Structure and Function of RNA Binding Proteins from Negative Sense-RNA Viruses

Georgia Marie Pangratiou

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds
The Astbury Centre for Molecular and Cellular Biology

October 2020

Acknowledgements

Firstly, I would like to thank everyone involved in the collaborative Bunyaviral project. Thanks go to all contributing authors for their help towards my first named author paper and the contribution of their work in this thesis in Chapter 4.

A special thanks to my supervision team Professor Thomas Edwards, Dr. John Barr and Professor Julian Hiscox for their advice and guidance. All three were instrumental in driving this PhD forward throughout the years and I wouldn't have finished without their enthusiasm and support.

I would like to thank members of the School of Molecular and Cellular Biology 8.59, Virology 8.61 and the wider Astbury Centre, for their continuous guidance and support and use of their equipment. The work presented here would not be possible without the use of Astbury Centre facilities and their members of staff. Namely, I would like to thank Dr. Chi Trinh for his guidance during the X-ray crystallography process and Dr. Nasir Khan aid during circular dichroism experiments. Moreover, Dr. Rachel George for performing mass-spectrometry experiments and analysis. Specifically, I would like to thank Dr. Selvaraj Muniyandi for staying up overnight with me to guide me through collecting X-ray diffraction data and Dr. Francis Hopkins who would always sit in the dark EM rooms with me whilst I collected data. Francis was also fundamental in the processing of micrographs and 2D data collections. I would also like to thank Sue Matthews who was our laboratory manager and technician and made an abundance of reagents readily available for me. Lastly, Timen Passchier my 'work husband' who I kept constantly distracted whilst in the office but whom kept me sane.

I wish to thank all of my family for their support and encouragement through the highs and lows of my PhD. Without the support of my parents, Andy and Sheilagh Pangratiou, I wouldn't have finished this PhD. Thank you for always trying to pretend you understand my work when I sent you documents to read. Thank you to my aunty, Anna Pangratiou and my uncle, Russel Fordham for letting my use your home as if it were mine in order to write this thesis.

Thank you to all of my friends who made my 8 long years at the University of Leeds bearable, I wish I could name you all, but I fear I'll be here forever! Becky Foster you're the best friend a girl could wish for, whom somehow put up with my OCD cleaning habits! I will miss binge watching Netflix with you whilst eating chocolate till we are sick. You and got me through some of the toughest times in Leeds and I'd be a lost soul without you. Thank you to my Leeds parents Eleanor Todd and Gavin Murry, for adopting me and taking me under their wings. Elizabeth Katie Banes, I thank you for keeping me sane during my final year in Leeds. Lastly, thank you to my friends from home; Sophie Harrison, Sorcha Byrne, Aaron (double A-ron) Hill, Claudio Taibi and Shayan (Shaniqua) Ansari for always making the upmost effort to visit me and cheer me up when science got me down.

Θα ήθελα να αφιερώσω αυτή την διδακτορική εργασία στον παππού μου Μιχάλη Παγκρατίου; Γνωρίζω ότι πάντα με προστατευείς και είσαι δίπλα μου, ελπίζω να σε έκανα περίφανο.

Abstract

Simbu orthobunyavirus (SIMV) is a member of the *Peribunyaviridae* family and is a segmented negative-sense RNA virus. Transmitted by mosquitos this family of viruses can cause outbreaks across the globe infecting humans and animals causing fever, birth defects and great economic loss. Despite on-going outbreaks there is no antiviral or FDA-approved vaccine for use in humans. The nucleoprotein (NP) functions to encapsidate the viral RNA genome for protection and is a suitable target for structure-based drug design. This thesis presents five high resolution crystal structures of NP in complex with RNA from two viruses of the *Peribunyaviridae* family; SIMV and Akabane orthobunyavirus (AKAV). We critically analysed the interactions formed between NP and RNA and further assessed their RNA binding contributions in fluorescence anisotropy (FA) assays with comparison to closely related Oropuche Orthobunyavirus (OROV). The structures presented here lay the foundations for structure-based drug design of small molecules that interfere with critical residues for ribonucleoprotein functioning.

Ebolavirus (EBOV) is a non-segmented, negative sense RNA virus within the *Filoviridae* family of the *Mononegavirales* order. EBOV is a highly contagious and virulent pathogen, transmitted by mosquitos and bodily fluids. The largest EBOV outbreak in 2013-2016 (40% mortality) highlighted the need for an effective therapeutic. Structure-based drug design has arisen as an investigation method for antiviral drug candidates. A better understanding of structure, multimerization and interactions of viral proteins will aid therapeutic development. Viral protein (VP) -35 functions within the ribonucleoprotein complex, interacting with NP to maintain NP in a monomeric state for newly synthesised RNA encapsidation. VP30 is an activator of viral transcription and is regulated by phosphorylation. The oligomeric state of both proteins is currently debated within the literature. This thesis optimises the expression and purification of both VP35 and VP30 from EBOV and closely related Marburg marburgvirus (MARV) to confirm their oligomeric states and interactions with RNA by FA.

Human orthopneumovirus (hRSV) is a non-segmented, negative sense RNA virus classified within the *Mononegavirales* order and *Pneumovirade* family, transmitted by aerosol droplets. hRSV is the leading cause of lower respiratory tract illness in infants and the immunocompromised, causing over 250,000 death annually. Up to 79% of deaths are reported in children under five, with 99% of the mortality occurring in developing countries. The M2-1 protein of hRSV represents a promising potential anti-viral target. M2-1 is a transcription anti-terminator with an essential role in viral gene expression; binding both viral RNA and the polymerase co-factor phosphoprotein (P). Here, we describe the optimised protocol for M2-1 and its known binding partner P_{90-160} purification, and the less well characterised binding partner M (matrix protein). Structural studies using X-ray crystallography revealed a 1.99 Å structure of M but no ligand. We confirmed the interaction of M2-1 and M. We implemented structural studies by electron microscopy with M2-1:P_{90-160} complex to fully characterise this interaction.

Contents

Figures

Tables

Abbreviations

ACU	Asymmetric cell unit
BUNV	Bunyamwera virus
CC	Coiled-coil domain
CD	Circular dichroism
CTD	C-terminal domain
CME	Clathrin-mediated endocytosis
CP	Convalescent plasm
CTD	C-terminal domain
CV	Column volume
CWB	Convalescent whole blood
DDD	Direct detector device
DED	Direct electron detector
DLS	Diamond Light Source
DNA	Deoxyribonucleic acid
dRI	Differential refractive index
EM	Electron microscopy
EBOV	Ebolavirus
F	Fusion protein
FA	Fluorescence anisotropy
FDA	Food and drug administration
FI	Fluorescent
FP	Forward primer
GE	Gene end
GJ	Gene junction
GP	Glycoprotein
GS	Gene start
hMPV	Human metapneumovirus
h.p.i.	Hours post infection
hRSV	Human orthopneumovirus
IBs	Inclusion bodies
IBAGs	Inclusion body-associated granules
ICTV	International Committee on Taxonomy and Viruses
IGRs	Intergenic regions
IL	Interleukin
LACV	La Crosse encephalitis virus
Le	Leader sequence
LN_2	Liquid nitrogen
LRTI	Lower respiratory tract infection
M	Matrix protein
M2-1	Matrix-like protein 1
M2-2	Matrix-like protein 2
MARV	Marburgvirus
MW	Molecular weights
MWCO	Molecular weight cut off
NP	Nucleoprotein
NMR	Nuclear magnetic resonance spectroscopy
NS	Non-structural protein
NTD	N-terminal domain
nts	Nucleotides
OD	Oligomerisation domain
ORF	Open reading frame
pA/C/G/U	Polynucleotide-A/C/G/U RNA
PA	Polymerase acidic
PCR	Polymerase chain reaction
PDB	Protein data bank

pI	Isoelectric point
PPI	Protein-protein interaction
RdRp	RNA dependent RNA polymerase
RMSD	Root mean square deviation
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
RP	Reverse primer
RT	Reverse transcriptase
SBDD	Structure based drug design
SEC	Size exclusion chromatography
SEC-MALLS	SEC-multiple angle laser light scattering
SDM	Site directed mutagenesis
SIMV	Simbu virus
SH	Small hydrophobic protein
Tm	Melting temperature
Tr	Trailer
Vo	Void volume
VSV	Vesicular stomatitis virus
VP30	Viral protein 30
VP35	Viral protein 35
WT	Wild type
ZBD	Zinc binding domain

Chapter 1 Introductions

1.1 Introduction to *Peribunyaviridae*

Simbu virus (SIMV) is a member of the *Orthobunyavirus* genus of the *Peribunyaviridae* family of the *Bunyavirales* order. The *Bunyavirales* includes 12 families of viruses that are single-stranded negative sense RNA viruses. Typically, these viruses termed 'bunyaviruses', have three negative sense segments. These zoonotic viruses cause a range of disease in humans and animals including fever and encephalitis and have arthropod vectors (Shi et al., 2009).

At present, there are no reported cases of SIMV infection in humans, however SIMV presents as an emerging arbovirus, and comparison can be drawn between the closely related AKAV and OROV (Figure 1.1) from the Simbu serocomplex which have caused outbreaks in ruminants and humans respectively (Table 1.1). This introduction focuses on the Simbu serocomplex to highlight the effects of these viruses, as well as viruses from the *Orthobunyavirus* family to highlight the impact of this family of viruses. Amino acids are highly conserved across the *Peribunyaviridae* family, SIMV share 81% and 72% sequence identity to AKAV and OROV respectively.

```
AKAV  MANQFIFNDVPQRNAATFNPDAGYVAFISKYGQQFNFTVARVFFLNQKKAKMVLHKTPQP  60
OROV  -MSEFIFNDVPQRTTSTFDPEAAYVSFEARYGQVLNAGVVRVFFLNQKKAKDVLRKTSRP  59
SIMV  MANQFIFEDVPQRNLSTFSPEAGYVAFIGRYGQQLNFSVVRVFFLNQKKAKMVLHKTAQP  60
       .:***:*****.  :**.*:*.**:*  .:*** :* *.********** **:** :*
AKAV  SVDLTFAGVKFTVVNNHFPQYTANPVSDTAFTLHRISGYLARWVAEQCKANQIKFAEAAA  120
OROV  MVDLTFGGVQFAMVNNHFPQFQSNPVPDNGLTLHRLSGYLARWAFTQMRS-PIKQAEFRA  118
SIMV  NVDLTFGGVKFTLVNNNFPQYTANPVPDNALTLHRLSGYLARWTAEQVKNNQVKLAEATA  120
        *****.**:*::***:***: :*** *...:****:*******. * : :* ** *
AKAV  TIVMPLAEVKGCTWSDGYAMYLGFAPGAEMFLETFEFYPLVIDMHRVIKDGMDVNFMRKV  180
OROV  TVVVPLAEVKGCTWNDGDAMYLGFAAGAEMFLQTFTFFPLVIEMHRVLKDGMDVNFMKKV  178
SIMV  AIVMPLAEVKGCTWNDGYTMYLGFAPGAEMFLETFEFFPLVIDMHRVLKDGMDVNFMRKA  180
       ::*:**********.** :****** ******:** *:****:****:**********:*.
AKAV  LRQRYGQLTAEEWMTSKLDAVKAAFGSVAQISWAKSGFSPAARAFLAQFGIQI  233
OROV  LRQRYGQKTAEQWMREEIVAVRAAFEAVGTLAWARTGFSPAARDFLRQFGIGI  231
SIMV  LRQRYGLLTAEQWMTQKIVEVKAAFDAVGQIAWAKSGFSPAARAFLQQFGFTG  233
       ****** ***:**  .:: *:*** :*. ::**::******* ** ***:
```

Figure 1.1 Sequence Alignment for AKAV, OROV and SIMV NP

* conserved residue.
: conservation between groups of strongly similar properties.
. conservation between groups of weakly similar properties.
Alignment made using Clustal Omega (Sievers et al., 2011).

Table 1.1 Simbu Serogroup Viruses

Virus	Year Isolated	Geographical Origin	Source
Aino	1964	Japan	Mosquitoes
Akabane*	1974	Japan	Cattle
Buttonwillow	1962	USA	Rabbits
Douglas	1978	Australia	Cattle
Facey's Paddock	1974	Australia	Mosquitoes
Ingwavuma	1959	South Africa	Birds
Jatobal	1985	Brazil	Rodent
Kaikalur	1971	India	Mosquitoes
Mermet	1964	USA	Monkey
Oropuche*	1955	Trinidad	Human
Peaton	1976	Australia	Midges
Sabo	1966	Nigeria	Goat
Sango	1965	Nigeria	Cattle
Sathuperi	1957	India	Mosquitoes
Shamonda	1965	Nigeria	Cattle
Shuni	1966	Nigeria	Cattle
Simbu*	1955	South Africa	Mosquitoes
Schmallenberg	2011	Europe	Ruminants
Tinaroo	1978	Australia	Midges
Yaba-7	1963	Nigeria	Mosquitoes

*viruses worked on in this thesis
Adapted from Saeed *et al*., 2001

Schmallenberg virus (SBV) (*Peribunyaviridae* family, Simbu serocomplex) is closely related to SIMV and is an emerging arbovirus. In 2011 the largest outbreak of SBV occurred in Germany spreading across Europe causing great economic loss. In domestic and wild ruminants, SBV causes stillbirths, abortions or congenital abnormalitie. SBV is transmitted by *Culicoides* midges.

Despite global outbreaks there is currently no antiviral or vaccine against these pathogens. The nuceleoprotein (NP) is a suitable target for structure-based drug design, understanding the assembly, structure and function of the NP will aid this.
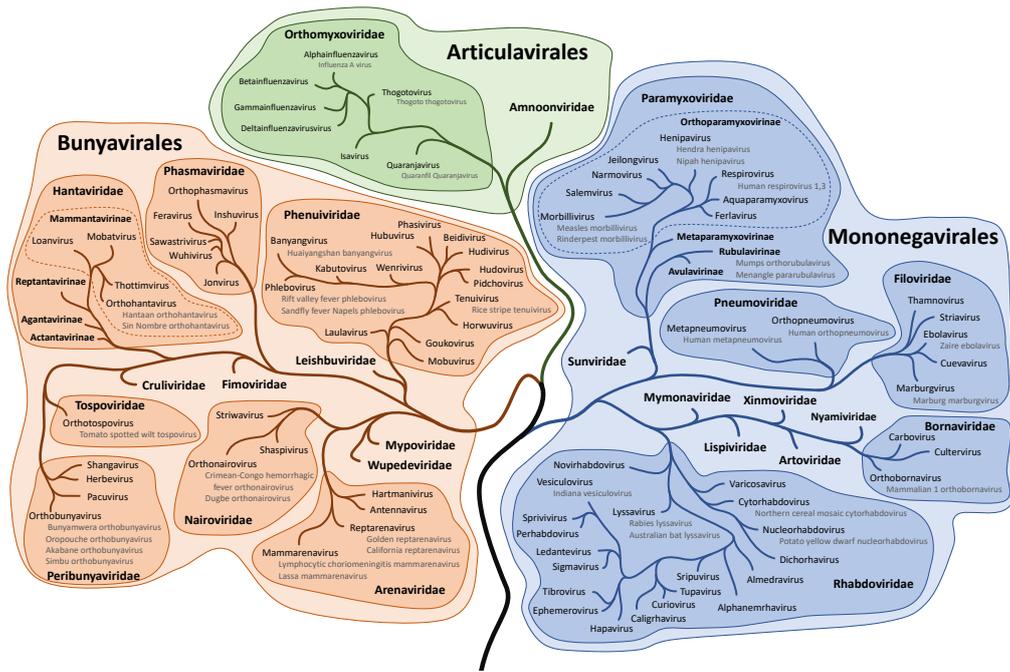
Figure 1.2 Taxonomy of the *Bunyavirales*
Figure made using data from International Committee on Taxonomy and Viruses, Virus Taxonomy: 2020 Release (Walker *et al.*, 2020),

1.1.1 Classification

The *Bunyavirales* order named after the prototype Bunyamwera virus (BUNV). BUNV was isolated from *Aedes* mosquitoes in Uganda's Semliki Forest during a study on yellow fever in 1943 (Smithburn *et al.,* 1946). BUNV was the first Bunyavirus genome to be sequenced and the first segmented negative-sense RNA virus generated from complementary DNA (cDNA) (Lees *et al.,* 1986; Elliott, 1989). In 1975 the *Bunyaviridae* family was established and the International Committee on Taxonomy and Viruses (ICTV) approved the Bunyavirus genus in 1980 (Porterfield, 1973; Bishop *et al.*, 1980). In 1995 the Bunyavirus genus was renamed *Orthobunyavirus* in order to distinguish five distinct genera.

The assignment of species into these genera was based on: lack of serologic cross-reactivity with other genera members, virion size (80-129 nm in diameter), genomic segments, gene expression strategy and conserved nucleotide sequences within the genomic RNA. However, transcriptomic sequencing and viral isolations highlighted unseen bunyaviral genetic diversity; the classic tri-segmented single-stranded RNA genome is no longer shared by all order members and taxonomic assignments should be based on viral nucleotide sequences. Bunyaviruses phylogeny is based on the sequence of conserved motifs within the viral RNA dependent RNA polymerase (RdRp). The *Bunyavirales* RdRp N-terminal HxxxxPDxxxE/DxxxK is a conserved motif which has also been seen in infleuza virus ('flu) (Maes *et al.*, 2019). Information is still unknown on newly classified viruses that have not been isolated including molecular and biochemical characteristics.

The ICTV consequently initiated taxonomic revision to incorporate unclassified Bunyavirus-like species placing all related viruses within a new taxon the *Bunyavirales* order. This created four additional families for the newly-identified species. Now ten families, five cause vertebrate-infections: *Hantaviridae, Nairoviridae, Peribunyaviridae, Phenuiviridae* and *Arenaviridae* families.

The *Peribunyaviridae* family is the largest, including four genera of tri-segmented viruses: *Orthobunyavirus, Herbevirus, Pacuvirus* and *Shangavirus* genera of which 82/89 species belong to the *Orthobunyavirus* genus, which is the only genus to include animal and human pathogens. The *Orthobunyavirus* genus contains over 200 viruses which are transmitted by mosquitoes and have amplification cycles in vertebrate hosts (Calisher, 1996; Shchetinin *et al.*, 2015; Adams *et al.*, 2017). Important human pathogens include Oropouche virus (OROV) which causes debilitating febrile illness, while Akabane virus (AKAV) and Schmallenberg virus (SBV) cause disease in domestic animals. At present there are 20 serogroups within the genus; Anopheles A, Anopheles B, Bakau, Bunyamwera, Bwamba, California, Capim, Gamboa, Group C, Guama, Koongol, Mapputta, Minatitlan, Nyando, Olifanstlei, Patois, Simbu, Tete, Turlock, and Wyeomyia; which have been separated by serologic relatedness of complement fixing antibodies (mediated by the NP) and hemagglutinating and neutralising antibodies (mediated by the glycoprotein (GP)). The Simbu serogroup includes AKAV, OROV, SBV and SIMV mentioned in this thesis, whereas the prototype virus BUNV is from the Bunyamwera serogroup.

1.1.2 Zoonotic Life Cycle

Arboviruses are transmitted by the bite and/or blood feeding of infected arthropods. The *orthobunyaviruses*, *phleboviruses*, *noroviruses* and *topsoviruses* and are transmitted in this way. *Hantaviruses* on the other hand have rodent vectors but infection occurs through aerosolised rat excreta (Brackney *et al.*, 2010).

Due to mosquitoes replicating in stagnant water, epidemics of mosquito-vectored viruses occur after periods of heavy rain. In mosquitoes, virion amplification occurs in the midgut epithelium ensuring hemocoel escape via the hemolymph and salivary glands when blood-feeding (Horne and Vanlandingham, 2014). Meanwhile, in infected animal hosts, virions spread to striated muscles for amplification (Elliott, 2014). High virus titres allow virions to cross the blood-brain barrier targeting neurons. After a ~5-day incubation period acute symptoms arise. The *Orthobunyaviruses* also infect livestock typically causing birth defects and stillbirth, however in humans mild flu-like symptoms persist. OROV can cause Oropouche fever characterised by fever, chills, headaches and myalgia.

Human to human transmission of *Orthobunyavirales*, is rare as humans are considered dead-end hosts as viral titres do not typically reach high enough blood titres. However, this has been documented for the Andes hantavirus (*Hantavirus*) and Crimean-Congo haemorrhagic fever virus (CCHFV *Phleobvirus*) (Padula *et al.*, 1998; Mardani *et al.*, 2009).

1.1.3 Impact, Epidemiology and Symptoms

The first outbreak of AKAV between 1972 and 1975 resulted in over 42,000 birth defects in calves in Japan, causing ¥5 billion in economic loss (~£35 million) (Tang *et al.*, 2019). Since, AKAV infections have been noted to be seasonal in temperate regions; bovine epizootic (postnatal) encephalomyelitis is endemic in Japan and South Korea, outbreaks have been reported annually since 2000 (Rovid Spickler, 2017). The morbidity rate of AKAV remains unclear, in 2000 an outbreak in Korea affected 30% of ruminants. In cattle, a foetus can be affected any time after two months, and severe defects occur between 80-150 days of gestation. Whereas sheep and goats are highly susceptible between 28-36 days of gestation (Rovid Spickler, 2017). Infections are typically asymptomatic resulting in stillbirths, premature births and congenital malformations of the brain.

The SBV emerged in 2011, at the border between Germany, the Netherlands and Belgium (Hoffmann *et al.*, 2012), the rate of spread was estimated between 0.9-1.5 km per day (Balmer *et al.*, 2014). SBV was initially identified due to farmers reporting, fever, diarrhoea and a reduction in milk yields (up to 50%) (Hoffmann *et al.*, 2012), however then reported congenital malformation in cattle and ewe offspring. Congenital malformation was reported in cattle herds was 4% and 8% in sheep flocks, with farms reporting 40-50% reduction in expected prolificacy (Harris *et al.*, 2014).

Central and South America have seen over 30 epidemics of OROV with over 0.5 million people infected in the last 60 years. The OROV incubation period is typically 3-8 days with symptoms including fever (known as Oropouche fever), headache and myalgia (Sakkas *et al.*, 2018). Oropouche fever is the second most frequent arbovirus febrile disease in Brazil (after Dengue virus) (Sakkas *et al.*, 2018), however since the first outbreak in 1960, no deaths have been reported.

1.1.4 Treatment

At present, there is no FDA approved vaccine or treatment for Bunyavirus infection.

Two inactivated vaccines are available for SBV cattle infection: Bovilis Ringvac (Merck Sharp and Dohme) and SBVvox (Merial Animal Health) both came to market with a provisional licence. AKAV vaccines induce neutralising antibodies and have been commercially available since 1990. Treatment for OROV is symptomatic, typically ribavirin is used for similar viral infections as it is a broad-spectrum anti-viral; however ribavirin exerts no effect on OROV infections (Sakkas *et al.*, 2018).

1.1.5 Virion Morphology

Virions appear spherical to pleomorphic and 80-120 nm in diameter. The surface GP protrude 5-10 nm and are embedded in the host-derived Golgi lipid envelope that is 5-7 nm thick. More specifically, BUNV virions are ~108 nm in diameter and pleomorphic (Bowden *et al.*, 2013). Mature virions are not always packaged with equal numbers of

RNP segments and unequal complements of RNPs may contribute to virion size (Talmon *et al.*, 1987; Battisti *et al.*, 2011) (Figure 1.3 A and B).

The tripod-like GP spikes are arranged with 3-fold symmetry in patches with gaps in between (Figure 1.3 A). On the other hand, Hantavirus (HTNV) and Tula hantavirus (TULV) display spike projections with 4-fold symmetry proposed to be maintained by homo-oligomerisation of Gn by interactions between Gn tetramers and Gc dimers (Huiskonen *et al.*, 2010; Battisti *et al.*, 2011; Hepojoki *et al.*, 2012; Bowden *et al.*, 2013). The GP spikes are essential for cell attachment and membrane fusion. Life cycle stage-specific activation of GP is pH- and ion-dependent. These pH- and ion-dependent conformational changes have been visualised for *Orthobunyaviruses*. pH and potassium ion (K^+) concentration changes expose hydrophobic regions on Gc that facilitate fusion between viral endosomal membranes (Punch *et al.*, 2018).

Interior features of the virion have also been seen by cryo-electron tomography, revealing electron density for the filamentous RNA segments as ribonucleoprotein (RNP) complexes. These RNPs are located close to viral envelopes with possible interactions with the GP cytoplasmic tail (Bowden *et al.*, 2013).
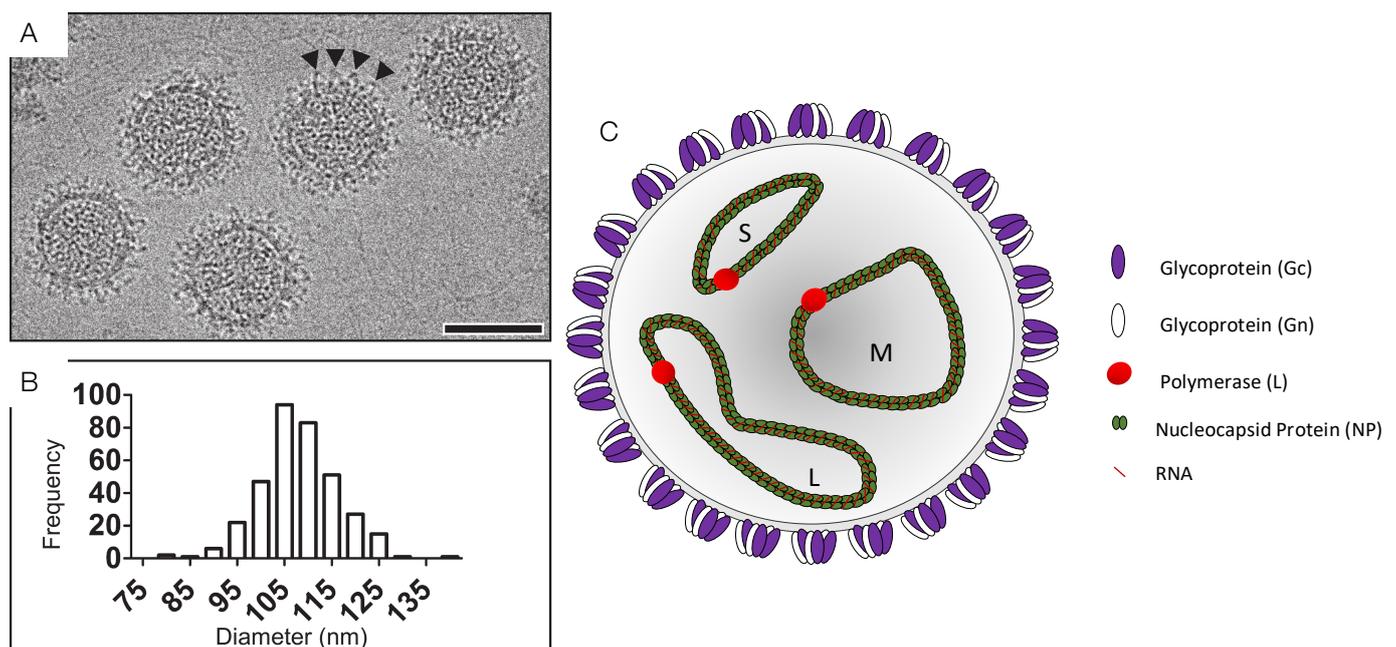
Figure 1.3 *Bunyavirales* Virions

A- Micrograph of BUNV virions. GP spikes highlighted by triangles protrude 5-10 nm.
B- BUNV virion diameters range from 80-120 nm.
C- Tri-segmented RNA is encapsidated by NP located close to the viral envelope which is derived from the host cell membrane. The virion is coated in heterodimers of glycoproteins Gc and Gn.
A-B Adapted from Punch *et al.,* 2018.

1.1.6 Genome

The *Peribunyaviridae* family, negative sense RNA virus genome is split into 3 segments, small (S), medium (M) and large (L) which are encapsidated by the nucleoprotein (NP) and known as an RNP complex (Figure 1.3 C). The segments possess a short stretch of 8-11 nucleotides at the 3' and 5' ends which are highly conserved within families (Table 1.2). It is through inter-terminal base-pairing of these conserved nt that form the pseudocircular structures and drive circularisation of the RNA (Pettersson and von Bonsdorff, 1975). Bunyaviral RNPs are flexible, being visualised as either 'beads on a string' or an ordered helix (Raymond *et al.*, 2010; Ferron *et al.*, 2011; Ariza *et al.*, 2013; Reguera *et al.*, 2013; Wang *et al.*, 2016).

These segments are liable to re-assort during co-infection giving rise to emerging pathogens. For example, re-assortment between the prototypic virus BUNV and Ngari virus caused an outbreak in eastern Africa between 1997-1998 (Briese et al., 2006). The tri-segments are discussed in short below.

Table 1.2: 3' and 5' Nucleotide Sequences of Genomic RNA from families within the *Bunyavirales* order

Peribunyaviridae (*Orthobunyavirus* genus)	3' UCAUCACAUGA..................UCGUGUGAUGA 5'
Hantaviridae (*Orthohantavirus* genus)	3' AUCAUCAUCUG.....................AUGAUGAU 5'
Nairoviridae (*Orthonairovirus* genus)	3' AGAGUUUCU.....................AGAAACUCU 5'
Phenuiviridae (*Phlebovirus* genus)	3' UGUGUUUC........................GAAACACA 5'
Tospoviridae	3' UCUCGUUAG.....................CUAACGAGA 5'

1.1.6.1 L Segment

The L segment gene codes for the RdRp, or L, essential for transcription and replication of the genome. Approximately 6,500 nt in length the L segments of orthobunyaviruses, hantaviruses, phleboviruses and banyangviruses are similar in size.

1.1.6.2 M Segment

M segment length can vary between 3600-5300 nucleotides and encodes a single polyprotein precursor that is translated on membrane-bound ribosomes, from a single ORF that is cleaved by cellular peptidases (Lappin *et al.*, 2019). Encoding Gn and Gc (referring to amino- or carboxy-terminal precursor position) glycoproteins involved in viral entry and NSm a non-structural protein, the function of which still remains elusive. NSm is also cleaved during Gn and Gc processing.

1.1.6.3 S Segment

The S segment of the *Peribunyaviridae* family is the shortest of the order and code for the NP protein and in some cases from a second ORF in the ambisense which encodes a small non-structural protein NSs (non-structural s) (Van Knippenberg and Elliott, 2015). S-segment products are translated on free ribosomes. NSs from bunyaviruses have low sequence homology but have similar roles regarding the host innate immune system and apoptosis.

1.1.7 Proteins

1.1.7.1 Nucleoprotein

The 1 kB S segment encodes NP protein (25 kDa) which is the most abundant viral protein in virions and infected cells. The main function of NP is in RNA encapsidation; however, NP also interacts with Gn and Gc to mediate virion assembly (Spiegel, Plegge and Pöhlmann, 2016). Interactions with L potentially drive RNP formations and mediate encapsidation (Gerlach *et al.*, 2015).

NP is required for both transcription and replication. BUNV NP mutants were either defective in antigenome synthesis but not mRNA transcription (K228T and G230R) or defective in replication but transcriptionally competent (M105T, N74S and S96G) in mini-genome and viral infectivity assays. Different domains of NP are therefore associated with RNA replication and transcription; it is unclear if these mutations cause conformational changes within NP that modulate the function of NP, or if transcription and replication require different binding partners that interact on differing locations on NP (Eifan and Elliott, 2009; Walter *et al.*, 2011; Ariza *et al.*, 2013).

The BUNV and SBV NP crystal structures were both solved by Ariza et al., in 2013. NP protein expressed in *E. coli* crystallised as an almost entirely α-helical globular tetramer, with oligomerisation occurring via an N-terminal extension (Figure 1.4). The N-terminal arm drives NP multimerization and RNP formation via interactions with adjacent NP monomers (Ariza *et al.*, 2013).

Although BUNV and SBV are from different serogroups their NP crystal structures are similar (RMSD 1.3 Å). The biggest difference between the two structures is in the position of the N-terminal arm and C- terminal helix. Upon superposition, at the N-terminus residue 3 are 8.74 Å apart; residue 233 of the C-terminal α-helices is transposed by 6.35 Å during superposition, whereas residue 211 is 0.69 Å apart, suggesting the core is similar but the N-terminal arm and C-terminal helix are in different positions within the crystal. In addition, SBV tetramers possess 2-fold symmetry in the crystal lattice forming a 'squashed' rhombus; whereas BUNV NP + RNA possesses a 4-fold symmetry as in a square; suggesting that RNA induces a conformational change from a 2-fold to 4-fold symmetrical tetramer. Neither crystal structure has density for the full N-terminal flexible arm (residues 11-15) (Figure 1.4).
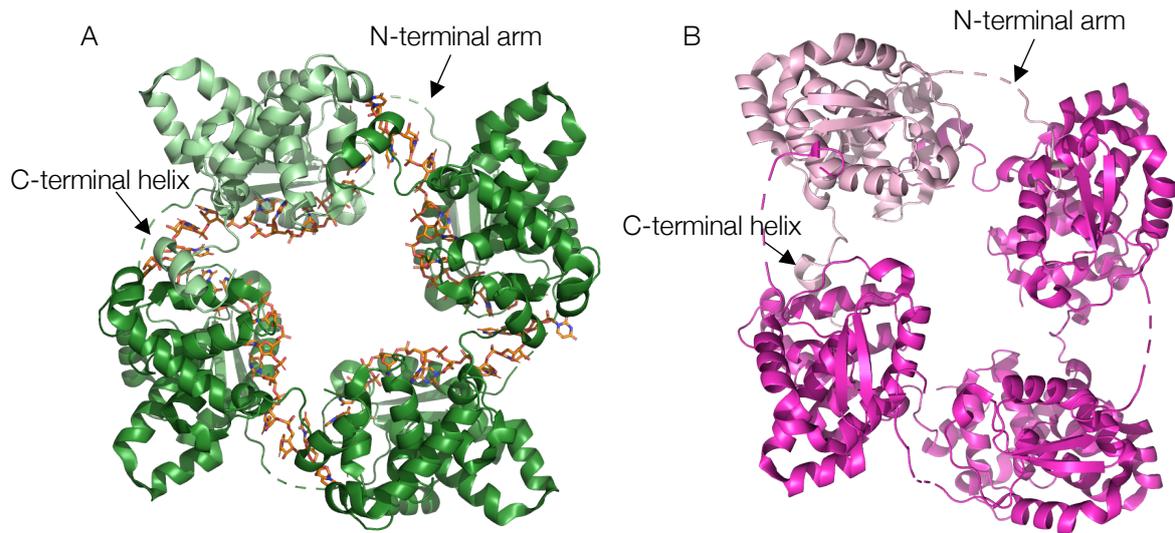
Figure 1.4 Crystal Structure of BUNV and SBV NP Tetramer

A- Crystal structure of BUNV tetramer (green) in complex with *E.coli* host RNA, monomer shown in light green.
B- crystal structure of SBV tetramer (pink), monomer shown in light pink.
Figures made in PyMol (version 2.3.2).
PDB: 3ZLA (BUNV), 3ZL9 (SBV).
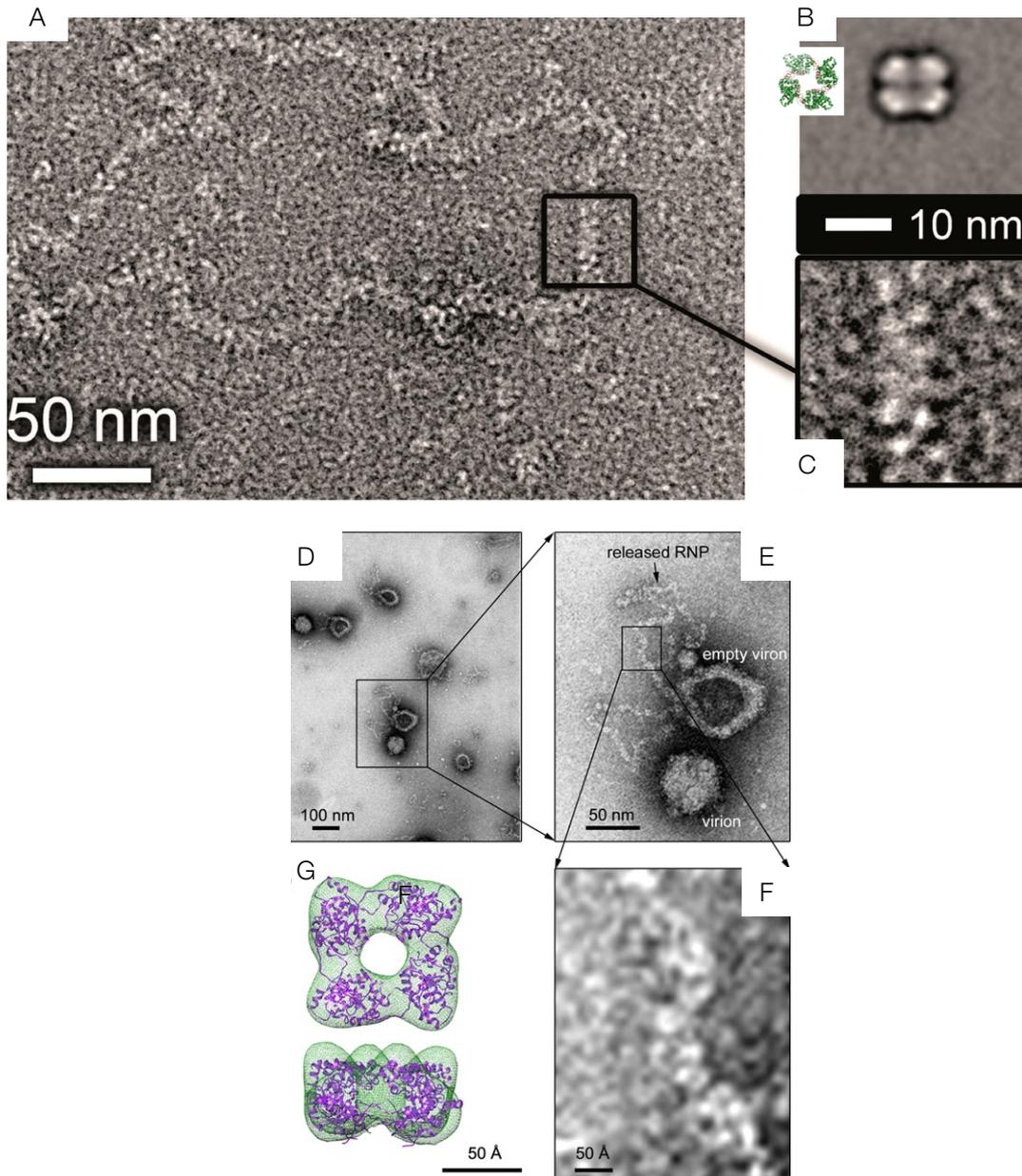
Figure 1.5 Negative Stain Electron-Microscopy of Live Virus RNPs and Purified Tetramers

A-Micrograph of BUNV RNP purified from live virus.
B- single particle averaging for recombinant SBV tetramers + RNA (synthetic 48 mer), with BUNV + host RNA in insert.
C- zoom of A.
D – micrograph f BUNV RNP extracted from live virus.
E-F- zoom of D.
G- 3D reconstruction of BUNV tetramer with crystal docking.
A-C adapted from Ariza *et al*., 2013.
D-F adapted from Li e*t al*., 2003.

The published micrographs of BUNV RNPs have led to conflicting hypothesis on the overall architecture. Purified live virus BUNV RNPs, are helical with a constant width of 10 nm in EM suggesting the RNP is a helical tetramer, as this width corresponds to two monomers or one side of an NP tetramer (Figure 1.5 A-C) (Ariza et al., 2013). Ariza et al., hypothesised that the BUNV RNP is a helical repeat of a four monomer unit, with an interior positively charged RNA binding groove. However, Li et al., suggests that RNPs from live virus are actually 5 nm taking on a filamentous appearance in EM (Li et al., 2013) corresponding to an NP monomer, with a more filamentous RNP architecture; suggesting the NP:RNA monomer to be the building block of an RNP giving a 'beads on a string' appearance (Figure 1.5 D-G).

There is, debate in the contemporary literature about which architecture of the RNPs reflects that in virions (helical tetramer or filamentous 'beads on a string'). The 5 nm wide "beads on a string" RNPs provides a model for RNPs that is not helical; whereas 10 nm wide RNPs may be helical formed with a base structure of a tetramer in a split ring conformation (Figure 1.5 D-G and A-C respectively). It is possible that both do indeed reflect a 'real' scenario, with helical RNPs being the standard not replicating form, but this has to expand or 'loosen' in order to be replicated/transcribed (the RdRp must be able to access the RNA as it processes along the RNA). The flat tetramers seen in crystal structures may represent what happens in the absence of the polymerase; when L protein is present, it may actively add monomeric NP to a growing helical RNP. More EM experiments are ongoing at the University of Leeds within the Edwards/Barr laboratories to attempt to resolve this dichotomy.

1.1.7.1.1 RNA Encapsidation

Crystal structures of apo SBV NP and BUNV NP in complex with host RNA revealed a unique protein fold with a positively charged channel in which RNA resides in encapsidated RNPs (Figure 1.6) (Ariza *et al.*, 2013). This RNA binding groove bisects a bi-lobed globular core. Each NP monomer contacts 9-10 nt through ionic and polar interactions to the RNA phosphate backbone. RNA is buried within the cleft on the inner face of the tetramer, protecting the RNA, as bases are inaccessible from the outside. The RNA binding groove is lined with residues that are able to from hydrogen bond interactions to bases and positively charged residues including K55, H93, R94, K127 and K179 which provide affinity by binding the negative RNA backbone. No base stacking interactions were seen to accommodate RNA binding apart from Y176 which forms two deep cavities, however, these are not explored by the RNA (Ariza *et al.*, 2013). The mechanism of RNA encapsidation during genome replication is poorly understood. The current model is that RNA transiently dissociates from NP during polymerisation in order to expose RNA for transcription and/or replication, the nascent RNA is encapsidated during exit from the RdRp exit channel by readily available free NP monomers (Reguera *et al.*, 2013).
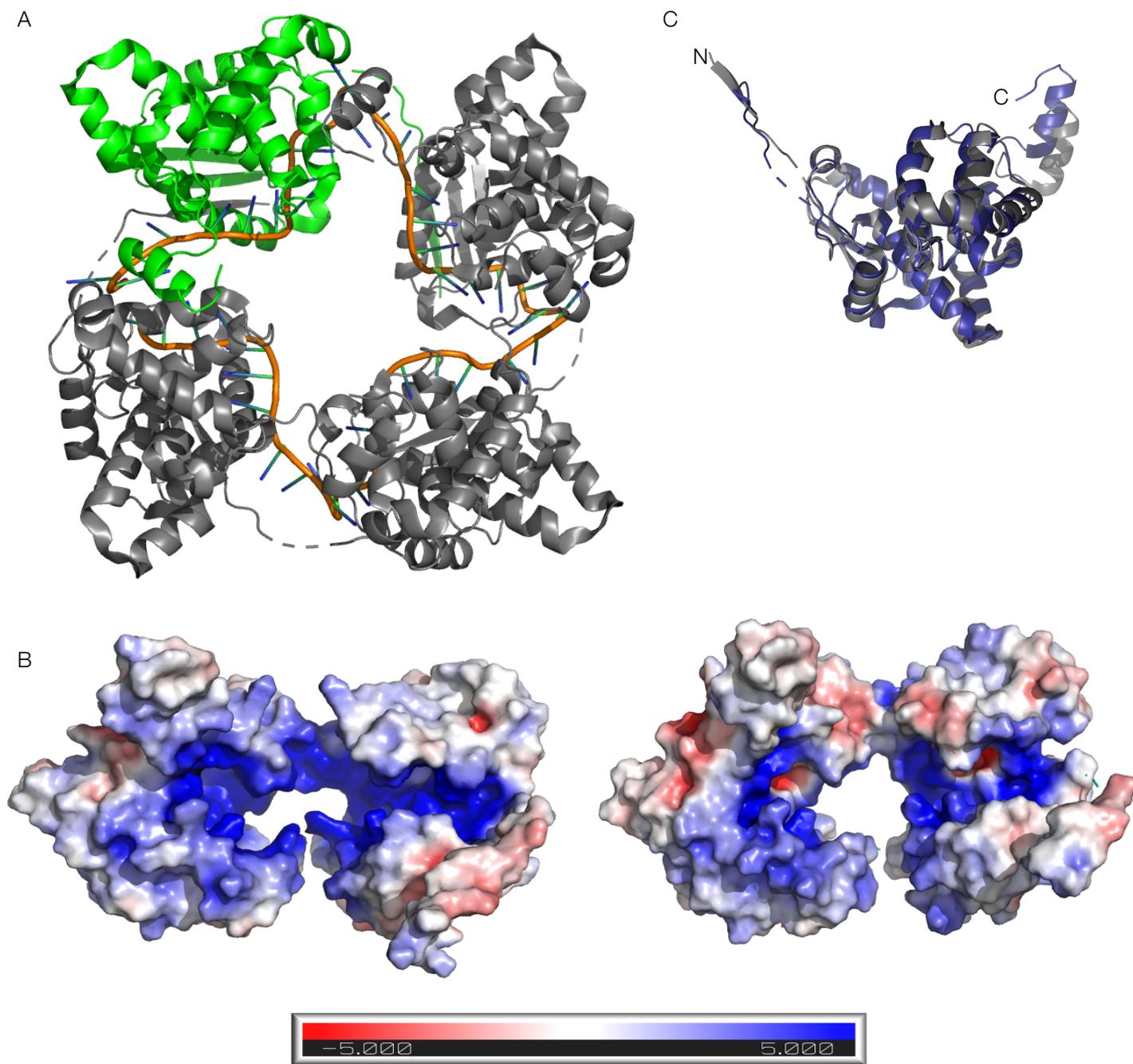
Figure 1.6 Crystal Structure of BUNV and SBV NP

A- BUNV NP tetramer (grey, monomer shown in green) bound to RNA (orange).
B- Electrostatic surface potential of BUNV (left) and SBV (right) NP dimer reveals an electropositive RNA-binding groove.
C- Superposition of BUNV (grey) and SBV (deep blue) NP monomer. RMSD 0.896 Å.
Figures made in PyMol (version 2.3.2) with APBS Electrostatic plugin.
PDB: 3ZLA (BUNV), 3ZL9 (SBV).

1.1.7.2 NSs

1.1.7.2.1 Function

Some viruses within the *Orthobunyavirus* genus encode an NSs protein, including SIMV, BUNV and SBV. SBV NSs is encoded by an ORF between nucleotides 48-323 at the antigenomic RNA position +1. Typically, around 10 kDa, NSs is a non-structural protein and a major virulence factor. SBV and related LACV NSs promotes proteasomal degradation of the RNA polymerase II Rpb1 complex *in vitro* inhibiting cellular transcription and protein synthesis which may in turn cause blockade of the interferon (IFN) innate immune response. Rpb1 expression was decreased 8-12 hours post infection (h.p.i.) in SBV infected cells (Gouzil *et al.*, 2017). Inhibition of cellular transcription reduces IFN production, favouring SBV replication and pathogenesis. On the other-hand, SBV and LACV NSs enhance apoptotic cell death shown through increased caspase-3 and -7 expression, perhaps to enhance virion release, however this was not seen for BUNV. BUNV NSs reduces/delays cell death through countering the activity of IFN regulatory factor 3 mediated induction of early cell death (Kohl *et al.*, 2003; Barry *et al.*, 2014).

NSs localises within the nucleus of virally infected cells, targeting the nucleolus through the nucleolar localisation signal of residues 33-51 which has two basic rich stretches, crucial for nucleolar targeting through mediating interactions with the nuclear pore complex (Cautain *et al.*, 2015). This central domain is highly disordered suggesting that NSs is intrinsically disordered. Fluorescence microscopy of eGFP-NSs33-51 showed co-localisation with nucleophosmin (B23), which in-turn is redistributed to the nucleoplasm causing virus-induced nuclear disruption (Gouzil *et al.*, 2017). This phenotype was seen as early as 4 h.p.i suggesting nuclear re-organisation occurs early on in SBV infection, western blot analysis did not show a decrease in B23 expression.

