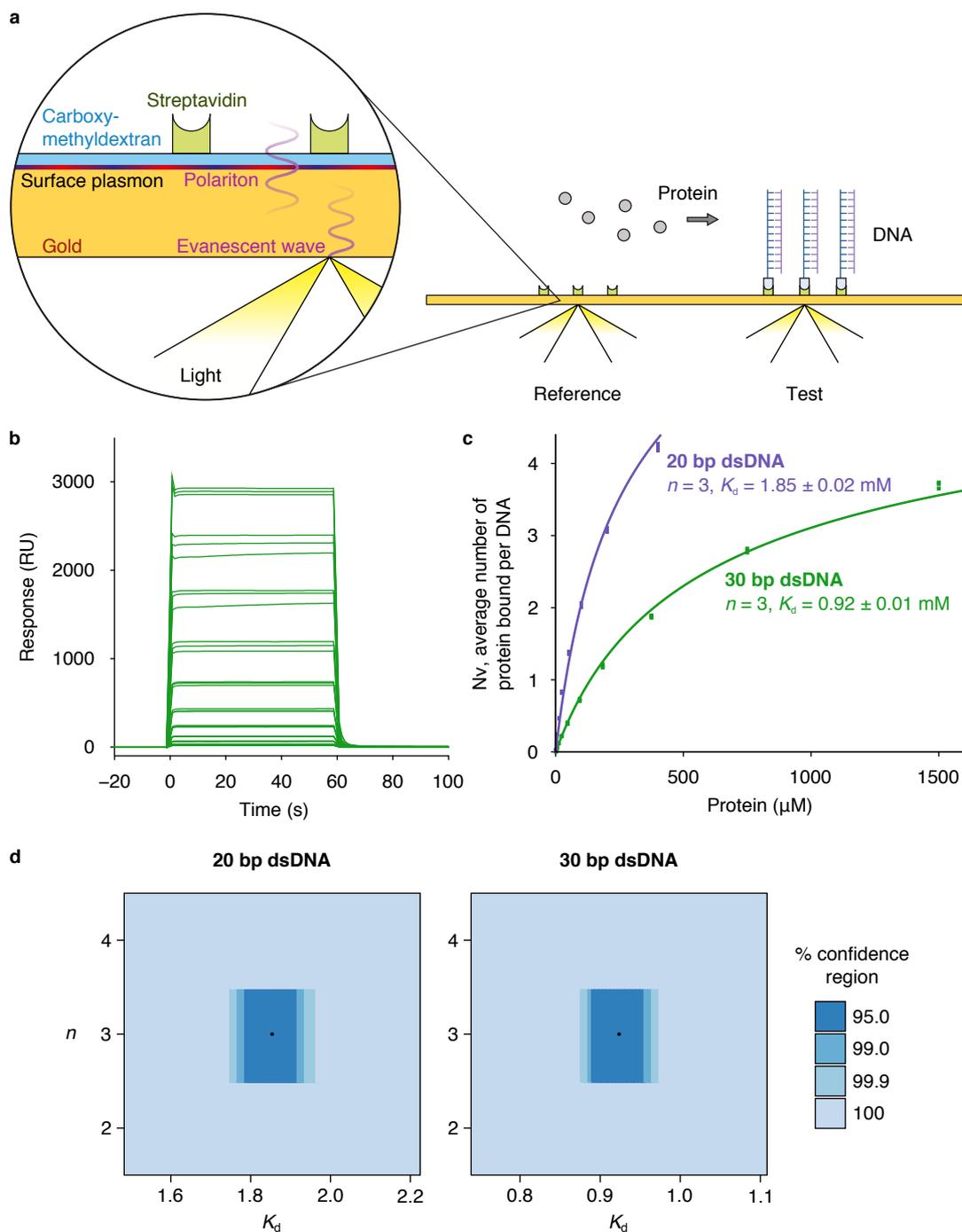


Figure 3.9. Surface plasmon resonance analysis of the DNA-binding properties of the SF6 small terminase N-terminal domain. (a) Schematic of the experiment. DNA is immobilised via streptavidin-biotin to a carboxymethyl dextran-modified gold surface. The change in the angle of incidence for maximum surface plasmon resonance is monitored as protein is injected over the DNA. Bulk refractive index changes during protein injection are measured using a reference surface with no immobilised DNA and corrected for. (b) Reference-corrected responses from injecting protein (0–1.5 mM) over 30 bp DNA in 20 mM HEPES, pH 7.0, 150 mM KCl, 0.05% (v/v) Tween 20. (c) Equilibrium response data analysed under the Tsodikov-modified McGhee–von Hippel model. Reported are the standard errors of regression. (d) Confidence regions based on residual sum of squares as a function of K_d and n . Points indicate the least-squares estimates.

3.5.2 Identification of the DNA binding surface by NMR

To enable NMR analysis of the SF6 N-terminal domain, a series of spectra were acquired using ^{13}C , ^{15}N -labelled protein. Backbone resonances were assigned in sequence based on cross peaks in the CBCA(CO)NH, CBCANH spectra, and the HNCO, HN(CA)CO spectra (Figure 3.10). In these spectra, the $\text{C}\alpha$, $\text{C}\beta$ and CO atoms of the previous and current residues are correlated through the amide groups of each residue. Side chain amide resonances were assigned based on cross peaks in the ^{15}N - ^1H TOCSY-HSQC and ^{15}N - ^1H NOESY-HSQC spectra, which measure the correlation between hydrogen atoms of the same residue and correlation between the hydrogen atoms of one residue and nearby hydrogen atoms in space, respectively. ^1H , ^{13}C and ^{15}N backbone resonances were assigned for all residues except proline and the first three residues. Side chain amide resonances were assigned for residues N36, N45 and N63 (Figure 3.11).

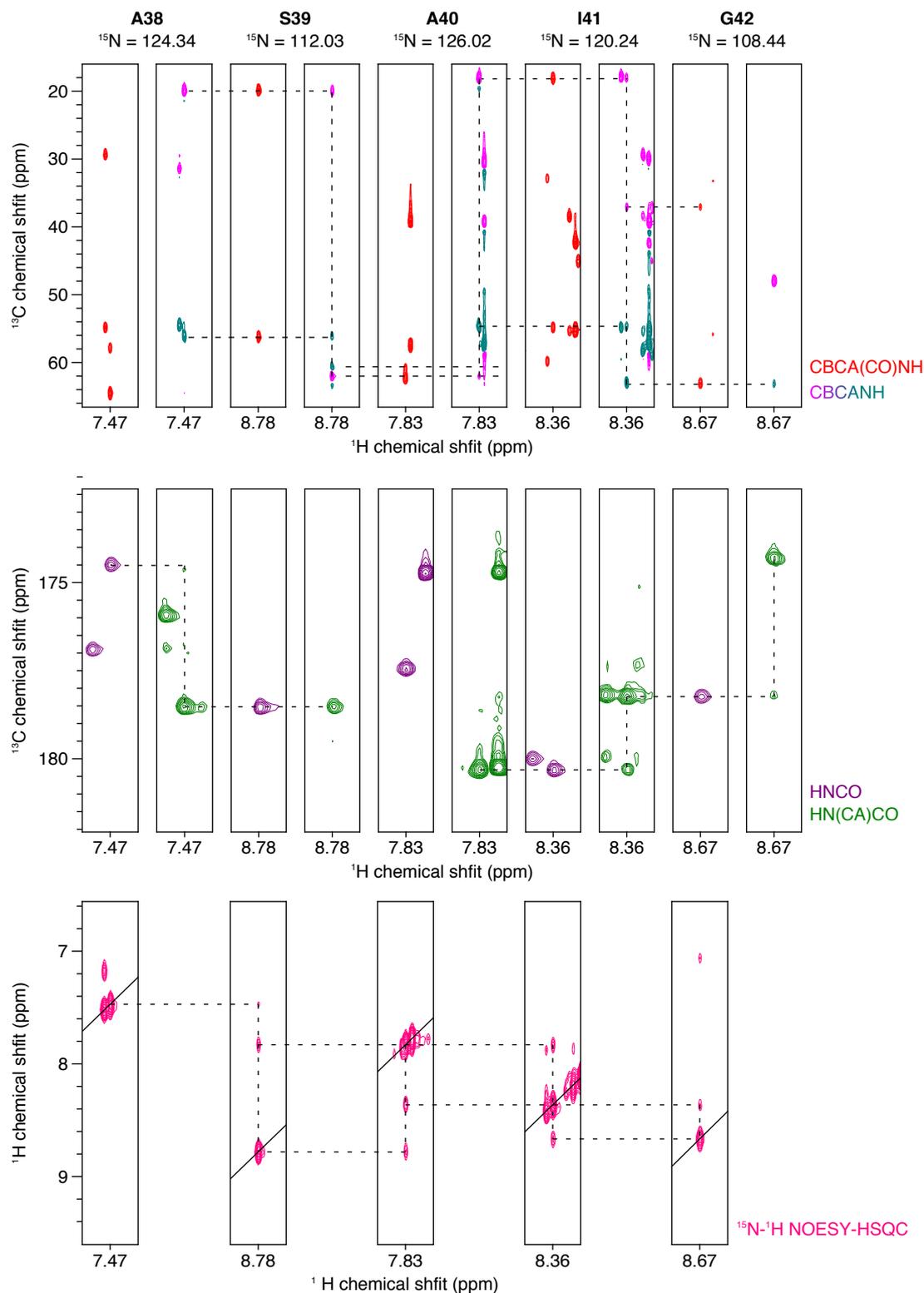


Figure 3.10. Backbone assignment of the SF6 small terminase N-terminal domain. Resonances are assigned by correlating the H, N, $\text{C}\alpha$ and $\text{C}\beta$ chemical shifts of each residue and its preceding residue in CBCA(CO)NH and CBCANH spectra, and H, N and CO chemical shifts in HNCO and HN(CA)CO spectra. Assignments in helical regions are validated by cross peaks in the ^{15}N - ^1H NOESY-HSQC spectrum corresponding to the $i \pm 1$ residue. Strip plots are shown to illustrate the process. The ^{15}N chemical shifts of the displayed planes are indicated.

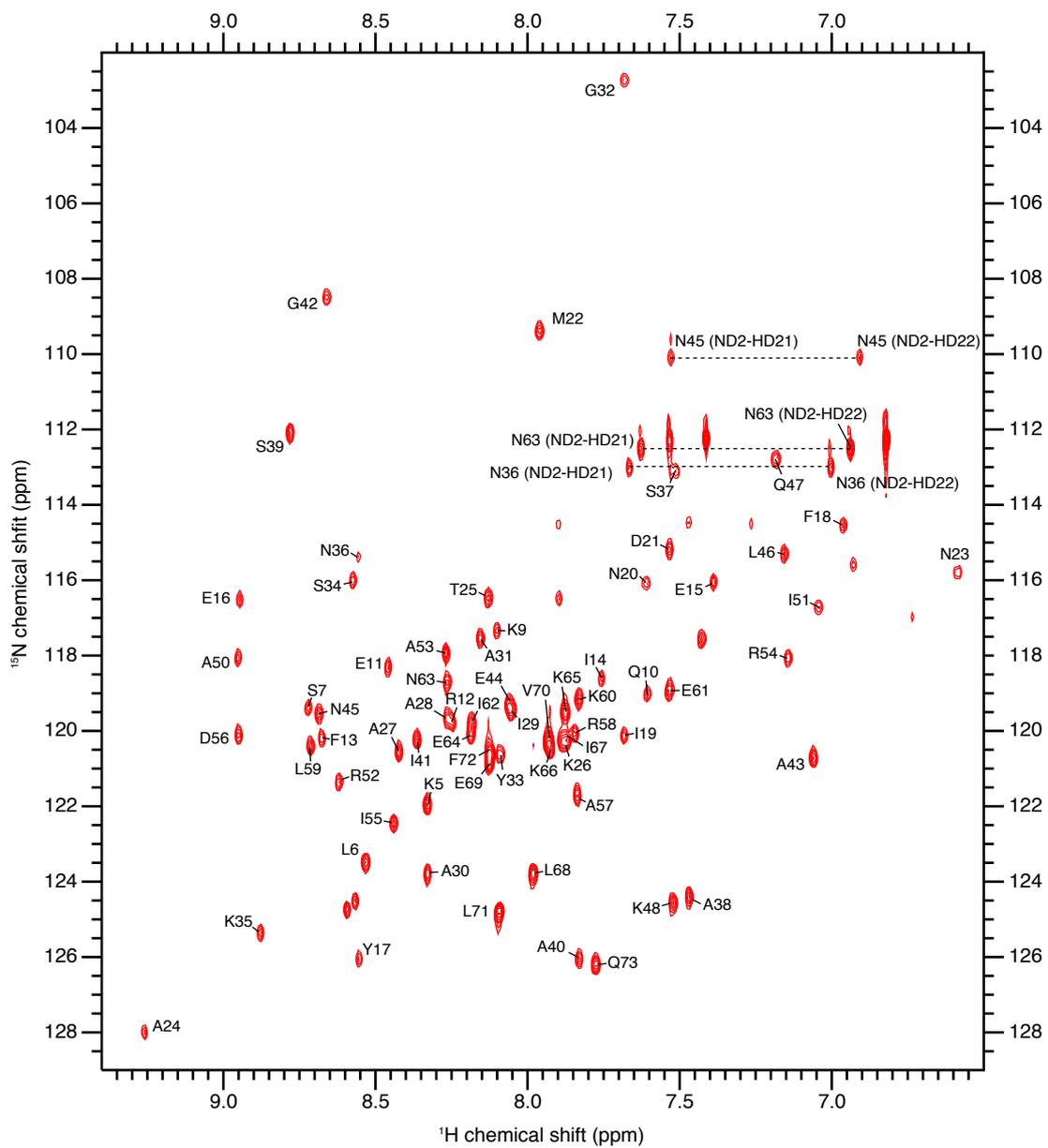


Figure 3.11. Amide assignment indicated on the ^1H - ^{15}N HSQC spectrum.

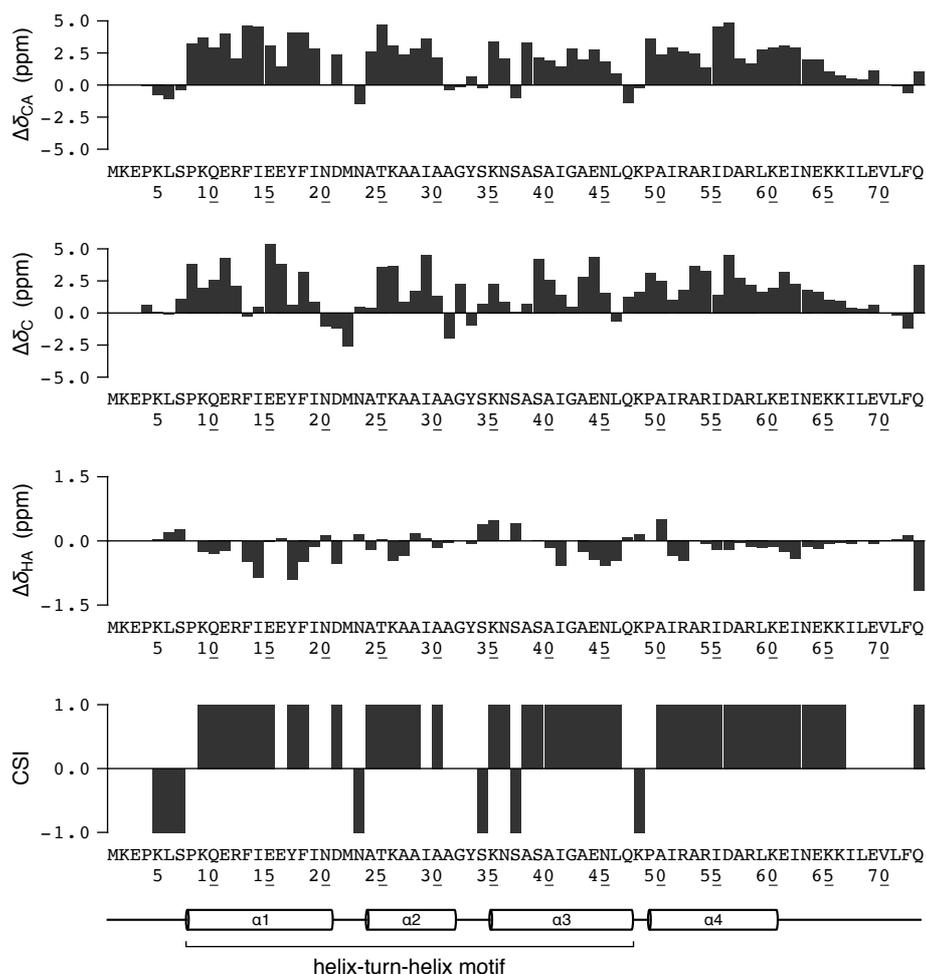


Figure 3.12. Secondary structure indicators. A positive $C\alpha$ or CO secondary chemical shift, or a negative $H\alpha$ secondary chemical shift indicates the presence of a helix. A chemical shift index (CSI) (215) can be calculated based on majority rule, where 1 represents a helix and -1 represents a strand. The secondary structure observed by X-ray crystallography (PDB code 3ZQQ) is shown underneath.

Because the structure of the SF6 small terminase is known, it was possible to validate the assignments. Residues in a helix would have peaks in the ^{15}N - 1H NOESY-HSQC spectrum corresponding to amide hydrogen atoms of the previous and next residue (Figure 3.10). Secondary structure predictions based on secondary chemical shifts, that is, the difference in chemical shift relative to that of a random coil, were also compared to the known secondary structure (Figure 3.12). Overall, assignments agree with the crystal structure.

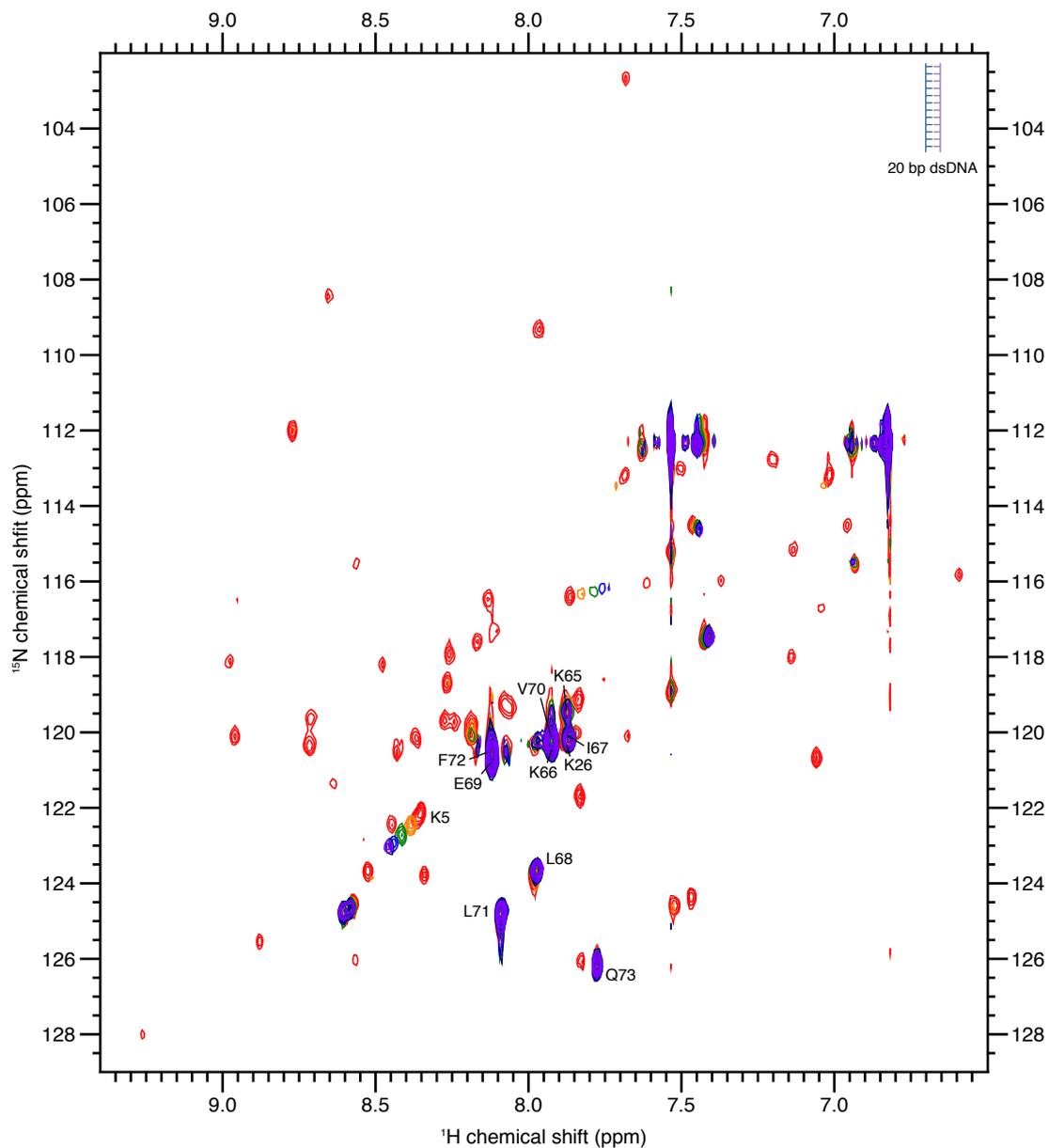


Figure 3.13. ^1H - ^{15}N HSQC titration series of 20 bp dsDNA into the SF6 N-terminal domain. Protein was titrated against DNA (red to purple) to a final DNA/protein ratio of 1.5 in 20 mM MES, 150 mM KCl, pH 6.0, 10% (v/v) D_2O , at 25 °C.

To identify the DNA binding surface on the protein, ^{15}N -labelled protein was titrated against the 20 bp DNA used in previous SPR experiments and chemical shifts monitored via ^1H - ^{15}N HSQC spectra. However, significant line broadening was observed, likely due to protein binding to DNA at different positions and in varying numbers (Figure 3.13). To overcome this problem, a 5-bp-forming hairpin DNA substrate was designed such that only 1:1 interactions were possible. The stoichiometry of the interaction was confirmed by sedimentation velocity analytical ultracentrifugation (178).

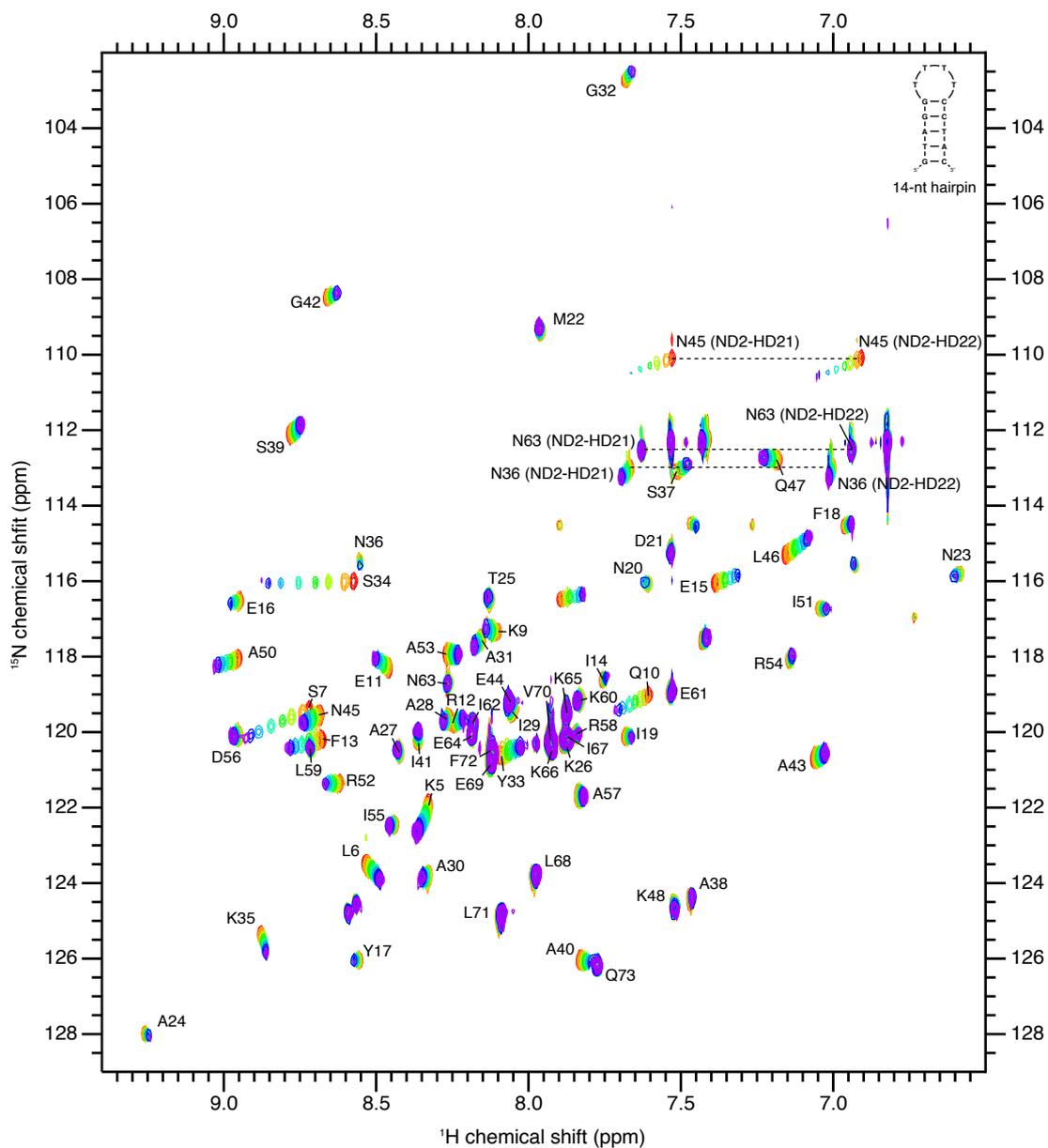


Figure 3.14. ^1H - ^{15}N HSQC titration of a 14-nt DNA hairpin into the SF6 N-terminal domain. Protein was titrated against DNA (red to purple) to a final DNA/protein ratio of 19.2 in 20 mM MES, 150 mM KCl, pH 6.0, 10% (v/v) D_2O , at 25 °C.

Titration of protein against this hairpin DNA substrate gave rise to measurable chemical shift perturbations (Figure 3.14). Mapping these perturbations to the crystal structure, the largest chemical shift differences occurred in the N-terminus and the helix-turn-helix motif (Figures 3.14, 3.15). Most affected residues lie on the same face of the protein, marking a potential DNA binding surface. The same surface in other HTH-containing proteins studied by NMR also showed significant perturbations when titrated against DNA (Figure 3.16). These data confirm that helix-turn-helix motifs found in small terminase proteins truly can mediate interactions with DNA, whether the protein is from a *cos* or a *pac* virus.

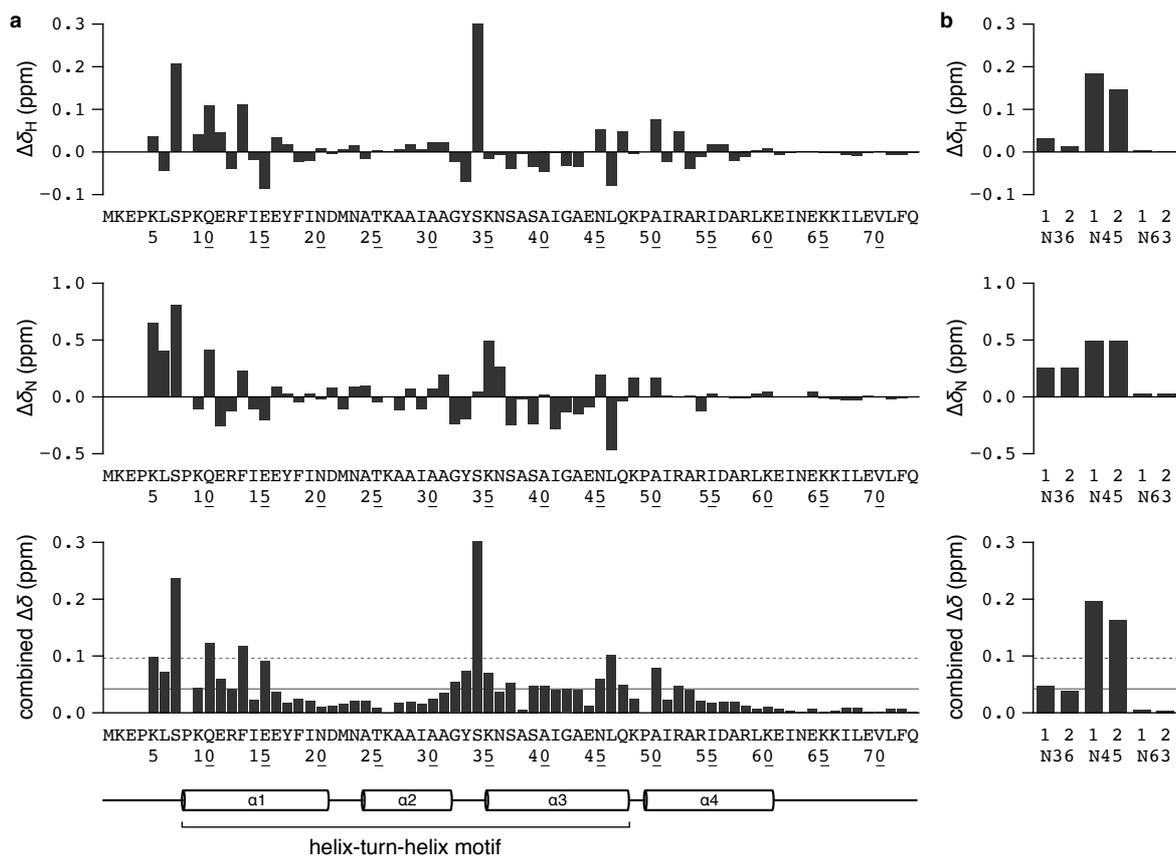


Figure 3.15. Chemical shift perturbation in response to the DNA hairpin. ^1H and ^{15}N chemical shift differences in the (a) backbone and (b) sidechain amide groups of the SF6 N-terminal domain at a DNA/protein ratio of 19.2 are shown. A combined chemical shift difference magnitude (205) is calculated. Solid line indicates the mean combined magnitude, and dashed line indicates one standard deviation above the mean.

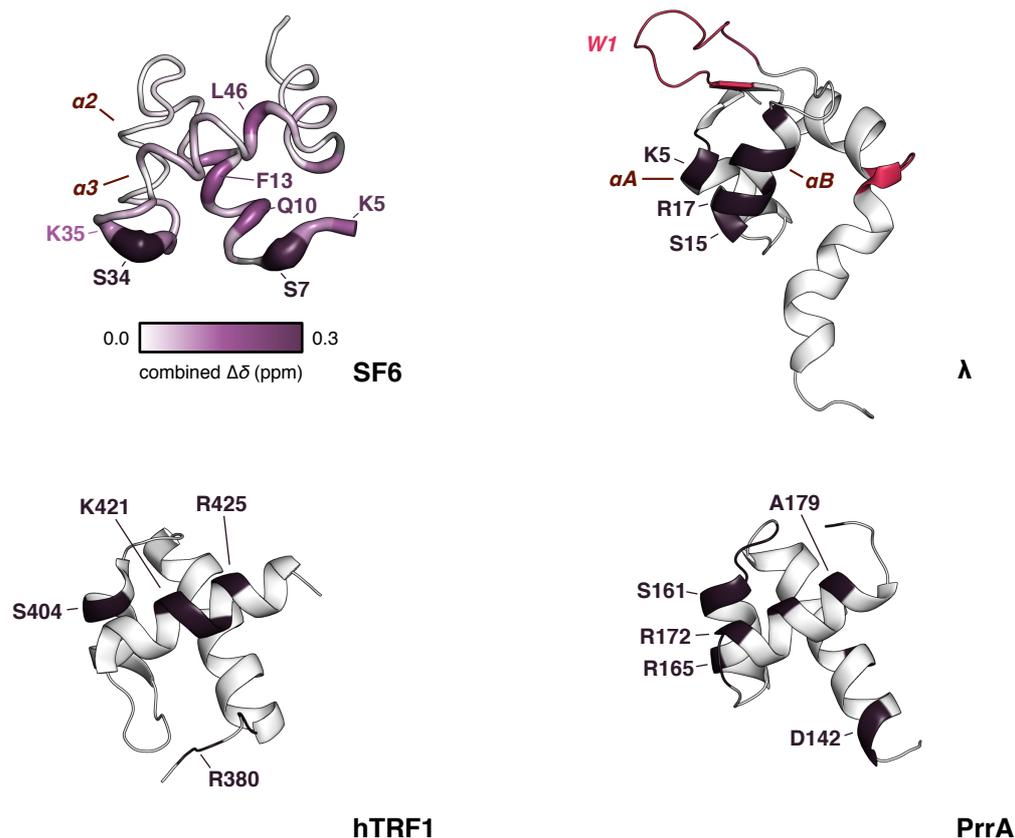


Figure 3.16. Mapping of chemical shift perturbation. Putty representation of the SF6 N-terminal domain based on combined chemical shift difference magnitudes. Residues are also coloured based on magnitudes from white to purple. Unassigned residues are coloured in grey. Also shown are structures of helix-turn-helix-containing proteins that have also been studied by NMR: the N-terminal domain of λ small terminase (95), human telomerase repeat factor binding protein 1 (hTRF1) (216), and the *Rhodobacter sphaeroides* transcription factor PrrA(217). Structures were aligned to the SF6 N-terminal domain and displayed in the same orientation. Residues showing significant chemical shift differences and line broadening on titration against DNA are coloured in dark purple and red, respectively.

