CryoEM analysis of the Nipah virus ribonucleoprotein complex

De-Sheng Ker

PhD

University of York

Chemistry

September 2021

Abstract

The Nipah virus (NiV) is a highly pathogenic RNA virus and a member of the *Paramyxoviridae* family. Its genomic RNA is wrapped by the nucleocapsid (N) protein in a helical manner forming a long flexible ribonucleoprotein (RNP) assembly. This assembly serves to protect the viral genome from ribonuclease degradation and the host immune response. The RNP assembly also serves as a functional template for the viral RNA synthesis by the synergistic action of the viral RNA dependent RNA polymerase (RdRp) complex which made up of viral large RNA polymerase (L protein) and the viral cofactor phosphoprotein (P protein). However, the molecular details of this viral RNA synthesis remain elusive. This research aims to better understand the initial step of viral RNA synthesis by elucidating the molecular interaction of the NiV RNP assembly and its viral cofactor P protein. The bacterial expression of the NiV nucleocapsid protein led to the formation of the nucleocapsid-like helical assembly containing bacterial cellular RNA. The NiV nucleocapsid protein-RNA complex was purified and its structure was determined using cryo electron microscopy (cryoEM). The structure reveals the nonspecific binding of the RNA in a "three-bases-in, three-bases-out" conformation by the nucleocapsid protein, as well as the role of the N- and C-terminal segments of the nucleocapsid protein in stabilising the RNP assembly. A stable protein complex between the NiV RNP assembly and the P protein was prepared using bacterial co-expression systems. The cryoEM analysis of the NiV N/P heterocomplex suggests the uncoiling of the RNP helical assembly upon interaction with the P protein. Structural data on the NiV nucleocapsid assembly reported here provide molecular insight into the assembly and structure of the NiV RNP complexes and will guide further work towards understanding the mechanisms of viral RNA synthesis.

List of Contents

List of Table

List of Figure

Acknowledgements

Firstly, I would like to thank my supervisor, Fred Antson, for the opportunity to work on the Nipah virus project as my PhD. I am grateful for his trust and confidence in me for granting me the freedom to decide the direction of my PhD project. This has in turn prepared and trained me for my next stage of scientific career on becoming an independent researcher. I also wish to extend my special thanks to the Tony Wild Fund Scholarship for supporting me by paying the full international student tuition fees on my behalf during my three years PhD study.

I wish to show my appreciation to Sandra Grieve, Huw Jenkins, and Maria Chechik for their insightful comments and suggestions that assist in my experimental designs. Thank you to Sandra for her continuous support and motivation boost, particular during the COVID19 pandemic times. Thank you to Huw for his immense knowledge and troubleshooting ability in solving most of my computer software related problems. Thank you to Maria for her technical support in the lab to ensure all my experiments were running smoothly.

I also wish to show my appreciation to the YSBL and TF technician team. Thank you to Sam Hart and Johan Turkenburg for their assistance in ensuring the cryoEM facility was up and running. Thank you to Simon Grist, Louise Haigh, and Juliet Borgia for their assistance in ensuring the lab has sufficient consumables and reagent buffers for all my experiments. I would like to thank all the past and present members of YSBL, for creating the friendly and supportive community in the department. It has been a great pleasure to be able to work alongside them. Thank you to Oli, Dorothy, Jake, Pooja, Rowan, and Katie for brightening up the lab place with their casual chats and jokes. Thank you to Tim Kirk for organising the weekly YSBL badminton session. Thank you to Clement, Zoe, and Alex for introducing the climbing sport to me. It has been one of my favourite sports now when I want to distract myself away from science.

Lastly, I would like to thank my family members and friends for their boundless support and care throughout my three years of study. Thank you to my parents and brother for keeping me occupied and motivated particularly during the COVID19 pandemic times. Thank you to my friends, Ram, Efendy and Jinn Chuan for always being there for a game session whenever I need a break from science.

Declaration

I declare that this thesis is a presentation of original work, and I am the sole author. This work has not previously been presented for an award at this, or any other, University. The work of Chapter 3 has been published as an article in:

Ker DS, Jenkins HT, Greive SJ, Antson AA (2021) **CryoEM structure of the Nipah virus nucleocapsid assembly**. PLoS Pathog 17(7): e1009740. https://doi.org/10.1371/journal.ppat.1009740

All sources are acknowledged as References.

Chapter 1   Introduction

Viruses constitute the most abundant and diverse biological entity on Earth. They lack the functional organelles for metabolic activity rendering them incapable of self-replicating, thus requiring them to hijack the living hosts' metabolism for their own propagation. They are very small and are not visible under a light microscope, which has a resolution limit of 200 nm. Indeed, the first virus, tobacco mosaic virus, was discovered as an infectious sub-microscopic entity that could pass through the Chamberland filter that has a pore size of $0.1 - 1$ μm[1,2]. The name virus was coined from the Latin word meaning "slimy liquid".

Since the discovery of the tobacco mosaic virus, viruses have been discovered to be infectious to all kingdoms of life. Viruses come in different sizes and shapes, and they are found in a wide variety of habitats. Pithoviruses are a giant DNA virus that infects amoebas and it is currently the largest virus, measuring 1.5 μm in length, making it visible under a light microscope[3]. Extremophile viruses have been isolated from acidic hot springs, where they infect extremophile archaea living in these hot conditions[4]. Foot and mouth disease viruses were the first animal viruses to be discovered and it is a highly infectious agent that infects cloven-hoofed (two-toed) mammals like cattle and pigs[5]. The yellow fever virus is transmitted to humans via mosquito bite, and it was the first human virus to be discovered[6].

Despite being small, viruses are the causative agent for many serious illnesses in humans. Some viruses are highly contagious and often result in a rapid spread of the viruses across regions or the whole world. Notable examples are the 1918 influenza pandemic, AIDS epidemic, Ebola epidemic and most recently, the Covid-19 pandemic, which have resulted in millions of deaths, causing immeasurable suffering. The transmission of viruses to humans can occur through inhalation, ingestion, injection (arthropod bites), sexual contact, or congenitally. During the past century, there has been tremendous progress in controlling the spread of infectious diseases through improved sanitation and safer water supplies. Combined with the development of antiviral treatments and vaccines, these have drastically reduced the mortality rate from the pathogenic viruses. In particular, the use of vaccines had successfully led to the global eradication of the smallpox and rinderpest viruses.

However, viral diseases continue to represent a major threat to public health worldwide. The viruses constant change through mutation or recombination have made the eradication efforts difficult. For example, influenza vaccines require reformulation every year to keep up with antigenic drift and shift of the circulating strains[7]. This is further complicated by the frequent emergence of previously unknown zoonotic viruses, through spill over from their animal hosts to humans. As an example, the Sin Nombre virus, a previously unknown member of the hantavirus genus associated with rodents, caused a 1993 outbreak in US of hantavirus pulmonary syndrome with a 60% case fatality rate[8]. Nipah virus, a previously unknown bat virus, appeared in 1998 in Malaysia with a 40% case fatality rate[9]. SARS virus, a previously unknown bat virus, caused an epidemic in 2002-2003 resulting in more than 700 deaths worldwide[10]. Early diagnosis of the causative agent to identify the viral threat and prompt intervention are extremely important to put the outbreaks under control.

1.1   Virus structure

Development of the transmission electron microscope (TEM), which has a theoretical resolution limit of better than 0.1 nm, has allowed the visualisation of the physical nature of the viruses that was not possible with the light microscope[11,12]. The introduction of the negative staining, where heavy metal salts (typically, uranyl acetate or phosphotungstic acid) were added into a virus suspension on a specimen grid, further improved the quality of observation of the viral particles by clearly distinguishing the viral particles from the background[13]. In many cases, it was possible to derive morphological information such as the symmetry and capsomer arrangements within the capsid, and facilitate virus identification and viral classification based on their morphologies. TEM has also been instrumental in elucidating the causative agent during the outbreak of Ebola virus in Zaire in 1976[14], and during the outbreak of Nipah virus in Malaysia in 1998[9].

A remarkable diversity of virus structures has been observed in negatively stained samples. The majority of the viruses can be divided into two groups, either helical or icosahedral, based on the symmetry of their nucleocapsids. Examples of viruses with helical symmetry are the single-stranded RNA viruses, like tobacco mosaic viruses, influenza, paramyxoviruses, bunyaviruses and filoviruses. These viruses form a long

rod-shaped nucleocapsid, in which multiple nucleoprotein units are arranged as a helix. Example of viruses with icosahedral symmetry are flaviviruses, poliovirus, hepatitis A, B and C, and adenovirus. In the case of icosahedral symmetry, the capsid proteins assemble into an approximately spherical shaped capsules with 12 pentameric and 20 triangular faces. In many viruses, the major capsid protein can form both pentameric and hexameric capsomers, owing to the principle of quasi-equivalence that was put forward by Caspar and Klug[15]. With the number of pentamers within the icosahedral capsid being fixed at 12, the capsid size can be expanded by increasing the number of hexons (T-number). This strategy is often used by viruses for accommodating larger genomes. Some viruses with large genomes, like poxviruses have a more complicated architecture and are asymmetric, they are categorized as complex with capsids being neither icosahedral nor helical.

Visualisation of the molecular details of virus structure using TEM is also possible if the samples are rapidly frozen, and the vitrified specimens are examined by cryo electron microscopy (cryoEM). Structural information obtained by cryoEM can be combined with the data derived from X-ray crystallography. In particular, low-resolution EM structures of large assemblies can be fitted with X-ray structures derived for individual components, resulting in "pseudo-atomic" structure. As an example, such structural information has revealed the location of antigenic sites on the surface of virions, allowing better understanding of the processes for viral attachment and penetration of cells[16]. The knowledge of the structure of viral surface proteins also allows us to better design vaccines and drugs to combat the spread of viruses. For instance, the availability of detailed structural information of the haemagglutinin and neuraminidase of influenza viruses has in turn informed the development of anti-influenza vaccines and anti-influenza viral drugs, respectively[17]. More recently structures of various viral RNA polymerases have been determined and these have enabled better understanding of the molecular mechanism of viral transcription and replication[18].

1.2 Virus genome

The virus genomes can be encoded in either DNA or RNA, and either in the form of a single stranded or double stranded nucleic acid conformations. Depending on the types of viruses, the genome could be in a circular or linear arrangement, monopartite

(all viral genes located within a single molecule of nucleic acid) or multipartite (viral genes are distributed over several molecules of nucleic acid). Upon infection, the viral proteins are synthesised by host translation machinery, and the viral genomes are replicated to produce more progeny. These can be accomplished via various different strategies depending on the nature of the viral genome.

1.2.1. DNA viruses

The genome size of DNA viruses normally ranges from 1 kb to 2 Mb. The presence of the very large viral genome is enabled by the less error prone nature of the DNA polymerase. Currently, the Pandoravirus, a family of viruses that infects the amoeba, has the largest viral genome[19]. At 2 Mb, the genome of Pandoravirus is larger than the genome of some eukaryotic microorganisms, obliterating the line between cells and viruses in terms of genome size and complexity. The majority of the thermophilic viruses also have a DNA genome, possibly owing to its better stability over an RNA genome in harsh environmental conditions, such as a high temperature environment.

Among the animal DNA viruses, the viral genome size usually ranges from 1 kb to 300 kb with the herpesvirus and poxvirus having the largest genome at about 300 kb. They can be either enveloped viruses, which have a lipid membrane that is derived from the host cell, or non-enveloped viruses, which lack a membrane. During infection, DNA viruses express different genes at different stages of the viral replication cycle. For DNA viruses which replicate in the nucleus, RNA synthesis is carried out by the host RNA polymerase II. As for the DNA viruses (Poxviridae and Asfarviridae) that replicate in the cytoplasm, they need to carry their own viral DNA-dependent RNA polymerase to generate mRNA for viral protein synthesis. The early viral genes, that are first transcribed and translated before viral DNA replication, encode viral proteins involved in repressing the cellular nucleic acid and protein synthesis, as well as regulating the expression of the viral genome and enzymes necessary for the replication of the viral nucleic acid. The late viral genes are transcribed and translated after DNA replication, and they are mostly encoding structural proteins involved in the viral assembly.

Viral genome replication of most DNA viruses occurs in the nucleus of the cell, and it can be done by utilizing either the host DNA polymerase or the viral encoded DNA polymerase. However, as DNA polymerase cannot initiate the synthesis of nascent

DNA, but only can extend synthesis of a short primer, this creates additional complications in replicating the primer binding sites. Most DNA viruses overcome this by removing the genome termini by having a circular genome, while others have a linear genome with complementary termini that serve as DNA primers. There are further variations, for example adenoviruses resolve this problem by having a protein primer that is covalently attached to the 5'-end of each DNA strand and used it to initiate the DNA synthesis[20].

While having a DNA genome allows the DNA viruses to utilize the host DNA synthesis machinery to replicate the viral DNA genome in the nucleus, additional limitations exist in that the cells only actively synthesise DNA during the mitosis stage. To circumvent this, some DNA viruses only infect cells at the active mitosis stage, while others produce molecules that stimulate cell division. For instance, large DNA viruses (Herpesviruses and Poxviruses) encode virokines that mimic host cytokines to stimulate cell proliferation[21].

1.2.2. RNA viruses

RNA virus's genome sizes range from 1.7 kb to 32 kb and can be further subcategorized into double stranded RNA viruses, single stranded positive sense RNA viruses and single stranded negative sense RNA viruses. In the positive strand RNA viruses, the RNA genome is of the same sense as the functional mRNA and can be used directly by the host translational machinery for generation of viral proteins (Figure 1.1A). As for the negative strand RNA viruses, the viral genome cannot be used directly for viral protein synthesis and therefore there is a need for the generation of a viral mRNA by the viral RNA dependent RNA polymerase (RdRp) (Figure 1.1B). The negative strand viruses can also be further categorised into two major groups: viruses with non-segmented RNA genome (*Mononegavirales*), which have their genome in a single strand of RNA, and viruses with segmented RNA genomes (*Multinegavirales*) which have their genomes in multiple segments. Due to the absence of RdRp-like polymerase in the host cell and the inability of the cell to utilise the negative sense viral genome directly as mRNA for the synthesis of viral protein, the negative strand and double stranded RNA viruses are required to carry a viral RdRp in the assembled virion.

Aside from RdRp, other proteins such as an RNA-helicase and NTPase, can be found coming together to form a replication complex that functions to produce copies of infectious viral genomes. The number and type of proteins in the replication complex vary from one virus family to another. In some cases, host cell proteins, like the stress granule protein and chaperone heat shock protein 70, are involved in the assembly of the replication complex and facilitate the synthesis of viral genomic RNA[22].

Within the viral genomic RNA, there are untranslated regions (UTR) found normally in the 5' end and the 3' end of the genome, which are highly conserved within the same virus family. The length of these UTR can be as long as several hundred nucleotides (nt), as in the case of picornaviruses and hepaciviruses, or as short as a few base paired hairpin, as in the case of flaviviruses[23]. In most cases, the UTR regions have been shown to play a crucial role in viral transcription and translation. For instance, in nonsegmented negative strand RNA viruses, the 3' end of the genome serves as a promoter for transcription and replication of the viral genome. In picornavirus and hepatitis C virus, the 5' end contains the internal ribosome entry site (IRES) which directs cap-independent internal initiation of protein translation[24]. For most positive strand RNA virus genomes, the 5' and the 3' ends of the genome are also capped and polyadenylated, respectively, to mimic the host mRNA.

Replication of many RNA viruses occurs in specialized intracellular compartments formed during viral infection. These structures serve as a platform which facilitates viral replication and concurrently protects the viral genome from host cellular defence mechanisms. In the case of positive strand RNA viruses, the viral replication sites are often associated with virus-specific membranes. The replication complexes are associated with virus induced membrane structures that are derived from different host membranes. For instance, the replication complexes of members of the flaviviruses, picornavirus, and coronavirus are associated with membranes from the rough endoplasmic reticulum. Nodaviruses are associated with mitochondria membranes, whereas togaviruses are associated with membranes of endocytic origin[25]. In contrast, the viral replication site of negative strand RNA viruses is in the cytoplasm and often localized into a region within the cytoplasm by forming an inclusion body probably via liquid-liquid phase separation[26]. The RNA genome of negative strand RNA viruses is encapsidated in the helical ribonucleoprotein (RNP)

assembly throughout the viral replication cycle, thus requiring the RdRp to unveil the RNP and translocate along the encapsidated RNA genome during replication. Following viral RNA replication, the newly synthesised RNA genome is assembled together with structural proteins into a vesicle which buds out from the host cell.
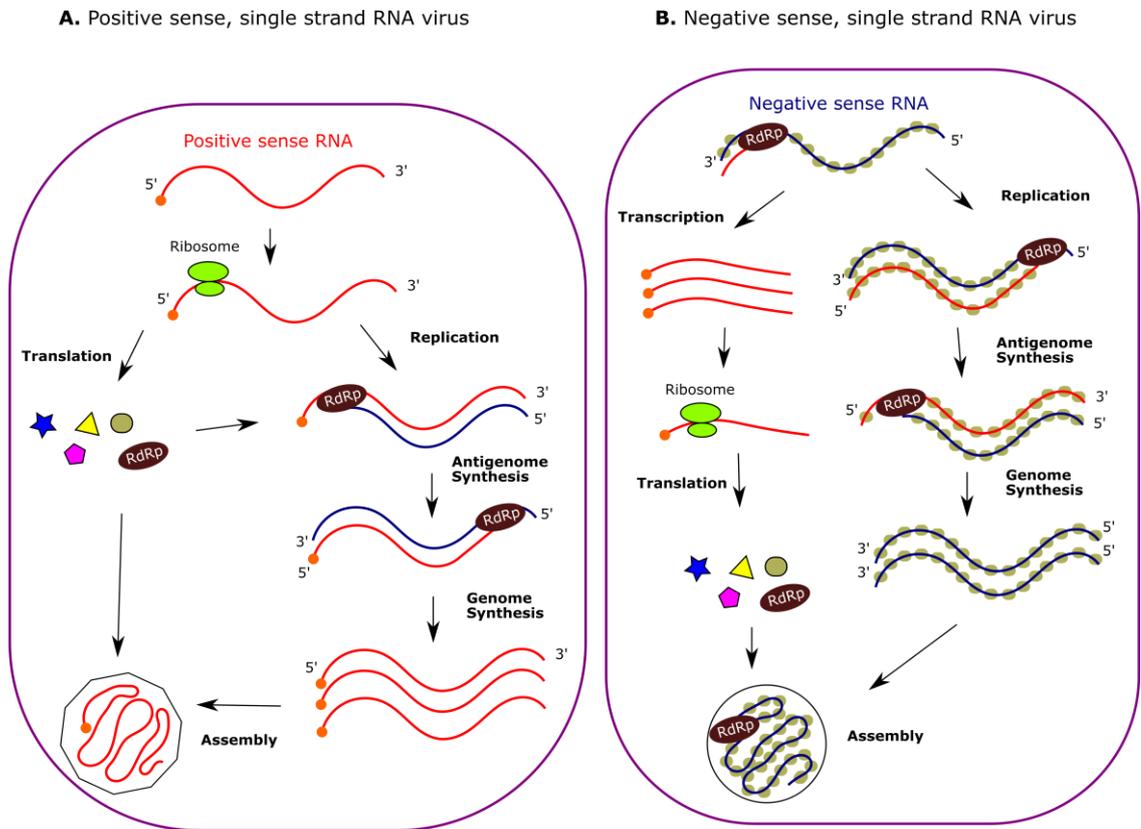
Figure 1.1. Replication strategies of positive and negative sense RNA viruses. (A) The positive sense RNA virus genome has the same sense as the host mRNA and therefore can be used directly by the host cellular translational machinery to produce viral proteins, including viral RNA dependent RNA polymerase (RdRp). During replication, the viral RdRp copies the positive sense RNA genome into complementary negative sense RNA genome (antigenome), which in turn serves as a template for the new positive sense RNA genome synthesis. (B) The negative sense RNA virus genome is encapsidated by the nucleocapsid proteins (dark green) throughout the viral replication cycle. The nucleocapsid bound viral genome must first be transcribed into mRNA in order for it to be used to synthesise more viral proteins through translation. The transcription is performed by the RdRp that is packaged in the virion. During replication, the negative sense RNA genome is transcribed into complementary positive sense RNA genome (antigenome), which is also encapsidated by the nucleocapsid proteins. This encapsidated positive sense RNA genome (antigenome) can then be used as the template for production of the nascent negative sense genomic RNA.

1.3 Diseases caused by RNA Viruses

The RNA viruses encompass many members responsible for significant human and animal diseases. Substantial efforts have been made in elucidating the life cycle, structure, and host interaction of these viruses. This has allowed great progress in designing antiviral and prophylactic approaches in preventing morbidity and mortality. Nevertheless, most viruses continue to prove problematic with respect to human health, further emphasising the need for continuous efforts in research to understand these pathogens. Due to the general lack of proofreading capability of RNA polymerase, the RNA viruses have a higher mutation rate thus enabling these viruses to quickly develop resistance toward developed antiviral drugs. Furthermore, some of the RNA viruses are capable of rapidly adapting to a new host and changing environments, making complete eradication of these viral diseases near impossible. Over the last century, the frequency of spill over of pathogenic viruses from arthropods, birds or nonhuman mammals has increased drastically. This is in part due to urbanization and globalization, which have led to the increasing contact between humans and wildlife, accelerating the risk of zoonotic disease outbreaks in humans. The transfer of the zoonotic viruses to humans can happen directly from the reservoir hosts, or indirectly via intermediate hosts.

1.4 Nipah virus

Nipah virus is a highly pathogenic negative sense RNA virus that belongs to the *Paramyxoviridae* family. The fruit bats of the *Pteropus* genus are recognized as the natural reservoir of the Nipah virus[27]. Nipah virus has a remarkably broad host tropism and is capable of causing systemic disease in cats, dogs, ferrets, hamsters, guinea pigs, and nonhuman primates. It was first recognized in 1998 during an encephalitis outbreak in Malaysia, with a case fatality rate of 40%. During this outbreak, the Nipah virus was transmitted from fruit bats to humans through domestic pigs as the intermediate host. As a measure to control the Nipah virus outbreak, approximately one million pigs were culled, causing devastating economic and social consequences. Following the initial outbreak, separate sporadic outbreaks continue to occur in Bangladesh and India, which involves direct transmission of the Nipah virus from bats to humans. This likely happens via the consumption of contaminated fresh date palm

sap by bat droppings. Human-to-human transmission of Nipah virus also has also been documented in Bangladesh and India cases[28].

The clinical symptoms of the Nipah virus infection are broad, ranging from asymptomatic to very severe cases. In the initial outbreak in Malaysia, approximately 27% of the pig farmer families were asymptomatic[29]. Upon infection with Nipah virus, encephalitic syndrome and respiratory symptoms are the two most common symptoms in humans. The encephalitic syndrome is characterised by headache, pyrexia, and other neurological symptoms; while the respiratory symptoms are characterised by cough, cold and dyspnoea. There was a higher rate of respiratory infections observed in the Bangladesh outbreaks compared to the initial outbreak in Malaysia. This may be partly contributed by the presence of human-to-human transmission of Nipah virus, and subsequently cause the higher transmission rate of the Nipah virus in the Bangladesh outbreaks[30]. Among the recovered patients, there may also be long-term neurological impairment, including personality changes, speech impairment and cognitive dysfunction[31].

1.4.1. Virus structure

Nipah viruses are enveloped pleomorphic viruses with a virus particle size ranging from 120 nm to 500 nm in diameter[9]. Within the virus particle core there is the single stranded viral RNA genome bound tightly by the nucleocapsid protein (N) forming a helical ribonucleoprotein (RNP). The helical RNP is closely associated by the RdRp complex, which consists of the large RNA polymerase (L) and phosphoprotein (P). On the surface of the Nipah virus particle, there are two types of transmembrane glycoproteins, the attachment glycoprotein (G) and the fusion glycoprotein (F). The matrix (M) proteins are found to form a layer of protein shell between the viral lipid membrane and the viral nucleocapsid complex (Figure 1.2 A).

1.4.2. Genome organization

The Nipah virus genome is a nonsegmented single-stranded negative sense RNA molecule that consists of six genes encoded from 3' end to 5' end, namely nucleoprotein (N), phosphoprotein (P), matrix protein (M), fusion glycoprotein (F), attachment glycoprotein (G) and the large RNA polymerase protein (L) (Figure 1.2 B). The genome length of Nipah virus is 18,246 nt, conforming to the "rule of six", in which the genome length is a multiple of six. However, compared to most paramyxoviruses, the Nipah virus genome is about 3000 nt longer, and this is due to the extended P gene and longer tracts of untranslated regions (UTRs) that flank each viral gene with the exception of the L gene. It also contains the classic non-coding regions, found in all paramyxovirus, such as the 3' leader and 5' trailer regions, gene-start, gene-end, RNA-editing and intergenic sequences[32] (Figure 1.2 B). The bipartite promoter for both transcription and replication is found at the 3' end of the genome[33]. In Nipah virus, mutations within the two conserved promoter elements (nt 1-12 and 79-91) result in a loss of minigenome function[34]. The gene-start (GS) and gene-end (GE) sequences are found in the beginning and end of each viral gene, respectively. In the Nipah virus, each GE and GS signal is connected by a conserved GAA nucleotide triplet named the intergenic sequence[32]. As there is only one promoter, expression of all viral genes is dependent on the viral RdRp complex initiating at the promoter and transcribing the preceding genes. However, unlike in most paramyxoviruses, the GE, IG and GS regions of Nipah virus do not seem to be involved in transcriptional attenuation (see section 1.5 below). Instead, it is the UTR after the GS sequence that plays a role in generating the transcriptional gradient in the Nipah virus[35]
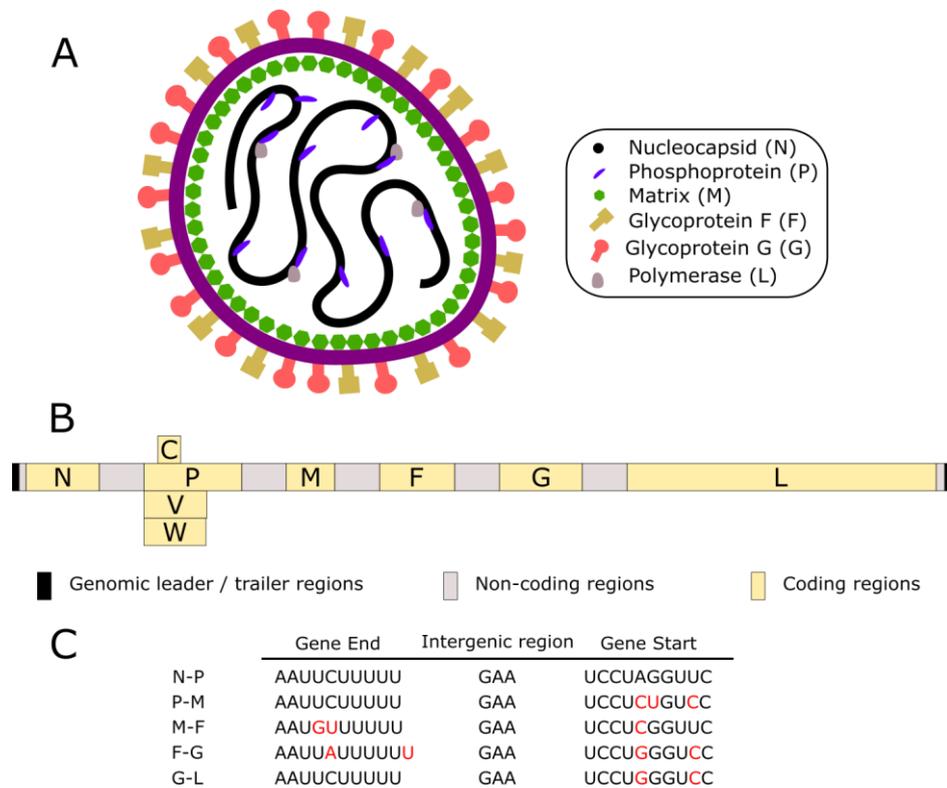
Figure 1.2 Schematic diagram of the (A) Nipah virus structure and (B) genome organization. (C) The sequences for Gene End, Intergenic region, and Gene Start in the Nipah genome are listed. Adapted from Sugai *et. al*. 2007

1.4.3. Virus replication cycle

The Nipah virus life cycle begins when the Nipah virus attachment glycoprotein (G) binds to the host ephrin B2 or ephrin B3 receptors on the cell surface (Figure 1.3). The ephrin B2 receptors are found expressed abundantly on neurons, and endothelial cells, while ephrin B3 receptors are found on cells of the brainstem and heart[36–38]. Upon binding to the ephrin B2/3 by the G protein, the fusion glycoprotein (F) mediates the virus-cell membrane fusion, allowing virus entry with its viral genome, as an RNP complex, being released into the host cytoplasm. During the early stage of viral infection, the viral RdRp complex uses the RNP as a template for transcription to generate viral mRNAs which will be translated into viral proteins by the host translational machinery. All viral proteins are synthesised in the cytoplasm except for the viral glycoproteins. The viral G glycoprotein is synthesised in the endoplasmic reticulum (ER) and matures during passage through the Golgi network to the cell membrane. Similarly, the F glycoprotein is synthesised in the ER and transported to the plasma membrane but as an inactive precursor form termed F0, with further maturation of the F glycoprotein occurring in the endosome[39]. When a high abundance

of viral proteins has been synthesised, the viral RdRp complex, through an unidentified mechanism, stops the transcription process and switches into the genome replication process. During viral replication, the negative sense viral genome is first used as template for generation of the positive sense antigenome, which is in turn used as a template to generate the new negative sense viral genome. The nascent viral genome and all viral structural proteins, including the viral RdRp complex, are assembled in the cell membrane and bud out from the host cell, coordinated by the M proteins[40].
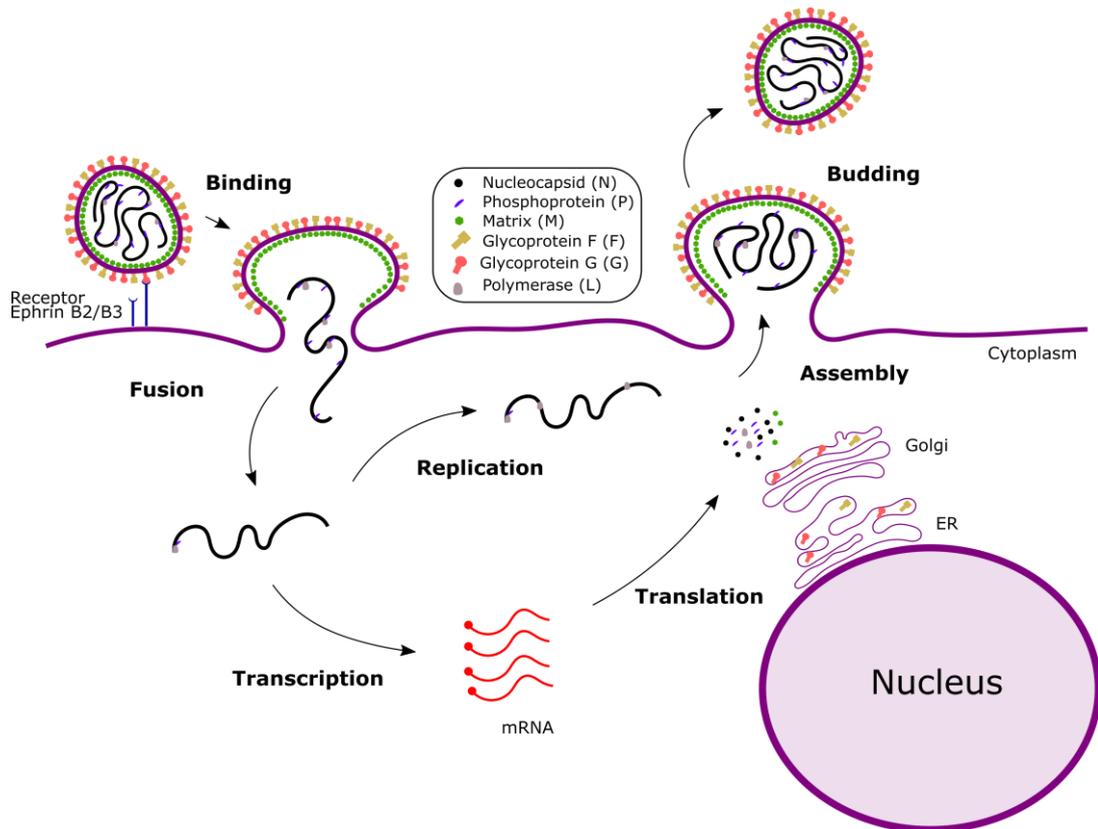
Figure 1.3 Nipah virus replication cycle. Nipah virus attaches to the Ephrin B2/B3 receptor and enters the cell. The viral genome is released and is transcribed into viral mRNAs which are in turn translated into viral proteins via host translational machinery. The new viral genome and proteins are assembled and released from the cell via budding. The Nipah virus RNA genome is encapsidated by the nucleocapsid proteins throughout the viral replication cycle. The thick black lines represent the RNA genome encapsidated by the N protein.

1.4.4. Nipah virus Proteins

1.4.4.1 Attachment Glycoprotein (G)

The Nipah virus attachment glycoprotein is 602 amino acids long and is a type II membrane protein, which is tetramerized through the N-terminal α-helical stalk domain. The G protein binds to the cellular receptor using their C-terminal globular head domains. However, unlike the Paramyxoviral attachment glycoprotein, the Nipah virus attachment glycoprotein lacks both hemagglutinin and neuraminidase activities, instead it attaches to the host cells by recognizing either the ephrin B2 or B3 receptor[36–38]. Ephrin B2 receptors are widespread in the endothelial cells and neurons of many different vertebrate species, which may account for the neurological involvement of Nipah virus, and its associated wide host range. The crystal structure of the ephrinB2 and ephrinB3 in complex with the G protein provide a comprehensive picture of critical residues (Y581 and I588) composing the hydrophobic binding cleft in the G protein which binds with high affinity to the ephrinB2 and ephrinB3[41] (Figure 1.4A). Upon binding of the ephrinB2, a receptor binding activation site in the stalk region of G protein triggers the conformational change of the fusion glycoprotein (F) and promotes viral fusion with the host cell[42].

1.4.4.2 Fusion Glycoprotein (F)

The Nipah virus F protein is a 546 amino acid type I transmembrane protein that forms trimers found on the virion surface. These trimers are responsible for mediating the fusion of virus-cell and cell-cell membranes during the virus life cycle (Figure 1.4B). The F protein is synthesised in the endoplasmic reticulum (ER) and transported to the plasma membrane through the secretory pathway as inactive precursors called F0. The proteins are then endocytosed and cleaved by the endosomal protease cathepsin L to yield the biologically active disulfide-linked F1 and F2 heterodimer[43,44]. After cathepsin cleavage, the F1 and F2 heterodimer is recycled back to the plasma membrane where it is incorporated into nascent virions[45]. During cell entry, after receiving the F-triggering signal from the receptor bound G protein, F protein inserts its hydrophobic fusion peptides into the target host membrane[44], undergoing an irreversible pre-fusion to post-fusion conformational change that drives the merger of the viral and host membranes to form a pore for insertion of the viral genome into the cell.

1.4.4.3 Matrix protein (M)

The Nipah M protein is a 352 amino acid structural protein, with pI of 9.3. The M protein functions by providing rigidity and structure to the virion through its interactions with the cytoplasmic tail of the F protein, and RNP near the plasma membrane of infected cells. In addition, the M protein also introduces sufficient membrane curvature for effective formation of a progeny virion during the viral budding process[46]. M protein deficient recombinant Nipah viruses displayed a severe budding defect, and relied mostly on cell-cell fusion (syncytium) for viral spreading[47]. Nipah M protein also contains nuclear import and export signals and undergoes nuclear transit, as revealed by a live cell confocal microscopy in both human- and bat-derived cell lines[48]. The evidence for the M protein nuclear localization is further supported by the identification of the importin-α, which is a nuclear import receptor, among the protein interactome of the Nipah M protein[49]. Within the nucleus, the M protein residue K258 is monoubiquitinated and this post-translational modification is important for nuclear export, membrane association and viral budding[49,50]. The Nipah M protein also contains two late (L) domain motifs, [62]YMYL[65] and [92]YPLGVG[97], which function by interacting with host cell proteins to facilitate viral budding[40]. Deletion or mutation of either of these late domain motifs of Nipah M protein resulted in abrogation of viral budding[40]. Aside from its viral assembly roles, the Nipah M protein is also capable of disrupting the Interferon-I (IFN-I) signalling to block host antiviral response by interacting with the host E3 ubiquitin ligase TRIM6[51].

1.4.4.4 Nucleocapsid protein (N)

The Nipah N protein is 532 amino acids long, and is responsible for encapsidating the viral genome to form a long herringbone-like structure visible under the electron microscope, the main function of which is to protect the Nipah viral genome from degradation and host immune detection. The Nipah N protein consists of a structured N-terminal domain, termed N-core (Figure 1.4C), which is involved in the viral RNA binding, and a C-terminal intrinsically disordered domain, termed as N-tail[52,53]. It has a known phosphorylation site at residue S451, and mutation of this residue into alanine or glutamate resulted in significant reduction of Nipah virus minigenome activity[54]. Two Nipah virus P protein binding regions were identified: the first binding region corresponds to amino acids 1-54 and it is located within the structured N-core

domain[53] (Figure 1.4C); second binding region corresponds to amino acids 468-532 and it is located within the unstructured N-tail of Nipah N protein[55,56]. In addition to that, at the Nipah N protein C-terminal disordered N-tail, the sequence [523]NDLDFV[528] has been shown to be involved in directing the N protein into the budding Nipah virus like particles (VLPs)[57]. Aside from its structural functions, the Nipah N protein has been shown to be able to hamper the IFN-1 signalling by inhibiting the nuclear transport of both signal transducer and activator of transcription I (STAT1) and 2 (STAT2)[58].

1.4.4.5 Phosphoprotein (P) gene products

The Nipah P protein is a 709 amino acid protein that is required as a cofactor for Nipah virus replication. It forms a crucial tethering component of the Nipah virus polymerase machinery by linking the nucleocapsid-associated viral RNA genome and large RNA polymerase protein together. Structurally, P protein consists of an exceptionally long disordered region at its N terminal region and two structured regions at its C-terminal (Figure 1.4D & Figure 1.5). It has two identified phosphorylation sites, at residues S240 and S472, with no detectable phosphotyrosine or phosphothreonine modifications reported[59]. Similar to the Nipah N protein, there are also two identified N protein binding sites in the P protein. The first site is found at the N-terminus (aa 1-50) and it serves as a chaperone to maintain the monomeric N protein to ensure the proper assembly of the RNP[53]. The second site is located at the C-terminus (aa 655-709), and it plays roles in binding to the C-terminal disordered tail of an assembled N protein. Aside from viral replication, other binding domains in the P protein that are involved in auxiliary functions have also been identified (Figure 1.5). These include the binding domain for STAT1 (aa 114-140)[60], binding site for Polo-like kinase 1 located near residue 130[61], and a nuclear export signal (aa 174-192) that allows cytoplasmic accumulation of the protein[62].
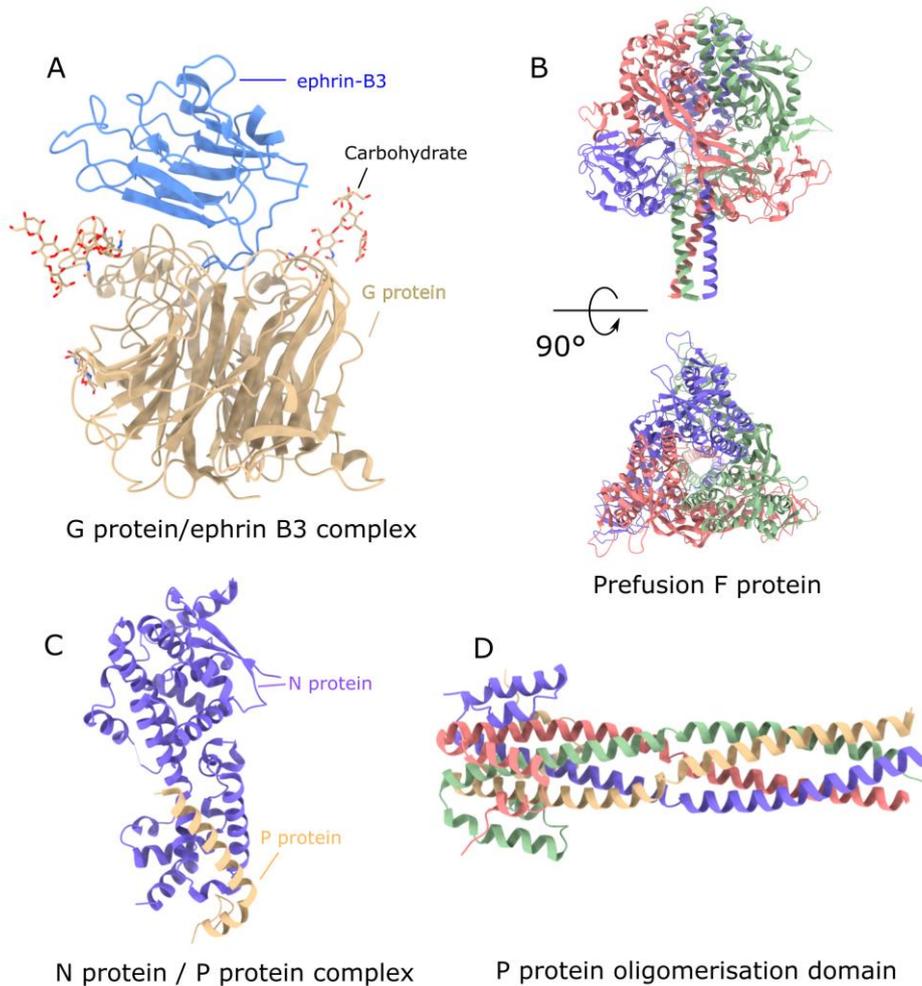
Figure 1.4: Structures of Nipah virus proteins. (A) Crystal structure of the Nipah virus G protein/ephrin-B3 complex (PDB:3D12). The carbohydrate moieties are shown as stick models. (B) Crystal structure of the Nipah virus F protein (PDB:5EVM). (C) Crystal structure of the RNA free Nipah virus N protein/P protein complex (PDB:4CO6). (D) Crystal structure of the Nipah virus P protein oligomerisation domain (PDB:4N5B).
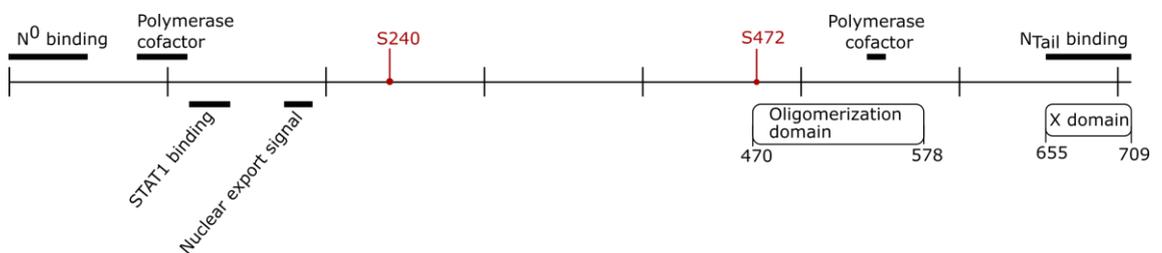
Figure 1.5 Schematic map of the NiV P protein showing the viral / host protein interaction / binding sites. Boxes show the known location of the oligomerization and X domains. Known phosphorylation sites in NiV P protein are designated by red dots. Lines show the location of the known functional / binding regions.