NSs plays a key role in viral pathogenesis; in SBVΔNSs infected mice, there is an attenuated phenotype characterised by a delay in death compared to WT infection. (Varela *et al.*, 2013).

1.1.7.3 M Segment Products: Gn, Gc and NSm

The M-segment is 4.5 kb and encodes Gn, Gc and NSm. Co-translational cleavage of the glycoprotein precursor (GPC) generates Gn (32-35 kDa) and Gc (110-120 kDa) glycoproteins that heterodimerise to form 'glycoprotein spikes' on the virion cell surface. In some *Orthobunyaviruses* including BUNV GPC cleavage also generates NSm (16-18 kDa) with the single ORF arranged Gn, NSm and Gc. Gn and Gc are preceded by signal peptide sequence, with cleavage mediated by host signal peptidase in the ER (Fazakerley and Ross, 1989).

Gn and Gc have asparagine-linked oligosaccharides important for correct folding and function; Gn has a complex oligosaccharides whereas Gc has mostly high-mannose glycans. Glycosylation occurs at one site on Gn and at two sites on Gc (Madoff and Lenard, 1982; Novoa *et al.*, 2005; Shi, Brauburger and Elliott, 2005).

These are class I transmembrane proteins with their N-terminus exposed on the virion surface and anchored by their C-terminus. Gn has Golgi targeting and retention signals, functioning as a Gc chaperone which contains a class II fusion domain and is therefore thought to be the key player in membrane fusion (Shi *et al.*, 2009).

NSm is hydrophobic but the function remains unclear: NSm is thought to be a membrane-bound protein that facilitates virion assembly due to its localisation to the Golgi with Gn and Gc (Shi *et al.*, 2006; Fontana *et al.*, 2008).

1.1.7.3.1 Structure

Glycoprotein spikes are formed from trimers of Gn and Gc heterodimers that extend from the virion membrane. Each protomer extends 18 nm from the virion surface; spanning a 'stalk' region between protein contacts. Contacts located close to the membrane, named the 'floor' region extend 5 nm from the viral membrane forming a flat triangle base; it is thought that Gn is localised here anchoring the spike to the membrane via a C-terminal transmembrane region. The second contact is located distally (the 'head' region), where Gc resides, which binds to Gn in the floor region (Bowden *et al.*, 2013) (Figure 1.7 A). Altering the pH to mimic endosome acidification induced an altered GP structure. GP spikes order was lost and Gc hydrophobic regionsexposed (Figure 1.8).
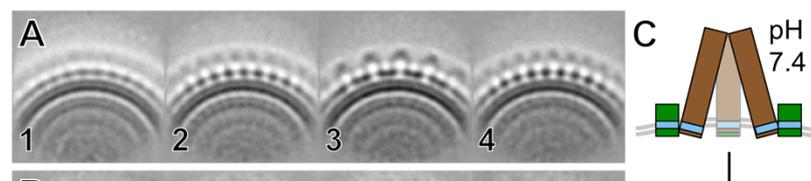
Figure 1.7 pH-Dependent Conformational GP Changes

A-B- class averages of GP spikes are pH 7.4 and 5.1 respectively.
C- model of pH effect on GP spikes. Acidification causes loss of order of GP spikes.
pH-decrease exposes hydrophobic regions on Gc (brown) that facilitate fusion.
Gn coloured green, transmembrane region blue and membrane shown in grey.
Adapted from Punch *et al*., 2018.
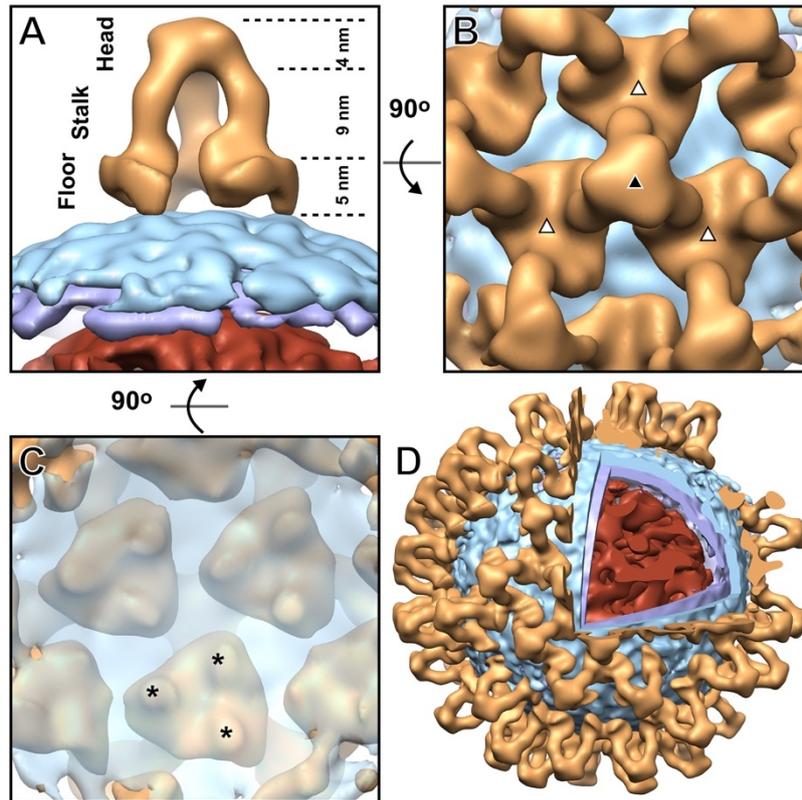
Figure 1.8 BUNV GP Spike is a Trimeric Assembly of Gn-Gc

A- side view of GP spike has a novel tripodal architecture. Spike protomers are elongated extending 18 nm from the virion surface.
B- top view of GP Spike. Triangles highlight membrane:GP contacts.
C- bottom view (below the membrane) of GP spikes. Stars highlight transmembrane contacts.
D- open model of a BUNV virion. GP shown in orange, membrane outer and inner leaflet are cyan and purple respectively and RNP in red.
Adapted from Bowden *et al.*, 2013..

1.1.7.4 L

1.1.7.4.1 Function

L proteins from *Peribunyaviridae* genus are ~240 kDa and mediate mRNA and viral RNA synthesis from copy RNA (cRNA) templates. L therefore possesses numerous enzymatic activities through the following domains: endonuclease, transcriptase, replicase and cap-binding domain.

The RdRp lacks a proof reading/repair mechanism, making it highly error prone. This is responsible for the rapid evolution of *Bunyavirales* (and all RNA viruses) (Duffy, Shackelton and Holmes, 2008), which may have modified pathogenicity and can lead to rapid drug resistance.

1.1.7.4.2 Structure

The LACV (La Crosse encephalitis virus) L protein (residues 1-1750 (77% of the polypeptide)) crystal structure has been solved, adopting a canonical RdRp fold; a right handed palm with fingers and thumb domains containing catalytic core motifs (Figure 1.9) (Gerlach *et al.*, 2015). The N-terminus of LACV L (residues 1-184) contains an endonuclease domain, producing 5' capping (and cap snatching) of mRNA. The N-terminus of LACV L is also a structural homologue to Flu's endonuclease within polymerase acidic (PA) subunit. This endonuclease domain is linked to a PA-like region by LACV L residues 185-270, which form an extended flexible linker. Residues 271-759 form two lobes that interact in a sequence-specific manner to two viral RNA termini at separate binding sites and the thumb and palm of the RdRp core. The central core 760-1432 builds an internal active site chamber, accessible by four solvent tunnels; two of which are for RNA entry and exit. Adjacent to each other, this tunnel forces the template through an internal path via the active site and is 20 nt in length. A third tunnel is for nascent RNA exit and the fourth for NTP entry. Channel proximity (template entry and exit channel) allows RNA to be extracted and replaced within the NP:RNA binding grove of the RNP without breaking NP:NP interactions, allowing for minimal disruption (Gerlach *et al.*, 2015; Swale *et al.*, 2016).
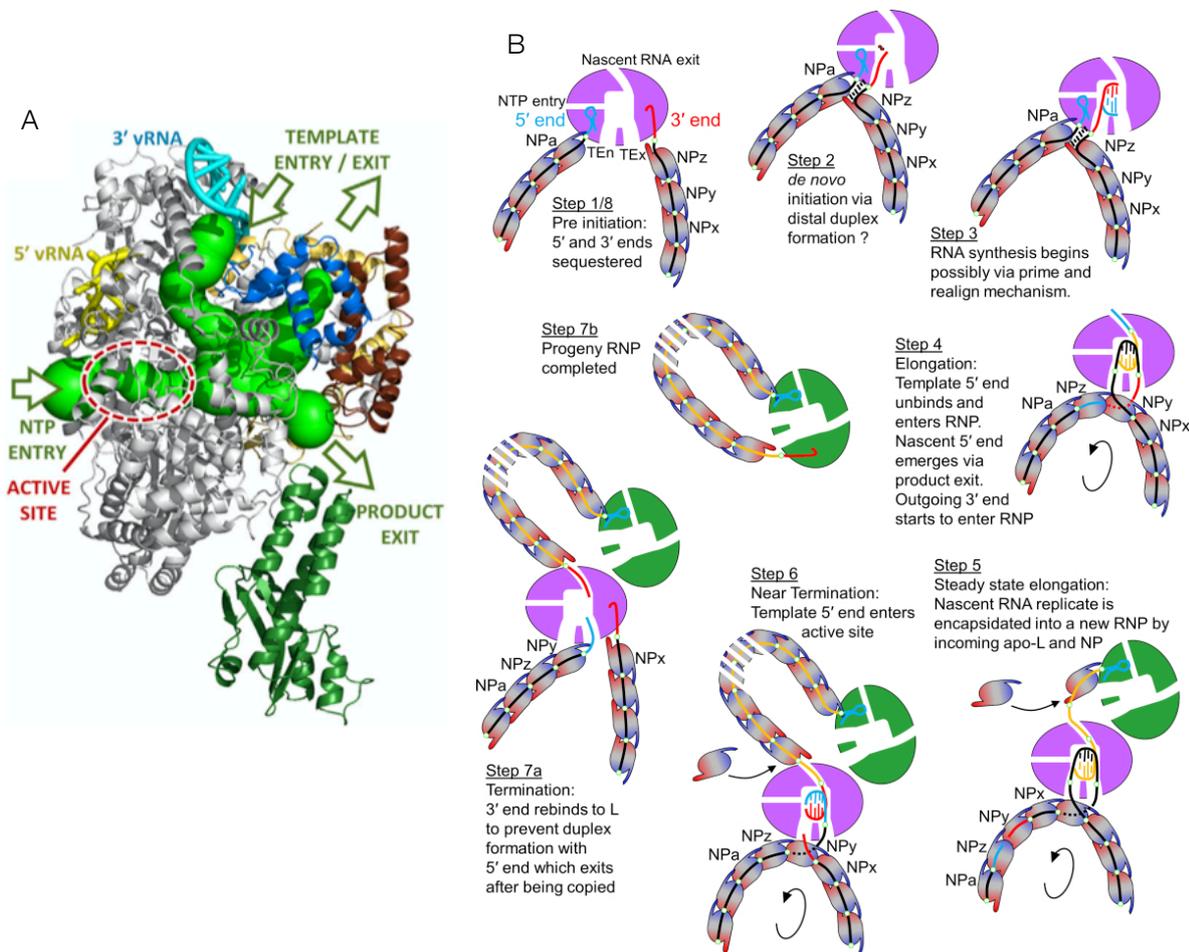
Figure 1.9 LACV Polymerase Crystal Structure and Viral RNA Replication Model

A- LACV polymerase (grey) with 5' and 3' viral RNA (vRNA) in yellow and cyan respectively. Four solvent tunnels (green) allow NTP, template entry and exit and product exit (arrows). A short template-product duplex is formed in close to the active site in the interior cavity. Template strand exits at the front of the polymerase and is re-integrated into the RNP. The nascent (product) strand exits to the rear of the polymerase where product processing occurs. Endonuclease (forest green), bridge (blue), thumb-ring (gold) and lid (brown) are also shown.

B- A LACV RNP is schematically represented with the polymerase (purple or green, can sequester 20-22 nucleotides), with template entrance (TEn), template exit (TEx), NTP entry and nascent RNA exit channels as marked, interacting with the viral RNA (black or yellow) and proximal NPs (ellipses coloured with a blue-to-red gradient). The complementary 5′ and 3′ viral RNA (vRNA) ends are, respectively, cyan and red.

1- In the inactive state, whether after vRNP assembly or in virions, both ends of the genomic RNA are sequestered into the specific 5′ and 3′ RNA binding sites of the polymerase, thus circularizing the RNP. 2- For RNA synthesis the 3′ end is relocated into the polymerase active site for initiation, by an unknown mechanism. 3- With the 5′ end bound to the allosteric site for the activation of the RNA synthesis, a nascent cRNA begins to be synthesized. 4- As elongation proceeds, the template dissociates from the proximal NP and is channelled into the active site. Because of the proximity of the entrance and exit channels the disruption of the RNA-NP assembly may only affect one NP. Early on, the 5′ end is detached from its specific binding site on the polymerase and enters the RNP by loading onto NPz. As incoming template is released from NPy on one side, the outgoing 3′ end is loaded on it from the other side. More generally, the RNA being pulled into the cavity by the polymerase motor detaches from the proximal NP which is pulled to the left thus pushing the NP-RNA array in the direction of the arrow. This model would imply that 5′ end binding is only required to activate initiation. This would be a difference from the influenza situation where the maintenance of 5′ end binding is required, at least during transcription, for self-polyadenylation to occur. 5- Once the nascent c5′ end emerges from the exit channel it can recruit an incoming apo-polymerase as the first step in encapsidating the progeny cRNP with incoming apo-NPs. 6- Approaching termination, the template 5′ end would be copied and the template 3′ end (bound to NPy) would approach its starting point. 7a- At termination the template 3′ end rebinds to its specific binding site on the polymerase to avoid base pairing with the emerging template 5′ end which subsequently rebinds to its polymerase binding site, thus completing the replication cycle. 7b-) Due to polymerase dimer formation, the nascent c3′ end, which emerges last from the product exit channel, can easily find and bind to specific 3′ binding site on the green polymerase, thus completing progeny cRNP formation. Without polymerase dimer formation being maintained throughout replication (or other mechanism for keeping the polymerases in close proximity), it is unclear how the c3′ could find and bind to the correct polymerase which may have diffused far away.
Adapted from Gerlach *et al*., 2015.

1.1.8 Life Cycle

1.1.8.1 Entry

Bunyaviruses have a typical nsRNA virus life cycle with host-cell entry being mediated by Gn and Gc glycoproteins (Figure 1.10). Bunyaviruses are able to bind to a range of host-cell receptor including DC-SIGN (dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin, a C-type lectin receptor) inducing clathrin-mediated endocytosis and endosome acidification (Lozach *et al.*, 2011). Bunyavirus entry is thought to be modulated by cholesterol; by facilitating viral fusion with endosomal membranes (Umashankar *et al.*, 2008). Upon endosome acidification the glycoproteins undergo conformational changes causing the fusion of viral-host membranes facilitating the release of viral RNPs; blocking endosomal acidification by ammonium chloride treatment blocked entry (Plassmeyer *et al.*, 2005, 2007; Santos *et al.*, 2008; Shi *et al.*, 2009). Bunyavirus infection at low pH has also been shown to mediate syncytia formation (Gonzalez-Scarano, Pobjecky and Nathanson, 1984; Ogino *et al.*, 2004; Whitfield, Ullman and German, 2005; Filone *et al.*, 2006; Shi *et al.*, 2007).

Virions are trafficked through early endosomes to late endosomes where fusion is triggered (Lozach, Huotari and Helenius, 2011). In order for viral RNPs to be released into the cytoplasm from endosomes an influx of potassium ions is required. This is dependent on cellular potassium ion channels within endosomal membranes (Punch *et al.*, 2018). Blocking potassium channels prevents viral RNP release into the cytoplasm and virions are further trafficked along the endosomal pathway and/or into lysosomes where they are inactivated by low pH (Hover *et al.*, 2016).

The literature suggests that Gc is involved in membrane fusion; anti-body binding studies, detergent partitioning experiments and protease sensitivity assays suggest that LACV). Gc undergoes conformational changes during pH changes (Gonzalez-Scarano *et al.*, 1985). Recombinantly expressed LACV Gc from Vaccinia virus was not able to cause cell fusion suggesting Gn is also required for activation. BUNV Gn cytoplasmic tail mutations affected viral fusion highlighting that both are required for host cell infection (Jacoby *et al.*, 1993; Shi *et al.*, 2007).
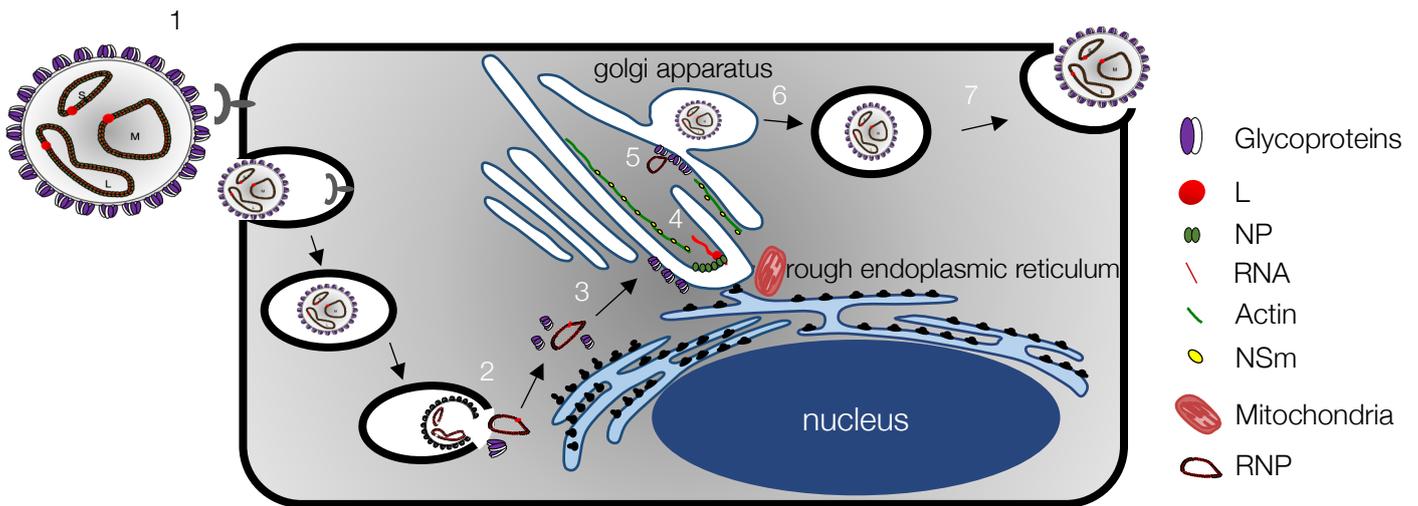
Figure 1.10 *Bunyavirales* Life Cycle Schematic

1- Entry is mediated by Gn and Gc glycoproteins which bind to host cell receptors. Clathrin-mediated endocytosis is induced, and endosomes acidify.
2- endosomal acidification causes conformation changes in Gn and Gc allowing membrane fusion and the release of RNPs.
3- viral RNPs are trafficked to the Golgi apparatus.
4- Viral transcription and replication occurs mediated by L.
5- encapsidated genomic RNA is transported through the Golgi apparatus where virions form from the Golgi membrane.
6- Golgi vesicles migrate via actin.
7- progeny virus buds from the host cell membrane.

1.1.8.2 Transcription and Replication

During transcription the bunyavirus RNP-associated RdRp produces a positive sense mRNA from the negative sense genome. This positive mRNA is transported to ribosomes on the rough endoplasmic reticulum (RER). Here, 'cap-snatching' takes place. This involves bunyavirus mRNA 'stealing' the 5' cap of host mRNAs for transcription initiation. Cap-snatching has also been outlined in 'Flu shown to require cap binding and endonuclease activities found in the L subunits of bunyaviruses. In bunyaviruses 5' capping occurs via an endonuclease domain (found within the N-terminal domain (NTD)), that belongs to the PD-D/ExK superfamily of cation dependent nucleases (Patterson, Holloway and Kolakofsky, 1984; Reguera, Weber and Cusack, 2010; Klemm *et al.*, 2013; Devignot *et al.*, 2015; Fernández-García *et al.*, 2016; Holm *et al.*, 2018). These endonuclease domains across segmented RNA viruses are structurally similar. For the *Peribunyaviridae* these contain a canonical catalytic histidine (Holm *et al.*, 2018). No cap binding domain for bunyaviruses has been identified as of yet. Comparisons with 'Flu segmented RdRp suggest this will reside within the CTD of *Bunyavirales* L. The distance between the cap binding domain and endonuclease domain is thought to dictate the length of the snatched primer, approximately 10-20 nt long; whilst the cap snatched by flu is 10-13 nucelotides (Garcin *et al.*, 1995; Duijsings, Kormelink and Goldbach, 2001; Cheng and Mir, 2012; Datta *et al.*, 2013).

Unlike most mRNAs the 3' end of *Bunyavirales* messages do not possess a poly(A) tail, with the exception of Sin Nombre virus (SNV) a *Orthohantavirus* (Hutchinson, Peters and Nichol, 1996). Unexpectedly, BUNV M- and L-segment mRNA termination site mapping did not reveal U-rich regions (Barr, Rodgers and Wertz, 2006). BUNV 3' non-translated

mRNA secondary structures enhance translatability, possibly functioning to replace the usual polynucleotide-A (pA) tail (Blakqori, van Knippenberg and Elliott, 2009).

Transcriptional termination in *Orthobunyaviruses* is well characterised for the M- and S-segments; here mRNA terminates between 40-100 nucleotides before the end of the RNA template. Whereas L-segment mRNA is terminated by run-off (Patterson and Kolakofsky, 1984; Eshita *et al.*, 1985; Cunningham and Szilágyi, 1987; Jin and Elliott, 1993; Coupeau *et al.*, 2013). The BUNV S-segment transcription termination signal was mapped to 33 nucleotides within the 5' non-translated region and included a conserved hexanucleotide sequence: 3'GUCGAC-5' plays a crucial role. A second motif 3'-UGUCG-5' was also found further downstream in the S- and L-segment but not the M-segment. S-segment sequence comparisons with *Othrobunyaviruses* revealed high sequence conservation with the mRNA 3' end sequence having the potential to form short stem loops as well as the conserved six nucleotide motif (Barr, Rodgers and Wertz, 2006).

Gn, Gc and NSm of the M-segment are transcribed and translated as a polyprotein which is co-translationally cleaved by cellular peptidases (Shi *et al.*, 2009). Mature glycoproteins are produced via the dimerisation of Gc and Gn in the RER prior the transportation to host membranes where they insert ready for viral budding.

Nucleotides in the 3' and 5' non-translated regions of the *Orthobunyavirus* genus contain signals for mRNA transcription and genome regulation; in BUNV 17 nucleotides from the 3' and 5' termini and required for RNA replication (Mohl and Barr, 2009) and this is sequence specific for the RdRp. Interestingly, a conserved mis-match nucleotide in BUNV is nucleotides nine at the 3' termini; for optimal transcription this must a U base, and this is independent of the corresponding 5' base at nucleotides nine (Barr and Wertz, 2005). On the other hand, the BUNV transcriptional promoter sequence compromises nucleotides within the 3' and 5' termini, suggesting that the RdRp requires both termini for transcriptional initiation (Barr and Wertz, 2004; Kohl *et al.*, 2004; Barr, Rodgers and Wertz, 2005).

The crystal structure of LACV RdRp in complex with the 3' and 5' terminal RNA shows two RNA binding sites on the RdRp. The 3' terminus binds as a single-strand at nucleotides 1-8 whereas 5' nucleotides 1-10 form a stem-loop structure. This interactions is thought to be mediated by the RdRp rather than inter-terminal base pairing (Gerlach *et al.*, 2015). It is possible that this interaction is not mutually exclusive, rather that both forms are needed at different viral replication stages.

The switch between viral transcription and replication is not yet understood. Replication begins with the production of an anti-genomic strand which acts as a template for replication which is then protected and coated by NP. Initiation of primer-independent RNA polymerisation at the 3' termini of a segment occurs, producing a full-length copy of a template.

Cycloheximide (a protein synthesis inhibitor) treatment prevented transcription and genome replication, suggesting that continuous protein synthesis is required for genomic replication; probably due to the need for 'free' NP required for newly synthesised viral

RNA encapsidation. This has been shown to be true for other negative sense RNA viruses including vesicular stomatitis (VSV) virus where NP supply is crucial and is recognised as a mediator between the transcription and replication switch (Noton and Fearns, 2015).

RNP complex formation models have been proposed: Mohl et al., examined if encapsidation of RNA by BUNV NP requires secondary RNA structures by M-fold analysis. Firstly Mohl et al., identified that BUNV segments are active for replication with 18/25 nt from the 3' anti-genomic NTR and 17/25 nt from the 5' genomic NTR; implying that encapsidation signals reside within corresponding 5' nucleotides in the nascent replication products. These nucleotides were therefore subjected to M-fold analysis giving delta G values of -1.6 kcal/mol and -1.2 kcal/mol respectively (Mohl and Barr, 2009). In support of this, the 5' residues have been shown to form a short stem-loop structure proposed for NP binding (Osborne and Elliott, 2000). On the other-hand, high-resolution crystal structures of NP +/- RNA do not provide structural evidence that RNA encapsidation is sequence specific, suggesting that encapsidation signals may not exist (Ariza *et al.*, 2013).

Encapsidated genomic RNA is then transported through the Golgi apparatus prior to the migration of Golgi vesicles via actin filaments to the host cell surface where new progeny virus form and bud from the host cell membrane (Elliott, 2014).

It has been suggested that NP preferentially binds genomic RNA, however this is not well characterised. Secondary structures within the 5' stem-loop structures of BUNV have been identified as preferentially NP binding sites and/or signals. It has been suggested that capped mRNAs have lost and/or disrupted this stem loop therefore giving preference to genomic RNA (Osborne and Elliott, 2000).

1.2 Introduction to Filoviruses

1.2.1 Discovery

Outbreaks of *Ebolavirus* (EBOV) were first identified in 1976 in the Democratic of Republic of Congo, where an individual treated for malaria, after returning from Zaire suffered from fever and chills before death from haemorrhagic fever. This first outbreak infected 318 individuals with 280 deaths (88% fatality rate). In the same year an outbreak in Sudan caused 151 deaths (53% fatality rate). Virologic investigations found the virus to be similar to the previously identified *Marburg marburgvirus* (MARV) cultivated from green monkey kidney cells in Marburg, Germany. Peter Piot (Institute of Tropical Medicine, Antwerp, Belgium) dubbed this newly identified virus as EBOV after the Ebola river in Zaire (Pattyn *et al.*, 1977).

1.2.2 Classification

First established in 1991, *Monongeavirales* order comprise related viruses that have non-segmented single-stranded negative sense RNA genomes. In 2017, the taxonomy was updated to comprise eight families; *Bornaviridae, Filoviridae, Mymonaviridae, Nyamiviridae, Paramyxoviridae, Pneumoviridae, Rhabdoviridae* and *Sunviridae*. In 2019 the *Monongeavirales* was further updated and is now formed of eleven families. The additional three families include: *Artoviridae*, *Lispiviridae* and *Xinmoviridae* (Amarasinghe *et al.*, 2019). Classification into this order was based on: a linear non-segmented single-stranded negative sense RNA genome, specific genome organisation: 3' UTR leader -> core protein genes -> envelope protein genes -> L -> trailer 5' UTR, 3' single promoter which produces monocistronic mRNAs by a stop-start mechanism, genome replication occurs via antigenomes and the RdRp is highly conserved amongst *Mononegavirales*.

The *Filoviridae* family includes six genera: *Cuevavirus*, *Dianlovirus*, *Ebolavirus*, *Marburgvirus*, *Striavirus* and *Thamnovirus* and 11 species. EBOV and MARV are currently known to infect humans whereas *Cuevavirus* has only been recovered from bats. There are five species in the *Ebolavirus* genus, four of which cause haemorrhagic fever; *Bundibugyo* (BDBV) causes 40% fatality in humans, *Reston* (RESTV) causes asymptomatic infection in humans, *Sudan* (SUDV) (41-65% fatality), *Tai Forest* (TAFV) is currently only known to have caused two non-fatal human infections and *Zaire* (ZEBOV, recently renamed EBOV) has the highest mortality rate of 57-90%. In 2018, *Bombali ebolavirus* (BOMV) was discovered after samples from bats in Sierra Leone were recovered. BOMVs nucleotide and amino acid identities to other *Ebolaviruses* are 55-59% and 64-72% respectively. The five other *Filoviridae* genera all have a single species (Figure 1.11). Filoviruses typically use bats as reservoir hosts (Messaoudi, Amarasinghe and Basler, 2015). Bats are often reservoirs for many viruses including the recent SARS-CoV-2 outbreak (COVID-2019). It is thought that bats support viruses as long term-persistent infections, rather than transient pathologies (Plowright *et al.*, 2016). Bats antiviral state induced by the IF pathway protects bat cells from mortality in cell culture and enhances establishment of long-term persistent infection (Brook *et al.*, 2020). Bats have been shown to constitutively express IFN in the absence of viral RNA or DNA and

without inflammation (P. Zhou *et al.*, 2016). Viruses then evolve rapid replication rates under bat antiviral defences.

The *Mononegavirales* share five common proteins: NP, viral proteins 35 and 40 (VP35 and VP40), GP and L whereas viral proteins 30 and 24 (VP30 and VP24) are unique to the *Filoviridae*.
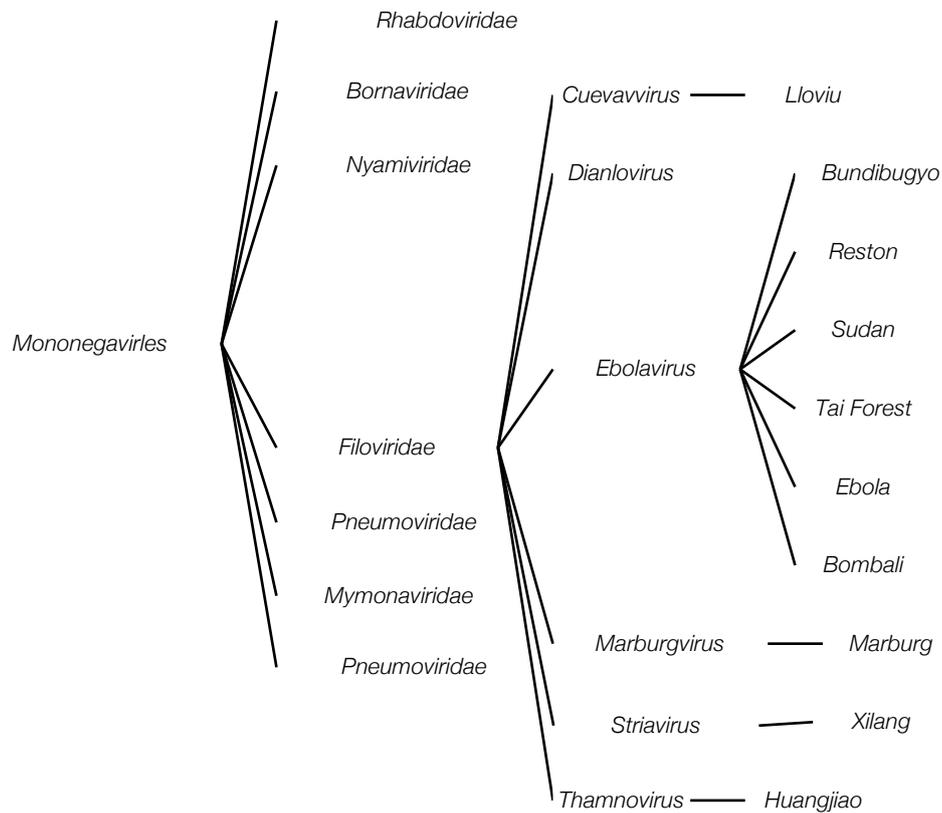
Figure 1.11 Taxonomy of Filoviruses

1.2.3 Zoonotic Life Cycle

EBOV route of infection occurs via mucosal surfaces, abrasions and parental transmission (Feldmann and Geisbert, 2011; Falasca *et al.*, 2015). Primarily, EBOV infects epithelial cells but causes a systemic infection in most organs and glands. Symptoms of infection include fever, haemorrhage and shock. Studies have shown disease progression occurs due to an inflammatory response leading to respiratory distress and hypotension (Hoenen, Groseth, Falzarano, Feldmann, *et al.*, 2006).

1.2.4 Impact, Epidemiology and Symptoms

The 2014-2016 outbreak cost an estimated $2.8 billion in gross domestic product in Liberia, Guinea and Sierra Leone. There was loss in private sector growth, cross-border trade and agricultural loss. In total this outbreak caused over 28,000 cases and 11,000 deaths (39% mortality); amongst healthcare workers there were 881 infections and 513 deaths (58% mortality). The cost of response was $3.6 billion with the Unites States of America, United Kingdom and Germany being the top donors (Evans, Goldstein and Popova, 2015). The second largest outbreak recorded in the Democratic Republic of Congo (2018-2020) reported 3,481 total cases with 2,299 deaths (66% mortality).

Since 1976, there have been 34,000 reported cases of EBOV infections of which 15,200 resulted in death (44% mortality, excluding the current 2020 outbreak in DRC). 22 outbreaks have occurred and 17 minor cases in 19 countries from 1976-2020. (Centres for Disease Control and Prevention (CDC)., 2020). Currently the DRC is facing its 11th outbreak of EBOV in the past 40 years. WHO has announced that there is no link between the 2018-2020 outbreak and the current outbreak. As of 2nd of September 2020 there have been 110 cases and 47 deaths (WHO, 2020).

Ebola virus haemorrhagic fever is characterised by an imbalance in the immune system causing multi-organ shock. Disordered coagulation and tissue damage lead to severe bleeding. Infections have been reported to have fatality rates as high as 90%, on average however fatality rates are 65%. MARV infections where mortality from Marburg haemorrhagic fever is between 23-90% but averages at 50%; from 1967-2014 there have been 12 outbreaks of MARV but only 465 reported cases (CDC, 2019).

1.2.5 Treatment

1.2.5.1 Vaccines and Immuno-Therapy

Despite multiple outbreaks since 1976, until recently no FDA vaccine or antiviral treatment has been approved for treatment in humans. Due to an emergency response to the 2013-2016 epidemic, there have been several vector-based vaccine candidates.

In 2015, the efficacy of recombinant adenovirus type-5 vector-based Ebola (Ad5-EBOV) vaccine was performed as a phase-2 clinical trial in Sierra Leone. The vaccine expressed the viral GP of the Makona strain. Non-human primates showed a 77% protection against

death but GP-specific antibodies elicited were not long lasting or robust in humans (Zhu *et al.*, 2017).

Recently, the Ad6-EBOV vaccine was approved by the FDA in July 2020. Ad6-EBOV uses the EBOV Mayinga strain GP on the surface of an adenovirus virion. This is boosted by a second dose of modified vaccinia ankara, which produces virus like particle using a pox virion, with a TAVF NP and EBOV, SUDV and MBOV GP (FDA, 2020).

The 2018-2020 outbreak in the Democratic Republic of the Congo saw the World Health Organisation (WHO) officials administrating 303,000 patients with the recombinant vesicular stomatitis virus-ZBOV (rVSV-ZBOV or Ervebo) vaccine which was been shown to be 100% effective in the 7,651 individuals vaccinated in Guinea in 2015. In December 2019 this approved by the FDA for adults aged 18 years and over (FDA, 2019).

ZMapp, a cocktail of three chimeric, humanised, monoclonal antibodies target different sites of EBOV GP. ZMapp, was used as a complementary therapy for the treatment of EBOV patients during the 2013-2016 outbreak, however the efficacy was not assessed due to co-administration with convalescent plasma (CP) transfusion (Davey *et al.*, 2016). Convalescent whole blood (CWB) and CP from recovered patients carry specific antibodies and have been used to treat patients, however the effectiveness till remains elusive (Mendoza, Racine and Kobinger, 2017).

Monoclonal antibody Mab114 and triple monoclonal antibody REGN-EB3, efficacy was assessed in a randomised, controlled trial of EVD during the 2018 outbreak, and compared to ZMapp. Both were superior to ZMapp in the 681 patients enrolled. Death occurred in 35.1% of Mab114 patients, and 33.5% patients receiving 33.5%, compared to 51.3% receiving ZMapp (Mulangu *et al.*, 2019).

1.2.5.2 Small Molecule Inhibitors

Favipiravir (6-fluoro-3-hydroxy-2-pyrazinecarboxamide) is a broad-spectrum inhibitor for viral RNA polymerases. Favipiravir has shown a 100-fold decrease in viral RNA determined by qPCR and increased survival rates. However, monotherapy is unlikely to be effective due to patients having high viral loads (Ct value <20), and rapid viral mutation rates (Bai *et al.*, 2016; Sissoko *et al.*, 2016).

Remdesivir (GS-5734, Gilead Sciences) is a monophosphoramidate prodrug pf a C-adenosine nucleoside analogue, terminating viral RNA synthesis by inhibiting the polymerase (Cho *et al.*, 2012). Remdesivir was in clinical trials for the treatment of EVD in the Democratic Republic of Congo during 2014 and showed selectivity for the viral RdRp over mammalian polymerases determined by two defined ten-amino acid motifs in the polymerase (Hoenen, Groseth and Feldmann, 2019). Remdesivir showed positive results with 67% survival at 28 days after a positive PCR test (Mulangu *et al.*, 2019).

1.2.6 Virion Morphology

EBOV virions form filamentous rods, composed of a central helical nucelocapsid (NC) wrapped around viral genomic RNA. Virions are 970-1,200 nm in length and EBOV is 90

nm in diameter whereas MARV is 92 nm in diameter (Bharat *et al.*, 2012). The NC is composed of nucleoprotein (NP) and viral RNA. The NC together with viral protein (VP) 35, VP30 and an RNA-dependent RNA polymerase (L), form a ribonucleoprotein (RNP) complex. This is surrounded by the matrix space and a layer of the matrix proteins VP24 and VP40 encased by a lipid envelope derived from a host cell plasma membrane saturated in glycoprotein (GP) (Goeijenbier *et al.*, 2014; Falasca *et al.*, 2015; Biedenkopf, Lier and Becker, 2016).

1.2.7 Genome

The EBOV RNA genome is single-stranded (ss) non-segmented and in the negative polarity. EBOV's 19 kb genome has seven genes arranged tandemly; 3'-leader-*NP-VP35-VP40-GP-VP30-VP24-L*-trailer-5' (Figure 1.12). *GP* also encodes the small glycoprotein and the soluble small glycoprotein (sGP and ssGP respectively). sGP, (the non-edited form of GP) is secreted and may play a role in immune evasion (Mühlberger 2007). Whereas it is the RNA edited form of GP (ssGP) that is found membrane-bound, and is the only protein on the virion surface (Mühlberger *et al.*, 1999). GP is composed of two subunits GP_1 and GP_2.

The 3'-leader region contains a bipartite replication promoter and initial transcription start site, shown to form stem-loop structures (Sanchez *et al.*, 1993; Mühlberger *et al.*, 1996). Each gene encodes the respective ORF and each is flanked by highly conserved transcriptional gene start (GS) and gene end (GE) signals which are separated by intergenic regions (IGRs). IGRs differ in length; for example, *NP-VP35* is five nucelotdies whereas *VP30-VP24* is 144 nucleotides (Figure 1.12). IGRs also overlap at three boundaries sharing nucleotide sequence ATTAA between genes: *VP35-VP40*, *GP-VP30* and *VP24-L* (Figure 1.12). Interestingly, it has been shown the length of the intergenic regions for EBOV are not essential for transcription of the downstream gene (Brauburger et al. 2016). GS signals directly to L and initiates mRNA synthesis, 5'-7-methylguanosine cap addition and methylation. Replication promoters are located within the 3' end of the genome and antigenome therefore L can only access the RNA template at the 3' end. Promoter sequences here are 155 and 176 nucleotides in length for the genome and antigenome respectively (Calain, Monroe and Nichol, 1999). GE signals contain a stretch of 5-6 uridines which are copied by L forming a pA tail via a stuttering mechanism. This mechanism involves the RdRp inserting an adenosine and then moving 'backwards' one nucleotide (with the nascent mRNA chain intact) resuming transcription of a second adenosine and the process is repeated giving rise to the classic pA tail before the nascent mRNA chain is released. Moreover, GE signals terminate the transcription of genes via the transcription stop sequence 3'UAAUUC. Filovirus transcription signals have been predicted to form stable RNA secondary structures. Lastly, the 5'-trailer region codes for a complementary sequence for the replication promoter utilised by L (Schlereth *et al.*, 2016).

The viral RNA genome is encapsidated by NP forming the NC. It is here that the polymerase complex interacts via VP35 and NP to transcribe and replicate the genome. The polymerase complex is composed of VP35 homo-oligomer, L and VP30 homo-

oligomers needed for transcription. VP30 is not essential for replication. (Biedenkopf, et al. 2016).
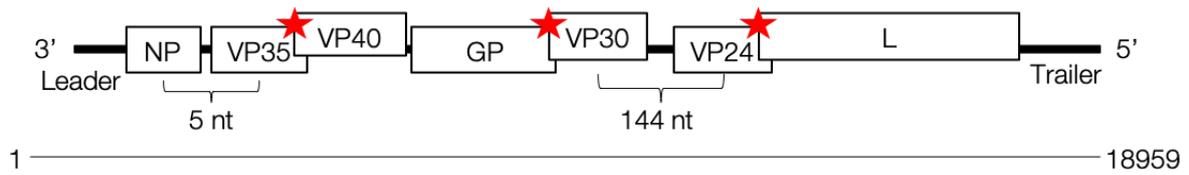
Figure 1.12 Genetic Map of EBOV

EBOV encodes eight proteins from seven genes. From 3' to 5' end; nucleoprotein (NP), viral protein 35 (VP35), viral protein 40 (VP40), glycoprotein (GP produces two proteins via mRNA editing), viral protein 30 (VP30), viral protein 24 (VP24), RNA-dependent RNA polymerase (L). IGRs overlap (red star) and share a ATTAA nucleotide sequence. IGRs between ORFs differ in length; 5 nt between NP and VP35 and 144 nt between VP30 and VP24.

1.2.8 Proteins

The *Filoviridae* produce eight protein from seven genes, their function is discussed in more detail below.

Table 1.3: Filovirus Protein Function

Protein	Function
NP	Encapsidates viral RNA forming an RNase-resistant nucleocapsid.
VP35	Polymerase cofactor. Type I IFN antagonist.
VP40	Matrix protein of the inner membrane. Key role in membrane budding.
ssGP	Class I membrane protein. Mediates fusion and receptor binding during viral entry. sGP is formed from non-edited mRNA, secreted acting as an anti-inflammatory protein.
VP30	Transcriptional activator.
VP24	Minor matrix protein. Involved in NC formation and assembly. Counteracts type 1 IFN response.
L	RNA-dependent RNA polymerase, enzymatic activity required for transcription and replication.

1.2.8.1 VP35

The polymerase co-factor VP35 (354 amino acids, 35 KDa), is positioned second in the genome and shares sequence homology with the phosphoprotein (P) of other *Mononegavirales* including hRSV P (discussed in 1.3.8.2); similarities lie mostly within the N-terminal domain (monomeric NP binding domain) and the central oligomerisation domain which is thought to form coiled-coil domain in both proteins (Figure 1.14). VP35 has a flexible NTD important for NP interaction, a central oligomerisation domain which contains a coiled-coil domain and an IFN inhibitory domain (IID) (Figure 1.13). Structural analysis of the EBOV VP35 coiled coil domain revealed a tetramer while (Bruhn *et al.*, 2017), by contrast, the NTD crystal structure of MARV VP35 shows a trimer. These differences may reflect the subdomains expressed recombinantly or be a true reflection of the oligomeric state of each protein.

```
VP35 ---MTTRTKGR-----GHTVATTQNDRMPGPELSGWISEQLMTGRIPVNDIFCDIENNPG 52
P    MEKFAPEFHGEDANNRATKFLESIKGKFTSPKDPK-----KKDSIISVNSIDIEVTKESP 55
         ::  . :*.  .  ..  :  :.::  .*: .  *  **.*  ::  . ::
VP35 LCYASQ----------------MQQTKPNPKMRNSQTQTDPICNHSFEEVVQTLASLATV 96
P    ITSNSTIINPTNETDDTAGNKPNYQRKPLVSFKEDPTPSDNPFSKLYKETIETFDNNEEE 115
            :  *  *  **  .:::.  *  :*  .:  ::*.::*:  .
VP35 -----VQQQTIASESLEQRITSLENGLKPVYDMAKTISSLNRVCAEMVAKYDLLVMTTGR 151
P    SSYSYEEINDQTNDNITARLDRIDEKLSEILGMLHT---------------LVVASAGP 159
          :  :  :.:.:  *:  ::: *.  :  .*  :*  *:*  ::*
VP35 ATATAAATEAYWAEHG---QPPPGPSLYEESAIRGKI-----------ESRDETVPQSV 196
P    TSARDGIRDAMIGLREEMIEKIRTEALMTNDRLEAMARLRNEESEKMAKDTSDEVSLNPT 219
     ::*  .  .:*  .  :  :  :*  :.  :..  ::  **.  :  .
VP35 REAFNNLDSTTSLTEENFGKPDISAKDLRNIMYDHLPGFGTAFHQLVQVICKLGKDSNSL 256
P    SEKLNNL------LEGNDSDNDLSLEDF--------------------------- 241
       *  :***  *  *  ..  *:*  :*:
VP35 DIIHAEFQASLAEGDSPQCALIQITKRVPIFQDAAPPVIHIRSRGDIPRACQKSLRPVPP 316
P    ---------------------------------------------------------- 241
VP35 SPKIDRGWVCVFQLQDGKTLGLKI 340
P    ----------------------- 241
```

Figure 1.14 Protein Sequence Alignment of EBOV VP35 and hRSV P

* conserved residue.
: conservation between groups of strongly similar properties.
. conservation between groups of weakly similar properties.
Alignment made using Clustal Omega (Sievers *et al*., 2011)
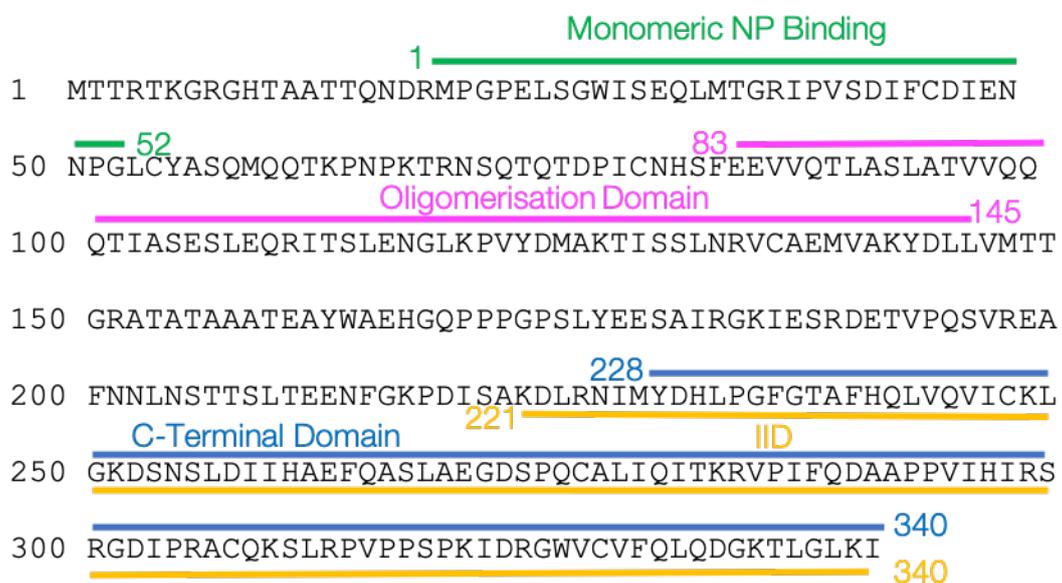
Figure 1.13 Functional Regions of EBOV VP35

Functional regions have been mapped for full-length VP35 (1-340). N-terminal domain binds monomeric NP (1-52, green). Oligomerisation domain (83-145, pink) forms coiled coils. C-terminal domain (228-340) overlaps with the interferon inhibitory domain ((IID) 221-240, yellow) which binds double-stranded RNA.

