Structural and functional studies on two ATP-driven nucleic acid machines from viruses

Mengyu Ge

Doctor of Philosophy

University of York

Chemistry

September 2019

Abstract

ATP-driven motors exist globally in almost all prokaryotic and eukaryotic organisms. The motors in viruses are essential for viral replication and assembly. In this thesis, two ATP-driven nucleic acid motors from viruses were studied: a DNA packaging ATPase present in *Thermus thermophilus* phage phiKo and an NS3 helicase from Zika virus.

The production and purification of the DNA packaging ATPase (gp14) from the phage phiKo was investigated. The purified gp14 protein was crystallised for determination of the native structure as well as the complexes with several non-hydrolysable ATP analogues, representing a precise picture of the ATP hydrolysis cycle. Structural analysis reveals that gp14 possesses an FtsK-like fold and likely translocates DNA similarly to other FtsK/HerA superfamily translocases. Moreover, an additional catalytic sensor residue from the adjacent subunit is required for ATP hydrolysis and protein assembly. The appropriate assembly of gp14 may be achieved by co-assembling with an auxiliary protein component into a higher-order hetero-oligomer complex. Further characterisations, including thermal unfolding, ATPase activity and DNA binding assays elucidate that gp14 is thermodynamically stable, but its ATPase activity decreases over time.

The purified Zika virus NS3 helicase (NS3h) was crystallised in its apo form and as complexes with several nucleotides, including pre-hydrolysis substrate, transition state, and post-hydrolysis product analogues, as well as in a complex with a single-stranded RNA (ssRNA) segment containing a 5’ phosphate. The structural observations illustrate how a catalytically important motif mediates the coupling of ATP hydrolysis with RNA translocation. Due to the transition state analogue ADP-$MgF_3(H_2O)^-$ not being observed before, this complex structure was subjected to careful examination by crystallography and ^{19}F NMR. Further functional characterisations, including protein stability, RNA binding and ATPase activity assays, demonstrate this NS3 helicase possesses weak RNA-binding mode and its ATPase activity can be stimulated by ssRNA.

List of contents

List of tables

List of figures

Acknowledgements

First of all, I want to express sincere gratitude to Professor Fred Antson for his supervision and inspiration over the four years of my PhD study. It's been a privilege to work under his expert guidance. Fred was always very supportive and generously shared his ideas and provided invaluable advice during my PhD research, in particular, crystallographic experiments, such as crystal handling and X-ray data collection. Moreover, I greatly appreciate that he carefully read my thesis and offered helpful advice, comments and suggestions for improvements.

I also want to thank the Antson group members for their help towards my PhD research project. I would like to give special thanks to Maria Chechik for helpful advice on cloning and protein purification as well as training on related chromatography equipment. I'm especially grateful to Dr. Sandra Grieve for the encouraging discussions on biochemical assays and NS3 helicase manuscript preparation with helpful advice and comments on scientific writing, as well as kind help with proofreading the entire thesis. I would like to thank Dr. Huw Jenkins for advice and practical demonstrations on X-ray data processing, crystal handling and structure determination, Dr. Evgeny Klimuk and Dr. Oliver Bayfield for discussions on phage isolation and purification, Dr. Simon Tanley for assistance for setting up thermal shift assays and Dr. Ben Bax for conversations and comments on the NS3 helicase manuscript.

I would also like to thank my thesis advisory panel member, Dr. Daniela Barilla, for her inspiration and advice during my TAP meetings. Also, I would like to thank Dr. Yi Jin for kindly sharing her expertise and for assistance with NMR experiments and helpful discussions on production and analysis of the transition state analogue complex of the NS3 helicase, and to thank professor Michael Blackburn for valuable advice on examination of the transition state analogue structure and his contribution in preparing the NS3 helicase manuscript. In addition, I want to thank Paul Bond for kindly helping with proofreading the crystallographic section of my thesis with helpful comments and suggestions. I would also like to thank Dr. Saba Dehghani-Tafti and Dr. Cyril Sanders for discussions on RNA binding activity of the NS3 helicase. I'd like to thank Dr. Johan Turkenburg and Sam Hart for help with X-ray data collection and thank YSBL and Biology L0 lab members who helped me over the past four years of my PhD study.

Besides, I would like to thank the Chinese Scholarship Council (CSC) scholarship and Wild Fund studentship for financial support for my PhD study at York.

Finally, I would like to express deep gratitude to my family for all of their support.

Declaration

I declare that this thesis is a presentation of my original work. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as references and cited in the text accordingly.

Any contribution from others is clearly acknowledged in the text and as follows. Maria Chechik in the Antson group cloned the Zika NS3 helicase recombinant constructs, pYM436 and pYM535, and constructs pYM351and pYM350 of phiKo gp14. Owen Jarman obtained the initial crystals for phiKo gp14-phosphate and phiKo gp14-ADP complexes during his summer research project under the supervision from Maria Chechik in the Antson group. Dr. Huw Jenkins in the Antson group solved the initial model of the native phiKo gp14 and refined the Zika virus NS3h-pRNA$_9$ complex structure.

In addition, sections 3.3.3.3, 3.3.4 and 3.3.5 in this thesis have been prepared, re-composed and combined with QM data from collaborators into a manuscript for submitting to a scientific journal. Sections 3.3.3.1, 3.3.3.2 and 3.3.6 have been prepared and re-composed with data from collaborators into a manuscript, being submitted to a scientific journal.

1 Introduction

1.1 Viruses

Generally, a complete virus particle, known as a virion, consists of a nucleic acid genome enclosed by a protein shell, named a capsid. Viruses have either RNA or DNA as their genetic material. The nucleic acid may be single- or double-stranded. Viruses can encode various numbers (from 4 to ~ 200) of proteins and can infect susceptible cells and direct the cell machinery to produce more viruses but cannot self-reproduce. The host cell range that a virus can infect is usually restricted and serves as a basis for classifying viruses. Viruses that infect only bacteria are called bacteriophages, or simply phages, while viruses which can infect animal or plant cells are generally called eukaryotic viruses (1).

1.1.1 Bacteriophages

Bacteriophages exist widely in the ecosphere, and the total numbers may exceed 10^{30} virions (2, 3). The majority of phages are composed of protein and nucleic acid with a head-tail morphology. Polyhedral, filamentous or pleomorphic phages comprise less than 4% of the bacteriophage population including both number and species, and a minority of these, including *Tectiviridae* family phages, have a lipid component in addition to nucleic acid and proteins (4). Notably, there are two energy-dependent genome packaging strategies recruited by these phages, i.e. the portal-terminase and FtsK/HerA type packaging systems, as described in section 1.2.1 below. The tailed dsDNA bacteriophages and their evolutionary related eukaryotic viruses, such as Herpesviruses, package their genomes similarly using the portal-terminase type genome packaging system (5). The lipid containing *Tectiviridae* phage utilises the FtsK/HerA type packaging system for packaging genome into the pre-formed procapsids. More importantly, the same FtsK/HerA type packaging system is recruited by a larger group of eukaryotic dsDNA viruses, nucleocytoplasmic large DNA viruses (NCLDVs), including giant viruses and *Poxviruses* (6–8). The evolutionary similarities indicate the importance of the study of these FtsK/HerA type packaging system containing phages.

Generally, DNA bacteriophages can be classified into three groups in terms of their genome type, size and virion morphology (9–11). The first group comprises tailed double-stranded (ds) DNA phages containing a larger genome ranging from ~ 16 kb to ~ 500 kb; the group includes the *Podoviridae*, *Siphoviridae* and *Myoviridae* families (Figure 1-1). The second group contain ssDNA phages possessing smaller genome ranging from approximately 3.5 kb to 10 kb; it includes the *Leviviridae*, *Microviridae* and *Inoviridae* families (Figure 1-1). The third group is comprised by bacteriophages that have no protein tail attached to the mature virion and contain a middle-sized dsDNA genome (ranging from ~ 10 kb to 16 kb); this group includes

Corticoviridae, Plasmaviridae and *Tectiviridae* families (Figure 1-1). Phages belonging to these families are lipid-containing, and the lipids can be either forming the outer layer of the virion or be located internally, enclosed within the virus capsids (4). The phage capsids usually have icosahedral symmetry with the exception of filamentous (*Inoviridae* family) or pleomorphic (*Plasmaviridae* family) phages (Figure 1-1) (2).

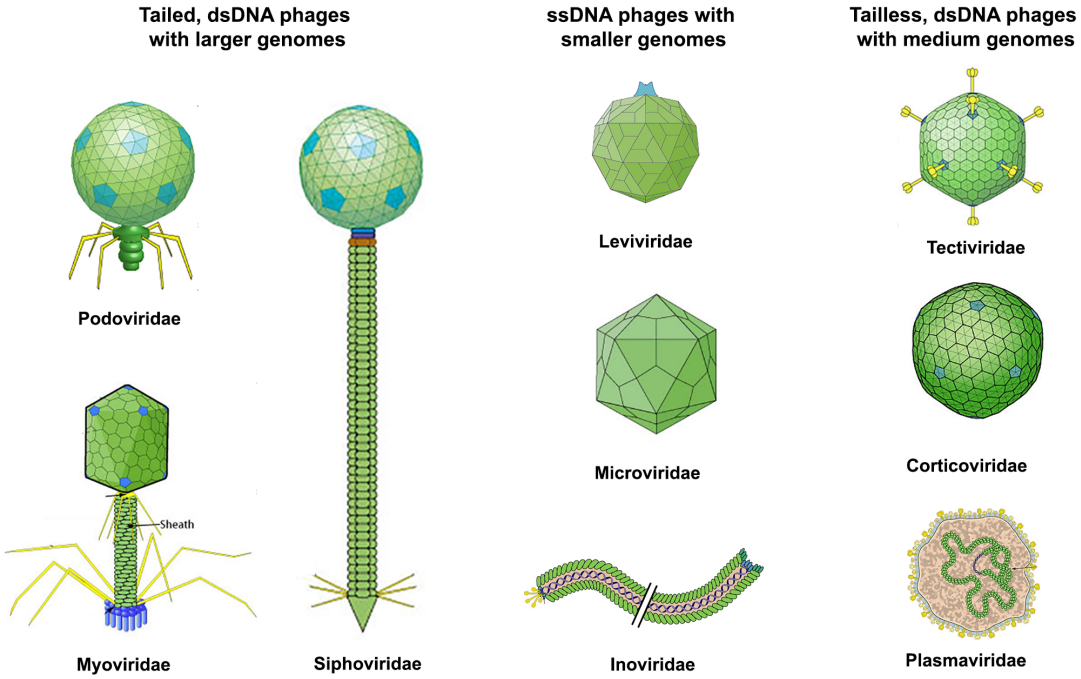


Figure 1-1 Several DNA bacteriophage families.

Figure panels were reproduced and collated, with permission, from ViralZone (12) (www.expasy.org/viralzone, SIB Swiss Institute of Bioinformatics).

Tectiviridae family

Tectiviridae (from Latin 'tectus', meaning covered) family bacteriophages feature a linear unit-length dsDNA genome and a protein-associated internal lipid-membrane, which is enclosed within an icosahedral protein capsid. The *Enterobacteria* phage PRD1 is a well-studied member of the *Tectiviridae* family (13). The major coat protein of phage PRD1 forms an icosahedral shell with a pseudo T = 25 lattice (14). The major coat protein monomer is folded into two eight-stranded β-barrels (the jelly-roll fold), with the loops between the β-strands facing the virion exterior. The minor coat protein dimers stretch out from one vertex to another cementing the major coat protein facets together (15). The majority of the vertices are occupied by flexible spikes composed of the receptor-binding protein, the spike protein and the penton protein (16). The spike complex is anchored to the membrane via transmembrane proteins (4). The virions are normally tailless, where the viral membrane is

transformed into a tubular structure used for genome delivery upon infection (17). One vertex of the mature virion differs from others and is used for the phage DNA packaging during virion assembly. This packaging vertex contains the packaging ATPase and three other proteins, including two membrane-associated proteins (9, 17). The packaging ATPase in *Tectiviridae* phage PRD1 is a structural protein and remains associated with the mature virion (18). This packaging ATPase and other packaging associated protein form a dodecameric portal complex external to the inner membrane, surrounded by ten major capsid protein trimers (14).

1.1.2 Single-stranded RNA viruses

Single-stranded RNA (ssRNA) viruses are viruses that utilise ssRNA as their genetic material. In virology, a positive-sense (5' to 3') viral RNA signifies that a particular viral RNA sequence may be directly translated into the coded proteins. Eight eukaryotic virus families are known to possess single-stranded, positive-sense RNA genomes. These include *Picornaviridae*, *Astroviridae*, *Caliciviridae* and *Hepeviridae* families containing non-enveloped capsids, as well as *Flaviviridae*, *Togaviridae*, *Arteriviridae* and *Coronaviridae* families containing enveloped capsids. Viruses from these eight families use their own genomes as messenger RNAs (mRNAs), from which they synthesise one or several polyproteins that are subsequently cleaved into individual proteins by viral or cellular proteases. The RNA replication of these viruses is catalysed by an RNA-dependent RNA polymerase (RdRP). This RdRP transcribes the positive RNA strand as well as the complementary negative RNA strand which arises as an intermediate product of genome replication. The initial transcription product of the RdRP is a double-stranded complex of the negative strand intermediate with the infecting positive-strand RNA. During this process, the new genomic RNA molecules are generated from the second transcription step (19).

Flaviviridae family

Flaviviridae (from the Latin 'flavus', meaning yellow) family viruses are positive-sense, single-stranded, enveloped eukaryotic RNA viruses with genomes that comprise approximately 9 - 13 kb genomic RNAs. The virions are spherical with diameters of 40 - 60 nm and contain an external lipid envelope (Figure 1-2). Viral proteins are produced as part of a single long polyprotein that is cleaved by a combination of host and viral proteases. The structural proteins are encoded within the N-terminal portion of the polyprotein with the non-structural (NS) proteins present in the remainder of the polypeptide, as described in below section 1.2.3.2.

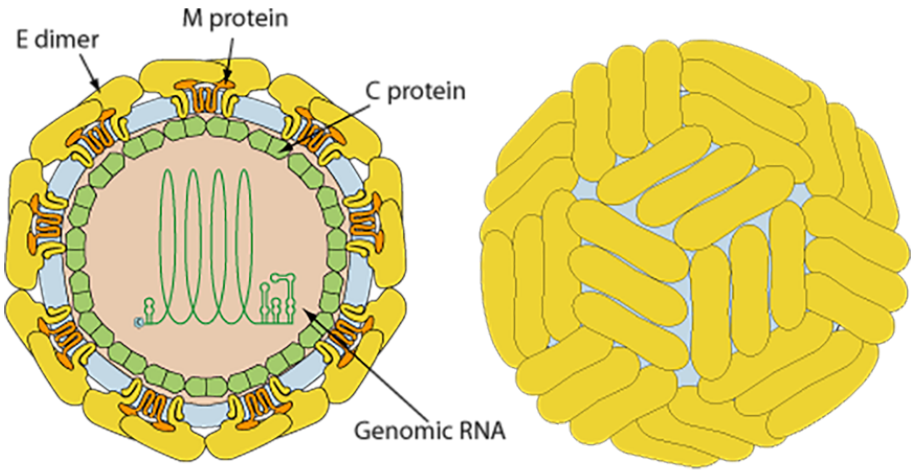


Figure 1-2 Virion of *Flaviviridae* family member Zika virus.

The external lipid envelope is shown as the blue layer. This figure was adapted with permission from ViralZone (12) (www.expasy.org/viralzone, SIB Swiss Institute of Bioinformatics).

The *Flaviviridae* family is comprised of four genera: *Flavivirus* (with 53 species); *Hepacivirus* (with 14 species); *Pegivirus* (with 11 species) and *Pestivirus* (with 4 species) (20), all having somewhat different characteristics. Virion of this family member has a single, small basic capsid (C) and two (in *Flavivirus, Hepacivirus* and *Pegivirus*) or three (in *Pestivirus*) membrane-associated envelope proteins. The *Flavivirus* genome consists of a 5' type-1 RNA cap structure (5' CAP1, me^7-GpppA-me^2) (21, 22), whereas the *Hepacivirus* lacks the 5' CAP1 structure (21, 22). In addition, the *Flavivirus* has seven non-structural proteins, NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5, while the *Hepacivirus* polyprotein comprises six non-structural proteins, NS2, NS3, NS4A, NS4B, NS5A and NS5B (22). Nevertheless, the non-structural proteins contain characteristic sequence motifs of the serine protease, RNA helicase and RNA-dependent RNA polymerase (RdRP) that are encoded at similar locations in the genomes of all genera (20, 23). Although helicase activities have been demonstrated or predicted for proteins of numerous positive-strand RNA viruses, their precise role in RNA replication remains unclear.

Several members of the *Flavivirus* and *Hepacivirus* genera affect human health, making a substantial impact on human population. *Flaviviruses* cause acute and self-limiting human infections, for instance, Zika virus has recently spread worldwide, and its infections have been linked to Guillain-Barré syndrome in adults and multiple neurodevelopmental defects in infants (24). Hepatitis C virus (HCV), the prototypic *Hepacivirus*, causes both acute and chronic infections and is the primary human pathogen causing chronic liver disease worldwide (25).

1.2 Virus replication and assembly

Viral replication is the formation of new virions during the infection process in the host cells. A virion is an entire virus particle consisting of an outer protein shell called a capsid and an inner core of nucleic acid, as mentioned in section 1.1. A virus cannot replicate its genome and synthesise viral proteins without the machinery and metabolism of a host cell. Viral replication is highly varied and depends on the type of genes encoded by the viruses. In general, viral replication can be subdivided into several steps: absorption and entry, viral genome replication, viral protein synthesis, as well as virion assembly and release. Virus assembly is a vital process during which the newly synthesised genomes (nucleic acids) and proteins are assembled to form new virus particles. For most enveloped viruses, virus assembly takes place either in the nucleus, cytoplasm or at the plasma membrane of the host cell.

1.2.1 Viral genome packaging systems

Viral genome packaging systems can be tentatively divided into three large groups based on the strategy employed and the relationship of the components: (i) energy-independent packaging system (condensation), where viruses perform genome packaging via protein-protein or protein-nucleic acid interactions in an energy-independent way. Human immunodeficiency virus (HIV) and *Flavivirus* are examples of this type of packaging system; (ii) portal-terminase system, in which virion possesses a unique vertex with a ring-shaped portal protein and a terminase that packages genome into the preformed capsid in an ATP-dependent manner. This type of packaging system is found in icosahedral dsDNA and dsRNA bacteriophages and Herpesviruses; (iii) FtsK/HerA-type genome packaging system (26–28). Viruses belonging to the last group package dsDNA using a conserved FtsK/HerA-like packaging ATPase, the motor assembly that is used by nucleocytoplasmic large DNA viruses (NCLDVs) and dsDNA bacteriophages belonging to the *Tectiviridae* family (6–8), as mentioned in section 1.1.1. The DNA packaging ATPases share the conserved FtsK/HerA fold (6, 8), suggesting that they may translocate dsDNA in a similar manner to other FtsK/HerA superfamily DNA translocases.

1.2.2 Phage replication and assembly

The *Tectiviridae* family bacteriophages can be classified into two major genera, according to the types of their host cells. The first group of phages, including PRD1, infect gram-negative bacteria, and the second group of phages, such as Bam35, infect gram-positive bacteria (29). Phage phiKo, studied in this thesis, infects gram-negative bacterium, *Thermus thermophilus* HB27 (30). Therefore, the replication and assembly of phages that infect gram-negative bacteria are discussed only in this thesis.

For the tailless *Tectiviridae* bacteriophage that infects gram-negative bacterium, the host cell is selected by specific recognition of a receptor on the host cell surface. In *Tectivirus* PRD1, the DNA delivery protein P11 binds to the host outer membrane and is responsible for the outer membrane (OM) penetration (31). The lytic trans-glycosylase protein P7 is an integral viral membrane protein and probably responsible for peptidoglycan digestion (31). Then, several unidentified viral DNA delivery proteins are transferred onto the cytoplasmic membrane (CM) to facilitate penetration (31). After absorption, the phage genome is injected into the cell cytosol, leaving the capsid outside (4, 13). Viral genomes and protein components are then synthesised in the cytoplasm of the host cells, and the viral membrane associated proteins are incorporated into the host cell membrane to form the viral inner membrane portion (14).

After the formation of each viral component, phage assembly begins. In general, there are three major steps in dsDNA phage assembly: (i) synthesis of the protein components and procapsid formation, (ii) viral genome encapsidation (genome packaging) and (iii) virion maturation.

1.2.2.1 Procapsid formation

In tailless *Tectiviridae* family viruses, such as PRD1, the synthesis of protein components takes place in parallel with DNA replication. The newly synthesised viral membrane proteins are transported to the cytoplasmic membrane of the host cell and form a transmembrane pore in a newly assembled viral membrane (14). The correct folding of the major capsid proteins and the formation of the procapsid are facilitated by the host GroEL-GroES chaperonin and viral scaffolding proteins (14, 18). Integral membrane-associated proteins assemble to form a transmembrane channel on the inner membrane (Figure 1-3a). The capsid proteins then assemble along the viral membrane and establish the unique packaging vertex to complete the procapsid formation (Figure 1-3b).

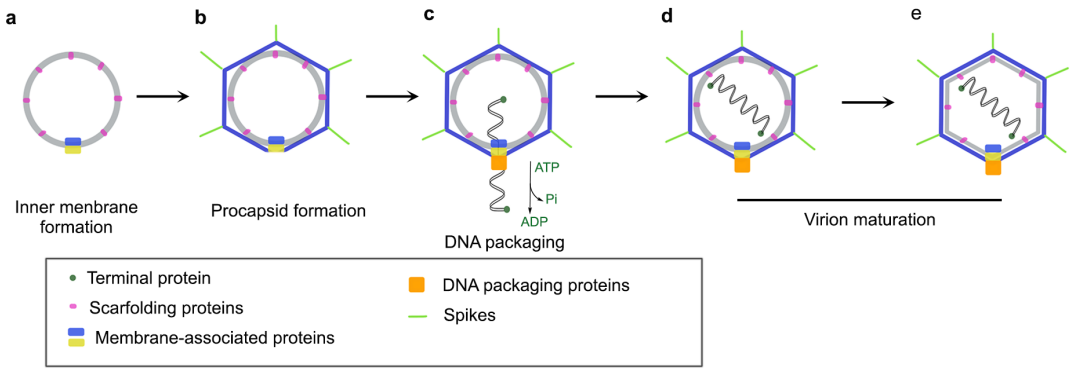


Figure 1-3 Schematic diagram of the tailless *Tectiviridae* family phage assembly.

a. Viral inner membrane formation; b. Procapsid formation; c. Genome packaging; d and e. Virion maturation. The viral capsid protein shell is shown as the blue icosahedral shell, the membrane is in grey. The key protein components are indicated in the figure.

1.2.2.2 DNA replication

Tectiviridae family phages replicate their genomes using a protein-primed DNA replication mechanism, similar to other viruses with linear unit-length dsDNA genomes, including adenovirus and the phi29-type phages (32–34). *Tectiviridae* genomes contain around 100 bp inverted terminal repeats (ITR) covalently linked to 5' terminal proteins that are used as primers during replication (13, 32). In these viruses, the genome replication starts from both ends of the terminal protein-containing DNA (TP-DNA) by forming a chemical bond between the first nucleotide to a hydroxyl group of a serine, threonine or tyrosine residue of the TP and progresses asymmetrically from both ends. Processive genome replication can be carried out *in vitro*, often only with two proteins, the DNA polymerase and the TP, although additional factors, such as single- and double-stranded DNA binding protein P12 in phage PRD1 that binds to both single- and double-stranded DNA (35), strongly stimulate the replication and are essential *in vivo* (33, 35–37).

In several other tailed phages, in parallel with virion assembly, DNA replication and recombination generate concatemeric viral DNA (covalently linked end-to-end polymers of virus chromosomes), using DNA primed DNA replication mechanism. The genomic DNA is then recognised by a terminase small subunit, which initiates its packaging into procapsid (26).

1.2.2.3 Genome packaging and virion maturation

The crucial step in virus assembly is genome packaging. This is usually accomplished by biological molecular motors that convert the chemical energy from ATP hydrolysis into conformational changes that drive DNA translocation into the procapsid. During DNA packaging of *Tectiviridae* PRD1, the DNA packaging ATPase P9 (6, 38) and the packaging efficiency factor P6 (39) form a 12-fold portal complex (14). No analogue of a portal protein, which is essential for DNA packaging in tailed bacteriophages, has been identified in phage PRD1 (14, 40, 41). Once the DNA packaging proteins P6 and P9 together become assembled at the right position within the unique vertex, DNA packaging can begin. The newly synthesised viral genomic DNA with terminal proteins fused at both 5' ends is packaged through the unique vertex fuelled by the hydrolysis of ATP (Figure 1-3c). When genome packaging is completed, the packaging ATPase remains on the mature virion (Figure 1-3d). The terminal protein may play a role as a protein valve to seal the pore in the vertex (14). After

genome packaging, the inner membrane is pushed against the icosahedral protein shell as the last step in virus maturation (Figure 1-3e) (14).

1.2.3 *Flavivirus* replication and assembly

Flaviviruses are positive-sense single-stranded RNA viruses, as described above in section 1.1.2. *Flavivirus* replication includes four major steps: viral particles absorption and entry, viral protein components synthesis, viral genome replication, as well as genome packaging and virion maturation. The replication and assembly of *Flaviviruses* are described in the sections 1.2.3.1 - 1.2.3.4 below.

1.2.3.1 Absorption and entry

Flaviviruses bind to the surface of host cells and first interact with receptors or attachment factors that are required to tether the virion onto the cell surface (22). Receptors or attachment factors identified in relevant target cells include the C-type lectin CD209 antigen (42, 43), the mannose receptor (44) as well as members of the TIM (T cell immunoglobulin mucin domain protein 1) and TAM [tyrosine protein kinase receptor 3 (TYRO3)-AXL-MER] family of phosphatidylserine receptors (45, 46). *Flavivirus* particles predominantly enter via a clathrin-dependent entry pathway, exposing viral particles to an acidic endosomal compartment that triggers conformational rearrangements of the envelope glycoproteins to allow fusion of viral and endosomal membranes, resulting in the release of the viral RNA into the cytoplasm (22, 47).

1.2.3.2 Viral protein synthesis and processing

After genomic RNA entry, the newly released positive-strand (+) RNA is recognised by ribosomes, initiating translation at the endoplasmic reticulum (ER) membrane, producing a single polyprotein. The viral genome mimics the cellular mRNA molecule in all aspects except for the absence of the poly-adenylated (poly-A) tail. This feature allows the *Flavivirus* to exploit cellular apparatus to synthesise both structural and non-structural proteins during replication (48). Viral and cellular proteases catalyse the co-translational and post-translational cleavage of the polyprotein into three structural and seven non-structural proteins (Figure 1-4), almost all of which are associated with intracellular membranes (22). The structural proteins, capsid protein (C), the major envelope protein (E) and membrane protein (prM), are constituents of the virion and envelope (49–51).

The non-structural proteins of *Flaviviruses* are essential for viral replication. NS1 is a glycoprotein peripherally associated with membranes, which forms homodimers apparently functional for virus replication (52). The N-terminal domain of NS3 forms the viral serine

protease complex along with NS2B that is involved in processing the polyprotein (53). The C-terminal portion of NS3 is a helicase that is essential for melting the secondary structure in the genomic RNA, to initiate negative-strand replication and for subsequently separating the two strands to facilitate the synthesis of the nascent positive ssRNA strands (54). NS3 helicase also possesses an RNA 5' triphosphatase (RTPase) activity that is required for viral assembly (21, 22, 54, 55), as described in section 1.2.3.3. NS4A and NS4B proteins are essential components of the ER membrane-associated replication complex (56). NS5 acts as the viral RNA-dependent RNA polymerase (RdRp) and also possesses methyltransferase activity involved in the modification of the viral cap structure (57).

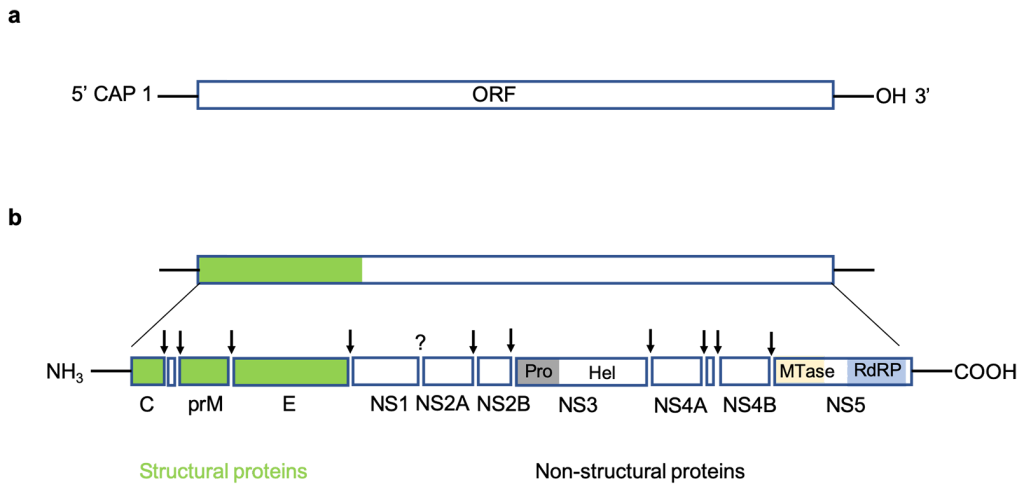


Figure 1-4 Diagram of *Flavivirus* genome and polyprotein organisation.

a. *Flavivirus* genome structure; b. *Flavivirus* polyprotein organisation. Structural proteins are in green and non-structural proteins are in white; the activity of certain proteins is highlighted on the figure. 'Pro' indicates protease, 'Hel' means helicase, 'MTase' designates methyltransferase and 'RdRP' indicates RNA-dependent RNA polymerase. Vertical down narrows indicate the polyprotein cleavage by cellular signal peptidases, the viral protease or by the furin protease. The '?' indicates the polyprotein cleavage performed by unknown protease.

1.2.3.3 Genome replication

In parallel with viral protein synthesis, *Flaviviruses* replicate (+) RNA genome in the cytoplasm, as mentioned in section 1.1.2. The viral (+) RNA is recognised and replicated into the uncapped negative-strand RNA [(-) RNA] intermediate by the RNA dependent RNA polymerase NS5 (Figure 1-5) on the endoplasmic reticulum (ER) membrane in the host cells (21, 22, 57–59). NS3 helicase is potentially involved in RNA replication by unwinding the newly synthesised duplex RNA during transcription (Figure 1-5). Within the transcribed

double-strand (ds) RNA, the (-) RNA acts as a template to produce a nascent (+) RNA (21, 59, 60). After transcription, the newly synthesised (+) RNA is modified at its 5' end with a type-1 RNA cap structure (CAP 1, me^7-GpppA-me^2) by NS3 helicase and NS5 methyltransferase (Figure 1-5) (21, 22). The 5' CAP 1 structure of cellular RNA is formed by a process involving RNA triphosphatase, guanyl-transferase, N^7-methyltransferase and 2′-O methyltransferase. NS3 helicase has RNA triphosphatase activity and uses the ATP binding site to remove the terminal phosphate from the 5' end of the RNA (21, 22, 54, 55). The NS5 methyltransferase has N^7-methyltransferase and guanyl-transferase activities, as well as the 2'-O methyltransferase activities, which are necessary for forming mature RNA CAP 1 structures (21). After 5' capping, the newly transcribed mature (+) RNA can be either packaged into immature viral particles or used for transcription or translation processes.

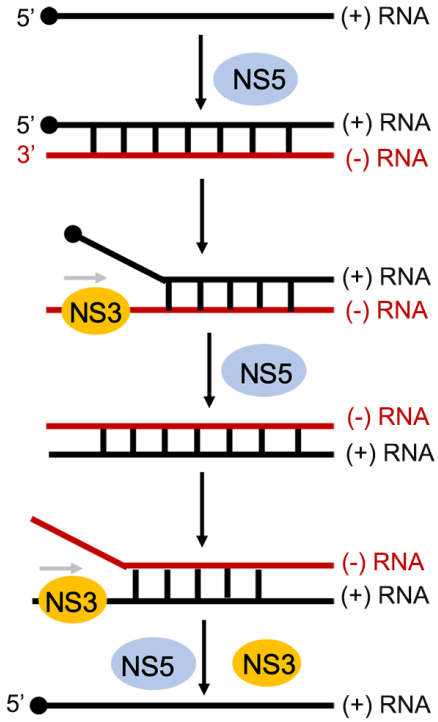


Figure 1-5 *Flavivirus* RNA replication.

NS3 and NS5 proteins are indicated on the figure, positive-strand (+) RNA is in black and the complementary negative-strand (-) RNA is in dark red. The 5' CAP 1 structures are shown as black dots.

1.2.3.4 Genome packaging and virion maturation

The *Flaviviruses* employ an energy-independent packaging system for genome encapsidation, as mentioned in section 1.2.1. The newly synthesised (+) RNA molecules can be incorporated into viral particles by a process involving RNA packaging and budding into the lumen of the

endoplasmic reticulum (ER) (61). Maturation of *Flavivirus* particles containing precursor membrane protein (prM) occurs along the secretory pathway by furin-mediated cleavage of prM (62, 63). The prM is cleaved by furin protease only after exposure to low pH, indicating that a conformational change is necessary before virus maturation can occur (64). *Flavivirus* particles are presumed to exit the cell through the conventional secretory pathway, although the details of how this process occurs and is involved in virus replication are still unclear (22).

1.3 ASCE group ATPases

In this thesis, the DNA packaging ATPase present in a *Tectiviridae* family phage phiKo, and the NS3 helicase from the Zika virus, both belonging to the 'Additional Strand Catalytic glutamate (E)' (ASCE) group ATPases, were studied.

The ASCE group is one of the largest and most functionally diverse groups of P loop NTPases (65). The ASCE division includes the AAA^+ (ATPases Associated with various cellular Activities), ABC (ATP-binding cassette), PilT, RecA and FtsK/HerA superfamily ATPases, superfamilies 1 and 2 (SF1 and SF2) helicases, as well as several additional, less confidently classified ATPases (7, 66). The ASCE ATPases are characterised by an additional strand in the central β-sheet located between the Walker A and Walker B motifs (65, 67, 68), as seen in section 1.3.1. The Walker A motif, or phosphate-binding loop, consists of a stretch of hydrophobic residues followed by the conserved sequence GxxxxGKS/T, where x is any amino acid. Walker B motif consists of an aspartate preceded by a stretch of hydrophobic amino acids (68, 69). ATP hydrolysis by the ASCE division ATPases typically depends on conserved catalytic glutamate polarising a water molecule for nucleophilic attack on the γ-phosphate of ATP substrate (Figure 1-6) (6–8). Generally, the conserved Walker A motif is responsible for nucleotide binding, and the conserved aspartate in the Walker B motif and the serine or threonine in the Walker A coordinate the catalytic metal (generally is Mg^{2+}) for ATP hydrolysis (Figure 1-6) (6, 68, 70, 71). The 'arginine finger' facilitates the formation of the transition state during ATP hydrolysis by insertion into the ATPase catalytic site of a neighbouring subunit or subdomain, in response to a conformational change in the ATPase catalytic site that follows substrate binding (72).

ASCE group ATPases are sometimes referred to as 'RecA-like' proteins, due to the similarity of their fold to the recombination factor RecA (73, 74). However, there are critical differences between the 'RecA-like' proteins and other ASCE family proteins, for example, the core nucleotide-binding folds of RecA-like and AAA^+ proteins are related, but differ in the details of their topology, conserved ATP binding residues, and the orientation of individual motor domains in higher-order quaternary states (6, 75, 76).

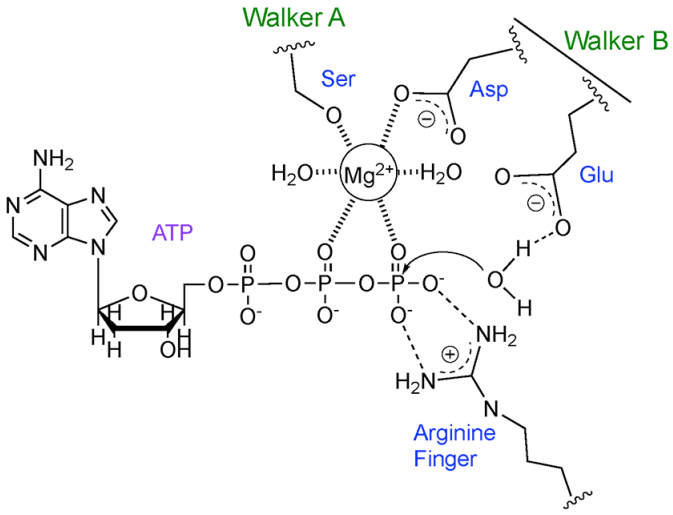


Figure 1-6 Schematic representation of the ATP hydrolysis mechanism.
Key residues of the Walker A and Walker B motifs and the 'arginine finger' are indicated on the figure. The H-bonds are displayed as dashed lines and metal-coordinated interactions are shown as hashed lines.

1.3.1 RecA-like core

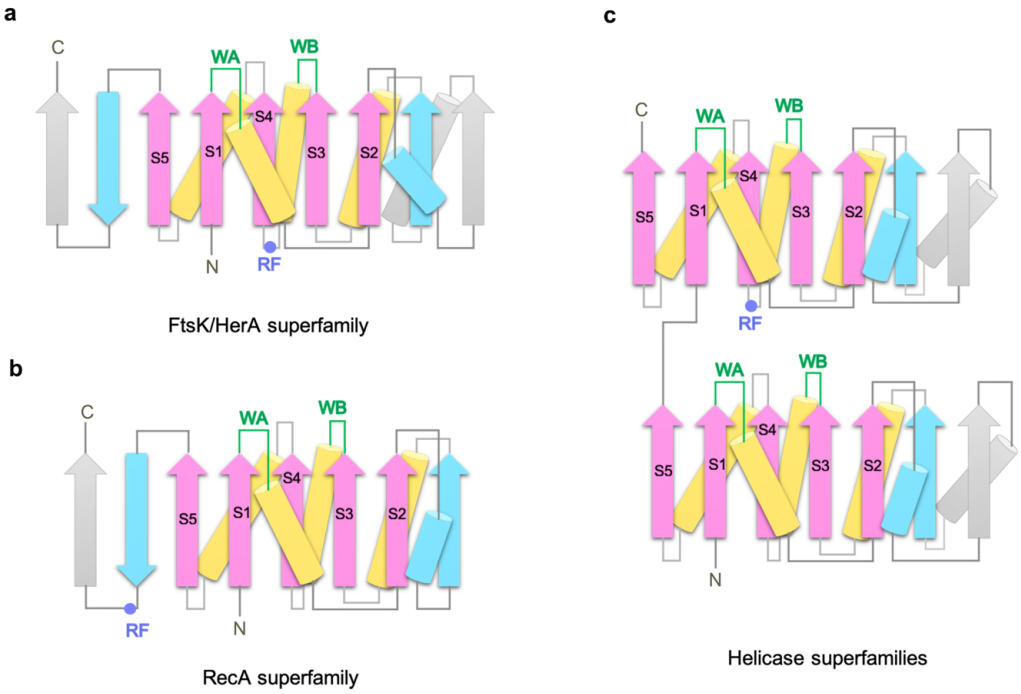


Figure 1-7 Topology diagram of the proteins of FtsK/HerA, RecA and helicase surperfamilies.

The ASCE core (RecA-like core) β-strands are labelled in the order of 51432. The structurally conserved central β-strands are in pink and α-helices in dark yellow. The helices and strands, that are slightly less conserved among ASCE group ATPases than the central β-sheet region, are in blue, and structural elements that are not conserved among ASCE group ATPases are in grey.

The fold of the RecA-like core or ASCE core is characterised by central stranded β-sheets flanked by α-helices (74) and an additional strand (S4) located between the Walker A and Walker B motifs (Figure 1-7). At the primary sequence level, the RecA-like motors belong to different classes of P-loop NTPases, including the FtsK/HerA, RecA, AAA^+ and ABC families, along with SF1, SF2 and SF3 helicases (66, 77). The structure conservation of the RecA-like ATPases indicates that they likely hydrolyse ATP using a similar mechanism.

Different RecA-like ATPases contain different numbers of RecA-like cores (73). Two RecA-like cores in a single polypeptide chain are found in many SF1 and SF2 helicases, such as PcrA (78) and NS3 helicases (79), as well as ABC transporters (80). In ABC transporters, two ATP molecules are bound at the interface (80), while other RecA-like core containing ATPase active site has one ATP molecule bound between two subunits or subdomains. Five RecA-like cores, only one subset of which bind and hydrolyse ATP, are found in the clamp loaders, the bacterial homologue of eukaryotic replication factor C that loads the sliding clamp (homologous to proliferating cell nuclear antigen) onto DNA (81). Six RecA-like cores forming a hexamer are found in many proteins, for instance, the FtsK translocase (72), the hydrophilic portion of ATP synthase F1 ATPase (82), E1 DNA helicase (83) and T7 gp4 DNA helicase (84).

1.3.2 FtsK/HerA superfamily

FtsK/HerA superfamily ATPases comprise the conserved ASCE fold, as discussed in sections 1.3.1 above. The common feature of the FtsK/HerA superfamily ATPases is a single functional domain with a central nine stranded β-sheet, including one antiparallel β-strand, that is flanked by six α-helices (6, 85), as seen in Figure 1-7a. The FtsK/HerA superfamily has been subdivided into six major clades based on the sequence homology: HerA, VirB4, VirD4/TrwB, FtsK, ssDNA phage packaging ATPases and A32-like viral dsDNA packaging ATPase (6). All proteins in the six clades contain a unique set of conserved residues found only in the FtsK and HerA homologues and not in any other previously characterised group of ATPases (6). The packaging ATPases of a variety of DNA viruses, including the P9 protein of bacteriophage PRD1 and homologues from nucleocytoplasmic large DNA viruses (NCLDVs) typified by the Vaccinia virus A32 protein, contain a conserved FtsK-like domain with ASCE fold (6),

indicating that they likely translocate dsDNA in a similar way to other members of the FtsK/HerA superfamily.

1.3.3 FtsK translocase

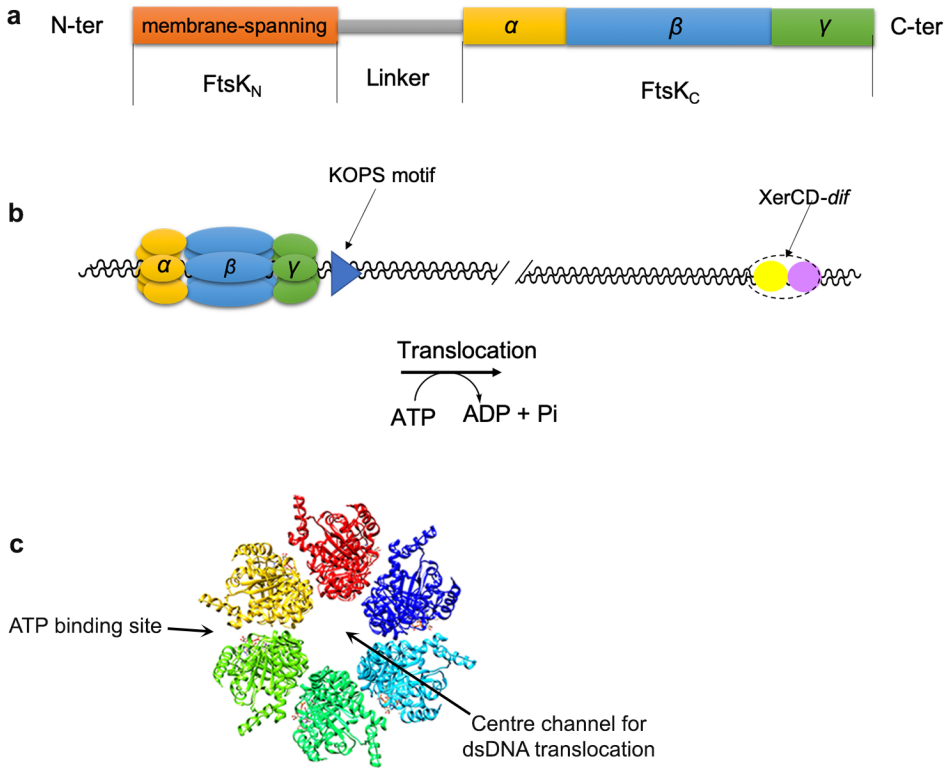


Figure 1-8 FtsK protein composition and motor domain translocation.
a. Organisation of the *E. coli* FtsK protein; b. Diagram of the translocation activation of the FtsK$_C$ hexamer. The KOPS motif is highlighted as a blue triangle. XerC and XerD are in yellow and purple, respectively. c. FtsK ASCE-fold domain hexamer assembly [PDB code 2IUU, (72)].

'Fts' is the abbreviation for Filamentous Temperature-Sensitive which describes a series of proteins, including FtsK, that emerged in 1995 from a screen for temperature-sensitive mutants of *Escherichia coli* with a filamentous phenotype (86). FtsK is a highly conserved dsDNA translocase, which assembles at the division septum in a wide range of bacteria during segregating of sister chromosomes of cell division (87–90). FtsK consists of three domains: an essential N-terminal membrane-spanning domain (FtsK$_N$) that anchors it to the division septum, a C-terminal motor domain (FtsK$_C$), and a middle linker domain (Figure 1-8a) (88). FtsK$_C$ is composed of *α*-, *β*-, and γ-subdomains (Figure 1-8a) (72). The *α*- and *β*- subdomains together form a hexamer motor domain; the *α*-subdomain is unique for FtsK while the *β*-

subdomain contains an ASCE fold that forms the ATP-binding pocket for ATP hydrolysis (72). The γ-subdomain plays a regulatory role in the recognition of the FtsK orienting the polar sequence (KOPS) that directs the orientation of FtsK translocation on DNA (Figure 1-8b) (87, 91) and activates XerCD-*dif* recombination (Figure 1-8b) (87, 90). The motor domain (*α*- and *β*-subdomains) forms a hexamer ring and translocates dsDNA through the centre channel using energy derived from ATP hydrolysis (Figure 1-8c) (72).

1.3.3.1 Proposed mechanism for FtsK

A rotary inchworm model for dsDNA translocation by the hexameric FtsK motor domain has been proposed (72). The FtsK hexamer ring accommodates dsDNA in its centre channel and moves DNA around the hexamer, with protein-DNA interactions involving one or two of the six subunits. One subunit of FtsK undergoes an ATP hydrolysis cycle initiating the DNA translocation. This translocation, with little rotation of DNA relative to protein, brings the DNA backbone into register with the DNA binding loops of the next adjacent FtsK subunit around in the hexamer while preventing further DNA binding by the first subunit. The second subunit undergoes an ATP hydrolysis cycle, and the DNA is then bound by the next subunit around to enable DNA translocation. Thus, the helical structure of DNA (B-form) allows a cycle of subunits of FtsK binding around the hexamer to pump DNA through the centre channel (72).

1.3.4 RNA helicases

RNA helicases can enzymatically unwind double-stranded RNA (dsRNA) by disrupting the hydrogen bonds that maintain two strands together using energy derived from NTP hydrolysis. The RNA helicases are present in almost all biological organisms, ranging from microbes to humans (92). Most eukaryotic RNA helicases are incapable of unwinding long double-stranded regions *in vitro*, as opposed to the highly processive activity of some of the DNA helicases. Indeed, a continuous dsRNA is rare in biological systems. However, a possible exception is the viral RNA helicases, where an RNA helicase may need to resolve dsRNA intermediates formed during transcription or replication of the viral RNA (93), as described in section 1.2.3.3 above. In this case, the viral RNA helicase may act more like a temporary clamp to prevent RNAs from re-associating, therefore, allowing other RNA-RNA or RNA-protein interactions to occur (92).

All of the RNA helicases contain the conserved Walker A and B motifs (also named as motifs I and II in helicases) (68, 94, 95). The structural conservation implies equal functional conservation. In general, RNA helicases comprise two conserved RecA-like domains (Domain 1 and Domain 2) tethered by a short flexible linker and a C-terminal extension which usually

forms an additional domain (Figure 1-7c), as mentioned in section 1.3.1 above. However, the C-terminal additional domain is structurally conserved at least in some cases, but its length and sequence vary wildly among RNA helicases (92). Some of these helicases, such as the eukaryotic initiation factor eIF4A and the DEAD-box RNA helicase Vasa, do not contain the third C-terminal domain (96, 97).

1.3.5 Superfamily 2 (SF2) helicases

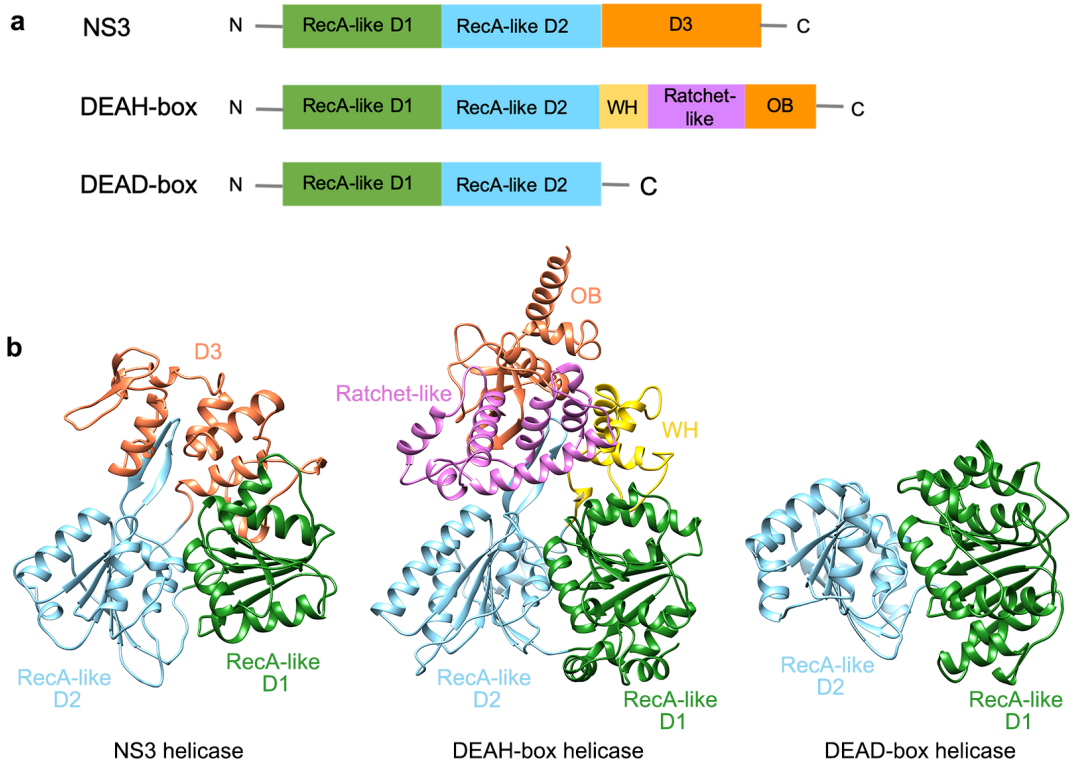


Figure 1-9 Comparisons of NS3, DEAH-box and DEAD-box helicases.

a. Diagram of protein organisation of NS3, DEAH-box and DEAD-box helicases. b. Structural comparisons of NS3, DEAH-box and DEAD-box helicases. NS3 helicase [PDB code 5MFX, this work], DEAH-box helicase [Prp43, PDB code 5LTA, (107)] and DEAD-box helicase [Vasa, PDB code 2DB3, (97)]. 'D' indicates domain, 'WH' stands for winged helix subdomain, and 'OB' indicates oligonucleotide/oligosaccharide binding subdomain.

Helicases can be classified into six superfamilies (SF1 - SF6), based on their primary sequences, structures, and mechanisms (77). Superfamily 2 (SF2) is the largest and most diverse superfamily among all DNA and RNA helicases. In addition, SF2 helicases are involved in transcription, DNA repair, and chromatin rearrangement and almost all aspects of RNA metabolism (98, 99). Generally, SF2 helicases appear to have nucleic acid stimulated

ATPase activity (100, 101). SF2 helicases comprise a conserved helicase core consisting of two similar protein domains that resemble the RecA-like fold (95, 102), as mentioned in section 1.3.1 above. It has been further divided into families of RecQ-like, RecG-like, Rad3/XPD, Ski2-like, type I restriction enzyme, RIG-I-like, NS3/NPH-II, DEAH/RHA, DEAD-box, and Swi/Snf families based on sequence homology (95, 101). The DEAD-box helicases are named after the strictly conserved sequence aspartate (D) - glutamate (E) - alanine (A) - aspartate (D) (103, 104). Similarly, DEAH-box helicases are named after the conserved motif of aspartate (D) - glutamate (E) - alanine (A) - histidine (H) (105). Notably, NS3 also contains the conserved specific signature (DExH) in the helicase motif II (95). However, the C-terminal domains differ significantly between DEAH-box and NS3 helicases (Figure 1-9a). The C-terminal domain 3 of NS3 helicase possesses a short helical-rich conformation, in contrast, the DEAH-box helicase has a more extended C-terminal domain comprising a winged-helix, ratchet and an oligonucleotide/oligosaccharide binding subdomains (Figure 1-9b) (95, 106–108).

1.3.6 Proposed mechanisms for NS3 helicase

In this thesis, the NS3 helicase (NS3h) from Zika virus, belonging to SF2 helicases, was studied. NS3h can enzymatically unwind duplexed RNA using the energy derived from the ATP hydrolysis, as mentioned in section 1.3.4. The NS3 helicase (NS3h) has three domains: RecA-like domains 1 and 2, as well as domain 3. There is an ATP-binding cleft between the two domains 1 and 2, and an RNA binding groove among the three domains (as discussed in sections 1.3.4 and 1.3.5). HCV NS3h can unwind both RNA and DNA (109–111). To describe a generic model of strand separation, the substrate will be referred to below as nucleic acid (NA). Previous studies of HCV NS3 helicase resulted in several proposed unwinding and translocation mechanisms that are briefly discussed in this section.

The spring-loaded mechanism

Domains 1 and 2 of NS3h move along the tracking strand (3’ to 5’) one nucleotide at a time per ATP hydrolysis cycle, where ATP binding and ADP release induce closure and opening of the domains 1 and 2, respectively. The domain 3 of NS3h lags behind by anchoring itself to the NA strand until three such steps have taken place. Then, the spring-loaded domain 3 moves forward in a burst motion, unwinding 3 base pairs (bp) as a consequence. NS3h continues unwinding in 3 bp steps up to about 18 bp unless it encounters an apparent barrier, suggesting that NS3h is not highly efficient in longer sequence unwinding (> 18bp). Upon encountering a barrier or after unwinding > 18bp, NS3h snaps or slips backwards rapidly and repeats this unwinding event many times in succession. This relatively low unwinding

processivity may help to keep secondary structures resolved during viral genome replication (112).

Inchworm type mechanism

The NS3h monomer has two RNA-binding sites surrounding an ATPase site for which ATP binding coordinates RNA translocation and unwinding in the fashion of an inchworm, similar to the structural homologues, SF1 PcrA (78) and Rep helicases (113). The inchworm model proposes that the two domains of the helicase (domains 1A and 2A in PcrA or domains 1 and 2 in NS3h) switch alternatively between tight and weak NA-binding states and move relative to each other as the helicase binds and hydrolyses ATP. In the initial stage, the single-stranded nucleic acid (ssNA) is bound at both domains 1 and 2. Binding of ATP to the ATPase active site of NS3h induces a cleft closure of domains 1 and 2. In order for this cleft closure to happen while retaining contact with ssNA, it is evident that one of the domains must release its grip on the NA strand and slide along it. ATP hydrolysis results in a destabilisation of the cleft closure because contacts mediated through the γ phosphate are broken. As the ATP binding cleft opens, domain 2 weakens its binding to NA and slides along the strand while the domain 1 remains bound to the NA tightly (78, 110).

Brownian motor mechanism

Another proposed model for NS3h is the Brownian motor mechanism which postulates that HCV NS3h helicase translocates single-stranded nucleic acid (ssNA) as a Brownian motor with a simple two-stroke cycle (111). The directional movement step is driven by ssNA binding, while ATP binding allows for a short period of random movement that prepares the helicase for the next cycle. HCV NS3h cycles between a tightly bound and a weakly bound state under the control of the ATP hydrolysis cycle and slides along the ssNA strand without directional preference (111). In the weak binding state, ATP binding induces reduced affinity of NS3h for single-stranded (ss) nucleic acid. Brownian motion allows NS3h to slide forward or backwards in the weakly NA bound state. The model predicts that preventing backward motion should increase the efficiency of the helicase. The backward motion can be prevented by either a second helicase molecule bound to the NA behind the leading helicase molecule or by interaction with a molecule bound to the opposite strand (111).

Brownian ratchet mechanism

An alternative mechanism is the Brownian ratchet mechanism, in which NS3h diffuses along the nucleic acid (NA) as a result of ATP-dependent cycling between tight and weak binding states, same as the proposed Brownian motor model (111). However, the Brownian ratchet mechanism is different from Brownian motor model, as the NS3h translocates the ssNA with a direction. Movement in the 3’ to 5’ direction is favoured due to the ratchet effect presumably

due to W501, which reduces the probability of the enzyme sliding backwards (114). The W501 ('rachet' tryptophan) was proved to be critical for HCV NS3h function (109). The importance of an aromatic residue at this position is also demonstrated by the fact that it can be substituted with phenylalanine, whereas alanine disrupts NA unwinding (115).

1.4 X-ray Crystallography

In this study, the phiKo gp14 and Zika virus NS3 helicase structures are solved by X-ray crystallography, specifically, by molecular replacement. X-ray crystallography is a technique used to determine the molecular structure from a crystal via transforming the recorded X-ray diffraction patterns to electron density maps. Some basic information involved in X-ray diffraction and structural determination is described below in this section.

1.4.1 Vapour diffusion crystallisation

Determination of a protein structure by X-ray crystallography requires the preparation of crystals. Several methods can be used to produce protein crystals, including vapour diffusion, micro-batch, micro-dialysis and free-interface diffusion. Hanging and sitting drop vapour diffusion experiments are widely used for the growth of protein crystals (Figure 1-10a and 1-10b). In this method, droplets containing purified protein, buffer and precipitant are allowed to equilibrate against a larger reservoir containing similar buffers and precipitants at higher concentrations (116). When the crystallisation system is at equilibrium, the crystal loses protein molecules at the same rate at which protein molecules re-join the crystal. The concentration of proteins in the solution at equilibrium is the equivalent to the protein solubility (117).

The first step of crystallisation is to form nuclei. Initially, the drops of protein solution contain comparatively low precipitant and protein concentrations. Nucleation usually occurs when the protein concentration increases through dehydration-driven reduction of the solution volume caused by equilibration through vapour diffusion from the protein-containing droplet to the more hygroscopic reservoir solution (Figure 1-10a and 1-10b). Once nucleation conditions are achieved, the drop solution remains supersaturated so that both nucleation and rapid crystal growth can occur simultaneously (118). Since more protein molecules form crystals and less protein remains in solution, protein concentration in solution will decrease to some extent.

The solubility of a protein varies with the concentration of a precipitant, as shown schematically in Figure 1-10c. Crystals dissolve in the unsaturation region, where the concentration is below the protein solubility, and grow in the supersaturation region. The supersaturation region can be divided into three zones, metastable, nucleation and precipitation. Since there is an energy barrier, nucleation (the process of forming a crystallisation nucleus)

takes time. In the 'metastable zone', the supersaturation is too small, and the nucleation rate will be too slow to form crystals. In the 'nucleation' zone, the supersaturation is large enough to lead to spontaneous nucleation. If the supersaturation is too large, then disordered assemblies, such as aggregates or precipitates, may form. The 'precipitation zone' is unfavourable for crystal formation, because the disordered aggregates and precipitates form faster than the ordered crystals. Since these zones are related to kinetic phenomena, the boundaries between the zones are not well-defined. In contrast, the solubility curve is clearly defined by the equilibrium between solution and crystal (117).

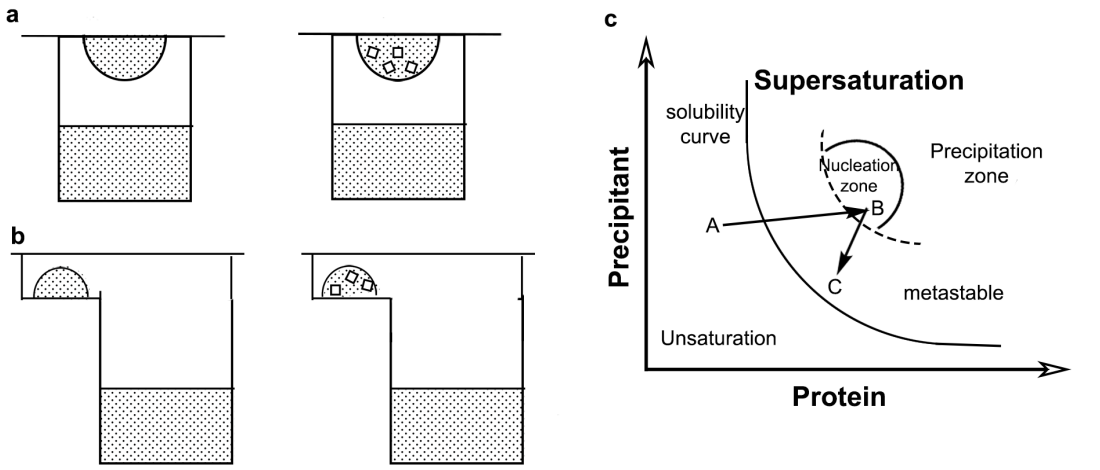


Figure 1-10 Schematic diagram of crystallisation by vapour diffusion.

a. Hanging drop vapour diffusion; b. Sitting drop vapour diffusion; in a and b, the squares indicate crystals. c. Protein crystallisation phase diagram. A to B signifies protein transition from the unsaturated state to the supersaturation level where nucleation and initial growth occur by vapour equilibration of the drop against the reservoir solution. B to C transition corresponds to changes in soluble protein concentration in the drop that is likely to decrease supersaturation over the time course of the experiment.

1.4.2 Micro-seeding

Some crystals obtained in initial screens are not of sufficient quality for structure determination by X-ray diffraction, and the conditions must be optimised in subsequent rounds of experiments to produce the desired quality crystals. Micro-seeding is a crystallisation method in which crystals or crystalline material generated in one experiment is transferred to a new experiment, for crystal optimisation. The crystalline material acts as a nucleation site for crystal growth in the new experiment, removing the need for *de novo* nucleation. This approach is rationalised by the hypothesis that the optimal conditions for crystal nucleation and crystal growth can be quite different (119, 120).

1.4.3 Diffraction theory

1.4.3.1 Bragg's Law

Laue, Friedrich and Knipping first observed X-ray diffraction by a crystal in 1912. In 1913, William Lawrence Bragg and his father, William Henry Bragg, proposed a description by considering the diffracted beams as reflections from sets of parallel planes that pass-through lattice points of the crystal (121). Though this description of 'reflections' is not correct in a physical sense, as planes of atoms do not reflect X-rays as such, it is correct in a geometrical sense and provides the simple expression for the analysis of crystal structure. Bragg's Law provides the condition for a plane wave to be diffracted by a family of lattice planes (Figure 1-11), according to the equation:

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

where d is the lattice spacing, θ is the angle between the wave vector of the incident plane wave and the lattice planes, λ is the wavelength and n is an integer, the order of the reflection (121). It is equivalent to the diffraction condition in reciprocal space and the Laue equations. The reciprocal lattice is constituted of the set of all possible linear combinations of the basis vectors a^*, b^*, c^* of the reciprocal space. The reciprocal and real spaces are related by a Fourier transform, and the reciprocal space is also called Fourier space or phase space. A point (node), H, of the reciprocal lattice is defined by its position vector:

$$OH = r^*_{hkl} = ha^* + kb^* + lc^*$$

Each vector OH of the reciprocal lattice is associated with a family of direct lattice planes (122).

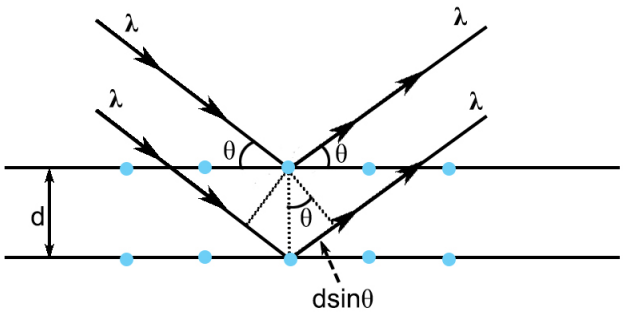


Figure 1-11 Bragg's Law.

The d is the lattice spacing, θ is the angle between the vector of the incident wave and the lattice planes, λ is the X-ray wavelength. Lattice points in crystal are displayed as blue dots.

1.4.3.2 Ewald sphere

The Ewald sphere is a geometrical expression of Bragg's Law which relates the reciprocal lattice and a 'sphere of reflection'. The Ewald sphere is a sphere of a radius of 1/λ (where λ is the X-ray wavelength) with the crystal at the centre (A), passing through the origin (O) of the reciprocal lattice, and the reciprocal lattice vector is d^*_{hkl}, as seen in Figure 1-12. The incident X-ray beam's direction is along a radius of the sphere, AO (Figure 1-12). If Bragg's Law is satisfied, it may be shown that the vector OP, connecting the point where the direct beam exits from the sphere with the point where the diffracted beam exits from the sphere, is identical to d^*_{hkl}, the spacing in the reciprocal space. In the Ewald construction, the reciprocal lattice origin is shifted from the centre of the sphere (A) to the point where the direct beam exits from the sphere (O) (Figure 1-12). Bragg's Law is equivalent to the statement that the reciprocal lattice point for the reflecting planes (*hkl*) should lie on the surface of the sphere. In contrary, if the reciprocal lattice point does not intersect the sphere, then Bragg's Law is not satisfied, and no diffracted beam occurs (123).

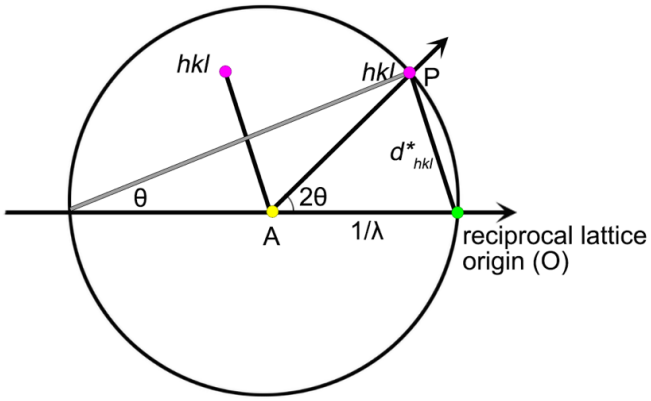


Figure 1-12 Ewald sphere for a set of planes at the correct Bragg angle.
The radius of the sphere is 1/λ, λ is the X-ray wavelength, d^*_{hkl} is the reciprocal lattice vector, *hkl* is the reflecting plane, O is the origin of the reciprocal lattice, A is the centre of the sphere where the crystal is placed.

1.4.4 X-ray diffraction

X-ray diffraction is the scattering of X-ray photons by atoms in a periodic lattice. The scattered monochromatic X-rays that are in phase give constructive interference. The intensities of X-ray diffraction peaks are determined by the atomic positions and nature of the atoms within the lattice planes. X-ray diffraction can be utilised to study single crystal or polycrystalline materials (124, 125). The single-crystal X-ray diffraction experiment is performed by shooting

an X-ray through crystals to produce diffraction patterns, as illustrated in Figure 1-13. The crystal is mounted on a goniometer. A collimated beam of X-rays is directed at the crystal. Most X-rays are transmitted through the crystal, but a small fraction is diffracted by the crystal lattice. X-rays diffracted by sets of planes satisfying the Bragg's Law at a given mutual orientation of the crystal and X-ray beam, are recorded using a 2-dimensional (2-D) X-ray detector. During the collection of a complete data set, the crystal is rotated around φ° (Figure 1-13) to bring all sets of planes into diffracting position. For practical reasons, a total crystal rotation of around 180° around the rotation axis is used during a routine data collection at a synchrotron, even though a significantly smaller total rotation is usually sufficient for recording a complete data set, especially in cases with high internal symmetry of the crystal.