In addition to the P protein, the P gene also encodes at least three non-structural proteins (Figure 1.2 B). The V and W proteins are generated upon addition of either one (V protein) or two (W protein) non-templated guanosines at the editing site of the P mRNA. The C protein is encoded by a second frameshifted open reading frame (ORF) downstream of the translational initiation site for the P protein ORF. All three accessory proteins are key players in the evasion of the interferon (IFN) mediated host-antiviral response. They can antagonise IFN signalling, inhibiting the IFN induction. The V protein does this by binding on STAT1 and STAT2 proteins, in the cytoplasm, forming a tripartite complex that results in inhibition of nuclear translocation of the STAT proteins and efficient phosphorylation of STAT proteins[62]. In contrast, the W protein can be found in the nucleus where it sequesters STAT1 to prevent transcription of IFN-dependent genes. Nipah virus C protein serves as an IFN antagonist by binding to IKKα to antagonise TLR7/9 dependent IFN-α induction[63]. In addition, the Nipah virus C protein was found to enhance the budding of Nipah virus M protein by recruiting the ESCRT (endosomal sorting complex required for transport) factor Tsg101[64].

1.4.4.6 Large Polymerase protein (L)

The Nipah L protein is 2244 amino acids long, and it is the largest and the least abundant viral protein found in an infected cell. The L protein contains all enzymatic activities necessary for RNA synthesis throughout the viral infection, including the phosphodiester bond formation, polyadenylation, mRNA capping, and cap methylation. Therefore, the L protein is involved in initiation, elongation, and termination of the mRNA transcription as well as genome replication. Across all the non-segmented negative strand RNA viruses, the L protein (RNA polymerase) has six conserved amino acid motifs, known as conserved regions (CR) I to VI (Figure 1.6). Recently the cryoEM structure of the L protein from the Parainfluenza virus 5[65], a member of Paramyxovirus family, has been determined. This structure provided information on the domain organization of the Paramyxovirus L protein (Figure 1.6). The RNA dependent RNA polymerase (RdRp) domain, which carries out the core enzymatic activity of the protein, is located at the end N-terminal of the protein. In the Nipah L protein, a ^{831}GDNE^{834} motif is found at the putative RdRp catalytic site, as opposed to the conserved GDNQ motif found in the RdRps of most negative sense RNA viruses. Site directed mutagenesis of this residue E834 to either Gln(Q), Ala(A) or Ile(I)

resulted in an active polymerase suggesting that residue E834 may be involved in maintaining structure rather than catalytic activity[66]. A recent biochemical study based on purified L proteins also investigated the role of other residues within the ^{831}GDNE^{834} motif and found that the D832A/N833A mutant led to an inactive protein[67].

Aside from the catalytic RdRp domain, the L protein also has additional two catalytic domains, a capping domain, and a methyltransferase (MT) domain. The capping domain possesses polyribonucleotide transferase (PRNTase) activity and catalyses the formation of the GTP-capped pre-mRNA. Within the capping domain, two highly conserved motifs, HR and GxxT, are found. A mutation within the ^{1347}HR^{1348} motif, H1347A, of Nipah L protein resulted in ~50% reduction in Nipah L protein activity in an *in vitro* RNA synthesis assay[67]. This suggested there is some level of cooperativity between the RdRp domain and capping domain during the early stage of RNA synthesis. The MT domain possesses both guanine-N7-methyltransferase and nucleoside-2'-O-methyltransferase activities, and it functions in methylate the GTP cap of viral mRNAs, first at the 2'-O and then at the N7 position[68].

The Paramyxovirus L protein also contains two structural domains with no known catalytic activity. These are the connector domain (CD), which can be found linking the RdRp-Cap module with the MT domain, and the C-terminal domain (CTD). Compared to the catalytic domains, both of these structural domains are weakly conserved at the sequence level[18].
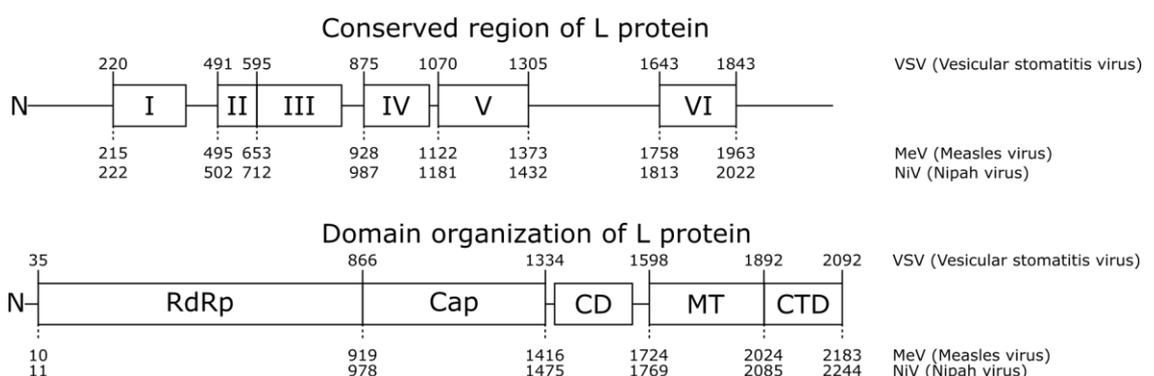
Figure 1.6 Schematic map of negative sense RNA virus L protein showing the location of conserved region and domain organization. The boundary for each region and domain is shown for vesicular stomatitis virus (VSV), Measles virus (MeV) and Nipah virus (NiV). Domains are designated as follows: RdRp, RNA dependent RNA polymerase domain; Cap, Capping domain; CD, Connector domain; MT, Methyltransferase domain; CTD, C-terminal domain.

1.5   Paramyxovirus Viral Replication

The genome of the nonsegmented negative strand RNA viruses is encapsidated by multiple copies of nucleocapsid protein to form a long filament complex with a helical configuration[69]. Depending on the virus family, these nucleocapsid complexes sometimes form higher order structures. For example, in Rhabdovirus, the nucleocapsid complex forms a bullet-like shaped nucleocapsid[70,71]. In the case of Paramyxovirus, the nucleocapsid complex is present as a loosely coiled flexible structure, where each monomer of the N protein binds to 6 nucleotides, and each helical turn made up of 13 nucleocapsid protein monomers[72,73]. The Paramyxoviral replication machinery mainly comprises the viral genome, nucleocapsid (N) protein, large RNA polymerase (L) and its cofactor the phosphoprotein (P). Upon viral infection, the RNP is used by the RdRp complex as a template for transcription and viral replication (Figure 1.7). At the early stage of infection, when the viral N protein in the host cell is scarce, the RdRp complex preferentially performs mRNA transcription, with the RdRp complex then favouring genome replication when the level of N protein is high[74].

1.5.1. Initiation

The recruitment of L protein to the RNP is aided by the cofactor P protein, and it binds at the 3' end of the viral genome. At the 3'end, the viral genome contains the 40-55 nt leader (*le*) region and the bipartite promoter which the RdRp complex recognizes and binds to for the generation of viral mRNAs and antigenomes. The first promoter element is located at the first 12 nt of the genome within the *le* region, and the second promoter element is located around 77-96 nt of the genome within the 5' UTR mRNA region of the first viral gene[75]. As a single helix turn of the Paramyxovirus helical RNP assembly consists of 78 nt (13 subunits of N protein where each N protein has 6 nt), both of the promoter elements are positioned such they are on the same axial face of the nucleocapsid helix at the 3' end of the genome, allowing the concerted recognition by the RdRp complex[76]. However, the bipartite promoters are bound by the N protein, therefore in order for the RdRp complex to have a direct contact with the RNA, the N proteins are expected to release the RNA during the RNA synthesis initiation step. It is likely that the initiation step for both the transcription and replication processes are the same.
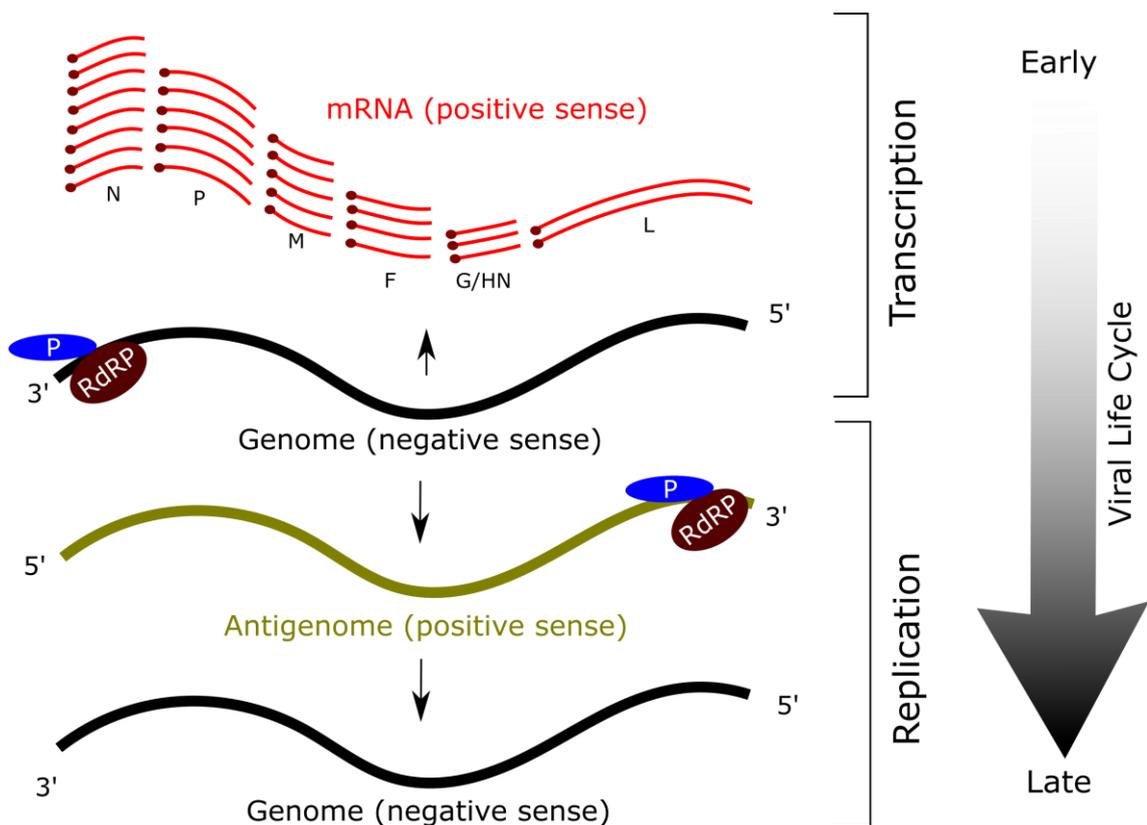
Figure 1.7 Transcription and Replication of Paramyxovirus. The RNP serves as a template for both transcription and replication. At the start of the viral infection, the RdRp complex prioritises transcription. The transcription starts from the 3' end of the RNA genome to produce capped and polyadenylated mRNAs, corresponding to each gene. At the end of each viral gene, the polymerase stops at a Gene End sequence, ignores the intergenic region, and restarts mRNA transcription at the Gene Start sequence. As the polymerase sometimes disengages from the genome during the transcription, and it is required to re-engage transcription from the 3' end of the genome again, a transcription gradient is generated with an attenuation of mRNA copies occurring in the direction of 3'-5'. When abundant viral proteins are available, the RdRp complex switches from transcription to replication. During replication, the RdRp complex ignores the signals from the Gene Start and Gene End sequences to produce a full length antigenome (green) which is also encapsidated. This encapsidated antigenome then serves as a template for the synthesis of additional copies of genomic RNA, which is also encapsidated co-transcriptionally.

1.5.2. mRNA Transcription

The transcription process starts when the RdRp complex, which consists of L and cofactor P proteins, initiates synthesis of the first nucleotide at the 3' end of the viral genome, and began with the synthesis of the short uncapped *le* RNA which is about 40-55 length long[33]. Subsequently, synthesis of a separate strand of mRNA for the N gene starts when the RdRp complex recognises the Gene Start (GS) site, where the RdRp complex initiates transcription at a 5' guanosine which is methylated to create the mRNA cap at the first nucleotide[77]. At the end of the viral gene is the Gene End (GE) site which contains a poly U tract and the stuttering of the RdRp complex on this U tract leads to polyadenylation of the mRNA[78]. It has been proposed that the

generation of the *le* RNA is necessary to remove the UTR region containing the encapsidation signal and prevent encapsidation of viral mRNA[79]. This unencapsidated mature mRNA is then released from the genomic RNP bound RdRp complex. The RdRp complex then transverses the intergenic region of the viral genome to reinitiate transcription at the next GS site of the sequential downstream gene. However, sometimes the RdRp complex can disengage from the genome resulting in a transcription gradient where the first gene is transcribed as mRNA to the greatest extent (Figure 1.7).

During the transcription process, transcriptional slippage can also occur resulting in several different forms of mRNAs. The transcription slippage of the RdRp complex within a short poly G tract within the P protein gene, where additional G residues are inserted, results in a translational frameshift for the production of mRNAs for the V and W proteins[80]. As a result, the V and W protein share the identical N-terminal region with different C-terminal regions. In addition, the mRNAs for one or more C proteins can also be generated from an additional open reading frame (ORF) within the P protein gene region due to leaky scanning of the RdRp complex[81,82].

1.5.3. Genome Replication

For the replication process, it is presumed that the same population of RdRp complex, that performed the transcription, attaches to the same *le* region promoter at the 3' end of the genome. However, in the replication process, the RdRp complex ignores the GE and GS sites and transcribes the entire length of the genome, including the *le*, intergenic, and *tr* regions, to produce the antigenome. The 3' end of the antigenome (positive sense) is the complement of the *tr* region and contains the antigenome promoter. This allows the RdRp complex to bind to the antigenome and use it as a replication template to produce the new viral genome. Furthermore, the antigenome promoter has been shown to be a stronger promoter than the *le* promoter in the *le* region of the genome, allowing the virus to generate a greater number of viral genome copies than for the antigenome[83]. The newly synthesized genome and antigenome are both encapsidated cotranscriptionally by the N protein to form the RNP.

The replication process can also be erroneous and result in RNA viral products that form defective interfering (DI) virus particle bound by N protein and assembled into

viral particles[84,85]. A copyback DI particle is formed during replication when the RdRp complex disengages from the template genome and re-forms a replication complex containing the nascent RNA strand on an RNA template of a different polarity (antigenome), forming a truncated genome with complementary ends. On the other hand, a deletion DI particle is formed when the polymerase disengages the template RNA genome and re-joins the same template elsewhere to form a shortened genome with standard *le* and *tr* ends[84,86]. Interestingly, knockout of the accessory C protein in the paramyxoviruses has been shown to result in a high frequency of copyback DI particles[87].

1.5.4. Regulation of the RdRp polymerase activity

In Paramyxovirus, both the transcription and replication processes are likely to be performed by the same RdRp complex. However, the mechanism in regulating the transition between transcription and genome replication by the same RdRp complex is poorly understood. Early studies have shown that there are a few time points at which the paramyxovirus RdRp complex activity is regulated. Firstly, the RdRp complex is regulated between the transcription and replication processes, and this transition from transcription to replication is dependent on the concentration of N protein. The current model suggested that, when the level of N protein is high enough to concurrently encapsidate the *le* RNA from RdRp complex, this stabilises the RNP complex and commits the RdRp complex into replicase mode to ignore all the *cis* acting elements to synthesise a full length antigenome[33,88]. Furthermore, all mRNA editing activity, including the mRNA transcription slipping of the RdRp complex needs to be regulated and suppressed during the replication process. Secondly, the RdRp complex is regulated between synthesis of antigenome and genome RNAs, to ensure accumulation of a high proportion of genomic RNA for virion assembly. Aside from having a stronger antigenome promoter, as described above, this transition is also likely to be regulated by a *trans* acting protein. For instance in Sendai virus, the C protein has been shown to modulate viral RNA synthesis by interacting with the RdRp complex and manipulating the RdRp complex behaviour with the promoter[89–91].

1.5.5. Formation and Dynamics of the Replication complex

The 3D structure of nucleocapsid (N) protein-RNA complex and a RdRp complex (L-P complex) for several representative paramyxoviruses are available[65,72,73]. These N-RNA structures show a conservation of the six RNA bases bound per N protein monomer, and a helical turn made up of 13 N protein monomers[72,73]. The recent cryoEM structure of the RdRp complex of Parainfluenza virus 5 (PIV5) has provided new structural insight for the molecular interaction between the P protein and the L protein[65]. Two regions of P protein were identified to interact with the L protein: (i) A 25 residue stretch at the C-terminal end of P protein oligomerization domain and its extension and (ii) the P protein X domain helices.

Despite the availability of structural details for the N-RNA complex and RdRp complex, the mechanistic insight on the interaction between the RNP and RdRp complex remains limited. These interactions are highly dynamic as the encapsidated genome needs to be de-encapsidated locally, to allow the RdRp complex to access and transcribe the RNA template, and re-encapsidation after the RdRp complex has transcribed it. Furthermore, as there is no additional viral nucleocapsid present in the host cell before the transcription and translation, the nucleocapsid protein, that was released during transcription, needs to be retained locally in order for it to re-encapsidate the RNA after the RdRp complex has transcribed it.

The paramyxoviral nucleocapsid protein in the helical RNP contains an unstructured N-tail at its C-terminal and it is likely that these disordered N-tails play a role in regulating this dynamic interaction that occurs during the transcription and replication processes[92,93]. Within the nucleocapsid N-tail, there is the molecular recognition element (MoRE) domain that has been shown to interact with the P protein X domain (XD), which in turn interacts with the L protein. High resolution structures reveal that the P protein XD is made up of three helices ($\alpha1$, $\alpha2$, and $\alpha3$) forming a triangular prism. The interface formed by helices $\alpha2$ and $\alpha3$ interacts with the N protein MoRE[94], while the interface formed by helices $\alpha1$ and $\alpha3$ interact with the L protein RdRp domain[65] (Figure 1.8). In the current RdRp complex procession model, it was proposed that the RdRp complex binds to the RNP and cartwheels along the helical RNP through repeated association and dissociation of the P protein XD with the N protein N-tail[76,94]. Interestingly, despite the ability of the P protein XD to bind to the N protein and L

protein from two different sites, the interaction has been shown to be mutually exclusive, likely due to the steric hindrance of the bulky protein complexes of N and L protein preventing the P protein X domain from binding to both of the proteins at the same time[95].
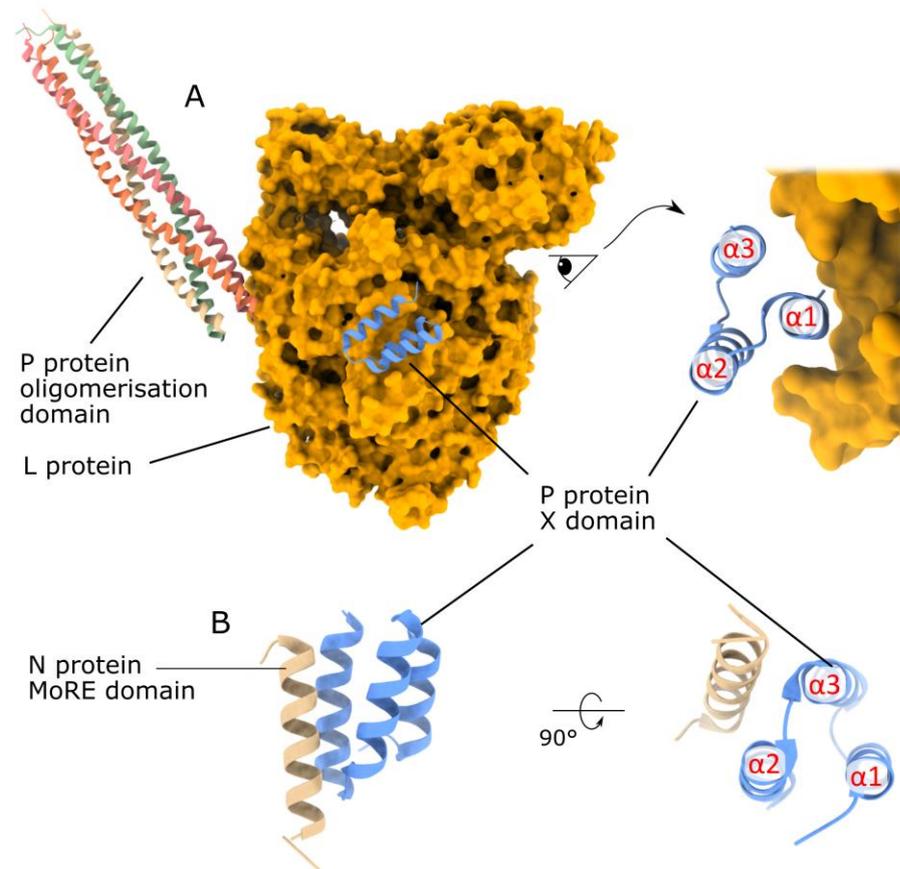
Figure 1.8: Interaction of the P protein X domain with the L and N protein. (A) The cryoEM structure of the L-P protein complex (PDB:6V85). The side view of the protein complex showing the α1 and α3 helices of the P protein X domain interacting with the L protein. (B) The crystal structure of the N protein and P protein complex (PDB:1T6O). The α2 and α3 helices of the P protein X domain interact with the N protein MoRE domain.

Surprisingly, the minireplicon activity assay showed that the N protein MoRE domain is not necessary for initial loading of the RdRp complexes onto the RNP template, but it is necessary to prevent premature termination of the advancing RdRp complex during transcription replication process[96,97]. It is currently still unclear how the RdRp complex initially binds to the RNP template in the absence of MoRE domain. One possibility is the presence of a direct / indirect association between RNP and the RdRp complex, aided by host cell cofactor(s). For example, tubulin has been shown to act as a positive transcription factor to stimulate the *in vitro* synthesis of leader RNA and N protein

mRNA and addition of the anti-tubulin antibody inhibits these *in vitro* transcription[98,99]. This observation suggested that tubulin may play a crucial role in the initiation of transcription at the 3' end of the paramyxovirus genome.

1.6   Cryo Electron microscopy

As described in section 1.1 above, the electron microscope has been an essential instrument in the virology field. While the use of negative staining has allowed the discovery and identification of various new viruses, the staining and drying processes tend to introduce artifacts such as flattening and limit the resolution to ~20 Å due to the stain granularity[100]. In order to determine the structures at high resolution, macromolecules needed to be preserved in a native hydrated state, and this can be achieved using cryo electron microscope (cryoEM) – a microscope equipped with a cryo-stage that allows preservation of the sample in a frozen state during imaging. For cryoEM, the sample is applied onto an EM grid which contains multiple tiny holes. The excess sample is then blotted using filter paper to form a thin layer that is rapidly frozen using liquid ethane (~180°C) to form a layer of vitrified water[101]. Multiple images, each containing hundreds of vitrified particles can then be recorded and used to reconstruct a 3D structure, provided that there is a good distribution of particle orientations in the grid.

More importantly, cryoEM has enabled structural insight into challenging large and flexible macromolecular complexes[102]. Several technical developments on the new direct electron detector technology[103] and sample preparation methods[104,105] have led to improved cryoEM image quality. Together with the new developments of cryoEM imaging processing algorithms, cryoEM offers the possibility of studying macromolecules with significant conformational flexibility and heterogeneity. The sorting of the vitrified particles into more homogenous subsets has allowed the determination of multiple discrete conformational states within a heterogeneous sample[106]. More recently, analysis of the continuous flexibility and heterogeneity has also become possible through the implementation of neural networks into the image processing software[107,108].

1.7   Research Objectives

Understanding molecular interactions of the RNP with the cofactor P protein and the L protein is an area of active ongoing research in several laboratories. Despite the general understanding of the overall pattern of regulation of paramyxovirus transcription and replication, little is known about molecular interactions between individual protein components. In particular, the molecular details pertaining to the recruitment and attachment of the RdRp complex onto the helical RNP are limited. Such lack of detailed information is largely owing to technical challenges associated with structure determination of large and flexible assemblies, such as the viral replication complexes.

In this research project, the Nipah virus has been chosen as the model pathogen for structural investigation of the viral replication complex. Despite the available three dimensional structures for the Nipah virus N protein monomer[53] and the P protein oligomerization domain[109], no structural information was available for the N-protein complex with RNA and it remained unknown how exactly the RNA is bound by the N protein, as well as how the N-RNA complex interacts with the P protein and other viral proteins. Due to the conformational flexibility of this viral replication complex, cryoEM technique was exploited for determination of the structure of this complex. The Nipah virus has been listed as one of the pathogens with great epidemic threats needing urgent research and development action by the World Health Organization (WHO). Currently, there is no approved treatment and vaccine available for Nipah virus disease. The availability of the structural information on the viral replication complex would enable a better understanding on the molecular mechanisms associated with viral RNA synthesis and in turn, open the door for accelerating the development of antiviral drugs for this dangerous pathogen.

The aims of this research project were:

1. To prepare and characterise the Nipah virus nucleocapsid-RNA complex
2. To determine cryoEM structure for the Nipah virus nucleocapsid-RNA complex
3. To obtain a stable of the Nipah virus nucleocapsid-like particles in complex with the phosphoprotein.
4. To characterise interactions of the Nipah virus nucleocapsid-like particles with the phosphoprotein using EM.

Chapter 2   : Materials and Methods

2.1   Cloning

The NiV N gene and P gene were gifted by Prof. Wen Siang Tan at the Universiti Putra Malaysia, Malaysia. The genes were PCR amplified using primers listed in Table 2.1. The successful gene amplification was confirmed by subjection of 1 µL of the PCR products to 1% w/v agarose gel electrophoresis. Upon confirmation, the remaining PCR products were purified using NucleoSpin Gel and PCR clean-up kit (Macherey-Narel) following the manufacturer's guidelines. The linearised expression vectors were obtained by PCR using primers listed in Table 2.2. After PCR reaction, the linearised expression vectors were loaded into 1% w/v agarose gel electrophoresis. The DNA band corresponding to the linearised expression vector was cut and purified using NucleoSpin Gel and PCR clean-up kit (Macherey-Narel) following the manufacturer's guidelines. The DNA concentration of the purified PCR products and linearised expression vector were determined using a NanoDrop UV Spectrophotometer (Thermo Scientific).

The purified PCR products were cloned into their respective linearised expression vector using In-fusion cloning kit (Takara Bio). For PCR products which have DNA length shorter than its expression vector, a 3:1 molar ratio of Insert:Vector was used; For PCR products which have DNA length longer than its expression vector, a 1:1 molar ratio of Insert:Vector was used. The In-fusion reactions were incubated at 37°C for 15 min followed by incubation at 50°C for 20 min. After the incubation, the In-fusion reactions were chilled on ice and transformed into chemically competent Stellar cells (Takara Bio) by a 55 sec heat shock at 42°C. The transformants were selected on Lysogeny broth (LB) agar plates containing antibiotic appropriate for the expression vector. After overnight incubation at 37°C, the individual colonies from each LB agar plate were isolated and positive colonies were identified by colony PCR. Recombinant plasmid was isolated from positive transformants using a GeneJET Plasmid Miniprep kit (Thermo Fisher). The constructs were verified by Sanger DNA sequencing (Eurofins Genomics).

Table 2.1 List of primers for Nipah virus gene amplification. The start and stop codons are marked by red colour in the primer sequences.

Protein name (Plasmid Name)	Vector	Forward Primer (5' – 3') Reverse Primer (5' – 3')
His-NiV N (**pYM663***)	YSBL 3C Lic+	CCAGGGACCAGCAATGAGTGATATCTTTGAAGAG GAGGAGAAGGCGCGTTACACATCAGCTCTGACGAA
His-NiV P (**pYM665***)	YSBL 3C Lic+	CCAGGGACCAGCAATGGATAAATTGGAACTAGTC GAGGAGAAGGCGCGTTAAATGTTACCGTCAATGAT
His-NiV P & NiV N (**pYM705***)	pETDUET	<u>MCS I</u> CACCATCATCACCACATGGATAAATTGGAACTAGTC TGCGGCCGCAAGCTTTTAAATGTTACCGTCAATGAT <u>MCS II</u> GAAGGAGATATACATATGAGTGATATCTTTGAAGAGGC TTTACCAGACTCGAGTCACACATCAGCTCTGACGAA
His-NiV N & NiV P (**pYM727***)	pETDUET	<u>MCS I</u> CACCATCATCACCACATGAGTGATATCTTTGAAGAGGC TGCGGCCGCAAGCTTTTACACATCAGCTCTGACGAA <u>MCS II</u> GAAGGAGATATACATATGGATAAATTGGAACTAGT TTTACCAGACTCGAGTTAAATGTTACCGTCAATGAT

*This is the accession number used in the Antson's laboratory plasmid archive.

Table 2.2 List of primers for linearising vector

Vector	Forward Primer (5' – 3') Reverse Primer (5' – 3')	Antibiotic resistance	Affinity tag / Protein fusion
YSBL 3C Lic+	CGCGCCTTCTCCTCACATATGGCTAGC TTGCTGGTCCCTGGAACAGAACTTCC	Kanamycin	3C protease cleavage N-terminal His tag
pETDUET	<u>MSC I</u> AAGCTTGCGGCCGCATAATGCTTAAG GTGGTGATGATGGTGATGGCTGCTGCC <u>MSC II</u> CTCGAGTCTGGTAAAGAAACCGC CATATGTATATCTCCTTCTTATAC	Ampicillin	Non-cleavage N-terminal His tag for gene cloned in MSC I. Gene cloned in MSC II is tagless

Table 2.3 Reaction component for one Infusion reaction

Reaction Component	Volume	Concentration
Purified PCR products	1 µL	10-50 ng
Linearized vector	1 µL	10-50 ng
5x In-Fusion HD Enzyme Premix	1 µL	-
Deionized water	2 µL	-

2.2　　Protein expression

The expression vector was transformed into *E. coli* BL21 Gold (DE3) Rosetta pLysS using heat shock transformation. A single colony was used to inoculate 5 mL of LB media supplemented with appropriate antibiotic depending on the expression vector in a sterile 50 mL falcon tube and was grown overnight at 37°C with shaking (180 rpm). For *E. coli* BL21 Gold (DE3) Rosetta pLysS host strain, chloramphenicol (30 µg/mL) was also added to maintain the pLysS plasmid carrying the T7 lysozyme gene, in addition to the antibiotic required to maintain the expression vector.

For small scale expression check, the overnight culture was inoculated into 10 mL of LB media supplemented with appropriate antibiotics in a sterile 50 mL falcon tube and the bacteria culture was grown at 37°C with shaking (180 rpm). Upon reaching OD_{600} value of ~0.6, the bacteria culture was induced using 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and was further cultured at 37°C or 16°C for 4 hours and overnight, respectively. 1 mL of bacterial culture was harvested by centrifugation for 1 min at 13,000 x g at room temperature. The cell pellets were resuspended in 200 µL of lysis buffer and lysed by sonication using a SONOPULS ultrasonic homogeniser (BANDELIN) with the setting of 50% amplitude and 30% pulse cycle for 10 sec. 20 µL of the lysate was taken as total protein and the remaining cell lysate was subjected to centrifugation for 2 min at 13,000 x g at room temperature. The pellet and supernatant corresponding to the insoluble and soluble proteins were analysed using 12% SDS PAGE, along with the total protein.

For large scale expression, 2 L of LB media was used instead. Upon reaching OD_{600} value of ~0.6, the bacteria culture was induced by addition of IPTG (1 mM) and further cultured at 37°C for 4 hours. For the 16°C overnight expression, the bacteria culture was chilled using an ice water bath for an hour before induction by addition of IPTG (1 mM). After shaking (180 rpm) incubation, the culture was harvested using centrifugation (5500 x g, 4°C, 20 min) and stored frozen at -80°C until further use.

2.3 Protein purification

2.3.1. Purification of the NiV N protein

5 g of cell pellets were resuspended in 50 mL of lysis buffer containing 20 mM Tris-HCl, pH 8.0, 1 M NaCl, 1 M Urea, 50 mM Imidazole, and 10% (v/v) glycerol (buffer composition adapted from Alayyoubi et al[72]). Cells were lysed by sonication using a SONOPULS ultrasonic homogeniser (BANDELIN) with the setting of 50% amplitude and 50% pulse cycle for 10 min. The lysate was clarified by centrifugation at 25,000 g for 30 min at 4°C. The supernatant was then applied to a HisTrap FF 5 mL column (GE Healthcare) which had been equilibrated with 5 column volumes (CV) of binding buffer containing 20 mM Tris-HCl, pH 8.0, 1 M NaCl, 50 mM Imidazole, and 10% (v/v) glycerol. The column was washed with 10 CV of binding buffer containing 50 mM Imidazole, followed by 6 CV of binding buffer containing 100 mM Imidazole. The protein was eluted using a 20 CV linear gradient of imidazole from 100 mM to 500 mM. Eluted proteins were concentrated using a Vivaspin 20 (100 kDa MWCO) ultrafiltration device and further purified by size exclusion chromatography (SEC) (Superose 6, GE Healthcare) in buffer containing 20 mM Tris pH 8.0, 500 mM NaCl.

For the cleavage of the N-terminal His tag, the His tagged 3C protease and DTT (2mM final concentration) was added to the HisTrap purified N protein. The N protein was dialysed overnight at 4°C against the dialysis buffer (20mM Tris-HCL pH 8.0, 1 M NaCl). After overnight dialysis, the proteins were re-applied to the HisTrap HP 5 mL column to remove the uncleaved N protein and 3C protease. The collected flow-through proteins were concentrated and further purified by SEC using a Superose 6 Increase 10/300 column equilibrated with a buffer containing 20 mM Tris pH 8.0, 500 mM NaCl.

2.3.2. Purification of the NiV P protein

5 g of cell pellets were resuspended in 50 mL of HisTrap buffer supplemented with 1 mM AEBSF, 0.5 µg/ml leupeptin, and 0.7 µg/ml pepstatin. The cell suspension was then subjected to sonication and purification as in section **2.3.1** above using buffer listed in Table 2.4.

Table 2.4 NiV P protein purification buffer

Condition	HisTrap buffer	SEC buffer
1 M NaCl	25 mM Tris pH 8 1 M NaCl 10% Glycerol 20 mM Imidazole 2 mM BME	25 mM Tris pH 8 1 M NaCl 2 mM BME
500 mM NaCl	25 mM Tris pH 8 500 mM NaCl 10% Glycerol 20 mM Imidazole 2 mM BME	25 mM Tris pH 8 500 mM NaCl 2 mM BME

To remove the bound nucleic acid, the HisTrap purified P protein was diluted to a salt concentration of ~100 mM NaCl by addition of 25mM Tris pH 8 buffer. The diluted sample was injected into the HEPARIN HP 5mL column equilibrated with 5 CV of a HEPARIN buffer containing 25 mM Tris pH 8, 100 mM NaCl, 2 mM BME. The column was further washed with HEPARIN buffer for 8 CV. The protein was eluted using a linear gradient from 100 mM to 1000 mM NaCl over 20 CV. Eluted proteins were concentrated using a Vivaspin 20 (100 kDa MWCO) concentrator and further purified by SEC (Superose 6 Increase 10/300, GE Healthcare) in SEC buffer containing 25 mM Tris pH 8, 250 mM NaCl, 2 mM BME.

For the cleavage of the N-terminal His tag, the His tagged 3C protease and DTT (2mM final concentration) were added to the HisTrap purified P protein. The P protein was dialysed overnight at 4°C against the dialysis buffer (20mM Tris-HCL pH 8.0, 150 mM NaCl). After overnight dialysis, the protein was re-applied to the HisTrap HP 5 mL column to remove the uncleaved P protein and 3C protease. The collected flow-through proteins were concentrated and further purified by SEC using Superose 6 Increase 10/300 column equilibrated with 20 mM Tris pH 8.0, 150 mM NaCl.

2.3.3. Purification of the co-expressed NiV N and NiV P

3 g of cell pellets were resuspended in 50 mL of HisTrap buffer supplemented with 1 mM AEBSF, 0.5 µg/ml leupeptin, and 0.7 µg/ml pepstatin. The cell suspension was then subjected to sonication and purification as in section **2.3.1** above using buffers listed in Table 2.5.

Table 2.5 Protein purification buffer for co-expressed protein complex

Condition	HisTrap buffer	SEC buffer
His-NiV P + NiV N		
250 mM NaCl	25 mM Tris pH 8 250 mM NaCl 10% Glycerol 20 mM Imidazole 2 mM BME	25 mM Tris pH 8 250 mM NaCl 1 mM EDTA 2 mM BME
200 mM KCl	25 mM Tris pH 8 200 mM KCl 0.1% Tween20 20 mM Imidazole 2 mM BME	25 mM Tris pH 8 200 mM KCl 0.1% Tween20 5 mM $MgCl_2$ 1 mM DTT
His-NiV N + NiV P		
300 mM NaCl	25 mM Tris pH 8 300 mM NaCl 10% Glycerol 20 mM Imidazole 2 mM BME	25 mM Tris pH 8 300 mM NaCl 1 mM DTT
200 mM KCl	25 mM Tris pH 8 200 mM KCl 10% Glycerol 20 mM Imidazole 2 mM BME	25 mM Tris pH 8 200 mM KCl 1 mM DTT

2.4   Protein concentration determination

The protein concentration was determined using the Bradford Assay (Thermo Fisher Scientific). 3 µL of N protein was added to 1 mL of Bradford solution and mixed. The mixtures were incubated for 5 min at room temperature and the absorbance reading at 595 nm was measured using DU720 UV Vis spectrophotometer (Beckman). A standard curve was generated using the BSA standard with known concentration. This standard curve was used to deduce the protein concentration.

2.5   Polyacrylamide Gel Electrophoresis

SDS-PAGE was run using Hoefer SE250 Mini vertical Protein Electrophoresis Unit in a discontinuous polyacrylamide gel containing sodium dodecyl sulphate (SDS). The molecular weight was calculated with a pre-stained molecular ruler (Thermo Scientific) or an unstained molecular ruler (Bio-rad). The protein bands were visualised by staining the gel with 0.1% (w/v) Coomassie blue. Native PAGE was run using a Hoefer SE260 Deluxe Mini vertical Protein Electrophoresis Unit in a continuous polyacrylamide gel. For samples containing nucleic acid, the native PAGEs were first stained with

ethidium bromide (1 µg/mL) and the nucleic acids were visualised by exposure to UV light. The native PAGEs were then stained with 0.1% (w/v) Coomassie blue to visualise protein bands. Urea PAGE was run using a Hoefer SE250 Mini vertical Protein Electrophoresis Unit in a continuous polyacrylamide gel containing 8M urea. The length of nucleic acid was calculated with an Ultra-Low Ladder. The gels were stained with ethidium bromide (1 µg/mL) and the nucleic acids were visualised by exposure to UV light.

Table 2.6 Electrophoresis running buffer

SDS PAGE		Urea PAGE	
SDS Running Buffer		**TBE Buffer**	
25 mM Tris		100 mM Tris	
190 mM Glycine		100 mM Boric Acid	
0.1% SDS		2 mM EDTA	
Native PAGE			
Tris Glycine Buffer pH 8	**Tris Glycine Buffer pH 8 + 0.1 M NaSO$_4$**	**Tris CAPS Buffer pH 10**	**Tris CAPS Buffer pH 10 + 0.1 M NaSO$_4$**
25 mM Tris	25 mM Tris	60 mM Tris	60 mM Tris
190 mM Glycine	190 mM Glycine	40 mM CAPS	40 mM CAPS
	100 mM NaSO$_4$		100 mM NaSO$_4$

Table 2.7 Recipes for resolving / continuous gels (10mL)

Component	7.5%	10%	12%	15%
Water	5.0 mL	4.0 mL	3.2 mL	2.4 mL
1.5 M Tris pH 8.8[#]	2.5 mL	2.5 mL	2.5 mL	2.5 mL
30% Acrylamide[†]	2.5 mL	3.3 mL	4.2 mL	5.0 mL
10% APS	50 µL	50 µL	50 µL	50 µL
TEMED	8 µL	8 µL	8 µL	8 µL

For SDS PAGE, the 1.5 M Tris pH 8.8 used contains 0.4% SDS
† For SDS PAGE, 37.5:1 Acrylamide/Bis was used. For Native PAGE, 99:1 Acrylamide/Bis was used.