VP35 is able to inhibit IFN -α and -β via retinoic acid-inducible gene I (RIG-I)-dependent signalling inhibition. Normally RIG-I would activate the transcription factor IFN regulator factor (IRF) -3 and IRF-7, which in turn regulate IFN-α/-β gene expression respectively. The ability of VP35 to bind double stranded (ds) RNA has been directly correlated to VP35's function as a virulence factor (Leung *et al.*, 2010).

As the EBOV genome is encapsidated in NP, it suggests that VP35 which is an essential component of the RNP complex, associates with L and guides it to the NC interacting with NP. VP35 is also thought to function in a similar way to hRSV P (discussed in detail in 1.3.8.2), chaperoning NP to viral RNAs in order to prevent association with host cellular RNAs. In minigenome assays absence of VP35 abolished reporter gene expression. Co-immunoprecipitation assays have shown VP35 to interact with both L and NP (Becker *et al.*, 1998; Möller *et al.*, 2005; Prins *et al.*, 2010). A binding pocket between helices 13 and 15 of NP binds VP35, in assembled helices this is then occupied by an adjacent NP subunit consistent with the idea that VP35 maintains NP in its monomeric state. VP35 is then displaced by an NP N-terminal helix in tandem with RNA encapsidation (Wan *et al.*, 2017). The crystal structure of closely related vesicular stomatitis virus (VSV) from the *Rhabdoviridae* family, shows the binding site of P overlapping with the NP RNA-binding site (Figure 1.15) and NP N-terminal arm domain, preventing both RNA binding and NP oligomerisation respectively when soluble NP is RNA free (Leyrat *et al.*, 2011).

Leung et al., tested truncation mutants in minigenome assays to map the specific VP35 region required for NP binding. Minigenome analysis revealed residues 20-52 (termed NPBP) as crucial for replication. Further analysis by isothermal titration calorimetry (ITC) revealed that residues 20-42 comprised the most critical region for high affinity NP binding (K_D = 19.4 ± 0.4 nM). Competitive RNA binding studies revealed that NPBP can compete with ssRNA for the NP binding site with an IC_{50} of 4 μM (Leung *et al.*, 2015).

When subjected to size exclusion chromatography (SEC), ΔNP_{NTD} and NPBP form a heterodimer indicating that this interaction prevents NP oligomerisation (Leung *et al.*, 2015).
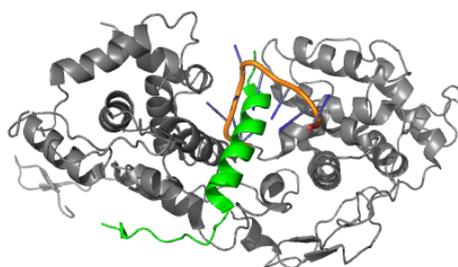
Figure 1.15 Crystal Structure of Vesicular Stomatitis Virus NP_{22-422} in Complex P_{6-35}

VSV NP (grey) and P (green) binding site overlaps with the RNA (orange) binding site.
RNA superimposed in PyMol
Adapted from Leyrat *et al.*, 2011.
Figure made using PyMol (version 2.3.2).
PDB: 3PMK.

The crystal structure of ΔNP$_{NTD}$:NPBP heterodimer was solved to 3.7 Å and revealed a ring-like assembly, with a 6 nm inner diameter formed of eight copies of ΔNP$_{NTD}$; this complex forms back-to-back double rings with 16 copies of the heterodimer. From this structure two lobes were identified for ΔNP$_{NTD}$, a head lobe (residues 38-240) and a foot lobe (residues 244-383) (Figure 1.16). The lobes are connected via a flexible linker in the head lobe which also contains 12 α-helices and two parallel β-strands. On the other-hand the foot lobe contains 10 α-helices and 2 short anti-parallel β-strands (Leung *et al.*, 2015). NPBP forms two helices (residues 26-36 and 40-42) and interacts with the foot lobe via hydrogen bonds and non-bonded contacts.
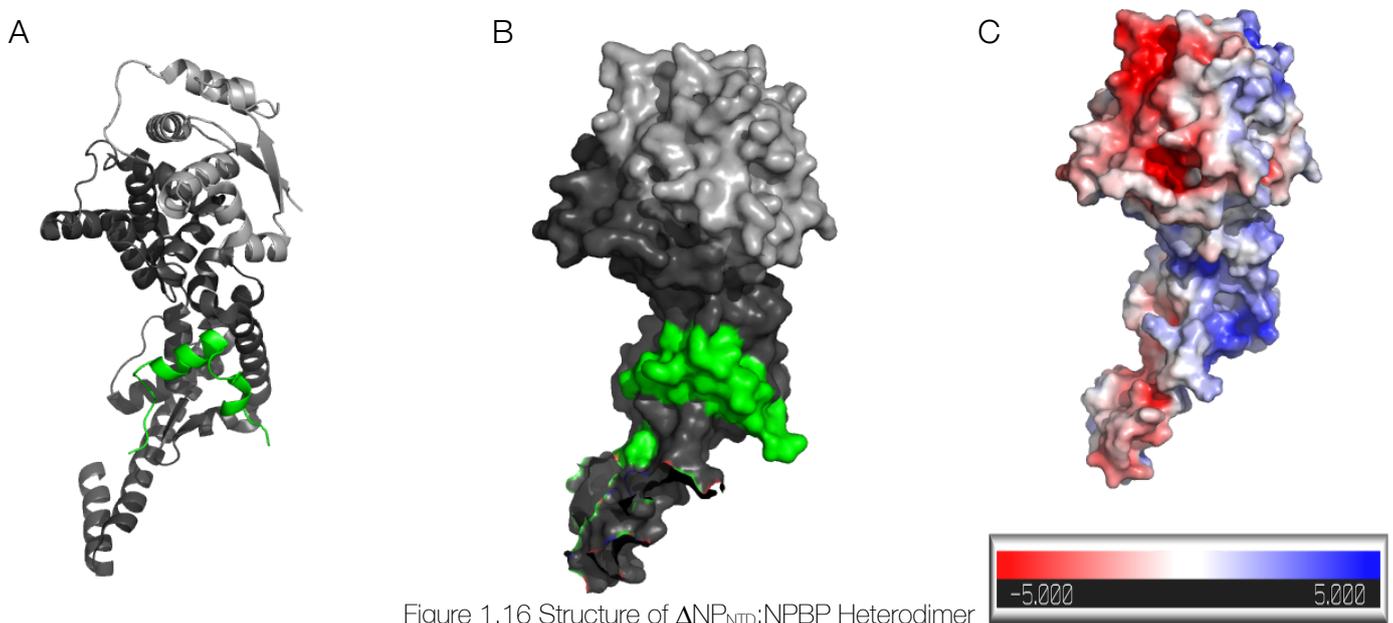
Figure 1.16 Structure of ΔNP$_{NTD}$:NPBP Heterodimer

A- Structure of the heterodimer. Head-lobe (residues 37-146, light grey) and foot-lobe (residues 240-285, dark grey). NPBP shown in green.

B- Surface representation

C- Electrostatic surface representation. Shows complementary and extensive hydrophobic interactions at the interface. Red, white and blue represent negative, neutral and positive potential respectively.

Figures made using PyMol (version 2.3.2).

PDB: 4YPI.

1.2.8.2 VP30

VP30 is a multifunctional protein, and acts as an essential activator of viral transcription regulated by phosphorylation. The 26 kDa protein has been shown to form hexamers in solution and infectious particles (Hartlieb et al. 2007). Disruption of the hexamerisation domain causes a loss of RNA binding and loss of VP30 function (Biedenkopf *et al.*, 2016).

Transcriptional activation is dependent on an RNA hairpin at the transcription start site within the first gene (NP) leader region (nucleotides 54-80) (Modrof *et al.*, 2002; Weik *et al.*, 2002; John *et al.*, 2007). VP30 promotes read-throughs at this hairpin enhancing transcription (Weik *et al.*, 2002) and acts as a transcription activator supporting transcription re-initiation at gene ends (Modrof *et al.*, 2002). It has been shown by John et al., that VP30 does not directly bind these stem-loop structures. Rather that it is the 3' or 5' extensions that are required for binding, implying that VP30 interacts outside of the stem-loop structure. Deletions within the leader region cause VP30-independent viral RNA transcription (Weik *et al.*, 2002; John *et al.*, 2007)

VP30 has 3 domains; a zinc-binding domain, phosphorylation sites and an RNA binding site. VP30 Filovirus sequence analysis revealed a motif similar to the zinc finger domain first characterised in Nup475 (mammalian nuclear protein) (Mühlberger et al., 2003). The Zn^{2+}-is co-ordinated by the conserved cysteines and histidines. This motif (Cys3-His comprising residues 68-95) is highly conserved across *Filoviruses*, and a similar motif has also been identified in the M2-1 protein of human orthopneumovirus (hRSV) (discussed in 1.3.8.1) (Modrof, Becker and Mühlberger, 2003; Tanner *et al.*, 2014). RNA binding activity of VP30 is dependent on the presence of zinc (Zn^{2+}) but this motif alone does not have RNA binding activity. This motif is located within the N-terminus of VP30 which suggests that the RNA interacting domain is also in the N terminus. Deletions within this region render VP30 inactive for RNA binding (John *et al.*, 2007). Specifically, residues 26-40 are important for the VP30:RNA interactions. Residues here contain two tyrosines at residues 35 and 39 and four arginines at residues 28, 32, 36 and 40.

EBOV VP30 is phosphorylated at its N-terminus within two serine clusters (amino acids 29-31 and 42, 44 and 46), (Martínez *et al.*, 2008) and also at threonine 143 and 146 (Modrof *et al.*, 2002). These phosphorylation sites overlap with the arginine-rich RNA binding site on VP30. On the other hand, in MARV VP30 a stretch of seven serines (residues 40-51), which represent the main RNA binding domain, and threonines (145 and 150) are yet to be identified as phosphorylation sites. Phosphorylation regulates EBOV VP30 activity during transcription, through the regulation of the association between EBOV VP30 and the NC, altering the balance of viral transcription and RNA replication. Mutations within these phosphorylation sites alter interactions with NP-induced inclusion bodies and RNA (Modrof *et al.*, 2002; Martínez *et al.*, 2008). Phosphatase 1 (PP1) and PP2A have been identified to dephosphorylate VP30 however kinase(s) are still unknown. Inhibition of PP1 in animals blocked viral proliferation due to VP30 hyperphosphorylation which did not support viral transcription (Ilinykh et al., 2014; Lier et al., 2017). Lier et al., have suggested that cellular phosphatases maintain VP30 in its transcriptionally active dephosphorylated state (Lier, Becker and Biedenkopf, 2017). Mutating six serine to aspartates or alanine showed differential affinities to WT VP30 for

RNA. The serine to asparagine mutations had a 2-fold lower affinity and serine to alanine showed increased affinity. An increase in negatively charged residues impairs VP30s RNA binding function (Biedenkopf, Lier and Becker, 2016).

Additionally, Biedenkopf *et al.*, described that the dynamic phosphorylation of a single serine at position 29 is sufficient to activate primary viral transcription. Thus, in EBOV it is through a series of phosphorylation and de-phosphorylation events that trigger the binding and release of VP30 from the RNP complex, and NC essential for full functionality of VP30 (Biedenkopf, Lier and Becker, 2016). Interestingly, MARV VP30 has been shown not to have a central role in minigenome assays as NP, L and VP35 represented the minimal requirement for transcription and replication, but is still required for virus rescue (Mühlberger *et al.*, 1999; Fowler *et al.*, 2005). Data shows a correlation between VP30 phosphorylation and VP35 interactions, suggesting a switch between transcription and replication due to weakened interactions with the RNP complex. However, is it the absence of VP30 converting the transcriptase complex to replicase? Or do the replicase complexes have preferred access to NP-RNA templates over transcriptase? (Biedenkopf, Lier and Becker, 2016)

Minigenome system analysis has revealed possible mechanisms that VP30 may use in its transcriptional regulatory role. It has been suggested that VP30 interacts with one or more of L, NP, VP35 to promote increased stability of a transcriptional complex (John *et al.*, 2007). This would yield higher levels of mRNA due to an increase in the number of active transcription initiation complexes and/or increased stability of the transcriptional elongation complex (Mühlberger *et al.*, 1999; John *et al.*, 2007).

On the other hand, it is thought that VP30 may interact directly with viral RNAs in order to regulate transcription. Bioinformatic analysis revealed possible RNA-binding motifs and 3 intrinsically disordered regions. Residues 26-32 of an N-terminal distorted region are thought to be involved in ligand-binding (John *et al.*, 2007) and are suggested to be an RNA-binding domain due to the zinc-finger domain and arginine rich sequence.

Moreover, phosphorylation of VP30 is thought to decrease VP30:VP35 interaction and VP30:RNA interaction favouring transcription.

Phosphorylated mutants of VP30 have been shown to be concentrated within NP-induced inclusion bodies, known sites of viral replication, while dephosphorylated VP30 mutants were diffused within the cytoplasm (Biedenkopf et al, 2016).

The crystal structure of VP30 CTD was solved to 2.2 Å, and crystallised in space group $P2_12_12_1$. The crystal lattice contained two monomers per asymmetric unit. Monomers fold into a compact helical core with a seventh helix that points away from the core interacting with the adjacent monomer (Figure 1.17 A). Contacts are polar between side chains, backbone amides and carbonyls between the loop region that connects helices six and seven. This loop lies within a groove on the neighbouring monomer between helices one and two. At the dimerisation interface, hydrogen bonds are found and a single salt bridge between E152 and K149 (Figure 1.17 B).

Figure 1.17 Crystal Structure of VP30 C-Terminal Domain
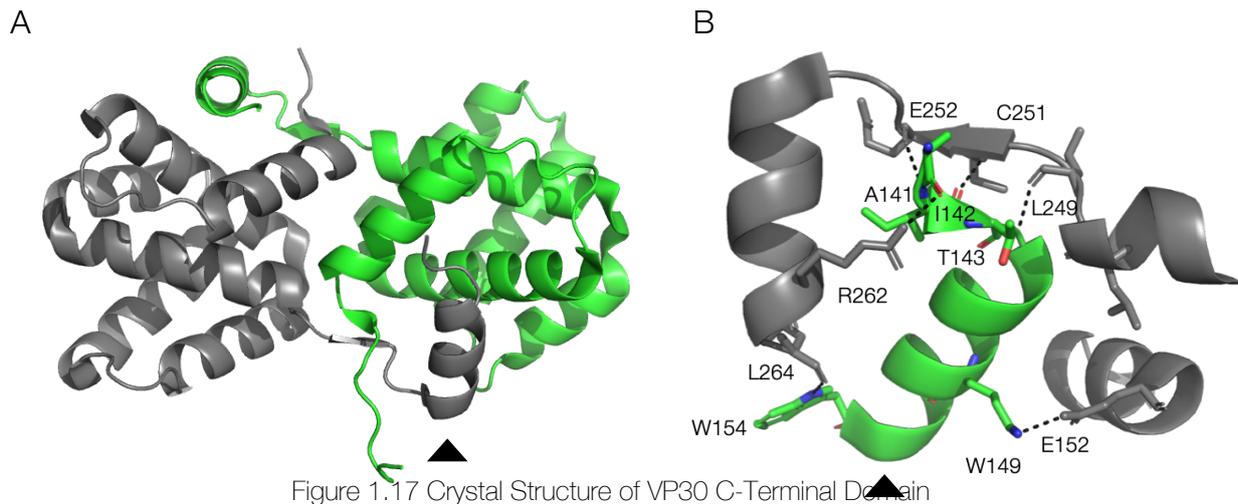
A- VP30 CTD dimer. Monomers represented by grey and green. Monomers are dominated by a compact core composed of six helices. Dimerisation forms via a head-to-head mechanism linked by helix 7 that points away from the core (black arrow).
B- Dimerisation interface. Interactions between helix seven and the loop region connecting helix six and seven and opposing helix 1.
Figures made using PyMol (version 2.3.2).
PDB: 2l8B.

Hartlieb et al., transiently expressed VP30, VP30$_{NTD}$ and VP30$_{CTD}$ in HUH-7 cells and showed that all forms of VP30 co-localised with NP inclusion bodies, interestingly only VP30 and VP30$_{CTD}$ were found in VLPs suggesting that the CTD is essential for VP30:NP interaction. VP30$_{CTD}$ mutational studies showed that glutamic acid 197 and a basic cluster (arginine179, lysine 180 and 183) are important or NP co-localisation and VLP incorporation. Hartileb et al., then went on to show that it is the basic cluster that is important for transcription.

VP30 interacts with NP through its NTD and CTD suggesting that the NTD interacts loosely with NP-RNA helical-coils that are transcription competent and a tight CTD interaction with packaged NP-RNA that's incorporated into VLP (Hartlieb *et al.*, 2007).

Kirchdoerfer et al., described that VP30$_{CTD}$ interacts with a short peptide within the NP C-terminal residues 360-739. This interaction has an affinity of 21.3 ± 4.5 µM. Biolayer interferometry showed that NP peptides containing conserved residues 600-617 is a major binding site for VP30. This interaction was shown to bind at 5.69 ± 0.04 µM. The

combination of these two sites and the avidity effect from oligomeric interactions will also make this interaction tighter so affinities may be stronger than reported.

Electrophoretic mobility shift assays (EMSA) have been used in attempting to quantify the K_D of the RNA:VP30 interaction. Biedenkopf *et al*., used 3' 1-154 nucleotide genomic RNA; this encodes essential signals such as the transcriptional start site, replication, encapsidation and a 56-158 nucleotide antigenomic RNA that encodes the mRNA start region. $VP30_{CTD}$ only bound weakly to the genomic RNA 3' 1-154 with an apparent K_D of 40 μM. A truncated $VP30_{8-272}$ was used for further analysis. EMSA with RNA 3' 1-154 nt yielded a K_D of 0.35 μM and the antigenomic RNA 0.49 μM. Schlereth et al., then carried out a series of deletions on the 3' 1-154 genomic RNA and tested these truncations for RNA binding. A deletion of the first 79 nucleotides, which forms a hairpin and encodes a transcriptional start site, bound to VP30 with a K_D of 0.99 μM, a 3-fold drop compared to the WT $VP30_{8-272}$. This hairpin structure was also shown to be important for RNA binding when an EMSA was performed using antigenomic RNA, suggesting that these internal hairpins contribute to VP30 binding. Moreover, deletion of the first 54 nucelotides reduced binding affinity to 0.39 μM. From this work the minimal genomic RNA substrate is ~40 nt long however optimal binding includes the upstream hairpin structure. VP30 exhibits high affinity binding to a ssRNA stretch linked to a hairpin structure (Figure 1.18). A proposed model of VP30:RNA binding suggests that the hairpin structure directs binding of RNA to a hexameric VP30 (trimer of dimers) as VP30 mutations preventing hexamerisation and abolishes RNA binding. It has been suggested that hexameric VP30 forms an RNA binding interface, a prerequisite for RNA binding (Biedenkopf et al. 2016; Schlereth et al. 2016).

A

1-154 nt ——————— $K_D = 0.35 \pm 0.07\ \mu M$

79-154 nt ——————— $K_D = 0.99\ \mu M$

55-154 nt ——————— $K_D = 0.39\ \mu M$

B   1-154 nt optimal substrate

C   55 -154 nt minimal substrate

Figure 1.18 Truncation Effects on EBOV VP30 RNA Binding

A- RNA sequence compared to the RNA truncations. Transcriptional start site is shown in red. Spacer element is shown in orange.
Replication promoter elements are shown in green (1-55 nt and 79-154 nt).
B- Features of the optimal RNA substrate for VP30 binding
C- Features of the minimal RNA substrate for VP30 binding
Adapted from Schlereth *et al.*, 2016

1.2.8.3 Other Filovirus Proteins

NP is the most abundant protein found in infected cells and NC (~3200 NP per virion) (Dziuba et $al.$, 2014), at 739 amino acids NP is the largest NP of the Mononegavirales. NP has a 450 amino acid hydrophobic N-terminal domain (NTD), and 150 amino aicd hydrophilic C-terminal domain (CTD) (Watanabe, Noda and Kawaoka, 2006). The NTD of NP is important for NP:NP self-assembly and formation of NP tubular structures (Watanabe, Noda and Kawaoka, 2006), transcription and replication. The N-terminal homo-oligomeric interaction has been mapped to residues 1-450, conserved amongst filoviruses.

Truncation mutants of NP all formed aggregates and oligomers of NP when in low NaCl concentrations. However, truncation of residues 25-457 (ΔNP_{NTD}) produced monomeric protein in 500 mM NaCl (Leung et $al.$, 2015). The ability of ΔNP_{NTD} to bind RNA was assessed via a dot blot assay. Here, ΔNP_{NTD} was shown to preferentially binds ssRNA (K_D = 620 nM) with lysines 160, 171 and 174 are essential in the RNA binding site.

The NTD of NP also contains two lobes N- and C- which are both mostly α-helical. These lobes clamp the positive RNA binding groove (Mühlberger et $al.$, 1999; Watanabe, Noda and Kawaoka, 2006; Leung et $al.$, 2015). A hydrophobic coiled-coil domain forms a pocket that is thought to accommodate VP35 (Leung et $al.$, 2015).

The C-terminal domain (CTD) 451-739 of NP has a low amino-acid conservation sequence among species with conservation occurring in the last 95 residues (641-739) (

Figure 1.19). This is unexpected, due to the CTD being required for multiple viral protein:protein interactions (Dziubańska et $al.$, 2014). The $CTD_{601-739}$ is required for interaction with VP40 and incorporation of the NC into virions, but is not required for NC formation, replication or transcription (Yang et $al.$, 1998; Noda et $al.$, 2007; Beniac et $al.$, 2012; Dziuba et $al.$, 2014).
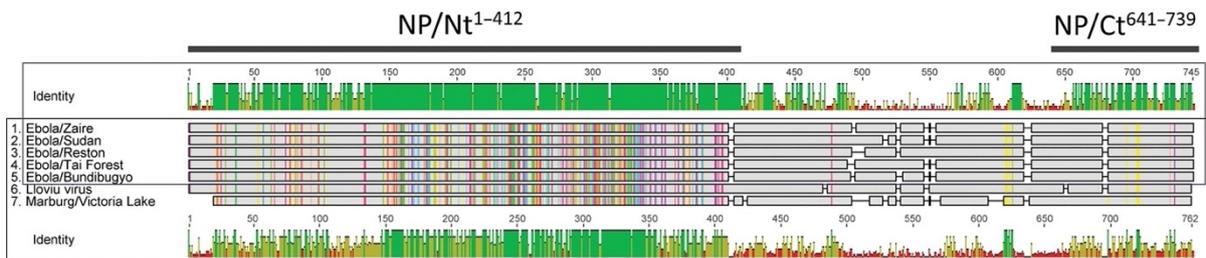
Figure 1.19 Graphical Representation of Amino Acid Conservation of Filoviridae Nucleoprotein

Top graph EBOV species, bottom graph MARV and LLOV species. 0-100% identity levels represented by red and green respectively. Identical residues are shown as coloured lines, running through the sequences. Adapted from Dziubańska *et al.*, 2014.

Phosphorylation of MARV NP has also been studied. MARV NP reportedly has seven phosphorylation sites within its CTD which have been suggested to modulate RNA binding (Lötfering *et al.*, 1999; DiCarlo *et al.*, 2011).

Computational analysis of LxxIxE motifs essential for binding of PP2A regulatory subunit B56, were found in NP proteins of all filoviruses in close proximity to the PPxPxY motif which was shown to be the VP30 binding site, suggesting a common conserved function. Interestingly, in EBOV and MARV the LxxIxE motif is N-terminal to the PPxPxY motif however in LLOV it is C-terminal, but all are separated by only 15-45 amino acids. PP2A regulates many cellular and signalling pathways and constitutes the most abundant source of cell phosphatase activity (Kruse *et al.*, 2018). B56 is the largest regulatory subunit subfamily from the B subunit family and comprises five human isoforms. Inhibition of the LxxIxE interaction suppress EBOV transcription and infection. NP co-localises with B56α in inclusion bodies, known sites of EBOV replication. The NP LxxIxE motif is required for PP2A-B56 mediated VP30 dephosphorylation. NP is thought to bridge the interaction between VP30 and B56 to allow dephosphorylation of VP30 to support replication.

The EBOV NP CTD X-ray crystal structure revealed alternating α-helices with β-sheets. On the other hand, NMR data of MARV NP suggests differential folding (Baker *et al.*, 2016). The N-terminal α-helical hairpin in EBOV, is unstructured in MARV. The second β-hairpin in EBOV is replaced by a short α-helix.

EBOV NP encapsidates RNA by a C-terminal α-helix which forms a long extended α-helix on the outside of NP clamping the RNA. This clamping-helix is positively charged and forms a stabilization ribbon along the NP helix by extending along the penultimate a-helix in the neighbouring NP subunits. The outer diameter is 41 nm and the inner hollow channel 16 nm (Bharat *et al.*, 2011, 2012). The model of RNA-NP (RNP complex) predicts that there are 13 nucelotides per NP monomer with most virions containing one copy of the genome (Beniac *et al.*, 2012). When additional proteins including VP24 and VP35 form complexes with NP the diameter of the NC increases to 50 nm. However, VP30 does not have this effect. It has been suggested that VP30 therefore binds on the interior of the NC and is not essential for NC formation. On the other hand VP35 and VP24 binding stabilises and protects the NC (Beniac *et al.*, 2012).

The matrix protein, VP40, coordinates virion assembly at the host cell plasma membrane (Madara *et al.*, 2015). VP40 exists in three forms; a dimer associates with the plasma membrane whereas hexamers form filaments making the viral matrix and support new virions (Ruigrok *et al.*, 2000; Bornholdt *et al.*, 2013a; Gerstman and Chapagain, 2017). Octameric rings also persist which bind RNA in negative stain EM (Timmins *et al.*, 2003).

Structurally VP40 contains two domains composed of β-folds. The two domains are connected by a flexible linker. Studies have shown both the CTD and N-terminal domain (NTD) are required for plasma membrane binding. VP40s NTD mediates oligomerisation and dimerisation. On the other hand, the CTD mediates membrane budding. When VP40 CTD interacts with the plasma membrane, conformational changes within the protein occur, VP40 forms a zig-zagged hexamer. This structure goes on to form the matrix coat in virions (Adu-Gyamfi *et al.*, 2014). Inhibition of the hexamerisation event disrupts the formation of virions and their subsequent release (Adu-Gyamfi et al. 2012).

Interestingly, VP40 can form VLPs that bud from host cell plasma membranes in the absence of the other proteins; emphasising that VP40 may be essential for viral budding; it is the hexameric form of VP40 that is required for budding (Bornholdt *et al.*, 2013; Johnson *et al.*, 2016). During association with the plasma membrane, VP40 inserts a CTD hydrophobic loop which allows lipid binding required for structural re-organisation of VP40.

Direct transcription of the GP gene produces sGP which is secreted and not included in virions. Mature GP is produced via an mRNA editing (stuttering) mechanism. During transcription of the GP gene an additional A is inserted into the nascent mRNA forming a stretch of eight adenosines at nucleotides 1019-1026. This causes a frame shift producing the membrane-anchored GP. Transcriptional editing of GP accounts for 20% of transcription (Mühlberger 2007). Guinea pigs infected with a recombinant EBOV, which contained mutations within the GP gene editing site to only allow the production of mature GP was shown to be less pathogenic to the guinea pigs *in vivo*. This was thought to be due to early death of host infected cells, limiting virion production, budding and spread and suggest why only 20% of GP gene mRNA transcripts encode mature GP (Volchkova et al. 2015).

Mature GP is formed from a heterodimer of GP_1 and GP_2 that then associate forming a trimer. GP is heavily glycosylated which protects from neutralising antibodies. On the GP_1 surface there is a N-glycan-containing cap region that is a heavily N- and O-glycosylated mucin-like-domain (Moller-Tank & Maury 2015). The GP crystal structure revealed that GP_2 forms the base of the protein, anchoring GP to the virion and GP_1 forms a cup. This arrangement makes GP a type-1 membrane protein. GP_2 also contains a fusion domain, responsible for the fusion of membranes during host-cell entry (Gregory *et al.*, 2011).

GP is primed by endosomal cathepsins for membrane fusion. GPs regular trimeric spike changes conformation (also dependent on low endosomal pH, pH 5.5). Conformational changes cause GP_2 to bend ~90° exposing an increased number of hydrophobic residues and GP_2, now with increased α-helices, inserts its fusion loop into the host-cell endosomal membrane. GP_2 folds in half bringing viral and host cell membranes together allowing the fusion event to occur (Gregory *et al.*, 2011).

ssGP also called Δ-peptide is produced by furin cleavage of sGP. Structural and biochemical analysis between sGP and ssGP suggest similar functions (Mehedi *et al.*, 2011).

Little is known about VP24. It has been suggested that VP24 is able to inhibit IFN-α and -β signalling however the exact mechanisms remains unclear. Reid et al., in 2006 suggested that VP24 blocks signal transducer and activator of transcription 1 (STAT1) nuclear localisation, thereby blocking STAT1 downstream events and ultimately the expression of IFN-α and -β (Reid *et al.*, 2006). VP24 binds to karyopherin-α (KPNα) preventing KPNα binding to STAT1 which normally would re-locate to the nucleus to switch on gene transcription of IFN-α and -β.

Han et al., progressively deleted N-terminal VP24 residues, this resulted in aggregation suggesting these residues are important for VP24 structure and stability. This work utilised western blots and SEC has suggested that VP24 forms tetramers in solution, however as VP24 is suggested to be a minor matrix protein, the formation of higher structures should not be ruled out. Han et al., also showed that VP24 interacts with plasma membranes suggesting a function in viral budding roles (Rahman Oany *et al.*, 2011).

The largest protein, L, is 2200 amino acids and 250 kDa. L is the catalytic component of the filoviruses and contains an RdRp, possessing methyltransferase, guanylytransferase, and synthetase activity. As well as replicating viral RNAs, L transcribes EBOV genomic RNA and completes mRNA capping and polyadenylation (Oany et al. 2011).

Sequence analysis revealed three conserved domains within the L gene: an RNA binding element at residues 554-571, an RNA template recognition and/or phosphodiester bond formation domain at residues 738-744 and an ATP/purine ribonucleotide triphosphate binding domain, residues 1815-1841. Moreover, the presence of cysteine residues are thought to stabilise important secondary structures for active site formation (Mühlberger *et al.*, 1992)

Trunschke et al., determined the binding domain for VP35 on L via deletion constructs. Their work mapped the binding domain to the first 380 N-terminal residues. Moreover, L homo-oligomerisation domain is also located within the N-terminus and overlaps with the VP35 binding domain (Trunschke *et al.*, 2013).

L interacts with the genome at a 3' polymerase binding site. Genome transcription occurs via a 'start-stop' mechanism at gene junctions which are recognised by L. Transcription always starts at the 3' end and due to L intermittently 'falling-off' the genome, genes at the 3' end of the genome are transcribed more abundantly than the 5' end. After transcription of the genome has occurred, L polyadenylates the mRNA transcript via the aforementioned stuttering mechanism, it is unknown whether L recognises the series of U's, if gene-ends signal to L, or interaction and/or signals with other viral proteins within the RNP complex cause this phenomenon to occur.

1.2.9 Life Cycle

1.2.9.1 Entry

EBOV virions are thought to bind to non-specific receptors on host cell surfaces, such as mannose-binding lectins and DC-SIGN, via the N- and O-linked glycans on GP_1. Virions being able to bind to an array of host cell surface receptors leads to enhanced infectivity (Khataby *et al.*, 2016). The exact mechanism of virion internalisation via these interactions remains unclear, however, GP also acts as a fusion protein. Virions are internalised by micropinocytosis into an endosome. During endosome acidification, cathepsin B and L cleave GP into GP_1 and GP_2. GP_2 receptor binding domain becomes exposed and binds to Niemann-Pick C1 internal receptor (NPC1), essential for cell membrane fusion. Fusion of virion and host cell membranes is mediated by GP_2 via a hydrophobic fusion loop that is normally buried beneath GP_1. These hydrophobic residues at the tip of GP_2 insert into host-cell endosome membranes. The GP_2 trimer unwinds and refolds into a six helix bundle forming a fusion pore via a series of conformational changes, which allows for NC release into the cytoplasm. The NC acts as a template for transcription. When sufficient viral proteins have been produced there is a switch to replication of full-length antigenome which is then replicated producing the full-length RNA negative sense genome (Figure 1.20). (Moller-Tank & Maury 2015; Gregory et al. 2014; Gregory et al. 2011; Weissenhorn et al. 1998).

GP

L

NP

\ -ve RNA Genome

\ +ve RNA Antigenome

VP40

VP24

VP30

VP35
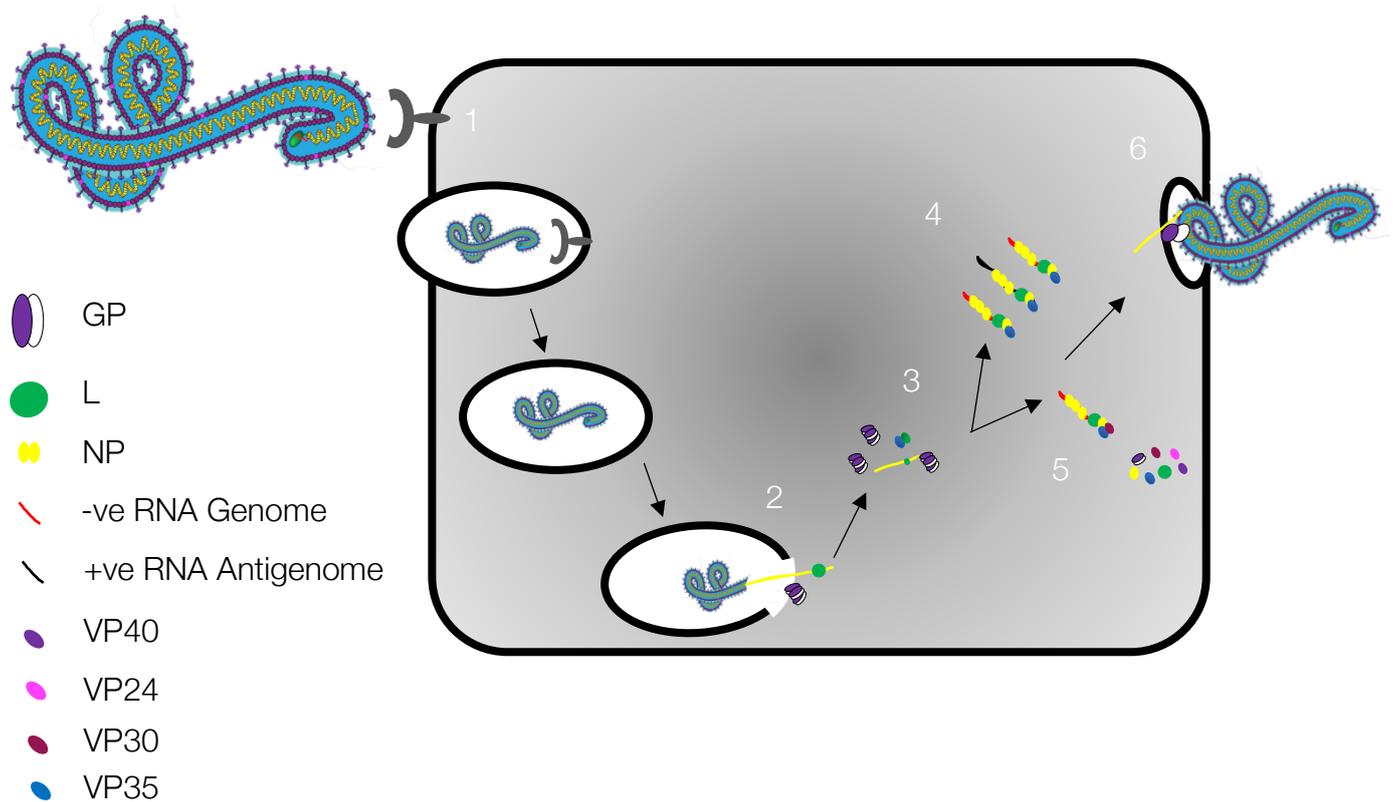
Figure 1.20 Schematic Representation of the Filoviral Life Cycle

1- Virions bind host-cell receptors and are internalised via micropinocytosis.
2- Endosomes acidification occurs and cathepsins cleave GP allowing fusion events occur. NC is released into the cytoplasm.
3- Primary transcription occurs.
4- Replication occurs from a positive sense antigenome.
5- Secondary transcription occurs.
6- NC formation occurs and viral budding via the host cell membrane.

1.2.9.2 Replication and Transcription

Little is currently known regarding the molecular mechanisms of EBOV infection transcription and regulation (Biedenkopf, Lier and Becker, 2016). Replication and secondary transcription occur when L switches to 'replication mode', with an increase of translated viral proteins. A transcriptional gradient is presented in infected cells, where messages from 3' genes are more abundant than those located towards the 5' end (Shabman *et al.*, 2013; Albariño *et al.*, 2018) because of polymerase fall off at the gene junctions.

Both EBOV and MARV have four RNP proteins in comparison to most other members of Mononegavirales which have three: NP, VP35, L and the extra protein VP30. NC proteins have dual functions; all are involved in viral morphogenesis as structural components and form a complex with viral RNAs catalysing RNA transcription (John *et al.*, 2007).

EBOV replication occurs in the cytoplasm of infected cells via the formation of inclusion bodies (IBs). IBs contain several viral RNA complexes consisting of viral RNA, NP, VP35,

L, and VP30. During replication, NCs are delivered to viral assembly sites where they are released as mature virions at the cell surface via budding. Budding occurs through an interaction with VP40 and endosomal sorting complex required for transport (ESCRT) pathway (Messaoudi, Amarasinghe and Basler, 2015).

In order to replicate the nsRNA genome a positive-sense antigenome is synthesised. The antigenome thereby acts as a template for EBOV replication and is also protected by NP.

A single promoter site is located within the 3' leader region and directs L to begin transcription. The release of the nascent mRNA is directed by gene-end sequences. At these gene ends the polymerase can potentially 'fall off'. To produce all seven transcripts, L needs to remain attached to the RNA genome at these gene junctions to then begin transcription again at the next promoter. Here, L is obliged to start at the 3' end and progress in a linear fashion. Transcription of the negative sense genome produces seven monocistronic mRNA species. These seven mRNAs are capped and polyadenylated by L. However, exactly how L does this is currently unknown.

Due to EBOV being a biosafety level 4 virus the replication has been studied utilising minigenome assays; the minimal protein requirements for replication of viral RNA include NP, VP35 and L and transcription must also include VP30. Minigenome analysis suggests that it's the non-phosphorylated form of VP30 and VP35 that is required for transcription initiation; VP30 delivered in trans supports transcription of the minigenome delivered by infectious virus-like particles (VLPs) (Sanchez *et al.*, 1993; Martínez *et al.*, 2008). Phosphorylation of VP30 leads to a weakened interaction with VP35 and/or viral RNA, excluding VP30 from the transcription complex. It is the transient phosphorylation of VP30 that is required for primary transcription (Biedenkopf, Lier and Becker, 2016).

1.2.9.3 Virion Assembly and Release

Transport of the NC is dependent on NP, VP35 and VP40 which form transport-competent NC-like structures. The current model for virion release is: NP self-assembles on RNA into helical tubes, VP35 and VP24 interact with NP forming NCs and VP40 mediates transport of NC-like particles via microtubules towards the cell surface. The NC-like particles are then incorporated into virions which bud from the cell surface membrane (Noda *et al.*, 2006; Wan *et al.*, 2017).

1.3 Introduction to hRSV

1.3.1 Discovery

Human orthopneumovirus (hRSV) (formerly human respiratory syncytial virus) is the leading cause of lower respiratory tract infections in the young, elderly and immunocompromised. Isolated from infants in 1956, hRSV caused a 'sponge-like' cytopathic effect in monkey kidney cells and was renamed human respiratory syncytial virus after being found to be indistinguishable from chimpanzee coryza agent virus (Chanock et al., 1957).

1.3.2 Classification

hRSV is a single stranded negative sense RNA virus. In the order of *Mononegavirales* which also includes deadly pathogens such as EBOV (*Filoviridae* family) and rabies virus (*Rhaboviridae* family).

The family of *Pneumovirade* comprises two genera; *Orthopneumovirus* and *Metapneumovirus*. These genera differ in their genome organisation. *Orthopneumoviruses* possess two non-structural proteins (NS1 and NS2). Both genera have two genotypes, A and B, that co-circulate and are distinguishable genetically and serologically (Gaunt *et al.*, 2011) variation typically occurs within the virion glycoproteins. GP varies 45% between the two genotypes (Meyer et al., 2008) hRSV-A is more common than -B, and for this reason has increased morbidity.

1.3.3 Impact, Epidemiology and Symptoms

Transmitted via respiratory droplets, hRSV is extremely contagious and can be fatal due to chronic obstructive pulmonary disease, resulting in over 250,000 deaths annually (Chung *et al.*, 2013; Kiss *et al.*, 2014). The highest mortality rates are seen in infants >6 months. Bronchiolar and alveola epithelium infections results in lower respiratory tract disease including bronchitis and pneumonia. Between 66,000-199,000 deaths are children under five, with 99% of this mortality occurring in developing countries (Nair *et al.*, 2011). Almost all children have been infected before the age of two and lifelong re-infection is common due to the ability of hRSV to evade the immune system through IgA B cell memory (Glezen et al., 1986) and incomplete/short-lived protective immunity. However, these values are likely to be underestimated as hRSV is often undetected in hospitalisation cases of pneumonia and lower respiratory tract infections. In the USA, hRSV mortality is estimates to be 11-17,000 per annum and 110-170,000 infections requiring hospital administration. Moreover, hRSV accounts for ~50% of all reported pneumonia cases and 90% of bronchiolitis in infancy (Domachowske and Rosenberg, 1999).

Early infection is associated with recurrent wheezing and asthma in later life, and increases susceptibility to bronchiolitis and pneumonia predominantly in the elderly and immunocompromised (El Omari *et al.*, 2011). This is due to an increase in interleukin (IL)-33 and downstream IL-4/5/13 that are implicated with asthma genesis (Saravia *et*

al., 2015). Moreover, high-levels of IL-33 are also found within hRSV site of infection in bronchi epithelial cells (Sigurs *et al.*, 2005). Typically, hRSV hospitalisation occurs during the winter months (December-January (northern hemisphere)).

Although mortality is low the economic burden is great. Despite many years of and ongoing research on hRSV there is no vaccine and/or antivirals. Immunoprophylaxis for hRSV treatment are extremely costly and incompletely protective, driving the need to develop antiviral drugs and vaccines. This can be done using structure-based drug design (SBDD) whereby obtaining high resolution structures via crystallography or electron microscopy can aid the synthesis of small molecule binders/inhibitors that stop viral progression and in turn disease progression.

1.3.4 Transmission

hRSV is transmitted in aerosol droplets contaminating nasal and conjunctival mucosa. The hRSV incubation period is typically four-five days, with viral replication beginning in the nasopharynx, triggering macrophages and the toll-like receptor 4 (TLR-4) response from fusion protein F recognition (Kurt-Jones *et al.*, 2000). Severe cases of hRSV infection occur in the lower respiratory tract, where replication is in ciliated epithelial cells and type I and II alveolar pneumocytes. hRSV spreads from the upper to the lower respiratory tract directly along the epithelium and through aspiration of nasopharyngeal secretions. Moreover, hRSV can spread cell-cell by inducing cell fusion and syncytia. This causes inflammation and necrosis (Domachowske and Rosenberg, 1999; de Waal *et al.*, 2018). Lower respiratory tract infection symptoms (such as wheezing) typically appear 1-3 days post rhinorrhoea.

hRSV-neutralising antibodies are present in new-borns via transplacental transfer of maternal antibodies both at similar titres. Infection occurs due to the natural decline of neutralising antibodies in the first few months of life.

1.3.5 Treatment

High prevalence and continued research efforts over the last ~60 years have not led to a suitable treatment for hRSV infection even though live attenuated and inactivated vaccine for bovine RSV (bRSV) have been available for 25 years. Vaccine development became dramatically limited after 2 infants died during a clinical trials in 1966 with a formalin-inactivated hRSV which developed higher titres of non-neutralising antibodies against formalin altered F and GP surface glycoproteins and caused an altered T-helper lymphocyte response (Murphy et al., 1986)

Nebulized ribavirin (a broad-spectrum anti-viral, nucleoside analogue) is the only FDA approved drug for hRSV infection. Ribavirin is unfavourable as it is not cost-effective and possibly toxic. Administration usually occurs overnight for up to 18 hours per day; often in a negative-pressure room. Side effects include coughing, nasal congestion, dyspensia haemolytic anaemia and gastrointestinal effects (Ariza-Heredia et al., 2012; DeVincenzo et al., 2014; Ebbert & Limper, 2005; Li et al., 2012; Shah & Chemaly, 2011).

Palivizumab (trading name Synagis, MedImmune) is a humanised mouse monoclonal neutralising antibody (mAb) against hRSV fusion protein F (Figure 1.35). Palivizumab provides short-term protection for high-risk patients. However, palivizumab is expensive and only reduces hospital administrations by 60% (DeVincenzo *et al.*, 2014). The total cost per infant receiving palivizumab treatment course (5 monthly intramuscular injections) is over £5000 and is only readily available to higher-income countries. In 2010, motavizumab (AstraZeneca) a second-generation mAb (based on palivizumab) was discontinued in phase III clinical trials after the FDA antiviral drugs advisory committee declined an endorsement of motavizumab licencing.

hRSV vaccination strategies are largely focused on the induction of the humoral immune response against F, however efforts for a successful vaccine are complicated by F's conformational diversity. Most human neutralizing antibodies recognise pre-fusion F providing insights for vaccine design. F-specific neutralizing antibodies bind before virus-host membrane fusion (Figure 1.35). Corti et al. isolated 30 hRSV neutralising antibodies, 85% (26/30) of which bound exclusively to pre-fusion F, one of these included MPE8. MPE8 a human monoclonal antibody, potently cross-neutralised hRSV, human metapneumovirus (hMPV) (33% F protein sequence identity) and also animal viruses; bRSV and pneumonia virus of mice (PVM) *in vivo*. Compared to palivizumab, MPE8 is 8-fold more potent in neutralizing hRSV with a 50% spreading inhibitory capacity. MPE8 was shown to the neutralises the four paramyxoviruses and prevented viral spreading via inhibition of syncytia formation. MPE8's high potency could be developed as a prophylaxis for new-borns and immunocompromised (Corti *et al.*, 2013).

Post-fusion hRSV F is also a current vaccine candidate. Post-fusion RSV F elicits high anti-hRSV antibody titres and protection in animal models thus giving promising results for human trials furthermore denoted from the crystal structure of post-fusion F key neutralising binding sites providing the basis for eliciting high neutralising antibody titres (Swanson *et al.*, 2011). Recent nanoparticle vaccines with the antigenic site II (recognised by palivizumab) (NOVAX E202) was trialled in women and deemed safe and immunogenic. Phase 2 and 3 clinical trials were also successful and development is towards using E202 for maternal immunisation to protect infants (Glenn *et al.*, 2016).

It is important to recognise that although F is the predominantly encountered antigen, that other vaccines could be designed utilising internal proteins such as NP and M, which if delivered by recombinant vectors may improve vaccine-induced T-cell mediated immunity (Anderson *et al.*, 2013). ALN-RSV01 (Alnylam Pharmaceuticals) Phase IIb clinical trials have given positive results for the treatment of severe hRSV infection causing progressive bronchiolitis obliterans syndrome (BOS). Within the first five days of symptoms ALN-RSV01 had a strong effect in 80% of patients, and at 180 days post infection ALN-RSV01 showed an eight-fold decrease in the risk of BOS development.

More recently, vaccine development has focused on live attenuated vaccines with deletions of NS, SH and M2-2 genes. The intranasal live attenuated virus termed hRSV

MEDI ΔM2-2 lacks the M2-2 gene and has been trialled in seronegative children. The vaccine produced higher titres of neutralising antibodies (Karron *et al.*, 2015).