3.6 Interactions of the HK97 small terminase with DNA

3.6.1 The DNA-binding properties of the HK97 small terminase

Data on the λ and SF6 small terminase suggest that *cos* DNA and *pac* DNA, respectively, wrap around the protein oligomer through interactions with its HTH motifs. Consequently, the length of DNA bound is determined by the footprint of individual motifs, the number of participating motifs and the spacing between them. For both systems, this length was estimated to be ~100 bp based on *in vitro* packaging and DNase footprinting experiments (89,106).

To identify the segment of DNA the HK97 small terminase recognises, the HK97 *cos* region was screened in fragments using an electrophoretic mobility shift assay (Figure 3.17a). A non-standard binding and running buffer containing 100 mM Na₂SO₄ was used because free protein in standard Tris-Borate buffer would precipitate at concentrations greater than 1 μM. On addition of 5 μM protein oligomer, strong shifts were observed for fragment II (Figure 3.17b). Some shifts were observed for fragment III but little was observed for fragment I. However, when enough protein was added, shifts could also be induced for fragment I (Figure 3.17c). A similar shift was observed for fragment II at the same protein concentration, indicative of a non-specific, high-order interaction regime. A window existed between 0.4 and 2 μM protein concentration where fragment II was preferentially bound. This bound DNA had a different mobility, likely resulting from a specific protein-DNA interaction. From these data, it can be concluded that the cognate DNA of HK97 small terminase lies in the region between -80 and +472.

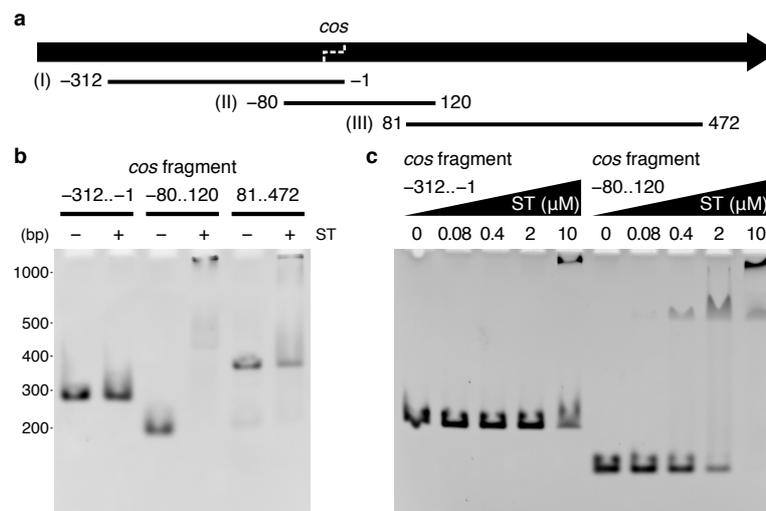


Figure 3.17. Electrophoretic mobility shift assay. (a) Segments of the HK97 *cos* region, designated I, II and III, were prepared by PCR from a *cos*-containing pUC18 vector. Dotted line indicates the site of cleavage during genome packaging. (b) 50 nM DNA was incubated with 5 μM small terminase nonamer before electrophoresis on a polyacrylamide-agarose composite gel (3%:0.5%) in 25 mM Tris-Cl, 192 mM glycine, pH 8.3, 100 mM Na₂SO₄, 10 mM MgSO₄, at 4 °C. (c) DNA was titrated against protein (0–10 μM) under the same conditions.

3.6.2 Preparation of a 95-bp DNA substrate for co-crystallisation studies

It remains unclear how small terminase interacts with DNA. Thus, attempts were made to co-crystallise the HK97 protein with DNA. No crystals have been obtained so far but the methods explored are documented here to facilitate future work. The small terminase

oligomer has an outer van der Waals diameter of ~ 85 Å. Accounting for the diameter of DNA, 97 bp of DNA would be required to circumscribe the protein. In a first attempt, a 95-bp segment of the HK97 *cos* region was recreated. The segment was assembled from synthetic DNA oligonucleotides, which were phosphorylated using T4 polynucleotide kinase, annealed and ligated using T4 DNA ligase (Figure 3.18a). Non-denaturing polyacrylamide gel electrophoresis revealed that the ligation reaction did not go to completion (Figure 3.18b). Purification of the full-length segment by size exclusion and anion exchange chromatography produced pure DNA with a 16% recovery compared to the theoretical yield (Figure 3.18c,d). Crystallisation trials with protein and DNA have so far yielded no crystals. The design of the DNA and conditions for complex assembly likely require more investigation.

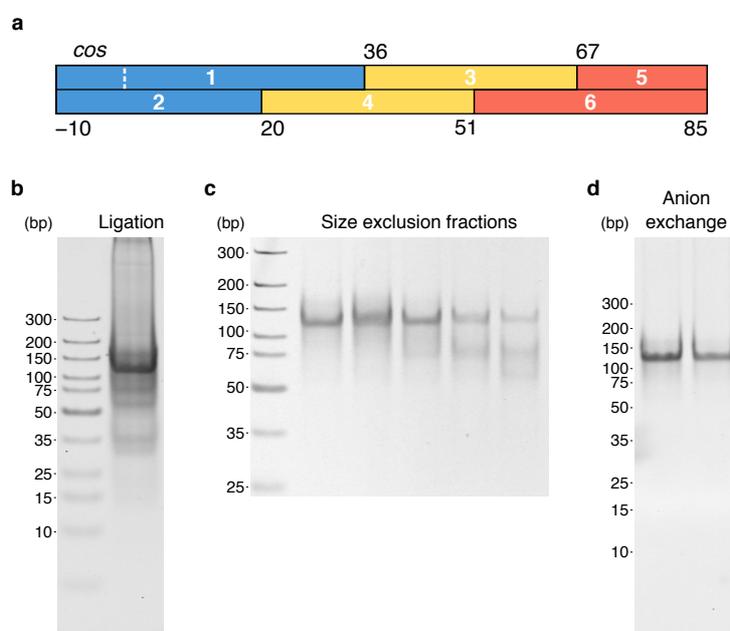


Figure 3.18. Production of a 95-bp *cos* segment for crystallisation with small terminase. (a) Oligonucleotides were annealed in the indicated pairs (blue, yellow, orange) and ligated together in a one-pot reaction using T4 DNA ligase. Oligonucleotides 2–4 were phosphorylated using T4 polynucleotide kinase before annealing. Dotted line indicates the site of cleavage during genome packaging. (b) Product of the ligation reaction. The DNA was purified by (c) size exclusion and (d) anion exchange chromatography.

3.7 Discussion

Small terminase appears to be essential for viral DNA recognition during packaging initiation and termination in *cos* viruses, and during packaging initiation in *pac* viruses. Because small terminase oligomerises to form a circular assembly, two mechanisms have

been proposed for the recognition of DNA by small terminase: DNA may be selected based on its ability to wrap around the protein; or, DNA threads through the central tunnel of the protein until the cognate sequence is bound.

DNase footprinting and *in vitro* packaging data suggest that binding of DNA by small terminase is interspersed over ~100 bp in the *pac* and *cos* region for phages SF6 and λ , respectively. In phage SF6, the spacing between protected regions coincides with the distance between consecutive N-terminal domains. Analysis of the N-terminal domain by SPR and NMR confirm that the domain can bind DNA and that it binds DNA through the HTH motif and N-terminus (Figures 3.8–3.15). The *pac* region has a high AT content, which favours kinks in the DNA. Coupled with the ring of N-terminal domains, a ‘DNA wrapping’ mechanism is plausible. In phage λ , IHF binding is thought to introduce bends, which favours interactions between the *cos* region and circularly arranged wHTH domains (92). The HK97 small terminase also has circularly arranged HTH elements (Figure 3.6). These HTH elements contain conserved positive residues that are implicated in DNA binding in the homologues (Figure 3.7). This supports a ‘DNA wrapping’ mechanism.

The central tunnel of the HK97 protein is lined with positive charges but it is long enough to accommodate only 16 bp of DNA. If the minimum DNA length for a specific interaction is longer than this, then the ‘DNA tunnel’ mechanism is unlikely. This can be investigated by DNase footprinting and electrophoretic mobility shift assays. Alternatively, cysteine mutations can be introduced on the tunnel surface, bulky chemical groups attached via maleimide labelling and the DNA-binding activity of the protein tested. In parallel, alanine mutations can be introduced in the HTH motif to test the ‘DNA wrapping’ mechanism.

Subtle differences exist between the proposed mechanisms for λ and SF6. Analytical ultracentrifugation data suggest that the λ wHTH domains bind DNA as a dimer and that the full-length protein forms a tetramer of dimers. The SF6 N-terminal domain is not known to dimerise, and no internal two-fold rotational symmetry is observed in the full-length structure. Similarly, subunits in the HK97 oligomer are related only by a nine-fold rotational symmetry (Figures 3.2, 3.5). The same oligomeric state was observed in solution (Figure 3.1). Unless the contacts observed between the HTH elements and oligomerisation helices are broken, it is unlikely that adjacent HTHs will be flexible enough to dimerise. If DNA does wrap around the HK97 protein, then local interactions between the protein and DNA will likely resemble those for SF6 rather than λ .

Two modes of protein-DNA interaction were observed for the HK97 small terminase: a specific interaction requiring a specific *cos* segment at 0.4–2 μM protein concentration, and a higher-order, non-specific interaction at 10 μM protein concentration (Figure 3.17). Two similar modes have also been observed for the λ protein. In fact, most DNA-binding proteins that recognise a specific sequence can interact with DNA non-specifically, only with a lower affinity. This is thought to help the proteins search for their cognate sequence via facilitated diffusion. Though small terminase was not essential for packaging initiation in the *in vitro* system (Figure 2.5c), non-specific interactions might have contributed to the higher packaging activity observed when the protein was present by helping to recruit large terminase.

Given a ‘DNA wrapping’ mechanism, an obvious question is why the HK97 *cos* region is preferentially bound. There is little sequence similarity between the HK97 and λ *cos* regions, or between the HK97 *cos* region and the SF6 *pac* region. However, between positions –17 and 78, the GC content is lower than 40% (Figure 3.19a). Kinks may occur in this region. No sequence matching the consensus IHF recognition sequence (218) is found in a 200 bp vicinity of the cleavage site. Nevertheless, the role of IHF in *cos* DNA recognition can be tested using electrophoretic mobility shift assays. Direct repeats of the form TTTNNNTTAA are located at positions 2 and 28. (Figure 3.19b). These may help to set the register of the interaction. A structure of the HK97 protein bound to *cos* DNA would elucidate the role of these DNA elements.

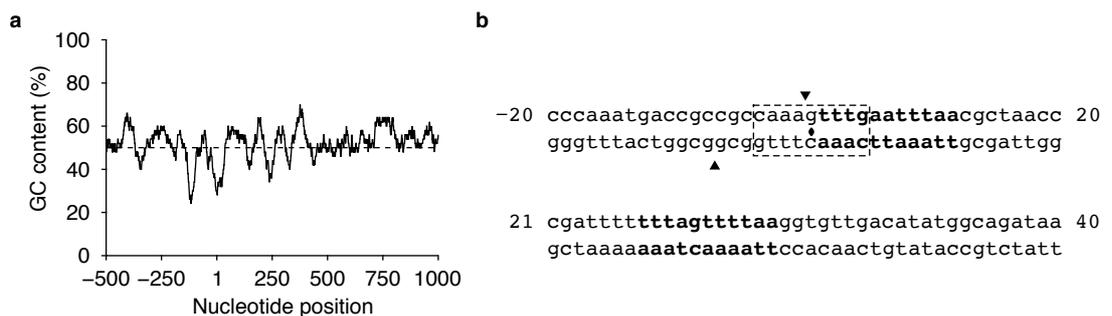


Figure 3.19. The HK97 *cos* region. (a) GC content, with a window size of 50 bp from the indicated position. (b) Sequence around the cleavage site during genome packaging. The sites of cleavage are indicated by arrows. Direct repeats of the form TTTNNNTTAA are indicated in bold. Dyad symmetry is present around the site of cleavage in the top strand, indicated by the dotted box.

For structure determination, the DNA should be long enough to enable specific protein-DNA interaction but short enough to minimise mobile elements and variation in register.

The type of DNA ends is another variable to explore. To prepare a large amount of DNA, a different approach such that used for the reconstitution of nucleosome core particles (219) may be considered, where tandem repeats of the substrate flanked by restriction sites are cloned into a vector. An advantage is that the DNA will be methylated as in the host during genome packaging. Procedures for assembling the protein-DNA complex may also require optimisation, for example, dialysis against decreasing salt concentrations to ensure a uniform 1:1 assembly. If the DNA length exceeds 60 bp, the complex will have a molecular weight of at least 200 kDa. Structure determination by cryo-EM will be possible with a Volta phase plate. Alternatively, small-angle X-ray scattering (SAXS) can yield information about the shape of the small terminase-DNA complex.

The interactions of the small terminase with large terminase remain uncharacterised. The C-terminal region of small terminase, which is absent in the HK97 crystal structure, is the putative site of recruitment for large terminase. This hypothesis can be tested in future pull-down experiments with deletions and mutations. It is anticipated that the interactions between the two proteins will be different during packaging initiation and termination. It is also unknown whether small terminase remains associated with large terminase during DNA translocation, or it is simply encountered again at the downstream *cos* signal. These questions may be addressed by further structural efforts, and perhaps by localising fluorescently labelled small terminase at *cos* sites on an extended single DNA molecule undergoing packaging.

4. Structural and biophysical analysis of the HK97 packaging motor

4.1 Introduction

The structure of a *cos* DNA packaging motor is unknown. Biochemical data suggest that the HK97 motor consists of at least the prohead and an ATPase assembly formed by large terminase. Symmetry mismatch in the ϕ 29 and T4 systems between the 12-subunit portal and the capsid at the 5-fold special vertex, and between the portal and the pentameric ATPase, has generated uncertainty over the nature of interfaces within the assembly. Based on the λ maturation complex (94), which features during packaging initiation, it may be speculated that the HK97 motor also contains four large terminase subunits. Work described in this chapter suggests that this is not the case. The HK97 motor was visualised by cryo-EM and reconstructed asymmetrically. Stoichiometry measurements made using total internal reflection fluorescence microscopy (TIRFM) on the stalled motor complement this structural analysis.

4.2 Materials and Methods

4.2.1 Protein and DNA production

The following construct was designed, from the N- to C-terminus: a short His-tag; followed by emerald GFP, which contains mutations S65T, F64L, S72A, N149K, M153T and I167T; a 10-residue G/S linker; and finally large terminase. The GFP coding sequence was amplified from a home-made plasmid using primers that have extensions for insertion via In-Fusion into the His-large terminase fusion construct. The reverse primer contained also the coding sequence for the G/S linker. The resulting construct was overproduced and purified as described for the His-SUMO fusion but protected from light where possible. For photobleaching experiments, DNA from position -80 to +150 of the HK97 *cos* region was amplified from a vector using unmodified forward and 5'-biotinylated reverse primers and purified using GeneJET PCR Purification Kit (Thermo Fisher).

4.2.2 Electron microscopy

Protein with native termini were used to reconstitute the motor. A packaging reaction was prepared as described, deposited on carbon-coated copper grids (Agar Scientific), stained with 1% (w/v) uranyl acetate (Fisher Scientific) or phosphotungstic acid, adjusted to pH

7.0 with NaOH (Sigma-Aldrich), and imaged on a Tecnai 12 G2 microscope with a tungsten filament source using an ISS MegaView III CCD camera at 120 kV at the University of York, or a Tecnai TF20 microscope with a FEG source using a Gatan Ultrascan 4000 CCD camera at 200 kV at the University of Pittsburgh. For cryo-imaging, reactions containing 125 nM prohead, 125 nM *ScaI*-linearised pUC18 DNA, 1 μ M large terminase monomer and 2 μ M small terminase monomer were prepared. ATP was added to 1 mM final concentration and the reaction was deposited on a Quantifoil R 2/2 grid for plunge-freezing using a FEI Vitrobot Mark III. Blotting times of 12–15 s were used.

Data for reconstruction were collected on a Polara G2 microscope using a Falcon II direct electron detector at 300 kV in nanoprobe mode. A condenser 2 aperture size of 70 μ m and nominal magnification of 109 000 \times were used. The calibrated pixel size, determined using recombinant HK97 Prohead II and a Mag*I*Cal standard (Electron Microscopy Sciences) containing alternating layers of Si and SiGe with defined thicknesses, was 1.37 \AA at this magnification. Four micrographs were taken using EPU (FEI) per hole with a defocus range of 2–6 μ m; 30 frames were collected per micrograph with a dose rate of 4 $e/\text{\AA}^2/\text{s}$ and total exposure time of 1.65 s.

Motion correction was performed using MotionCor2 (220). Given a spherical aberration of 2 mm and amplitude contrast of 0.10, defocus values were estimated for each micrograph using CTFFIND4 (221). Prohead particles were first picked manually in RELION and then auto-picked based on 2D classes (222). Motor particles were picked manually in X3D (223). Starting models were generated using the RMC procedure implemented in AUTO3DEM with 200 particles (224). Classification of prohead particles and prohead reconstructions were performed using RELION 2.1 (225) on a computer with an NVIDIA Tesla K80 dual-GPU processor at the York Advanced Research Computing Cluster. Out of 72 236 picked particles, 19 576 and 15 383 particles were used for reconstruction of classes 1 and 2, respectively. The HK97 motor was reconstructed as described later using AUTO3DEM 4.05.2 (223,224) on a desktop computer with an Intel Core i7-4790 CPU @ 3.60 GHz.

4.2.3 Subunit counting by total internal reflection fluorescence microscopy

Flow cells were constructed and functionalised by applying 25 μ L 1 mg/mL biotin-BSA (Sigma-Aldrich) with \sim 0.1 mg/mL 5- μ m non-functionalised silica microspheres (Bangs

Laboratories) to a quartz slide, covering with a coverslip (No. 1, 22 mm × 64 mm, Menzel-Gläser) and sealing with nail varnish over the short sides. After 10 min, unbound biotin-BSA was washed out and the flow cell equilibrated with two 25- μ L volumes of imaging buffer: 20 mM Tris-Cl, 10 mM MgSO₄, 30 mM potassium glutamate, 1 mM ATP γ S, 0.1 mM β -mercaptoethanol, 0.5 mM Trolox (Sigma-Aldrich), 0.1 mg/mL acetylated BSA (Sigma-Aldrich), 0.25% (w/v) PEG 6000 (Santa Cruz Biotech), 380 nM BOBO-3 stain (Thermo Fisher). Separately, 100 nM DNA was incubated with 3.7 μ M BOBO-3 stain (Thermo Fisher), 100 nM streptavidin tetramer, 200 nM proheads, 1 μ M GFP-large terminase and 2 μ M small terminase for 20 min at room temperature, that is, 20–22 °C. Adding ATP to 1 mM final concentration, the mixture was immediately chased with 10 volumes of imaging buffer. Immediately afterwards, 25 μ L of this diluted mixture was drawn into the flow cell and the long sides sealed subsequently.

The sample was visualised by prism-coupled total internal reflection fluorescence microscopy on a modified inverted IM35 microscope (Carl Zeiss AG). Fluorophores were excited with 488-nm and 561-nm lasers (Coherent) operating at 10 mW and 30 mW, respectively. Incident 488-nm light was circularly polarised using a 488-nm zero-order quarter-wave plate (Edmund Optics) to minimise orientation-dependent excitation. Fluorescence emission was captured through a Plan-Apochromat 100 \times /NA 1.4 oil-immersion objective (Carl Zeiss AG). A dual-view image splitter (OptoSplit II, Cairn Research) with 1.6 \times magnification, 580 nm long-pass emission dichroic (Zeiss), and bandpass filters for GFP (ET525/50M, Chroma) and BOBO-3 (ET605/20M, Chroma) was used to separate the image into two fluorescence emission channels. Video data were recorded using an Evolve 512 electron-multiplying CCD camera (Photometrics), cooled to –70 °C and operated through MicroManager (226) with 33 ms exposure at 200 electron multiplier gain. Pixel width in the magnified image was 97.7 nm, determined using a USAF calibration target (Edmund Optics). The vector relating the two channels was determined using TetraSpeck microspheres (ThermoFisher).

To locate DNA, the average of the first 500 frames in the red channel was binarised and analysed by connected-component labelling (227). The intensity-weighted centroids of identified components were taken to be the DNA coordinates. Coordinates fewer than 10 px apart were rejected. For each coordinate, given the vector between channels, the corresponding pixel in the green channel was identified. A circle 3.5 px in radius was

drawn around the centre of this pixel. The mean intensity of this circle over time was extracted and fitted for steps using the ‘Progressive Idealization and Filtering’ (PIF) algorithm (228) with Chung-Kennedy filtering (229). For each fit, a counter-fit was generated where steps are instead placed between the original steps (230). A better chi-squared statistics in the counter-fit indicates over-fitting and thus the fit was rejected. Protein and DNA were considered to co-localise only if the intensity-weighted centroid in the green channel was within experimental error of that in the red channel. Experimental error was defined to be two standard deviations in the position measurement of one GFP monomer relative to another over time (231). Binominal distributions were fitted to the distribution of step counts under the assumption that all photobleaching events occurred independently of one another within and across all motor complexes. All data analysis routines were implemented as custom scripts in MATLAB R2016b.

4.3 Negative-stain and cryo-EM imaging of an active HK97 DNA packaging motor

The structure of a complete *cos* motor is unknown to date. Single-particle cryo-EM has been the method of choice to visualise viral DNA packaging motors, given their size and heterogeneity in solution. In the first instance, the HK97 motor was imaged by negative-stain EM. An active assembly was prepared using *ScaI*-linearised pUC18 DNA which did not contain the *cos* region, and ATP. The preparation was then spotted immediately on carbon-coated copper grids, washed with packaging buffer and stained with uranyl acetate. Density could be observed in the resulting images on the exterior of capsids which could correspond to a complex between large terminase and DNA (Figure 4.1a,b). Distortions and expansion of capsids were sometimes observed with uranyl acetate staining.

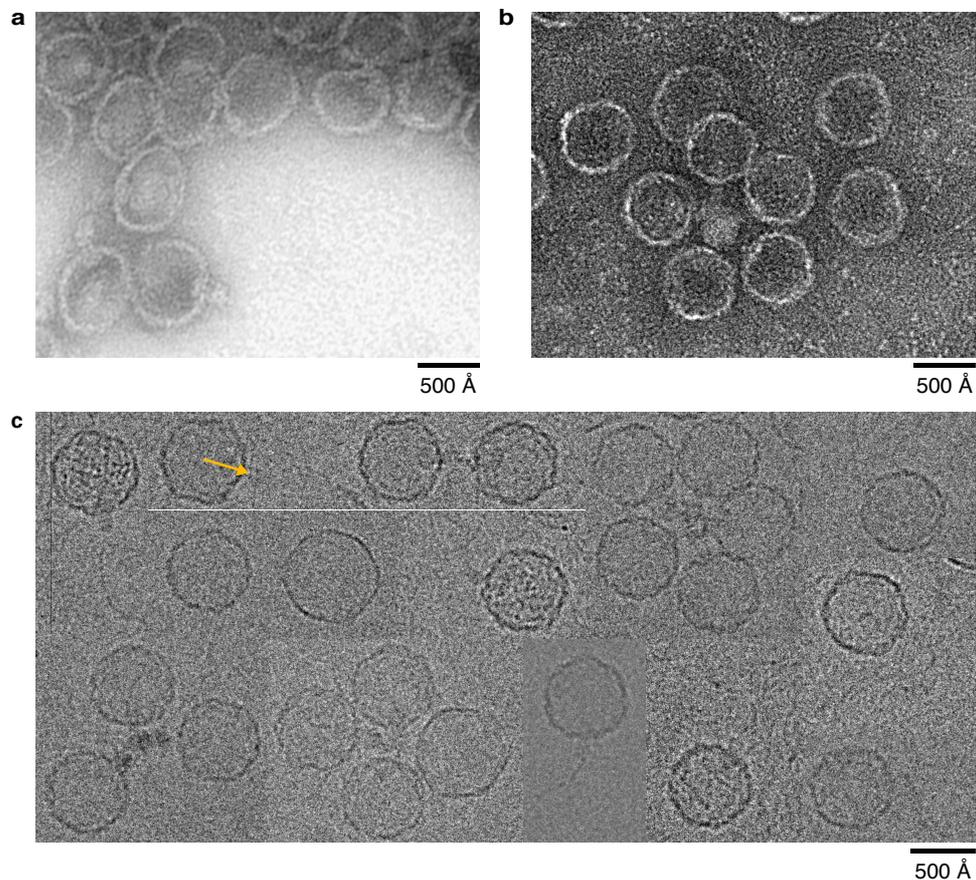


Figure 4.1. Negative-stain and cryo-images of the HK97 motor. (a) Uranyl acetate-stained sample imaged on a (a) Tecnai 12 G2 microscope using a ISS MegaView III CCD camera at 120 kV; (b) Tecnai TF20 using a Gatan Ultrascan 4000 CCD camera at 200 kV. (c) Cryo-images taken on a Polara G2 microscope using a Falcon II direct electron detector at 300 kV. Arrow indicates the vector defined during manual picking.

Conditions were subsequently optimised for cryo-EM imaging. Active HK97 assemblies were ultimately prepared using 125 nM DNA, 125 nM prohead, 1 μ M large terminase monomer and 2 μ M small terminase monomer, and 1 mM ATP. The assemblies were spotted on a Quantifoil holey carbon-coated copper grid, blotted and plunge-frozen. Because of the high viscosity of the sample, grids had to be blotted either once for 15 s, or twice for 4 s each time, to produce ice of an appropriate thickness for imaging. Free proheads, some expanded, dominated the micrographs. However, clear density for DNA and a prohead-bound macromolecular complex could be observed in some particles (Figure 4.1c). The number of these particles was low. The relative concentrations of components were likely suboptimal for motor assembly, and assembled motors might be destabilised at the air-water interface formed after blotting. Addition of non-hydrolysable analogues such as ATP γ S at 10 mM final concentration did not increase the number of motor-like particles. Nevertheless, three-dimensional reconstructions were performed using the available particles.

4.4 Cryo-EM reconstructions of the HK97 prohead and DNA packaging motor

4.2.1 Cryo-EM reconstructions of the HK97 prohead

The sizes of the observed proheads, whether free or DNA-bound, were unexpected. They were intermediate between Prohead II (77) and Head II (66), the assumed states of the HK97 capsid before and after DNA packaging, respectively. To determine if the apparent expansion was caused by preparation of a packaging reaction, the proheads were imaged again, but without terminase, DNA or ATP. Similarly sized proheads were observed, indicating that the observed conformation was inherent to the prohead preparation.

Pooling all free prohead particles, intact particles free of ice were selected via reference-free 2D classification in RELION (Figure 4.2a). A 3D starting model was generated using AUTO3DEM by assigning 200 particles with random orientations and box centres as origins and refining under icosahedral symmetry. Based on this model, two 3D classes, with diameters 600 Å and 604 Å, were identified using RELION (Figure 4.2b). These classes refined to 6.7 Å and 8.9 Å resolution, respectively (Figure 4.2c).

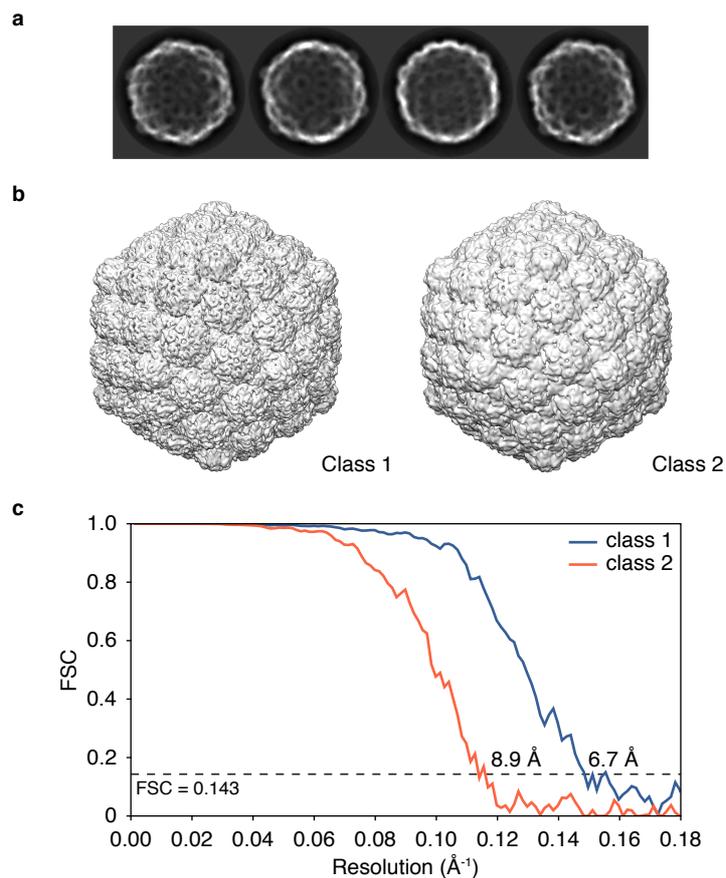


Figure 4.2. Reconstruction of the HK97 prohead. (a) 2D classes of particles. (b) 3D classes refined and low-pass filtered according to overall resolution. B factors of -543 and -615 Å² were applied, respectively. (c) Fourier shell correlation (FSC) as a function of resolution. Artefacts from masking were corrected for by subtraction of FSC following phase randomisation (232) beyond 9.2 and 13.5 Å, respectively.

Prohead II and Head II have diameters 540 Å and 659 Å, respectively. Both structures were determined using recombinant particles formed from self-assembling capsid protein. There was no portal in these particles. On the contrary, the proheads imaged were derived from a bacteriophage infection and they contain an embedded portal. The presence of a portal may have caused the capsid to expand slightly, in the absence of a molecular trigger such as packaging, before reaching a metastable state.

The HK97 capsid, when it first assembles, has skewed hexamers. Molecular dynamics-based fitting of the Prohead II structure into the 6.7 Å resolution volume suggests that hexamers become more symmetric and expanded before DNA packaging (Figure 4.3). Pentamers remain relatively unexpanded until further expansion is triggered. This agrees with a previous proposal based on portal-free EI-1 capsids (233), which were generated by mutagenesis to represent the first expansion intermediate during packaging. The capsids

have similarly expanded and symmetric hexamers. Thus, it would seem that the EI-I state occurs naturally, though it probably occurs before packaging rather than during packaging. As proposed, expansion during packaging is likely dependent on E-loop rearrangement and concomitant release of conformational strain in the spine helix.

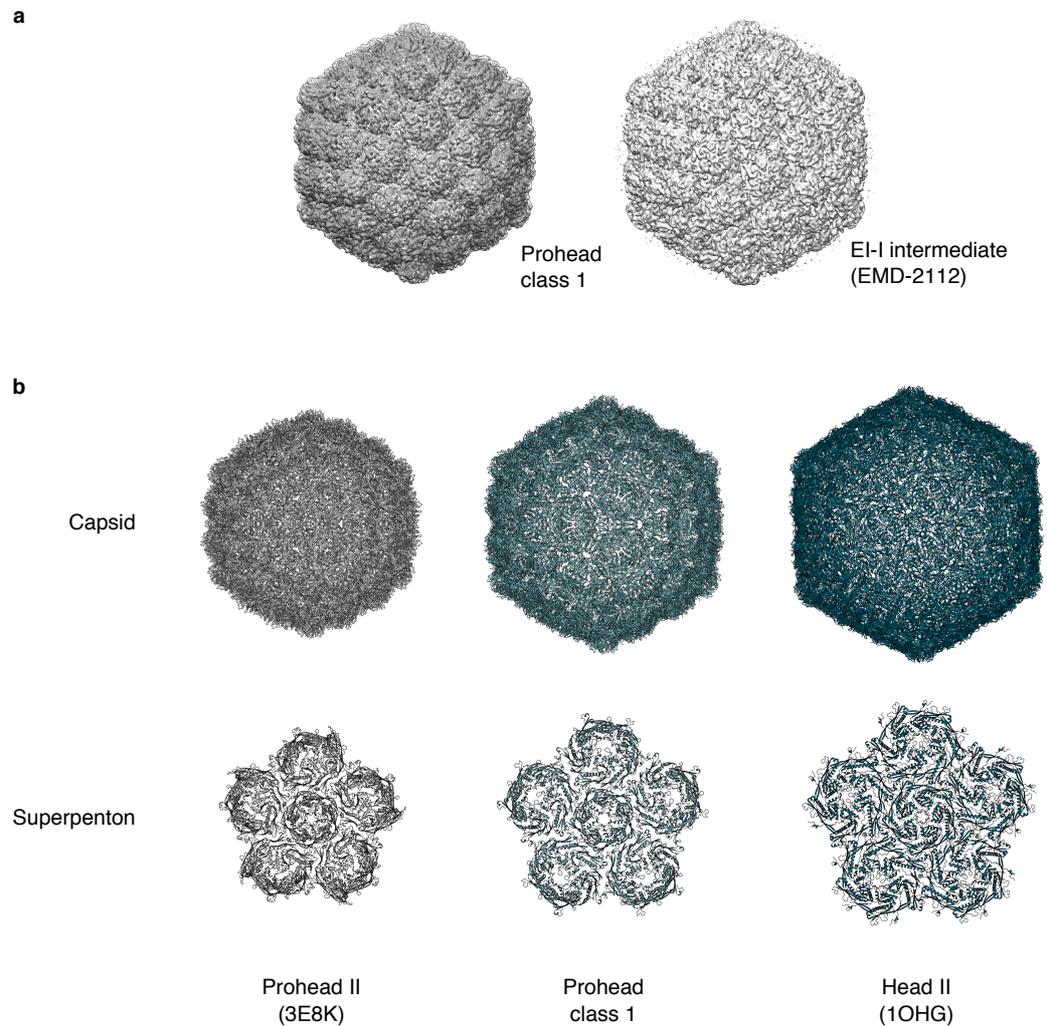


Figure 4.3. Flexible fitting of the prohead structure. (a) Prohead II structure (PDB code 3E8K) fitted into prohead reconstruction class 1. Cryo-EM reconstruction of the EI-I intermediate (EMD 2112) is shown alongside. (b) The HK97 superpenton, that is, a pentamer surrounded by five hexamers, of Prohead II, prohead reconstruction class 1 and Head II (PDB code 1OHG).