2.6   *In vitro* assembly of NiV N/P heterocomplex

For agarose gel electrophoresis analysis, *in vitro* assembly of NiV N/P heterocomplex was done at 20 µL scale using His tagged NiV N and His tagged NiV P. The molar concentration (µM) was calculated based on the protein mass concentration (µg/mL), determined from Bradford Assay above, using the simple formula:

$$µM = (µg/mL)/(MW \text{ in kDa})$$

Both of the His tagged NiV proteins were mixed at three different molar ratios (Table 2.8) and incubated for 30 min at room temperature. After incubation, the protein mixtures were analysed by 1% agarose gel electrophoresis. The agarose gel

electrophoresis was performed using TBE as a running buffer and it was run overnight at 10 mA at 4°C in a cold room. The gel was stained with 0.1% (w/v) Coomassie blue to visualise protein bands.

Table 2.8 Reaction compositions for the *in vitro* assembly of NiV N/P heterocomplex

N:P Molar ratio	6:1	2:1	2:3
NiV N protein volume (1.2 mg/mL, 20 μM)	15 μL	10 μL	5 μL
NiV P protein volume (0.8 mg/mL, 10 μM)	5 μL	10 μL	15 μL

For purification of the NiV N/P heterocomplex, *in vitro* assembly was performed at 500 μL scale using His tag cleaved NiV N and His tag cleaved NiV P. To saturate the N protein (1 mg/mL) with the P protein, both of the His tag cleaved NiV proteins were mixed at 1:1 molar ratio and incubated for 30 min at room temperature. After incubation, the protein mixture was diluted by addition of 4.5 mL SEC buffer as in Table 2.9. The protein mixtures were concentrated back to 500 μL using protein concentrator Vivaspin 20 (100 kDa MWCO) and further purified by SEC (Superose 6 Increase 10/300, GE Healthcare) in SEC buffer.

Table 2.9 SEC buffer for *in vitro* assembly

Condition	SEC Buffer
500 mM NaCl	25 mM Tris pH 8 500 mM NaCl
250 mM NaCl	25 mM Tris pH 8 250 mM NaCl

2.7 Digestion of NiV N by Proteinase K, RNase A, and DNase

10 μg of purified NiV N protein was subjected to Proteinase K, RNase A and DNase digestion in 20 μL digestion buffer (50 mM Tris pH 8, 150 mM NaCl, 20 mM $MgCl_2$, 2 mM $CaCl_2$). To release the bound nucleic acid, the N protein was first treated with 0.5 mg/mL (final concentration) proteinase K at room temperature. After overnight incubation, all proteins including the proteinase K were digested due to the non-specific digestion nature of proteinase K. The sample was further heated at 60°C for 5 min and chilled to room temperature before addition of RNase A or DNase to the

sample (0.5 mg/mL final concentration). To check for the nucleic acid susceptibility to nuclease degradation, RNase A or DNase was added to the N protein sample and incubated for an hour at room temperature. The nuclease activity was inhibited by addition of SDS to the final concentration of 0.5% (v/v), before overnight proteinase K digestion. The nucleic acid was subjected to 10% Native PAGE and visualised with ethidium bromide post-staining. The sample was also subjected to 12% SDS PAGE to verify that all protein was digested.

For the determination of the RNA length, the 15% urea PAGE was first pre-run at 200 V for an hour at RT. The purified N protein was mixed with 2X RNA loading dye (8 M urea, 20% (v/v) glycerol, 0.05% (w/v) bromophenol blue) and heated for 5 min at 95°C. The combination of urea and heat works to denature the N protein and release the bound nucleic acid. After heating for 5 min, the sample was rapidly chilled on ice for 1 min before being subjected to 15% urea PAGE. The nucleic acid was visualised with ethidium bromide post-staining.

2.8   Negative stained Transmission Electron Microscopy

For negative staining, 3 μL of sample (~0.05 mg/mL) was applied to a glow discharged Formvar 200 mesh carbon support grid (Agar Scientific). The grid was rinsed twice with MilliQ water before being stained using the 2% (w/v) uranyl acetate solution. Negatively stained grids were imaged on a Tecnai 12 BioTWIN G2 transmission electron microscope (FEI) operating at 120 keV and equipped with SIS Megaview III CCD camera. Images were recorded at a magnification of x68,000 with the defocus set at ~1 μm.

2.9   CryoEM screening and data collection of NiV N

3 μl of NiV N protein (0.4 mg/mL) was applied to glow-discharged UltraAuFoil R1.2/1.3 gold support grids (Quantifoil), blotted for 2 s with -10 force setting, and vitrified by plunging the grid into liquid ethane using the FEI Vitrobot Mark IV at 4°C and 100% relative humidity. Initial imaging to optimise the protein concentration and grid blotting setting was done using the Tecnai 12 BioTWIN G2 transmission electron microscope (FEI) operating at 120 keV using the Fischione cryo transfer holder and SIS Megaview III CCD camera. Grids were loaded into the microscope by Dr Sam Hart. Grids prepared using the optimised condition were then clipped and further screened

using Tecnai Arctica microscope at University of Sheffield with the help from Dr Svetomir Tzokov.

Micrographs for structure determination were collected at the Diamond eBIC facility on a Titan Krios microscope (FEI) operating at 300 keV and equipped with K2 camera and an energy filter slit width of 20 eV (Gatan). Automated data collection was performed using the FEI EPU software. 1879 movies with a total electron dose of ~41 e-/$Å^2$ were recorded in counting mode over 11 s (40 frames) with a pixel size of 1.048 Å. The defocus range chosen for automatic collection was 0.5 to 2.1 μm.

2.9.1. CryoEM Image processing

All datasets were processed using RELION 3.0[110] unless stated otherwise. The raw movie frames collected were first aligned to account for beam-induced sample motion and stage movement using MotionCor2[111]. Micrographs that failed to be aligned were inspected and micrographs that contain just empty ice were identified and removed by modifying the *corrected_micrographs.star* file. Contrast transfer function (CTF) parameters were estimated using CTFFIND4[112] with the following parameters:

Use exhaustive search?:	YES
FFT box size (pix):	512
Minimum resolution (Å):	30
Maximum resolution (Å):	3
Minimum defocus value (Å):	3000
Maximum defocus value (Å):	30000
Defocus step size (Å):	200

Micrographs that had an estimated resolution of >25 Å were inspected and micrographs that contained just empty ice or ice contamination were removed by modifying the *micrographs_ctf.st*ar file. Approximately 6000 NiV N particles were manually picked for generation of initial averaged projections that served as references for Autopicking using RELION. After the Autopicking by RELION, all the micrographs were manually inspected to ensure the picking of all visible rare views. A total of 217,522 particles were extracted with a binned sized 4X (4.192 Å/pixel, 88-pixel box) and subjected to reference-free 2D classification to remove particles associated with noisy or contaminated classes. Attempts to generate a de novo 3D model using stochastic gradient descent (SGD) from the 2D particles was not successful, likely due to the preferential orientation of particles within the dataset.

Instead, the resulting 189,662 particles were subjected to 3D classification using a map generated from the Measles N protein (EMDB:0141)[113], trimmed to single turn helix using UCSF Chimera's "Volume Eraser" function[114] and low-pass filtered to 60 Å, as a reference model. The resulting best 3D class was low-pass filtered to 30 Å and used as a reference model for a new round of 3D classification against the same initial set of particles. Two distinct conformations were identified among the NiV N particles: spiral assembly and clam shaped assembly.

For the spiral assembly, initial 3D refinement was focused on NiV N particles that only form a single spiral turn by excluding all multi-turn NiV N particles. However, the resulting cryoEM map was highly anisotropic, due to the lack of side-views. Therefore, the 3D refinement of the final spiral map was achieved by using all unbinned particles (1.048 Å/pixel, 384-pixel box) from the spiral classes (124,891 particles) with a 13-protomer spiral turn solvent mask yielding a less anisotropic cryoEM map. Finally, local symmetry was imposed to further improve the cryoEM map quality.

For the local symmetry, masks around all 13 protomers were created and low-pass filtered to 15 Å. The 3D transformation matrices, which were needed to superimpose protomers onto each other within the group, were calculated from the search feature of the *relion_local_symmetry* that finds and optimizes the 3D transformation matrices in real space. This was done by fitting the PDB model into every set of identical regions and converting them into MRC density maps of those local regions using UCSF Chimera's "molmap" command[114]. The MRC map was used to generate a soft-edged, binary mask for each set of identical regions. Several rounds of searches with gradually decreased searching ranges and steps were then performed until the refined values for 3D transformation matrices were converged to required precision (~0.1° and ~0.1 pixels)(Table 2.10)[115]. This final 3D transformation matrices were firstly verified and inspected to ensure all the local symmetry operators and masks provided were accurate, before it being used for 3D refinement using the option *--local_symmetry*.

Table 2.10 Searching scheme for 3D transformation matrices needed for the local symmetry

Round	Angular ranges (°)	Angular sampling (°)	Translational ranges (pix)	Translational samplings (pix)
1	360, all angles	5	--	--
2	10	1	--	--
3	2	0.5	2	0.5
4	2	0.1	--	--
5	0.5	0.1	0.5	0.1

At each iteration of the 3D refinement (every step of expectation-maximization algorithm), the local symmetry was applied according to the input mask and 3D transformation matrices. This symmetrisation was performed after the maximisation step in real space, to ensure that the signal to noise ratio gain from the local symmetry is not considered when calculating the Fourier shell correlation (FSC) between the two independent half-map reconstruction in 3D refinement[116]. Therefore, during the 3D refinement the empirical regularization T-value, *--tau2_fudge*, was set to 13 to account for the expected gain in signal during the refinement. Similar processing strategy has been used during the 3D reconstruction of the cryoEM structure of Fanconi anaemia ligase complex[116]. Subsequent per-particle CTF refinement and Bayesian polishing in RELION 3.1beta led to a final map at 3.5 Å resolution, estimated by the 0.143 FSC criterion. The maps were postprocessed in RELION 3.1beta[117] and are shown after B-factor sharpening.

The remaining 64,771 non-spiral particles were further subjected to 3D classification using the "clam-shaped" 3D class, obtained from previous 3D classification, as a reference model. Two major 3D classes, a spiral clam-shaped assembly (23,029 particles, 1.572Å/pixel, 256-pixel box) and a semi-spiral clam assembly (18,979 particles, 1.572Å/pixel, 256-pixel box), were selected. Subsequent per-particle CTF refinement in RELION 3.1beta, 3D refinement of these two 3D classes resulted in final maps at 4.3 Å (spiral clam-shaped assembly) and 5.2 Å (semi-spiral clam-shaped assembly) resolution, respectively. The local resolution maps were calculated using RELION 3.1beta.

2.9.2. Model building, refinement, and analysis

Atomic model building of the NiV nucleocapsid spiral assembly was performed using the previously reported crystal structure of the RNA free NiV nucleocapsid (pdb:4co6)[53] as an initial model, which was docked as a rigid body into the 3.5 Å resolution CryoEM maps using the UCSF Chimera's "Fit in map" function[114]. The RNA chain was modelled as poly-uridine into the cryoEM map, and the whole protein-RNA model was adjusted manually using Coot[118]. Poly-uridine was chosen because it has the smallest possible nucleobase amongst RNA bases.

For the refinement, the RNA chain was first subjected to refinement using ERRASER2[119] in the absence of protein chain. Despite the fact that ERRASER2 has the ability to work with protein-RNA complexes, attempts to use the whole protein-RNA model as an input resulted in "insufficient memory" error, on a 256 GB RAM computer cluster. Given that the ERRASER2 software only resamples RNA from the input pdb file, and other protein components are held fixed and only contributed to the scoring "environment" of the resampled RNA, the decision was made to just use the isolated RNA chain as the input, and it was successfully run on a 16GB RAM desktop. The cryoEM map of the RNA chain was isolated using the "vop zone" command in UCSF's Chimera protein with a radius of 5 Å. The fasta file of the RNA chain was generated using the *pdb2fasta.py* script provided in the Rosetta *tools/rna_tools/pdb_util* directory.

The command to run the ERRASER2 was as below:

```
erraser2 -s [input].pdb -edensity:mapfile [input].mrc -fasta [fasta].fasta -edensity::mapreso
[map resolution] -score:weights stepwise/rna/rna_res_level_energy4.wts -set_weights
elec_dens_fast 10.0 cart_bonded 5.0 linear_chainbreak 10.0 chainbreak 10.0 fa_rep 1.5
fa_intra_rep 0.5 rna_torsion 10 suiteness_bonus 5 rna_sugar_close 10 -rmsd_screen 3.0 -mute
core.scoring.CartesianBondedEnergy
```

The refined RNA chain from the ERRASER2's output was rigid body fitted back into the cryoEM map together with the protein model. The whole protein-RNA model was further refined using ISOLDE[120] with default settings. All the identified Ramachandran and rotamer outliers were manually adjusted and further refined within ChimeraX using ISOLDE. The final refined model from the ISOLDE output was subjected to ProSMART to generate a *self_restraint.txt* file which was then used as a set of restraints parameter during refinement using REFMAC5[121]. Setting for the REFMAC5 were as shown below:

```
Refmac cycles = 20
Auto weight = True
Sharpen / Blur = 50
Jelly body = False
Add hydrogen = False
Use restraints = ProSMART_self_restraint.txt
Auto symmetry = None
```

For the "spiral clam" CryoEM maps, the monomeric model of the RNA-bound NiV N protein (taken from NiV N spiral assembly) was fitted as a rigid body into the maps[114]. No protein models were fitted into the seam regions, due to the lack of interpretable map features in the CryoEM maps. Models were refined using Refmac5[121] with the same setting as above.

Protein interfaces were analysed using the COCOMaps server[122]. Protein domain motion was analysed using the DynDom server[123]. Multiple sequence alignments were performed using Clustal Omega[124] and visualised in JalView[125]. Phylogenetic tree analysis was conducted in MEGA X[126]. The sequence identity matrix was calculated using MatGAT[127]. The conserved regions in the structure were analysed using ConSurf server[128]. Figures showing protein/RNA structure were created using UCSF ChimeraX[129]. Structural comparison between the NiV N protein structure in the RNA-free and RNA-bound states was done using ChimeraX's MatchMaker feature, which first creates pairwise sequence alignments, then fits the aligned residue pairs. To calculate the electrostatic potential, the PDB format files were converted into the PQR format with the PDB2PQR server using the PARSE force field and assigned protonation states at pH 7.0. The file was applied to the APBS server by including 0.15 M of ions in the calculation[130].

Table 2.11 Statistics of CryoEM data collection and processing of NiV N protein

	Spiral	Spiral Clam
Data collection		
Voltage (kV)	300	
Detector	Gatan K2 Summit	
Electron exposure (e-/Å^2)	41.2	
Defocus range (μm)	0.5 to 2.1	
Pixel size (Å)	1.048	
Data processing		
Symmetry imposed	C1	C1
Final particle images (no.)	124,891	23,029
Map resolution (Å)	3.5	4.3
FSC threshold	0.143	0.143
Map sharpening B factor (Å^2)	-69	-10
Model composition		
Non-hydrogen atoms	40200	24128
Protein residues	5088	5532
Nucleic acid residues	78	90
Validation		
RMS Bond lengths (Å)	0.013	0.013
RMS Bond angles (°)	2.2	1.9
MolProbity score	1.7	1.1
Clashscore	6.53	0.87
Rotamer outliers (%)	1.5	0.0
Ramachandran Favoured (%)	96.5	95.6
Ramachandran Outliers (%)	0.18	0.33
RNA average suiteness	0.560	0.566

2.10   CryoEM screening and data collection for the NiV N/P heterocomplex

3 μl of SEC purified NiV N/P heterocomplex (0.1 mg/mL) was applied to glow-discharged Cu 300 Mesh, 1.2/1.3 (Quantifoil), blotted for 3 s with -5 force, and vitrified by plunging the grid into liquid ethane using the FEI Vitrobot Mark IV operated at 4°C and 100% relative humidity. Initial grid screening and data collection was done using a 200 kV Glacios microscope (Thermo Scientific) equipped with a Falcon-IV counting direct electron detector and a Schottky field emission gun (FEG). The grids were screened using the Glacios microscope and optimum grids were selected for cryoEM data collection. Automated data collection was performed using the FEI EPU software. 6796 movies with a total electron dose of 50.11 e-/Å^2 were recorded in counting mode

over 5.73 s (1379 internal frames) with a pixel size of 0.58 Å. The defocus range chosen for automatic collection was 0.5 to 2.0 μm.

2.10.1.   CryoEM Image processing

All datasets were processed using RELION 3.1.2[110] unless stated otherwise. The 1379 internal frames from the EER dataset were grouped by 29 to get 47 fractions in which each fraction has 1.053 e-/Å^2. The EER files were then rendered in 4K and pre-processed as described in section **2.9.1** above**.**

The initial particles were selected using RELION the LoG-based auto-picking feature and the particles were sorted using reference-free 2D classification. The two best class averages with different orientations each were selected as reference template for RELION Auto-picking. A total of 629,683 particles were extracted with a binned sized 7X (4.54Å/pixel, 88-pixel box) and subjected to reference-free 2D classification to remove particles associated with noisy or contaminated classes. The resulting 311,217 particles were subjected to 3D classification using the cryoEM map from NiV N protein (EMD-12581)[131], low-pass filtered to 60 Å, as a reference model. One good 3D class was identified as a rigid helical turn assembly. Further 3D refinement of this assembly was done using particles with a binned sized 3X (1.74Å/pixel, 230-pixel box), leading to an 8 Å resolution map.

Two distinct conformations were further identified among the particles of the NiV N/P heterocomplex after an additional round of 3D classification: spiral assembly and an uncoiling loose helical assembly. Each of the assemblies were selected and subjected to additional 2D and 3D classifications. Within the spiral assembly particles, 6302 particles were identified as clam shaped assembly from the 2D classification. Several rounds of 3D classification of the remaining 54925 spiral assembly particles were done by varying the "number of classes", but no reasonably distinct 3D classes were identified. This was likely due to the lack of angular coverages which in turn resulted in the poor angular assignment of the projection during 3D reconstruction. For the uncoiling loose helical assembly, the 3D classification identified a low-resolution distinct 3D class. Further 3D refinement of this 3D class was done using particles with a binned sized 3X (1.74Å/pixel, 230-pixel box), leading to a 12 Å resolution map.

Table 2.12 Statistics of CryoEM data collection and processing of NiV N/P heterocomplex

	Rigid spiral	Loose spiral
Data collection		
Voltage (kv)	200	
Detector	Falcon 4	
Electron exposure (e-/Å^2)	50.11	
Defocus range (μm)	0.5 to 2.0	
Pixel size (Å)	1.048	
Data processing		
Symmetry imposed	C1	C1
Final particle images (no.)	15,149	33,033
Map resolution (Å)	8	12
FSC threshold	0.143	0.143

2.10.2.   CryoEM map fitting

The monomeric model of the RNA-bound NiV N protein (taken from NiV N spiral assembly) was fitted as a rigid body into the 12 Å loose spiral cryoEM maps using the "Fit in map" feature in Chimera[114]. The optimum model fitting was obtained by correlation optimisation during map-in-map fitting. The difference map was obtained by using the volume operation (vop) feature of Chimera. The 1-36 aa region of NiV P protein (pdb:4co6)[53] and the X domain of HeV P protein (pdb:4heo)[132] were fetched from PDB and fitted as a rigid body into the difference map.

Chapter 3   CryoEM structure of the Nipah virus nucleocapsid assembly

3.1   Introduction

Paramyxoviruses store their genomic RNA by tightly encapsidating it in a helical array formed with the nucleocapsid (N) protein, thus protecting the viral genome from host ribonuclease degradation and host cell immune detection[133]. Besides its protective role, the N protein concentration has also been shown to be an important factor in regulating the switch between the transcription and the replication process during viral infection as described earlier in Chapter 1. Most recently, structures of nucleocapsid-like assemblies of several paramyxoviruses, assembled as helical[73], ring[72] or clam-shaped[134] complexes, have been determined by either X-ray crystallography or cryo electron microscopy. However, the nucleocapsid of NiV shares only 32% sequence identity with the nucleocapsid of the Measles virus, the closest homologue with an available structure. Although the overall fold would be expected to be preserved, the extent of structural differences due to this considerable evolutionarily divergence would make it challenging to precisely map the biochemical findings to the structure, using only a homologous structure. The only structural information for the NiV nucleocapsid available is for a truncated single subunit lacking N-terminal (residues 1-31) and C-terminal (residues 384-532) regions, which was determined in complex with a 50 amino acid peptide of the P protein[53]. No structural information is available for the oligomeric assembly of the NiV nucleocapsid, despite the ability of the recombinant protein expressed in bacteria[135], yeast[136] and insect cells[137] to form nucleocapsid-like helical structures containing cellular RNA.

This chapter reports cryoEM structures of several different types of assemblies formed by recombinantly produced NiV N protein, elucidating detailed information about protein-RNA interactions. These structures also reveal how the N- and C-terminal segments of the NiV N protein, which were not present in the construct used to determine the previous N protein structure, stabilise the assembly, by interacting with the same surfaces of adjacent subunits, that were previously shown to interact with the P-protein[53]. These new data permit the analysis of similarities and differences with

other members of the *Paramyxoviridae* family as well as more distantly related members from the *Mononegavirales* order of ssRNA viruses.

3.2   Results

3.2.1. Purification and characterization of the Nipah N protein

The NiV N protein was soluble in *E. coli* when expressed at 16°C. The protein was successfully purified using two purification steps, nickel affinity chromatography and size exclusion chromatography (SEC). The N protein was eluted as a major peak with an apparent molecular weight of larger than 660 kDa from a Superose 6 column (Figure 3.1). The UV absorbance readings for SEC showed an A_{260}/A_{280} of ~1, indicating the presence of nucleic acid within the purified sample, presumably due to the nonspecific binding of the N protein to *E. coli* host nucleic acid. The A_{260}/A_{280} ratio for pure protein with minimal nucleic acid contamination is ~0.6[138]. Direct imaging of negatively stained purified samples using transmission electron microscope (TEM) showed that the N-RNA protein is purified as ring like structures with a single ring structure predominating.

The N protein has been observed to undergo a time dependent degradation into a protein band size around 50 kDa and smaller, when stored at 4°C (Figure 3.2). The N protein is predicted to have a 132 aa long disordered region at its C-terminal N-tail region, and protein degradation of the N-tail region of Nipah N protein has been reported previously[135,139]. Mass spectrometry analysis indicates that both protein bands (60 kDa and 50 kDa) are N protein and there is a missing peptide hit in the region of N-tail in the 50 kDa protein band which provides supportive indication that the missing 10 kDa may be due to the degradation of the 100 aa C-terminal disordered region (Figure 3.2).
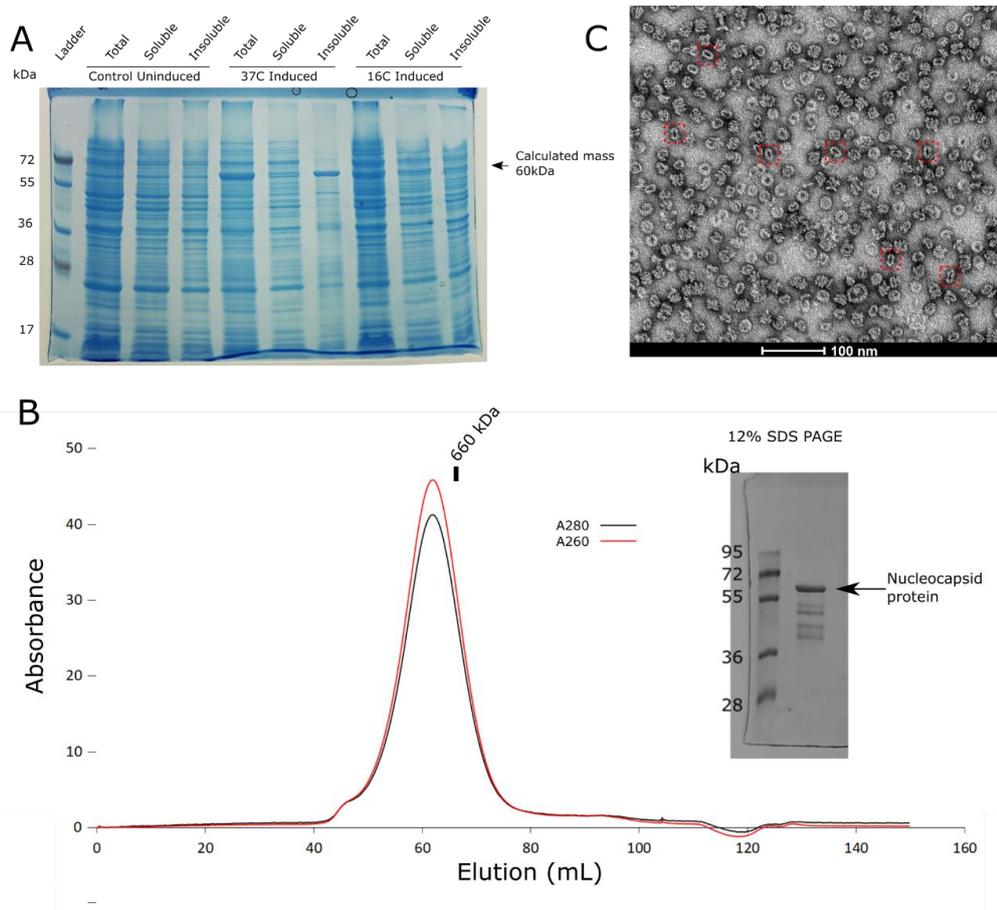
Figure 3.1 Expression and purification of the NiV N protein. (A) Small scale expression of the NiV N protein at 16°C and 37°C. (B) SEC purification and SDS PAGE profile of Nipah N protein using a Superose 6 16/600 column. Elution retention volume of thyroglobulin is indicated. (C) Negative stained EM micrograph of the SEC purified N-RNA protein from (B). Dashed red boxes represent side-views of the clam-shaped assembly.

As the N protein is known to protect the RNA from RNase degradation[140], proteinase K was used to degrade the protein and release the bound RNA. The susceptibility of the bound RNA toward the nuclease digestion was also tested by the addition of DNase or RNase A prior to the Proteinase K digestion. The result clearly showed that the bound nucleic acid is RNA, as the intensity of the nucleic acid band faded significantly after RNase A digestion (Figure 3.3A). Interestingly, addition of RNase A prior to Proteinase K digestion appeared to have degraded the bound RNA. Early works on isolating Paramyxoviruses clearly shown that the RNA within the N protein is highly resistant to 20-30 µg/mL of RNase digestion at any salt concentration[140,141]. However, work on Mump virus (MuV) N protein had reported the successful removal of bound RNA by incubated the MuV N protein at high concentration of RNase A (1 mg/mL)[142]. Similar observations were made in this work when NiV N protein was treated with 0.5 mg/mL RNase A. Although it cannot be ruled out that the RNase may remained partially active

after the addition of 0.5% SDS, and may have digested the RNA upon N protein digestion by Proteinase K (Figure 3.3A). Electrophoresis using urea PAGE strongly indicated that the nucleic acid bound in the N protein is a single stranded RNA with a length ranging from 60 nt to 150 nt (Figure 3.3C). A single ring of the N protein is known to bind to a 78 nt long single stranded RNA[72], and the electrophoresis suggests that the sample has a high heterogeneity, with a majority of the RNA being longer than 78 nt.

Protein sequence coverage: 21%

Matched peptides shown in **bold red**.

```
  1 MGSSHHHHHH SSGLEVLFQG PAMSDIFEEA ASFRSYQSKL GRDGRASAAT
 51 ATLTTKIRIF VPATNSPELR WELTLFALDV IRSPSAAESM KVGAAFTLIS
101 MYSERPGALI RSLLNDPDIE AVIIDVGSMV NGIPVMERRG DKAQEEMEGL
151 MRILKTARDS SKGKTPFVDS RAYGLRITDM STLVSAVITI EAQIWILIAK
201 AVTAPDTAEE SETRRWAKYV QQKRVNPFFA LTQQWLTEMR NLLSQSLSVR
251 KFMVEILIEV KKGGSAKGRA VEIISDIGNY VEETGMAGFF ATIRFGLETR
301 YPALALNEFQ SDLNTIKSLM LLYREIGPRA PYMVLLEESI QTKFAPGGYP
351 LLWSFAMGVA TTIDRSMGAL NINRGYLEPM YFRLGQKSAR HHAGGIDQNM
401 ANRLGLSSDQ VAELAAAVQE TSAGRQESNV QAREAKFAAG GVLIGGSDQD
451 IDEGEEPIEQ SGRQSVTFKR EMSISSLANS VPSSSVSTSG GTRLTNSLLN
501 LRSRLAAKAA KEAASSNATD DPAISNRTQG ESEKKNNQDL KPAQNDLDFV
551 RADV
```

Protein sequence coverage: 19%

Matched peptides shown in **bold red**.

```
  1 MGSSHHHHHH SSGLEVLFQG PAMSDIFEEA ASFRSYQSKL GRDGRASAAT
 51 ATLTTKIRIF VPATNSPELR WELTLFALDV IRSPSAAESM KVGAAFTLIS
101 MYSERPGALI RSLLNDPDIE AVIIDVGSMV NGIPVMERRG DKAQEEMEGL
151 MRILKTARDS SKGKTPFVDS RAYGLRITDM STLVSAVITI EAQIWILIAK
201 AVTAPDTAEE SETRRWAKYV QQKRVNPFFA LTQQWLTEMR NLLSQSLSVR
251 KFMVEILIEV KKGGSAKGRA VEIISDIGNY VEETGMAGFF ATIRFGLETR
301 YPALALNEFQ SDLNTIKSLM LLYREIGPRA PYMVLLEESI QTKFAPGGYP
351 LLWSFAMGVA TTIDRSMGAL NINRGYLEPM YFRLGQKSAR HHAGGIDQNM
401 ANRLGLSSDQ VAELAAAVQE TSAGRQESNV QAREAKFAAG GVLIGGSDQD
451 IDEGEEPIEQ SGRQSVTFKR EMSISSLANS VPSSSVSTSG GTRLTNSLLN
501 LRSRLAAKAA KEAASSNATD DPAISNRTQG ESEKKNNQDL KPAQNDLDFV
551 RADV
```

Figure 3.2 Degradation of the purified NiV N protein. SDS PAGE of the N protein which had been stored at 4°C for 3 days and 27 days, respectively. Protein bands were sent for peptide fingerprinting using MALDI TOF. Mass spectrometry sequence coverage of tryptic digests of two protein bands were shown. Identified peptides are shown in bold red. Sequence and numbering are for the recombinantly NiV N containing the N-terminal 6xHis tag.
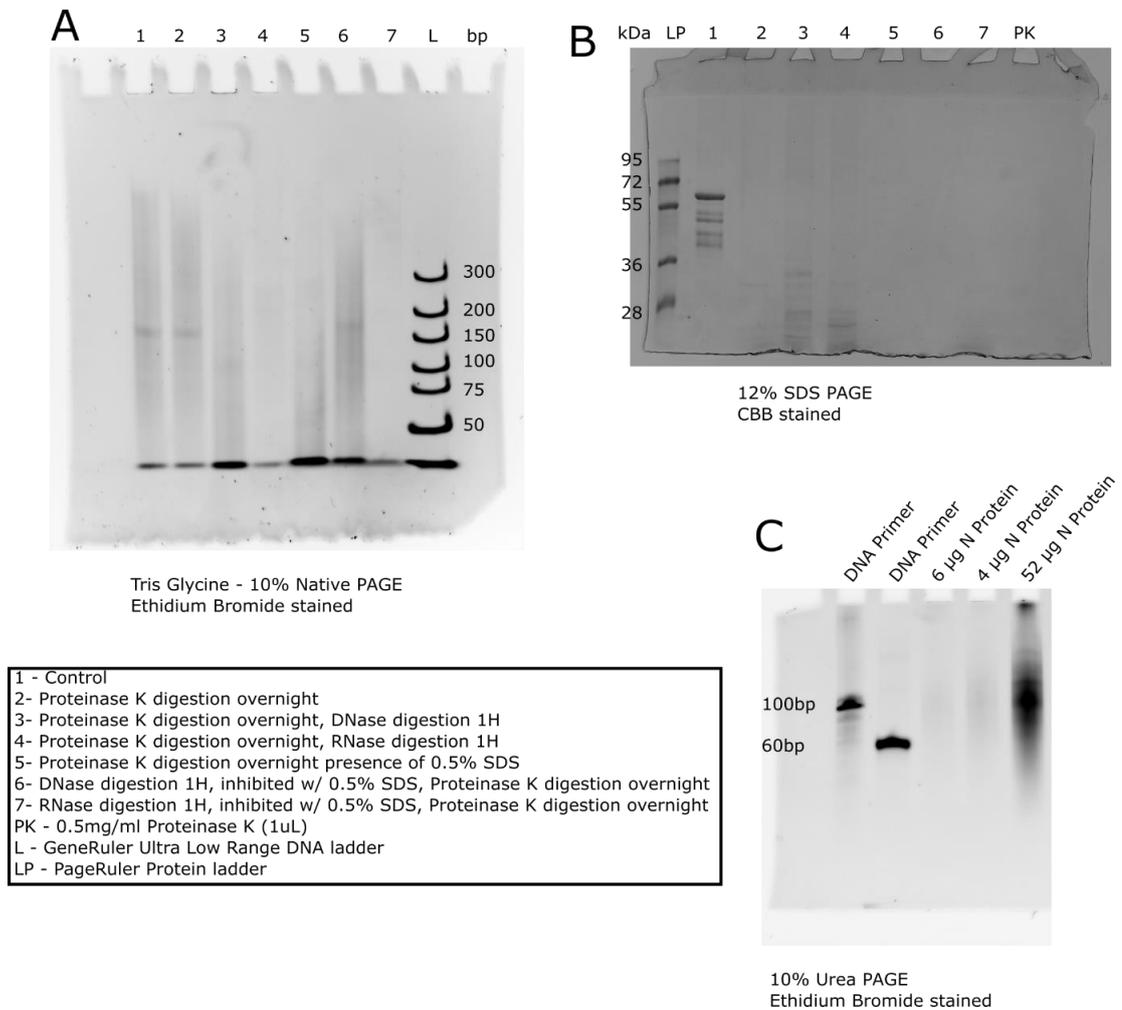
Figure 3.3 Identification of the bound nucleic acid by NiV N protein. (A) Native PAGE and (B) SDS-PAGE of the NiV N protein sample by DNase or RNase digestion prior or after overnight Proteinase K digestion. For the Native PAGE, it was stained with ethidium bromide to visualise the nucleic acid. (C) Urea PAGE of the purified NiV N protein at various protein concentrations. CBB, Coomassie brilliant blue.

3.2.2. Structure of the helical assembly

Using CryoEM single-particle analysis, the spiral assembly of the full length NiV nucleocapsid protein, bound to *E. coli* cellular RNA, was determined. De novo 3D model generation was not successful due to the somewhat preferential orientation of the particles on the grid (see section **2.9.1** above). Therefore, the EM map of the nucleocapsid from the evolutionarily related measles virus (EMDB-0141) was used as an initial model after it was low-passed filtered at 60 Å (Figure 3.4).

2D and 3D classification showed that the majority of the particles represent a spiral assembly of 13 subunits per turn, with minor populations of particles representing a longer spiral with multiple turns, and a clam-shaped, face-to-face assembly of two short spirals[134]. Reconstruction of the spiral assembly (65% of the particles) was done

using a solvent mask corresponding to a 13-subunit spiral turn, resulted in a 3.5 Å resolution CryoEM map. To further improve the CryoEM map, local symmetry was imposed during the 3D reconstruction by averaging the 13 subunits of the NiV N protein complex (see section **2.9.1** above). The local symmetrised map showed better interpretation of the map when compared to the asymmetric NiV N protein (Figure 3.5A). Angular distribution analysis (Figure 3.5C) demonstrated that, there was a preferred orientation for the particles (viewed along the central axis of the helical assembly), explaining the failure of de novo 3D map generation for the construction of an initial model. Despite preferential orientation of the particles, there was still an acceptable distribution of particles across alternative orientations, including side-views (Figure 3.5C).

For model building, the NiV N monomer structure (pdb:4CO6)[53] was docked into the 3.5 Å CryoEM maps and served as a template for model building. Using COOT software, the N protein monomer was manually adjusted to fit to the CryoEM map, and the missing residues 4- 31, 116-121, 186-189 and 372-399 from the NiV N monomer structure template were manually built. The additional density was assigned as the RNA chain, and it was modelled as poly-uridine. The model was refined using Phenix[143] and Refmac5[121] to ensure proper geometry. However, the manually built RNA chain has a very poor geometry based on the MolProbity 's "RNA backbone" score[144]. Therefore, the RNA chain was subjected to remodelling using ERRASER2 to improve the geometry of the RNA chain (see section 2.9.2 above). The model built also has a very high crashscore (~30), and this crashscore was managed to reduce to ~5 using ISOLDE software which utilises Molecular Dynamic for structure refinement[120]. Due to the poorly defined density, the residues 116-121 and 369-399 are modelled as alanine residue.