Alternatively, suboptimal nucleotide substitution in nine of 11 hRSV ORFs resulted in a vaccine candidate with improved attenuation and higher immunogenicity (Le Nouën *et al.*, 2017).

Lastly, the crystal structure of the NP N-terminal domain (NTD) in complex with residues from the P C-terminal domain (CTD) revealed that phenylalanine 241 of P is buried deep within a previously identified NP pocket, acting as an anchor for P with this interaction stabilised via electrostatic interactions of aspartic acid 240 of P. These two interactions represent an ideal druggable pocket on the NP, as it is easily accessible and conserved within *Pneumovirinae*. Compound M76 (1-(2,4-dichlorobenzyl)-1H-pyrazole-3, 5-dicarboxylate) bound to the same site of the NP as P but did not exert inhibitory activity. The M76 prodrug (M76-diAM, via the addition of acylal groups on both carboxylates, deemed M76 pH sensitive, allowed for cell internalisation) caused a replication decrease in a dose-dependent manner (IC_{50} 122 +/-7 µM) however, caused high cell toxicity (CC_{50} 226 +/- 7 µM). M76-diAM represents a prodrug that showed viral inhibition *in cellulo* but still needs improving due to its toxicity and release of formaldehyde upon hydrolysis (Ouizougun-Oubari *et al.*, 2015).

1.3.6 Virion Morphology

hRSV is an enveloped virion whose morphology appears to be in both spherical and filamentous forms by cryo-EM (100 nm-1 µM and 200 nm – 2 µM; width 70 – 190 nm respectively), the relevance of the two morphologies is not understood; however, filamentous virions exhibit reduced infectivity (Liljeroos *et al.*, 2013). These forms may represent an artefact of the virus preparation. The host-derived lipid membrane is studded with an irregular array of F, GP and SH surface glycoproteins. M is associated with the internal membrane. The core of the virion is formed by the left-handed RNP complex (NP encapsidated RNA and L, P and M2-1) (Figure 1.21). Secondary density (green arrows Figure 1.21 D) was thought to be M2-1. This spacing is supported by Kiss et al., who also described the regularly spaced density between the M and the RNP to be M2-1 in their tomograms. Zernike phase contrast cryo-electron tomography (ZPC-cyroEM, is an emerging technique that produces higher image contrasting than conventional cryoEM) revealed small densities between the layer of M and the RNP to be ~12 nm from the membrane and 12.6 nm between each other (+/- 2.1 nm). Laser scanning confocal microscopy showed M2-1 to be present in the cytoplasm but also in viral filaments with M, supporting the hypothesis that M2-1 is part of the structure of a mature hRSV virion (Kiss *et al.*, 2014)
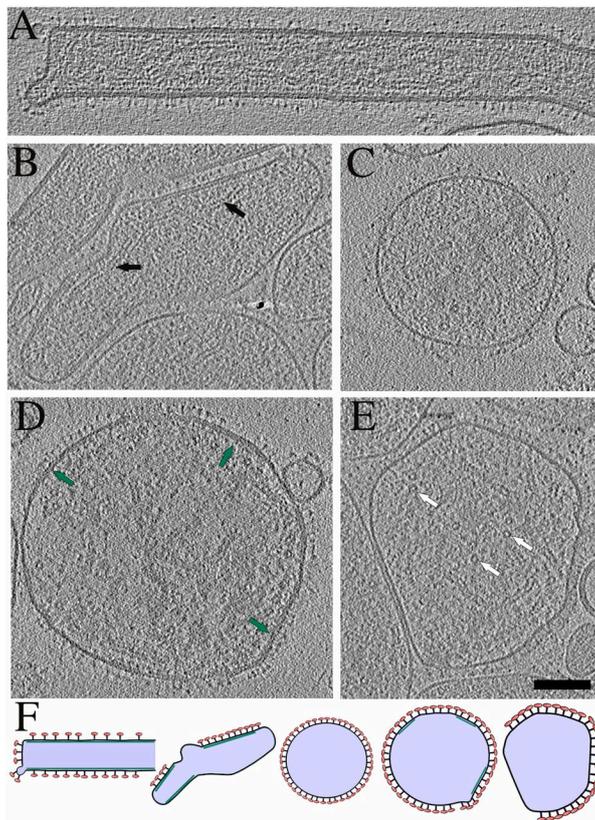
Figure 1.21 hRSV Virion Morphology

A-E- Tomography of hRSV virions (A2 strain), obtained using 3.8 nm thick tomograph slices. A- filamentous virion.
B and D- intermediate virion morphology present both filamentous and spherical features respectively.
C- spherical virion.
E- deformed spherical particle when in close proximity to neighbouring particles, glycoprotein spike free.
White and black arrows: RNP's. Green arrows: secondary density assumed to be M2-1.
F- schematic representation of tomographs.
Scale bar 100 nm.
Adapted from Liljeroos *et al.*, 2013.

1.3.7 Genome

hRSV single stranded non-segmented RNA genome is 15.2 kb, encoding nine structural and two non-structural proteins from ten genes. Of these eleven proteins, six are essential to the virus life cycle, representing potential anti-viral targets. The GP gene also encodes sGP protein that functions in immune evasion similar to that of sGP of the *Filoviridae*.

The leader region (le), found at the 3' end is well conserved and is 44 nucleotides in length. The first 11 nucleotides are important for polymerase recruitment and RNA synthesis initiation. Each gene encodes the respective ORF flanked by gene start (GS) and gene end (GE) sequences separated by intergenomic regions. The M2 gene has two ORFs producing M2-1 and M2-2 and is discussed in detail in 1.3.8.1 and 1.3.8.4. The first 34 nucleotides of the 5' ends trailer sequence (tr) are required for antigenome synthesis and RNA encapsidation (Fearns et al., 2002) (Figure 1.22)
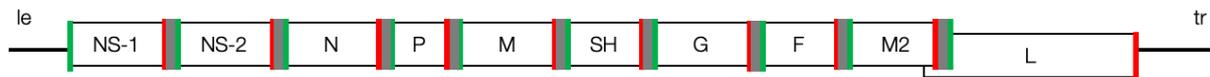
Figure 1.22 hRSV Genome

The 10 genes of hRSV produce 11 proteins. The gene start (GS) are shown in green and gene end (GE) in red separated by intergenomic regions shown in grey.

1.3.8 Proteins

1.3.8.1 M2-1

M2-1 is an anti-termination transcription factor that allows for the full-transcription of viable viral mRNAs and is essential for the rescue of infectious virions. M2-1 appears to enhance transcriptional recognition of GS/GE sequences; In the absence of M2-1 mostly short non-functional mRNAs are produced, or GS/GE elements are ignored to allow for viral RNA replication.

Mini-genome studies supplied with plasmids encoding M2-1, P, N and L showed increased viral mRNA transcripts but no effect on replication. Moreover, primer extension showed mRNAs produced in the absence of M2-1 were mostly truncated and rarely possessed intact 3' end sequences. The RdRp was capable of transcribing NS-1 and NS-2 only in the absence of M2-1 as these proteins are only a few hundred nucleotides in length.

M2-1's known binding partners include P and RNA which interact directly and in a competitive manner, however it is unclear if this interaction is mutually exclusive or occurs simultaneously on M2-1 tetramers. Our previously published model (Figure 1.23) suggest the latter is occurring due the M2-1:P protein interaction being tetrameric. 1 or more monomers of M2-1 and P protein can interact simultaneously while 1 or more monomers of M2-1 can interact with the nascent mRNA chain.
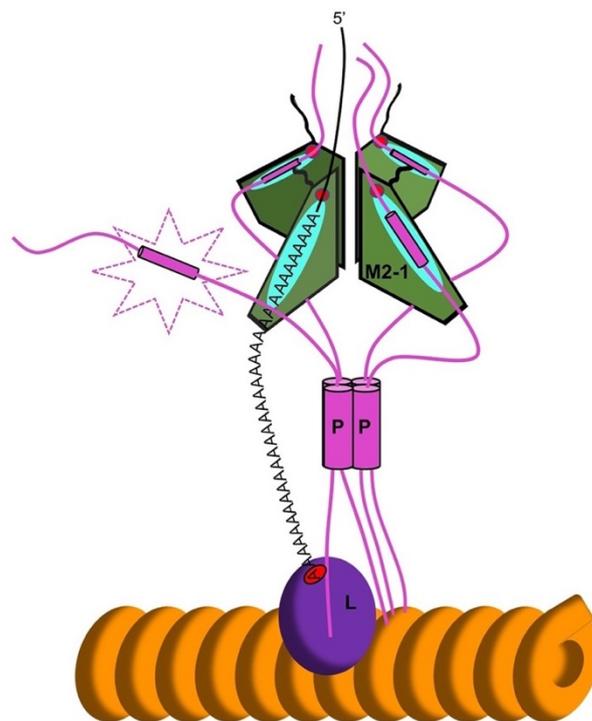
Figure 1.23 Transcriptional RNP Complex Schematic for hRSV

P (pink) interacts with L (purple) that is transcribing RNA that is encapsidated by NP (orange). A-rich mRNA displaces P 90-110 peptide on M2-1 (green). M2-1 is recruited to the complex by P. Adapted from Selvaraj et al., 2018.

In 2014 Tanner *et al*., (University of Leeds) solved the full-length crystal structure of M2-1 in its stable tetrameric form to 2.5 Å (Figure 1.24). The M2-1 (22 kDa) crystal structure reveals a novel fold comprising nine α-helices. M2-1 protein monomer contacts extend to all other monomers in the tetramer. The last 20 residues of M2-1 are not resolved in the crystal structure and are assumed to be unstructured. Deletions of these residues allowed infectious hRSV rescue and these are therefore thought to be dispensable, although their evolutionary maintenance implies a hidden function (Tanner et al., 2014). The structure supported previous cross linking and negative stain EM data of M2-1's physiological oligomeric state (Blondot et al., 2012).

Protomers of M2-1 form three distinct regions; an N-terminal Cys3-His (zinc-binding domain (ZBD)) (Figure 1.24), an oligomerisation domain and a core domain which are discussed below.

The ZBD (residues 7-25), is responsible for the incorporation of a structural zinc; mutations of conserved residues within this region result in loss of M2-1's anti-termination activity and NP binding (Hardy and Wertz, 2000). Moreover, the ZBD stabilises the M2-1 tetramer by interacting with an adjacent promoter (Tanner et al. 2014). The ZBD of M2-1 lies on the N-terminal face in close proximity to the RNA binding domain. A Cys_3-His_1 motif is also present in the structurally related EBOV VP30 protein (discussed in section 1.2.8.2) where mutation of residues for zinc coordination result in reduced transcription (Modrof, Becker and Mühlberger, 2003).

The oligomerisation domain, also known as the tetramerisation helix comprises a single α-helix (residues 32-50). This central helix that controls oligomerisation buries four hydrophobic residues (leucine 36 and 43, isoleucine 46 and methionine 50) on one helix face within a four-helix bundle interacting with the other M2-1 monomers. Linking the oligomerisation domain to the core domain, is a highly flexible linker (residues 52-67) which also encompasses phosphorylation sites on serine 58 and 6; this region was poorly resolved in the crystal structure due to the flexibility (Tanner *et al.*, 2014).

The core domain (residues 69-172), is largely globular (residues 75-171), comprising six α-helices with a ligand binding groove that can bind either P protein or RNA. Helices one and two and five and six form a helix bundle, while helix three and four form a hairpin stacked upon helix six. Analysis of the core domain revealed serine 58, lysine 92 and 150 and arginine 151 and 159 as essential modulators of the M2-1 protein antitermination function. However, viral RNA binding is also influenced by arginine three and four which are not found in the core domain, but are still critical for the antitermination function; suggesting that the RNA binding surface extends beyond that of P (Tanner *et al.*, 2014).

Residues 174-194, presumed to be disordered, were not resolved in the crystal structure of M2-1. Deletion mutants (Δ127, Δ148 and Δ177) were created to assess the requirement of the CTD to M2-1's function by attempting to rescue infectious virions from cDNA. Although infectious virus was recovered, a reduction in replication efficiency was seen *in vivo* and vitro for M2-1Δ177 when the extreme CTD 17 nucleotides had been deleted. Moreover, M2-1Δ177 showed a 50% reduction in mRNA synthesis of the β-galactosidase reporter gene compared to WT M2-1 when

M2-1 processivity was measured. M2-1Δ127 and -Δ148 abolished synthesis and did not allow for infectious virus recovery (Tang *et al.*, 2001). Northern blot analysis of viral RNA expression showed that M2-1Δ177 did not abolish gene junction read-throughs during hRSV replication and supported viral RNA transcription (Tang *et al.*, 2001).

Figure 1.24 Structure of hRSV M2-1
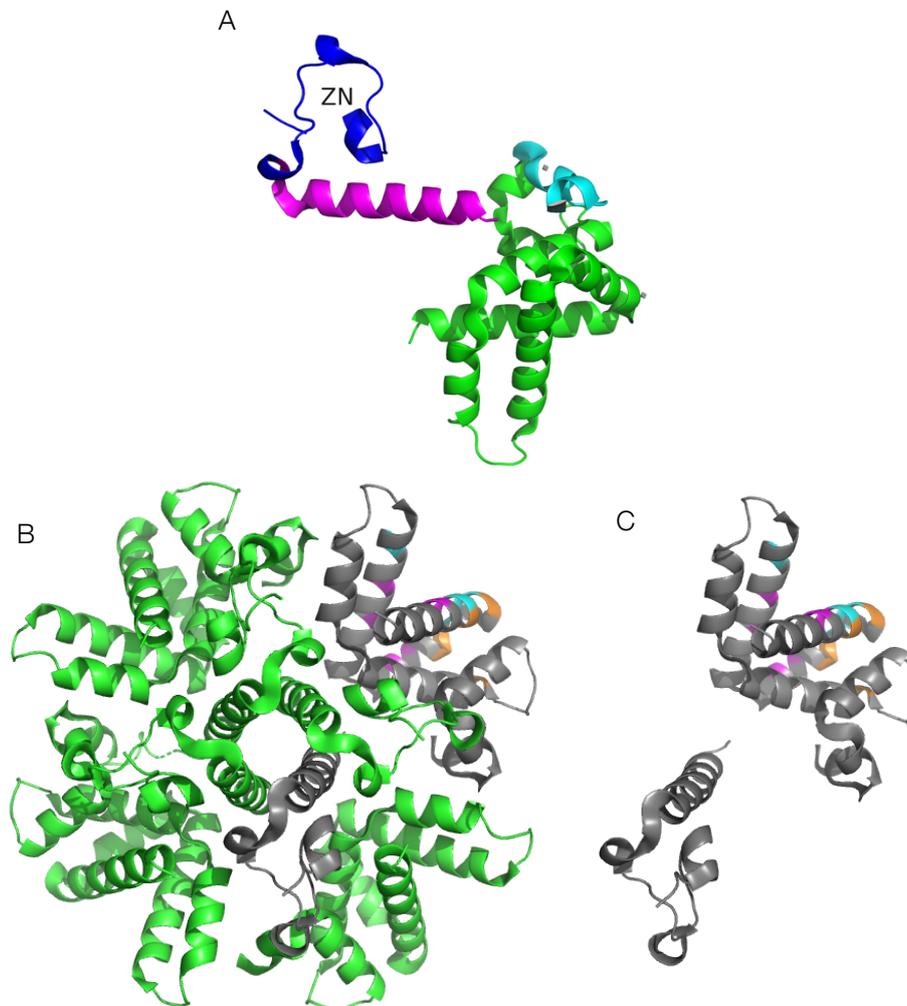
A- M2-1 monomer. M2-1 has 3 distinct domains: ZBD (blue), tetramerization helix (magenta), core domain (green). The tetramerization helix and core domain are linked via a flexible linker (cyan).
B- M2-1 tetramer with 3 protomers (green) and a monomer in grey with residues for RNA and P protein binding coloured (see C). NTD and CTD labelled for grey monomer.
C- M2-1 monomer with residues important for RNA binding (orange), P binding (magenta) or both (pale cyan).
Figure made in PyMol (version 2.3.2),
PDB: 4C3B.

Phosphorylation of M2-1 was first observed in 2000 with different migrating patterns on an SDS-PAGE gels and is thought to be phosphorylated by a host cell CK1 kinase (serine/threonine protein kinase) (Hardy and Wertz, 2000; Cartee and Wertz, 2001).

In mini-genome studies phosphoablatant (S58A and S61A) and phosphomimetic (S58D and S61D) mutants decreased M2-1s antitermination function but both phosphorylated and non-phosphorylated forms were detected in hRSV infected cells suggesting that the activity of M2-1 is dynamically regulated by phosphorylation/dephosphorylation, potentially for optimal switching of function (Zhou, Cheng and Jin, 2003; Tanner *et al.*, 2014). Perhaps phosphorylation regulates switching between an anti-terminator in transcription vs non-recruitment to the polymerase complex in replication?

The surface of M2-1 is abundant in positively charged residues forming four positive regions. Initial RNA binding studies showed M2-1 bound to long RNAs with no sequence specificity (700-1300 nucelotides). However, sequences >80 nucleotides exerted sequence specificity with highest affinities shown for anti-genomic leader sequences (Cuesta *et al.*, 2000). More recently, binding to viral mRNAs that are A-rich has also been shown (Cartee and Wertz, 2001). This was supported by Blondot *et al.*; M2-1 bound to hRSV-specific genomic and anti-genomic RNA but also showed that M2-1's core domain (residues 58-177) had a preference for A-rich sequences when testing short RNAs (10-15 nucleotides). Fluorescence anisotropy experiments further supported Blondot's findings. Higher affinity binding to A-rich regions (pA binding K_D = 19.1 nM) is posited to be important for a possible M2-1 function in polyadenylation (Blondot *et al.*, 2012; Tanner *et al.*, 2014).

Analysis of M2-1's RNA binding ability was initially performed using M2-1 mutants in serine 58 and 61, as well as residues within the ZBD (Cuesta *et al.*, 2000). However, Tran et al., defined M2-1's RNA binding domain to be central to the core domain residues 59-153 (Tran *et al.*, 2009). NMR perturbation experiments showed that core residues arginine 126, valine 127, asparagine 129, threonine 130, serine 133, leucine 152 and valine 156 were also critical for RNA binding (Blondot *et al.*, 2012). M2-1's core residues serine 58 and 61, lysine 92 and 150, arginine 151 and lysine159 were mutated to either alanine or aspartic acid and showed a reduction in RNA binding ability and a reduction in activity within the mini-genome system. This highlights that M2-1 RNA binding activity is crucial for its function during infection. Moreover arginine 3 and 4 within the N-terminal arm of M2-1 were also mutated to alanine and this also resulted in loss of RNA binding (Tanner *et al.*, 2014). This work supported previous findings that the M2-1 core domain (residues 58-177) exerted weakened RNA binding affinity compared to full-length. The M2-1 core domain NMR structure is similar to full-length M2-1, showing 2 oppositely charged faces with the positively charged face proposed for RNA and P binding (Blondot *et al.*, 2012; Tanner *et al.*, 2014).

The recent crystal structure of hRSV M2-1 in complex with SH7 RNA (positive-sense GE of the SH gene) at 2.7 Å revealed previously unidentified residues involved in RNA binding. Base stacking interactions occur with F28 and hydrogen bonds interactions with C7 K8 and F9 between nucleotide A6, which flips out from the main fold of RNA. F28 also forms hydrogen bonds with the 2'OH of the RNA ribose. Gao *et al.*,

suggested that the ZBD of M2-1 may recognises the nucleobase of RNA with these interactions favouring, but not limited to A (Gao *et al.*, 2020).

M2-1 in complex with SH7 crystallised as a tetramer, and aligned well with previously published structures for apo M2-1 (Tanner *et al.*, 2014) and M2-1 in complex with P_{90-110} (Selvaraj *et al.*, 2018), root-mean-square deviation (RMSD) 0.7 and 0.9 Å respectively, suggesting no significant conformational changes occur upon RNA binding.

Tran et al., initially believed that RNA/P compete for the same binding site on M2-1, but this was before the characterisation of tetrameric M2-1 (with four binding sites per tetramer). NMR studies then showed binding surfaces for RNA and P which partially overlap. Loss of the M2-1:P interaction decreased transcription and prevented M2-1 localisation to IBs; therefore it is thought that P interactions enable transport of M2-1 to provide RNA binding functions in IBAGs (Blondot *et al.*, 2012; Rincheval *et al.*, 2017).

α-helices 4-6 of M2-1 were shown to interact with P with contacts at serine 133, tyrosine 134. valine 165, lysine 159, asparagine 163 and lysine 162 (Figure 1.25). P 100-120 were thought to interact with M2-1 with residues 101, 102 and 109 being important when subjected to affinity chromatography (Mason et al., 2003). Deletions of residues 100-120 showed 6% binding but did not affect P:NP binding. Deletions of residues 120-140 and 140-160 decreased binding to M2-1 and NP, however these residues compromise P's coiled-coil domain and loss of binding may be due to disruption of the natural tetrameric oligomeric state of P. The M2-1 binding region of P (residues 100-120) was further analysed by double and single alanine mutagenesis. Single mutations of leucine 101, tyrosine 102 and phenylalanine 109 resulted in decreased luciferase output in an M2-1 dependent luciferase mini-genome however in an M2-1 independent chloramphenicol acetyltransferase system reduced output was not seen (Mason et al., 2003).

The recent crystal structure of M2-1:P_{90-110} elucidates the orientation of M2-1 and P monomers (Figure 1.26). P_{90-110} is a single α-helix that lies along an M2-1 surface cleft formed by α-helices 7, 8 and 9 of the M2-1 globular core domain. P's N-terminus faces the N-terminal surface of M2-1. Amino acid side chains within the M2-1:P interface drive this complex formation. P_{90-110} hydrophobic reissued proline 97, phenylalanine 98, leucine 101 and isoleucine 106 face towards this M2-1 cleft. P binds to previously identified M2-1:P binding residues that overlap with M2-1's RNA binding domain via ionic, hydrophobic and hydrogen bond interactions with no apparent changes to the M2-1 backbone or side changes apart from arginine 126 and tyrosine 134 which move to allow for hydrogen bonds. 15 out of the 16 M2-1 residues that interact with P_{90-110} in the crystal structure are conserved amongst Orthopneumoviruses (Selvaraj et al., 2018).

This crystal structure allows the identification of the orientation of M2-1's RNA binding surface in relation to the RdRp active site and RNA exit channel as the M2-1:P/RNA binding site overlap. This is supported by the hMPV M2-1 crystal structure. This model therefore suggests that M2-1 is positioned to interact with the nascent mRNA chain (Figure 1.27). The higher binding affinities seen through fluorescence anisotropy

suggest that M2-1 favours RNA binding over P and RNA binding displaces P from their shared binding site. It has been suggested that only one monomer of P is displaced during this mechanism of action whilst the RdRp transcribes one RNA molecule therefore the stoichiometry is 4 M2-1:3 P:1 RdRp:1 mRNA (Selvaraj *et al.*, 2018).

Mini-genome analysis revealed that M2-1 mutants R126E and L148A decreased GFP expression by 30% compared to WT M2-1. Arginine 126 forms electrostatic interactions with P residues 104 and 107 (glutamic acids) and this decrease is likely due to the loss of these interactions. L148A decrease is likely due to hydrophobic interactions with leucine 101 from P_{90-110}. P mutants F98A, Y102A and also decreased mini-genome activity as these residues mediate M2-1: P_{90-110} interactions (Selvaraj *et al.*, 2018).

Co-immunoprecipitation assays showed that M2-1 interacts with the NTD of M and this co-localisation occurs in the absence of other viral proteins. M2-1 is thought to mediate in the interaction between M and RNP complexes for inclusion into IBs (D. Li *et al.*, 2008). In a transcription reaction using purified RNPs, when M was blocked using an M-specific antibody viral transcription was increased. This suggests that M can inhibit transcription and upregulate/initiate viral assembly (Meanger *et al.*, 2002).
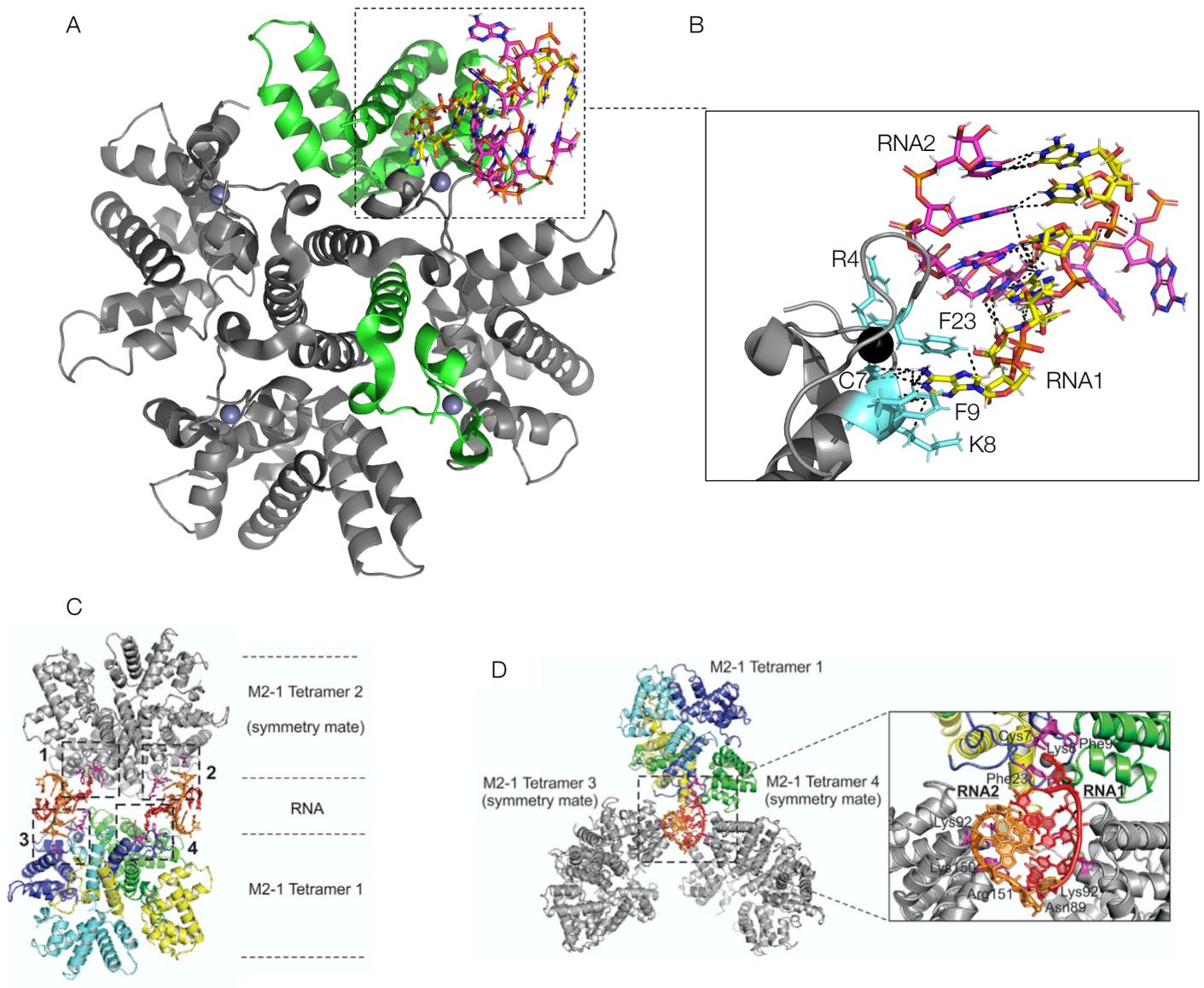
Figure 1.25 Crystal Structure of M2-1 in Complex with SH7 RNA

A- M2-1 crystallised as a tetramer with monomer shown in green. RNA shown in yellow (RNA1) and pink (RNA2).
B- M2-1 interacting residues shown in cyan. A6 is flipped out to interact with the ZBD (black sphere). Interactions shown by black dashes.
C- Symmetry mate of M2-1 shown in grey. M2-1 monomers coloured.
Adapted from Gao *et al*., 2020.
Figure A-B made in PyMol (version 2.3.2).
PDB: 6PZQ.
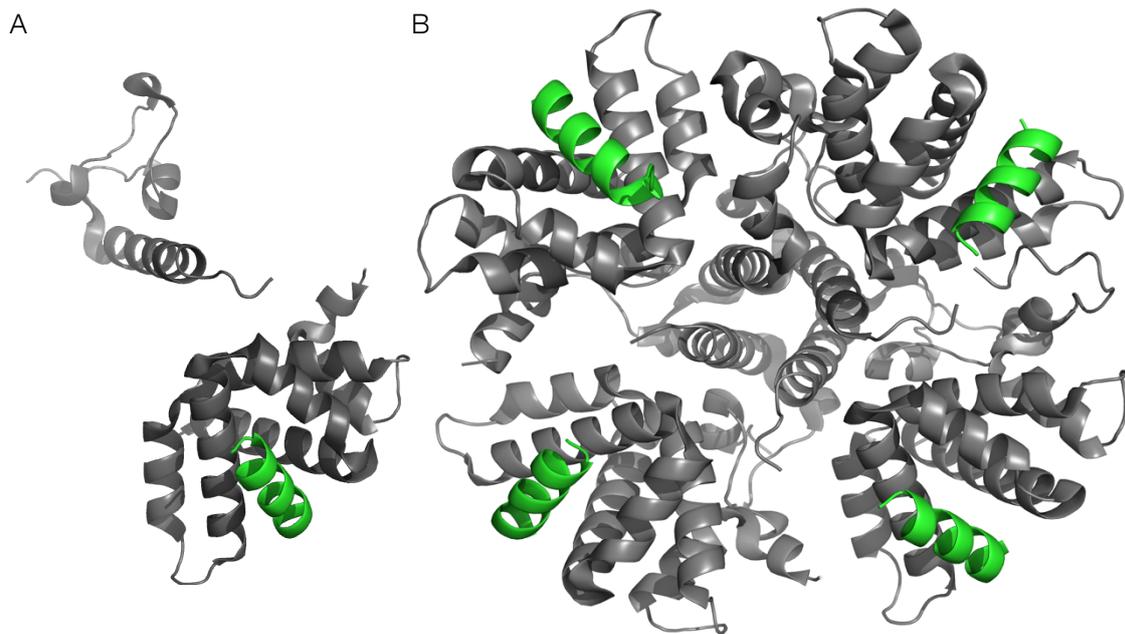
Figure 1.26 Structure of hRSV M2-1 and P$_{90-110}$

A- M2-1 monomer in complex with P 90-110 (green).
B- M2-1 tetramer (grey) in complex with P 90-110 (green).
Adapted from Selvaraj *et al*., 2018.
Figure made in PyMol (version 2.3.2).
PDB: 6G0Y.

M2-1 is also thought to have a structural role within mature virions. Electron cryo-tomography data shows an internal layer of electron density (discovered to be M2-1 by laser stacking confocal microscopy) under M; linked to the inner leaflet of the viral envelope adjacent to the RNP (Liljeroos *et al.*, 2013; Kiss *et al.*, 2014). M2-1 crystal structure does not support M2-1 forming a 2D layer within the virion. It has however, been suggested that without the incorporation of zinc, M2-1 is found monomeric and it is monomeric M2-1 that plays a structural role. The hMPV M2-1 crystal structure was also tetrameric but core domains could 'flip out' (Figure 1.27) forming a 2D layer (Leyrat *et al.*, 2014).
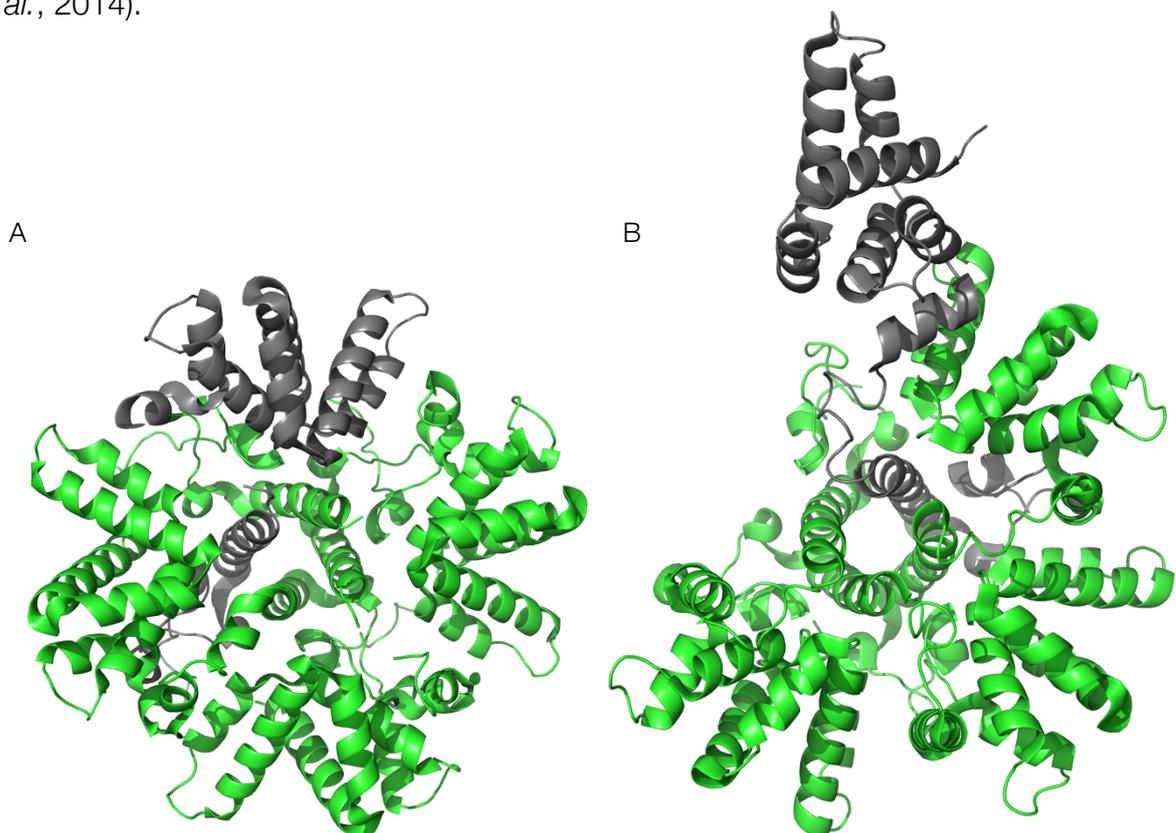
Figure 1.27 Comparison of hRSV and hMPV M2-1 Structures

A- tetrameric hRSV M2-1 (green), monomer highlighted in grey.
B- tetrameric hMPV M2-1 (green), 'flipped out' monomer highlighted in grey hypothesised to bind M. P and/or RNA binding results in the 'closed' M2-1 conformation (A).
Figure made in PyMol (version 2.3.2).
PDB: 4C3B (hRSV) and 4CS7 (hMPV).

hMPV is the most closely related *Orthopneumovirus* to hRSV. Also encoding an M2-1 protein that is 38% identical and 64% similar to hRSV M2-1. Unlike hRSV, the M2 gene does not encode a second overlapping ORF and hMPV M2-1 is not essential to viral transcription as it can be deleted, and infectious virus can still be rescued. The structure of hMPV M2-1 was also tetrameric with the core of one protomer 'flipped-out' by 60 Å, forming an 'open' conformation whereby the core domain as dissociated away from the tetramer (Figure 1.27). It was also bound in complex with an AMP molecule which hMPV M2-1 favours. AMP interacts with lysine eight and phenylalanine 23. Hydrogen bonds occurred to lysine eight backbone and proline six carbonyl oxygen (Figure 1.28 B) or the sulphur atom of cystine seven (Figure 1.28 C) (dependent on bond orientation). M2-1 crystal soaking with DNA or RNA lead to a decrease in diffraction apart from soaking; apart from DNA with a sequence of AGTTA. AG nucleotides were clearly resolved in the structure between the exposed zinc finger and the core domain between symmetry related M2-1 molecules. The adenosine nucleotide interacts with lysine 22 side chain via its phosphate and hydrogen bonds with alanine five backbone nitrogen. Guanine nucleotide did not share atomic contacts. TT nucleotides interacted at the same binding pocket as AMP however, the second T was partially disordered but showed hydrogen bonds to the M2-1 backbone (Figure 1.28 D). These structures show that the binding of nucleotides to the core domain is stabilised between phosphates and positively charged residues including lysine 91 and arginine 149 (Leyrat *et al.*, 2014).

SAXS analysis and molecular dynamic simulations showed a dynamic structure with nucleotide binding favouring the closed symmetric tetramer like hRSV M2-1. Residues that bind RNA and P are conserved between both viruses. A model of hMPV M2-1:RNA binding showed 13 nucleotides per monomer which coincides with the consensus of hRSV GE sequences and is similar to the hRSV transcriptional RNP complex model. (Leyrat *et al.*, 2011, 2014; Blondot *et al.*, 2012; Tanner *et al.*, 2014).
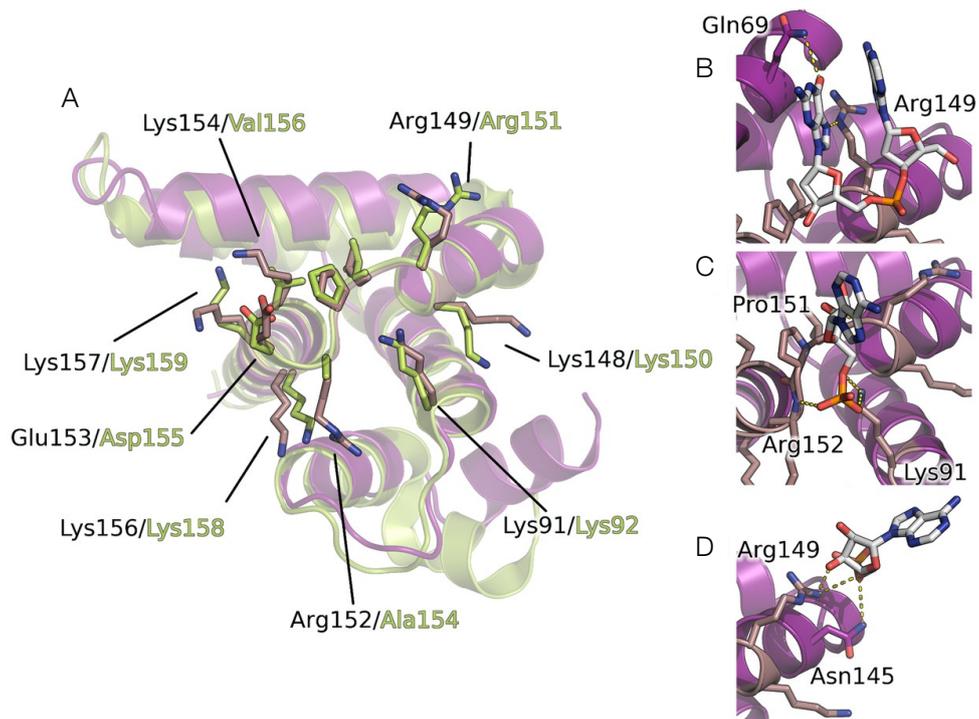
Figure 1.28 hMPV M2-1 Core Domain Interactions

A- superimposition of hRSV (green) and hMPV (purple) M2-1 core domains.
B- hMPV M2-1 core domain abound to DNA nucleotides AG.
C/D- AMP bound to M2-1 core domain (purple). Previously identified NMR RNA binding
residues side chains shown in brown.
Adapted from Leyrat *et al*., 2014.
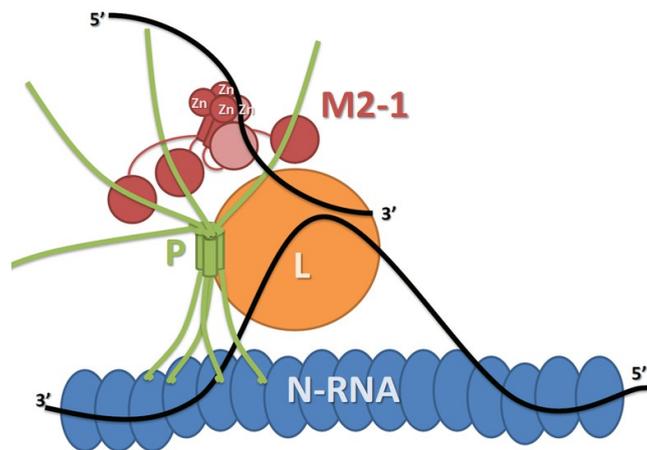
Figure 1.29 Transcriptional RNP Complex Schematic for hMPV

The NC (blue) associates with RNA (black) and tetrameric P protein (green). P binds L (orange) and tetrameric M2-1 (red). M2-1 undergoes confirmation changes upon RNA binding and recognition of A-rich RNA sequences (gene ends and/or viral mRNA pA tails) highlighted by M2-1 monomer (pale red). Adapted from Leyrat *et al.*, 2011.

EBOV VP30 (discussed in detail 1.2.8.2) is a structural homologue of hRSV M2-1. VP30 is regulated by dynamic phosphorylation events and is involved in transcriptional initiation via interaction with a conserved stem loop within the NP gene (Hoenen, Groseth, Falzarano and Feldmann, 2006). VP30 has been shown to form hexamers where as M2-1 is tetrameric however both contain a ZBD essential for protein function. The CTD VP30 crystal structure show similarities in α- helices two, five and six despite low sequence homology (9%) (Figure 1.30). These similarities suggest the potential for an anti-termination function.
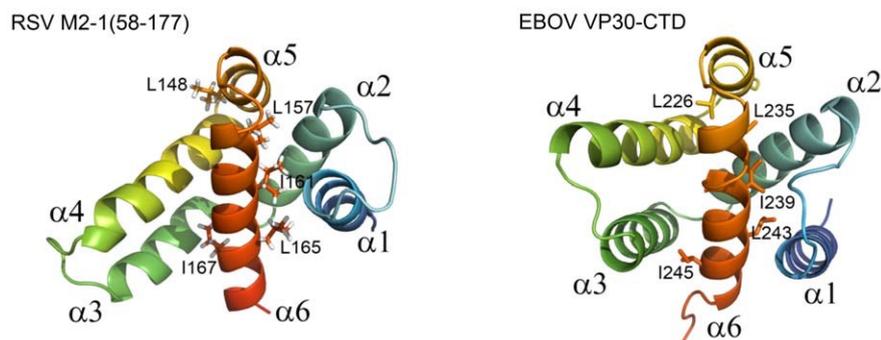
Figure 1.30 Structural Comparison of hRSV M2-1$_{58-177}$ and EBOV VP30 CTD

hRSV M2-1 core domain (58-177) and EBOV VP30 CTD structural alignment RMSD 3.9 Å despite low sequence homology. Identical hydrophobic residues in helices 5 and 6 represented with sticks. Adapted from Blondot *et al.*, 2012.

As an essential protein to the viral life cycle. M2-1 has promise as an attractive target for anti-viral compounds. Cyclopamine was discovered from a screen of bioactive hRSV replication inhibition compounds and inhibits hRSV *in cellulo*. Moreover, cyclopamine treatment reduced M2-1 expression within a mini-genome system and reduced viral lung titres in mice. Serial passages of hRSV + cyclopamine resulted in a resistant escape mutant with a single mutation within M2-1 (R151K) - within the RNA and P overlapping binding sites (Blondot *et al.*, 2012; Tanner *et al.*, 2014; Bailly *et al.*, 2016). This confirms the mode of action of cyclopamine via inhibition of M2-1 and confirming M2-1 as a valid anti-viral target.

1.3.8.2 P

The Phosphoprotein (P) is the polymerase co-factor playing important roles in RNA replication and mRNA transcription via L, N and M2-1 interactions. The 33 kDa protein is the smallest of the *Paramyxovirinae* lacking alternative ORFs and RNA editing mechanisms. P is phosphorylated at serine 116, 117, 119, 232 and 237 by casein kinase II. Phosphorylation is important for regulation of transcription and replication via an unknown mechanism (Barik et al., 1995; Mason et al., 2003). P protein may play a role in 'placing' L onto the RNP template, and phosphorylation may regulate this and other protein-protein interactions.

P has low sequence identity across the *Mononegavirales* order, however, all have 3 distinct domains. An NTD which interacts with monomeric NP; the CTD interacts with the RNP complex; a central domain that contains the M2-1 binding and the coiled coil oligomerisation domain responsible for P tetramerization. NTD residues 11-26 of P contain a short conserved MoRe site (soyuz 1) which has been reported to prevent N self-binding (Galloux *et al.*, 2015) i.e. P chaperones monomeric N at the replication complex where it will be needed to immediately encapsidate newly synthesised genome/anti-genome.

hRSV P is suggested to be intrinsically disordered at a variety of regions along the polypeptide; this is common amongst replicative co-factors as the proteins fold upon interaction with binding partners. With the exception of the coiled coil domain, P is predicted to be mostly unstructured, but folds upon its interactions with NP, L and M2-1 (Selvaraj *et al.*, 2018).

The P-L cryo-EM structure (discussed in further detail below) revealed interactions between P monomers. Interactions of monomers within the P protein oligomerisation domain were expected, however, the C-terminal α-helices of P_{2-3} rest on top of P_1, forming a three-helix bundle. P_1 alanine 169, isoleucine 173 and 178, are capped by $P_{2/3}$ hydrophobic residues. Residues in P between 131-189 and residue 205 were mutated to alanine in an minigenome replication assay to test their importance. Results showed less than 20% activity compared to WT P. 16 mutants were mapped to the oligomerization domain (within the coiled-coil domain of P), and eight mutants mapped to P:P contacts in the cryo-EM structure that are not found in the oligomerization domain. A further eight mapped to L binding domain (Figure 1.31) (Gilman *et al.*, 2019).
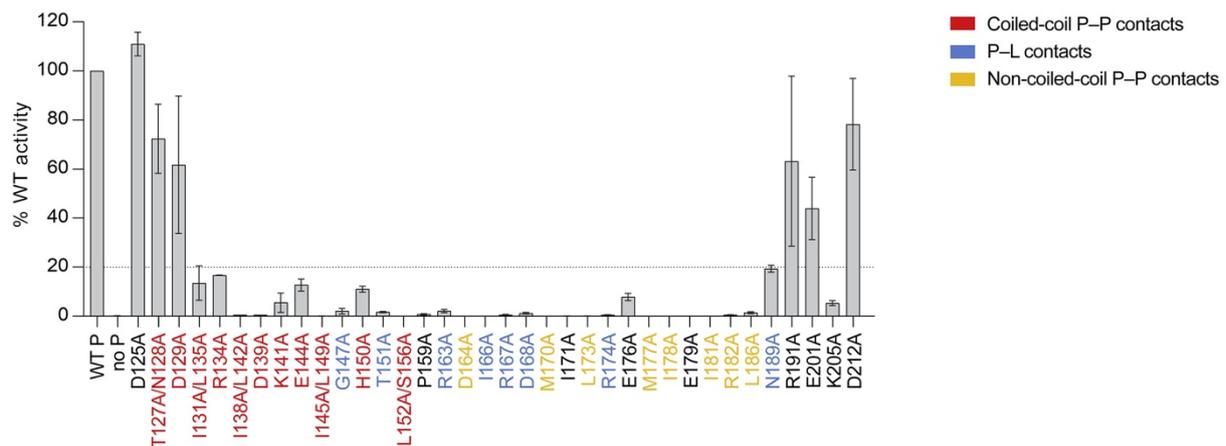
Figure 1.31 $P_{131\text{-}189}$ is Critical for Polymerase Function

Viral RNA synthesis was quantified by luciferase activity 24 hours after transfection in a minigenome system.
Dotted line represents 20% of wild-type activity.
Adapted from Gilman et al., 2019.

In order to allow transcription and replication via the RdRp, P and L interact; however discovery of the mechanism of binding is currently elusive. Initially a mini-genome system and coimmunoprecipitation studies using ^{35}S methionine labelled L and P showed that the deletion of the P coiled coil domain (residues 121-160) prevented L interaction (Yunus et al., 2001). Whilst other data showed a 50% reduction in binding when residue 120-150 were deleted. Moreover, coimmunoprecipitation was largely reduced when residues 203-241 were deleted (Asenjo et al., 2008). It is thought that P and L binding can occur in the absence of the coiled coil domain however this is essential for full P function in viral synthesis. CTD P deletions 216-241 and 203-241, resulted in the loss of the P and L interaction. The coiled coil domain of P is shorter than other viruses and it has been suggested that L binds to a distal region of P as well as the coiled coil domain (Asenjo et al., 2008; Sourimant et al., 2015).