4.2.2 Cryo-EM reconstructions of the HK97 DNA packaging motor

Because the motor preparation was active at the point of freezing, the motor particles imaged were likely asymmetric. As there is also uncertainty over the stoichiometry of packaging motors, care was taken to reconstruct the motor asymmetrically.

To reconstruct the motor, the portal vertex needs to be identified and particles aligned accordingly. Initial efforts suggested that it would be difficult to identify the vertex using standard approaches which perform a global angular search. This could be due to signal from the capsid dominating during the search. Because the capsid is symmetric, it would be possible to obtain good alignments to the reference model in different particle orientations. In RELION, this meant that the true orientation would not be weighed strongly enough during calculation of the 3D volume, resulting in an averaging of densities among the 5-fold vertices, only one of which would contain portal, terminase and DNA. Thus, a symmetric relaxation approach was adopted with the use of *a priori* knowledge.

A priori knowledge was introduced by defining a vector away from the portal vertex, following the protruding density, during manual particle picking in X3D. Particles were split into half-sets according to the gold-standard procedure (234), and origins were located while applying icosahedral symmetry using AUTO3DEM. Local angular refinement, starting from the specified orientation, against the capsid with fixed origins under 5-fold symmetry allows the closest 5-fold vertex to be identified (Figure 4.4). Finally, both origins and orientations were refined in the absence of symmetry and half-maps compared.

A final reconstruction was obtained at 22.8 Å resolution from 1024 particles. Densities corresponding to the portal, terminase and DNA were observed (Figure 4.5a). Density extending inwards marks the location of the embedded portal. Density protruding outwards would accommodate at least a large terminase oligomer and DNA. Oscillation in the Fourier shell correlation between half-datasets is likely due to the limited range of defocus values in the particles selected (Figure 4.5b). Due to disorder in the terminase region, it is unclear how many large terminase subunits constitute the motor. Nevertheless, the overall architecture of the motor is conserved across phages HK97, T4 and ϕ 29 (Figure 4.5c).

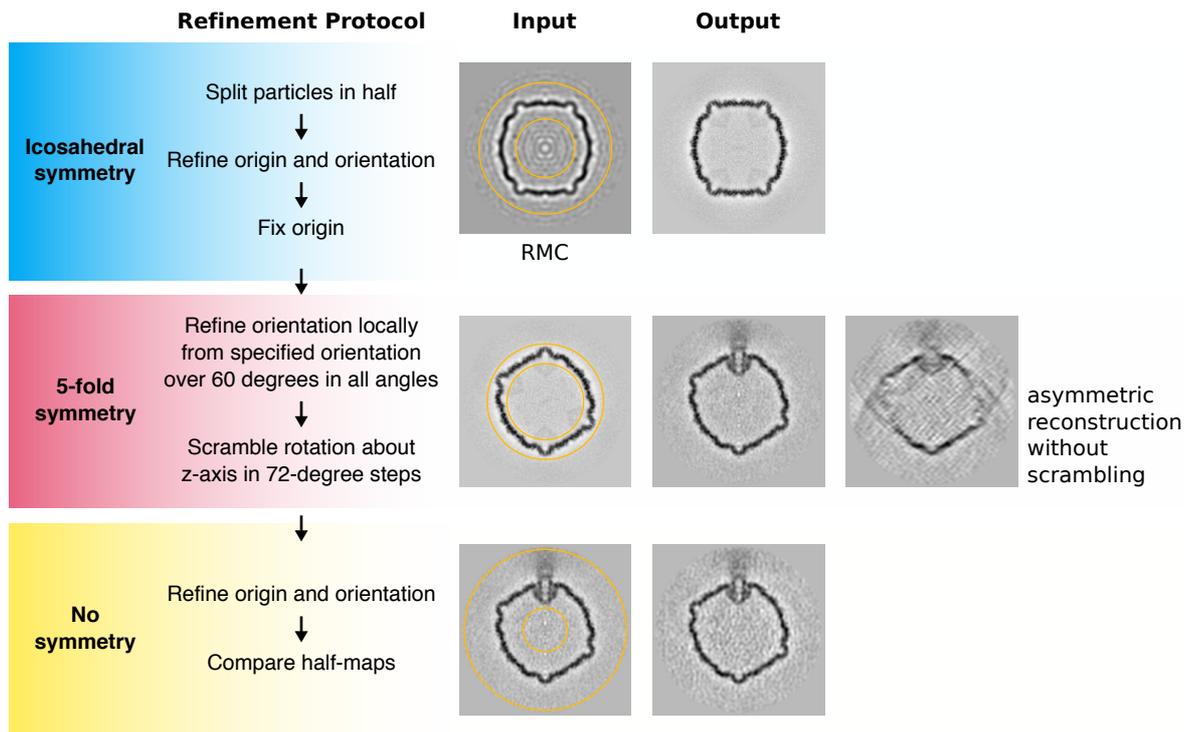


Figure 4.4. Refinement protocol for the HK97 motor. For illustration purposes, central slices of maps in the xz plane, calculated using particles from both half-sets, are shown, but maps and particles were kept separate throughout the procedure. Circles indicate boundaries of the soft spherical mask applied to the input map. The starting model was generated using the RMC algorithm in AUTO3DEM using 200 particles. Due to an arbitrary starting orientation, orientations were under-sampled in the xy-plane in the 5-fold step. This was resolved by scrambling the rotation about the z-axis in 72-degree steps.

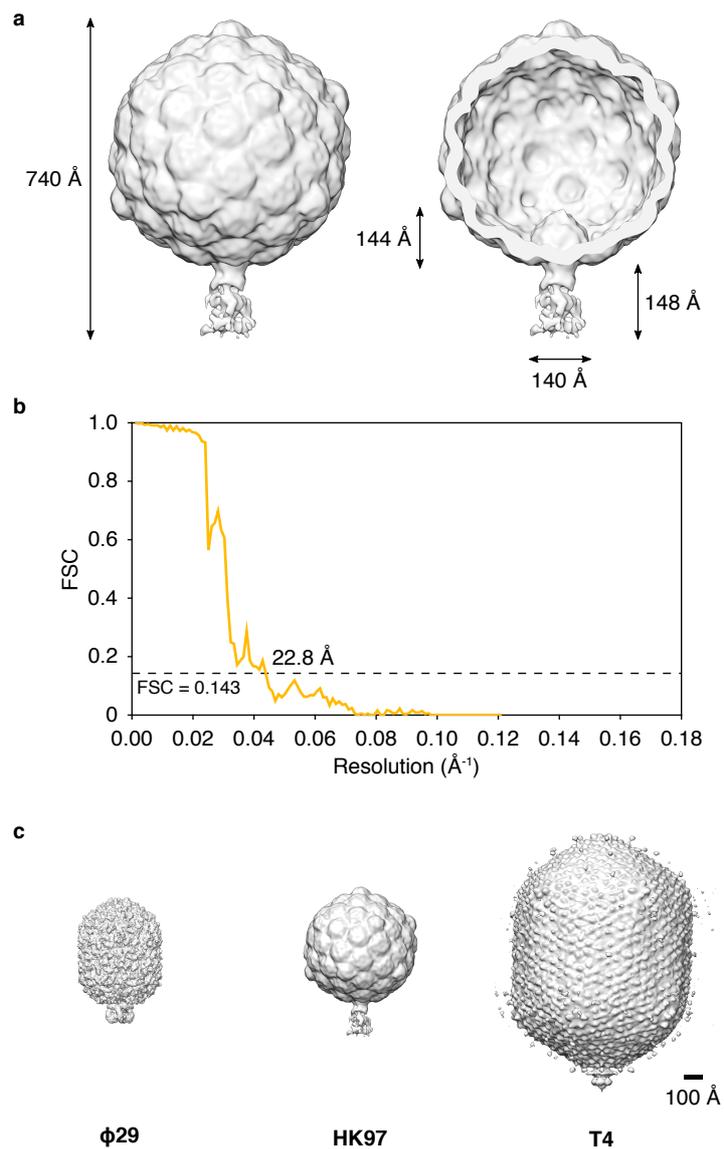


Figure 4.5. Asymmetric reconstruction of the HK97 motor. (a) Map calculated from half-maps generated by AUTO3DEM, low-pass filtered and sharpened with a B factor of -200 \AA^2 using RELION. (b) FSC as a function of resolution. (c) The ϕ 29, HK97 and T4 motors, shown to scale (EMD 1573, 6560).

4.5 Measurement of subunit stoichiometry for the HK97 motor by total internal reflection fluorescence microscopy

4.5.1 Photobleaching assay development

To measure the number of large terminase subunits in the HK97 motor, an N-terminal His-GFP-large terminase fusion was generated. By reconstituting the motor with this fusion and BOBO-3-stained 5'-biotinylated DNA, stalling with ATP γ S and tethering to a streptavidin-coated surface, the motor could be visualised by total internal reflection fluorescence microscopy (TIRFM; Figure 4.6a). Despite fusion with GFP and staining with BOBO-3, the reconstituted motor still possessed DNA packaging activity (Figure 4.6b).

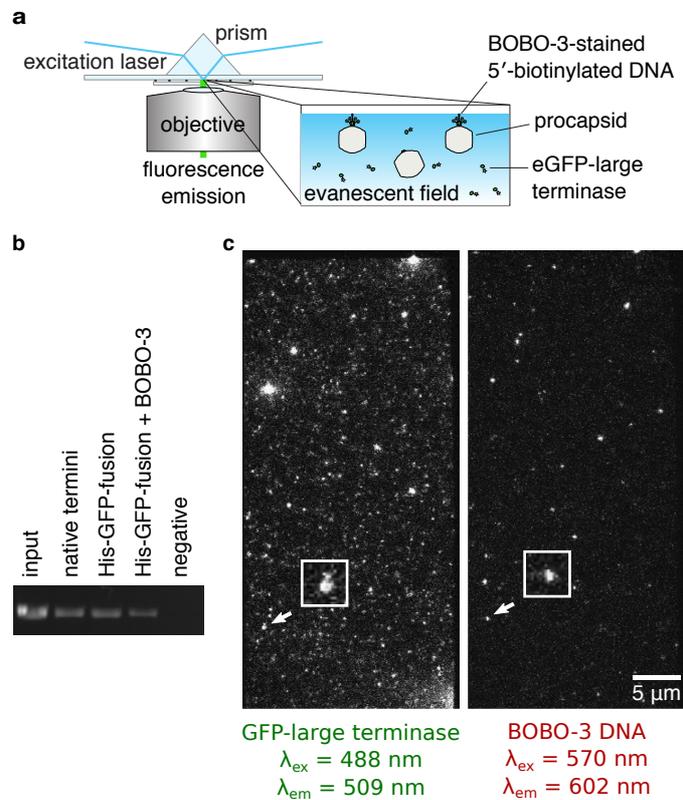


Figure 4.6. Design of the TIRFM photobleaching assay. (a) BOBO-3-stained 5'-biotinylated DNA was pre-incubated with prohead, small terminase and His-GFP-large terminase protein. ATP was added to initiate translocation and chased with ATP γ S. The complex was immobilised on a streptavidin-coated surface for visualisation by TIRFM. (b) DNA packaging by the His-GFP-large terminase fusion in the presence of BOBO-3 stain. Reactions were subjected to a DNase protection assay as described in Chapter 2. (c) Still image of the green and red channels during the TIRFM experiment. Insets, enlarged view of an event colocalising in the green and red channels.

On excitation, protein would fluoresce green and DNA red. This enables identification of motor assemblies by co-localisation. The translation vector relating the green and red imaging channels was determined by dual-excitation of surface-adsorbed TetraSpeck microspheres, which would fluoresce in both channels (Figure 4.7a). Microspheres were identified in a single-frame image by binarisation and connected-component labelling. Coordinates were determined based on intensity-weighted centroids and the average vector between each pair of observations was calculated. This calibration was performed before every experiment. In addition, it provided a measure of precision in the localisation of events based on intensity-weighted centroids (Figure 4.7b).

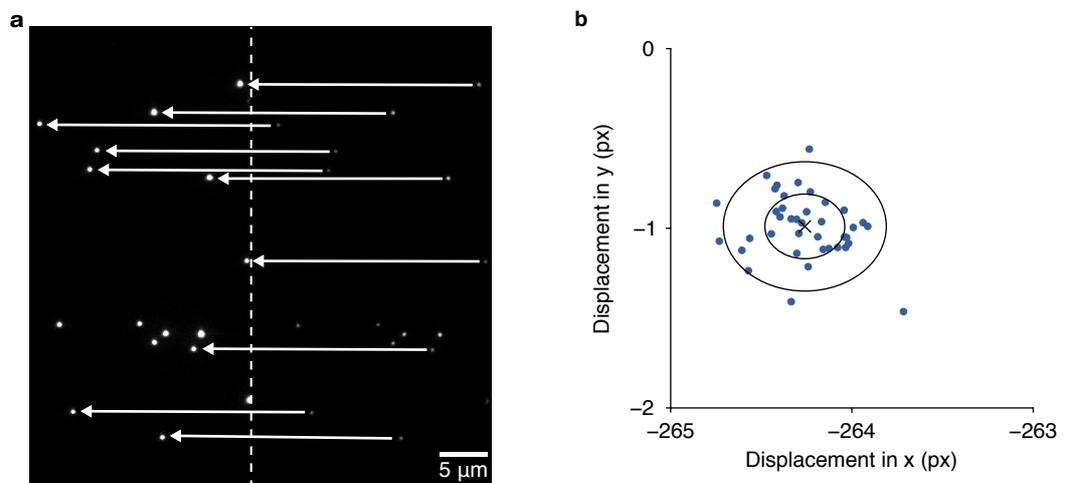


Figure 4.7. Alignment of image channels using TetraSpeck microspheres. (a) Still image. Arrows indicate the vector between matching point spread functions. (b) Precision in the calibration vector.

Photobleaching is a stochastic, irreversible event. It occurs when a fluorophore transitions from an excited singlet state or an excited triplet state, and undergoes a chemical change to a dark state, which is unable to fluoresce (235). When a spot containing multiple fluorophores is illuminated, individual fluorophores will photobleach at random times, resulting in a series of steps in the recorded intensity over time. The number of steps would correspond to the number of emitting fluorophores originally present. Using this approach, the number of large terminase subunits in individual motor assemblies was counted.

4.5.2 Routines for photobleaching data analysis

Different step-finding algorithms were trialled to count steps in a consistent and unbiased manner. A moving two-sample t -test (236) and chi-squared minimisation-based method (230), used originally for the tracking kinesin and microtubule by bead, were adapted to analyse the recoded intensities. However, both algorithms resulted in a systematic under-fitting or over-fitting of steps, due to the low signal-to-noise ratio of the experiment. A suitable algorithm was eventually identified — the PIF algorithm, which was developed to measure the stoichiometry of membrane proteins (228), combined with Chung-Kennedy filtering to reduce noise (229). To guard against over-fitting, a criterion introduced previously for microtubule tracking was implemented (230). For each fit, a counter-fit was generated, where steps are placed in between the identified steps and the original steps eliminated. The chi-squared statistics of this counter-fit and the fit were compared, and fits rejected if the counter-fit had a smaller chi-squared statistic value.

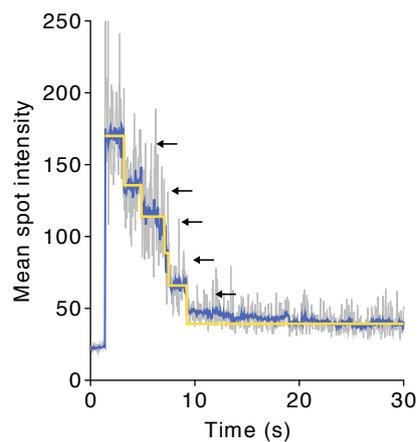


Figure 4.8. GFP fluorescence of a protein spot co-localising with DNA under constant illumination. Grey, raw intensities; blue, Chung-Kennedy-filtered intensities; yellow, fitted steps using the PIF algorithm.

A criterion for co-localisation was also introduced. An event was accepted only if the intensity-weighted centroid of the protein spot analysed was within experimental error of the centroid for the DNA. A centroid-based measure was used because the density of events in the green channel was high and, in turn, background pixels surrounding each event could not be reliably defined for the fitting of Gaussian distributions. To adopt the method used by Gelles et al. (231), experimental error was defined as two standard deviations with which the x - and y -position of one GFP monomer could be assigned in

time relative to another monomer. For the current configuration of optics and camera, two standard deviations measured 80.8 nm, less than the calibrated size of a pixel.

4.5.3 Photobleaching data analysis

In a control experiment, a flow chamber was prepared where DNA was immobilised but only GFP-large terminase was applied and subsequently washed away. Acetylated BSA and PEG 6000 were added to passivate the surface, but significant adsorption of the GFP-large terminase to the imaging surface still occurred. This gave rise to a background of mostly single-step spots which did not co-localise with DNA throughout the experiments (Figure 4.6c, Table 4.1, Figure 4.9).

In order to observe multi-step photobleaching events that co-localised with DNA, it was necessary to add an excess of large terminase and proheads during slide preparation to reconstitute the motor. Still, the proportion of protein co-localising with DNA was low (Table 4.1), similar to what was observed by EM. A distribution of step counts was observed (Figure 4.9). Fitting a binomial distribution where the motor has five subunits, by maximum likelihood, each subunit has a 72% probability of being fluorescent (Figure 4.10). This probability is widely reported in the literature (228), as GFP can misfold or photobleach before an experiment. Higher-order assemblies were observed, too, which did not co-localise with DNA. It could be that the protein oligomerised spontaneously, or that DNA was present but not stained.

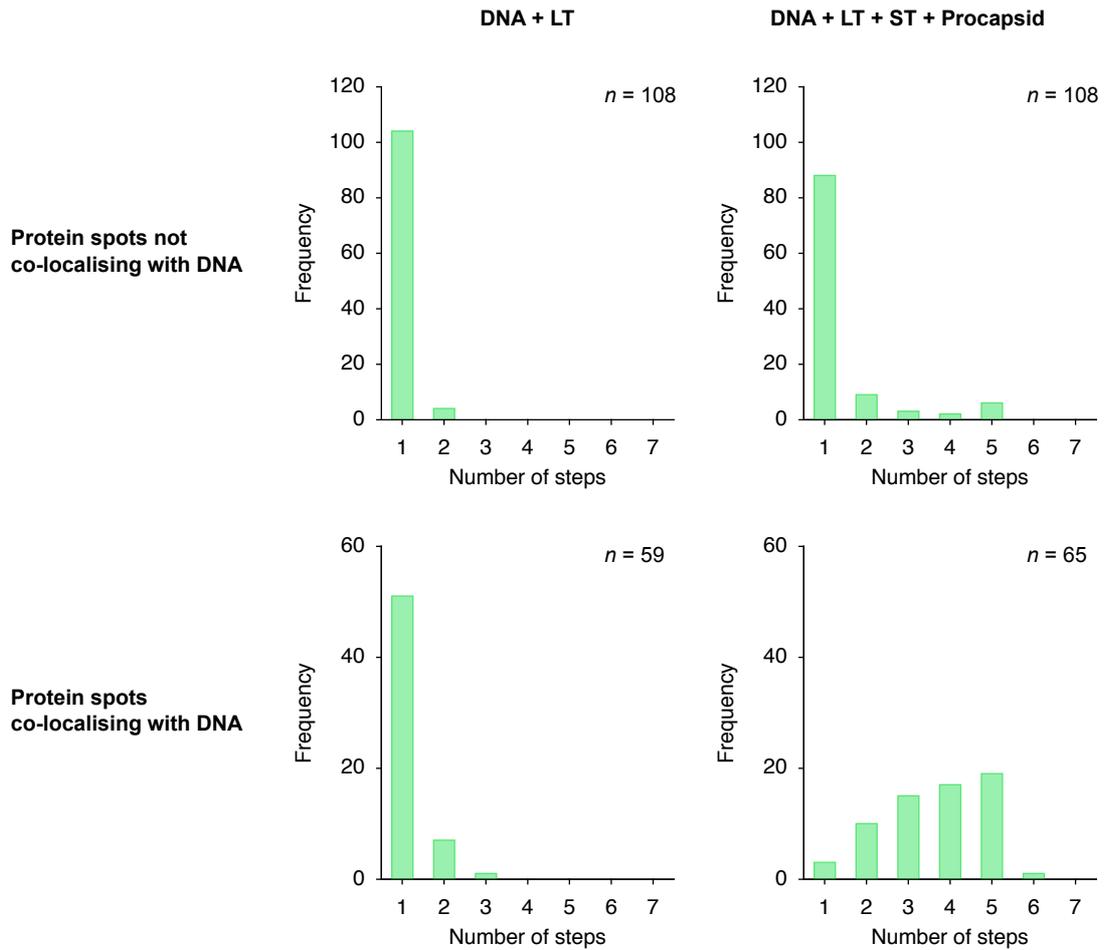


Figure 4.9. GFP photobleaching step counts. Left, control experiment with immobilised DNA and GFP-tagged large terminase (LT). Right, immobilised motor complexes. Events co-localising and not co-localising with DNA are displayed in separate histograms.

Table 4.1. Event statistics. Events were discarded if the counter-fit of steps had a smaller chi-squared value.

	DNA + LT	DNA + LT + ST + Prohead
Number of DNA spots	699	367
with no co-localising protein signal	634	259
with co-localising protein signal, discarded	6	43
with co-localising protein signal, analysed	59	65

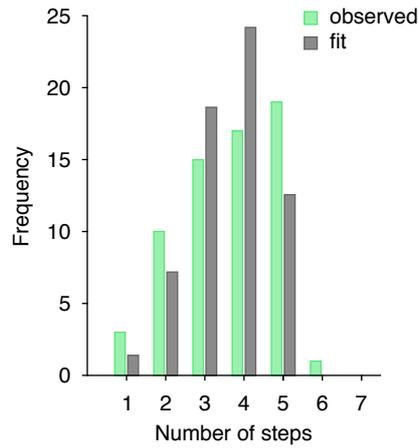


Figure 4.10. Binomial distribution of step counts for a motor containing five large terminase subunits. The probability of observing an emitting GFP was estimated to be 72% by maximum likelihood.

Due to errors in the experiment, the true distribution of step counts would be convolved with additional probability distributions to give the observed histogram. Given a measured photobleaching rate of 0.10 s^{-1} , which is typical of GFP proteins (237), and the low signal-to-noise ratio, two photobleaching steps could be mistaken for one if the steps occurred within several video frames of each other (238). The probability of this would increase with the number of fluorophores present. Also, motor events could be coincident with other passively adsorbed monomers, resulting in additional steps. The small number of lower-order species observed in the control experiment could further contribute to the shape of the histogram in the motor experiment. It would be difficult to model these errors. However, the clear drop in the number of counts from five to six convincingly places the upper limit for the number of large terminase subunits in the HK97 motor at five.

4.6 Discussion

Cryo-EM reconstructions of the HK97 motor revealed a molecular architecture similar to those of phages T4 and ϕ 29 (Figure 4.5). In all studied systems, an oligomeric terminase assembly associates with the portal vertex of a prohead to form the packaging motor.

The stoichiometry of viral DNA packaging motors has been a subject of long debate. Analytical ultracentrifugation data on the λ maturation complex, which features during packaging initiation, suggests that the complex is a tetramer of heterotrimers, each trimer comprising two small and one large terminase subunits. Although this would complement the established 12-fold symmetry of a capsid-embedded portal, it is unclear whether large terminase remains a tetramer in the λ motor, as the motor has not yet been visualised.

TIRFM photobleaching data suggest that the HK97 large terminase is a pentamer in the packaging motor (Figure 4.10). This agrees with the stoichiometry observed in *pac* and ϕ 29-like packaging motors. However, it raises again the issue of symmetry mismatch between motor components, since the terminase pentamer would be associating with a 12-subunit portal at a 5-fold vertex of the capsid. The apparent mismatch is partially resolved in ϕ 29 because five pRNA molecules bridge between the capsid and ATPase subunits (124). A recent structure of phage P22 portal indicates that portals can adopt a pseudo 5-fold symmetric conformation, eliminating mismatch (143). A similar strategy may be adopted in HK97.

It is unclear where small terminase forms a part of the motor during DNA translocation. Because the sample imaged was heterogeneous, density in the cryo-EM reconstruction distal to the prohead could correspond to DNA in different conformations, or small terminase. To resolve this ambiguity, motors assembled in the absence of small terminase (Figure 2.4) could be imaged, reconstructed, and a difference density map calculated.

Ultimately, a higher-resolution cryo-EM reconstruction will help to elucidate the stoichiometry and molecular detail of the HK97 motor. Currently, resolution is limited by heterogeneity in the complex and low probability of motor formation. Methods for stalling and stabilising the motor need to be developed. The motor may be stalled using roadblocks. This may be streptavidin bound to a biotin label on the DNA, or small terminase on a *cos*-containing substrate at limiting ATP concentrations (Figure 2.10).

Adding a mixture of ATP, ADP, ground state analogue ATP γ S, and transition state analogues such as ADP-AlF₄ may help to select for a particular motor state. Enrichment may be possible by reconstituting the motor on biotinylated DNA, pulling down with streptavidin-coated beads followed by a restriction digest to liberate the complex. To prevent dissociation of the complex due to equilibrium effects, crosslinking may be necessary at the same time. Grafix is a promising method which has already been used for structure determination by cryo-EM (239). Alternatively, photoreactive protein-DNA crosslinkers such as psoralen (240) may be used to fix the complex upon motor assembly. With more particles and greater uniformity, 3D classification of particles would be possible. Particles with large terminase pointing into and out of the imaging plane might become identifiable, increasing the range of orientations sampled. The resolution of the reconstruction might improve as a result. Combined with focussed refinement of individual components, more detailed insights would be gained into the mechanism of DNA translocation by the HK97 motor.

5. Structure and function of the HK97 large terminase

5.1 Introduction

The large terminase proteins of *cos* and *pac* viruses have multiple functions. They cleave DNA, and they translocate DNA. TIRFM data in the previous chapter suggest that the HK97 large terminase forms a pentameric assembly in the packaging motor. As such, it has the same oligomeric state as the T4 large terminase and ϕ 29 packaging ATPase. The mechanisms by which these proteins are regulated are generally unclear. Viral packaging ATPases form a unique family in the ASCE superfamily of ATPases. There is debate over whether the ATPase is activated by a *cis*-acting arginine or a *trans*-acting arginine finger. To date, there is no structure of a *cos* large terminase. Only two full-length structures are available, both derived from a *pac* virus. Thus, the *apo* structure of HK97 large terminase was determined, described here in this chapter. Mutagenesis experiments coupled with ATPase and packaging assays reveal a third possible mode of ATPase regulation.

5.2 Materials and methods

5.2.1 Protein production

The coding sequence for HK97 large terminase was sub-cloned into a pET22a-based vector using the In-Fusion Cloning system (Clontech) to generate a fusion construct with a non-cleavable N-terminal His-tag (MGSSHHHHHH). Site-directed mutagenesis of the His-fusion construct was performed using an adapted QuikChange protocol (241) where primers are extended in the 3' direction beyond the region in which they overlap. For selenomethionine labelling, cells were grown in M9 minimal medium supplemented with 30 μ g/mL kanamycin, ammonium chloride, glucose, vitamins and trace metals (179). At an OD₆₀₀ of 0.6, lysine, threonine and phenylalanine were added to a concentration of 100 μ g/mL, and leucine, isoleucine and valine to a concentration of 50 μ g/mL to suppress methionine biosynthesis (242). Selenomethionine was subsequently added to a concentration of 200 μ g/mL and overexpression was induced with 0.4 mM IPTG. For crystallisation, proteins were purified as described for the N-terminal His-SUMO-fusion, except that no SUMO protease was added during dialysis. Mutants were purified using loose Ni-NTA Agarose resin (Invitrogen), washing and eluting in gravity flow columns, followed by dialysis, anion exchange chromatography on a MonoQ 5/50 column, buffer-

exchanged into 20 mM Tris, 300 mM potassium glutamate, 2 mM DTT, pH 7.5, and concentrated to ~5 mg/mL. Mutants were placed into potassium glutamate instead of NaCl to avoid NaCl poisoning in the packaging assay given their low stock concentrations.

5.2.2 Crystallisation and structure determination

Native crystals of His-tagged large terminase (20 mg/mL in 20 mM HEPES, 250 mM NaCl, 5% (v/v) glycerol, 2 mM DTT, 20 mM MgSO₄, 10 mM AMP-PNP, pH 7.5) were obtained by hanging drop vapour diffusion against 1.6 M ammonium sulfate, pH 7.6, with streak-seeding. Crystals of selenomethionine-labelled protein were obtained against 1.5 M ammonium sulfate pH 7.6, by streak-seeding first with native crystals then with its own crystals. The selenomethionyl crystals were cryo-protected in 10 mM HEPES, pH 7.5, 5 mM MgSO₄, 4 mM AMP-PNP, 1.9 M ammonium sulfate, 21% (v/v) glycerol, before freezing in liquid nitrogen. Final native crystals were soaked in 10 mM HEPES, pH 7.5, 1.8 M ammonium sulfate, 0.5 M KCl, 320 mM KBr, 10 mM MgSO₄, 5 mM AMP-PNP, 20% (v/v) glycerol. Data were collected at 100 K at Diamond Light Source beamlines I02 and I03, integrated and scaled with XDS (181). For MAD experiments, data were collected from the same crystal at the peak and inflection point of the Se absorption edge with exposure times of 0.04 s in 0.1° oscillations over 360°. Native data were collected over 200° with a less attenuated beam, which produced stronger diffraction. Substructure solution, phase calculation, and iterative density modification and chain tracing were carried out using SHELX (182). More of the model was built and side chains added using Buccaneer (185). The model was manually completed in Coot (186) and refined against the final native data using REFMAC (187). SAD log-likelihood gradient maps for Br were calculated using Phaser (183).

5.2.3 Size exclusion chromatography with multi-angle laser-light scattering analysis

The HK97 large terminase (100 µL 1.2 mg/mL) was applied to a Superdex 75 10/300 GL column (GE Healthcare) in 20 mM HEPES, 300 mM KCl, 1 mM DTT, pH 7.0. Eluate was analysed using an on-line Dawn HELEOS-II light scattering detector and Optilab rEX refractometer (Wyatt). Data were processed based on a refractive index increment of 0.183 mL/g and scattering intensities at seven different angles in ASTRA 5.3.4 (Wyatt).

5.2.4 Sedimentation velocity analytical ultracentrifugation

Epon double-sector centrepieces were filled with protein (2.5–20 μM in 20 mM HEPES, 300 mM KCl, 1 mM DTT, pH 7.0) and buffer, and centrifuged in an An-60 Ti rotor at 50000 rpm, 20 $^{\circ}\text{C}$, for 7.5 h using a Beckman Optima XL/I ultracentrifuge. Absorbance scans were acquired at 280 nm at 7.5 min intervals. A protein partial specific volume estimate of 0.73747 mL/g, buffer density of 1.01390 g/mL and buffer viscosity of 0.01002 P were calculated using Sednterp (243). Data were analysed under the continuous $c(s)$ distribution model using Sedfit (244).

5.2.5 Measurement of ATPase activity

Reactions containing 1 μM large terminase, with or without 25 nM ScaI-linearised *cos*-containing pUC18 DNA, 25 nM procapsid and 2 μM small terminase, were prepared in 100 μL at room temperature, that is, 20–22 $^{\circ}\text{C}$. For arginine finger complementation assays, mutant E154A was added to 1 μM final concentration. The initial rate of phosphate release at different ATP concentrations was measured using the EnzChek Phosphate Assay Kit (Molecular Probes). Briefly, purine nucleoside phosphorylase was used to convert 2-amino-6-mercapto-7-methylpurine ribonucleoside (MESG) and inorganic phosphate to 2-amino-6-mercapto-7-methylpurine, which absorbs at 360 nm, and ribose 1-phosphate (Figure 5.1). Absorbance was proportional to concentration between 2 μM and 75 μM phosphate concentration. Reactions were monitored over 5–30 min using a Biochrom Lightwave II UV-Vis spectrophotometer in a 50 μL Hellma quartz cuvette with a 1-cm light path.