The final model shows that 13 nucleocapsid monomers bind to the single-stranded RNA forming a left-handed spiral turn with outer and inner diameters of 204 Å and 56 Å, respectively (Figure 3.6). Assuming the NiV N-RNA protein forms a continuous spiral with the same symmetry, the pitch for NiV N-RNA protein is calculated to be 54 Å with 13.4 subunits per turn (Figure 3.7). Each N protein monomer is comprised of two main globular N-terminal and C-terminal N-core (Nucleocapsid core) domains, with each domain flanked by the N-terminal arm (NT-arm, residues 1-31) and the C-

terminal arm (CT-arm, residues 373-398) subdomains (Figure 3.6B). The C-terminal arm is associated with high atomic displacement parameters (ADP) and this is correlate with its low local resolution of the map, suggesting that greater flexibility in this region (Figure 3.8).The CT-arm is followed by a disordered C-terminal region (residues 399-532)[52] for which there is no clearly defined density. The C-terminal region is also prone to degradation and despite purifying the protein in the presence of protease inhibitors, minimal degradation is still observable (Figure 3.1B). Therefore, it is likely that the poorly resolved CryoEM map density in the C-terminal region is also partly contributed by the missing peptide in this region due to protein degradation.
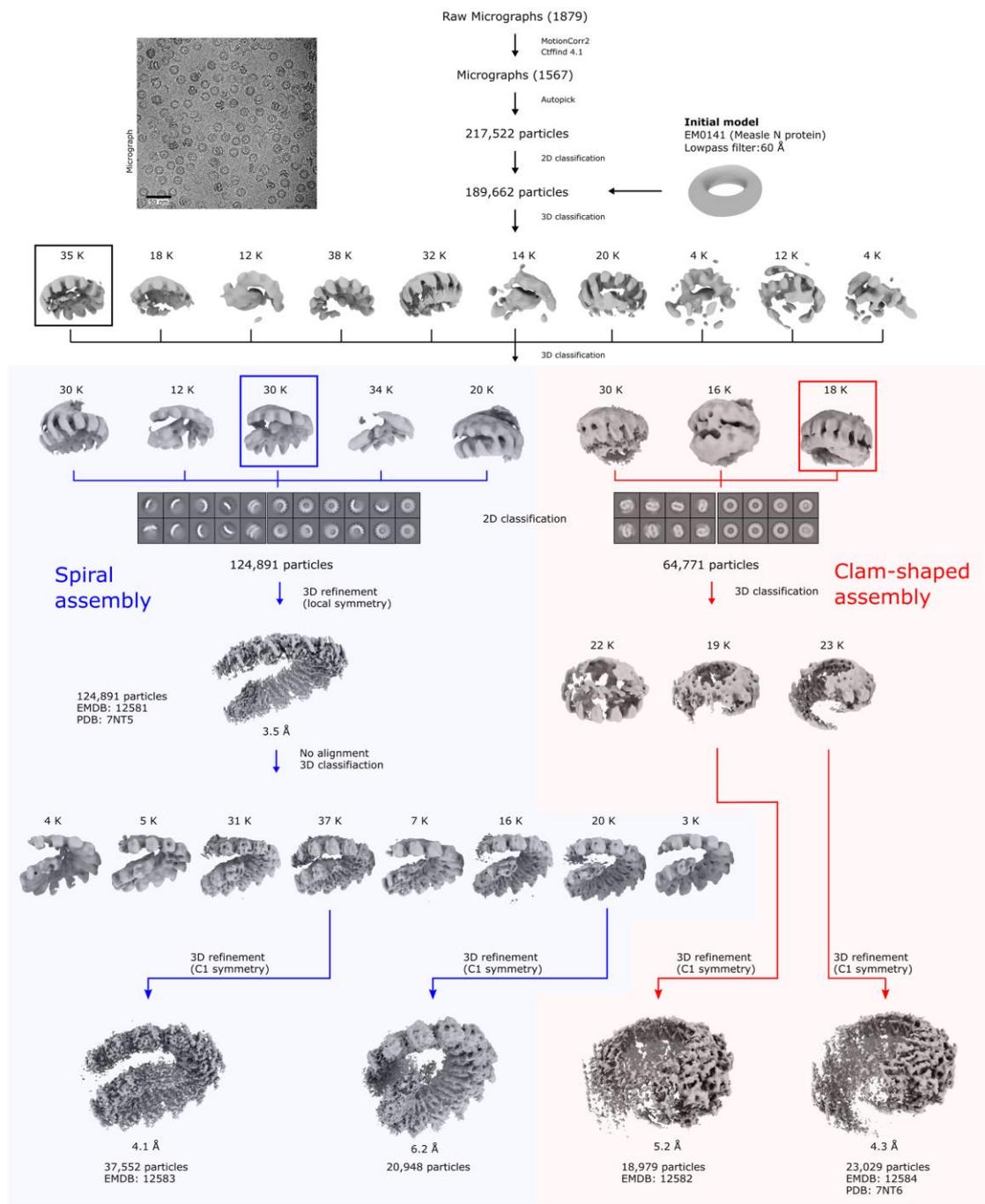
Figure 3.4 CryoEM data processing workflow. After the 2D classification, 189,662 particles were initially sorted by 3D classification using an initial model generated from EM0141 low-pass filtered to 60 Å. The best 3D class (squared on the figure) was used as a reference for a new round of 3D classification, to sort the 189,662 particles into the spiral assembly (65%) and clam-shaped assembly (35%) groups. An additional 2D classification was performed to inspect the selected particles and 2D class averages for each type of assembly. For the spiral assembly, a mask representing a single turn of the spiral assembly was applied to the map, leading to a 3.5 Å structure. Further 3D classification without alignment resulted in several spiral assembly maps, with different conformations and compositions (Figure 3.17). For the clam-shaped assembly, further 3D classification resulted in a semi-spiral and a spiral clam shaped assemblies.
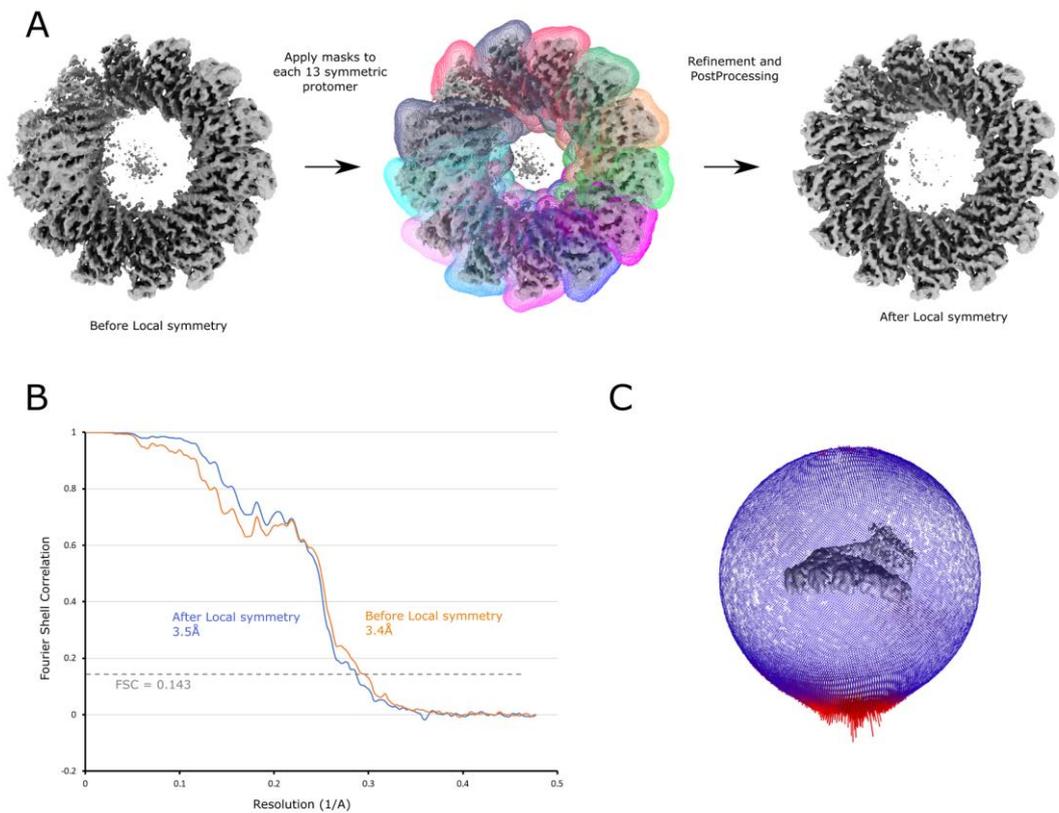
Figure 3.5 3D local symmetry refinement of the Nipah N-RNA protein spiral assembly. (A) Local symmetry refinement workflow. (B) "Gold-standard" FSC plot before and after local symmetry refinement. (C) Angular distribution plot for the spiral assembly.
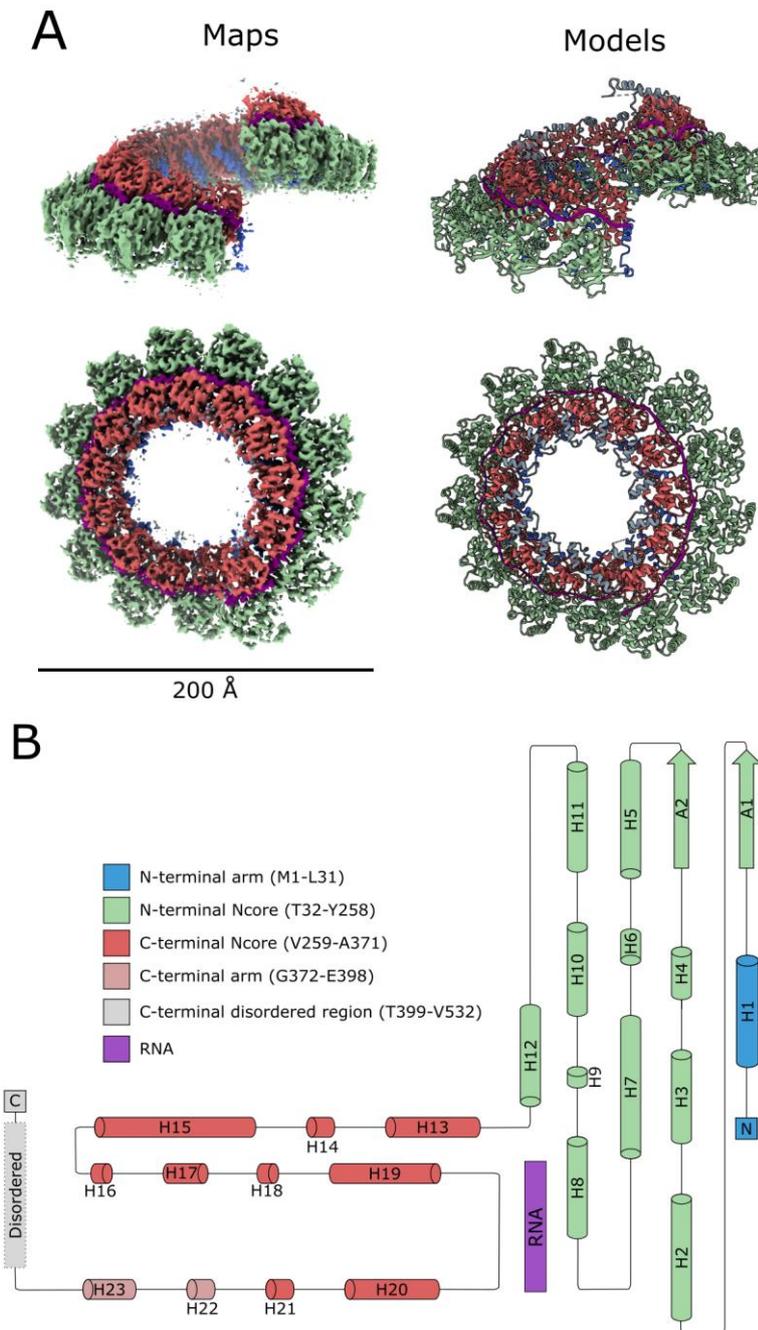
Figure 3.6. Structure of the NiV nucleocapsid protein-RNA complex. CryoEM structure of the nucleocapsid protein-RNA complex, determined at 3.5 Å resolution. Two orthogonal views of the CryoEM map (left) are shown next to the corresponding molecular models (right). (B) Schematic representation of the nucleocapsid protein, with the colour-coding as in (A). Boxed segments correspond to regions with defined structure. Dashed line corresponds to the disordered region.

Spiral Assembly

~54 Å

13.4 subunits per turn

Clam-shaped Assembly

~44 Å

~45 Å

13.1 subunits per turn
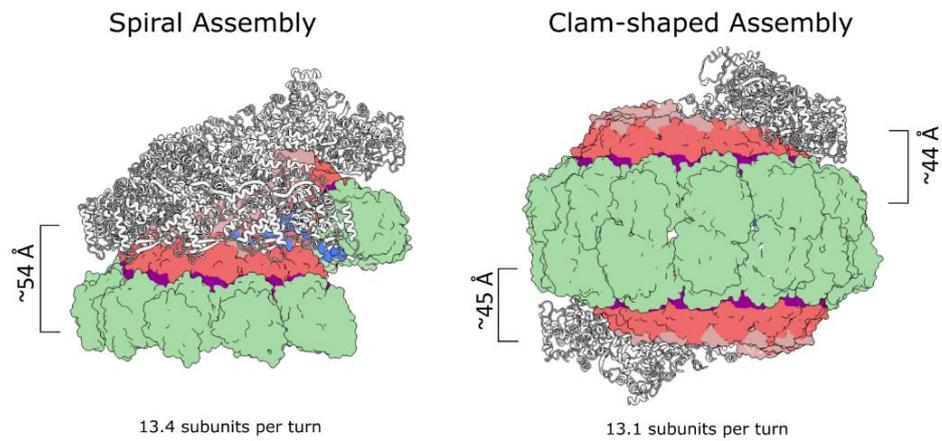
Figure 3.7. Models of the helical and clam shaped assemblies of the NiV N-RNA protein. (A) Ribbon model in white is a duplicate of the helical assembly, generated from the original structure, aligned using Chimera MatchMaker feature, so that its first subunit matches the last subunit of the original structure. The resulting spiral turn was used for calculation of the pitch and the number of subunits per turn.
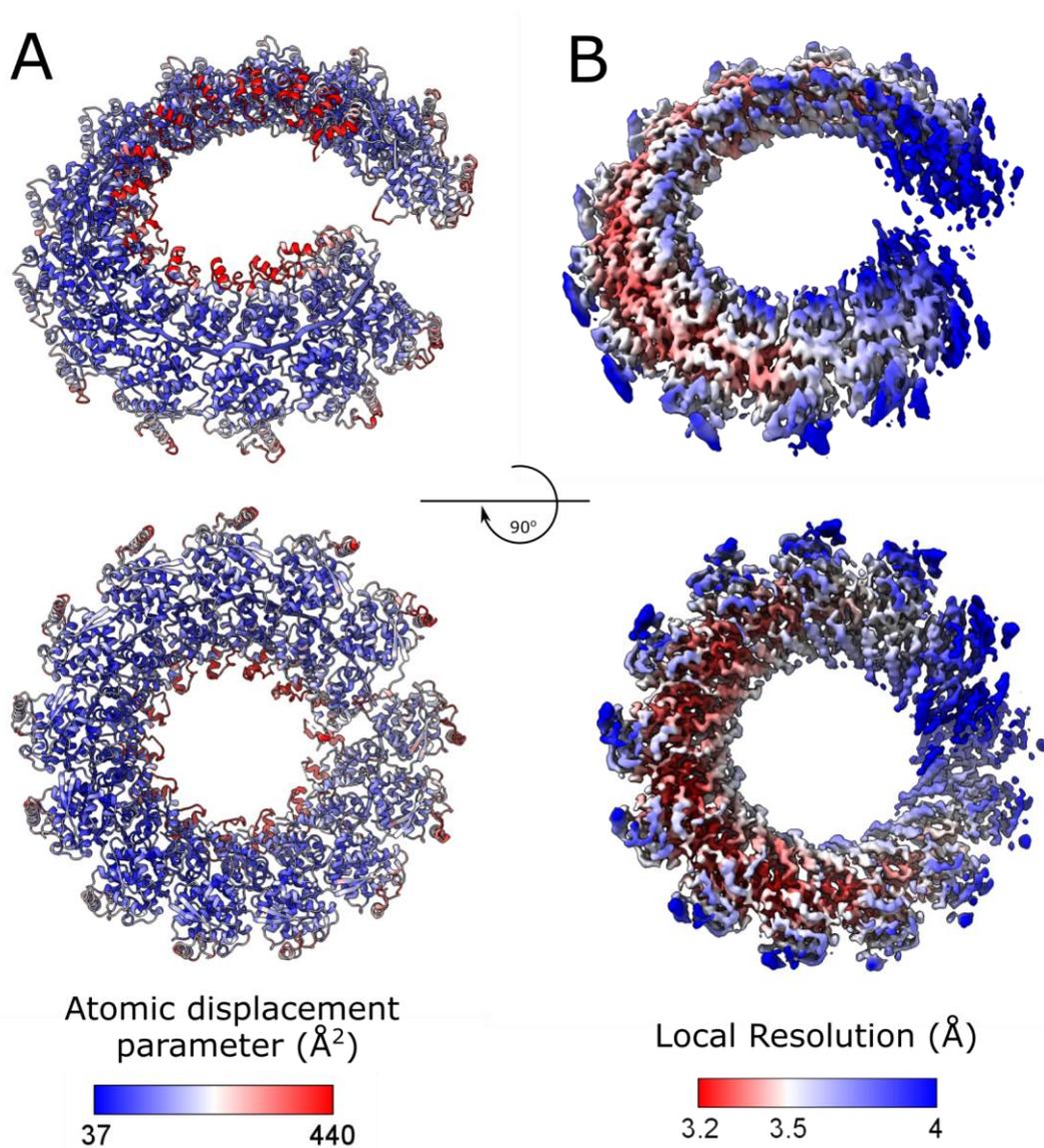
Figure 3.8. Analysis of the NiV N-RNA protein cryoEM model and map. (A) Colour coded atomic displacement parameter (B-factor) distribution of the cryoEM model. (B) Colour coded local resolution distribution for the cryoEM map.

3.2.3. Protomer-protomer interactions within the helical assembly

The spiral assembly of the NiV N-RNA protein is primarily formed through lateral contacts over a calculated interface area of ~3000 $Å^2$ between two adjacent protomers (Table 3.1). The contact area includes a hydrophobic core comprising an aromatic residue (F11) from one protomer and a triad of aromatic residues (F267, F268, Y301) in the adjacent protomer (Figure 3.9). All of these aromatic residues are well conserved in the *Paramyxoviridae* facilitating similar protomer-protomer interactions across all

family members[73]. The NT and CT arms, which have been reported to play a role in the spiral assembly of nucleocapsid[53,145], occupy a hydrophobic groove in the C-terminal N-core domain of the adjacent protomer. In the crystal structure of the NiV N protein monomer (pdb:4co6), the hydrophobic groove is occupied by a 50 amino acid segment of the NiV phosphoprotein (P) which is essential in maintaining the N protein in its RNA-free, monomeric state (Figure 3.9B)[53]. Structures of the N-terminal (residues 4-31) and C-terminal (residues 372-398) segments were not resolved in the crystal structure of the monomeric RNA-free N protein, which was based on a protein construct comprising residues 32-383[53]. Both of these segments, missing in the earlier study, are critical for the oligomer assembly, making bridging interactions with adjacent subunits (Figure 3.9B).

Table 3.1 Interface area between adjacent protomers of the Nipah virus (NiV), Parainfluenza virus 5 (PIV5) and Measles virus (MeV). The interface area was calculated using solvent accessible surface area for the helical assembly of each virus.

	NiV (This study)	PIV5 (PDB:4xjn)[72]	MeV (PDB:6h5q)[113]
Interface area (Å^2)	2998 (100%)	2918 (100%)	2931 (100%)
Polar interface area (Å^2)[†]	2011 (67%)	1931 (66%)	1812 (62%)
Nonpolar interface area (Å^2)[‡]	986 (33%)	987 (34%)	1119 (38%)

[†]Polar interface area is the area contributed by polar interface atoms (N, O, and S containing).

[‡]Nonpolar interface area is the area contributed by non-polar interface atoms (C containing).

A NP$_{n+1}$ NP$_n$ NP$_{n-1}$

N-terminal arm

NP$_{n+1}$ F268 Y301 NP$_n$ F267

F11 I271 M298

B N-terminal arm subunit n+1

C-terminal arm subunit n-1

P protein segment

60°

N-terminal arm subunit n+1

C-terminal arm subunit n-1

C

	11	267 271	298 301
NIPAH	ASFRS	⋯ AGFFATIRF ⋯	SLMLLYRE
MEASLES	ALFKR	⋯ ASFILTIKF ⋯	SLMNLYQQ
ParaInfluenzeVirus5	ERFTL	⋯ GGFFLTLKY ⋯	SLMALYQT
SENDAI	DTFSS	⋯ ASFMNTIKY ⋯	SLIDTYLS
MUMP	ERFTI	⋯ TAFFLTLKY ⋯	SLMMLYRD
NewcastleDiseaseVirus	EQLLA	⋯ TAFFLTLKY ⋯	QLMRLYRM

Figure 3.9 Subunit-subunit interactions within the NiV nucleocapsid protein-RNA assembly. (A) Three adjacent protomers, where the two outside protomers are presented as ribbons and the central protomer is shown in surface representation, with the NT-arm in blue, N-terminal N-core in green, C-terminal N-core in coral and CT-arm in pink, as in Fig. 1(B). A magnified view of the molecular interaction of the NT-arm and the C-terminal N-core domain is shown on the right with interacting residues (sticks) displayed within the CryoEM map. (B) The P protein segment (yellow ribbon) from the structure of the monomeric NiV N complex (pdb:4co6) superimposed onto the central subunit in (A), shown in white surface representation. The two adjacent protomers are presented as dark blue (n+1) and red (n-1) ribbons. (C) Alignment of interacting residue segments, shown in the magnified view in (A), for N proteins from several Paramyxoviruses, with conserved residues highlighted using the ClustalX colour scheme.

3.2.4. Protein-RNA interactions

The CryoEM map shows a clear density for the single stranded RNA, modelled as a poly-uridine chain, wrapped around the nucleocapsid. The RNA molecule is bound to the protein in the classical "3-base-in, 3-base-out" conformation[73], where the RNA chain twists about 180° every 3 nucleotides to place three consecutive nucleotides with the bases facing the protein, followed by 3 nucleotides with exposed bases. The structure shows that the nucleic acid lies within the charged cleft of the N protein at the interface between the N-terminal N-core and the C-terminal N-core domains. This groove is lined by the residue segments K178-Q200 and S344-Y354 that are positioned at the outer edge of the spiral assembly (Figure 3.10). Within the RNA binding cleft, a series of basic (K178, R192, R193, R352) and polar (T181, Q319, S344) residues, with well-defined density, are within hydrogen-bonding distance from the RNA sugar-

phosphate backbone. Residues Q199 and Q200 from helix H8 also have well-defined density, with their side chains projected toward the RNA bases. These two amino acids are conserved in the nucleocapsid proteins of the measles virus (MeV) and parainfluenza virus 5 (PIV5), where they make similar interactions with RNA bases [72,113]. At the interface between the two protomers, aromatic residues Y258 and Y354, one from each adjacent protomer, are positioned in close proximity to the RNA chain (Figure 3.10C), facilitating the twist in its conformation. This twist in the sugar-phosphate backbone is assisted by a series of additional protein-RNA interactions contributed by polar residues lining the RNA binding cleft. The second twist in the RNA conformation, spaced by three nucleotides from the first one, occurs in a cleft within a single protomer and is facilitated by steric hindrance from the side chain of L348 (Figure 3.10D). The majority of residues interacting with the RNA within the RNA-binding cleft are highly conserved among Paramyxovirus N proteins (Figure 3.10E) indicating a similar mechanism for RNA binding.
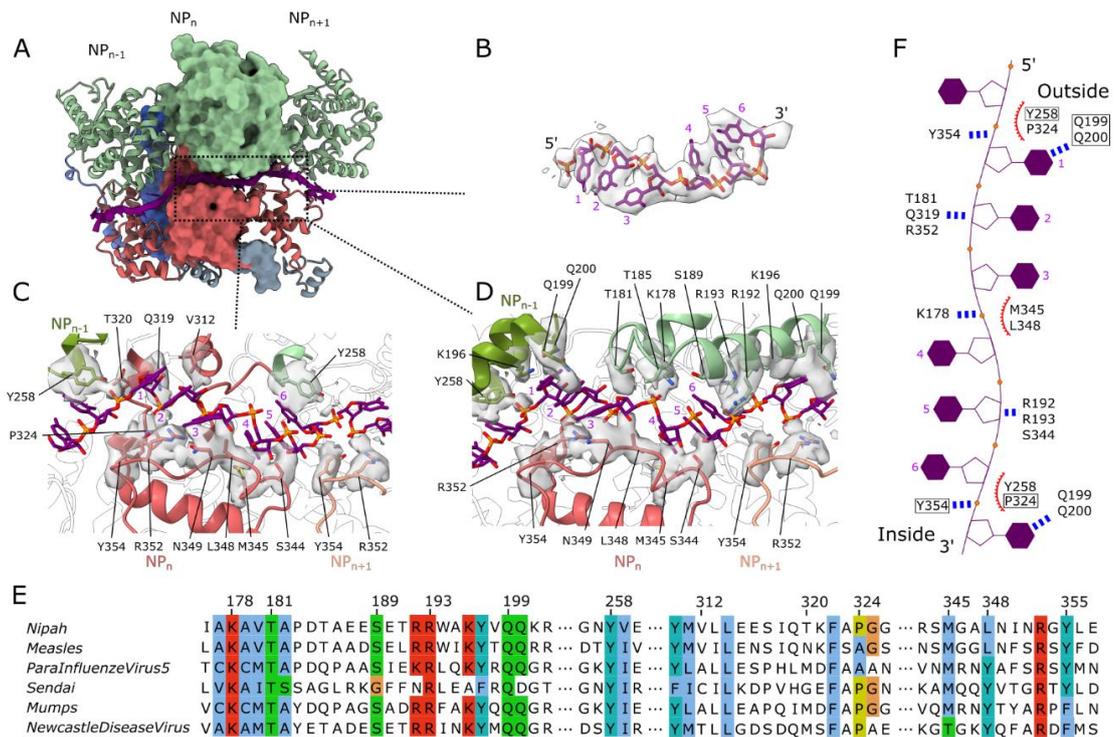
Figure 3.10 Protein-RNA interactions. (A) Three adjacent protomers shown as in Figure 3.9, with RNA (purple) shown in ribbon and sticks. (B) CryoEM map corresponding to the RNA, with a fitted poly-uridine RNA model (sticks). (C-D) Two different views at the protein-RNA interface, detailing protein-RNA interactions. CryoEM density corresponding to the side chain atoms of interacting residues, is shown with a 3 Å distance cut-off. (E) Alignment of RNA-binding residue segments from several Paramyxoviral N proteins with conserved residues highlighted using the ClustalX colour scheme. (F) Schematic of the ssRNA conformation in complex with the N protein. The residues in close proximity to RNA are labelled. Boxed residues indicate those from the neighbouring protomer; the blue dotted lines indicate putative hydrogen bonding interactions; red curves indicate putative hydrophobic interactions.

Several residues connecting the two RNA binding segments (K178-Q200 and S344-Y354) with the rest of the N-protein are poorly defined. It is likely that the flexible nature of these regions serves to provide plasticity to accommodate and interact, in a non-sequence specific manner, with the varying sequence along the entire length of the RNA strand. This flexibility may also allow the RdRp complex to access the ssRNA while bound within the nucleocapsid assembly, for RNA synthesis.

In the crystal structure of the RNA-free monomeric NiV N protein, the flexible loop, residues A180-R192, was mostly disordered and positioned such that it would block access to the RNA binding cleft, suggesting that this loop needs to move out of the cleft to permit RNA binding. As seen from structure comparison, RNA binding is also accompanied by an approximately 28° rotation of the N-terminal and C-terminal N-

69

core domains towards each other, around a hinge region formed by the H12-H13 loop[53], H15-H16 loop and helix H17 (Figure 3.11). Similar conformational changes have also been observed for the nucleocapsid of MeV[73].
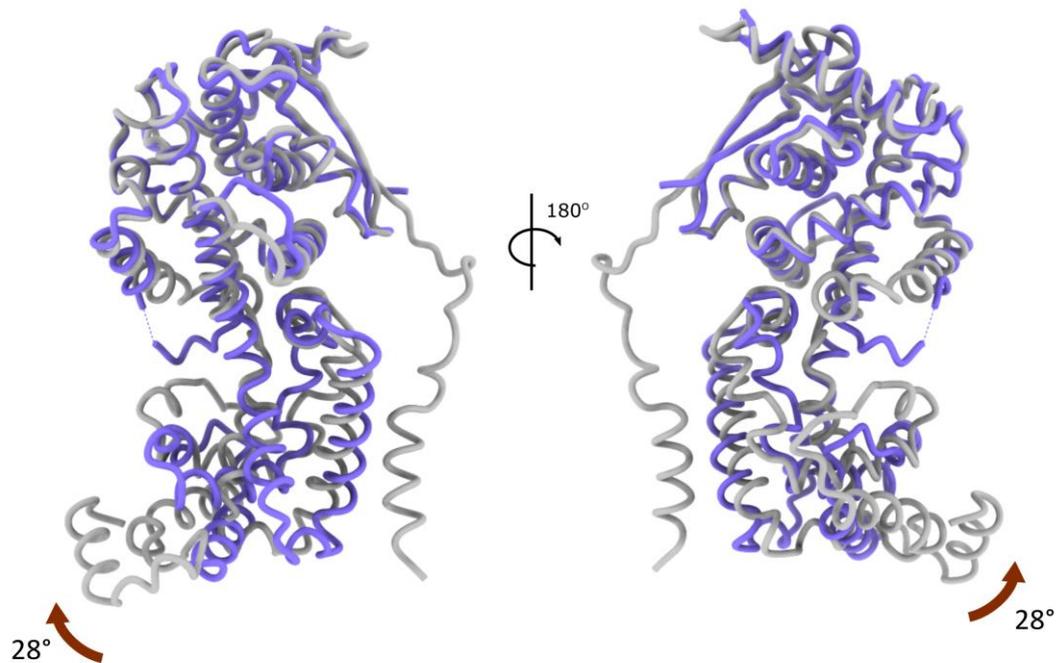
Figure 3.11. Comparison of the Nipah N protein structure in the RNA-free and RNA-bound states. Superimposed models are presented as cartoons. The RNA-free N protein (pdb:4co6) is in light purple, while the RNA-bound protein is in light grey.

3.2.5. Comparison with the structure of N protein from other Paramyxoviruses

Within the Paramyxovirus family, the NiV N protein shares about 56%-92% sequence identity with other members of the *Henipavirus* genus, with the N proteins from Hendra virus (92% sequence identity) and Cedar virus (59% sequence identity) being the closest relatives (Table 3.2). Mapping these sequence differences to the structure (Figure 3.12) shows that the RNA-binding surfaces are the most conserved, not only among the Henipaviruses, but also for nucleocapsids of more distantly related viruses such as Measles. In contrast, the most variable region is the N-terminal N-core domains of N protein, particularly within the 51-amino acid segment 106-156. This segment has a well-defined structure, but its conformation displays significant variability among the paramyxovirus family members, with considerable differences in the position and length of helix H5, residues 123-139 in NiV N (Figure 3.12, inset).

Within the N-terminal N-core domain, the closely related NiV and Hendra virus (HeV) N proteins contain four amino acid substitutions: V70I, V108L, D137E and I236M (Figure 3.13), with V108 and D137 being exposed at the outermost edge.

Table 3.2 Pairwise sequence identity between N protein of different Paramyxoviral N protein. The genus for each virus is indicated in brackets.

	NiV	HeV	CeV	NarV	BeiV	RPV	MV	PIV5
NiV (*Henipavirus*)								
HeV (*Henipavirus*)	92.1							
CeV (*Henipavirus*)	59.4	59.8						
NarV (*Narmovirus*)	37.2	37.4	34.9					
BeiV (*Jeilongvirus*)	33.7	33.6	34.9	39.6				
RPV (*Morbillivirus*)	33.5	33.0	32.1	37.5	35.9			
MV (*Morbillivirus*)	32.1	32.5	32.8	38.7	36.3	74.5		
PIV5 (Orthorubulavirus)	27.9	27.7	28.1	28.5	26.4	27.3	26.3	
NDV (Orthoavulavirus)	28.8	28.7	28.3	26.6	27.2	25.2	25.5	34.6
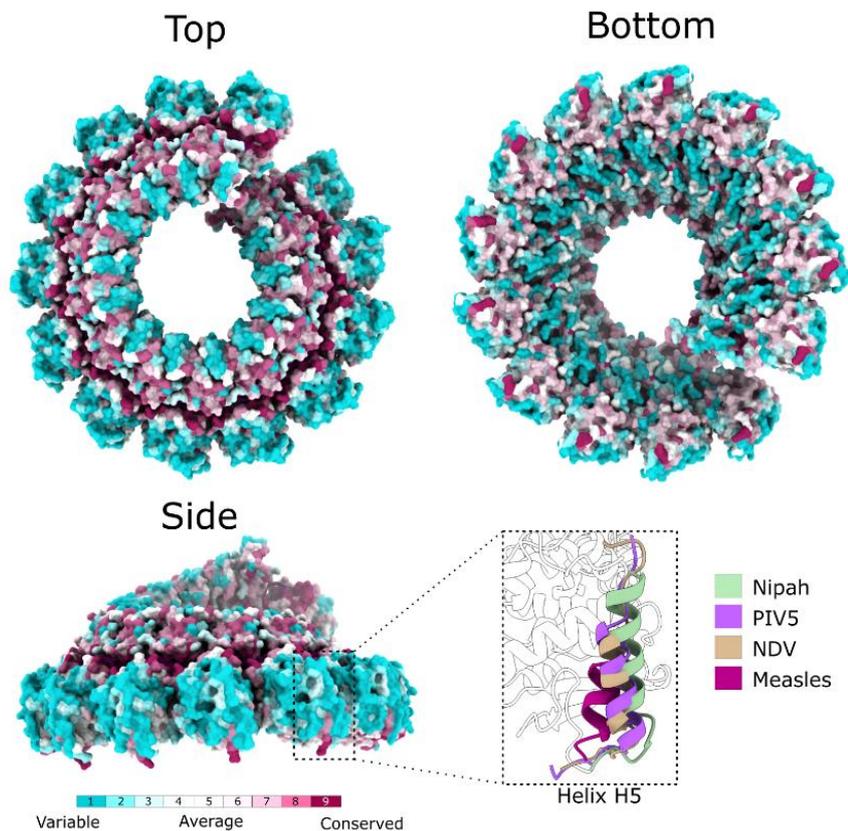
Figure 3.12 Mapping sequence variation to the structure. The surface of the NiV nucleocapsid protein-RNA assembly is coloured according to sequence conservation among Paramyxoviral N proteins based on ConSurf analysis, visualised at three different orientations. Inset (bottom right) compares the conformation of the residue segment 118-139 containing helix H5, in four different Paramyxoviruses: NiV, Parainfluenza virus 5 (PIV5, PDB code 4xjn), Newcastle Disease virus (NDV, PDB code 6jc8) and Measles (PDB code 6h5q).
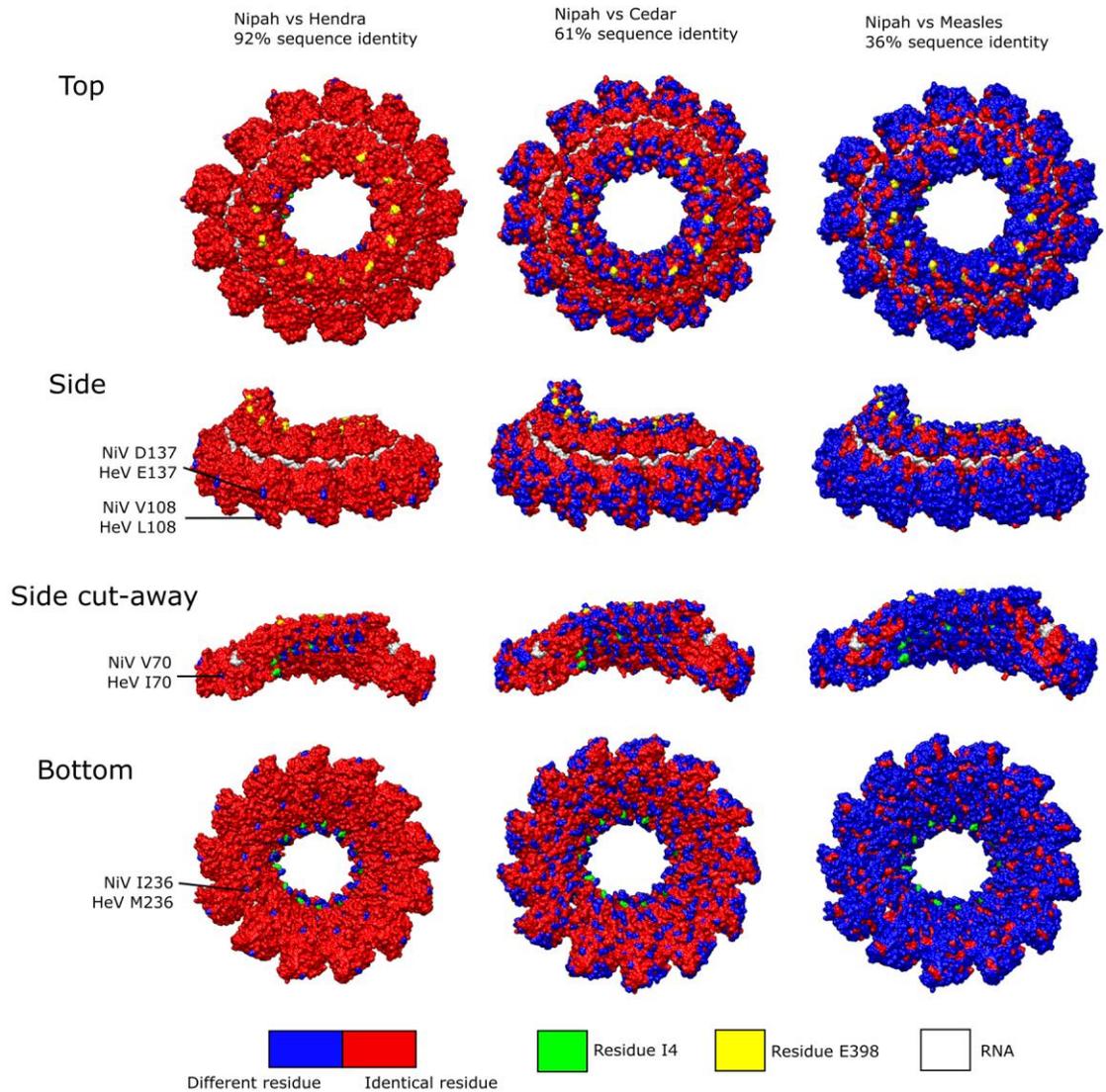
Figure 3.13 Sequence differences between Nipah virus (NiV), Hendra virus (HeV), Cedar virus (CeV) and Measles virus (MV) mapped onto the Nipah N-RNA protein assembly surface. Amino acids differing between the NiV and HeV N proteins, are labelled.

3.2.6. Clam-shaped assemblies of recombinant NiV N protein

Aside from the common helical assembly, about 35 % of N-RNA protein particles were found as clam-shaped assemblies which can be further subclassified into two distinct primary conformations, a spiral clam-shaped assembly, and a semi-spiral clam-shaped assembly. The spiral clam-shaped assembly is composed of two N-RNA protein spirals stacked face to face, as seen for the Newcastle Diseases virus (NDV) N-RNA protein assembly[134]. In contrast, the semi-spiral clam-shaped assembly features one 14-subunit ring and one 13-subunit N protein spiral stacked as shown on Figure 3.14. Asymmetric reconstruction of both assemblies leads to 4.3 Å and 5.2 Å resolution maps, respectively. Models for the spiral assemblies were built by rigid-body fitting

and real space refinement of the N protein protomer taken from the protein-RNA complex described above. The structure of the N protein monomer within these clam-shaped assemblies remains largely the same as in the spiral assembly, with an overall RMSD of 0.9 Å calculated over Cα atoms.
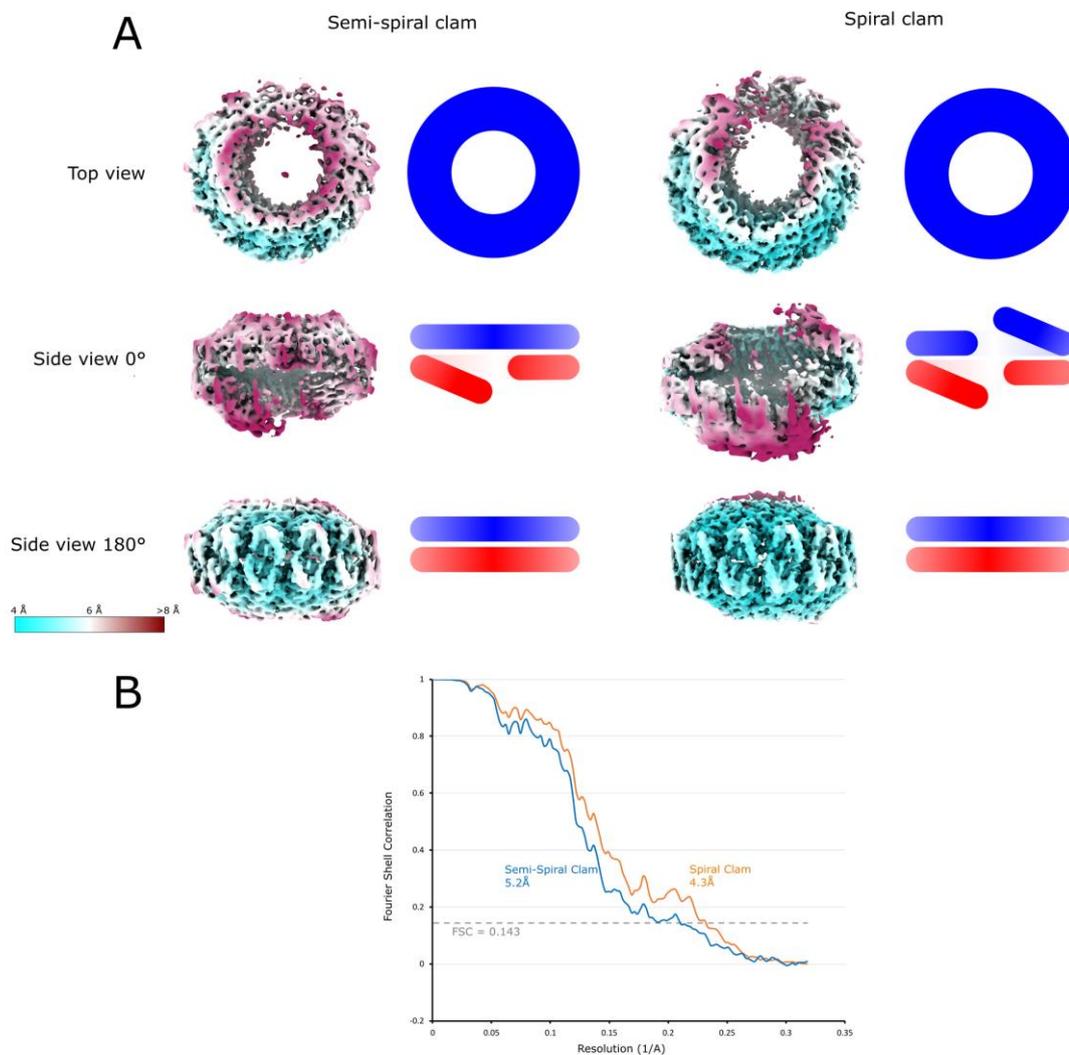
Figure 3.14 CryoEM maps for the two major types of clam-shaped assemblies. (A) Local resolution of CryoEM maps for each of the two assemblies are shown in three different views. Cartoon illustrations for each view are also presented. (B) "Gold-standard" FSC plots for the semi-spiral clam and spiral clam-shaped assembly.