Gilman et al., characterised the structure of hRSV P bound to L via cryo-EM. P was shown to be dynamic, with unliganded P not visible in the structure suggesting these parts remain unstructured and disordered in solution. The same residues of P monomers were shown to form distinct conformations, indicating that hRSV P shares properties with 'transformer proteins' (Gilman et al., 2019). Transformer proteins have previously been identified in EBOV; EBOV VP40 utilises different oligomeric states in order to assemble the viral matrix and regulate viral transcription (Bornholdt et al., 2013). This maximises viral efficacy when genomic content is limited.

Gilman et al., built portions of the four P monomers into the cryo-EM structure which ranged in length from 53-98 residues. The longest monomer of P built included residues 131-228 which correspond to the oligomerisation domain and most of the CTD (Gilman et al., 2019). The interaction with L stabilised residues within the CTD of P. There was large degree of variation between the P monomer conformations ($P_{1\text{-}4}$), which allowed P to wrap around the RdRp of L in a tentacular fashion. P_1 and $_4$ made extensive contacts with L, while P_3 displayed minimal contacts, P_2 on the other-hand, interacted with the other P monomers. This organisation of the P tetramers contacted 11 regions of L from 100 P residues and buried 4,000 A^2 of surface areas on L.

The CTD of P is less understood, Gilman et al., showed that contacts between L and P also encompass the P proteins oligomerisation domain and surrounding regions via hydrophobic and electrostatic interactions. Particularly, residues 167-179 of P_1 which forms a charged region. Arginine 167 174, form a salt bridge with L asparagine 718 and glutamine 722. Moreover, P_1 glutamine 176 179 also form salt bridges with arginine 523 and lysine 529 of L.

P_4 monomer was the largest monomer for which residues could be resolved in the structure. P_4 isoleucine 181 and leucine 198 form a hydrophobic cap above phenylalanine 452 on the β6 stand of L. Moreover, P_4 also forms a β-hairpin that hydrogens bonds with β6 of L forming a small β sheet. These same residues in $P_{1/2}$ found within a α helix and in P_3 are not resolved. Thus, highlighting the flexibility of P.

The base of the loop connecting the P_4 β stands hairpin, arginine 163 forms a salt ridge with glutamine 765 of L and main chain glycine 165 hydrogen bond with tyrosine 72 of L. P_4 CTD makes contacts with L; hydrogen bonds form between P asparagine 217 and 362 and threonine 362 of L.

Figure 1.32 The hRSV P-L Interaction
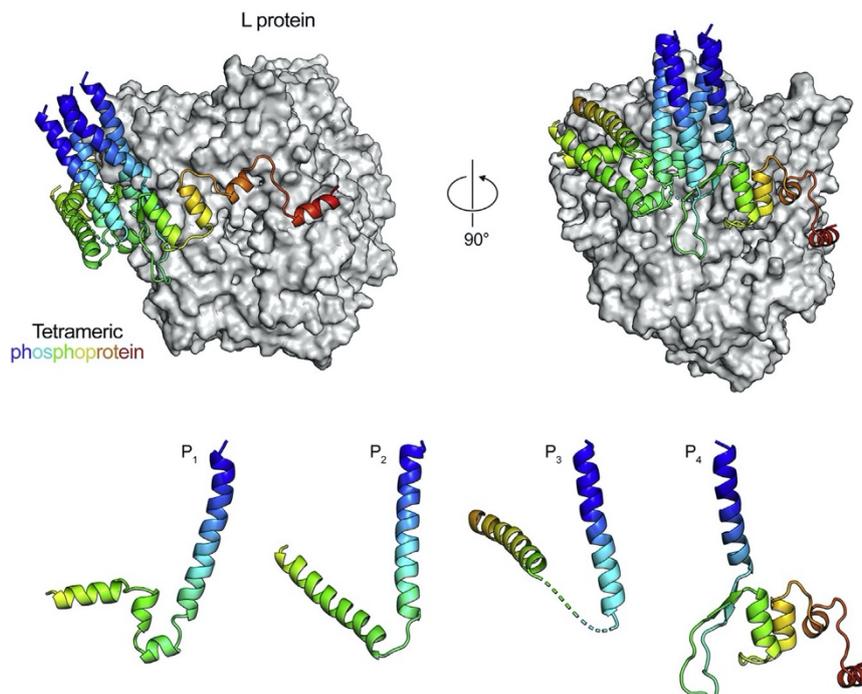
hRSV P tetramer wraps around L in a tentacular fashion. L is shown in white and P monomers shown in ribbons. The P_4 monomer is coloured as a rainbow (blue to red N to C respectively). P_{1-3} are coloured to match P_4.
Adapted from Gilman et al., 2019.

The NTD of P (residues 1-28), binds monomeric NP in order to maintain the availably of 'free' NP for newly synthesis viral RNA, acting as a chaperon protein. P also prevents NP oligomerisation and host cellular RNA binding. The binding of P to the RNP complex also allows RNP uncoating for L to access RNA for transcription and replication.

The CTD of P (residues 233-241), are the RNP complex binding region. This region has low sequence similarity to the closely related hMPV P protein. A second interaction site is within the coiled coil domain of P (residues 161-180). In bovine RSV (bRSV) the removal of this part of the coiled coil domain and the CTD completely disrupted RNP complex binding. Whilst the L and NP binding sites overlap on P, it is thought that both can bind simultaneously to different monomers within the tetramer (Yunus et al., 2001; Sourimant et al., 2015).

Monomeric and RNA-free NP binds the NTD of P (residues 1-29). NP binding to the CTD is inhibited by the NTD of P binding to NP (Mallipeddi et al., 1996; Galloux et al., 2015), here the previously mentioned MoRe site prevents NP self-binding.

Figure 1.33 Binding Regions of hRSV P

Functional regions have been mapped to full length P (1-241). N-terminal domain (1-29) binds monomeric NP (green), M2-1 binding domain (90-110 blue), coiled-coil domain (121-161) essential for P tetramerization (pink), L binding domain (203-241, red), C-terminal domain involved in RNP binding (233-241, orange).

1.3.8.3 M

M is a 25 kDa hydrophobic protein essential for virion morphogenesis. M is a critical structural component of hRSV forming a layer under the virion envelope. Here, M interacts with the cytoplasmic tail of F in order to maintain a filamentous tubular virion. Moreover, M plays a role in particle assembly and virion budding.

M has also been implicated in the silencing of viral transcription in preparation for RNP packaging (discussed previously in 1.2.1.3.4). Moreover, M is localised within the nucleus during early infection linked to host cell transcription decrease. It is thought that residues 120-170 of M are essential to this function as these are required for non-specific RNA binding. hRSV M transport to the nucleus is via direct binding to importin β1 (Meanger *et al.*, 2002; Ghildyal *et al.*, 2005)

The hRSV M crystal structure solved to 1.6 Å shows two compact N- and C-terminal domains separated by a 13-residue linker region that is largely unstructured (Figure 1.34). The CTD is a flattened beta barrel. The electrostatic surface potential reveals a continuous positive patch that covers one face of the monomer which mediates membrane association (Money *et al.*, 2009) while the NTD has a negatively charged area thought to be essential for the interaction of positively charged viral proteins. M is structurally similar to EBOV VP40 protein.
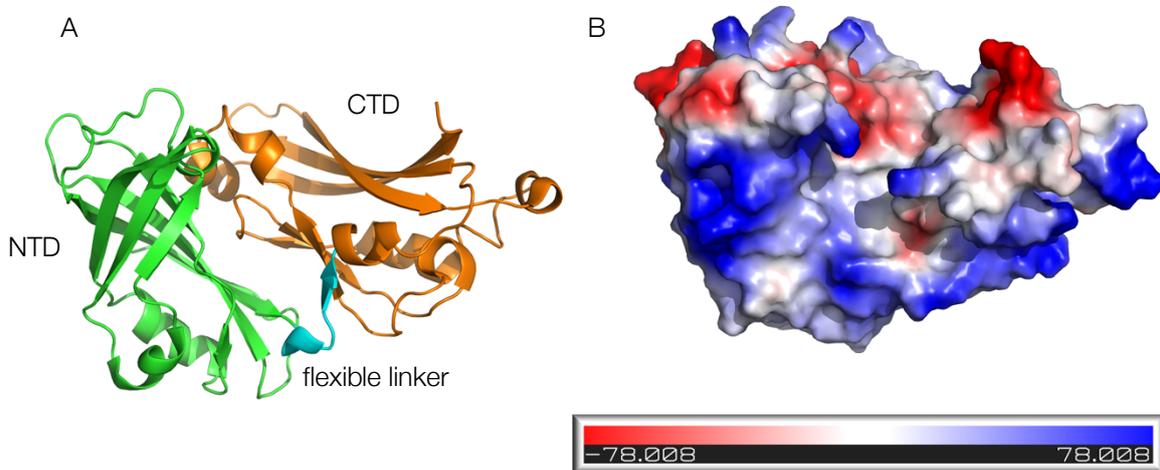
Figure 1.34 Structure of hRSV M

A- M monomer. The N-terminal domain (green) is connected to the C-terminal domain (orange) by a flexible linker (cyan). B- electrostatic surface potential of M. N-terminal domain is negatively charged, with a continuous positive patch on one face.
Figure made in PyMol (version 2.3.2).
PDB: 4V23.

1.3.8.4 M2-2

The 90 amino acid M2-2 protein is encoded by the M2 gene with overlapping M2-1 ORF. The function of M2-2 still remains elusive and deletion from the genome has no effect highlighting M2-2 as non-essential. However M2-2 deleted viruses are growth attenuated and present possible vaccine candidates (Ahmadian, Randhawa and Easton, 2000; Teng *et al.*, 2000).

1.3.9 Life Cycle

1.3.9.1 Host Cell Infection

How hRSV infects target cells has not yet been demonstrated. Initially, it was thought that clathrin mediated endocytosis of hRSV occurs (Gutiérrez-Ortega, Sánchez-Hernández and Gómez-García, 2008). Inhibition of actin, EGFR (epidermal growth factor receptor), Na^+/H^+ exchanger (NHE), protein kinase C (PKC), non-muscle myosin II and Rab5 reduced hRSV infection by 60-95% however, none of these inhibitors affected F binding. Furthermore, when clathrin mediated inhibitory experiments were completed no significant reduction of hRSV infection was shown (Krzyzaniak et al., 2013).

hRSV F protein interacts with host-cell nucleolin (*in vitro* and *in vivo*) (Tayyari *et al.*, 2011; Mastrangelo and Hegele, 2013), found at epithelial cells apical surface. Nucleolin acts a molecular shuttle between host cellular compartments; having a half-life of one hour, being a suitable target for hRSV as it is rapidly internalised and replaced via Ca^{2+} dependent clathrin-mediated endocytosis. It is thought that nucleolin is concentrated within lipid rafts, favoured for viral entry. Up to three molecules of nucleolin are able to bind to trimeric pre-fusion F (F_0), this binding induces a conformation change to the extended (intermediate) form of F (Mastrangelo and Hegele, 2013) and initiates endocytosis. F is then inserted into the host-cell membrane and a second host cell protease cleavage event triggers further irreversible conformation changes to F's post-fusion six-helix-bundle form. The lower energy state of this conformation drives this change, membrane fusion occurs and the subsequent release of the RNP complex in to the host cells (Krzyzaniak *et al.*, 2013) (Figure 1.35).

hRSV infection is a two-step process occurring through the attachment to the host cell membrane and syncytia mediated by F. Initial attachment however is electrostatically driven by GP to host cell surface carbohydrates that are negatively charged. GP, is heavily N- and O- glycosylated and removal of these results in reduced infectivity by 97% *in vitro* (Lambert, 1988). GP binds to proteoglycans by enhanced electrostatics helping to stabilise F's interaction with nucleolin.

SH has been suggested to be a viroporin causing membrane permeability acting as an ion channel (Fuentes et al., 2007). Acting similarly to 'Flu A viroporin M2 which forms tetramers while SH forms pentamers, as characterised by NMR structural studies (Schnell and Chou, 2008; Gan *et al.*, 2012).

Post-infection syncytia also occur, whereby infected cells express F on the cell surface and fuse with neighbouring cells forming multinucleated cells which allows for viral spreading. Formation of syncytia is the hallmark of the cytopathic effect of hRSV infection *in vitro*.
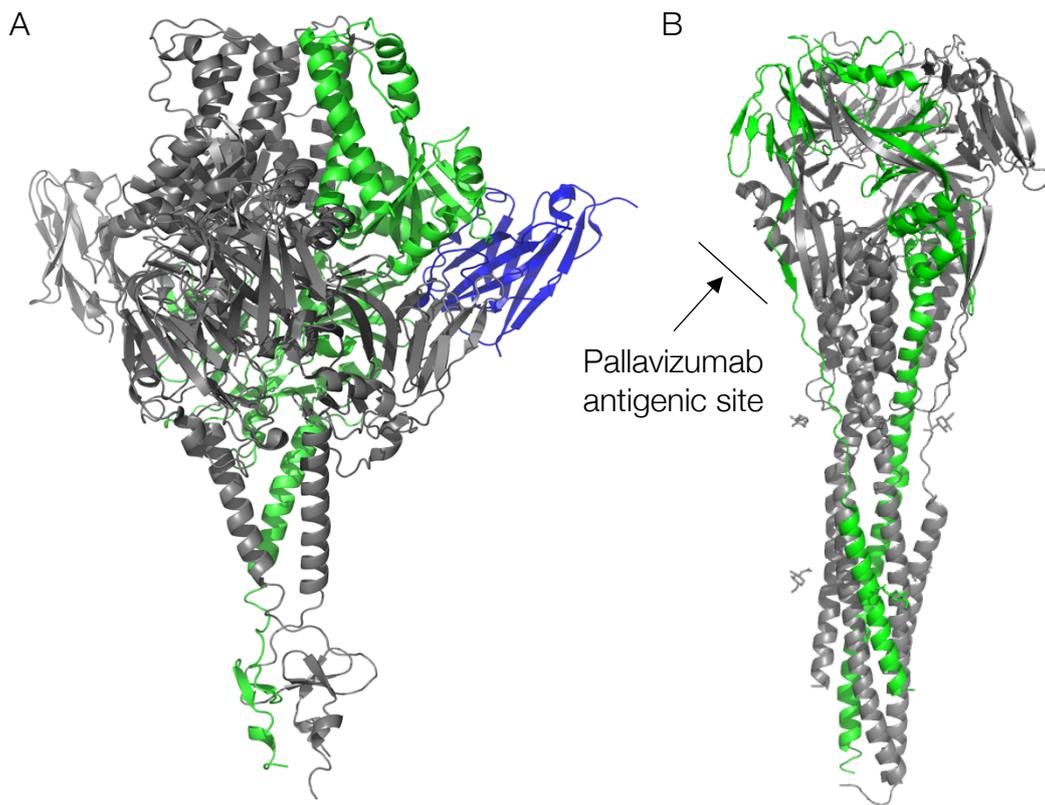
Figure 1.35 Structure of hRSV F

A- Neutralising antibody VHH-L66 (blue) bound to hRSV F, maintaining F in its pre-fusion state; 2 molecules
of the hRSV F trimer (grey) with monomer in green.
B- Post-fusion F trimer (grey) with monomer in green.
Figures made in PyMol.version (2.3.2).
PDB: 5TOJ (pre-fusion) and 3RRR (post-fusion).

Table 1.4: hRSV Encoded Proteins

Protein	Function
Non-structural protein 1 (NS-1)	Type I and III IFN antagonist. IFN α/β antagonist mediating antiviral state, suppressing maturation of dendritic cells and T-lymphocyte response. Inhibits phosphorylation of IFN response element 3 disrupting binding to IFN promoter and decreases STAT2 production through degradation (Wu *et al.*, 2012).
Non-structural protein 2 (NS-2)	Type I IFN antagonist. Causes degradation of STAT2 and interacts with RIG-I to suppress IFN synthesis (Wu *et al.*, 2012).
Nucleoprotein (NP)	RNA encapsidation. Associates with RNA forming the ribonucleocapsid (RNP) complex.
Phosphoprotein (P)	RdRp cofactor that interacts with the RNP complex to place RdRp onto the RNA. Also interacts with M2-1 and NP.
Matrix Protein (M)	Drives hRSV assembly and budding (Kiss *et al.*, 2014); vital for virus particle formation, having positive and hydrophobic domain important for cytoplasmic membrane binding (Money *et al.*, 2009).
Short hydrophobic protein (SH)	Forms a pentameric ion channel, and is able to inhibit tumour necrosis factor alpha (TNFα) signalling, perhaps helping HRSV evade the immune system (Gutiérrez-Ortega et al., 2008)
Glycoprotein (GP/sGP)	Involved in viral attachment to the host cell. Also appears in a soluble form playing a role in immune evasion through inhibition of LR3/4 mediated IFNβ induction (Shingai *et al.*, 2008)
Fusion protein (F)	Required for fusion of host cell via membranes and promotes syncytia (Blanco *et al.*, 2010)
M2-1	Essential for efficient synthesis as inhibits transcription termination at gene ends and premature transcriptional termination of long mRNAs (Leyrat *et al.*, 2014) and is therefore a second cofactor of the RNP complex. The core domain of M2-1 is responsible for the recruitment of M2-1 to the RNP complex, resulting in a high-affinity, non-globular complex which is controlled by the phosphorylation of threonine 108 of P (Blondot *et al.*, 2012).
M2-2	Inhibits viral transcription up-regulating RNA replication therefore mediates the regulatory switch from transcription to RNA replication (Bernstein *et al.*, 2012).
L	Essential replicative machinery, catalytic component containing an RNA dependent RNA polymerase (RdRp)

A summary of similarities between hRSV and EBOV proteins can be found in Table 1.5

1.3.9.2 Replication and Transcription

hRSV induces formation of amorphous or granular IBs within the cytosol where proteins forming the RNP complex are concentrated and RNA is synthesised. Similarly this has also been shown for Marburg virus (MBOV) indicating that IBs are viral replication and transcription factories (Dolnik et al., 2015). Within these IBs are concentrated areas termed inclusion body-associated granules (IBAGs) were M2-1 and viral mRNA are concentrated post-transcriptionally before release into the cytoplasm. Interestingly, IBAG formation is dependent on RdRp activity. The role of IBAGs is thought to be in viral mRNA sorting (Rincheval *et al.*, 2017).

Each gene within the hRSV genome is flanked by cis-acting conserved 9-10 nucleotide promoter (gene-start (GS)) sequence and a semi-conserved 12-13 nucleotide gene-end sequence (GE), of which four-seven nucleotides are uridines (Figure 1.22), which are recognised by the RdRp to produce mRNA via a 'stop-start' mechanism. GS sequence is 3' CCCCGUUUA 5' and differs only for L (3' CCCUGUUUU 5'). The less conserved GE sequence has a conserved 3' UCAAU 5' motif followed by three-four nucleotides of U for the generation of the 3' pA tail (described below). GS and GE sequences are separated by intergenic regions (IGRs) and together form a gene-junction (GJ). IGRs are not sequence conserved and vary in length however always end in an 'A'.

Transcription of hRSV RNA genome always starts at the 3' end (similarly to EBOV transcription and replication); GS directs the RdRp to begin transcription of each gene and the release of the nascent mRNA is directed by GE sequences; this is known as the 'stop-start' mechanism whereby transcription of the downstream gene requires termination of the upstream at the GE sequence. At GE the polymerase can 'fall off' once the nascent mRNA has been released. In order to produce full length transcripts the RdRp needs to remain attached to the RNA genome at these GJs and then begin transcription again at the next promoter, however, exactly how the RdRp does this is currently unknown. It is thought that once the RdRp recognises a GE sequence the nascent mRNA is released and RdRp can move bidirectionally to find a GS sequence, a process known as 'scanning'. This is also true for the M2 gene GE which overlaps with L GS by 68 nucleotides. Here, the polymerase releases the nascent M2 mRNA and moves upstream of the M2 GE sequence in order to 'find' the L GS. The M2 GE is essential for L transcription even though the 64 nucleotides between L GS and M2 GE do not modulate transcription and are not sequence conserved (Fearns & Collins, 1999; Kuo et al. , 1996).

The ratio of transcription of each gene is dependent on the RdRp, with genes towards the end of the genome producing fewer transcripts due the polymerase "fall off" at gene-end sequences. This explains the organisation of the genome; less RdRp is needed to over-throw the host immune system and produce functioning virions, therefore L at the 5' end of the genome is transcribed last, however with sufficient levels of L available for transcription and replication purposes. By contrast, a high abundance of NS1 and NS2 protein is needed quickly for host immune evasion. Viral mRNAs are detected within host cell cytoplasmic four-six h.p.i. Gene ends direct polyadenylation (Fearns 2002; Stillman & Whitt 1997; Barr et al. 1997) which is carried

out by the RdRp via a stuttering mechanism. During mRNA production, the RNA polymerase can one- extend the nascent chain by one nucleotide, or two- pause. The latter increases the probability that the active site of the polymerase along with the nascent mRNA chain are repositioned one-two nucelotides upstream and the template is copied a second time (Phane Hausmann *et al.*, 1999)

M2-1 is thought to play a role here and is an essential transcription factor, without M2-1, there is an abundance of short non-functional mRNAs that are produced. In a hRSV negative-sense minigenome system which used NS-1-NS-2-CAT (chloramphenicol acetyltransferase) system (a chimera of the first 1125 nucleotide of hRSV), where NS-1 and NS-2 genes were identical to the 3' sequence of hRSV; NS-1 and NS-2 production was not dependant on M2-1 protein, however in the absence of the M2-1 protein, CAT mRNA was not produced. The inclusion of M2-1 in the minigenome assay increased NS-1 and NS-2 production 3-fold but CAT mRNA 27-fold. The dependence of CAT transcription is thought to be due to its length and position within the genome. In WT infection, the third gene in hRSV transcription would be of the NP. The NP is shorter than CAT and is therefore assumed that NP transcription would also be heavily dependent on the presence of M2-1 (Fearns and Collins, 1999). The M2 gene is transcribed as a single mRNA and translated into M2-1 and M2-2 from the overlapping open-reading frames.

Translated viral proteins remain within the cytoplasm until a level of free NP concentration is reached and viral genome replication is switched on.

The polymerase cofactor P, is essential for L activity. The P is required for the RdRp to move along the tightly, NP encapsidated RNA genome. The P plays an important role in the switch from transcription to genome replication by a currently unknown mechanism (potentially via interactions with M2-1) and/or interaction with the RdRp that ignores the aforementioned GS and GE sequences in order to produce full-length positive-sense RNA (an anti-genome). There are thought to be two different replication switch models: one- 3' leader region is encapsidated by NP and RNA synthesis if not terminated, allowing synthesis to occur along the whole genome. This model is thought to be dependent on the concentration of 'free' NP. Two- there are two different 'forms' of polymerase complexes; a transcriptase and replicase. The first composed of L, P and M2-1 proteins and the latter L and P only.

The cryo-EM model of hRSV NP in complex with RNA supports model 1; showing the 3' polymerase recruitment sequence and GS sequence are within close proximity (Bakker *et al.*, 2013) (Figure 1.36). On the other hand, VSV has also shown two forms of polymerase complexes (Qanungo et al., 2004; Tawar et al., 2009). However, more recent publications indicate that it is more complex than this; the polymerase is able to initiate at position three+ as well as one+ due to a GS-like sequence within the promoter suggesting different initiation sites differentiate transcription and replication respectively. When initiation occurs at 3+ mRNA is prematurely terminated at a 5' signal within the Le; thus, the polymerase complex is allowed to 'scan' further downstream and re-initiate at the 3' GS sequence. Termination Le transcripts have been identified in RNA purified from hRSV infected cells, supporting this theory (Tremaglio et al., 2013).

The replicative components consist of RdRp, P and N forming the essential proteins for replication. During replication the RdRp produces a single continuous anti-genomic RNA which is again encapsidated by NP and then replicated again producing copies of negative sense genomic RNA. The NP binds to P which keeps monomeric NP close to the replication complex for concurrent RNA encapsidation (Stokes et al., 2003). It is the 5' Le that encodes the 3' reverse complement leader promoter which allows for antigenome synthesis (Curran & Kolakofsky, 2008; J. Li et al., 2008).

The RNP complex serves a role in stability protecting the genomic RNA. N also provides protection from toll-like receptors (TLRs) and 'hides' RNA helices that initiate innate immune response through IFN and nuclear factor κβ (NF-κβ) responses (Collins and Graham, 2008).
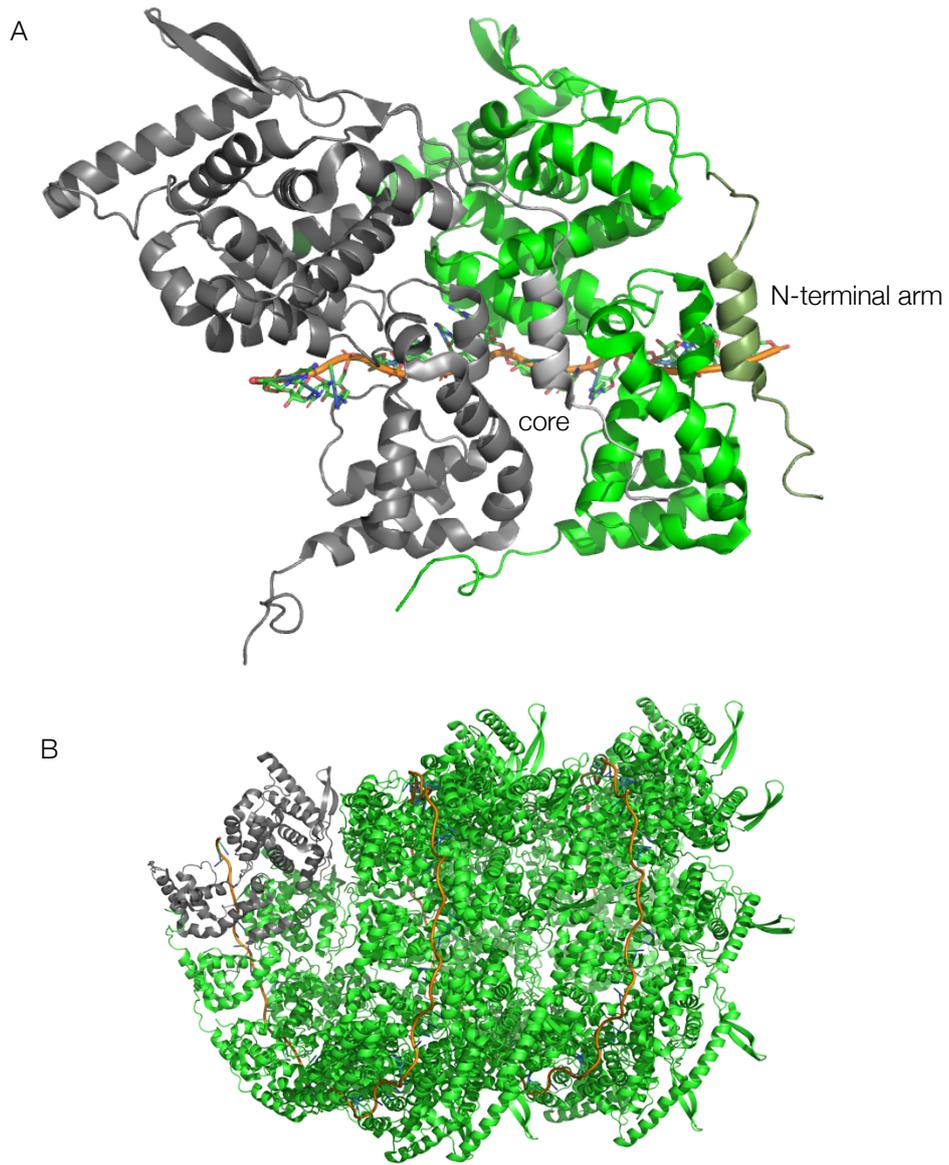
Figure 1.36 Atomic Model of hRSV NP:RNA Interaction

A- NP:NP interactions (monomers shown in grey and green). N-terminal arm (light grey and pale green of
adjacent monomer) binds core of adjacent monomer (green). RNA shown in orange.
B- RNP complex forms a lefthanded helix. RNA in orange is orientated on the outer side of NP.
Figure made in PyMol (version 2.3.2).
PDB: 4BKK.

1.3.9.3 Translation of the M2-2 Gene

Unique to *Pneumoviridae* the M2 gene is transcribed into a single mRNA and translated into 2 proteins from two overlapping open reading frames (ORFs) M2-1 and M2-2. M2-2 has three possible start codons and is upstream of the M2-1 stop codon. Translation is by a coupled translation mechanism. M2-1 termination is required for M2-2 translation and the reverse translocation of the host ribosome for re-initiation at the M2-2 start codon(s). M2-1 truncation analysis revealed that full length M2-1 ORF was required for wild-type (WT) reporter expression from the second ORF, suggesting RNA secondary structures are involved in this coupling process; moving the M2-1 stop codon towards the 5' end of the M2-2 start codons also prevents M2-2 translation. Mfold predictions support this (Ahmadian, Randhawa and Easton, 2000; Gould and Easton, 2005).

1.3.9.4 Assembly

Assembly of new virions occurs at the polarised apical surfaces of epithelial cells. Hypothesised budding is thought to utilise apical recycling endosomes (ARE) which are enriched with Rab11a. Unlike filoviruses hRSV does not utilise the ESCRT (endosomal sorting complexes for transport) -pathway for virion budding; hRSV budding occurs in the presence of ESCRT-pathway inhibitors and is therefore thought to be independent of this.

Amongst the newly packaged viral proteins a number of host proteins are also found within virions including lipid raft microdomains enriched with caveolin-1, CD5 and CD58, and actin (Sugrue *et al.*, 2002; Jeffree *et al.*, 2007; Collins and Graham, 2008). Inhibitors of the actin associated heat-shock protein (HSP)-90 in hRSV infected calls resulted in the inhibition of virion assembly due to the inability of actin to allow assembly and release (Radhakrishnan *et al.*, 2010).

1.4 Protein Comparisons between hRSV and EBOV

In addition to the aforementioned structural similarities between proteins including M2-1 / VP30 and P / VP35 of hRSV and EBOV respectively, there are more common similarities between the two.

Both viral RNA genomes are in the negative polarity and are non-segmented. hRSV and EBOV share a common 5 genes that include: NP, polymerase co-factor P or VP35, matrix protein M or VP40, glycoprotein GP (filovirus GP also share similarities with hRSV F), transcriptional regulator M2-1 or VP30, and L. Function similarities are highlighted in Table 1.5. Unsurprisingly, the genomes are arranged in a similar manner thereby proteins expressed in a similar order during infection.

Table 1.5: Similarities between hRSV and Filoviral Proteins

Protein	hRSV Function	EBOV Function	Protein
NS-1	Type I and III IFN antagonist.	N/A	N/A
NS-2	Type I IFN antagonist.	N/A	N/A
NP	Encapsidated viral RNA forming an RNP	Encapsidates viral RNA forming an RNP.	NP
P	Tetrameric polymerase cofactor.	Trimeric and/or tetrameric polymerase cofactor. Type I IFN antagonist.	VP35
M	Drives assembly and budding	Drives assembly and membrane budding	VP40
N/A	N/A	Minor matrix protein. Involved in NC formation and assembly. Counteracts type I IFN response.	VP24
SH	Pentameric ion channel.	N/A	N/A
GP/sGP	Cell surface attachment. mRNA edited sGP plays a role in immune evasion	Type 1 membrane protein. Mediates fusion and receptor binding during viral entry. sGP is formed from a non-edited mRNA, secreted acting as an anti-inflammatory.	GP/sGP
F	Required for fusion of host cell via membranes and promotes syncytia (Blanco *et al.*, 2010)	See GP	
M2-1	Tetrameric anti-termination factor regulated by phosphorylation. RNP complex co-factor.	Hexameric transcriptional activator regulated by phosphorylation. RNP complex co-factor	VP30
M2-2	Mediates the regulatory switch between transcription and regulation	N/A	N/A
L	Essential replicative machinery for transcription and replication via an RdRp	RdRp, enzymatic activity required for transcription and replication.	L

1.5 Project Goals

This thesis is split into three sections representing each viral family outlined in the introduction.

The aim of the *Orthobunyavirus* chapter is to elucidate the structure of AKAV and SIMV NP with and without synthetic RNA. The synthetic RNA is represented by pA/C/G/U and 10 bases in length . The literature supports the hypothesis that each NP monomer from this genus binds 8-11 bases of RNA and therefore 10 bases was a suitable choice for RNA here. Once crystal structures are fully refined, residues that interact with the RNA will be mutated to glutamine in order to assess their importance to NP binding to RNA.

Secondly, the aim of the Filovirus chapter is to elucidate the structure of full-length EBOV and MARV VP30 and compare the two structures to each other but also to M2-1 of hRSV, a structural homologue. Solving the VP30 crystal structure will also elucidate the oligomeric state as this still remains elusive and is thought to be tetrameric or hexameric. We further aimed to understand the interaction between VP30 and VP35 and whether or not this interaction is dependent on RNA.

Lastly, a chapter on hRSV aims to further understand the interaction between M2-1 and its binding partner P and also M using X-ray crystallography and electron microscopy.

Chapter 2 Introduction to Methods Used

2.1 Principles of Chromatography

Chromatographic techniques such as affinity chromatography purify proteins based on ligand specificity whereas size exclusion chromatography (SEC) separates proteins based on size. For NP, VP35, VP30, P, M2-1 and M protein purification both techniques were utilised.

Affinity chromatography using nickel and histidine affinity utilised HisTrap columns pre-packed with cross-linked agarose beads and nickel ions (Ni^{2+}) coupled to the chelating matrix. NP, VP30, P and M fused to a six histidine (His6) purification tag bind to the nickel ions, whilst non-specific proteins are eluted. Tagged protein is eluted with imidazole. HisTrap columns are re-usable up to 10 times when washed with H_2O and stored correctly in 20% ethanol. Columns can then be stripped with 50 mM EDTA and re-charged with 0.1 mM nickel sulphate.

Glutathione S-transferase (GST) fused VP35 and M2-1 can be purified similarly. A glutathione (GSH) sepharose support medium, which has the glutathione substrate covalently attached, binds GST and allows non-specific proteins to be eluted. Tagged protein is eluted via incubation with GSH and the medium returned to its unbound from.

SEC media is formed from a matrix of resin which contains spherical particles, filled with buffer that penetrates the channels within the particles of the matrix. Higher molecular weight proteins do not pass through the pores and are eluted earlier, whereas lower molecular weight proteins spend time within the pores. Consequently, molecules are separated by size. By using a series of standard molecular weight proteins, one can generate a calibration curve and use this to approximate the molecular weight of the protein in aqueous solution. Thus, one can estimate oligomeric state of a protein in solution, which is not possible when, for instance, running an SDS PAGE where the protein is unfolded, and quaternary structure destroyed. One caveat of this method is that the shape of the molecule can affect elution volumes as well as the molecular weight. Non-spherical proteins may not elute at the predicted point vs the calibration curve.

2.2 Fluorescence Anisotropy (FA)

One method used in this thesis to measure the affinity of the interactions between proteins and their binding partners (ssRNA and dsRNA) was fluorescence anisotropy (FA).

FA measures the diffusional tumbling rate of the fluorescein (Fl)-labelled RNA. It is often chosen for biological studies as it is high throughput, semi-automated and requires low quantities of material enabling the estimation of binding affinities up to the µM range.

Moreover, compared to classical methods such as gel shift assays, FA is a real-time and true-equilibrium method.

After electrons of Fl are excited by photons of plane polarised light, photons fluoresce with a degree of polarisation inversely proportional to the tumbling rate of the molecule. Unbound Fl-RNA is small, having a high tumbling rate in solution resulting in increased depolarisation of light as the 3'Fl-RNA is in a 'new' orientation when this fluorescent event occurs. When bound to protein, the complex is larger, therefore tumbles slower and emits more polarised light (Figure 2.1) (Pollard, 2010). The degree of polarisation is measured by emission filters that are parallel and perpendicular measuring emission intensities.
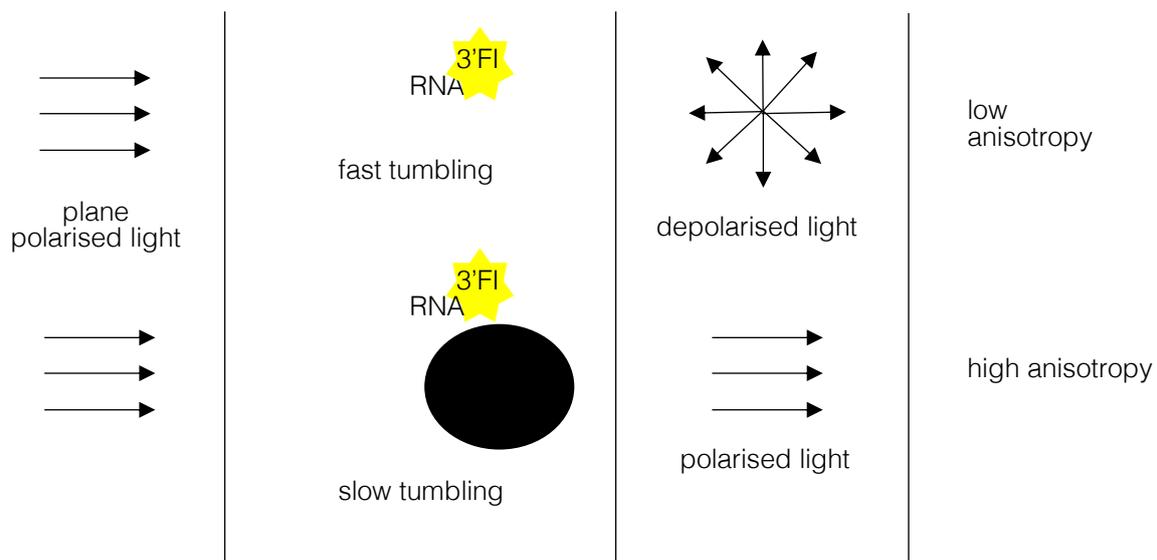
Figure 2.1 Fluorescence Anisotropy Schematic

Plane polarised light excites a fluorescently labelled RNA, the light is depolarised due to rapid tumbling in solution. Upon binding to protein (black circle), tumbling rates decrease polarised light it emitted to an extent

2.3   Circular Dichroism (CD)

Circular Dichroism (CD) spectroscopy can determine protein secondary structure in the far-ultraviolet (UV) spectrum (190-250 nm) (Nakanishi, Berova and Woody, 1994) and tertiary structure in the near-UV region (240-350 nm) (Kelly and Price, 2000) as well as folding and binding properties of a protein (Greenfield, 2006). Circular polarised light is absorbed around a chiral centre (peptide bond) which acts as a chromophore. The chromophore absorbs right and left polarised light differently which is measured by the spectropolarimeter. The spectropolarimeter switches between right and left polarised light and compares the difference between the two absorbance signals, known as ellipticity (θ). CD data is used to deconvolute secondary structural type; treated as the sum of characteristic individual spectra arising from each secondary structure present in the sample. From this the mean residue ellipticity (MRE) is calculated when ellipticity is combined with the mean residue molar concentration. A distinct MRE signal is seen when peptide bonds are arranged regularly and folded. Thereby α-helices, β-sheets

and random coil secondary structures have a unique signal (Nakanishi, Berova and Woody, 1994). This thesis focuses on proteins that are largely α-helical (represented in Figure 2.2 in green). α-helical proteins are typically characterised by troughs at 208 and 222 nm (Figure 2.2), the ratio between 222/208 nm (>1) provides evidence for the formation of a coiled-coil structure. Protein with well-defined antiparallel β-sheets have
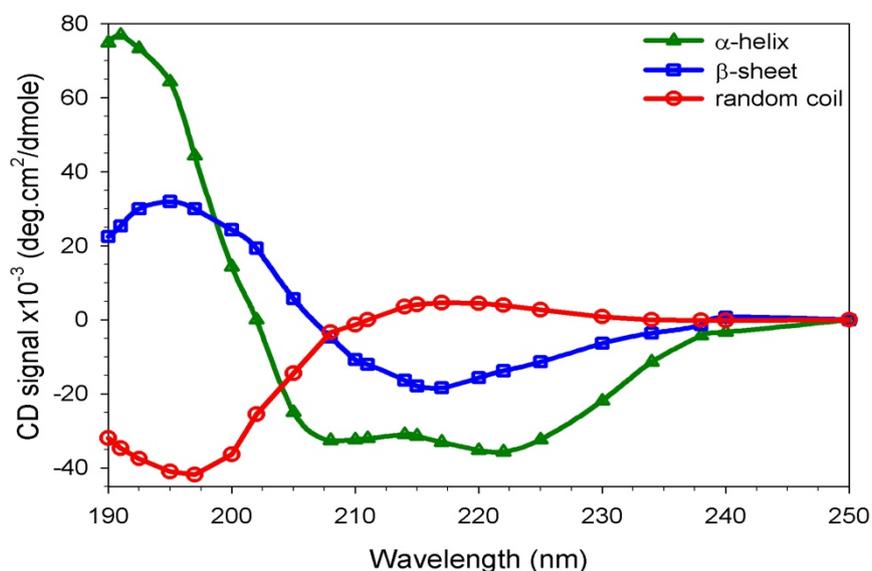
Figure 2.2 Circular Dichroism of Secondary Structures

Distinct MRE signals are seen for proteins with three particular secondary structure elements; Three different peptide bonds (green α-helical, blue β-sheet and red random coil) when arranged regularly and correctly folded.

a trough at 218 nm. Analysis of CD spectra to estimate secondary structures used in this thesis was DICHROWEB (Whitmore and Wallace, 2004).

2.4 SEC-MALLS

The analysis of molecular weight by SEC relies on two assumptions: the protein shares the same relationship between diffusion properties and molecular weight, and, does not interact with the column; this is true for non-globular proteins and disordered proteins. When combines with multi-angle light scattering (MALS) and differential refractive index (dRI) detectors, molecular weight determination is more reliable. SEC is performed firstly on the protein sample, for separation. After, the protein sample pass through a MALLS detector probed by a laser beam. MALLS signal, UV absorbance and dRI signal are analysed to quantify the protein properties. MALS detector measures light scattered relative to the incident laser beam and dRI detectors determine concentration based on the change in solution refractive index (Some *et al.*, 2019). Molecular weight is therefore determined independently from elution time using the following equation:

$$MW = \frac{R(0)}{Kc\left(\frac{dn}{dc}\right)2}$$

Whereby molecular weight (MW) is determined by the: Rayleigh ratio (R(0), the amount of light scattered by the protein relative to the laser intensity), the concentration (c) determined by the UV or dRI detector, dn/dc, the refractive index and K, the optical constant.

2.5   Mass-Spectrometry

Mass spectrometry (MS) can measure the different molecules (therefore oligomeric states) within a sample. As the sample needs to be charged it is ionized by the addition or removal electrons from a neutral sample, this typically occurs in the gas phase (Siuzdak, 2004). Electrospray ionization (ESI) was used to produce charged sample for this thesis. ESI produces gaseous ionizes molecules directly from liquid, creating a fine spray of highly charged ions in the presence of an electromagnetic field (Siuzdak, 2004). The charged sample is the sorted in mass by acceleration and deflection. Positive ions accelerate towards negative plates, the speed at which the occurs is dependent on mass. These are then deflected by a magnetic field, which occurs before reaching a detector. Each peak in a mass spectrum shows a component of unique mass:charge ratio (m/z) in the sample, and heights of the peaks connote the relative abundance of the various components in the sample. Ionised molecules within the sample have an m/z of 1 which is equal to the mass of the molecule (Baker, 2010). Adjacent peaks on a mass-spectrum often differ by a single charge, equivalent of the addition of a single proton, highlighting how the mass of the molecule within the same sample remains the same, but the m/z varies. During native MS, the number of acquire charges usually correlates with the surface area of a globular proteins. Several charge states within a sample require the calculation of molecular weight from adjacent peaks.

2.6 Protein Structure Analysis

After a protein production protocol has been established, this high-quality protein can be used for structural studies including nuclear resonance spectroscopy (NMR), X-ray crystallography and electron microscopy (EM). Depending on the information required each technique has its advantages, some are as follows; recent advances have allowed cryo-EM to compete with X-ray crystallography for high resolution structures and the use of imaging in EM overcomes the phase problem in X-ray crystallography (a diffraction method). Furthermore, crystallography requires the formation of protein crystal rather than EM and NMR that both require protein in solution, and potentially provide a more physiologically relevant structure of protein(s). EM and NMR have protein size limitations; negative stain EM requires proteins larger than 75 kDa, cryo-EM greater than ~150 kDa and NMR less than 20 kDa (or 75 kDa if selectively labelled). X-ray crystallography is material intensive requiring protein at high concentrations (~10 mg/mL), and NMR requires similar concentrations, but also requires the production of ^{15}N labelled protein or even ^{13}C ^{15}N double labelled protein. On the other hand, EM requires much lower (~0.1 mg/mL) protein.

X-ray crystallography and negative-stain EM techniques were used in this thesis and are outlined below.

2.6.1 X-Ray Crystallography

2.6.1.1 Crystal Formation

Protein crystal formation is required for X-ray crystallography from >95% pure, soluble and concentrated protein. Each crystal contains multiple (billions of) copies of the protein which are arranged regularly within the crystal lattice which allows the detection of diffracted X-rays. A mathematical formula (Fourier transform) can then be used to calculate a three-dimensional electron density map of the building blocks of the crystal – the protein. Protein crystals form during protein supersaturation, overcoming the energy required for nucleation whereby a cluster of protein molecules form a crystal together (Chernov, 1997). Nucleation occurs in the labile stage of protein crystallography (Figure 2.3 A), with crystal growth occurring in the metastable stage. Supersaturation however can also proceed forming protein aggregates and precipitation within the precipitation stage; there is a delicate balance in finding conditions where the protein does not aggregate or precipitate but crystallises (Chernov, 1997; Dessau and Modis, 2011).