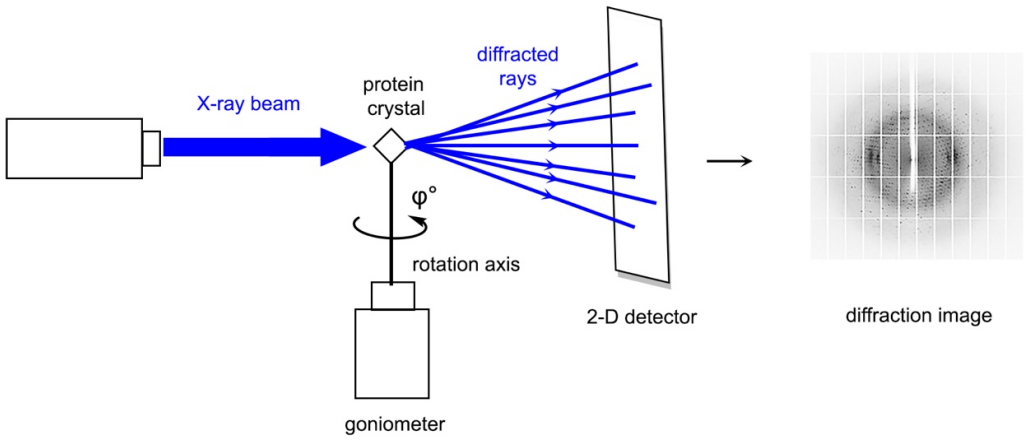


Figure 1-13 Single-crystal X-ray diffraction experiment.
This figure is detailed in the text.

1.4.5 Data collection

A diffraction data set forms an image of 3-D reciprocal space. This 3-D image consists of a series of 2-D diffraction images, each of them representing a different, curved slice of reciprocal space (126). Each 2-D diffraction image is collected while rotating the crystal by a small angle (typically between 0.2 and 2°) about a fixed axis (Figure 1-13). The rotation method with minimal rotation angles per image (fine φ-slicing) has been developed to provide improved signal to noise for weakly diffracting samples (127). If a fine φ-slicing approach is applied, there is generally no advantage in reducing $\Delta\varphi$ to less than half of the crystal mosaicity. In crystallography, the mosaicity is a measure of the spread of crystal plane orientations. A mosaic crystal is a simplified model of a real crystal imagined to consist of numerous small crystallite blocks that are, to some extent, randomly misoriented (122). For very large mosaicity values and large cell dimensions, reflection overlaps may occur no matter how small the rotation increment is (128).

To determine the crystal lattice symmetry, it is necessary to index the collected reflections from sets of different planes. Software for autoindexing, such as the program iMOSFLM (129), is usually able to use just one or a few images to match the spatial location of spots with the dimensions of the unit cell and give the likely Bravais lattice, as well as determine its orientation with respect to the experimental apparatus (127). The unit cell is defined as the smallest repeating unit having the full symmetry of the crystal structure, and the Bravais lattice is an arithmetic crystal class with matrix group of lattices (122).

1.4.6 Data processing

After collecting sufficient diffraction data, the next step of structural determination is data processing. The analysis and reduction of single-crystal diffraction data consist of seven major steps. These are (i) visualization and preliminary analysis of the unprocessed detector data, (ii) indexing of the diffraction pattern, (iii) refinement of the crystal and detector parameters, (iv) integration of the intensity of diffraction peaks, (v) finding the relative scale factors between individual measurements, (vi) precise refinement of crystal parameters using the whole data set, and (vii) merging and statistical analysis of the measurements related by space-group symmetry (126).

The quality of diffraction data is usually evaluated by the merging R factor (R_{merge}), based on reflection intensities. R_{merge} gives the average ratio of the spread of intensities of the multiply measured symmetry-equivalent reflections to the estimated value of the reflection intensity:

$$R_{merge} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$$

where $I_i(hkl)$ indicates the individually measured intensities of all reflections equivalent by the point group symmetry to $I(hkl)$ and $\langle I(hkl) \rangle$ is the average of all such intensities. The value of R_{merge} is not a statistically objective indicator since it strongly depends on the data multiplicity. Consequently, it is always higher for data in high-symmetry space groups than those in low symmetry. The R_{merge} value increases with increasing multiplicity (130, 131). Other versions of R factor have been proposed, such as R_{meas} (132):

$$R_{meas} = \frac{\sum_{hkl} n \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} (n-1) \sum_i I_i(hkl)}$$

where n is a number of the symmetry equivalent contributors to the average. R_{meas} is an improvement over R_{merge}, as it is relatively insensitive to data multiplicity (132, 133).

A useful statistic that comes from scaling is the ratio of the measured signal to noise, $I/\sigma(I)$. In principle, as long as the ratio of average intensity to the associated estimated error is higher than 1.0, the data contain some information (134). This criterion can be used to define the

diffraction resolution limit for a crystal, i.e. the resolution at which the data may be truncated without losing significant information. In addition, the correlation coefficient between random half data sets (CC1/2) is a primary indicator for selecting high resolution cut-off for data processing. It is related to the effective signal to noise of the data (135, 136). The advantage of a correlation coefficient is that it has a well-defined range: 1.0 for a good correlation and 0.0 for no correlation (135). Typically, CC1/2 is near 1.0 (or 100%) at low resolution and drops smoothly toward 0 as the signal-to-noise ratio decreases. The value of CC1/2 between ~ 0.1 and ~ 0.4 can be roughly equated to $I/\sigma(I)$ values between ~ 0.5 and ~ 1.5 (136).

1.4.7 Structure factors

The structure factor is the central concept in structure determination by X-ray diffraction. The wave that results from scattering is called a structure factor, $\mathbf{F}_{hkl}$, which is described in a vector form. Mathematically, the structure factor $\mathbf{F}_{hkl}$ can be described by the following functions:

$$\mathbf{F}_{hkl} = F_{hkl} \exp(i\alpha_{hkl})$$

$$= \sum_j f_j \exp\left[2\pi i\left(hx_j + ky_j + lz_j\right)\right]$$

$$= \sum_j f_j \cos\left[2\pi(hx_j + ky_j + lz_j)\right] + i\sum_j f_j \sin\left[2\pi(hx_j + ky_j + lz_j)\right]$$

where α_{hkl} is the phase of the diffracted beam; the F_{hkl} is the amplitude; the summation is over all atoms in the unit cell; x_j, y_j, z_j are the positional coordinates of the jth atom and f_j is the scattering factor of the jth atom (122, 137). In structure determination, the intensity of a diffracted beam is directly related to the structure factor amplitude. However, the phases are estimated, and an initial description of the positions and anisotropic displacements of the scattering atoms is deduced. The calculated structure factors of this initial model are compared with the experimentally observed structure factors. Iterative refinements used to minimise the difference between calculated and observed structure factors until a satisfactory fit is obtained (122).

1.4.8 The phase problem

The square of the structure factor amplitudes is related to the diffracted intensities and can be converted using the following equation:

$$F_{hkl}^2 = I_{hkl}(kLpA)^{-1}$$

Where I_{hkl} is intensity, k is a scale factor, Lp is the Lorentz-polarization correction factor, and A is the transmission factor representing the absorption of X-rays by the crystal (122).

X-ray structure determination aims to obtain the electron density map for the protein crystal. The electron density is the Fourier transform of the structure factors. To calculate the electron

density at a position (*xyz*) in the unit cell of a crystal, the following summation over all the *hkl* planes is required:

$$\rho(xyz) = 1/V \sum F_{hkl} \exp(i\alpha_{hkl}) \, exp\,[-2\pi i(hx + ky + lz)]$$

where V is the volume of the unit cell and α_{hkl} is the phase associated with the structure-factor amplitude F_{hkl}. The amplitudes can be measured, but the phases of the diffracted amplitudes cannot be experimentally recorded, this represents the phase problem of X-ray crystallography. The phases are essential in producing the correct electron density and carrying structural information (138). Without knowing the phases, it is not possible to reconstruct the atomic structure. Therefore, estimating the phases is an essential step in successful structure determination (122, 139).

In protein crystallography, generally, phases are derived either by using the atomic coordinates of a structurally similar protein, i.e. molecular replacement (MR), or by finding the positions of heavy atoms that are intrinsic to the protein or that have been added, such as multiple isomorphous replacement (MIR) and single-wavelength/multi-wavelength anomalous dispersion (SAD/MAD) or combinations of these (138). Phase information is further improved during structure refinement, a mathematical process that minimises differences between the observed and calculated (from the current model) amplitudes while imposing geometric restraints on the atomic model.

1.4.8.1 Multiple Isomorphous Replacement (MIR)

Initial phases for new macromolecular structures have originally been derived by the method of multiple isomorphous replacement (MIR) (140). Information about unknown phases may be obtained by making a known change to the contents of the unit cell and measuring the effect on the diffraction pattern. In practice, a reactive group containing a heavy metal ion is introduced to the protein without disturbing the overall structure and packing of molecules in the crystal, i.e. the two structures must be 'isomorphous'. Due to the heavy atoms scattering X-ray more strongly than the rest of the atoms in the structure, it is usually possible to locate the heavy atoms alone by direct methods or using Patterson maps. Once the location of the heavy atoms is known, the scattering from those atoms may be calculated both in magnitude and in phase (139). In general, the phase information from a single isomorphous derivative is ambiguous. Thus, multiple isomorphous replacements (MIR) or supplementary information, such as from anomalous scattering, molecular averaging, or solvent flattening is used to resolve the phase ambiguity (141).

Nevertheless, isomorphous replacement has several problems: non-isomorphism between crystals (unit-cell changes, reorientation of the protein, conformational changes, changes in

salt and solvent ions), problems in locating all the heavy atoms, problems in refining heavy-atom positions, occupancies and thermal parameters and errors in intensity measurements (138, 142). In some cases, SAD/MAD is ideal for solving the phase problem when there is no perfect isomorphism (142), as described below.

1.4.8.2 Single-/Multi-wavelength Anomalous Dispersion (SAD/MAD)

Anomalous scattering

The atomic scattering factor contains three components: a real scattering term f_0 that is dependent on the Bragg angle and two terms, f' and f'', that are dependent on wavelength but not scattering angle. If X-rays can excite electrons surrounding the atomic nucleus such that they are able to jump from lower to higher energy shells, an auxiliary resonant anomalous signal is observed, and the atomic form factor can be expressed as a complex number $f' + if''$. Generally, f'' is proportional to the atomic absorption of the X-rays and their fluorescence and f' follows the derivative of this function, according to the Kramer-Kronig transformation (143). In contrast to the normal atomic scattering factor f', the anomalous dispersion corrections f' and f'' depend only on the wavelength λ of the X-rays used for the diffraction experiment and do not diminish with the diffraction angle (142).

The full atomic form factor is:

$$f(\theta, \lambda) = f_0(\theta) + f'(\lambda) + if''(\lambda)$$

The dispersive term f' modifies the real scattering factor, whereas the absorption term f'' is 90° advanced in phase. In the absence of anomalous scattering, the intensities of reflections with indices h, k, l and -h, -k, -l are equal and form Friedel pairs. Friedel's law states that the members of Friedel pairs have the same amplitude, but the opposite phase (Figure 1-14a). However, in the presence of anomalous scattering, this relationship is no longer true because there is an additional contribution to the phase from the anomalous scatterers, breaking both the amplitude and phase relationship between the Freidel pairs (Figure 1-14b).

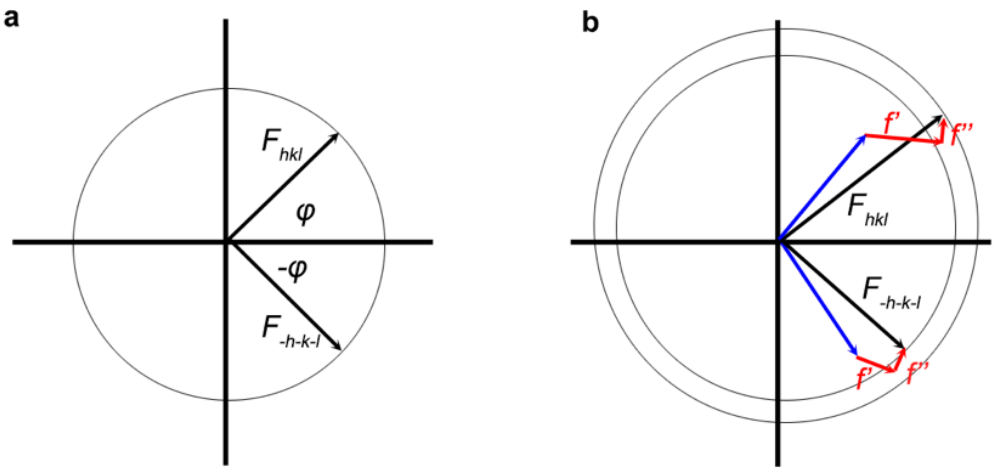


Figure 1-14 Breaking of Friedel's law.

a. The normal reflection. b. The effect of anomalous scattering is added to a, breaking the relationship between F_{hkl} *and* F_{-h-k-l}. The whole non-anomalous part of the structure factors (not include any anomalous component) is blue, and the anomalous parts due to the f' and f'' values of the heavy atoms are in red.

The anomalous difference can be used in the same way as the isomorphous difference in Patterson or direct methods to locate the anomalous scattered waves. Phases for the native structure factors can then be derived in a similar way to the MIR case. Anomalous scattering can be used to solve the phase ambiguity in a single isomorphous replacement experiment, leading to single isomorphous replacement with anomalous scattering (SIRAS) (138).

Single-/multi-wavelength Anomalous Dispersion (SAD/MAD)

Single wavelength anomalous dispersion (SAD) is an approach applied in protein structure determination to obtain interpretable electron density maps from intensities containing the anomalous signal within a single dataset recorded using only one X-ray wavelength. Multiwavelength anomalous dispersion (MAD) is an approach used by comparing the anomalous intensities collected at different wavelengths. MAD uses only the wavelength-dependence of the atomic structure factor of the anomalously scattering atoms for solving the phase problem (141, 144, 145). In this approach, several data sets are collected at various wavelengths around the absorption edge of the anomalous scatterer present in the crystal and the differences in the f' and f'' contributions are utilised for phase calculation (142).

The scattering from an atom is usually largely independent of wavelength. However, each atomic type has a few 'absorption edges' around which the scattering varies rapidly (in amplitude and phase) with wavelength. By differing the wavelength around the absorption edge for an atom, the contribution from those atoms to the total scattering can be different. If

the positions of the anomalously scattering atoms are known, such as a few heavy atoms in the structure, phase information may be recovered by using the same method as for MIR data. In theory, it is possible to obtain an unambiguous phase estimate for a structure factor using only a single crystal, by measuring the scattering at a single or several wavelengths at the absorption edge (139, 142). Though SAD is simpler than MAD, MAD usually give a better initial phase estimate. In addition, the SAD/MAD approach is particularly well suited to proteins where methionine residues can be replaced by seleno-methionine derivatives. In practice, the selenomethionyl proteins can be produced in the expression cells B834 (DE3) cultured in seleno-methionine containing cell media (146–148). Selenium has a sufficiently strong anomalous scattering effect allowing phasing of a macromolecule (122).

1.4.8.3 Molecular replacement

Molecular replacement (MR) is a phasing method that uses previous information from known structures that are related or homologous to protein components in the crystal. The first recorded use of the term 'molecular replacement' was in 1972 (149). Since no additional experimental procedures or data are required, MR is usually applied for structure determination when a suitable search model is available. In principle, MR involves rotational and translational searches over many possible placements of a molecular model within the unit cell of an unknown structure. MR makes use of the fact that proteins with similar sequences typically form similar three-dimensional structures. The conventional approach for searching structurally related models in MR is to use the sequence of the target protein to search against a series of solved structures. Sequence similarity often correlates with structural similarity. Given sufficient similarity, a known structure correctly positioned in the unit cell of the target structure by MR can provide approximate phase information for the structure of interest. An alternative approach (protein sequence independent) for identifying homologous structures is to exploit the measured data directly, comparing the lattice parameters searched against the entire Protein Data Bank (PDB) or applying a brute-force search of a nonredundant PDB database (150). If the crystal diffraction dataset has sufficiently high resolution, fragments as small as single atoms of elements usually found in proteins can yield *ab initio* solutions of macromolecular structures, including some that elude traditional direct methods (151).

In this thesis, the crystal structures of the two ATP driven nucleic acid motors, gp14 from the *Thermus thermophilus* bacteriophage phiKo and the NS3 helicase from Zika virus, were determined by molecular replacement, as described in sections 2.2.6 and 3.2.3.

2 Structure and function of the DNA packaging ATPase from *Thermus thermophilus* phage phiKo

2.1 Introduction

Double-stranded DNA (dsDNA) viruses use ATP-driven motors for loading viral genomes into pre-formed viral capsid during virus assembly, in a process known as viral DNA packaging. Genome packaging is a crucial step in virus assembly and usually accomplished by biological molecular motors that convert the chemical energy of ATP or NTP hydrolysis into conformational changes that drive DNA packaging (6, 8, 26). A large group of dsDNA viruses, including *Poxviruses*, giant viruses (such as Pandoravirus), Polintoviruses and bacterial *Tectiviruses* as well as virophages, utilise an FtsK-like ATPase for packaging their genome into a pre-formed procapsid, detailed in sections 1.1.1 and 1.2.1.

This chapter describes structural and functional characterisations of the putative DNA packaging ATPase from the bacteriophage phiKo, belonging to the *Tectiviridae* family. The aim of this part of the project was to achieve insights into the structure and functional properties of this DNA packaging ATPase.

2.2 Methods and materials

2.2.1 Cloning

The gene of the putative DNA packaging ATPase was derived from the genomic DNA of the *Thermus thermophilus* phage phiKo, namely the fourteenth coding fragment of the forward strand of the genome. The product encoded by this gene is here-on referred to as gp14. In order to generate a sufficient amount of gp14 for crystallisation and functional characterisations, several expression vectors were used for over-expressing the target proteins *in vitro* to improve yield. The target gene was introduced into the expression vectors by Ligation Independent Cloning (LIC).

LIC is a technique developed in the early 1990s as an alternative to restriction enzyme/ligase cloning (152). LIC makes use of the 3’ to 5’ activity of a DNA polymerase to create specific single-stranded complementary overhangs (of at least 12 bases) on the target vector and the DNA fragment for insert created by the polymerase chain reaction (PCR). PCR products with complementary overhangs are introduced using primers that contain the appropriate 5’ sequence extensions. Inserts are normally amplified by PCR, and linearized vectors can be made either by restriction enzyme digestion or by PCR. The ligation-independent In-Fusion Cloning or HiFi assembly systems were used for the DNA assembly reaction in this thesis.

2.2.1.1 Preparation for linearized LIC vectors by PCR

In this study, five expression vectors, YSBL-LIC (-), YSBL-LIC3C, pET22b, pET 6 × His GFP TEV LIC vector (GFP_TEV) and Champion™ pET SUMO, were used for cloning constructs of phiKo gp14. YSBL-LIC expression vectors are derived from the pET-28a plasmid (Novagen) and have the same expression characteristics (153). The primers for annealing the expression vectors of YSBL-LIC (-), YSBL-LIC3C, pET22b and Champion™ pET SUMO (listed in Appendix 1) were provided by Maria Chechik in the Antson group. The primers for linearization of the plasmid GFP_TEV were designed in this study and are given in Appendix 1. All linearized vectors were created by PCR, except for PET22b was linearized by restriction enzymes, NdeI and XhoI (Thermo Scientific™). The PCR protocol is given in Table 2-1.

Table 2-1 PCR protocols for linearization of the expression vectors.

Plasmids	PCR reaction composition		PCR program			Steps
YSBL-LIC (-)	50 µl reaction:				cycles	
			98 ˚C	30 s	1	Initial denaturation
	DNA template	1 µl (~	98 ˚C	15 s	35	Denaturation
	Forward and reverse primers (2 µM for each)	10 µl x 2	60 ˚C	30 s	35	Annealing
	CloneAmp™ HiFi PCR Premix (2×)	25µl	72 ˚C	120 s	35	Elongation
	Ultra-pure water	Top up to 50 µl	72 ˚C	420 s	1	Final elongation
YSBL-LIC3C	50 µl reaction:				cycles	
			98 ˚C	30 s	1	Initial denaturation
	DNA template	1 µl (~	98 ˚C	10 s	35	Denaturation
	Forward and reverse primers (2 µM for each)	10 µl x 2	55 ˚C	30 s	35	Annealing
	CloneAmp™ HiFi PCR Premix (2×)	25 µl	72 ˚C	40 s	35	Elongation
	Ultra-pure water	Top up to 50 µl	72 ˚C	420 s	1	Final elongation
GFP_TEV	25 µl reaction:				cycles	
			98 ˚C	30 s	1	Initial
	DNA template (pYM350 plasmid, amount <100ng)	0.5 µl	98 ˚C	15 s	40	Denaturation
	Forward and reverse primers (2 µM for each)	5 µl x 2	62 ˚C	20 s	40	Annealing
	CloneAmp™ HiFi PCR Premix (2×)	12.5 µl	72 ˚C	180 s	40	Elongation

	Ultra-pure water	Top up to 25 µl	72 ˚C	420 s	1	Final elongation

DpnI restriction enzyme was used to degrade template DNA (plasmid) after the PCR reactions, as no methylases were present in the reaction and the PCR products lacked methylation. The DpnI restriction enzyme binds to and digests methylated 5'-GmeATC-3' sites (154). PCR products were digested by 1.5 µl (10 U/µl) DpnI (Thermo Scientific™) in 1 × Tango buffer (Thermo Scientific™), and the reaction was incubated at 37 ˚C for 3 h. The linearized plasmid was further purified using the NucleoSpin™ Gel and PCR Clean-up kit and stored at - 20 ˚C for further use.

2.2.1.2 Amplification of genes of interest

The DNA sequence of the full length phiKo gp14 protein (residues 1 - 203) was amplified from the genomic DNA of *Thermus thermophilus* phage phiKo and cloned into expression vectors using either NEBuilder® HiFi DNA Assembly or In-Fusion® HD Cloning Plus system to generate N-terminal or C-terminal tag fusions. The genomic DNA of phage phiKo (NCBI accession code: MH673671, unpublished) was provided by Anna Lopatina and Konstantin Severinov (Skoltech, Moscow). Twelve constructs were cloned to investigate the phiKo gp14 protein (Table 2-2), and PCR protocols for each construct are given in Table 2-3. Primers for amplification of the genes of interest were designed based on LIC (as described above in section 2.2.1) and for single-point mutations were designed based on the QuickChange protocol (Takara). All primers are shown in Appendix 1. Primers were synthesised by Eurofins Genomics Germany GmbH (purification level: salt free).

Table 2-2 List of constructs of phiKo gp14.

Construct	Fusion protein name	Vector	Protein residues	Type	Antibiotics resistance
non-cleavable tag fusion constructs					
pYM351	N-terminal 6 × His tagged gp14	YSBL-LIC (-)	M1-E203	wild type	Kanamycin, Chloramphenicol
pYM418	N-terminal 6 × His tagged gp14	YSBL-LIC (-)	G7-E203	wild type	Kanamycin, Chloramphenicol
pYM419	C-terminal 6× His tagged gp14	pET22b	G7-E203	wild type	Ampicillin, Chloramphenicol

pYM537	7 amino acid insertion for linking two N-terminal 6 × His tagged gp14	YSBL-LIC (-)	M1-E203	wild type	Kanamycin, Chloramphenicol
pYM546	13 amino acid insertion for linking two N-terminal 6 × His tagged gp14	YSBL-LIC (-)	M1-E203	wild type	Kanamycin, Chloramphenicol
pYM569	10 amino acid insertion for linking two N-terminal 6 × His tagged gp14	YSBL-LIC (-)	M1-E203	wild type	Kanamycin, Chloramphenicol
cleavable tag fusion constructs					
pYM350	N-terminal SUMO tagged gp14	Champion™ pET SUMO	M1-E203	wild type	Kanamycin, Chloramphenicol
pYM409	N-terminal 6xHis tagged gp14	YSBL-LIC3C	M1-E203	wild type	Kanamycin, Chloramphenicol
pYM635	N-terminal SUMO tagged gp14	Champion™ pET SUMO	M1-E203	K22A mutant	Kanamycin, Chloramphenicol
pYM636	N-terminal SUMO tagged gp14	Champion™ pET SUMO	M1-E203	D100A mutant	Kanamycin, Chloramphenicol
pYM637	N-terminal SUMO tagged gp14	Champion™ pET SUMO	M1-E203	R124A mutant	Kanamycin, Chloramphenicol
pYM638	N-terminal GFP tagged gp14	GFP_TEV	M1-E203	wild type	Kanamycin, Chloramphenicol

Table 2-3 PCR protocol for amplification for phiKo gp14.

Constructs	PCR reaction composition		PCR program		
pYM 409	25 µl reaction:				cycles
			98 ˚C	30 s	1
	DNA template	0.5 µl (~ 20	98 ˚C	10 s	35
	Forward and reverse primers (2 µM for each)	5 µl × 2	55 ˚C	30 s	35
	CloneAmp™ HiFi PCR Premix	12.5 µl	72 ˚C	40 s	35
	Ultra-pure water	Top up to 25 µl	72 ˚C	420 s	1
pYM418,	50 µl reaction:				cycles

pYM419			98 ˚C	30 s	1
	DNA template	1 μl (~ 40 ng)	98 ˚C	10 s	35
	Forward and reverse primers (2 μM for each)	10 μl × 2	55 ˚C	30 s	35
	CloneAmp™ HiFi PCR Premix	25 μl	72 ˚C	40 s	35
	Ultra-pure water	Top up to 25 μl	72 ˚C	420 s	1
pYM537	25 μl reaction:				cycles
			98 ˚C	30 s	1
	DNA template	0.5 μl (~ 20	98 ˚C	15 s	35
	Forward and reverse primers (2 μM for each)	5 μl × 2	60 ˚C	30 s	35
	CloneAmp™ HiFi PCR Premix	12.5 μl	72 ˚C	40 s	35
	Ultra-pure water	Top up to 25 μl	72 ˚C	300 s	1
pYM546	25 μl reaction:				Cycles
			98 ˚C	30 s	1
	DNA template	0.3 μl (~ 20	98 ˚C	15 s	40
	Forward and reverse primers (2 μM for each)	5 μl × 2	65 ˚C	30 s	40
	CloneAmp™ HiFi PCR Premix	12.5 μl	72 ˚C	50 s	40
	Ultra-pure water	Top up to 25 μl	72 ˚C	300 s	1
pYM569	25 μl reaction:				Cycles
			98 ˚C	30 s	1
	DNA template	0.3 μl (~ 20	98 ˚C	15 s	40
	Forward and reverse primers (2	5 μl × 2	65 ˚C	30 s	40
	CloneAmp™ HiFi PCR Premix	12.5 μl	72 ˚C	50 s	40
	Ultra-pure water	Top up to 25 μl	72 ˚C	300 s	1
pYM635, pYM636, pYM637	25 μl reaction:				cycles
			98 ˚C	30 s	1
	DNA template (pYM350 plasmid)	0.5 μl (~ 20	98 ˚C	10 s	16
	Forward and reverse primers (2 μM for each)	4 μl × 2	55 ˚C	30 s	16
	CloneAmp™ HiFi PCR Premix	12.5 μl	72 ˚C	180 s	16
	Ultra-pure water	Top up to 25 μl	72 ˚C	420 s	1
pYM638	25 μl reaction:				cycles
			98 ˚C	30 s	1
	DNA template (pYM569 target gene PCR product)	0.5 μl (~ 20 ng)	98 ˚C	15 s	40
	Forward and reverse primers (2 μM for each)	4 μl × 2	68 ˚C	30 s	40
	CloneAmp™ HiFi PCR Premix	12.5 μl	72 ˚C	50 s	40
	Ultra-pure water	Top up to 25 μl	72 ˚C	300 s	1

PCR product, produced using plasmid as a template, was digested by DpnI (Thermo Scientific™) as described in section 2.2.1.1. To ensure bands of approximately the correct size were formed, 0.5 µl of the digestion product was resolved by electrophoresis using a 1% agarose gel, and 1 µl 1 × DNA gel loading dye (Thermo Scientific™) was added to the sample before loading it onto the gel. The gel was run in 1× TAE buffer at 100 V for 60 min, stained with 1 × SYBR™ safe dye (diluted from 10,000-fold stock) and then visualised using the GelDoc system (BioRad). 1 × TAE buffer contains 40 mM Tris pH 8.3, 20 mM acetic acid and 1 mM EDTA. The PCR products with correct lengths (sizes) were further purified using the NucleoSpin™ Gel and PCR Clean-up kit and stored at - 20 °C for subsequent use.

2.2.1.3 Cloning of engineered homodimer constructs of phiKo gp14

Since gp14 protein possesses an FtsK-like fold and equivalent conserved Walker A and B motifs, based on sequence alignment and structure analysis, discussed in results section 2.3.4, gp14 is expected to form an oligomer *in vivo*, like other FtsK/HerA family members (as discussed in sections 1.3.2, 1.3.3 and 1.3.3.1). However, the fusion proteins produced in this study were purified as a monomer and crystallised as a monomer species, as discussed in sections 2.3.2 and 2.3.4 below. Therefore, a protein engineering approach was adopted in an attempt to engineer two gp14 monomers in one polypeptide chain with a flexible linker. The linkers of the engineered homodimer of phiKo gp14 were designed based on a method previously used for an AAA^+ family ATPase (155). The homodimer of phiKo gp14 with an inserted polypeptide chain linker was modelled on the basis of the crystal structure of *P. Aeruginosa* FtsK hexamer [PDB code 2IUU, (72)]. Such modelling showed that a 7-amino acid (aa) linker has sufficient length for connecting the C-terminus of one subunit to the N-terminus of the next adjacent subunit in the clockwise direction. Modelling also suggested that a linker length with 7 - 13 amino acids is not able to connect the N terminus of the adjacent counter-clockwise subunit.

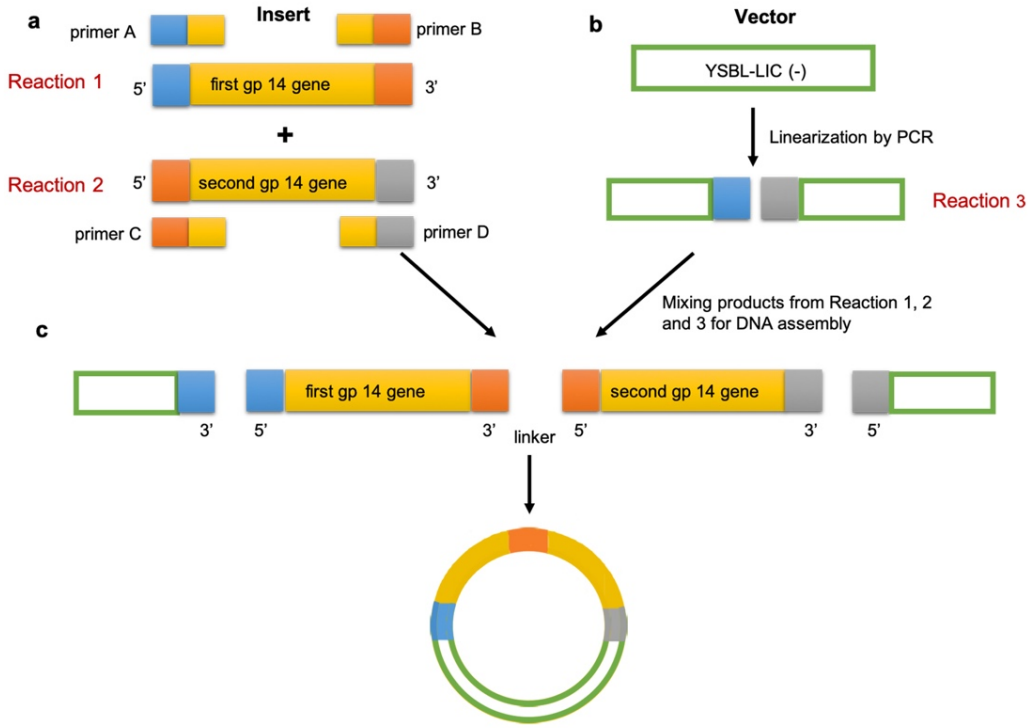


Figure 2-1 Cloning protocol of engineered homodimer constructs.

a. Reactions 1 and 2 for producing gp14 genes with unique 5' and 3' ssDNA overhangs, coloured in blue, orange and grey, respectively. b. Linearization of the expression vector of YSBL-LIC (-); c. DNA assembly of the recombinant plasmid expressing the engineered homodimer constructs.

In order to investigate methods to improve oligomerisation of the phiKo gp14, three engineered constructs, pYM537, pYM546 and pYM569, were cloned by linking the genes for two gp14 monomers with a DNA sequence coding for a short flexible linker. The phiKo gp14 gene was amplified from the genomic DNA and assembled into a tandem construct, or 'double-gene', separated by the linker sequence (Figure 2-1). The PCR primers were designed to generate unique specific single-stranded (ss) overhangs on the 5' end and 3' end of the gp14 gene in two separate reactions, respectively, based on DNA assembly protocols to introduce the recombinant fragments into the YSBL-LIC(-) expression vector (Figure 2-1). To generate the fusion between the tandem genes, primers in this segment were designed to create an overlapping region that included the 3' region of the gene for the first subunit, the linker DNA sequence (encoding for 7, 10 or 13 amino acids, respectively) and the 5' region of the second DNA segment in the engineered 'double-gene'. Specifically, primer A in reaction 1 was used for introducing the 5' overhang of the amplified insert for further assembly onto the expression vector YSBL-LIC(-), and primer B in reaction 1 was used for introducing a 3' overhang for further generating a portion of the linker coding region. Primer C in reaction 2 was used for

introducing a 5' overhang for generating a part of the linker coding sequence, and primer D used in reaction 2 was designed to introduce a vector specific 3' overhang to the insert for further assembly onto the expression vector YSBL-LIC (-) (Figure 2-1). Primers used in this section are listed in Appendix 1.

2.2.1.4 Target gene and linearized LIC vectors ligation

Ligation was performed based on the NEBuilder™ HiFi DNA assembly or In-Fusion® HD cloning (Clontech) protocol. Detailed reaction compositions for each commercial kit are given in Table 2-4. The reactions were incubated according to the commercial protocols.

Table 2-4 DNA assembly protocol for fusion proteins of phiKo gp14.

NEBuilder™ HiFi DNA Assembly protocol		**In-Fusion® HD Cloning protocol (Clontech)**	
10 µl reaction:		6 µl reaction:	
Linearized expression vector	50 -100 ng	Linearized expression vector	50 -100 ng
DNA fragment 1*	50 - 100 ng	DNA fragment 1*	50 - 100 ng
DNA fragment 2*	50 - 100 ng	DNA fragment 2*	50 - 100 ng
2 × NEBuilder HiFi DNA Assembly master mix	5 µl	6 × In-Fusion® HD Cloning master mix	1 µl
Recommended DNA Molar Ratio	Vector : insert = 1:2	Recommended DNA Molar Ratio	Vector : insert = 1:1

* DNA fragments 1 and 2, as well as the linearized plasmid YSBL-LIC (-) were used for generating the DNA for coding for the protein product that linked two gp14 proteins in one polypeptide chain. DNA fragments 1 and 2 corresponded to the products from reactions 1 and 2, respectively, in Figure 2-1.

2.2.1.5 Transformation

Before transformation, the commercial Stellar™ chemical competent cells (Clontech) were placed on ice to thaw. 3 µl of the product from DNA assembly reaction (in section 2.2.1.4) was transferred into 50 µl of thawed competent cells. The cell mixture was kept on ice for 20 min. The transformation reactions were incubated for 45 s at 42 °C for a 'heat shock' and then immediately placed on ice for 2 minutes. 240 µl SOC media (Clontech) was then added to the transformation reaction which was incubated at 37 °C with shaking (220 rpm) for 1.5 h. Approximate 300 µl of each reaction was plated on pre-incubated (at 37 °C) Luria-Bertani (LB) agar plates containing the appropriate antibiotic as necessary (30 µg/ml Kanamycin and/or 34 µg/ml Chloramphenicol or 100 µg/ml Ampicillin) and incubated at 37 °C overnight.

2.2.1.6 Recombinant plasmid validation by colony PCR

Single colonies were carefully picked to produce template DNA for colony PCR, and the remainder of the same colony on the original plate was kept for further use. YSBL-LIC vectors are derived from the pET28a expression vector, which is under control of a robust bacteriophage T7 promoter, whose transcription is induced by T7 RNA polymerase expressed in the host cell. Hence, a portion of the T7 promoter sequence was used as the forward primer and the target gene's (gp14 gene) reverse primer was used for verifying the recombinant plasmid by colony PCR (Table 2-5). Alternative primers for colony PCR were the target gene forward primer, and a portion of T7 transcription terminator sequence (sequences are given in Appendix 1). The single colony containing recombinant plasmid as defined by a PCR product of the correct size was picked and transferred into the 20 - 30 ml LB media supplemented with appropriate antibiotics and cultured at 37 ˚C overnight with 200 rpm shaking.

Table 2-5 Colony PCR protocols for recombinant plasmids of phiKo gp14.

PCR reaction composition		**PCR program**		
20 µl reaction:				cycles
		94 ˚C	60 s	1
DNA template	One single colony	94 ˚C	20 s	25
Forward and reverse primers (2 µM)	4 µl x 2	55 ˚C	30 s	25
DreamTaq DNA Polymerase (5 U/µL) (Thermo Scientific™)	12.5 µl	72 ˚C	100 s	25
dNTP (25 µM)	0.2 µl	72 ˚C	300 s	1
10x PCR reaction buffer (Thermo Scientific™)	2 µl	10 ˚C		hold
Ultra-pure water	Top up to 20 µl			

The overnight culture containing the recombinant plasmid was centrifuged to pellet the cells, and the supernatant was discarded. The cell pellets were then used for extraction and purification of the recombinant plasmids which were purified using the NucleoSpin™ plasmid purification Kit. Purified DNA was stored at -20 ˚C for future use. 10 µl of the purified plasmid (50 - 80 ng/µl) was used for further validation by sequencing (GATC Biotech) using the T7 and pRSET-RPnew primers. The successfully cloned recombinant plasmids with the correct sequences were transferred into expression cells via transformation as described above (in section 2.2.1.5), with slight modification depending on the properties of the cells and vectors used. Unless otherwise stated, the Rosetta™ (DE3) pLysS cells were used for protein expression in this study.

2.2.2 Protein expression

Rosetta™ host strains are *E. coli* BL21 derivatives designed to enhance the expression of eukaryotic proteins that contain rare codons in *E. coli*. In Rosetta™ (DE3) pLysS, the rare tRNA genes are present in the same plasmids that carry the T7 lysozyme gene. DE3 indicates that the host cell contains a lysogen of λDE3 and carries the T7 RNA polymerase gene under control of the *lacUV5* promoter. pLysS strains further suppress basal expression of T7 RNA polymerase before induction (156). When vectors containing the T7*lac* promoters are used in DE3 cells, the *lac* repressor acts both at the lac*UV5* promoter in the host chromosome to repress transcription of the T7 RNA polymerase gene by the host polymerase and at the T7*lac* promoter in the vector to block transcription of the target gene by any T7 RNA polymerase that is made. The transcription of the T7 RNA polymerase gene is controlled by the introduction of a lactose derivative, Isopropyl-*β*-D-thio-galactoside (IPTG). When IPTG is added to the culture, it binds to the *lac* repressor and releases the repressor from the *lac* operator and enables transcription of T7 RNA polymerase, which in turn binds to the T7 promoter on the vector and transcribes the cloned gene (157).

Freshly transformed colonies containing the recombinant expression plasmids were picked and transferred to fresh autoclaved LB media (400 ml), containing 30 μg/ml Kanamycin or 100 μg/ml Ampicillin and 34 μg/ml Chloramphenicol, and cultured overnight at 37 °C with shaking (200 rpm). The LB media was prepared by dissolving one LB powder capsule (Thermo Scientific™) into 1 L of deionised water. The next day, 60 ml of the overnight culture was transferred to 6 L freshly prepared, sterile LB media with corresponding antibiotics. The cell mixture was incubated at 37 °C with shaking (at 220 rpm) until the OD600 reached 0.6 - 0.8, then IPTG was added to the cell culture to a final concentration of 1 mM. Overnight expression of the protein was carried out at 16 °C with 200 rpm shaking. Induction optimisation tests suggested that expression at 16 °C overnight for each construct appeared to improve yield with less precipitation of the phiKo gp14 protein, compared to 37 °C (data not shown). The induction tests for pYM350 and pYM351 constructs in Rosetta™ (DE3) pLysS expression cells were performed by Owen Jarman in the Antson group during his undergraduate summer research project. The cell pellets were harvested by centrifugation in an Avanti J-26S high-speed centrifuge (Beckman) for 20 min at 5000 g, 4 °C.

2.2.2.1 Solubility test

Cell pellets from 1 ml of overnight IPTG induction culture were harvested and resuspended in 250 μl buffer. Results were resolved with SDS-PAGE by loading 15 μl of each sample onto a 15% polyacrylamide gel. Thirteen buffers were used for N-terminal non-cleavable 6 × His tagged protein ($gp14_{NNH1\text{-}203}$) solubility test, and buffer details are given in Appendix 2.

2.2.3 Protein purification strategies

In this study, nickel affinity chromatography, size exclusion chromatography and ion exchange chromatography were applied for protein purification. Some basic information of these approaches is discussed briefly in the sections 2.2.3.1 – 2.2.3.3 below.

2.2.3.1 Nickel affinity chromatography

Metal affinity chromatography is a method based on the interactions of a transition metal ion, such as nickel (Ni^{2+}), immobilised on a matrix and specific amino acid sidechains (158). Ni^{2+} is electrochemically stable under the experimental conditions used in protein chromatography (159). Histidine residues exhibit the most potent interaction with immobilised metal ion matrices, as electron donor groups on the histidine imidazole ring readily form coordination bonds with the immobilised transition metal. Affinity tag containing 6 × histidine is commonly used in immobilised nickel affinity chromatography and are efficiently retained on the column matrices. Following washing of the matrix material to remove contaminating proteins, polypeptides containing poly-histidine sequences can be eluted, as the main product, by adding free imidazole to the column (158). Imidazole is a strong unidentate ligand/ligate displacer and useful in desorption of proteins from the immobilised Ni^{2+} ions. Imidazole at low concentrations (0.001 - 0.03 M) is the eluting agent of choice for removing proteins that are weakly adsorbed to the Ni^{2+} ions, and it is also excellent as a protein displacer when used in either step-wise or gradient elution protocols (159).

2.2.3.2 Size exclusion chromatography

Size exclusion chromatography (also called gel filtration chromatography) is a technique for separating proteins and other biological macromolecules based on molecular size. The packed chromatography resin is equilibrated with a buffer which fills the pores of the matrix and the space between the particles. The liquid inside the pores, or stationary phase, is in equilibrium with the liquid (buffer or mobile phase) outside the particles. Results from size exclusion chromatography show the distribution in sample components as they are eluted from the column in order of their molecular sizes. Molecules that are larger than the largest pores in the matrix cannot enter the matrix and are eluted directly through the column (void volume). Molecules with partial access to the pores of the matrix are separated and eluted from the column in order of decreasing size (160).

2.2.3.3 Ion exchange chromatography

Ion exchange chromatography separates molecules based on differences in their surface net charge. Molecules vary considerably in their charge properties and will show different

interactions with charged chromatography media according to differences in their overall charge, charge density, and surface charge distribution. The reversible interactions between charged molecules and oppositely charged chromatography media are controlled in order to favour binding or elution of specific molecules and achieve separation. A protein has no net charge at a pH value equivalent to its isoelectric point (pI) and will not interact with the charged medium. However, at a pH above its pl, a protein will bind to positively charged medium, and conversely, will bind to negatively charged medium at pH values below its pI (161).

2.2.4 Protein purification for phiKo gp14 constructs

2.2.4.1 Fusion proteins with cleavable expression tag

In this thesis, six fusion constructs with cleavable expression tags, $gp14_{NCS1-203}$, $gp14_{NCH1-203}$, $gp14_{NCS1-203}$_K22A, $gp14_{NCS1-203}$_D100A, $gp14_{NCS1-203}$_R124A and $gp14_{NCG1-203}$ were made in an attempt to obtain a sufficient amount of soluble and stable phiKo gp14 protein that are suitable for structural and biochemical studies (see details of each construct in Table 2-6).

Table 2-6 Fusion proteins of phiKo gp14 with cleavable tags.

Fusion protein	Recombinant plasmid	Fusion tag
$gp14_{NCS1-203}$	pYM350	SUMO
$gp14_{NCH1-203}$	pYM409	6 × His
$gp14_{NCS1-203}$_K22A	pYM635	SUMO
$gp14_{NCS1-203}$_D100A	pYM636	SUMO
$gp14_{NCS1-203}$_R124A	pYM637	SUMO
$gp14_{NCG1-203}$	pYM638	GFP_TEV

Nickel affinity chromatography

Construct $gp14_{NCS1-203}$ is a fusion protein of phiKo gp14 with a cleavable N-terminal SUMO tag. The harvested cell pellets were resuspended in lysis buffer in a ratio of 5 g cell pellets to 50 ml buffer. Lysis buffer contains 50 mM Tris pH 7.5, 1 M NaCl, 10% glycerol and 5 mM Dithiothreitol (DTT), 100 μg/ml lysozyme and protease inhibitors (1 mM AEBSF and 0.7 μg/ml pepstatin). Resuspended cells were placed in a glass beaker and disrupted by sonication on ice to minimise thermal damage to the cell lysate. The cell debris was removed by fixed angle centrifugation at 18,000 rpm for 50 minutes at 4 °C using a Sorvall SS34 rotor in a Sorvall Evolution high-speed centrifuge (Beckman Coulter). The supernatant was collected and loaded onto a 5 ml HisTrap column (GE Healthcare) which was pre-equilibrated with the binding buffer for nickel affinity chromatography (50 mM Tris 7.5, 1 M NaCl, 10% glycerol,

20 mM imidazole and 1 mM DTT). The bound protein eluted with a linear gradient of 20 – 500 mM imidazole in the same binding buffer on an ÄKTA Fast protein liquid chromatography (FPLC) system (GE Healthcare).

The eluted fractions were verified as containing the protein of the correct size by SDS-PAGE in the SDS running buffer containing 25 mM Tris, 192 mM glycine and 1% SDS at 200 V for 60 min. 1 × loading buffer containing 10% Glycerol, 62.5 mM Tris-HCl pH 6.8, 2% SDS, 0.01% bromophenol blue and 1.25% beta-mercaptoethanol was added to the protein sample which was then heated at 95 $^{\circ}$C for 5 min before loading on to the gel. In addition, a standard protein marker was run alongside the samples of interest to determine the approximate molecular weights of the protein components of the sample.

SDS-PAGE

Polyacrylamide gel electrophoresis (PAGE) is a technique for separating molecules by size and charge. SDS is a detergent that denatures secondary and non-disulphide linked tertiary protein structures and covers them with negative charges. Therefore, the charge/mass ratio of proteins in the sample is uniform, and separation is mostly determined by their molecular size. The commonly used SDS-PAGE system was first described by Laemmli (162), and is also named as Laemmli SDS-PAGE system. Typically, the gel used in this system is discontinuous with an upper stacking gel and lower resolving gel that have different pH values and polyacrylamide concentrations. The upper stacking gel has a lower percentage of polyacrylamide and low pH buffer (~ 6.8) allowing proteins to move through quickly and 'stack' into a tight band before entering into the resolving gel which contains a higher percentage of polyacrylamide and higher buffer pH (~ 8.8) for separation. Glycine, in gel running buffer (pH ~ 8.3), can exist in three different charge states, positive, neutral or negative, depending on the pH; therefore, control of the charge state of the glycine by the different buffers is the key to the PAGE. When the gel is running, the negatively charged glycine in the pH ~ 8.3 running buffer are forced to enter the stacking gel (pH ~ 6.8) and switches predominantly to the neutral state. The Cl^- ions in the stacking gel, move much faster and migrate ahead of the glycine. All of the proteins in the gel sample have electrophoretic mobility that is intermediate between the mobility of the glycine and Cl^-, when the two fronts, glycine and Cl^-, sweep through the sample well, the proteins are concentrated into the narrow zone between the Cl^- and glycine fronts. When this front reaches the resolving gel (pH ~ 8.8), the glycine molecules are mostly negatively charged and can migrate much faster than the proteins. Thus, the glycine front accelerates past the proteins. The proteins are then 'stacked' in a very narrow band at the interface of the stacking and resolving gels. The resolving gel has a higher polyacrylamide concentration for the separation of the protein components (163). In addition, the concentrations of polyacrylamide in the resolving gel can be optimised for the

size range of molecules present in the sample. Small proteins will move through the resolving gel faster than large proteins (164).

The relatively pure fractions with protein bands of the approximately correct size were pooled and dialysed against the buffer of 50 mM Tris 7.5, 1 M NaCl, 10% glycerol and 1 mM DTT to remove imidazole. The sample (in low or no imidazole buffer) was transferred into 50 ml falcon tubes. 5 mM DTT and SUMO protease (in a 1:50 mass ratio of SUMO protease to $gp14_{NCS1\text{-}203}$ protein) were added to the collected fractions and mixed gently and thoroughly. The protein mixture was incubated at room temperature overnight for SUMO protease digestion, and the digested product was verified by SDS-PAGE using a 15% polyacrylamide gel. The successfully digested protein was re-purified over a 5 ml HisTrap column which was pre-equilibrated with the same binding buffer (described above). The unbound protein fractions were collected as the target protein with the tag removed and were checked by SDS-PAGE. The pure fractions were concentrated to a small volume (0.5 ml - 5 ml) by centrifugation of samples at 4500 rpm, 4 °C using Vivaspin® 20 concentrators (MWCO 10,000, Sartorius) or by PEG solution concentrating using a 10 kDa cut-off dialysis membrane with slow stirring in a 50% solution of PEG 35000 in the buffer containing 20 mM Tris 7.5, 1 M NaCl, 10% glycerol and 1 mM DTT.

Size exclusion chromatography (SEC)

In order to obtain high purity and homogeneous protein as a standard requirement for crystallisation, size exclusion chromatography (SEC) was performed. The concentrated protein of interest (in a volume of ~ 3 - 5 ml) was loaded onto a HiLoad 16/600 Superdex 75 prep grade column (GE Healthcare). The column was pre-equilibrated with SEC buffer containing 20 mM Tris 7.5, 1 M NaCl, 10% glycerol, 0.2 mM DDM and 1 mM DTT. The peak fractions were collected and validified by SDS-PAGE. The fractions containing a single band of the correct molecular weight were pooled and concentrated for immediate use in the following experiments or stored at - 80 °C after flash freezing in liquid nitrogen for future use.

$Gp14_{NCH1\text{-}203}$ comprises an N-terminal cleavable 6 × His tag which can be digested by 3C protease. In this study, 1:50 mass ratio of 3C protease to protein sample was used for protein digestion. The fusion protein $gp14_{NCH1\text{-}203}$ was purified similarly to $gp14_{NCS1\text{-}203}$, but different buffers were used for purification. The nickel affinity binding and SEC buffers were made according to the properties of the $gp14_{NCH1\text{-}203}$ protein, as detailed in Appendix 3. In addition, the N-terminal cleavable SUMO tagged fusion constructs of phiKo gp14 protein mutants, $gp14_{NCS1\text{-}203}$_K22A, $gp14_{NCS1\text{-}203}$_D100A and $gp14_{NCS1\text{-}203}$_R124A, were purified similarly to $gp14_{NCS1\text{-}203}$. Buffers for purification of each construct were optimised according to their intrinsic properties and are listed in Appendix 3.

2.2.4.2 Fusion proteins with non-cleavable tags

In this study, six fusion constructs with non-cleavable tags, $gp14_{NNH1-203}$, $gp14_{NNH7-203}$, $gp14_{CNH7-203}$, $gp14_{NNHL7}$, $gp14_{NNHL13}$, $gp14_{NNHL10}$, $gp14_{CNG1-203}$, were made in an attempt to harvest soluble phiKo gp14 protein and to investigate the oligomerization state. Details of each construct are given in Table 2-7.

Table 2-7 Fusion proteins of phiKo gp14 with non-cleavable tags.

Fusion protein	Recombinant plasmid	Fusion tag
$gp14_{NNH1-203}$	pYM351	6 × His
$gp14_{NNH7-203}$	pYM418	6 × His
$gp14_{CNH7-203}$	pYM419	6 × His
$gp14_{NNHL7}$	pYM537	6 × His
$gp14_{NNHL13}$	pYM546	6 × His
$gp14_{NNHL10}$	pYM569	6 × His

$Gp14_{NNH1-203}$ is the fusion construct of phiKo gp14 ATPase with a non-cleavable N-terminal 6 × His tag. The harvested cell pellets were resuspended in lysis buffer in a ratio of 25 g cell pellets to 200 ml buffer. The protein of $gp14_{NNH1-203}$ is purified similarly to the $gp14_{NCS1-203}$ for the first nickel affinity chromatography, as described above (in section 2.2.4.1), and purification buffers used for $gp14_{NNH1-203}$ are given in Appendix 3. When purifying the protein sample for ATPase activity and DNA binding assays, the lysis and binding buffers were made without $MgCl_2$. The eluted fractions were verified by SDS-PAGE and were further purified by size exclusion chromatography in a similar protocol described above for $gp14_{NCS1-203}$ in section 2.2.4.1, using slightly different SEC buffer (Appendix 3).

The other five protein constructs with N-terminal non-cleavable 6 × His-tag were purified using a similar purification strategy as described above, while the purification buffers were adjusted according to the properties of each protein constructs. Buffers for purifying each construct are listed in Appendix 3.

2.2.5 Crystallisation screens and optimisations

Since protein crystallisation requires relatively large amounts of protein with high concentration and purity, only the construct of $gp14_{NNH1-203}$, phiKo gp14 with an N-terminal non-cleavable 6 × His tag fusion, most satisfied these requirements (for details in section 2.3.3). This protein construct ($gp14_{NNH1-203}$) was used for crystallisation in this study.

The sitting-drop vapour diffusion method (this method is detailed in section 1.4.1 above) was used for initial crystallisation screens. The screens were performed using 96-well SWISSCI®

MRC 2-drop crystallisation plates. Droplets were prepared by mixing 150 nl protein solution with 150 nl of reservoir solution (or 300 nl protein solution and 150 nl reservoir solution for optimisation) using the Mosquito® liquid handling instrument (TTP Labtech). The crystallisation drops were equilibrated against 54 µl of reservoir solution at 20 °C. The initial crystal trays were performed using commercial screens: Index HT (Hampton Research); PACT Premier HT (Molecular Dimensions); and MPD (Qiagen). Further optimisation was performed to generate single well-diffracting crystals of > 10 µm in size. Unliganded protein sample was in a buffer solution containing 20 mM Tris 8.0, 0.5 M NaCl, 5% glycerol, 20 mM $MgCl_2$ and 2 mM DTT. The complexes of phiKo gp14 with ADP, AMPPNP and AMPPCP were formed by adding 10 mM ADP, AMPPNP and AMPPCP, respectively, to 2.2 - 2.5 mg/ml phiKo gp14 (in the same buffer) and incubating at room temperature for at least 20 min prior to crystallisation. Crystals of complexes of phiKo gp14-phosphate, gp14-ADP, gp14-AMPPNP and gp14-AMPPCP grew in similar optimised conditions containing 0.1 M citric acid pH 4.5 or 5.0 and 6% - 15% MPD. Subsequently, in this study, these crystallisation conditions were optimised in 96-well plates with optimal reservoir solutions. Crystals were tested using a Rigaku RU-H3R X-ray generator with a rotating anode, equipped with Osmic multilayer optics and a MAR345 (Mar Research) imaging-plate detector.

2.2.6 X-ray data collection, structure determination and model building

X-ray data of the well diffracting crystals were collected using synchrotron radiation at the I02 beamline at the Diamond light source, Didcot. Sam Hart and Dr. Johan Turkenburg (University of York) assisted in X-ray data collected from the well-diffracting crystals. Data of phiKo gp14-phosphate, gp14-ADP, gp14-AMPPNP and gp14-AMPPCP were processed using XDS (165). The initial model of phiKo gp14 was solved by molecular replacement using Phenix.mr-rosetta (166) and Phaser (version 2.6.0)(167) and was re-built in SHELXE (168) and ARP/wARP (169). Molecular models of phiKo gp14 were manually built in Coot (170) and refined by Refmac5 (171) using the CCP4i2 suite (172). Structures of phiKo gp14 complexes with ADP, AMPPNP and AMPPCP were refined using phiKo gp14-phosphate as the initial model, respectively. Unless otherwise mentioned, figures for structures in this thesis were generated using CCP4MG (173) or Chimera (174).

2.2.7 Nano differential scanning fluorimetry (NanoDSF)

Thermal unfolding curves, utilising the intrinsic tryptophan and tyrosine fluorescence of the phiKo gp14 protein in different buffer conditions, were measured by Nano differential scanning fluorimetry (NanoDSF). The basis of NanoDSF is that changes in polarity of the microenvironment around the aromatic amino acids (predominantly tryptophan), and protein unfolding leads to a difference in the average emission wavelength of the intrinsic fluorescence

of the aromatic amino acids. Typically, a shift occurs from 330 nm to 350 nm in a protein indicating that the local environment of the tryptophan changes from fully buried to completely exposed to water or solvent, resulting in an increase in the ratio of fluorescence intensities at 350 nm and 330 nm ($F_{350/330}$). Generally, the $F_{350/330}$ yields data with well-defined transitions on protein unfolding (175).

10 µl soluble proteins of the untagged gp14 ($gp14_{NCS1-203}$, the gp14 protein with a native N-terminus by cleaving SUMO tag) were placed in the standard capillaries (NanoTemper Technologies). A temperature gradient of 1 °C/min from 25 to 95 °C was applied for the measurements, and the intrinsic protein fluorescence emission at 330 and 350 nm was recorded using the Prometheus NT.48 instrument (NanoTemper Technologies). Data were analysed by the PR.ThermControl software to determine the thermal unfolding transition midpoint or melting temperature (T_m, °C) and interpreted by OriginPro 2017 (OriginLab, Northampton, MA). Each assay was performed with two replicates.

2.2.8 Malachite green ATPase activity assay

The Malachite Green ATPase activity assay is a sensitive spectrophotometric measurement for detecting inorganic phosphate produced by ATP hydrolysis, employing the reaction of phosphomolybdate with the malachite green to form a coloured complex (176), as seen in Figure 2-2. Malachite green is a basic dye belonging to the triphenylmethane group. The colour formation from the reaction can be measured on a spectrophotometer (at 600 - 660 nm).

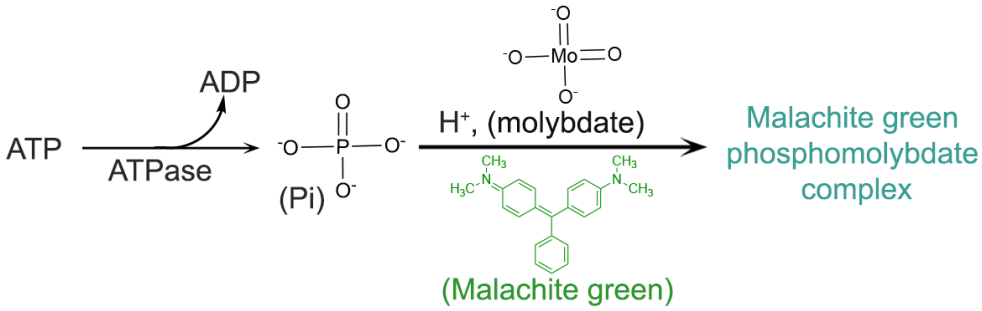


Figure 2-2 Schematic diagram of the Malachite green reaction.
ATP hydrolysed by ATPase, i.e. gp14 in this thesis. Pi indicates the inorganic phosphate. The reaction product, malachite green phosphomolybdate complex, absorbs strongly at a wavelength of 600 – 660 nm.

To compare the ATPase hydrolysis rate of each construct of phiKo gp14, the soluble fraction of untagged phiKo gp14 ($gp14_{NCS1-203}$ with a native N-terminus by cleaving SUMO tag), untagged phiKo gp14 with the post-cleavage G-P-A residues at the N-terminus ($gp14_{NCH1-203}$) and N-terminal 6 × His tagged phiKo gp14 ($gp14_{NNH1-203}$) were used for the ATPase activity

assay. Final concentrations of 0.5 μM, 1 μM and 2 μM of each protein were used for the measurements. Reactions were performed in a buffer containing 20 mM Tris 8.0, 250 mM NaCl, 5% glycerol, 0.2 mM DDM, 5 mM magnesium acetate, 1 μM TCEP and 5 μg/ml BSA were initiated by adding 0.4 mM ATP. Reactions were incubated at 50 ˚C for 30 min. Each sample has duplex replicates.

To probe the factors that can affect gp14 ATPase activity, 2 μM untagged phiKo gp14 ($gp14_{NCS1-203}$ with a native N-terminus by cleaving the SUMO tag) was used for the following assays, due to it being relatively more active compared to the other two constructs: (i) For the assays of the effect of ATP concentration on ATPase activity, reactions were performed in a buffer containing 10 mM citric acid pH 5.0, 5% glycerol, 100 mM NaCl, 0.2 mM DDM, 5 mM magnesium acetate, 1 μM TCEP and 5 μg/ml BSA and were initiated by adding ATP and incubated at 50 ˚C for 30 min; (ii) To assay the effect of pH on the ATPase activity, the applied reaction conditions containing 10 mM buffer (pH 5 - 9), 250 mM NaCl, 5% glycerol, 0.2 mM DDM, 5 mM magnesium acetate, 1 μM TCEP and 5 μg/ml BSA. Reactions were initiated by adding 0.5 mM ATP and incubated at 70 ˚C for 1 h; (iii) For the assays on the effect of reaction temperature on the ATPase activity, the performed reaction conditions containing 10 mM Tris 8.0, 100 mM NaCl, 5% glycerol, 0.2 mM DDM, 5 mM magnesium acetate, 1 μM TCEP and 5 μg/ml BSA. The reaction was initiated by adding 0.4 mM ATP and was incubated at 50, 60 or 70 ˚C for 1 h; (iv) To determine the effect of NaCl concentration (from 100 mM to 550 mM) on ATPase activity, reactions containing 10 mM Tris 7.5, 5% glycerol, 0.2 mM DDM, 5 mM magnesium acetate, 1 μM TCEP and 5 μg/ml BSA as well as different concentrations of NaCl, were utilised. The reactions were initiated by adding 0.4 mM ATP and were incubated at 70 ˚C for 1 h. (v) To assay the effects of reaction/incubation time on ATPase activity, reactions were performed in a buffer of 10 mM Tris 7.5, 1 μM TCEP, 50 mM sodium citrate, 5% glycerol, 5 mM magnesium acetate, 0.2 mM DDM and 5 μg/ml BSA. The sample reactions were incubated at 50 ˚C for 10 min, 20 min, 30 min and 1 h, respectively. Each sample has duplex replicates, except for the incubation time effects assay has triplex replicates.

All assays were performed using colour reagent from ATPase/GTPase Activity Assay Kit (Sigma) on Corning® 96-well flat-bottom plates. KH_2PO_4 was used to generate an inorganic phosphate (Pi) concentration standard curve (Appendix 5). Data (absorbance at the wavelength of 620 nm, A620) were recorded at room temperature in an Epoch Microplate Spectrophotometer (BioTek).

2.2.9 Electrophoretic mobility shift assay (EMSA)

Electrophoretic mobility shift assay (EMSA) is a common affinity electrophoresis technique under non-denaturing (native) conditions for the separation of a protein-nucleic acid mixture (177). The EMSA technique is based on the observation that protein–nucleic acid complexes migrate more slowly than free linear DNA fragments when subject to non-denaturing polyacrylamide or agarose gel electrophoresis. Because the DNA migration rate is shifted or retarded when bound to protein, the assay is also referred to as a gel shift assay. EMSA buffers do not contain the denaturing anionic detergent component SDS; hence, EMSA separates proteins according to both their intrinsic charge and volume. The relatively low ionic strength of the electrophoresis buffer helps to stabilise transient interactions, permitting even labile complexes to be resolved and analysed by this method (177).

The binding of dsDNA to the N-terminal 6 × His tagged phiKo gp14 protein ($gp14_{NNH1\text{-}203}$) was probed by mixing 20 nM 30 bp Cy5-labelled dsDNA, with a series of protein dilutions, 1 mM ATPγS, 5 mM $MgCl_2$ in reaction buffer containing 10 mM Tris 7.5, 50 mM NaCl, 0.2 mM DDM and 0.1 mM TCEP. The reactions were incubated at 37 ˚C for 1 h. Results were resolved on a 0.6% agarose gel.

The DNA binding of the untagged phiKo gp14 ($gp14_{NCS1\text{-}203}$) protein was performed by mixing 10 nM 30 bp Cy5-labelled dsDNA, with a series of protein dilutions, 1 mM ATPγS, 20 mM Magnesium acetate, 5 mM sodium citrate in the reaction in the buffer containing 10 mM Tris 7.5, 166 mM NaCl, 0.2 mM DDM and 0.1 mM TCEP. The reaction was incubated at 50 ˚C for 0.5 h. Results were analysed on a 2.5% native polyacrylamide gel.

The DNA binding of 3C-cleaved phiKo gp14 that contained the remaining G-P-A residues at the N-terminus ($gp14_{NCH1\text{-}203}$) was performed by mixing 20 nM 30 bp Cy5-labelled dsDNA, and protein dilutions in the reaction buffer containing 10 mM Tris 8.0, 50 mM $NaSO_4$, 5 mM $MgSO_4$, 1 mM DDM and 1 mM DTT. Each reaction was incubated at 37 ˚C for 0.5 h. Results were separated on a 2.5% native polyacrylamide gel.

The dsDNA conjugated with a 5’ Cy5 fluorophore (excitation/emission at maximum wavelength of 650/670 nm) was ordered from Eurofins Genomics Germany GmbH. The sequences of assayed dsDNA are given in Appendix 6. 10 µl of each sample was mixed with 5% glycerol before loading on to the native gel. In this study, either 1 × TBE buffer containing 89 mM Tris pH 7.6, 89 mm Boric acid and 2 mM EDTA or 1 × THE buffer containing 15 mM Tris pH 7.5, 50 mM HEPES pH 7.5, 6.7 mM magnesium acetate and 2.2 mM tri-sodium citrate was used as the running buffer. The 0.6% agarose gel was made by dissolving 0.6 g agarose in 100 ml 1 × gel running buffer. The 2.5% native polyacrylamide gel was made by mixing

2.5 ml 30% (100:1) acrylamide/Bis-acrylamide, 7.5 ml 1 × running buffer, 80 µl 10% APS and 10 µl TEMED. Agarose gel was run at 100 V in a cold room (4 °C) for 60 min, and native polyacrylamide gel was run at 100 V, for 2.5 h at room temperature or 4 °C. Gels were visualised at an excitation wavelength of 635/650 nm in a Typhoon™ 8610 scanner (York Bioscience Technology Facility).