For both assemblies, there is a significant surface area buried at the interface between the two halves of the clam shell, with up to ~670 $Å^2$ of buried area per monomer (Figure 3.15A). Interactions across this interface are mediated by hydrogen bonding and polar interactions made by loop segments A1-H2, A2-H5 and H6-H7 (Figure 3.15B) from each opposing protomer. The surface area, buried at the clam-shell interface of each monomer, is about five times larger than seen in the clam-shaped assembly of

NDV, where only one protein loop (residues 104-124) is involved in the interaction[134]. This likely creates a significantly closer interaction between the halves of the NiV clam shell as compared to the NDV assembly. Indeed, this tight interaction has also been observed in the recent clam-shaped assembly from Sendai virus (SeV)[146]. Interestingly, although sequences of these clam-shell interface loops are not conserved between NiV, NDV and SeV, or other members of the *Paramyxoviridae*, these loops are rich in glycine and proline residues (Figure 3.15C). Proline and glycine can facilitate protein dynamics as flexible hinges and confer conformational flexibility upon the protein regions that incorporates them[147].
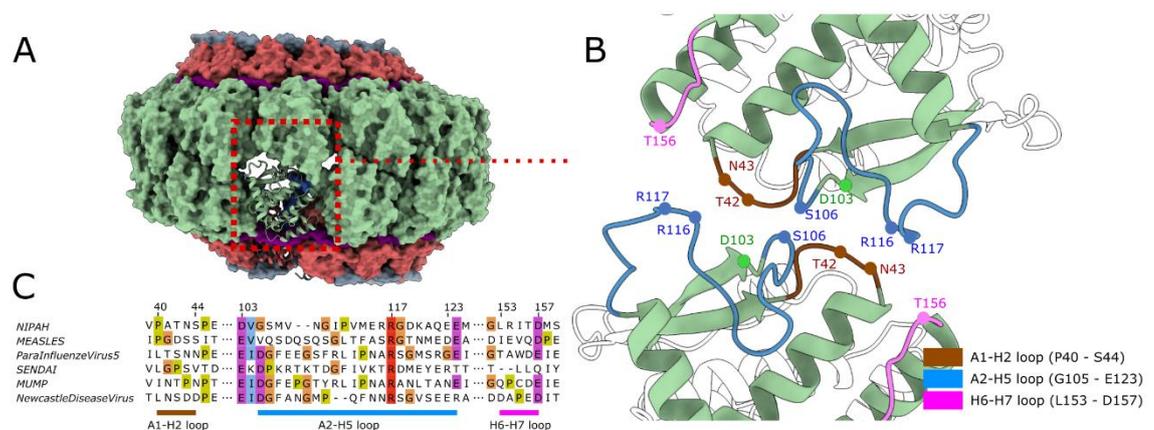
Figure 3.15 Clam-shaped nucleocapsid assembly. (A) Model of the clam-shaped nucleocapsid assembly, presented as a molecular surface, showing the interaction between two opposing, top and bottom N-RNA protein spirals, with one of the protomers shown as ribbon. (B) The clam-shaped interaction is primarily mediated by protein loops from the N-terminal N-core domain. Putative residues involved in the interaction between the two halves of the shell are indicated. (C) Alignment of the three interacting loop regions highlighted in (B) for N proteins from several Paramyxoviruses, with glycine, proline and surrounding conserved residues highlighted using the ClustalX colour scheme.

3.3   Discussion

Formation of a helical RNP assembly, that safeguards the viral genome and serves as a template for RNA replication, is a unifying feature of negative strand RNA viruses[148]. When the NiV nucleocapsid proteins are expressed recombinantly in the *E. coli*, it is likely that the nucleocapsid-like helical assembly is formed with the E. coli cellular RNA. However, it is also possible that the N protein binds to the NiV N gene transcript, as it is expected to be one of the highly abundant unstructured RNA in the *E. coli* cell, due to the T7 expression system[149]. While the encapsidation specificity of the N protein is currently not well understood, it is likely that the N protein does not encapsidate the

mRNA in order to allow the mRNA to be available for translation by the ribosome. The exclusion of the mRNA encapsidation by the N protein can be done through detection of the 5' cap of the nascent mRNA by the RdRp complex. As the prokaryotic mRNA lacks the 5' cap found in the eukaryotic mRNA, it is therefore possible for the N protein to encapsidate the NiV N mRNA during the recombinant expression in *E. coli*.

Here, we determined the CryoEM structure for the assembly of full-length recombinant NiV N protein with RNA, presenting the first high resolution structural data on the nucleocapsid assembly for this virus and the *Henipavirus* genus. The RNA strand, accommodated in the groove between the N- and C-terminal N-core domains, binds with the bases of consecutive nucleotide triplets alternatively exposed and buried (Figure 3.10). The spiral assembly is stabilised, not only by the protein-RNA interactions, but also by subunit-subunit interactions between contiguously bound N proteins (Figure 3.9).

Comparison with the structure of an RNA-free monomer in complex with a segment of the P-protein[73] indicates a mechanism by which the P-protein can modulate assembly of the nucleocapsid during viral replication. The P-protein segment is bound in a groove which is occupied by the N-terminal helix of one adjacent subunit and a C-terminal region of another adjacent subunit. Thus, binding of the P-protein would directly compete with the nucleocapsid assembly formation, as has been proposed earlier on the basis of the nucleocapsid-RNA structure of the measles virus[73].

The helical assembly has 13.4 subunits per turn, in common with Measles, Parainfluenza virus 5, and Newcastle disease virus, the three distant homologues of *Paramyxoviridae* with available structures of nucleocapsid assemblies[72,73,134] in which NiV N protein shares 32%, 28% and 29% sequence identity (Table 3.2), respectively. Areas with the highest sequence conservation are found at the RNA-binding surfaces and also at subunit-subunit interfaces, with the outer exposed surfaces of the nucleocapsid displaying the most sequence variation (Figure 3.12).

The oligomeric assembly, protein-RNA interactions and the conformation of bound RNA are conserved among the different genera of *Paramyxoviridae*[113,134] (Figure 3.16). Moreover, the N protein of Ebola virus, a representative of *Filoviridae*, binds RNA in a conserved manner, wrapping it around the outer edge of the oligomeric N protein

helix, with similarities observed in protein-RNA interactions and even in the conformation of the bound RNA[150,151] (Figure 3.16). N proteins of other *Mononegavirales*, from families that are more distant than *Filoviridae*, also display a highly conserved protein fold, with the highest similarity observed in the N-terminal domain and more limited, but still detectable, fold similarity within the C-terminal domain[69]. Available structural information indicates that although N proteins of all *Mononegavirales* bind RNA within a groove between adjacent domains, a significant variation is observed in the number of RNA nucleotides bound per protein protomer (6 to 9 nucleotides) and also in the relative arrangement of subunits, where the number of subunits per helical turn can vary from 10 to 24, depending on the virus. Most strikingly, some viruses, like rabies virus and vesicular stomatitis virus, members of *Rhabdoviridae* family, bind the RNA on the inner edge of helically assembled N protein[152,153], with the N protein tilted slightly away from the central helical axis[154,155]. This is in contrast to the viruses of the *Paramyxoviridae*[73], *Filoviridae*[150,156] and *Pneumoviridae*[157,158] families, where the RNA is wrapped around the outside of helical assembly. It is interesting to note that, compared to *Paramyxoviridae* and *Pneumoviridae*, *Rhabdoviridae* also adopt a different strategy to inhibit non-specific RNA binding of its N protein, by inserting the P protein directly into the RNA binding groove [53,159,160].
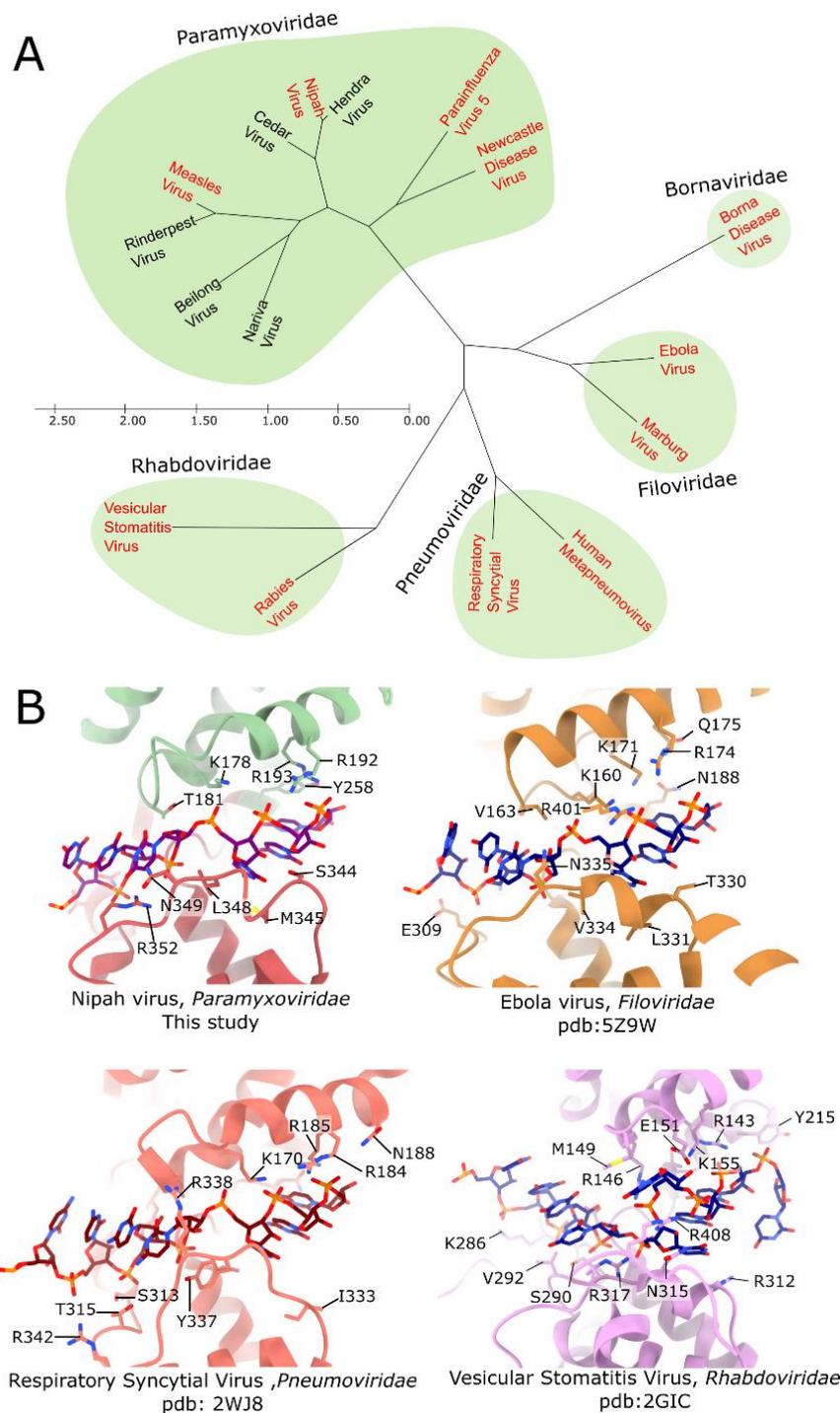
Figure 3.16 Phylogenetic relation of the N protein within the Mononegavirales. (A) Phylogenetic tree based on the N protein sequences of selected Mononegavirales. N proteins with available protein structures are coloured in red. (B) Zoomed-in view of RNA binding region of Nipah virus, Ebola virus Respiratory Syncytial Virus, and Vesicular Stomatitis Virus showing the conserved protein-RNA interaction among these viruses. Putative residues involved in the protein-RNA interaction are shown as sticks.

Apart from significant differences between the nucleocapsid assemblies of different members of the *Mononegavirales*, structural differences are observed also for individual viruses. In particular, for Paramyxoviruses, variations have been observed in

the helical pitch and diameter of the assembly, depending on the pH and salt concentration[140], as well as, the presence of the C-terminal disordered N-tail region[161]. Likewise, the recombinant full length NiV N-RNA protein helical assembly is highly flexible, with further 3D classification showing significant conformational and compositional heterogeneity in the NiV N-RNA protein spiral assembly within this CryoEM dataset. While most of the particles (roughly 85%) represent assemblies with a single helical turn containing 13 subunits, there is a smaller proportion of particles containing a larger number of subunits with up to ~2.5 helical turns. Further classification of particles revealed notable variation in the helical pitch, ranging from 53 Å to 55 Å and ~13.4 subunits per turn of the helix (Figure 3.17). For comparison, the Measles nucleocapsid protein, the closest homologue with an available structure, forms helical assemblies with a 50-66 Å pitch and 12.8-13.5 subunits per turn[161]. Such flexibility in the nucleocapsid assembly may facilitate conformational rearrangements required for RdRp complex access to the RNP during mRNA transcription and genome replication, and/or for binding of host proteins that regulate this process.

The viral genomic RNA within the RNP assembly, despite being protected from degradation by nucleases, is still accessible to the viral RdRp complex, suggesting a yet to be understood mechanism that governs encapsidation while allowing transcription. These regulatory roles may be facilitated by the C-terminal disordered N-tail region. In most Paramyxoviruses, this C-terminal disordered N-tail region has been shown both to be the site for P protein interaction[56,162] and to play a role in regulating transcriptional activity[92,93]. Furthermore, viral RNP genomes are selectively incorporated into budding virion particles[163,164], and this is likely driven by the interaction between the N protein N-tail region with M proteins[57]. In the NiV N-RNA protein, the last defined residue in the structure, E398, although situated at the inner core, points towards the outer edge of the assembly, suggesting that the disordered N-tail region, residues 399-532, could be at least partially exposed, consistent with its potential role in facilitating RdRp complex and M protein interaction during viral replication cycle.

Besides the typical spiral assembly, particles that resemble a clam-shaped assembly are also present, with a calculated pitch of ~45 Å with ~13.1 subunits per turn and at least two distinct conformations observed (Figure 3.7). Since the clam-shaped

assembly has not yet been observed during replication *in vivo* for NiV, it is also possible that the formation of these assemblies was induced by the conditions used during production of recombinant N-RNA protein. However, the consistent presence of such assemblies during purification of the NiV nucleocapsid (Figure 3.1C), indicates an ordered, stable complex, suggesting that they may indeed serve some, as yet unidentified, biological purpose. Indeed, the N protein is known to bind to a short leader transcript and thought to form short ring-like assemblies during the very early stage of viral mRNA transcription and protein translation[74,165]. Two of these short ring-like assemblies may assemble in a face-to-face manner to form a highly stable complex. It is also possible that the interactions underpinning the clam-shaped assembly serve to protect the 5' end of the genomic nucleocapsid through interaction with an N-bound-leader RNA ring, as such assemblies have been purified from isolated MeV virions[165]. Similarly, these assemblies may also facilitate the formation of circular nucleocapsid that was observed in recombinant MeV nucleocapsid[161], as well as lead to the budding of virions which contain multiple RdRp-nucleocapsid assemblies[166]. Similar clam-shaped assemblies were observed for NDV and SeV (Figure 3.18), both for recombinantly produced protein and for nucleocapsids purified from virions, and it has been proposed they may act as a seed for formation of a double headed spiral assembly[134,146], although the possibility cannot be excluded that these NDV and SeV assemblies may have formed from damaged fragments of virion nucleocapsids[167]. Previous mutational analysis, in NDV and SeV, of residues corresponding to the A2-H5 loop (Figure 3.15B), which facilitated formation of the clam-shaped assembly, has resulted in up to 100% loss of *in vitro* replication while retaining the helical assembly capability of the N protein[134,168]. Nevertheless, as NiV N shares only 28% and 29% sequence identity with the NDV and SeV nucleocapsids, respectively, the observation of similar clam-shaped assemblies indicates their potential biological significance, necessitating further research.
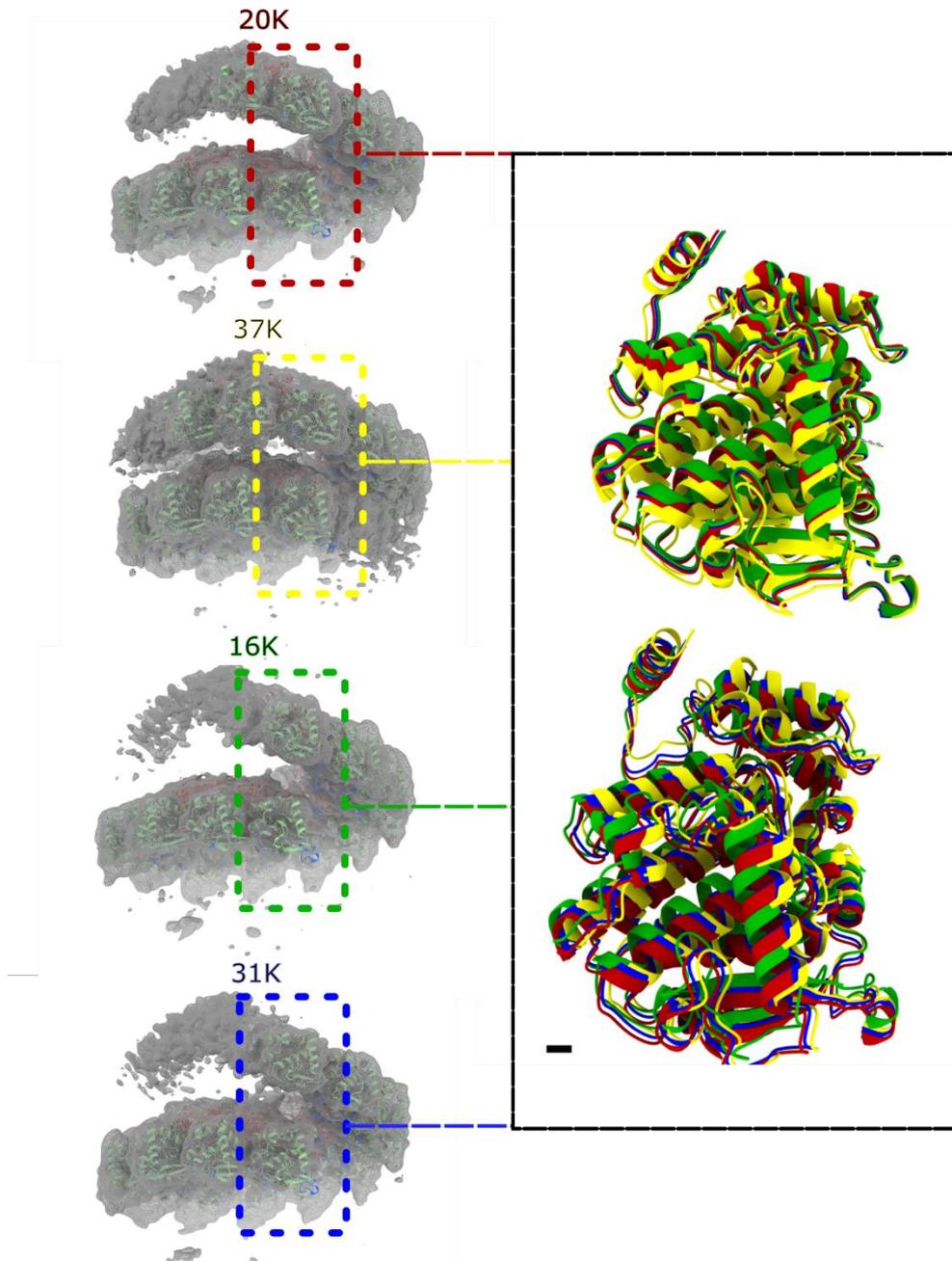
Figure 3.17 CryoEM maps of the top four 3D classes from the classification of NiV N protein helical assembly (Figure 3.4), shown along with ribbon diagrams of fitted N protein subunits. Number of particles contributed to each respective class is indicated above each model. Overlay of all the fitted NiV N protein models reveals a subtle variation in the seam region of the helical turn. Overlay of single subunits from the four different assemblies, calculated and shown for two single subunits of each subunit, taken from two different positions of the helical assembly. Scale bar, 1 Å.
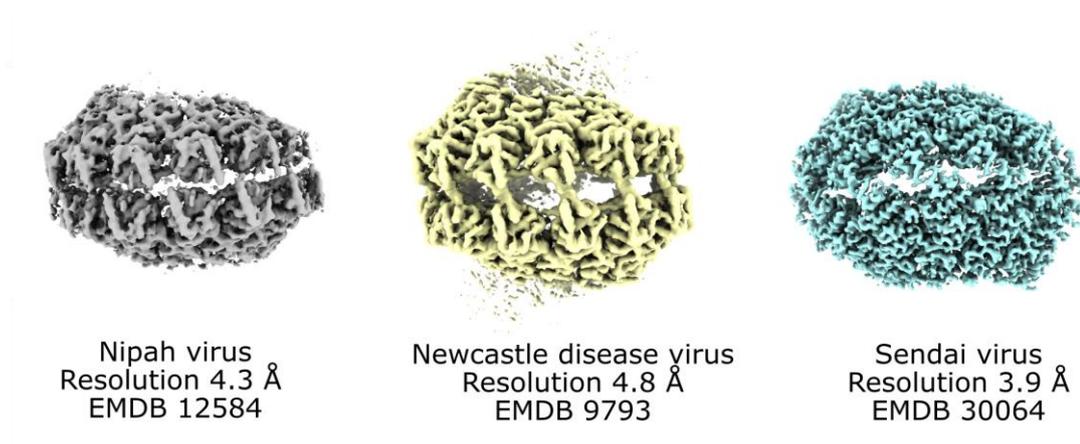
Figure 3.18. Comparison of clam-shaped assemblies formed by the Nipah virus, Newcastle disease virus, and Sendai virus

The first high-resolution structure of the Nipah nucleocapsid assembly reported here, determined in complex with RNA, will inform the design of inhibitors that disrupt subunit-subunit or protein-RNA interactions. Future studies on the N-RNA protein interaction with other factors such as the P[56,162] and M[57] proteins will allow understanding of the full scale of molecular events that occur during nucleocapsid assembly and viral replication.

Chapter 4 Interaction of the Phosphoprotein with the Nucleocapsid-RNA assembly

4.1 Introduction

Viral transcription and genome replication are essential processes in the RNA virus life cycle. Using the viral genome as the template, viral transcription and viral replication are initiated to generate viral protein and genome for the new progeny virus. In Paramyxovirus, the initiation of the RNA synthesis is executed by the binding of the RdRp complex to the *le* promoter in the 3' end of the viral RNP genome. This is followed by the localised uncoating of the viral genome from the nucleocapsid (N) protein to allow the RNA polymerase (L protein) to access and synthesise new viral RNA. However, how the L protein attaches to the 3' end of the viral RNP genome remains an enigma.

The phosphoprotein (P protein) is an important co-factor protein in the viral life cycle and has been shown to interact with N protein and L protein (see section **1.5** above). Therefore, it is likely that the P protein plays the role of tethering the N protein and the L protein together. However, despite the wealth of knowledge about the region of NiV P protein that is crucial for RdRp complex activities, it remains unknown what structural elements of NiV P are involved in the binding to NiV helical RNP assembly. This chapter reports the structural investigation of the protein-protein interaction between the NiV nucleocapsid-like particles and NiV P protein using electron microscopy. The information reported here provides preliminary insight into the mechanism regulating the viral RNA synthesis.

4.2 Results

4.2.1. Expression and Purification of NiV P protein

The full-length P protein was successfully expressed as a soluble protein in *E. coli* when induced by IPTG at 37°C (Figure 4.1A). Interestingly, SDS PAGE analysis showed the NiV P protein to be about 20 kDa larger than its calculated molecular weight (78 kDa). A similar phenomenon has been observed in previous work on the NiV P protein[56,169],

and it is likely that the NiV P protein experiences a retarded mobility in SDS PAGE and therefore ends up being interpreted as being larger than its calculated molecular weight. Protein sequence analysis of the NiV P protein shows that the protein has a calculated pI of 4.39, with about 18% of acidic amino acids (Asp/Glu). The IUPred3 analysis also shows that the first 450 aa region of the P protein is predicted to be disordered (Figure 4.1B), and this is in agreement with previous NMR studies on the NiV P protein[170,171]. The highly charged protein surface is known to be part of the contributing factors for the retardation of protein mobility in SDS-PAGE. Likewise, several proteins such as Gir2, an acidic protein from yeast[172], and Def, a nucleolar protein from zebrafish[173], have been reported to run higher than their expected molecular weights. In both of these examples and in the case of NiV P protein, the apparently higher molecular weight is likely the result of poor interaction with the SDS due to the highly charged residues, resulting in slower migration during the electrophoresis.
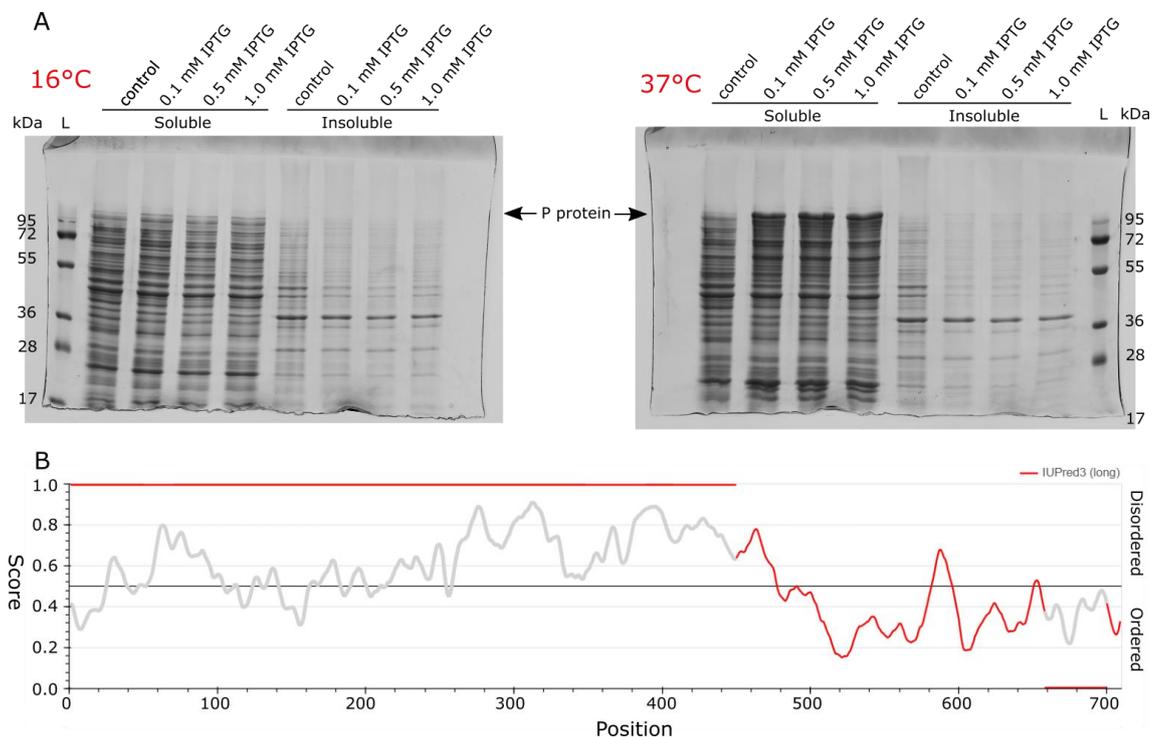
Figure 4.1 Recombinant expression of the Nipah virus phosphoprotein. (A) Expression of the NiV P protein in *E. coli* at 16°C and 37°C. Three different concentrations of IPTG (0.1mM, 0.5mM, 1.0mM) were tested for induction of protein expression. (B) Disorder prediction of NiV P protein using IUPred3. The IUPred3 returns a score between 0 and 1 for each residue. Red lines at 0 and 1 annotate those regions as true ordered or true disordered, based on experimental data, while the grey line represents the prediction score of IUPred3 on those corresponding regions.

Initial purification of the NiV P protein showed that the P protein bound to some nucleic acid as judged by the A_{260}/A_{280} ratio UV absorbance. Furthermore, the P protein was eluted in SEC as a macromolecule larger than the protein standard thyroglobulin (669 kDa, 86 Å) (Figure 4.2A). The molecular weight of the monomeric P protein is calculated to be 78 kDa and it has been reported to form a ~312 kDa tetramer in solution[174,175]. SEC separates the protein based on its hydrodynamic radius, suggesting that the tetrameric NiV P having a hydrodynamic radius larger than 86 Å, compared a typical globular thyroglobulin due to its highly disordered nature. Interestingly, attempts to remove the bound nucleic acid from P protein using high salt (1M NaCl) were not successful (Figure 4.2B). Instead, an A_{260}/A_{280} ratio higher than 1.0 was observed, suggesting that at high salt concentration, the P protein binds more favourably to nucleic acid. To verify that it was indeed the presence of nucleic acid that contributed to the increase of the 260 nm absorbance, the purified P protein was subjected to 12% native PAGE followed by ethidium bromide post staining (Figure 4.2C). The ethidium bromide staining clearly indicated that it was the presence of nucleic acid in the sample that contributed to the high A_{260}/A_{280} ratio observed during SEC purification (Figure 4.2C). While the nucleic acids were shown to have migrated toward the anode, the proteins were mostly remained in the native PAGE wells. The absence of co-migration indirectly suggested a weak interaction between the NiV P and nucleic acid.

To further purify and improve the homogeneity of the P protein sample, the P sample was desalted and subjected to an additional purification step using a Heparin column. This additional step removes most of the nucleic acid and resulted in a more homogenous NiV P protein as judged from the SEC profile (Figure 4.3). Interestingly, the elution volume of the NiV P protein in SEC was the same regardless of the presence or absence of nucleic acid, suggesting that the binding to nucleic acid did not result in any major conformational changes particularly in the N-terminal disordered region of the NiV P protein, as the overall hydrodynamic radius of the NiV P remained the same. As the NiV P protein was eluted as a large sized protein from the SEC, the NiV P protein was subjected to negative stain TEM imaging. However, the TEM imaging revealed no visible structured protein particles in agreement with its highly disordered nature (Figure 4.3).

Despite the addition of protease inhibitors during the protein purification process, multiple bands of the P protein were still observed by SDS PAGE indicating the high susceptibility of the NiV P protein to degradation. In particular, after overnight His tag cleavage using 3C protease, a significant portion of degraded P protein was eluted at a size of ~100 kDa in SEC, which are shown as multiple bands sized around 60 kDa in SDS PAGE. The identity of both the 100 kDa and 60 kDa protein bands in SDS PAGE as the NiV P protein were further confirmed by mass spectrometry following in-gel trypsin digestion (Figure 4.4). The mass spectrometry sequence coverage for both of the protein bands is very similar, in which it has about 15% coverage. However, for the truncated NiV P protein bands, only peptides mapped to the 450 aa disordered region were detected. This suggested that the 60 kDa truncated NiV P protein is mostly disordered lacking a tertiary structure[170,171].
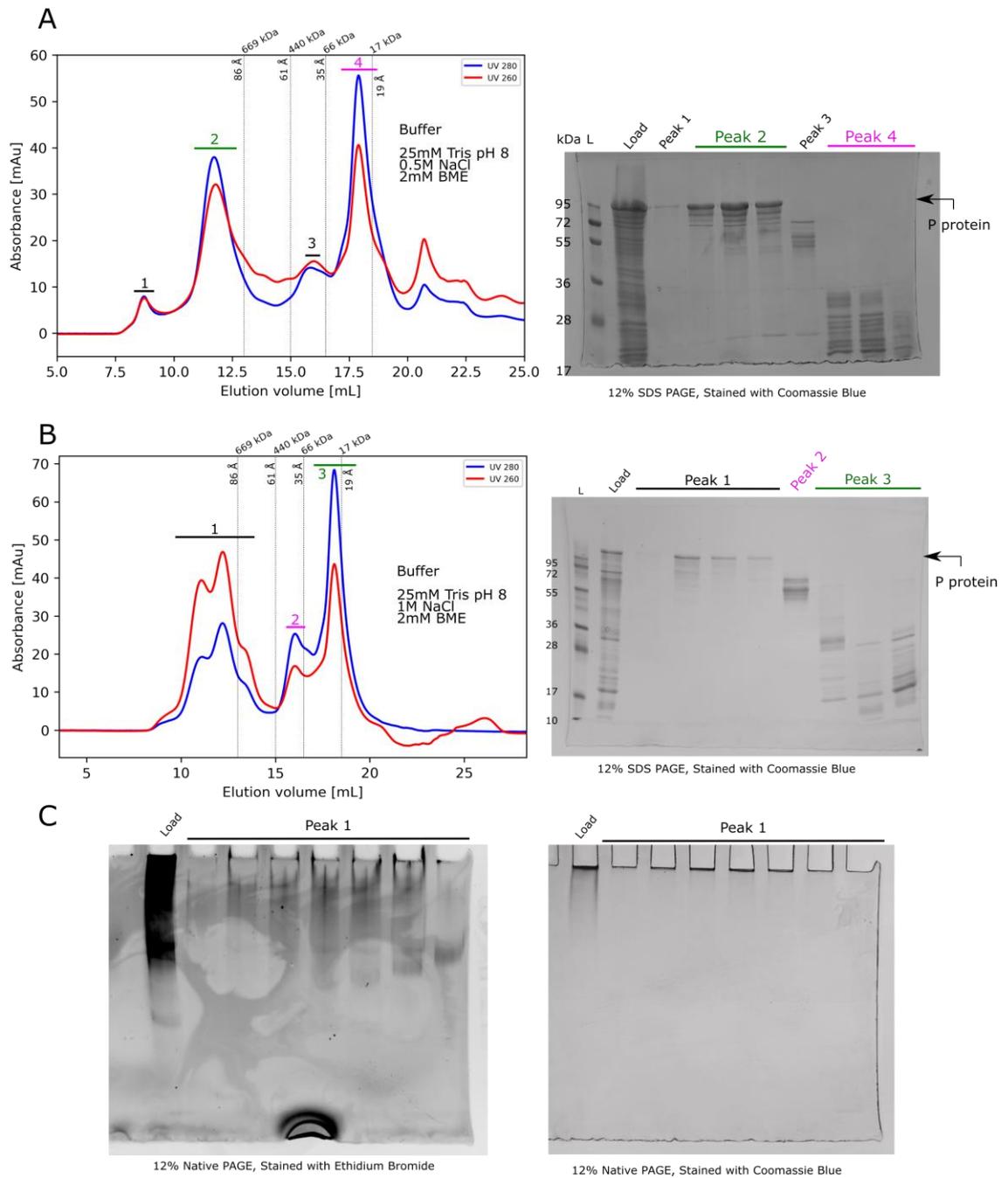
Figure 4.2. NiV P protein purification. The NiV P protein was subjected to SEC purification using a Superose 6 Increase 10/300 column in 0.5 M NaCl (A) and 1.0 M NaCl (B) after the Nickel affinity purification step. (C) The fractions from Peak 1 from (B) were further subjected to 12% Native PAGE separation for 1 hour at 150 V. The Native PAGE was first stained with ethidium bromide (left) followed by Coomassie blue (right).

Figure 4.3 NiV P protein purification using the heparin column. SEC profile of NiV P protein using a Superose 6 Increase 10/300 column after Nickel affinity purification and Heparin purification is shown. The TEM image is shown for the negatively stained sample corresponding to the fraction from peak 3. The red circles designate the putative NiV P particles.

A

Buffer
25mM Tris pH8
150mM NaCl
1mM DTT

Protein sequence coverage: 17%
Matched peptides shown in **bold red**.

```
  1 MGSSHHHHHH SSGLEVLFQG PAMDKLELVN DGLNIIDFIQ KNQKEIQKTY
 51 GRSSIQQPSI KDQTKAWEDF LQCTSGESEQ VEGGMSKDGG DVERRNLEDL
101 SSTSPTDGTI GKRVSNTRDW AEGSDDIQLD PVVTDVVYHD HGGECTGYGF
151 TSSPERGWSD YTSGANNGNV CLVSDAKMLS YAPEIAVSKE DRETDLVHLE
201 NKLSTTGLNP TAVPFTLRNL SDPAKDSPVI AEHYYGLGVK EQNVGPQTSR
251 NVNLDSIKLY TSDDEEADQL EFEDEFAGSS SEVIVGISPE DEEPSSVGGK
301 PNESIGRTIE GQSIRDNLQA KDNKSTDVPG AGPKDSAVKE EPPQKRLPML
351 AEEFECSGSE DPIIRELLKE NSLINCQQGK DAQPPYHWSI ERSISPDKTE
401 IVNGAVQTAD GQRPGTPMPK SRGIPIKKGT DAKYPSAGTE NVPGSKSGAT
451 RHVRGSPPYQ EGKSVNAENV QLNASTAVKE TDKSEVNPVD DNDSLDDKYI
501 MPSDDFSNTF FPHDTDRLNY HADHLGDYDL ETLCEESVLM GVINSIKLIN
551 LDMRLNHIEE QVKEIPKIIN KLESIDRVLA KTNTALSTIE GHLVSMMIMI
601 PGKGKGERKG KNNPELKPVI GRDILEQQSL FSFDNVKNFR DGSLTNEPYG
651 AAVQLREDLI LPELNFEETN ASQFVPMADD SSRDVIKTLI RTHIKDRELR
701 SELIGYLNKA ENDEEIQEIA NTGNDIIDGN I
```

B

Protein sequence coverage: 13%
Matched peptides shown in **bold red**.

```
  1 MGSSHHHHHH SSGLEVLFQG PAMDKLELVN DGLNIIDFIQ KNQKEIQKTY
 51 GRSSIQQPSI KDQTKAWEDF LQCTSGESEQ VEGGMSKDGG DVERRNLEDL
101 SSTSPTDGTI GKRVSNTRDW AEGSDDIQLD PVVTDVVYHD HGGECTGYGF
151 TSSPERGWSD YTSGANNGNV CLVSDAKMLS YAPEIAVSKE DRETDLVHLE
201 NKLSTTGLNP TAVPFTLRNL SDPAKDSPVI AEHYYGLGVK EQNVGPQTSR
251 NVNLDSIKLY TSDDEEADQL EFEDEFAGSS SEVIVGISPE DEEPSSVGGK
301 PNESIGRTIE GQSIRDNLQA KDNKSTDVPG AGPKDSAVKE EPPQKRLPML
351 AEEFECSGSE DPIIRELLKE NSLINCQQGK DAQPPYHWSI ERSISPDKTE
401 IVNGAVQTAD GQRPGTPMPK SRGIPIKKGT DAKYPSAGTE NVPGSKSGAT
451 RHVRGSPPYQ EGKSVNAENV QLNASTAVKE TDKSEVNPVD DNDSLDDKYI
501 MPSDDFSNTF FPHDTDRLNY HADHLGDYDL ETLCEESVLM GVINSIKLIN
551 LDMRLNHIEE QVKEIPKIIN KLESIDRVLA KTNTALSTIE GHLVSMMIMI
601 PGKGKGERKG KNNPELKPVI GRDILEQQSL FSFDNVKNFR DGSLTNEPYG
651 AAVQLREDLI LPELNFEETN ASQFVPMADD SSRDVIKTLI RTHIKDRELR
701 SELIGYLNKA ENDEEIQEIA NTGNDIIDGN I
```

Figure 4.4 His tag cleavage for the NiV P protein. (A) SEC profile of the NiV P protein after removal of the N-terminal His tag using 3C protease. Th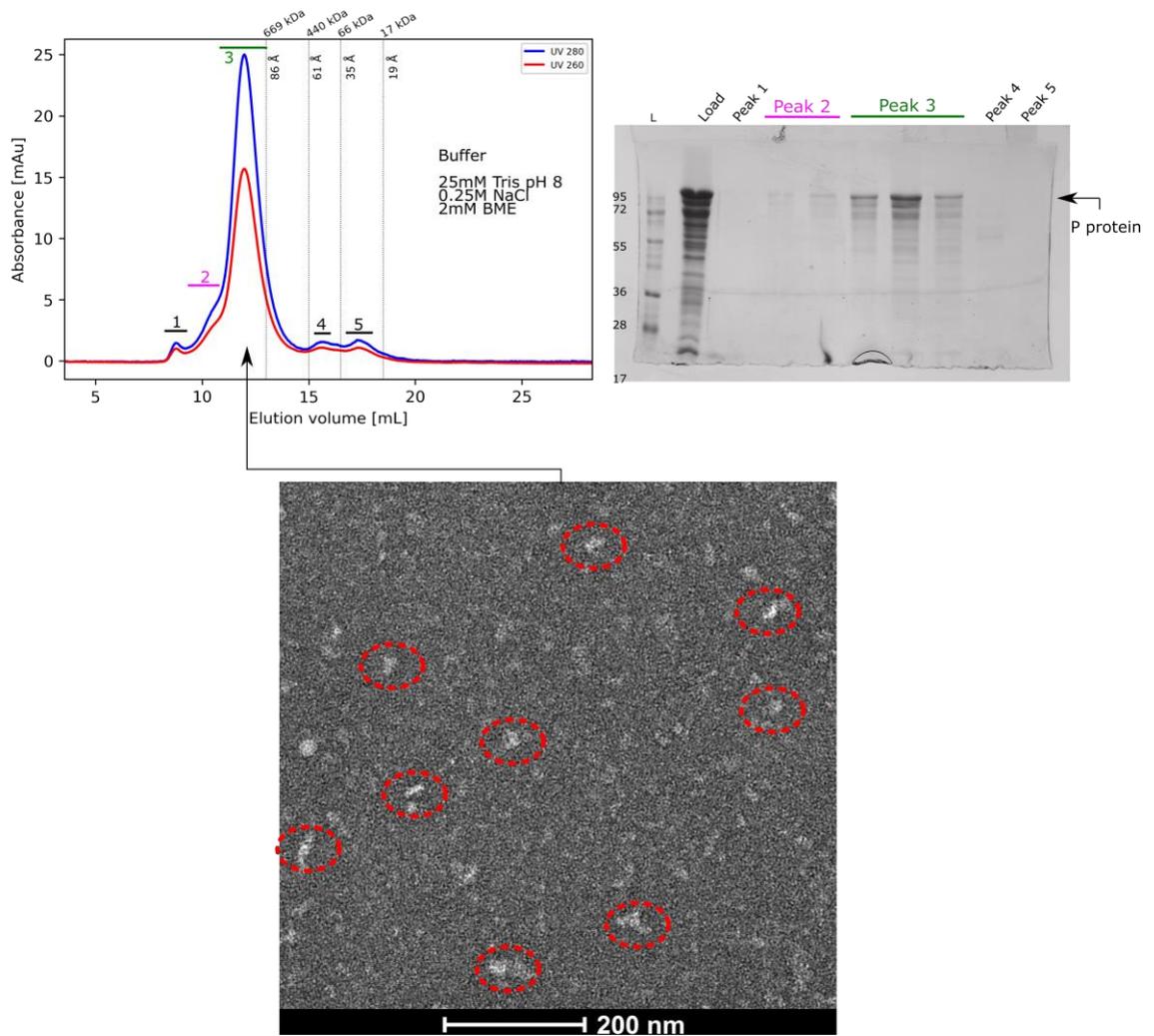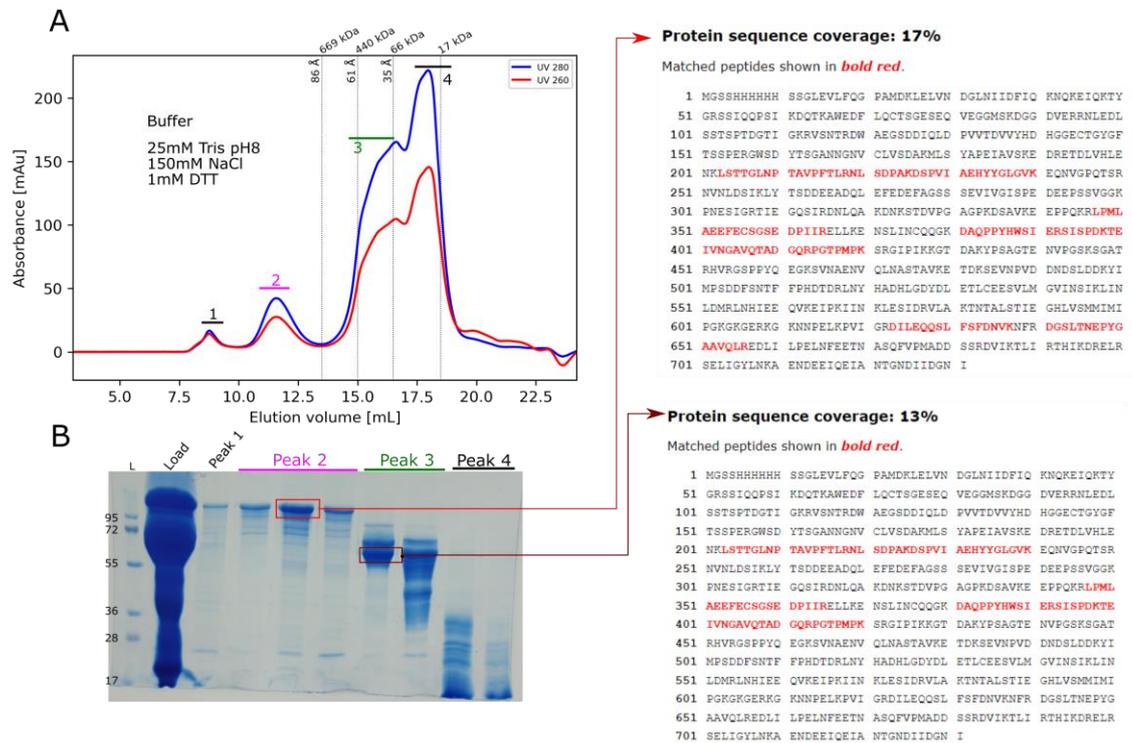e His tag cleaved NiV P protein was subjected to SEC purification using a Superose 6 Increase 10/300 column. (B) The SDS PAGE of the His tag cleaved NiV P protein. Mass spectrometry sequence coverage of tryptic digests of two protein bands were shown. Identified peptides are shown in bold red. Sequence and numbering are for the recombinant NiV P protein containing the N-terminal 6xHis tag.