Commercially available screens available contain a diverse range of conditions for protein precipitation and are known as mother liquor. Mother liquors are biased toward conditions published for previously crystallised proteins (McPherson and Gavira, 2014). This thesis utilised two crystal formation methods: vapour diffusion hanging drop and sitting drop. While batch and dialysis diffusion techniques are also commonly used. During vapour diffusion (whether it be hanging or sitting drop) the protein:precipitant drop is suspended and equilibrated with a precipitant reservoir (Figure 2.3 C) such that the volatile solvent (water) moves from drop to mother liquor, concentrating the protein

and precipitate and moving the solution toward saturation then supersaturation (Dessau and Modis, 2011). Sometimes this produces crystals (in a non-predictable way) therefore many different conditions must be screened in order to find the few conditions that produce useful crystals.
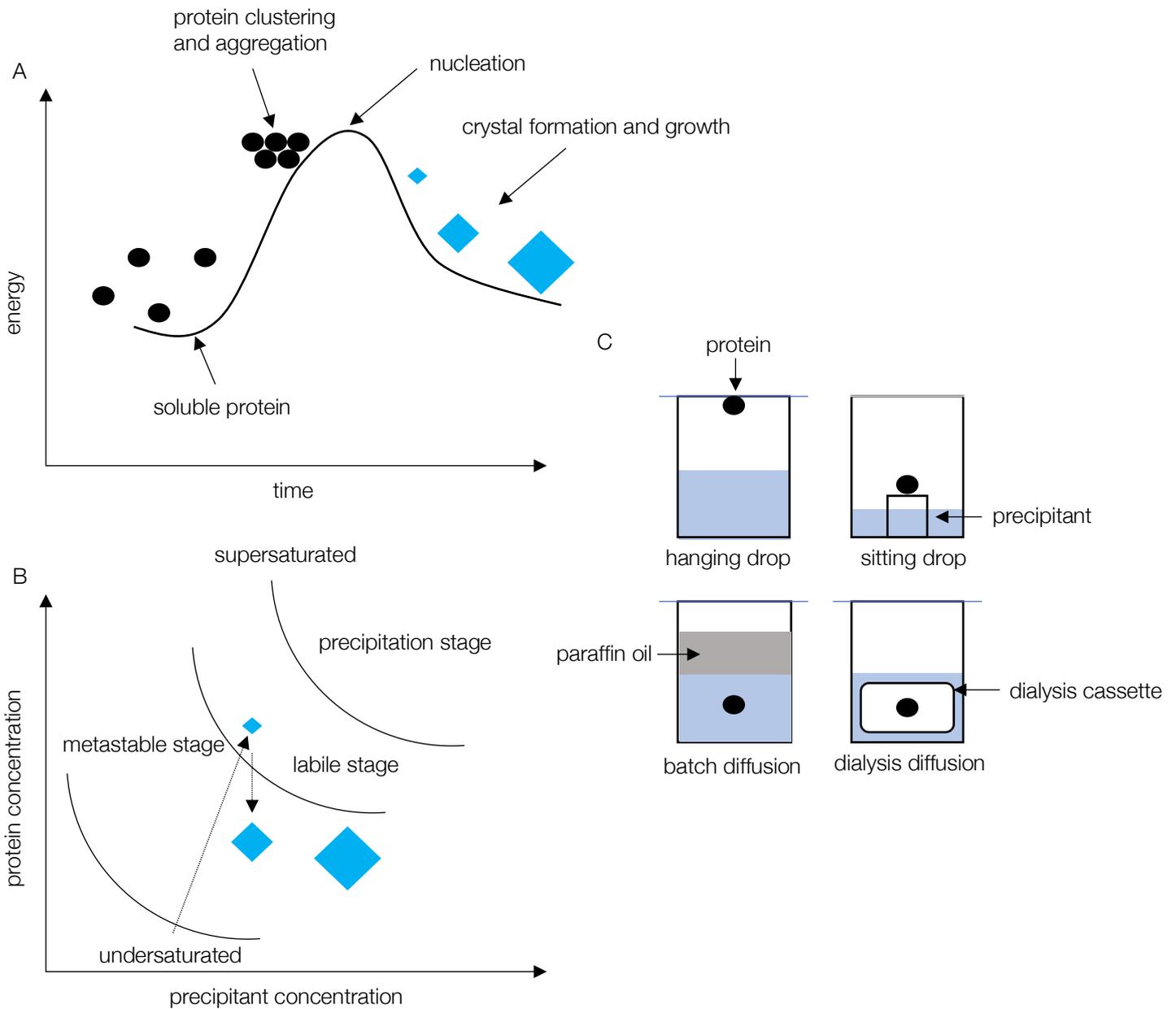
Figure 2.3 Solubility Phase Diagram

A- nucleation events for crystal formation
B- the phase moves from undersaturated to supersaturated as protein precipitation increases. During the liable stage, spontaneous nucleation of protein crystals occurs, lowering the concentration of protein in solution. The metastable stage is then reached which supports crystal growth. The precipitation stage does not support crystal growth but causes disordered aggregation of protein.
C- crystal growth by vapour diffusion in different formats.

2.6.1.2 Solving the Crystal Structure

Data is collected using X-rays due to their wavelength being in the same order of magnitude as the interatomic distances that need to be resolved during a crystallographic experiment (1 Å = 0.1 nm). During data collection, photons from the X-ray waves mostly pass through the crystal. Occasionally however, photons interact with electrons of protein atoms in the crystal lattice and are 'scattered'. These scattered X-ray waves emerge from the protein crystals with different phases and scattering angles. For X-rays to constructively interfere the waves must be in phase, having peaks and troughs at the same time. If the second wave is out of phase, this interferes and cancels out the wave, known to be destructive and no photons will be produced and in turn no intensities (known as reflections) are recorded. These reflections form a diffraction pattern which is the raw data of X-ray crystallography experiment (Smyth and Martin, 2000).

The scattered X-ray waves must meet two conditions; the Ewald sphere and Bragg's law in order for diffraction to be recorded. Bragg's law ($n\lambda=2d\sin\theta$) (Figure 2.4), relates to the angle (θ) of incidence between lattice planes and two parallel waves that are in phase scattered from the two planes within the crystal lattice, and the distance between the planes (d). The path difference between two waves is $2d \sin\theta$. Scattered X-ray waves remain in phase when the difference between the two waves paths lengths is equal to a whole number of wavelengths ($n\lambda$), i.e. spots can be observed when Braggs law is met, when $n\lambda=2d\sin\theta$ (Bragg and Bragg, 1913).
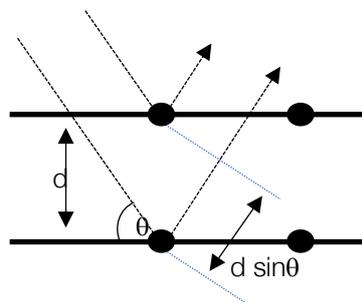
Figure 2.4 Bragg's Law

If two parallel waves are a distance (d) apart, with an angle of incidence (θ), these waves path distance will remain in phase if the second wave travels an addition distance (d sinθ) that is an integer of the wave.

The way in which Bragg's planes intersect each unit cell are described by Miller indices (hkl). Miller indices are assigned to each plane that contributes to a diffraction pattern. Each reflection in this diffraction pattern is therefore assigned a corresponding Miller indicex during diffraction data indexing.

All X-ray waves that contribute across the whole crystal for a single reflection can be added together in the structure factor equation (Fhkl). The structure factor F is a vector where the amplitude is directly proportional to the square root of the diffraction intensity (the measured reflection intensity and their position). Each diffraction pattern only provides information about the crystal in a single direction, therefore the crystal must be

rotated in ordered to produce a complete dataset. Different sets of X-ray waves will therefore be recorded during crystal rotation.

Phases cannot be recorded during data collection creating the 'phase problem', whereby the phase of X-ray waves that generate diffraction spots are unknown. To estimate phases, 3 methods are commonly used; 1) heavy metal soaking, 2) molecular replacement and 3) incorporation of atoms with "anomalous dispersion" such as selenomenthionine, which is now more common than heavy metal soaking. Molecular replacement was used in this thesis as a reference structure was available for model building. Typically, molecular replacement can be used for structurally similar proteins, or where proteins share >30% sequence identity. Molecular replacement works by fitting the known reference structure into the same orientation and position of the unknown structure. The phases from the reference structure can be used in the structure factor equations for the unknown structure along with its experimental structure factor amplitudes. A Fourier transform can then be applied to generate an electron density map Fo-Fc. Fo-Fc electron density map show where model has been built. The 2Fo-Fc electron density map shows the difference between the observed and calculated phases, highlighting where model needs to be built into and/or moved and/or removed.

should show features missing from the original model than can be built into from the unknown structure.

Patterson maps are derived from the Fourier transformation of the structure factor amplitudes, and do not require phases. Patterson maps therefore show the vectors between atoms not the position of atom within a molecule. Intra-molecular vectors represent the distance between two atoms within a molecule, and inter-molecular vectors represent the distance between two atoms of neighbouring molecules. During molecular replacement, the orientation of the unknown structure is compared to the reference structure through intra- and inter-molecular vectors of their corresponding Patterson maps.

2.6.1.2.1 Refinement

Refinement seeks to minimise the difference between the calculated structure factor amplitudes of the reference structure and the experimental structure factor amplitudes (R_{value}) derived from the diffraction pattern. This generates the most accurate electron density map. REFMAC5 (Murshudov, et al. 2011) was used for the refinement process. Improving the model also improves the calculated phases, which are in turn used to improve the electron density map, which is used in the cycle to improve the unknown structure. COOT (Emsley and Cowtan 2004) was then used to manually adjust the position of each residue in the protein chain to improve electron density fit, torsion angles and bond lengths. Iterative cycles of refinement and rebuilding improve the model and the Rfactor drops as the calculated and experimental amplitudes converge

This cycling however, can produce bias or errors in the model. To overcome this, assessment of the refinement process is determined by calculating the R_{value} for ~5% of data that was omitted during refinement. This is known as R_{free} which is used as a cross

validation that the model is improving and that the calculated structure factor amplitudes are moving towards the original experimental data, independent of the refinement.

Table 2.1 Quality Statistics used in X-ray Crystallography

Quality Indicator	Summary
R_{value}	Measure of the difference between the calculated and observed structure factor amplitudes, based on the refined model. A perfect R_{value} would be 0, however in practice this should roughly be < 10% of the resolution e.g. at 2.1 Å R_{value} should be <0.21
R_{free}	Measure of how well the refined model can predict the original structure factor amplitudes from the 5% data that was omitted from the refinement process. The difference between R_{value} and R_{free} should be less than 5-7%
B factors	Measurement of the degree of movement of atomic positions in the model. Could vary due to: thermal motion (vibration of an atom around its rest position) and disorder in the crystal. High B factors indicate atomic position uncertainty
RMSD bond lengths	Root mean squared deviation of model bond lengths, should be <0.02 Å
RMSD bond angles	Root mean square deviation of model bond angles, should be < 4 Å

2.6.2 Transmission Electron Microscopy

Recent technical advancements have enabled high resolution structural determination via electron microscopy (EM), which is becoming an increasingly popular alternative to X-ray crystallography. Transmission EM (TEM) transmits electrons through a sample generating a 2D image from the electron:sample interaction allowing atomic-resolution structures to be produced (Merk *et al.*, 2016).

Figure 2.5 outlines a typical TEM existing of an electron source, lenses, specimen holder and a detection system. TEMs are vacuumed in order to minimise electron loss and gas atoms. Liquid nitrogen $(LN)_2$ is used as a cold trap, cooling a copper rod which traps water vapour to prevent contamination (Scarff *et al.*, 2018).

The electron source is typically lanthanum hexaboride (LaB_6) or tungsten single crystal or filament, connected to a high voltage source. The T12 at the University of Leeds was used to produce micrographs for this thesis. The T12 operates with a 120 keV LaB_6 electron source. An electrical current causes electron emission, due to high temperatures (thermionic emission) or high potential difference between two electrodes (field emission) (Campbell *et al.*, 2014; Herzik, Wu and Lander, 2017).

Magnets produce an electromagnetic field which in turn produce lenses to focus the electron beam. In TEM there are typically 3 lenses: condenser, objective and projector. The condenser focuses the electron beam to the sample. The objective lens focuses the beam after it passes through the sample and the projector lens expands the electron beam onto the detection system in order to produce an image.

Image production occurs via electron scattering elastically or inelastically. Elastic electron scattering does not result in energy loss when the electron path is altered. On the other hand, inelastic scattering transfers energy from the electron to the sample. This phenomenon occurs frequently (4:1 vs elastic) and causes sample damage (Baker and Rubinstein, 2010).

Electrons are detected now by direct electron detectors (DEDs, commonly also known as direct detection device (DDD)). Detecting a voltage drop across a capacitor DEDs combine speed and efficiency compared to older methods which include film and charge coupled devices. DEDs allow for the production of multiple frames in the same field of view. The efficiency of a detector is expressed as the quantum efficiency which is the ratio of output signal-to-noise ratio (SNR) and input SNR. Film provided a high number of pixels with a small pixel size but a high background noise. Charged couple devices increased speed of detection compared to film but had a lower quantum efficiency (McMullan *et al.*, 2009).

electron source

condenser lens

condenser aperture

sample holder

objective lens

objective aperture

projector lens
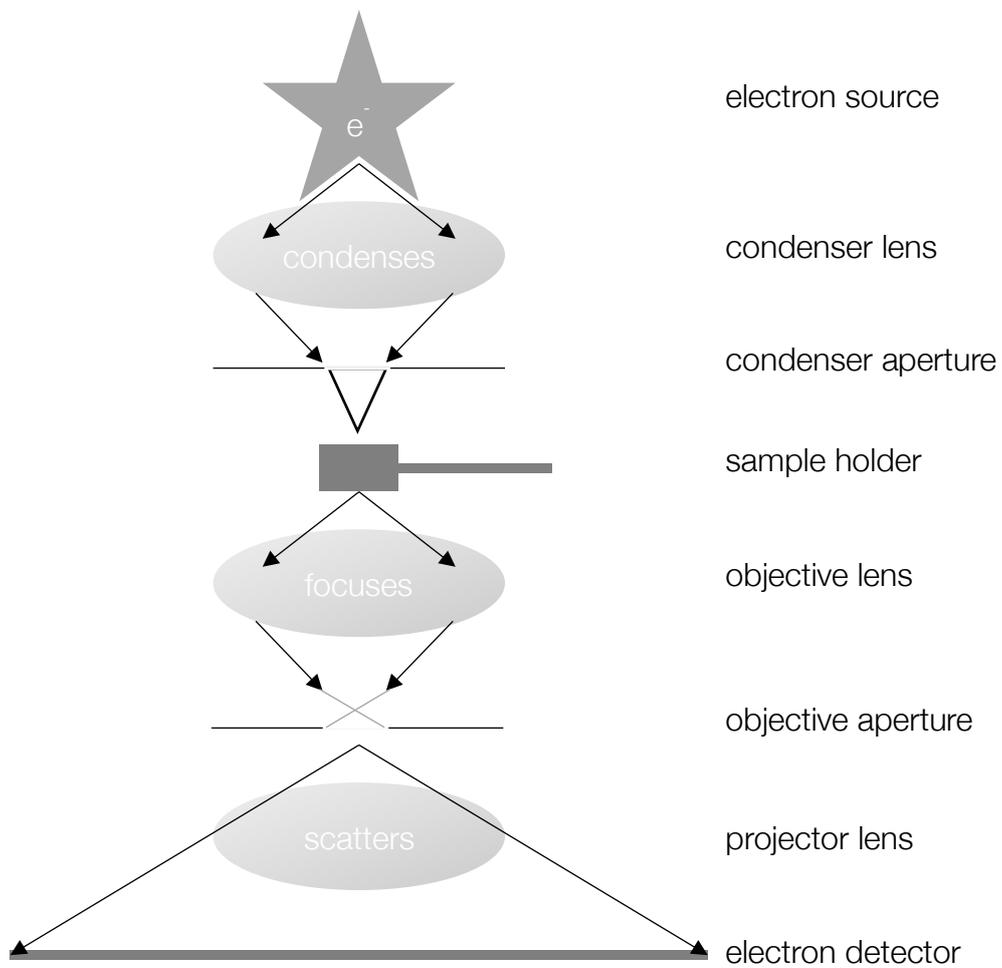
electron detector

Figure 2.5 Transmission Electron Microscopy Schematic

Electron source generates a current causing electron emission. Electron beam shown by arrows and thin grey lines. Electromagnetic field produces lenses. Scattered electrons detected, and image produced.

2.6.2.1 Negative Stain Electron Microscopy

Prior to sample loading onto an EM copper grid (usually ~10 nm thick carbon coated) the grids were glow discharged. This process coats the carbon surface of the grid with negatively charged ions making it hydrophilic and more amenable to sample loading. 2% uranyl acetate (UA) is a common stain that allows for the negative visualisation of the sample as heavy metal atoms absorb electrons and, is used to preserve the sample for TEM. UA staining creates a contrast between the absence of stain (where sample is) and the surrounding shell of heavy metal atoms. Positive stain can be observed whereby the sample loaded becomes stained, however this produces less informative micrographs (Orlova and Saibil, 2011). Moreover, stain can also induce artefacts and/or uranyl acetate crystals on the EM grid distorting sample staining contrast and structure. Despite this however, negative-stain EM is a rapid technique for the visualisation of sample, data generation and 2D reconstruction. If averaging and computational alignment is used, information can be increased, and 3D reconstructions determined (though this is still difficult and time consuming), however this is typically limited to 20 Å resolution. If high resolution 3D information is required then cryo-EM would normally be the next step, but again this is time consuming and extremely sample dependent. Negative-strain EM micrographs however can be analysed by single particle analysis or helical reconstruction. Single particle analysis was used in this thesis and in short relies on a protein and/or macromolecular complex adopting multiple conformations; these conformations are represented in the collected micrographs which contain multiple copies of the conformations at different orientations (Cheng *et al.*, 2015). Particles are selected from micrographs by defining the centre of the particle, the saved x,y coordinates extract information from a user-defined box surrounding the particle centre. Particles can be picked manually or automatically; manual particle picking is slow however provides greater control and confidence for input data. Whereas, automatic picking is high throughput but care must be taken not to introduce bias (He and Scheres, 2017). The extracted particle represents the single, angular orientation of a 3D macromolecular complex as a 2D projection. To produce a 3D reconstruction, alignments and classifications are performed by maximum likelihood; the user defines the number of classes. 2D classes are used to remove poor quality particles before 3D reconstruction (He and Scheres, 2017). A Fourier transform can then be used to produce a 3D structure (Cheng *et al.*, 2015).

Chapter 3 Materials and Methods

3.1 Materials

3.1.1 Recipes

Reagents from Sigma-Aldrich were used unless otherwise stated.

Table 3.1 General Recipes

Luria Broth (LB)	10 g tryptone, 5 g yeast, 10 g NaCl per 1 L
Agar	16 g tryptone, 10 g yeast extract, 5 g NaCl, 15 g agar, per 1 L supplemented with selective antibiotic (ampicillin (100 μg/mL), kanamycin (40 μg/mL) or chloramphenicol (50 μg/mL))
2xYT	16 g tryptone, 10 g yeast, 10 g NaCl per 1L
Coomassie Blue Stain	45% methanol (v/v), 45% H_2O, 10% acetic acid (v/v), 0.25% (w/v) Coomassie brilliant blue R-250
Coomassie Blue De-Stain	30% (v/v) methanol, 60% H_2O, 10% acetic acid (v/v)
SDS-Page Gel	15% resolving gel: 15 mL 30 % acrylamide mix (w/v), 7.5 mL 1.5 M Tris-Cl pH 8.8, 300 μL 10% SDS (w/v), 300 μL APS (v/v), 12 μL TEMED 5% stacking gel: 1.7 mL 30% acrylamide mix (w/v), 1.5 mL 1 M Tris-Cl pH 6.8, 100 μL SDS (w/v), 100 μL APS (v/v), 10 μL TEMED
SDS-Page Running Buffer	25 mM Tris-HCL pH 7.5, 0.1% (w/v) SDS, 190 mM glycine
SDS-Page Loading Dye	100 mM Tris, 4% (w/v), 20% (v/v) glycerol. 200 mM DTT, 0.3% (w/v) bromophenol blue pH 6.8
0.5 X Tris/Borate/EDTA (TBE) Buffer	40 mM Tris, 45 mM boric acid, 1 mM EDTA, pH 8
Agarose Gel	1% gel: 0.5 g agarose per 50 mL of TBE, 1 μL 10,000 X SYBR® Safe
RNA Binding Buffer	20 mM Tris pH 7.5, 150 mM NaCl, 0.01% (v/v) Triton X-100
Circular Dichroism Buffer	50 mM Na_3PO_4
Cryoprotectant	25% glycerol, 25% polyethylene glycol (PEG)-400, 25% 2-Methyl-2, 4-pentanediol (MPD), 25% ethylene glycol

Table 3.2 AKAV, OROV, SIMV NP Purification Buffers

Lysis Buffer	5% glycerol (w/v), 2% Triton X-100 (v/v), 10 mM MgCl$_2$, 200 mM NaCl, 20 mM Tris, 10 mM Imidazole, pH 8
Wash Buffer	200 mM NaCl, 20 mM Tris, 20 mM Imidazole, pH 8
High Salt Buffer	1 M NaCl, 20 mM Tris, 40 mM Imidazole
Elution Buffer	200 mM NaCl, 20 mM Tris, 150 – 500 mM Imidazole, pH 8
Dialysis Buffer	200 mM NaCl, 20 mM Tris, 10mM Imidazole, pH 8
Size-Exclusion Buffer	200 mM NaCl, 20 mM Tris pH 7.5, 1 mM DTT, 5% glycerol

Table 3.3 hRSV M2-1 Purification Buffers

Lysis Buffer	1 M NaCl, 25 mM Tris pH 7.5, 10 mM MgCl$_2$
GS4B Wash Buffer	150 mM NaCl, 25 mM Tris pH 7.5
Elution Buffer	150 mM NaCl, 25 mM Tris, 0.01% Triton X-100, 5 mM glutathione, pH 7
Cleavage Buffer	150 mM NaCl, 25 mM Tris pH 7.5, 0.01% Triton X-100
Diluting Buffer	25 mM Tris pH 7.5
SP Sepharose Wash Buffer	15 mM NaCl, 25 mM Tris pH 7.5
50 mM NaCl Elution Buffer	50 mM NaCl, 25 mM Tris pH 7.5
High NaCl Buffer	90 mM NaCl, 25 mM Tris pH 7.5
Size-Exclusion Buffer	150 mM NaCl, 25 mM Tris pH 7.5, 1 mM DTT, 5% glycerol

Table 3.4 hRSV P90-160 Purification Buffers

Lysis Buffer	150 mM NaCl, 25 mM Tris, 10 mM Imidazole, 10 mM MgCl2, pH 8
Wash Buffer	150 mM NaCl, 25 mM Tris, 20 mM Imidazole, pH 8
High Salt Buffer	1 M NaCl, 20 mM Tris, 40 mM Imidazole, pH 8
Elution Buffer	150 mM NaCl, 20 mM Tris, 150/300/500 mM imidazole, pH 8
Size-Exclusion Buffer	150 mM NaCl, 25 mM Tris pH 7.5, 1 mM DTT, 5% glycerol

Table 3.5 hRSV M Purification Buffers

Lysis Buffer	50 mM Na_3PO_4, 300 mM NaCl, 10 mM Imidazole, 10 mM $MgCl^2$, pH8
Wash Buffer	50 mM Na_3PO_4, 300 mM NaCl, 20/40/60/80 mM Imidazole, pH8
Elution Buffer	50 mM Na_3PO_4, 300 mM NaCl, 150/300/500 mM Imidazole, pH8
Dialysis Buffer	50 mM Na_3PO_4, 300 mM NaCl, 10 mM Imidazole, pH8
Size-Exclusion Buffer	50 mM Na_3PO_4, 300 mM NaCl, 1 mM DTT, 5% glycerol

Table 3.6 VP35 Purification Buffers

Lysis Buffer	500 mM NaCl, 50 mM Tris, 20 mM Imidazole, 0.1% triton, 10 mM $MgCl^2$, 15% glycerol, pH 8
Wash Buffer	500 mM NaCl, 50 mM Tris, 40 mM Imidazole, 15% glycerol, pH 8
High Salt Buffer	2M NaCl, 50 mM Tris, 80 mM Imidazole, 15% glycerol, pH 8
Elution Buffer	500 mM NaCl, 50 mM Tris, 150-500 mM Imidazole, 15% glycerol, pH 8
Amylose Column Buffer	300 mM NaCl, 20 mM Tris pH 7.5, 15% glycerol
Amylose Elution Buffer	300 mM NaCl, 20 mM Tris pH 7.5, 10 mM maltose
Size-Exclusion Buffer	500 mM NaCl, 25 mM Tris H 7.5, 1 mM DTT, 5% glycerol

Table 3.7 VP30 Purification Buffers

Lysis Buffer	200 mM NaCl, 20 mM Tris, 10 mM Imidazole, 0.1% triton, 10 mM $MgCl^2$, 5% glycerol, pH 8
Wash Buffer	200 mM NaCl, 20 mM Tris, 20 mM Imidazole, pH 8
High Salt Buffer	1M NaCl, 20 mM Tris, 40 mM Imidazole
Elution Buffer	200 mM NaCl, 20 mM Tris, 150/300/500 mM Imidazole pH 8
Size-Exclusion Buffer	200 mM NaCl, 20 mM Tris, 1 mM DTT, 5% glycerol, pH 8

3.1.2 Consumables

3.1.2.1 Vectors

pET28a-SUMO (pET-SUMO) – pET28a (Novagen) derivative with an N-terminal 6-histidine-SUMO (His6-SUMO) purification tag. The His6-SUMO tag is cleaved by 1P SUMO protease (Ubl-specific protease 1 from Saccharomyces cerevisiae) which recognises the tertiary structure of SUMO.

pGEX-6P-2 (GE Healthcare) – PGEX (GE Healthcare) derivative with an N-terminal glutathione S-transferase (GST) purification tag. The GST purification tag is cleaved by PreScission protease at the cleavage site: Leu-Glu-Val-Leu-Phe-Gln*Gly-Pro (GST-tagged human rhinovirus 3C protease, GE Healthcare). Cleavage occurs between Gln and Gly leaving two non-native amino acids in the recombinant protein expressed and purified.

pET-MAL-28a-Precision (pET-MAL) – pET28a (Novagen) derivative with an N-terminal 6-histidine-maltose binding protein (MBP) purification tag. PreScission protease cleavage site.

cDNA of AKAV, OROV and SV NP was synthesised by Genewiz in pUC57 and sub-cloned using restriction sites BamHI and XhoI into pET-SUMO

hRSV A2 strain M2-1 open reading frame was gifted from Gail Wertz and sub-cloned into pGEX6P2 by Sian Tanner with 5 additional non-native residues (Gly-Pro-Leu-Gly-Ser) which remain after cleavage. cDNA of P90-160 was sub-cloned into a pET28a modified vector to include a non-cleavable C-terminal 6-histidine tag. cDNA of M (A2 strain) was synthesised and cloned into pET-SUMO by Genewiz.

cDNA of EBOV and MBOV VP30 and VP35 was custom synthesised by Genewiz flanked by BamHI and XhoI restriction sites. Constructs were sub-cloned using restriction sites BamHI and XhoI (NEB) into all 3 of the aforementioned vectors.

Nucleotide sequences were confirmed by DNA sequencing (Genewiz).

3.1.2.2 Bacterial Strains

Plasmid DNA was transformed into DH5α competent cells (details in section 3.2.1) (Life Technologies) for amplification or BL21 (DE3) derivatives (Table 3.8) for recombinant protein expression inducible by isopropyl β-D-1-thiogalactopyranoside (IPTG), which included Gold, Star, Rosetta, CodonPlus and Lemo21 cell strains.

Table 3.8 Bacterial Strain Genotypes

E. coli Competent Cells	Features	Genotype
DH5α	Endonuclease I (endA gene) inactivated (prevents plasmid DNA degradation) recA1 inactivated from of RecA to reduce RNA recombination leackage and cell lysis gyrA96 mutation gives ccdB resistance relA1 differs lipid structure to reduce protein	E. coli B F⁻ φ80lacZΔM15 Δ(lacZYA-argF)U169 recA1 endA hsdR17(r_K^-, m_K^+) phoA supE44 λ⁻thi-1 gyrA96 relA1
Gold	E. coli phenotype increases transformation efficiency Endonuclease I (endA gene) inactivated (prevents plasmid DNA degradation) lon gene naturally lack and outer membrane (OmpT gene) protease deletion (reduction in protein degradation) dcm+ unable to methylate cytosine hsdS(r_B^- m_B^-) DNA methylation deficient	E. coli B F⁻ ompT hsdS_B(r_B^- m_B^-) dcm⁺ Tet^r gal endA E. coli
Star	RnaseE gene (rne131) mutation (reduces endogenous Rnases and mRNA degradation) lon gene and outer membrane (OmpT gene) protease deletion (reduction in protein degradation) hsdS_B(r_B^- m_B^-) DNA methylation deficient	E. coli B F⁻ompT hsdS_B (r_B^-, m_B^-) gal dcm+ (DE3)
Rosetta	Supply tRNAs for AGG, AGA, AUA, CUA, CCC and GGA codons in chloramphenicol-resistant plasmid dcm+ unable to methylate cytosine pRARE chlormaphencol resitance	E. coli B F⁻ ompT hsdS_B(r_B^- m_B^-) gal dcm+ (DE3) pRARE (Cam^R)
CodonPlus	E. coli phenotype increases transformation efficiency endA gene inactivated outer membrane (OmpT gene) protease deletion (reduction in protein degradation)	E. coli B F⁻ ompT hsdS_B(r_B^- m_B^-) dcm+ Tet^r gal λ(DE3) endA E. coli [argU proL BB Cam^r] [argU ileY leuW Strep/Spec^r]

	dcm+ unable to methylate cytosine	
Lemo21	Tuneable expression via variation of lysozyme level (lysY) controlled by chloramphenicol resistance outer membrane (*OmpT* gene) protease deletion (reduction in protein degradation)	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS/ pLemo*(Cam^R)

3.1.2.3 RNA synthesis

Oligoribonucleotides (unlabelled and 3'fluorescein labelled) were synthesised by Dharmacon (GE Healthcare). The oligoribonucleotides were desalted and high-performance liquid chromatography (HPLC) purified (Dharmacon, Horizon Discovery), prior to resuspension in nuclease free water in RNase free tubes at 1 mM stock concentration.

3.1.2.4 Primers

Primers were synthesised by Sigma-Aldrich and dissolved in nuclease free water at 100 μM stocks.

3.2 Methods

3.2.1 Transformation of Competent Escherichia coli with Plasmid DNA

50 μL of competent *E. coli* were transformed with 1 μL of 50 ng/μL of plasmid DNA containing a selective antibiotic resistance gene (ampicillin or kanamycin). Samples were incubated on ice for 20 minutes (min) and heat-shocked at 42°C for 45 seconds (sec) before returning to ice for 2 min. 500 μL of sterile LB was added before recovery at 37°C for 45 min with orbital shaking at 180 revolution per minute (rpm) (Stuart SI5000). Transformed E. coli were plated out onto an antibiotic selective agar plate and incubated at 37°C for 16 hours. BL21 (DE3) Gold, Rosetta, Rosetta II, Codon Plus, Star and Lemo21 for protein expression. Note: Lemo21 cells contain a second plasmid for chloramphenicol resistance.

3.2.2 Bacterial Starter Cultures

5 mL of sterile LB supplemented with 5 μL of 100 mg/mL selective antibiotic was inoculated with a single colony and incubated at 37°C for 16 hours with orbital shaking 180 rpm (Stuart SI5000).

3.2.3 Plasmid DNA Extraction from Escherichia coli

5 mL starter cultures were pelleted at 4000 rpm (Fisher Scientific accuSpin 1R) for 5 minutes. Bacterial cell pellets were lysed, and DNA extracted using Qiagen Miniprep Kit to the manufactures protocol.

3.2.4 Optical Density$_{600}$

A referenced spectrophotometer was used to measure cell culture optical density (OD) using 1 cm cuvettes at 600 nm.

3.2.5 Induction of Protein Expression

For small scale expression trials, 1 mL of overnight starter culture (16 hours growth) was used to inoculate 10 mL of 2xYT for protein expression. For large scale expression, 10 mL of starter culture was used to inoculate 1 L of 2xYT. Protein expression was induced at OD$_{600}$ 0.6-0.8, with specified concentrations of IPTG for each individual protein (typically 0.5 mM), and growth temperature reduced from 37°C to 18°C.

3.2.6 Small Scale Expression

10 mL expression culture was pelleted at 4000 rpm (Fisher Scientific accuSpin 1R) for 5 minutes. Bacterial cells were lysed with 500 μL of Bacterial Protein Extraction Reagent (B-PER) (Thermo Fisher Scientific) and 3 rounds of freeze-thawing (FT). Soluble fractions were collected by centrifugation at 13,000 rpm for 10 minutes. Insoluble pellet was resuspended in 500 μL of 0.1% triton and phosphate buffered saline (PBS)

3.2.7 Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE gels were hand casted using a Mini-PROTEAN tetra cell casting stand and clamp (Bio-Rad), the thickness of the gel was 1.5 cm and was left to polymerise at room temperature for 45 minutes. Sample was mixed with an equal volume of loading dye before boiling at 90°C for 2 minutes. 10 μL of boiled sample was loaded per well (SDS-PAGE were run at 180 volts (V) in 1 X SDS running buffer for 45 minutes before staining in Coomassie blue stain for 1 hour and de-stained for 40 minutes.

3.2.8 Agarose Gel Electrophoresis

Agarose gels were hand casted using a cassette and masking tape. The agarose gel buffer was heated in a 700-watt microwave for 10 seconds to help the agarose dissolve. SYBR® safe was added once cooled prior to setting in the cassette. 5 μL of sample mixed with 1 μL 6 x purple loading dye (NEB). 3 μL of sample:dye mix was loaded and the agarose gel was run at 100 V in TBE buffer for 90 minutes or until sufficient separation.

3.2.9 Agarose Gel Purification

DNA was visualised (Safe Imager, Invitrogen) and extracted using Qiagen DNA Extraction Kit to the manufactures protocol.

3.2.10   Cloning

3.2.10.1   Q5 Site Directed Mutagenesis for SV NP E Mutants

For glutamic acid (E) mutations into SV NP, a Q5 site directed mutagenesis (SDM, NEB) was used to the manufacture protocol. Primers were designed using the NEB online tool with nucleotide changes represented in lower case in Table 3.10.

Table 3.9 Q5 Site Directed Mutagenesis Primers for SV NP Mutations

Mutation		3'-5' Forward Primer	3'-5' Reverse Primer
K51E		SV NP K51E F	SV NP K51E R
K = AAA	E = GAA	GAAGAAAGCGgAAATGGTGCTG	CTGGTTCAGGAAGAACACAC
N77E		SV NP N77E F	SV NP N77E R
N = AAC	E = GAG	GGTGAACAACgAgTTTCCGCAGTA	AGGGTGAACTTCACACCA
R95E		SV NP R95E F	SV NP R95E R
R = CGT	E = GAA	GACTCTGCACgaaCTGTCCGGCTAC	AGAGCGTTATCCGGAACT
R182E		SV NP R182E F	SV NP R182E R
R = CGT	E = GAA	CAAAGCTCTGgaaCAGCGTTACG	CGCATAAAGTTCACGTCC
Q183E		SV NP Q183E F	SV NP Q183E R
Q = CAG	E = GAG	AGCTCTGCGTgAGCGTTACGGTC	TTGCGCATAAAGTTCACGTC

3.2.10.2   DNA Amplification

Polymerase Chain Reaction (PCR) was performed with a total reaction volume of 45 μL: 0.25 μL Phusion HF Polymerase, 5 μL of 10 μM forward primer, 5 μL of 10 μM reverse primer, 5 μL dNTPs mix (25 mM of each), 10 μL Phusion Buffer, 1 μL DNA template (100 ng/μL).

Table 3.10 PCR Design

Stage 1	Stage 2 x 20 cycles	Stage 3 x 20 cycles	Stage 4
94°C (2 m)	94°C (30 s) *Primer annealing temperature specific (30 s)	94°C (30 s) 70°C (30 s) 72°C (60 s)	72°C (30 s per kB)

| | 72°C (60 s) | | |

PCR products were visualised on a 1% agarose gel and correct size inserts purified using a Qiagen Gel Extraction or Qiagen PCR Purification kit (Qiagen) to the manufactures protocol.

3.2.10.3   Restriction Digests

1 μg plasmid containing no insert or PCR product was digested with 1 μL of BamHI and XhoI (NEB) with 1 x Cut Smart Buffer (NEB) to a final reaction volume of 50 μL for 3 h at 37°C to create complementary sticky ends. Cut plasmid and PCR product (insert) were agarose gel purified using Qiagen Gel Extraction Kit to the manufactures protocol.

3.2.10.4   Ligations

Insert and plasmid were agarose gel purified and ligated using 5 μL of instant sticking end master mix (NEB) with a 3:1 ratio and a minimum of 50 ng of DNA to a final volume of 10 μL.

3.2.10.5   VP30 and VP35 Truncations

EBOV VP30$_{8-272}$, full-length VP35 (1-340) (Appendix 9.1.1) and MBOV VP30$_{15-281}$, and full length VP35 (1-329) (Appendix 9.1.3) coding sequences were codon optimised for expression in E. coli and purchased from GeneWiz in pUC57-kanamycin cloning vector with BamHI and XhoI restriction endonuclease sites. Each plasmid was PCR amplified, cut with BamHI and XhoI and ligated into PGEX6P2, pET-SUMO, and pET-MAL prior to transformation into DH5α. 8 colonies were picked for analysis of successful ligation via restriction digest. 5 mL of starter culture plasmid DNA was extracted using Qiagen Miniprep Kit to the manufacturers protocol. Restriction products were analysed on a 1% agarose gel. Secondary analysis was performed via DNA sequencing (GeneWiz). Correctly sub-cloned plasmids were transformed into expression strains,).

3.2.10.5.1 Primer Design

Primers were designed to amplify EBOV and MBOV VP30 and VP35 DNA prior to cloning.

Table 3.11 VP30 and VP35 Primer Design

Truncation	3'-5' Forward Primer	3'-5' Reverse Primer
EBOV VP30 55 -265	EBOV VP30 55-265 F	EBOV VP30 55-265 R
	AATTGGATCCCTGACCGTTCC GCCTG	AATTCTCGAGCTACAGGGTACG CAGACCAC
EBOV VP30 87 -265	EBOV VP30 87-265 F	EBOV VP30 87-265 R
	AATTGGATCCAGCCTGACCGA TCGCCT	AATTCTCGAGCTACAGGGTACG CAGACCAC
	MBOV VP30 17-273 F	MBOV VP30 17-273 R

MBOV VP30 17-273	AATTGGATCCCCGACCATCTACCATGAAAC	AATTCTCGAGTTACGGCAGGATGAAGTGATC
MBOV VP30 79-273	MBOV VP30 79-273 F	MBOV VP30 79-273 R
	AATTGGATCCCGCAATATGAAAATTGGCC	AATTCTCGAGTTACGGCAGGATGAAGTGATC
MBOV VP30 101-273	MBOV VP30 101-273 F	MBOV VP30 101-273 R
	AATTGGATCCCTGACCAATCGTGAGCTG	AATTCTCGAGTTACGGCAGGATGAAGTGATC
MBOV VP35 60-329	MBOV VP35 60-329 F	MBOV VP35 60-329 R
	ATTGGATCCGATATTGTTTGGGACCAACTGATCG	AATTGAGCTCTTAGATTTTCAGGGCC

3.2.10.5.2 Cloning of VP30 Truncations

Purified PCR product plasmid DNA was restriction digested and ligated prior to transforming DH5α. Colonies from each ligation plate were picked for starter cultures. DNA was extracted via Qiagen miniprep to the manufacturer protocol and digested with BamHI and XhoI for 3 h at 37°C prior to analysis via 1% agarose gel. Nucleotide sequence was confirmed by DNA sequencing (GeneWiz) prior to transformation into expression strains.

3.2.11   Large Scale Purification

3.2.11.1.1 Large Scale Protein Expression

1 L of LB was inoculated with 10 mL of starter culture and selective antibiotic (100 mg/mL ampicillin, 40 mg/mL kanamycin, 50 mg/L chloramphenicol). Cells were incubated at 37°C at 180 rpm (Infors HT Multitron) until OD_{600} 0.6-0.8 was reached. Cells were induced for protein expression with 500 mM IPTG with the addition of 50 μM $ZiSO_4$ for VP30 and M2-1, and 0-2000 μM L-rhamnose when using Lemo21 cells. Cells were incubated at 18°C 16 h with orbital shaking. Cells were pelleted at 5000 x g (Fisher Scientific accuSpin 1R) for 15 minutes. Pellets were stored at -80°C until protein purification.

3.2.11.1.2 Cell Lysis

1 cell pellet (1L of cell growth) was resuspended with 10 mL of lysis buffer and the addition of, 1 protease inhibitor (cOmplete, EDTA-free tablet protease inhibitor cocktail, Merck, Sigma-Aldrich), and 500 μL of B-PER (Thermo Fisher Scientific), 20 μL of RNase at 10 mg/mL, 1 μg/mL Dnase 1 and 100 mg of chicken egg white lysozyme. Cells were lysed using high pressure homogenisation at 1000 psi (Avestin C3 Cell Disruptor, ATA Scientific). The addition of 0.25% w/v 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate hydrate (CHAPS) to M lysate was incubated for 10 minutes before centrifugation. Lysates were centrifuged at 43000 x g for 1 hour at 4°C (Sorvall Evolution, SS-34)

3.2.11.1.3 Histidine Affinity Chromatography (AKAV, OROV and SIMV NP WT and mutants, hRSV P$_{90\text{-}160}$, EBOV and MARV VP35 and VP30)

Soluble bacterial supernatant was syringe filtered (Ministart 0.45 μM Sartorius Stedim Biotech) applied to a pre-equilibrated pre-packed 5mL HisTrap HP column (GE Healthcare) attached to a peristaltic pump. The pump was used at room temperature and run at 1 mL/minute. The column was washed with 50 mL wash buffer, 50 mL high salt wash buffer (10 column volumes if appropriate), and 25 mL of each elution buffer (five column volumes, 150-500 mM imidazole). 25 mL eluate from each step was collected. Typically, in the Edwards/Barr group proteins eluted with 300 and 500 mM imidazole using the HisTrap HP columns. Fractions containing protein of interest that had a 260:280 ratio below 1 (indicating RNA free protein) were pooled, cleaved from the purification and dialysed to remove imidazole. hRSV P$_{90\text{-}160}$ fractions from the first affinity chromatography step were concentrated to 10 mL for SEC as no cleavage was necessary. For VP30 cleavage caused aggregation and precipitation of protein therefore cleavage was not performed, and the fusion protein was subjected to SEC, but dialysis still occurred to remove the high concentrations of imidazole.

3.2.11.1.4 Glutathione S-transferase Tagged Affinity Chromatography (hRSV M2-1)

Soluble bacterial supernatant was syringe filtered (Ministart 0.45 μM Sartorius Stedim Biotech) applied to a pre-equilibrated column containing Glutathione Sepharose Superflow (GS4B) resin (GE Healthcare). GST-tagged protein was left to bind to GS4B resin for 1 hour at 4°C with agitation. The resin was washed with 50 mL of lysis buffer followed by 50 mL of wash buffer; 25 mL of elution buffer was used to during on-column cleavage; 500 μL of 6 mg/mL PreScision Protease (made by Sue Matthews, technician in the Edwards group) was used for cleavage at 4°C for 16 hours. Cleaved eluted protein was found in the flow through. Cleaved protein was diluted from 150 mM to 50 mM NaCl with dilution buffer for ion exchange chromatography.

3.2.11.1.5 Ion Exchange Chromatography (hRSV M2-1)

Diluted protein was applied to a pre-equilibrated SP Sepharose column. A 2,5 x 10 cm glass Econo-Column® chromatography column was used with 5 mL of SP Sepharose beads. The column was washed with 15 mM NaCl SP Sepharose wash buffer and eluted with 60 – 90 mM NaCl high salt buffer. Eluates with protein of interest determined by SDS-PAGE were concentrated to 10 mL with a 10 kDa molecular weight cut-off centrifugal concentrator (Merk Millipore) at 4000 rpm (Fisher Scientific, aacuSpin 1R) and syringe filtered (0.22 μM Millex) prior to SEC purification.

3.2.11.1.6 Size Exclusion Chromatography

Syringe filtered (0.22 μM Millex) and de-gassed SEC buffer was used to pre-equilibrate a 330 mL HiLoad 26/60 Superdex 75 pg column (S75, GE Healthcare), HiLoad 26/60 Superdex 200 pg (S200, GE Healthcare) or a Sephacryl S-400 High Resolution column (S-400, GE Healthcare) at 4°C with an AKTA Prime pump liquid chromatography system measuring absorbance at 280 nm. Filtered protein was loaded into a 10 mL loop and

flowed through the column at 2 mL/min (or 1.3 mL/min for S-400) collecting 3 mL fractions after the initial 90 mL void volume. Fractions containing pure protein were concentrated, flash frozen in liquid nitrogen and stored at -80°C.

Protein concentration was measured using a NanoDrop One Microvolume UV-Vis Spectrophotometer (ThermoFisher Scientific), calculated from absorption at 280 nm and calculated extinction coefficients for each protein from ProtParam (Gasteiger *et al.*, no date).

3.2.12   Protein Analysis

3.2.12.1   Mass Spectrometry

Native mass spectrometry was used to determine the oligomeric state of protein. Pure protein was buffer exchanged into 10 mM ammonium acetate pH 7.5. Buffer exchange occurred using a 5 mL HiTrap Desalting column prepacked with Sephadex G-25 to the manufactures protocol (GE Healthcare). 200 μL of sample at 1 μM was given to the mass spectrometry facility at University of Leeds. Native mass spectrometry was performed as a service by facility manager Dr. James Ault and technician Rachel George using a nano-electrospray-ionization mass spectrometry with an LCT Premier mass spectrometer. Samples were ionised using a NanoMate (Advion) chip with ionization voltage at +1.75 kV and 30V sample cone voltage.

3.2.12.2   Circular Dichroism

Circular dichroism (CD) experiments were performed on a Chirascan-plus (Applied Photophysics) spectopolarimeter purged with nitrogen gas as a service by Dr. Nasir Khan (University of Leeds).. Pure protein sample was diluted to 0.1 mg/mL and buffer exchanged into CD buffer. Buffer exchange occurred using a 5 mL HiTrap Desalting column prepacked with Sephadex G-25 to the manufactures protocol (GE Healthcare). Sample temperature was controlled using a Peltier system and data collected as ellipticity θ in units of millidegrees. To determine to mean residue ellipticity the following equation was used

$$\text{Ellipticity } \theta = \frac{\text{raw signal}}{\text{protein concentation x (peptide bonds} - 1)\text{x pathlength (m)}}$$

3.2.13 Structural Studies

3.2.13.1 Crystallisation

3.2.13.1.1 Sparse Matrix Screens

Initial crystallisation trials were set up using commercially available sparse matrix screens JCSG Core 1-4 (Hampton Research) using an NT8 liquid handling robot (Formulatrix) in MRC 96-well plates (Molecular Dimensions) with 25μL reservoir solution. 0.1 μL drops at 50:50 ratio protein:reservoir were made with protein at 10-20 mg/mL. Plates were sealed with Viewseal pressure adhesive transparent seals (Grenier Bio-One) and incubated at 18°C.

3.2.13.1.2 Crystal Optimisation

Initial trials were monitored from 3-21 days for crystal growth at 18°C. Hits were optimised in 24-well plates (Hampton Research) using hanging-drops. 1 μL of protein was diluted with 1 μL of reservoir buffer on a 22 mm square glass cover slide (Jena Bioscience). Each well was filled with 999 μL of reservoir and the glass cover slide placed over the well, sample side down. The well was sealed with silicon grease (Jena Bioscience). Reservoir conditions varied precipitant concentration and buffer pH.

3.2.13.1.3 Crystal Harvesting

Crystals were picked with appropriately sized nylon-loops (Hampton Research) under a light microscope. The loops were then Immersed in reservoir buffer with added cryoprotectant prior to cryo-cooling in liquid nitrogen. The crystal loops were stored in a cryo-cooled universal puck (MiTeGen) and shipped to the DLS in a dewar (CX100 Taylor-Wharton) filled with liquid nitrogen.

3.2.13.1.4 Data Collection

X-ray diffraction data was collected on beamline I04 at the Diamond Light Source (DLS) at 100 K. Crystals were initially screened for diffraction using 3 images between 0-90° using 12,658 eV X-rays (0.9795 Å) before data collection from 0-360°. Xia2 DIALS (Diffraction Integration for Advanced Light Sources) (Grosse-Kunstleve *et al.*, 2002; Evans, 2006; Winter *et al.*, 2018), was used to integrate X-ray diffraction data and the autoprocessed .mtz file was downloaded from ISPyB (Delageniere *et al.*, 2011).

3.2.13.1.5 Structure Determination and Refinement

Bunyavirus NP monomer (unpublished data) was used to solve the structure of SV NP apo-protein and co-complex with RNA using CCP4i2 (Potterton *et al.*, 2018) via molecular replacement (Phaser) (McCoy *et al.*, 2007). Refinement utilised manual model building and COOT (Emsley and Cowtan, 2004) and REFMAC5 (Murshudov *et al.*, 2011).

3.2.13.2 Negative Stain Electron Microscopy

3.2.13.2.1 Glow Discharge

Carbon-coated EM grids were made by technician Martin Fuller (Astbury Biostructure Laboratory, University of Leeds) and were rendered hydrophilic prior to sample application by glow discharging (PELCO easyGlow). EM grids were placed carbon side up onto a glass slide coated in parafilm prior to glow discharging for 45 seconds.

3.2.13.2.2 Grid Preparation

5 μL of sample was applied to an EM grid for 1 minute. Excess sample was blotted using filter paper (Whatman, No.1). EM grids were washed twice firstly with 10 μL of dilution of buffer and then a water wash. Blotting was performed in between. 10 μL of 2% of uranyl acetate stain was applied for 30 sec and blotted. Grids were left to air-dry.