2.2.10 Phage isolation, purification and DNA packaging assay

2.2.10.1 Growth of the *Thermus thermophilus* phage phiKo

TB media was prepared for phiKo phage growth and purification. The TB media was made by dissolving 0.8% (w/v) Tryptone, 0.4%(w/v) NaCl, 0.2% (w/v) yeast extract, 0.5 mM $MgCl_2$ and 0.5 mM $CaCl_2$ in Evian® mineral water which contains essential mineral for growth. Several cell colonies of *Thermus thermophilus* HB 27 (HB27 cells in the following text) were picked and transferred to 5 ml freshly prepared TB media. The HB 27 cells were incubated with 180 rpm shaking overnight at 70 °C. The next day, 4 ml of the overnight culture of HB27 cells were transferred to 400 ml freshly prepared TB media, a 1:100 ratio for scaling up the culture. The cell culture was incubated at 70 °C with shaking (200 rpm) until the OD600 reached 0.2 - 0.3 as measured in a DU 720 UV/Visible Spectrophotometer (Beckman Coulter). Then, 100 µl of the phage stock (with titre at 10^8 *pfu*/ml) or 1 - 3 single plaques of phage phiKo was added to the 400 ml HB27 cell culture (at OD600 of ~ 0.25) as the starting point (0 h). The culture reached the highest OD600 around 3 hours post infection when cell lysis began to occur. About 5 h after injection of the phage, the cell culture was fully lysed. The OD600 of the phage culture was measured every hour during phage growth at 70 °C with 200 rpm shaking. The amount of phage in the fully lysed cell culture was estimated using the phage plaque assay.

Phage plaque assay

Phage plaque assays were used for determining the quantity of the harvested phage or phage stock. The solution of phage phiKo was diluted 10^3 - 10^9-fold using autoclaved TB media. A 50 µl aliquot of HB27 cells overnight culture was added to 5 ml freshly prepared and pre-warmed (at 70 °C) top TB media containing 0.8% agar, and then 100 µl of each phage dilution was added to each top TB agar media. Each sample was mixed thoroughly by vortex and poured onto a pre-warmed bottom TB agar plate (containing 2% agar). Once the top agar had solidified, the plates were incubated at 70 °C overnight. The next day, the phage plaques were counted, and the titre (in plaque forming units, *pfu*) were calculated according to the equation as follow:

$$phage\ titre\ (pfu/ml) = \frac{Number\ of\ plaques\ on\ a\ plate \times dilutions}{Volume\ of\ diluted\ phage\ added\ to\ a\ plate}$$

2.2.10.2 Phage isolation

Once the cells fully lysed and the cell debris was removed by centrifugation at 8,000 rpm at 8 °C for 20 min, the supernatant was collected and subjected to differential ultracentrifugation at 30,000 rpm at 8 °C for 1.5 - 2 h using the Ti45 rotor (Beckman Coulter). Pellets at the bottom containing phage were resuspended in a buffer of 50 mM sodium phosphate pH 7.2, 50 mM NaCl, 10 mM $MgCl_2$ and 2 mM DDM by gentle agitation. The resuspended phage solution was placed at 4 °C for further purification. The amount of isolated phage was estimated by phage plaque assay or by SDS-PAGE.

2.2.10.3 Phage purification

The isolated phage phiKo sample was purified by density gradient ultracentrifugation followed by size exclusion chromatography, as described in this section.

Density gradient ultracentrifugation

The resuspended phage sample was first purified by rate zonal density gradient ultracentrifugation. In rate zonal centrifugation, particles are separated in terms of their distinct sedimentation velocities in the liquid gradient medium, reflected by their sedimentation coefficients:

$$v = \frac{2r_s^2[\rho_s - \rho_1]}{9\eta} F$$

where ρ_s is the particle density and ρ_1 is the liquid density, η is the viscosity of the liquid medium, and $F = \omega^2 r$, ω is the rotor speed (s^{-1}) and r_s is the distance between the particle and the centre of rotation. The sample solution is layered on top of the density gradient medium. Under centrifugal force, the particles will begin to sediment through the gradient in distinct zones (178).

A continuous density gradient of either 10 - 30% or 40% glycerol in buffer containing 50 mM sodium phosphate pH 7.2, 150 mM NaCl and 1 mM $MgCl_2$ was prepared using the Gradient Master (BioComp). 1 - 2 ml of the isolated phage sample was loaded at the top of the pre-cooled density gradient column. Rate zonal ultracentrifugation was performed at 24,000 rpm, 10 °C for 2 h using the SW28 rotor (Beckman Coulter). The fractionated sample fractions (zones) were collected separately and the content determined by SDS-PAGE (corresponding to the capsid protein molecular weight) and the relevant samples reserved for additional purification by size exclusion chromatography.

Size exclusion chromatography

The collected fractions containing phages from the density gradient ultracentrifugation were further purified by size exclusion chromatography. Up to 5 ml of the phage solution was loaded on to a HiPrep Sephacryl S-500 HR column (GE Healthcare) pre-equilibrated with buffer containing 50 mM sodium phosphate pH 7.2, 150 mM NaCl and 1 mM $MgCl_2$. Eluted fractions were resolved by SDS-PAGE, the fractions containing the approximately correct size (corresponding to the phiKo capsid protein) were further analysed by negative stain transmission electron microscopy (TEM). The purified phage sample was concentrated by differential ultracentrifugation using Vivaspin® centrifugal concentrators (MWCO 100 kDa, Sartorius) or by adsorption against solid PEG 35,000. Aliquots of the concentrated sample were used immediately for DNA packaging assays or supplemented with 15% (v/v) glycerol and stored at - 80 °C after flash freezing in liquid nitrogen.

2.2.10.4 Negative stain electron microscopy

Surface cleaning of carbon coated copper grids (Agar Scientific™) was performed by glow discharging for 20 s in a plasma cleaner (Harrick Plasma). 5 µl of phage solution was deposited onto the prepared grids and incubated for 5 min. The grids were washed with ultrapure water 3 times and stained using 5 µl of 5% (w/v) uranyl acetate. The stained grids were dried at room temperature and imaged on a Tecnai™ 12 Transmission Electron Microscope (York Bioscience Technology Facility).

2.2.10.5 Isolation of phage genomic DNA

The phage isolation was performed in the same way as described in section 2.2.10.2. The pellets containing phage were collected and resuspended in the buffer of 20 mM potassium phosphate pH 7.2, 10 mM $MgCl_2$ and 0.2 mM DDM. 2 µg/ml DNAase I (Sigma) was added to the resuspended phage solution, and the mixture was incubated at 37°C for 1 h to remove contaminating host DNA. Then, 50 mM EDTA was added to the sample mixture and the reaction was incubated at 75 °C for 10 min to inactivate the DNAase I. The reaction mixture was then incubated at 37 °C for 60 min with 0.2 vol. of 10% SDS added to disrupt phage capsids. The genomic DNA was extracted with 1 vol. of phenol/chloroform/isoamyl alcohol (pH 8.0) solution four times (spinning at 13,300 rpm for 10 min each time). Finally, the isolated phage genomic DNA was precipitated from the aqueous phase by the addition of 0.1 vol. of 3 M sodium acetate and 2 vol. of 100% ethanol pre-cooled on ice. The pelleted DNA was resuspended in TE buffer (10 mM Tris pH 7.5 and 0.1 mM EDTA, 4 mM NaCl and 5% glycerol) and stored in aliquots at 4°C or - 20 °C for future use. The 2-Log DNA ladder (New England Biolabs, Inc.) comprising fragments ranging from 100 bp to 10 kb was used for the gel running. The isolated genomic DNA sample was verified by electrophoresis using a 0.8%

agarose gel, and the gel was stained by SYBR™ safe dye and imaged by GelDoc system (Bio-Rad).

2.2.10.6 DNA packaging assay

The composition of the DNA packaging assay is given in Table 2-8. The assays were performed in 20 µl reaction volumes containing 3% PEG 8,000, 10 mM Mg Cl_2, 10 - 100 mM NaCl, 20 mM Bis-Tris pH 7.0 or Tris pH 7.2, 0.1 mM TCEP and 2 mM ATP. The purified empty capsids of phage phiKo (in section 2.2.10.3) were dialysed against buffer containing 20 mM Bis-Tris pH 7.0 or Tris pH 7.2 and 50 mM NaCl prior to the assay. After adding all the reaction components without ATP, the reaction was first incubated at room temperature for 5 min, before being initiated by adding 2 mM ATP and further incubated at room temperature, 37 °C, 50 °C or 70 °C for 20 min. Subsequently, 1 µg/ml DNAase I (Sigma) was added to the reaction, and then incubated at room temperature or 37 °C for 20 - 30 min to remove unpackaged DNA. The reaction was terminated by adding 25 mM EDTA and incubation at 70 °C for 15 min to inactive the DNAase I. The protein shell was removed by addition of 5 mM $CaCl_2$ and 0.5 mg/ml proteinase K (Promega) to the reaction and further incubation at 42 °C or 50 °C for 30 - 60 min. Finally, 1 µl of 100 mM AEBSF was added to the reaction mixture to inactive the protease K before the reactions were resolved by electrophoresis using 0.6% agarose gel visualised with SYBR™ Gold gel stain and imaged by GelDoc system (Bio-Rad).

Table 2-8 Composition of DNA packaging assay for phiKo gp14.

Components	1	2	3	4	5	6	7	8
gDNA (ng/µl)	4	0	0	4	4	4	4	4
capsids (A280 = 0.128) (µl)	0	2.5	5	2.5	5	2.5	5	0
phiKo gp14 (µM)	0	0	0	0	0	0.5	0.5	0.5

The purified N-terminal 6 × His tag fusion phiKo gp14 ($gp14_{NNH1-203}$) was used in these reactions.

2.3 Results

2.3.1 The phiKo gp14 protein possesses low solubility

The recombinant protein constructs were analysed for expression in a variety of *E. coli* expression strains, and the protein gp14 was expressed best in Rosetta™ (DE3) pLysS cells compared to others (data not shown). Solubility tests determined that the construct with the

N-terminal non-cleavable 6 × His tagged phiKo gp14 fusion ($gp14_{NNH1-203}$) expressed in Rosetta™ (DE3) pLysS cells exhibited poor solubility in a series of buffers (Figure 2-3). Likewise, the rest of the constructs of phiKo gp14 displayed a small but promising amount of soluble material (data not shown). The 6 × His tagged gp14 protein ($gp14_{NNH1-203}$) comprises 213 amino acids and has a molecular weight of 24.8 kDa, as calculated by ExPASy ProtParam (179). The solubility test was performed in 16 buffers (Appendix 2). The crude cell extracts from 1 ml overnight induced culture were sonicated in each buffer and were resolved by SDS-PAGE (Figure 2-3). No significant differences were observed among the tested 16 buffers, and a considerable amount of the protein appeared in the insoluble fractions (Figure 2-3). Buffers 1 and 16 were chosen for further purification of the 6 × His tagged phiKo gp14 ($gp14_{NNH1-203}$), due to the slightly better yield of soluble materials obtained in these buffers.

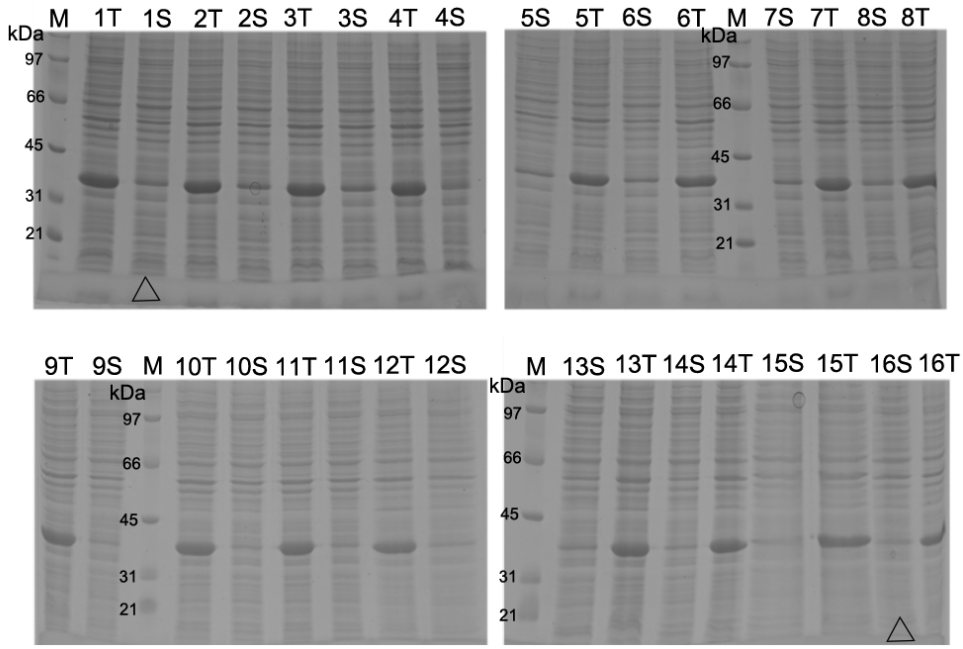


Figure 2-3 Solubility test for the 6 × His tagged phiKo gp14 ($gp14_{NNH1-203}$).

'M' is the standard protein marker, 'T' indicates total cell lysate containing both soluble and insoluble protein, and S indicates the soluble protein fraction. Numbers 1-16 designate the 1-16 buffers with compositions are given in Appendix 3. Lanes containing samples in buffer 1 and 16 are highlighted with triangles on the figure.

2.3.2 Protein purification

2.3.2.1 PhiKo gp14 fusions with cleavable tags

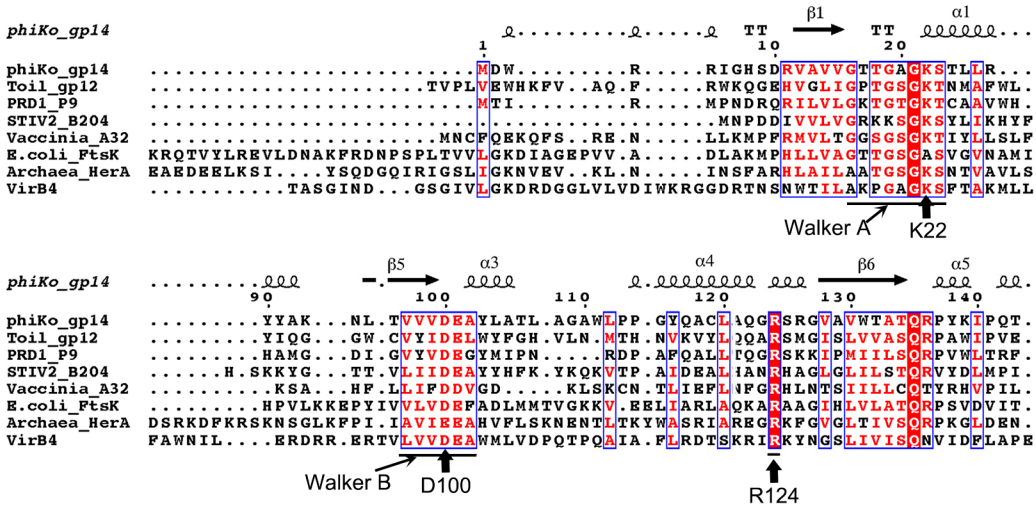


Figure 2-4 Sequence alignment of several FtsK/HerA family ATPases.

Tectiviridae family phage DNA packaging ATPases phiKo gp14, Toil gp12 (180) and PRD1 P9 (181); *Turriviridae* phage Sulfolobus turreted icosahedral virus 2 (STIV2) B204 (182); Vaccinia virus A32 (183) and *Escherichia coli* FtsK (72) were used for sequence alignment. The sequence alignment was performed with Clustal Omega (184), and the figure was produced in ESPript (185). Residues K22, D100 and R124 are highlighted with black arrows on the figure.

Five constructs of phiKo gp14 were fused to N-terminal cleavable tags, i.e. SUMO tag and 6 × His tag (protein details are given in Table 2-6). The $gp14_{NCS1\text{-}203}$ and $gp14_{NCH1\text{-}203}$ are wild type proteins, and $gp14_{NCS1\text{-}203}$_K22A, $gp14_{NCS1\text{-}203}$_D100A and $gp14_{NCS1\text{-}203}$_R124A are mutants, respectively. As seen from sequence alignment (Figure 2-4), the conserved K22 residue is part of the Walker A motif, the conserved D100 is part of the Walker B motif, and R124 is predicted as the 'arginine finger' (68, 72).

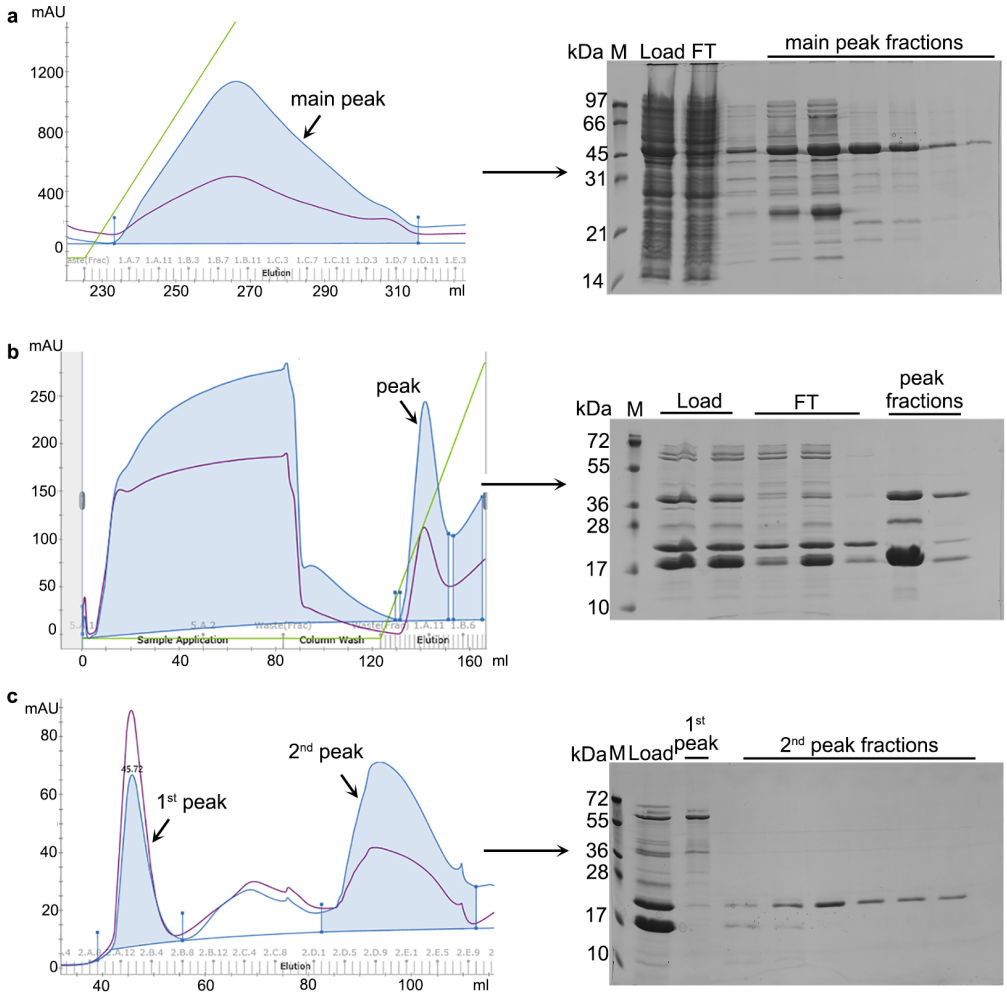


Figure 2-5 Nickel affinity and size exclusion chromatography of gp14$_{NCS1-203}$.

a. Purification of gp14$_{NCS1-203}$ by nickel affinity chromatography; b. Purification of SUMO protease digested gp14$_{NCS1-203}$ by nickel affinity chromatography; c. Purification of gp14$_{NCS1-203}$ by size exclusion chromatography. The eluted peak fractions are highlighted on the chromatograms on the left, and the SDS-PAGE results are shown on the right. 'M' indicates a lane containing protein size marker and 'FT' indicates the flow through fractions. 'Load' is the loaded sample. The blue lines are the sample absorbance at 280 nm (A280), and the purple lines are the absorbance at 254 nm (A254). The green lines in panels a and b indicate the percentage of the elution buffer containing 500 mM imidazole.

The size of each SUMO tagged phiKo gp14 fusion (gp14$_{NCS1-203}$, gp14$_{NCS1-203}$_K22A, gp14$_{NCS1-203}$_D100A and gp14$_{NCS1-203}$_R124A, Table 2-6) is ~ 36.9 kDa (protein sequences are given in Appendix 7). The target protein was purified by Ni-affinity chromatography and was eluted by 30 - 40% imidazole (500 mM) in the binding buffer (Figure 2-5a) that was collected for SUMO protease digestion. The digested product was further purified using a nickel affinity

column, where the target protein was eluted in the flow through fractions (FT) (Figure 2-5b). Due to the poor solubility, the protein was concentrated against a solution of PEG 35,000, a gentle concentrating method, at room temperature, overnight. Following this, the target protein was centrifuged to remove aggregated protein and further purified using a size exclusion column, due to the low purity of the samples collected from the second round of nickel affinity chromatography (Figure 2-5b). Two major peaks eluted from the size exclusion column. Protein contaminants were eluted in the first peak fractions, while the second peak fractions contain protein of approximately the correct size (Figure 2-5c). The additional band (at ~ 17 kDa) was possibly the digested SUMO tag protein. The purified tag-cleaved phiKo gp14 protein (with a molecular size of 23.6 kDa) eluted as a monomer from the size exclusion column (Figure 2-5c). Fractions from the second peak, which eluted at a volume of 90 - 115 ml and contained protein corresponding to a molecular mass of ~ 20 kDa, were pooled together. Notably, the SUMO tagged phiKo gp14 fusion proteins were quite soluble. However, after cleaving the SUMO tag, the protein became much less soluble, and could only be concentrated to ~ 50 µM in the SEC buffer (buffer details are given in Appendix 3).

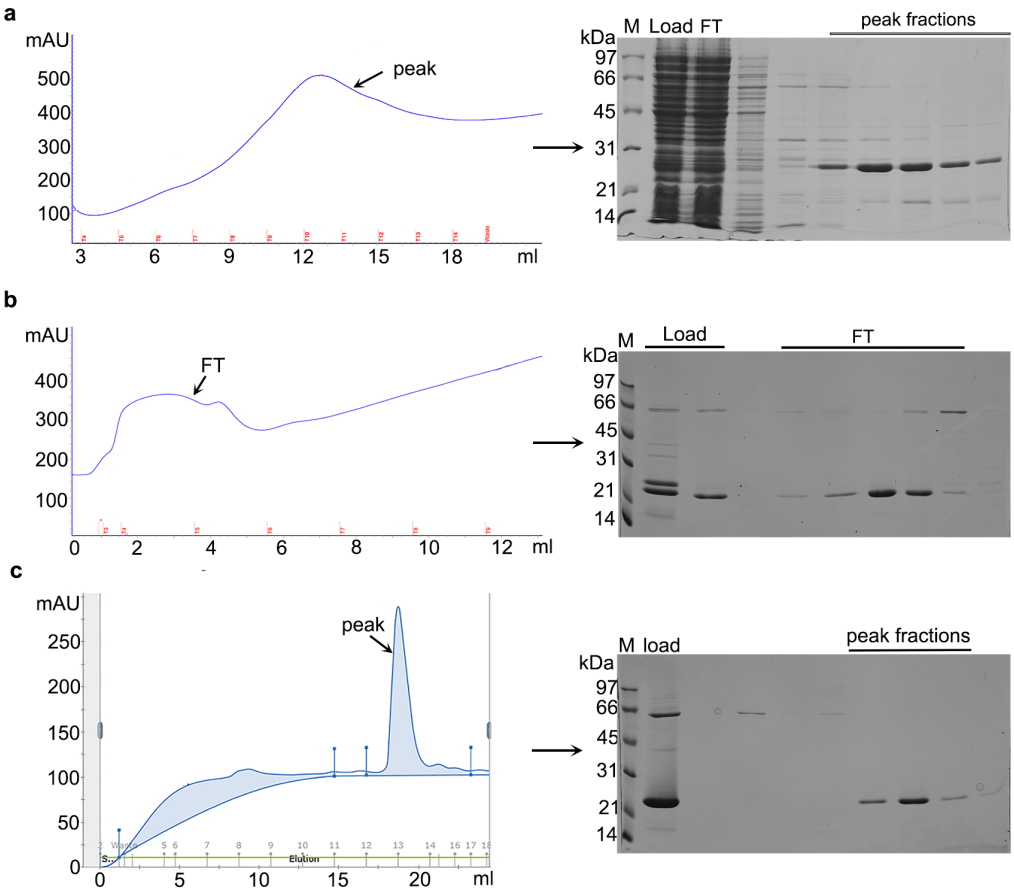


Figure 2-6 Purification procedure for gp14$_{NCH1-203}$.

a. Purification of $gp14_{NCH1-203}$ by nickel affinity chromatography; b. Purification of the 3C protease digested $gp14_{NCH1-203}$ by nickel affinity chromatography; c. Purification of $gp14_{NCH1-203}$ by size exclusion chromatography. The eluted peaks are highlighted on the chromatograms on the left, and the SDS-PAGE results are displayed on the right. 'M' indicates the protein size marker. 'Load' is the loaded sample. 'FT' indicates the flow through. The blue lines on chromatograms are the absorbance at 280 nm (A280).

The results of the nickel affinity chromatography of phiKo gp14 (with cleavable 6 × His tag, $gp14_{NCH1-203}$) show that the major peak fractions contain a reasonable amount of target protein. This peak was eluted by 30 - 50% imidazole (500 mM) in the same binding buffer (Figure 2-6a). The fractions containing the target protein were pooled and digested by 3C protease overnight at room temperature to cleave the N-terminal 6 × His tag. A considerable amount of target protein precipitated during digestion, likely due to the N-terminal 6 × His tag improving the protein solubility. The digested product was centrifuged at 15,000 rpm, 4 °C for 15 min to remove insoluble materials and the soluble fraction in the supernatant was collected and further purified by nickel affinity chromatography (Figure 2-6b) followed by size exclusion chromatography (Figure 2-6c). The target protein eluted in the peak fractions, as verified by SDS-PAGE (Figure 2-6c, right).

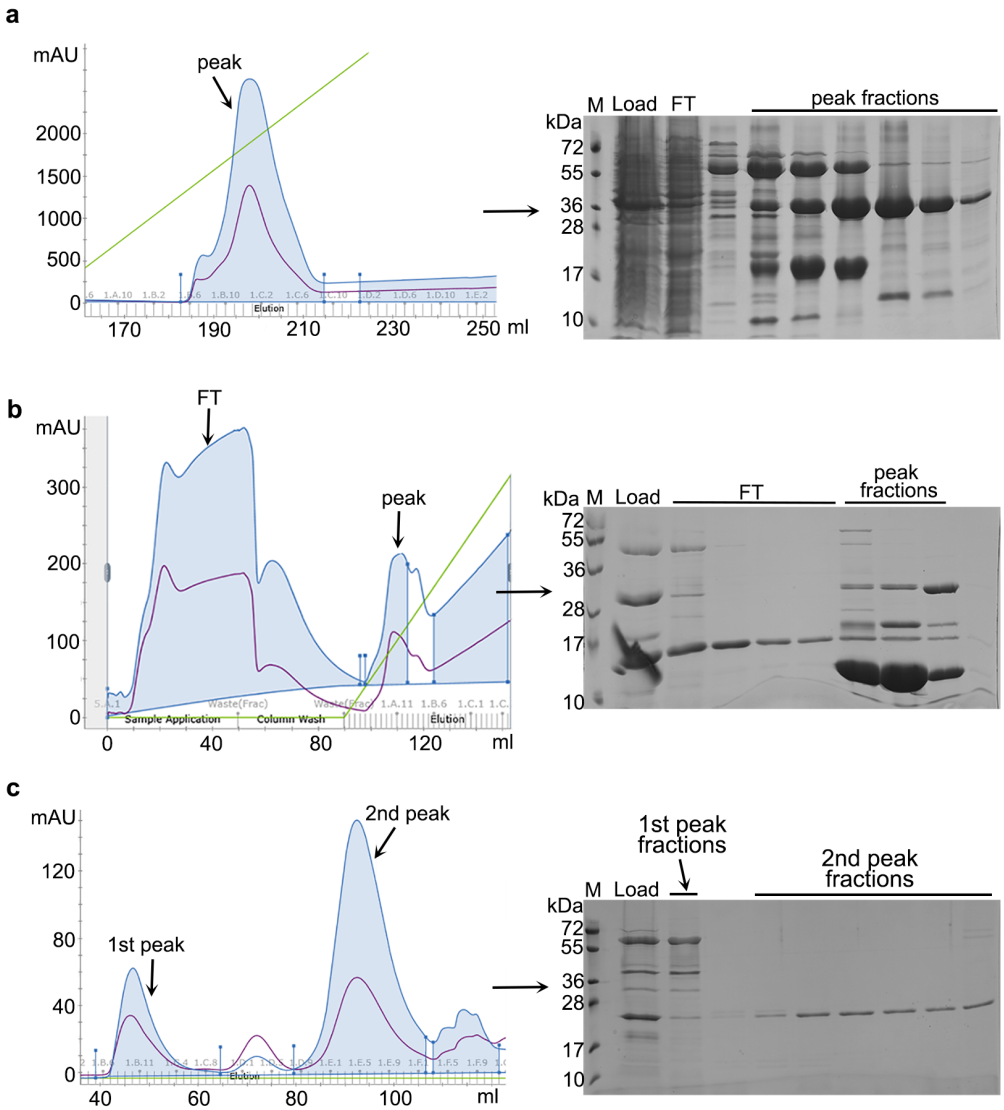


Figure 2-7 Nickel affinity chromatography and gel filtration for gp14$_{NCS1-203}$_K22A.
a. Purification of gp14$_{NCS1-203}$_K22A by nickel affinity chromatography; b. Purification of SUMO protease digested gp14$_{NCS1-203}$_K22A by nickel affinity chromatography; c. Purification of gp14$_{NCS1-203}$_K22A by size exclusion chromatography. The chromatograms are on the left, and SDS-PAGE results are shown on the right. 'M' indicates the protein size marker. 'FT' indicates the flow through. On the chromatograms, the blue lines are the absorbance at 280 nm, the purple lines are the absorbance at 254 nm, and the green lines indicate the percentage of the buffer containing 500 mM imidazole.

Results from nickel affinity chromatography of SUMO tagged gp14$_{NCS1-203}$_K22A mutant showed that the major peak fractions contained a considerable amount of protein with the approximately correct size (Figure 2-7a). This peak was eluted by 30 - 50% imidazole (500 mM) in the same binding buffer (Figure 2-7a), similar to gp14$_{NCS1-203}$ (Figure 2-5a). To remove the N-terminal SUMO tag, the fractions containing target protein were collected and digested

by SUMO protease overnight at room temperature. Like for the gp14$_{NCS1-203}$ construct, the presence of the SUMO tag improved the solubility of the gp14$_{NCS1-203}$_K22A protein, and extensive precipitation of the target protein (cleaved tag) was observed during digestion, with only a limited amount of soluble material being harvested. The digested protein was subjected to centrifuging at 15,000 rpm, 4 ˚C for 15 min, and the soluble material was collected and further purified by nickel affinity chromatography (Figure 2-7b) followed by size exclusion chromatography (Figure 2-7c). The target protein eluted in the second peak on the size exclusion chromatogram (Figure 2-7c, left), as verified by SDS-PAGE (Figure 2-7c, right). The phiKo gp14$_{NCS1-203}$_K22A mutant is more insoluble than the wild type gp14 protein.

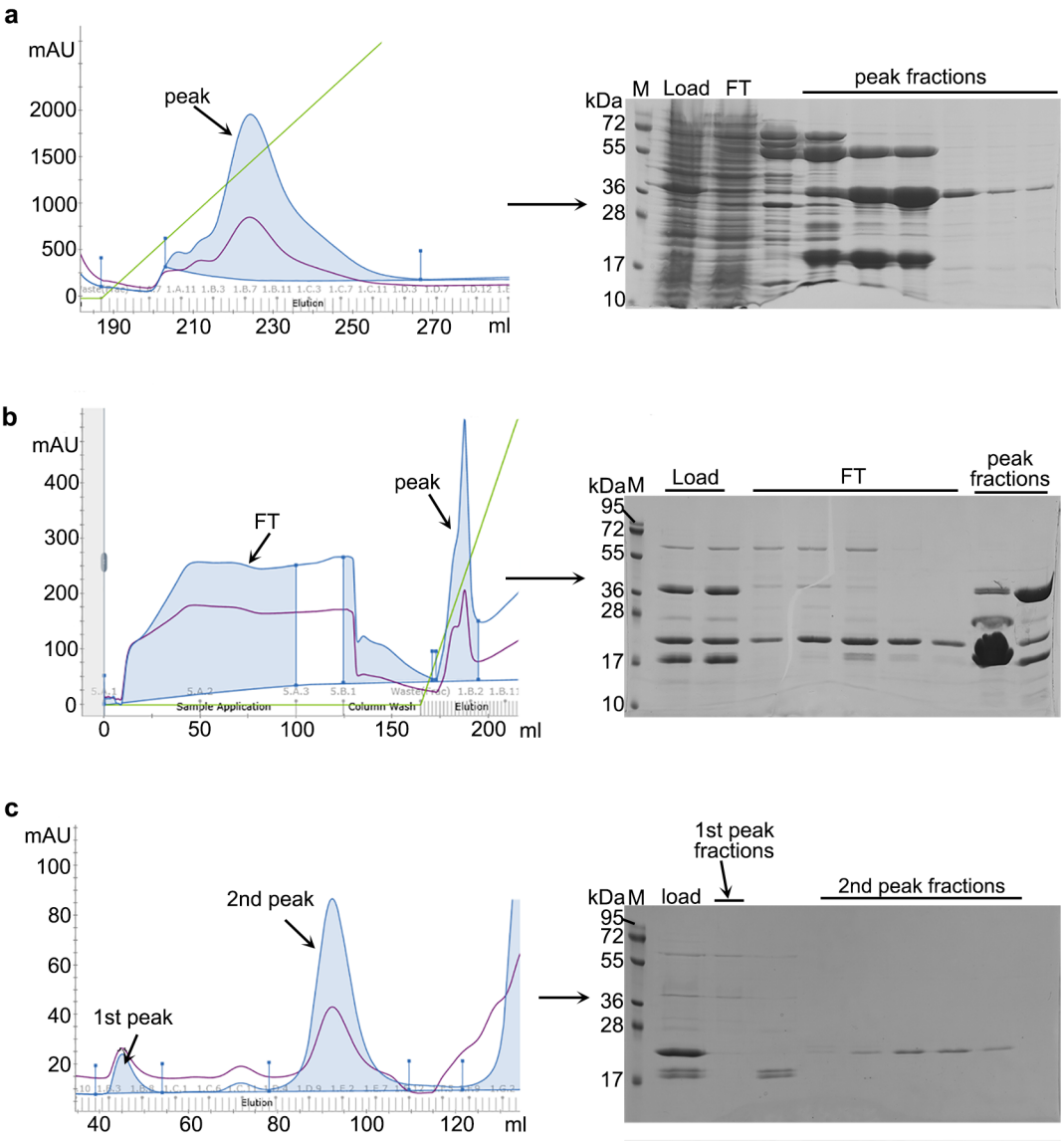


Figure 2-8 Nickel affinity chromatography and gel filtration for gp14$_{NCS1-203}$_R124A.

a. Purification of gp14$_{NCS1-203}$_R124A by nickel affinity chromatography; b. Purification of SUMO protease digested gp14$_{NCS1-203}$_R124A by nickel affinity chromatography; c. Purification of gp14$_{NCS1-203}$_R124A by size exclusion chromatography. The eluted peaks are indicated on the chromatograms (left), and the SDS-PAGE results are displayed on the right. 'M' indicates protein size marker and the size standards are highlighted at the left-hand side of the gels. 'FT' indicate the fractions containing the sample pooled from the flow through. The blue lines are the absorbance of the sample at 280 nm, the purple lines indicate absorbance of the sample at 254 nm, and the green lines indicate the percentage of buffer containing 500 mM imidazole on the chromatograms.

The SUMO tagged gp14$_{NCS1-203}$_R124A mutant was purified in a procedure similar to the other SUMO tagged gp14$_{NCS1-203}$ and gp14$_{NCS1-203}$_K22A proteins (Figure 2-5, 2-7 and 2-8). The solubility of this mutant protein after tag cleavage was similar to the SUMO cleaved wild type gp14 (gp14$_{NCS1-203}$).

2.3.2.2 PhiKo gp14 fusions with non-cleavable tags

Six constructs of phiKo gp14 were fused to N-terminal or C-terminal non-cleavable 6 × His tags for investigation of gp14 (Table 2-8). The purification of each protein construct is described in this section.

Two peaks were observed in the elution profile of nickel affinity chromatography for phiKo gp14$_{NNH1-203}$, where the second peak contained the target protein (Figure 2-9a). Fractions of the second peak (2nd peak) containing the target protein were eluted by 30 – 40% of buffer containing 500 mM imidazole and pooled together. So far, the N-terminal non-cleavable 6 × His tagged phiKo gp14$_{NNH1-203}$ appeared to be the most soluble one among all the constructs of gp14 produced.

To obtain high purity protein and investigate the oligomeric state of phiKo gp14, the soluble portion of the pooled fractions were further purified by size exclusion chromatography (Figure 2-9b). The first peak contained mostly contaminants, while the target protein was present in the fractions of the second peak. As seen in Figure 2-9b, the second peak eluted at 100 - 110 ml volume from the Superdex 200 16/60 column (GE Healthcare), which corresponds to the approximate ~20 kDa. The molecular weight of the target protein of gp14$_{NNH1-203}$ is 24.8 kDa as calculated by ExPASy ProtParam (179), indicating that the eluted protein is probably monomeric. Pure fractions in the second peak were pooled and concentrated to the maximum soluble concentration limit of 2.5 mg/ml, due to a large amount of the gp14$_{NNH1-203}$ precipitating during the concentrating process. Part of the freshly concentrated protein sample

was used for crystallisation, and the rest was stored at - 80 °C after flash freezing in liquid nitrogen.

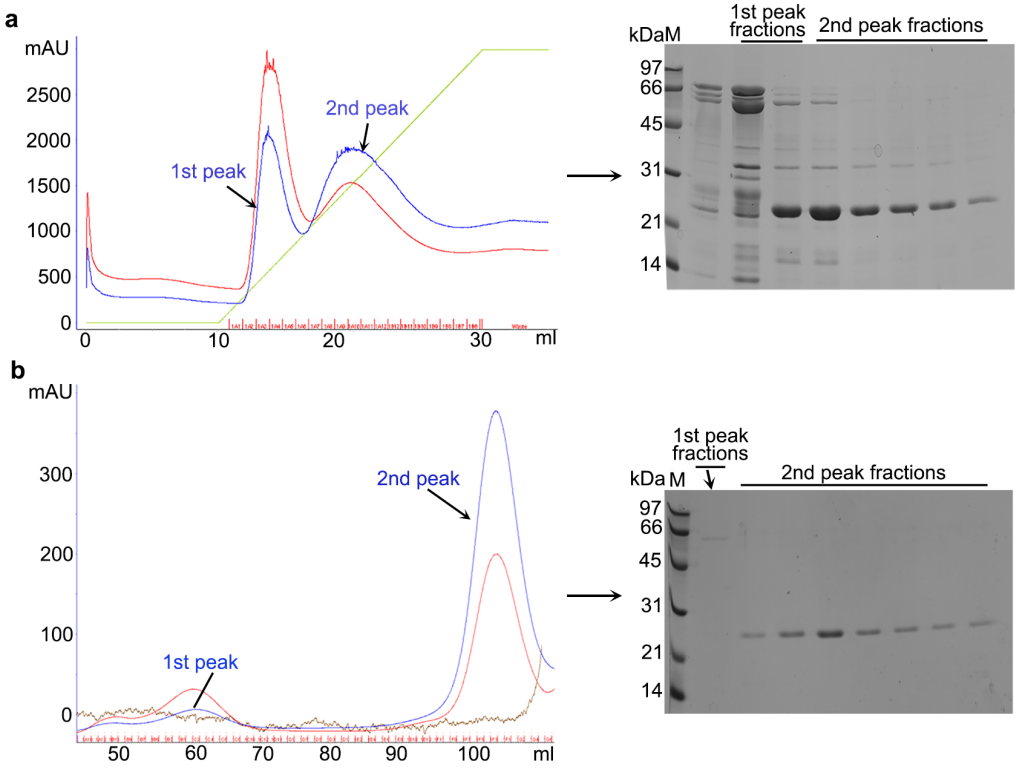


Figure 2-9 Nickel affinity and size exclusion chromatography of gp14$_{NNH1-203}$.
a. Purification of gp14$_{NNH1-203}$ by nickel affinity chromatography. The green line indicates the percentage of the buffer containing imidazole (500 mM), and the brown line in panel b is the conductivity of the buffer; b. Purification of gp14$_{NNH1-203}$ by size exclusion chromatography. The eluted peaks are indicated on the chromatograms on the left, and SDS-PAGE results are displayed on the right. On the chromatograms, the blue lines show absorbance at 280 nm, and the red lines indicate absorbance at 254 nm. The 'peak fractions' labelled on gels indicate the lanes containing sample fractions eluted from the 1st and 2nd peaks corresponding to the chromatograms. 'M' indicates protein size marker and size standards are highlighted on the gel panels.

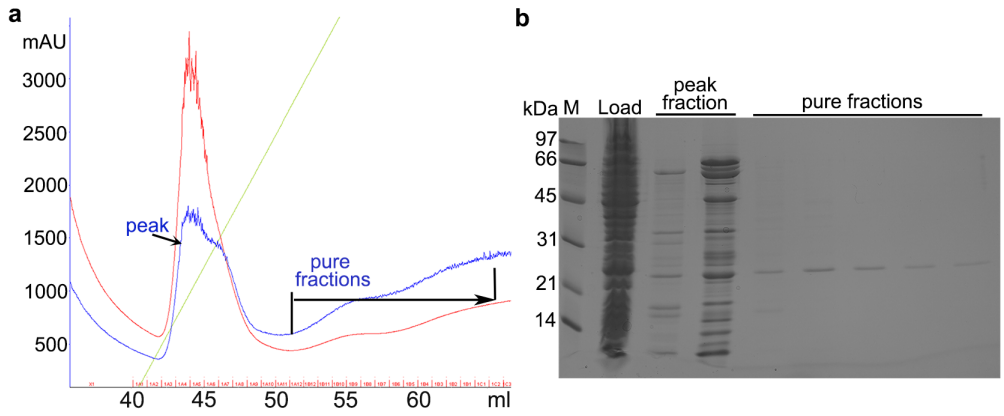


Figure 2-10 Nickel affinity chromatography of gp14$_{NNH7-203}$.

a. The gp14$_{NNH7-203}$ fusion protein was purified by nickel affinity chromatography. The peak and pure fractions containing target protein (gp14$_{NNH7-203}$) are indicated on the chromatogram. The blue lines are absorbance at 280 nm, the red lines show absorbance at 254 nm, and the green line indicates the percentage of the buffer containing 500 mM imidazole. b. Elution fractions were resolved by SDS-PAGE. The 'peak fractions' and 'pure fractions' indicate the lanes containing sample fractions eluted from the peak and pure fraction region corresponding to the chromatogram (panel a). 'M' indicates protein size marker. 'Load' indicates the sample loaded.

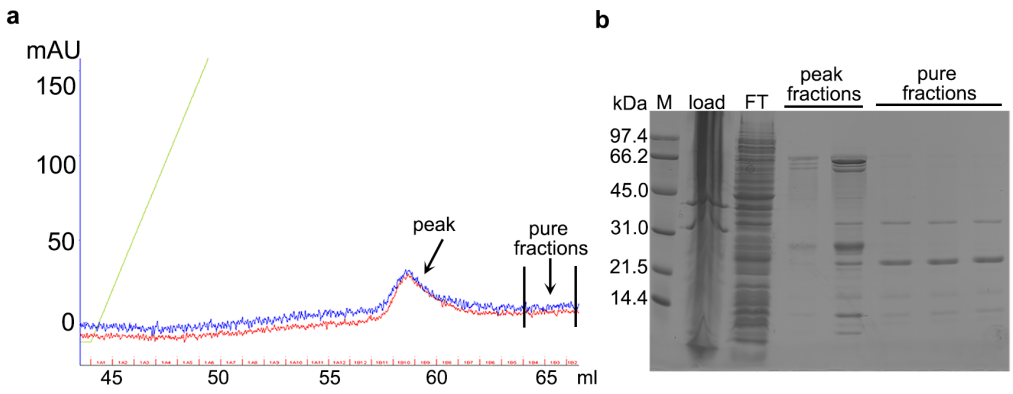


Figure 2-11 Nickel affinity chromatography of gp14$_{CNH7-203}$.

a. The gp14$_{CNH7-203}$ fusion protein was purified by nickel affinity chromatography. The blue lines are the absorbance of the sample at 280 nm, the red lines are the absorbance of the sample at 254 nm, and the green line in a indicates the percentage of the elution buffer containing 500 mM imidazole. b. Eluted fractions were resolved by SDS-PAGE. 'M' indicates a lane containing protein size marker, 'FT' indicates the flow through. The 'peak fractions' and 'pure fractions' indicate the lanes containing sample fractions eluted from the peak and pure fractions in panel a.

According to the disorder prediction by PrDOS (186) (Appendix 4), the N-terminal six residues (MDWRRI) are likely to be disordered. Thus, two truncated constructs, $gp14_{NNH7-203}$ and $gp14_{CNH7-203}$ (Figure 2-10 and 2-11), both comprising 196 residues (residue 7-203) of gp14, were cloned and expressed in *E. coli* cells in an attempt to improve crystallisation. The $gp14_{NNH7-203}$ construct is phiKo $gp14_{7-203}$ with an N-terminal non-cleavable 6 × His tag, and the $gp14_{CNH7-203}$construct is phiKo $gp14_{7-203}$ fusion with a C-terminal non-cleavable 6 × His tag. The theoretical pI of $gp14_{NNH7-203}$ is 8.61, and its molecular weight is 23.9 kDa, while the theoretical pI of $gp14_{CNH7-203}$ is 8.05 and its molecular weight is 23.8 kDa. The soluble materials from crude cell extracts of $gp14_{NNH7-203}$ and $gp14_{CNH7-203}$ were purified using nickel affinity chromatography. However, only a tiny amount of the target $gp14_{NNH7-203}$ or $gp14_{CNH7-203}$ protein was harvested (Figure 2-10 and 2-11), this was particularly evident for $gp14_{NNH7-203}$ or $gp14_{CNH7-203}$ during the concentration step, resulting in a small yield of purified protein. Optimising buffer pH and increasing salt concentration did not improve the solubility of these two proteins (data not shown).

2.3.2.3 Engineering homodimers of the phiKo gp14

Since gp14 is a putative FtsK-like ATPase, based on sequence alignment (Figure 2-4) and structure observations (in section 2.3.4), it is expected to function in an oligomeric form, like other FtsK/HerA family ATPases, as discussed in sections 1.3.2 and 1.3.3 above. Recombinant proteins of phiKo gp14 of all truncated or full-length constructs made in this study were purified as monomeric species (Figure 2-5 - 2-11). Since the protein is expected to function as an oligomer, construction of an obligate oligomeric species was attempted by fusing two monomers to form a homo-dimeric polyprotein, to attempt to facilitate the formation of higher-order oligomeric species. Thus, a protein engineering approach was attempted to connect two subunits of phiKo gp14 into a single polypeptide chain with a glycine-rich flexible linker (short connecting polypeptide) as shown in Table 2-9. However, it was proven that all of the three constructs were too insoluble and hence impossible to purify in reasonable quantities (Figure 2-12 - 2-14). A possible reason for this is protein misfolding and subsequent precipitation. Co-expression of chaperone proteins using Chaperone Plasmid Set (Takara) was adopted in an attempt to facilitate protein folding. However, it did not improve the solubility and yield of these engineered fusion proteins ($gp14_{NNHL7}$, $gp14_{NNHL13}$ and $gp14_{NNHL10}$). Further investigation using different linker peptides might be helpful in improving the folding and solubility.

Table 2-9 Constructs of engineered dimers of phiKo gp14.

Fusion protein	Recombinant plasmid	Constructs	MW (kDa)	Theoretical pI
$gp14_{NNHL7}$	pYM537	MGSSHHHHHH-gp14(M1-E203)-**GGGSEGG**-gp14	48.9	8.94
$gp14_{NNHL13}$	pYM546	MGSSHHHHHH-gp14(M1-E203)-**GGGSEGSEGGSEG**-gp14	49.4	8.61
$gp14_{NNHL10}$	pYM569	MGSSHHHHHH-gp14(M1-E203)-**GGGSEGGGSE**-gp14	49.2	8.80

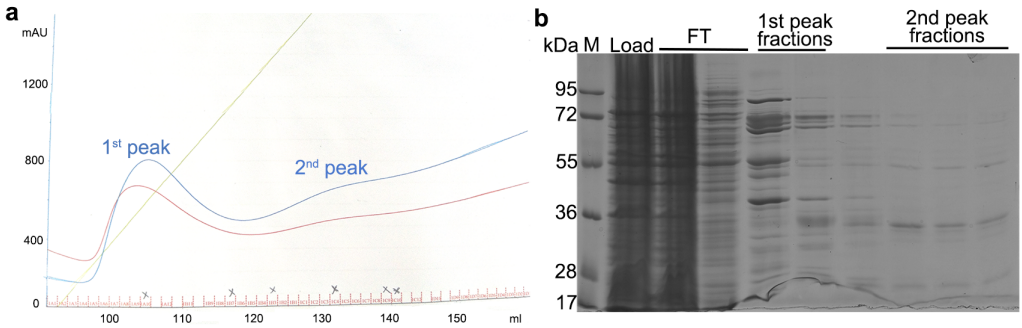


Figure 2-12 Nickel affinity chromatography for $gp14_{NNHL7}$.

a. The $gp14_{NNHL7}$ fusion protein was purified by nickel affinity chromatography. b. Elution fractions were resolved by SDS-PAGE. 'M' indicates a lane containing protein size marker. 'Load' indicates the sample loaded. 'FT' indicates the lanes containing fractions from flow through. The '1st and 2nd peak fractions' in panel b indicate the fractions that were collected from the peaks corresponding to the 1st and 2nd peaks (blue line, wavelength of 280 nm) on the chromatogram.

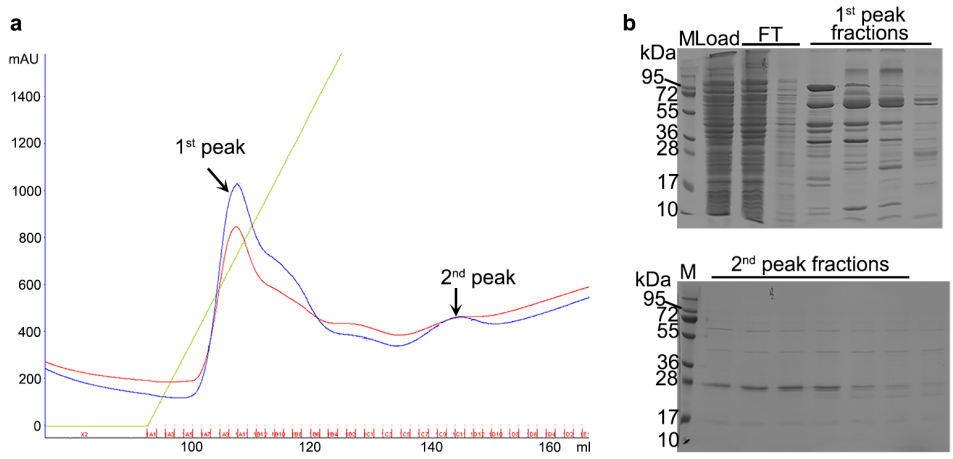


Figure 2-13 Nickel affinity chromatography for $gp14_{NNHL13}$.

a. The gp14$_{NNHL13}$ fusion protein was purified by nickel affinity chromatography. The blue line on the chromatogram indicates the eluted sample absorbance at 280 nm, and the red curve indicates eluted sample absorbance at 254 nm. The green line indicates the percentage of 500 mM imidazole in the buffer. The 1st and 2nd peaks of the sample absorbance at 280 nm (blue line) are indicated; b. Eluted fractions were resolved by SDS-PAGE. 'M' indicates a lane containing protein size marker, 'FT' indicates the fractions collected from flow through. The 1st and 2nd peak fractions in panel b are the fractions collected under the peaks corresponding to the 1st and 2nd peaks in panel a.

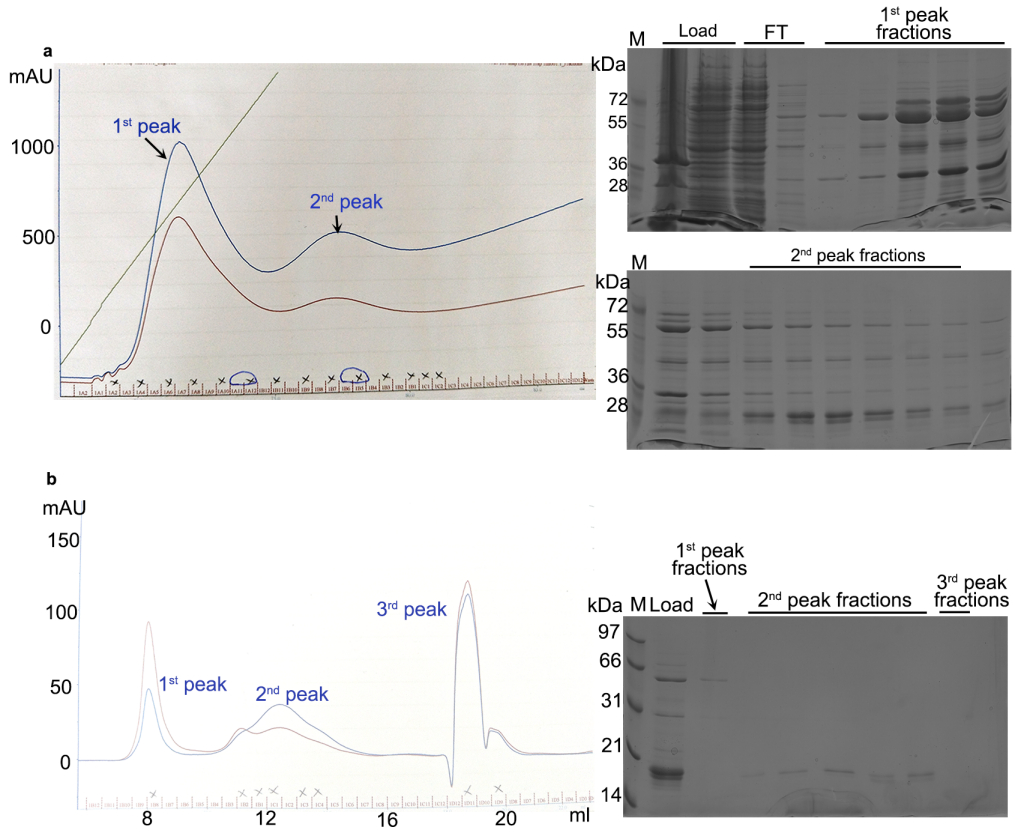


Figure 2-14 Nickel affinity and size exclusion chromatography for gp14$_{NNHL10}$.
a. The gp14$_{NNHL13}$ fusion protein was purified by nickel affinity chromatography, and the elution chromatogram is shown in the left panel. Eluted fractions were resolved by SDS-PAGE (right panel). b. The gp14$_{NNHL13}$ fusion protein was purified by size exclusion chromatography, and the chromatogram is in the left panel. Eluted fractions were analysed by SDS-PAGE (right panel). On the chromatograms, the lines of absorbance at 280 nm are in blue, and the lines of the absorbance at 254 nm are in red. The green line in panel a indicates the percentage of the elution buffer containing 500 mM imidazole. 'M' indicates a lane containing protein size marker, and 'FT' indicates the fractions of flow through. 'Load' indicates the loaded sample of gp14$_{NNHL13}$ for purification.

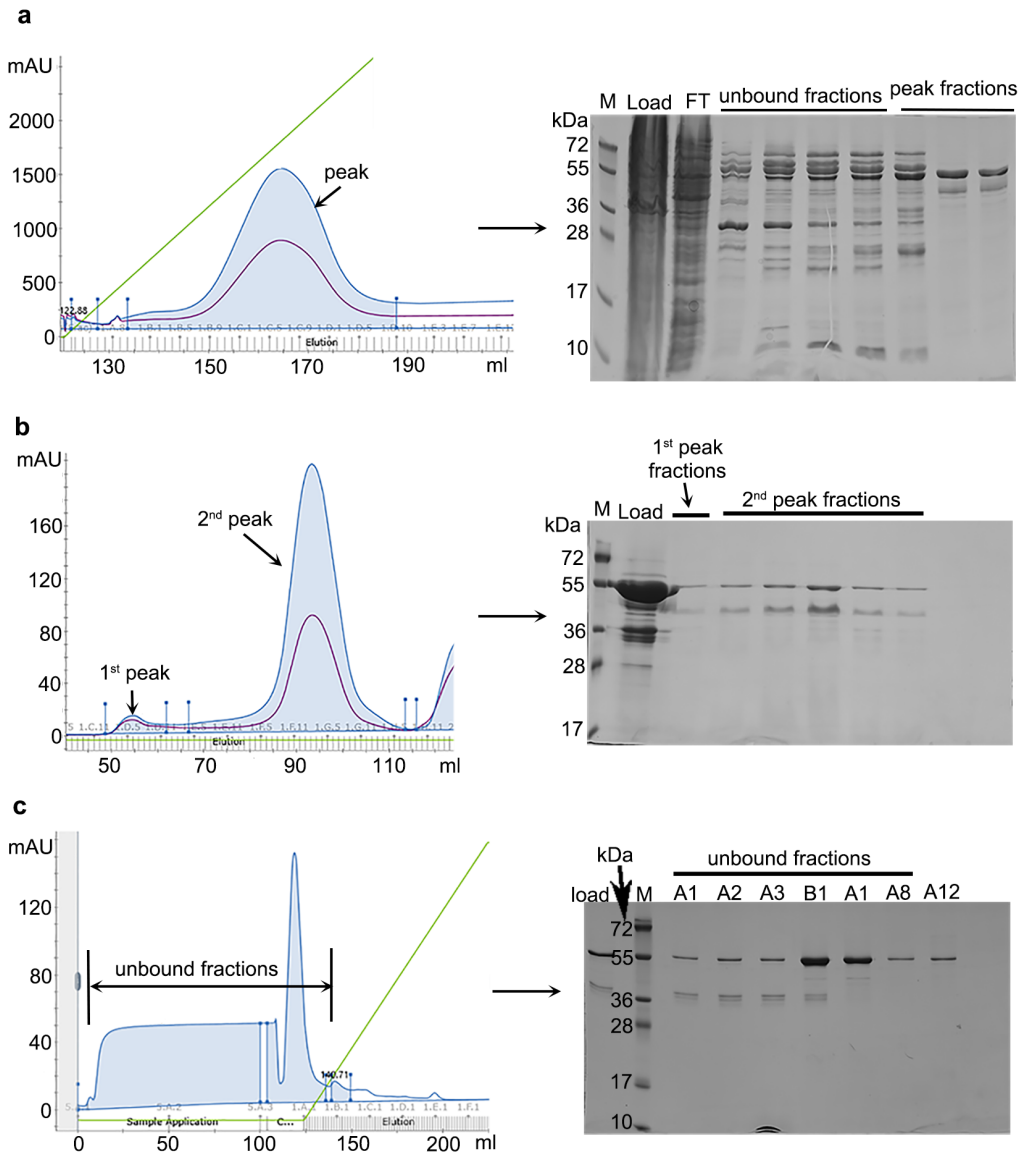


Figure 2-15 Purification procedure for gp14$_{NCG1-203}$.

a. Purification of gp14$_{NCG1-203}$ by nickel affinity chromatography; b. Purification of gp14$_{NCG1-203}$ by size exclusion chromatography; c. Purification of gp14$_{NCG1-203}$ by ion-exchange chromatography. On the chromatograms (left), the blue lines are the absorbance of the eluted sample at 280 nm, the purple lines are the absorbance of the eluted sample at 254 nm, and the green lines in panels a and c indicate the percentage of elution buffers containing imidazole and NaCl, respectively. The peak fractions are the fractions pooled from the peaks on the chromatograms (left). 'M' indicates protein molecular weight marker and 'FT' indicates the flow through fractions, and 'unbound fractions' indicate the fractions eluted by buffer containing zero or low (30 - 40 mM) imidazole concentration.

Since the low solubility of gp14 protein limits the sensitivity of biochemical characterisations, and the N-terminal SUMO tag improved the solubility of gp14 fusion protein, an N-terminal

GFP tagged construct ($gp14_{NCG1-203}$, fusion protein sequence is given in Appendix 7) was cloned and expressed. However, the recombinant fusion protein, $gp14_{NCG1-203}$ was insoluble, similar to the other phiKo gp14 constructs made in this study. The soluble fractions of $gp14_{NCG1-203}$ were purified by nickel affinity chromatography, followed by size exclusion chromatography (Figure 2-15a and 2-15b). The fractions from the two peaks observed in the chromatogram (Figure 2-15b) were both composed of bands with the same size (Figure 2-15b). The first peak is likely an oligomeric state of $gp14_{NCG1-203}$, though it is not known whether the GFP tag or gp14 protein promotes this oligomerisation, due to the poor solubility and yield limiting further investigations. Since two bands were observed on the SDS-PAGE gel, to obtain a protein with high purity and homogeneity, ion exchange chromatography was applied to further purify the $gp14_{NCG1-203}$ protein. However, ion exchange chromatography failed to isolate a single species in the eluted fractions (Figure 2-15c). The protein contaminants likely came from the expression cells and remained bound to the target protein.

2.3.3 Crystallisation of phiKo gp14

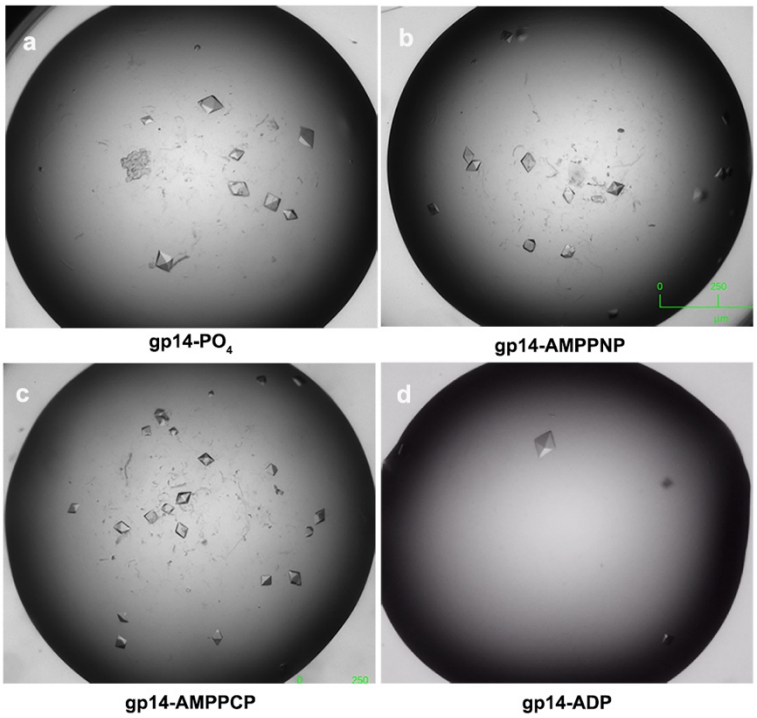


Figure 2-16 Crystals of phiKo gp14 protein complexes.
Crystals of phiKo gp14 complexes with phosphate (a), AMPPNP (b), AMPPCP (c), and ADP (d). Largest crystals in (a) had the size up to 0.1 × 0.06 × 0.03 mm^3.

As the gp14 fusion with an N-terminal non-cleavable 6 × His tag ($gp14_{NNH1-203}$) had superior solubility (maximum protein concentration of 2.2 - 2.5 mg/ml) compared to the untagged

protein (see section 2.3.2), it was used for crystallisation screens. Crystals were obtained in sitting drops with reservoir liquid containing 0.1 M citric acid pH 4.0 - 5.0, with slightly larger crystals grown at higher concentrations (9% -16% v/v) of 2-Methyl-2,4-pentanediol (MPD), see details in section 2.2.5. The complexes, gp14-ADP, gp14-AMPPNP and gp14-AMPPCP, were formed by adding 20 mM $MgCl_2$ and 10 mM ADP or AMPPNP or AMPPCP in the protein samples, and the sample mixtures were incubated for 20 - 30 min at room temperature prior to crystallisation. Crystals with a maximum length of 50 - 60 μm and a tetrahedral shape appeared within 2 days (Figure 2-16). Promising crystals were frozen and stored in liquid nitrogen for further X-ray data collection.

2.3.4 Structure determination

X-ray structures of the gp14 ATPase in complex with phosphate, ADP, AMP-PNP and AMPPCP were determined at resolutions ranging between 1.8 and 2.2Å (Table 2-10). All crystals belong to the space group $P4_32_12$ (tetragonal unit cell). In crystallography, space group represents a description of the symmetry of a crystal, and the unit cell is the smallest crystal volume that contains a collection of asymmetric units (molecules) that are related by crystallographic symmetry (187). The crystals have relatively high solvent content (67%) estimated by Matthew's coefficient calculations (188, 189), corresponding to one molecule in an asymmetric unit.

Table 2-10 Crystallographic and refinement statistics.

	gp14-phosphate	**gp14-ADP**	**gp14-AMPPNP**	**gp14-AMPPCP**
Data collection				
Wavelength (Å)	0.97922	0.97922	0.97957	0.97957
Space group	$P4_3\ 2_1\ 2$	$P4_3\ 2_1\ 2$	$P4_3\ 2_1\ 2$	$P4_3\ 2_1\ 2$
Cell dimension				
a, b, c, Å	73.2, 73.2, 130.3	72.7, 72.7, 132.3	72.7, 72.7, 131.8	72.5, 72.5, 132.4
Resolution, Å	63.80 - 2.20	63.8 - 1.87	63.8 - 1.86	48.90 - 1.80
R_{meas} [a]	0.083 (1.853)	0.172 (1.287)	0.102 (1.759)	0.091 (2.035)
$I/\sigma I$ [b]	13.5 (0.7)	8.5 (1.1)	14.1 (1.1)	14.6 (1.1)
Completeness (%) [c]	98.7 (92.0)	99.9 (99.0)	99.8 (99.1)	99.8 (99.0)
CC (1/2)	0.999 (0.414)	0.381	0.999 (0.506)	0.998 (0.531)

Multiplicity	6.8 (5.0)	12.0 (6.6)	12.6 (13.0)	12.8 (13.1)
Refinement				
Resolution, Å	48.70 – 2.20	48.99 – 1.90	48.86 – 1.88	48.94 – 1.80
No. reflections	17491	27402	28069	31848
R_{work}/R_{free}	0.198 / 0.237	0.172 / 0.215	0.179 – 0.211	0.173 / 0.205
No. atoms				
Protein	1654	1680	1668	1663
Ligand/ion	10	39	34	34
Water	93	189	140	203
Mean B factor [d]				
Protein	54.2	40.7	45.8	42.7
Ligand/ion	58.2	45.4	53.9	72.8
Water	53.9	51.6	55.4	55.3
RMSD				
Bond lengths, Å	0.011	0.011	0.010	0.010
Bond angles, °	1.7	1.7	1.9	1.6

a. R_{meas}: $\sum_{hkl} n \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}(n-1)\sum_i I_i(hkl)$, where n is a number of the symmetry equivalent contributors to the average; b. $I/\sigma I$: Signal to noise ratio for merged intensities; c. Completeness for unique reflections; d. Mean B factor for atoms of protein, ligand/ion and water, respectively. Values for the highest resolution shell are given in brackets.

R factor is a measure of agreement between the amplitudes of the structure factors calculated from a crystallographic model and those from the original X-ray diffraction data, indicating the quality of the atomic model obtained from the crystallographic data. The R factor is calculated during each cycle of least-squares structure refinement to assess progress (122). Before refinement begins, about 10% of the experimental observations are removed from the data set. Then, refinement of the model is performed using the remaining 90% of the data, and the R_{work} is calculated from these reflections. The R_{free} value is calculated by seeing how well the model predicts the 10% of reflections that were not used in refinement (190). The final values of R_{work}/R_{free} for refined phiKo gp14 structures are 0.198/0.237, 0.172/0.215, 0.179/0.211 and 0.173/0.205 (Table 2-10), showing good correspondence between the calculated and observed structure factors.