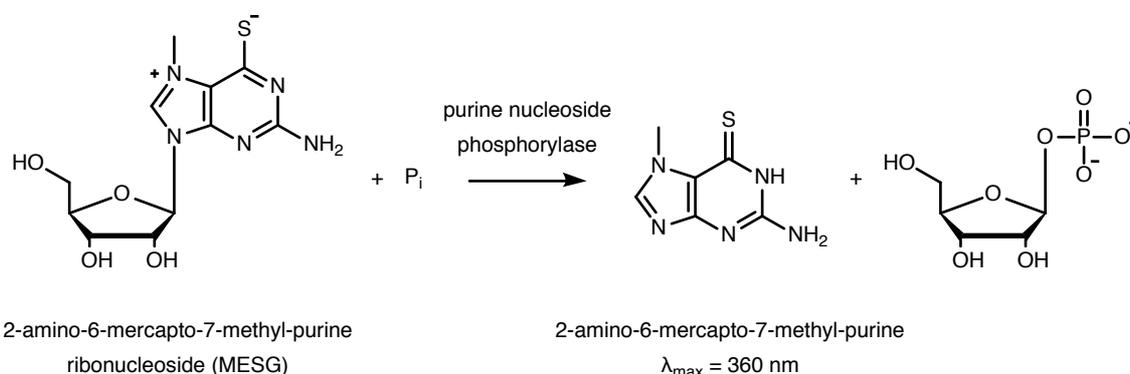


Figure 5.1 Enzyme-coupled ATPase assay. The amount of inorganic phosphate released is measured in real time by conversion of MESG to 2-amino-6-mercapto-7-methylpurine, which absorbs at 360 nm.

Kinetic data were analysed by fitting the Michaelis-Menten equation:

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

where v is the rate of reaction, V_{\max} is the maximum rate of reaction, $[S]$ is the substrate concentration, and K_m is the concentration at which the rate of reaction is half-maximal; and the Hill equation:

$$v = \frac{V_{\max} [S]^n}{K^n + [S]^n}$$

where n is the Hill coefficient. Values of n greater than 1 indicate cooperativity. Non-linear least-squares fitting was performed using Origin v2017 (OriginLab).

5.2.6 Far-UV circular dichroism spectropolarimetry

Far-UV circular dichroism spectra were recorded using a JASCO J-810 spectropolarimeter at the University of York, or a Applied Photophysics Chirascan Plus spectropolarimeter at the University of Leeds, at 20 °C with a 1-mm pathlength, 1-nm bandwidth, 50-nm/min scan rate, 1-s response time, averaged over five scans and buffer-corrected. Protein was buffer-exchanged into 20 mM Tris sulfate, 50 mM Na₂SO₄, 10 mM MgSO₄, 0.5 mM Tris(2-carboxyethyl)phosphine hydrochloride, pH 7.5, using Micro Bio-Spin polyacrylamide P-6 columns (Bio-Rad), and diluted to final concentrations of 2–5 μM.

5.3 Crystal structure of the HK97 large terminase

5.3.1 Crystallisation

For crystallisation, an N-terminal His-fusion with no intervening linker was used. Because protease cleavage was no longer necessary or possible, most of the overproduced protein could be recovered after purification. The first crystals of the protein were obtained by sitting drop vapour diffusion against 1.4 M ammonium sulfate, pH 7.0, 0.5 M LiCl (Figure 5.2a). These grew in stacks and diffracted poorly. Single crystals with better morphology were eventually obtained in hanging drops by micro-seeding against 1.4 M ammonium sulfate, pH 7.0, 0.5 M LiCl (Figure 5.2b) and streak-seeding against 1.6 M ammonium

sulfate, pH 7.6 (Figure 5.2c). A similar approach, based on cross-seeding and streak-seeding, was employed to obtain diffracting selenomethionyl crystals (Figure 5.2d).

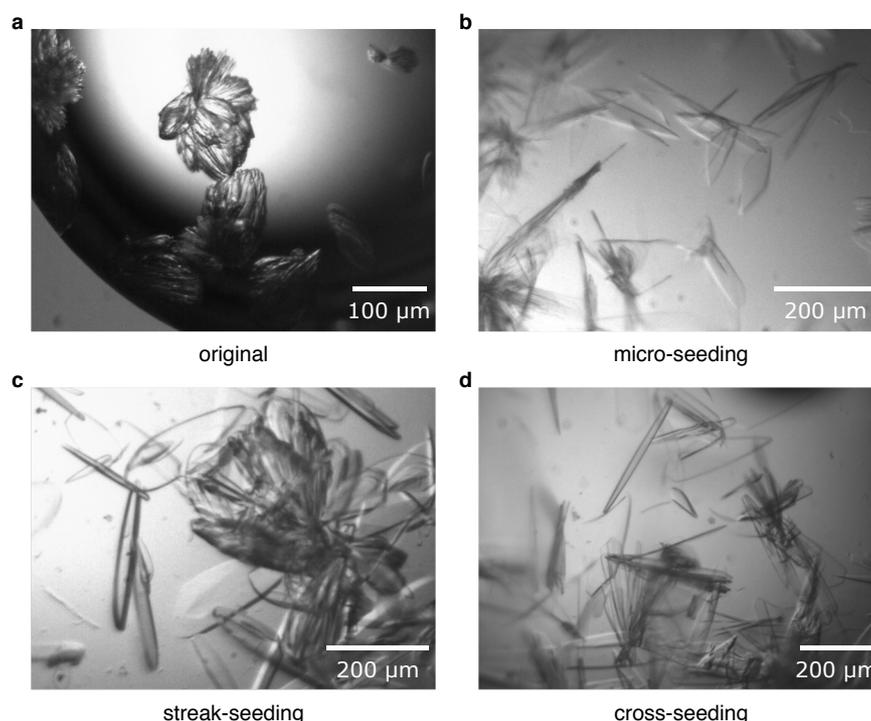


Figure 5.2. Optimisation of HK97 large terminase crystals. (a) Native crystal in sitting drop against 1.4 M ammonium sulfate, pH 7.0, 0.5 M LiCl. (b) Crystal in hanging drop with micro-seeding, equilibrated against 1.4 M ammonium sulfate, pH 7.0, 0.5 M LiCl. (c) Crystal with streak-seeding, equilibrated against 1.6 M ammonium sulfate, pH 7.6. (d) Selenomethionyl crystal obtained by cross-seeding with native crystals.

It was noted initially that crystals were more likely to appear in the presence of AMP-PNP. Thus AMP-PNP was added to the protein before each experiment. Later biochemical efforts would reveal that the concentrations used were insufficient for good occupancy of the ligand in the ATPase active site. As a ground state analogue, the ligand might have helped to stabilise a particular ATPase conformation. It might have also simply acted as a hydrotrope, altering interactions between the protein and the solvent (245).

In the first structures determined, the C-terminal nuclease domain was partially disordered (Figure 5.3f). Better-ordered densities were obtained by soaking native crystals in cryoprotectant containing 0.5 M KCl and 320 mM KBr. This might have altered the hydration shell surrounding the protein and consequently protein dynamics in solution.

5.3.2 Structure determination

Molecular replacement using differently pruned ATPase and nuclease domain structures did not yield a good solution. Rebuilding these solutions with MR-Rosetta did not improve the phases. The fusion protein has 17 methionine residues out of a total of 514. Adequate anomalous signal should result given sufficient labelling. Thus, selenomethionyl protein was overproduced, crystals grown and data collected on a tuneable-wavelength beamline for experimental phasing. The peak and inflection point of the Se absorption edge were identified by X-ray fluorescence (Figure 5.3a) and a MAD dataset was collected (Table 5.2). The anomalous signal was sufficiently strong for substructure solution using SHELXD (Figure 5.3b,c). Peaks corresponding to all but the N-terminal methionine residues were present in the anomalous difference Fourier map. Densities were traced and the model completed using SHELXE and Buccaneer, respectively. The C-terminal domain was partially disordered at this point (Figure 5.3d); only 438 of 517 residues could be built. The C-terminal domain became ordered after soaking in cryoprotectant containing 0.5 M KCl and 320 mM KBr (Figure 5.3e). The remaining residues were built manually in Coot. Knowledge of the location of methionine residues from the MAD experiment helped to determine the register during model building. The final structure contains all residues except the N-terminal His-tag, residues 19–22 and residues 484–504 of the HK97 large terminase protein. Bromide ions were placed according to the local chemical environment and peaks in the Phaser log-likelihood gradient map for Br, calculated using phases from the model without solvent molecules.

Table 5.1. Crystallographic data collection and refinement statistics.

	SeMet		KCl / KBr soak
	<i>peak</i>	<i>inflection</i>	
Data collection			
Wavelength (Å)	0.9795	0.9797	0.9198
Space group	<i>C2</i>	<i>C2</i>	<i>C2</i>
Cell dimensions			
<i>a, b, c</i> (Å)	211.5, 39.3, 66.0	211.9, 39.3, 66.1	211.8, 39.2, 65.5
β (°)	103.3	103.2	103.4
Resolution (Å) [†]	49.59–2.00 (2.05–2.00)	49.72–2.40 (2.49–2.40)	45.53–2.20 (2.27–2.20)
R_{merge}	0.122(1.233)	0.100 (1.249)	0.136 (0.842)
$I / \sigma(I)$	10.1 (1.0)	13.6 (1.6)	7.1 (1.8)
CC _{1/2} (%)	99.7 (36.5)	99.9 (52.9)	99.1 (59.5)
CC _{anom} (%)	69.9 (2.7)	35.4 (–0.5)	25.5 (5.8)
Completeness (%)	98.3 (85.3)	99.9 (99.7)	99.6 (98.2)
Multiplicity	6.2 (4.1)	6.6 (6.4)	3.3 (3.3)
Wilson B (Å ²)	43.3	60.5	37.6
No. unique reflections	35605 (2257)	21206 (2198)	26796 (2242)
Refinement			
Resolution (Å)			45.53–2.20
No. reflections			
Working			25551
Free			1244
$R_{\text{work}}/R_{\text{free}}$			0.205/0.254
No. atoms			
Protein			3522
Water			93
Ligand			18
B factors			
Protein			38.9
Water			33.3
Ligand			53.1
R.m.s. deviations			
Bond lengths (Å)			0.008
Bond angles (°)			1.3
Ramachandran			
Favoured (%)			97.5
Outlier (%)			0.0

[†] Values in parentheses correspond to the highest resolution shell.

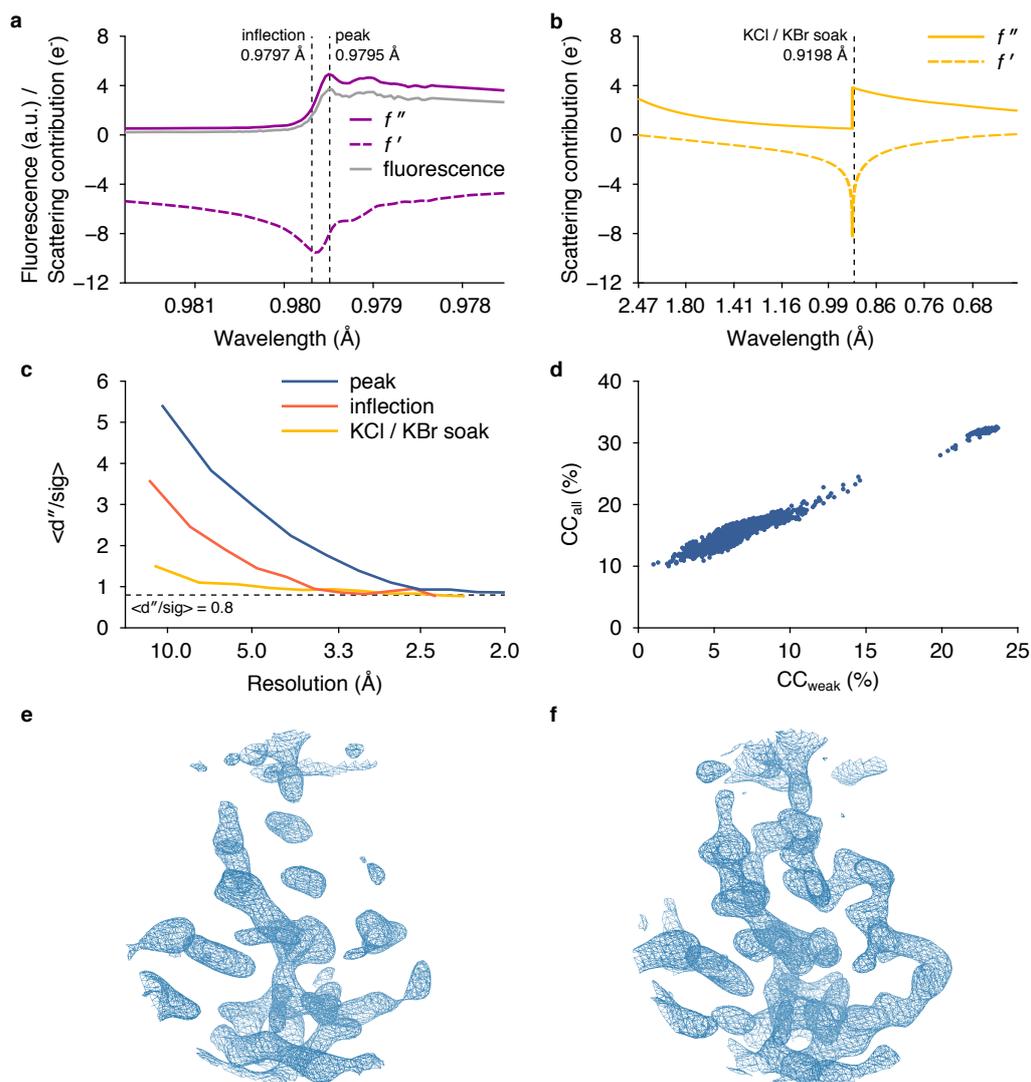


Figure 5.3. Experimental phasing and model building. (a) Fluorescence spectrum of the selenomethionyl crystal and experimental anomalous scattering contribution, calculated using CHOOCH (246). (b) Theoretical anomalous scattering contribution of bromine (209,210). Dotted lines indicate the wavelengths at which data were collected. (c) $\langle d''/\text{sig} \rangle$ as a function of resolution. (d) CC_{all} and CC_{weak} of SHELXD trials for MAD phasing. (e,f) Electron density maps in the C-terminal nuclease domain, with and without soaking in 0.5 M KCl and 320 mM KBr.

5.3.3 The HK97 large terminase crystal structure

The HK97 large terminase protein crystallised as a monomer (Figure 5.4). No higher-order functional assemblies were apparent in the crystal packing. The structure comprises an ATPase and a nuclease domain (Figure 5.5). The ATPase domain has a three-helix lid and an ATPase core. The ATPase core is an eight-stranded β -sheet with six parallel strands and two anti-parallel strands. The C-terminal nuclease domain has a RNase H-like topology.

A short linker connects the two domains. Following the linker is a unique β -strand that is absent in other large terminase structures (Figure 5.5). This strand forms a β -sheet with the conserved hairpin of the nuclease domain, known to control DNA accessibility for DNA cleavage (109,125). As such, the strand may be involved in crosstalk between domains, forming part of a switch that toggles between translocase and nuclease activities during packaging. Additional secondary structure elements are present in the nuclease domain. Compared to other nuclease structures, there is an $\alpha\beta$ insertion after the anti-parallel section of the β -sheet (Figures 5.5, 5.6). This may be a feature of *cos* phages, or a feature specific to HK97-like viruses, which act in tandem with an endonuclease cofactor such as GP74.

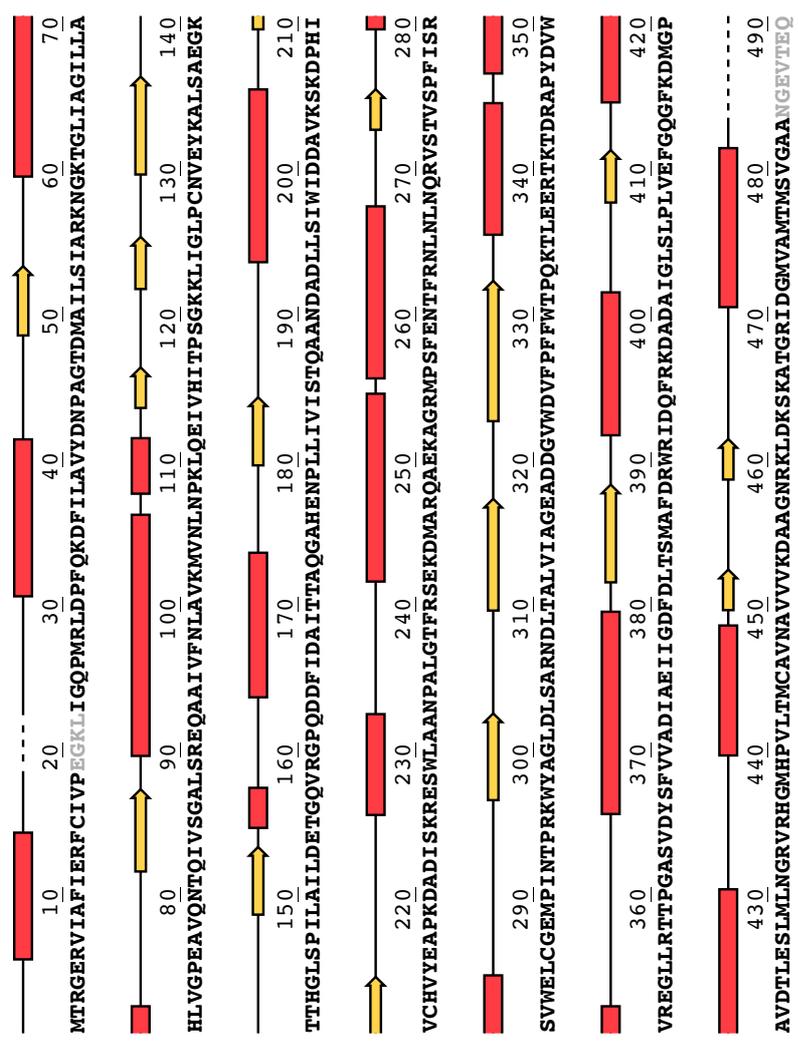
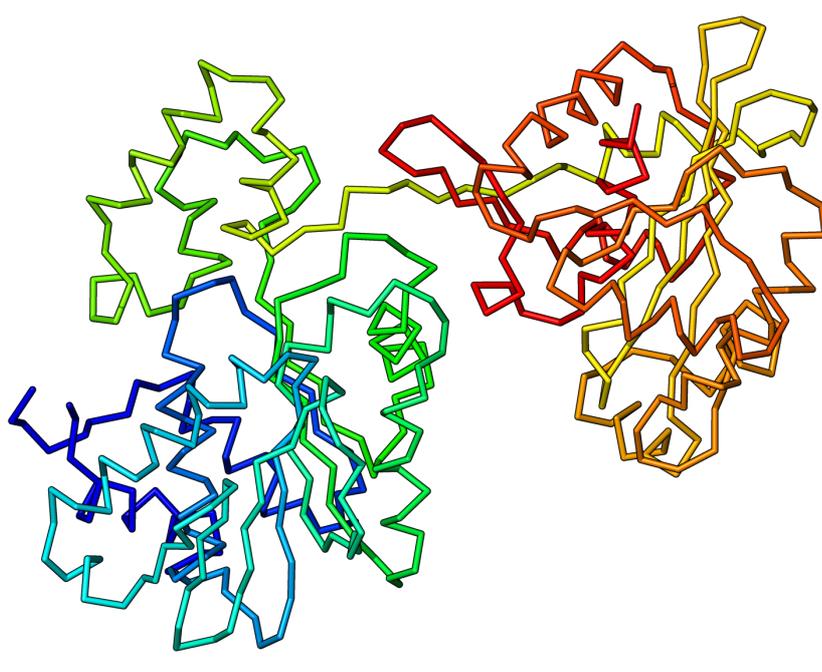


Figure 5.4. Secondary structure profile and ribbon representation of HK97 large terminase. Helices are indicated by red rectangles, strand by yellow arrows, random coil by a solid line, and unresolved residues by a dotted line. The ribbon diagram is coloured blue to red from N- to C-terminus.

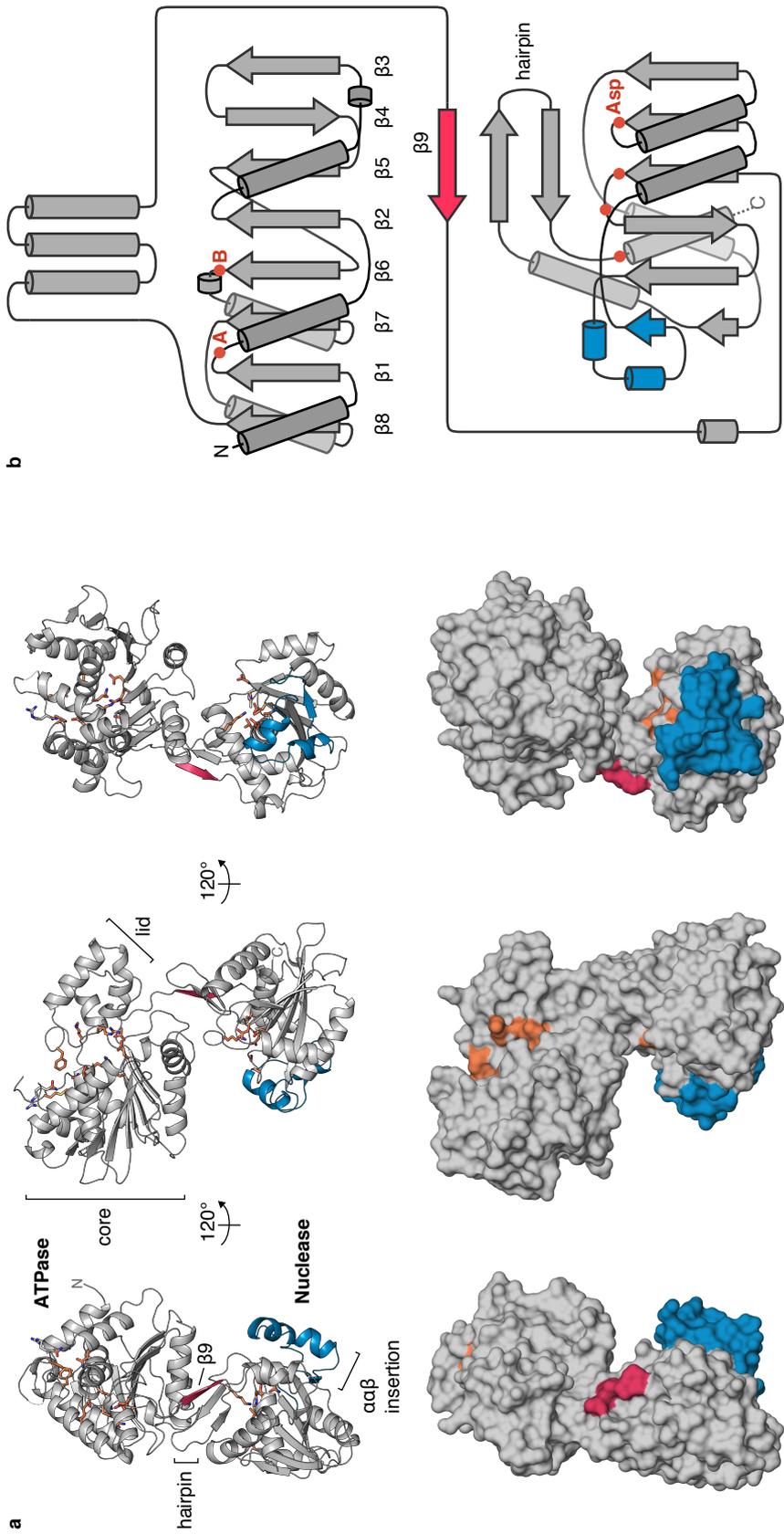


Figure 5.5. Crystal structure of the HK97 large terminase. (a) Ribbon diagrams and solvent-excluded surface representations. Orange sticks depict the putative ATPase and nuclease active sites. Secondary structure elements unique to the HK97 protein are coloured in red and blue. (b) Topology diagram of the HK97 protein. The locations of Walker A and B residues, and aspartate residues in the nuclease active site are marked.

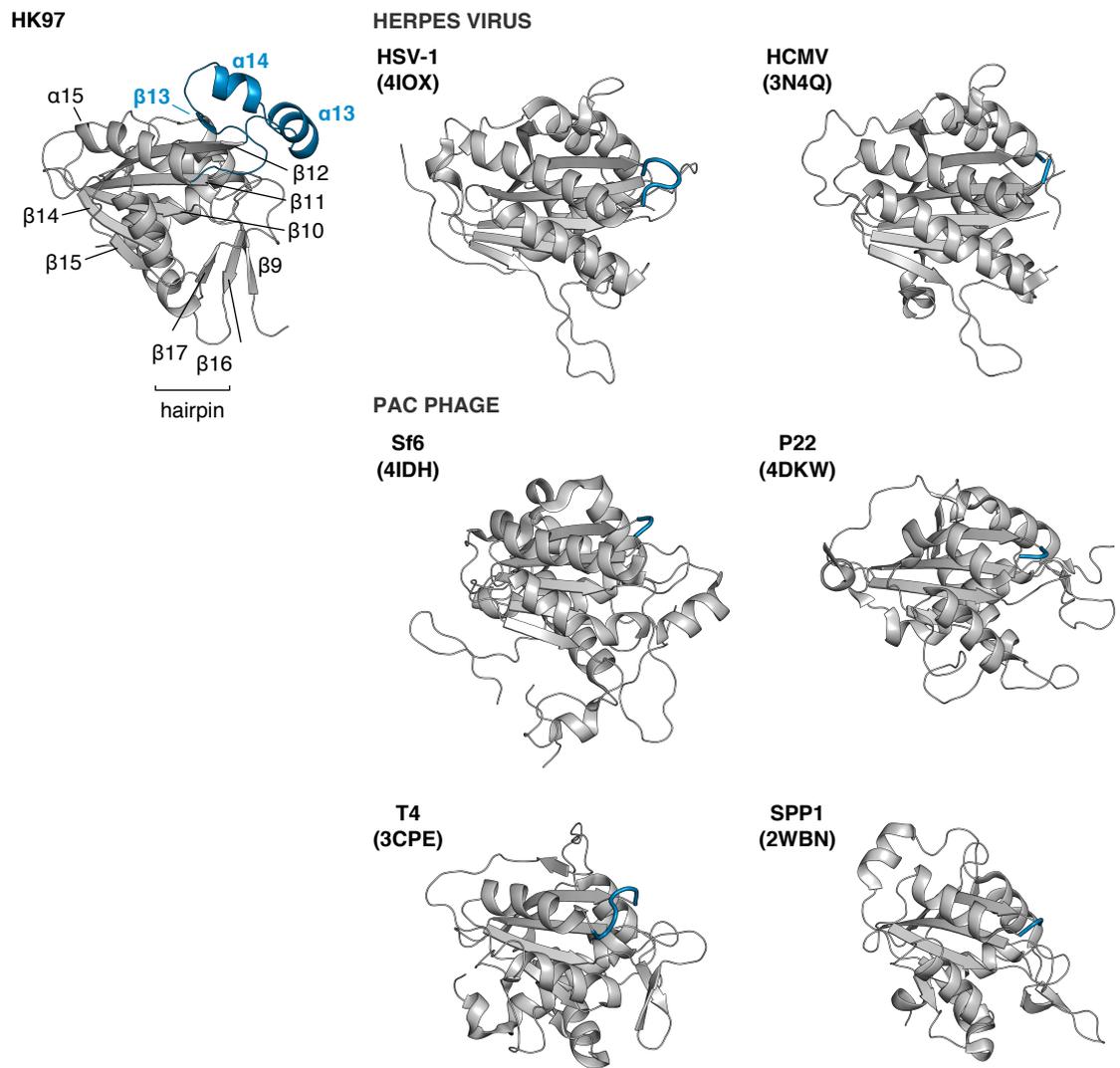


Figure 5.6. Large terminase nuclease domains. The HK97 $\alpha\alpha\beta$ insert is coloured in blue. The corresponding regions in herpes and *pac* homologues are indicated. PDB codes are listed in brackets.

The Walker A motif of the ATPase active site is located between strands $\beta 1$ and $\beta 2$. A lysine residue, K60, occupies the classical position (Figure 5.7b). Another lysine residue, K57, is present upstream. This residue was later found to be important for both ATPase and DNA packaging activities (Figure 5.10). The critical arginine residue conserved across most large terminase proteins, R56, is also present. Catalytic residues D153 and E154 are found in the Walker B motif at the end of strand $\beta 6$. Further downstream is residue T186, which superposes with T287 of T4 large terminase gp17, thought to act as a sensor which mediates conformational changes within the protein during ATP turnover. A putative adenosine-binding pocket is formed between residue Q33 and hydrophobic residues F239 and M27. Movement between the lid and the ATPase core as a result of ATP turnover is thought to drive DNA translocation (108,126,136).

In the nuclease active site, residues D302, D388 and D471 form a conserved catalytic centre next to the conserved nuclease hairpin (Figure 5.7c). Residue K463, which follows the hairpin, reaches into the active site, coordinating a water molecule. A similarly placed lysine in phage Sf6 is proposed to act as a switch for metal binding (108), in tandem with the hairpin, which alters DNA accessibility (247).

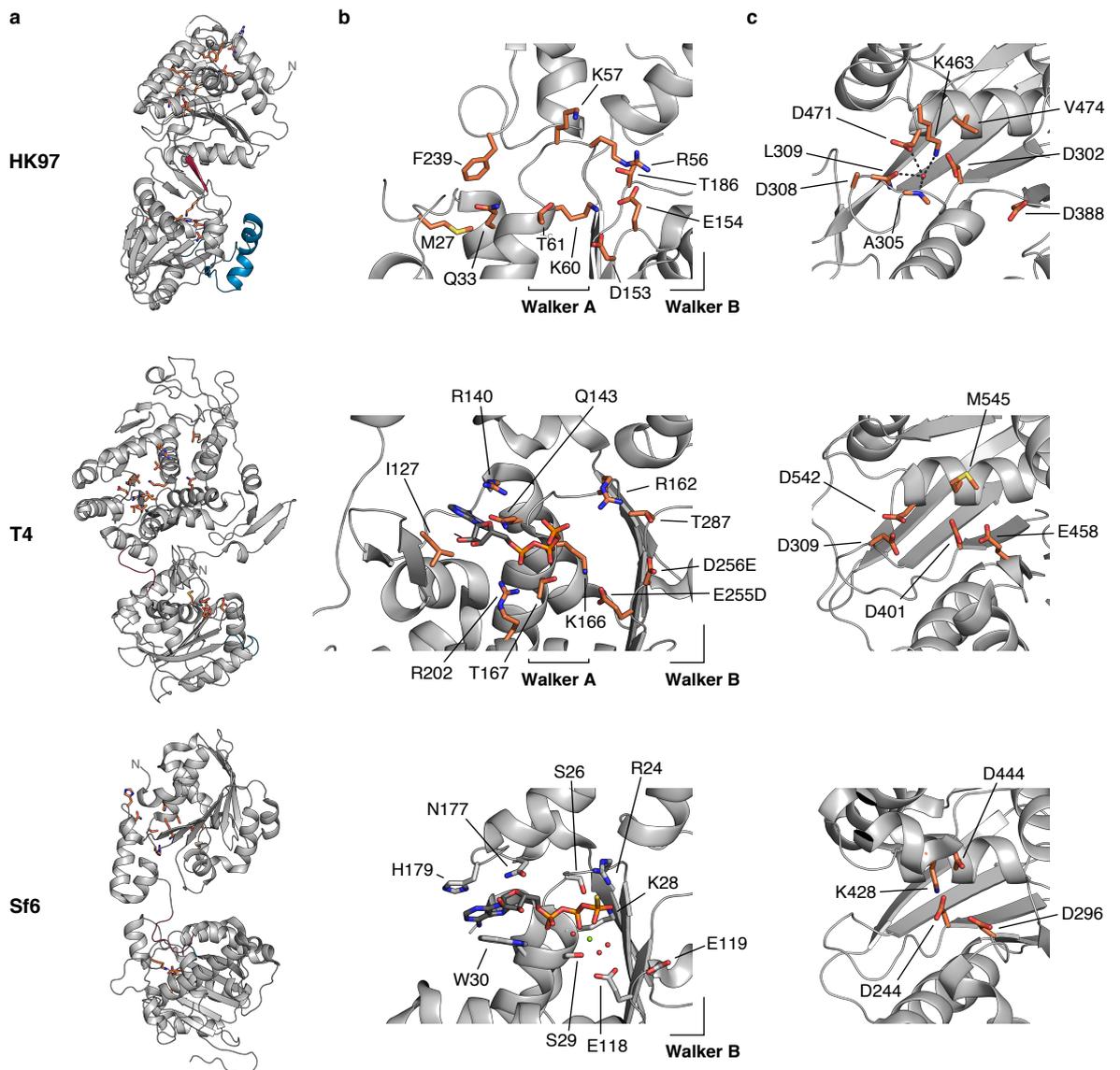


Figure 5.7. Large terminase active sites. (a) Full-length *apo* structures of the HK97, T4 and *Shigella* phage Sf6 large terminase. Elements unique to HK97 and their corresponding regions in T4 and Sf6 (PDB codes 3CPE, 4IDH) are coloured in red and blue. Active site residues are displayed as orange sticks. (b) ATPase active sites. Structures shown for T4 and Sf6 are that of a Walker B double mutant co-crystallised with ATP (PDB code 2O0H) and that of wild-type protein co-crystallised with ATP γ S (PDB 4IEE), respectively. (c) Nuclease active sites. Aspartate residues are present to coordinate magnesium or manganese ions for DNA cleavage. K428 in Sf6 is thought to be a switch for metal ion binding.