3.2.13.2.3 Micrograph Acquisition

EM grids were loaded into a single tilt holder and inserted into a FEI Tecnai T12 microscope with an LaB$_6$ electron source at 120 kV. A Gatan US4000/SP CCD detector was used. A magnification of x 69,000 (69 K). Micrographs were collected with one second exposures at -1.5 μM defocus. For 2C classification, a series of micrographs were collected, the microscope was refocused and defocused back to -1.5 μM and the field of view moved between each micrograph collected to ensure all particles were unique.

3.2.13.2.4 Micrograph Processing

Micrograph processing of negative stain EM data was analysing using RELION3 software (Zivanov *et al.*, 2018). Initial particle picking occurred manually generating 100 particles prior for reference-free 2C classification. The most abundant classes were selected and used as a reference for auto-picking of 1000 particles for a final set of 2D classes.

3.2.14 Fluorescence Anisotropy

3.2.14.1.1 Direct Binding

Binding experiments were carried out in triplicate in a black 384-well optiplate (Perkin Elmer). 20 μL of RNA binding buffer was added to each well. 40 μL of protein at 1 mg/mL was titrated across each row of 24. 20 μL of 3' fluorescein labelled RNA (pA/C/U 10 mer) at 5 nM or RNA binding buffer (controls) was also added. The plate was incubated for 45 minutes prior to data collection on a Spark 10M Multimode Microplate Reader (Tecan) with a 485nm (2 m, bandwidth) excitation filter and parallel (S) and perpendicular (P) channel emission filters at 535 nm (25 nm bandwidth). Anisotropy values were calculated using the raw emission values for S and P signals with the following equations

$$Intensity = (2 \, x \, P) + S$$

$$Anisotropy = \frac{S - (Px1)}{intensity}$$

All graphs were plotted using OriginPro 9.1 (Origin Lab) according to:

$$y = \frac{A1 - A2}{1 + \left(\frac{\chi}{\chi 0}\right) p} + A2$$

Where y is the fraction of RNA bound, χ is the protein concentration, A1 is the initial anisotropy value, A2 is the final anisotropy value, $\chi 0$ is dissociation constant and p is the Hill coefficient.

Chapter 4 Structure and Function of the Nucelocapsid Protein from *Orthobunyaviruses*

4.1   Chapter Introduction

The *Orthobunyavirus* genus contains highly virulent pathogens with no FDA-approved therapeutic available yet. Outbreaks are common across Europe (e.g. SBV) and Japan (AKAV) causing great economic loss highlighting the importance for a better understanding of viral transcription and replicative machinery in order to develop new approaches for therapeutics.

SIMV has a tripartite genome that encodes 5 proteins. The viral proteins carry out several functions but still require the interaction, modulation and/or hijacking of host cell machinery for virus life cycle completion. The segmented RNA is encapsidated by the viral NP for protection and immune evasion, making NP an antiviral target to be taken into account. Uncoating of RNA segments and coating of newly synthesized RNA by NP occurs in the Golgi apparatus. This process is likely to be transient in order to maintain RNA protection. Interestingly, the BUNV:RNA structure showed host cell RNA to be buried deep within an narrow cleft that would prevent nucleotide base interactions via Watson-Crick base pairing (Ariza *et al.*, 2013). Questioning how the pan-handle RNP architecture is maintained within the host cell and virions. Moreover, it was previously shown that BUNV NP mutant R94A (found at the rim of the RNA-binding groove) was unable to bind RNA, with poor activity in RNA replication and mRNA transcription (Ariza *et al.*, 2013).

In this chapter, X-ray crystallography was used to elucidate NP:RNA interactions to decipher the role of functionally important NP residues. The previously published structure for the related virus BUNV, was not at a high enough resolution to distinguish the identity of RNA bases. This was thought to be due to the RNA electron density being both 4-fold averaged due to the 4-fold symmetry of the BUNV NP tetramer, and derived from the recombinant E. coli host (Ariza *et al.*, 2013). Novel hydrogen bond and charge interactions were discovered and validated using fluoresce anisotropy RNA binding assays.

4.1.1 Objectives

Viruses require NP to encapsidate their RNA for protection against host cell RNA degradation factors and immune evasion; the polymerase also needs encapsidated genome as a template and will not recognise naked RNA. This chapter focuses on three closely related *Orthobunyaviruses* AKAV, OROV and SIMV. We aimed to elucidate the binding properties of NP from the *Orthobunavirus* genus using X-ray crystallography. High resolution structures will give insights into how NP from the *Orthobunyavirus* family recognise RNA with no apparent sequence specificity, as every base in the entire genome must be encapsidated. Structural studies focused on AKAV and SIMV NP only. The Edwards/Barr group have unpublished data for OROV NP and OROV NP in complex with pC10 RNA from previous projects, it was advised that efforts are focused on NP from AKAV and SIMV.

The second aim of this chapter was to validate and assess key amino acids that interact with RNA bases in the crystal structure and elucidate their contributions towards RNA binding. The OROV NP mutants had been previously made by Daniel Ven (undergraduate student, University of Leeds) prior to this project commencing, we therefore used OROV and SIMV for binding assays, due to the SIMV NP pA20 structure providing the greatest confidence for the interactions with RNA.

Note: Cloning of expression vectors for NP (AKAV, OROV, SIMV and SIMV mutants) worked on in this chapter were performed by Daniel Van (undergraduate student, University of Leeds). OROV NP mutagenesis was performed by Sue Matthews (Edwards group technician, University of Leeds). Sample preparation for AKAV NP was completed by Christopher Smith (PhD student, University of Leeds).

4.2 Results

4.2.1 NP Expression and Purification

4.2.1.1 Expression Vector Cloning

Codon-optimised complementary DNA (cDNA) sequence encoding AKAV NP, OROV NP and SIMV NP was cloned into pET-SUMO-28a vectors for expression of the viral protein with an N-terminal His6-SUMO tag. Production and validation of the cloned insertion inside the backbone vector was checked by Sanger sequencing and performed by Daniel Van (undergraduate student, University of Leeds). The vector was transformed into competent cells and grown for plasmid production (DH5α cells). Amino acids to mutate from OROV NP were decided by Professor Thomas Edwards (group leader, University of Leeds) based on known amino acids to interact with RNA from the unpublished OROV NP pC10 structure. Amino acids from SIMV NP to mutate was based on known interactions from the SIMV NP pA20 structure (discussed in 4.2.3), and a selection chosen across the length of the RNA binding groove (Figure 4.7 B). The mutants are listed in Table 4.1 and all NP expressed and purified in this thesis utilised the His6-SUMO purification tag as all were cloned into pET-28a-SUMO (Table 4.1).

Table 4.1 *Orthobunyavirus* NP Constructs

Construct		
AKAV NP	OROV NP	SIMV NP
WT	WT	WT
	K50A	K51E
	K47A	N77E
	R94A	R95E
	R180A	R182E
	R181A	Q183E
	R182A	

4.2.1.2 Expression

NP from AKAV, OROV and SIMV expressed at high yields reproducibly apart from mutant OROV and SIMV NP K50A and K51E respectively.

4.2.1.3 Affinity Chromatography

Initial purification was performed via affinity chromatography (Figure 4.1 A). Fusion-protein (His6-SUMO-NP WT or mutant) was eluted with increasing concentrations of imidazole. Eluates were tested to ensure a 260/280 ratio was below one, an indication of RNA-free protein (Table 4.2). Fractions containing RNA free protein were pooled (300 mM and 500 mM imidazole eluates). SUMO-protease was allowed to cleave for 16 hours before concentrating the protein mix for SEC.

Table 4.2 260/280 Ratio for NP Eluates

	260/280	
Protein	300 mM Imidazole Eluate	500 mM Imidazole Eluate
AKAV NP WT	0.65	0.53
OROV NP WT	0.67	0.65
OROV NP K47A	0.49	0.54
OROV NP R94A	0.44	0.43
OROV NP R180A	0.64	0.55
OROV NP R181A	0.66	0.70
OROV NP R182A	0.59	0.49
SIMV NP WT	0.67	0.44
SIMV NP N77E	0.59	0.62
SIMV NP R95E	0.63	0.63
SIMV NP R182E	0.60	0.42
SIMV NP Q183E	0.60	0.41

4.2.1.4 Size Exclusion Chromatography

To remove contaminants such as the purification tag, size-exclusion chromatography was performed (Figure 4.1 C) and analysed by SDS-PAGE on a 15% gel (Figure 4.1 B). All *Orthobunyavirus* NP proteins purified here were purified in the same SEC buffer Multiple peaks were seen for each protein with the largest peak eluting at ~175 mL. Pure tetrameric protein was determined by a single peak at 175 mL and SDS-PAGE, with a single band for NP (25 kDa) in the lane. Protein was aliquoted and flash frozen in liquid nitrogen prior to storing a -80˙C. Expression and purification of NP produced protein reproducibly and did not require further optimisation.

4.2.1.5 Summary

NP from AKAV, OROV and SIMV including mutants expressed well. A 2-step purification technique was applied to produce pure protein. The initial affinity chromatography step utilised the His6-fusion tag that bound to beads charged with nickel within the HisTrap column. SEC separated out any remaining contaminants and the fusion tag based on size. A peak eluting at 175 mL according to the calibartation contained protein tetramers and were seen for all NP proteins.
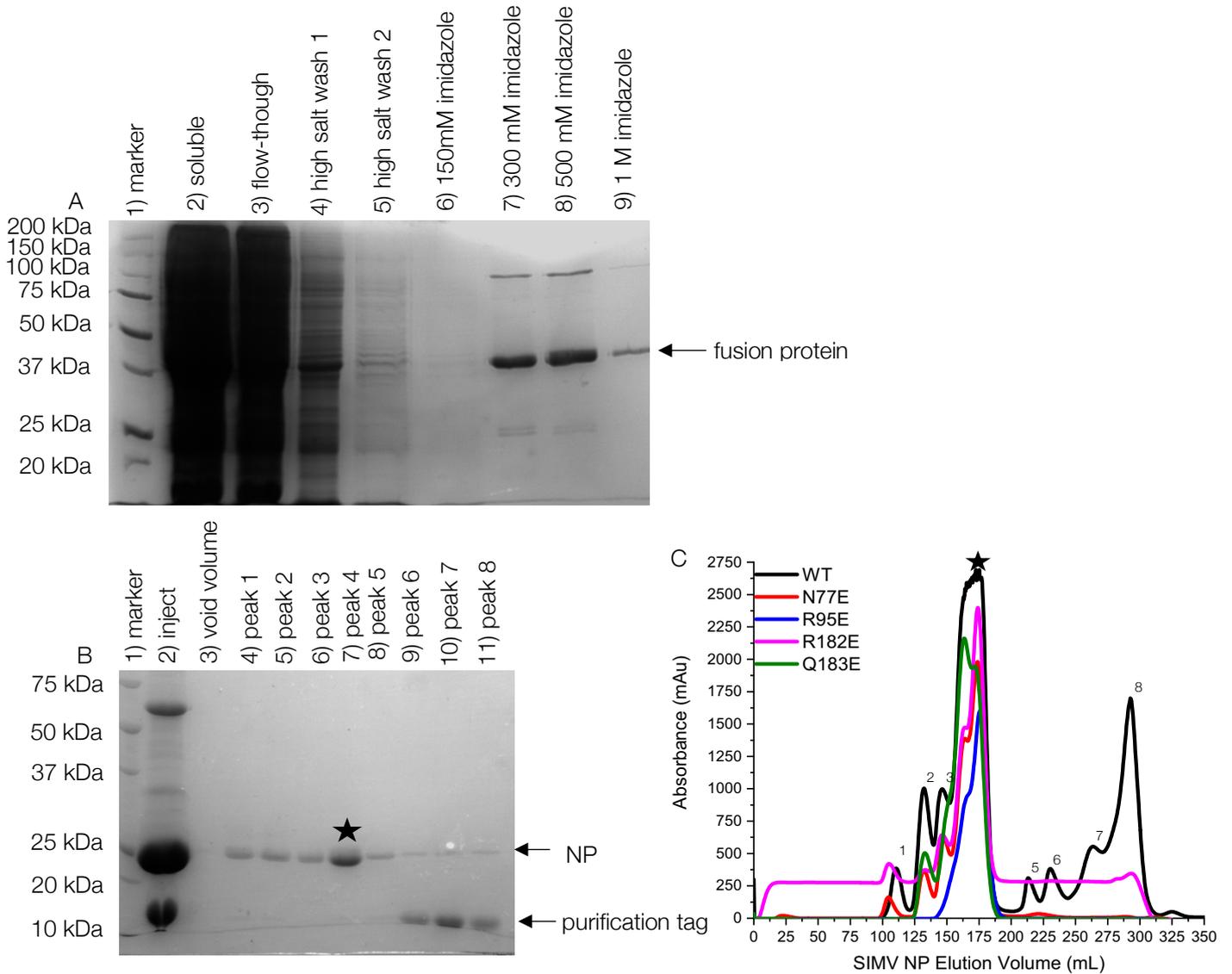
Figure 4.1 Purification of WT SIMV NP and Mutants

A- SDS-PAGE of affinity chromatography for WT SIMV NP. 1) marker, 2) soluble lysate, 3) flow-through, 4) wash 1, 5) wash 2, 6) 300 mM imidazole eluate, 8) 500 mM imidazole eluate, 9) 1 M imidazole wash

B- SDS-PAGE of gel filtration chromatography corresponding to each peak for SIMV WT. 1) marker, 2) loaded sample, 3) peak 1 4) peak 2, 5) peak 3, 6) peak 4 (star), 7) peak 5, 8) peak 6, 9) peak 7, 10) peak 8 Fractions containing pure tetrameric protein (star) were pooled and used for further experiments.

C- size exclusion chromatography of purified WT SIMV NP (black) and mutants. Star represents fractions containing tetrameric protein. Before the divider presents higher oligomeric states and aggregated protein. After is peak for 6His-SUMO tag and contaminating proteins.

Fusion protein = 39 kDa.

NP = 26 kDa.

Purification tag = 13 kDa.

4.2.2 Crystallography

4.2.2.1 Initial Screening

Following obtaining pure protein, initial crystallisation screens were performed using apo NP and NP in complex with pA/C/G/U 20-10 mer RNA. Initial screens used commercially available sparse matrix screens JCSG Core 1-4. Trials were set up using 1:1 μM ratio of protein:RNA. Crystal 'hits' were seen after 48 hours in a range of conditions centring around ethylene glycol (EG), 0.1 M sodium acetate or 0.1 M sodium citrate for SIMV (Figure 4.2). AKAV crystal hits were seen in varied range of conditions shown in Table 4.3 with similar morphologies to SIMV NP (rod-like crystals) (Figure 4.3).

Table 4.3 Crystal Conditions for AKAV NP in Complex with RNA

Screen	Conditions
JCSG Core I G5	0.2 M ammonium sulphate, 0.1 M citric acid pH 5, 10% (w/v) PEG 6000
JCSG Core I G9	0.2 M ammonium sulphate, 0.1 M sodium acetate pH 4.6, 20% (w/v) PEG 2000 MME
JCSG Core I H5	0.2 M sodium chloride, 10% (w/v) PEG 3000, 0.1 M phosphate-citrate pH 4.2
JCSG Core IV A1	0.2 M lithium sulphate, 0.1 M CHAPS pH 10.5, 2.0 M ammonium sulphate
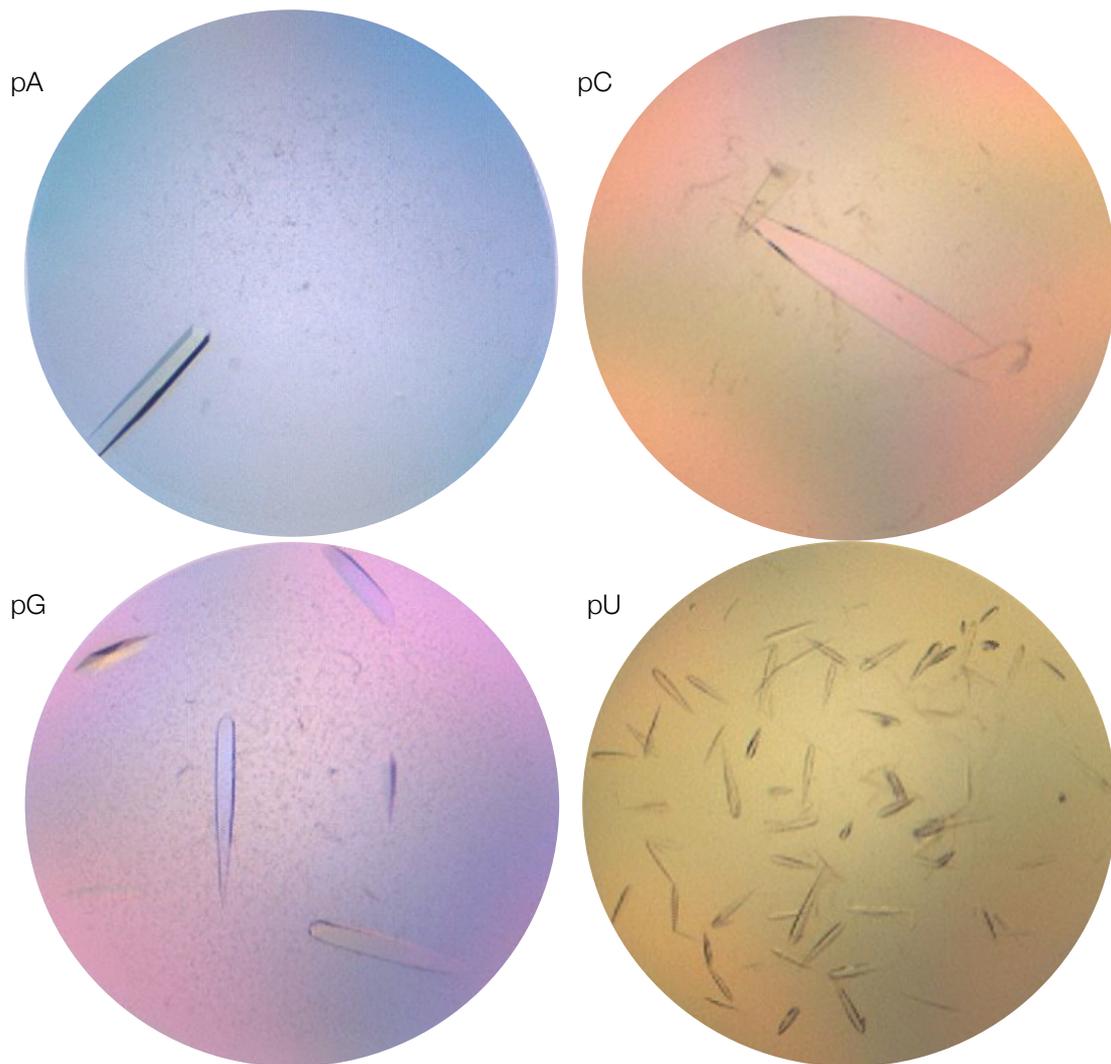
Figure 4.2 SIMV NP Crystal Morphology

Crystallisation trials were set down with 1:1 μM ratio of protein:RNA and incubated at 18°C. Crystals formed in JCSG Core II H3 commercial screen (25% (v/v) ethylene glycol) with a rod-like morphology of different sizes.

pA

pC

pG

pU
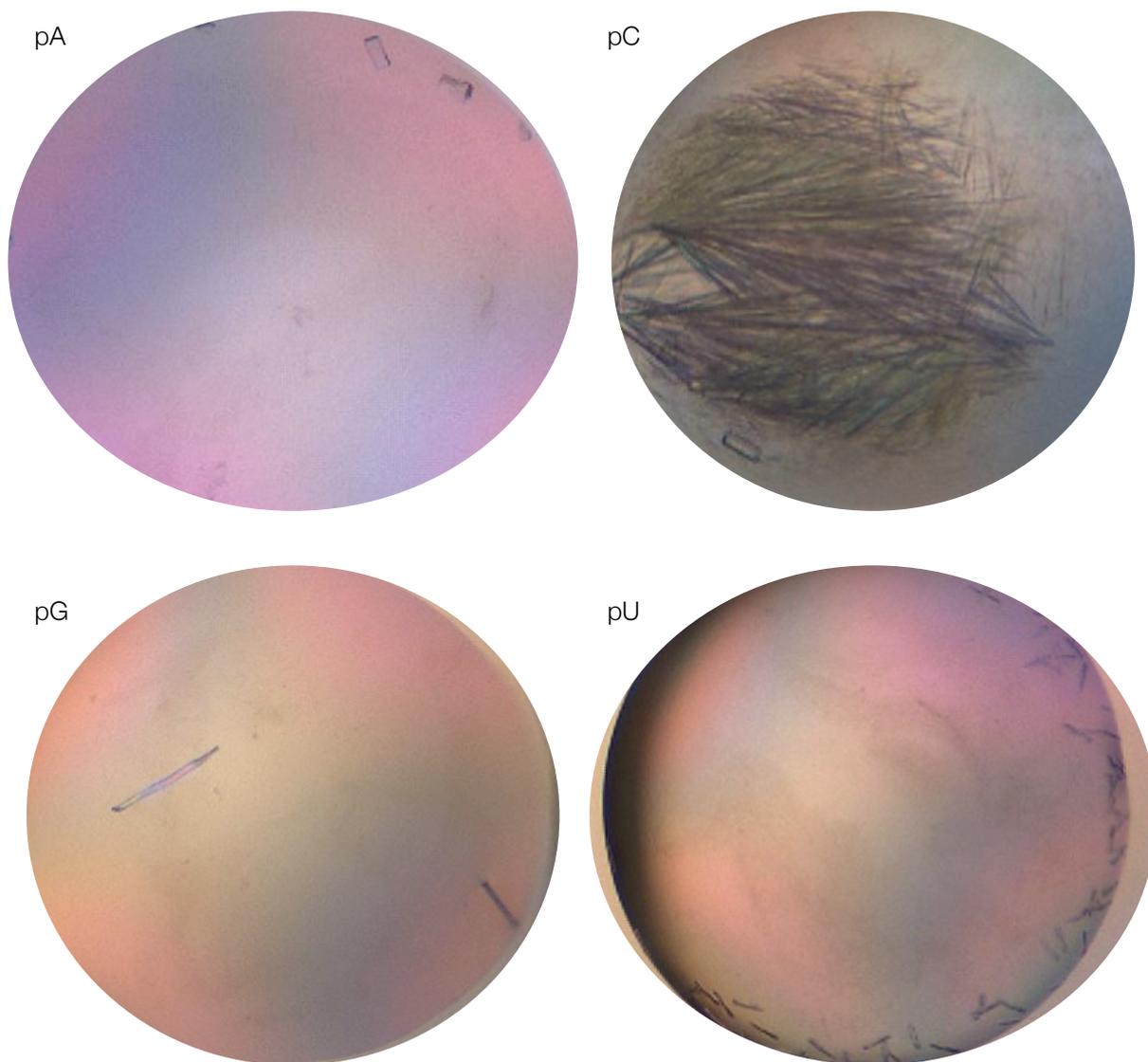
Figure 4.3 AKAV NP Crystal Morphology

Crystallisation trials were set down with 1:1 μM ratio of protein:RNA and incubated at 18°C.
pA- JCSG Core I G5, 0.1 M citric acid pH5, 10% (w/v) PEG 6000).
pC- G9 0.2 M ammonium sulphate, 0.1 M sodium acetate pH 4.0, 20% (w/v) PEG 2000 MME
pG- H5 0.2 M sodium chloride, 10% (w/v) PEG 3000, 0.1 M phosphate citrate pH 4.2.
pU- JCSG Core IV A1 0.2 M lithium sulphate, 0.1 M CHAPS pH 10.5 and 2.0 M ammonium sulphate.

4.2.2.2   SIMV NP Optimisation of Crystallisation Conditions

JCSG Core 1-4 screens that produced crystal hits were optimised in 24-well plates using hanging-drops with 1 µL of protein or protein:RNA at 50:50 ratio protein:reservoir. Ethylene-glycol concentration 30-100%, 0.1 M sodium acetate pH 4-5 or 20-50% polyethylene glycol (PEG) 3000 and 0.1 M sodium citrate pH 3.5-6 were altered. Crystals formed in all conditions reproducibly.

Table 4.4 Optimisation Conditions 1

EG (%) →	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
0.1 M CH$_3$COONa pH								4.0							
								4.1							
								4.2							
								4.3							
								4.4							
								4.5							
								4.6							
								4.7							
								4.8							
								4.9							
								5.0							

Table 4.5 Optimisation Conditions 2

PEG 3000 (%) →	20	25	30	35	40	45	50
0.1 M Na$_3$C$_6$H$_5$O$_7$ pH			3.5				
			3.75				
			4.0				
			4.25				
			4.5				
			4.75				
			5.0				
			5.25				
			5.5				
			5.75				
			6.0				

4.2.2.3   Crystal Harvesting

Crystals typically appeared rod-shaped for SIMV NP and similar for AKAV NP although smaller (Figure 4.2 and Figure 4.3 respectively). Crystals were harvested in appropriate size loops. Additional cryo-protectant for SIMV NP crystals were not added due the high concentrations of EG and PEG 3000. The cryoprotectant mix was added to the AKAV NP mother liquors to a final concentration of 5%. Multiple crystals were harvested and flash frozen in liquid nitrogen from the same condition and multiple conditions picked for apo crystals and those with NP:RNA mix. Crystals were shipped to the DLS for screening and data collection.

4.2.2.4   Data Collection at the Diamond Light Source

Diffraction pattern collection occurred on beamline I24 (AKAV NP pU19), I03 (SIMV NP pA20), I04 (SIMV NP pC/G/U 10) to a maximum resolution of 2 Å for each data set collected. All crystals were screened for diffraction at 0°, 70° and 140°, to assess the diffraction limits. 1 data set for each crystal was collected at 0.2 oscillations for a total of 360°, exposure time was 2 seconds. No data was obtained for SIMV NP apoprotein or AKAV NP apoprotein in complex with pA/C/G.

4.2.2.5   Processing Diffraction Data

Auto-processed .mtz files were downloaded from ISPyB. The highest diffracting crystal was SIMV NP pA20 which was solved by molecular replacement using Phaser against the previously published BUNV NP model (3ZLA) (Figure 4.4 A). This was used to build an initial model which showed a single tetramer per asymmetric unit. Cycles of REFMAC5 (Murshudov *et al.*, 2011) used for structure refinement and manual model building in COOT occurred before a final REFMAC refinement. The high resolution SIMV NP pA20 structure was used as the molecular replacement model to solve the pC10, pG10 and pU10 structures for SIMV and AKAV NP pU19 followed by cycles of refinement and model building in REFMAC and Coot. The crystals diffracted from 1.73 to 2.98 Å in resolution (Figure 4.4). Statistics for data collection and refinement are shown below (Table 4.6 and Table 4.7).
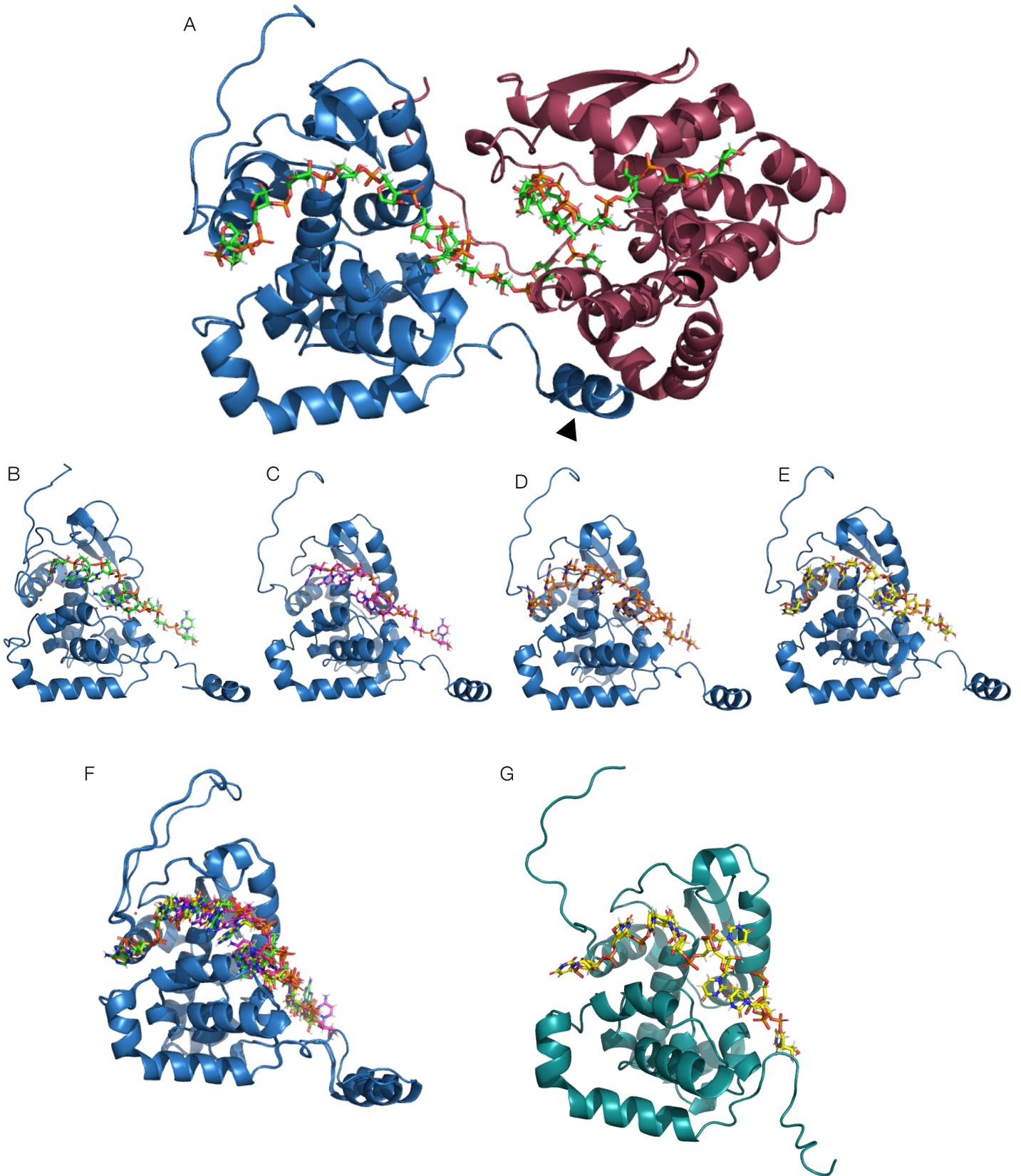
Figure 4.4 Crystal Structures of SIMV and AKAV NP in Complex with RNA

A – SIMV NP dimer in complex with pA 20 (green). NP monomers highlighted in blue and ruby. C-terminal helix (arrow) drives oligomerisation.

B-D – SIMV NP monomer (blue) bound to B- pA20 (green), C- pC10 (magenta), D- pG10 (orange) and E- pU10 (yellow)

F – SIMV NP monomers (blue) aligned. RNA represented as above. Differences seen in the N-terminal arm and C-terminal helix which are known to be flexible.

G – AKAV NP monomer (deep teal) in complex with pU19 RNA (yellow).

Table 4.6 Data Collection Statistics

	AKAV pU 19	SIMV pA 20	SIMV pC 10	SIMV pG 10	SIMV pU 10
Wavelength (Å)	0.9686	0.9763	0.96861	0.96861	0.96861
High Resolution Limit	2.11 (2.15 - 2.11)	1.73 (1.76 - 1.73)	2.65 (2.70-2.65)	2.98 (3.03 - 2.98)	2.75 (2.80 - 2.75)
Low Resolution Limit	75.18 (2.15 - 2.11)	75.43 (1.76 - 1.73)	43.79 (2.65-2.70)	75.42 (2.98 - 3.03)	43.83 (2.75 - 2.80)
Completeness	99.87	98.50	99.98	98.49	100.00
Multiplicity	24.46	11.99	12.85	3.37	12.81
CC-Half	0.9749	0.9990	0.9985	0.9586	0.9987
I/sigma	9.95	14.37	9.86	6.65	12.18
Rmerge (I)	0.2346	0.0744	0.7931	0.1069	0.0981
Anomalous Completeness	99.99	98.52	99.87	77.00	99.79
Anomalous Multiplicity	13.25	6.15	6.69	1.89	6.69
A	106.31	106.69	107.069	47.745	106.942
B	106.32	106.69	107.069	79.029	106.942
C	48.01	48.04	47.992	79.128	48.048
α	90.0	90.000	90.000	84.741	90.000
β	90.0	90.000	90.000	72.396	90.000
γ	90.0	90.000	90.000	72.538	90.000
Space Group	P4 21 2	I4	I 4	I 4	I 4
Twinning Score	2.97	2.20	2.84	3.01	3.18

Table 4.7 Refinement Statistics

	AKAV pU 19	SIMV pA 20	SIMV pC 10	SIMV pG 10	SIMV pU 10
Resolution (Å)	2.11	1.73	2.6	2.7	2.75
No. reflections	16393	55650	7920	7054	7106
R-factor/R-free	0.246/0.289	0.179/0.199	0.299/0.335	0.272/0.282	0.266/0.288
RMSD Bonds	0.006	0.004	0.004	0.007	0.004
RMSD Angles	0.96	0.885	0.832	1.240	0.811
B-factors ($Å^2$)	67.0	37.6	135.0	142.7	133.8
Ramachandran outliers (%)	2.3	0.22	0.44	1.32	0.88
MolProbity Score	2.21	0.79	1.49	2.05	1.39

4.2.3 Structures of NP in Complex with RNA

The crystal structures of *Peribunyaviridae* NP in this thesis were solved in complex with RNA representing bases adenine, cytosine, guanine and uridine (Figure 4.4). The crystal structures also displayed a positive groove proposed for RNA binding (Figure 4.5). Charge and hydrogen bond interactions play a key role in RNA recognition (Figure 4.6).

Table 4.8 outlines which residues from SIMV NP and AKAV NP interact with RNA. AKAV and SIMV NP exhibit high amino acid identity (82%) and are classified in the same serogroup, the RMSD between SIMV NP pU10 and AKAV NP pU19 structures was 0.752 Å for all atoms.

AKAV NP crystallised as one monomer per asymmetric unit in complex with RNA. The synthesised RNA is wrapped within the proposed RNA-binding groove. Nine bases of uridine are visible in the structure per monomer (Figure 4.4 G).
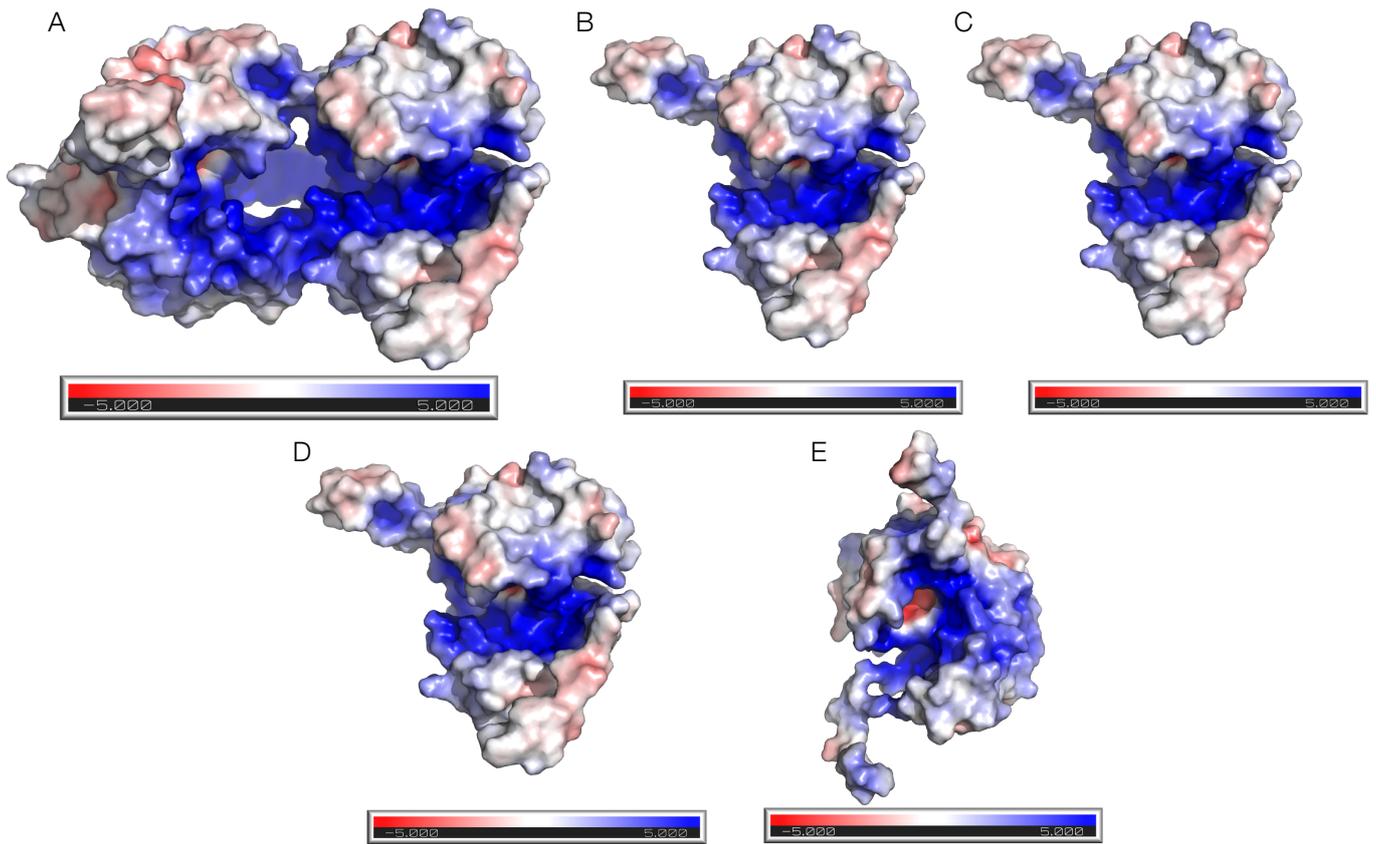
Figure 4.5 NP Electrostatic Surface Potential

Positive RNA binding groove (blue) found in AKAV pU19 (E)and SIMV NP (A-D)
crystal structures in this thesis.
A- SIMV NP pA20.
B- SIMV NP pC10.
C- SIMV NP pG10.
D-SIMV NP pU10.
E- AKAV NP pU19.
Figure made in PyMol (version 2.3.2) using APBS Electrostatic plugin.
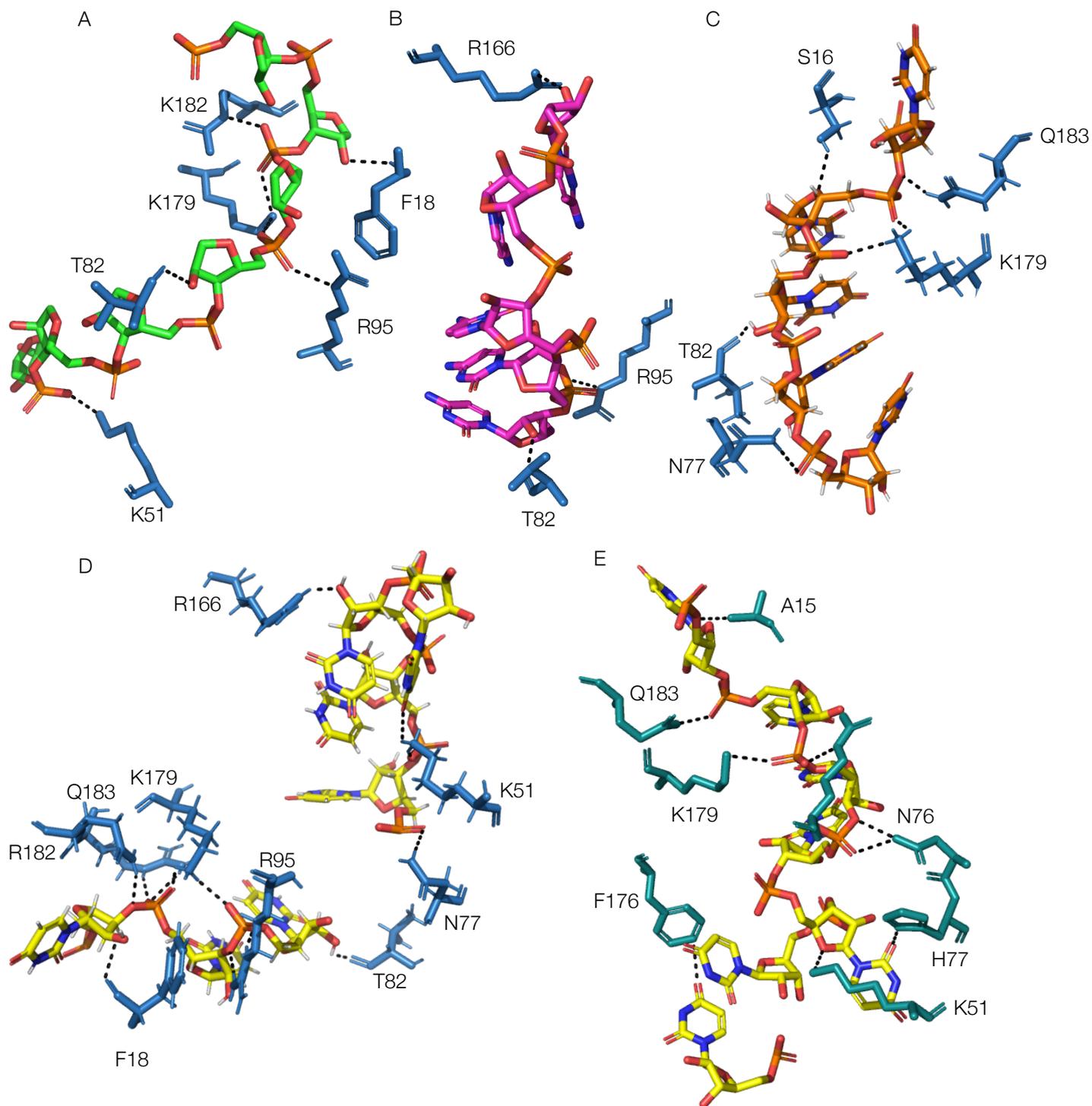
Figure 4.6 SIMV and AKAV NP Residues involved in RNA Binding

Crystal structures were analysed for residue interaction with RNA bases, ionic interactions and hydrogen bonds (represented by black dashed lines) to sugars and phosphates (orange). Two base stacking interactions were seen: SIMV NP pG10 F179 and AKAV NP pU19 F176.
A- SIMV NP (blue), pA20 RNA shown in green.
B- SIMV NP (blue), pC10 RNA shown in pink.
C- SIMV NP (blue), pG10 RNA shown in orange.
D- SIMV NP (blue), pU10 RNA shown in yellow.
E- AKAV NP (deep teal), pU19 RNA shown in yellow.
Figures made in PyMol (version 2.3.2).

Eight AKAV NP residues interact with the RNA (Table 4.8). A15 forms a weak hydrogen bond from a carboyl-terminal hydrogen to a phosphate; K51 also form a hydrogen bond with a sugar. N76 forms two hydrogen bonds to a phosphate. H77 forms a hydrogen bond with a phosphate. R95 and K179 interact ionically with phosphates and Q183 forms a hydrogen bond with a phosphate. F176 interacts with the uridine base via base stacking (Figure 4.6 E).

SIMV NP crystallised as one monomer per asymmetric unit in complex with pC/G/U10 RNA and as a dimer in complex with pA20. The synthesised RNA is wrapped within the proposed RNA-binding groove of each monomer facing inwards. All 20 RNA bases in the SIMV NP pA20 crystal structure are visible, with each monomer binding encapsidating 10 bases each. In the SIMV NP pG10 crystal structure all ten bases of RNA are also visible. However, only eight bases of RNA were visible in the SIMV NP pC10 and pU10 structures (Figure 4.4 B and D respectively).

Hydrophobic F18 in SIMV NP pA20 structure is buried within the protein chain forms weak hydrogen bonds with the carboxyl-terminus double bonded oxygen atom with a pentose sugar. Positively charged K51 and K179 interact with a negatively charged single phosphate. Polar T82 forms hydrogen bonds also to a phosphate. Positively charged R95 and R182 interact with a negatively charged phosphate. The RNA binding interactions are therefore charge based and via hydrogen bonds, overall 7 interactions are seen from 6 residues (Figure 4.6 A).

The SIMV NP pC10 structure only shows three interactions; 1 hydrogen bond formed by T82 and 2 ionic interactions by R95 and R166. Here, the hydrogen bond formed by T82 carboxyl-terminus double bonded oxygen to a hydrogen found on a pentose ring opposed to the phosphate in the RNA backbone of the pA20 structure (Figure 4.6 B).

S16 in the SIMV NP pG10 structure forms two hydrogen bonds to a sugar. N77 forms a hydrogen bond to a phosphate too. T82 forms a hydrogen bond to a sugar similar to pG10 structure. K179 forms a bond with a phosphate and lastly Q183 a hydrogen bond to a phosphate. Overall this structure showed 11 interactions to 7 SIMV NP residues and was the first structure to show RNA base and residue interactions. S16 was not shown to interact with RNA in the three other structures (Figure 4.6 C).

Lastly, nine residues of SIMV NP in the pU10 structure were shown to interact with RNA (Figure 4.6 D). F18 again forms a hydrogen bond with a sugar. N77 and T82 both for hydrogen bonds to sugars. R95 forms ionic interactions with 2 phosphates. R166 forms three interactions; two hydrogen bonds to an RNA base and one ionic interaction to a sugar and one to a phosphate. This was the second residue to display interactions directly with the RNA base. K19 and R182 each interreacted ionically with different phosphates and Q183 forms hydrogen bonds to a phosphate.

Table 4.8 SIMV and AKAV NP Residue Interactions with RNA

HB = hydrogen bond interactions
BS = base stacking interactions

		Structure				
	Residue	AKAV NP pU19	SIMV NP pA20	SIMV NP pC10	SIMV NP pG10	SIMV NP pU10
	A15	1 x HB to phosphate	/	/	/	/
	S16	/	/	/	2 x HB to sugar	/
	F18	/	1 x HB to sugar	/	/	1 x HB to sugar
	K51	1 x ionic to phosphate	1 x ionic to phosphate	/	/	1 x ionic to phosphate
	N77	(N76) 2 x HB to phosphate	/	/	1 x HB to phosphate	1 x HB to sugar
	H77	1 x HB to phosphate	/	/	/	/
	T82	/	1 x HB to phosphate	1 x HB to sugar	/	1 x HB to sugar
	R95	1 x ionic to phosphate	1 x ionic to phosphate	1 x ionic to phosphate	/	2 x ionic to phosphate
	R166	/	/	1 x ionic to phosphate	/	1 x ionic to phosphate 2 x HB to base
	F176	1 x BS to base	/	/	/	/
	K179	1 x ionic to phosphate	2 x ionic to phosphate	/	1 x ionic to phosphate	1 x ionic to phosphate
	R182	/	1 x ionic to phosphate	/	/	1 x ionic to phosphate
	Q183	1 x HB to phosphate	/	/	1 x HB to phosphate	1 x HB to phosphate
Total	Ionic	2	5	2	1	5
	HB to sugar	1	1	1	2	3
	HB to phosphate	6	1	0	2	1

4.2.4 Analysis of NP residues Involved in RNA binding

4.2.4.1 Fluorescence Anisotropy

To assess the RNA-binding contribution of NP residues, residues know to interact with RNA in OROV NP (from an unpublished crystal structure in complex with pC10 within the Edwards/Barr group (previously mutated to alanine by Sue Matthews, Edwards group technician)) and residues that interacted with the RNA across the crystal structures (SIMV NP pA20, pC10, pG10, pU10), were selected and mutated to alanines (OROV NP mutants) or glutamates (SIMV NP mutants) (Figure 4.7 A and B respectively). OROV NP was chosen for binding studies due to the time scale of experiments, the OROV NP mutants were already available and AKAV NP pU19 X-ray diffraction data had not been processed. Binding affinity of each mutant was assessed through fluorescence anisotropy for OROV and SIMV NP.