PhiKo gp14 in complex with phosphate was crystallised under similar conditions to complexes with other non-hydrolysable ATP analogues. The bound phosphate is likely to be a contaminant coming from the purification and crystallisation procedure. Though 5 μM of a 12-mer dsDNA (the sequence is shown in Appendix 6) was added to the protein sample for crystallisation, no DNA density was observed in electron density map (Table 2-10).

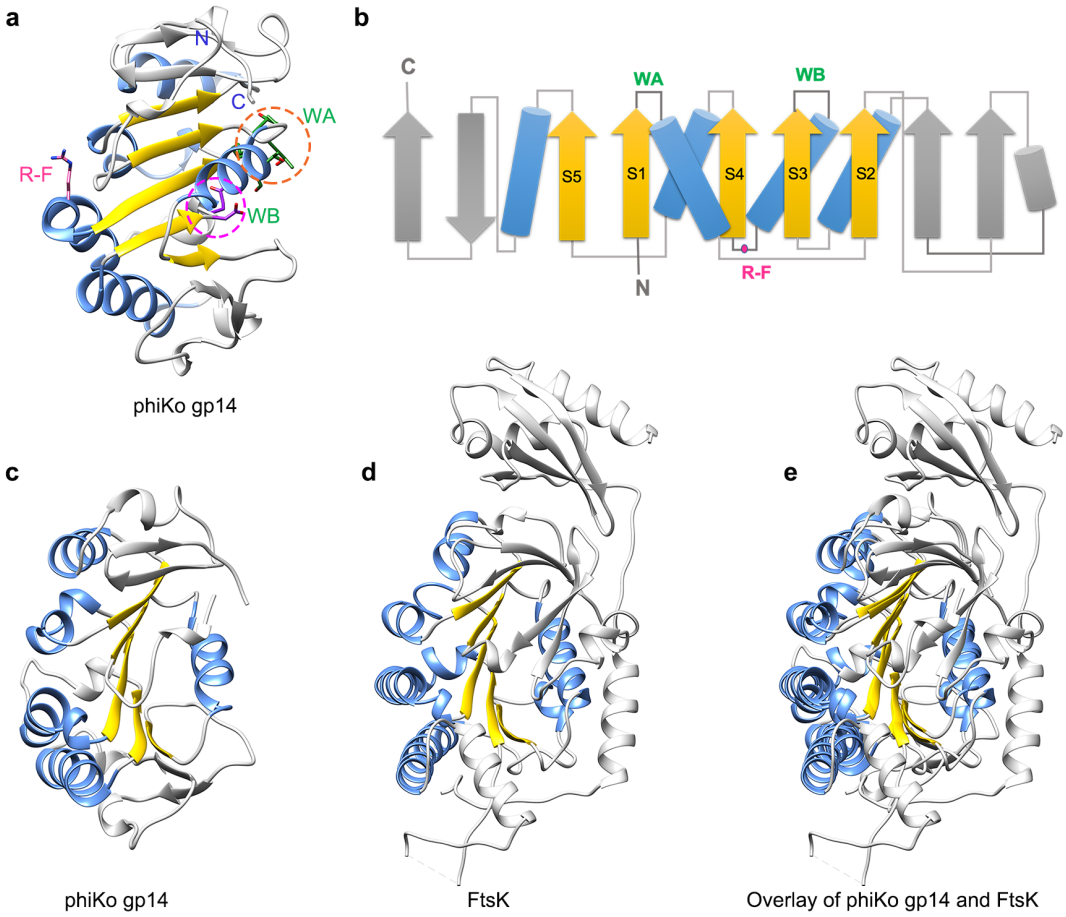


Figure 2-17 Comparison of phiKo gp14 and FtsK structures.

a. The overall structure of phiKo gp14. The core RecA-like fold is in yellow. WA stands for Walker A and is highlighted in an orange dashed circle; WB indicates Walker B and is highlighted in a magenta dashed circle. R-F indicates the 'arginine finger', i.e. R124 in phiKo gp14. b. Topology diagram of phiKo gp14. Strands are shown as arrows with the arrowhead at the C-terminus, helices are shown as cylinders. α-helices are in blue. *β*-strands conserved across the ASCE group are numbered (according to the ASCE core fold or RecA-like core fold) and in yellow, with the rest of the *β*-strands shown in grey. The non-conserved region across the FtsK/HerA family is in grey. c-d. Overall structures of phiKo gp14 (c) and FtsK (d) in the same view. E. Overlay of phiKo gp14 and FtsK. FtsK [PDB code 2UIS, (72)].

The full length phiKo gp14 protein comprises 203 residues with a molecular mass of 23.6 kDa and possesses a conserved FtsK/HerA-like fold (6) with a nine β-strands sandwiched between α-helices (Figure 2-17a and 2-17b). In addition, it has an anti-parallel β-strand found at the C-terminus of gp14 (Figure 2-17a and 2-17b), as expected for the FtsK/HerA family ATPases (Figure 2-17c - 2-17e)(6). The four structures of phiKo gp14 complexes are well superimposable, with the root mean square deviations (RMSD), calculated for the main-chain atoms, of 0.31Å (gp14-phosphate and gp14-AMPPCP), 0.33 Å (gp14-phosphate and gp14-AMPPNP), 0.25 Å (gp14-ADP and gp14-AMPPNP), 0.14 Å (gp14-AMPPCP and gp14-AMPPNP) and 0.43Å (gp14-phosphate and gp14-ADP), indicating that ATP binding induces subtle changes of the overall structural conformation. These subtle conformational changes induced by nucleotide binding is probably either due to the nucleotide bound complexes of gp14 not being in the functional conformation or the crystal lattice restraints or both. The N- and C- termini of the phiKo gp14 are located close to each other (Figure 2-17a). Moreover, the N- and C-termini locates close to the ATP binding pocket (Figure 2-17a), indicating that both N- and C-termini play a role for ATP hydrolysis or protein oligomerisation. The residue, E203, at the C- terminal end is not visible in the gp14 crystal structures. The N-terminal non-cleavable 6 × His tag is highly disordered in the gp14 structures. However, the residues at the N-terminus of gp14 are visible in the crystal structure of phiKo gp14, despite the prediction suggesting that the six N-terminal residues would be disordered (Appendix 4).

The conserved ASCE core (or RecA-like core) fold is shown in Figure 2-17a and 2-17b. The structure-based sequence alignment shows the conserved Walker A and Walker B motifs (68) as well as the 'arginine finger' are present in phiKo gp14 (Figure 2-17a and 2-17b), suggesting gp14 is an FtsK-like ATPase as expected. The Walker A motif is located between the 1st structurally conserved β-strand and the 1st α-helix, and the Walker B motif lies in between the structurally conserved 5th β-strand and 3rd α-helix (Figure 2-17a and 2-17b). The region between the 6th and 7th β-strands is conserved across FtsK/HerA superfamily ATPases from viruses, including bacteriophage PRD1. It has been named as P9-specific motif (Figure 2-18), although the exact function of this region is not yet known (6, 41, 85).

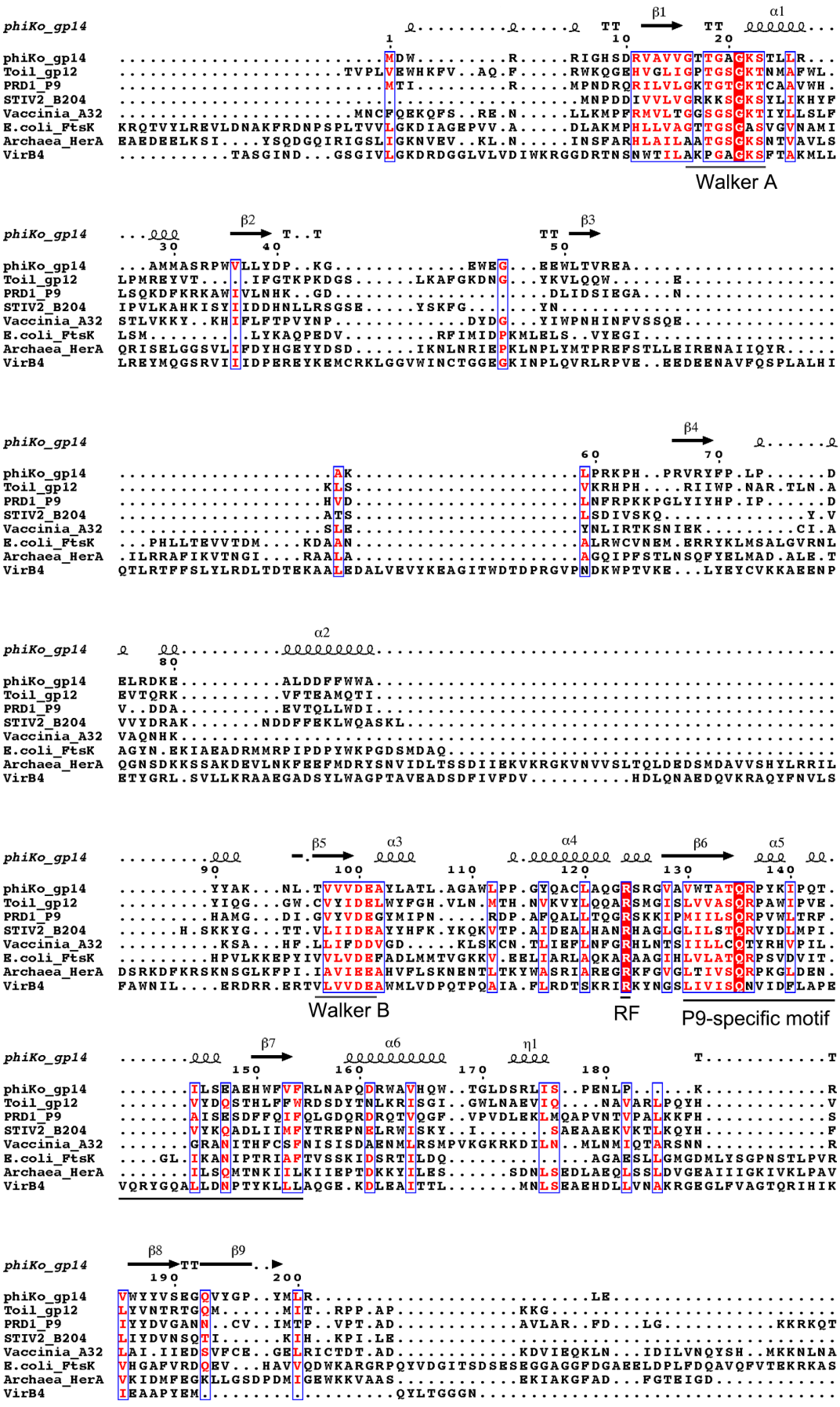

phiKo_gp14
phiKo_gp14
Toil_gp12
PRD1_P9
STIV2_B204
Vaccinia_A32
E.coli_FtsK
Archaea_HerA
VirB4
Walker A
Walker B
RF
P9-specific motif

Figure 2-18 Structure based sequence alignment of phiKo gp14.

Protein sequences used for the alignment includes phiKo gp14; Toil gp12 (180); PRD1 P9 (181); *Turriviridae* phage Sulfolobus turreted icosahedral virus 2 (STIV2) B204 (182); Vaccinia virus A32 (183); and *Escherichia coli* FtsK (72). The sequence alignment was performed by Clustal Omega (184), and the figure was made using ESPript (185). Walker A and B motifs are indicated on the figure, and RF indicates the 'arginine finger'.

2.3.5 Electrostatic surface potential analysis

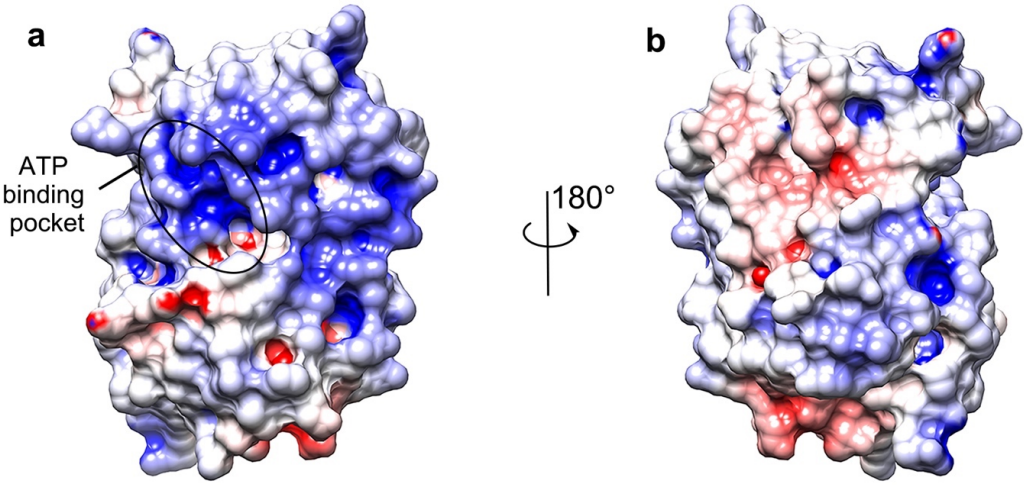


Figure 2-19 Electrostatic surface potential of phiKo gp14.

a. Front views of phiKo gp14; b. Back views of phiKo gp14. The solvent excluded surface of phiKo gp14 protein was calculated in Chimera (174). Models with ligands and water molecules removed were used for calculating the electrostatic surface potential. The chemical groups in the structure of gp14 were assigned charges by PDB2PQR (191) using the SWANSON force field (192). Electrostatics potential surfaces of phiKo gp14 structures (red, negative to blue, positive, ± 10 kT/e) were calculated by APBS (193) in Chimera (174).

The oligomeric state of a functional assembly of phiKo gp14 is unknown. Several fusion constructs of gp14 did not form oligomers in solution (see details in section 2.3.2). In order to investigate the possibility of oligomerisation of phiKo gp14, the electrostatic surface potential of the fully refined structures of complexes of phiKo gp14 with phosphate was calculated and analysed (Figure 2-19). The conserved Walker A and B motifs form a mostly positively charged pocket for stabilising ATP (Figure 2-19). Other positively charged regions of phiKo gp14 may potentially be involved in binding DNA. Little difference was observed of the

electrostatic potential surfaces of the three structures of gp14-phosphate, gp14-AMPPMP and gp14-ADP (data not shown).

2.3.6 Mechanism of ATP hydrolysis

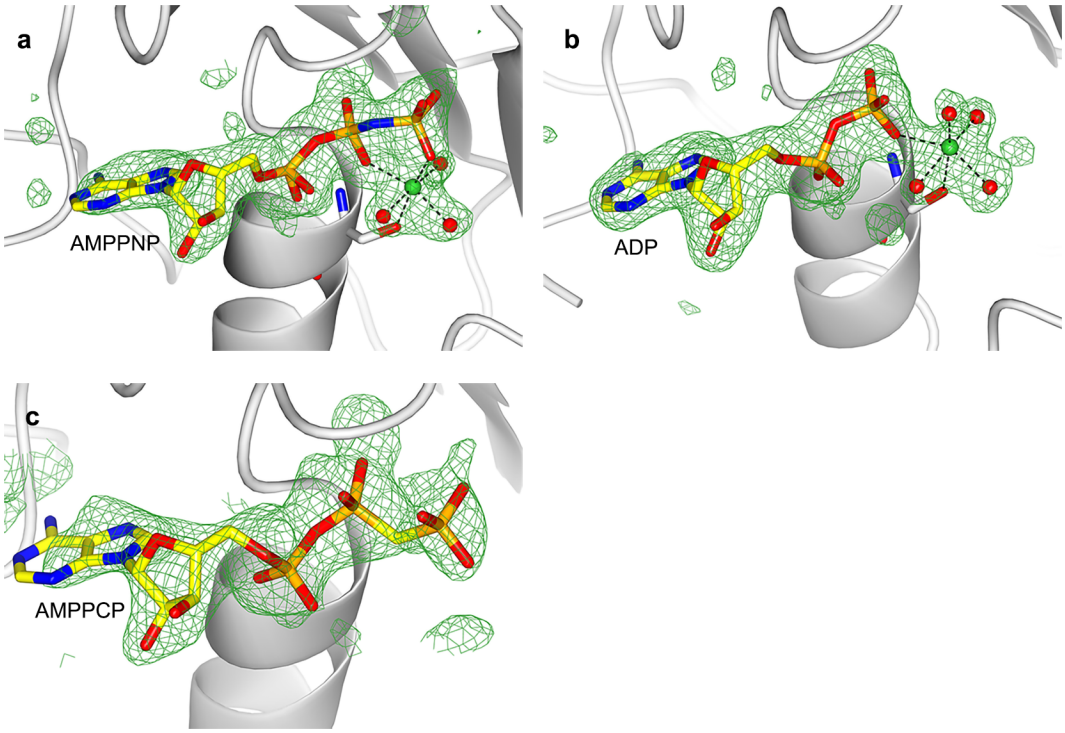


Figure 2-20 Omit maps for structures of phiKo gp14 complexes with ATP analogues. a. Omit map (mF_O-DF_C) of phiKo gp14-AMPPNP complex contoured at 2.5 σ level; b. Omit map (mF_O-DF_C) of phiKo gp14-ADP complex contoured at 3 σ level; c. Omit map (mF_O-DF_C) of phiKo gp14-AMPPCP complex contoured at 2 σ level. Omit maps were contoured by CCP4MG (173). The magnesium ion is represented by the lime green sphere, and water molecules are displayed as red spheres.

To investigate the ATP hydrolysis mechanism, the structures of phiKo gp14 complexes of pre-hydrolysis state (with AMPPNP or AMPPCP bound) and post-hydrolysis state (with ADP bound) were compared (Figure 2-20). The bound AMPPNP, ADP and AMPPCP moieties were unambiguously defined in the omit density maps, respectively, inside the ATP binding pocket (Figure 2-20). The structures of phiKo gp14 in complex with AMPPNP and ADP contain the catalytic magnesium in octahedral coordination (Figure 2-20a and 2-20b). In construct, the structure with AMPPCP does not contain a catalytic metal ion (Figure 2-20c). The absence of catalytic metal ion in the gp14-AMPPCP complex along with high mean B factors of the AMPPNP moiety (Table 2-10) suggests weak interactions of AMPPCP with the ATP binding pocket, possibly corresponding to a pre-binding state.

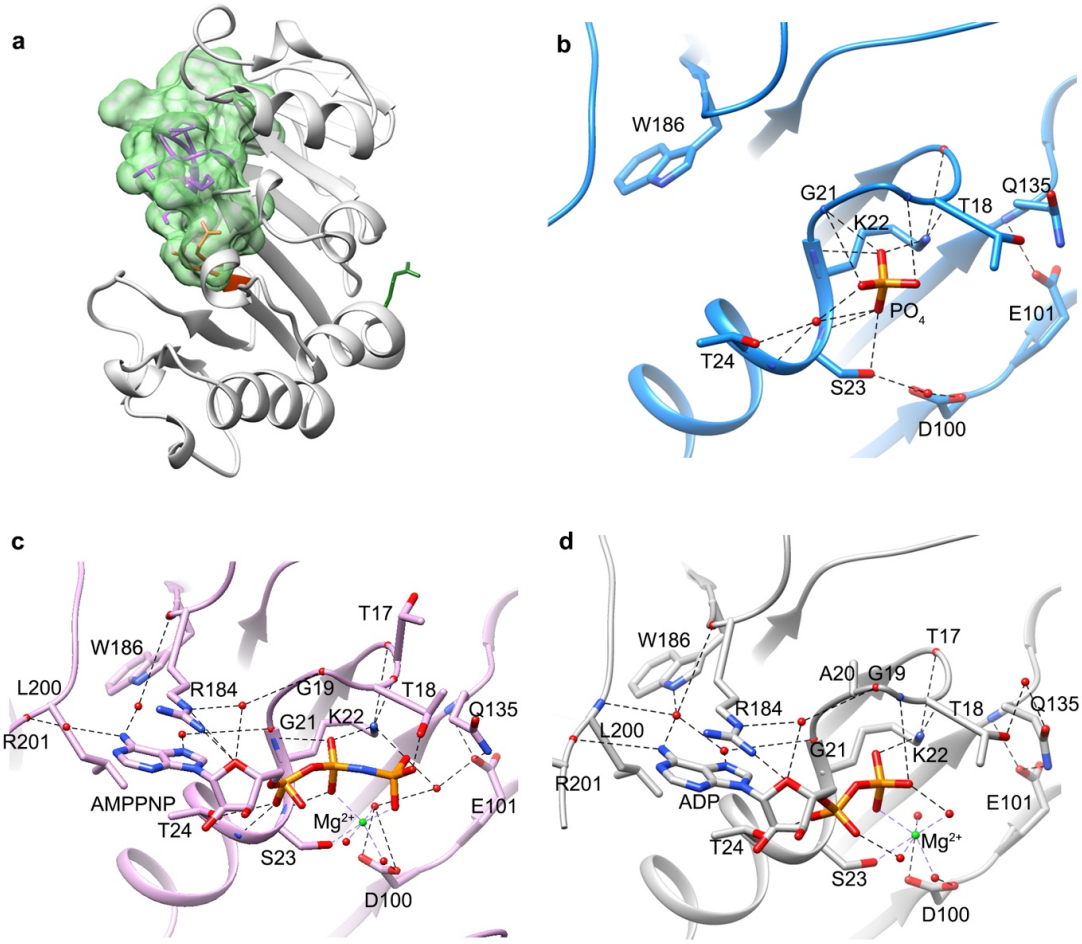


Figure 2-21 ATP binding pocket of phiKo gp14 complex structures.

a. ATP binding pocket of phiKo gp14. The solvent excluded surface of the ATP binding pocket is calculated and coloured in semi-transparent light green using Chimera (174). The Walker A motif is coloured in purple; D100 and E101 of the Walker B motif are coloured in orange. The conserved R124 is highlighted in dark green. b. ATP binding site geometry of the gp14-phosphate complex; c. ATP binding site geometry of the gp14-AMPPNP complex; d. ATP binding site geometry of the gp14-ADP complex. Atoms are coloured as follows: oxygen, red; nitrogen, blue; phosphor, orange; magnesium, green.

Next, configurations of the ATP binding site for complexes with phosphate, AMPPNP and ADP were investigated (Figure 2-21). The ATP binding pocket is formed by both the Walker A (including residues G16, T17, T18, G19, A20, G21, K22, S23 and T24) and Walker B motifs (including residues V97, V98, V99, D100 and E101). In addition, residues K42, E44, Q135, R184, V185, W186, M199, L200 and R201 also contribute to the establishment of the ATP binding pocket (Figure 2-21a). In the structure of the phiKo gp14 complex with phosphate, the phosphate binds to the Walker A motif residues G21, K22, S23 and T24, through direct or

water-mediated hydrogen-bonding interactions (Figure 2-21b). The phosphate is positioned similarly to the β-phosphate of the bound AMPPNP and ADP (Figure 2-21c and 2-21d).

In both the AMPPNP and ADP bound gp14 structures, the sidechains of R184 and L200 stack onto the adenine moiety by non-polar interactions (Figure 2-21c and 2-21d). In the phiKo gp14-AMPPNP structure, an intensive H-bonding network is formed to stabilise the AMPPNP (Figure 2-21c). The sidechain amino group of K22 stabilises both the β- and γ-phosphate of AMPPNP via H-bonds (Figure 2-21c). Moreover, the sidechain of T18 forms a H-bond to the γ-phosphate and the sidechain of T24 interacts with the α-phosphate of AMPPNP. The main chain amide of S23 is H-bonding to the β-phosphate. The sidechain of R184 stabilises the sugar pucker of AMPPNP via a direct H-bond. In addition, the N^6 of the adenine interacts with the main chain of R201 and R184 via water-mediated H-bonds, respectively, and the N^7 of the adenine contacts the mainchain hydroxyl of G21 (Figure 2-21c). The conserved E101 is H-bonded to a water molecule which contributes two H-bonds to the γ-phosphate and a metal coordinated water (Figure 2-21c), indicating that E101 might be the general base to polarise the water molecule for nucleophilic attack, in common with other ASCE group ATPases (6, 194, 195). The sidechain of Q135 forms a salt bridge to the sidechain of E101, likely facilitating proton transfer during ATP hydrolysis (Figure 2-21c). The catalytic magnesium is coordinated by oxygen atoms from the β- and γ-phosphates, two water molecules and the sidechain hydroxyl of S23 (Figure 2-21c).

The binding mode of ADP in the active site is similar to the gp14-AMPPNP structure, but several differences are noted. Due to the lack of γ-phosphate, the H-bonding interactions between protein and the γ-phosphate are broken (Figure 2-21c and 2-21d). In particular, the T18 residue has reoriented its sidechain to establish a water-mediated H-bond with the sidechain hydroxyl group of Q135 (Figure 2-21d). The adenine moiety of ADP forms a slightly more enhanced H-bonding network to the protein by involving in W186 to stabilise the adenine moiety of ADP. The N^6 of the adenine forms a H-bond to the mainchain oxygen of R201.

Moreover, the N^6 atom of the adenine of ADP makes water-mediated interactions with the main chain of L200 as well as the sidechain of W186 (Figure 2-21d). The mainchain amide of G19 is H-bonding to the β-phosphate (Figure 2-21d). In addition, the D100 of the conserved Walker B motif mediates the coordination of waters to the catalytic magnesium but not directly to the magnesium ion in both gp14-AMPPNP and gp14-ADP structures (Figure 2-21c and 2-21d), whereas the conserved aspartate of Walker B motif directly coordinates the catalytic metal ion in some other ASCE ATPases (6, 196). The S23 residue coordinates the catalytic

magnesium in both gp14-AMPPNP and gp14-ADP structures, similar to other ASCE ATPases (6, 85).

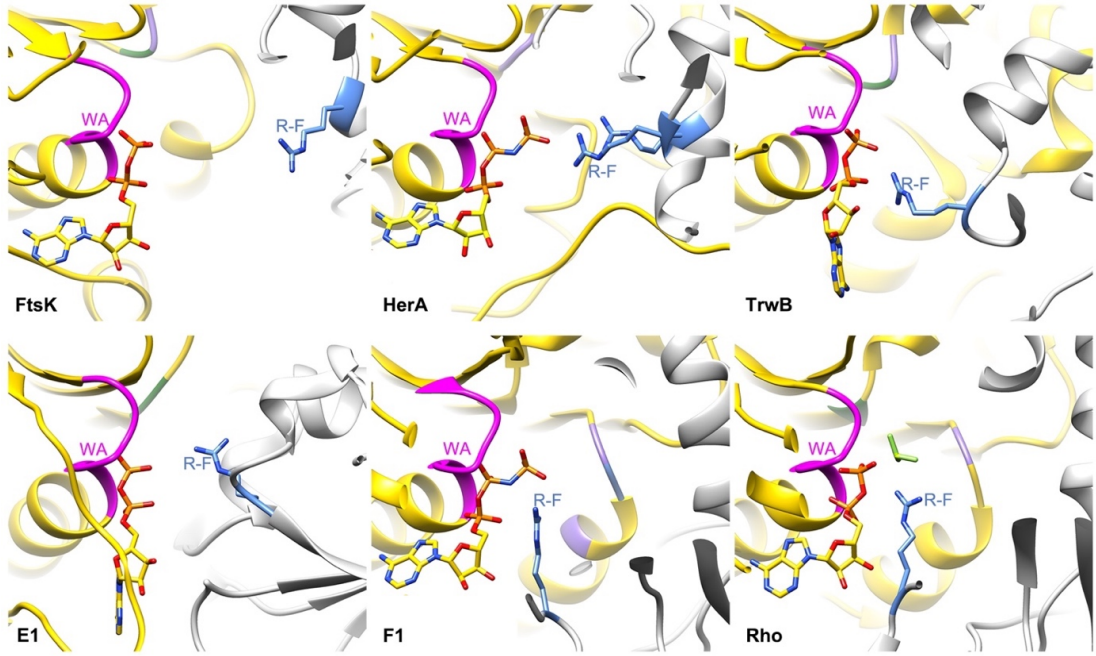


Figure 2-22 The 'trans-arginine finger' of hexametric ASCE group ATPases.
The ATP binding sites of the FtsK translocase [upper left, PDB code 2IUU, (72)], HerA ATPase [upper central, PDB code 4D2I, (197)) and TrwB ATPase [upper right, PDB code 1GKI, (198)] show the location of catalytic residues common to all FtsK/HerA family ATPases. The ATP binding site of E1 helicase [lower left, PDB code 2GXA] highlighting the location of the catalytic residues common to all hexametric AAA^+ family ATPases. The ATP binding sites of F1-ATPase β subunit [lower central, PDB code 2HLD, (199)] and Rho helicase [lower right, PDB code 3ICE, (200)] shows the composition of the catalytic residues common to all hexametric RecA-like ATPases. The P-loop Walker A motif is labelled as WA and coloured in magenta, the 'trans-arginine-finger' (R-F; blue). The catalytic glutamate of Walker B is in purple, and the aspartate involved in catalysis is in dark green. The two adjacent subunits are coloured in yellow (left) and light grey (right). Nucleotides, including ADP, AMPPMP and ADP-BeF_3, are displayed in sticks.

According to the accepted mechanism of ATP hydrolysis by ASCE group hexametric ATPases, an additional positively charged residue (arginine) is required as a catalytic sensor to facilitate ATP hydrolysis (Figure 2-22). However, in the monomeric gp14-AMPPNP structure, K22 is the only positively charged residue stabilising the γ-phosphate (Figure 2-21c). Hence, it is likely that the catalytic positively charged residue is provided by an 'adjacent'

subunit, following the assembly of the oligomeric state, in common with other ASCE group hexametric ATPases (6, 72, 197, 198).

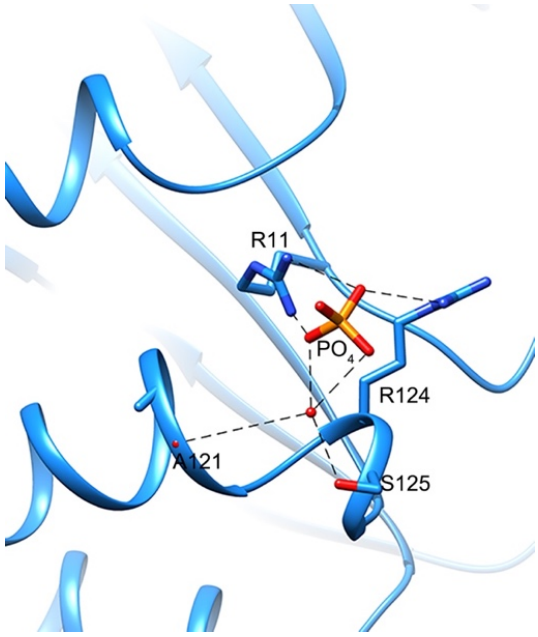


Figure 2-23 A phosphate binds to the conserved R124 in the structure of the gp14-phosphate complex.

The protein is shown as ribbon and coloured in blue. The phosphate ion and coordinated sidechains are displayed as sticks with atoms coloured as follows: oxygen - red, phosphor - orange, nitrogen - blue. H-bonds shown by dashed lines.

In the gp14-phosphate structure, a phosphate is bound to the sidechains of R11 and R124 (Figure 2-23), indicating that these two arginine residues likely serve as a phosphate binding site. This phosphate is also bound to A121 and S125 through water-mediated interactions (Figure 2-23). Interestingly, R124 is conserved across FtsK/HerA family ATPases (Figure 2-17a and 2-18), suggesting this conserved arginine is likely the 'trans-acting arginine finger' playing the role of a catalytic sensor for assisting the scissile phosphate release during ATP hydrolysis, in common with other hexametric ASCE ATPases (Figure 2-22).

2.3.7 A model of the phiKo gp14 hexamer

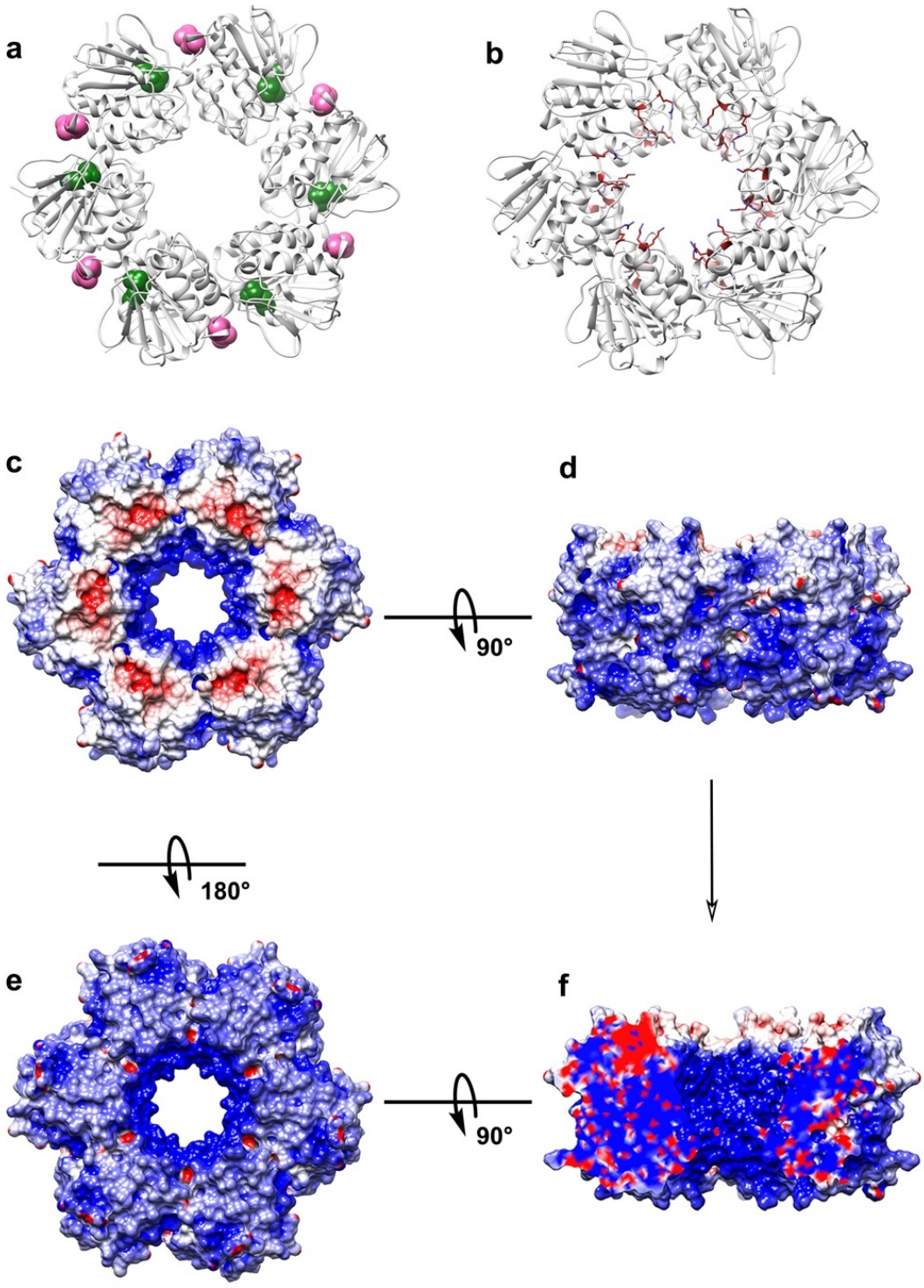


Figure 2-24 Predicted model of the phiKo gp14 homo-hexamer.

a. Ribbon diagram of the model for the hexamer. The homo-hexamer of gp14 was predicted by GalaxyTongDock (201), with the protein in light grey and solvent excluded surfaces of the conserved R124 and K22 in pink and forest green, respectively. b. Ribbon diagram with the positively charged residues (arginine and lysine) located at the surface

of central channel shown as sticks and coloured in brown. c-f. Orthogonal views of the electrostatic surface potential of the hexamer. The docked structure was assigned charges by PDB2PQR (191) using SWANSON force field (192). Electrostatics potential surfaces of phiKo gp14 structures (red, negative to blue, positive, ± 10 kT) were calculated by APBS (193) in Chimera (174).

To investigate the oligomerisation possibilities of phiKo gp14, a protein-protein docking method to generate a homo-oligomer was adopted for further analysis. The homo-oligomer of gp14 was modelled by rigid-body docking (202) with the monomeric gp14 structure as the input (Figure 2-24). The possible homo-hexamer of phiKo gp14 manifests the ATP binding site at the interface between two adjacent subunits, and the conserved R124 (Figure 2-18) protrudes towards the ATP binding pocket (Figure 2-24a). Positively charged residues, R77, K79, K139, R162 and R174, are located at the surface of the centre channel with sidechains exposed into the solvent (Figure 2-24b). One or several of these residues are likely to contact with DNA during DNA translocation.

To further analyse the docked hexameric model of the phiKo gp14, the charge distribution of the protein surface was investigated (Figure 2-24c - 2-24f). The top surface of the hexamer is mostly negatively charged (Figure 2-24c), and the side and the bottom surfaces of phiKo gp14 are largely positively charged (Figure 2-24d and 2-24e). The central channel of gp14 hexamer is also highly positively charged, indicating the potential for protein-DNA interaction (Figure 2-24f).

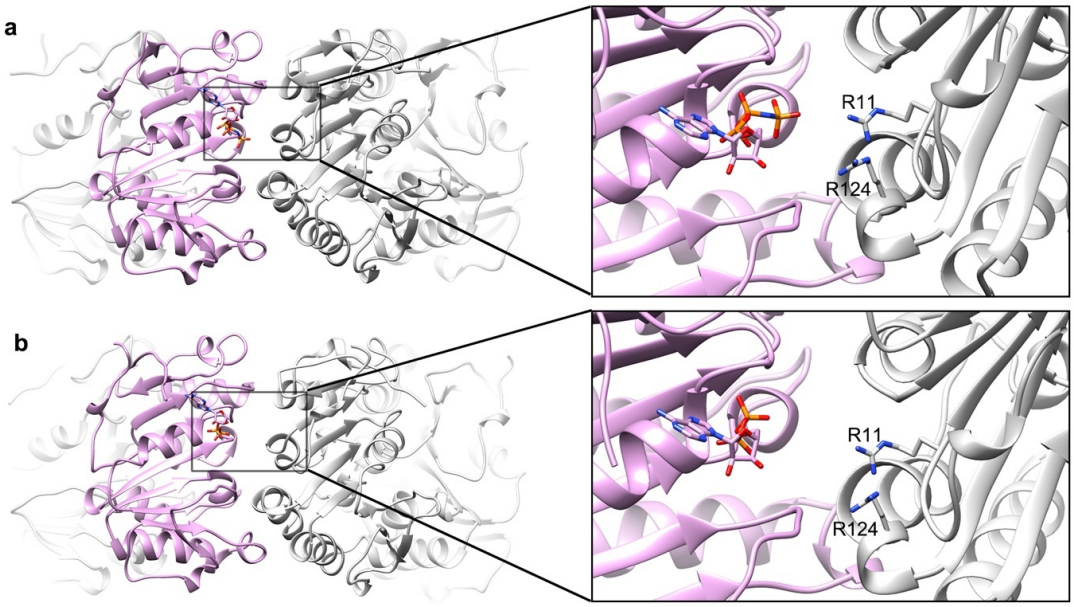


Figure 2-25 Location of R124 in the docked model of phiKo gp14 homo-hexamer.

Side views at the two adjacent subunits of the hexamer. a. gp14-AMPPNP complex; b. Same view as (a) but for the complex with ADP. The left subunit is in pink, and the right subunit is in light grey.

Significantly, the sidechain of R124 from the neighbouring subunit is facing towards the scissile phosphate of the bound AMPPNP and is close to the bound ADP in the hexamer models (Figure 2-25). This structural feature likely illustrates the possible linkage between ATP hydrolysis and protein oligomerisation, in common with other FtsK/HerA ATPases (72, 197). In addition, R11 residue may also be involved in ATP hydrolysis as a catalytic sensor, similar to several other ASCE group ATPases that use two '*trans*-arginine fingers' for ATP hydrolysis (83, 197).

2.3.8 Characterisations of protein stability of phiKo gp14

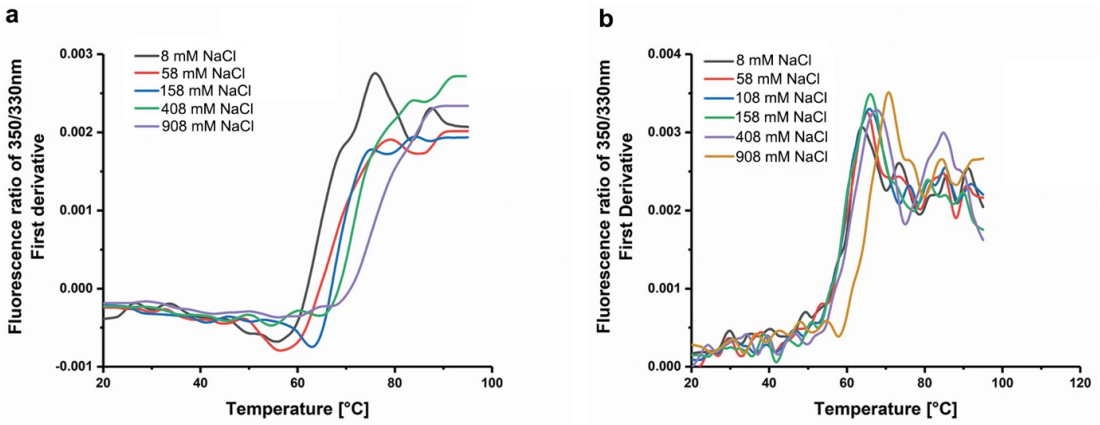


Figure 2-26 Effects of salt concentration on thermal unfolding profiles of phiKo gp14. a. Effects of salt concentration on thermal unfolding profiles of phiKo gp14 at pH 7.5. b. Effects of salt concentration on thermal unfolding curves of phiKo gp14 at pH 5.0.

Table 2-11 Thermal unfolding midpoint temperatures (T_m) of phiKo gp14.

	Tris buffer pH 7.5	**Citrate buffer pH 5.0**
NaCl concentration	T_m (°C)	T_m (°C)
8 mM	55.1	64.5
58 mM	57.0	64.9
108 mM	-	65.8
158 mM	62.3	65.9
408 mM	53.7	67.4

908 mM	56.7	70.8

Data derived from the same measurements in Figure 2-26.

The thermostability of phiKo gp14 was measured by NanoDSF, as described in section 2.2.7. The label-free nanoDSF technique determines the thermal unfolding transition midpoint (melting point, T_m) by measuring the intrinsic dual-UV fluorescence change in tryptophan and tyrosine residues at emission wavelengths of 330 and 350 nm (203). The phiKo gp14 protein contains 12 tryptophan and 11 tyrosine residues in its 203 amino acids, hence, the recorded changes in the intrinsic fluorescence are expected to accurately reflect T_m, which was obtained for each sample from the plot of inflection points in the first derivative curve of the ratio of emission intensities ($F_{350/330}$) against temperature (Figure 2-26). The results of the thermal unfolding profiles show that phiKo gp14 has higher T_m values in citrate buffer pH 5.0 than in Tris buffer pH 7.5 (Table 2-11), suggesting that the protein is more stable at pH 5.0. Consistently, crystals of phiKo gp14 were grown in the condition containing 0.1 M citric acid pH 4.0 - 5.0. Furthermore, at pH 5.0, a higher concentration of NaCl results in enhanced stabilisation of phiKo gp14 by increasing the melting temperature from 64.5 to 70.8 ˚C (Table 2-11).

2.3.9 Characterisations of ATPase activity of phiKo gp14

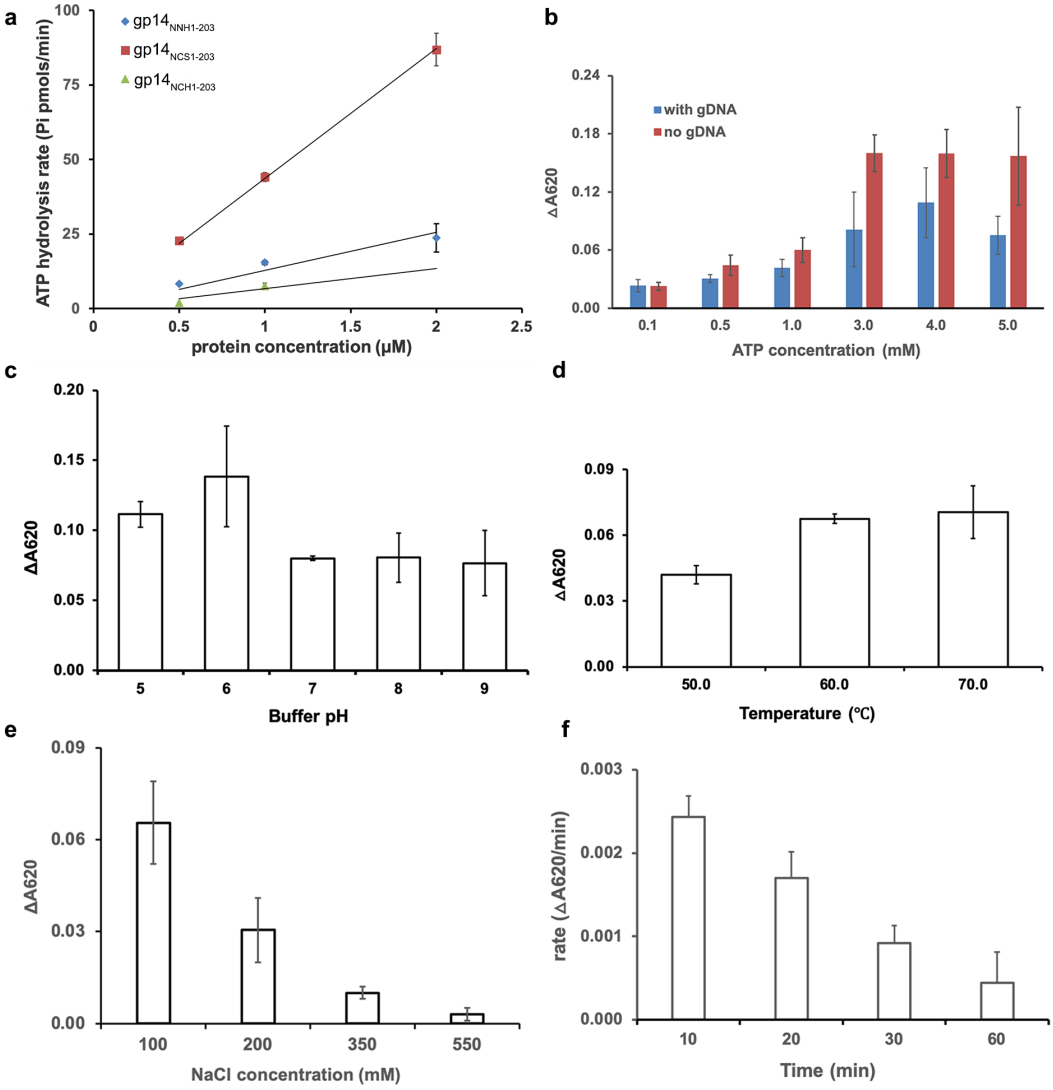


Figure 2-27 ATPase activity of phiKo gp14.

a. ATP hydrolysis rate of three protein constructs: untagged phiKo gp14 with native N-terminus ($gp14_{NCS1-203}$, post-cleaved by SUMO protease), untagged phiKo gp14 with the post-cleavage G-P-A residues at N-terminus ($gp14_{NCH1-203}$) and 6 × His tagged phiKo gp14 ($gp14_{NNH1-203}$). The linear fit for each protein construct was set to pass through the zero intercept. The icons show the averaged data of sample replicates, and the error bars show plus/minus standard deviation (SD); b. Effects of ATP concentration on the ATPase activity of phiKo gp14; c. Effect of pH on ATPase activity of phiKo gp14; d. Effects of reaction temperature on ATPase activity of phiKo gp14; e. Effects of NaCl concentration on ATPase activity of phiKo gp14; f. Effects of reaction/incubation time on ATPase activity of phiKo gp14. The columns in the graph denote the mean of each datapoint replicates while the error bars on each column show plus (panel f) or both plus and minus

(panel e) standard deviations (SD). Each sample has two replicates, except for the data in (f), which has three replicates.

Based on the structural observations of phiKo gp14, the N- and C- termini are quite close to each other (Figure 2-17), and the C-terminus is involved in nucleotide-binding (Figure 2-21). To find out whether the addition of the expression tag protein at the N-terminus affects phiKo gp14 ATPase activity, an initial experiment was performed by comparing three constructs of phiKo gp14: untagged phiKo gp14 with native N-terminus ($gp14_{NCS1-203}$), untagged phiKo gp14 with remaining post-cleavage G-P-A residues at the N-terminus ($gp14_{NCH1-203}$) and a 6 × His tagged phiKo gp14 ($gp14_{NNH1-203}$) (Figure 2-27a). The ATP hydrolysis rate was determined from the absorbance value at 620 nm (A620) using an inorganic phosphate (Pi) standard curve (Appendix 5). KH_2PO_4 was used to generate the Pi standard curve (Appendix 5). The results show that the untagged phiKo gp14 ($gp14_{NCS1-203}$) is the most active protein amongst the three, while the non-tagged protein with remaining G-P-A residues at N-terminal end is the weakest (Figure 2-27a), showing that the additional residues from the fusion construct slightly impact the ATPase activity of gp14. These subtle changes in the ATPase activity of gp14 indicate that the N-terminal end of gp14 may play a role in ATP hydrolysis. Also, a protein concentration-dependent increase in the rate of ATP hydrolysis was observed for gp14 (Figure 2-27a), though the highest hydrolysis rate is still low (43.5 pmols/μM/min), compared to other ATPases (85, 204). This low ATP hydrolysis rate may be partially due to the poor solubility of phiKo gp14.

Accordingly, further experiments were performed with the untagged phiKo gp14 construct with native N-terminus ($gp14_{NCS1-203}$). An increase in absorbance at 620 nm (A620) was observed when increasing the ATP concentration (Figure 2-27b). The phiKo gp14 is most active when adding 3 mM ATP, while the A620 value drops slightly when the ATP concentration is greater than 3 mM. The decrease in A620 may be due to the higher concentration of ATP, causing the protein to precipitate or form inappropriate assemblies. Unexpectedly, the addition of phiKo genomic DNA slightly decreased the activity of phiKo gp14 (Figure 2-27b).

The optimal activity of phiKo gp14 was observed at pH 5 and 6 (Figure 2-27c). In addition, the measured A620 increased when the incubation temperature was increased from 50 ˚C to 70 ˚C (Figure 2-27d). When considering the thermal unfolding profiles of phiKo gp14 (Figure 2-26 and Table 2-11), the optimal reaction temperature is 50 ˚C as it is lower than the T_m values of phiKo gp14. Although the higher concentration of NaCl thermodynamically stabilises phiKo gp14 protein (Figure 2-26 and Table 2-11), a high concentration of NaCl is harmful to its ATPase activity (Figure 2-27e) by causing a decrease in the ATP hydrolysis rate. Finally, increasing the incubation time from 10 min to 60 min, resulted in a 5-fold decrease of the ATP

hydrolysis rate, indicating that phiKo gp14 might precipitate or aggregate over time or, potentially, misassemble during ATP turnover.

2.3.10 Characterisation of DNA binding of phiKo gp14

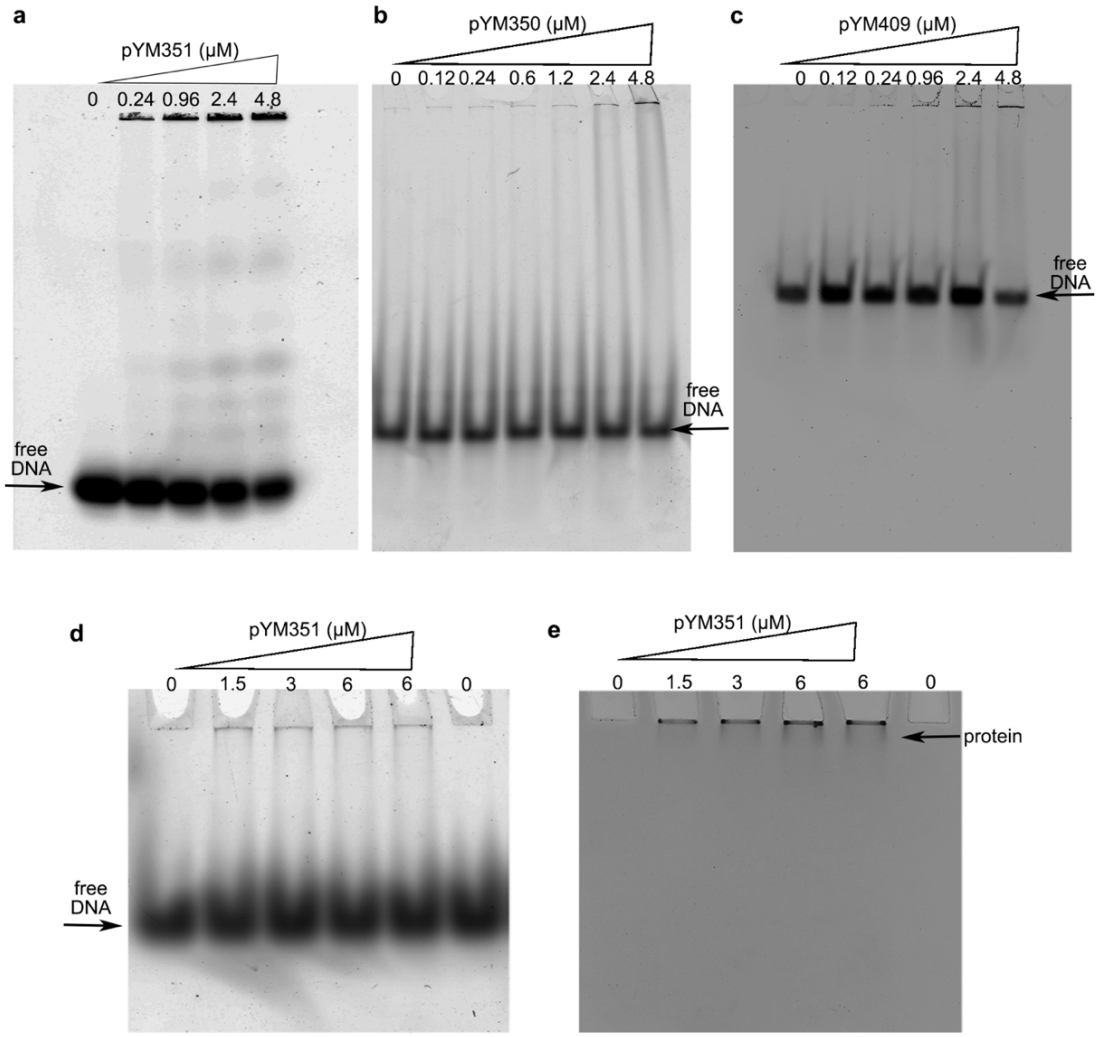


Figure 2-28 DNA binding activity of phiKo gp14.

a. Probing DNA binding of 6 × His tagged phiKo gp14 ($gp14_{NNH1\text{-}203}$) by Electrophoretic mobility shift assay (EMSA); b. DNA binding of untagged phiKo gp14 ($gp14_{NCS1\text{-}203}$) protein; c. DNA binding of untagged phiKo gp14 with the post-cleavage G-P-A residues at the N-terminus ($gp14_{NCH1\text{-}203}$); d and e. DNA binding assays for series of dilutions of 6 × His tagged phiKo gp14 ($gp14_{NNH1\text{-}203}$) visualised at a fluorescence emission wavelength of 635/650 nm (panel d) and stained in Coomassie blue dye (panel e).

In order to investigate the DNA binding activity of phiKo gp14, the electrophoretic mobility shift assay (EMSA) was performed for the purified gp14 protein and dsDNA (the sequences

are given in Appendix 6). First, binding of the 6 × His tagged phiKo gp14 ($gp14_{NNH1-203}$) to a 30 bp dsDNA was investigated. The result shows that the binding activity of phiKo gp14 is too weak to quantify. When the protein concentration was increased, a considerable amount of protein precipitated in the wells and did not run into the gel, similar to the other fusion proteins of phiKo gp14, $gp14_{NCH1-203}$ and $gp14_{NNH1-203}$ (Figure 2-28). Multiple weak bands are observed above the free DNA band (Figure 2-28a), may suggest that multiple protein molecules bind to the same DNA, and protein recognises DNA non-specifically. The weak binding is probably due to the hydrophobicity of the surface of phiKo gp14 (Figure 2-29) that may potentially cause improper protein assembly and protein precipitation. For the three constructs ($gp14_{NNH1-203}$, $gp14_{NCH1-203}$ and $gp14_{NNH1-203}$) of phiKo gp14, a smeary region was observed above the free DNA band on the 2.5% polyacrylamide gel (Figure 2-28b -2-28d), indicating that phiKo gp14 may dissociate from the DNA gradually, or the protein precipitation somewhat prevents the free DNA migration into gel during electrophoresis. The 0.6% agarose gel appears to be better than the native polyacrylamide gel for this assay, as it contains larger pores for protein-dsDNA complex migration during electrophoresis. 16 bp and 20 bp dsDNA were also used for binding assays, but only a weak smear staining in the lane was observed, similar to the results of 30 bp dsDNA (data are not shown).

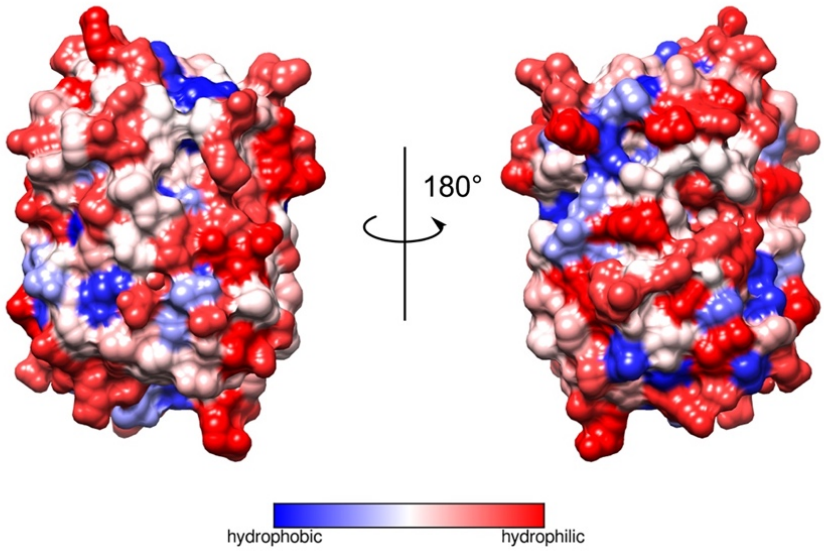


Figure 2-29 Hydrophobicity surface of phiKo gp14.

Blue is for the most hydrophobic residues, and red is for the most hydrophilic residues. The hydrophobicity of the surface was calculated in Chimera (174).

2.3.11 Characterisation of the DNA packaging activity of phiKo gp14

2.3.11.1 Isolation of phage phiKo

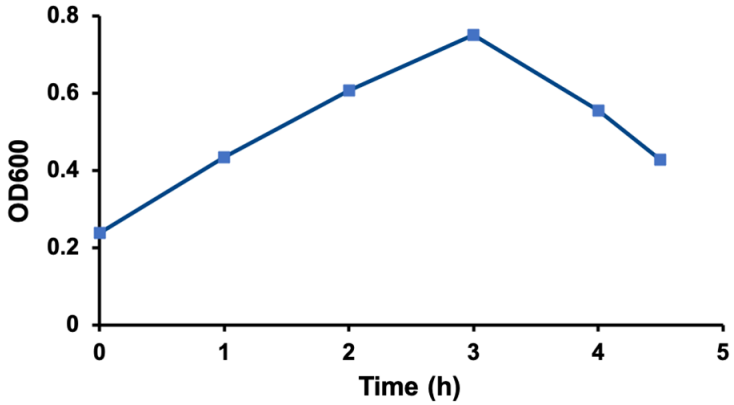


Figure 2-30 Growth curve of phage phiKo.

OD600 of the post-infection phage culture was measured at different time points (every 1 hour). 0 h was the time point when phage stock or phage plaques were injected into the host cell culture.

In order to characterise the DNA packaging activity of phiKo gp14, the production of empty capsids and mature virions of phage phiKo was investigated. Phage growth experiments of phage phiKo were conducted using *Thermus thermophilus* HB 27 cells as the host. The latent period of phiKo was about 3 h (Figure 2-30), and phiKo is a lytic phage, like other *Tectiviridae* family viruses (205). The latent period is the time taken by a phage particle to reproduce inside the host cell after entry. It begins at the point that the virions attach to the host cell and/or nucleic acid uptake and ends with infection termination and virion release (206, 207).

2.3.11.2 Purification of phage phiKo

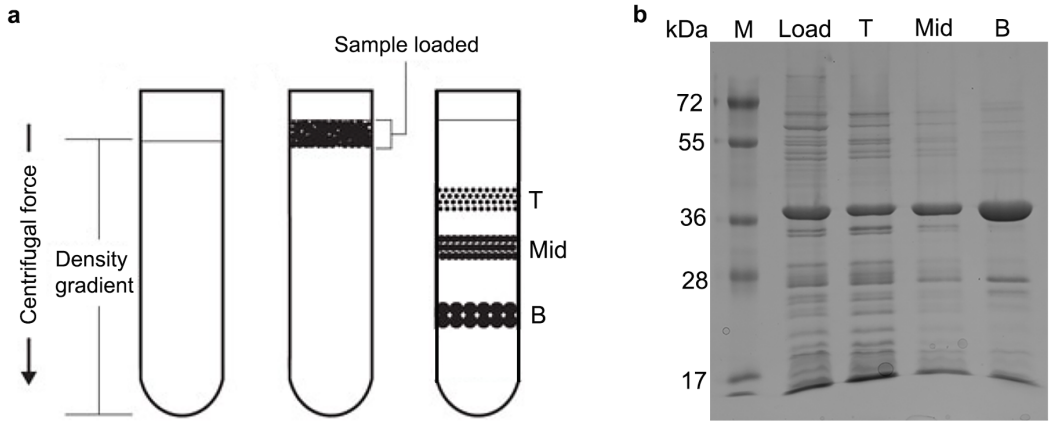


Figure 2-31 Purification of phage phiKo by density gradient ultracentrifugation.

a. Diagram of density gradient ultracentrifugation for phage phiKo (not to scale); b. Fractions collected from density gradient ultracentrifugation were resolved by SDS-PAGE. The SDS-polyacrylamide gel was stained with Coomassie brilliant blue dye. M - protein molecular mass marker, Load – the loaded sample before ultracentrifugation, T – sample of the top zone (fraction), Mid – sample of the middle zone (fraction) and B – sample of the bottom zone (fraction).

The phage sample was isolated as described in the section 2.2.10.2 and was initially purified by rate zonal density gradient ultracentrifugation. Three zones were observed on the density gradient column as depicted in the diagram above (Figure 2-31a). The samples were collected from each fraction (zone) and resolved by SDS-PAGE. All three zones contained a band at ~ 36 kDa (Figure 2-31b), which was considered to be the major capsid protein [(molecular weight of 37.2 kDa as calculated by ExPASy ProtParam (179)], showing all these fractions contained a considerable amount of the capsids. The samples from the middle and bottom zones contained fewer contaminating proteins than the top zone (Figure 2-31b). The materials collected from each fraction were separately further purified by size exclusion chromatography.

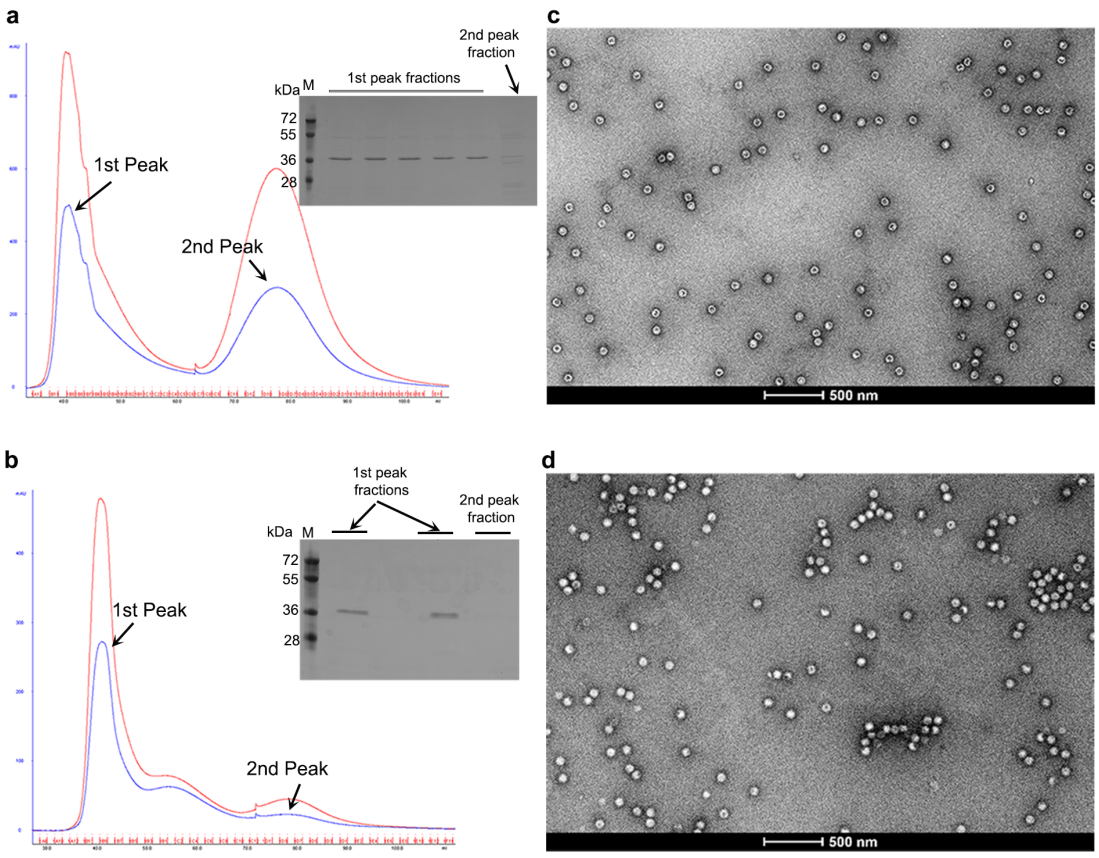


Figure 2-32 Purification of phage phiKo by size exclusion chromatography.

a. Chromatogram of the middle zone of phiKo phage sample obtained from density gradient ultracentrifugation and the eluted fractions were analysed by SDS-PAGE; b. Chromatogram of the bottom zone of phiKo phage sample obtained from density gradient ultracentrifugation and the eluted fractions were viewed by SDS-PAGE; c. Electron microscopy image of the eluted fractions of (a) containing phage phiKo (mostly empty capsids); d. Electron microscopy image of the eluted fractions of (b) containing phage phiKo (mostly filled capsids).

Phage phiKo belongs to the *Tectiviridae* family and likely has a similar virion conformation to other *Tectiviruses*, such as PRD1 (14, 15). Phage PRD1 has an external pseudo T = 25 icosahedral capsid shell composed of ~ 720 copies of the major capsids protein (14, 15). Accordingly, the assembled capsid of phage phiKo is also expected to contain ~ 720 capsids protein copies with the total estimated molecular weight of ~ 26 MDa. The phage phiKo possess 11 kb dsDNA linear genome, and its molecular weight is ~ 6.9 MDa. In addition, the fractionation range of the column (HiPrep Sephacryl S-500 HR column, GE Healthcare) applied in this study is 40 kDa – 20 MDa, smaller than the size of the phage particle, therefore, the phage particles are expected to be eluted in the void volume (mobile phase) of the column.