5.4 Biochemical and biophysical characterisation of the HK97 large terminase

In solution, the HK97 protein exhibits an interesting behaviour. The protein as an isolated species is predominantly a monomer, measuring 52.7 kDa and 52.5–56.7 kDa in mass by SEC-MALLS and sedimentation velocity analytical ultracentrifugation (AUC), respectively (Figure 5.8a,b). The protein also crystallised as a monomer. However, cryo-EM and TIRFM data indicated that the protein would oligomerise during motor assembly.

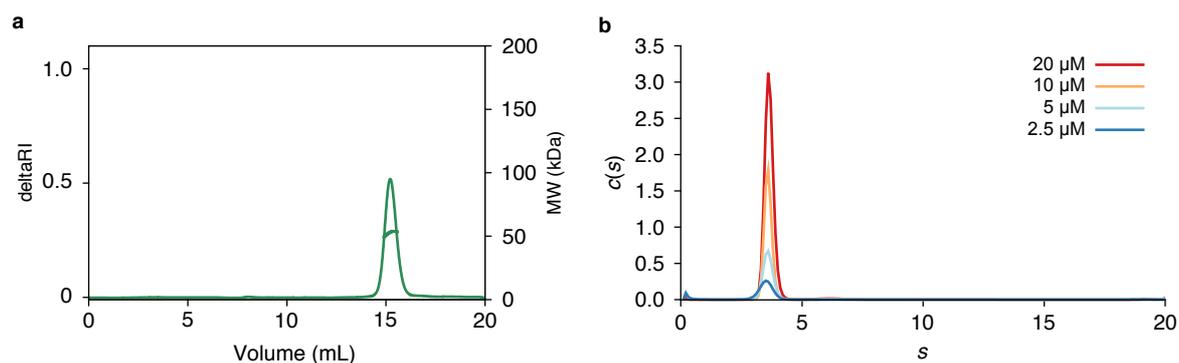


Figure 5.8. Molecular weight measurements of HK97 large terminase. (a) SEC-MALLS analysis using a Superdex 200 10/300 column in 20 mM HEPES, 300 mM KCl, 1 mM DTT, pH 7.0. (b) Sedimentation coefficient distribution of protein (2.5–20 μM) in 20 mM HEPES, 300 mM KCl, 1 mM DTT, pH 7.0, during sedimentation velocity AUC. Frictional ratio estimates of 1.32–1.37 and standard sedimentation coefficients of 3.66–3.81 S were obtained, contributing to over 90% of the signal.

At physiological ATP concentration (248), large terminase in isolation was a poor ATPase, based on enzyme-coupled assays where the rate of inorganic phosphate release turnover was measured (Figure 5.9a). There were also signs of substrate inhibition at higher ATP concentrations. Yet, in the presence of DNA, small terminase and prohead, the affinity for ATP was such that greater than half-maximal activity was achieved at physiological ATP concentration. Under these conditions, ATP turnover may be weakly cooperative. A fit of the Hill equation estimated K_m and n to be 0.32 mM and 1.38, respectively. However, it should be noted that not all large terminase molecules would be engaged in a motor. Free molecules would contribute to the apparent activity observed at high ATP concentration. Thus, the true value of K_m would deviate from the current value. It is also unclear whether there is truly cooperativity in the system. Still, it was clear that large terminase was more active under conditions that favoured DNA packaging. This would have direct implications on the efficiency of ATP usage during virus assembly. To understand the molecular basis of ATPase activation, a series of mutagenesis experiments were subsequently performed.

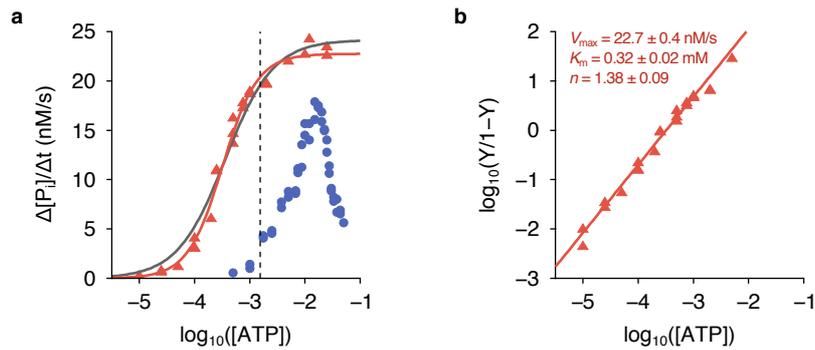


Figure 5.9. Regulation of ATPase activity. (a) Initial rate of inorganic phosphate release for free large terminase (blue) and large terminase in the presence of DNA, small terminase and prohead (red), as a function of ATP concentration. Black line indicates a fit of the Michaelis-Menten equation by non-linear regression. Red line indicates a fit of the Hill equation. Dotted line indicates the average *E. coli* ATP concentration, 1.5 ± 1.2 mM, mean \pm sd (248). (b) Hill plot for large terminase in the presence of DNA, small terminase and prohead, where Y is the rate of inorganic phosphate release relative to V_{\max} according to the Hill fit (red line). The mean and the standard error of fit are indicated for each parameter.

5.5 Molecular basis of HK97 large terminase ATPase activity

Alanine mutations were introduced to identify residues in the ATPase domain that are critical for ATPase and packaging activities. The rate of inorganic phosphate release as a result of ATP turnover was measured for both free protein and the packaging motor. The extent of DNA packaging was also monitored by DNase protection.

Mutation of Walker A residue K60 and Walker B residue E154 to alanine abrogated both DNA packaging and ATPase activities (Figure 5.10a). Mutation of T186 had little effect on activity, suggesting that the residue does not act as a sensor I as proposed for T4 gp17. Mutation of R56 affected function significantly. R56 corresponds to R162 of T4 gp17, said to be a *cis*-acting arginine finger (126). As a *cis* finger, the arginine is repositioned on protein oligomerisation to coordinate the γ -phosphate of ATP for hydrolysis (125). The same residue in P74-26 TerL, R39, is said to be a sensor II-like residue, which coordinates conformational changes during ATP turnover for mechanical movement .

In P74-26 and ϕ 29, a different arginine residue is implicated in catalysis (124,136) — a *trans*-acting arginine finger found on the side of the β -sheet distal to the active site (Figure 5.10g). Mutation of K140, which aligns with the P74-26 finger, did not impair DNA packaging or ATPase activity (Figure 5.10h). Mutation of K204 and K206, which align

with the AAA+ *Thermus thermophilus* ClpB fingers (249), also did not impair activity (Figure 5.10i).

To search for a *trans*-acting finger, alanine mutants were generated for every arginine residue in the ATPase domain and every lysine residue on the surface of the ATPase core expected to have the finger. In case of redundancy, double mutants were also generated. Of all mutants, only R240A, R254A and R270A showed reduced activities (Figure 5.10b). Circular dichroism spectropolarimetry suggests that secondary structure is retained in these mutants (Figure 5.11). Their activities could not be rescued with a Walker B E154A mutant (Figure 5.10d,e), which would have donated an intact finger, suggesting that the residues do not act in *trans*. Given the local environment and effects of their mutation on activity, it is more likely that R240 and R270 mediate movement between the ATPase core and lid through salt bridges and hydrogen bond interactions during ATP turnover (Figure 5.10j,k). Disruption of movement may in turn slow down turnover and translocation. Though not essential for hydrolysis in the free protein, R254 might have a role in inter-subunit coordination or contacting DNA, hence the total lack of DNA packaging.

Contrary to P74-26 and ϕ 29, the HK97 protein contains a second lysine residue upstream of the Walker A lysine. This residue, K57, is essential for packaging and ATPase activities (Figure 5.10c). In the absence of a *trans*-acting finger, K57 could provide the additional charge required for catalysis. Without a *trans*-acting finger, which would have directly contributed to hydrolysis, activation of the ATPase is likely to occur through allostery. A finer dissection of the hydrolysis cycle will be required to understand how activity is coordinated between subunits.

Other systems exist where the Walker A motif donates three positive charges to the ATPase active site, including phage T5, herpes simplex and cytomegalovirus (Figure 5.12). In these systems, an arginine residue is present in a similar position to K57. Data for the HK97 system would predict that this residue could substitute for a *trans*-acting finger.

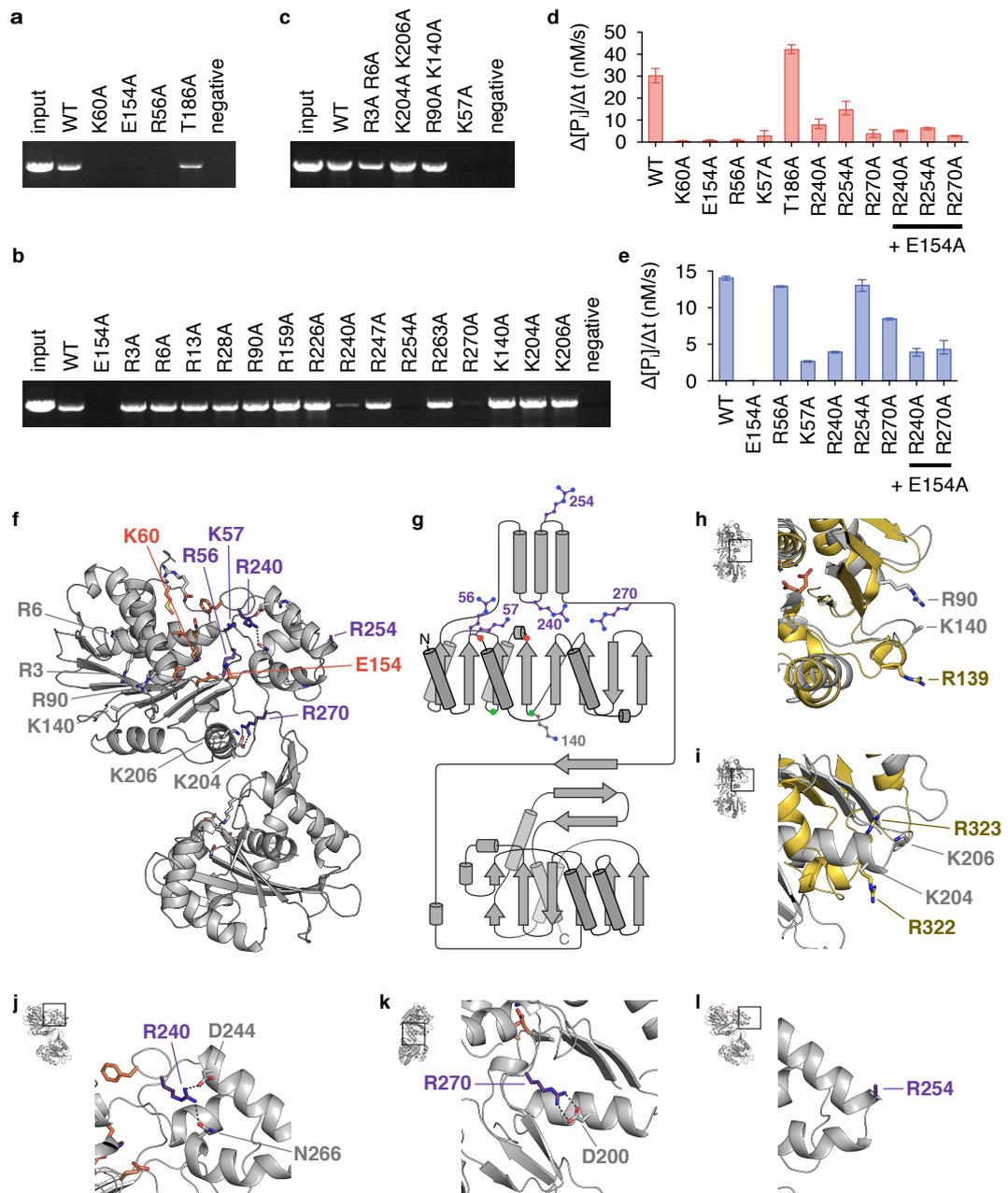


Figure 5.10. Critical residues of the HK97 large terminase ATPase domain. (a, b, c) DNA packaging by alanine mutants. (d) ATPase activity in the presence of DNA, small terminase and prohead at 1 mM ATP concentration. (e) ATPase activity of free mutant proteins at 10 mM ATP concentration. For arginine finger complementation, equimolar amounts of mutant E154A were added. (f, g) Ribbon representation and topology diagram of the crystal structure. ATPase active site residues are coloured in orange. Additional arginine or lysine residues important for function are coloured in purple. Red dots indicate the locations of the Walker A and B sequences. Green dots indicate potential arginine finger positions based on $\phi 29$ and P74-26. Superposition of the HK97 structure with the arginine fingers of (h) P74-26 large terminase (gold, PDB code 4ZNK) and (i) *Thermus thermophilus* ClpB NBD1 (gold, PDB code 1QVR). (j, k, l) The local chemical environment of residues R240, R270 and R254. The side chain of R254 is disordered in the crystal structure.

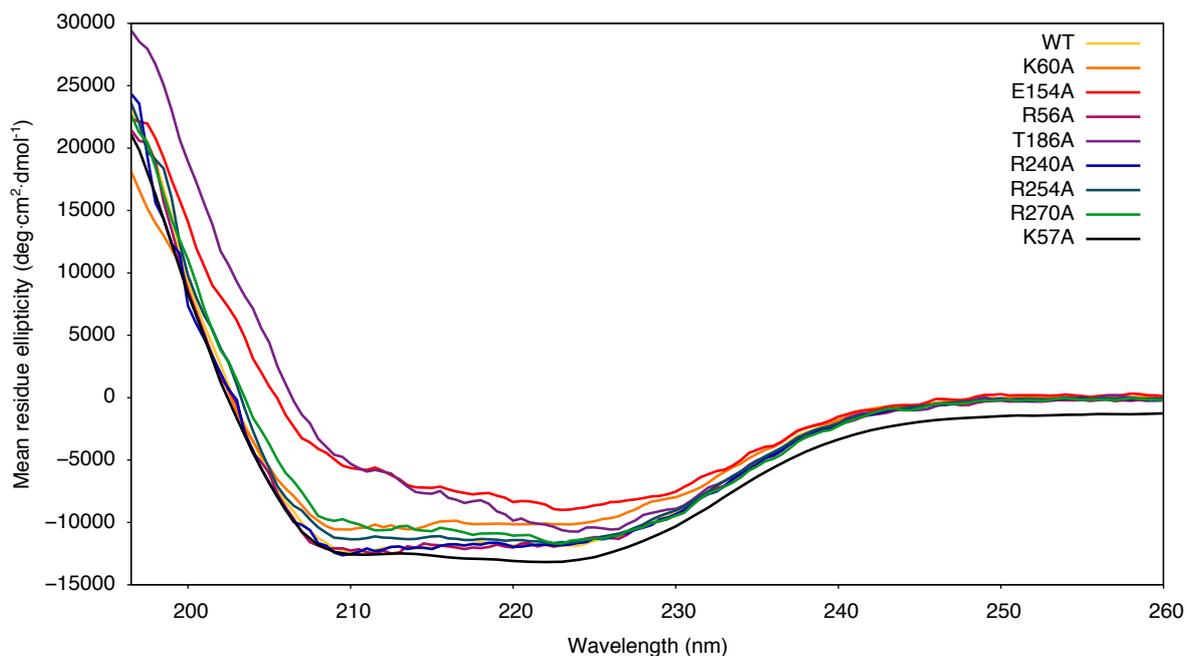


Figure 5.11. Far-UV circular dichroism spectra of wild-type HK97 large terminase and mutants. Spectra were recorded for protein (2–5 μ M) in 20 mM Tris sulfate, 50 mM Na_2SO_4 , 10 mM MgSO_4 , 0.5 mM Tris(2-carboxyethyl)phosphine hydrochloride, pH 7.5, at 20 °C.

	Walker A																				
HK97_GP2	50	A	I	L	S	I	A	R	K	N	G	K	T	G	L	I	A	G	I	L	68
HSV1_UL15		T	V	F	L	V	P	R	R	H	G	K	T	W	F	L	V	P	L	I	
HCMV_UL89		T	V	F	L	V	P	R	R	H	G	K	T	W	F	I	I	P	I		
HSV2_UL15		T	V	F	L	V	P	R	R	H	G	K	T	W	F	L	V	P	L	I	
HHV8_AAB08384		S	V	F	L	I	P	R	R	H	G	K	T	W	I	V	V	A	I		
T4_GP17		T	V	C	N	L	S	R	Q	L	G	K	T	T	V	V	A	I	F	L	
T7_GP19		F	I	L	Q	A	F	R	G	I	G	K	S	F	I	T	C	A	F	V	
T5_AAU05290		V	T	A	C	V	S	R	V	G	K	S	F	I	A	Y	T	L	G		
SPP1_GP2		Y	V	L	K	G	G	R	G	S	A	K	S	T	H	I	A	M	W	I	
Sf6_GP2		K	V	A	K	G	G	R	G	S	G	K	S	W	A	I	A	R	L	L	
Phi29_GP16		N	F	V	I	G	A	R	G	I	G	K	S	Y	A	M	K	V	Y	P	
P22_GP2		R	C	F	M	A	G	N	Q	L	G	K	S	F	T	G	A	A	E	V	
Lambda_GPA		V	N	V	V	K	S	A	R	V	G	Y	S	K	M	L	L	G	V	Y	

Figure 5.12. Multiple sequence alignment of large terminase Walker A motifs. Virus and gene product names are indicated. Highlighted in grey is the classical lysine residue; red, the critical arginine residue of large terminase proteins; yellow, the additional positive residue in HK97, T5 and herpes viruses. Lambda gpA has an unusual Walker A sequence with two lysine residues upstream and downstream of the classical position, both required for ATPase and DNA packaging activities (250,251).

5.6 Model of the HK97 DNA packaging motor

Based on the cryo-EM reconstruction of the HK97 motor, a pseudo-atomic model can be built by combining the flexible-fitted prohead structure, the structure of a portal protein, and a pentamer model of the HK97 large terminase (Figure 5.13a). The portal protein from *Corynebacterium diphtheriae* has 17% sequence identity to the HK97 portal. The large terminase pentamer was created by superposition of the HK97 structure to the phage P74-26 pentamer model. The P74-26 model was generated by molecular docking (136) and contains only the ATPase domain. By superposing a structure comprising both nuclease and ATPase domains, the model is now extended. Given a similar domain orientation in the motor as in the crystal structure, the ATPase lid of one subunit would insert between the ATPase core and nuclease domain of an adjacent subunit. The lid may thus act as an anchor while the core part of the ATPase moves to translocate DNA (Figure 5.13b).

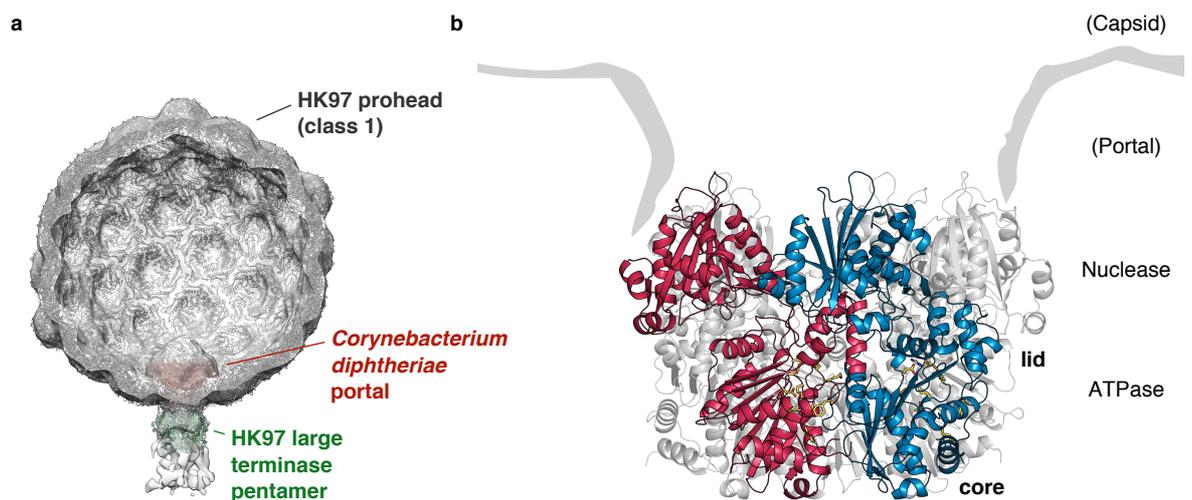


Figure 5.13. Model of the HK97 DNA packaging motor. (a) Electron density from asymmetric reconstruction of the HK97 motor, fitted with the prohead structure, portal protein (PDB code 3KDR) and the HK97 large terminase pentamer. (b) The HK97 large terminase pentamer, based on the P74-26 model. Two adjacent subunits are coloured in red and blue, and the remainder are coloured in grey.

5.7 Discussion

The crystal structure of the HK97 large terminase shows that *cos* large terminase proteins have the same domain arrangement as their *pac* counterparts (Figure 5.6). Both ATPase and nuclease active sites contain a conserved set of catalytic residues (Figure 5.8). However, additional secondary structure elements in the linker region and nuclease domain suggest that the regulation of ATPase and nuclease activities and the mechanism of DNA cleavage, respectively, may differ for HK97 and related systems (Figures 5.4–5.7).

Activation of the HK97 ATPase domain appears to be dependent on the formation of a packaging motor on DNA (Figure 5.9). However, a *trans*-acting arginine finger could not be found (Figure 5.10). Instead, an additional lysine residue was identified in the active site which is essential for ATPase activity. This residue is not present in P74-26 or ϕ 29, for which *trans*-acting arginine fingers have been identified (136). The lysine residue possibly substitutes for the arginine finger in catalysis. Given that motor formation is required for ATPase activation, in the absence of an arginine finger, the HK97 ATPase is likely activated by conformational changes induced by adjacent subunits.

To understand the coordination between motor subunits and to investigate whether or not there is cooperativity, a series of kinetics experiments is proposed here. These experiments will make use of fluorescent 2'/3'-O-N-methylanthraniloyl (mant) nucleotides. The assumption is made that all large terminase molecules can be driven to form motors, in the absence of nucleotides or in the presence of ADP or ATP γ S. Using a stopped-flow apparatus, ADP off-rates can be measured by competing off pre-bound mant-ADP using non-fluorescent ADP and the change in mant fluorescence monitored (252). ATP hydrolysis rates can be measured by adding mant-ATP to nucleotide-free motors and stopping the reaction in a quench-flow system; the amount of mant-ADP produced can be quantified by HPLC (253). ATP binding rates can be measured by changes in mant-ATP γ S fluorescence. Alternatively, the affinity for ATP γ S can be determined by fluorescence polarisation, off-rate by competition with unlabelled nucleotide, and the on-rate calculated. These kinetics data would reveal the sequence of events during ATP turnover.

Optical tweezers experiments on the ϕ 29 motor suggest that the motor cycles between a burst phase and a dwell phase (132). During each burst, only four of five subunits would hydrolyse ATP, resulting in four translocation steps. Data suggest that the fifth subunit

would also bind and hydrolyse ATP. The number of ATP molecules bound in the HK97 motor can be measured in an equilibrium binding experiment using mant-ATP γ S and nucleotide-free motors (252). In addition to providing mechanistic insights into motor function, the stoichiometry and kinetics of nucleotide binding will inform future structural efforts to visualise the motor at different points in the ATP hydrolysis cycle.

A full investigation of the nuclease domain was beyond the scope of this work. However, mutations and deletions can be introduced to identify residues that are involved in DNA cleavage, whether directly through the two-metal ion mechanism or indirectly through interactions with GP74. The hypotheses that residue K463 acts as a switch for metal ion binding and that the unique β -strand regulates nuclease activity can also be tested. In particular, residues in the β -strand may be substituted with residues with low β -sheet-forming propensities such as proline (254).

It will be of interest to identify the residues that are responsible for contacting DNA in the HK97 large terminase. However, this may not be a trivial task. Observations by TIRFM, EM and preliminary electromobility shift assays indicate that free large terminase has little affinity for DNA. A packaging motor or complex for initiation or termination will likely need to be assembled for any significant DNA binding to be observed. Mass spectrometry analysis of crosslinked protein-DNA complexes and cryo-EM reconstructions may provide a faster route to understanding how the HK97 motor contacts DNA.

6. Conclusions and future perspectives

Tailed bacteriophages encode a DNA packaging motor to translocate their genome into empty capsids for self-replication. While the proteins responsible have been identified, structural details are lacking for the molecular assemblies that form during initiation, translocation and termination. This is especially the case for *cos* phages, because no structure is available yet for the complete motor or any of the terminase components that drive packaging. The goal of this work was thus to elucidate the mechanisms of DNA packaging in *cos* phages by establishing a new *in vitro cos* system based on *E. coli* phage HK97 and determining its structural, biophysical and biochemical properties.

Given a free DNA end, a minimal motor could be assembled using purified proheads and large terminase protein in the presence of magnesium and ATP. Small terminase was not essential but improved the efficiency of packaging. Under ATP-limited conditions, the motor would stall on encountering an internal *cos* site in the forward direction. The stalling was dependent on small terminase; small terminase bound to the *cos* site likely acted as a roadblock. This *in vitro* event perhaps captures the transition from DNA translocation to termination *in vivo*. Assays with the endonuclease cofactor GP74 showed that the protein led to greater processing efficiency at the *cos* site but it also gave rise to a low background of non-*cos*-specific cuts. A fine balance of packaging and processing would be critical to successful genome packaging. It is clear that small terminase is an important element during packaging initiation and termination. Small terminase-induced stalling perhaps contributes to the specificity of a *cos* system during termination. It is also clear that GP74 contributes to *cos* site processing during initiation or termination or both. Future work focused on isolating the molecular assemblies at play and mapping the protein-protein interactions within will shed light on the role of the two proteins during these processes.

For a structural understanding of packaging in a *cos* phage, the crystal structures of both small terminase and large terminase were determined. The HK97 small terminase features a circular array of nine N-terminal helix-turn-helix motifs, similar to the small terminase proteins of *pac* phages. NMR assignments and chemical shift perturbation studies with the helix-turn-helix motif-containing N-terminal domain of *pac* phage SF6 showed that the motif was important for DNA binding. For initiation, *cos* DNA likely wraps around the HK97 small terminase in a manner similar to how *pac* DNA wraps around the SF6 small terminase. Wrapping at downstream *cos* sites may also create a roadblock for the

packaging motor, contributing to termination. Because there are no long sequence repeats or IHF binding sites in the HK97 *cos* region as there are in the λ *cosB* region, it is unclear how register is determined. A structure of the protein-DNA complex would provide some insights. Given that non-specific DNA binding can occur at high protein concentrations, identifying the minimal *cos* region for small terminase binding would facilitate future structural efforts. The ATPase and nuclease active sites of the HK97 large terminase contain the same catalytic residues as large terminase proteins from *pac* phages. A unique β -strand between the ATPase and nuclease domains offers a potential mechanism for the regulation of ATPase and endonuclease activities during the different stages of packaging. Unique inserts in the nuclease domain adjacent to the active site reveal a potential surface for GP74 interaction or structure necessary for specific *cos* DNA cleavage. Assays probing the protein-protein interactions between large terminase and GP74, and protein-DNA interactions between large terminase and *cos* DNA would reveal the role of these additional structural elements. Isolation of the termination complex, which is feasible given the stalling behaviour *in vitro* induced by small terminase, would enable further understanding of the process.

Cryo-EM reconstruction of the HK97 packaging motor revealed the *cos* packaging motor had a similar architecture to *pac* and ϕ 29 packaging motors. Single-molecule total internal reflection fluorescence microscopy-based counting of GFP-labelled large terminase suggested that motor contained five large terminase subunits. The similar stoichiometry to *pac* and ϕ 29 motors indicated that the ATPase subunits are likely assembled in a similar manner. However, biochemical data suggest that ATPase activity is regulated differently in the HK97 large terminase pentamer. No *trans*-acting arginine fingers were found. Instead, an additional lysine residue present in the P loop of the active site was found to be essential. This residue could donate a positive charge in place of the finger for catalysis. As activity is stimulated by motor formation, ATPase activity is likely regulated allosterically. A similar positive residue is present in herpes viruses and phage T5 (Figure 5.12) and thus will be of interest in the study of these viruses.

A more thorough understanding of the packaging motor can be gained from measuring its kinetic and thermodynamic properties. The use of fluorescent ATP analogues combined with stopped-flow and quenched-flow techniques would provide parameters with which to construct a model for the coordination of events between subunits during ATP turnover.

The existing set-up for subunit counting by single-fluorophore photobleaching can be (255,256). This would allow the stoichiometry and turnover of large terminase in individual, actively translocating motors to be followed in real-time. Step sizes, rotation and the duration of different motor states may be measured as a function of mechanical force and torsion in these single-molecule assays, which would enable direct comparison with the ϕ 29 system (133). With the new biochemical and biophysical information, motor complexes can be assembled with greater homogeneity and in greater amounts. Atomic details of the packaging motor will become available then by cryo-EM. Given enough particles and appropriate classification routines (257,258), and perhaps using time-resolved spraying technology (259,260), it would be possible to visualise the active motor in its different states. From these representative states, manifold embedding techniques recently applied to ribosomes (261) can be used to construct an energy landscape, and from this the complete trajectory of the motor during DNA translocation can be inferred.

The molecular components for DNA packaging, their biochemistry and structures are documented here for phage HK97. The stage is set for higher-resolution, more detailed studies. With the characterisation of this system, and others such that comparisons are possible, we may ultimately arrive at a mechanistic understanding of DNA packaging in viruses, a basic biological process which has contributed to the evolution of life as we know it today.

Appendix A. DNA and protein sequences

DNA oligonucleotides used in this study are listed in the following table.

Table A.1. DNA oligonucleotides. Primers for plasmid amplification for In-Fusion cloning are indicated. Numbering on *cos* primers indicates the genome position of the first complementary nucleotide.