4.2.2. In vitro N+P heterocomplex assembly

In an attempt to obtain structural information on how the NiV P protein attaches to the NiV N-RNA helical complex, *in vitro* assembly of NiV N/P heterocomplex was performed using the purified NiV P protein and the purified NiV N-RNA complex (Chapter 3). The formation of the *in vitro* assembly of NiV N/P heterocomplex was first investigated using polyacrylamide gel electrophoresis. An initial attempt of using Native PAGE to study the protein-protein interaction was not successful, as both the NiV P and NiV N proteins did not migrate into the Native PAGE (Figure 4.2B and Figure 4.5). Therefore, further optimization of the Native PAGE procedure was performed focussing primarily on the NiV N protein. The pore size of the polyacrylamide gel is dependent on (i) the total percentage of the acrylamide in the gel and (ii) ratio of acrylamide to bis-acrylamide (crosslinker) in the gel. A higher total percentage of acrylamide leads to a smaller pore size gel that can be used to separate smaller molecules. A higher ratio of acrylamide to bis ratio results in a larger pore size as there are fewer crosslinkers[176]. However, despite using polyacrylamide with a larger pore

size, prepared using a low percentage (8%) polyacrylamide with an acrylamide:bis ratio of 99:1, the NiV N-RNA protein remained mostly in the gel wells. As native PAGE also separates proteins according to their net charge, in addition to their size and shape, different types of native PAGE running buffers were tested. Conditions based on two different pH values (pH 8 and pH 10) of native PAGE in the presence or absence of 0.1 M $NaSO_4$ were tested (Figure 4.5). However, regardless of the types of native PAGE buffer used, the migration of the NiV N-RNA protein into the gel was minimal at best (Figure 4.5A-D). The ethidium bromide and Coomassie blue staining indicated that the NiV N protein-RNA complex remained mostly in the well of the native PAGE. The calculated pI of the NiV N protein is 6.06, and together with the presence of RNA in the protein complexes, the protein complexes should have an overall negative net charge at pH 8 and higher. Possible reasons for the poor protein migration in the native PAGE may be due to protein aggregation during the electrophoresis process or that the protein complexes are too large for native PAGE. To overcome this, the protein-protein interaction between these two viral proteins was investigated using agarose gels, which have a larger pore size compared to polyacrylamide.

Using agarose gel electrophoresis, both the NiV N-RNA and NiV P protein successfully migrated into the agarose gel, suggesting that the large particle size of NiV N-RNA and NiV P might be the reason for poor migration in Native PAGE (Figure 4.5E). The agarose gel electrophoresis clearly showed NiV N-RNA protein interacting with the NiV P protein, as judged by the shift of the NiV N-RNA protein bands in the presence of NiV P (Figure 4.5E). Although each subunit of N protein contains a MoRE region which can interact with the P protein XD (see section **1.5.5** above), the NiV N proteins are held together in close proximity by binding to the RNA in the helical assembly, therefore, it is highly likely that not all N proteins are interacting with and/or are bound to the P protomers due to restriction by steric hindrance. Regardless, the agarose gel electrophoresis confirmed that the full length NiV N-RNA protein interacts with full length NiV P.
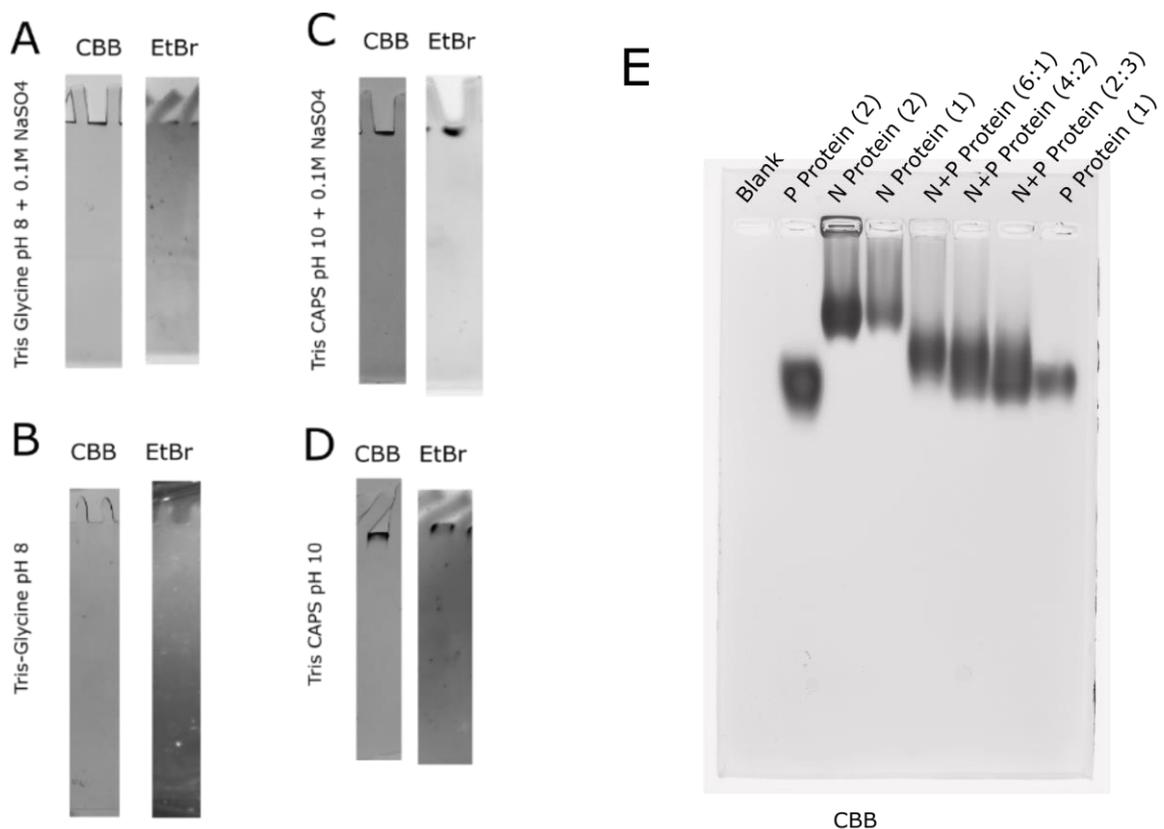
Figure 4.5 Gel electrophoresis of NiV N and NiV P. (A-D) Different Native PAGE buffers tested to improve the mobility of the NiV N during Native PAGE analysis. The Native PAGE was run at 4°C at 10mA overnight. (E) 1% Agarose gel electrophoresis of the NiV N interactions with NiV P. The number in the lane labels represents the molar ratio of the protein used in each respective well. 1 molar ratio represents 10 µM.

To obtain the N/P heterocomplex suitable for structural determination using cryoEM, additional purification was done using SEC. The N-terminal His-tags for both of the proteins were cleaved off to ensure that the presence of the affinity tag does not interfere with the N/P heterocomplex formation[177,178]. As described previously in Chapter 3, the NiV N-RNA protein is prone to degradation. Here, the additional His tag cleavage step has resulted in significantly more degraded NiV N-RNA protein that co-eluted with the full length NiV N-RNA protein (Figure 4.6A). Unexpectedly, the degradation of the NiV N-RNA protein appears to increase following the protein storage at -70°C, resulting in mostly degraded NiV N-RNA protein after *in vitro* assembly of NiV N/P heterocomplex (Figure 4.6B). Therefore, to minimize the protein degradation of the NiV N-RNA protein, the *in vitro* assembly of NiV N/P heterocomplex was performed immediately after the second HisTrap purification, following His tag cleavage by the 3C protease, to remove uncleaved N protein.

The NiV N/P heterocomplex mostly elutes close to the void volume (~8 mL) of the Superose 6 column. TEM imaging of negatively stained samples for the fraction from the void volume observed mostly just the nucleocapsid protein, with some visible bulky non-helical particles, which are likely to be the NiV P protein (Figure 4.7). While there were no obvious NiV N/P heterocomplexes identified, the co-elution of the N-RNA and P protein from the SEC represents an encouraging protein binding result. It is also possible that the protein complexes fall apart during negative staining with uranyl acetate, as the grid preparation process involves imposing a low pH condition onto the sample followed by a drying step. Therefore, cryoEM imaging of the protein complex was explored, however, attempt to increase the concentration of the N/P heteroprotein complex sufficient for cryoEM has been challenging. It turned out that concentrating the NiV N/P heterocomplex from the elution fraction of SEC often led to a lower than expected protein concentration. The protein concentration of the NiV N/P heterocomplex eluted from the SEC was around 0.01 mg/mL, however, reducing the protein solution volume by 20 times using centrifugal protein concentrator increased the protein concentration only to around 0.05 mg/mL, instead of the expected 0.2 mg/mL.

Recently, it was shown that when N-RNA and P proteins from the measles virus (MeV) were mixed, the resulting mixture was undergoing a phase separation *in vitro*[179]. Although no phase separation occurrence was observed directly from the SEC elution fraction, likely due to its low concentration, it is still possible that the NiV N-RNA and P mixture underwent phase separation and/or precipitated during the protein concentrating process. Regardless, the low yield of the purified full length NiV N-RNA protein and full length NiV P protein, as well as their tendency for degradation / precipitation have made the *in vitro* assembly of NiV N/P heterocomplex unfeasible.
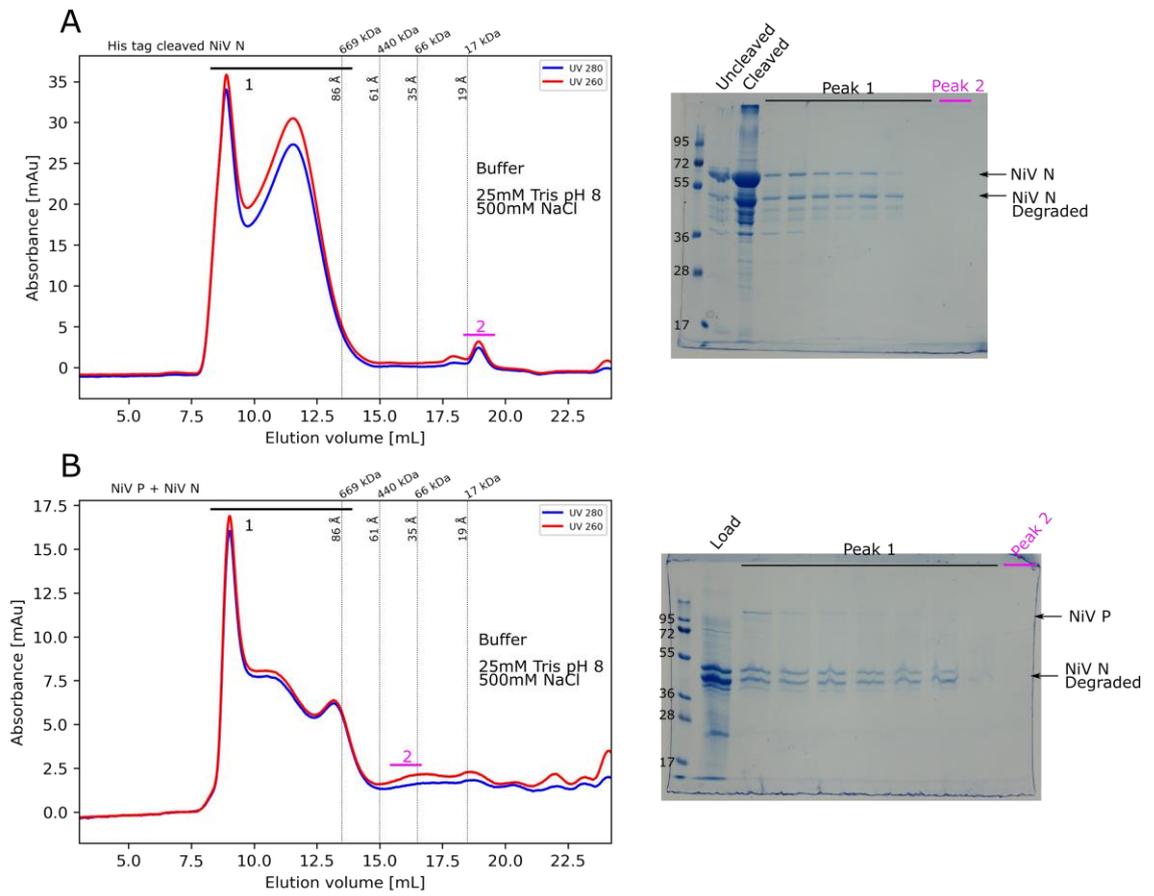
Figure 4.6 Initial purification of the NiV N and NiV P protein complex. (A) SEC profile of the His tag cleaved NiV N using Superose 6 Increase 10/300, (B) SEC profile of the NiV N in complex with NiV P using Superose 6 Increase 10/300. The protein mixture was incubated at RT for 30 min before subjected to SEC purification.
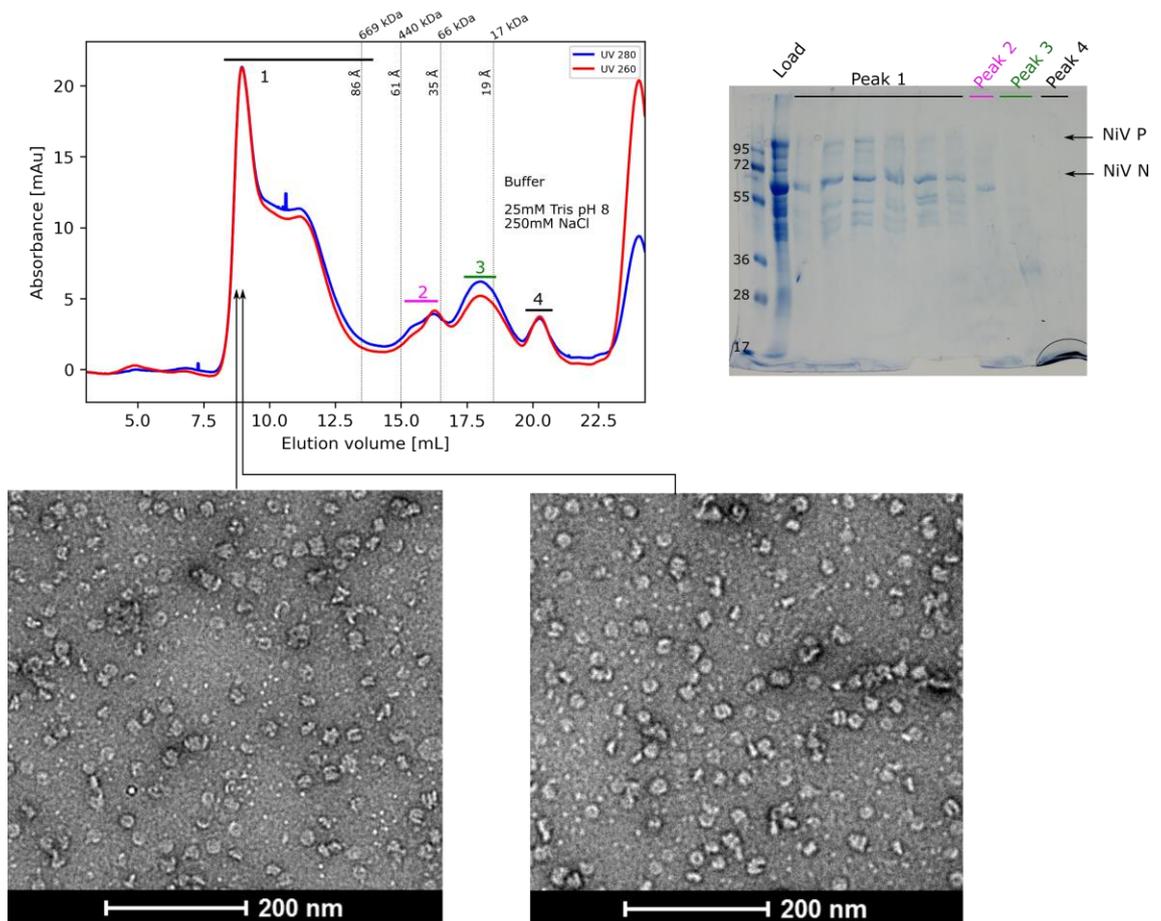
Figure 4.7 SEC profile of the NiV N/P heterocomplex obtained *in vitro*. The protein assembly was separated using Superose 6 Increase 10/300 (top). TEM images of negatively stained sample from the fraction eluted at 8mL retention volume (bottom).

4.2.3. In vivo assembly of N/P heterocomplex

As attempts to assemble the NiV N/P heterocomplex *in vitro* were unsuccessful, the *in vivo* assembly of the NiV N/P heterocomplex using co-expression in *E. coli* was explored. Two vector constructs were designed using the pET-DUET plasmid.

4.2.3.1 His-NiV P and NiV N

The first vector construct involved co-expression of a N-terminal His-tagged NiV P protein and a tagless N protein. Initial two step purification using Nickel affinity chromatography and SEC did not result in any N protein bands observed in the SDS PAGE, suggesting there was no protein-protein interaction between the NiV P and NiV N (Figure 4.8A). It is likely that the protein interaction between N protein and P protein is very weak, and the protein complex dissociated during the cell lysis and purification steps. As the site of NiV replication occurs in the cytoplasm of eukaryotic cells which has potassium as the main cation[180,181], the purification was also repeated using KCl as a salt to mimic the eukaryotic cytoplasmic environment, and hopefully promote better protein interaction between N protein and P protein. However, no NiV N protein was detected after the Nickel affinity chromatography step, purified in the presence of KCl salt (Figure 4.8B). Interestingly, in both purifications, the NiV P protein was again co-purified together with nucleic acid, as judged from the A_{260}/A_{280} ratio. In particular, when it was purified in buffer containing 200mM KCl salt, the NiV P was able to be purified to high homogeneity, resulting in a single sharp peak in SEC. Negative stain TEM imaging of the fraction from the SEC peak shows the presence of irregular ball-like structures (Figure 4.8C). Further analysis using urea PAGE analysis suggested that the nucleic acid, co-purified with NiV P, is longer than 300 nt (Figure 4.8D). In common with previous NiV P purification, it was possible to remove the bound nucleic acid with an additional purification step, in this case, using anion exchange chromatography.
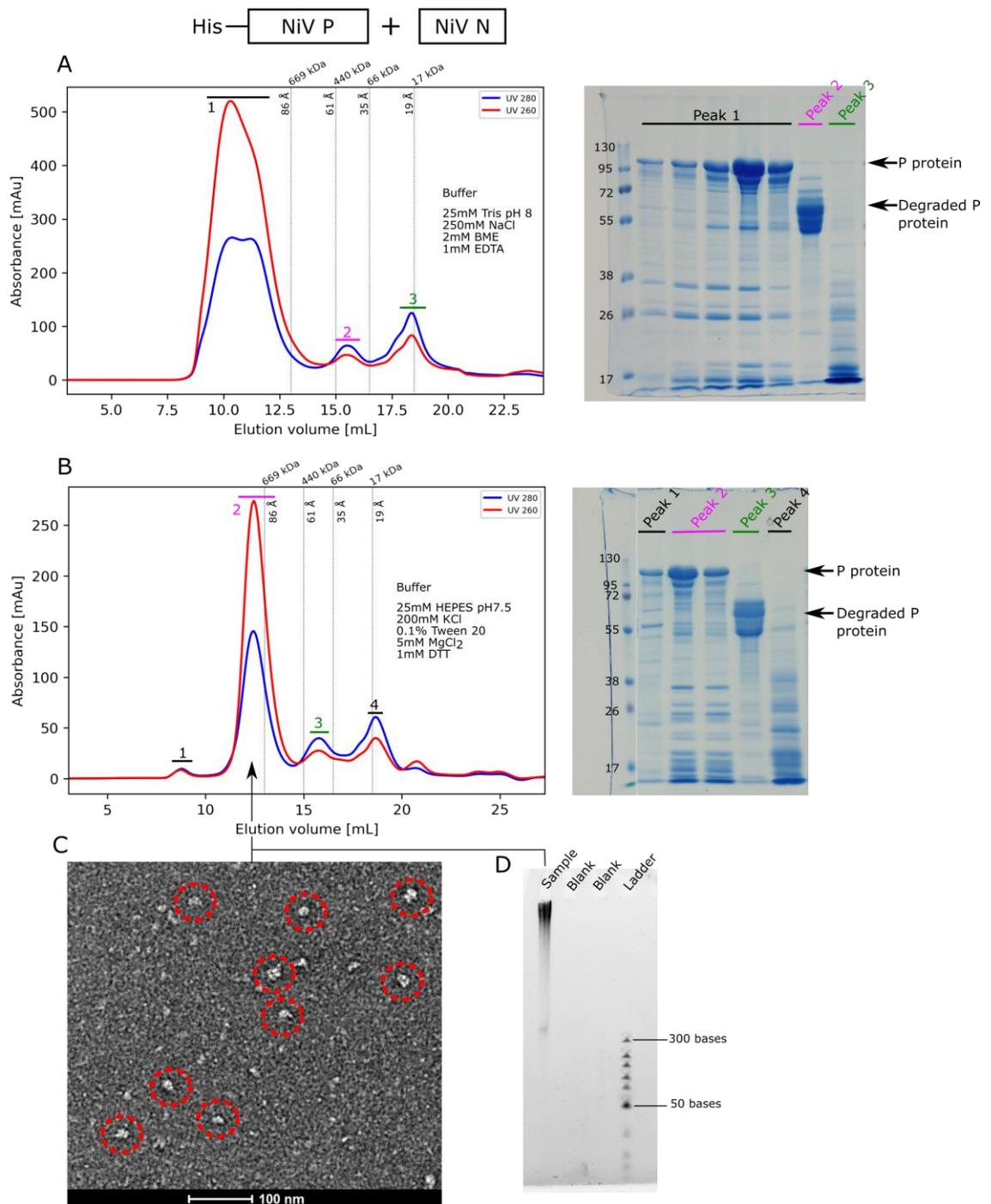
Figure 4.8 SEC profile for the purification of the His-NiV P + N co-expression. The protein sample after Nickel affinity chromatography purification was subjected to SEC using a Superose 6 Increase 10/300 column at two different salt conditions, (A) 250mM NaCl, and (B) 200mM KCl. (C) Negative stained TEM image of the Peak 2 fraction from Figure 4.8B. The putative NiV P particles are highlighted by red circles. (D) 10% Urea PAGE of the Peak 2 fraction from Figure 4.8B stained by ethidium bromide.

4.2.3.2 His-NiV N + NiV P

The second vector construct involved co-expression of a N-terminal His tagged NiV N and a tagless NiV P. Initial purification of protein expressed from this vector construct was very promising. In particular, both the NiV N-RNA and NiV P were detected as being co-eluted during two step purification using Nickel affinity chromatography and SEC (Figure 4.9A). As the protein complex mostly eluted in the void volume when Superose 6 Increase column was used, a Sephacryl S500 HR column was used to further separate the higher molecular weight protein complex (Figure 4.9B). Compared to Superose 6 Increase, the Sephacryl S500 HR has a higher upper exclusion limit, and it is commonly used for purification of large macromolecules such as viruses or DNA plasmid. Further SEC purification using Sephacryl S500 shows a single broad peak corresponding to larger species than the protein standard thyroglobulin (660 kDa, 86 Å). However, despite the additional SEC purification step, the negative stain TEM images showed the presence of mostly random aggregates with a minor population of helical nucleocapsid (Figure 4.9). The negative stain images closely resemble the negative stain images of the isolated NiV P protein as seen in Figure 4.2, suggesting this random aggregate is likely due to the presence of excess NiV P protein in the sample. It remains unclear how this excess of tagless NiV P protein remains bound onto NiV N-RNA protein after Nickel affinity chromatography and SEC purification steps. It is possible that the NiV P was dissociated from the NiV N/P heterocomplex owing to centrifugal forces used during the protein concentrating process.

Similar to the previous His-NiV P + NiV N vector construct, the purification was repeated using KCl as the salt to better mimic the eukaryotic host cell cytoplasm conditions. In the Nickel affinity chromatography, the NiV N-RNA was found to elute from 50 mM to 400 mM imidazole, and the NiV P was found to co-elute with NiV N-RNA with decreasing affinity as the concentration of imidazole increases (Figure 4.10A). Protein fractions eluted from 100 mM to 200mM Imidazole were collected and subjected to SEC purification (Figure 4.10B). Negative stained TEM imaging of the SEC purified sample showed similar aggregate formation, as seen previously, likely due to the presence of excess of NiV P. In an attempt to reduce non-specific binding of NiV P toward NiV N, the protein fractions from 300 mM to 400 mM imidazole were collected and subjected to SEC purification (Figure 4.10C). TEM imaging of negatively stained

samples revealed mostly helical nucleocapsid assemblies. Interestingly, attempts to increase the protein concentration of the corresponding SEC fraction were successful with the NiV P protein band remaining intact after concentrating the protein mixture to 0.1 mg/mL.
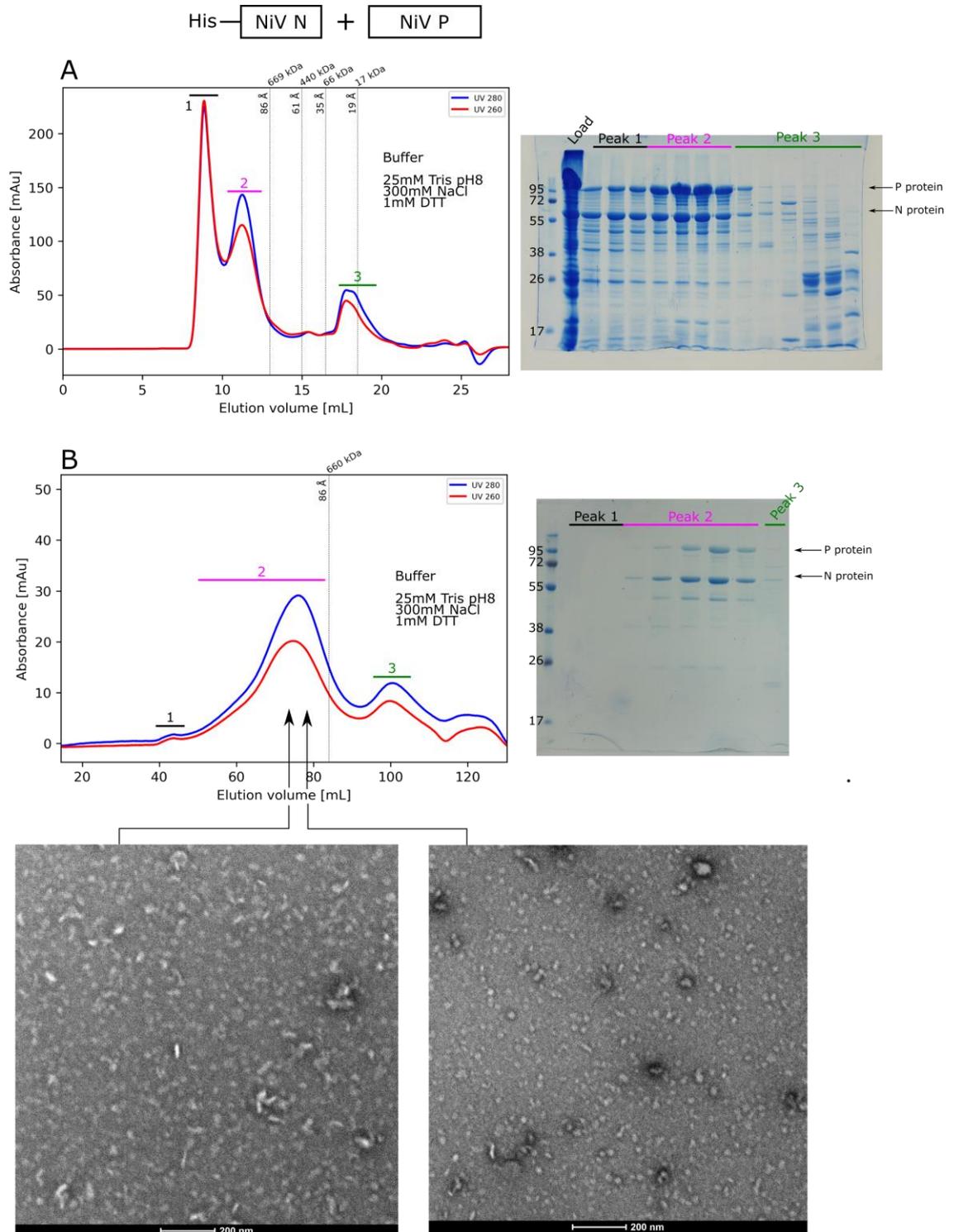
Figure 4.9 SEC profile of the His-NiV N + NiV P co-expression. (A) SEC purification using a Superose 6 Increase 10/300 column. (B) SEC purification using a HiPrep 16/60 Sephacryl S-500 HR column. Bottom: TEM images of negatively stained sample corresponding to the fractions from peak 2.
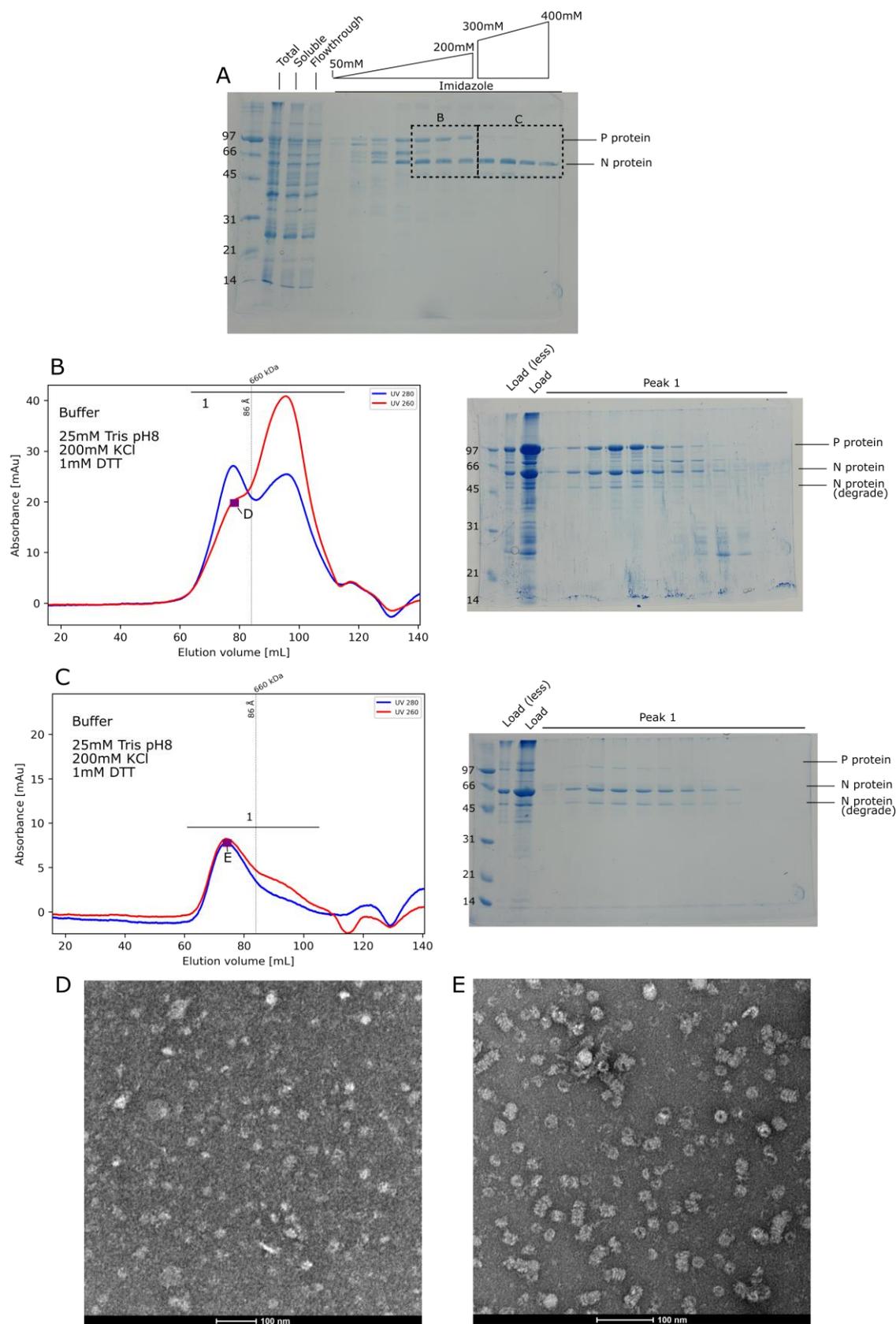
Figure 4.10 Purification of the His-NiV N + NiV P co-expression using 200 mM KCl buffer. (A) Fractions from nickel affinity chromatography purification using HisTrap FF 5 mL column. (B) SEC purification of protein sample eluted using ~100 mM Imidazole. (C) SEC purification of protein sample eluted using ~350mM Imidazole. (D) The TEM image of negatively stained sample corresponding to fractions from SEC in (B). (E) The TEM image negatively stained sample corresponding to fractions from SEC in (C).

4.2.4. CryoEM analysis of His-NiV N + NiV P complex

With sufficient protein sample and concentration (0.1 mg/mL) from the His-NiV N + NiV P construct purified using the buffer containing 200 mM KCl (section **4.2.3.2** above, Figure 4.10C), several cryoEM grids were prepared and a data collection was performed. The two-day cryoEM data collection was conducted using the 200 kV Glacios instrument, equipped with a Falcon 4 detector, at the University of York. A total of 6796 micrographs were collected as EER (Electron Event Representation) format, a new movie format for the Falcon 4 detector. The EER format records electron events at a very high frame rate (248 Hz) and generally has a smaller file size compared to the MRC movie format[182]. The EER dataset was processed using RELION 3.1.2 as described in the previous Chapter 3, following the conversion of the EER data into the MRC movie. A total of 629,683 particles were selected using RELION template-based Auto-picking and subjected to 2D classification. The majority of the classes showed particles that resemble a single helical turn of the nucleocapsid, as described earlier in Chapter 3 (Figure 4.11B). In addition, many of the class averages also contained particles which resemble the side views of the nucleocapsid protein-RNA complex, however, they are very noisy, likely due to the presence of the disordered P protein. Regardless, the initial 2D classification did not detect any obvious P protein, either as an individual protein or in complex with the N-RNA protein.
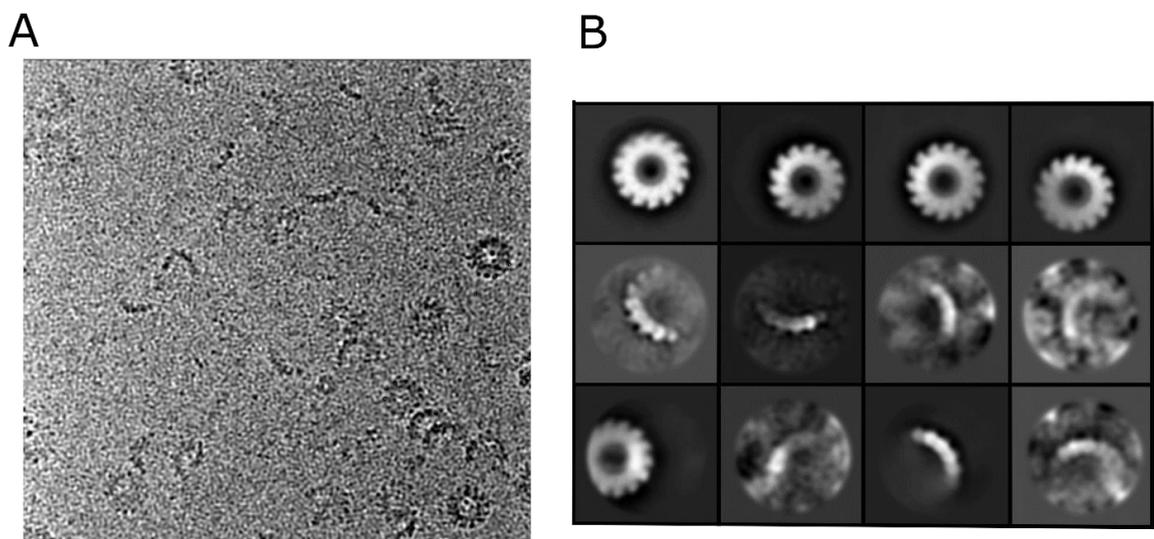
A
B

Figure 4.11 CryoEM data collection for the NiV N/P heterocomplex. (A) A representative cryoEM image of the NiV N/P heterocomplex. (B) Representative views of the 2D class averages of the picked particles.

Further sorting of the particles using the 3D classification feature in RELION, was used to identify any 3D class averages with the presence of P protein among the selected particles. The NiV N-RNA complex cryoEM map from previous Chapter 3 was used as a reference model for the 3D classification (Figure 4.12). Initial 3D classification identified 15,149 particles corresponding to a single good 3D class, and 3D refinement of this 3D class resulted in an 8 Å resolution cryoEM map. This class resembles a rigid helical turn of the nucleocapsid protein formed by about 15 subunits of N protein (Figure 4.13A). Superimposition of cryoEM map of the NiV N-RNA assembly from Chapter 3 onto the cryoEM of this helical turn nucleocapsid protein showed nearly identical fitting between the two maps suggesting that both of the protein assemblies share similar helical pitch and angle (Figure 4.14).