The eluted fractions of the samples from the middle and bottom zones of the density gradient column were first resolved by SDS-PAGE (Figure 2-32a and 2-32b), showing that the first peak that eluted at the early stage of the void volume contained a band with the approximately correct size for capsid protein, while the second peak contained protein contaminants. The fractions of the chromatography first peak were further analysed by negative stain transmission electron microscopy, as seen in Figure 2-32c and 2-32d. Transmission electron microscopic (TEM) images of purified phage from different fractions revealed that the empty capsids primarily existed in the middle zone of the density gradient column, and the filled phage particles were found mainly in the bottom zone (Figure 2-32c and 2-32d), due to the increased density of DNA packaged virions compared with the empty ones. Moreover, the TEM images demonstrate that phage phiKo is a tailless bacteriophage and morphologically identical to the widely studied *Tectivirus* phage PRD1 (Figure 2-33) (see section 1.1.1 for details of phage PRD1). The mature virion of phiKo has an outer protein shell (diameter of ~ 65 - 70 nm) and an inner membrane, encapsulating a genome (Figure 2-33).

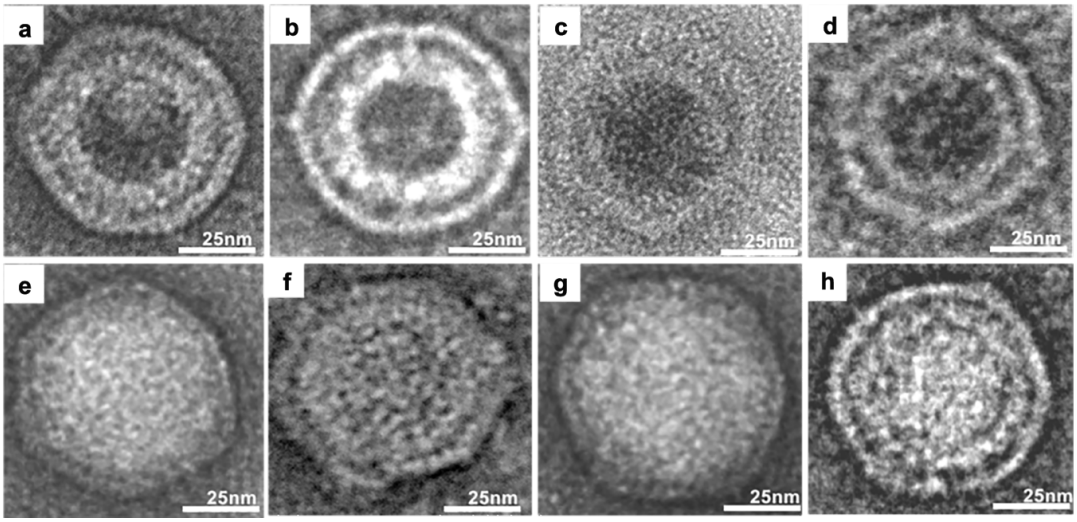


Figure 2-33 Morphology of phage phiKo.

a-d. Empty capsids of phage phiKo; e-h. Filled capsids of phage phiKo.

2.3.11.3 Characterisation of DNA packaging of phiKo gp14

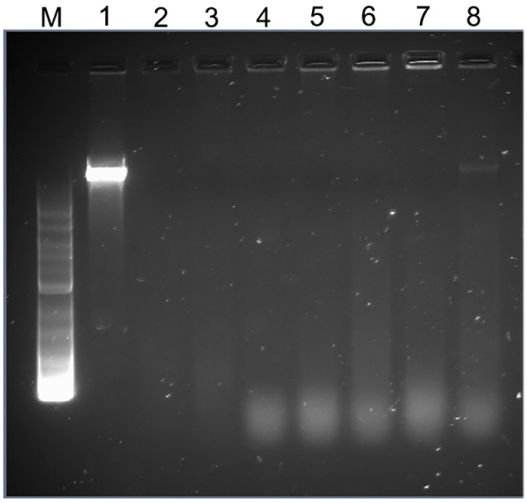


Figure 2-34 DNA packaging assay for phiKo gp14.

M indicates the DNA ladder, and gDNA indicates genomic DNA. 1. gDNA (input, negative control); 2 - 3. negative controls (empty capsids only); 4 - 5. gDNA + empty capsids; 6 - 7. gDNA + empty capsids + phiKo gp14 protein; 8. gDNA + phiKo gp14 protein.

The genomic DNA (gDNA) of phage phiKo was isolated and purified in this study (in Method and materials section 2.2.10.5), based on the method described previously for phage PRD1 gDNA isolation (32). The size of phiKo gDNA is ~ 11 kb (Figure 2-34). The purified empty

capsids in section 2.3.11.2 were used for DNA packaging assays. Despite intensive optimisation of the conditions for DNA packaging assays, no packaging activity of phiKo gp14 was observed as shown in the example displayed in Figure 2-34. Since the optimal growth temperature for the phage is 70 °C and the optimal temperature for ATPase activity of gp14 is 50 - 60 °C (in section 2.3.9), assays were carried out at 50, 60 and 70 °C, yet no packaging activity has been observed for gp14 so far. The poor solubility of phiKo gp14 may also limit the packaging activity, although the capsid protein is very soluble in low salt (even no salt) buffers. Since the DNA packaging ATPase, packaging efficiency factor and two membrane associated proteins, form the packaging unique vertex on the viral particles, like the *Tectivirus* PRD1 (14, 39, 41), as discussed in section 1.2.2.3 above. It is likely that one or several of these 'partner' protein components facilitate DNA packaging along with the oligomeric assembly of packaging ATPase gp14 in phage phiKo. Therefore, identification of the potential 'partner' protein is essential for understanding how this packaging machinery of phiKo works.

2.4 Discussion

Solubility of phiKo gp14 is low

The phiKo gp14 protein was successfully overexpressed, purified and crystallised in this study. The full length phiKo gp14 protein is relatively small, comprising 203 residues and has a molecular weight of 23.6 kDa, as seen in section 2.3.2. The protein solubility of phiKo gp14 is low, and the soluble materials of purified phiKo gp14 can be concentrated only to up to 2.5 mg/ml. Attempts at generating different fusion constructs (Table 2-2) to improve solubility have not been successful so far, nor was the experiment to produce oligomeric phiKo gp14, as described in sections 2.3.2.1 - 2.3.2.3. Although the recombinant cleavable SUMO-tagged and 6 × His tagged fusions helped to solubilize the phiKo g14 protein, a considerable amount of phiKo gp14 precipitated upon cleavage of the tag from the target fusion protein. The shorter non-cleavable 6 × His tag only slightly increased protein solubility, compared to the other cleavable SUMO and 6 × His tagged fusion constructs.

PhiKo gp14 is thermodynamically stable but loses ATPase activity over time

Protein stability usually refers to two parameters: kinetic and thermodynamic stabilities. The kinetic stability is characterised by an enzyme's half-life and is determined by following the decrease in enzyme catalytic activity over time (208). The thermodynamic (or conformational) stability is given by the resistance of a folded protein to denaturation by thermal unfolding, reflected in the melting temperature (T_m).

Since the phiKo gp14 is a putative DNA packaging ATPase, ATPase and DNA binding activities were investigated to obtain a better understanding of its function and stability. PhiKo is a *Thermus thermophilus* bacteriophage with an optimal growth temperature of 70 °C, as seen in section 2.2.10.1. Therefore, the putative packaging ATPase phiKo gp14 is expected to be stable at a relatively high temperature. Indeed, the thermal unfolding profiles show that phiKo gp14 is relatively stable, with a melting temperature that is higher than 53 °C in Tris pH 7.5 buffer and higher than 64.5 °C in citrate buffer pH 5.0 (Figure 2-26 and Table 2-11). Unexpectedly, the melting temperature of phiKo gp14 is lower than the optimal phage growth temperature, suggesting that other protein components may potentially stabilise gp14 within the same phage or its host cells at higher temperatures. In addition, well-diffracting crystals grew from the crystallisation conditions containing citrate ions at low pH 4.0 - 5.0, further suggesting that citrate and low pH buffer are potentially vital for stabilising gp14. Notably, buffers containing high NaCl concentration also increase the melting temperature of phiKo gp14 (Table 2-11), however, NaCl at high concentration is harmful to the ATPase activity, resulting in decreased ATP hydrolysis rates (Figure 2-27e). The effects of NaCl is likely due to the high ionic strength inhibiting phiKo ATPase activity. For future characterisation of the ATPase activity and DNA packaging activity of gp14, optimisation of conditions that balance salt concentration and gp14 ATPase activity may be needed.

The ATPase activity assays illustrate that phiKo gp14 loses ATPase activity over time (Figure 2-27f), likely due to poor protein solubility causing irreversible precipitation during the reaction, and to lack of interacting phage protein partners required for protein assembly. In addition, the DNA binding assays show that phiKo gp14 probably bind dsDNA weakly and non-specifically (Figure 2-28), though a large amount of phiKo gp14 precipitated and remained in the well of the native gel during the electrophoresis (Figure 2-28), preventing quantification of the binding data. Also, it has been challenging to reproduce the DNA binding activity of phiKo gp14, partially due to the poor solubility likely induced by mis-assembly of phiKo gp14. For future investigations of the ATPase and DNA binding activities of phiKo gp14, the presence of other protein partners may also be needed for appropriate oligomerisation of gp14.

Potential oligomerisation of phiKo gp14 protein

The phiKo gp14 was purified as a monomer in solution (as seen in sections 2.3.2.1 and 2.3.2.2) and appeared as a monomeric species in the crystals (Table 2-10). Analysis of protein docking of the monomeric phiKo gp14 structure suggests that protein oligomerisation likely requires some other factors, such as a catalytic sensor residue involved in transition state formation during ATP hydrolysis, projecting from the neighbouring subunit for the appropriate assembly of the phiKo gp14, or a ‘helper’ protein component within the same phage.

Since phiKo gp14 has a conserved FtsK/HerA fold (Figure 2-17), it is expected to function as an oligomer *in vivo*, most likely a ring-shape homo-hexamer, similar to other FtsK/HerA family ATPases (as mentioned in section 1.3.3) (6, 72, 197, 209). Therefore, a hexamer of phiKo gp14 was generated by molecular docking. The centre channel of the model hexamer of phiKo gp14 is large enough to be able to accommodate a B-form dsDNA with a ~ 23 Å diameter. In addition, previous cryoEM reconstruction analysis for the evolutionary related phage PRD1 suggested that the packaging ATPase and the packaging efficiency factor form a dodecameric portal complex (14). However, based on structural modelling analysis, phiKo gp14 is unlikely to form ring-like homo-dodecamers, due to the central channel of the modelled homo-dodecamer ring being ~ 80 Å (data not shown), unrealistically large for efficient DNA translocation. It is possible that the appropriate assembly of phiKo gp14 is completed by forming a higher-order complex with an auxiliary protein component within the same phage to assemble into a hetero-dodecamer.

N- and C- terminal segments of phiKo gp14 may facilitate oligomerisation

Structural observations show that C-terminal residues are involved in nucleotide binding (Figure 2-21). Moreover, the ATPase activity assays demonstrate that the phiKo gp14 construct with an N-terminal 6 × His tag and the gp14 protein with three post-cleavage N-terminal residues showed lower activity than the untagged phiKo gp14 protein (Figure 2-27a), suggesting that the N-terminus of gp14 may play a role in ATP hydrolysis. Future studies may benefit from expressing a native untagged construct of phiKo gp14 to probe their potential influence on oligomerisation. It is also possible that gp14 oligomerisation requires other 'helper' protein components and the appropriate cellular context, as the protein assembly may happen in cells.

Characterisation of DNA packaging of phiKo gp14

For tailless *Tectiviruses*, the DNA packaging vertex, located on the capsid, is comprised of four components, including two membrane associated proteins, a DNA packaging ATPase and a packaging efficiency factor (as discussed in section 1.2.2.3). The two membrane associated proteins are assembled on the unique vertex of the procapsid, and the DNA packaging ATPase and packaging efficiency factor proteins are attached to the packaging vertex on the capsids. They may initiate the DNA packaging, as mentioned in section 1.2.2.3. After packaging DNA into the procapsid, the DNA packaging ATPase remain on the virion (14, 41). Phage phiKo belongs to the *Tectiviridae* family and is expected to employ a similar unique vertex for DNA packaging.

In this thesis, a DNA protection assay was performed in an attempt to probe the DNA packaging activity of phiKo gp14, as described in section 2.2.10.6. The phage sample

containing empty capsids were purified and used for the DNA packaging assay. The observed lack of DNA packaging activity was likely due to the absence of the packaging efficiency factor or another packaging associated auxiliary protein. Since the packaging efficiency factor protein lacks sequence conservation, it has not yet been identified in the phage phiKo. Future study would benefit from the identification of the DNA packaging efficiency factor.

In addition, *Tectiviridae* family viruses, such as phage PRD1, use protein-primed DNA replication mechanism (as described in section 1.2.2.2), in which the terminal protein remains covalently attached to the 5' genome after DNA replication. The complex of terminal protein - genomic DNA is packaged into the procapsid through the unique vertex while the genome terminal protein may function as a valve that closes the channel once the genome is inside in the bacteriophage PRD1 (14). Likewise, in phage phiKo, there should also be a 'terminal protein' with functions similar to the one in PRD1. However, due to the lack of sequence conservation between terminal proteins, the putative terminal protein in phiKo has not yet been identified. The unsuccessful packaging assay may also be due to lack of terminal protein; hence, further investigation of the potential 'helper' protein would be helpful for investigation of the DNA packaging activity of gp14.

3 Structure and function of the NS3 helicase from Zika virus

3.1 Introduction

Zika virus NS3 helicase (NS3h) belongs to the superfamily 2 (SF2) helicases having the conserved helicase motifs spread over two RecA-like domains (69, 75, 79, 97, 112). Most helicases belong to superfamilies 1 and 2 and contain most or all of the conserved helicase motifs I-VI (69, 75). The conserved helicase motifs I (P-loop), Ia, Ib, Ic, II (DExx motif) and III are in domain 1, and helicase motifs IV, V and VI are in domain 2 (75, 79, 97, 112), as seen in section 3.3.3. These conserved helicase motifs are either responsible for RNA binding or NTP hydrolysis (79, 112). Moreover, the helicase motif V was thought to be responsible for coupling ATP hydrolysis and RNA unwinding in a DEAD-box helicase,Vasa, based on mutagenesis analysis (210).

Metal fluoride analogues of the phosphoryl group (PO_3^-) have been applied in a wide range of phosphoryl transfer reactions. These analogues provide insights into a mechanism for phosphoryl transfer with in-line geometry in the transition state (TS) (211, 212). The tetrahedral BeF_3^- ground state analogue (GSA) with octahedral AlF_4^- and trigonal bipyramidal MgF_3^- transition state analogue (TSA) complexes have enabled observation of molecular events that couple the catalytic steps of phosphoryl transfer to conformational changes (212, 213).

This chapter describes the structural basis for RNA translocation and NTP hydrolysis by the Zika virus NS3h. The NS3h protein was successfully overexpressed and purified *in vitro*. The purified NS3h protein was used for crystallisation. Since the co-crystallisation of NS3h with the AMPPMP and AMPPCP yielded apo form structure, lacking bound ATP analogues. In an attempt to obtain crystals for pre-hydrolysis and transition state complexes, the more stable and commonly used metal fluoride mimics were used for co-crystallisation. Crystal structures of NS3h were determined in its apo form and in complexes with several nucleotides, including pre-hydrolysis substrate analogues (ATPγS and ADP-BeF_3^-), a transition state analogue [ADP-$MgF_3(H_2O)^-$] and a post-hydrolysis product compound (ADP), as well as in a complex with a single-stranded RNA segment containing the 5' phosphate ($pRNA_9$). Due to the transition state analogue ADP-$MgF_3(H_2O)^-$ not having been observed before, this complex was subjected to careful examination by crystallography and ^{19}F NMR, to validate this unique ligand. Furthermore, additional biochemical characterisations, including protein stability, RNA binding and ATPase activity assays, were performed. The data presented here, unveil the structural and chemical events that accompany RNA unwinding.

3.2 Methods and materials

3.2.1 Protein production and purification

The full-length NS3 protein is multifunctional, comprising an N-terminal proteinase domain and a C-terminal helicase domain (as mentioned in section 1.2.3.2). The C-terminal helicase domain (residues 171-617 and 183-623) of NS3 from Zika virus was investigated in this thesis. The C-terminal helicase domain of the NS3 protein (residues 183 – 623, $NS3h_{183-623}$) corresponding to a Zika virus strain isolated in Brazil (NCBI accession code: ANC90426.1) (214) was produced using synthetic DNA (GeneArt, Thermofisher Scientific) cloned into YSBL-LIC (-) vector (153). The construct of the NS3h (residues 171 - 617, $NS3h_{171-617}$) was amplified from the same DNA of $NS3h_{183-623}$.

6 × His-tagged NS3h fusion proteins were expressed in Rosetta™ 2 (DE3) cells. Colonies carrying the recombinant target vector were picked and transferred in 10 ml freshly prepared LB medium containing 30 μg/ml kanamycin and 33 μg/ml chloramphenicol for overnight culture at 37 °C with shaking (180 rpm). The next day, the overnight culture was transferred to 10 L freshly prepared LB medium containing 30 μg/ml kanamycin and 33 μg/ml chloramphenicol. Cell cultures were incubated with shaking (220 rpm) at 37 °C. When the OD600 reached 0.6 - 0.8, IPTG was added to a final concentration of 1 mM and the culture incubated at 16 °C for over 16 h. Cells were harvested by centrifugation at 5,000 rpm, at 4 °C and resuspended in lysis buffer (Appendix 3) supplemented with 0.1 mg/ml lysozyme and the protease inhibitors (0.5 mg/ml leupeptin, 100 mM AEBSF and 0.7 mg/ml pepstatin A). The resuspended cells were disrupted by sonication.

The soluble materials were purified by Nickel affinity chromatography using a HisTrap HP column (GE Healthcare) pre-equilibrated with binding buffer (Appendix 3) and eluted with a gradient of imidazole (30 - 500 mM) in the same binding buffer. The 6 × His-tagged NS3h fusion protein was further purified by size exclusion chromatography using a Superdex 200 HiLoad 16/60 column (GE Healthcare) equilibrated with the SEC buffer (Appendix 3). Protein purity was verified by SDS-PAGE. Since the $NS3h_{183-623}$ construct contains six cysteine residues, 2 mM of the reducing agent, DTT, was added to the purification buffers. The half-life of DTT is about 1.4 hours at pH 8.5, 20 °C, and it decreases further with increasing temperature (215). Bearing this in mind, DTT was added freshly in the buffers at the beginning of each purification step. The relatively pure fractions of target protein were collected and concentrated using a Vivaspin™ concentrator (MWCO 30 kDa, Sartorius) at 4°C, 4,500 rpm in a centrifuge (Sigma). Protein concentration was determined by absorbance at 280 nm (A280) using a NanoDrop™ spectrophotometer (Thermo Scientific) based on the theoretical extinction

coefficient of 65890 $M^{-1}cm^{-1}$ for $NS3h_{183-623}$, same with $NS3h_{171-617}$, calculated by ExPASy ProtParam (179).

The production and purification procedure of $NS3h_{171-617}$ is similar to the $NS3h_{183-623}$ construct with slightly different buffers adjusted based on the protein properties (buffers are given in Appendix 3).

3.2.2 Crystallisation screens and optimisations

In this study, the purified 6 × His-tagged $NS3h_{183-623}$ fusion protein construct was used for crystallisation. All complex crystals were obtained by co-crystallisation or by soaking with the details described below.

The crystallisation screens were performed by sitting-drop vapour-diffusion methods using 96-well SWISSCI® MRC 2-drop crystallisation plates. 10 mg/ml $NS3h_{183-623}$ protein was mixed with 1 mM AMPPCP, 4 mM $MgCl_2$ in the buffer of 20 mM Tris 7.5. 0.17 M NaCl and 2 mM DTT before setting up crystallisation. Crystallisation droplets containing 150 nl sample mixture and 150 nl reservoir solution were prepared by Mosquito® liquid handling instrument (TTP Labtech) and equilibrated against 54 µl reservoir solution at 20 °C. Crystals grew in 10 days in the condition of 20% (w/v) polyethene glycol 3350, 0.1 M Bis-Tris propane 7.5 and 0.2 M Tri-sodium citrate. Then, they were mounted directly from the drop before being frozen in liquid nitrogen. X-ray data from the crystals grown in this condition was solved to reveal the apo form of NS3h.

The NS3h-$pRNA_9$ complex was formed by mixing $NS3h_{183-623}$ and $pRNA_9$ (sequence is given in Appendix 6) in a 1:1 molar ratio in buffer containing 20 mM Tris pH 7.5, 170 mM NaCl, 2 mM $MgCl_2$, 5 mM ADP and 2 mM DTT. The protein concentration in the final complex was 5 mg/ml. Crystals of NS3h-$pRNA_9$ complex grew in 20% (w/v) ethylene glycol, 10% (w/v) polyethylene glycol 8000, 0.1 M MES/Imidazole pH 6.5, and 20 mM of each of following salts: sodium formate, ammonium acetate, tri-sodium citrate, sodium potassium $_L$-tartrate and sodium oxamate. The NS3h-ADP complex was formed by adding 5 mM ATPγS, 5 mM $MnCl_2$ and 100 µM of a short hairpin RNA (RNA sequence is given in Appendix 6) into 5 mg/ml $NS3h_{183-623}$ in a buffer of 20 mM Tris 7.5, 150 mM NaCl and 2 mM DTT. The mixture was incubated at room temperature for 60 min before setting up crystallisation. The NS3h-ADP complex crystals grew in the crystallisation condition containing 20% (w/v) ethylene glycol, 10 % (w/v) polyethylene glycol 4000, 0.1 M MES/Imidazole pH 6.5, and 20 mM of each of following salts: sodium formate, ammonium acetate, tri-sodium citrate, sodium potassium $_L$-tartrate and sodium oxamate. Crystallisation trials for the NS3h-$pRNA_9$ and NS3h-ADP complex crystals were set up in a similar way as described for the apo NS3h crystals.

The crystal of NS3h-ADP-BeF_3^- complex was obtained by soaking the NS3h-ADP complex crystals in the solution containing 2 mM $MnCl_2$, 2 mM $BeCl_2$, 4 mM NH_4F, 2 mM ADP, 20% (w/v) glycerol, 10% (w/v) polyethylene glycol 4000, 0.1 M MES/imidazole pH 6.5, 20 mM sodium formate, ammonium acetate, 20 mM trisodium citrate, and 20 mM sodium potassium L-tartrate. Crystals were soaked for 5 min before flash-freezing in liquid nitrogen.

The NS3h-ADP-$MgF_3(H_2O)^-$ complex was pre-formed by mixing 5 mg/ml $NS3h_{183-623}$ with 20 mM ADP, 40 mM NH_4F, 0.3 M $MgCl_2$ and 2 mM TCEP. The sample mixture was incubated at room temperature for 30 min before setting up crystallisation trials. Micro-seeding was applied to obtain well-diffracting crystals and trials were set up using an Orxy8® protein crystallisation robot (Douglas Instruments Ltd.). The seeding stock was made using crystals of the NS3h-ADP complex. The crystallisation droplets containing 150 nl protein sample mixture, 50 nl crystal seeding stock and 100 nl reservoir solution were equilibrated against 54 µl of reservoir solution at 20 °C. Crystals of the NS3h-ADP-$MgF_3(H_2O)^-$ complex were formed in the crystallisation reservoir containing 15% PEG3350, 0.16 M sodium citrate and 9% ethylene glycol. To eliminate the possibility of Al^{3+} contamination in the crystallisation trials, the NS3h-ADP-$MgF_3(H_2O)^-$ complex crystals were fished from the original crystallisation drop and soaked in a solution containing 10 mM deferoxamine (a strong Al^{3+} chelator), 20 mM ADP, 40 mM NH_4F, 40 mM $MgCl_2$, 20 mM sodium citrate, 20 mM Tris pH 7.5, 150 mM NaCl, 10% PEG3350 and 20% glycerol. X-ray diffraction data were collected from crystals that grew in the original crystallisation condition and those soaked in the solution containing deferoxamine. The difference maps of the active site containing $MgF_3(H_2O)^-$ were compared between the X-ray datasets collected from deferoxamine treated and untreated crystals and showed no geometry differences.

3.2.3 X-ray data collection, structure determination and model building

X-ray data were collected at Diamond Light Source (Didcot), beamlines I02 and I03. X-ray datasets were processed with DIALS (216) for ZIKV NS3h-ADP-$MgF_3(H_2O)^-$ and NS3h-ADP complexes, and with XDS (165) for the apo NS3h, NS3h-$pRNA_9$ and NS3h-ADP-BeF_3^- complex structures. The structures of the NS3h-ADP-BeF_3^- and NS3h-$pRNA_9$ complexes were solved by molecular replacement with Phaser (167) using the structure of a Dengue virus NS3 helicase [PDB code 2JLQ (79)] as a search template followed by automated rebuilding with ARP/wARP (217). The apo NS3h, NS3h-ADP-$MgF_3(H_2O)^-$ and NS3h-ADP complex structures were refined using the structure of the NS3h-ADP-BeF_3^- complex as the starting model. All models were further improved through alternate cycles of manual rebuilding with Coot (170) and refinement with Refmac5 (171) using isotropic B factors for the ADP-BeF_3^- and NS3h-$pRNA_9$ complexes and using anisotropic B factors for the apo NS3h as well as

NS3h-ADP-$MgF_3(H_2O)^-$ and NS3h-ADP complex structures. Crystallographic calculations were conducted with CCP4i (218) and CCP4i2 (172).

For ligand structure refinement, the ligand dictionaries for BeF_3^- and $MgF_3(H_2O)^-$ were generated by JLigand (219) and were tuned based on previously described data (220–223). The defined ligand dictionary of ADP was created by Grade (224). The final ligand dictionaries for ADP-BeF_3^- and ADP-$MgF_3(H_2O)^-$ were generated by LIBCHECK in CCP4i suite (218). The $MgF_3(H_2O)^-$ ligand was fitted with a defined Mg-F bond length of 1.985 Å and Mg-O bond length of 2.069Å, resulting in Mg-F bond lengths of 1.86 Å, 1.86 Å and 1.92 Å and the Mg-O bond length of 2.02 Å after refinement. All figures for protein structure were generated in either CCP4MG (173) or Chimera (174).

3.2.4 ^{19}F NMR spectroscopy

In this study, fluorine-19 nuclear magnetic resonance spectroscopy (^{19}F NMR) was adopted for monitoring the formation of ZIKV NS3h-ADP-BeF_3^-, NS3h-ADP-AlF_4^- and NS3h-ADP-$MgF_3(H_2O)^-$ complexes.

Fluorine possesses an intrinsic nuclear magnetic resonance (NMR) sensitivity almost as high as protons (83%) (225). A fluorine nucleus in a molecule is, on average, surrounded by nine electrons, rather than a single electron, as is the case with hydrogen. The range and sensitivity of fluorine chemical shifts to the details of the local environment are much higher than hydrogen. A fluorine spin typically is nearly within van der Waals contact of hydrogen nuclei when the fluorine is present in a biological macromolecule (226). For a fluorinated small molecule that binds at an active site, the electric fields, short-range contacts and hydrogen-bonding possibilities experienced by the fluorine are different in the free versus ligand-bound states, and these will be reflected in a change of the chemical shift value (225).

In addition, a ^{19}F signal from a water-soluble molecule will experience chemical shift changes when the solvent is changed from water (H_2O) to the deuterium oxide (D_2O). By replacing H_2O with D_2O, all exchangeable protons become replaced by deuterons, leading to a slightly changed chemical environment which may also bear influence on the ^{19}F H_2O/D_2O isotope shift (225). Solvent-induced hydrogen/deuterium primary isotope shifts (SIIS) are sensitive probes of the coordination environment of fluorine atoms. The number of H-bond donors can be assigned based on SIIS values of ^{19}F NMR resonances.

The NS3h-ADP-$MgF_3(H_2O)^-$ complex was formed by mixing 500 μM NS3h with 10 mM $MgCl_2$, 20 mM NH_4F, and 10 mM ADP or GDP in the buffer containing 20 mM Tris pH 7.5, 300 mM NaCl and 2 mM DTT. The NS3h-ADP-$MgF_3^-(H_2O)$-ssRNA complex was formed by

adding 0.13 mM ssRNA (RNA sequence is given in Appendix 6) into the sample in the same reaction condition but containing 0.1 mM NS3h. Each reaction volume was 500 µl.

^{19}F NMR spectra were recorded at 298 K on a Bruker Avance 500 MHz spectrometer (operating at 470.38 MHz for fluorine) equipped with a 5 mm QXI probe with z-axis gradients. Spectra recorded using 1 mM TCEP to replace DTT showed no difference. ^{19}F NMR spectra were recorded in Manchester Institute of Biotechnology with the assistance from Dr. Yi Jin (University of Cardiff) and Dr. Matthew Cliff (University of Manchester). The solvent-induced isotope shifts (SIIS) of the ^{19}F resonances were measured by comparing spectra for samples separately prepared in buffers containing either 10% D_2O or 100% D_2O. SIIS is defined as the chemical shift difference for [$\delta\,^{19}F$ (90% H_2O buffer) - $\delta\,^{19}F$ (100% D_2O buffer)].

The ^{19}F NMR spectra of the $AlCl_3$ titration experiment to convert NS3h-ADP-$MgF_3(H_2O)^-$ into NS3h-ADP-AlF_4^- complexes were carried out by mixing the addition of 1 mM $AlCl_3$ and 5 mM $AlCl_3$ into a sample of NS3h-ADP-$MgF_3(H_2O)^-$ (buffer composition is the same as described above). Higher concentrations (greater than 5 mM) of $AlCl_3$ cannot be achieved, given the severe precipitation of protein in the NMR tube caused by acidification resulting from $AlCl_3$ addition.

3.2.5 ATPase active site cavity measurement

To investigate the differences of ATPase active site with $MgF_3(H_2O)^-$ bound and with MgF_3^- bound, the active site cavity volume of ZIKV NS3h-ADP-$MgF_3(H_2O)^-$ and RhoA-GDP-MgF_3^--RhoGAP complex structures were calculated and compared. RhoA is a GTPase and regulates phosphorylation pathways that control a range of biological functions including cytoskeleton formation and cell proliferation. RhoGAP, a GTPase-activating protein that stimulates Rho family proteins and forms a complex with RhoA during hydrolysis of GTP (227). The solvent-exclusive surfaces of the ATP binding site of the NS3h-ADP-$MgF_3(H_2O)^-$ and RhoA-GDP-MgF_3^--RhoGAP (PDB code 1OW3) complexes were calculated with selected central residues of the ATPase active site (Table 3-1). The volume of the ATPase active site cavity for carrying scissile phosphate and β-phosphate was measured by Chimera (174).

Table 3-1 Residues selected for analysing the ATP/GTP β- and γ-phosphate binding pockets.

NS3h-ADP-$MgF_3(H_2O)^-$ TSA complex	RhoA-GDP-MgF_3^--RhoGAP TSA complex
PDB code 6S0J (this work)	**PDB code 1OW3**
P196.a	G12.b

G197.a	D13.b
A198.a	G14.b
G199.a	A15.b
K200.a	C16.b
T201.a	G17.b
R202.a	K18.b
V228.a	T19.b
E231.a	Y34.b
D285.a	V35.b
E286.a	P36.b
T316.a	T37.b
A317.a	D59.b
M414.a	T60.b
G415.a	A61.b
A416.a	G62.b
N417.a	Q63.b
Q455.a	R85.a
R459.a	
R462.a	

a is chain A and b is chain B in both structures. NS3h-ADP-$MgF_3(H_2O)^-$ TSA complex structure (PDB code 6S0J, this work) and RhoA-GDP-MgF_3^--RhoGAP TSA complex structure [PDB code 1OW3, (228)].

3.2.6 NADH-coupled microplate photometric assay

The NADH-coupled microplate photometric assay was applied to investigate the ATPase activity of NS3h since the attempt to obtain more accurate measurements with the Malachite green assay was unsuccessful which was partially due to NS3h being relatively thermodynamically unstable (as discussed in below section 3.3.6.1). The NADH-coupled ATPase assay is based on quantifying the decrease in absorbance of NADH (which absorbs strongly at 340 nm) as it is oxidized to NAD^+ (which absorbs strongly at 260 nm). It is called a 'coupled assay' because each time an ATP molecule is hydrolysed and a molecule of NADH is rapidly oxidized (Figure 3-1). Therefore, ATP hydrolysis and NADH oxidation are coupled (229). The assay makes use of the well-studied properties of the enzymes pyruvate kinase and lactate dehydrogenase, with the substrates ADP, phosphoenolpyruvate (PEP), and NADH. The reactions involved in the assay are given in Figure 3-1.

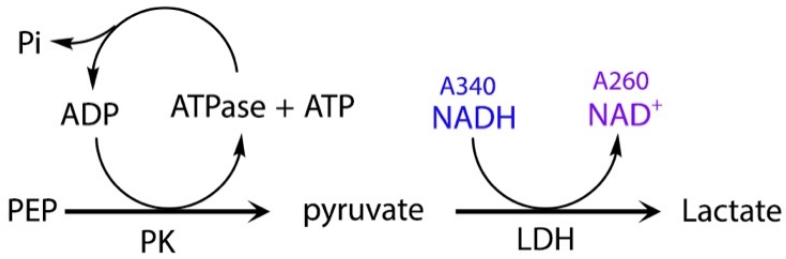


Figure 3-1 NADH-coupled ATPase activity assay.

PEP indicates phosphoenolpyruvate, PK indicates pyruvate kinase, Pi indicates inorganic phosphate, LCH indicates L-lactate dehydrogenase, NADH indicates the reduced form of Nicotinamide adenine dinucleotide (NAD) and NAD^+ indicates the oxidised form of NAD. ATP turnover by an ATPase, i.e. NS3 helicase, in this case, is linked to the oxidation of NADH to NAD^+ to observe the time-dependent decrease in absorbance at 340 nm.

To investigate whether freezing or storage influences the ATPase activity, fresh $NS3h_{171-617}$ and thawed $NS3h_{171-617}$ were utilised for NADH-coupled ATPase activity measurements. Reactions were performed in the condition containing 25 mM HEPES pH 7.2, 70 mM KCl, 5 mM $MgCl_2$, 1 mM TCEP, 5 µg/ml BSA, 6 - 10 U/ml pyruvate kinase (PK), 10 - 14 U/ml Lactic Dehydrogenase (LDH), 1 mM PEP and 0.5 or 1 mM NADH was used, with a total reaction volume of 100 µl. Fresh $NS3h_{171-617}$ indicates the freshly purified $NS3h_{171-617}$, and thawed $NS3h_{171-617}$ indicates the freshly purified $NS3h_{171-617}$ which was flash frozen in liquid nitrogen and stored at - 80 °C and then defrosted. Kinetics of 200 nM fresh and thawed $NS3h_{171-617}$ were determined by titrating different concentrations (0 to 1000 µM) of ATP in each well. For RNA stimulation assays, 50 nM NS3h was mixed with 5 µM 5' phosphorylated or non-phosphorylated $ssRNA_9$ in the same condition as described above. Assays were performed in a transparent flat-bottom 96-well microplate (Costar, Corning incorporated, USA) and recorded at 25 °C for 30 min with 30 s interval time in a Tecan Infinite 200 Pro microplate spectrophotometer (Switzerland). Buffer optical density values were corrected and used to calculate the NADH concentration using an extinction coefficient of 6.22 mM^{-1} cm^{-1}.

A calibration curve was used to convert the NADH absorbance readings at 340 nm into moles of ATP based on the 1:1 coupling of NADH to the ATP consumption in the course of the PK/LDH assay. The path-length of the 96-well plate was measured using 0.25 mM NADH using a modified procedure as described previously (230). The path-length curve was given in Appendix 8. The absorbance data measured at 340 nm (2 - 3 min with 20 - 30 s intervals) were averaged and used for analysis using the Michaelis-Menten equation (231), as implemented in OriginPro 2017 (OriginLab, Northampton, MA).

Michaelis–Menten kinetics is one of the best known models for understanding enzyme kinetics. It is named after German biochemist Leonor Michaelis and Canadian physician Maud Menten (231). The velocity of an enzymatic reaction, in terms of the rate of product formation, is given by the Michaelis-Menten equation:

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

Where v_{max} represents the maximum rate achieved by the system (which happens at saturating substrate concentration), [S] - the concentration of a substrate. The value of the Michaelis constant K_m is numerically equal to the substrate concentration at which the reaction rate is half of v_{max} (232).

3.2.7 Thermal shift assay

Two micromolar (μM) $NS3h_{183-623}$ was mixed with 2 mM $MgCl_2$, and 1 mM, 2 mM or 5 mM ATP analogue (ADP, AMPPNP, AMPPCP, ATPγS, AMP, GDP or GMPPNP), respectively. The volume of the sample mixture was 12.5 μl. The reaction buffer contained 20 mM Tris 7.5, 0.3 M NaCl, 2 mM DTT. 12.5 μl 2.5-fold dilution of SYPRO™ Orange dye (diluted from 5000-fold original stock, Molecular Probes, Inc.) was added to each reaction. Each sample had three replicates. Assays were recorded by increasing temperature from 25°C to 95 °C using the Stratagene Mx3005P system (Agilent Technologies). Data were analysed in IBM SPSS Statistics software.

3.2.8 Nano differential scanning fluorimetry (NanoDSF)

In order to measure the thermal unfolding profiles of NS3h in the presence of nucleotide, 2 mM ADP and ATPγS were added to 5 μM $NS3h_{171-617}$ in buffer containing 25mM HEPES 7.2, 70 mM NaCl, 1 mM TCEP and 5 mM $MgCl_2$, respectively. The measurements were carried out similarly to the measurements for gp14 as described above in section 2.2.7. Thermal unfolding curves were plotted in OriginPro 2017 (OriginLab, Northampton, MA). Each sample has duplex replicates.

3.2.9 RNA binding assays

3.2.9.1 Electrophoretic mobility shift assay (EMSA)

The RNA binding for NS3h was probed using Electrophoretic mobility shift assay (EMSA), as seen in section 3.3.6.3. In each reaction for EMSA, 25 μM $pRNA_9$ and 0 - 10 μM ZIKV $NS3h_{183-623}$ were mixed thoroughly in binding buffer containing 20 mM Tris 7.5, 150 mM $MgCl_2$ and 2 mM DTT. Reactions were incubated at RT for 20 min. 6% (w/v) polyacrylamide

gel (37.5:1 of acrylamide:bisacrylamide) was used for visualising results. The native gel was pre-run at 80 V for 40 min at 4 °C before loading the sample and was run at 80 V for 80 min at 4 °C and stained in 1 × SYBR™ gold dye (diluted from 10000 × stock) at 4 °C for 1 h in the dark. The gel was imaged using the GelDoc system (Bio-Rad) and further stained to visualise protein using the standard Coomassie brilliant blue solution.

3.2.9.2 Intrinsic tryptophan fluorescence quenching assay

Since the RNA binding assay performed using EMSA was unsuccessful (discussed in section 3.3.6.3), a relatively more sensitive approach, intrinsic tryptophan fluorescence quenching assay (TFQ)(233), was utilised for determining the NS3h-RNA binding equilibrium dissociation constant (K_d). Typically, a shift occurs from 330 nm to 350 nm in a protein indicating that the change of tryptophan from being entirely buried change to wholly exposed to water, increasing the fluorescence intensity ratio ($F_{350/330}$) (175). The increased $F_{350/330}$ is derived from the free RNA interacting with the protein near where tryptophan is located. When the increasing $F_{350/330}$ reaches a steady state, this indicates that the RNA-bound NS3h is saturating. The TFQ data was monitored by Nano differential scanning fluorimetry (NanoDSF), and the recorded data were analysed and fitted using the Hill equation (234).

The Hill coefficient is commonly used to estimate the number of ligand molecules that are required to bind to a receptor to produce a functional effect (235). The Hill equation is:

$$y = y_{max} \frac{x^n}{k^n + x^n}$$

Where k is the ligand concentration, at which half the receptors are ligand-bound, x is the ligand concentration, y is the effect elicited by the ligand at k, y_{max} is the maximal effect that can be induced by the given ligand in the given system, n is the Hill coefficient (234). When n = 1, the equation can be described by Michaelis-Menten kinetics (in section 3.2.6).

For RNA binding measurements, 5 μM $NS3h_{171-617}$ was titrated with 0 μM - 250 μM $pRNA_9$ in a buffer containing 25 mM HEPES 7.2, 70 mM NaCl, 1 mM TCEP and 5 mM $MgCl_2$. The readings for the changes in the ratio of fluorescence intensity 350/330 nm were averaged over 1 °C range (20 - 21°C) for each tested concentration of $pRNA_9$. Since there are no tryptophan or tyrosine residues located inside the RNA binding pocket, the decrease in the $F_{350/330}$ ratio is caused by a local conformational change of the tryptophan and tyrosine residues that occur during $pRNA_9$ binding, resulting in a shift in the fluorescence spectra. Each sample has six replicates. Data from three separate experiments were used to determine the mean and standard errors of the mean (SE), and these data were fitted into the Hill equation (236) to determine the equilibrium dissociation constant (K_d) and the Hill coefficient n using OriginPro 2017 (OriginLab, Northampton, MA).

3.3 Results

3.3.1 Production of the Zika virus NS3h

3.3.1.1 Solubility screening for NS3h

Cells expressing 6 × His tagged NS3 helicase protein (residues 183-623, $NS3h_{183-623}$) were disrupted by sonication using 20 different lysis buffers (Appendix 9). The solubility of this $NS3h_{183-623}$ protein in different buffers is similar (data not shown), except for buffers at pH 4, where the protein primarily precipitated. In order to obtain sufficient soluble material and optimal buffer condition for crystallisation for the $NS3h_{183-623}$ construct, the buffer containing 20 mM Tris 7.5, 0.3 M NaCl and 2 mM DTT was chosen for large scale purification.

3.3.1.2 Protein purification of NS3h

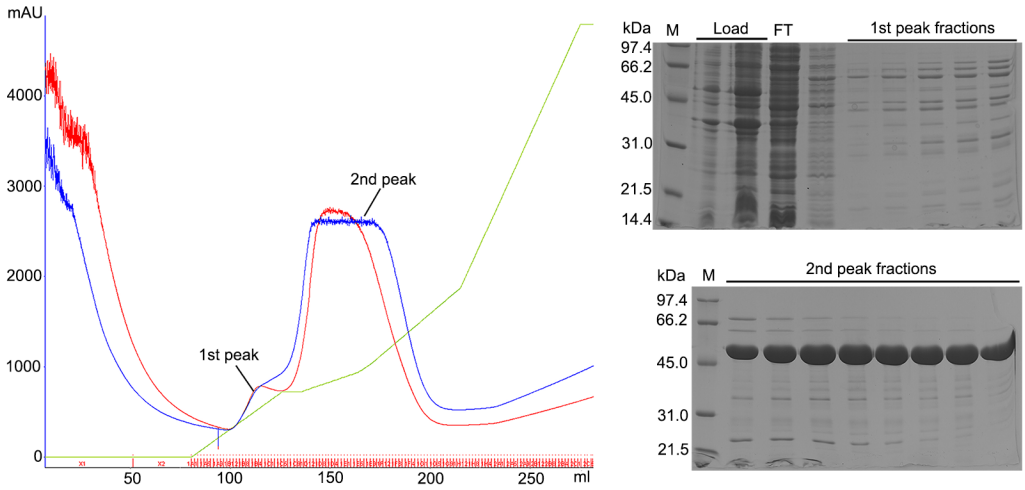


Figure 3-2 Nickel affinity chromatography of $NS3h_{183-623}$.
Left panel: Elution profile for $NS3h_{183-623}$ purification by Nickel affinity chromatography; right panel: the eluted fractions were resolved by SDS-PAGE. 'M' is the standard marker. 'Load' is the sample loaded for purification. 'FT' indicates the flow through. Samples of fractions collected from the first and second peaks corresponding to the left panel, are highlighted on the top of the gels.

The 6 × His-tagged $NS3h_{183-623}$ protein was initially purified by Nickel affinity chromatography. The protein sample with relatively accurate molecular size was eluted under the second peak by 30 - 40% imidazole on the chromatogram and verified by SDS-PAGE (Figure 3-2). The molecular weight of the 6 × His-tagged $NS3h_{183-623}$ protein is 50666 Da calculated by ExPASy ProtParam (179). The eluted fractions containing the approximately

expected size of the target protein were pooled together and concentrated to 1 - 5 ml volume for the next step purification.

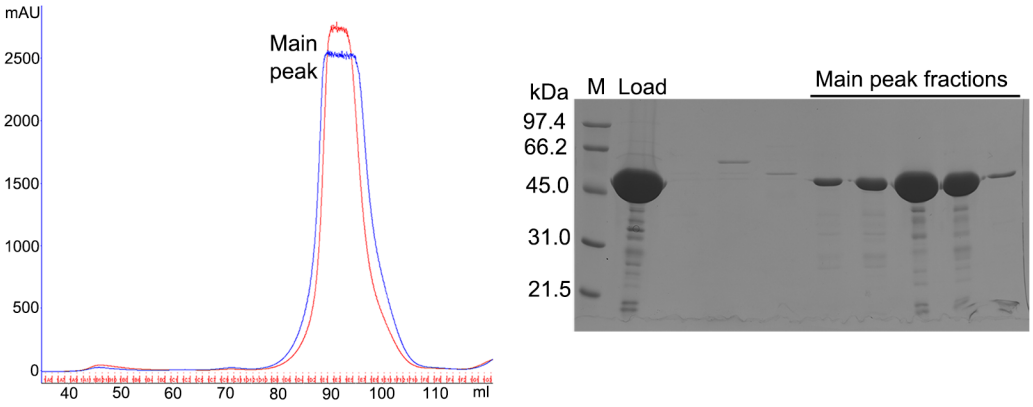


Figure 3-3 Purification of NS3h$_{183-623}$ by size exclusion chromatography.
Left panel: Size exclusion chromatogram of NS3h$_{183-623}$, right panel: SDS-PAGE for fractions later pooled for storage. 'M' is the standard marker, and the size standards are labelled next to the gel. 'Load' is the sample loaded for purification. Samples of fractions collected from the main peak corresponding to the left panel, are highlighted on the top of the gel.

The concentrated 6 × His-tagged NS3h$_{183-623}$ protein was further purified by size exclusion chromatography. The purified NS3h$_{183-623}$ protein was eluted at a column volume of 85 - 105 ml, as seen in the elution chromatogram (Figure 3-3, left panel), corresponding to the molecular mass of ~ 50 kDa, suggesting that NS3h$_{183-623}$ eluted as a monomer. The eluted peak fractions containing target protein (Figure 3-3, right panel) were pooled and concentrated to ~ 55 mg/ml. The purification yielded ~ 110 mg of NS3h$_{183-623}$ from 10 L cell cultures. Since the NS3 helicase protein was not particularly stable at room temperature, exhibiting slow precipitation, it was frozen in liquid nitrogen and stored in small 25 µl aliquots at - 80 $^{\circ}$C for subsequent experiments.

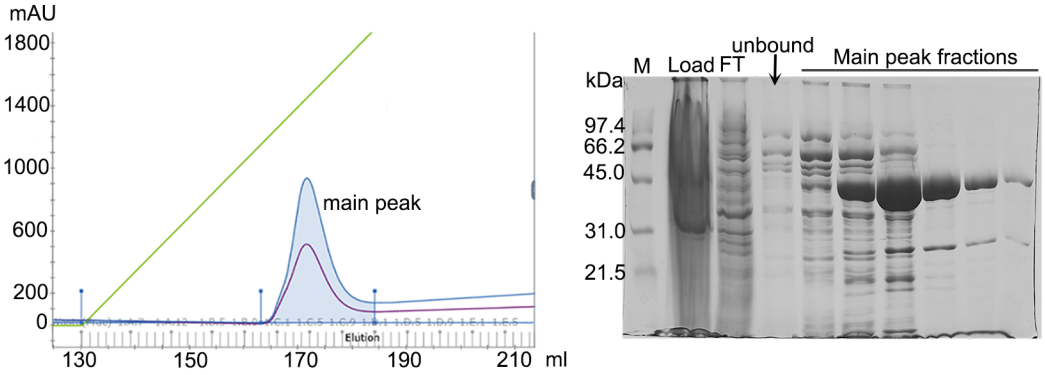


Figure 3-4 NS3h$_{171-617}$ purification by nickel affinity chromatography.

Left panel: Nickel affinity chromatogram of NS3h$_{171-617}$, right panel: analysis of eluted fractions by SDS-PAGE. 'M' is the standard marker. 'Load' is the sample loaded for purification. 'FT' indicates the flow through, 'unbound' indicates the lane containing unbound fractions. Samples of the fractions collected from the main peak referring to the left panel chromatogram, are highlighted on the top of the gel.

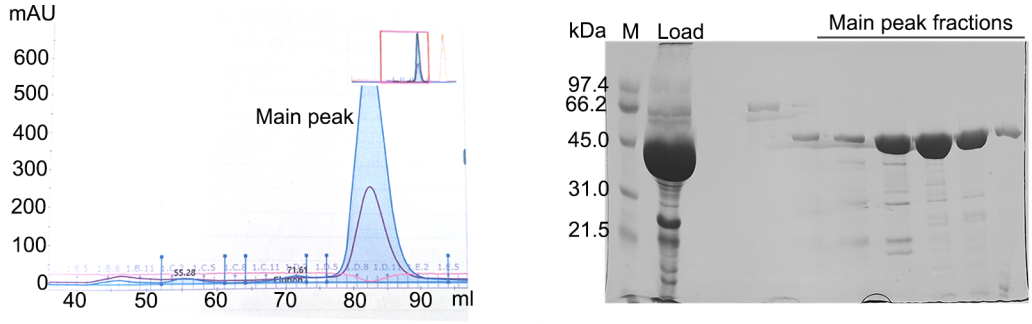


Figure 3-5 NS3h$_{171-617}$ purification by size exclusion chromatography.

Left panel: Size exclusion chromatogram for NS3h$_{171-617}$, right panel: SDS-PAGE for pooled fractions. 'M' is the standard marker, and the size standards are labelled next to the gel. 'Load' is the sample loaded for purification. Samples of the fractions collected from the main peak are highlighted on the top of the gel.

The 6 × His tagged 6 × His tagged NS3 helicase protein (residues 171-617, NS3h$_{171-617}$) was purified similarly to the construct for NS3h$_{183-623}$. The eluted fractions from the HisTrap HP column (GE Healthcare) containing approximately bands of ~ 50 kDa were collected (Figure 3-4). The concentrated target protein was further purified by size exclusion chromatography. As expected, the NS3h$_{171-617}$ protein was eluted as a monomer (Figure 3-5), similar to the NS3h$_{183-623}$ construct. Pure protein fractions were pooled and concentrated to ~ 5.9 mg/ml (~ 2.5 ml) from 2 L cell culture. Fresh protein was used for conducting the following biochemical

characterisations, and the rest of the protein sample was flash frozen using liquid nitrogen and stored at - 80 ˚C.

3.3.2 Crystallography

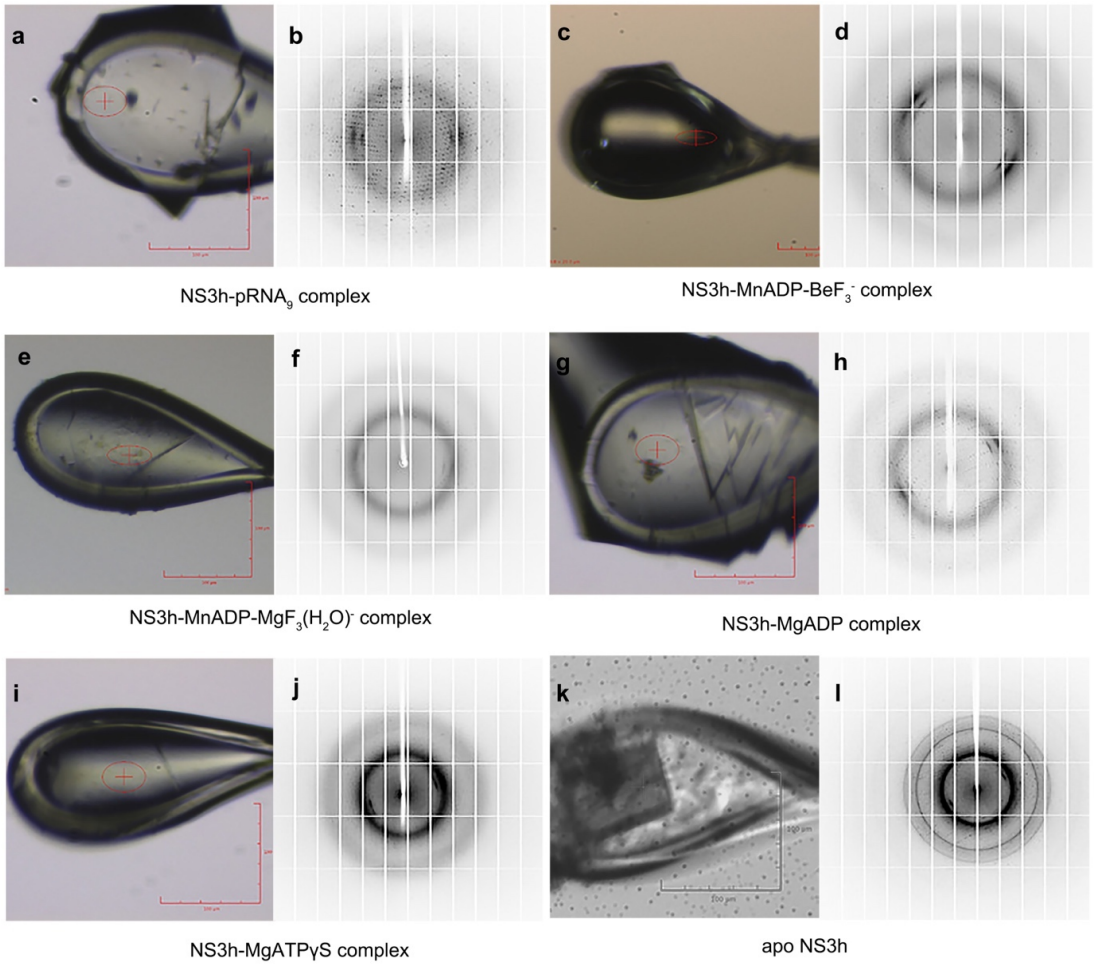


Figure 3-6 Crystals and X-ray diffraction images of the ZIKV NS3h.

a, c, e, g, i and k show crystals vitrified in nylon loops; b, d, f h, j and l are diffraction images.

The purified $NS3h_{183-623}$ protein was utilised to screen for suitable crystallisation conditions. In total, about 200 crystallisation trials were set up to identify conditions that produced X-ray quality crystals. The well-diffracting crystals of native NS3h, and complexes with $pRNA_9$, ADP, ATPγS, ADP-BeF_3^- and ADP-$MgF_3(H_2O)^-$ were obtained and applied for further crystallographic studies. The ADP is the product of ATP hydrolysis, the ATPγS and ADP-BeF_3^- are the substrate (ATP) mimics and the ADP-$MgF_3(H_2O)^-$ is the transition state mimic of ATP, as discussed in section 3.3.3.3 below. The complex crystals of NS3h-$pRNA_9$, NS3h-ADP-BeF_3^- and NS3h-ADP were obtained by co-crystallisation, NS3h-ADP-$MgF_3(H_2O)^-$ crystals were obtained by micro-seeding and the crystals of NS3h-ATPγS and NS3h-ADP-

BeF_3^- were produced by soaking (see details in section 3.2.2). Crystals of the native NS3h formed within ~ 10 days and crystals of NS3h complexes grew within 2 days. Most crystals grew as big thin plates, and the apo form NS3h crystals had a cubic shape (Figure 3-6).

For NS3h-ADP-$MgF_3(H_2O)^-$ complex, before micro-seeding, the complex crystals were too tiny to mount and gain useful diffracting data (Figure 3-7, left panel). Following micro-seeding, crystals grew as large thin plates which exhibited good diffraction quality (Figure 3-7, right panel).

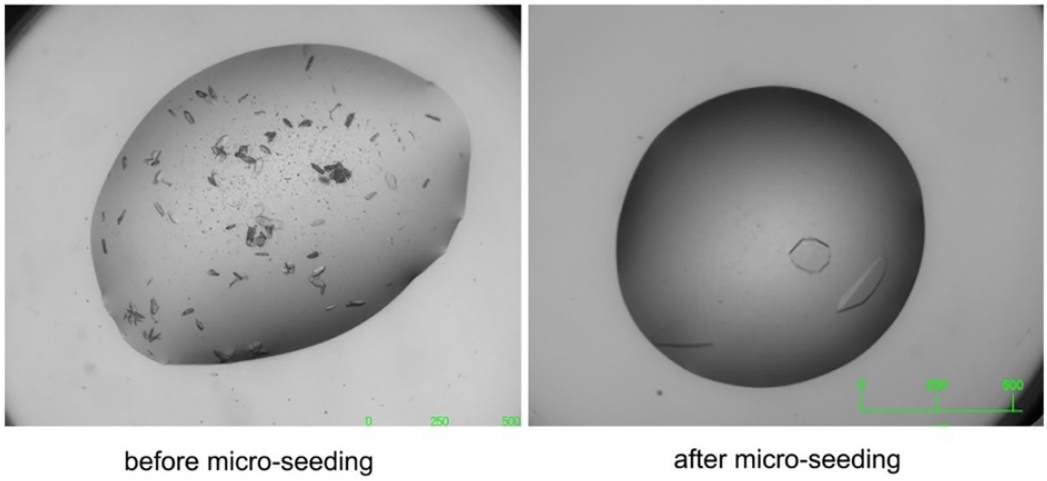


Figure 3-7 Crystals of the ZIKV NS3h-ADP-$MgF_3(H_2O)^-$ complex obtained by micro-seeding.

Left panel: crystals of NS3h-ADP-$MgF_3(H_2O)^-$ complex grew from initial crystallisation screen; right panel: complex crystals of NS3h-ADP-$MgF_3(H_2O)^-$ formed after micro-seeding.

Table 3-2 Statistics table of data collection and refinement.

	apo NS3h	**NS3h-pRNA$_9$**	**NS3h-ATPγS**	**NS3h-ADP**	**NS3h-ADP-BeF_3^-**	**NS3h-ADP-$MgF_3(H_2O)^-$**
PDB code	-	5MFX	-	-	6RWZ	6S0J
Data collection						
Space group	*P1*	*P2$_1$*	*P2$_1$*	*P2$_1$*	*P2$_1$*	*P2$_1$*
Cell dimensions						
a, *b*, *c* (Å)	44.28, 48.47, 57.90	52.36, 73.04, 58.70	53.33,69.22, 57.80	53.28, 69.27, 57.72	52.92, 71.36, 57.35	52.94, 69.60, 57.75

Resolution (Å)	43.26 - 1.42	45.69 - 1.60	69.22 - 1.40	69.27 – 1.58	57.23 - 1.70	69.58 - 1.50
R_{meas} [a]	0.048 (0.955)	0.049 (0.928)	0.061 (1.713)	0.074 (0.844)	0.065 (1.508)	0.055 (1.149)
I/σ (I) [b]	10.37 (1.04)	14.95 (1.46)	11.0 (0.9)	8.4 (1.0)	13.1 (1.0)	10.1 (1.2)
CC (1/2)	0.998 (0.544)	0.999 (0.484)	0.999 (0.372)	0.997 (0.562)	0.999 (0.444)	0.998 (0.419)
Completeness (%) [c]	86.2 (86.9)	98.8 (96.2)	99.9 (99.9)	99.4 (91.5)	99.4 (94.3)	99.7 (99.0)
Multiplicity	1.9 (1.9)	3.0 (2.9)	4.0 (4.0)	3.0 (2.6)	4.1 (4.0)	2.9 (2.9)
Refinement						
Resolution (Å)	43.26 - 1.42	45.69 - 1.60	57.65 - 1.40	57.50 - 1.55	57.23 - 1.70	57.65 - 1.50
No. reflections	66505	57734	82349	56746	44518	63340
R_{work} / R_{free}	18.73 / 22.09	16.19 / 19.14	0.151/ 0.192	0.144 / 0.194	0.167/ 0.206	0.145 / 0.189
No. atoms						
Protein	3214	3357	3420	3416	3478	3439
RNA	-	92	-	-	-	-
Ligand/ion	1	17	32	29	32	33
Water	187	387	430	391	353	463
Mean B factors (Å^2) [d]						
Protein	30.85	26.56	23.1	27.0	33.6	28.3
RNA	-	23.84	-	-		
Ligand/ion	19.99	41.92	41.6	32.6	34.6	31.4
Water	32.67	36.76	37.2	39.4	41.7	43.6
R.M.S. deviations						
Bond length (Å)	0.025	0.012	0.011	0.010	0.013	0.010
Bond angle (°)	2.2	1.5	1.7	1.4	1.5	1.6

a. R_{meas}: $\sum_{hkl} n \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}(n-1)\sum_i I_i(hkl)$, where n is a number of the symmetry equivalent contributors to the average; b. $I/\sigma I$: Signal to noise ratio for merged intensities. c. Completeness for unique reflections, highest resolution shells in brackets. d. B factor (temperature factors) averaged for atoms of protein, ligand/ion and water, respectively. *Values of highest resolution shell in brackets. Structure deposited in the public domain are indicated by PDB codes in the table, while the structures not available in the PDB database are indicated as '-' in the table.

Crystal structures of the NS3h complexes were determined at 1.50 - 1.70 Å resolution using crystals belonging to the $P2_1$ space group (Table 3-2), and the apo crystal structure was determined at 1.42 Å resolution in the *P1* space group. All the crystals contained a single NS3h molecule per asymmetric unit. The apo NS3h crystallised in the *P1* space group (Table 3-2), unlike several other structures of the apo NS3h which were previously determined in $P2_1$ space group (237–239). These observations suggest that the apo NS3h exhibits more flexibility than NS3h complexes bound with ATP analogues.

3.3.3 Structure determination of ZIKV NS3h

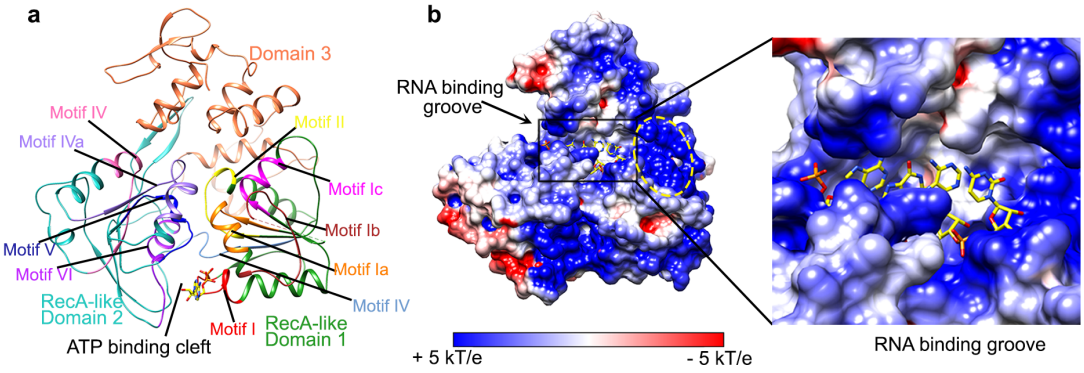


Figure 3-8 Overall structure of ZIKV NS3h.

a. Structure of the ZIKV NS3h complex with ADP. Domains 1-3 are coloured in forest green, light sea green and orange, respectively. The motifs I - VI are coloured differently and are designated on the figure. b. Molecular surface coloured according to the electrostatics potential for the NS3h-pRNA$_9$ complex structure. NS3h structure was assigned charges using the SWANSON force field (192), and the electrostatics surface potential of the NS3h was computed with APBS (193) in Chimera (174). Protein surface coloured according to the electrostatics potential (red, negative to blue, positive, ± 5 kT/e), and RNA shown as sticks in yellow. The yellow dashed circle indicates the potential RNA binding region of domain 1.

ZIKV NS3h (residues 183-617) comprises three domains, the two RecA-like core containing domains 1 (residues 183-324) and domain 2 (residues 325 - 480) as well as C-terminal domain 3 (residues 481-617) (Figure 3-8a). The NTP-binding cleft lies in between domains 1 and 2, and the RNA binding groove locates in the centre amongst the three domains (Figure 3-8). Domain 1 consists of a six-stranded β-sheet flanked by four α-helices, while domain 2 contains eight β-strands, including an anti-parallel β-strand (Figure 3-8a). Six of eight β-strands are located in the core of domain 2 and are flanked by five α-helices, while the β-hairpin, protrudes out towards domain 3 from domain 2 (Figure 3-8a). NS3h contains the conserved I - VI helicase motifs (Figure 3-8a), based on previously defined conserved helicases motifs (69). The RNA binding tunnel is constructed by motifs Ia, Ib and Ic of domain 1, and motifs IV, IVa and V of domain 2, as well as segments of domain 3 (Figure 3-8a). The NTP-binding pocket is formed by the helicase motifs I (P loop Walker A motif), II (DEAH motif) and VI (Figure 3-8a). The electrostatics surface potential of NS3h demonstrates that the RNA binding and NTP binding cleft are largely positively charged, suggesting that these regions are able to bind nucleic acid or nucleotide (Figure 3-8b). On top of that, a positively charged region lies on domain 1 (Figure 3-8b), indicating that this region may interact with nucleic acid.

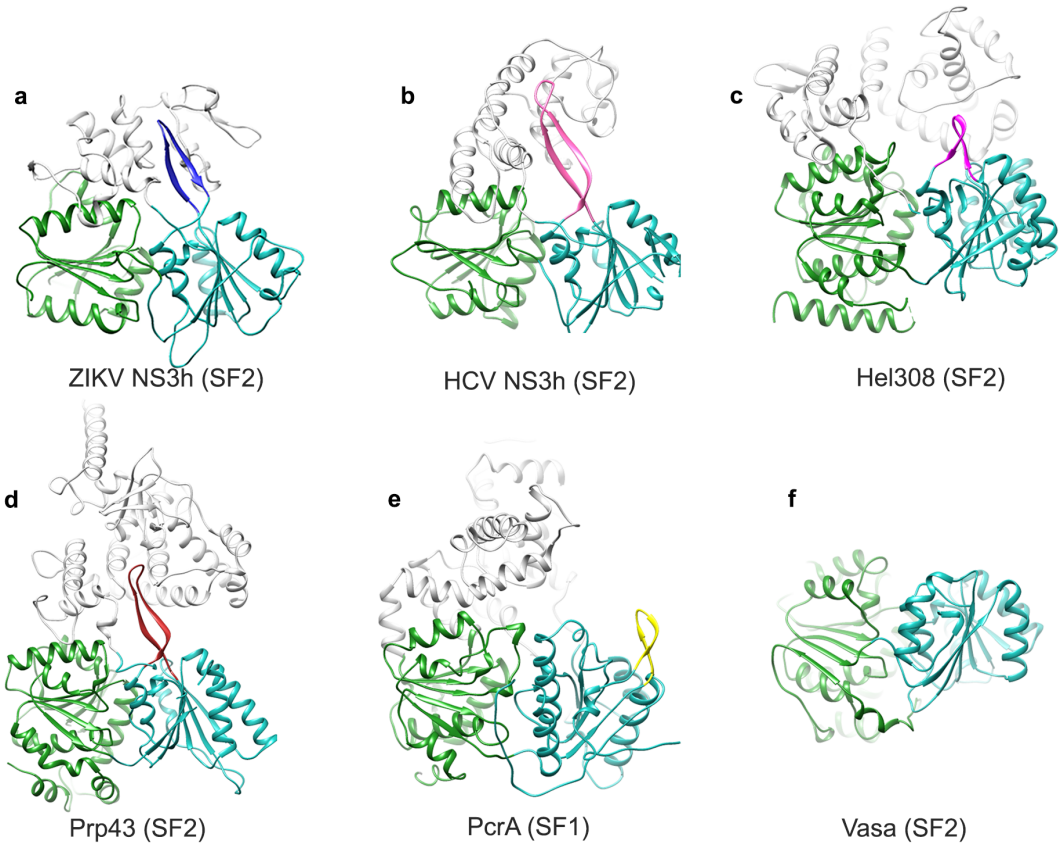


Figure 3-9 Comparisons of the β-hairpin of the ZIKV NS3h with equivalent structural elements found in the SF1 and SF2 helicases.

a. ZIKV NS3h (SF2, PDB code 5MFX, this work), β-hairpin is coloured in blue; b. HCV NS3h [SF2, PDB code 3O8C, (240)], β-hairpin is coloured in hot pink; c. Hel308 [SF2, PDB code 2P6R, (241)], β-hairpin is in magenta; d. Prp43 [SF2, PDB code 5LTJ, (107)], β-hairpin is in brown; e. PcrA [SF1, PDB code 3PJR, (78)]; f. Vasa [SF2, PDB code 2DB3, (97)]. Ribbon diagrams are shown with domain 1 in forest green, domain 2 in light sea green and C-terminal region in grey.