<i>In-Fusion for His-SUMO construct</i>	
SUMO_GP1_for	GAACAGATTGGTGGTGCAGATAAACGAATCCGTTCCG
SUMO_GP1_rev	TACCTAAGCTTGTCTTTATCCGTGCTTGGGAAAGGC
SUMO_GP2_for	GAACAGATTGGTGGTATGACGCGAGGTGAGCG
SUMO_GP2_rev	TACCTAAGCTTGTCTTCACATGCTCAGCGGTCTG
Plasmid_for	AGACAAGCTTAGGTATTTATTCGG
Plasmid_rev	ACCACCAATCTGTTCTCTGTGA
<i>In-Fusion for His-fusion construct</i>	
His_GP2_for	ATCACCACCACCACATGACGCGAGGTGAGCG
His_GP2_rev	TGAGGAGAAGGCGCGTCACATGCTCAGCGGTCTG
Plasmid_for	CGCGCCTTCTCTCACTGTTCCAGGGGCCCAT
Plasmid_rev	TGTGGTGGTGGTGTATGATGGCTGCTGCCC
<i>Site-directed mutagenesis</i>	
GP2_R3A_for	CGGCAGGTGAGCGTGTAATAGCGTTCATT
GP2_R3A_rev	CTCACCTGCCGTCATGTGGTGGTGG
GP2_R6A_for	GTGAGGCCGTAATAGCGTTCATTGAGCG
GP2_R6A_rev	TATTACGGCCTCACCTCGCGTCATGTGG
GP2_R13A_for	ATTGAGGCCTTTTGCATCGTGCCAGAAG
GP2_R13A_rev	CAAAAGGCCTCAATGAACGCTATTACACGC
GP2_R28A_for	CTATGGCGTTGGACCCCTTTCAGAAAGATTT
GP2_R28A_rev	TCCAACGCCATAGGTTGGCCGATAAGC
GP2_R56A_for	CGCCGCGAAAAATGGTAAGACTGGCC
GP2_R56A_rev	TTTTTCGCGGCGATGGAGAGGATCGC
GP2_R90A_for	GCGCGAACAGGCGGCCATCG
GP2_R90A_rev	TGTTCCGCGCTGAGTGCACCGCT
GP2_R159A_for	GGTTGCGGGCCCGCAGGATGATTTTATC
GP2_R159A_rev	GCCCGCAACCTGCCCTGTTTCATCG
GP2_R226A_for	AGTAAAGCCGAGTCTGGCTGGCTGC
GP2_R226A_rev	GACTCGGCTTTACTGATATCAGCGTCTTTTGG
GP2_R240A_for	CATTCGCGTCAGAAAAAGACATGGCGC
GP2_R240A_rev	CTGACGCGAATGTTCCCAGTGCCG
GP2_R247A_for	GCGGCCAGGCTGAGAAAGCTGG
GP2_R247A_rev	CTGGGCCCATGTCTTTTTCTGACC
GP2_R254A_for	CTGGCGCAATGCCAAGCTTCGAAAACAC
GP2_R254A_rev	CATTGCGCCAGCTTTCTCAGCCTGGC

GP2_R263A_for	CTTCGCGAACCTCAACCTCAATCAGCG
GP2_R263A_rev	AGGTTTCGCGAAGGTGTTTTCGAAGCTTG
GP2_R270A_for	TCAGGCCGTGTCTACCGTATCGCCG
GP2_R270A_rev	ACACGGCCTGATTGAGGTTGAGGTTTCG
GP2_K140A_for	AAGGTGCGACGACGCACGGCC
GP2_K140A_rev	CGTCGCACCTTCTGCGGATAAAGCC
GP2_K204A_for	GTCGCGTCGAAAGATCCGCACATCG
GP2_K204A_rev	TTCGACGCGACCGCATCATCAATCCAG
GP2_K206A_for	AATCGGCCGATCCGCACATCGTGTG
GP2_K206A_rev	GATCGGCCGATTTGACCGCATCATCAA
GP2_T186A_for	CAGTGCAGGCAGCAGCAAACGATGCTGA
GP2_T186A_rev	CCTGCGCACTGATAACGATTAGCAGCGG
GP2_K60A_for	GGTGC GACTGGCCTGATTGCCGGAAT
GP2_K60A_rev	CCAGTCGCACCATTTTTTCGGGCGATG
GP2_E154A_for	TCGATGCGACAGGGCAGGTTAGGGG
GP2_E154A_rev	CTGTGCGATCGAGAATGGCCAGAATGG
GP2_K57A_for	CCCGAGCGAATGGTAAGACTGGCC
GP2_K57A_rev	ATTCGCTCGGGCGATGGAGAGGATC
<i>In-Fusion for GFP/linker insertion</i>	
GFP_for	ATCACCACCACCACATGGTGAGCAAGGGC
GFP_rev	CCAGATCCGCCGCTGCCACCCTTGTACAGCTCGTCCAT
Plasmid_for	CAGCGGCGGATCTGGCGGTTCCGGTATGACGCGAGGTGAG
Plasmid_rev	TGTGGTGGTGGTGGTGGTGGCTGCTGCCC
<i>Overlap-extension and restriction cloning of HK97 cos region</i>	
cos-312BamHIfor	GTGGATCCGACGGTAAGTTTGTTCAGC
cos1strev	CAAACCTTTGGCGGCGGTCATTTG
cos2ndfor	GACCGCCGCCAAAGTTTGAATTTAAC
cos+472EcoRIrev	GTGAATTCCTGTCATCACGATCG
<i>Generation of 5' biotinylated DNA</i>	
cos-80for	ATCAAATGAGAATGAATCGCATC
cos+150rev	AAACCTGCATGGGACG-BIO
<i>HK97 cos DNA truncation series</i>	
cos-312for	GACGGTAAGTTTGTTCAGCTGG
cos-20for	CCCAAATGACCGCCG
cos+81for	CGGCGGCAGCGG
cos+120for	ACACCATCGATCCGCC
cos-1rev	TTTGGCGGCGGTCATT
cos+49rev	ATGTCAACACCTTAAACTAAAAA
cos+120rev	TCCACAGCTGCATTTTTTC
cos+150rev	AAACCTGCATGGGACG
cos+472rev	CCGTGCATCACGATCGTTC

Sequences of proteins overproduced are listed below.

> Small terminase (GP1)

MADKRIRSDSSAAAVQAMKNAAVDTIDPPSHAGLEKKAEPFWHDNIRSKALDSWTPADL
LA
AVELANNQLYITVLRKDLRKEERIRGEERDEGLIKDLRKQIVELQRILAQRDLQIHS
ATNGESRDQKKRNQNDRDARNTKNEHQDQDDNLIAPKHG

> Large terminase (GP2)

MTRGERVIAFIERFCIVPEGKLIQPMRLDPFQKDFILAVYDNPAGTDMAILSARKNGK
TGLIAGILLAHLVGPEAVQNTQIVSGALSREQAAIVFNLAVKMVNLPKLQEIVHITPSG
KKLIGLPCNVEYKALSAEGKTTHGLSPILAILDETGQVRGPQDDFIDAITTAQGAHENPL
LIVISTQAANDADLLSIWIDDAVKSKDPHIVCHVYEAPKDADISKRESWLAANPALGTFR
SEKDMARQAEEKAGRMPSFENTFRNLNLNQRVSTVSPFISRSVWELCGEMPINTPRKWYAG
LDLSARNDLTALVIAGEADDGVWDVFPFFWTPQKTLEERTKTDRAPYDVVWVREGLLRTT
P
GASVDYSFVVADIAEIHGDFDLTSMADFDRWRIDQFRKDADAIGLSLPLVEFGQGFKDMGP
AVDTLESLMLNGRVRHGMHPVLTMCVNAVVKDAAGNRKLDKSKATGRIDGMVAMT
MSV
GAANGEVTEQGGDFDDFIFRPLSM

> N-terminal His-tag

MGSSHHHHHH

> His-SUMO

MGSSHHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKI
KKTTPRLRLMEAFKRQGKEMDSLRFYDGIQADQTPEDLDMEDNDIIEAHREQIGG

> His-GFP-G/S

MGSSHHHHHHMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTCLKFICT
TGKLPVPWPTLVTTLTLYGVQCFAFYDPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKT
R
AEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHKVYITADKQKNGIKVNFKTRH
NIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGI
TLGMDELYKGGSGGSGSG

> His-TEV-GP74

MGSSHHHHHHSSGRENLYFQGHMNKEPRVYGSRWDKARLRFLQQHPLCVMCEQQGRIT
PA
TVVDHIVPHKCLKDALKSGNPLAISKAQLLFWSKENWQPLCKAHDSTKQRMEKSGAVIG
C
DANATRSILRLTGARNERPHH

> SUMO protease (Ulp1)

MRGSHHHHHHGLVPRGSLVPELNEKDDDQVQKALASRENTQLMNRDNIEITVRDFKTLA
P
RRWLNDTIIIEFFMKYIEKSTPNTVAFNSFFYTNLSERGYQGVRWRMCRKKTQIDKLDKIF
TPINLNQSHWALGIIDLKKTIGYVDSLSNGPNAMSFALDQKYVMEEKHTIGEDFD
LIHLDCPQQPNGYDCGIYVCMNTLYGSADAPLDFDYKDAIRMRRFIAHLILTDALK

Appendix B. Theory

B.1 X-ray crystallography

B.1.1 Diffraction theory

In X-ray crystallography, molecular structures are determined based on the scattering of X-rays by electrons in a crystal. Most of the scattering is due to elastic Thomson scattering, where electrons are only accelerated by the incident photon, not excited. A small amount of scattering is due to Compton scattering, which is inelastic and is responsible for the diffuse background observed during a diffraction experiment.

Take a crystal. The unit cells of this crystal are related by discrete translations along primitive vectors \mathbf{a} , \mathbf{b} and \mathbf{c} of the Bravais lattice. Families of planes passing through the unit cell are described by Miller indices h , k , and l . The reciprocal lattice in reciprocal space is defined by the axes or primitive vectors:

$$\mathbf{a}^* = \frac{\mathbf{b} \times \mathbf{c}}{\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})}, \quad \mathbf{b}^* = \frac{\mathbf{c} \times \mathbf{a}}{\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})}, \quad \mathbf{c}^* = \frac{\mathbf{a} \times \mathbf{b}}{\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})}$$

The reciprocal lattice vector or diffraction vector, $\mathbf{S}(hkl) = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$, is perpendicular to the family of planes (hkl) . Its length is the inverse of the spacing between planes.

Diffraction occurs for a given family of planes when Bragg's law is satisfied:

$$2d \sin \theta = n\lambda$$

where d is the lattice spacing, θ is the angle between the incident wave and the set of lattice planes, λ is the wavelength of the incident wave, and n is an integer. When Bragg's law is satisfied, waves reflecting off successive planes are in phase and they undergo constructive interference. The amplitude and phase of the resultant wave is described by a structure factor, which is the sum of scattering contributions by all atoms in the unit cell. Each scattering contribution has an amplitude and a phase. The amplitude is determined by the atomic scattering factor, which is dependent on the wavelength and angle of the incident wave and atom type. The phase is determined by the position of the atom relative to the lattice plane. The structure factor can be expressed in terms of Miller indices \mathbf{h} and

fractional coordinates \mathbf{x} , since $\mathbf{S} \cdot \mathbf{r} = \mathbf{h} \cdot \mathbf{x}$ where $\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$ describes the position of a point scatterer relative to the origin in real space. This structure factor has the form

$$\mathbf{F}(\mathbf{h}) = F_{hkl} e^{2\pi i \varphi_{hkl}} = \sum_j f_j e^{2\pi i \mathbf{h} \cdot \mathbf{x}_j} = \sum_j f_j e^{2\pi i (hx_j + ky_j + lz_j)}$$

where j represents an atom. Given scattering is elastic, Friedel's law applies:

$$\mathbf{F}(\mathbf{h}) = \mathbf{F}^*(-\mathbf{h})$$

Integrating over the unit cell, the structure factor is a Fourier transform of electron density inside the unit cell.

$$\mathbf{F}(\mathbf{h}) = \int_{\text{cell}} \rho(\mathbf{x}) e^{2\pi i \mathbf{h} \cdot \mathbf{x}} d\mathbf{x}$$

Therefore, electron density can be calculated using an inverse Fourier transform.

$$\rho(\mathbf{x}) = \frac{1}{V} \sum_{\mathbf{h}} \mathbf{F}(\mathbf{h}) e^{-2\pi i \mathbf{h} \cdot \mathbf{x}} = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} e^{2\pi i \varphi_{hkl}} e^{-2\pi i (hx + ky + lz)}$$

In a diffraction experiment, the intensities of diffracted X-rays are measured. The intensity of each reflection is related to the square of its structure factor amplitude. The phases of structure factors are unknown, but they may be estimated using direct methods, molecular replacement or experimental phasing. Initial phases allow a model to be built. During refinement, the model is altered so that differences between the observed and calculated structure factor amplitudes are minimised. Because of their relevance to the current work, the principles of experimental phasing based on anomalous diffraction are described below. A full description of X-ray crystallography can be found in (262).

B.1.2 Anomalous scattering

As the energy of incident radiation approaches the binding energy of an electron, resonant or anomalous scattering sometimes occur. The electron absorbs a photon, becoming excited, and de-excites, emitting a photon of identical energy. This process causes a perturbation in the overall scattering. The perturbation has real and imaginary components, f' and f'' , respectively. The atomic scattering factor can be described by $f = f^\circ + f' + i f''$,

where f° is normal scattering due to Thomson scattering. After absorption, fluorescence with a Stokes shift can also occur, manifesting as a background in the diffraction pattern.

Values of f' and f'' are dependent on wavelength and, to an almost negligible extent, scattering angle. These values can be calculated using quantum mechanics but also determined experimentally from fluorescence absorption spectra, since fluorescence is proportional to absorption, which is in turn proportional to f'' . Experimental values can deviate from theoretical ones because of the local chemical environment. Thus, they can be critical to the choice of wavelengths in an anomalous diffraction experiment.

For a given reflection, the structure factor contribution of all anomalous atoms is given by

$$\mathbf{F}_A = {}^\circ\mathbf{F}_A + (f'/f^\circ){}^\circ\mathbf{F}_A + i(f''/f^\circ){}^\circ\mathbf{F}_A = {}^\circ\mathbf{F}_A + \mathbf{F}'_A + \mathbf{F}''_A$$

where ${}^\circ\mathbf{F}_A$ represents the normal scattering component with amplitude $|{}^\circ\mathbf{F}_A|$ and phase φ_A .

The total structure factor is given by

$$\mathbf{F} = {}^\circ\mathbf{F}_N + \mathbf{F}_A = {}^\circ\mathbf{F}_N + {}^\circ\mathbf{F}_A + \mathbf{F}'_A + \mathbf{F}''_A = {}^\circ\mathbf{F}_T + \mathbf{F}'_A + \mathbf{F}''_A$$

where ${}^\circ\mathbf{F}_N$ is the normal scattering contribution from all atoms except the anomalous atoms, and ${}^\circ\mathbf{F}_T$ is the sum of ${}^\circ\mathbf{F}_N$ and ${}^\circ\mathbf{F}_A$, that is, the normal scattering contribution from all atoms including the anomalous atoms. ${}^\circ\mathbf{F}_T$ has amplitude $|{}^\circ\mathbf{F}_T|$ and phase φ_T .

Consider a Friedel pair (Figure B.1). Complex conjugates of ${}^\circ\mathbf{F}_N$ and ${}^\circ\mathbf{F}_A$ contribute to each member of the pair. However, \mathbf{F}''_A always has a phase shift of 90° relative to ${}^\circ\mathbf{F}_A$. Friedel's law is thus broken. The intensities of the Friedel pair are given by

$$|\mathbf{F}^\pm|^2 = |{}^\circ\mathbf{F}_T|^2 + a|{}^\circ\mathbf{F}_A|^2 + b|{}^\circ\mathbf{F}_T||{}^\circ\mathbf{F}_A|\cos\alpha \pm c|{}^\circ\mathbf{F}_T||{}^\circ\mathbf{F}_A|\sin\alpha \quad (1)$$

where $a = (|f' + if''|/f^\circ)^2$, $b = 2(f'/f^\circ)$, $c = 2(f''/f^\circ)$ are wavelength-dependent and $\alpha = \varphi_T - \varphi_A$ is the phase difference between the normal scattering contribution from all atoms and the anomalous contribution from the anomalous atoms (263,264).

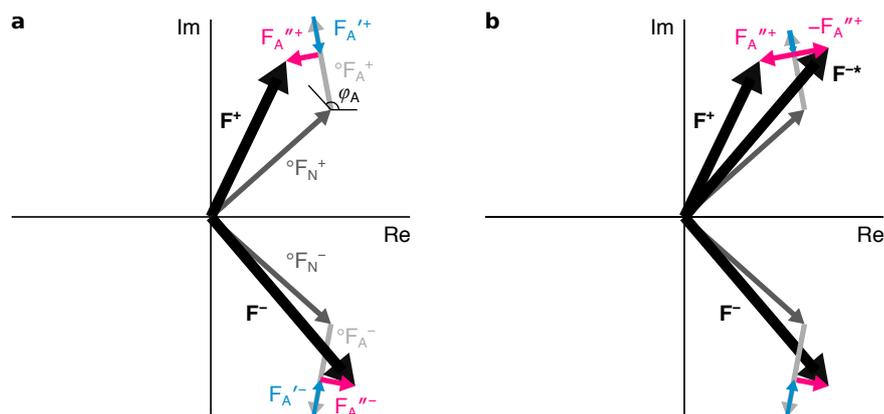


Figure B.1. Breaking Friedel's law. (a) F_A'' always has a phase shift of $+\pi/2$ relative to ${}^\circ F_A$. As a result, structure factor amplitudes $|F^+|$ and $|F^-|$ are no longer equal and phases are no longer opposite. (b) The complex conjugate of F^- , F^{-*} , is useful for phase calculation. F^{-*} is related to F^+ by $2 F_A''$.

The set of $|{}^\circ F_A|$ from all reflections can be used to determine the atomic coordinates of anomalous scatterers. Phases calculated from the substructure, $\phi_{A,\text{calc}}$, can be used in turn to estimate the phases corresponding to all atoms. These phases are improved upon in density modification, which also enables the correct hand of the solution to be selected.

1.11.3 Substructure solution

A Patterson map can be calculated based on $|{}^\circ F_A|^2$ which describes the interatomic vectors between all anomalous scatterers. In a multi-wavelength anomalous diffraction (MAD) experiment, equation (1) can be solved for values of $|{}^\circ F_A|$. In a single-wavelength anomalous diffraction (SAD) experiment, $|{}^\circ F_A|$ can be approximated by assuming that the anomalous scattering is small relative to the total scattering. Under this assumption, $|F_T|$ can be estimated as $(|F^+| + |F^-|) / 2$. Substituting into equation (1) and subtracting,

$$|F^+| - |F^-| \approx c |{}^\circ F_A| \sin \alpha$$

In SHELX, the coefficient c is ignored due to the use of normalised structure factor amplitudes, or E values. Because only reflections with the largest anomalous differences are used and α is usually close to 90° or 270° for these reflections, good normalised estimates of $|{}^\circ F_A|$ can be obtained from $|F^+| - |F^-|$.

A dual-space method is employed in SHELXD for substructure solution (182). In each trial, a pair of atoms is placed according to a random strong peak in the Patterson map. The vector between the atoms is used to calculate a Patterson superposition minimum function,

which contains fewer peaks than the original Patterson map. Additional atoms are placed based on these peaks, and the algorithm cycles between reciprocal space and real space for phase refinement and atom placement, respectively. In the final stage, occupancies of the placed atoms are refined. Correlation coefficients above the noise level between observed and calculated E values for all reflections and unused reflections indicate a good solution. Because the Patterson function is centrosymmetric, both hands of the substructure solution are valid solutions. The correct hand with interpretable density in real space is chosen later.

1.11.4 Phasing and hand determination

Given structure factor amplitudes $|\mathbf{F}^+|$ and $|\mathbf{F}^-|$, and ${}^\circ\mathbf{F}_A$ from one hand of the substructure, φ_T can be determined for each reflection. The problem can be understood using a Harker diagram (Figure B.2). Two solutions are possible in SAD, resulting in a phase ambiguity. One approach to resolving the ambiguity is to take the average phase (265). This is implemented in SHELX, where α is set to 90° when $|\mathbf{F}^+| \gg |\mathbf{F}^-|$ and 270° when $|\mathbf{F}^+| \ll |\mathbf{F}^-|$. A maximum likelihood approach is used in Phaser (266). The probability of $|\mathbf{F}^+|$ and $|\mathbf{F}^-|$ given \mathbf{F}_H^+ and \mathbf{F}_H^- calculated from the substructure is expressed as a SAD likelihood function (Figure B.3). Variance terms in the function and the positions, occupancies and B factors of atoms in the substructure can be refined by maximising the likelihood. The amplitude of the final maximum likelihood estimate relative to F_{obs} is taken as a figure of merit (267). The approach applies to other experimental phasing methods, for which likelihood functions have been constructed (268).

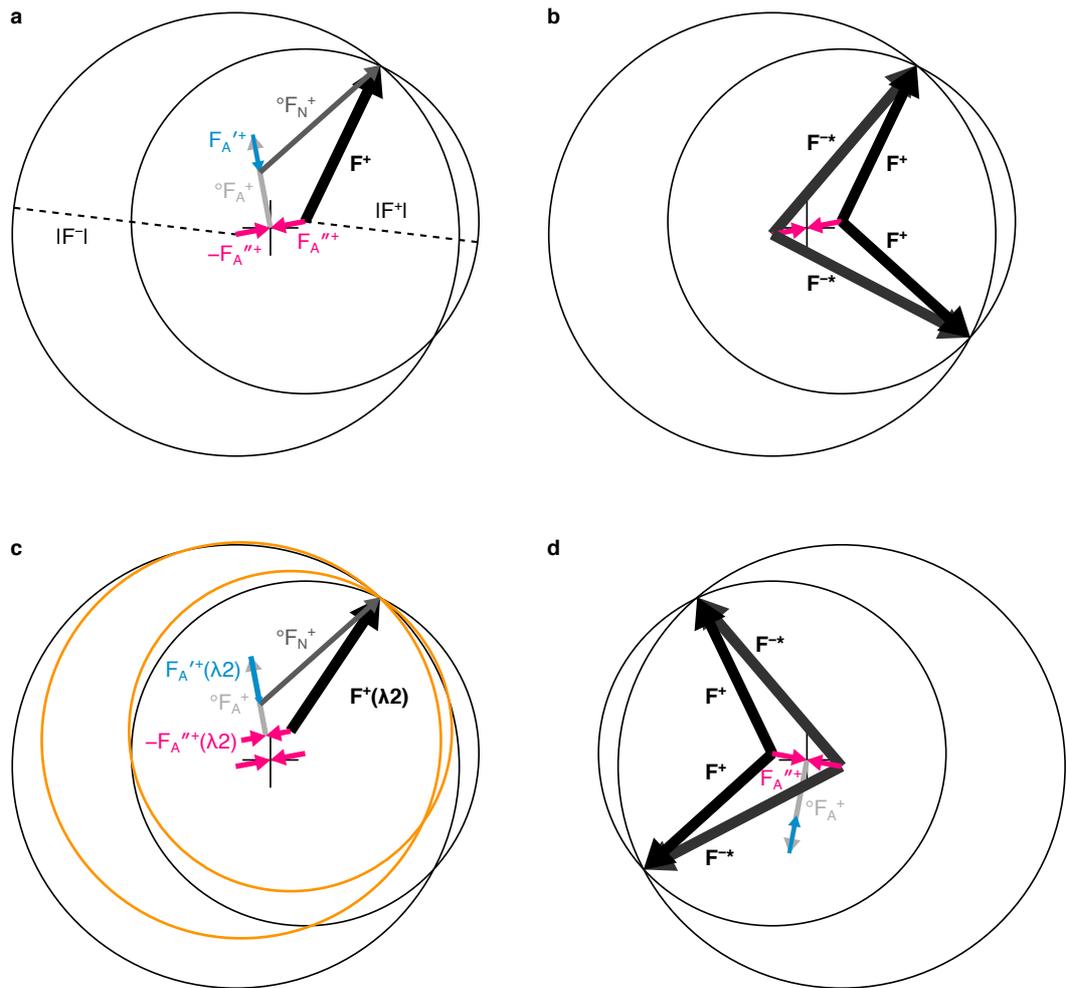


Figure B.2. Harker constructions for anomalous diffraction experiments. (a) SAD construction. Circles of radii $|F^-|$ and $|F^+|$ with centres $\pm F_A''$ away from an arbitrary origin are drawn. Vectors from the origins of the circles to the intersections formed satisfy the condition $F = {}^\circ F_N + {}^\circ F_A + F_A' + F_A''$. (b) Phase ambiguity from two possible solutions in SAD. (c) MAD construction. ${}^\circ F_N$ and ${}^\circ F_A$ are constant but the amplitudes of F_A' and F_A'' are wavelength-dependent. The pair of circles corresponding to the second set of F_A' and F_A'' are coloured orange. Errors aside, the phase solution is unambiguous. (d) SAD construction with a substructure of opposite hand. Solutions are reflected through the imaginary axis.

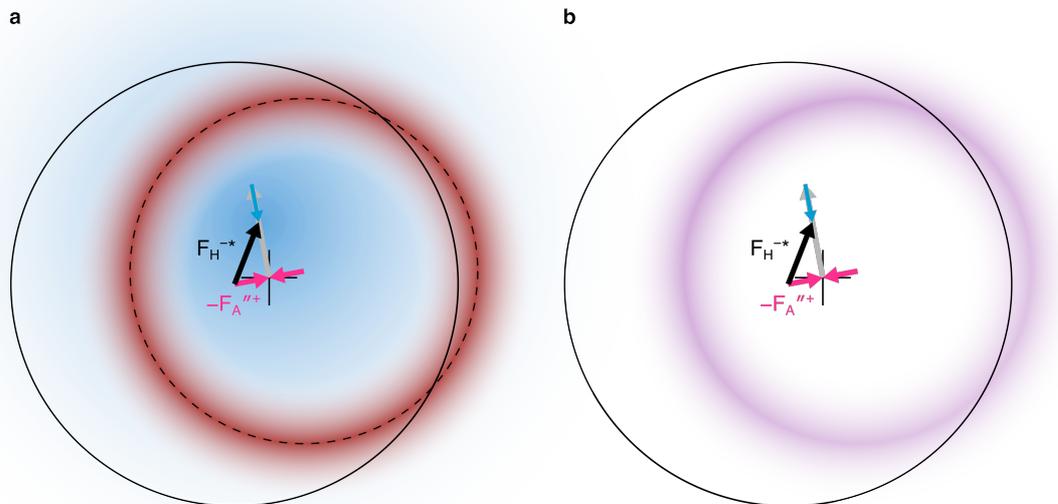


Figure B.3. Probabilistic SAD Harker construction. (a) The probability $P(F^{-*} | F_H^{-*})$ in blue is a Sim distribution (269) centred on F_H^{-*} . The probability $P(|F^+| | F^{-*}, F_H^+, F_H^{-*})$ in red is centred around the dotted circle of radius $|F^+|$. (b) The product of the two probabilities is shown in purple. Likelihood is given by integration of this product under the black circle of radius $|F^-|$.

The opposite hands of the substructure have opposite phases (Figure B.2d). Their structure factors are complex conjugates of each other. In SAD, the final calculated structure factors for density modification will be related by a reflection through the imaginary axis (270). This is equivalent to taking the complex conjugate and applying a 180° phase shift. The corresponding Fourier maps are thus mirror images of each other with densities of opposite signs. The correct hand will produce positive density resembling true density superposed with noise, whereas the positive density for the incorrect hand contains only noise. Density modification often allows the correct hand to be identified.

1.11.5 Density modification

A number of density modification methods are routinely used for phase improvement to generate interpretable maps for model building after phase determination. These are:

Solvent flattening. A molecular envelope is defined, and density in the solvent region is set to a constant while negative protein density is set to zero (271). The molecular envelope is determined by constructing a map where the density at each grid point is the weighted sum of densities within a sphere of radius 8–12 Å. A threshold is then chosen such that the

volume below the threshold corresponds to the expected solvent content of the crystal, thus defining the solvent region. This calculation was performed originally in real space, but it is now implemented in reciprocal space (272).

Histogram matching. Density values are scaled and shifted so that their distribution matches that of a simulated map based on reference data (273). Reference data are truncated and rescaled; then, a density map is calculated based on the structure and adjusted figures of merit. Finally, a solvent mask is applied and density values in the protein region are used to construct the reference histogram (184).

Pairwise-weighted non-crystallographic symmetry (NCS) averaging. The local correlation between NCS-related regions relative to the standard deviation of correlation between unrelated regions is used to generate a weighted averaging mask (184). A weighted average of densities from related grid points is calculated according to the mask.

Gamma correction. Bias is removed by subtracting the contribution of the initial structure factor from the final structure factor (274). In the case of solvent flattening, this results in solvent flipping. This method has been generalised to yield better phase estimates following an arbitrary density modification procedure (275).

These methods are implemented in Parrot. Density modification in SHELXE is carried out differently (276). First, 5% of grid points with the strongest density are assigned to the protein region. The remaining points are ranked according to the variance in density in a 2.42 Å radius of these points. A crossover region is centred at the expected solvent content level. Points with lower variances before the region are designated solvent and flipped. Points with higher variances are designated protein and weighted according to variance. Points in the crossover region receive a weighted sum of the treatments. This density modification procedure is combined iteratively with chain tracing for even better phase improvements (277).

B.2 Nuclear magnetic resonance spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy provides information about the local chemical environment and dynamics of nuclei in a molecule. A nucleus of spin quantum number I can have spin angular momentum with a z -component $m\hbar$, where $m = -I, -I + 1, \dots, I$. In a magnetic field \mathbf{B}_0 oriented along the z -axis, the nuclear spin will have magnetic moment $\mu = \gamma m\hbar$ according to the gyromagnetic ratio γ and have energy

$$E = -\mu \cdot \mathbf{B}_0 = -\gamma m\hbar B_0$$

where B_0 is the strength of the magnetic field.

For ^1H , ^{13}C and ^{15}N nuclei, $I = 1/2$. In a magnetic field, two energy states are possible: a lower energy state, α , with $m = 1/2$, and a higher energy state, β , $m = -1/2$. At equilibrium, nuclei are distributed between the two states according to the Boltzmann distribution:

$$\frac{N_\beta}{N_\alpha} = e^{-(E_\beta - E_\alpha)/kT} = e^{-h\nu/kT}$$

where ν is the frequency associated with the energy difference between states. ν is also the Larmor frequency at which a spin precesses about the z -axis. The Larmor frequency of a spin is dependent on its local chemical environment. A nucleus adjacent to an electron-withdrawing group will become deshielded and have a higher frequency or a downfield chemical shift.

At equilibrium, the magnetic moments of individual nuclei are oriented such that there is a net magnetisation along the applied field, which is described by a magnetisation vector pointing along z . The magnetisation vector can be rotated onto the xy plane by excitation of nuclear spins using radio frequency. As the vector precesses, currents are induced in the coils of the spectrometer placed in the xy plane. Relaxation during acquisition produces a free induction decay (FID). Fourier transform of the FID gives the NMR spectrum.

The NMR signal of a protein can be resolved by correlating nuclei through bonds in a multidimensional heteronuclear experiment. The basic experiment involves excitation of nuclei in the direct dimension, usually ^1H , transfer of magnetisation to nuclei in the indirect dimension, allowing spins to evolve during a period of time t_1 , and transfer again

back to the direct dimension for acquisition. By collecting a series of spectra with different t_1 values, the precession of correlated spins in the indirect dimension for each frequency in the direct dimension can be determined. Subsequent Fourier transforms produces the NMR spectrum. Transfer of magnetisation through bonds is possible due to scalar or J -coupling. Every bond type has an associated coupling constant. Maximum transfer occurs when the delay time in the INEPT pulse sequence equals one-half of the reciprocal of the constant. Magnetisation transfer through dipolar interactions allows nuclei to be correlated through space, such as in a nuclear Overhauser effect spectroscopy (NOESY) experiment, where cross-relaxation between spins during the mixing period generates cross peaks (201).

By correlating different sets of nuclei in a protein, the connectivity between spins can be established. Spins can then be assigned to sequence based on the characteristic chemical shifts of certain residue types, for example, the low $C\beta$ shift of alanine, the high $C\beta$ shifts of serine and threonine, and the low $C\alpha$ shift of glycine with an inverted sign. Once assigned, chemical shift perturbation on interaction of the protein with DNA, for example, can be mapped. Perturbations are often followed via a series of heteronuclear single quantum coherence (HSQC) spectra, which will have cross peaks for backbone amides, and the $N\delta$ - $H\delta_2$, $N\epsilon$ - $H\epsilon_2$ and $N\epsilon_1$ - $H\epsilon_1$ groups of asparagine, glutamine and tryptophan residues, respectively (204).

B.3 Electron microscopy

Objects are visualised in transmission electron microscopy by their scattering of electrons. The observed 2D projections can be used to reconstruct the 3D structure of the object. Macromolecular objects were first visualised by negative staining (278). With the development of cryo-imaging methods (279), the introduction of direct electron detectors (280), phase plates (281), and the development of new reconstruction routines (224,225,282), the three-dimensional structure of macromolecular complexes can be determined to resolutions higher than 4 Å.

The wave describing electrons at the exit surface of a specimen is given by

$$\tau(\mathbf{r}) = \tau_0 e^{i\phi(\mathbf{r})}$$

$$\phi(\mathbf{r}) = \pi\lambda \int U(\mathbf{r}, z) dz$$

where τ_0 is the incoming wave, $\phi(\mathbf{r})$ is the phase shift due to scattering, which is related to electric potential in the specimen (283). Under the weak-phase-object approximation, applicable to thin samples, only a small number of electrons interact with the specimen. Thus, $\phi(\mathbf{r}) \ll 1$. In the absence of absorption, the wave can be approximated as

$$\tau(\mathbf{r}) = \tau_0 [1 + \phi(\mathbf{r})]$$

Image intensity is given by $\tau(\mathbf{r})\tau^*(\mathbf{r})$. At zero defocus, this is a constant. Changing the defocus introduces an additional phase shift, which produces phase contrast. To account for high-angle scattering and inelastic scattering, an amplitude component is introduced:

$$\tau(\mathbf{r}) = \tau_0 [1 + \phi(\mathbf{r}) + u(\mathbf{r})]$$

Assuming that the Fourier transforms of $\phi(\mathbf{r})$ and $u(\mathbf{r})$ are related by a constant amplitude contrast ratio A , the Fourier transform of image intensity can be written as a product of the Fourier transform of $\phi(\mathbf{r})$, the contrast transfer function (CTF) and envelope functions plus additive noise (284). Envelope functions attenuate at high frequencies. They account for the effects of a finite source size (285), defocus fluctuations (286), specimen charging (287), and detector modulation transfer function (288).