Additional rounds of 3D classification of the remaining particles further identified two distinct assemblies (Figure 4.12). The first type of assembly resembles the spiral assembly of the nucleocapsid protein; however, the preferential orientation of these particles, lacking side projections, hampered a high resolution 3D reconstruction of this assembly. Interestingly, additional 2D classification of this set of spiral assembly further detected a subset of about 6000 particles that correspond to the clam shaped assembly particles, as judged from 2D class averages. Compared to the clam shaped assembly reported earlier in the Chapter 3, where about 35% of the particles in the NiV N cryoEM dataset have been classified as clam shaped assembly, here, in the NiV N + P cryoEM dataset, only about 2% of the particles were identified as the clam shaped assembly.
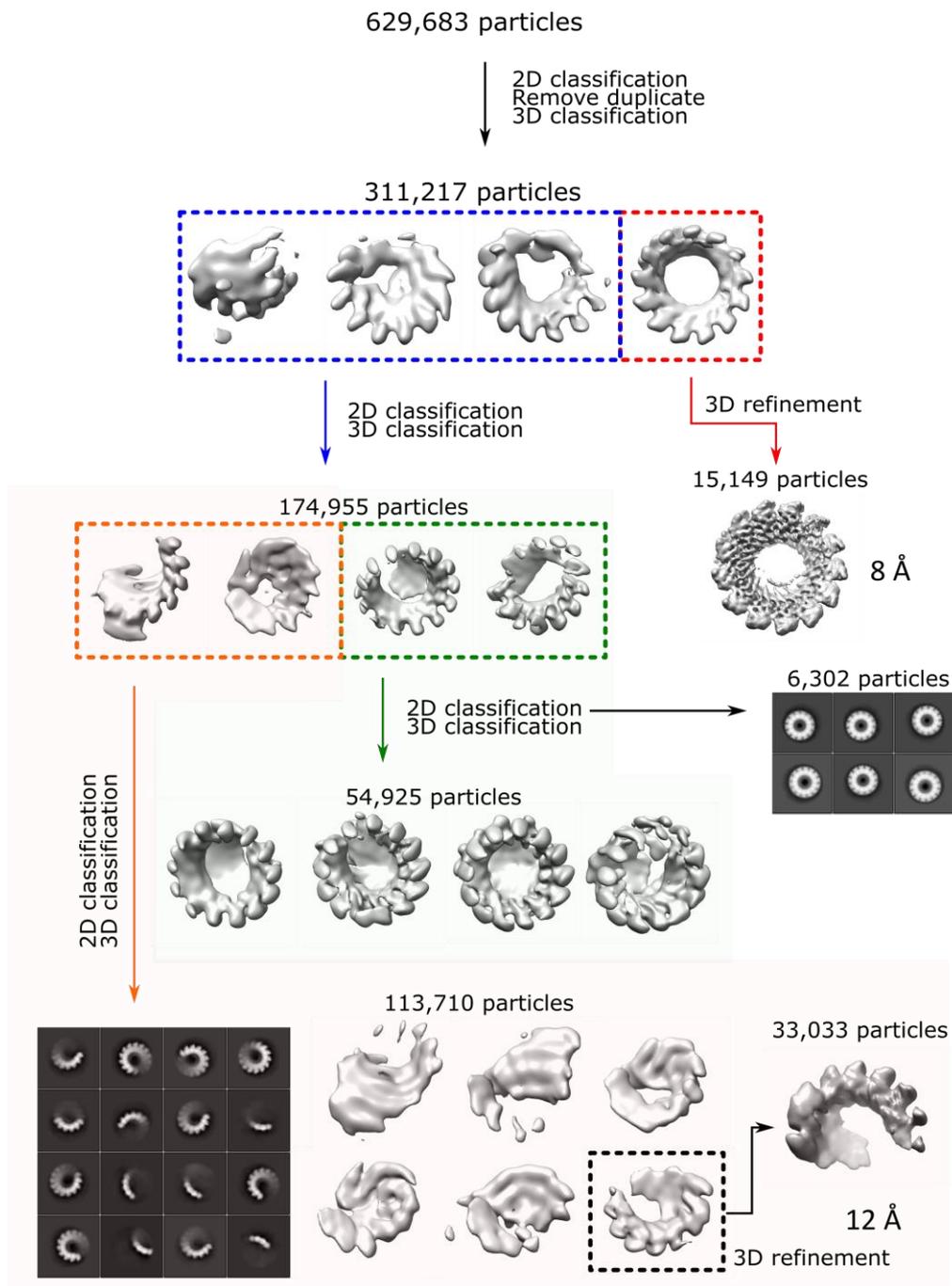
Figure 4.12 CryoEM data processing of NiV N/P heterocomplex. 629,683 particles which were picked by RELION's Auto-picking, were subjected to 2D classification to remove junk particles and duplicated particles. They were then subjected to 3D classification using an initial model generated from the NiV N-RNA EM map (EMD-12581) low-pass filtered to 60 Å. One good 3D class (red box) was identified, and further 3D refinement led to an 8 Å cryoEM map. The remaining bad classes (blue box) were further sorted using 2D classification and 3D classification into two distinct assemblies, the spiral assembly (green box) and the uncoiling assembly (orange box). All of the 3D classification was done using C1 symmetry. Inset image shows the 2D class averages of the clam shaped assembly and uncoiling assembly (orange box).

The second type of assembly resembles a structurally more flexible "loose" spiral nucleocapsid assembly. 2D classification of this set of particles showed that the particles were present in different orientations with a great proportion of the particles representing a side view of the nucleocapsid assembly. However, despite the availability of the side view particles, the 3D reconstruction for the "loose" spiral nucleocapsid assembly proved to be extremely challenging due to the dynamic and heterogeneity of these assemblies[183]. 3D classification identified only 33,033 particles corresponding to a relatively well defined 3D class while the remaining particles were classified as junk (Figure 4.12). 3D refinement of this respective good class led to a 12 Å resolution cryoEM maps, and 2D class averages for this type of spiral assembly contained particles show particles that, unlike the nucleocapsid structure reported earlier, resembled a more relaxed, loose helix with partially uncoiled state of the spiral assembly (Figure 4.13B).
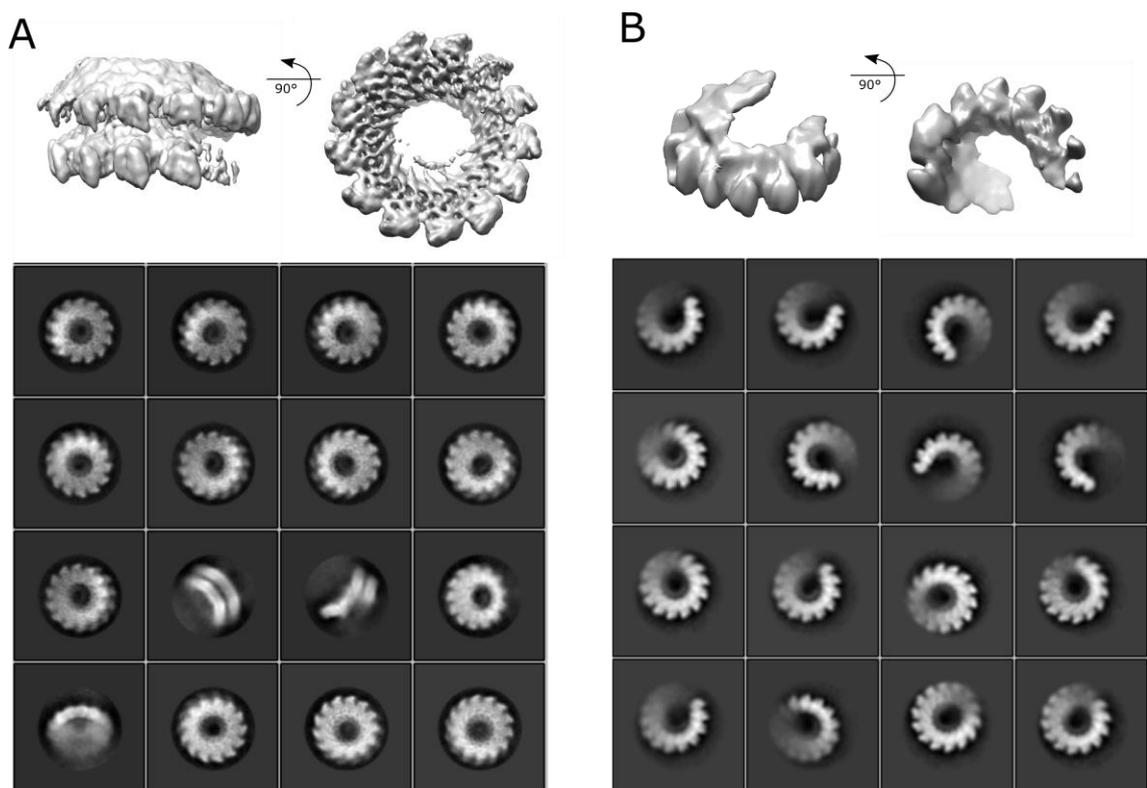
Figure 4.13 Particle classification of the NiV N/P heterocomplex. (A) 3D map represents a double helical turn of the N protein. The representative 2D class averages of the particles contributed to the map are also shown. (B) 3D map represents an uncoiling loose spiral nucleocapsid assembly. The representative 2D class averages contributed to the map are also shown.
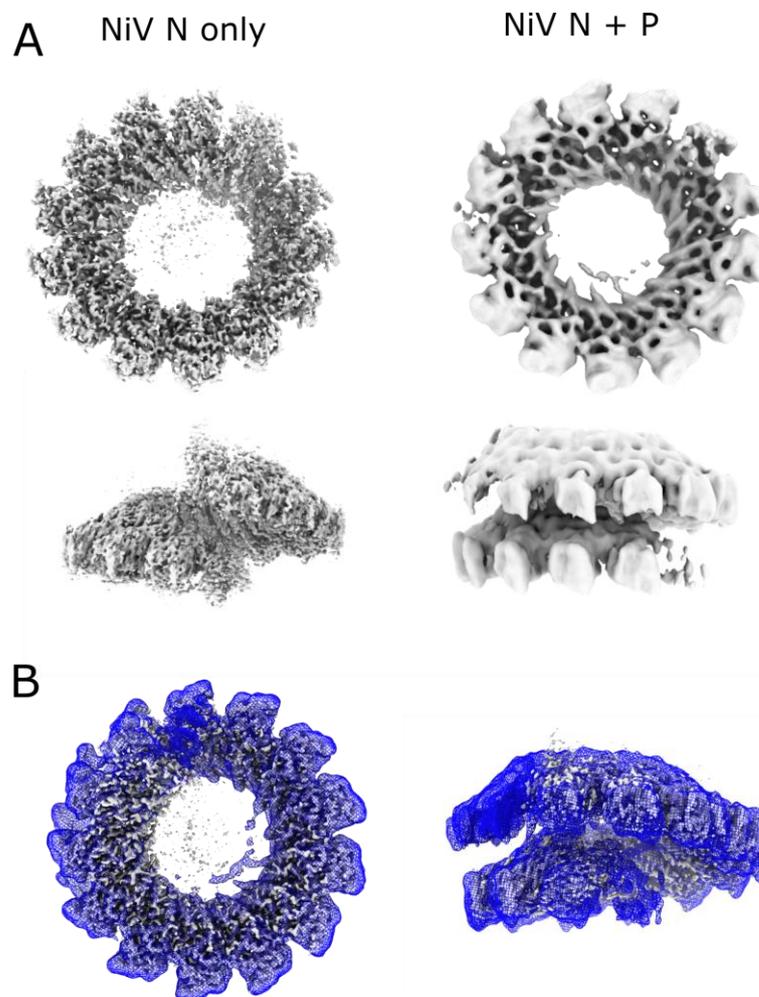
Figure 4.14 Comparison of the cryoEM maps of the NiV N and NiV N/P heterocomplex assemblies. (A) The side-by-side comparison of cryoEM maps of the NiV N only (left, Chapter 3) and NiV N + P (right). (B) Fitting of the NiV N only cryoEM map (grey) into the NiV N + P cryoEM map (blue mesh).

To gain structural insight about the "loose" spiral assembly formation, the structure of NiV N monomer was fitted into the cryoEM map of the "loose" nucleocapsid spiral assembly. Upon fitting the NiV N monomeric models into the maps, there was a weak unmodelled density observed near the N-terminal N-core domain of the NiV N protein. This unmodelled density appeared to interact with the A1-H2 and A2-H5 loop regions of the NiV N protein (Figure 4.15). The additional unmodelled density is large enough for accommodating the 1-36 aa region of the NiV P protein and the X domain of HeV P protein crystal structures (Figure 4.15), while the oligomerisation domain of NiV P protein is too large to fit in. Overall, the low resolution of the cryoEM map, prevents the identification of approximate regions of the NiV P protein that potentially bind to the NiV N protein.
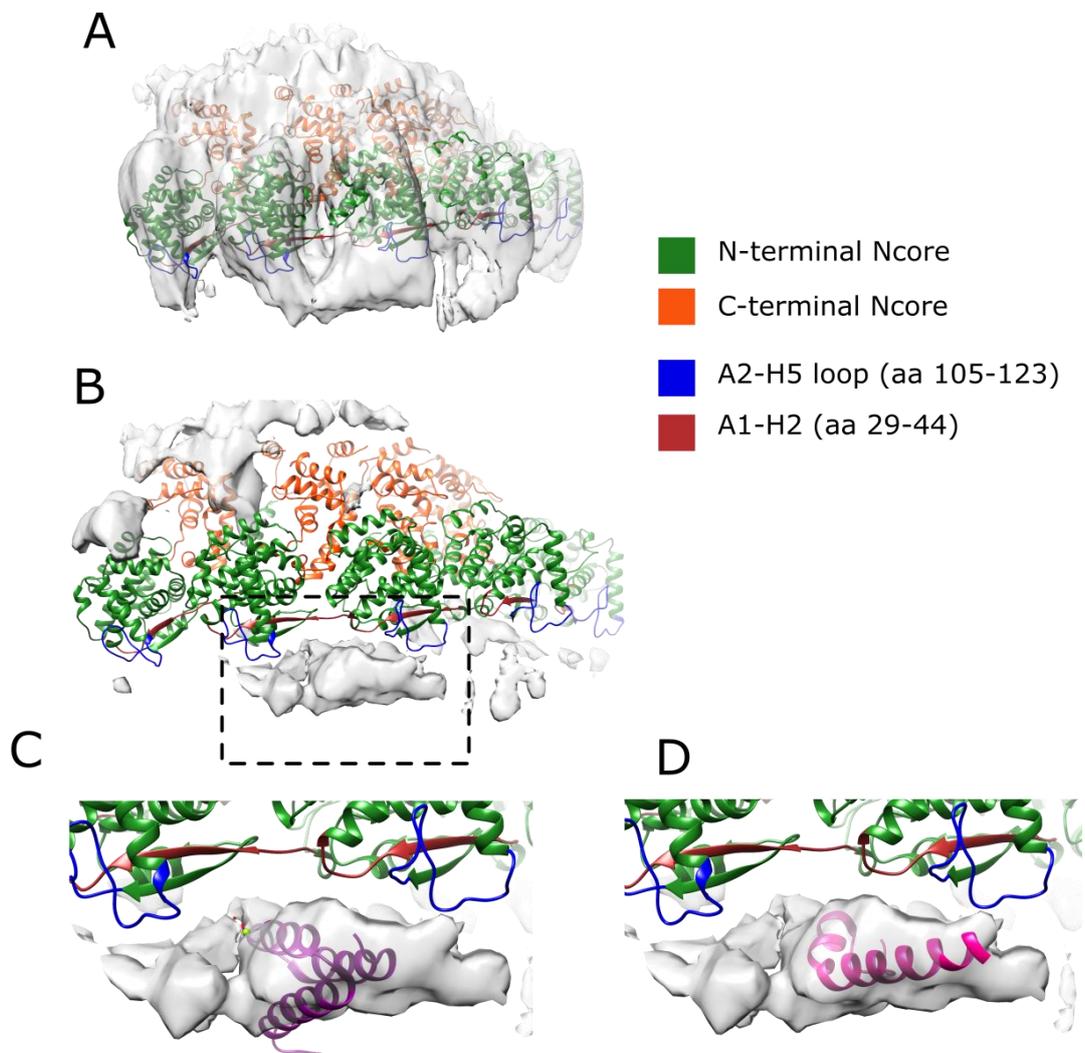
Figure 4.15 Fitting of the NiV N and NiV P model into the cryoEM map of the "loose" spiral nucleocapsid assembly. (A) The NiV N monomeric model (pdb:7NT5) was fitted into the cryoEM map of the "loose" spiral nucleocapsid assembly. (B) Difference map (cryoEM map minus the model map) using the fitted NiV N model. (C) Hendra virus (HeV) P-XD crystal structure (pdb:4HEO, purple) fitted into the unmodelled difference map density. (D) 1-36 region of the NiV P protein model (pdb:4CO6, pink) fitted into the unmodelled difference map density.

4.3 Discussion

NiV P protein expression and purification using the *E coli* expression system have been reported earlier by Salvamani *et. al.*[169] and most recently by Jensen *et. al.*[175]. Here the purification of the NiV P protein expressed in *E. coli* showed that the NiV P protein tended to co-purify together with nucleic acid as judged by the A_{260}/A_{280} ratio and gel electrophoresis. The removal of the nucleic acid can be achieved with an additional purification step using a Heparin column. Unexpectedly, attempts to reduce the non-specific nucleic acid binding by the NiV P protein by using a high salt purification buffer resulted in an opposite effect, with more nucleic acid being co-purified together with

the P protein. Protein-nucleic acid interactions are generally mediated via attraction between the positively charged residues in the protein and the negatively charged phosphate backbone of the nucleic acid. However, as the NiV P protein has a low pI value of 4.4, the high salt concentration might screen the negative charge at the protein surface thus facilitating the positively charged protein residues to interact with the nucleic acid[184]. Regardless of the presence or absence of the nucleic acid, the NiV P protein was eluted at a retention volume larger than hydrodynamic radius of thyroglobulin (82Å) from SEC. This suggested that the binding of the nucleic acid did not result in a more rigid assembly, as the hydrodynamic radius remained mostly the same.

The NiV P protein contains a tetramerization domain and it has been shown to form homotetramers (~320 kDa) in solution[174,175]. As the SEC measures hydrodynamic radius of macromolecules as a function of their size and shape, it is likely that the NiV P protein has a larger hydrodynamic radius than a normal globular protein due to its highly disordered nature, thus resulting in it being eluted earlier than expected during SEC. This lack of structured features in the NiV P protein was evident in the TEM images of negatively stained NiV P protein samples (Figure 4.3), where the SEC purified P protein mostly existed as aggregate-like particles.

Subsequently, the assembly of the NiV N-RNA and NiV P proteins was attempted. Purification of the *in vitro* assembled NiV N/P heterocomplex often resulted in low yield, and surprisingly, concentrating the NiV N/P heterocomplex, always resulted in lower than expected protein concentration. This phenomenon is likely explainable by the formation of liquid-liquid phase separation (LLPS). LLPS occurs when macromolecules come together and condense into a dense phase that coexists with a dilute phase[185]. LLPS formation has increasingly been recognised as one of the key principles for formation of membrane-less intracellular organisation[186,187]. Upon infection of the host cell, NiV is known to induce the formation of cytoplasmic inclusion bodies[180], and these inclusion bodies might be biomolecular condensates formed by LLPS. Apart from NiV, the formation of LLPS has been reported for cells infected by the Measles virus (MeV)[188] and more recently it has been shown that *in vitro* mixing of the MeV N and MeV P proteins also led to LLPS formation[179]. Although formation of phase separation was not observed directly upon mixing NiV N-RNA and

NiV P, likely due to the low concentration used during the *in vitro* assembly, it is still possible that the LLPS might have occurred, and the protein complexes pelleted during the centrifugation prior to protein purification by SEC. The occurrence of the LLPS during *in vitro* assembly of NiV N/P heterocomplex greatly hindered the purification of the NiV N/P heterocomplexes.

The *in vivo* assembly of the NiV N/P heterocomplex was conducted using protein expressed from two different co-expression vector constructs with each construct yielding completely different results. Purification of the NiV N/P heterocomplex using Nickel affinity chromatography was only observed when the N-terminal His tag was introduced on the NiV N protein, but not when the N-terminal His tag was on the NiV P protein. It is possible that the introduction of the N-terminal His tag at the NiV P protein may have interfered with the protein interaction between the NiV N and NiV P. This indirectly suggested that the N-terminal region of the NiV P protein may involve in a binding interaction with the NiV N-RNA helical complex. While the first 50 aa of the NiV P are known to interact with the NiV N to form a RNA free N[0]P complex[53], this 50 aa region of NiV P is not known to interact with helical N-RNA, therefore it is highly unlikely that the N-terminal region of NiV P bind to NiV N-RNA helical complex. The next possible explanation is that the introduction of the N-terminal His tag on the NiV P may have contributed to the nucleic acid binding ability of the NiV P protein, thus inhibiting the complex formation between the NiV N-RNA and NiV P. Paramyxovirus phosphoproteins are known to be an essential co-factor during RNA synthesis, however, direct phosphoprotein interaction with any form of RNA has not been reported before. Generally, the *Mononegavirales* phosphoprotein serves as a noncatalytic factor of the RNA polymerase and as a major interferon (IFN) antagonist[189,190]. Furthermore, non-specific nucleic acid binding due to the presence of recombinant His tag has also been reported[191]. It is possible that the non-specific binding to the nucleic acid resulted in some minor conformational changes that shielded the N protein binding region in the NiV P protein from interacting with the NiV N-RNA protein. Interestingly, the co-purification results shown here are of great contrast with the previous co-purification of the Mumps virus (MuV) N and P proteins[192]. In the cases of MuV N protein, it was possible to co-purify the MuV N-terminal His-tagged P protein and the untagged MuV N-RNA protein complex by

affinity chromatography, although most of the MuV N protein was found eluted during the washing step. However, the MuV P protein is about 391 aa long, which is 318 aa shorter than the NiV P protein. It is possible that the presence of the additional N-terminal ~100 aa long disordered region accounts for the inability of the NiV P protein to bind to the NiV N-RNA protein while having an N-terminal His-tag (Figure 4.16).
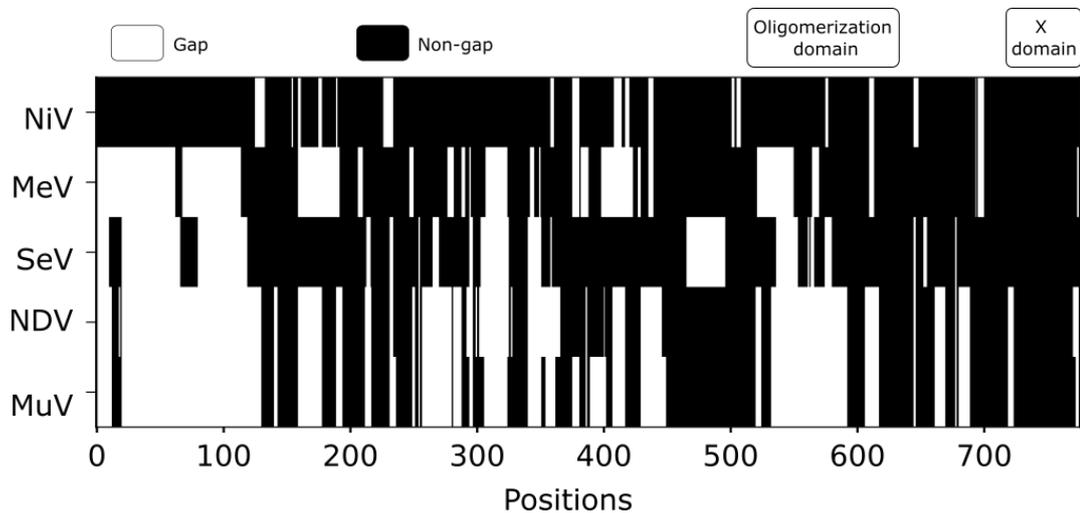
Figure 4.16 Multiple sequence alignment analysis (MSA) of the Paramyxovirus P protein. The MSA was performed using Clustal Omega. The overview of the MSA is shown here with the aligned residues represented as black and the alignment gaps represented as white area. The oligomerisation domain and the X domain for the NiV P protein are also indicated. NiV, Nipah virus; MeV, Measles virus; SeV, Sendai virus; NDV, Newcastle disease virus; MuV, Mumps virus.

During the natural NiV infection, the transcription and replication processes occur in the cytoplasm of the host cell. The cytoplasm is a highly crowded space, occupied by various macromolecules and these abundant macromolecules are known to influence the biomolecular behaviour of the protein assembly[193]. For instance, the intrinsically disordered protein can become more compact under crowded conditions, gaining a secondary structure in the form of a molten globular domain[194]. Both the NiV N and NiV P proteins contain intrinsically disordered regions, and the presence of the macromolecular crowding may help stabilise these intrinsically disordered regions and promote proper and stable interaction between the N and P proteins during the *in vivo* assembly.

Interestingly, the NiV N + P heterocomplex cryoEM dataset collected is more heterogeneous and noisier than the NiV N-RNA cryoEM dataset described in the previous Chapter 3, likely due to the presence of the NiV P protein in the sample. It is possible that the presence of the disordered P protein in the sample might have

contributed to the noise and affected the final reconstruction of the map. In contrast to the aforementioned rigid spiral ring and clam shaped assemblies, the N-RNA protein in the sample of the NiV N/P heterocomplex is found to also form a "loose" spiral assembly. The 2D class averages for this "loose" spiral assembly showed that one end of the spiral particle is gradually faded into the ice, and this can be associated with the "uncoiling" of the N-RNA helix turn (Figure 4.13B). 3D reconstruction of the "loose" spiral nucleocapsid assembly observed density that could be attributed to a segment of the NiV P protein.

In NiV N protein, aside from the residues 1-54 and 468-532 segments which are known as the NiV P binding regions in NiV N (section **1.4.4.4** above), the residues 135 – 146 segments in the N-terminal N-core of the NiV N protein has also been identified as a crucial NiV P protein binding region, that is required for proper replication in a minigenome assay[195]. However, this residues 135 – 146 NiV P binding region of NiV N was not reproducible in a separate study by an independent group[56]. Both the residue 1-54 and 135-146 binding regions are located within the structured N-core domain of NiV N, whereas the residue 468-532 binding region is located as part of the unstructured Ntail of NiV N. Here the cryoEM results suggest another potential NiV P binding region within the N-terminal N-core domain of NiV N protein. It is possible that the binding of the NiV P protein to the NiV N-RNA protein prevented formation of the more "rigid" spiral assembly. Furthermore, it is also possible that the binding of the NiV P protein to the NiV N-RNA protein sterically blocks interaction regions that are required for the clam shaped formation, in turn resulting in fewer clam shaped assembly particles observed in this NiV N/P heterocomplex cryoEM dataset.

In Mumps virus (MuV), the N-terminal region of MuV P protein has been shown to interact with the helical N-RNA protein and induce the uncoiling of the helical N-RNA protein[196]. It is likely that the N-terminal region of the NiV P protein interacts with the helical N-RNA protein resulting in a looser helical assembly (Figure 4.13B). In contrast, the C-terminal region of the P protein, the X domain, is known to also interact with the helical nucleocapsid protein. However, it has been shown that, in both MuV and Hendra virus (HeV), this region did not trigger any major conformational changes in the helical nucleocapsid assembly[132,196]. The primary role of the C-terminal region of the P protein is to allow the L protein to recognize and bind to the helical RNP. This is

followed by the uncoiling of the helical RNP by the N-terminal region of P protein, allowing the L protein to access the viral genome[196]. While the results obtained here suggest that a single binding event of the NiV P to the NiV N-RNA spiral assembly led to the partial uncoiling of the NiV N-RNA spiral assembly, it is unknown if the binding of just a single NiV P to the helical RNP is sufficient to trigger the uncoiling of the helical RNP. Furthermore, not all NiV N-RNA particles in this dataset are in the "uncoiling" state, as NiV N-RNA in the rigid spiral assembly and clam shaped assembly are also observed. This suggests that the uncoiling state of the RNP is a transient conformational state that may require the binding of addition viral or host proteins, in order to maintain the uncoiling state for viral replication. Regardless, the preliminary result here will pave the way for future structural investigation of the NiV N-RNA protein in complex with the NiV P protein.

Chapter 5   Final Remarks and Conclusion

Throughout the viral infection cycle, the viral RNA genome of the non-segmented negative strand RNA virus is wrapped by the nucleocapsid (N) protein in a helical symmetry forming a long filament of ribonucleoprotein (RNP). This RNP serves as the functional template for RNA synthesis by the viral RNA dependent RNA polymerase (RdRp) complex. In addition, the RNP also provides a protective shield for the viral RNA against nuclease degradation and the host's immune response during the viral infection[133]. Several structural studies have been performed for the nucleocapsid like particles of several negative strand RNA viruses, including Measles virus[73], Ebola virus[151], Vesicular stomatitis virus[152] and Hantaan virus[197], describing the molecular role of the N protein in the N-RNA interaction that contributed to the RNP assembly. The work reported in this thesis further extends the existing insights on the RNP assembly of negative strand RNA viruses to the *Henipavirus* genus of *Paramyxoviridae* family.

In Paramyxovirus, the viral genome follows the "rule of six", where each of the nucleocapsid protomers is bound to six nucleotides in a "three-base-in and three-base-out" conformation[72,73]. Likewise, in NiV, the RNA is bound in a "three-base-in and three-base-out" conformation and this conformation is promoted by the steric hindrance by the bulky protein residues in combination with a series of protein-RNA interaction within the RNA binding region[131]. However, the *in vivo* functionality of this "three-base-in and three-base-out" conformation is unknown. It is likely that this conformation is important for the stability and initiation of the helical nucleocapsid formation, as this "three-base-in and three-base-out" RNA binding conformation is also observed in the Filoviruses family, which also binds to six nucleotide per nucleocapsid protomer[150]. However, compared to paramyxoviruses, filoviruses does not strictly follow the "rule of six" as varying the genome length of the Filovirus does not influence the activity of minigenomes[198]. This indicates that the "three-base-in and three-base-out" conformation, that is found in both virus families, does not directly regulate the genome length[199].

Interestingly, the NiV and other Paramyxovirus nucleocapsid-like-particles purified from bacterial expression systems are also found to assemble into a clam shaped assembly[131,134,146], in addition to the helical assemblies. No such assemblies have been reported yet for nucleocapsid-like-particles purified from eukaryotic expression systems. Similarly, the presence of clam shaped N protein assemblies was also not observed for *in vitro* assembled nucleocapsid-like samples, which are assembled using chemically synthesized RNA and N protein purified from *E. coli* expression system[113,200]. Therefore, it is possibly that the formation of a clam shaped assembly, found from bacterial expression systems, might be dependent on the nature of the encapsidated bacteria cellular RNA or the nature of the encapsidated NiV N gene mRNA transcript. Interestingly, only a small fraction of the nucleocapsid-like particles purified from bacterial co-expression of the NiV N and NiV P protein were the clam-shaped assembly, suggesting that the clam-shaped assembly can be easily manipulated by the presence of viral or host factors.

For the initiation of the RNA synthesis, the RNA polymerase has to bind to the 3' end of the encapsidated RNA genome and recognise the bipartite promoter. The co-factor phosphoprotein protein is known to function as a tethering protein to promote the interaction between the nucleocapsid (N) protein and RNA polymerase (L protein). Addition of the P protein to the helical N protein has been shown to result in the relaxation or uncoiling of the helical nucleocapsid assembly. In MuV, addition of the N-terminal region of MuV P protein resulted in the uncoiling of nucleocapsid helical assembly[196]. In line with previous discoveries, the cryoEM analysis of NiV N/P heterocomplex protein samples (section **4.2.4** above) also hinted at the presence of a more relaxed helical assembly, which was not observed in the cryoEM analysis of the protein sample of NiV N-RNA alone (section **3.2.2** above). This uncoiling event is likely to happen due to binding of the NiV P onto the N-terminal N-core region of the NiV N spiral assembly (Figure 4.15). However, due to the flexibility and heterogeneity of the NiV N/P heterocomplexes cryoEM dataset, the exact domain or structural feature of P protein that promotes this flexibility in the NiV N protein assembly could not be identified. In most of the Paramyxovirus, the current model for RdRp complex recruitment to the encapsidated RNA genome assumes an inaugural interaction between a MoRE domain in the N-tail and the X domain in the viral P protein[96].

However, the binding of the Hendra virus P-XD to the N-tail of the helical nucleocapsid did not result in any uncoiling or relaxation of the helical nucleocapsid, suggesting that the P-XD subdomain itself is insufficient to trigger the uncoiling of encapsidated genome[132].

In order for the viral RNA to be more accessible to the RdRp complex, a structural reorganisation of the helical nucleocapsid assembly is required. Based on the observation from this work and previously published works on other paramyxovirus replication complexes of other paramyxoviruses, a model for the structural rearrangement of the nucleocapsid assembly for viral RNA release was proposed (Figure 5.1). The L protein first interacts with the P protein to form an RdRp complex (L-P complex)[65]. The RdRp complex then binds to the 3' end of the viral genome and promotes the relaxation of the helical RNP assembly. Following this relaxation of the helical RNP assembly, the viral RNA, in the case of NiV, is locally released via the destabilization of the loop helix α8 (residue 174-200), and the 28° rotation of the N-terminal and C-terminal N-core domains away from each other[131]. Upon RNA release, the N protein is likely to retain its loose helical assembly[142], thus keeping the RNA-free N protein in place and ready to re-encapsidate the viral RNA after it has been transcribed by the RdRp complex. These local conformational changes occur throughout the rest of the viral genome as the RdRp complex propagates along the helical RNP assembly transcribing the viral RNA genome.
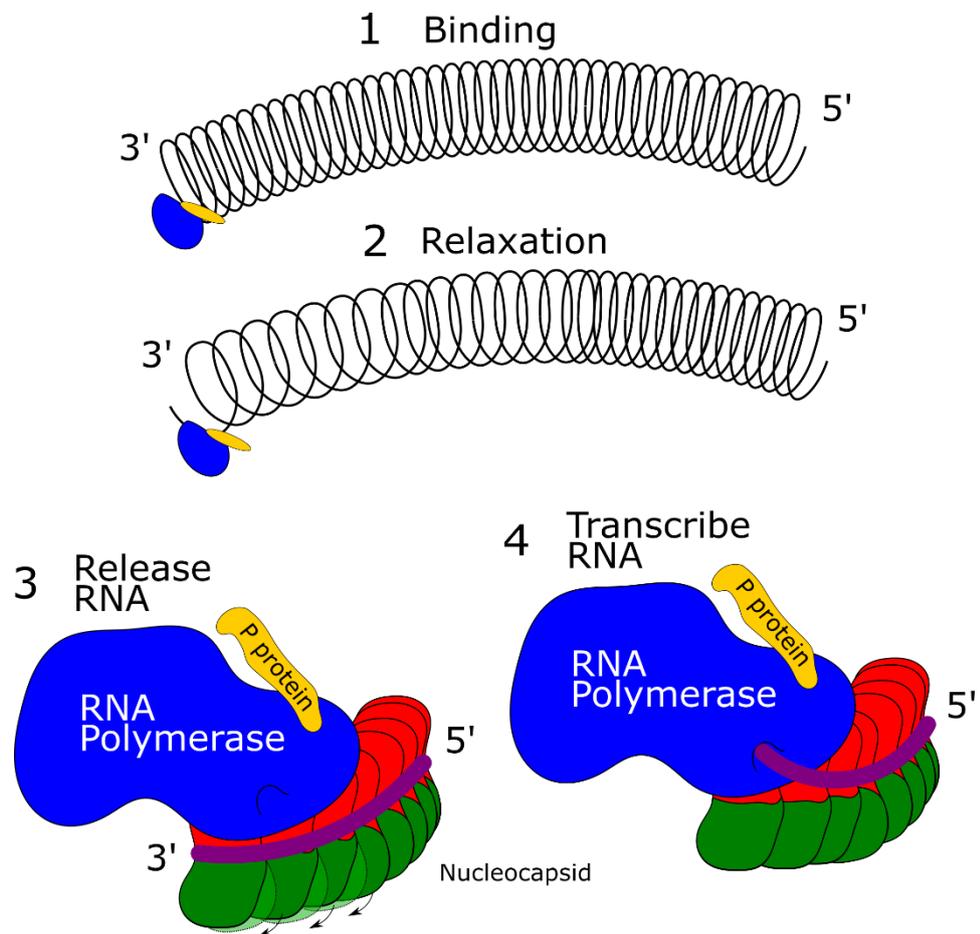
Figure 5.1 Model of the initiation step during viral RNA synthesis. The RNA polymerase L-P complex first attaches onto the 3' end of the encapsidated viral RNA genome. This interaction triggers the relaxation of the nucleocapsid helical assembly where the 3' end side of the helical assembly is looser than the 5' end. This is followed by the release of viral RNA from the nucleocapsid protein via N-terminal N-core domain rotation. The released viral RNA then enters the template entry tunnel of the RNA polymerase for transcription. It hypothesizes that the RNA-free nucleocapsid remains bound in its helical assembly, allowing the RNA-free nucleocapsid to re-encapsidate the viral RNA after it has been transcribed by the RNA polymerase.

The NiV N protein comprises an additional 150 aa long region exhibiting high levels of intrinsic disorder. The existence of an unstructured N-tail was found conserved among the paramyxovirus, bornavirus, and filoviruses but absent among the pneumovirus and rhabdovirus nucleocapsid proteins, questioning the necessity of the N-tail in Mononegavirus viral replication[92]. However, the complete removal of the N-tail in Paramyxovirus often resulted in abolishment of the nucleocapsid bioactivity[96], indicating that a Paramyxovirus utilises a slightly different viral replication mechanism to that of a pneumovirus and rhabdovirus. Minireplicon assays using MeV N protein demonstrated that the removal of just the MoRE region within the N-tail does not completely abolish the N bioactivity, indicating that the RdRp complex is able to

interact with other contact domains, aside from the MoRE region, in the N protein during the initial loading of RdRp complex onto RNP [93,96]. Whether this observation is applicable to NiV or Henipavirus in general, remains to be discovered. Differences in the N and P protein interactions among the Paramyxovirus family have been observed. For instance, in the case of the *Rubulavirinae* genus, which comprises the MuV, interaction between P and the N proteins occur not in the N-tail region, but are mediated by other contact domains of N protein[201]. Indeed, the repositioning of the MoRE region of MeV N from the N-tail to N-core domain still retained the bioactivity of a functional virus, demonstrated the positional flexibility of the MoRE region[92].

The N-tail region is also known to undergo phosphorylation[54,202,203]. Alanine mutation of the phosphorylation site in N protein has resulted in a decrease in minigenome activity for MeV and NiV[54,202], but caused an increase in minigenome activity for MuV[203], demonstrating the varied effect of phosphorylation of N protein among the Paramyxovirus family. Similarly, the Paramyxoviral P proteins are also phosphorylated, however the precise role of these modifications remain mostly elusive. The exception is the Parainfluenza virus 5 (PIV5) P protein, which shows a reduction in viral mRNA transcription upon alanine mutation on the phosphorylation site[204], while most of the mutations of the phosphorylation site in Paramyxoviral P protein do not result in any significant differences in terms of viral RNA synthesis[205]. In all cases, mutation of the phosphorylation sites does not disrupt the formation of the N/P heterocomplexes, indicating that the changes in the minigenome bioactivities are not related to the binding affinity between the RNP and P protein. The exact mechanism of the viral RNA synthesis regulation by the phosphorylation remains to be discovered. All the NiV N and P proteins described in this thesis were expressed recombinantly using an *E. coli* expression system, which has limited capacity for post translational modification. Recombinant expression of the NiV N and P protein using insect cell or mammalian cell expression systems could be used to promote proper phosphorylation of the NiV N and P proteins for future protein characterisation. It is speculated that the phosphorylation of the N or P proteins might enhance the association of the viral protein with a host protein, that promotes the viral RNA synthesis[204]. Aside from regulation by the aforementioned post translational modifications, the activity of the

viral replication can also be influenced by host factors, such as heat shock protein 90 (HSP90)[206] and tubulin[98], as well as the viral accessory C protein[87].

Infection of the host cell by paramyxoviruses is characterised by the formation of cytoplasmic inclusion bodies[207]. In MeV infected cells, these membraneless biomolecular condensates have been shown to serve as the site for viral replication by concentrating all the viral RNA replication machinery together, including the RNP genome, RdRp complex (L-P complex), and C protein[207–209]. The formation of MeV inclusion bodies has been associated with liquid-liquid phase separation (LLPS) which is initiated by the interaction between the N-tail and P proteins[179,188]. Post translational modifications have been associated with modification of the protein's LLPS behaviour. For instance, phosphorylation greatly regulates the dynamic of human ribonucleoprotein (RNP) granules, and it can have either a suppressive or promoting effect on the RNP granule formation via LLPS[210]. Given the growing evidence of phosphorylation site identified in the N and P proteins, it is likely that phosphorylation might also play a regulatory role in the formation of inclusion bodies which in turn influences the viral RNA synthesis. The LLPS formation is also heavily influenced by the cellular crowding and the interaction with host cellular macromolecules, making *in vitro* study of this process extremely challenging[26,209]. Future structural insights on the viral replication complexes in action directly in the virus infected cell will be extremely valuable in unveiling the mechanism of virus replication, and this might be possible to achieve in future by cryo-electron tomography (cryoET), owing to the ongoing technological advancement in this technique[211,212].

The regulation mechanism of the Paramyxovirus replication complexes is complicated and not well understood. Here I reported the cryoEM analysis of the NiV nucleocapsid-RNA complex and the NiV N/P heterocomplex. The stable assembly of the NiV N/P heterocomplex was only possible upon *in vivo* assembly of NiV N/P heterocomplex through co-expression of both proteins using *E. coli* expression systems. The cryoEM analysis of the NiV N/P heterocomplex suggested a potential NiV P binding region in the NiV N-RNA protein, that could trigger uncoiling of the RNP. However, the highly dynamic nature of the NiV N/P heterocomplex hampered the high-resolution structure determination for this important binding interaction. Future work in identifying the potential host cofactors to stabilise this dynamic complex can be carried out by co-

express NiV N and P protein in human cells and performing an affinity purification mass spectrometry[213]. Stabilisation of this dynamic protein complex by the use of chemical cross-linkers like glutaraldehyde[214,215] may also be helpful for biochemical characterisation, to provide better mechanistic understanding on Paramyxoviral RNA synthesis.

Abbreviations

Å	Angstrom
aa	Amino acid
BME	Beta-mercaptoethanol
CAPS	N-cyclohexyl-3-aminopropanesulfonic acid
CBB	Coomassie Brilliant Blue
CryoEM	Cryo electron microscopy
CTF	Contrast transfer function
CV	column volume
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EER	Electron Event Representation
ER	endoplasmic reticulum
ESCRT	endosomal sorting complex required for transport
EtBr	Ethidium Bromide
FSC	Fourier Shell Correlation
GE	Gene end
GS	Gene start
HEPES	4-(2-hydroxyelthyl)-1-piperazineethanesulfonic acid
HeV	Hendra virus
IFN	interferon
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KCl	Potassium chloride

kDa	kilo-Dalton
L protein	Large RNA polymerase protein
le	leader
LLPS	Liquid-liquid phase separation
MeV	Measles virus
MoRE	molecular recognition element
MuV	Mumps virus
MWCO	Molecular weight cut off
N protein	Nucleocapsid protein
NaCl	Sodium chloride
NDV	Newcastle disease virus
NiV	Nipah virus
nt	nucleotide
OD_{600}	Optical density at 600 nm
ORF	open reading frame
P protein	Phosphoprotein
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PIV5	Parainfluenza virus 5
RdRp	RNA dependent RNA polymerase
RNA	Ribonucleic acid
RNP	ribonucleoprotein
rpm	Rotation per minute

RT	Room temperature
SDS	Sodium dodecyl sulphate
SEC	Size exclusion chromatography
SeV	Sendai virus
STAT	signal transducer and activator of transcription
TEM	Transmission electron microscope
UTR	Untranslated region
UV	Ultraviolet
WHO	World Health Organisation
XD	X domain

Reference

1.   Ivanofsky, D. Concerning the Mosaic Disease of the Tobacco Plant. *St. Petersbg. Acad. Imp. Sci. Bull* **35**, 67–70 (1892).