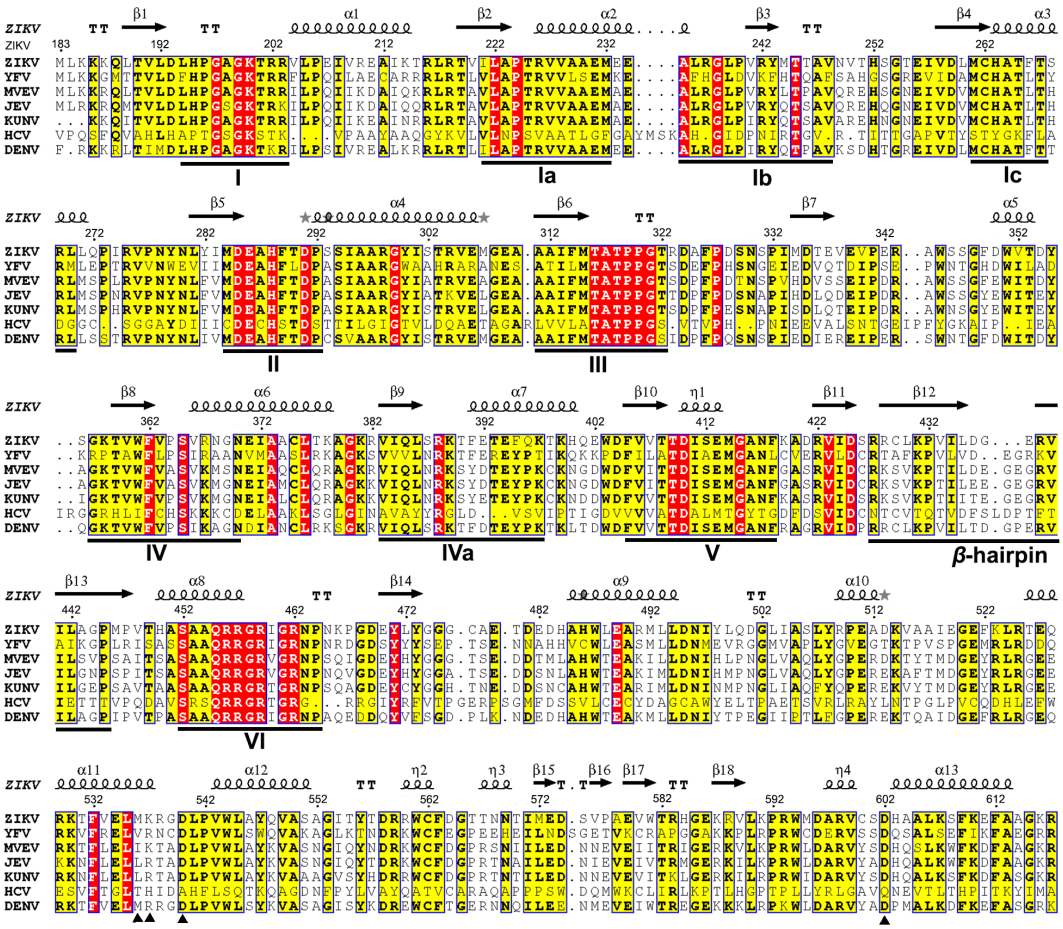


Figure 3-10 Structure based sequence alignment of *Flavivirus* NS3hs.
Alignment was performed for seven NS3 helicases (NS3hs) present in Zika virus [ZIKV, PDB code 5MFX, this work], Dengue virus 4 [DENV4, PDB code 2JLQ, (79)], Yellow Fever virus [YFV, PDB code 1YKS, (242)], Murray Valley encephalitis virus [MVEV, PDB code 2WV9, (243)], Kunjin virus [KUNV, PDB code 2QEQ, (244)], Japanese encephalitis virus [JEV, PDB code 2Z83, (245)] and Hepatitis C virus [HCV, PDB code 3O8C, (240)]. The conserved motifs I - VI are labelled accordingly on the figure.

The amino acid β-hairpin conserved amongst *Flaviviridae* family NS3hs (Figure 3-9 and 3-10) does not correspond to the previously identified SF2 helicase motifs (69), but is present in

3' → 5' SF2 family helicases (95). The structural conservation indicates that this amino acid hairpin is expected to play a crucial role in unwinding (Appendix 10), though it lacks sequence conservation among SF2 family helicases (Figure 3-9). Domains 1 and 2 of SF2 family helicases shared conserved RecA-like core as well as the conserved helicase motifs, as discussed in sections 1.3.4 and 1.3.5, while several SF2 helicases, such as DEAD-box Vasa helicase, lack a β-hairpin that may function as a wedge in nucleic acid unwinding (Figure 3-9f, Appendix 10) (95, 246), indicating these helicases may use different unwinding mechanisms from other SF2 helicases, including NS3h.

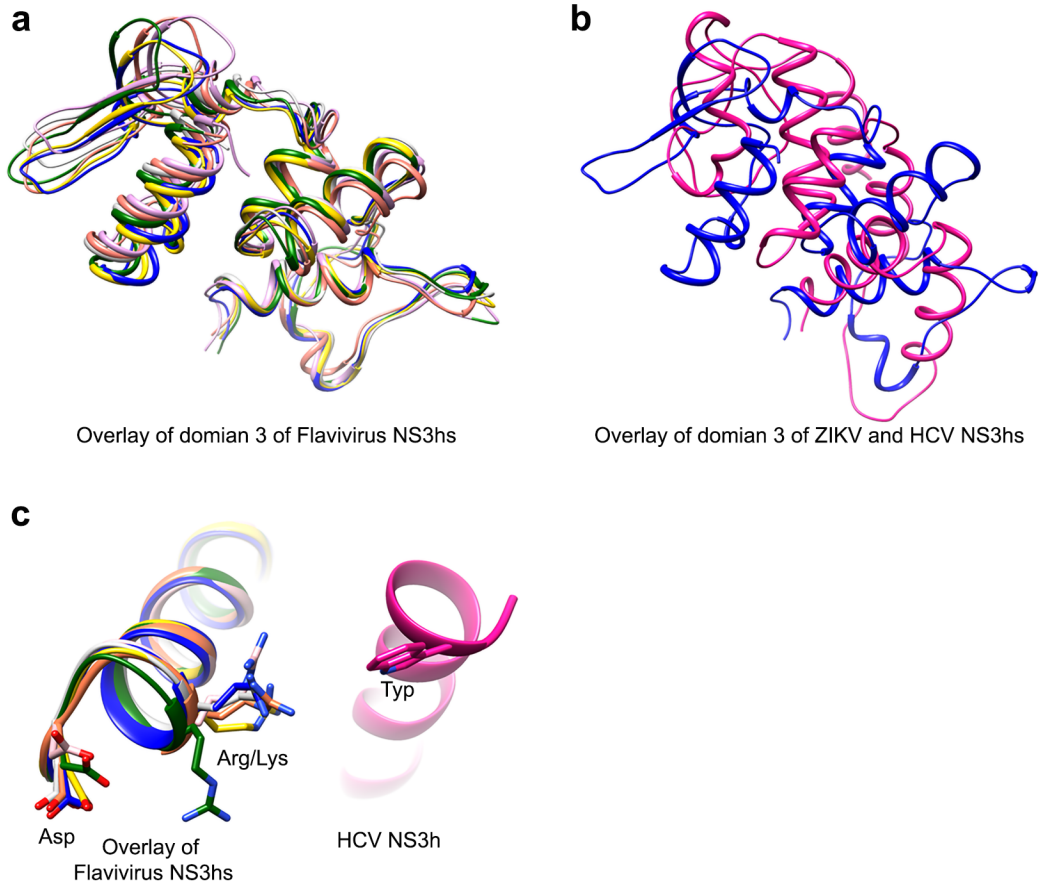


Figure 3-11 Structural comparison of domain 3 of *Flavivirus* and HCV NS3hs.
a. Overlay of domain 3 of *Flavivirus* NS3hs; b. Overlay of domain 3 of ZIKV and HCV NS3hs. Domain 3 from *Flavivirus* NS3hs were superimposed using domain 1 as the reference (domain 1 is not shown in the figure); c. Comparison of structurally relevant α-helices of *Flavivirus* NS3hs (left) and HCV NS3h (right). The conserved aspartate and arginine/lysine and W501 (in HCV NS3h) that is thought to act as a ratchet for DNA translocation, are shown as sticks. ZIKV NS3h (PDB code 5MFX, this work) is in blue, DENV4 NS3h [PDB code 2JLQ, (79)] is in pink, YFV NS3h [PDB code 1YKS, (242)] is in coral, MVEV NS3h [PDB code 2WV9, (243)] is in grey, KUNV NS3h [PDB code

2QEQ, (244)] is in dark green, JEV NS3h [PDB code 2Z83, (245)] is in yellow, and HCV NS3h [PDB code 3O8C, (240)] is in dark pink.

Interestingly, domain 3 of the *Flavivirus* NS3 helicase (residues 481 to 617 in ZIKV NS3h) is predominantly formed by four α-helices (Figure 3-11a). The sequences of domain 3 are conserved among *Flavivirus* NS3hs (Figure 3-10). However, domain 3 of ZIKV NS3h is significantly different from HCV NS3h (Figure 3-11b), suggesting this domain 3 in ZIKV NS3h may be functionally different from HCV NS3h. Moreover, there is no so-called 'ratchet' tryptophan (or an aromatic residue) observed in the domain 3 of all *Flavivirus* NS3hs (Figure 3-11c and Appendix 11), while it is present in the HCV NS3h and some other SF2 helicases for preventing RNA or DNA backsliding during nucleic acid translocation and unwinding (109, 241), as discussed in section 1.3.6. In addition, there is no 'ratchet' subdomain essentially present in DEAH-box family helicases observed in *Flavivirus* NS3hs (as described in section 1.3.5). Domain 3 of NS3h is smaller and structurally simpler, resulting in a shorter RNA binding tunnel (Figure 3-9a - 3-9d, Appendix 11), compared to other SF2 family helicases (108), suggesting that the NS3h is likely utilising a different RNA translocation strategy from these SF2 helicases and may require a 'helper' protein for RNA unwinding.

3.3.3.1 Crystal structure of NS3h-pRNA$_9$ complex

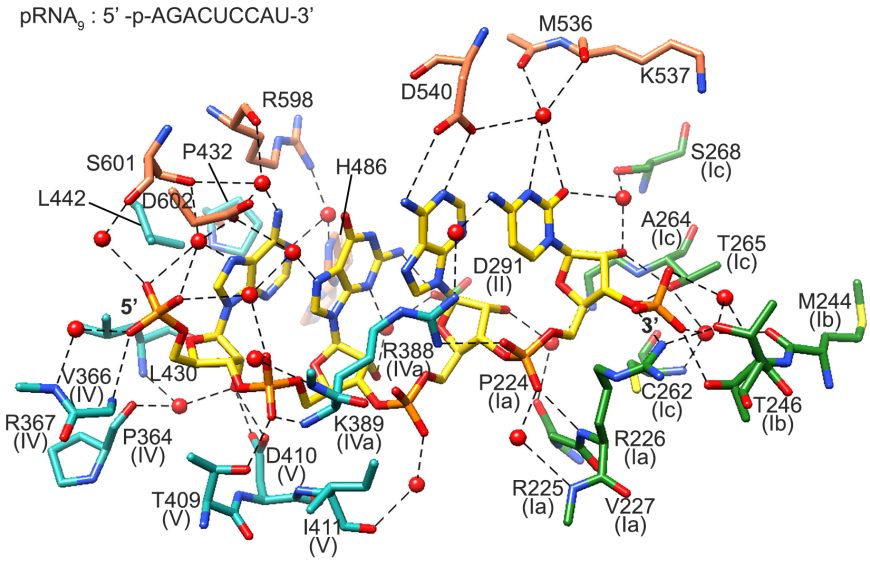


Figure 3-12 Hydrogen bonding interactions of ZIKV NS3h with pRNA$_9$.
Hydrogen bonds (less than 3.5 Å) are displayed with dashed lines, protein residues of Domain 1 - 3 are in forest green, light sea green and orange, respectively. Atoms are coloured as follow: oxygen, red; nitrogen, blue; carbon, gold; and phosphorus, coral.

The crystal structure of the ZIKV NS3h in complex with $pRNA_9$ was determined at 1.60 Å resolution (Table 3-2). In common with previously published structures of NS3h-ssRNA complexes (79, 247), most interactions between RNA and protein in the ZIKV NS3h-$pRNA_9$ are water-mediated (Figure 3-12). P432 and L442 in the β-hairpin stack with the 5' adenine of RNA via hydrophobic interactions (Figure 3-12). The RNA is stabilised by multiple direct or water-mediated hydrogen-bonding interactions with the protein (Figure 3-12). Significantly, most (10 of 13) direct H-bonds are observed for the phosphodiester backbone and ribose 2' hydroxyl group of RNA (Figure 3-12), suggesting the non-sequence specific binding of RNA to NS3h, as required for unwinding processivity. D291 of motif II stabilises the guanine of G2 (Figure 3-12). Three residues, T409, D410 and I441, from helicase motif V interacts with RNA via either direct or water-mediated H-bonds (Figure 3-12), demonstrating that this motif V is inherently involved in RNA binding, like other NS3hs (112). Moreover, the carboxylic sidechain of D410 is H-bonded to the 2' hydroxyl of A1 nucleotide (Figure 3-12). In contrast, this conserved aspartate (equivalent to D412 in HCV NS3h) does not interact with DNA (109), suggesting that ZIKV NS3h binds preferentially to RNA rather than DNA.

The $pRNA_9$ used in this study was designed to have a 5' phosphate group, representing a complete section of the unwound strand of the natural substrate (Figure 3-12 and 3-13). The protein structures in the ZIKV NS3h-$pRNA_9$ (this work) and ZIKV NS3h-ssRNA complexes [PDB code 5GJB, (247)] are well superimposable (Figure 3-13a), with the root-mean-square deviation (RMSD) calculated for main-chain atoms (residues 183 - 617) of only 0.4 Å. However, the 5' phosphate group induces notable remodelling of hydrogen bonding interactions between the protein and $pRNA_9$ (Figure 3-13c - 3-13f), while not significantly altering the overall position of bound RNA (Figure 3-13a and 3-13b). The remodelled hydrogen bonding network indicates the enhanced RNA binding is consistent with the approximately 2-fold increase in affinity of RNA binding observed for a 5' phosphate containing ssRNA in the case of DENV NS3 helicase (248). Importantly, D602 in domain 3 contributes to water-mediated hydrogen-bonding interactions with the 5' phosphate group and the A1 base (Figure 3-13c). Conversely, D602 does not participate in RNA binding in the absence of the 5' phosphate (Figure 3-13e). Likewise, main chain oxygen of R598 is hydrogen bonding to the A1 base through a water molecule only in the presence of 5' phosphate (Figure 3-13c and 3-13e).

Additionally, differences in protein-RNA interactions are observed for the third, fourth and fifth nucleotides (Figure 3-13c - 3-13f). K537 does not interact with the $pRNA_9$, whereas, in the NS3h-ssRNA structure, its sidechain amino group forms an H-bond with the ribose hydroxyl of the fifth nucleotide (Figure 3-13d and 3-13f). The main chains of K537 and M536 interact with the fourth nucleotide of RNA via water-mediated interactions in the presence of

5’ phosphate (Figure 3-13d), on the contrary, these two residues interact with the fifth nucleotide in the absence of the 5’ phosphate (Figure 3-13f). Moreover, the carboxylic sidechain of D540 is hydrogen bonded to the exocyclic amino group of the A3 base of the $pRNA_9$ (Figure 3-13d). However, this aspartate makes a water-mediated interaction with the base of the fourth nucleotide in the absence of the 5’ phosphate group, and its main chain oxygen is hydrogen bonded to the A3 base through a mediating water molecule (Figure 3-13f). The sidechain of K431, a residue in the β-hairpin, is partially disordered in the presence of the 5’ phosphate (Figure 3-13d), whereas it interacts with the A1 base in the absence of the 5’ phosphate (Figure 3-13f).

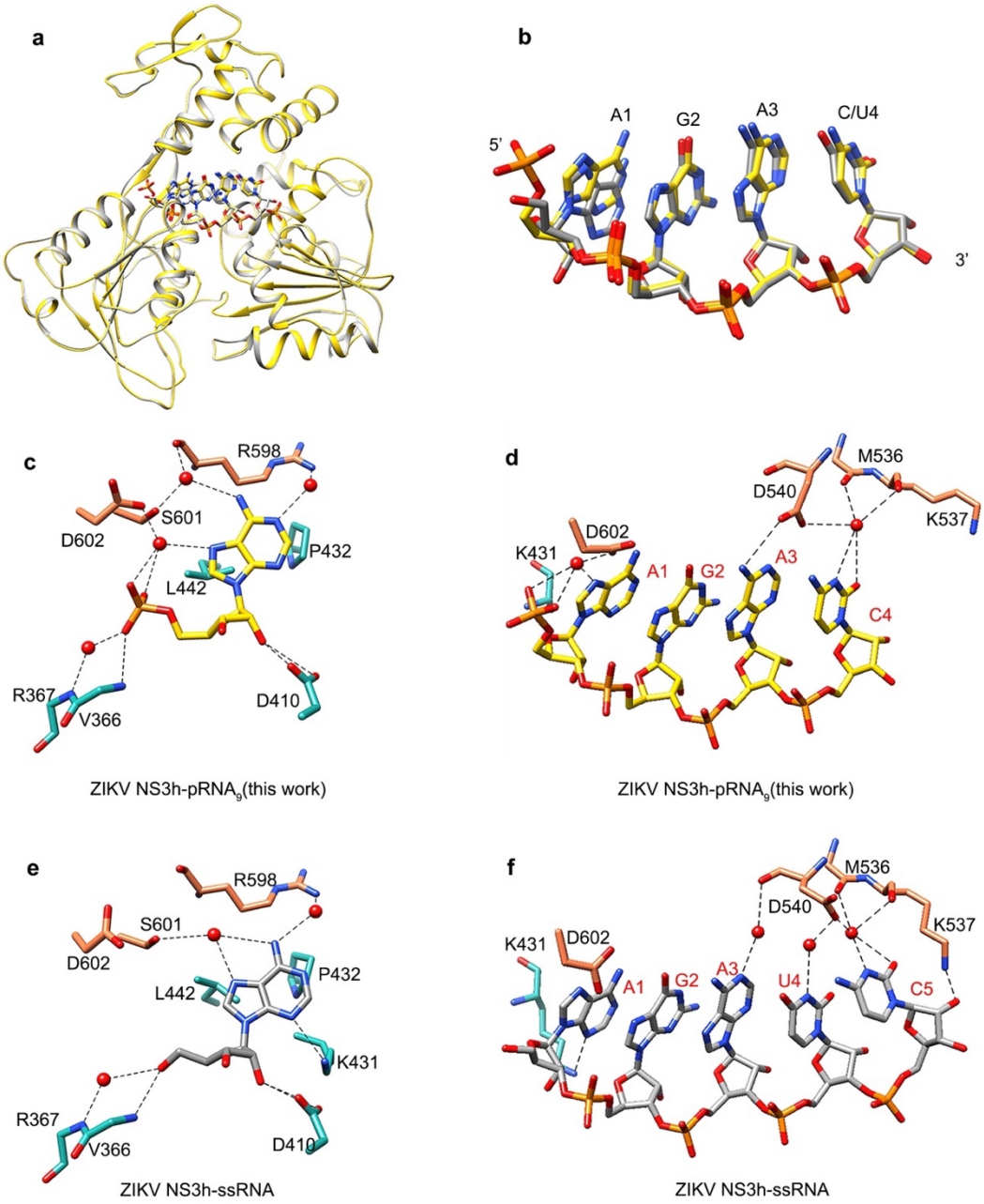


Figure 3-13 Comparison of protein-RNA interactions in the presence and absence of the 5' phosphate of ssRNA.

a. Superposition of ZIKV NS3h-pRNA$_9$ and ZIKV NS3h-ssRNA structures; b. Overlay of the pRNA$_9$ and ssRNA with four nucleotides shown in the figure panel; c. Hydrogen bonding network for the A1 nucleotide in the ZIKV NS3h-pRNA$_9$ structure; d. Key hydrogen bonding interactions for the RNA in the structure of the NS3h-pRNA$_9$ complex; e. Hydrogen bonding network between protein residues and A1 nucleotide in the NS3h-ssRNA structure; f. Key hydrogen bonding interactions formed with RNA in the structure of the NS3h-pRNA$_9$ complex. H-bonds (less than 3.5 Å) are displayed as dashed lines. Structures used are ZIKV NS3h-pRNA$_9$ and ZIKV NS3h-ssRNA [PDB code 5GJB, (247)]; Protein residues are coloured in forest green, light sea green and orange according to domains 1-3, respectively. Atoms are coloured as follow: oxygen, red; nitrogen, blue; carbon, gold; and phosphorus, coral.

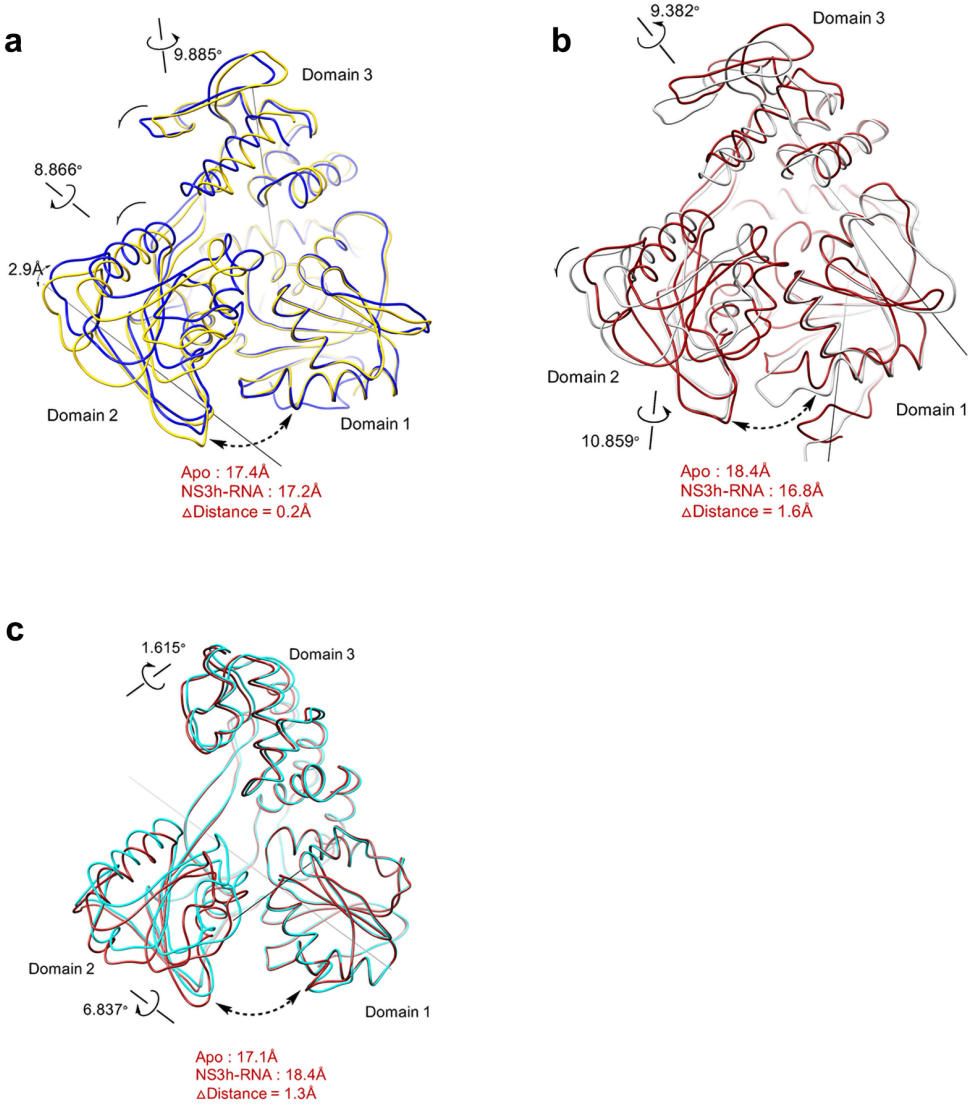


Figure 3-14 Domain rotations of NS3h helicases induced by RNA binding.

a. Overlay of the structures of the ZIKV NS3h-pRNA$_9$ complex (gold, PDB code 5MFX, this work) and its apo form [blue, PDB code 5JMT, (239)]. The crystal structures of ZIKV NS3h were determined in *P2$_1$* space group; b. Overlay of the structures of the DENV NS3h-ssRNA complex [dark red, PDB code 2JLU, (79)) with its apo form [grey, PDB code 2JLQ, (79)]. The crystal structures of the apo form of DENV NS3h were determined in *P2$_1$* space group, and the DENV NS3h-RNA crystals were in the *C2* space group; c. Overlay of the structures of HCV NS3h-ssRNA complex [brown, PDB code 3O8C, (240)] with its apo form [cyan, PDB code 3O8B, (240)]. The crystal structures of HCV NS3h were determined in *P2$_1$* space group.

To further investigate the conformational changes induced by RNA binding, rigid domain rotations were analysed based on comparison of the ZIKV NS3h-pRNA$_9$ and apo structures. Upon RNA binding, the domain 2 rotates ~ 8.9° towards domain 1, and domain 3 rotates 9.9° away from domains 1 and 2. Consequently, these domain rotations result in a more open RNA binding cleft and closing of the ATP binding cleft between domains 1 and 2 (Figure 3-14a). Similarly, as a result of RNA binding, the domain 2 of DENV NS3h turns 10.9° closer to domain 1, and the domain 3 rotates away from the domain 1 by 9.4° compared with its apo form structures (Figure 3-14b). In contrast, the comparison based on HCV NS3h structures shows that domain 2 rotates 6.8° closer to domain 1 upon RNA binding (Figure 3-14c), domain 3 of HCV NS3h moves slightly away from domain 1 by rotating 1.6° (Figure 3-14c). It is possible that the observed differences in conformational adjustments induced by RNA binding amongst ZIKV, DENV and HCV NS3 helicases are probably due to either structural divergence in domain 3 (Figure 3-11) or differences in crystal packaging.

3.3.3.2 Crystal structure of NS3h complex with ADP

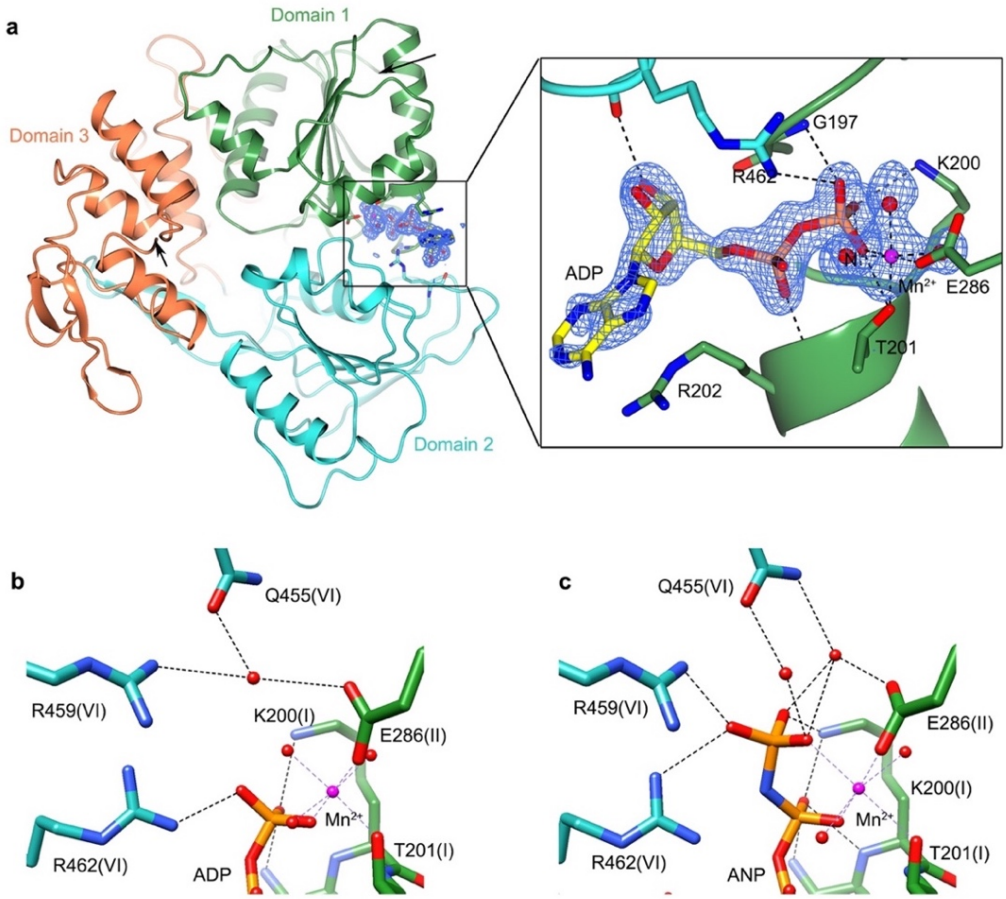


Figure 3-15 Crystal structure of ZIKV NS3h-ADP complex.

a. Structure of the ZIKV NS3h in complex with ADP. Left panel: ZIKV NS3h with ADP bound at the ATP binding pocket between the RecA-like domains 1 and 2. Right panel: Omit map (*mFo-DFc*) for ADP, Mn^{2+} and coordinated water molecules, contoured at 4σ by CCP4MG (173). b. Structural details of the ADP *β*-phosphate coordination (this work); c. Same as b but for the NS3h-AMPPNP complex [PDB code 5Y4Z, (249)].

Extensive attempts to co-crystallise the ZIKV $NS3h_{183\text{-}623}$ protein with both ATPγS and a partially duplexed hairpin RNA (Appendix 12), unexpectedly, resulted in crystals containing clearly defined electron density of an ADP moiety in the ATP binding pocket with no evidence of bound RNA (Figure 3-15a, left panel). The resulting structure of the NS3h-ADP complex was determined at 1.58 Å resolution (Table 3-2). The presence of ADP, rather than the expected ATPγS, was unambiguously defined by the omit electron density map (Figure 3-15a, right panel). The ADP moiety is stabilised by the phosphate binding loop (helicase motif I) of domain 1 and the R462 from domain 2 (Figure 3-15a, right panel), with no density for the γ-

phosphate. The catalytic manganese ion is in octahedral coordination with ADP, T201, E286 and three water molecules (Figure 3-15a, right panel).

Details of the ATP binding site for the structures of ZIKV NS3h in complex with ADP (this work) and AMPPNP [PDB code 5Y4Z, (249)] were compared, to understand the basal (nucleic acid free) ATP hydrolysis cycle of NS3h. Notably, R459 (motif VI) donates an H-bond to the γ-phosphate in the NS3h-AMPPNP complex, whereas it is making an H-bond to a water molecule in the NS3h-ADP complex (Figure 3-15b and 3-15c). The sidechain of Q455 makes two water-mediated H-bonds with the γ-phosphate of AMPPNP, while it interacts with E286 and R459 when the γ-phosphate is cleaved and released (Figure 3-15b and 3-15c). The observed significant reduction in interactions with the nucleotide, resulting from the γ-phosphate release (Figure 3-15b and 3-15c), demonstrates structural events that result in the reduction of the binding affinity of nucleotide upon γ-phosphate dissociation.

Compared with the apo form, the structure of the ZIKV NS3h in complex with ATP or ADP reveals slight conformational changes (Figure 3-16). Upon binding ATP, domain 2 rotates about 3.1° towards domain 1 as a rigid body around axis A, and domain 3 moves 2.8° away from domain 1 rotating around axis B (Figure 3-16, left panel). After γ-phosphate of ATP release, domain 2 moves apart from domain 1 by undergoing 0.7° rotation around axis A, and domain 3 rotates 1.8° around axis B, bringing it closer to domain 1 (Figure 3-16, middle panel). Upon releasing ADP, domain 2 rotates 2.4° around axis A moving away from domain 1, and domain 3 rotates 1.5° around axis B bringing it closer to domain 1 (Figure 3-16, right panel). Motions between domains 1 and 2 lead to a closure of the ATP binding cleft, while no significant differences between ATP bound and ADP bound NS3h complexes were observed. As ADP-AlF_4 binding results in significant domain adjustments in the HCV NS3h (Appendix 13), the vastly smaller domain adjustments observed for the ZIKV NS3h may be the result of the restraint imposed by the crystal packing.

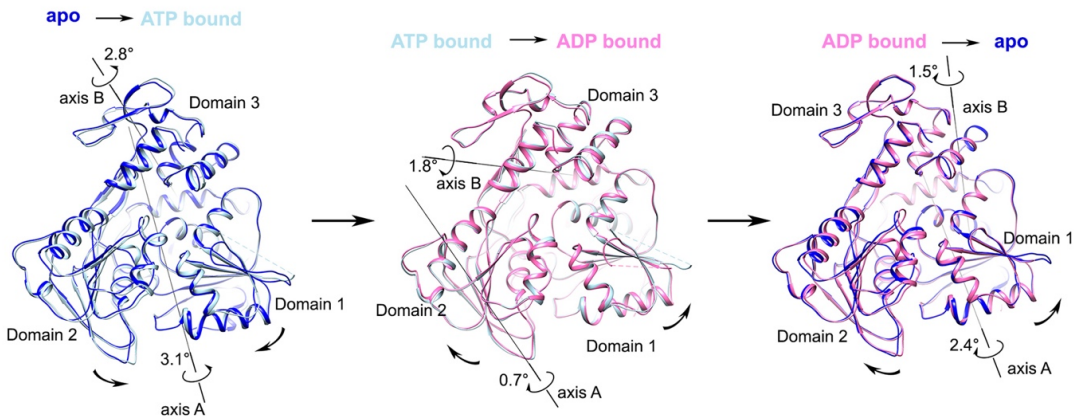


Figure 3-16 ATP hydrolysis induced conformational changes of ZIKV NS3h.
Left: Superimposition of the ZIKV NS3h-ATP complex structure [light blue, PDB code 5GJC, (247)] with the apo form [blue, PDB code 5JMT, (239)]. Middle: Superposition of the ZIKV NS3h-ADP (pink, this work) and ZIKV NS3h-ATP complexes [light blue, PDB code 5GJC, (247)]. Right: Superposition of the ZIKV NS3h-ADP complex structure (pink, this work) with the apo form [blue, PDB code 5JMT(239)].

3.3.3.3 NS3h-ADP-BeF_3^- and NS3h-ADP-$MgF_3(H_2O)^-$ complexes structures

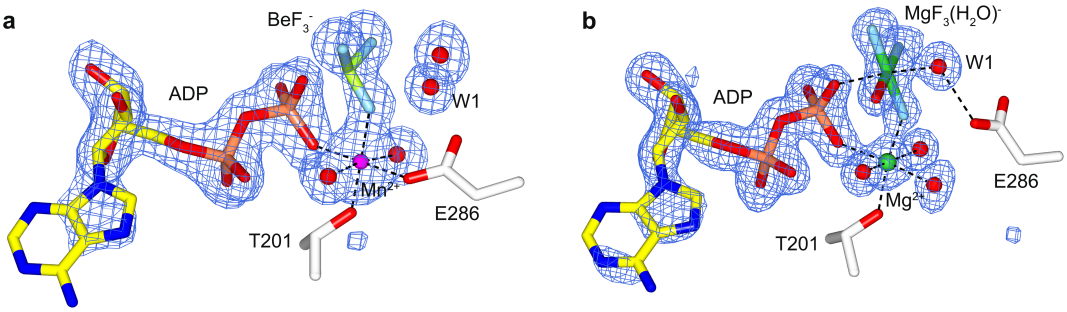


Figure 3-17 Omit maps (mF_o-DF_c) for ADP-BeF_3^- and ADP-$MgF_3(H_2O)^-$ structures.
a. Omit maps (mF_o-DF_c) for ADP-BeF_3^-, b. Omit maps (mF_o-DF_c) for ADP-$MgF_3(H_2O)^-$. Omit maps (*mFo-DFc*) are in blue meshes and were contoured at 4σ. W1 indicates the nucleophilic water. ADP carbon atoms are in yellow, and protein residues are in white. Beryllium is in lime green, and magnesium is in green and fluorine in light blue. Atoms are coloured as follows: oxygen, red; phosphorus, coral; nitrogen, blue.

The structure of the ZIKV NS3h ground state analogue (GSA) complex with ADP trifluoroberyllate was determined at 1.70 Å resolution (Table 3-2). The crystal of NS3h-ADP-BeF_3^- GSA complex was obtained by soaking Be^{2+} and F^- into the NS3h-ADP crystals to capture the pre-hydrolytic conformation of ADP with tetrahedral BeF_3^- (Figure 3-17a) (see details in Method and materials section 3.2.2). In this structure, oxygen of the hydrolytic water molecule (W1) is 3.7 Å from Be^{2+}, donating an anti-catalytic H-bond to F^1 (3.1 Å) and a second H-bond (2.8 Å) to the sidechain C=O of Q455 to give a near attack conformation (NAC) (Figure 3-17a). The observed H-bonds are somewhat longer than those seen in ssDNA-bound NS3h-ADP-BeF_3^- complex for HCV, at 3.5 Å (PDB codes 3KQU and 3KQN) to Be^{2+}, and 2.7 Å (PDB code 3KQU) or 2.8 Å (PDB code 3KQN) to the glutamine (109).

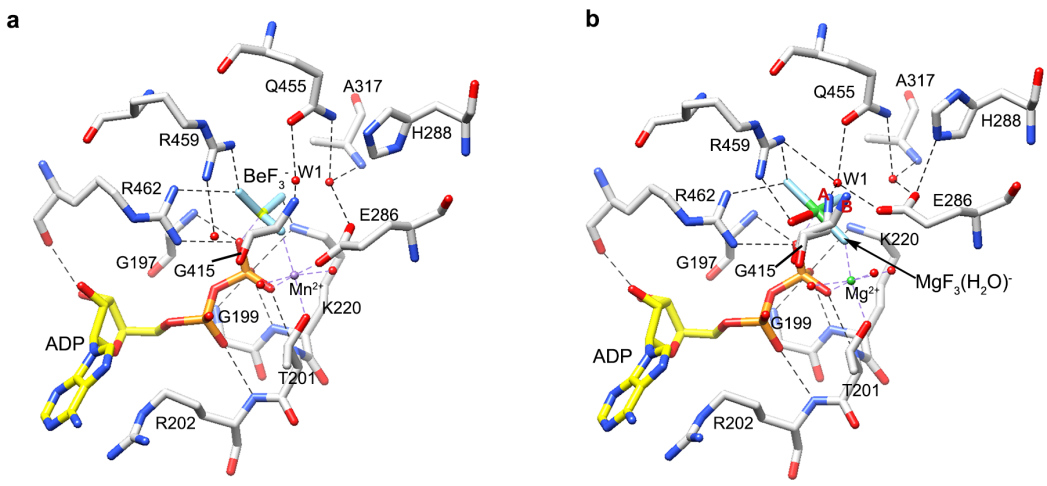


Figure 3-18 ATPase active site of NS3h-ADP-BeF_3^- and NS3h-ADP-$MgF_3(H_2O)^-$ complexes.

a. ADP-BeF_3^-. b. ADP-$MgF_3(H_2O)^-$. Carbon atoms of protein are in white and of carbons of ADP are in yellow. Beryllium is in lime green, and magnesium is in green and fluorine is in light blue. Metal coordination is displayed as purple dashes, and H-bonds (less than 3.4 Å) are displayed as black dashes.

The ZIKV NS3h-ADP-$MgF_3(H_2O)^-$ transition state analogue (TSA) was obtained by crystallising NS3h in the presence of 20 mM ADP, 150 mM Mg^{2+} and 20 mM F^-, and is the first high resolution structure for the Zika virus NS3h in complex with magnesium trifluoride, determined at 1.50 Å resolution (Table 3-2). Unexpectedly, the trifluoromagnesate moiety has octahedral geometry, in contrast to the generally observed trigonal bipyramidal (tbp) geometry. This structure has clearly defined density for the six coordinate Mg^{2+} in the transition state analogue (TSA) complex (Figure 3-17b). A square planar species located between β-phosphate of ADP and the hydrolytic oxygen of W1 is unambiguously defined in the omit electron density maps (Figure 3-17b). In addition, this TSA complex positions E286 of NS3h as a general base to activate the nucleophilic water by switching its interaction from the catalytic metal in the GSA complex structure (Figure 3-17a and 3-18a) to the nucleophilic water in the TSA complex structure (Figure 3-17b and 3-18b), similar to the conserved glutamate of DExx motif in other ATPases (74, 250–252).

3.3.4 Examination of the ADP-$MgF_3(H_2O)^-$ TSA structure

The arrangement of this square planar moiety of $MgF_3(H_2O)^-$ is similar to the AlF_4^- seen in the structure of a closely related homologue, HCV NS3h with ADP-AlF_4^- (PDB codes 3KQL and 5E4F) (109, 253). The observed octahedral moiety with a central magnesium atom is

unexpected, since all 24 PDB entries of trifluoromagnesate bound complexes have five coordinate MgF_3^- (Appendix 14) and are identified as isosteric transition state analogues for phosphoryl group transfer (211, 228, 254). To eliminate the potential contamination by AlF_4^-, deferoxamine, a strong aluminium chelator, was added to the crystallisation solution, as detailed in section 3.2.2, following a procedure described previously (255). The electron density map of this $MgF_3(H_2O)^-$ species was not changed at all (data not shown), despite the crystal of the complex being soaked overnight with 10 mM deferoxamine.

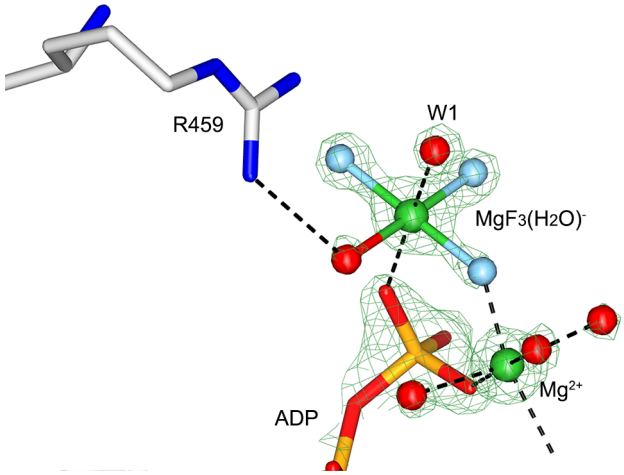


Figure 3-19 Examination of ADP-$MgF_3(H_2O)^-$ structure.

Omit map (mF_o-DF_c) of $MgF_3(H_2O)^-$ is in lawn green mesh and was contoured at 8 σ.

Examination of the omit map of this unusual complex shows weaker electron density at the position closest to R459 (Figure 3-19), suggesting that the atom at this position is oxygen rather than fluorine. After fitting a water molecule (H_2O) into this position and fluorines in the other three equatorial positions, the Mg-F bond lengths refine to 1.88 Å and the Mg-OH_2 bond length to 2.02 Å. The bond lengths fall into the range of bond lengths expected for a six-coordinated magnesium (223), while the axial W1-Mg-O^{3B} angle of ADP-$MgF_3(H_2O)^-$ is 175.6° and the distances from Mg^{2+} to W1 and the β-phosphate O^{3B} are 1.96 Å and 2.10 Å, respectively (Figure 3-20). Therefore, the $MgF_3(H_2O)^-$ is, by far, a new transition state mimic for the PO_3^- group, based on the dominant first-sphere coordination number of six for Mg^{2+} (256). The net negative charge of $MgF_3(H_2O)^-$ makes it a more fitting TSA than either MgF_4^{2-} or $MgF_3(OH)^{2-}$, and supports the 'priority of charge over geometry' concept for TSA complexes observed for phosphoryl transfer enzymes (255, 257).

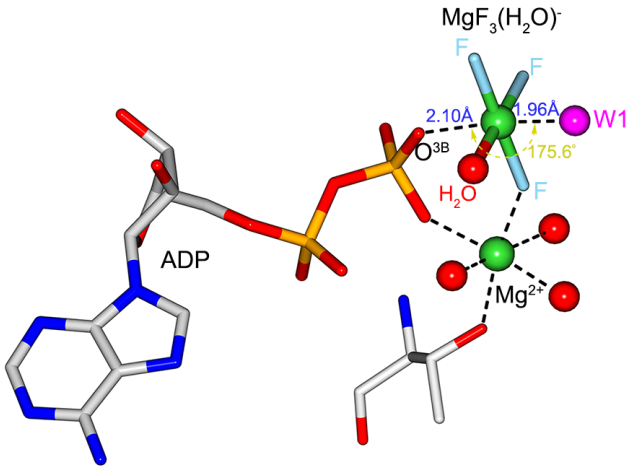


Figure 3-20 Structure of ADP-$MgF_3(H_2O)^-$.

Key bond lengths and angle were indicated on the figure. Protein and ADP carbon atoms are in grey. Magnesium is in green and fluorine in light blue. The hydrolytic water W1 is in magenta.

To provide independent validation of the $MgF_3(H_2O)^-$ moiety, further examinations of the NS3h-ADP-$MgF_3(H_2O)^-$ complex was performed using Fluorine-19 nuclear magnetic resonance spectroscopy (^{19}F NMR), detailed in section 3.2.4. ^{19}F NMR complements structural data with careful analysis of metal coordination. As expected, the solution spectra showed three ^{19}F resonances present in 1:1:1 ratio (Figure 3-21), establishing that the magnesium fluoride moiety has only three fluorines. The three fluorines were assigned using solvent induced isotope shift (SIIS) values (Table 3-3), due to the SIIS value accurately reflecting the number and orientation of H-bond donors around each fluorine (258), see details described in section 3.2.4. In the NS3h-ADP-$MgF_3(H_2O)^-$ complex, the F^1 (-175.2 ppm, SIIS 0.15 ppm, Table 3-3) is the most shielded ion with weak H-bonding interaction along with its coordination by magnesium. F^2 (-153.4 ppm; SIIS 1.44 ppm, Table 3-3, and Figure 3-21) is H-bonded to the sidechain amino group of K200 and to a water molecule that is H-bonded to E286. F^3 is the most deshielded fluorine (-146.1 ppm; SIIS 1.38 ppm, Table 3-3) with two H-bonds from the R459 and R462 guanidinium groups. These two arginine residues have been observed to neutralize the negative charge developed on the γ-phosphate during ATP hydrolysis (Figure 3-18), acting like a '*trans*-arginine finger' (259). Replacing ADP by GDP did not alter the ^{19}F spectra, demonstrating the absence of strict nucleotide specificity of NS3h (Figure 3-21).

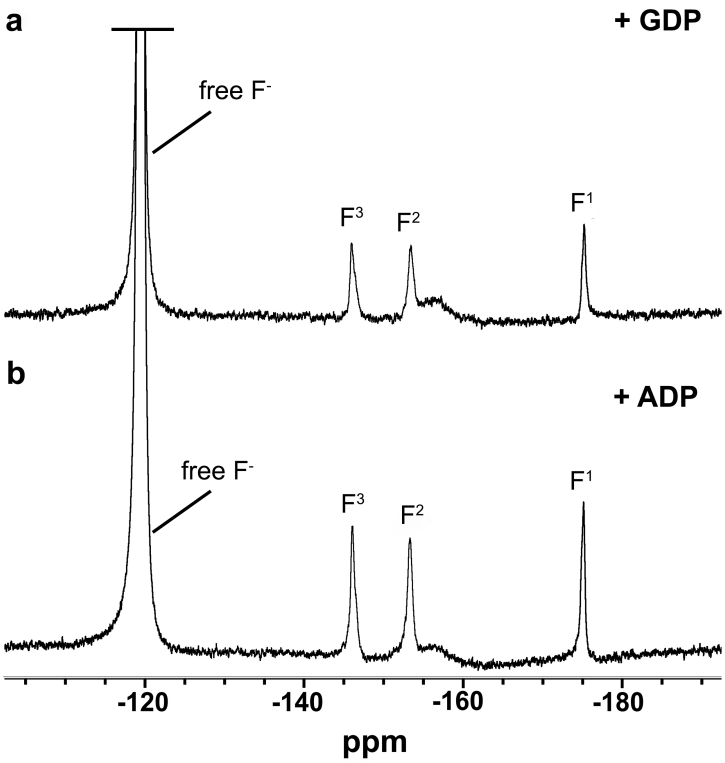


Figure 3-21 ^{19}F NMR spectra of ZIKV NS3h-ADP-$MgF_3(H_2O)^-$ and NS3h-GDP-$MgF_3(H_2O)^-$ TSA complexes.

a. NS3h-MgGDP-$MgF_3(H_2O)^-$ TSA in 90% H_2O buffer (F^3 = -146.1, F^2 = -153.4, F^1 = -175.2 ppm). b. NS3h-ADP-$MgF_3(H_2O)^-$ TSA in 90% H_2O buffer (F^3 = -146.1, F^2 = -153.4, F^1 = -175.2 ppm).

Since octahedral coordinate AlF_4^- has been found to form a more stable complex than MgF_3^- for many phosphoryl transfer enzymes, having a much lower dissociation constant (211, 212, 255, 257, 260, 261), in order to investigate active site affinity of this unexplored $MgF_3(H_2O)^-$ complex relative to AlF_4^-, a competition experiment was performed. Addition of 1 mM $AlCl_3$ to a sample of preformed NS3h-ADP-$MgF_3(H_2O)^-$ complex in the presence of 10 mM $MgCl_2$, resulted in a 5% decrease of the three ^{19}F resonances and the rise of a small rotationally averaged peak at -152.1 ppm, showing 5% of the active site switched from $MgF_3(H_2O)^-$ to tetrafluoroaluminate (Figure 3-22a and 3-22b). The rotationally averaged peak is AlF_4^- associated. Addition of $AlCl_3$ to 5 mM final concentration, gives a conversion of $MgF_3(H_2O)^-$ to AlF_4^- of only 50% (Figure 3-22a and 3-22c). This limited conversion to AlF_4^- further demonstrates that $MgF_3(H_2O)^-$ has NS3h TSA stability at least equivalent to that of AlF_4^-, since an AlF_4^- complex would only achieve full occupancy at much higher concentration of Al^{3+} (at least 20 mM).

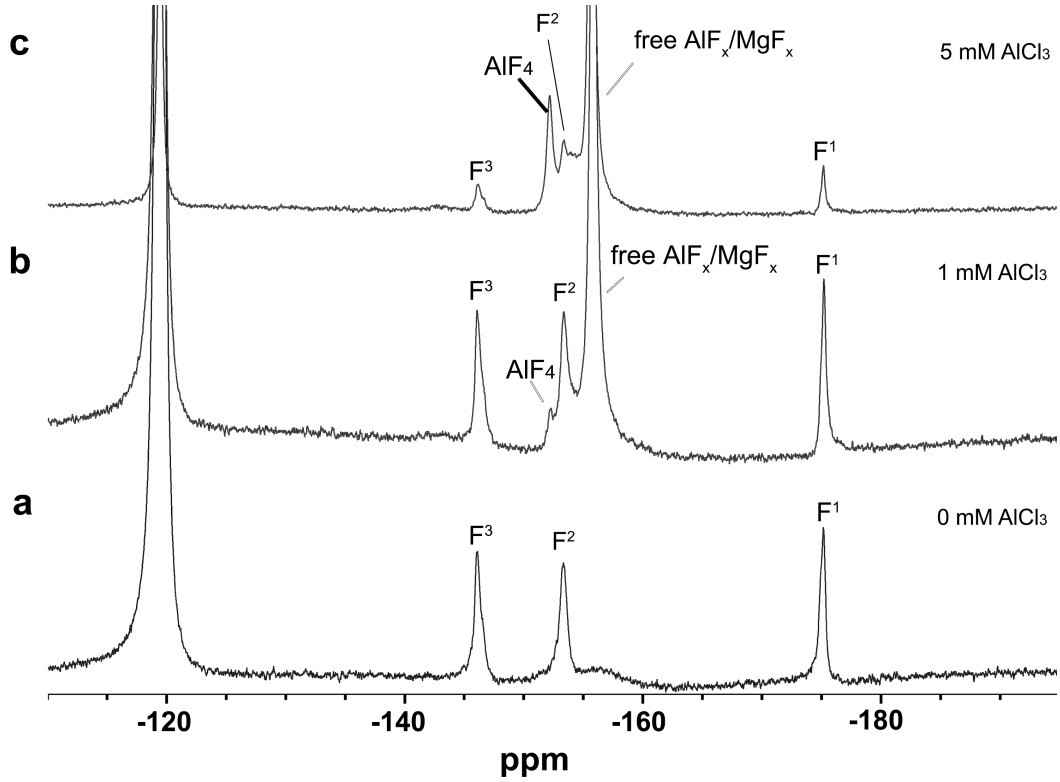


Figure 3-22 ^{19}F NMR spectra of the $AlCl_3$ titration to convert NS3h-ADP-$MgF_3(H_2O)^-$ into NS3h-ADP-AlF_4^- complexes.

a. NS3h-ADP-$MgF_3(H_2O)^-$ in 90% H_2O buffer (F^3 = -146.1 ppm, F^2 = -153.4 ppm, F^3 = -175.2 ppm). b and c. Addition of 1 mM $AlCl_3$ in panel b or 5 mM $AlCl_3$ in panel c resulted in a rotationally averaging peak (152.1 ppm) of bound AlF_4^-.

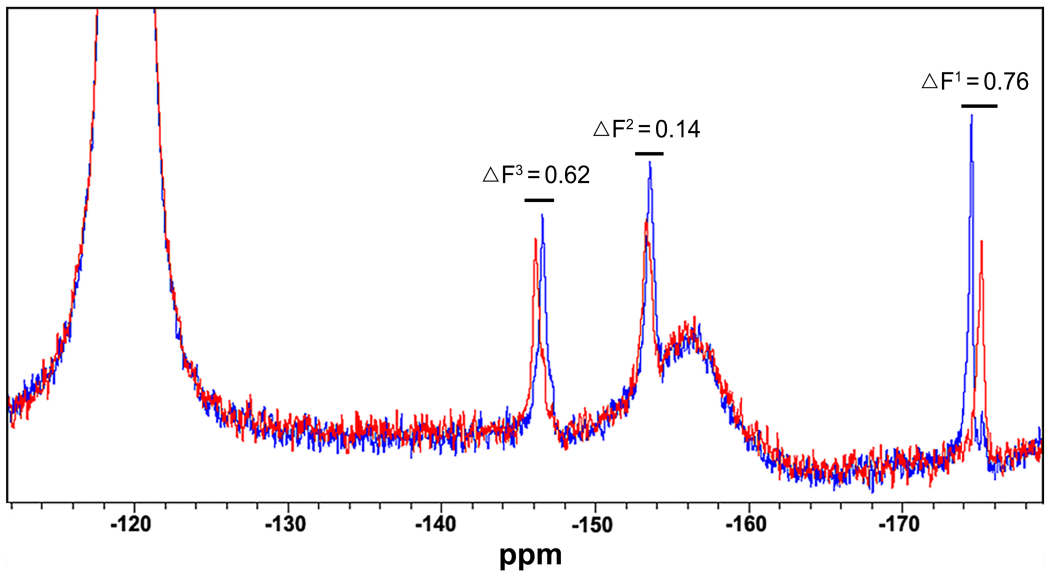


Figure 3-23 Complexes of ssRNA-free and ssRNA-bound NS3h-ADP-$MgF_3(H_2O)^-$.

^{19}F NMR spectra of ssRNA-free (red) and ssRNA-bound (blue) NS3h-ADP-$MgF_3(H_2O)^-$ TSA complexes. Chemical shift differences for each resonance are shown on the figure.

Table 3-3 Solvent induced isotope shifts (SIIS, ppm) for ^{19}F NMR signal of RNA-free and RNA-bound ZIKV NS3h-ADP-$MgF_3(H_2O)^-$ complexes.

NS3h-ADP-$MgF_3(H_2O)^-$		F^3	F^2	F^1
RNA-bound	$\delta^{19}F_{(100\% H2O)}$	-146.6	-153.6	-174.5
	SIIS	1.4	1.5	0.2
RNA-free	$\delta^{19}F_{(100\% H2O)}$	-146.1	-153.4	-175.2
	SIIS	1.4	1.4	0.2

SIIS = $\delta\ ^{19}F_{(90\% H2O\ buffer)} - \delta\ ^{19}F_{(100\% D2O\ buffer)}$.

In addition, NS3h is able to form a stable complex of ssRNA-bound NS3h-ADP-$MgF_3(H_2O)^-$ TSA complex in solution (Figure 3-23), though attempts to crystallise this ternary complex were unsuccessful. When ssRNA was added to the sample of the NS3h-ADP-$MgF_3(H_2O)^-$ TSA complex, the three fluorine resonances in the ^{19}F NMR changed only by 0.62 ppm for F^3, 0.14 ppm for F^2, and 0.76 ppm for F^1 (Figure 3-23). These shifts indicate only a small difference in the H-bonding network within the NS3h ATPase active site and subtle conformational changes upon ssRNA binding. Notably, all three fluorine resonances of the $MgF_3(H_2O)^-$ complex increase in intensity upon addition of ssRNA, most prominently for F^1 (Figure 3-23). These increases suggest that binding of ssRNA retards exchange between the $MgF_3(H_2O)^-$ moiety and free MgF_x resulting from tighter binding of the TSA complex. Likewise, a two-fold decrease of the K_d for ADP-AlF_4^- in the absence of ssDNA has been observed for the HCV NS3h by fluorescence polarization (253). Binding ssRNA also slightly increases the SIIS values for all three fluorines, consistent with overall H-bond shortening in the TSA complex (Table 3-3). The ^{19}F NMR spectra provide direct evidence of the conformational closure of the finely tuned H-bond network around the transferring phosphoryl group in the TS, indicating ssRNA stimulates NTPase activity of ZIKV NS3h, like HCV NS3h (253).

3.3.5 Structural basis for ATP hydrolysis mechanism of the NS3 helicase

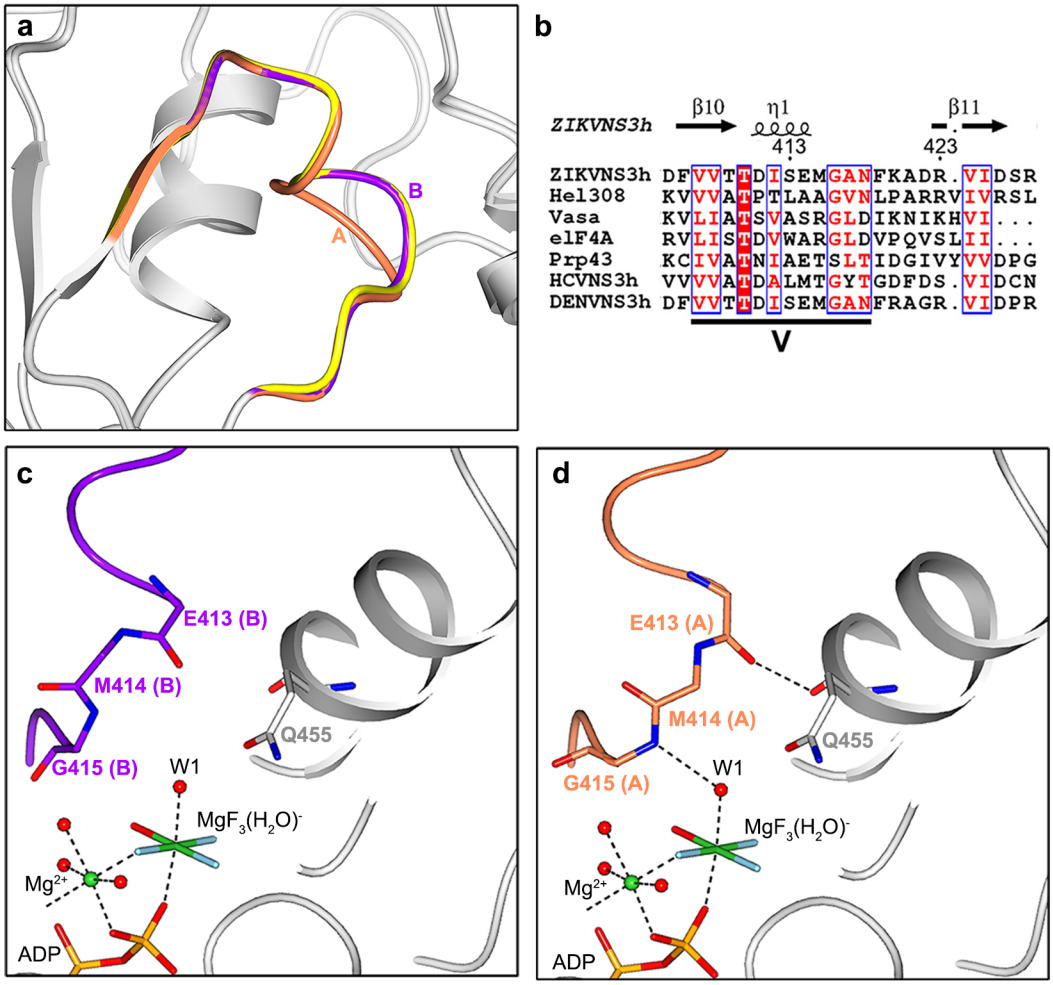


Figure 3-24 Two conformations of the conserved motif V in ZIKV NS3h.

a. Superposition of NS3h-ADP-$MgF_3(H_2O)^-$ (light grey, conformation A in coral, conformation B in purple) and NS3h-MnADP-BeF_3^- (light grey, motif V in yellow) structures. b. Sequence alignment of several SF2 helicases. ZIKV NS3h (PDB code 5MFX, this work), Hel308 [PDB code 2P6R, (241)], Vasa [PDB code 2DB3, (97)], elF4A [PDB code (262)], Prp43 [PDB code 5LTA, (107)], HCV NS3h [PDB code 3O8C, (240)] and DENV NS3h [PDB code 2JLU, (79)]. c and d. Side-by-side comparison of the two conformations in the NS3h-ADP-$MgF_3(H_2O)^-$ structure. Close-up views at the loop conformation B, purple (c) and A, coral (d) are shown with Mg^{2+} in green, fluorine in light blue, water molecules in red and H-bonds (≤ 3.1 Å) shown by dashed lines.

The single molecule of the NS3h, in the asymmetric unit of the crystal, presents two conformations, A and B, for the catalytically important motif V in the NS3h-ADP-$MgF_3(H_2O)^-$ structure (Figure 3-24a). Conformation B adopts the usual 'relaxed' position seen in the structures of NS3h-ADP-BeF_3^- (Figure 3-24a), where the amide of G415 is 4.0 Å from the

hydrolytic water W1 and donates a H-bond to the backbone carbonyl of E413 (Figure 3-24b). This motif V is conserved across SF2 helicases (Figure 3-24b). In conformation A, following the reorganization of the motif V, the amide of G415 moves 1.0 Å towards W1 and donates an H-bond (3.0 Å) to coordinate it (Figure 3-24c and 3-24d). This demonstrates that conformation A is involved directly in ATP hydrolysis in the RNA free complex of NS3h. As motif V is involved both in nucleic acid binding and in transition state formation for ATP hydrolysis (Figure 3-12, 3-24 and Table 3-4), it is likely that the structurally conserved motif V plays a vital role in coupling NTP hydrolysis with RNA translocation.

Table 3-4 X-ray structures of *Flaviviridae* family NS3 helicases in complex with transition state analogues (TSA).

PDB code	Protein	Resolution (Å)	Space Group	TSA	With RNA or DNA bound	Motif V loop conformation
6S0J (this work)	ZIKV NS3h	1.50	*$P2_1$*	ADP-$MgF_3(H_2O)^-$	no	A&B
2JLY	DENV NS3h	2.40	*C2*	ADP-PO_4^-	RNA bound	B
2JLX	DENV NS3h	2.20	*C2*	ADP-VO_4^-	RNA bound	A
5E4F	HCV NS3h	2.10	*P2*	ADP-AlF_4^-	no	B
3KQL	HCV NS3h	2.50	*P2*	ADP-AlF_4^-	DNA bound	A

Comparison of the motif V conformations of all *Flaviviridae* family viruses NS3h-TSA complexes structures which are available in the PDB by July 2018. The compared structures includes DENV NS3h-ssRNA-ADP-PO_4^- [PDB code 2JLY, (79)], DENV NS3h-ssRNA-ADP-VO_4^- [PDB code 2JLX, (79)], HCV NS3h-ADP-AlF_4^- [PDB code 5E4F, (253)] and HCV NS3h-ssDNA-ADP-AlF_4^- [PDB code 3KQL, (109)]. Conformations A and B of motif V refer to the conformations A and B in Figure 3-24a.

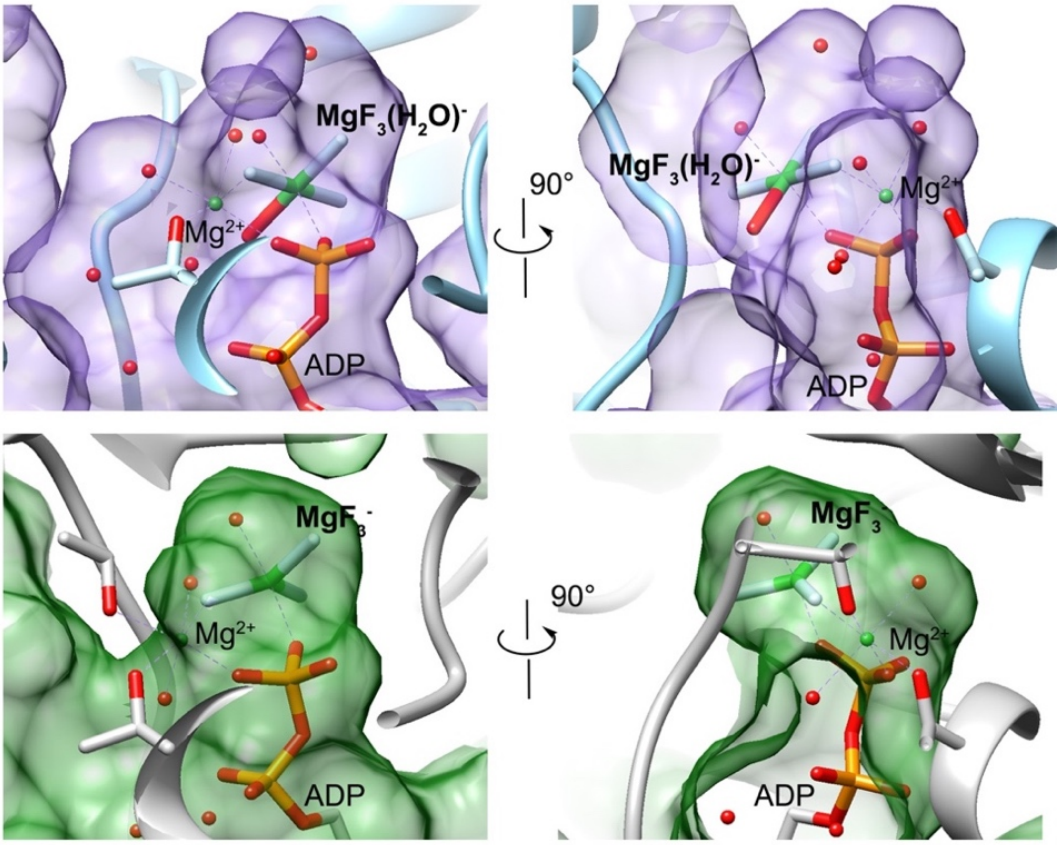


Figure 3-25 Comparisons of the ATP binding pockets for the transition state complex structures of NS3h-ADP-$MgF_3(H_2O)^-$ and RhoA-GDP-MgF_3^--RhoGAP.