The two-dimensional CTF (289) is given by

$$\text{CTF} = -w_1 \sin[\chi(\lambda, \mathbf{g}, \Delta f, C_s)] - w_2 \cos[\chi(\lambda, \mathbf{g}, \Delta f, C_s)]$$

$$\chi(\lambda, \mathbf{g}, \Delta f, C_s) = \pi \lambda |\mathbf{g}|^2 (\Delta f - 1/2 \lambda^2 |\mathbf{g}|^2 C_s)$$

$$\Delta f = \frac{1}{2} [\text{DF}_1 + \text{DF}_2 + (\text{DF}_1 - \text{DF}_2) \cos(2[\alpha_g - \alpha_{ast}])]]$$

$$w_1 = \sqrt{1 - A^2} \quad , \quad w_2 = A$$

where $\mathbf{g} = \mathbf{k}' - \mathbf{k}$ is the scattering vector in Fourier space, constants w_1 and w_2 are determined by amplitude contrast, C_s is the spherical aberration, Δf is the defocus, described by defocus values DF_1 and DF_2 . Astigmatism has magnitude $\text{DF}_1 - \text{DF}_2$ and azimuth α_{ast} , measured as the angle between the x -axis and the direction related to DF_1 . α_g is the angle between the x -axis and the scattering vector. Astigmatism can arise due to an elliptical beam cross-section, or tilt in the sample due to a difference in thermal expansion coefficient between the carbon support and copper grid.

The CTF causes phase reversals and amplitude modulations that manifest as Thon rings in the power spectrum of an image. Defocus and astigmatism are estimated in CTFFIND4 by maximising the correlation coefficient between calculated CTF and the background-subtracted power spectrum (221). The background, due to inelastic scattering, is calculated as the square root of the power spectrum after low-pass or mean filtering. Astigmatism is first estimated as half the angle through which to rotate the spectrum onto a mirrored version of itself (290). Defocus values are estimated using a grid search, and finally all three parameters are optimised using the conjugated gradient method. Given a number of images describing the same object at different defocus values, information near the zero crossings of the CTF in each image can be recovered by CTF correction using a Wiener filter (291).

The central section theorem states that the Fourier transform of a projection of a 3D object is the central 2D cross-section of the 3D Fourier transform perpendicular to the projection vector (292). Given sufficient angular sampling, the 3D object can be reconstructed. The orientation and origin of motion-corrected particles are refined using a maximum likelihood approach in RELION (225). The posterior distribution is optimised given prior

distributions assuming random rotations and a Gaussian distribution of origins. For 2D and 3D classification, an additional parameter is introduced, that is, the object to which the image corresponds. Every possible orientation and origin of each particle contributes to the reconstructed volume, weighted according to the posterior distribution. This helps to reduce bias from the reference volume (293). By dividing particles into independent half sets and performing reconstructions based on each half set independently, the gold standard procedure reduces over-fitting during the refinement process (234). The resolution of a reconstruction is determined by the Fourier shell correlation (FSC) between maps calculated from the two half sets (294).

$$\text{FSC} = \frac{\sum \mathbf{F}_1 \cdot \mathbf{F}_2^*}{\sqrt{\sum |\mathbf{F}_1|^2 \sum |\mathbf{F}_2^*|^2}}$$

where \mathbf{F}_1 and \mathbf{F}_2 correspond to the Fourier transform of the half maps. The agreement between the map calculated using the full dataset and a perfect map is given by

$$C_{ref} = \sqrt{\frac{2 \text{FSC}}{1 + \text{FSC}}}$$

These are similar to the correlation coefficients $CC_{1/2}$ and CC^* in X-ray crystallography (295). A FSC of 0.143 corresponds to a C_{ref} of 0.5, which corresponds to a phase error of 60° . This FSC value is used as a threshold for defining the resolution of the reconstruction.

B.4 Total internal reflection fluorescence microscopy

In total internal reflection fluorescence microscopy (TIRFM), fluorophores close to the boundary between two media are excited using an evanescent wave (296). Single-molecule events *in vitro* or *in vivo* close to the surface of cells can be followed.

The reflection and transmission of light when light of frequency ω strikes the boundary between two isotropic media with refractive indices n_i and n_t are described by the law of reflection and Snell's law, respectively:

$$\theta_i = \theta_r$$

$$n_i \sin \theta_i = n_t \sin \theta_t$$

where θ_i is the angle of incidence, θ_r is the angle of reflection, θ_t is the angle of refraction, measured from the normal of the boundary. When $n_i > n_t$ and θ_i is greater than the critical angle $\theta_c = \sin^{-1} (n_t / n_i)$, total internal reflection occurs. Given a boundary at $z = 0$ and incidence plane $y = 0$, the z -component of the transmitted wave is imaginary. The electric field associated with the transmitted wave is complex:

$$\mathbf{E}_t = \mathbf{E}_{t,0} e^{\pm z \xi} e^{i[(k_x n_i \sin \theta_i / n_t) - \omega t]} \quad , \quad \xi = \frac{k_t}{n_t} \sqrt{n^2 \sin^2 \theta_i - n_t^2}$$

The wave propagates in the x -direction but decays exponentially in the z -direction. The intensity of the wave has decay constant $1/d$, where d is the depth of penetration:

$$d = \frac{\lambda_0}{4\pi} \sqrt{n^2 \sin^2 \theta_i - n_t^2}$$

The depth of penetration is greatest when θ_i approaches the critical angle. Applying Fresnel's equations to separate \mathbf{E}_t into s -polarised and p -polarised components, parallel and perpendicular to the incidence plane, respectively, the intensities associated are

$$I_0^\perp = I_i^\perp \frac{4 \cos^2 \theta_i}{(1 - n^2)}$$

$$I_0^\parallel = I_i^\parallel \frac{4 \cos^2 \theta_i (2 \sin^2 \theta_i - n^2)}{n^4 \cos^2 \theta_i + \sin^2 \theta_i - n^2}$$

where $n = n_t / n_i$. The refractive index of quartz glass is 1.51 and that of water is 1.33. In a typical microscope, the intensity of the evanescent wave can be four or five times greater than that of the incident light when $\theta_i \approx \theta_c$. A prism or the objective itself can be used to deliver the incident light. The resulting evanescent wave typically has a penetration depth of 200 nm. Thus, it can be used to selectively excite fluorophores close to the boundary.

In any optical experiment, the true image is convoluted with the point spread function (PSF) of the microscope (297). Given a point source, light is emitted as a spherical wave. When a portion of this wave is captured by a perfect objective, an Airy function is observed in the image plane due to diffraction. The numerical aperture of the objective determines the full-width at half maximum of the PSF and thus resolution. The image is further modified due to optical aberrations and noise. Under photon-limited conditions, which are typical in single-molecule TIRFM experiments, the number of photons measured by a given detector element follows a Poisson distribution. For time resolution, short exposure times are often also necessary, as is the case for the work in this thesis. To increase the signal, electron multiplying CCD cameras are used. Charges from the CCD layer are passed or clocked through a multiplication register, where additional charges are created by impact ionisation when charges travel from one cell of the register to the next. This also amplifies noise. Consequently, data from single-molecule TIRFM experiments tend to have a low signal-to-noise ratio. Data analysis routines capable of extracting signal from noise are essential to the interpretation of TIRFM data.

Abbreviations

AEBSF	4-(2-Aminoethyl)benzenesulfonyl fluoride
AUC	Analytical ultracentrifugation
BMRB	Biological Magnetic Resonance Bank
BSA	Bovine serum albumin
CCD	Charge-coupled device
CD	Circular dichroism
DOC	Sodium deoxycholate
dsDNA	Double-stranded deoxyribonucleic acid
DSS	4,4-Dimethyl-4-silapentane-1-sulfonic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
EMDB	Electron Microscopy Data Bank
FSC	Fourier shell correlation
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
IPTG	Isopropyl β -D-1-thiogalactopyranoside
LB	Lysogeny broth
Mant	2'/3'-O-N-methylanthraniloyl
MES	2-(N-Morpholino)ethanesulfonic acid
MS	Mass spectrometry

NA	Numerical aperture
NMR	Nuclear magnetic resonance
OD ₆₀₀	Optical density at 600 nm
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
PFU	Plaque-forming unit
PSF	Point spread function
SDS	Sodium dodecyl sulfate
SEC-MALLS	Size exclusion chromatography with multi-angle laser-light scattering
SPR	Surface plasmon resonance
ssRNA	Single-stranded ribonucleic acid
TCA	Trichloroacetic acid
TCA	Trichloroacetic acid
TIRFM	Total internal fluorescence microscopy
Tris	Tris(hydroxymethyl)aminomethane
USAF	United States Air Force

References

1. Hershey, A. D., and Chase, M. (1952) Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J. Gen. Physiol.* **36**, 39–56
2. Crick, F. H., Barnett, L., Brenner, S., and Watts-Tobin, R. J. (1961) General nature of the genetic code for proteins. *Nature* **192**, 1227–1232
3. Weiss, B., and Richardson, C. C. (1967) Enzymatic breadage and joining of deoxyribonucleic acid. 3. An enzyme-adenylate intermediate in the polynucleotide ligase reaction. *J. Biol. Chem.* **242**, 4270–4272
4. Jackson, D. A., Symons, R. H., and Berg, P. (1972) Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **69**, 2904–2909
5. Lobban, P. E., and Kaiser, A. D. (1973) Enzymatic end-to end joining of DNA molecules. *J. Mol. Biol.* **78**, 453–471
6. Zinder, N. D., and Lederberg, J. (1952) Genetic exchange in *Salmonella*. *J. Bacteriol.* **64**, 679–699
7. Mann, R., Mulligan, R. C., and Baltimore, D. (1983) Construction of a retrovirus packaging mutant and its use to produce helper-free defective retrovirus. *Cell* **33**, 153–159
8. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., and Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712
9. Pennisi, E. (2013) The CRISPR craze. *Science* **341**, 833–836
10. Wilson, W. H., Schroeder, D. C., Allen, M. J., Holden, M. T., Parkhill, J., Barrell, B. G., Churcher, C., Hamlin, N., Mungall, K., Norbertczak, H., Quail, M. A., Price, C., Rabinowitsch, E., Walker, D., Craigon, M., Roy, D., and Ghazal, P. (2005) Complete genome sequence and lytic phase transcription profile of a Coccolithovirus. *Science* **309**, 1090–1092

11. Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., and Claverie, J. M. (2004) The 1.2-megabase genome sequence of Mimivirus. *Science* **306**, 1344–1350
12. Koonin, E. V., and Martin, W. (2005) On the origin of genomes and cells within inorganic compartments. *Trends Genet.* **21**, 647–654
13. Bamford, D. H. (2003) Do viruses form lineages across different domains of life? *Res. Microbiol.* **154**, 231–236
14. Krupovic, M., and Koonin, E. V. (2017) Multiple origins of viral capsid proteins from cellular ancestors. *Proc. Natl. Acad. Sci. USA* **114**, E2401–E2410
15. Twort, F. W. (1915) An investigation on the nature of ultra-microscopic viruses. *The Lancet* **186**, 1241–1243
16. D’Herelle, F. (1917) Sur un microbe invisible antagoniste des bacilles dysentériques. *C. R. Hebd. Seances Acad. Sci.* **165**, 373–375
17. Ellis, E. L., and Delbruck, M. (1939) The Growth of Bacteriophage. *J. Gen. Physiol.* **22**, 365–384
18. Luria, S. E., and Anderson, T. F. (1942) The identification and characterization of bacteriophages with the electron microscope. *Proc. Natl. Acad. Sci. USA* **28**, 127–130 121
19. Luria, S. E., Delbruck, M., and Anderson, T. F. (1943) Electron microscope studies of bacterial viruses. *J. Bacteriol.* **46**, 57–77
20. Lwoff, A. (1950) Recherches sur un *Bacillus megatherium* lysogène. *Ann. Inst. Pasteur* **78**, 711–739
21. Kozloff, L. M., and Putnam, F. W. (1950) Biochemical studies of virus reproduction: III. The origin of virus phosphorus in the escherichia coli T6 bacteriophage system. *J. Biol. Chem.* **182**, 229–242

22. Kozloff, L. M., Knowlton, K., Putnam, F. W., and Evans, E. A., Jr. (1951) Biochemical studies of virus reproduction. V. The origin of bacteriophage nitrogen. *J. Biol. Chem.* **188**, 101–116
23. Watson, J. D., and Crick, F. H. (1953) Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738
24. Meselson, M., and Stahl, F. W. (1958) The replication of DNA in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **44**, 671–682
25. Epstein, R. H., Bolle, A., Steinberg, C. M., Kellenberger, E., Boy de la Tour, E., Chevalley, R., Edgar, R. S., Susman, M., Denhardt, G. H., and Lielausis, A. (1963) Physiological studies of conditional lethal mutants of bacteriophage T4D. *Cold Spring Harbor Symp. Quant. Biol.* **28**, 375–394
26. Karamata, D., Kellenberger, E., Kellenberger, G., and Terzi, M. (1962) Sur une particule accompagnant le développement du coliphage λ . *Pathol. Microbiol.* **25**, 575–585
27. Campbell, A. (1961) Sensitive mutants of bacteriophage λ . *Virology* **14**, 22–32
28. Weigle, J. J. (1953) Induction of mutations in a bacterial virus. *Proc. Natl. Acad. Sci. USA* **39**, 628–636
29. Casjens, S., Hohn, T., and Kaiser, A. D. (1970) Morphological proteins of phage lambda: identification of the major head protein as the product of gene E. *Virology* **42**, 496–507
30. Georgopoulos, C. P., Hendrix, R. W., Casjens, S. R., and Kaiser, A. D. (1973) Host participation in bacteriophage lambda head assembly. *J. Mol. Biol.* **76**, 45–60
31. Hendrix, R. W., and Casjens, S. E. (1975) Assembly of bacteriophage lambda heads: Protein processing and its genetic control in petit λ assembly. *J. Mol. Biol.* **91**, 187–199
32. Hohn, B., and Hohn, T. (1974) Activity of empty, headlike particles for packaging of DNA of bacteriophage λ *in vitro*. *Proc. Natl. Acad. Sci. USA* **71**, 2372–2376

33. Kaiser, D., Syvanen, M., and Masuda, T. (1975) DNA packaging steps in bacteriophage lambda head assembly. *J. Mol. Biol.* **91**, 175–186
34. Weigle, J. (1966) Assembly of phage lambda *in vitro*. *Proc. Natl. Acad. Sci. USA* **55**, 1462–1466
35. Campbell, A. (1959) Ordering of genetic sites in bacteriophage λ by the use of galactose-transducing defective phages. *Virology* **9**, 293–305
36. Casjens, S., Hohn, T., and Kaiser, A. D. (1972) Head assembly steps controlled by genes F and W in bacteriophage λ . *J. Mol. Biol.* **64**, 551–563
37. Botstein, D., and Matz, M. J. (1970) A recombination function essential to the growth of bacteriophage P22. *J. Mol. Biol.* **54**, 417–440
38. Méndez, J., Blanco, L., Esteban, J. A., Bernad, A., and Salas, M. (1992) Initiation of phi 29 DNA replication occurs at the second 3' nucleotide of the linear template: a sliding-back mechanism for protein-primed DNA replication. *Proc. Natl. Acad. Sci. USA* **89**, 9579–9583
39. Kamtekar, S., Berman, A. J., Wang, J., Lazaro, J. M., de Vega, M., Blanco, L., Salas, M., and Steitz, T. A. (2006) The phi29 DNA polymerase:protein-primer structure suggests a model for the initiation to elongation transition. *EMBO J.* **25**, 1335–1343
40. Hershey, A. D., Burgi, E., and Ingraham, L. (1963) Cohesion of DNA molecules isolated from phage lambda. *Proc. Natl. Acad. Sci. USA* **49**, 748–755
41. Kaiser, A. D., and Hogness, D. S. (1960) The transformation of *Escherichia coli* with deoxyribonucleic acid isolated from bacteriophage lambda-dg. *J. Mol. Biol.* **2**, 392–415
42. Kaiser, A. D., and Inman, R. B. (1965) Cohesion and the biological activity of bacteriophage lambda DNA. *J. Mol. Biol.* **13**, 78–91
43. Strack, H. B., and Kaiser, A. D. (1965) On the structure of the ends of lambda DNA. *J. Mol. Biol.* **12**, 36–49

44. Better, M., and Freifelder, D. (1983) Studies on the replication of Escherichia coli phage lambda DNA. I. The kinetics of DNA replication and requirements for the generation of rolling circles. *Virology* **126**, 168–182
45. Taylor, K., and Węgrzyn, G. (1995) Replication of coliphage lambda DNA. *FEMS Microbiol. Rev.* **17**, 109–119
46. Dodson, M., Echols, H., Wickner, S., Alfano, C., Mensa-Wilmot, K., Gomes, B., LeBowitz, J., Roberts, J. D., and McMacken, R. (1986) Specialized nucleoprotein structures at the origin of replication of bacteriophage lambda: localized unwinding of duplex DNA by a six-protein reaction. *Proc. Natl. Acad. Sci. USA* **83**, 7638–7642
47. Belanger, K. G., and Kreuzer, K. N. (1998) Bacteriophage T4 initiates bidirectional DNA replication through a two-step process. *Mol. Cell* **2**, 693–701
48. Liu, J., and Morrical, S. W. (2010) Assembly and dynamics of the bacteriophage T4 homologous recombination machinery. *Virol. J.* **7**, 357
49. Walker, J. E., Auffret, A. D., Carne, A., Gurnett, A., Hanisch, P., Hill, D., and Saraste, M. (1982) Solid-phase sequence analysis of polypeptides eluted from polyacrylamide gels. An aid to interpretation of DNA sequences exemplified by the Escherichia coli unc operon and bacteriophage lambda. *Eur. J. Biochem.* **123**, 253–260
50. Medina, E. M., Andrews, B. T., Nakatani, E., and Catalano, C. E. (2011) The bacteriophage lambda gpNu3 scaffolding protein is an intrinsically disordered and biologically functional procapsid assembly catalyst. *J. Mol. Biol.* **412**, 723–736
51. Murialdo, H., and Becker, A. (1978) A Genetic analysis of bacteriophage lambda prohead assembly in vitro. *J. Mol. Biol.* **125**, 57–74
52. Medina, E., Wiczorek, D., Medina, E. M., Yang, Q., Feiss, M., and Catalano, C. E. (2010) Assembly and maturation of the bacteriophage lambda procapsid: gpC is the viral protease. *J. Mol. Biol.* **401**, 813–830

53. Hendrix, R. W., and Casjens, S. R. (1974) Protein fusion: a novel reaction in bacteriophage lambda head assembly. *Proc. Natl. Acad. Sci. USA* **71**, 1451–1455
54. Hohn, T., Flick, H., and Hohn, B. (1975) Petit lambda, a family of particles from coliphage lambda infected cells. *J. Mol. Biol.* **98**, 107–120
55. Chen, D. H., Baker, M. L., Hryc, C. F., DiMaio, F., Jakana, J., Wu, W., Dougherty, M., Haase-Pettingell, C., Schmid, M. F., Jiang, W., Baker, D., King, J. A., and Chiu, W. (2011) Structural basis for scaffolding-mediated assembly and maturation of a dsDNA virus. *Proc. Natl. Acad. Sci. USA* **108**, 1355–1360
56. Ganser, B. K., Li, S., Klishko, V. Y., Finch, J. T., and Sundquist, W. I. (1999) Assembly and Analysis of Conical Models for the HIV-1 Core. *Science* **283**, 80–83
57. Battisti, A. J., Yoder, J. D., Plevka, P., Winkler, D. C., Prasad, V. M., Kuhn, R. J., Frey, T. K., Steven, A. C., and Rossmann, M. G. (2012) Cryo-electron tomography of rubella virus. *J. Virol.* **86**, 11078–11085
58. Crick, F. H., and Watson, J. D. (1956) Structure of small viruses. *Nature* **177**, 473–475
59. Crick, F. H. C., and Watson, J. D. (1957) Virus Structure: General Principles. in *Ciba Foundation Symposium - The Nature of Viruses* (Wolstenholme, G. E. W., and Millar, E. C. P. eds.), John Wiley & Sons, Ltd., Chichester, UK. pp 5–18
60. Franklin, R. E., and Holmes, K. C. (1956) The helical arrangement of the protein subunits in tobacco mosaic virus. *Biochim. Biophys. Acta* **21**, 405–406
61. Bernal, J. D., and Carlisle, C. H. (1948) Unit cell measurements of wet and dry crystalline turnip yellow mosaic virus. *Nature* **162**, 139
62. Caspar, D. L., and Klug, A. (1962) Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* **27**, 1–24
63. Tao, Y., Olson, N. H., Xu, W., Anderson, D. L., Rossmann, M. G., and Baker, T. S. (1998) Assembly of a tailed bacterial virus and its genome release studied in three dimensions. *Cell* **95**, 431–437

64. Fokine, A., Chipman, P. R., Leiman, P. G., Mesyanzhinov, V. V., Rao, V. B., and Rossmann, M. G. (2004) Molecular architecture of the prolate head of bacteriophage T4. *Proc. Natl. Acad. Sci. USA* **101**, 6003–6008
65. Choi, K. H., Morais, M. C., Anderson, D. L., and Rossmann, M. G. (2006) Determinants of bacteriophage ϕ 29 head morphology. *Structure* **14**, 1723–1727
66. Helgstrand, C., Wikoff, W. R., Duda, R. L., Hendrix, R. W., Johnson, J. E., and Liljas, L. (2003) The refined structure of a protein catenane: the HK97 bacteriophage capsid at 3.44 Å resolution. *J. Mol. Biol.* **334**, 885–899
67. Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K., and Bricogne, G. (1978) Tomato bushy stunt virus at 2.9 Å resolution. *Nature* **276**, 368–373
68. Hendrix, R. W. (1978) Symmetry mismatch and DNA packaging in large bacteriophages. *Proc. Natl. Acad. Sci. USA* **75**, 4779–4783
69. Wikoff, W. R., Duda, R. L., Hendrix, R. W., and Johnson, J. E. (1998) Crystallization and preliminary X-ray analysis of the dsDNA bacteriophage HK97 mature empty capsid. *Virology* **243**, 113–118
70. Wikoff, W. R., Liljas, L., Duda, R. L., Tsuruta, H., Hendrix, R. W., and Johnson, J. E. (2000) Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* **289**, 2129–2133
71. Ding, Y., Duda, R. L., Hendrix, R. W., and Rosenberg, J. M. (1995) Complexes between chaperonin GroEL and the capsid protein of bacteriophage HK97. *Biochemistry* **34**, 14918–14931
72. Xie, Z., and Hendrix, R. W. (1995) Assembly in vitro of bacteriophage HK97 proheads. *J. Mol. Biol.* **253**, 74–85
73. Duda, R. L., Martincic, K., and Hendrix, R. W. (1995) Genetic basis of bacteriophage HK97 prohead assembly. *J. Mol. Biol.* **247**, 636–647
74. Duda, R. L., Hempel, J., Michel, H., Shabanowitz, J., Hunt, D., and Hendrix, R. W. (1995) Structural transitions during bacteriophage HK97 head assembly. *J. Mol. Biol.* **247**, 618–635

75. Oh, B., Moyer, C. L., Hendrix, R. W., and Duda, R. L. (2014) The delta domain of the HK97 major capsid protein is essential for assembly. *Virology* **456-457**, 171–178
76. Conway, J. F., Duda, R. L., Cheng, N., Hendrix, R. W., and Steven, A. C. (1995) Proteolytic and conformational control of virus capsid maturation: the bacteriophage HK97 system. *J. Mol. Biol.* **253**, 86–99
77. Gertsman, I., Gan, L., Guttman, M., Lee, K., Speir, J. A., Duda, R. L., Hendrix, R. W., Komives, E. A., and Johnson, J. E. (2009) An unexpected twist in viral capsid maturation. *Nature* **458**, 646–650
78. Lander, G. C., Evilevitch, A., Jeembaeva, M., Potter, C. S., Carragher, B., and Johnson, J. E. (2008) Bacteriophage lambda stabilization by auxiliary protein gpD: timing, location, and mechanism of attachment determined by cryo-EM. *Structure* **16**, 1399–1406
79. Duda, R. L. (1998) Protein chainmail: catenated protein in viral capsids. *Cell* **94**, 55–60
80. Roos, W. H., Gertsman, I., May, E. R., Brooks, C. L., 3rd, Johnson, J. E., and Wuite, G. J. (2012) Mechanics of bacteriophage maturation. *Proc. Natl. Acad. Sci. USA* **109**, 2342–2347
81. Grimes, S., and Anderson, D. (1997) The bacteriophage phi29 packaging proteins supercoil the DNA ends. *J. Mol. Biol.* **266**, 901–914
82. Guo, P. X., Erickson, S., and Anderson, D. (1987) A small viral RNA is required for in vitro packaging of bacteriophage phi 29 DNA. *Science* **236**, 690–694
83. Zhao, W., Jardine, P. J., and Grimes, S. (2015) An RNA domain imparts specificity and selectivity to a viral DNA packaging motor. *J. Virol.* **89**, 12457–12466
84. Wang, J. C., and Kaiser, A. D. (1973) Evidence that the cohesive ends of mature lambda DNA are generated by the gene A product. *Nat New Biol* **241**, 16–17
85. Emmons, S. W. (1974) Bacteriophage lambda derivatives carrying two copies of the cohesive end site. *J. Mol. Biol.* **83**, 511–525

86. Syvanen, M. (1975) Processing of bacteriophage lambda DNA during its assembly into heads. *J. Mol. Biol.* **91**, 165–174
87. Gold, M., Becker, A., and Parris, W. (1983) The bacteriophage lambda terminase enzyme. *Methods Enzymol.* **100**, 183–191
88. Parris, W., Davidson, A., Keeler, C. L., and Gold, M. (1988) The Nu1 subunit of bacteriophage lambda terminase. *J. Biol. Chem.* **263**, 8413–8419
89. Shinder, G., and Gold, M. (1988) The Nul subunit of bacteriophage lambda terminase binds to specific sites in cos DNA. *J. Virol.* **62**, 387–392
90. Feiss, M., Kobayashi, I., and Widner, W. (1983) Separate sites for binding and nicking of bacteriophage lambda DNA by terminase. *Proc. Natl. Acad. Sci. USA* **80**, 955–959
91. Shinder, G., and Gold, M. (1989) Integration host factor (IHF) stimulates binding of the gpNu1 subunit of lambda terminase to cos DNA. *Nucleic Acids Res.* **17**, 2005–2022
92. Ortega, M. E., and Catalano, C. E. (2006) Bacteriophage lambda gpNu1 and Escherichia coli IHF proteins cooperatively bind and bend viral DNA: implications for the assembly of a genome-packaging motor. *Biochemistry* **45**, 5180–5189
93. Yang, T. C., Ortiz, D., Yang, Q., De Angelis, R. W., Sanyal, S. J., and Catalano, C. E. (2017) Physical and functional characterization of a viral genome maturation complex. *Biophys. J.* **112**, 1551–1560
94. Maluf, N. K., Gaussier, H., Bogner, E., Feiss, M., and Catalano, C. E. (2006) Assembly of bacteriophage lambda terminase into a viral DNA maturation and packaging machine. *Biochemistry* **45**, 15259–15268
95. De Beer, T., Fang, J., Ortega, M., Yang, Q., Maes, L., Duffy, C., Berton, N., Sippy, J., Overduin, M., Feiss, M., and Catalano, C. E. (2002) Insights into specific DNA recognition during the assembly of a viral genome packaging machine. *Mol. Cell* **9**, 981–991

96. Feiss, M., Min, J. Y., Sultana, S., Patel, P., and Sippy, J. (2015) DNA packaging specificity of bacteriophage N15 with an excursion into the genetics of a cohesive end mismatch. *PLoS One* **10**, e0141934
97. Ortega, M. E., Gaussier, H., and Catalano, C. E. (2007) The DNA maturation domain of gpA, the DNA packaging motor protein of bacteriophage lambda, contains an ATPase site associated with endonuclease activity. *J. Mol. Biol.* **373**, 851–865
98. Jackson, E. N., Jackson, D. A., and Deans, R. J. (1978) EcoRI analysis of bacteriophage P22 DNA packaging. *J. Mol. Biol.* **118**, 365–388
99. Sternberg, N., and Coulby, J. (1987) Recognition and cleavage of the bacteriophage P1 packaging site (*pac*): I. Differential processing of the cleaved ends in vivo. *J. Mol. Biol.* **194**, 453–468
100. Casjens, S., and Huang, W. M. (1982) Initiation of sequential packaging of bacteriophage P22 DNA. *J. Mol. Biol.* **157**, 287–298
101. Deichelbohrer, I., Messer, W., and Trautner, T. A. (1982) Genome of *Bacillus subtilis* bacteriophage SPP1: Structure and nucleotide sequence of *pac*, the origin of DNA packaging. *J. Virol.* **42**, 83–90
102. Wu, H., Sampson, L., Parr, R., and Casjens, S. (2002) The DNA site utilized by bacteriophage P22 for initiation of DNA packaging. *Mol. Microbiol.* **45**, 1631–1646
103. Zhao, H., Finch, C. J., Sequeira, R. D., Johnson, B. A., Johnson, J. E., Casjens, S. R., and Tang, L. (2010) Crystal structure of the DNA-recognition component of the bacterial virus Sf6 genome-packaging machine. *Proc. Natl. Acad. Sci. USA* **107**, 1971–1976
104. Büttner, C. R., Chechik, M., Ortiz-Lombardia, M., Smits, C., Ebong, I. O., Chechik, V., Jeschke, G., Dykeman, E., Benini, S., Robinson, C. V., Alonso, J. C., and Antson, A. A. (2012) Structural basis for DNA recognition and loading into a viral packaging motor. *Proc. Natl. Acad. Sci. USA* **109**, 811–816

105. Roy, A., Bhardwaj, A., Datta, P., Lander, G. C., and Cingolani, G. (2012) Small terminase couples viral DNA binding to genome-packaging ATPase activity. *Structure* **20**, 1403–1413
106. Chai, S., Lurz, R., and Alonso, J. C. (1995) The small subunit of the terminase enzyme of *Bacillus subtilis* bacteriophage SPP1 forms a specialized nucleoprotein complex with the packaging initiation region. *J. Mol. Biol.* **252**, 386–398
107. Sun, S., Gao, S., Kondabagil, K., Xiang, Y., Rossmann, M. G., and Rao, V. B. (2012) Structure and function of the small terminase component of the DNA packaging machine in T4-like bacteriophages. *Proc. Natl. Acad. Sci. USA* **109**, 817–822
108. Zhao, H., Christensen, T. E., Kamau, Y. N., and Tang, L. (2013) Structures of the phage Sf6 large terminase provide new insights into DNA translocation and cleavage. *Proc. Natl. Acad. Sci. USA* **110**, 8075–8080
109. Xu, R. G., Jenkins, H. T., Chechik, M., Blagova, E. V., Lopatina, A., Klimuk, E., Minakhin, L., Severinov, K., Greive, S. J., and Antson, A. A. (2017) Viral genome packaging terminase cleaves DNA using the canonical RuvC-like two-metal catalysis mechanism. *Nucleic Acids Res.* **45**, 3580–3590
110. Guo, P., Grimes, S., and Anderson, D. (1986) A defined system for in vitro packaging of DNA-gp3 of the *Bacillus subtilis* bacteriophage phi 29. *Proc. Natl. Acad. Sci. USA* **83**, 3505–3509
111. Grimes, S., Jardine, P. J., and Anderson, D. (2002) Bacteriophage phi 29 DNA packaging. *Adv. Virus Res.* **58**, 255–294
112. Hwang, Y., and Feiss, M. (1995) A defined system for *in vitro* λ DNA packaging. *Virology* **211**, 367–376
113. Rao, V. B., and Black, L. W. (1988) Cloning, overexpression and purification of the terminase proteins gp16 and gp17 of bacteriophage T4. Construction of a defined in-vitro DNA packaging system using purified terminase proteins. *J. Mol. Biol.* **200**, 475–488

114. Smith, D. E., Tans, S. J., Smith, S. B., Grimes, S., Anderson, D. L., and Bustamante, C. (2001) The bacteriophage straight phi29 portal motor can package DNA against a large internal force. *Nature* **413**, 748–752
115. Fuller, D. N., Raymer, D. M., Rickgauer, J. P., Robertson, R. M., Catalano, C. E., Anderson, D. L., Grimes, S., and Smith, D. E. (2007) Measurements of single DNA molecule packaging dynamics in bacteriophage lambda reveal high forces, high motor processivity, and capsid transformations. *J. Mol. Biol.* **373**, 1113–1122
116. Fuller, D. N., Raymer, D. M., Kottadiel, V. I., Rao, V. B., and Smith, D. E. (2007) Single phage T4 DNA packaging motors exhibit large force generation, high velocity, and dynamic variability. *Proc. Natl. Acad. Sci. USA* **104**, 16868–16873
117. Pease, P. J., Levy, O., Cost, G. J., Gore, J., Ptacin, J. L., Sherratt, D., Bustamante, C., and Cozzarelli, N. R. (2005) Sequence-directed DNA translocation by purified FtsK. *Science* **307**, 586–590
118. Perkins, T. T., Li, H.-W., Dalal, R. V., Gelles, J., and Block, S. M. (2004) Forward and reverse motion of single RecBCD molecules on DNA. *Biophys. J.* **86**, 1640–1648
119. Yin, H., Wang, M. D., Svoboda, K., Landick, R., Block, S. M., and Gelles, J. (1995) Transcription against an applied force. *Science* **270**, 1653–1657
120. Wang, M. D., Schnitzer, M. J., Yin, H., Landick, R., Gelles, J., and Block, S. M. (1998) Force and velocity measured for single molecules of RNA polymerase. *Science* **282**, 902–907
121. Cappello, G., Pierobon, P., Symonds, C., Busoni, L., Christof, J., Gebhardt, M., Rief, M., and Prost, J. (2007) Myosin V stepping mechanism. *Proc. Natl. Acad. Sci. USA* **104**, 15328–15333
122. Crozat, E., Meglio, A., Allemand, J. F., Chivers, C. E., Howarth, M., Venien-Bryan, C., Grainge, I., and Sherratt, D. J. (2010) Separating speed and ability to displace roadblocks during DNA translocation by FtsK. *EMBO J.* **29**, 1423–1433