2.   Beijerinck, M. W. Concerning a contagium vivum fluidum as a cause of the spot-disease of tobacco leaves. *Verh. Akad. Wetensch., Amsterdam, II* **6**, 3–21 (1898).

3.   Legendre, M. *et al.* Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proc. Natl. Acad. Sci.* **111**, 4274–4279 (2014).

4.   Rice, G. *et al.* From The Cover: The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. *Proc. Natl. Acad. Sci.* **101**, 7716–7720 (2004).

5.   Mahy, B. W. J. Introduction and History of Foot-and-Mouth Disease Virus. in *Foot-and-Mouth Disease Virus* 1–8 (Springer-Verlag). doi:10.1007/3-540-27109-0_1.

6.   Woolhouse, M., Scott, F., Hudson, Z., Howey, R. & Chase-Topping, M. Human viruses: discovery and emergence. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 2864–2871 (2012).

7.   Houser, K. & Subbarao, K. Influenza Vaccines: Challenges and Solutions. *Cell Host Microbe* **17**, 295–300 (2015).

8.   Nichol, S. *et al.* Genetic identification of a hantavirus associated with an outbreak of acute respiratory illness. *Science* **262**, 914–917 (1993).

9.   Chua, K. B. *et al.* Nipah virus: a recently emergent deadly paramyxovirus. *Science* **288**, 1432–5 (2000).

10.   LAM, W., ZHONG, N. & TAN, W. Overview on SARS in Asia and the World. *Respirology* **8**, S2–S5 (2003).

11.   Knoll, M. & Ruska, E. Das Elektronenmikroskop. *Zeitschrift für Phys.* **78**, 318–339 (1932).

12.   Kausche, G. A., Pfankuch, E. & Ruska, H. Die Sichtbarmachung von pflanzlichem Virus im Übermikroskop. *Naturwissenschaften* **27**, 292–299 (1939).

13.   Brenner, S. & Horne, R. W. A negative staining method for high resolution electron microscopy of viruses. *Biochim. Biophys. Acta* **34**, 103–110 (1959).

14.   Bowen, E. T. *et al.* Viral haemorrhagic fever in southern Sudan and northern

Zaire. Preliminary studies on the aetiological agent. *Lancet (London, England)* **1**, 571–3 (1977).

15. Caspar, D. L. D. & Klug, A. Physical Principles in the Construction of Regular Viruses. *Cold Spring Harb. Symp. Quant. Biol.* **27**, 1–24 (1962).

16. Burrell, C. J., Howard, C. R. & Murphy, F. A. Virion Structure and Composition. in *Fenner and White's Medical Virology* 27–37 (Elsevier, 2017). doi:10.1016/B978-0-12-375156-0.00003-5.

17. Xie, Y., Gong, J., Li, M., Fang, H. & Xu, W. The Medicinal Potential of Influenza Virus Surface Proteins: Hemagglutinin and Neuraminidase. *Curr. Med. Chem.* **18**, 1050–1066 (2011).

18. te Velthuis, A. J. W., Grimes, J. M. & Fodor, E. Structural insights into RNA polymerases of negative-sense RNA viruses. *Nat. Rev. Microbiol.* **19**, 303–318 (2021).

19. Philippe, N. *et al.* Pandoraviruses: Amoeba Viruses with Genomes Up to 2.5 Mb Reaching That of Parasitic Eukaryotes. *Science* **341**, 281–286 (2013).

20. Hoeben, R. C. & Uil, T. G. Adenovirus DNA Replication. *Cold Spring Harb. Perspect. Biol.* **5**, a013003–a013003 (2013).

21. Nathanson, N. & González-Scarano, F. Patterns of Infection. in *Viral Pathogenesis* 81–94 (Elsevier, 2016). doi:10.1016/B978-0-12-800964-2.00007-0.

22. Nagy, P. D. & Pogany, J. The dependence of viral RNA replication on co-opted host factors. *Nat. Rev. Microbiol.* **10**, 137–149 (2012).

23. Payne, S. Introduction to RNA Viruses. in *Viruses* 97–105 (Elsevier, 2017). doi:10.1016/B978-0-12-803109-4.00010-6.

24. Jaafar, Z. A. & Kieft, J. S. Viral RNA structure-based strategies to manipulate translation. *Nat. Rev. Microbiol.* **17**, 110–123 (2019).

25. Miller, S. & Krijnse-Locker, J. Modification of intracellular membrane structures for virus replication. *Nat. Rev. Microbiol.* **6**, 363–374 (2008).

26. Su, J. M., Wilson, M. Z., Samuel, C. E. & Ma, D. Formation and Function of Liquid-Like Viral Factories in Negative-Sense Single-Stranded RNA Virus Infections. *Viruses* **13**, 126 (2021).

27. Johara, M. Y. *et al.* Nipah Virus Infection in Bats (Order Chiroptera) in Peninsular Malaysia. *Emerg. Infect. Dis.* **7**, 439–441 (2001).

28. Gurley, E. S. *et al.* Person-to-Person Transmission of Nipah Virus in a Bangladeshi

Community. *Emerg. Infect. Dis.* **13**, 1031–1037 (2007).

29.   Parashar, U. D. *et al.* Case-Control Study of Risk Factors for Human Infection with a New Zoonotic Paramyxovirus, Nipah Virus, during a 1998–1999 Outbreak of Severe Encephalitis in Malaysia. *J. Infect. Dis.* **181**, 1755–1759 (2000).

30.   Sharma, V., Kaushik, S., Kumar, R., Yadav, J. P. & Kaushik, S. Emerging trends of Nipah virus: A review. *Rev. Med. Virol.* **29**, e2010 (2019).

31.   Ng, B.-Y., Lim, C. C. T., Yeoh, A. & Lee, W. L. Neuropsychiatric Sequelae of Nipah Virus Encephalitis. *J. Neuropsychiatry Clin. Neurosci.* **16**, 500–504 (2004).

32.   Harcourt, B. H. *et al.* Molecular Characterization of Nipah Virus, a Newly Emergent Paramyxovirus. *Virology* **271**, 334–349 (2000).

33.   Noton, S. L. & Fearns, R. Initiation and regulation of paramyxovirus transcription and replication. *Virology* **479–480**, 545–554 (2015).

34.   Walpita, P. & Peters, C. J. Cis-acting elements in the antigenomic promoter of Nipah virus. *J. Gen. Virol.* **88**, 2542–2551 (2007).

35.   Sugai, A., Sato, H., Yoneda, M. & Kai, C. Gene end-like sequences within the 3' non-coding region of the Nipah virus genome attenuate viral gene transcription. *Virology* **508**, 36–44 (2017).

36.   Bonaparte, M. I. *et al.* Ephrin-B2 ligand is a functional receptor for Hendra virus and Nipah virus. *Proc. Natl. Acad. Sci.* **102**, 10652–10657 (2005).

37.   Negrete, O. A. *et al.* EphrinB2 is the entry receptor for Nipah virus, an emergent deadly paramyxovirus. *Nature* **436**, 401–405 (2005).

38.   Negrete, O. A. *et al.* Two Key Residues in EphrinB3 Are Critical for Its Use as an Alternative Receptor for Nipah Virus. *PLoS Pathog.* **2**, e7 (2006).

39.   Diederich, S., Moll, M., Klenk, H.-D. & Maisner, A. The Nipah Virus Fusion Protein Is Cleaved within the Endosomal Compartment. *J. Biol. Chem.* **280**, 29899–29903 (2005).

40.   Ciancanelli, M. J. & Basler, C. F. Mutation of YMYL in the Nipah Virus Matrix Protein Abrogates Budding and Alters Subcellular Localization. *J. Virol.* **80**, 12070–12078 (2006).

41.   Bowden, T. A. *et al.* Crystal Structure and Carbohydrate Analysis of Nipah Virus Attachment Glycoprotein: a Template for Antiviral and Vaccine Design. *J. Virol.* **82**, 11628–11636 (2008).

42.   Zamora, J. L. R., Ortega, V., Johnston, G. P., Li, J. & Aguilar, H. C. Novel roles of

the N1 loop and N4 alpha helical region of the Nipah Virus Fusion glycoprotein in modulating early and late steps of the membrane fusion cascade. *J. Virol.* (2021) doi:10.1128/JVI.01707-20.

43. Pager, C. T. & Dutch, R. E. Cathepsin L Is Involved in Proteolytic Processing of the Hendra Virus Fusion Protein. *J. Virol.* **79**, 12714–12720 (2005).

44. Pager, C. T., Craft, W. W., Patch, J. & Dutch, R. E. A mature and fusogenic form of the Nipah virus fusion protein requires proteolytic processing by cathepsin L. *Virology* **346**, 251–257 (2006).

45. Popa, A. *et al.* Residues in the Hendra Virus Fusion Protein Transmembrane Domain Are Critical for Endocytic Recycling. *J. Virol.* **86**, 3014–3026 (2012).

46. Cox, R. M. & Plemper, R. K. Structure and organization of paramyxovirus particles. *Curr. Opin. Virol.* **24**, 105–114 (2017).

47. Dietzel, E. *et al.* Nipah Virus Matrix Protein Influences Fusogenicity and Is Essential for Particle Infectivity and Stability. *J. Virol.* **90**, 2514–2522 (2016).

48. McLinton, E. C. *et al.* Nuclear localization and secretion competence are conserved among henipavirus matrix proteins. *J. Gen. Virol.* **98**, 563–576 (2017).

49. Pentecost, M. *et al.* Evidence for Ubiquitin-Regulated Nuclear and Subnuclear Trafficking among Paramyxovirinae Matrix Proteins. *PLoS Pathog.* **11**, e1004739 (2015).

50. Wang, Y. E. *et al.* Ubiquitin-Regulated Nuclear-Cytoplasmic Trafficking of the Nipah Virus Matrix Protein Is Important for Viral Budding. *PLoS Pathog.* **6**, e1001186 (2010).

51. Bharaj, P. *et al.* The Matrix Protein of Nipah Virus Targets the E3-Ubiquitin Ligase TRIM6 to Inhibit the IKKɛ Kinase-Mediated Type-I IFN Antiviral Response. *PLOS Pathog.* **12**, e1005880 (2016).

52. Habchi, J., Mamelli, L., Darbon, H. & Longhi, S. Structural disorder within Henipavirus nucleoprotein and phosphoprotein: From predictions to experimental assessment. *PLoS One* **5**, (2010).

53. Yabukarski, F. *et al.* Structure of Nipah virus unassembled nucleoprotein in complex with its viral chaperone. *Nat. Struct. Mol. Biol.* **21**, 754–759 (2014).

54. Huang, M. *et al.* Determination of a phosphorylation site in Nipah virus nucleoprotein and its involvement in virus transcription. *J. Gen. Virol.* **92**, 2133–2141 (2011).

55.	Chan, Y. P., Koh, C. L., Lam, S. K. & Wang, L. F. Mapping of domains responsible for nucleocapsid protein-phosphoprotein interaction of henipaviruses. *J. Gen. Virol.* **85**, 1675–1684 (2004).

56.	Ranadheera, C. *et al.* The interaction between the Nipah virus nucleocapsid protein and phosphoprotein regulates virus replication. *Sci. Rep.* **8**, 15994 (2018).

57.	Ray, G., Schmitt, P. T. & Schmitt, A. P. C-Terminal DxD-Containing Sequences within Paramyxovirus Nucleocapsid Proteins Determine Matrix Protein Compatibility and Can Direct Foreign Proteins into Budding Particles. *J. Virol.* **90**, 3650–3660 (2016).

58.	Sugai, A., Sato, H., Takayama, I., Yoneda, M. & Kai, C. Nipah and Hendra Virus Nucleoproteins Inhibit Nuclear Accumulation of Signal Transducer and Activator of Transcription 1 (STAT1) and STAT2 by Interfering with Their Complex Formation. *J. Virol.* **91**, 1–15 (2017).

59.	Shiell, B. J., Gardner, D. R., Crameri, G., Eaton, B. T. & Michalski, W. P. Sites of phosphorylation of P and V proteins from Hendra and Nipah viruses: newly emerged members of Paramyxoviridae. *Virus Res.* **92**, 55–65 (2003).

60.	Ciancanelli, M. J., Volchkova, V. A., Shaw, M. L., Volchkov, V. E. & Basler, C. F. Nipah Virus Sequesters Inactive STAT1 in the Nucleus via a P Gene-Encoded Mechanism. *J. Virol.* **83**, 7828–7841 (2009).

61.	Ludlow, L. E., Lo, M. K., Rodriguez, J. J., Rota, P. A. & Horvath, C. M. Henipavirus V Protein Association with Polo-Like Kinase Reveals Functional Overlap with STAT1 Binding and Interferon Evasion. *J. Virol.* **82**, 6259–6271 (2008).

62.	Rodriguez, J. J., Cruz, C. D. & Horvath, C. M. Identification of the Nuclear Export Signal and STAT-Binding Domains of the Nipah Virus V Protein Reveals Mechanisms Underlying Interferon Evasion. *J. Virol.* **78**, 5358–5367 (2004).

63.	Yamaguchi, M., Kitagawa, Y., Zhou, M., Itoh, M. & Gotoh, B. An anti-interferon activity shared by paramyxovirus C proteins: Inhibition of Toll-like receptor 7/9-dependent alpha interferon induction. *FEBS Lett.* **588**, 28–34 (2014).

64.	Park, A. *et al.* Nipah Virus C Protein Recruits Tsg101 to Promote the Efficient Release of Virus in an ESCRT-Dependent Pathway. *PLOS Pathog.* **12**, e1005659 (2016).

65.	Abdella, R., Aggarwal, M., Okura, T., Lamb, R. A. & He, Y. Structure of a

paramyxovirus polymerase complex reveals a unique methyltransferase-CTD conformation. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 4931–4941 (2020).

66. Magoffin, D. E., Halpin, K., Rota, P. A. & Wang, L.-F. Effects of single amino acid substitutions at the E residue in the conserved GDNE motif of the Nipah virus polymerase (L) protein. *Arch. Virol.* **152**, 827–832 (2007).

67. Jordan, P. C. *et al.* Initiation, extension, and termination of RNA synthesis by a paramyxovirus polymerase. *PLoS Pathog.* **14**, e1006889 (2018).

68. Paesen, G. C. *et al.* X-ray structure and activities of an essential Mononegavirales L-protein domain. *Nat. Commun.* **6**, 1–10 (2015).

69. Luo, M., Terrell, J. R. & Mcmanus, S. A. Nucleocapsid Structure of Negative Strand RNA Virus. *Viruses* **12**, (2020).

70. Brown, J. C., Newcomb, W. W. & Wertz, G. W. Helical Virus Structure: The Case of the Rhabdovirus Bullet. *Viruses* **2**, 995–1001 (2010).

71. Riedel, C., Hennrich, A. A. & Conzelmann, K.-K. Components and Architecture of the Rhabdovirus Ribonucleoprotein Complex. *Viruses* **12**, 959 (2020).

72. Alayyoubi, M., Leser, G. P., Kors, C. A. & Lamb, R. A. Structure of the paramyxovirus parainfluenza virus 5 nucleoprotein–RNA complex. *Proc. Natl. Acad. Sci.* **112**, E1792–E1799 (2015).

73. Gutsche, I. *et al.* Near-atomic cryo-EM structure of the helical measles virus nucleocapsid. *Science* **348**, 704–707 (2015).

74. Blumberg, B. M., Leppert, M. & Kolakofsky, D. Interaction of VSV leader RNA and nucleocapsid protein may control VSV genome replication. *Cell* **23**, 837–845 (1981).

75. Tapparel, C., Maurice, D. & Roux, L. The Activity of Sendai Virus Genomic and Antigenomic Promoters Requires a Second Element Past the Leader Template Regions: a Motif (GNNNNN) 3 Is Essential for Replication. *J. Virol.* **72**, 3117–3128 (1998).

76. Kolakofsky, D. Paramyxovirus RNA synthesis, mRNA editing, and genome hexamer phase: A review. *Virology* **498**, 94–98 (2016).

77. Ogino, T., Kobayashi, M., Iwama, M. & Mizumoto, K. Sendai Virus RNA-dependent RNA Polymerase L Protein Catalyzes Cap Methylation of Virus-specific mRNA. *J. Biol. Chem.* **280**, 4429–4435 (2005).

78. Whelan, S. P. J., Barr, J. N. & Wertz, G. W. Transcription and Replication of

Nonsegmented Negative-Strand RNA Viruses. in 61–119 (2004). doi:10.1007/978-3-662-06099-5_3.

79. Conzelmann, K. K. Nonsegmented negative-strand RNA viruses: genetics and manipulation of viral genomes. *Annu. Rev. Genet.* **32**, 123–62 (1998).

80. Vidal, S., Curran, J. & Kolakofsky, D. A stuttering model for paramyxovirus P mRNA editing. *EMBO J.* **9**, 2017–2022 (1990).

81. Kolakofsky, D., Le Mercier, P., Iseni, F. & Garcin, D. Viral RNA polymerase scanning and the gymnastics of Sendai virus RNA synthesis. *Virology* **318**, 463–473 (2004).

82. Lo, M. K., Søgaard, T. M. & Karlin, D. G. Evolution and Structural Organization of the C Proteins of Paramyxovirinae. *PLoS One* **9**, e90003 (2014).

83. Keller, M. A. & Parks, G. D. Positive- and negative-acting signals combine to determine differential RNA replication from the paramyxovirus simian virus 5 genomic and antigenomic promoters. *Virology* **306**, 347–358 (2003).

84. Welch, S. R. *et al.* Inhibition of Nipah Virus by Defective Interfering Particles. *J. Infect. Dis.* **221**, S460–S470 (2020).

85. Lazzarini, R. A., Keene, J. D. & Schubert, M. The origins of defective interfering particles of the negative-strand RNA viruses. *Cell* **26**, 145–154 (1981).

86. Manzoni, T. B. & López, C. B. Defective (interfering) viral genomes re-explored: impact on antiviral immunity and virus persistence. *Future Virol.* **13**, 493–503 (2018).

87. Pfaller, C. K., Radeke, M. J., Cattaneo, R. & Samuel, C. E. Measles Virus C Protein Impairs Production of Defective Copyback Double-Stranded Viral RNA and Activation of Protein Kinase R. *J. Virol.* **88**, 456–468 (2014).

88. Vidal, S. & Kolakofsky, D. Modified model for the switch from Sendai virus transcription to replication. *J. Virol.* **63**, 1951–1958 (1989).

89. Horikami, S. M., Hector, R. E., Smallwood, S. & Moyer, S. A. The Sendai Virus C Protein Binds the L Polymerase Protein to Inhibit Viral RNA Synthesis. *Virology* **235**, 261–270 (1997).

90. Grogan, C. C. & Moyer, S. A. Sendai Virus Wild-Type and Mutant C Proteins Show a Direct Correlation between L Polymerase Binding and Inhibition of Viral RNA Synthesis. *Virology* **288**, 96–108 (2001).

91. Irie, T., Nagata, N., Yoshida, T. & Sakaguchi, T. Paramyxovirus Sendai virus C

proteins are essential for maintenance of negative-sense RNA genome in virus particles. *Virology* **374**, 495–505 (2008).

92. Cox, R. M., Krumm, S. A., Thakkar, V. D., Sohn, M. & Plemper, R. K. The structurally disordered paramyxovirus nucleocapsid protein tail domain is a regulator of the mRNA transcription gradient. *Sci. Adv.* **3**, (2017).

93. Thakkar, V. D. *et al.* The Unstructured Paramyxovirus Nucleocapsid Protein Tail Domain Modulates Viral Pathogenesis through Regulation of Transcriptase Activity. *J. Virol.* **92**, JVI.02064-17 (2018).

94. Kingston, R. L., Hamel, D. J., Gay, L. S., Dahlquist, F. W. & Matthews, B. W. Structural basis for the attachment of a paramyxoviral polymerase to its template. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 8301–8306 (2004).

95. Pont, V. Du, Jiang, Y. & Plemper, R. K. Bipartite interface of the measles virus phosphoprotein X domain with the large polymerase protein regulates viral polymerase dynamics. *PLoS Pathog.* **15**, 1–27 (2019).

96. Krumm, S. A., Takeda, M. & Plemper, R. K. The Measles Virus Nucleocapsid Protein Tail Domain Is Dispensable for Viral Polymerase Recruitment and Activity. *J. Biol. Chem.* **288**, 29943–29953 (2013).

97. Bloyet, L.-M. *et al.* Modulation of Re-initiation of Measles Virus Transcription at Intergenic Regions by PXD to NTAIL Binding Strength. *PLoS Pathog.* **12**, e1006058 (2016).

98. Moyer, S. A., Baker, S. C. & Lessard, J. L. Tubulin: a factor necessary for the synthesis of both Sendai virus and vesicular stomatitis virus RNAs. *Proc. Natl. Acad. Sci.* **83**, 5405–5409 (1986).

99. Moyer, S. A., Baker, S. C. & Horikami, S. M. Host cell proteins required for measles virus reproduction. *J. Gen. Virol.* **71**, 775–783 (1990).

100. Cheng, Y. *et al.* Single Particle Reconstructions of the Transferrin–Transferrin Receptor Complex Obtained with Different Specimen Preparation Techniques. *J. Mol. Biol.* **355**, 1048–1065 (2006).

101. Dubochet, J. *et al.* Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.* **21**, 129–228 (1988).

102. Villarreal, S. A. & Stewart, P. L. CryoEM and image sorting for flexible protein/DNA complexes. *J. Struct. Biol.* **187**, 76–83 (2014).

103. McMullan, G., Faruqi, A. R. & Henderson, R. Direct Electron Detectors. in

Methods in Enzymology 1–17 (2016). doi:10.1016/bs.mie.2016.05.056.

104. Darrow, M. C. *et al.* Enabling a Paradigm Shift in CryoEM Sample Preparation with chameleon. *Microsc. Microanal.* **27**, 524–525 (2021).

105. Ravelli, R. B. G. *et al.* Cryo-EM structures from sub-nl volumes using pin-printing and jet vitrification. *Nat. Commun.* **11**, 2563 (2020).

106. Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).

107. Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat. Methods* **18**, 176–185 (2021).

108. Punjani, A. & Fleet, D. J. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *J. Struct. Biol.* **213**, 107702 (2021).

109. Blocquel, D., Beltrandi, M., Erales, J., Barbier, P. & Longhi, S. Biochemical and structural studies of the oligomerization domain of the Nipah virus phosphoprotein: Evidence for an elongated coiled-coil homotrimer. *Virology* **446**, 162–172 (2013).

110. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, (2018).

111. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

112. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

113. Desfosses, A. *et al.* Assembly and cryo-EM structures of RNA-specific measles virus nucleocapsids provide mechanistic insight into paramyxoviral replication. *Proc. Natl. Acad. Sci.* **116**, 4256–4264 (2019).

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

115. He, S. Helical reconstruction in RELION. (University of Cambridge, 2018).

116. Shakeel, S. *et al.* Structure of the Fanconi anaemia monoubiquitin ligase complex. *Nature* **575**, 234–237 (2019).

117. Zivanov, J., Nakane, T. & Scheres, S. H. W. Estimation of High-Order Aberrations and Anisotropic Magnification from Cryo-EM Datasets in RELION-3.1. *bioRxiv*

(2019) doi:10.1101/798066.

118. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).

119. Leman, J. K. *et al.* Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat. Methods* **17**, 665–680 (2020).

120. Croll, T. I. ISOLDE : a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 519–530 (2018).

121. Murshudov, G. N. *et al.* REFMAC 5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 355–367 (2011).

122. Vangone, A., Spinelli, R., Scarano, V., Cavallo, L. & Oliva, R. COCOMAPS: a web application to analyze and visualize contacts at the interface of biomolecular complexes. *Bioinformatics* **27**, 2915–2916 (2011).

123. Lee, R. A., Razaz, M. & Hayward, S. The DynDom database of protein domain motions. *Bioinformatics* **19**, 1290–1291 (2003).

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

125. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

126. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).

127. Campanella, J. J., Bitincka, L. & Smalley, J. MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences. *BMC Bioinformatics* **4**, 29 (2003).

128. Ashkenazy, H. *et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016).

129. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).

130. Unni, S. *et al.* Web servers and services for electrostatics calculations with APBS and PDB2PQR. *J. Comput. Chem.* **32**, 1488–1491 (2011).

131. Ker, D.-S., Jenkins, H. T., Greive, S. J. & Antson, A. A. CryoEM structure of the Nipah virus nucleocapsid assembly. *PLoS Pathog.* **17**, e1009740 (2021).

132. Communie, G. *et al.* Atomic Resolution Description of the Interaction between the Nucleoprotein and Phosphoprotein of Hendra Virus. *PLoS Pathog.* **9**, e1003631 (2013).

133. Nilsson-Payant, B. E. *et al.* Reduced Nucleoprotein Availability Impairs Negative-Sense RNA Virus Replication and Promotes Host Recognition. *J. Virol.* **95**, (2021).

134. Song, X. *et al.* Self-capping of nucleoprotein filaments protects the Newcastle disease virus genome. *Elife* **8**, 1–19 (2019).

135. Tan, W. S., Ong, S. T., Eshaghi, M., Foo, S. S. & Yusoff, K. Solubility, Immunogenicity and Physical Properties of the Nucleocapsid Protein of Nipah Virus Produced in Escherichia coli. *J. Med. Virol.* **73**, 105–112 (2004).

136. Joseph, N. M. S., Tey, B. T., Tan, C. S., Shafee, N. & Tan, W. S. Production of long helical capsid of Nipah virus by Pichia pastoris. *Process Biochem.* **46**, 1871–1874 (2011).

137. Eshaghi, M., Tan, W. S., Ong, S. T. & Yusoff, K. Purification and Characterization of Nipah Virus Nucleocapsid Protein Produced in Insect Cells. *J. Clin. Microbiol.* **43**, 3172–3177 (2005).

138. Porterfield, J. Z. & Zlotnick, A. A simple and general method for determining the protein and nucleic acid content of viruses by UV absorbance. *Virology* **407**, 281–288 (2010).

139. Chong, F. C., Tan, W. S., Biak, D. R. A., Ling, T. C. & Tey, B. T. Modulation of protease activity to enhance the recovery of recombinant nucleocapsid protein of Nipah virus. *Process Biochem.* **45**, 133–137 (2010).

140. Heggeness, M. H., Scheid, A. & Choppin, P. W. Conformation of the helical nucleocapsids of paramyxoviruses and vesicular stomatitis virus: reversible coiling and uncoiling induced by changes in salt concentration. *Proc. Natl. Acad. Sci.* **77**, 2631–2635 (1980).

141. Lynch, S. & Kolakofsky, D. Ends of the RNA within Sendai virus defective interfering nucleocapsids are not free. *J. Virol.* **28**, 584–589 (1978).

142. Severin, C. *et al.* Releasing the Genomic RNA Sequestered in the Mumps Virus Nucleocapsid. *J. Virol.* **90**, 10113–10119 (2016).

143. Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and

crystallography. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 531–544 (2018).

144. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).

145. Milles, S. *et al.* Self-Assembly of Measles Virus Nucleocapsid-like Particles: Kinetics and RNA Sequence Dependence. *Angew. Chemie - Int. Ed.* **55**, 9356–9360 (2016).

146. Zhang, N. *et al.* Structure and assembly of double-headed Sendai virus nucleocapsids. *Commun. Biol.* **4**, 494 (2021).

147. Cordes, F. S., Bright, J. N. & Sansom, M. S. P. Proline-induced Distortions of Transmembrane Helices. *J. Mol. Biol.* **323**, 951–960 (2002).

148. Green, T. J. *et al.* Common Mechanism for RNA Encapsidation by Negative-Strand RNA Viruses. *J. Virol.* **88**, 3766–3775 (2014).

149. Tabor, S. Expression Using the T7 RNA Polymerase/Promoter System. *Curr. Protoc. Mol. Biol.* **11**, (1990).

150. Sugita, Y., Matsunami, H., Kawaoka, Y., Noda, T. & Wolf, M. Cryo-EM structure of the Ebola virus nucleoprotein–RNA complex at 3.6 Å resolution. *Nature* (2018) doi:10.1038/s41586-018-0630-0.

151. Wan, W. *et al.* Structure and assembly of the Ebola virus nucleocapsid. *Nature* **551**, 394–397 (2017).

152. Green, T. J., Zhang, X., Wertz, G. W. & Luo, M. Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science* **313**, 357–60 (2006).

153. Albertini, A. A. V. V. *et al.* Crystal Structure of the Rabies Virus Nucleoprotein-RNA Complex. *Science* **313**, 360–363 (2006).

154. Ge, P. *et al.* Cryo-EM model of the bullet-shaped vesicular stomatitis virus. *Science* **327**, 689–93 (2010).

155. Riedel, C. *et al.* Cryo EM structure of the rabies virus ribonucleoprotein complex. *Sci. Rep.* **9**, 9639 (2019).

156. Su, Z. *et al.* Electron Cryo-microscopy Structure of Ebola Virus Nucleoprotein Reveals a Mechanism for Nucleocapsid-like Assembly. *Cell* **172**, 966-978.e12 (2018).

157. Tawar, R. G. *et al.* Crystal Structure of a Nucleocapsid-Like Nucleoprotein-RNA Complex of Respiratory Syncytial Virus. *Science* **326**, 1279–1283 (2009).

158. Renner, M. *et al.* Structural dissection of human metapneumovirus

phosphoprotein using small angle x-ray scattering. *Sci. Rep.* **7**, 1–16 (2017).

159. Renner, M. *et al.* Nucleocapsid assembly in pneumoviruses is regulated by conformational switching of the N protein. *Elife* **5**, 1–12 (2016).

160. Leyrat, C. *et al.* Structure of the Vesicular Stomatitis Virus N0-P Complex. *PLoS Pathog.* **7**, e1002248 (2011).

161. Bhella, D., Ralph, A. & Yeo, R. P. Conformational Flexibility in Recombinant Measles Virus Nucleocapsids Visualised by Cryo-negative Stain Electron Microscopy and Real-space Helical Reconstruction. *J. Mol. Biol.* **340**, 319–331 (2004).

162. Habchi, J. *et al.* Characterization of the Interactions between the Nucleoprotein and the Phosphoprotein of Henipavirus. *J. Biol. Chem.* **286**, 13583–13602 (2011).

163. Blair, C. D. & Robinson, W. S. Replication of Sendai Virus. *J. Virol.* **5**, 639–650 (1970).

164. Mottet, G. & Roux, L. Budding efficiency of Sendai virus nucleocapsids: influence of size and ends of the RNA. *Virus Res.* **14**, 175–187 (1989).

165. Castaneda, S. J. & Wong, T. C. Leader sequence distinguishes between translatable and encapsidated measles virus RNAs. *J. Virol.* **64**, 222–230 (1990).

166. Loney, C., Mottet-Osman, G., Roux, L. & Bhella, D. Paramyxovirus Ultrastructure and Genome Packaging: Cryo-Electron Tomography of Sendai Virus. *J. Virol.* **83**, 8191–8197 (2009).

167. Hosaka, Y. & Shimizu, K. Lengths of the nucleocapsids of Newcastle disease and mumps viruses. *J. Mol. Biol.* **35**, 369–373 (1968).

168. Myers, T. M. & Moyer, S. A. An amino-terminal domain of the Sendai virus nucleocapsid protein is required for template function in viral RNA synthesis. *J. Virol.* **71**, 918–924 (1997).

169. Salvamani, S., Tey, B. T., Ng, W. C. & Tan, W. S. Production and purification of the phosphoprotein of Nipah virus in Escherichia coli for use in diagnostic assays. *Biotechnol. Bioprocess Eng.* **16**, 1166–1172 (2011).

170. Schiavina, M. *et al.* Ensemble description of the intrinsically disordered N-terminal domain of the Nipah virus P/V protein from combined NMR and SAXS. *Sci. Rep.* **10**, 19574 (2020).

171. Habchi, J., Mamelli, L., Darbon, H. & Longhi, S. Structural Disorder within Henipavirus Nucleoprotein and Phosphoprotein: From Predictions to

Experimental Assessment. *PLoS One* **5**, e11684 (2010).

172. Alves, V. S., Pimenta, D. C., Sattlegger, E. & Castilho, B. A. Biophysical characterization of Gir2, a highly acidic protein of Saccharomyces cerevisiae with anomalous electrophoretic behavior. *Biochem. Biophys. Res. Commun.* **314**, 229–234 (2004).

173. Guan, Y. *et al.* An equation to estimate the difference between theoretically predicted and SDS PAGE-displayed molecular weights for an acidic peptide. *Sci. Rep.* **5**, 13370 (2015).

174. Bruhn, J. F. *et al.* Crystal Structure of the Nipah Virus Phosphoprotein Tetramerization Domain. *J. Virol.* **88**, 758–762 (2014).

175. Jensen, M. R. *et al.* Structural Description of the Nipah Virus Phosphoprotein and Its Interaction with STAT1. *Biophys. J.* **118**, 2470–2488 (2020).

176. Holmes, D. L. & Stellwagen, N. C. Estimation of polyacrylamide gel pore size from Ferguson plots of linear DNA fragments. II. Comparison of gels with different crosslinker concentrations, added agarose and added linear polyacrylamide. *Electrophoresis* **12**, 612–619 (1991).

177. Goel, A. *et al.* Relative position of the hexahistidine tag effects binding properties of a tumor-associated single-chain Fv construct. *Biochim. Biophys. Acta - Gen. Subj.* **1523**, 13–20 (2000).

178. Majorek, K. A., Kuhn, M. L., Chruszcz, M., Anderson, W. F. & Minor, W. Double trouble-Buffer selection and His-tag presence may be responsible for nonreproducibility of biomedical experiments. *Protein Sci.* **23**, 1359–1368 (2014).

179. Guseva, S. *et al.* Measles virus nucleo- And phosphoproteins form liquid-like phase-separated compartments that promote nucleocapsid assembly. *Sci. Adv.* **6**, 1–12 (2020).

180. Ringel, M. *et al.* Nipah virus induces two inclusion body populations: Identification of novel inclusions at the plasma membrane. *PLoS Pathog.* **15**, e1007733 (2019).

181. Korolev, N. How potassium came to be the dominant biological cation: of metabolism, chemiosmosis, and cation selectivity since the beginnings of life. *BioEssays* **43**, 2000108 (2021).

182. Guo, H. *et al.* Electron-event representation data enable efficient cryoEM file

storage with full preservation of spatial and temporal resolution. *IUCrJ* **7**, 860–869 (2020).

183. Jonić, S. Cryo-electron Microscopy Analysis of Structurally Heterogeneous Macromolecular Complexes. *Comput. Struct. Biotechnol. J.* **14**, 385–390 (2016).

184. Kutnowski, N. *et al.* Specificity of protein–DNA interactions in hypersaline environment: structural studies on complexes of Halobacterium salinarum oxidative stress-dependent protein hsRosR. *Nucleic Acids Res.* (2019) doi:10.1093/nar/gkz604.

185. Alberti, S., Gladfelter, A. & Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419–434 (2019).

186. Crowe, C. D. & Keating, C. D. Liquid–liquid phase separation in artificial cells. *Interface Focus* **8**, 20180032 (2018).

187. Orti, F., Navarro, A. M., Rabinovich, A., Wodak, S. J. & Marino-Buslje, C. Insight into membraneless organelles and their associated proteins: Drivers, Clients and Regulators. *Comput. Struct. Biotechnol. J.* **19**, 3964–3977 (2021).

188. Zhou, Y., Su, J. M., Samuel, C. E. & Ma, D. Measles Virus Forms Inclusion Bodies with Properties of Liquid Organelles. *J. Virol.* **93**, (2019).

189. Pfaller, C. K., Cattaneo, R. & Schnell, M. J. Reverse genetics of Mononegavirales: How they work, new vaccines, and new cancer therapeutics. *Virology* **479–480**, 331–344 (2015).

190. Ito, N. *et al.* Role of Interferon Antagonist Activity of Rabies Virus Phosphoprotein in Viral Pathogenicity. *J. Virol.* **84**, 6699–6710 (2010).

191. Paul, N. K., Baksh, K. A., Arias, J. F. & Zamble, D. B. The impact of a His-tag on DNA binding by RNA polymerase alpha-C-terminal domain from Helicobacter pylori. *Protein Expr. Purif.* **167**, 105541 (2020).

192. Cox, R. *et al.* Characterization of a Mumps Virus Nucleocapsidlike Particle. *J. Virol.* **83**, 11402–11406 (2009).

193. André, A. A. M. & Spruijt, E. Liquid–Liquid Phase Separation in Crowded Environments. *Int. J. Mol. Sci.* **21**, 5908 (2020).

194. Roque, A., Ponte, I. & Suau, P. Macromolecular Crowding Induces a Molten Globule State in the C-Terminal Domain of Histone H1. *Biophys. J.* **93**, 2170–2177 (2007).

195. Omi-Furutani, M., Yoneda, M., Fujita, K., Ikeda, F. & Kai, C. Novel Phosphoprotein-Interacting Region in Nipah Virus Nucleocapsid Protein and Its Involvement in Viral Replication. *J. Virol.* **84**, 9793–9799 (2010).

196. Cox, R. M. *et al.* Structural studies on the authentic mumps virus nucleocapsid showing uncoiling by the phosphoprotein. *Proc. Natl. Acad. Sci.* **111**, 15208–15213 (2014).

197. Arragain, B. *et al.* High resolution cryo-EM structure of the helical RNA-bound Hantaan virus nucleocapsid reveals its assembly mechanisms. *Elife* **8**, 1–14 (2019).

198. Mühlberger, E. Filovirus replication and transcription. *Future Virol.* **2**, 205–215 (2007).

199. Gutsche, I., le Mercier, P. & Kolakofsky, D. A paramyxovirus-like model for Ebola virus bipartite promoters. *PLoS Pathog.* **16**, e1008972 (2020).

200. Zinzula, L. *et al.* Cryo-EM structure of the cetacean morbillivirus nucleoprotein-RNA complex. *J. Struct. Biol.* **213**, 107750 (2021).

201. Kingston, R. L., Baase, W. A. & Gay, L. S. Characterization of Nucleocapsid Binding by the Measles Virus and Mumps Virus Phosphoproteins. *J. Virol.* **79**, 10097–10097 (2005).

202. Hagiwara, K. *et al.* Phosphorylation of measles virus nucleoprotein upregulates the transcriptional activity of minigenomic RNA. *Proteomics* **8**, 1871–1879 (2008).

203. Zengel, J., Pickar, A., Xu, P., Lin, A. & He, B. Roles of Phosphorylation of the Nucleocapsid Protein of Mumps Virus in Regulating Viral RNA Transcription and Replication. *J. Virol.* **89**, 7338–7347 (2015).

204. Sun, D., Luthra, P., Xu, P., Yoon, H. & He, B. Identification of a Phosphorylation Site within the P Protein Important for mRNA Transcription and Growth of Parainfluenza Virus 5. *J. Virol.* **85**, 8376–8385 (2011).

205. Fuentes, S. M., Sun, D., Schmitt, A. P. & He, B. Phosphorylation of paramyxovirus phosphoprotein and its role in viral gene expression. *Future Microbiol.* **5**, 9–13 (2010).

206. Katoh, H. *et al.* The R2TP complex regulates paramyxovirus RNA synthesis. *PLoS Pathog.* **15**, e1007749 (2019).

207. Dolnik, O., Gerresheim, G. K. & Biedenkopf, N. New Perspectives on the

Biogenesis of Viral Inclusion Bodies in Negative-Sense RNA Virus Infections. *Cells* **10**, 1460 (2021).

208. Pfaller, C. K. *et al.* The C Protein Is Recruited to Measles Virus Ribonucleocapsids by the Phosphoprotein. *J. Virol.* **94**, 1–15 (2019).

209. Ma, D. *et al.* Upon Infection, Cellular WD Repeat-Containing Protein 5 (WDR5) Localizes to Cytoplasmic Inclusion Bodies and Enhances Measles Virus Replication. *J. Virol.* **92**, (2018).

210. Hofweber, M. & Dormann, D. Friend or foe—Post-translational modifications as regulators of phase separation and RNP granule dynamics. *J. Biol. Chem.* **294**, 7137–7150 (2019).

211. Zhang, P. Advances in cryo-electron tomography and subtomogram averaging and classification. *Curr. Opin. Struct. Biol.* **58**, 249–258 (2019).

212. Li, X. Cryo-electron tomography: observing the cell at the atomic level. *Nat. Methods* **18**, 440–441 (2021).

213. Luo, Y. & Muesing, M. A. Mass spectrometry-based proteomic approaches for discovery of HIV–host interactions. *Future Virol.* **9**, 979–992 (2014).

214. Stark, H. GraFix: Stabilization of Fragile Macromolecular Complexes for Single Particle Cryo-EM. in 109–126 (2010). doi:10.1016/S0076-6879(10)81005-5.

215. Adamus, K., Le, S. N., Elmlund, H., Boudes, M. & Elmlund, D. AgarFix: Simple and accessible stabilization of challenging single-particle cryo-EM specimens through crosslinking in a matrix of agar. *J. Struct. Biol.* **207**, 327–331 (2019).