Upper panel: The inner surface of the triphosphate binding site for the NS3h-ADP-$MgF_3(H_2O)^-$ TSA complex shown in semi-transparent purple. Protein is highlighted in light blue with $MgF_3(H_2O)^-$ shown in sticks. The size of the pocket for binding the TSA and β-phosphate is 256.2 $Å^3$. Lower panel: The inner surface of the triphosphate binding site for RhoA-GDP-MgF_3^--RhoGAP TSA complex [PDB code 1OW3, (228)] shown in semi-transparent green. The protein is in grey. MgF_3^- moiety is shown as sticks and waters as red spheres. The size of the pocket for binding the TSA and β-phosphate is 69.4 $Å^3$.

3.3.6 Functional characterisations of ZIKV NS3h

3.3.6.1 Protein stability determination for ZIKV NS3h

The thermal unfolding transition midpoint (melting temperature, T_m, °C) of $NS3h_{183-623}$ was derived using a thermal shift assay. The melting temperature of native $NS3h_{183-623}$ is 39.02 ± 0.04 °C in the buffer condition containing 300 mM NaCl (Table 3-5). The addition of ATP analogues resulted in significant increases of the melting temperature of NS3h. In particular, when adding 1 mM ATPγS to the protein sample, the melting temperature was increased by

3.21 °C. In addition, the melting temperature was increased to 46.68 ± 0.07 °C (Table 3-5), by adding 5 mM ATPγS to the protein sample. The rise of melting temperature likely indicates the addition of ATP analogue binds to the native NS3h.

Table 3-5 Thermal shift assay of NS3h with various ATP analogues.

	Melting temperature (°C)						
	ADP	**AMPPNP**	**AMPPCP**	**ATPγS**	**AMP**	**GDP**	**GMPPNP**
5 mM	40.4 ± 0.0	39.5 ± 0.1	38.5 ± 0.2	46.7 ± 0.1	38.5 ± 0.1	39.4 ± 0.1	37.9 ± 0.1
2 mM	39.9 ± 0.1	39.4 ± 0.1	38.7 ± 0.2	44.7 ± 0.1	38.7 ± 0.1	39.3 ± 0.1	38.4 ± 0.3
1 mM	39.4 ± 0.0	39.1 ± 0.2	38.7 ± 0.1	42.2 ± 0.0	38.8 ± 0.1	39.3 ± 0.1	38.8 ± 0.1
0 mM				39.0 ± 0.0			

Mean data of sample replicates and plus/minus standard deviations are shown in the table. Each sample has three replicates.

Due to the hydrophobic fluorescence dye contributing to the precipitation of NS3h, Nano differential scanning fluorimetry (NanoDSF) was applied to further characterise the protein solubility by monitoring the tryptophan intrinsic fluorescence emission shifts. NanoDSF records the fluorescence intensities at 350 nm and 330 nm (see described in section 3.2.8). Generally, $F_{350/330}$ yields data with well-defined transitions upon protein unfolding and can reduce the possibility that protein stability is altered by the fluorescent dye used in Thermofluor style assays.

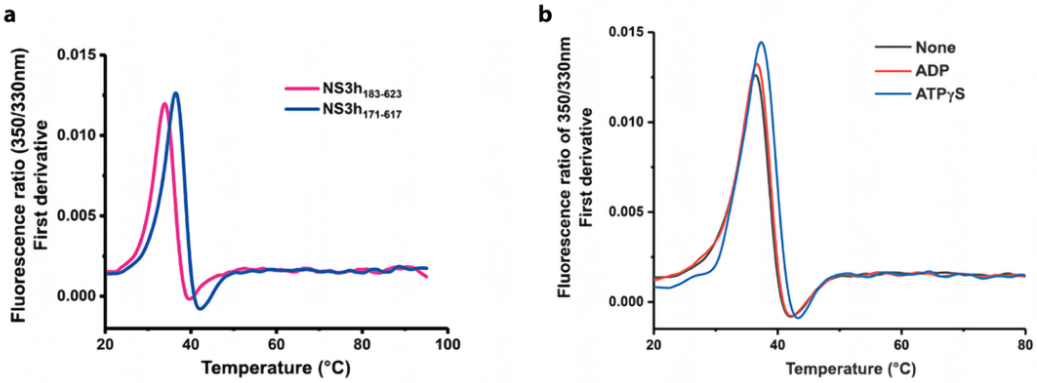


Figure 3-26 Thermal unfolding curves of ZIKV NS3h.

a. Thermal unfolding curves of $NS3h_{183-623}$ and $NS3h_{171-617}$. b. Thermal unfolding curves of ZIKV NS3h with ADP and ATPγS. Reactions were recorded at 20-95°C, 5 μM ZIKV $NS3h_{183-623}$ and $NS3h_{171-617}$ was used in each reaction.

Table 3-6 Melting temperatures of $NS3h_{183-623}$ and $NS3h_{171-617}$.

	T_{onset} [ºC]	T_m1 [ºC]	T_m2 [ºC]
$NS3h_{183-623}$	27.5 ± 0.3	33.6 ± 0.1	-
$NS3h_{171-617}$	27.9 ± 1.2	36.0 ± 0.0	42.9 ± 0.0
$NS3h_{171-617}$ + ADP	28.6 ± 0.7	36.2 ± 0.0	43.1 ± 0.2
$NS3h_{171-617}$ + ATPγS	31.3 ± 0.6	37.0 ± 0.1	44.1 ± 0.1

Mean data of sample replicates and plus/minus standard deviations are shown in the table. Each sample has duplex replicates.

Since the construct of $NS3h_{183-623}$ is not stable and tends to precipitate at room temperature, in particular at low salt concentrations (less than 100 mM), and the 183 – 617 residues (including partial N-terminal 6 × His) of $NS3h_{183-623}$ are visible in the crystal structures, the $NS3h_{171-617}$ construct was made to further investigate the protein stability. Since $NS3h_{171-617}$ and $NS3h_{183-623}$ have 9 tryptophan and 11 tyrosine residues in a total of 457 and 451 amino acids, respectively, the ratio of intrinsic fluorescence of tryptophan and tyrosine ($F_{350/330}$) are sufficient to reflect the conformational changes of NS3h proteins. The reaction onset temperature (T_{onset}) is the temperature where it is assumed that protein unfolding begins. The values of T_m1 and T_m2 indicate the first and second midpoint of the unfolding transition, respectively. The T_m1 of $NS3h_{183-623}$ is 33.6 ± 0.1 ˚C, and T_m1 of $NS3h_{171-617}$ is 36.0 ± 0.0 ˚C (Figure 3-26a and Table 3-6), suggesting that the $NS3h_{171-617}$ is slightly more stable than $NS3h_{183-623}$, though $NS3h_{171-617}$ is still unstable and tends to precipitate at room temperature (20 ˚C). The construct of $NS3h_{171-617}$ was used for following characterisations for ZIKV NS3h.

The ATPγS or ADP and RNA were used to further investigate NS3h stability (Figure 3-26b and Table 3-6). In the presence of ATPγS, the T_{onset}, T_m1, and T_m2 of NS3h are increased from 27.9 ± 1.2˚C, 36 ± 0.0˚C and 42.9 ± 0.0˚C to 31.3 ± 0.6˚C, 37 ± 0.1˚C and 44.1 ± 0.1˚C, respectively (Figure 3-26b and Table 3-6). However, in the presence of ADP, the T_{onset}, T_m1 and T_m2 of NS3h are only slightly increased to 28.6 ± 0.7˚C, 36.2 ± 0.0˚C and 43.1 ± 0.2˚C from 27.9 ± 1.2˚C, 36 ± 0.0˚C, respectively (Figure 3-26b and Table 3-6). This result suggests that ATPγS stabilises NS3h better than ADP, and the increases of T_{onset}, T_m1 and T_m2 as a result of binding of ATPγS to the native NS3h, consistent with the structural observations (in section 3.3.3.2).

3.3.6.2 Characterisations of ATPase activity of ZIKV NS3h

Since enzymes with poor thermal stability and a tendency to aggregate often show lability in their activities, the ATPase activity was compared in NADH-coupled photometric assays for

either freshly purified protein (fresh $NS3h_{171-617}$) and protein that was snap frozen in liquid nitrogen and stored at - 80˚C before immediate use (thawed $NS3h_{171-617}$). The K_m of fresh $NS3h_{171-617}$ was 82.52 ± 5.04 μM, and the k_{cat} was 4.42 ± 0.09 s^{-1}, while the K_m of thawed $NS3h_{171-617}$ was 89.68 ± 9.51 μM, and the k_{cat} was 2.49 ± 0.08 s^{-1} (Figure 3-27a). The enzyme efficiency (k_{cat}/K_m) of NS3h was about 2-fold decreased due to freezing and storage (Figure 3-27a).

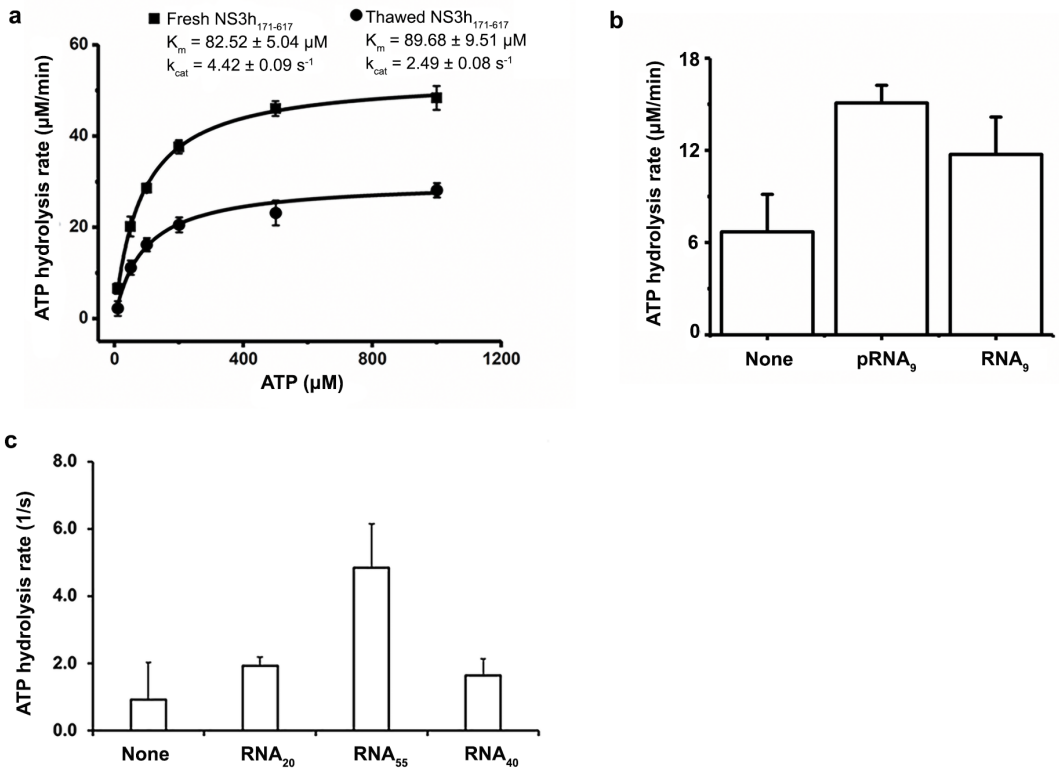


Figure 3-27 Characterisations of ATPase activity of ZIKV $NS3h_{171-617}$.

a. ATPase activity of fresh and thawed $NS3h_{171-617}$. Assays were carried out by titrating 0 - 1 mM ATP to 200 nM $NS3h_{171-617}$. The icons show the mean data of the sample replicates and error bars on the icons show the plus/minus standard mean errors (SE). b. 9-mer single-stranded (ss) RNA stimulation of NS3h ATPase activity assay. $pRNA_9$ and RNA_9 were utilised in the assays. c. RNA stimulation of ATPase activity assay. 50 nM $NS3h_{171-617}$ was mixed with 1 μM 20-mer (RNA_{20}), 40-mer (RNA_{40}) and 55-mer (RNA_{55}) ssRNA (RNA sequences are given in Appendix 6), respectively. The columns show the average data of sample replicates and the error bars indicate the plus standard mean errors (SE).

To further define if ssRNA was able to stimulate the ATPase activity of ZIKV NS3h, either $pRNA_9$ or $ssRNA_9$ were added to $NS3h_{171-617}$ for the NADH-coupled photometric assay. The addition of 5 μM (100:1 ratio of RNA to protein) $pRNA_9$ or RNA_9 resulted in a 2.25-fold and

1.75-fold increase of ATP hydrolysis rate, respectively, compared to the basal ATP hydrolysis rate of $NS3h_{171-617}$ (Figure 3-27b). There was no significant statistical difference ($p \leq 0.05$) among the three samples (as defined by the T-test), due to the variability in assays, despite using freshly thawed protein for each experiment. However, even short ssRNAs can clearly stimulate the ATP hydrolysis activity of ZIKV NS3h, as observed for other *Flavivirus* NS3hs (263–267).

To investigate whether RNA with different lengths and sequences stimulates ZIKV NS3h ATPase activity, three ssRNAs of different lengths were utilised for ATPase activity stimulation assays. The addition of a 55-mer single-stranded (ss) RNA_{55} significantly stimulates NS3h ATPase activity with a 5.2-fold increased hydrolysis rate (Figure 3-27c), while the addition of RNA_{40} and RNA_{20} resulting in 2.1-fold and 1.76-fold increase of the hydrolysis rate, respectively (Figure 3-27c). These results indicate that ssRNA can stimulate ATPase activity of NS3h, and longer lengths of RNA oligos appear to achieve slightly faster ATP hydrolysis rates.

3.3.6.3 Characterisations of RNA binding affinity for ZIKV NS3h

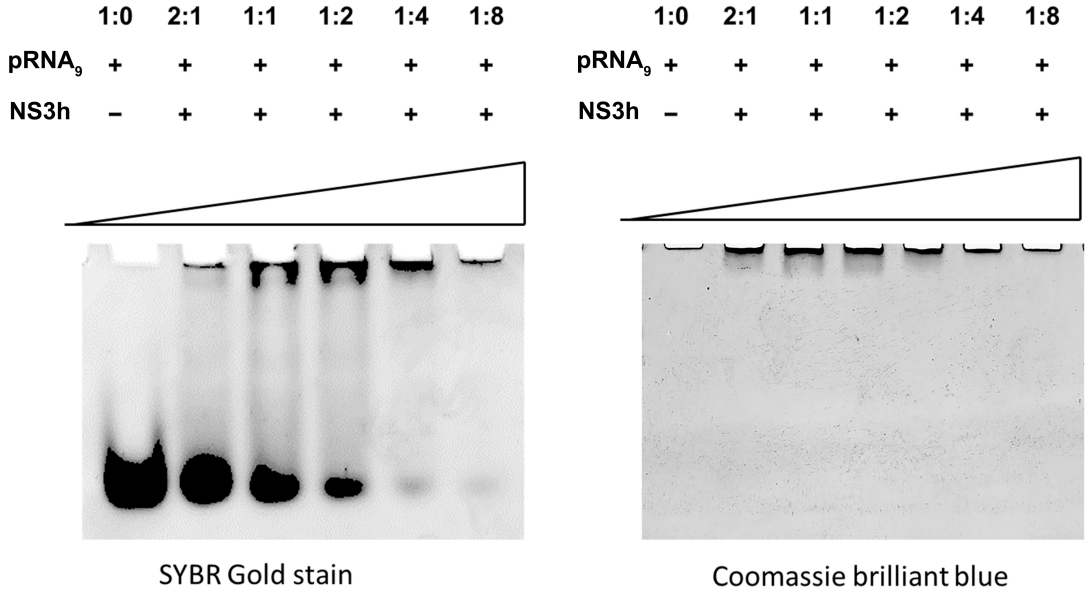


Figure 3-28 RNA binding assay by EMSA.

Left panel: the native polyacrylamide gel stained by SYBR Gold dye. Right panel: the native polyacrylamide gel stained further by Coomassie blue dye. The molar ratio of RNA to protein was shown on the top of the gel.

Electrophoretic mobility shift assays (EMSA) were performed to probe the binding of $NS3h_{183-623}$ and the $pRNA_9$. Considerable attempts to optimisation conditions failed to obtain sharper

band shifts on the native gels. A representative result is given in Figure 3-28, showing that the free RNA indeed decreased as protein concentration increased. However, most of the protein stuck at the top wells and could not migrate into the gel (Figure 3-28, left panel). Still, a little amount of protein entered into the gel near the top wells were detected by staining in Coomassie blue dye (Figure 3-28, right panel). The smeared region observed in the lane above the free RNA when mixing protein with $pRNA_9$ may indicate that NS3h binds to $pRNA_9$ in a weak binding manner, despite the difficulty of quantification of the binding affinity, due to the unstable protein molecules might be precipitating or aggregating during the reaction. In addition, the difficulty in obtaining sharp band shifts is possibly due to the poor stability of NS3h, in particular in buffer containing a low concentration of salt (less than 100 mM), resulting in a reduced concentration of soluble NS3h protein in solution. Moreover, the primarily positively charged surface of NS3h (Figure 3-8b) was protonated and carrying positive charges in the gel running buffer of pH 8.3 (i.e. buffer pH is less than protein pI of 8.6). The positively charged surface of protein may prevent the protein from migrating into the gel and the negative charges that $pRNA_9$ appears not enough to facilitate the protein complex to migrate into the gel.

In order to accurately define the ssRNA binding activity of ZIKV NS3h, several biophysical methods were attempted to quantify the equilibrium dissociation constant (K_d). The propensity for this protein tends to precipitate prevented accurate measurements using fluorescence anisotropy, and no signal was observed using isothermal titration calorimetry and microscale thermophoresis. Nonetheless, the intrinsic tryptophan fluorescence quenching assays were performed to probe the RNA binding of NS3h. To take the advantages of NanoDSF (as detailed in section 2.2.7 and 3.2.8), i.e. no additional fluorophore is needed in the reaction, the intrinsic fluorescence intensities at 350 nm and 330 nm of aromatic residues (mainly Tryptophan and Tyrosine) of ZIKV NS3h were recorded in the presence of a 5' phosphorylated 9-nucleotide ssRNA ($pRNA_9$, Appendix 6).

First, the thermal unfolding transition midpoint (T_m, $^{\circ}C$) of ZIKV $NS3h_{171\text{-}617}$ was measured by monitoring the tryptophan fluorescence emission shifts using NanoDSF upon addition of increasing $pRNA_9$. The $NS3h_{171\text{-}617}$ has two T_m values: the population of T_m2 (the second T_m value) is much smaller than the T_m1 peak (the first T_m value, Table 3-7 and Figure 3-29a). Transition midpoints and onsets are given in Table 3-7. Titration of $pRNA_9$ from 0 to 250 μM, with 5 μM NS3h protein, resulted in the T_m1 value of NS3h increasing by 1.6 $^{\circ}C$, and the T_m2 value increasing by 1.4 $^{\circ}C$ (Figure 3-29a and Table 3-7). These data indicate that $pRNA_9$ binds to NS3h and that the resulting complex has enhanced thermal stability.

Table 3-7 Thermal unfolding midpoints of ZIKV NS3h with different concentration of $pRNA_9$.

$pRNA_9$ (μM)	T_{onset} [°C]	T_m1 [°C]	T_m2 [°C]
0 μM	26.4 ± 1.6	35.9 ± 0.0	42.9 ± 0.0
5 μM	27.8 ± 0.1	35.9 ± 0.1	43.1 ± 0.0
10 μM	25.8 ± 1.9	36.0 ± 0.1	43.2 ± 0.0
15 μM	26.1 ± 0.1	36.0 ± 0.0	43.2 ± 0.2
20 μM	27.5 ± 0.1	36.1 ± 0.0	43.2 ± 0.0
25 μM	26.7 ± 0.2	36.1 ± 0.0	43.2 ± 0.0
30 μM	26.2 ± 0.8	36.1 ± 0.0	43.3 ± 0.0
35 μM	25.7 ± 0.8	36.2 ± 0.0	43.3 ± 0.0
40 μM	27.8 ± 0.8	36.3 ± 0.1	43.4 ± 0.0
45 μM	27.3 ± 0.6	36.4 ± 0.0	43.6 ± 0.0
50 μM	27.8 ± 1.3	36.5 ± 0.0	43.6 ± 0.0
125 μM	28.0 ± 0.0	37.1 ± 0.0	43.8 ± 0.0
250 μM	30.0 ± 0.3	37.5 ± 0.0	44.3 ± 0.0

ZIKV $NS3h_{171-617}$ was used in each reaction. Mean data of sample replicates and plus/minus standard deviations are given in the table. Each sample has six replicates.

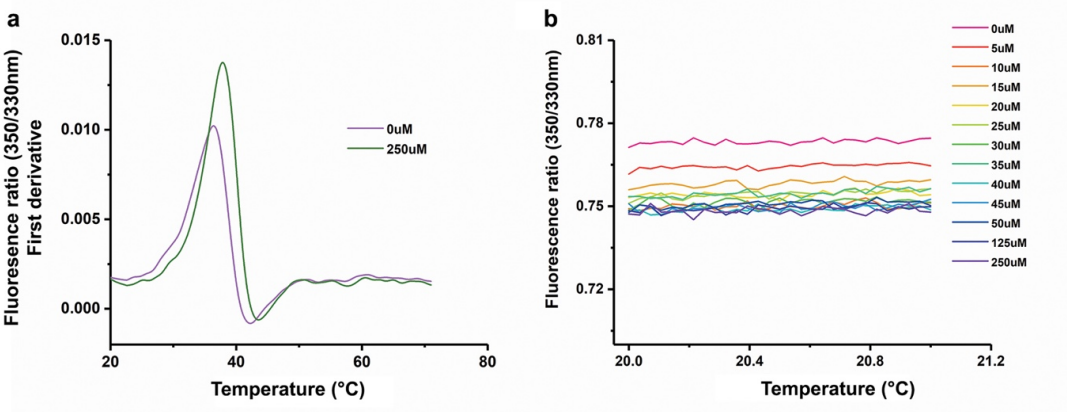


Figure 3-29 Thermal unfolding data of ZIKV NS3h with a titration of $pRNA_9$.

a. The first derivative of the fluorescence ratio of 350/330nm of ZIKV $NS3h_{171-617}$ with titration of 0 and 250 μM $pRNA_9$, b. Fluorescence ratio of 350/330nm of ZIKV $NS3h_{171-617}$ with titration of 1 - 250 μM $pRNA_9$. Data for reactions were recorded at 20 - 21 ˚C, 29 time points averaged for each replicate. 5 μM $NS3h_{171-617}$ was used in each reaction. Each sample has six replicates.

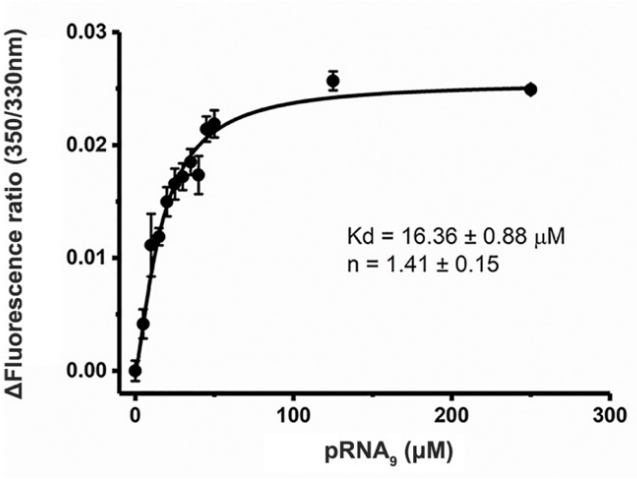


Figure 3-30 The binding affinity of NS3h with $pRNA_9$.

The binding curve of NS3h with $pRNA_9$, the curve is fitted with the Hill equation (234), and equilibrium dissociation constant (K_d) and Hill coefficient (n) are indicated in the figure. The black dots denote the mean data of the sample replicates and error bars on them show the standard mean errors (SE). Each sample has six replicates.

Second, the binding affinity for $NS3h_{171-617}$ was quantified using intrinsic tryptophan/tyrosine fluorescence quenching (TFQ) (233, 268) by adapting the NanoDSF protocol. The ratio of the fluorescence intensities ($F_{350/330}$) at 20 - 20.9 °C clearly showed a decreasing trend as $pRNA_9$ was titrated with protein (Figure 3-29b). The increased $F_{350/330}$ is derived from the free RNA interacting with the protein near where the tryptophans are located. When increased $F_{350/330}$ reaches a steady state, the RNA-bound NS3h is saturating. The TFQ data was analysed and fitted with the Hill equation (Figure 3-30) defining an equilibrium dissociation constant (K_d) of 16.36 ± 0.88μM, which falls into a similar range as DENV NS3h-ssRNA binding model (248). The Hill coefficient (n) is 1.41 ± 0.15, elucidating that the binding of $pRNA_9$ to ZIKV NS3h is probably positive cooperativity between adjacent proteins.

3.4 Discussion

The structural and biochemical data in this chapter provide insights into the ATP hydrolysis and RNA translocation mechanism of the ZIKV NS3h and are also applicable for the *Flaviviridae* family NS3hs.

The Motif V plays a vital role for bridging ATP hydrolysis and RNA translocation

The close agreement between spectroscopic and structural studies on ZIKV NS3h enables a rigorous analysis of the atomic details of the previously unobserved ADP-$MgF_3(H_2O)^-$ species, confirming that it is a new metal fluoride transition state analogue (TSA). Moreover, the ^{19}F NMR spectra demonstrated that a stable ZIKV NS3h-ssRNA-ADP-$MgF_3(H_2O)^-$ complex can

be formed in solution. The clear distinction between the RNA-free and RNA-bound TSA complexes is manifested in subtle yet significant differences in H-bonding network (see in section 3.3.4), suggesting the bound RNA tightens the binding of ADP-$MgF_3(H_2O)^-$. In addition, comparing the complex structures of ZIKV NS3h-ADP-BeF_3^- (PDB code 6RWZ, this work) and HCV NS3h-ssDNA-ADP-BeF_3^- [PDB codes 3KQU and 3KQN, (109)], the shorter H-bonds of the nucleophilic attacking water (W1) to Be^{2+} as a result of DNA binding suggests that nucleic acid binding tightens the pre-transition state analogue complex.

Motif V adopts two conformations in the NS3h-ADP-$MgF_3(H_2O)^-$ TSA complex structure. Significantly, conformation A has not been previously observed in nucleic acid-free TSA complex structures of NS3 helicases of *Flaviviridae* family viruses (Table 3-4). The two conformations adopted by the catalytically important motif V, in the presence of the $MgF_3(H_2O)^-$ TSA demonstrates how ATP hydrolysis affects protein conformation. The direct structural observations in the NS3h-ADP-$MgF_3(H_2O)^-$ complex structure illuminate how this motif V, particularly the conserved glycine (Figure 3-24d), is involved in ATP hydrolysis. In addition to the responsibility of nucleic acid binding in the ZIKV NS3h (in section 3.3.3.1), motif V plays a crucial role for linking the ATP hydrolysis with mechanical translocation of RNA, like other SF2 helicases (109, 210).

How is ZIKV NS3h accommodating the first unwound nucleotide of RNA?

The bound RNA, in the ZIKV NS3h-$pRNA_9$ complex, contains a 5' phosphate group. The structure reveals a more intensive H-bonding network for protein-RNA interactions than observed earlier for the complex with non-phosphorylated RNA (Figure 3-13). Although several interactions are the same as seen in the ZIKV NS3h-ssRNA structure (247), the 5' phosphate induced more compact interactions with the protein (Figure 3-12). Of particular interest is the domain 3 residue D602, which forms multiple water-mediated hydrogen bonding interactions with RNA in the presence of 5' phosphate (Figure 3-31a and 3-31b) and may have a role for stabilising the incoming unpaired nucleotide of RNA during RNA unwinding. Moreover, in the presence of the 5' phosphate, interactions of D540 and K537 shift in register from the fourth and fifth nucleotides to the third and fourth nucleotides of the bound RNA, respectively (Figure 3-31a and 3-31b). Coincidentally, the 5' phosphate group is situated at the position of the first unwound nucleotide in the structure of the archaeal Hel308 SF2 helicase determined in complex with DNA (241), further supporting the premise that the presence of the 5' phosphate likely mimics the first unpaired nucleotide. The 5' phosphate induced remodelling in the H-bonding pattern likely demonstrates how the NS3h protein accommodates the first incoming nucleotide of RNA during translocation.

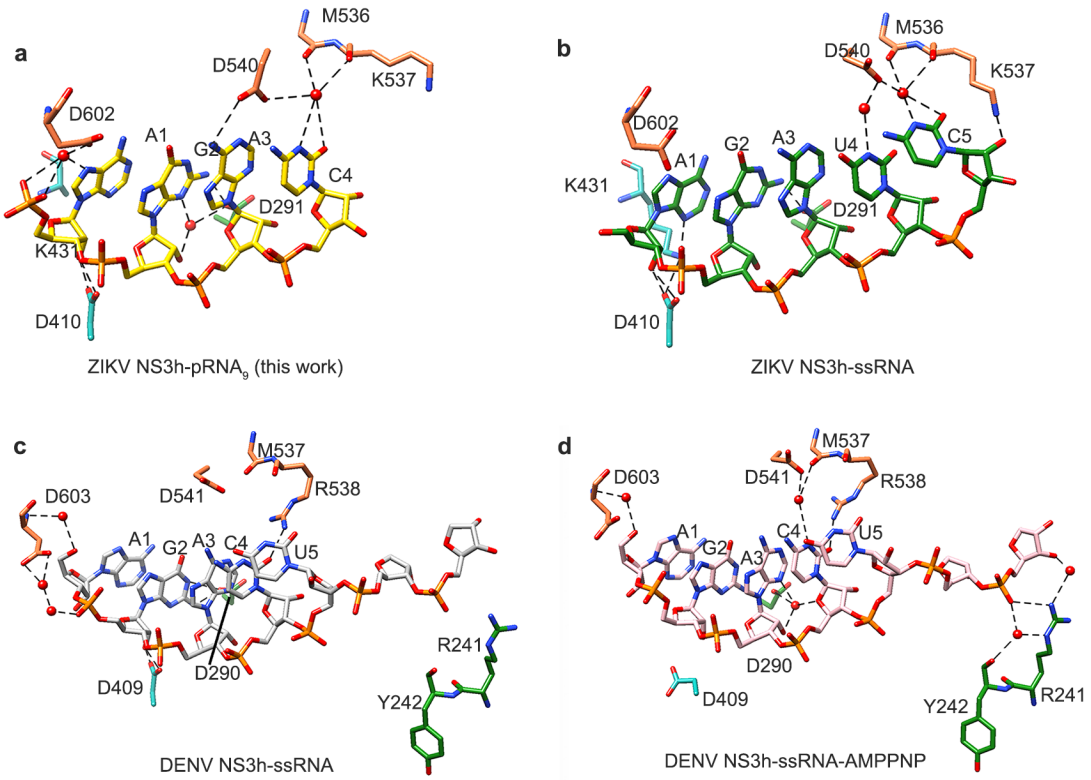


Figure 3-31 Comparisons of H-bonding interactions of RNAs in ZIKV and DENV NS3hs.

a and b. H-bonding interactions of key residues in the RNA binding site of ZIKV NS3h-pRNA$_9$ complex (PDB code 5MFX, this work) (panel a) and ZIKV NS3h-ssRNA complex [PDB code 5JGB, (247)] (panel b). c and d. H-bonding interactions of key RNA binding residues of DENV NS3h-ssRNA complex [PDB code 2JLU, (79)] (panel c) and DENV NS3h-ssRNA-AMPPNP complex [PDB code 2JLV, (79)] (panel d).

A proposed mechanism of RNA translocation of ZIKV NS3h

Despite the lack of a ternary NS3h-ssRNA-ATP complex structure for Zika virus, the combination of the data shows common features with the evolutionarily closely related homologues of ZIKV NS3h (Figure 3-31c and 3-31d). It suggests ZIKV NS3h utilises an inchworm like translocation mechanism. ZIKV NS3 helicase is expected to unwind RNA with 3'→ 5' directionality, like other evolutionary related SF2 helicases (95, 241). The RNA binding site is primarily established by key residues from the ATP-binding domains 1 and 2 (Figure 3-12), indicating RNA translocation is most likely proceeds as a switch of alternate RNA binding interactions between domains 1 and 2, driven by the ATP hydrolysis cycle. Indeed, comparison of the DENV NS3h-ssRNA and NS3h-ssRNA-AMPPNP complex structures shows that D409 (equivalent to D410 in domain 2 of ZIKV NS3h) is no longer

interacting with RNA upon AMPPNP binding (Figure 3-31c and 3-31d), suggesting that the conserved aspartate of motif V favours a 'loose' RNA binding mode when ATP is bound. In addition, R241 (equivalent to R242 in ZIKV NS3h) donates two H-bonds to the RNA when AMPPNP is bound while it does not interact with RNA in the absence of AMPPNP (Figure 3-31c and 3-31d), suggesting the conserved arginine of motif Ib (domain 1) tends to grip RNA firmly upon ATP binding during RNA translocation. Moreover, the interaction of sidechain D290 of motif II (equivalent to D291 in ZIKV NS3h) with RNA shifts one nucleotide towards 3', upon binding AMPPNP (Figure 3-31c and 3-31d), indicating that RNA likely slides over domain 1 when domain 2 rotates towards domain 1 upon ATP binding.

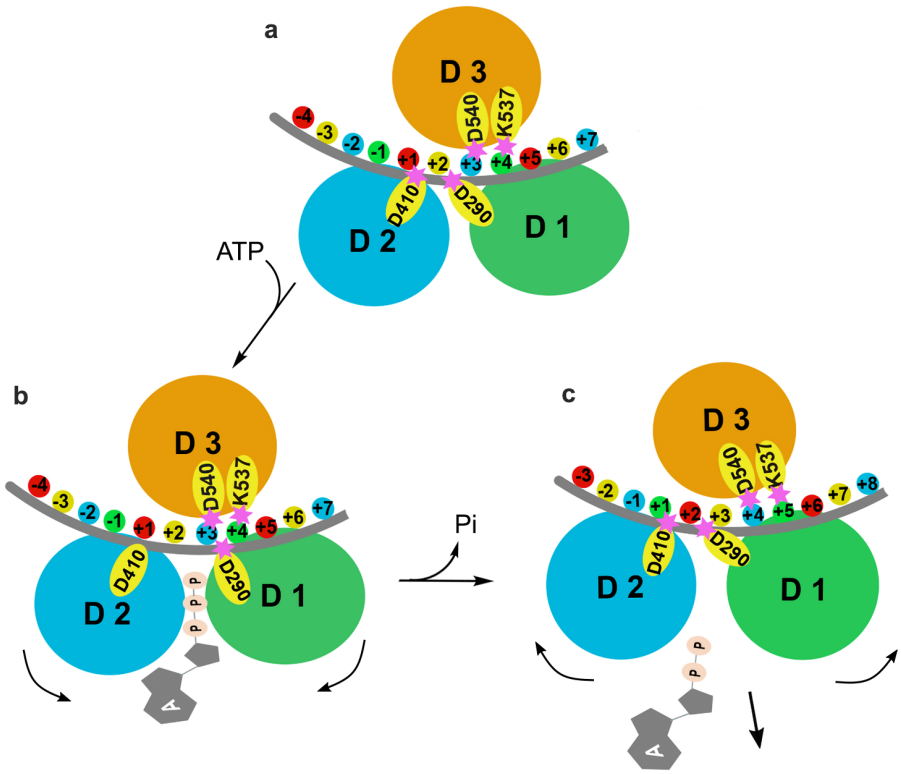


Figure 3-32 Proposed RNA translocation mechanism of ZIKV NS3h.
Domains 1, 2 and 3 (D1, D2 and D3) are in dark green, light blue and orange, respectively, with key residues and factors indicated on the figure.

The identical features of the fundamental conformational changes of the RNA and ATP binding sites shared by *Flavivirus* NS3hs suggest a plausible RNA translocation mechanism. Structural data on comparisons of domain rotations of ZIKV and DENV in the absence of ATP, upon binding to ssRNA, domains 2 and domain 3 move apart from each other to create an 'open' conformation of the RNA binding tunnel in order to accommodate the bound ssRNA (Figure 3-14, section 3.3.3.1), meanwhile, central residues in the RNA binding motifs of domain 2, form a firm grip on the RNA, especially around the 5' region (Figure 3-31a and

Figure 3-32a). The firm grip on the 5' end of the RNA is particularly evident in the description above of the 5' phosphate group induced interactions of the RNA binding (Figure 3-31a). RNA binding is thought to induce an allosteric dynamic effect on the ATP binding site, allowing NS3h to adapt a conformation more suitable for binding and hydrolysing ATP. Next, upon binding ATP, the RecA-like domains 1 and 2 are brought together as domain 2 rotates close to domain 1 (Figure 3-32b). Domain 2 likely remains tightly bound to the RNA strand while the comparatively weaker interaction with domain 1 allows the strand to slide, presumably by one nucleotide. When ATP is correctly situated in the binding pocket, hydrolysis leads to a destabilisation of the cleft closure between domain 1 and 2, in particular, between the 5' region of RNA and D410, resulting in a relatively loose binding mode on domain 2 (Figure 3-32b). Subsequently, domain 1 grips the RNA comparatively more tightly than domain 2, binding to the 3' region of the RNA and domain 2 switches to a relatively weaker binding mode to RNA, potentially 'priming' the pre-translocation complex for movement in this direction. After γ-phosphate release, the interaction of R459 with the nucleotide is broken (Figure 3-15, section 3.3.6.2), potentially allowing domain 2 to rotate away from domain 1. As the RecA-like domains 1 and 2 move apart, the ADP product is released, critical residues in motif V (in particular D410) re-establish contact with RNA, and therefore domain 2 becomes re-engaged in tight interaction with RNA in the new register (Figure 3-32c).

Domain 3 plays a role in RNA translocation and unwinding

Due to the lack of sequence and structural conservation, how domain 3 of NS3h is involved in nucleic acid unwinding and translocation remains unclear. *Flavivirus* NS3hs have no aromatic residue equivalent to W501 in HCV NS3h (109) and W599 in Hel308 (241) positioned for base stacking interactions with the RNA, as discussed in sections 1.3.6 and 3.3.3, suggesting that *Flavivirus* NS3hs likely recruit a different mechanism for RNA translocation. Nevertheless, the interactions of D540 and K537 in domain 3 in ZIKV NS3h with RNA show a one-nucleotide shift towards the 3' end of RNA in the presence of 5' phosphate (Figure 3-31a and 3-31b), demonstrating that conserved aspartate and lysine/arginine are likely the alternatives to the aromatic residue (as discussed in section 3.3.3.1 above). Moreover, D602 tightly binds to the 5' phosphate of RNA but does not interact with RNA in the absence of 5' phosphate (Figure 3-31a and 3-31b), suggesting this conserved aspartate likely plays an essential role in accommodating the incoming nucleotide of RNA during translocation, as discussed in section 3.3.3.1. Since domain 3 of NS3h is relatively smaller and simpler, compared to other SF2 helicases, as discussed in section 3.3.3.1, *Flavivirus* NS3hs may also function in concert with other 'helper' proteins for RNA unwinding, like several other SF2 DEAD-box helicases which do not have a domain 3 and require a 'partner' protein to facilitate unwinding (97, 269).

4 Conclusions and future perspectives

In this thesis, two ATP-driven motor proteins from viruses, the DNA packaging ATPase gp14 from the *Tectiviridae* family phage phiKo and the NS3 helicase (NS3h) from Zika virus, were studied, to investigate their structural and functional properties.

The DNA packaging ATPase phiKo gp14

Several types of ATP-driven motors are used by double-stranded DNA (dsDNA) viruses during their assembly, for packaging viral genomes into preformed capsids. A large group of dsDNA viruses, including eukaryotic *Poxviruses*, giant viruses (such as pandoraviruses), bacterial *Tectiviruses* as well as virophages and polintoviruses, use an FtsK/HerA type packaging system for genome packaging, as described in section 1.2.1. This type of packaging system typically comprises several components, the DNA packaging ATPase, packaging efficiency factor and two membrane associated proteins (14, 41), as discussed in section 1.2.2.3. *Thermus thermophilus* phage phiKo belongs to the *Tectiviridae* family and is expected to utilise the FtsK/HerA type packaging system. As is often the case for viruses, the phiKo sequence differs significantly from other members of the *Tectiviridae* family, except for the conserved major capsids protein and the DNA packaging ATPase, gp14. This work aimed to determine structural and biochemical properties for the phiKo DNA packaging gp14 protein, in order to obtain structural and mechanistic insights about the mechanism of ATP hydrolysis and DNA packaging by the FtsK/HerA type packaging motors.

The solubility of the phiKo gp14 was initially investigated by generating several different fusion constructs and optimising the purification buffer conditions (detailed in sections 2.2.1.2 and 2.3.2). Improving protein solubility proved challenging, with a considerable amount of protein precipitating during protein concentration. Nevertheless, the soluble material of the purified gp14 protein was successfully crystallised, with well-diffracting crystals obtained for the apo structure and its complexes with several different non-hydrolysable ATP analogues (detailed in section 2.3.3).

The crystal structure of the phiKo gp14 reveals that this protein possesses the classical FtsK/HerA fold with the conserved locations of the Walker A and Walker B motifs (discussed in section 2.3.4). The structural conservation indicates that the phiKo gp14 is an ATPase belonging to the FtsK/HerA superfamily that likely translocate DNA using a similar mechanism to other ring-shaped hexameric FtsK/HerA superfamily translocases (6, 72, 197, 209) (see sections 1.3.2 and 1.3.3). In addition, the structures of phiKo gp14 complexes with phosphate, AMPPNP, AMPPCP and ADP provided a precise picture of different functional states of ATP hydrolysis, including the pre-hydrolysis and post-hydrolysis states, as detailed in sections 2.3.4 and 2.3.6. The phiKo gp14 was purified as a monomer in solution and packed

as a monomeric species in the crystal (described in section 2.3.2 and 2.3.4), though the functional oligomeric state of this phiKo gp14 is expected to be a hexamer. Since ATPase active site is typically found at the interface between two adjacent subunits, in common with other FtsK/HerA family members (6, 72, 197, 209), as discussed in sections 1.3.2 and 1.3.3, ATP hydrolysis is expected to facilitate phiKo gp14 oligomerisation possibly by necessitating the arginine finger to protrude into the ATPase active site of the adjacent subunit and facilitating the ATP hydrolysis, as mentioned in section 1.3.3.

The ATPase activity measurements were carried out using the Malachite green assay (detailed in section 2.2.8 and 2.3.9), demonstrating that phiKo gp14 is able to hydrolyse ATP, but relatively weakly, with the ATP hydrolysis rate of 43.5 pmols/μM/min. Increasing NaCl concentration caused a significant decrease in the ATP hydrolysis rate, despite the fact that the high concentration of NaCl helps stabilise the gp14 protein. The activity assays also showed that short N-terminal extensions somewhat inhibited the ATPase activity of phiKo gp14. The construct with an N-terminal 6 × His tag and the gp14 protein with three post-cleavage N-terminal residues displayed lower activity than the untagged phiKo gp14 protein, suggesting that the N-terminus of phiKo gp14 may interfere with ATP hydrolysis. Consistently, the N-terminus is located close to the C-terminus that interacts with the nucleotide and is involved in forming the ATP binding pocket (as described in section 2.3.6), indicating that N-terminus may influence ATP hydrolysis. Future studies may benefit from expressing a construct of phiKo gp14 using a no-tag expression vector, for probing the potential influence on oligomerisation. The ATPase activity of phiKo gp14 decreased over time, likely due to the poor protein stability. Future biochemical characterisation of the gp14 may be directed at the optimisation of conditions that balance salt concentration (and protein solubility) as well as gp14 ATPase activity.

To further investigate the possibilities of oligomerisation of phiKo gp14, the potential oligomerisation by protein docking was adopted to generate a putative hexameric model of gp14. The modelling of gp14 (in section 2.3.7) indicated that the formation of the hexameric assembly might be facilitated by forming a higher-order complex with an auxiliary protein component, to assemble into a hetero-dodecamer. Indeed, the DNA packaging machinery assembled at the unique vertex of the viral capsid is expected to employ several other protein factors, in addition to the genome packaging ATPase (as discussed in sections 1.1.1 and 1.2.2). Based on previous 12 Å resolution data on *Tectiviridae* phage PRD1, the DNA packaging ATPase and the packaging efficiency factor form a dodecameric portal complex at the packaging vertex (14), though the detailed arrangement of protein subunits of this portal complex remains unknown. It is likely that the packaging ATPase gp14 of phage phiKo is forming a hexameric or a dodecameric assembly *in vivo* (see sections 1.2.2.3 and 1.3.3).

In addition, the protein stability and DNA binding activity were investigated to probe the functional properties of phiKo gp14 (detailed in sections 2.3.8 and 2.3.10). The protein stability was assayed in buffers containing different salt concentrations by NanoDSF, showing that the phiKo gp14 is relatively stable, with melting temperatures higher than 53 ˚C in Tris pH 7.5 buffer and higher than 64.5 ˚C in citrate buffer pH 5.0. The melting temperature of phiKo gp14 is lower than the optimal *Thermus thermophilus* bacteriophage growth temperature of 70 ˚C, suggesting that a potential 'partner' protein may be stabilising gp14 *in vivo*. In consistence with melting temperature measurements, well-diffracting crystals of gp14 grew from the crystallisation conditions containing citrate ions at pH 4.0 - 5.0, further suggesting that citrate ion or low pH buffer or both are potentially important for protein stability. The DNA binding activity of gp14 was assayed by EMSA (in section 2.3.10). These experiments were challenging due to a considerable amount of protein precipitating and remaining in the wells of the gel. The results demonstrate that the protein can bind dsDNA weakly and non-specifically. Further investigation of the DNA binding activity of gp14 should be performed with the oligomeric form of phiKo gp14, which could be stabilised by auxiliary protein factors.

This study also attempts to investigate the DNA packaging activity of the gp14, as detailed in section 2.3.11. Similar to other *Tectiviridae* family viruses, phage phiKo is expected to recruit a protein-primed DNA replication mechanism, in which the terminal protein remains covalently attached to the 5' genome after DNA replication (270), as described in section 1.2.2.2. The complex of terminal protein-genomic DNA is packaged into the procapsid through the unique vertex while the terminal protein may function as a valve to seal the channel of the packaging system once the genome is inside in the capsid (14). Likewise, in phage phiKo, there should also be a 'terminal protein' with functions similar to the one in phage PRD1. However, due to the lack of detectable sequence similarity amongst terminal proteins, the putative terminal protein of phiKo has not yet been identified. The lack of packaging activity in the assay, performed *in vitro* by adding viral genomic DNA, gp14 and ATP to the solution containing empty phiKo capsids, may thus be due to the lack of the terminal protein. Thus, a future study directed at identifying the potential 'terminal protein' and characterising its role in DNA packaging could be critical for further investigation of the DNA packaging activity of gp14.

As shown for other members of the *Tectiviridae* family, the DNA packaging ATPase remains on the mature virion after genome packaging (14, 41, 271). The mature virions were successfully isolated and purified (detailed in section 2.3.11). However, the genomic DNA leaked out from the purified mature virions over time (after approximately 1 week storage at 4 ˚C, data not shown), despite extensive buffer optimisations performed for stabilising the

packaged phage particles, suggesting that *in vitro* the protein components preventing the packaged DNA leakage or maintaining the capsid structure or both, might be dissociating during the storage at 4 °C or at room temperature. While storing capsids at - 80 °C after flash freezing in liquid nitrogen ensured phage preservation, the phage titre dropped during the long term, over months, storage. Future work on investigating phiKo gp14 assembly and the mechanism of DNA packaging would benefit from 3-D reconstruction of the unique vertex of the mature virion of phage phiKo at high resolution by cryoEM**.** The data presented here suggest that grids for cryoEM would have to prepared immediately after phage production and purification unless better conditions are identified for preserving the integrity of phage particles. Finally, obtaining pro-capsids that are competent for packaging, is required for establishing DNA packaging *in vitro*. The research may further benefit from generating phiKo phage mutants, such as gp14 defective mutant, that may facilitate producing packaging competent procapsids.

Zika virus NS3 helicase

Zika virus (ZIKV) NS3 helicase (NS3h) belongs to the superfamily 2 (SF2) helicases having the conserved helicase motifs spread over two RecA-like core domains (69, 75, 79, 97, 112), as shown in section 3.3.3. Such conserved helicase motifs are either responsible for RNA binding or NTP hydrolysis (79, 112). The NS3h can enzymatically unwind double-stranded RNA (dsRNA) using energy derived from the NTP hydrolysis. However, the mechanism for coupling RNA translocation to the NTP hydrolysis events of the ZIKV NS3h remains unclear. The goal of this work was to gain a better understanding of the mechanisms of RNA translocation, NTP hydrolysis as well as coupling between these two processes, by determining the structural and functional properties of this NS3 helicase.

X-ray structures of NS3h in complex with several ATP analogues, including the pre-hydrolysis (ATPγS and ADP-BeF_3^-), transition state [ADP-$MgF_3(H_2O)^-$] and post-hydrolysis (ADP) analogues, were determined, as detailed in section 3.3.3. These structures present the NS3 helicase in several different functional states of ATP hydrolysis. Notably, it turned out that the ADP-$MgF_3(H_2O)^-$ structure obtained in this study was previously unobserved and was validated by a combination of crystallography and ^{19}F NMR. Structural observations also revealed that the catalytically important helicase motif V possesses two conformations in the NS3h-ADP-$MgF_3(H_2O)^-$ complex (in section 3.3.5). Significantly, the conformation A has not been previously observed in nucleic acid-free TSA complexes of NS3 helicases from *Flaviviridae* family viruses (Table 3-4). In addition to its role in nucleic acid binding by motif V residues, T409, D410 and V411, in the NS3h-$pRNA_9$ complex structure, motif V also plays a crucial role in linking the ATP hydrolysis with mechanical translocation of RNA. Moreover, the NS3h-ssRNA-ADP-$MgF_3(H_2O)^-$ complex can be formed in solution and was verified by

the ^{19}F NMR spectra, though attempts to co-crystallise the ternary complex of NS3h-ssRNA-ADP-$MgF_3(H_2O)^-$ have not yet been successful.

The structure of the NS3h-$pRNA_9$ complex revealed that the presence of the 5' phosphate on the RNA induced more compact interactions with the protein, as seen in section 3.3.3.1. In particular, the 5' phosphate induced remodelling of the H-bonding network indicating how the NS3h protein accommodates the first incoming nucleotide of RNA during unwinding. Of particular interest are the domain 3 residues D602, D540 and K537 that adopt different conformations in the presence of the 5' phosphate, compared with the absence of the 5' phosphate. D602 tightly binds to the 5' phosphate of RNA but does not interact with RNA in the absence of 5' phosphate, suggesting this conserved aspartate likely plays an essential role in accommodating the incoming nucleotide of RNA during translocation, as discussed in sections 3.3.3.1 and 3.4. In addition, *Flavivirus* NS3hs have no aromatic residue equivalent to tryptophan residue in other members of the SF2 helicases, such as W501 in HCV NS3h (109) and W599 in Hel308 (241), positioned to form base stacking interactions with the RNA, as mentioned in section 1.3.6. These observations indicate that there are mechanistic variations in *Flavivirus* NS3hs. Furthermore, the interactions of D540 and K537 with RNA in the case of the Zika virus NS3h (this study) may reveal a one-nucleotide shift of RNA towards the 3' end in the presence of 5' phosphate, demonstrate that conserved aspartate and lysine/arginine likely play a similar role to the aromatic residue, as discussed in section 3.4.

The protein stability of NS3h was assayed by NanoDSF (in section 3.3.6.1), and the results indicate that the NS3h is not particularly stable, with the melting temperature of 36 $^{\circ}C$. Moreover, the protein tends to precipitate over time. The ATPase activity measurements were carried out by NADH-coupled ATPase activity assays (in section 3.3.6.2), demonstrating the ATPase activity of NS3h decreases over time and can be stimulated by ssRNA. The RNA binding assays were performed using intrinsic tryptophan fluorescence quenching (in section 3.3.6.3), illustrating that the NS3h binds RNA weakly and non-specifically with a dissociation constant (K_d) of 16.36 μM.

Taken together, similarities in the conformational changes associated with RNA and ATP binding sites, shared by *Flavivirus* NS3hs, suggest a plausible inchworm-like mechanism for translocation along the RNA and an accompanying unwinding of the duplex strands, as discussed in section 3.4. This mechanism is based on sequential cycling between alternatively tight and weak RNA binding states of domains 2 followed by domain 1. The structural and biochemical data are detailed in chapter 3, providing insights into the ATP hydrolysis and RNA unwinding mechanism of the ZIKV NS3h. The data are applicable to other members of the *Flaviviridae* family NS3hs. Future study would benefit from focusing on understanding the unwinding mechanism of this helicase in more detail. In particular, it would be valuable to

understand how NS3h performs duplex strand separation. Such studies would benefit from (i) obtaining high resolution structures of NS3h in complex with hairpin RNA or RNA replication fork, either by X-ray crystallography or cryoEM and from (ii) associated further biochemical characterisation.

Appendices

Appendix 1 Primers used in this study.

Primer name	Primer sequences 5' → 3'
pYM350gp14_For	GAACAGATTGGTGGTATGGACTGGAGGCGGATAG
pYM350 gp14_Rev	TACCTAAGCTTGTCTTCATTCTAACCTCAACATGTAAGG
pYM351 gp14_For	ATCACCACCACCACATGGACTGGAGGCGGATAG
pYM351 gp14_Rev	TGAGGAGAAGGCGCGTCATTCTAACCTCAACATGTAAGG
pYM409 gp14_For	CCAGGGACCAGCAATGGACTGGAGGCGGATAG
pYM409 gp14_Rev	TGAGGAGAAGGCGCGTCATTCTAACCTCAACATGTAAGG
pYM418 gp14_For	ATCACCACCACCACGGGCATAGCGATAGGGTTG
pYM418 gp14_Rev	TGAGGAGAAGGCGCGTCATTCTAACCTCAACATGTAAGG
pYM419 gp14_For	AAGGAGATATACATATGATGGGGCATAGCGATAGGGTTG
pYM419 gp14_Rev	GGTGGTGGTGCTCGAGTTCTAACCTCAACATGTAAGG
pYM537 gp14_A	ATCACCACCACCACATGGACTGGAGGCGGATAG
pYM537 gp14_B	TCCACCTTCGCTACCTCCACCTTCTAACCTCAACATGTAAGG
pYM537 gp14_C	GGTGGAGGTAGCGAAGGTGGAATGGACTGGAGGCGGATAG
pYM537 gp14_D	TGAGGAGAAGGCGCGTCATTCTAACCTCAACATGTAAGG
pYM546 gp14_A	ATCACCACCACCACATGGACTGGAGGCGGATAG
pYM546 gp14_B	TCCACCTTCGCTACCTCCACCTTCTAACCTCAACATGTAAGG
pYM546 gp14_C	AGTGAGGGAGGAAGTGAGGGCATGGACTGGAGGCGGATAG
pYM546 gp14_D	TGAGGAGAAGGCGCGTCATTCTAACCTCAACATGTAAGG
pYM569 gp14_A	ATCACCACCACCACATGGACTGGAGGCGGATAG
pYM569 gp14_B	CTCACTACCTCCACCTTCGCTACCTCCACCTTCTAACCTCAACATGTAA
pYM569 gp14_C	AGCGAAGGTGGAGGTAGTGAGATGGACTGGAGGCGGATAG
pYM569 gp14_D	TGAGGAGAAGGCGCGTCATTCTAACCTCAACATGTAAGG
pYM635 gp14_For	ACCACGGGGGCGGGGGCGTCCACTCTCCTTAGG
pYM635 gp14_Rev	CCTAAGGAGAGTGGACGCCCCCGCCCCCGTGGT
pYM636 gp14_For	CTGACGGTCGTTGTGGCCGAGGCTTATCTAGC
pYM636 gp14_Rev	GCTAGATAAGCCTCGGCCACAACGACCGTCAG
pYM637 gp14_For	TGTCTGGCACAAGGGGCAAGCCGAGGGGTGGCC
pYM637 gp14_Rev	GGCCACCCCTCGGCTTGCCCCTTGTGCCAGACA
pYM638 gp14_For	GATGAGCTCTACAAAATGGACTGGAGGCGGATAG
pYM638 gp14_Rev	GTGGTGGTGCTCGAGTCATTCTAACCTCAACATGTAAGG
Linearization for pYM638_For	CTCGAGCACCACCACCACCAC
Linearization for pYM638_Rev	TTTGTAGAGCTCATCCATGCC
T7	TAATACGACTCACTATAGGG

T7_term	CTAGTTATTGCTCAGCGGT
Linearization for YSBL-LIC(-)_For	CGCGCCTTCTCCTCACTGTTCCAGG
Linearization for YSBL-LIC(-)_Rev	TGTGGTGGTGGTGATGATGGCTGC
Linearization for YSBL-LIC3C _For	TTGCTGGTCCCTGGAACAGAACTTCC
Linearization for YSBL-LIC3C _Rev	CGCGCCTTCTCCTCACATATGGCTAGC
Linearization for pET SUMO _For	AGACAAGCTTAGGTATTTATTCGG
Linearization for pET SUMO _Rev	ACCACCAATCTGTTCTCTGTGA

'For' indicates the forward primer, and 'Rev' indicates the reverse primer. Primers A-D for pYM537, pYM546 and pYM 569 are detailed in section 2.2.1.3.

Appendix 2 Buffers used in solubility screens for phiKo gp14.

No.	Buffer composition
1	50 mM Tris 7.5, 5% glycerol, 0.5 M NaCl, 2 mM $MgCl_2$
2	50 mM Tris 7.5, 10% glycerol, 0.5 M NaCl, 2 mM $MgCl_2$
3	50 mM Tris 7.5, 15% glycerol, 0.5 M NaCl, 2 mM $MgCl_2$
4	50 mM Tris 7.5, 5% glycerol, 0.25 M NaCl, 2 mM $MgCl_2$
5	50 mM HEPES 7.5, 10% glycerol, 0.5 M NaCl, 2 mM $MgCl_2$
6	50 mM HEPES 7.5, 5% glycerol, 0.5 M NaCl, 2 mM $MgCl_2$
7	50 mM Tris 7.5, 15% glycerol, 0.25 M NaCl, 2 mM $MgCl_2$
8	50 mM Tris 7.5, 10% glycerol, 0.25 M NaCl, 2 mM $MgCl_2$
9	50 mM HEPES 7.5, 5% glycerol, 0.25 M NaCl, 2 mM $MgCl_2$
10	50 mM HEPES 7.5, 10% glycerol, 0.25 M NaCl, 2 mM $MgCl_2$
11	50 mM HEPES 7.5, 15% glycerol, 0.25 M NaCl, 2 mM $MgCl_2$
12	0.25 M K-phosphate 7.5, 5% glycerol, 0.5 M NaCl, 2 mM $MgCl_2$
13	0.25 M K-phosphate7., 10% glycerol, 0.5 M NaCl, 2 mM $MgCl_2$
14	50 mM MES 6.0, 15% glycerol, 0.5 M NaCl, 2 mM $MgCl_2$

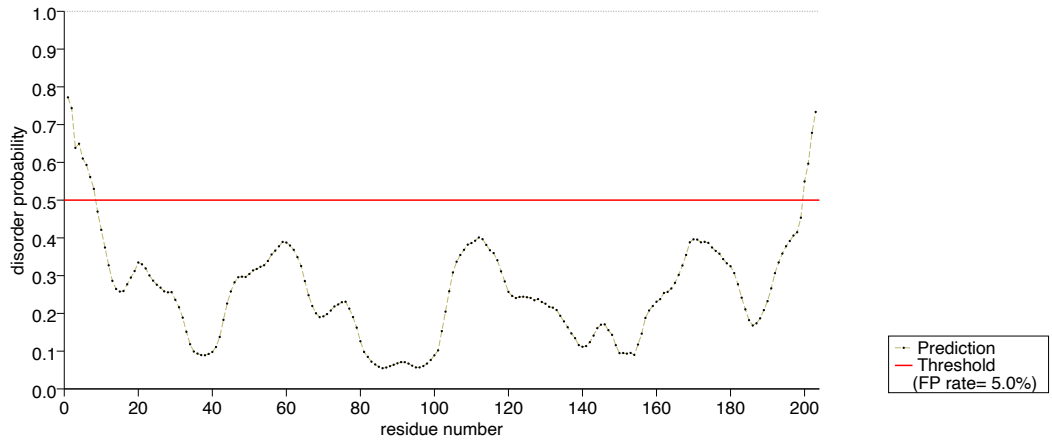
15	50 mM Tris 7.5, 15% glycerol, 0.5 M NaCl, 2 mM $MgSO_4$
16	50 mM Tris 8.0, 5% glycerol, 0.5 M NaCl, 20 mM $MgCl_2$

Appendix 3 Buffers used for protein purification.

Protein	Construct	Buffer name	Buffer composition
phiKo gp14	pYM350	Lysis buffer	50 mM Tris 7.5, 1 M NaCl, 10% glycerol and 5 mM DTT
		NAC binding buffer	50 mM Tris 7.5, 1 M NaCl, 10% glycerol, 20 mM imidazole and 1 mM DTT
		SEC buffer	20 mM Tris 7.5, 1 M NaCl, 10% glycerol, 0.2 mM DDM and 1 mM DTT
	pYM351	Lysis buffer	50 mM Tris 8.0, 0.5 M NaCl, 5% glycerol, 20 mM MgCl2 and 5 mM DTT
		NAC binding buffer	50 mM Tris 8.0, 0.5 M NaCl, 5% glycerol, 20 mM $MgCl_2$, 30 mM imidazole and 2 mM DTT
		SEC buffer	20 mM Tris 8.0, 0.5 M NaCl, 5% glycerol, 20 mM $MgCl_2$ and 2 mM DTT
	pYM409	Lysis buffer	50 mM Tris 7.5, 0.5 M NaCl, 20 mM imidazole and 2 mM DTT
		NAC binding buffer	50 mM Tris 7.5, 0.5 M NaCl, 20 mM imidazole and 2 mM DTT
		SEC buffer	20 mM Tris 7.5, 1 M NaCl, 0.07% OTG and 2 mM DTT
	pYM418	Lysis buffer	50 mM Tris 8.0, 5% glycerol, 0.5 M NaCl, 20 mM $MgCl_2$, 2 mM DTT and 20 mM imidazole
		NAC binding buffer	50 mM Tris 8.0, 5% glycerol, 0.5 M NaCl, 20 mM $MgCl_2$, 2 mM DTT and 20 mM imidazole
	pYM419	Lysis buffer	50 mM HEPES 7.0, 5% glycerol, 0.5 M NaCl, 20 mM $MgCl_2$, 2 mM DTT and 20 mM imidazole
		NAC binding buffer	50 mM HEPES 7.0, 5% glycerol, 0.5 M NaCl, 20 mM $MgCl_2$, 2 mM DTT and 20 mM imidazole
	pYM537	Lysis buffer	50 mM Tris 7.5, 5% glycerol, 1 M NaCl, 5 mM DTT and 20 mM imidazole
		NAC binding buffer	50 mM Tris 7.5, 5% glycerol, 1 M NaCl, 5 mM DTT and 20 mM imidazole
	pYM546	Lysis buffer	50 mM Tris 7.5, 5% glycerol, 1 M NaCl, 3 mM DTT and 20 mM imidazole

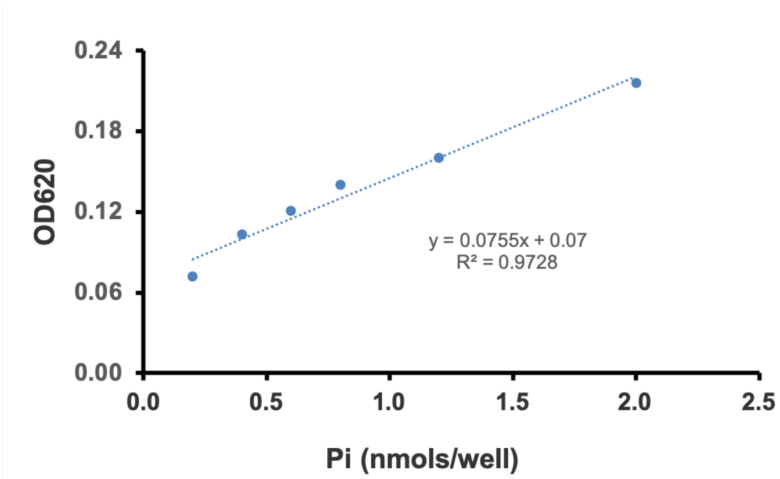
		NAC binding buffer	50 mM Tris 7.5, 5% glycerol, 1 M NaCl, 3 mM DTT and 20 mM imidazole
	pYM569	Lysis buffer	50 mM Tris 8.0, 1 M NaCl, 2 mM DTT, 0.07% OTG and 20 mM imidazole
		NAC binding buffer	20 mM Tris 8.0, 1 M NaCl, 2 mM DTT and 0.07% OTG
	pYM635	Lysis buffer	50 mM Tris 7.5, 1 M NaCl, 10% glycerol, 2 mM DTT and 20 mM imidazole
		NAC binding buffer	50 mM Tris 7.5, 1 M NaCl, 10% glycerol, 2 mM DTT and 20 mM imidazole
		SEC buffer	20 mM Tris 7.5, 1 M NaCl, 10% glycerol, 2 mM DTT and 0.2 mM DDM
	pYM637	Lysis buffer	50 mM Tris 7.5, 1 M NaCl, 10% glycerol, 2 mM DTT and 20 mM imidazole
		NAC binding buffer	50 mM Tris 7.5, 1 M NaCl, 10% glycerol, 2 mM DTT and 20 mM imidazole
		SEC buffer	20 mM Tris 7.5, 1 M NaCl, 10% glycerol, 2 mM DTT and 0.2 mM DDM
	pYM638	Lysis buffer	50 mM Tris 7.5, 1 M NaCl, 2 mM DTT and 20 mM imidazole
		NAC binding buffer	50 mM Tris 7.5, 1 M NaCl, 2 mM DTT and 20 mM imidazole
		SEC buffer	20 mM Tris 7.5, 0.5 M NaCl and 2 mM DTT
		IEC binding buffer	50 mM Tris 7.5, 100 mM NaCl, 2 mM DTT
Zika virus NS3 helicase	$NS3h_{183-623}$	Lysis buffer	20 mM Tris pH 7.5, 500 mM NaCl, 30 mM imidazole, 2 mM DTT
		NAC binding buffer	20 mM Tris pH 7.5, 500 mM NaCl, 30 mM imidazole, 2 mM DTT
		SEC buffer	20 mM Tris 7.5, 300 mM NaCl, 2 mM DTT
	$NS3h_{171-617}$	Lysis buffer	50 mM Tris 7.5, 500 mM NaCl, 2 mM DTT, 20 mM imidazole
		NAC binding buffer	50 mM Tris 7.5, 500 mM NaCl, 2 mM DTT, 20 mM imidazole
		SEC buffer	50 mM Tris 7.5, 300 mM NaCl,10% glycerol. 2 mM DTT

NAC stands for Nickel affinity chromatography; SEC is short for size exclusion chromatography and IEC indicates ion-exchange chromatography.



Appendix 4 Disorder prediction of phiKo gp14 protein by PrDOS.

The full-length sequence of phiKo gp14 protein was used for the prediction in web server PrDOS (186).



Appendix 5 Phosphate standard curve for Malachite green assay.

The standard procedure was applied to the samples containing 0.2 nmols to 2.0 nmols inorganic phosphate per well. KH_2PO_4 solution was used for the measurements.

Appendix 6 DNA and RNA oligos used in this study.