123. Morais, M. C., Koti, J. S., Bowman, V. D., Reyes-Aldrete, E., Anderson, D. L., and Rossmann, M. G. (2008) Defining molecular and domain boundaries in the bacteriophage phi29 DNA packaging motor. *Structure* **16**, 1267–1274
124. Mao, H., Saha, M., Reyes-Aldrete, E., Sherman, M. B., Woodson, M., Atz, R., Grimes, S., Jardine, P. J., and Morais, M. C. (2016) Structural and molecular basis for coordination in a viral DNA packaging motor. *Cell Rep.* **14**, 2017–2029
125. Sun, S., Kondabagil, K., Draper, B., Alam, T. I., Bowman, V. D., Zhang, Z., Hegde, S., Fokine, A., Rossmann, M. G., and Rao, V. B. (2008) The structure of the phage T4 DNA packaging motor suggests a mechanism dependent on electrostatic forces. *Cell* **135**, 1251–1262
126. Sun, S., Kondabagil, K., Gentz, P. M., Rossmann, M. G., and Rao, V. B. (2007) The structure of the ATPase that powers DNA packaging into bacteriophage T4 procapsids. *Mol. Cell* **25**, 943–949
127. Dixit, A., Ray, K., Lakowicz, J. R., and Black, L. W. (2011) Dynamics of the T4 bacteriophage DNA packasome motor: endonuclease VII resolvase release of arrested Y-DNA substrates. *J. Biol. Chem.* **286**, 18878–18889
128. Yeo, A., and Feiss, M. (1995) Specific interaction of terminase, the DNA packaging enzyme of bacteriophage lambda, with the portal protein of the prohead. *J. Mol. Biol.* **245**, 141–150
129. Hugel, T., Michaelis, J., Hetherington, C. L., Jardine, P. J., Grimes, S., Walter, J. M., Falk, W., Anderson, D. L., and Bustamante, C. (2007) Experimental test of connector rotation during DNA packaging into bacteriophage phi29 capsids. *PLoS Biol.* **5**, e59
130. Liu, S., Chistol, G., Hetherington, C. L., Tafoya, S., Aathavan, K., Schnitzbauer, J., Grimes, S., Jardine, P. J., and Bustamante, C. (2014) A viral packaging motor varies its DNA rotation and step size to preserve subunit coordination as the capsid fills. *Cell* **157**, 702–713

131. Chemla, Y. R., Aathavan, K., Michaelis, J., Grimes, S., Jardine, P. J., Anderson, D. L., and Bustamante, C. (2005) Mechanism of force generation of a viral DNA packaging motor. *Cell* **122**, 683–692
132. Moffitt, J. R., Chemla, Y. R., Aathavan, K., Grimes, S., Jardine, P. J., Anderson, D. L., and Bustamante, C. (2009) Intersubunit coordination in a homomeric ring ATPase. *Nature* **457**, 446–450
133. Chistol, G., Liu, S., Hetherington, C. L., Moffitt, J. R., Grimes, S., Jardine, P. J., and Bustamante, C. (2012) High degree of coordination and division of labor among subunits in a homomeric ring ATPase. *Cell* **151**, 1017–1028
134. Lin, S., Alam, T., Kottadiel, V., Van Gessel, C., Tang, W.-C., Chemla, Y., and Rao, V. (in press) Altering the speed of a DNA packaging motor from bacteriophage T4. *Nucleic Acids Res.*
135. Leipe, D. D., Koonin, E. V., and Aravind, L. (2003) Evolution and classification of P-loop kinases and related proteins. *J. Mol. Biol.* **333**, 781–815
136. Hilbert, B. J., Hayes, J. A., Stone, N. P., Duffy, C. M., Sankaran, B., and Kelch, B. A. (2015) Structure and mechanism of the ATPase that powers viral genome packaging. *Proc. Natl. Acad. Sci. USA* **112**, E3792–3799
137. Wiczorek, D. J., and Feiss, M. (2001) Defining *cosQ*, the site required for termination of bacteriophage lambda DNA packaging. *Genetics* **158**, 495–506
138. Cue, D., and Feiss, M. (2001) Bacteriophage lambda DNA packaging: DNA site requirements for termination and processivity. *J. Mol. Biol.* **311**, 233–240
139. Feiss, M., and Campbell, A. (1974) Duplication of the bacteriophage lambda cohesive end site: Genetic studies. *J. Mol. Biol.* **83**, 527–540
140. Tavares, P., Santos, M. A., Lurz, R., Morelli, G., De Lencastre, H., and Trautner, T. A. (1992) Identification of a gene in *Bacillus subtilis* bacteriophage SPP1 determining the amount of packaged DNA. *J. Mol. Biol.* **225**, 81–92
141. Casjens, S., Wyckoff, E., Hayden, M., Sampson, L., Eppler, K., Randall, S., Moreno, E. T., and Serwer, P. (1992) Bacteriophage P22 portal protein is part of

- the gauge that regulates packing density of intravirion DNA. *J. Mol. Biol.* **224**, 1055–1074
142. Padmanabhan, R., Wu, R., and Bode, V. C. (1972) Arrangement of DNA in lambda bacteriophage heads: III. Location and number of nucleotides cleaved from λ DNA by micrococcal nuclease attack on heads. *J. Mol. Biol.* **69**, 201–207
143. Lokareddy, R. K., Sankhala, R. S., Roy, A., Afonine, P. V., Motwani, T., Teschke, C. M., Parent, K. N., and Cingolani, G. (2017) Portal protein functions akin to a DNA-sensor that couples genome-packaging to icosahedral capsid maturation. *Nat. Commun.* **8**, 14310
144. Homa, F. L., and Brown, J. C. (1997) Capsid assembly and DNA packaging in herpes simplex virus. *Rev. Med. Virol.* **7**, 107–122
145. Johnson, D. C., Wisner, T. W., and Wright, C. C. (2011) Herpes simplex virus glycoproteins gB and gD function in a redundant fashion to promote secondary envelopment. *J. Virol.* **85**, 4910–4926
146. Abrescia, N. G., Cockburn, J. J., Grimes, J. M., Sutton, G. C., Diprose, J. M., Butcher, S. J., Fuller, S. D., San Martin, C., Burnett, R. M., Stuart, D. I., Bamford, D. H., and Bamford, J. K. (2004) Insights into assembly from structural analysis of bacteriophage PRD1. *Nature* **432**, 68–74
147. Hong, C., Oksanen, H. M., Liu, X., Jakana, J., Bamford, D. H., and Chiu, W. (2014) A structural model of the genome packaging process in a membrane-containing double stranded DNA virus. *PLoS Biol.* **12**, e1002024
148. Gillock, E. T., Rottinghaus, S., Chang, D., Cai, X., Smiley, S. A., An, K., and Consigli, R. A. (1997) Polyomavirus major capsid protein VP1 is capable of packaging cellular DNA when expressed in the baculovirus system. *J. Virol.* **71**, 2857–2865
149. Hurdiss, D. L., Morgan, E. L., Thompson, R. F., Prescott, E. L., Panou, M. M., Macdonald, A., and Ranson, N. A. (2016) New structural insights into the genome and minor capsid proteins of BK polyomavirus using cryo-electron microscopy. *Structure* **24**, 528–536

150. King, J. A., Dubielzig, R., Grimm, D., and Kleinschmidt, J. A. (2001) DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids. *EMBO J.* **20**, 3282–3291
151. Cotmore, S. F., and Tattersall, P. (2005) Encapsidation of minute virus of mice DNA: aspects of the translocation mechanism revealed by the structure of partially packaged genomes. *Virology* **336**, 100–112
152. Aoyama, A., and Hayashi, M. (1986) Synthesis of bacteriophage ϕ X174 in vitro: Mechanism of switch from DNA replication to DNA packaging. *Cell* **47**, 99–106
153. Dressler, D. H., and Denhardt, D. T. (1968) Mechanism of replication of phi-X-174 single stranded DNA. *Nature* **219**, 346–351
154. Ilang, L. L., Olson, N. H., Dokland, T., Music, C. L., Cheng, R. H., Bowen, Z., McKenna, R., Rossmann, M. G., Baker, T. S., and Incardona, N. L. (1995) DNA packaging intermediates of bacteriophage phi X174. *Structure* **3**, 353–363
155. Mindich, L. (2004) Packaging, replication and recombination of the segmented genome of bacteriophage Phi6 and its relatives. *Virus Res.* **101**, 83–92
156. McDonald, S. M., and Patton, J. T. (2011) Assortment and packaging of the segmented rotavirus genome. *Trends Microbiol.* **19**, 136–144
157. Jose, J., Taylor, A. B., and Kuhn, R. J. (2017) Spatial and temporal analysis of alphavirus replication and assembly in mammalian and mosquito cells. *mBio* **8**
158. Sundquist, W. I., and Krausslich, H. G. (2012) HIV-1 assembly, budding, and maturation. *Cold Spring Harb. Perspect. Med.* **2**, a006924
159. Stockley, P. G., Twarock, R., Bakker, S. E., Barker, A. M., Borodavka, A., Dykeman, E., Ford, R. J., Pearson, A. R., Phillips, S. E., Ranson, N. A., and Tuma, R. (2013) Packaging signals in single-stranded RNA viruses: nature's alternative to a purely electrostatic assembly mechanism. *J. Biol. Phys.* **39**, 277–287
160. Dhillon, E. K., Dhillon, T. S., Lai, A. N., and Linn, S. (1980) Host range, immunity and antigenic properties of lambdoid coliphage HK97. *J. Gen. Virol.* **50**, 217–220

161. Juhala, R. J., Ford, M. E., Duda, R. L., Youlton, A., Hatfull, G. F., and Hendrix, R. W. (2000) Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *J. Mol. Biol.* **299**, 27–51
162. Randall-Hazelbauer, L., and Schwartz, M. (1973) Isolation of the bacteriophage lambda receptor from *Escherichia coli*. *J. Bacteriol.* **116**, 1436–1446
163. Cumby, N., Edwards, A. M., Davidson, A. R., and Maxwell, K. L. (2012) The bacteriophage HK97 gp15 moron element encodes a novel superinfection exclusion protein. *J. Bacteriol.* **194**, 5012–5019
164. Kala, S., Cumby, N., Sadowski, P. D., Hyder, B. Z., Kanelis, V., Davidson, A. R., and Maxwell, K. L. (2014) HNH proteins are a widespread component of phage DNA packaging machines. *Proc. Natl. Acad. Sci. USA* **111**, 6022–6027
165. Yanofsky, C., and Ito, J. (1966) Nonsense codons and polarity in the tryptophan operon. *J. Mol. Biol.* **21**, 313–334
166. Boos, W., and Shuman, H. (1998) Maltose/maltodextrin system of *Escherichia coli*: transport, metabolism, and regulation. *Microbiol. Mol. Biol. Rev.* **62**, 204–229
167. Brown, A. (1956) A study of lysis in bacteriophage-infected *Escherichia coli*. *J. Bacteriol.* **71**, 482–490
168. Popa, M. P., McKelvey, T. A., Hempel, J., and Hendrix, R. W. (1991) Bacteriophage HK97 structure: wholesale covalent cross-linking between the major head shell subunits. *J. Virol.* **65**, 3227–3237
169. Dierkes, L. E., Peebles, C. L., Firek, B. A., Hendrix, R. W., and Duda, R. L. (2009) Mutational analysis of a conserved glutamic acid required for self-catalyzed cross-linking of bacteriophage HK97 capsids. *J. Virol.* **83**, 2088–2098
170. Slonczewski, J. L., Rosen, B. P., Alger, J. R., and Macnab, R. M. (1981) pH homeostasis in *Escherichia coli*: measurement by ³¹P nuclear magnetic resonance of methylphosphonate and phosphate. *Proc. Natl. Acad. Sci. USA* **78**, 6271–6275
171. Kaiser, D., and Masuda, T. (1973) In vitro assembly of bacteriophage lambda heads. *Proc. Natl. Acad. Sci. USA* **70**, 260–264

172. Griep, M. A., and McHenry, C. S. (1989) Glutamate overcomes the salt inhibition of DNA polymerase III holoenzyme. *J. Biol. Chem.* **264**, 11294–11301
173. Hofmeister, F. (1888) Zur lehre von der wirkung der salze: zweite mittheilung. *Naunyn Schmiedebergs Arch. Exp. Pathol. Pharmacol.* **24**, 247–260
174. Leirmo, S., Harrison, C., Cayley, D. S., Burgess, R. R., and Record, M. T., Jr. (1987) Replacement of potassium chloride by potassium glutamate dramatically enhances protein-DNA interactions in vitro. *Biochemistry* **26**, 2095–2101
175. Schultz, S. G., and Solomon, A. K. (1961) Cation transport in *Escherichia coli*. I. Intracellular Na and K concentrations and net cation movement. *J. Gen. Physiol.* **45**, 355–369
176. Measures, J. C. (1975) Role of amino acids in osmoregulation of non-halophilic bacteria. *Nature* **257**, 398–400
177. Ames, B. N., Dubin, D. T., and Rosenthal, S. M. (1958) Presence of polyamines in certain bacterial viruses. *Science* **127**, 814–815
178. Greive, S. J., Fung, H. K., Chechik, M., Jenkins, H. T., Weitzel, S. E., Aguiar, P. M., Brentnall, A. S., Glousieau, M., Gladyshev, G. V., Potts, J. R., and Antson, A. A. (2016) DNA recognition for virus assembly through multiple sequence-independent interactions with a helix-turn-helix motif. *Nucleic Acids Res.* **44**, 776–789
179. Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234
180. Cavaluzzi, M. J., and Borer, P. N. (2004) Revised UV extinction coefficients for nucleoside-5'-monophosphates and unpaired DNA and RNA. *Nucleic Acids Res.* **32**, e13
181. Kabsch, W. (2010) Xds. *Acta Cryst. D* **66**, 125–132
182. Sheldrick, G. M. (2008) A short history of SHELX. *Acta Cryst. A* **64**, 112–122

183. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674
184. Cowtan, K. (2010) Recent developments in classical density modification. *Acta Cryst. D* **66**, 470–478
185. Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Cryst. D* **62**, 1002–1011
186. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Cryst. D* **66**, 486–501
187. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst. D* **67**, 355–367
188. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **98**, 10037–10041
189. Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G., and Baker, N. A. (2007) PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **35**, W522–525
190. Swanson, J. M., Wagoner, J. A., Baker, N. A., and McCammon, J. A. (2007) Optimizing the poisson dielectric boundary with explicit solvent forces and energies: lessons learned with atom-centered dielectric functions. *J. Chem. Theory Comput.* **3**, 170–183
191. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797
192. McGhee, J. D., and von Hippel, P. H. (1974) Theoretical aspects of DNA-protein interactions: co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice. *J. Mol. Biol.* **86**, 469–489

193. Tsodikov, O. V., Holbrook, J. A., Shkel, I. A., and Record, M. T., Jr. (2001) Analytic binding isotherms describing competitive interactions of a protein ligand with specific and nonspecific sites on the same DNA oligomer. *Biophys. J.* **81**, 1960–1969
194. R Core Team. (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
195. Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995) A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.* **16**, 1190–1208
196. Grzesiek, S., and Bax, A. (1992) Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J. Magn. Reson.* **96**, 432–440
197. Clubb, R. T., Thanabal, V., and Wagner, G. (1992) A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue ¹HN, ¹⁵N, and ¹³C' chemical shifts in ¹⁵N, and ¹³C-labelled proteins. *J. Magn. Reson.* **97**, 213–217
198. Grzesiek, S., and Bax, A. (1992) An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J. Magn. Reson.* **99**, 201–207
199. Grzesiek, S., and Bax, A. (1992) Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* **114**, 6291–6293
200. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., and Clore, G. M. (1989) Overcoming the overlap problem in the assignment of proton NMR spectra of larger proteins by use of three-dimensional heteronuclear proton-nitrogen-15 Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 β . *Biochemistry* **28**, 6150–6156
201. Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., and Bax, A. (1989) Three-dimensional heteronuclear NMR of nitrogen-15 labeled proteins. *J. Am. Chem. Soc.* **111**, 1515–1517

202. Orekhov, V. Y., and Jaravine, V. A. (2011) Analysis of non-uniformly sampled spectra with multi-dimensional decomposition. *Prog. Nucl. Magn. Reson. Spectrosc.* **59**, 271–292
203. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293
204. Bodenhausen, G., and Ruben, D. J. (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* **69**, 185–189
205. Williamson, M. P. (2013) Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73**, 1–16
206. Rossmann, M. G., and Blow, D. M. (1962) The detection of sub-units within the crystallographic asymmetric unit. *Acta Cryst.* **15**, 24–31
207. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Cryst. D* **67**, 235–242
208. Dauter, Z., Dauter, M., and Rajashankar, K. R. (2000) Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. *Acta Cryst. D* **56**, 232–237
209. Cromer, D. T., and Liberman, D. (1970) Relativistic calculation of anomalous scattering factors for X rays. *J. Chem. Phys.* **53**, 1891–1898
210. Brennan, S., and Cowan, P. L. (1992) A suite of programs for calculating x-ray absorption, reflection, and diffraction performance for a variety of materials at arbitrary wavelengths. *Rev. Sci. Instrum.* **63**, 850–853
211. Steiner, T. (1998) Hydrogen-bond distances to halide ions in organic and organometallic crystal structures: up-to-date database study. *Acta Cryst. B* **54**, 456–463

212. Chou, P. Y., and Fasman, G. D. (1974) Conformational parameters for amino acids in helical, β -sheet, and random coil regions calculated from proteins. *Biochemistry* **13**, 211–222
213. Brandl, M., Weiss, M. S., Jabs, A., Suhnel, J., and Hilgenfeld, R. (2001) C-H... π -interactions in proteins. *J. Mol. Biol.* **307**, 357–377
214. Di Primo, C., and Lebars, I. (2007) Determination of refractive index increment ratios for protein-nucleic acid complexes by surface plasmon resonance. *Anal. Biochem.* **368**, 148–155
215. Wishart, D. S., and Sykes, B. D. (1994) The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J. Biomol. NMR* **4**, 171–180
216. Court, R., Chapman, L., Fairall, L., and Rhodes, D. (2005) How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: a view from high-resolution crystal structures. *EMBO Rep.* **6**, 39–45
217. Laguri, C., Phillips-Jones, M. K., and Williamson, M. P. (2003) Solution structure and DNA binding of the effector domain from the global regulator PrrA (RegA) from *Rhodobacter sphaeroides*: insights into DNA binding specificity. *Nucleic Acids Res.* **31**, 6778–6787
218. Hales, L. M., Gumpert, R. I., and Gardner, J. F. (1994) Determining the DNA sequence elements required for binding integration host factor to two different target sites. *J. Bacteriol.* **176**, 2999–3006
219. Dyer, P. N., Edayathumangalam, R. S., White, C. L., Bao, Y., Chakravarthy, S., Muthurajan, U. M., and Luger, K. (2004) Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol.* **375**, 23–44
220. Zheng, S. Q., Palovcak, E., Armache, J. P., Verba, K. A., Cheng, Y., and Agard, D. A. (2017) MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332

221. Rohou, A., and Grigorieff, N. (2015) CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221
222. Scheres, S. H. (2015) Semi-automated selection of cryo-EM particles in RELION-1.3. *J. Struct. Biol.* **189**, 114–122
223. Conway, J. F., Trus, B. L., Booy, F. P., Newcomb, W. W., Brown, J. C., and Steven, A. C. (1993) The effects of radiation damage on the structure of frozen hydrated HSV-1 capsids. *J. Struct. Biol.* **111**, 222–233
224. Yan, X., Sinkovits, R. S., and Baker, T. S. (2007) AUTO3DEM — an automated and high throughput program for image reconstruction of icosahedral particles. *J. Struct. Biol.* **157**, 73–82
225. Scheres, S. H. (2012) RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530
226. Edelstein, A. D., Tsuchida, M. A., Amodaj, N., Pinkard, H., Vale, R. D., and Stuurman, N. (2014) Advanced methods of microscope control using µManager software. *J. Biol. Methods* **1**, e10
227. Haralick, R. M., and Shapiro, L. G. (1992) Computer and Robot Vision. Addison-Wesley, Boston. pp 28–48
228. McGuire, H., Aourousseau, M. R., Bowie, D., and Blunck, R. (2012) Automating single subunit counting of membrane proteins in mammalian cells. *J. Biol. Chem.* **287**, 35912–35921
229. Chung, S. H., and Kennedy, R. A. (1991) Forward-backward non-linear filtering technique for extracting small biological signals from noise. *J. Neurosci. Methods* **40**, 71–86
230. Kerssemakers, J. W., Munteanu, E. L., Laan, L., Noetzel, T. L., Janson, M. E., and Dogterom, M. (2006) Assembly dynamics of microtubules at molecular resolution. *Nature* **442**, 709–712
231. Gelles, J., Schnapp, B. J., and Sheetz, M. P. (1988) Tracking kinesin-driven movements with nanometre-scale precision. *Nature* **331**, 450–453

232. Chen, S., McMullan, G., Faruqi, A. R., Murshudov, G. N., Short, J. M., Scheres, S. H., and Henderson, R. (2013) High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35
233. Veesler, D., Quispe, J., Grigorieff, N., Potter, C. S., Carragher, B., and Johnson, J. E. (2012) Maturation in action: CryoEM study of a viral capsid caught during expansion. *Structure* **20**, 1384–1390
234. Henderson, R., Sali, A., Baker, M. L., Carragher, B., Devkota, B., Downing, K. H., Egelman, E. H., Feng, Z., Frank, J., Grigorieff, N., Jiang, W., Ludtke, S. J., Medalia, O., Penczek, P. A., Rosenthal, P. B., Rossmann, M. G., Schmid, M. F., Schroder, G. F., Steven, A. C., Stokes, D. L., Westbrook, J. D., Wriggers, W., Yang, H., Young, J., Berman, H. M., Chiu, W., Kleywegt, G. J., and Lawson, C. L. (2012) Outcome of the first electron microscopy validation task force meeting. *Structure* **20**, 205–214
235. Song, L., Hennink, E. J., Young, I. T., and Tanke, H. J. (1995) Photobleaching kinetics of fluorescein in quantitative fluorescence microscopy. *Biophys. J.* **68**, 2588–2600
236. Carter, N. J., and Cross, R. A. (2005) Mechanics of the kinesin step. *Nature* **435**, 308–312
237. Pierce, D. W., Hom-Booher, N., and Vale, R. D. (1997) Imaging individual green fluorescent proteins. *Nature* **388**, 338
238. Ulbrich, M. H., and Isacoff, E. Y. (2007) Subunit counting in membrane-bound proteins. *Nat. Methods* **4**, 319–321
239. Stark, H. (2010) GraFix: stabilization of fragile macromolecular complexes for single particle cryo-EM. *Methods Enzymol.* **481**, 109–126
240. Sastry, S. S. (1997) Isolation and partial characterization of a novel psoralen-tyrosine photoconjugate from a photoreaction of psoralen with a natural protein. *Photochem. Photobiol.* **65**, 937–944

241. Liu, H., and Naismith, J. H. (2008) An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* **8**, 91
242. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L., and Clardy, J. (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* **229**, 105–124
243. Laue, T. M., Shah, B. D., Ridgeway, T. M., and Pelletier, S. L. (1992) Computer-aided interpretation of analytical sedimentation data for proteins. in *Analytical ultracentrifugation in biochemistry and polymer science* (Harding, S. E., Rowe, A. J., and Horton, J. C. eds.), The Royal Society of Chemistry, Cambridge, UK. pp 90–125
244. Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78**, 1606–1619
245. Patel, A., Malinowska, L., Saha, S., Wang, J., Alberti, S., Krishnan, Y., and Hyman, A. A. (2017) ATP as a biological hydrotrope. *Science* **356**, 753–756
246. Evans, G., and Pettifer, R. F. (2001) CHOOCH: a program for deriving anomalous-scattering factors from X-ray fluorescence spectra. *J. Appl. Crystallogr.* **34**, 82–86
247. Smits, C., Chechik, M., Kovalevskiy, O. V., Shevtsov, M. B., Foster, A. W., Alonso, J. C., and Antson, A. A. (2009) Structural basis for the nuclease activity of a bacteriophage large terminase. *EMBO Rep.* **10**, 592–598
248. Yaginuma, H., Kawai, S., Tabata, K. V., Tomiyama, K., Kakizuka, A., Komatsuzaki, T., Noji, H., and Imamura, H. (2014) Diversity in ATP concentrations in a single bacterial cell population revealed by quantitative single-cell imaging. *Sci. Rep.* **4**, 6522
249. Lee, S., Sowa, M. E., Watanabe, Y. H., Sigler, P. B., Chiu, W., Yoshida, M., and Tsai, F. T. (2003) The structure of ClpB: a molecular chaperone that rescues proteins from an aggregated state. *Cell* **115**, 229–240

250. Duffy, C., and Feiss, M. (2002) The large subunit of bacteriophage λ 's terminase plays a role in DNA translocation and packaging termination. *J. Mol. Biol.* **316**, 547–561
251. Hang, J. Q., Tack, B. F., and Feiss, M. (2000) ATPase center of bacteriophage λ terminase involved in post-cleavage stages of DNA packaging: identification of ATP-interactive amino acids¹¹Edited by M. Gottesman. *J. Mol. Biol.* **302**, 777–795
252. Kim, Y. C., Snoberger, A., Schupp, J., and Smith, D. M. (2015) ATP binding to neighbouring subunits and intersubunit allosteric coupling underlie proteasomal ATPase function. *Nat. Commun.* **6**, 8520
253. Toseland, C. P., Powell, B., and Webb, M. R. (2012) ATPase cycle and DNA unwinding kinetics of RecG helicase. *PLoS One* **7**, e38270
254. Minor, D. L., Jr., and Kim, P. S. (1994) Measurement of the beta-sheet-forming propensities of amino acids. *Nature* **367**, 660–663
255. Cross, S. J., Brown, C. E., and Baumann, C. G. (2016) Transverse magnetic tweezers allowing coincident epifluorescence microscopy on horizontally extended DNA. *Methods Mol. Biol.* **1431**, 73–90
256. Baumann, C. G., and Cross, S. J. (2011) Probing the mechanics of the complete DNA transcription cycle in real-time using optical tweezers. *Methods Mol. Biol.* **778**, 175–191
257. Zhao, J., Benlekbir, S., and Rubinstein, J. L. (2015) Electron cryomicroscopy observation of rotational states in a eukaryotic V-ATPase. *Nature* **521**, 241–245
258. Loveland, A. B., Demo, G., Grigorieff, N., and Korostelev, A. A. (2017) Ensemble cryo-EM elucidates the mechanism of translation fidelity. *Nature* **546**, 113–117
259. Chen, B., Kaledhonkar, S., Sun, M., Shen, B., Lu, Z., Barnard, D., Lu, T.-M., Gonzalez, Ruben L., and Frank, J. (2015) Structural dynamics of ribosome subunit association studied by mixing-spraying time-resolved cryogenic electron microscopy. *Structure* **23**, 1097–1105

260. Jain, T., Sheehan, P., Crum, J., Carragher, B., and Potter, C. S. (2012) Spotiton: a prototype for an integrated inkjet dispense and vitrification system for cryo-TEM. *J. Struct. Biol.* **179**, 68–75
261. Frank, J., and Ourmazd, A. (2016) Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-EM. *Methods* **100**, 61–67
262. Rupp, B. (2010) *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*, Garland Science, New York
263. Karle, J. (1980) Some developments in anomalous dispersion for the structural investigation of macromolecular systems in biology. *Int. J. Quantum Chem* **18**, 357–367
264. Hendrickson, W. A. (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51–58
265. Blow, D. M., and Rossmann, M. G. (1961) The single isomorphous replacement method. *Acta Cryst.* **14**, 1195–1202
266. McCoy, A. J., Storoni, L. C., and Read, R. J. (2004) Simple algorithm for a maximum-likelihood SAD function. *Acta Cryst. D* **60**, 1220–1228
267. McCoy, A. J., and Read, R. J. (2010) Experimental phasing: best practice and pitfalls. *Acta Cryst. D* **66**, 458–469
268. McCoy, A. J. (2004) Liking likelihood. *Acta Cryst. D* **60**, 2169–2183
269. Sim, G. (1959) The distribution of phase angles for structures containing heavy atoms. II. A modification of the normal heavy-atom method for non-centrosymmetrical structures. *Acta Cryst.* **12**, 813–815
270. Wang, J., Wlodawer, A., and Dauter, Z. (2007) What happens when the signs of anomalous differences or the handedness of substructure are inverted? *Acta Cryst. D* **63**, 751–758
271. Wang, B. C. (1985) Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol.* **115**, 90–112

272. Leslie, A. (1987) A reciprocal-space method for calculating a molecular envelope using the algorithm of B.C. Wang. *Acta Cryst. A* **43**, 134–136
273. Zhang, K. Y. J., and Main, P. (1990) Histogram matching as a new density modification technique for phase refinement and extension of protein molecules. *Acta Cryst. A* **46**, 41–46
274. Abrahams, J. P. (1997) Bias reduction in phase refinement by modified interference functions: introducing the gamma correction. *Acta Cryst. D* **53**, 371–376
275. Cowtan, K. (1999) Error estimation and bias correction in phase-improvement calculations. *Acta Cryst. D* **55**, 1555–1567
276. Sheldrick, G. M. (2002) Macromolecular phasing with SHELXE. *Z. Kristallogr. Cryst. Mater.* **217**, 644
277. Sheldrick, G. M. (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Cryst. D* **66**, 479–485
278. Brenner, S., and Horne, R. W. (1959) A negative staining method for high resolution electron microscopy of viruses. *Biochim. Biophys. Acta* **34**, 103–110
279. Adrian, M., Dubochet, J., Lepault, J., and McDowell, A. W. (1984) Cryo-electron microscopy of viruses. *Nature* **308**, 32–36
280. Kuhlbrandt, W. (2014) The resolution revolution. *Science* **343**, 1443–1444
281. Danev, R., Buijsse, B., Khoshouei, M., Plitzko, J. M., and Baumeister, W. (2014) Volta potential phase plate for in-focus phase contrast transmission electron microscopy. *Proc. Natl. Acad. Sci. USA* **111**, 15635–15640
282. Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I., and Ludtke, S. J. (2007) EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46
283. Wade, R. H. (1992) A brief look at imaging and contrast transfer. *Ultramicroscopy* **46**, 145–156

284. Zhu, J., Penczek, P. A., Schröder, R., and Frank, J. (1997) Three-dimensional reconstruction with contrast transfer function correction from energy-filtered cryoelectron micrographs: Procedure and application to the 70S *Escherichia coli* ribosome. *J. Struct. Biol.* **118**, 197–219
285. Frank, J. (1973) The envelope of electron microscope transfer functions for partially coherent illumination. *Optik* **38**, 519–536
286. Wade, R. H., and Frank, J. (1977) Electron microscope transfer functions for partially coherent axial illumination and chromatic defocus spread. *Optik* **49**, 81–92
287. Kenney, J. M., Hantula, J., Fuller, S. D., Mindich, L., Ojala, P. M., and Bamford, D. H. (1992) Bacteriophage $\phi 6$ envelope elucidated by chemical cross-linking, immunodetection, and cryoelectron microscopy. *Virology* **190**, 635–644
288. McMullan, G., Faruqi, A. R., Clare, D., and Henderson, R. (2014) Comparison of optimal performance at 300 keV of three direct electron detectors for use in low dose electron microscopy. *Ultramicroscopy* **147**, 156–163
289. Mindell, J. A., and Grigorieff, N. (2003) Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347
290. van Heel, M., Gowen, B., Matadeen, R., Orlova, E. V., Finn, R., Pape, T., Cohen, D., Stark, H., Schmidt, R., Schatz, M., and Patwardhan, A. (2000) Single-particle electron cryo-microscopy: towards atomic resolution. *Q. Rev. Biophys.* **33**, 307–369
291. Penczek, P. A. (2010) Image restoration in cryo-electron microscopy. *Methods Enzymol.* **482**, 35–72
292. Crowther, R. A., DeRosier, D. J., and Klug, A. (1970) The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. *Proc. Royal Soc. A* **317**, 319–340
293. Sigworth, F. J. (1998) A maximum-likelihood approach to single-particle image refinement. *J. Struct. Biol.* **122**, 328–339

294. Rosenthal, P. B., and Henderson, R. (2003) Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745
295. Karplus, P. A., and Diederichs, K. (2012) Linking crystallographic model and data quality. *Science* **336**, 1030–1033
296. Martin-Fernandez, M. L., Tynan, C. J., and Webb, S. E. (2013) A 'pocket guide' to total internal reflection fluorescence. *J. Microsc.* **252**, 16–22
297. Richards, B., and Wolf, E. (1959) Electromagnetic diffraction in optical systems. II. Structure of the image field in an aplanatic system. *Proc. Royal Soc. A* **253**, 358–379