Name	**Sequence (5' → 3')**	**fluorescence label**
12-mer dsDNA	CCTACGCGTAGG	No

16bp dsDNA	For: ACCAGGGCAGGGCGAC Rev: GTCGCCCTGCCCTGGT	5’ end of forward strand was labelled with Cy5
30bp dsDNA	For: ACGCTGGCTAAAGGGAAGGGCAATGACGAG Rev: CTCGTCATTGCCCTTCCCTTTAGCCAGCGT	5’ end of forward strand was labelled with Cy5
20bp dsDNA	For: GCAGGAGGACGTAGGGTAGG Rev: CCTACCCTACGTCCTCCTGC	5’ end of forward strand was labelled with Cy5
$pRNA_9$	p-AGACUCCAU	No
$ssRNA_9$	AGACUCCAU	No
Hairpin RNA with 3’ 9-mer overhang	CUGGGAAACCAGAGACUCCAU	No
$ssRNA_{20}$	GGGUACCGAGCUCGAAUUCG	No
$ssRNA_{40}$	CGAAUUCGAGCUCGGGUACC-A_{20}	No
$ssRNA_{55}$	CGAAUUCGAGCUCGGUACCC-U_{35}	No

F and R strands of each dsDNA were annealed at 95 °C for 10 min and slowly cool to RT to allow dsDNA formation; A_{20} stands for the sequence containing 20-mer polyA; U_{35} is the sequence containing 35-mer polyU.

Appendix 7 Sequences of gp14 fusion constructs.

> Full length gp14

MDWRRIGHSDRVAVVGTTGAGKSTLLRAMMASRPWVLLYDPKGEWEGEEWLTVREAAKLPRK
PHPRVRYFPLPDELRDKEALDDFFWWAYYAKNLTVVVDEAYLATLAGAWLPPGYQACLAQGR
SRGVAVWTATQRPYKIPQTILSEAEHWFVFRLNAPQDRWAVHQWTGLDSRLISPENLPKRVW
YYVSEGQVYGPYMLRLE

> N-terminal 6 × His His-tag

MGSSHHHHHH

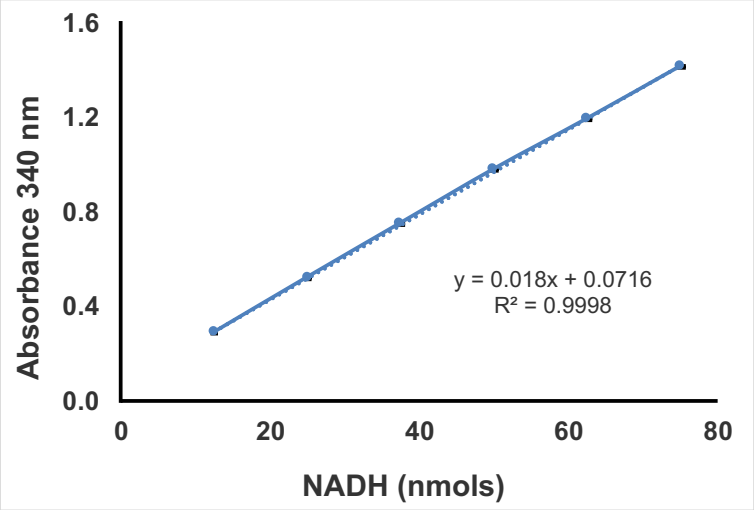
> N-terminal 6 × His SUMO tag

MGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKK
TTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG

> N-terminal 6 × His GFP-TEV tag

MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGSEFELRRRENLYFQSLSKGEELFTGVVPI
LVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPD
HMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILG
HKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHY
LSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMDELYK**SEGGGSE**

*The 6 × His tag is highlighted in bold and the linker between GFP and phiKo gp14 is highlighted in bold and with underscore.



Appendix 8 The path-length curve for NADH-coupled microplate photometric assay.

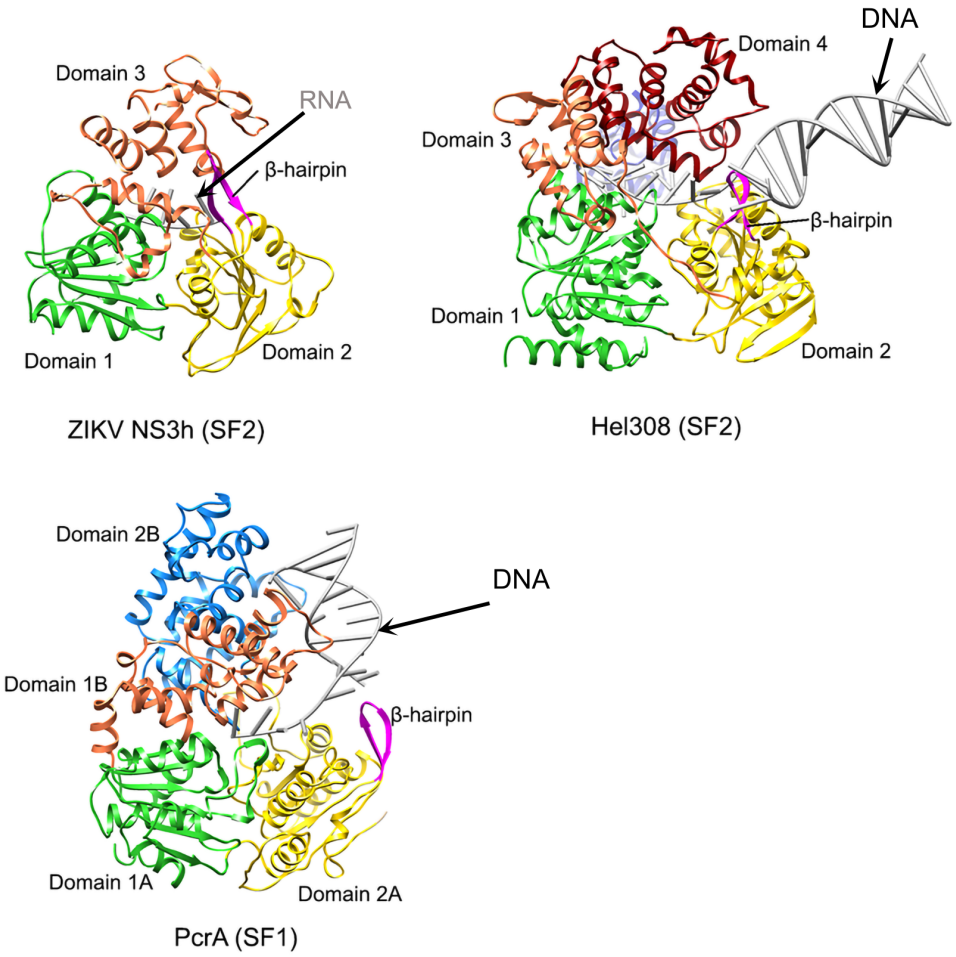
Path-lengths for different fill volumes were calculated from single absorbance readings using the molar absorption coefficient for NADH in water (6,200 OD × mol^{-1} × cm^{-1}) as a reference. Added 12.5, 25, 37.5, 50, 62.5 and 75 nmols (i.e. 50 µl, 100 µl, 150 µl, 200 µl, 250 µl and 300 µl of 250 µM) NADH in wells on a flat bottom 96 well plate, respectively. The absorbance at wavelength of 340 nm (A340) of each well were recorded. Each concentration has three replicates. The molar absorption coefficient for NADH in water is 6,200 OD × mol^{-1} × cm^{-1}. The pathlength of well (cm) = A340 / final NADH concentration / 6,200 OD × mol^{-1} × cm^{-1}.

Appendix 9 List of buffers for NS3h$_{183-623}$ solubility test.

No.	Buffers	Solubility
1	50 mM Tris 8.0, 5% glycerol, 0.5 M NaCl, 2 mM $MgCl_2$	Y
2	50 mM Tris 8.0, 5% glycerol, 0.25 M NaCl, 2 mM $MgCl_2$	Y

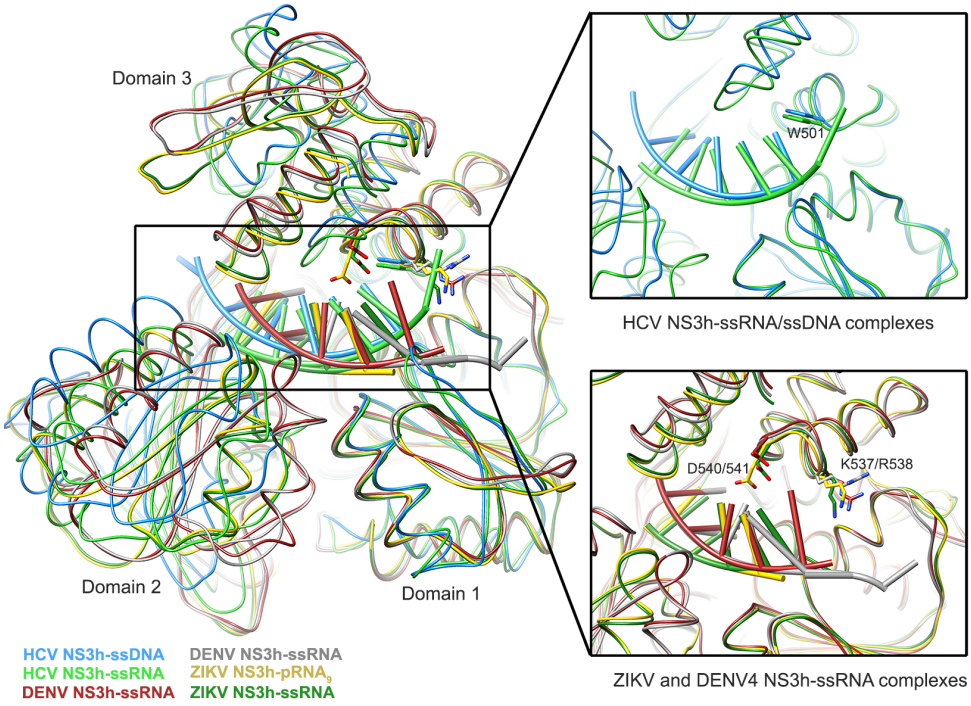
3	50 mM Tris 8.0, 5% glycerol, 0.5 M NaCl, 20 mM $MgCl_2$	Y
4	50 mM Tris 8.0, 5% glycerol, 0.25 M Na_2SO_4, 20 mM $MgCl_2$	Y
5	50 mM Tris 8.0, 5% glycerol, 0.5 M Na_2SO_4	Y
6	50 mM Tris 8.0, 5% glycerol, 1 M NaCl	Y
7	50 mM HEPES 7.0, 5% glycerol, 0.5 M NaCl, 2 mM $MgCl_2$	Y
8	50 mM HEPES 7.0, 5% glycerol, 0.25 M Na_2SO_4, 2 mM $MgCl_2$	Y
9	50 mM HEPES 7.0, 5% glycerol, 0.5 M Na_{Cl}, 20 mM $MgCl_2$	Y
10	50 mM HEPES 7.0, 5% glycerol, 0.25 M Na_2SO_4, 20 mM $MgCl_2$	Y
11	0.25 M K-phosphate 7.5, 5% glycerol, 0.5 M NaCl, 20mM $MgCl_2$	Y
12	0.25 M K-phosphate 7.5, 5% glycerol, 0.25 M Na_2SO_4, 20mM $MgCl_2$	Y
13	0.25 M K-phosphate 7.5, 5% glycerol, 0.5 M NaCl, 2 mM $MgCl_2$	Y
14	0.25 M K-phosphate 7.5, 5% glycerol, 0.25 M Na_2SO_4, 2 mM $MgCl_2$	Y
15	50 mM MES 6.0, 5% glycerol, 0.5 M NaCl, 20 mM $MgCl_2$	Y
16	50 mM MES 6.0, 5% glycerol, 0.25 M Na_2SO_4,20 mM $MgCl_2$	Y
17	50 mM Na-citrate 4.0, 5% glycerol, 0.5 M NaCl, 20 mM $MgCl_2$	N
18	50 mM Na-citrate 4.0, 5% glycerol, 0.25 M Na_2SO_4, 20 mM $MgCl_2$	N
19	50 mM Tris 7.5, 5% glycerol, 1 M NaCl	Y
20	50 mM Tris 7.5, 1 M NaCl	Y

Y – target protein was soluble in this buffer, N – target protein was not soluble in this buffer.



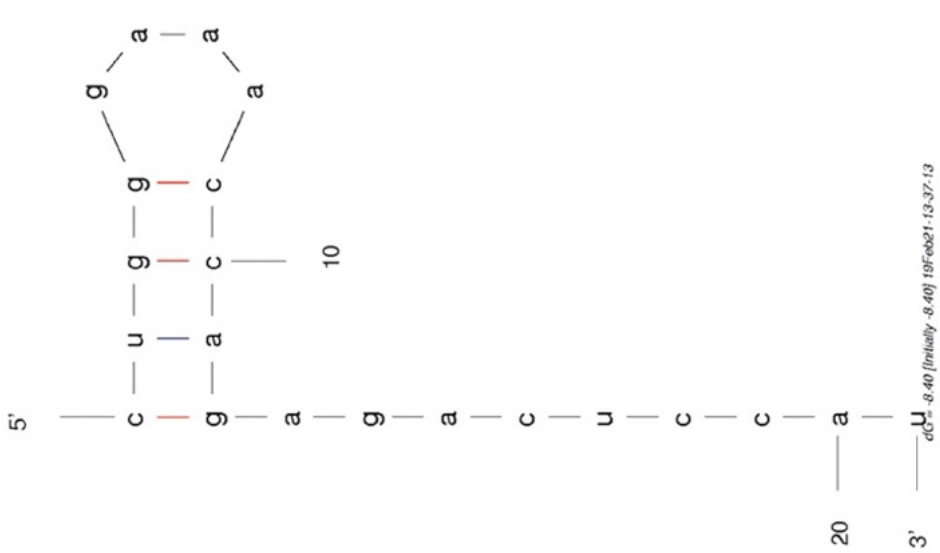
Appendix 10 Comparison of NS3h-pRNA$_9$, Hel308-DNA and PcrA-DNA complex structures.

Overall structure comparison of NS3h-pRNA$_9$ (PDB code 5MFX, this work), Hel308 SF2 helicase [PDB code 2P6R, (241)] and PcrA SF1 helicase [PDB code 3PJR, (78)]. The conserved RecA-like domains 1 and 2 are in green and yellow, respectively. The domain 3 is in coral. Proteins are shown in ribbon. Nucleic acids are in grey. The β-hairpin is highlighted in magenta.



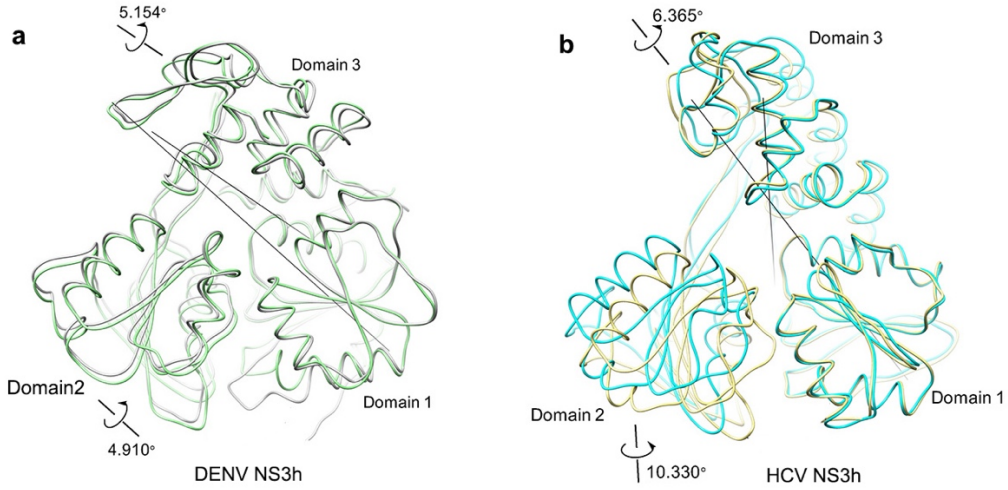
Appendix 11 Comparisons of ZIKV, DENV and HCV NS3h-RNA/DNA complex structures.

Left panel. Overlay NS3h-pRNA$_9$ structures of ZIKV (PDB code 5MFX, this work) is in gold and the ZIKV NS3-ssRNA [PDB code 5GJB (247)], DENV [PDB code 2JLQ and 5XC6 (79, 265)] are in brown and dark green, respectively, and HCV NS3h-DNA/RNA complexes [PDB code 3KQK and 3O8C (109, 240)] are in light blue and green. The conserved aspartate and arginine/lysine, as well as the W501 (in HCV NS3h) that is thought to act as a 'ratchet' for DNA translocation, are shown as sticks.



Appendix 12 Predicted secondary structure of the partially duplexed hairpin RNA.

The secondary structure of the partially duplexed (21-mer) hairpin RNA with 9-mer 3' overhang was predicted by Mfold web server (272). The RNA oligo was dissolved in buffer containing 10 mM HEPES 7.5, 100 mM NaCl and 0.5 mM EDTA. The annealing experiment was performed by heating the sample of 1 mM RNA at 95 ˚C for 5 min and then placing the RNA sample on ice for 15 min.



Appendix 13 Domain movements of DENV and HCV NS3hs upon nucleotide binding.

a. Overlay of DENV NS3h-AMPPNP complex [light green, PDB code 2JLR, (79)] and its apo form [grey, PDB code 2JLQ, (79)]; b. Overlay of HCV NS3h-ADP-AlF_4^- [khaki, PDB code 5E4F, (253)] and its apo form [cyan, PDB code 3O8B, (240)]. Structures were superposed on domain 1.

Appendix 14 List of protein complex structures with MgF_3^- bound.

PDB code	Space group	Protein name	Resolution	References
4C4R	$P2_12_12_1$	β-Phosphoglucomutase (βPGM)	1.10 Å	(211)
2WF5	$P2_12_12_1$	β-Phosphoglucomutase (βPGM)	1.30 Å	(273)
5OLX	$P2_12_12_1$	β-Phosphoglucomutase (βPGM)	1.38 Å	(274)
2WZB	$P2_12_12_1$	Human phosphoglycerate kinase	1.47 Å	(260)
4C4S	$P2_12_12_1$	β-Phosphoglucomutase (βPGM)	1.50 Å	(211)
5JAJ	$P2_12_12_1$	LGP2 (RNA binding protein)	1.50 Å	(275)

3MSX	C2 2 2_1	RhoA GTPase	1.65 Å	unpublished
4NV0	P 1 2_1 1	Cytosolic 5'-nucleotidase IIIB (cN-IIIB)	1.65 Å	(276)
5IRC	$P2_12_12_1$	RhoA GTPase	1.72 Å	(277)
4DLC	P 4_3 2_1 2	Trypanosoma brucei dUTPase	1.76 Å	(278)
1OW3	$P2_12_12_1$	RhoA GTPase	1.80 Å	(228)
3T9E	$P2_12_12_1$	Human diphosphoinositol pentakisphosphate kinase 2 (PPIP5K2)	1.90 Å	(279)
3QHW	P 1 2_1 1	CDK2 Kinase	1.91 Å	(280)
3ZOZ	$P2_12_12_1$	Human Phosphoglycerate Kinase	1.95 Å	unpublished
5OLY	P 1 2_1 1	β-Phosphoglucomutase (βPGM)	2.00 Å	(274)
3QHR	P 1 2_1 1	CDK2 Kinase	2.17 Å	(280)
2IS6	P 1 2_1 1	UvrD helicase	2.20 Å	(251)
5HPY	$P2_12_12_1$	RhoA GTPase	2.40 Å	(281)
5M6X	P 1 2_1 1	RhoA GTPase	2.40 Å	(282)
4BEW	$P2_12_12_1$	Sarcoplasmic/endoplasmic reticulum calcium ATPase 1	2.50 Å	unpublished
5MPM	P 1 2_1 1	Sarcoplasmic reticulum Ca2+-ATPase	3.30 Å	unpublished
4JSV	$P2_12_12_1$	mTOR kinase	3.50 Å	(283)
4BEV	$P2_12_12_1$	Copper efflux ATPase	3.58 Å	unpublished
3JBZ	-	ATM kinase	28.0 Å	(284)

24 entries for the MgF_3^- (PDB ligand code: MGF) available in the PDB database in May 2019 are listed in the table, 19 of the 24 entries have better than 2.50 Å resolution.

Abbreviations

AEBSF	4-(2-Aminoethyl) benzenesulfonyl fluoride
ADP	Adenosine-5'-diphosphate
APS	Ammonium persulphate
ATP	Adenosine-5'-triphosphate
BSA	Bovine serum albumin
DDM	n-Dodecyl-B-D-Maltoside
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded deoxyribonucleic acid
dNTP	Deoxyribonucleotides dATP, dTTP, dCTP and dGTP
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
FT	Flow through
GF	Gel filtration
gp	gene product
GSA	Ground state analogue
HEPES	4-(2-Hydroxyethyl) piperazine-1-ethanesulfonic acid
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB medium	Luria-Bertani medium
MES	2-(N-Morpholino) ethanesulfonic acid
MPD	2-methylpentane-2,4-diol
NMR	Nuclear magnetic resonance
OD600	Optical density at 600 nm
OTG	n-Octyl beta-D-thioglucopyranoside
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol

RNA	Ribonucleic acid
SDS	Sodium dodecyl sulfate
ssRNA	Single-stranded ribonucleic acid
SEC	Size exclusion chromatography
TCEP	Tris (2-carboxyethyl) phosphine
TEMED	N,N,N′,N′-Tetramethylethylenediamine
SOC	Super Optimal broth with Catabolite repression, a nutritionally rich bacterial culture medium
Tris	Tris(hydroxymethyl)aminomethane
TSA	Transition state analogue
T_{onset}	Onset temperature
T_m1	First thermal unfolding transition midpoint
T_m2	Second thermal unfolding transition midpoint
WT	Wild type

References

1. Lodish, H., *et al*. (2000) Viruses: structure, function, and uses. *Molecular Cell Biology*. W. H. Freeman & Co., New York.

2. Hyman, P. and Abedon, S. T. (2012) Smaller fleas: viruses of microorganisms. *Scientifica (Cairo)*, **2012**, 1–23.

3. Weinbauer, M. G. (2004) Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.*, **28**, 127–181.

4. Oksanen, H. M., *et al*. (2010) Bacteriophages: lipid-containing. *Encycl. Life Sci.*, John Wiley & Sons, Ltd, Chichester.

5. Koonin, E. V., *et al*. (2015) Origins and evolution of viruses of eukaryotes: the ultimate modularity. *Virology*, **479**–**480**, 2–25.

6. Iyer, L. M., *et al*. (2004) Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging. *Nucleic Acids Res.*, **32**, 5260–5279.

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

8. Krupovic, M. and Koonin, E. V. (2015) Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution. *Nat. Rev. Microbiol.*, **13**, 105–115.

9. King, A. M. Q., Adams, M. J., Carstens, E. B. and Lefkowitz, E. J. (eds.) (2011) *Virus Taxonomy: ninth report of the international committee on Taxonomy of viruses*. Elsevier Academic Press, London.

10. Abedon, S. T. (2011) Size does matter - distinguishing bacteriophages by genome length (and 'breadth'). *Microbiol. Aust.*, **32**, 95–96.

11. Abedon, S. T. (2009) Phage evolution and Ecology. *Advances in applied Microbiology*. Elsevier Inc.

12. Hulo, C., *et al.* (2011) ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res.*, **39**, 576–582.

13. Grahn, A. M., *et al*. (2006) PRD1: dissecting the genome, structure, and entry. *The bacteriophages*. Oxford University Press, New York.

14. Hong, C., *et al*. (2014) A structural model of the genome packaging process in a membrane-containing double stranded DNA virus. *PLoS Biol.*, **12**, e1002024.

15. Abrescia, N. G., *et al.* (2004) Insights into assembly from structural analysis of bacteriophage PRD1. *Nature*, **432**, 68–74.

16. Rydman, P. S., *et al*. (1999) Bacteriophage PRD1 contains a labile receptor-binding structure at each vertex. *J. Mol. Biol.*, **291**, 575–587.

17. Peralta, B., *et al*. (2013) Mechanism of membranous tunnelling nanotube formation in viral genome delivery. *PLoS Biol.*, **11**, e1001667.

18. Mindich, L., *et al*. (1982) Assembly of bacteriophage PRD1: particle formation with wild-type and mutant viruses. *J. Virol.*, **44**, 1021–1030.

19. Modrow, S., *et al*. (2013) Viruses with single-stranded, positive-sense RNA genomes. *Molecular Virology*. Springer, Berlin, Heidelberg.

20. King, A. M. Q., Adams, M. J., Carstens, E. B. and Lefkowitz, E. J. (eds.) (2011) Family - *Flaviviridae*. *Virus Taxonomy, ninth report of the international committee on Taxonomy of viruses*. Elsevier Academic Press, London.

21. Klema, V. J., *et al*. (2015) *Flaviviral* replication complex: coordination between RNA synthesis and 5'-RNA capping. *Viruses*, **7**, 4640–4656.

22. Neufeldt, C. J., *et al*. (2018) Rewiring cellular networks by members of the *Flaviviridae* family. *Nat. Rev. Microbiol.*, **16**, 125–142.

23. Miller, R. H. and Purcell, R. H. (1990) Hepatitis C virus shares amino acid sequence similarity with *Pestiviruses* and *Flaviviruses* as well as members of two plant virus supergroups. *Proc. Natl. Acad. Sci.*, **87**, 2057–2061.

24. Wikan, N. and Smith, D. R. (2016) Zika virus: history of a newly emerging arbovirus. *Lancet Infect. Dis.*, **16**, e119–e126.

25. (2019) Hepatitis C. World Health Organization, available at: https://www.who.int/en/news-room/fact-sheets/detail/hepatitis-c.

26. Rao, V. B. and Feiss, M. (2015) Mechanisms of DNA packaging by large double-stranded DNA viruses. *Annu. Rev. Virol.*, **2**, 351–378.

27. Chelikani, V., *et al*. (2014) Revisiting the genome packaging in viruses with lessons from the 'Giants'. *Virology*, **466**–**467**, 15–26.

28. Khromykh, A. A., *et al*. (2001) Coupling between replication and packaging of *Flavivirus* RNA: evidence derived from the use of DNA-based full-length cDNA clones of Kunjin virus. *J. Virol.*, **75**, 4633–4640.

29. Saren, A. M., *et al*. (2005) A snapshot of viral evolution from genome analysis of the *Tectiviridae* family. *J. Mol. Biol.*, **350**, 427–440.

30. Henne, A., *et al.* (2004) The genome sequence of the extreme thermophile *Thermus thermophilus*. *Nat. Biotechnol.*, **22**, 547–553.

31. Grahn, A. M., *et al*. (2002) Sequential model of phage PRD1 DNA delivery: active involvement of the viral membrane. *Mol. Microbiol.*, **46**, 1199–1209.

32. Bamford, D., *et al*. (1983) Identification of a protein bound to the termini of bacteriophage PRD1 DNA. *J Virol*, **47**, 311–316.

33. Méndez, J., *et al*. (1997) Protein-primed DNA replication: a transition between two modes of priming by a unique DNA polymerase. *EMBO J.*, **16**, 2519–2527.

34. Ikeda, J. E., *et al*. (1982) Adenoviral protein-primed initiation of DNA chains *in vitro*. *Proc. Natl. Acad. Sci.*, **79**, 2442–2446.

35. Pakula, T. M., *et al.* (1990) Characterisation of a DNA binding protein of bacteriophage

PRD1 involved in DNA replication. *Nucleic Acids Res.*, **18**, 6553–6557.

36. Berjón-Otero, *et al*. (2016) Disclosing early steps of protein-primed genome replication of the gram-positive *Tectivirus* Bam35. *Nucleic Acids Res.*, **44**, 9733–9744.

37. Caldentey, J., *et al*. (1992) *In vitro* replication of bacteriophage PRD1 DNA. Metal activation of protein-primed initiation and DNA elongation. *Nucleic Acids Res.*, **20**, 3971–3976.

38. Bamford, J. K. H., *et al.* (1991) Genome organization of membrane-containing bacteriophage PRD1. *Virology*, **183**, 658–676.

39. Karhu, N. J., *et al*. (2007) Efficient DNA packaging of bacteriophage PRD1 requires the unique vertex protein P6. *J. Virol.*, **81**, 2970–2979.

40. Stromsten, N. J., *et al*. (2003) The unique vertex of bacterial virus PRD1 is connected to the viral internal membrane. *J. Virol.*, **77**, 6314–6321.

41. Strömsten, N. J., *et al*. (2005) *In vitro* DNA packaging of PRD1: a common mechanism for internal-membrane viruses. *J. Mol. Biol.*, **348**, 617–629.

42. Tassaneetrithep, B., *et al.* (2003) DC-SIGN (CD209) mediates Dengue virus infection of human dendritic cells. *J. Exp. Med.*, **197**, 823–829.

43. Mitchell, D. A., *et al*. (2001) A novel mechanism of carbohydrate recognition by the C-type lectins DC-SIGN and DC-SIGNR. Subunit organization and binding to multivalent ligands. *J. Biol. Chem.*, **276**, 28939–28945.

44. Miller, J. L., *et al*. (2008) The mannose receptor mediates Dengue virus infection of macrophages. *PLoS Pathog.*, **4**, 47.

45. Lee, I., *et al.* (2018) Probing molecular insights into Zika virus-host interactions. *Viruses*, **10**, 233.

46. Sirohi, D. and Kuhn, R. J. (2017) Zika virus structure, maturation, and receptors. *J. Infect. Dis.*, **216**, S935–S944.

47. Modis, Y., *et al*. (2004) Structure of the Dengue virus envelope protein after membrane fusion. *Nature*, **427**, 313–319.

48. Wikipedia (2019) *Flavivirus*, available at: https://en.wikipedia.org/wiki/Flavivirus.

49. Rice, C. M., *et al.* (1985) Nucleotide sequence of Yellow fever virus: implications for *Flavivirus* gene expression and evolution. *Science*, **229**, 726–733.

50. Kuno, G., *et al*. (1998) Phylogeny of the genus *Flavivirus*. *J. Virol.*, **72**, 73–83.

51. Simmonds, P., *et al.* (2017) ICTV virus Taxonomy profile: *Flaviviridae*. *J. Gen. Virol.*, **98**, 2914–2915.

52. Mackenzie, J. M., *et al*. (1996) Immunolocalization of the Dengue virus nonstructural glycoprotein NS1 suggests a role in viral RNA replication. *Virology*, **220**, 232–240.

53. Falgout, B., *et al*. (1991) Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of Dengue virus nonstructural proteins. *J. Virol.*, **65**, 2467–2475.

54. Perera, R. and Kuhn, R. J. (2008) Structural proteomics of Dengue virus. *Curr. Opin. Microbiol.*, **11**, 369–377.

55. Patkar, C. G. and Kuhn, R. J. (2008) Yellow fever virus NS3 plays an essential role in virus assembly independent of its known enzymatic functions. *J. Virol.*, **82**, 3342–3352.

56. Zou, J., *et al*. (2015) Characterisation of Dengue virus NS4A and NS4B protein interaction. *J. Virol.*, **89**, 3455–3470.

57. Zhao, B., *et al.* (2017) Structure and function of the Zika virus full-length NS5 protein. *Nat. Commun.*, **8**, 14762.

58. Scaturro, P., *et al*. (2015) Dengue virus non-structural protein 1 modulates infectious particle production via interaction with the structural proteins. *PLoS Pathog.*, **11**, 1–32.

59. Selisko, B., *et al*. (2014) Regulation of *Flavivirus* RNA synthesis and replication. *Curr. Opin. Virol.*, **9**, 74–83.

60. Lindenbach, B. D., *et al*. (2007) *Flaviviridae*: The viruses and their replication. *Fields Virology*. Lippincott-Raven Publishers, Philadelphia.

61. Welsch, S., *et al*. (2009) Composition and three-dimensional architecture of the Dengue virus replication and assembly sites. *Cell Host Microbe*, **5**, 365–375.

62. Yu, I. M., *et al*. (2008) Structure of the immature Dengue virus at low pH primes proteolytic maturation. *Science*, **319**, 1834–1838.

63. Kuhn, R. J., *et al.* (2002) Structure of Dengue virus: implications for *Flavivirus* organization, maturation, and fusion. *Cell*, **108**, 717–725.

64. Mukhopadhyay, S., *et al*. (2005) A structural perspective of the *Flavivirus* life cycle. *Nat. Rev. Microbiol.*, **3**, 13–22.

65. Leipe, D. D., *et al*. (2003) Evolution and classification of P-loop kinases and related proteins. *J. Mol. Biol.*, **333**, 781–815.

66. Iyer, L. M., *et al*. (2004) Evolutionary history and higher order classification of AAA^+ ATPases. *J. Struct. Biol.*, **146**, 11–31.

67. Leipe, D. D., *et al*. (2002) Classification and evolution of P-loop GTPases and related ATPases. *J. Mol. Biol.*, **317**, 41–72.

68. Walker, J. E., *et al*. (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.*, **1**, 945–951.

69. Gorbalenya, A. E. and Koonin, E. V. (1993) Helicases: amino acid sequence comparisons and structure-function relationships. *Curr. Opin. Struct. Biol.*, **3**, 419–429.

70. Guo, P. and Lee, T. J. (2007) Viral nanomotors for packaging of dsDNA and dsRNA. *Mol. Microbiol.*, **64**, 886–903.

71. Hanson, P. I. and Whiteheart, S. W. (2005) AAA^+ proteins: have engine, will work. *Nat. Rev. Mol. Cell Biol.*, **6**, 519–529.

72. Massey, T. H., *et al*. (2006) Double-stranded DNA translocation: structure and mechanism of hexameric FtsK. *Mol. Cell*, **23**, 457–469.

73. Ye, J., *et al*. (2004) RecA-like motor ATPases - lessons from structures. *Biochim. Biophys. Acta*, **1659**, 1–18.

74. Story, R. M. and Steitz, T. A. (1992) Structure of the RecA protein-ADP complex. *Nature*, **355**, 374–376.

75. Caruthers, J. M. and McKay, D. B. (2002) Helicase structure and mechanism. *Curr. Opin. Struct. Biol.*, **12**, 123–133.

76. Thomsen, N. D. and Berger, J. M. (2008) Structural frameworks for considering microbial protein and nucleic-acid dependent motor ATPases. *Mol. Microbiol.*, **69**, 1071–1090.

77. Singleton, M. R., *et al*. (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.*, **76**, 23–50.

78. Velankar, S. S., *et al*. (1999) Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell*, **97**, 75–84.

79. Luo, D., *et al.* (2008) Insights into RNA unwinding and ATP hydrolysis by the *Flavivirus* NS3 protein. *EMBO J.*, **27**, 3209–3219.

80. Hopfner, K. P., *et al.* (2000) Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. *Cell*, **101**, 789–800.

81. Jeruzalmi, D., *et al*. (2001) Crystal structure of the processivity clamp loader gamma (γ) complex of *E. coli* DNA polymerase III. *Cell*, **106**, 429–441.

82. Abrahams, J. P., *et al*. (1994) Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. *Nature*, **307**, 621–628.

83. Enemark, E. J. and Joshua-Tor, L. (2006) Mechanism of DNA translocation in a replicative hexameric helicase. *Nature*, **442**, 270–275.

84. Singleton, M. R., *et al*. (2000) Crystal structure of T7 gene 4 ring indicates a mechanism for sequential hydrolysis of nucleotides. *Cell*, **101**, 589–600.

85. Happonen, L. J., *et al*. (2013) The structure of the NTPase that powers DNA packaging into Sulfolobus turreted icosahedral virus 2. *J. Virol.*, **87**, 8388–8398.

86. Begg, K. J., *et al.* (1995) A new *Escherichia coli* cell-division gene, FtsK. *J. Bacteriol.*, **177**, 6211–6222.

87. Bigot, S., *et al*. (2005) KOPS: DNA motifs that control *E. coli* chromosome segregation by orienting the FtsK translocase. *EMBO J.*, **24**, 3770–3780.

88. Barre, F. X. (2007) FtsK and SpoIIIE: the tale of the conserved tails. *Mol. Microbiol.*, **66**, 1051–1055.

89. Kaimer, C. and Graumann, P. L. (2011) Players between the worlds: multifunctional DNA translocases. *Curr. Opin. Microbiol.*, **14**, 719–725.

90. May, P. F. J., *et al*. (2015) Assembly, translocation, and activation of XerCD-*dif* recombination by FtsK translocase analyzed in real-time by FRET and two color tethered fluorophore motion. *Proc. Natl. Acad. Sci.*, **112**, e5133–e5141.

91. Sivanathan, V., *et al.* (2006) The FtsK γ domain directs oriented DNA translocation by interacting with KOPS. *Nat. Struct. Mol. Biol.*, **13**, 965–972.

92. Tanner, N. K. and Linder, P. (2001) DExD/H box RNA helicases: from generic motors to specific dissociation functions. *Mol. Cell*, **8**, 251–262.

93. Kadaré, G. and Haenni, A. L. (1997) Virus-encoded RNA helicases. *J. Virol.*, **71**, 2583–2590.

94. Koonin, E. V. (1992) A new group of putative RNA helicases. *Trends Biochem. Sci.*, **17**, 495–497.

95. Fairman-Williams, M. E., *et al*. (2010) SF1 and SF2 helicases: family matters. *Curr. Opin. Struct. Biol.*, **20**, 313–324.

96. Linder, P. and Jankowsky, E. (2011) From unwinding to clamping - the DEAD box RNA helicase family. *Nat. Rev. Mol. Cell Biol.*, **12**, 505–516.

97. Sengoku, T., *et al*. (2006) Structural basis for RNA unwinding by the DEAD-Box protein Drosophila Vasa. *Cell*, **125**, 287–300.

98. Bennett, R. J. and Keck, J. L. (2004) Structure and function of RecQ DNA helicases. *Crit. Rev. Biochem. Mol. Biol.*, **39**, 79–97.

99. Fuller-Pace, F. V. (2006) DExD/H box RNA helicases: multifunctional proteins with important roles in transcriptional regulation. *Nucleic Acids Res.*, **34**, 4206–4215.

100. Pyle, A. M. (2008) Translocation and unwinding mechanisms of RNA and DNA helicases. *Annu. Rev. Biophys.*, **37**, 317–336.

101. Jankowsky, E. and Fairman, M. E. (2007) RNA helicases - one fold for many functions. *Curr. Opin. Struct. Biol.*, **17**, 316–324.

102. Byrd, A. K. (2012) Superfamily 2 helicases. *Front. Biosci.*, **17**, 2070.

103. Linder, P. and Fuller-Pace, F. V. (2013) Looking back on the birth of DEAD-box RNA helicases. *Biochim. Biophys. Acta*, **1829**, 750–755.

104. Linder, P., *et al.* (1989) Birth of the DEAD box. *Nature*, **337**, 121–122.

105. Company, M., *et al*. (1991) Requirement of the RNA helicase-like protein Prp22 for release of messenger RNA from spliceosomes. *Nature*, **349**, 487–493.

106. He, Y., *et al*. (2010) Structural basis for the function of DEAH helicases. *EMBO Rep.*, **11**, 180–186.

107. Tauchert, M. J., *et al*. (2017) Structural insights into the mechanism of the DEAH-box RNA helicase Prp43. *eLife*, **6**, e21510.

108. Hamann, F., *et al*. (2019) Structural basis for RNA translocation by DEAH-box ATPases. *Nucleic Acids Res.*, **47**, 4349–4362.

109. Gu, M. and Rice, C. M. (2010) Three conformational snapshots of the Hepatitis C virus NS3 helicase reveal a ratchet translocation mechanism. *Proc. Natl. Acad. Sci.*, **107**, 521–528.

110. Dumont, S., *et al.* (2006) RNA translocation and unwinding mechanism of HCV NS3 helicase and its coordination by ATP. *Nature*, **439**, 105–108.

111. Levin, M. K., *et al*. (2005) A Brownian motor mechanism of translocation and strand separation by Hepatitis C virus helicase. *Nat. Struct. Mol. Biol.*, **12**, 429–435.

112. Myong, S., *et al*. (2007) Spring-loaded mechanism of DNA unwinding by Hepatitis C virus NS3 helicase. *Science*, **317**, 513–516.

113. Korolev, S., *et al.* (1997) Major domain swiveling revealed by the crystal structures of complexes of *E. coli* Rep helicase bound to single-stranded DNA and ADP. *Cell*, **90**, 635–647.

114. Raney, K. D., *et al*. (2010) Hepatitis C virus non-structural protein 3 (HCV NS3): a multifunctional antiviral target. *J. Biol. Chem.*, **285**, 22725–22731.

115. Lin, C. and Kim, J. L. (1999) Structure-based mutagenesis study of Hepatitis C virus NS3 helicase. *J. Virol.*, **73**, 8798–8807.

116. Rhodes, G. (2006) *Crystallography made crystal clear. 3rd ed*. Elsevier Inc.

117. Asherie, N. (2004) Protein crystallisation and phase diagrams. *Methods*, **34**, 266–272.

118. Weber, P. C. (1997) Overview of protein crystallisation methods. *Methods in Enzymology, Macromolecular Crystallography, Part A*. Academic Press.

119. Bergfors, T. (2003) Seeds to crystals. *J. Struct. Biol.*, **142**, 66–76.

120. D'Arcy, A., *et al*. (2014) Microseed matrix screening for optimization in protein crystallisation: what have we learned? *Acta Crystallogr. Sect. F*, **70**, 1117–1126.

121. Bragg, W. L. (1913) The diffraction of short electromagnetic waves by a crystal. *Proc. Cambridge Phil. Soc.*, **17**, 43–57.

122. Authier, A. and Chapuis, G. (eds.) (2017) *A little dictionary of crystallography, 2nd ed*. International Union of Crystallography.

123. Hammond, C. (ed.) (2015) X-ray diffraction: the contributions of Max von Laue, W. H. and W. L. Bragg and P.P. Ewald. *The Basics of Crystallography and Diffraction*. Oxford University Press.

124. Chatterjee, A. K. (2001) X-ray diffraction. *Handbook of Analytical Techniques in Concrete Science and Technology*. William Andrew Inc.

125. Suryanarayana, C. and Norton, M. G. (1998) *X-ray diffraction: a practical approach illustrate.* Springer Science & Business Media.

126. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods in enzymology, Macromolecular Crystallography, Part A*. Academic Press.

127. Leslie, A. G. W. (2006) The integration of macromolecular diffraction data. *Acta Crystallogr. Sect. D*, **62**, 48–57.

128. Pflugrath, J. W. (1999) The finer things in X-ray diffraction data collection. *Acta Crystallogr. Sect.*, **55**, 1718–1725.

129. Powell, H. R., *et al*. (2013) Autoindexing diffraction images with iMosflm. *Acta Crystallogr. Sect. D*, **69**, 1195–1203.

130. Dauter, Z. (1999) Data-collection strategies. *Acta Crystallogr. Sect. D*, **55**, 1703–1717.

131. Dauter, Z. (2005) Efficient use of synchrotron radiation for macromolecular diffraction data collection. *Prog. Biophys. Mol. Biol.*, **89**, 153–172.

132. Diederichs, K. and Karplus, P. A. (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Mol. Biol.*, **4**, 269–275.

133. Weiss, M. S. (2001) Global indicators of X-ray data quality. *J. Appl. Crystallogr.*, **34**, 130–135.

134. Dauter, Z. (1997) Data collection strategy. *Methods in Enzymology, Macromolecular Crystallography, Part A*. Academic Press.

135. Evans, P. R. and Murshudov, G. N. (2013) How good are my data and what is the resolution? *Acta Crystallogr. Sect. D*, **69**, 1204–1214.

136. Karplus, P. A. and Diederichs, K. (2015) Assessing and maximizing data quality in macromolecular crystallography. *Curr. Opin. Struct. Biol.*, **34**, 60–68.

137. Coppens, P. (2006) The structure factor. *International Tables for Crystallography, Volume B: Reciprocal space*. International Union of Crystallography.

138. Taylor, G. L. (2010) Introduction to phasing. *Acta Crystallogr. Sect. D*, **66**, 325–338.

139. Cowtan, K. (2003) Phase problem in X-ray crystallography, and its solution. *Encycl. Life Sci.*, 10.1038/npg.els.0002722.

140. Green, D. W., *et al*. (1954) The structure of haemoglobin - IV. Sign determination by the isomorphous replacement method. *Proc. R. Soc. Lond. A*, **225**, 287.

141. Hendrickson, W. A. (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science*, **254**, 51–58.

142. Dodson, E. (2003) Is it jolly SAD? *Acta Crystallogr. Sect. D*, **59**, 1958–1965.

143. James, R. W. (1958) The optical principles of the diffraction by X-rays London: Bell & Sons.

144. Phillips, J. C. and Hodgson, K. O. (1980) The use of anomalous scattering effects to phase diffraction patterns from macromolecules. *Acta Crystallogr. Sect. A*, **36**, 856–864.

145. Hendrickson, W. A. and Ogata, C. M. (1997) Phase determination from multiwavelength anomalous diffraction measurements. *Methods in Enzymology, Macromolecular Crystallography, Part A*. Academic Press.

146. Cowie, D. B. and Cohen, G. N. (1957) Biosynthesis by *Escherichia coli* of active altered

proteins containing selenium instead of sulfur. *Biochim. Biophys. Acta*, **26**, 252–261.

147. Wood, W. B. (1966) Host specificity of DNA produced by *Escherichia coli*: Bacterial mutations affecting the restriction and modification of DNA. *J. Mol. Biol.*, **16**, 118–133.

148. Leahy, D. J., *et al*. (1992) Structure of a fibronectin type III domain from tenascin phased by MAD analysis of the selenomethionyl protein. *Science*, **258**, 987–991.

149. Rossmann, M. G. (2001) Molecular replacement - historical background. *Acta Crystallogr. Sect. D*, **57**, 1360–1366.

150. Simpkin, A. J., *et al.* (2018) SIMBAD: a sequence-independent molecular replacement pipeline. *Acta Crystallogr. Sect. D*, **74**, 595–605.

151. McCoy, A. J., *et al*. (2017) *Ab initio* solution of macromolecular crystal structures without direct methods. *Proc. Natl. Acad. Sci.*, **114**, 3637–3641.

152. Aslanidis, C. and De-Jong, P. J. (1990) Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res.*, **18**, 6069–6074.

153. Fogg, M. J. and Wilkinson, A. J. (2008) Higher-throughput approaches to crystallisation and crystal structure determination. *Biochem. Soc. Trans.*, **36**, 771–775.

154. Mierzejewska, K., *et al.* (2014) Structural basis of the methylation specificity of R.DpnI. *Nucleic Acids Res.*, **42**, 8745–8754.

155. Martin, A., *et al*. (2005) Rebuilt AAA^{+} motors reveal operating principles for ATP-fuelled machines. *Nature*, **437**, 1115–1120.

156. Novagen (2019) Rosetta™ (DE3) pLysS Competent Cells, available at: https://www.merckmillipore.com/GB/en/product/RosettaDE3pLysS-Competent-Cells-Novagen,EMD_BIO-70956.

157. Novagen (2003) pET system manual. Available at: https://www.novagen.com.

158. Bornhorst, J. A. and Falke, J. J. (2000) Purification of proteins using polyhistidine affinity tags. *Methods Enzym.*, **326**, 245–254.

159. Porath, J. (1992) How to use immobilised metal ion affinity chromatography. *Protein Expr. Purif.*, **3**, 263–281.

160. (2018) Size exclusion chromatography principles and methods. *GE Healthcare Handbooks*.

161. (2016) Ion exchange chromatography: principles and methods. *GE Healthcare Handbooks*.

162. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680–685.

163. Oswald, N. (2016) How SDS-PAGE works. Available at: https://bitesizebio.com/580/how-sds-page-works/.

164. Brunelle, J. L. and Green, R. (2014) One-dimensional SDS-polyacrylamide gel electrophoresis (1D SDS-PAGE). *Methods in Enzymology*. Elsevier Inc.

165. Kabsch, W. (2010) XDS. *Acta Crystallogr. Sect. D*, **66**, 125–132.

166. Terwilliger, T. C., *et al*. (2012) Phenix.mr-rosetta: molecular replacement and model rebuilding with Phenix and Rosetta. *J. Struct. Funct. Genomics*, **13**, 81–90.

167. McCoy, A. J., *et al*. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.

168. Sheldrick, G. M. (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. Sect. D*, **66**, 479–485.

169. Langer, G. G., *et al*. (2013) Visual automated macromolecular model building. *Acta Crystallogr. Sect. D*, **69**, 635–641.

170. Emsley, P., *et al*. (2010) Features and development of Coot. *Acta Crystallogr. Sect. D*, **66**, 486–501.

171. Murshudov, G. N., *et al.* (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D*, **67**, 355–367.

172. Potterton, L., *et al.* (2018) CCP4i2: The new graphical user interface to the CCP4 program suite. *Acta Crystallogr. Sect. D*, **74**, 68–84.

173. McNicholas, S., *et al*. (2011) Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr. Sect. D*, **67**, 386–394.

174. Pettersen, E. F., *et al*. (2004) UCSF Chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.

175. Duy, C. and Fitter, J. (2006) How aggregation and conformational scrambling of unfolded states govern fluorescence emission spectra. *Biophys. J.*, **90**, 3704–3711.

176. Carter, S. G. and Karl, D. W. (1982) Inorganic phosphate assay with malachite green: an improvement and evaluation. *J. Biochem. Biophys. Methods*, **7**, 7–13.

177. Ausubel, F. M., *et al.* (2003) *Current protocols in molecular biology*, John Wiley & Sons, Inc.

178. Volkl, A. (2002) Ultracentrifugation. *Encyclopedia of life sciences*. John Wiley & Sons, Ltd.

179. Gasteiger, E., *et al*. (2005) Protein identification and analysis tools on the ExPASy server. *The Proteomics Protocols Handbook.* Humana Press.

180. Gill, J. J., *et al*. (2018) Characterisation of a novel *Tectivirus* phage Toil and its potential as an agent for biolipid extraction. *Sci. Rep.*, **8**, 1–11.

181. Bamford, D. H., *et al*. (1981) Comparison of the lipid-containing bacteriophages PRD1, PR3, PR4, PR5 and L17. *J. Gen. Virol.*, **57**, 365–373.

182. Happonen, L. J., *et al*. (2010) Familial relationships in hyperthermo- and acidophilic archaeal viruses. *J. Virol.*, **4**, 37–52.

183. Cassetti, M. C., *et al*. (1998) DNA packaging mutant: repression of the Vaccinia virus A32 gene results in non-infectious, DNA-deficient, spherical, enveloped particles. *J*

Virol, **72**, 5769–5780.

184. Sievers, F., *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, **7**, 539–545.

185. Robert, X. and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.*, **42**, W320–W324.

186. Ishida, T. and Kinoshita, K. (2007) PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.*, **35**, 460–464.

187. Papiz, M. Z. and Winter, G. (2017) X-ray crystallography, biomolecular structure determination methods. *Encyclopedia of Spectroscopy and Spectrometry*.

188. Matthews, B. W. (1968) Solvent content of protein crystal. *J. Mol. Biol.*, **33**, 491–497.

189. Weichenberger, C. X. and Rupp, B. (2014) Ten years of probabilistic estimates of biocrystal solvent content: new insights via nonparametric kernel density estimate. *Acta Crystallogr. Sect. D*, **70**, 1579–1588.

190. Kleywegt, G. J. and Jones, T. A. (1997) Model building and refinement practice. *Methods in Enzymology*.

191. Dolinsky, T. J., *et al*. (2007) PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.*, **35**, 522–525.

192. Swanson, J. M. J., *et al*. (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.

193. Baker, N. A., *et al.* (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci.*, **98**, 10037–10041.

194. Wendler, P., *et al*. (2012) Structure and function of the AAA^+ nucleotide binding pocket. *Biochim. Biophys. Acta*, **1823**, 2–14.

195. Cleland, W. W. and Hengge, A. C. (2006) Enzymatic mechanisms of phosphate and sulfate transfer. *Chem. Rev.*, **106**, 3252–3278.

196. Schmidt, H., *et al*. (2012) Insights into dynein motor domain function from a 3.3 Å crystal structure. *Nat. Struct. Mol. Biol.*, **19**, 492–497.

197. Rzechorzek, N. J., *et al*. (2014) Structure of the hexameric HerA ATPase reveals a mechanism of translocation-coupled DNA-end processing in archaea. *Nat. Commun.*, **5**, 5506–5521.

198. Gomis-Rüth, F. X., *et al.* (2001) The bacterial conjugation protein TrwB resembles ring helicases and F1-ATPase. *Nature*, **409**, 637–641.

199. Kabaleeswaran, V., *et al*. (2006) Novel features of the rotary catalytic mechanism revealed in the structure of yeast F1 ATPase. *EMBO J.*, **25**, 5433–5442.

200. Thomsen, N. D. and Berger, J. M. (2009) Running in reverse: the structural basis for translocation polarity in hexameric helicases. *Cell*, **139**, 523–534.

201. Park, T., *et al*. (2019) GalaxyTongDock: symmetric and asymmetric ab initio protein–protein docking web server with improved energy parameters. *J. Comput. Chem.*, **9999**, 1–5.

202. Katchalski-Katzir, E., *et al.* (1992) Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci.*, **89**, 2195–2199.

203. Alexander, C. G., *et al.* (2014) Novel microscale approaches for easy, rapid determination of protein stability in academic and commercial settings. *Biochim. Biophys. Acta*, **1844**, 2241–2250.

204. Amado, E., *et al*. (2019) The FtsK-like motor TraB is a DNA-dependent ATPase that forms higher-order assemblies. *J. Biol. Chem.*, **294**, 5050–5059.

205. Caldentey, J., *et al*. (1994) Gene XV of bacteriophage PRD1 encodes a lytic enzyme with muramidase activity. *Eur. J. Biochem.*, **225**, 341–346.

206. Sinha, S., *et al*. (2018) Modeling bacteria-phage interactions and its implications for phage therapy. *Advances in Applied Microbiology*. Elsevier Inc.

207. Hyman, P. and Abedon, S. T. (2009) Practical methods for determining phage growth parameters. *Bacteriophages: Methods and Protocols, Volume 1: Isolation, Characterisation, and Interactions*. Humana Press, a part of Springer Science+Business Media.

208. Polizzi, K. M., *et al*. (2007) Stability of biocatalysts. *Curr. Opin. Chem. Biol.*, **11**, 220–225.

209. Crozat, E. and Grainge, I. (2010) FtsK DNA translocase: the fast motor that knows where it's going. *ChemBioChem*, **11**, 2232–2243.

210. Liang, L., *et al*. (1994) Localization of Vasa protein to the Drosophila pole plasm is independent of its RNA-binding and helicase activities. *Development*, **120**, 1201–1211.

211. Jin, Y., *et al*. (2014) Fluorophosphonates reveal how a phosphomutase conserves transition state conformation over hexose recognition in its two-step reaction. *Proc. Natl. Acad. Sci.*, **111**, 12384–12389.

212. Jin, Y., *et al*. (2016) ^{19}F NMR and DFT analysis reveal structural and electronic transition state features for RhoA-catalyzed GTP hydrolysis. *Angew. Chemie Int. Ed.*, **55**, 3318–3322.

213. Jin, Y., *et al*. (2017) Metal fluorides as analogues for studies on phosphoryl transfer enzymes. *Angew. Chemie Int. Ed.*, **56**, 4110–4128.

214. Donald, C. L., *et al.* (2016) Full genome sequence and sfRNA interferon antagonist activity of Zika virus from Recife, Brazil. *PLoS Negl. Trop. Dis.*, **10**, e0005048.

215. Stevens, R., *et al*. (1983) The stabilities of various thiol compounds used in protein purifications. *Biochem. Educ.*, **11**, 70.

216. Winter, G., *et al.* (2018) DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr. Sect. D*, **74**, 85–97.

217. Langer, G., *et al*. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.*, **3**, 1171–1179.

218. Potterton, E., *et al*. (2003) A graphical user interface to the CCP4 program suite. *Acta Crystallogr. Sect. D*, **59**, 1131–1137.

219. Lebedev, A. A., *et al.* (2012) JLigand: a graphical tool for the CCP4 template-restraint library. *Acta Crystallogr. Sect. D*, **68**, 431–440.

220. Jin, Y., *et al*. (2017) Metal fluorides: tools for structural and computational analysis of phosphoryl transfer enzymes. *Top. Curr. Chem.*, **375**, 1–31.

221. Griffin, J. L., *et al.* (2012) Near attack conformers dominate phosphoglucomutase complexes where geometry and charge distribution reflect those of substrate. *Proc. Natl. Acad. Sci.*, **109**, 6910–6915.

222. Chabre, M. (1990) Aluminofluoride and beryllofluoride complexes: new phosphate analogs in enzymology. *Trends Biochem. Sci.*, **15**, 6–10.

223. Bock, C. W., *et al*. (1994) Coordination of water to magnesium cations. *Inorg. Chem.*, **33**, 419–427.

224. Smart, O. S., *et al*. (2011) grade, version 1.2.12. Global Phasing Ltd.

225. Kreutz, C. and Micura, R. (2008) Investigations on fluorine-labeled ribonucleic acids by ^{19}F NMR spectroscopy. *Modified Nucleosides: in Biochemistry, Biotechnology and Medicine*. John Wiley & Sons.

226. Gerig, J. T. (1994) Fluorine NMR of proteins. Pergamon Press.

227. Rittinger, K., *et al*. (1997) Structure at 1.65 Å of RhoA and its GTPase-activating protein in complex with a transition-state analogue. *Nature*, **389**, 758–762.

228. Graham, D. L., *et al*. (2002) MgF_3^- as a transition state analog of phosphoryl transfer. *Chem. Biol.*, **9**, 375–381.

229. Lindsley, J. E. (2001) Use of a real-time, coupled assay to measure the ATPase activity of DNA topoisomerase II. *Methods Mol. Biol.*, **95**, 57–64.

230. Kiianitsa, K., *et al*. (2003) NADH-coupled microplate photometric assay for kinetic studies of ATP-hydrolyzing enzymes with low and high specific activities. *Anal. Biochem.*, **321**, 266–271.

231. Johnson, K. A. and Goody, R. S. (2011) The original Michaelis constant: translation of the 1913 Michaelis–Menten paper. *Biochemistry*, **50**, 8264–8269.

232. Wikipedia (2019) Michaelis–Menten kinetics, available at: https://en.wikipedia.org/wiki/Michaelis–Menten_kinetics.

233. Robinson, K. E., *et al*. (2014) Mapping Hfq-RNA interaction surfaces using tryptophan fluorescence quenching. *Nucleic Acids Res.*, **42**, 2736–2749.

234. Gesztelyi, R., *et al*. (2012) The Hill equation and the origin of quantitative pharmacology. *Arch. Hist. Exact Sci.*, **66**, 427–438.

235. James N. Weiss (1997) The Hill equation revisited: uses and misuses. *FASEB J.*, **11**, 835–841.

236. Goutelle, S., *et al.* (2008) The Hill equation: a review of its capabilities in pharmacological modelling. *Fundam. Clin. Pharmacol.*, **22**, 633–648.

237. Cao, X., *et al.* (2016) Molecular mechanism of divalent-metal-induced activation of NS3 helicase and insights into Zika virus inhibitor design. *Nucleic Acids Res.*, **44**, 10505–10514.

238. Jain, R., *et al*. (2016) Structure of the NS3 helicase from Zika virus. *Nat. Struct. Mol. Biol.*, **23**, 752–754.

239. Tian, H., *et al.* (2016) The crystal structure of Zika virus helicase: basis for antiviral drug design. *Protein Cell*, **7**, 450–454.

240. Appleby, T. C., *et al.* (2011) Visualizing ATP-dependent RNA translocation by the NS3 helicase from HCV. *J. Mol. Biol.*, **405**, 1139–1153.

241. Büttner, K., *et al*. (2007) Structural basis for DNA duplex separation by a superfamily 2 helicase. *Nat. Struct. Mol. Biol.*, **14**, 647–652.

242. Wu, J., *et al*. (2005) Structure of the *Flavivirus* helicase: implications for catalytic activity, protein interactions, and proteolytic processing. *J. Virol.*, **79**, 10268–10277.

243. Assenberg, R., *et al.* (2009) Crystal structure of a novel conformational state of the *Flavivirus* NS3 protein: implications for polyprotein processing and viral replication. *J. Virol.*, **83**, 12895–12906.

244. Mastrangelo, E., *et al.* (2007) Crystal structure and activity of Kunjin virus NS3 helicase; protease and helicase domain assembly in the full length NS3 protein. *J. Mol. Biol.*, **372**, 444–455.

245. Yamashita, T., *et al.* (2008) Crystal structure of the catalytic domain of Japanese encephalitis virus NS3 helicase/nucleoside triphosphatase at a resolution of 1.8 Å. *Virology*, **373**, 426–436.

246. Jarmoskaite, I. and Russell, R. (2014) RNA helicase proteins as chaperones and remodelers. *Annu. Rev. Biochem.*, **83**, 697–725.

247. Tian, H., *et al.* (2016) Structural basis of Zika virus helicase in recognizing its substrates. *Protein Cell*, **7**, 562–570.

248. Gebhard, L. G., *et al*. (2014) Monomeric nature of Dengue virus NS3 helicase and thermodynamic analysis of the interaction with single-stranded RNA. *Nucleic Acids Res.*, **42**, 11668–11686.

249. Li, L., *et al*. (2018) Structural view of the helicase reveals that Zika virus uses a conserved mechanism for unwinding RNA. *Acta Crystallogr. Sect. F*, **74**, 205–213.

250. Zhang, X. and Wigley, D. B. (2008) The 'glutamate switch' provides a link between ATPase activity and ligand binding in AAA^+ proteins. *Nat. Struct. Mol. Biol.*, **15**, 1223–1227.

251. Lee, J. Y. and Yang, W. (2006) UvrD helicase unwinds DNA one base pair at a time by a two-part power Stroke. *Cell*, **127**, 1349–1360.

252. Orelle, C., *et al.* (2003) The conserved glutamate residue adjacent to the Walker-B motif is the catalytic base for ATP hydrolysis in the ATP-binding cassette transporter BmrA. *J. Biol. Chem.*, **278**, 47002–47008.

253. Gu, M. and Rice, C. M. (2016) The spring α-helix coordinates multiple modes of HCV (Hepatitis C virus) NS3 helicase action. *J. Biol. Chem.*, **291**, 14499–14509.

254. Baxter, N. J., *et al.* (2006) A trojan horse transition state analogue generated by MgF_3^- formation in an enzyme active site. *Proc. Natl. Acad. Sci.*, **103**, 14732–14737.

255. Jin, Y., *et al*. (2012) Charge-balanced metal fluoride complexes for protein kinase A with adenosine diphosphate and substrate peptide SP20. *Angew. Chemie Int. Ed.*, **51**, 12242–12245.

256. Bock, C. W., *et al*. (1999) Manganese as a replacement for magnesium and zinc: functional comparison of the divalent ions. *J. Am. Chem. Soc.*, **121**, 7360–7372.

257. Baxter, N. J., *et al.* (2008) Anionic charge is prioritized over geometry in aluminum and magnesium fluoride transition state analogs of phosphoryl transfer enzymes. *J. Am. Chem. Soc.*, **139**, 3952–3958.

258. Hansen, P. E., *et al*. (1985) Solvent-induced deuterium isotope effects on ^{19}F chemical shifts of some substituted fluorobenzenes. Formation of inclusion complexes. *J. Magn. Reson.*, **62**, 487–496.

259. Sampath, A., *et al.* (2006) Structure-based mutational analysis of the NS3 helicase from Dengue virus. *J. Virol.*, **80**, 6686–6690.

260. Cliff, M. J., *et al*. (2010) Transition state analogue structures of human phosphoglycerate kinase establish the importance of charge balance in catalysis. *J. Am. Chem. Soc.*, **132**, 6507–6516.

261. Liu, X., *et al*. (2011) Prioritization of charge over geometry in transition state analogues of a dual specificity protein kinase. *J. Am. Chem. Soc.*, **133**, 3989–3994.

262. Buchwald, G., *et al*. (2010) Insights into the recruitment of the NMD machinery from the crystal structure of a core EJC-UPF3b complex. *Proc. Natl. Acad. Sci.*, **107**, 10050–10055.

263. Gebhard, L. G., *et al*. (2012) Novel ATP-independent RNA annealing activity of the Dengue virus NS3 helicase. *PLoS One*, **7**, 27–29.

264. Incicco, J. J., *et al*. (2013) Steady-state NTPase activity of Dengue virus NS3: number of catalytic sites, nucleotide specificity and activation by ssRNA. *PLoS One*, **8**, e58508.

265. Swarbrick, C. M. D., *et al*. (2017) NS3 helicase from Dengue virus specifically recognises viral RNA sequence to ensure optimal replication. *Nucleic Acids Res.*, **45**, 12904–12920.

266. Mackintosh, S. G., *et al*. (2006) Structural and biological identification of residues on the surface of NS3 helicase required for optimal replication of the Hepatitis C virus. *J. Biol. Chem.*, **281**, 3528–3535.

267. Kuo, M. D., *et al.* (1996) Characterisation of the NTPase activity of Japanese encephalitis virus NS3 protein. *J. Gen. Virol.*, **77**, 2077–2084.

268. Lakowicz, J. R. (2006) Principles of fluorescence Spectroscopy, 3rd ed. Springer Science + Business Media, LLC.

269. Silverman, E., *et al*. (2003) DExD/H-box proteins and their partners: helping RNA helicases unwind. *Gene*, **312**, 1–16.

270. Savilahti, H., *et al*. (1989) Bacteriophage PRDl terminal protein: expression of gene VIII in *Escherichia coli* and purification of the functional P8 product. *Gene*, **85**, 45–51.

271. Žiedaite, G., *et al*. (2009) Purified membrane-containing procapsids of bacteriophage PRD1 package the viral genome. *J. Mol. Biol.*, **386**, 637–647.

272. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.

273. Baxter, N. J., *et al.* (2010) Atomic details of near-transition state conformers for enzyme phosphoryl transfer revealed by MgF_3^- rather than by phosphoranes. *Proc. Natl. Acad. Sci.*, **107**, 4555–4560.

274. Ampaw, A., *et al.* (2017) Observing enzyme ternary transition state analogue complexes by ^{19}F NMR spectroscopy. *Chem. Sci.*, **8**, 8427–8434.

275. Uchikawa, E., *et al*. (2016) Structural analysis of dsRNA binding to anti-viral pattern recognition receptors LGP2 and MDA5. *Mol. Cell*, **62**, 586–602.

276. Monecke, T., *et al*. (2014) Crystal structures of the novel cytosolic 5′-nucleotidase IIIB explain its preference for m^7GMP. *PLoS One*, **9**, e90915–e90915.

277. Amin, E., *et al*. (2016) Deciphering the molecular and functional basis of RhoGAP family proteins: a systematic approach toward selective inactivation of Rho family proteins. *J. Biol. Chem.*, **291**, 20353–20371.

278. Hemsworth, G. R., *et al*. (2013) On the catalytic mechanism of dimeric dUTPases. *Biochem. J.*, **456**, 81–88.

279. Wang, H., *et al*. (2011) Structural basis for an inositol pyrophosphate kinase surmounting phosphate crowding. *Nat. Chem. Biol.*, **8**, 111–116.

280. Bao, Z. Q., *et al*. (2011) Briefly bound to activate: transient binding of a second catalytic magnesium activates the structure and dynamics of CDK2 kinase for catalysis. *Structure*, **19**, 675–690.

281. Yi, F., *et al*. (2016) Noncanonical Myo9b-RhoGAP accelerates RhoA GTP hydrolysis by a dual arginine finger mechanism. *J. Mol. Biol.*, **428**, 3043–3057.

282. Jin, Y., *et al*. (2017) Assessing the influence of mutation on GTPase transition states by using X-ray Crystallography, ^{19}F NMR, and DFT approaches. *Angew. Chemie Int. Ed.*, **56**, 9732–9735.

283. Yang, H., *et al*. (2013) mTOR kinase structure, mechanism and regulation. *Nature*, **497**, 217–223.

284. Lau, W. C. Y., *et al.* (2016) Structure of the human dimeric ATM kinase. *Cell Cycle*, **15**, 1117–1124.