Table 4.9: NP Mutants

| OROV NP	SIMV NP
K47A	/
K50A	K51E
/	N77E
R94A	R95E
R180A	R182E
Q181A	Q183E
R182A	/
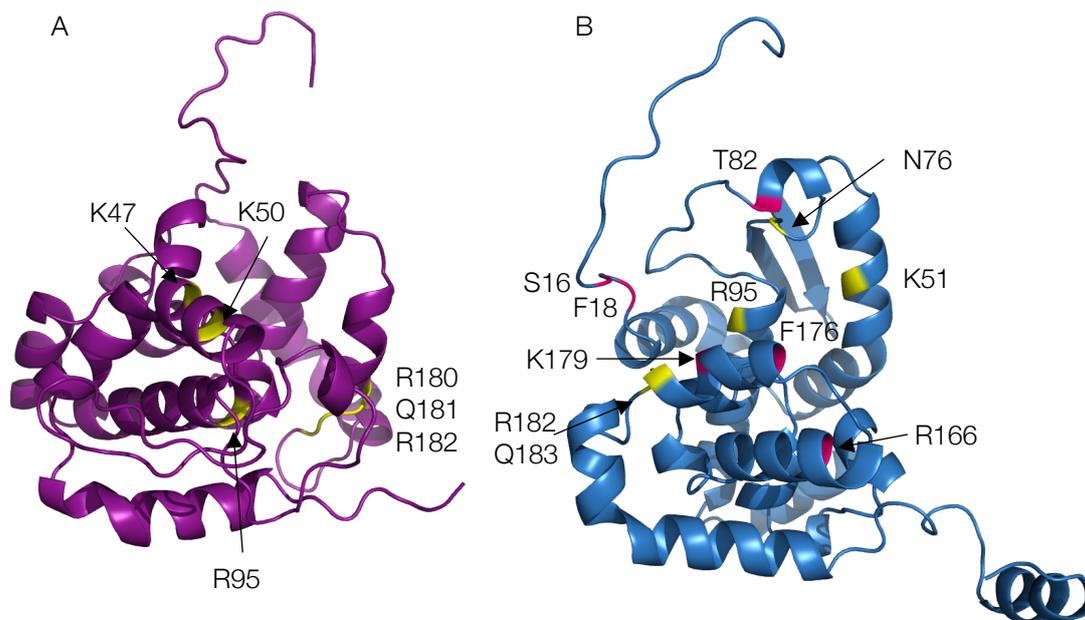
Figure 4.7 Structure-Function Analysis of OROV and SV NP

A- Structure of OROV NP monomer (purple, unpublished data Edwards/Barr group). Residues mutated for RNA binding assays shown in yellow.
B- Structure of SIMV NP monomer bound (blue). Residues mutated for RNA binding assays shown in yellow and residues shown to interact to RNA shown in pink.
Figures made in PyMol (version 2.3.2).

OROV NP K50A and SIMV NP K51E did not express soluble protein. For both OROV and SIMV recombinant NP expression produced small bacterial pellets (> 1 g) compared to WT and other mutants (average 5 g).

NP:RNA binding was detected using 3'fluorescein labelled oligonucleotides representing each base (A,C and U). Unfortunately, G could not be assessed, possibly due to secondary structures of the pG RNA. Binding to 10mer, 3'FITC labelled pA, C and U RNA was analysed by FA in order to determine if NP from AKAV, OROV and SIMV preferentially bind RNA bases (Figure 4.8).

The highest binding affinities were seen for AKAV NP for all three oligonucleotides tested (pA 1.510 nM, pC 4.640 nM and pU 0.039 nM) (Figure 4.8 A-C respectively, black curve).

No binding was seen for SIMV NP mutant R95E to pC and pU RNA; weak binding was observed to pA and the experiment was repeated with a higher starting concentration of R95E NP in order to observe a binding curve (Figure 4.8 D cyan curve) (1.845 μM compared to 3 nM for WT). The corresponding mutant in OROV NP (R94A) exerted significantly reduced binding when compared to WT; 40 μM vs 7.22 nM pA, 1.8 μM vs 0.8 μM pC and 0.34 μM vs 0.04 nM for pU) (Figure 4.8 D-F respectively, blue curve).

Table 4.10 Table of Affinities (apparent K_D) for NP RNA Binding Assays

| Protein	RNA		
	pA10	pC10	pU10
	Apparent K_D (μM)		
AKAV WT	0.002 ± 0.001 (1.510 nM)	0.005 ± 0.001 (4.640 nM)	0.00004 (0.039 nM)
OROV WT	0.007 ± 0.001 (7.220 nM)	0.798 ± 0.257	0.0002 (0.024 nM)
OROV K47A	0.010 ± 0.001 (10.430)	5.184 ± 1.862	0.0002 (0.220 nM)
OROV R94A	40.293 ± 34.552	1.878 ± 4.675	0.393 ± 0.040
OROV R180A	0.542 ± 0.109	5.434 ± 3.111	0.005 ± 0.001 (4.820 nM)
OROV Q181A	0.024 ± 0.002 (24.000 nM)	3.207 ± 0.937	0.001 ± 0.001
OROV R182A	0.128 ± 0.014	4.525 ± 1.529	0.001 (0.544 nM)
SIMV WT	0.003 ± 0.001 (3.000 nM)	0.084 ± 0.005 (84.000 nM)	0.020 ± 0.001 (20.000 nM)
SIMV N77E	0.090 ± 0.002	0.308 ± 0.008	0.078 ± 0.003
SIMV R95E	1.845 ± 0.036	/	/
SIMV R182E	0.070 ± 0.002	0.222 ± 0.005	0.108 ± 0.002
SIMV Q183E	0.331 ± 0.017	0.947 ± 0.114	0.473 ± 0.021
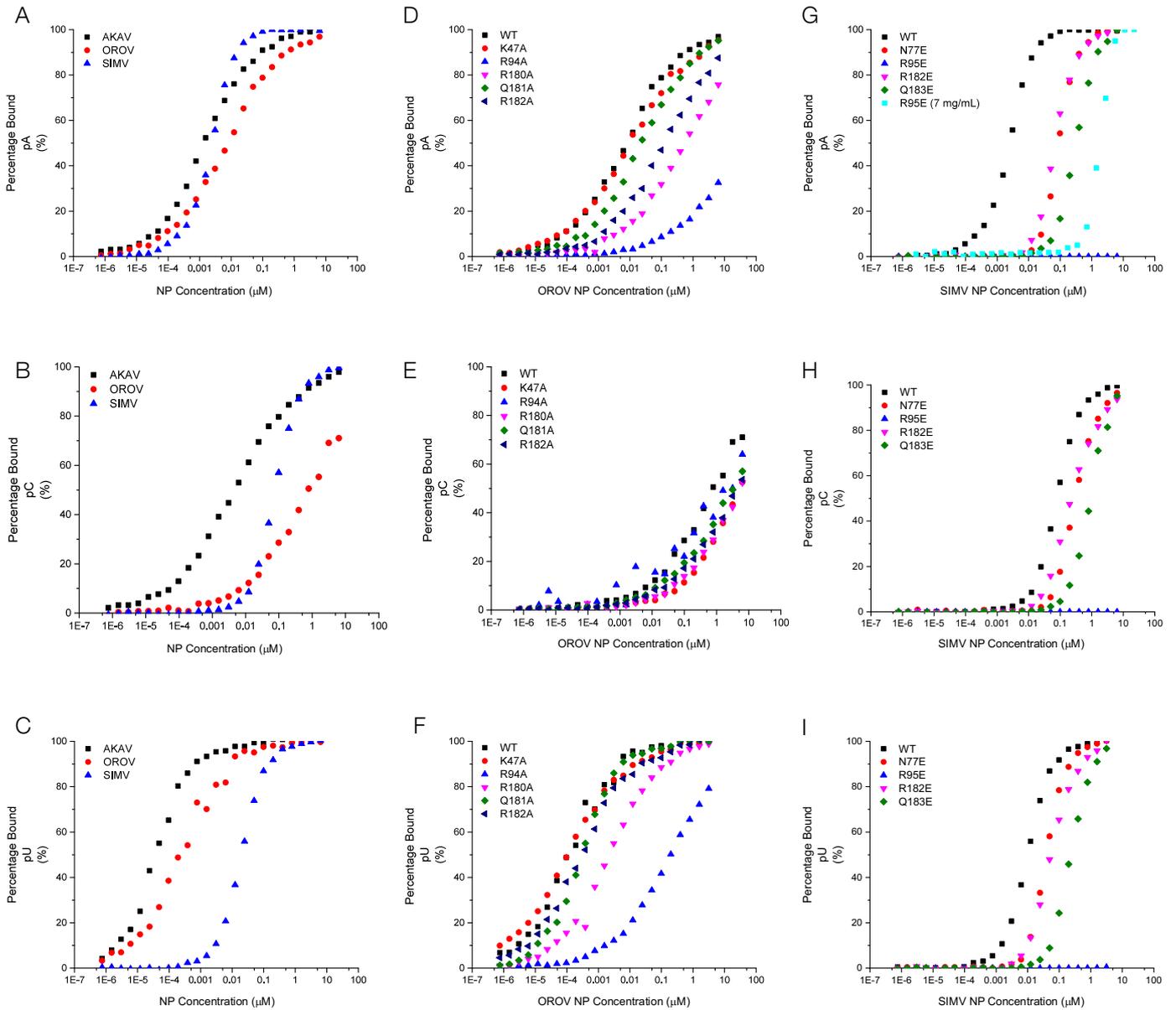
Figure 4.8 Direct Binding of *Orthobunyavirus* NP to RNA

Fluorescence anisotropy RNA binding assay. NP were tested for binding to 3' fluorescein-labelled 10 mer RNAs pA, pC and pU..
A-C- WT AKAV, OROV and SIMV NP.
D-F- WT and mutant OROV NP.
G-I- WT and mutant SIMV NP.

4.3 Discussion

4.3.1 Crystal Structures

4.3.1.1 AKAV NP

It was not expected that AKAV NP in complex with pU19 would crystallise as one monomer per asymmetric unit (Figure 4.4 G). A dimer of AKAV NP was expected due to the length of RNA used; as each monomer binds 9-11 bases of RNA you would therefore expect two protomers to be bound with the pU19 RNA, but only one monomer and 9 bases of RNA are seen in the asymmetric unit. Refinement statistics suggest that something is wrong regarding this structure. For a high-resolution structure of 2.1 Å, the R-factors remain high. There was no obvious improvement needed when looking at electron density maps for the protein chain. XTriage (Phenix) also does not suggest that anything is wrong with the data, data was not twinned, and no ice rings were present in diffraction images. The space group for all four SIMV NP crystal structures is I4 however AKAV NP space group is P4 21 2, suggesting the space group could potentially be wrong; however POINTLESS (Evans, 2006) gave no other suitable solution. The suggested R-factor for the AKAV NP pU19 structure should be less than 10% of the resolution. The AKAV NP pU19 structure is 2.1 Å, R-factor should therefore be less than 0.21. Here, the R-factor is 0.246 and R-free 0.2885. A typical R-free value will be ± 5-7% of the R-factor e.g. 0.228-0.263 R-free. This suggests that initial indexing could be wrong, possibly due to pseudosymmetry and/or translational elements, however an indexing issue was not suggested by XTriage,

The electron density for the RNA chain is poor (Figure 4.10 E). We suggest that the RNA is degrading due to the high pH of the mother liquor used for crystal growth (pH 10.5). From the SIMV NP pA20 model you can see bases of RNA that are not completely encapsidated by NP (between two protomers), if this was also true for AKAV NP pU19 crystal then RNA degradation can occur here. This may have 'broken' the RNA in two and therefore allows for a single monomer to be present in the asymmetric unit.

Due to the poor density for the RNA we cannot be confident with the positions of RNA within the structure and therefore the interactions between NP and RNA are probably an over interpretation.

4.3.1.2 SIMV NP

Subtle differences can be observed between SIMV NP structures; the N-terminal arm and the C-terminal helix between the crystal structures (Figure 4.4 F), both exert a great deal of flexibility, with probable functional relevance in RNP assembly and disassembly, and differences seen here are assumed to reflect this. The NP backbone otherwise does not move to accommodate the different bases (Figure 4.4 F). Table 4.11 highlights the degree of similarity between the four structures. Typically, an RMSD over 1 Å would indicate a structural difference.

Hydrogen bonds to sugars are likely to discriminate between DNA and RNA and contribute to how NP recognises RNA. Again, due to the density for the RNA being poor for pC10, pG10 an pU10 structures (Figure 4.9 B-D respectively), we cannot be confident with the positioning of the RNA bases. We analysed the NP:RNA interactions as you would do for crystal structures with better density.

Table 4.11 RMSD Values for SIMV NP Crystal Structures
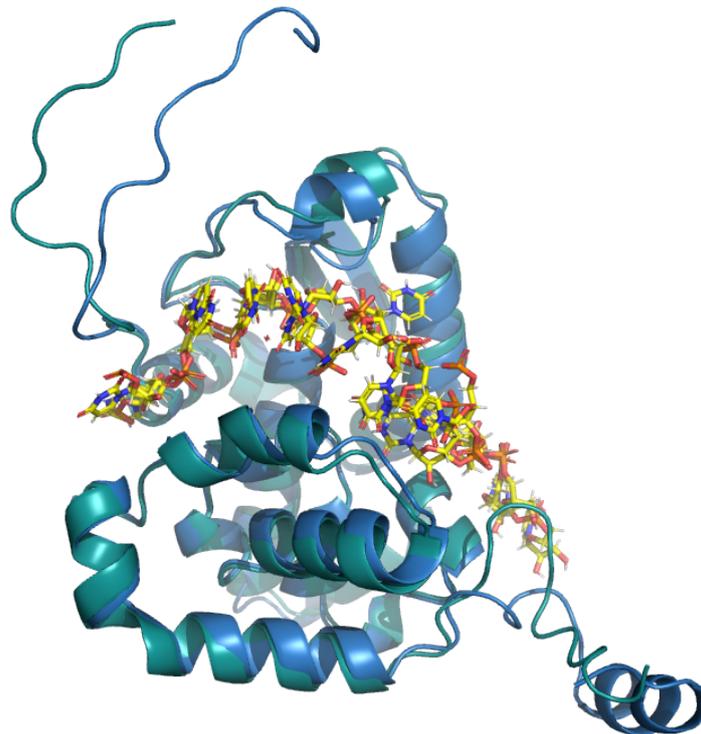
Figure 4.9 Structural Superposition of AKAV and SIMV NP in Complex with pU

AKAV NP (deep cyan) and SIMV NP (blue) in complex with pU RNA (yellow) were aligned in PyMol.
RMSD 0.792 Å.
Figure made in PyMol (version 2.3.2) using superposition function.

| | RMSD (Å) | | | |
	pA	pC	pG	pU
pA		0.172	0.274	0.307
pC			0.219	0.245
pG				0.226
pU				
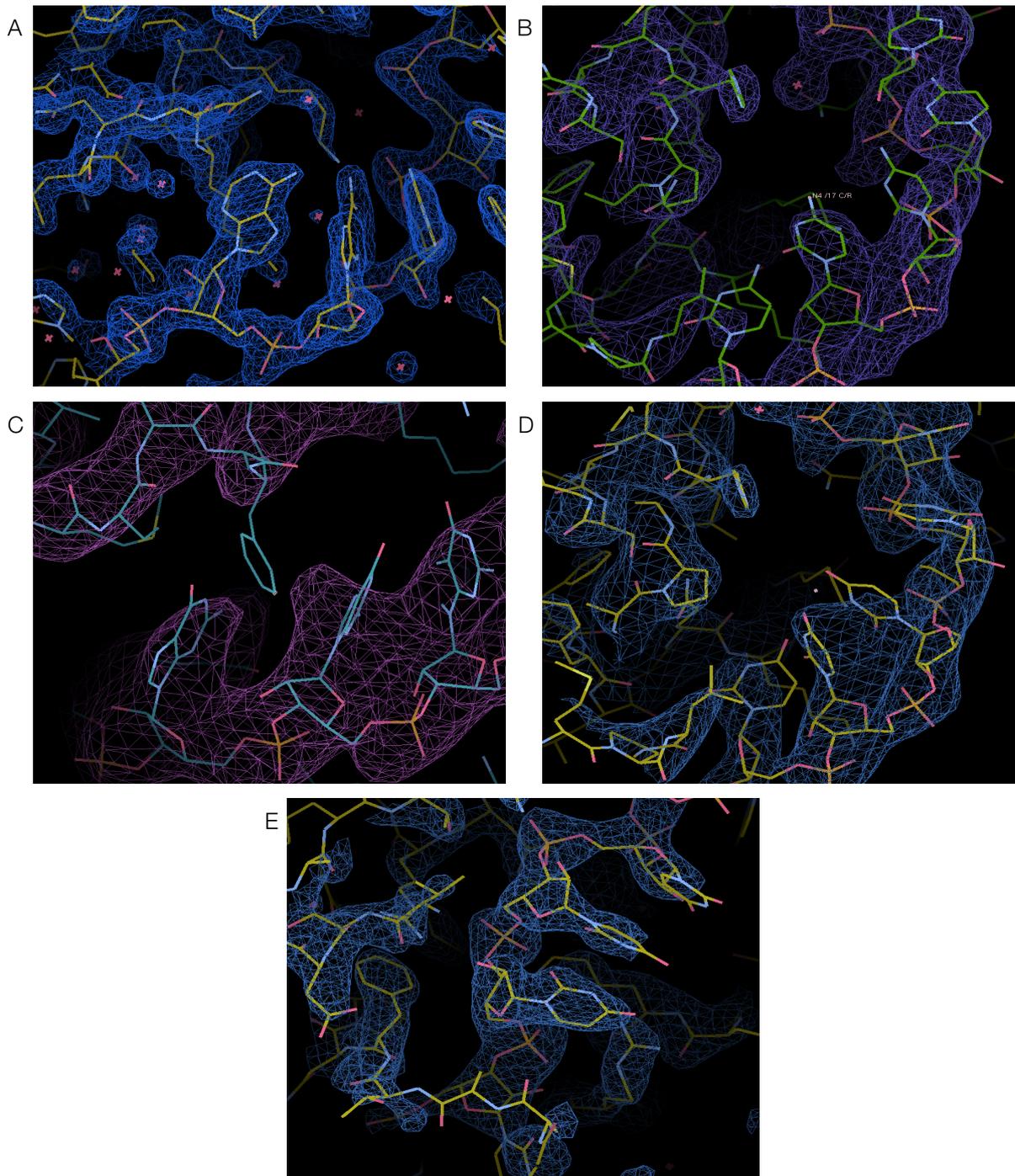
Figure 4.10 Electron Density for NP Crystal Structures in Complex with RNA

2Fo-Fc (1.0 sigma) electron density maps for SIMV and AKAV NP structures.
A- SIMV NP pA20.
B- SIMV NP pC10.
C- SIMV NP pG10.
D- SIMV NP pU10.
E- AKAV NP pU19.

4.3.1.2.1 Structural Similarities between other *Peribunyaviridae* Viruses

The monomeric NP of BUNV, SBV and SIMV are remarkably similar (Figure 4.11 D and E respectively), with an RMSD of 0.892 Å and 0.521 Å respectively. This is unsurprising as all three viruses are from the same family and SBV and SIMV are from the same sero complex (Simbu complex). The BUNV and SIMV NP monomers align similarly, the majority of the SBV and SIMV NP monomer overlap, apart from the N-terminal arm and C-terminal helix. These are known to be flexible and therefore the differences seen in the crystal structures are likely to be due to this.

The superposition of tetramers of NP from the three viruses however differ slightly. BUNV and SIMV in complex with RNA present a more symmetric shaped tetramer. On the other hand, SBV apo tetramer is more rhombic in shape compared to BUNV and SIMV in complex with RNA. It was previously suggested that this difference is due to RNA binding and/or crystal packing. The BUNV and SIMV NP tetramer helices and β-sheets overlap in most of the structure. Whereas the SBV and SIMV tetramers are not overlapping but align next to each other. This suggest that RNA binding impacts tetramer formation and enforces small conformation changes. These changes create a more tightly packed tetramer (BUNV and SIMV) which has biological relevance for RNA protection.
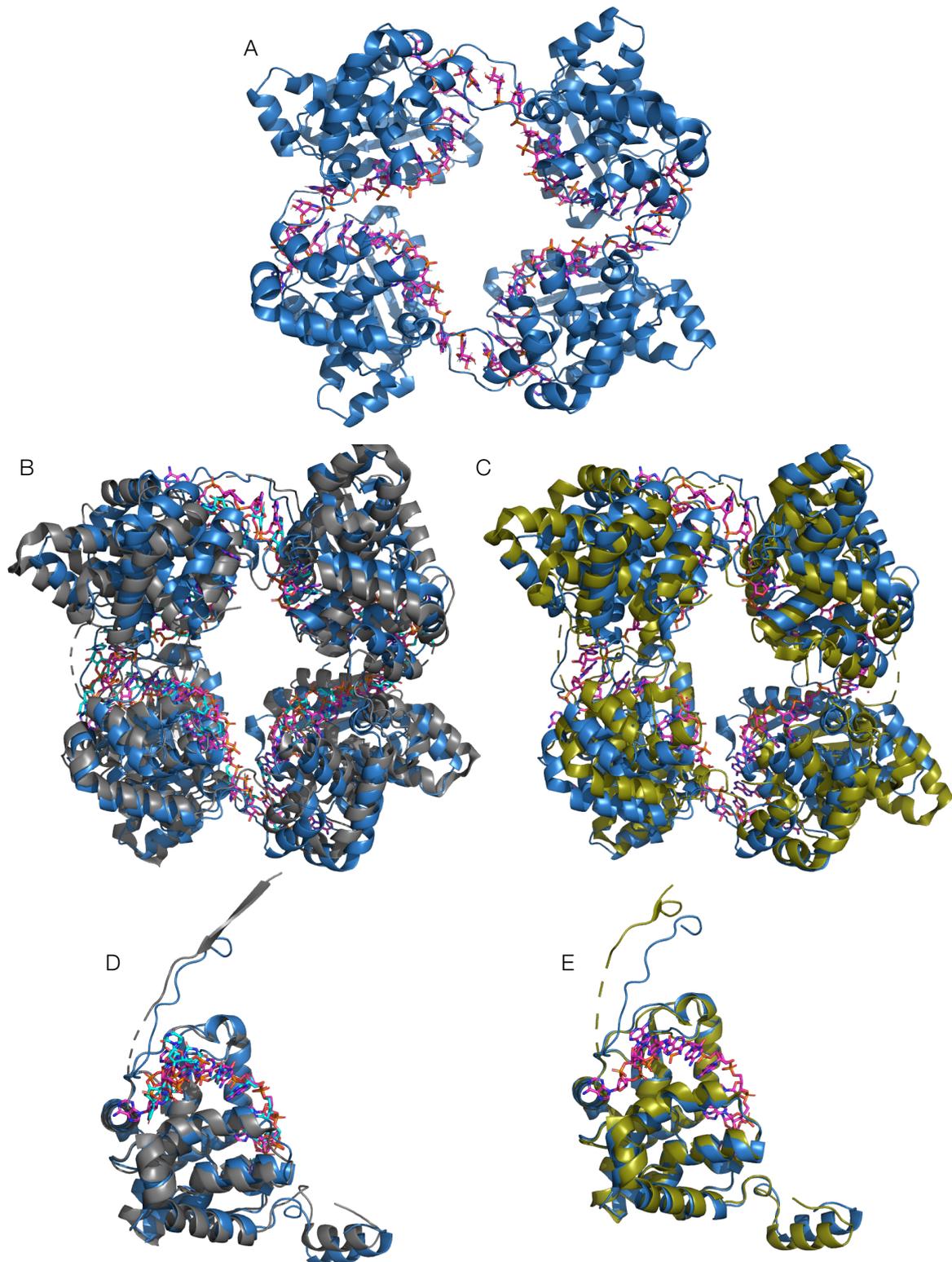
Figure 4.11 Structural Comparisons of SIMV to BUNV and SBV NP Tetramers

A- SIMV NP tetramer (blue) in complex with pA20 RNA (pink).
B- SIMV NP tetramer aligned with BUNV NP tetramer (grey) in complex with E. coli host RNA (cyan) and
D- monomers aligned.
C- SIMV NP tetramer aligned with SBV NP tetramer (olive) and E- monomers aligned.
Figure made in PyMol (version 2.3.2) using superposition function.
PDB: 3ZLA (BUNV) 3ZL9 (SBV).

4.3.2 RNA Binding

From the crystal structures it appears that only SIMV NP R166 (pU10 crystal structure) binds to RNA base. It is interesting why NP interacts differently with the sugar-phosphate backbone in the different crystal structures if NP is encapsidating RNA in a sequence independent manner. As we are not confident on the specific RNA interactions true comparisons cannot be made on why different NP residues interact differently with RNA. However, RNA binding to pC exerted the lowest affinities for all three viral NP. Weaker affinity binding to pC is supported by the pC10 crystal for SIMV NP as only three interactions were seen compared to seven for pA, five for pG and 12 for pU.

RNA binding data suggests that AKAV and OROV NP preferentially bind uridine nucleotides (apparent K_D 1.51 nM and 7.22 nM for pU10 vs 4.64 nM and 0.80 μM pC10 and 1.51 nM and 7.22 nM pA10 respectively). SIMV NP however, preferentially binds adenosine nucleotides (apparent K_D 3 nm pA10, 84 nM pC10 and 20nM pU10). From the crystal structures you may expect the binding affinity to pU10 to be higher than pA10 for SIMV NP, due to the number of interactions seen in the pU10 structure (6 interactions to pA20 compared to 12 in the pU10 structure). AKAV NP pU19 crystal structure shows 12 interactions between RNA and protein, however crystal structures for the other RNA bases will be needed to draw any comparisons, and the high possibility that the pU19 RNA was degrading brings uncertainties to the analysis for this structure.

The literature has always suggested that NP from the *Orthobunyavirus* genus recognises RNA in a sequence independent manner. However, data presented in this thesis suggests otherwise. If NP is able to recognise RNA in a sequence dependant manner, it is interesting that the different viruses preferentially bind different RNA bases.

The RMSD value between the two pU structures for AKAV and SIMV NP is 0.752 Å, suggesting that NP does not move to incorporate the RNA. AKAV NP forms 9 interactions from 8 residues whereas SIMV NP forms 12 interactions from 9 residues. Suggesting that the number of interactions formed does not increase the binding affinity to RNA.

Of the mutants R94A from OROV NP and respective residue, R95E from SIMV NP cause the biggest reduction in RNA binding suggesting that this arginine is most important for NP RNA binding activity. A94 (OROV NP) and respective residue A95 (SIMV NP) are buried deep within the RNA binding groove. This is supported by the previously published BUNV NP mutant R94A (found at the rim of the RNA-binding groove) was also unable to bind RNA, with poor activity in RNA replication and mRNA transcription (Ariza *et al.*, 2013).

SIMV NP R95 interacts with a phosphate in the pA, pC and pU10 structures (but not pG) and this change in charge repels the RNA, the detection of RNA binding goes below the limits of this assay, however highlights that R95 is important for RNA binding. It is possible that the change in charge within the buried surface of the RNA binding

groove may disrupt overall NP fold however circular dichroism analysis did not show this (Figure 4.12).

OROV NP R180A, Q181A and R182A and SIMV NP R182E and Q183E displayed reduced levels of RNA binding compared with WT NP (75-fold, 3-fold and 17-fold decrease and 23-fold and 110-fold decrease respectively). These residues are located within the wall of the RNA binding groove and reduced binding affinity seen for RNA highlights their importance for RNP function.

The number of interactions seen from RNA base to amino acid in the crystal structure does not correlate with the affinities, and this problem is yet to be solved. Discussion with Dr. Thomas Edwards (supervisor, University of Leeds), we suggest that this may be due to dynamics and/or water mediated. This could be measured by H-bonds and other non-covalent interactions. Conversations with Dr. Megan Wright (academic fellow, University of Leeds), suggested in order to do this higher resolution structures would be needed where the RNA is degraded.

Unfortunately, the RNA binding capabilities of NP to the 3'flurescin label was not assessed. However, typically NP only recognises single-stranded RNA, therefore it is unlikely that NP bound fluorescein.
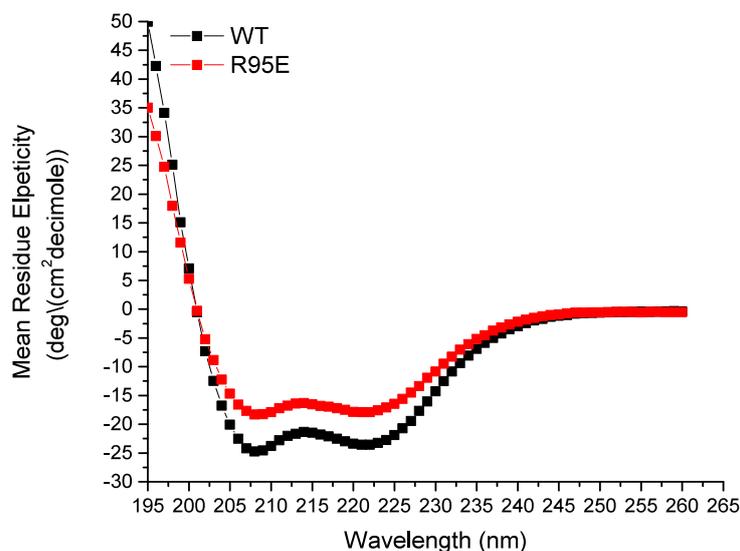
Figure 4.12 Protein Quality Check for SIMV NP R95E Mutant

CD analysis of SIMV WT NP and R95E mutant showed folded protein that is α-helical.

4.4   Chapter Summary

This chapter outlines the optimised purification method for *Orthobunyavirus* NP. High yields of AKAV, OROV and SIMV NP were produced by a 2-step purification method that produced pure tetrameric protein for X-ray crystallography trials and fluorescence anisotropy binding assays.

SIMV NP crystal structures were solved to high resolution (1.73 – 2.75 Å) in complex with four different bases of polynucleotides (pA20, pC10, pG10 and, pU10). A further crystal structure was also solved of AKAV NP in complex with pU19. The differences seen in the crystal structures cannot yet explain the differences seen for binding affinities to the different polynucleotides tested. Dynamics and water mediated interactions may play roles here. We are thinking of collaborations to perform molecular simulations however this goes beyond the speculation of this thesis.

4.5   Future Directions

Usually we can learn a lot about the function of a protein from its structure. However, the *Orthobunyavirus* NP structures unfortunately do not yet explain differences observed in experimental data for RNA binding affinities. NP functions to encapsidate the entire genome suggesting that binding independent of sequence. We were therefore surprised to see different preferences for the sequences tested, which cannot yet be fully-explained by our high-resolution crystal structures. In order to make a full comparison for AKAV NP, crystal structures will need to be solved in complex with pA/C/G RNA. We will also need to obtain crystal structures with better density for RNA in order to full elucidate which NP residues are interacting with RNA. The crystal structures presented here provide a possible target for structure-based drug design to interfere with critical residues for RNP functions.

Chapter 5 Protein Purification of VP35 and VP30 protein of Ebolavirus and Marburgvirus (EBOV and MARV) and Oligomeric State Elucidation

5.1 Chapter Introduction

The 2013-2016 outbreak in Africa highlighted the need for effective therapeutics against Ebolavirus disease (EVD). At present, efforts are being made to develop and implement a vaccine against EBOV, whilst it is still important to remember that antiviral treatments are needed to diminish viral load in infected patients, preventing the spread of EVD. Structure based drug design has arisen as an investigation method for antiviral drug candidates. Small molecule inhibitors directed against viral proteins have the potential for broad spectrum or pan-antiviral activity due to structural similarities within closely related viral proteins. A better understanding of the structure, multimerization and interactions of viral proteins will aid therapeutic development.

EBOV and Marburgvirus (MARV) have a 19 kB genome that encodes nine proteins from seven genes. The ribonucleoprotein (RNP) complex responsible for replication is composed of viral RNA, encapsidated by nucleoprotein (NP), the RNA-dependent-RNA polymerase L, and, viral protein 35 (VP35) a co-factor of L. For transcription, the aforementioned proteins are required as well as an addition protein, VP30. Transcription and replication of RNA occur in the cytoplasm of infected cells.

VP35 (a homolog of the hRSV phosphoprotein P) plays important roles within the RNP complex and during immune evasion. Interacting with NP, VP35 maintains NP in a 'free' state for newly synthesised RNA encapsidation, similarly to hRSV P. On the other-hand, VP35 decreases IFN production and impairs dendritic cell maturation (these functions may be homologous to those of NS1 and NS2 in RSV – Ebola does not encode the equivalent of NS proteins).

VP30s transcriptional activity is regulated by phosphorylation and through its interaction with RNA. It is thought that VP30 promotes read-throughs at gene-ends, similarly to hRSV M2-1 which preferentially binds A-rich mRNA sequences and promotes the transcription of full-length hRSV genes. M2-1 function is also regulated by phosphorylation. During infection, the replication complex of single-stranded negative-sense RNA viruses must switch between replication and transcription, and potentially phosphorylation is important in regulating that switch (Biedenkopf, Lier and Becker, 2016).

5.1.1 Objectives

In order to fully address VP35 and VP30 functions within the viral life cycle, it is important to understand if and how these proteins oligomerise. Understanding the oligomeric state will also enable comparisons to be made between related viruses.

The globular core domain of M2-1 is structurally similar to the CTD of VP30. As M2-1 is a tetramer and has similar functions to VP30 one might expect VP30 to also multimerise as a tetramer. Similarly, P and VP35 share functional similarities. P has also shown to be tetrameric having a 1:1 interaction with tetrameric M2-1. Therefore, if VP30 and VP35 interact in a similar manner, a similar 1:1 interaction may also occur here.

Structure and function studies require high quantities of homogenous and functionally active protein. A common approach to generate this is through the overexpression of recombinant proteins using an *E. coli* host. The cDNA representing the open reading frame (ORF) of the protein of interest (VP30 and VP35) is inserted behind an IPTG inducible promoter within a specialised expression plasmid, creating a fusion protein with a selected purification tag. Gene expression by RNA polymerase produces high expression levels and, inductions in large scale cultures allows for high quantities of recombinant protein expression and purification to homogeneity using chromatographic techniques.

In this chapter the oligomeric states of VP35 and VP30 from both Ebolavirus and Marburgvirus are determined using SEC, SEC-MALLS and mass spectrometry analysis.

5.2 VP35

The oligomeric state of VP35 remains unknown and is debated within the literature. The current crystal structures of the N-terminal oligomerisation domain suggest trimers for both EBOV VP35 (85-145) (Zinzula *et al.*, 2019) and MARV VP35 (60-135) (Bruhn *et al.*, 2017). The oligomerisation domain of MARV VP35$_{60-135}$ formed an elongated trimeric coiled coil. Similarly, the crystal structure for EBOV VP35$_{85-145}$ revealed 12 protein chains within the asymmetric unit (Zinzula *et al.*, 2019). Each chain formed a continuous α-helix also; the 12 chained structure was formed by the antiparallel association of two trimers. On the other-hand, the oligomerisation domain of the most distantly related to EBOV within the genus, Reston virus (RESTV) VP35 (71-109) revealed a four-helix structure. A salt bridge between Arg 110, Glu 108 and Glu 115 in EBOV and Arg 99, GLu 97 and Glu 104 in RESTV modulates oligomeric states. These residues are conserved amongst ebolaviruses. EBOV VP35$_{85-145}$ mutant R110A formed tetramers in SEC-MALLS and data suggested that EBOV VP35 can form tetramers and trimers as these oligomeric states are thought to be separated by low-energy barriers (Zinzula *et al.*, 2019).

Previous work on the EBOV VP35 IID (221-340) showed the IID to be monomeric in solution by dynamic light-scattering. However, the recent crystal structure of the EBOV VP35 IID (PDB: 3FKE) revealed two monomers within the asymmetric unit. This interaction was thought to be week due to the buried surface area between the two monomers being only 490 Å^2 (Leung *et al.*, 2009), typical of a crystal contact rather than dimerization interface.

The aim of this chapter was to express and purify full-length VP35 from both EBOV and MARV in order to elucidate the native oligomeric state of both proteins using both biophysical techniques and structural work.

5.2.1 Cloning

The cDNA of VP35 from EBOV and MARV was codon optimised for *E. coli* expression, synthesised and cloned into pGEX-6P-2 expression vector by Genewiz. Due to potential oligomerisation states (thought to be either trimer or tetramer) the cDNA of VP35 was also sub cloned into pET-28a-SUMO and pET-MAL-PreScission using restriction enzyme cloning utilising custom designed Bam HI and Xho I recognition sites. GST is naturally a dimer in solution, therefore this may have complicated analysis.

5.2.1.1 DNA Amplifications

DNA amplification of VP35 cDNA was performed via PCR to truncate the N-terminus, based on results from I-TASSER, secondary structure prediction server and previously published CD analysis showing the N-terminal residues of VP35 are intrinsically disordered in solution; disordered domains are often problematic for soluble protein expression (Leung *et al.*, 2015).

Initial 'touchdown' PCR for EBOV VP35$_{79-329}$ did not produce any PCR product. A temperature gradient PCR was therefore performed (Figure 5.1A). Analytical agarose gel revealed the optimal temperature for primer annealing and DNA amplification to be 55°C (Figure 5.1 A, lane 5). PCR amplification for EBOV VP35$_{79-329}$ was repeated on a larger scale (Figure 5.1 B, lane 2). PCR product was digested with Bam HI and Xho I restriction enzymes for ligation with pET-MAL-PreScission plasmid. Ligation product was transformed into DH5α competent cells, single colonies were grown, and DNA extracted prior to plasmid verification by sequencing. Ligation was largely unsuccessful with sequencing often showing 'empty' vectors for EBOV VP35$_{79-329}$. MARV VP35$_{60-329}$ was cloned successfully and verified by sequencing.

As EBOV VP35$_{80-145}$ oligomerisation domain was expressed, purified and subjected to SEC-MALLS analysis previously (Zinzula *et al.*, 2019), and the EBOV VP35 IID crystal structure (Leung *et al.*, 2009), in light of time restraints and problems with EBOV VP35$_{79-329}$ cloning, we decided to focus the project on MARV VP35$_{60-329}$ from here out in order to draw a comparison between the two viral species.
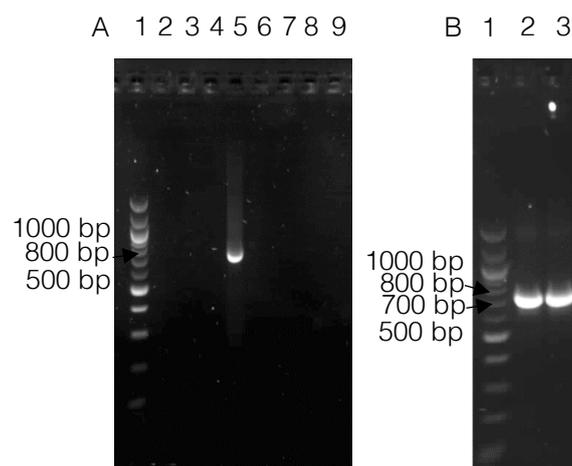

Figure 5.1 PCR Cloning of VP35 Truncations

A- temperature gradient PCR for EBOV VP35$_{79-329}$; 1) 100 bp marker, 2) 45°C, 3) 50°C, 4) 55°C, 5) 60°C, 7) 65°C, 8) 70°C, 9) 75°C.
B- large scale PCR; 1) 100 bp marker, 2) PCR product for EBOV VP35$_{79-329}$, 3) PCR product for MARV VP35$_{60-329}$.
EBOV VP35$_{79-329}$ = 753 bp.
MARV VP35$_{60-329}$ = 810 bp.

5.2.2 Protein Expression of VP35

The MARV VP35 protein oligomerisation domain (60-130) was previously expressed in the literature using an N-terminal histidine purification tag (Bruhn *et al.*, 2017) and EBOV VP35 with a C-terminal His6 purification tag (Zinzula *et al.*, 2019). Moreover, the IID of EBOV VP35 was also expressed and purified with a maltose binding protein (MBP) purification tag. Protein purification therefore needed to be optimised. A protocol was needed to allow for large yields of MARV VP35 protein to be purified in order to perform functional, biophysical and structural studies.

Several possible stages within any expression and purification protocol highlighted as being potentially problematic and responsible for low yields and poor protein quality (purity and homogeneity), including; buffer composition, lysis method, chromatography techniques and concentration methods. Other reasons for low yields include expression strains, growth temperature and induction methods which are discussed below. Codon usage can also affect protein yields; bacteria favour different codon usage for the same amino acid. This phenomenon increases translational efficiency and gene expression by accommodating the codon bias of the organism (Z. Zhou *et al.*, 2016). VP35 (and VP30) constructs worked on in this chapter were all codon optimised for bacterial cells expression in order to overcome this.

For VP35 the standard growth temperatures was chosen (37°C reduced to 18°C for 16 hours) as for most of the previous proteins purified within the Edwards/Barr group were also performed this way. Secondly, 0.5 mM IPTG was also chosen for the same reasoning. The lysis method chosen was cell disruption as it is gentle in comparison to sonication, where samples often heat up, and more reliable and robust when compared to freeze-thaw methods. SEC was also chosen to ensure oligomers of VP35 were separated from aggregates.

5.2.2.1 Small Scale Expression

Expression plasmid expressing WT EBOV and MARV VP35 (including MARV VP35 truncation 60-329) was transformed into all chemically competent expression strains of *E. coli* available within the Edwards/ Barr group. Five constructs were checked for protein expression; EBOV and MARV VP35 in pET-28a-SUMO and pET-MAL-PreScission and MARV VP35$_{60-329}$ in pET-MAL- PreScission only.

Cultures were harvested, and protein extracted using the freeze-thaw method. Insoluble and soluble fractions were run on a 15% acrylamide gel to identify cell strains that expressed the highest levels of soluble fusion protein (~ 75 kDa). Many of the proteins are expressed as insoluble protein. MARV VP35$_{60-329}$ produced higher levels of soluble protein than both full-length constructs. Large scale expression was carried out in BL21 DE3 Gold *E. coli* cells, soluble His6-MBP-fusion expression highlighted by the red box (Figure 5.2 D lane 13, red box). Although high protein expression was also seen for MARV His6-SUMO-VP35 (Figure 5.2 D lane 8), this was largely insoluble, and thought that the NTD would cause downstream purification problems. It was therefore decided to use the truncated MARV His6-MBP-VP35$_{60-329}$ construct.

A — Rosetta 1

1) marker
2) insoluble
3) soluble
4) insoluble
5) soluble
6) insoluble
7) soluble
8) insoluble
9) soluble
10) insoluble
11) soluble
12) insoluble
13) soluble

150 kDa
100 kDa
75 kDa
50 kDa
37 kDa
25 kDa
20 kDa
10 kDa

MARV VP35$_{60-329}$ MBP   MARV VP35 SUMO   EBOV VP35 SUMO

B — Rosetta 2

MARV VP35$_{60-329}$ MBP   MARV VP35 SUMO   EBOV VP35 SUMO

C — Star

EBOV VP35 SUMO   MARV VP35 SUMO   MARV VP35$_{60-329}$ MBP

D — Gold

EBOV VP35 SUMO   MARV VP35 SUMO   MARV VP35$_{60-329}$ MBP

E — Gold   Rosetta 2   Rosetta 1

EBOV VP35 MBP   MARV VP35 MBP

F — Rosetta 1   Rosetta 2   Gold

EBOV VP35 MBP   MARV VP35 MBP

G — Star

EBOV VP35 MBP   MARV VP35 MBP

141

Figure 5.2 Small Scale Expression of VP35 Constructs

A- Rosetta 1 cells; 1) marker, 2-5) MARV VP35$_{60-329}$ pET-MAL; uninduced/insoluble, uninduced/soluble, induced/insoluble, induced/soluble. 6-9) MARV VP35 pET-SUMO; uninduced/insoluble, uninduced/soluble, induced/insoluble and induced/soluble, 10-13) EBOV VP35 pET-SUMO; uninduced/insoluble, uninduced/soluble, induced/insoluble and induced/soluble.

B- Rosetta 2 cells; 1) marker, 2-5) MARV VP35$_{60-329}$ pET-MAL; uninduced/insoluble, uninduced/soluble, induced/insoluble, induced/soluble. 6-9) MARV VP35 pET-SUMO; uninduced/insoluble, uninduced/soluble, induced/insoluble and induced/soluble, 10-13) EBOV VP35 pET-SUMO; uninduced/insoluble, uninduced/soluble, induced/insoluble and induced/soluble.

C- Star cells; 1) marker, 2-5) EBOV VP35 pET-SUMO; uninduced/insoluble, uninduced/soluble, induced/insoluble and induced/soluble, 6-9) MARV VP35 pET-SUMO; uninduced/insoluble, uninduced/soluble, induced/insoluble and induced/soluble, 10-13) MARV VP35$_{60-329}$ pET-MAL; uninduced/insoluble, uninduced/soluble, induced/insoluble, induced/soluble.

D- Gold cells; 1) marker, 2-5) EBOV VP35 pET-SUMO; uninduced/insoluble, uninduced/soluble, induced/insoluble and induced/soluble, 6-9) MARV VP35 pET-SUMO; uninduced/insoluble, uninduced/soluble, induced/insoluble and induced/soluble, 10-13) MARV VP35$_{60-329}$ pET-MAL; uninduced/insoluble, uninduced/soluble, induced/insoluble, induced/soluble.

E- VP35 pET-MAL constructs; 1) marker, 2-5) EBOV VP35 Gold cells; uninduced/insoluble, uninduced/soluble, induced/insoluble, and induced/soluble. 6-13) MARV VP35 6-9) Rosetta 2; uninduced/insoluble, uninduced/soluble, induced/insoluble, and induced/soluble, 10-13) Rosetta 1 cells; uninduced/insoluble, uninduced/soluble, induced/insoluble, and induced/soluble.

F- VP35 in pET-MAL constructs; 1) marker, 2-5) EBOV VP35, Rosetta 1 cells; uninduced/insoluble, uninduced/soluble, induced/insoluble, and induced/soluble, 6-9) EBOV VP35, Rosetta 2; uninduced/insoluble, uninduced/soluble, induced/insoluble, and induced/soluble, 10-13) MARV VP35, Gold cells; uninduced/insoluble, uninduced/soluble, induced/insoluble, and induced/soluble.

G- VP35 in pET-MAL constructs, star cells; 1) marker, 2-5) EBOV VP35; uninduced/insoluble, uninduced/soluble, induced/ insoluble, uninduced/soluble, 6-9) MARV VP35; uninduced/insoluble, uninduced/soluble, induced/ insoluble, uninduced/soluble.

VP35 = 35 kDa

MARV VP35$_{60-329}$ = 29 kDa

His6-SUMO = 13 kDa

His6-MBP = 45 kDa

Key: - - divider between EBOV and MARV

Uninduced sample

Induced sample

5.2.2.2 Large Scale Expression

MARV VP35$_{60-329}$ (His6-MBP fusion protein) expressed well in Gold *E. coli* cells and was used for large scale protein expression. 1 L of Gold cells expressing MARV VP35$_{60-329}$ was induced with 0.5 mM IPTG and expressed as before. *E. coli* cells were harvested by centrifugation and lysed with VP35 specific lysis buffer and cell disruption.

5.2.2.3 Affinity Chromatography 1

VP35 was expressed as a fusion protein with an N-terminal His6 purification allowing for lysate purification via immobilised metal affinity chromatography, using nickel ions in a HisTrap column. Soluble lysate was flowed over the immobilised ions to allow histidine binding (*Figure 5.4*A lane 2-3). The column was washed (Figure 5.4 A lane 4-7), and protein eluted with increasing concentrations of imidazole (Figure 5.4 A lane 8-10). High salt washes were performed to remove potential expression host bound RNA. Purity was analysed by 15% SDS PAGE gel (Figure 5.4 A).

5.2.2.4 Fusion Tag Cleavage and Dialysis

Eluates containing fusion protein were checked for a 260/280 ratio below 1 indicating RNA free protein was purified (Table 5.1). These eluates were pooled and cleaved for 16 h with the addition of PreScission protease that cleaves at the 3C recognition site producing His6-MBP and VP35$_{60-329}$. Cleavage occurred in dialysis buffer to remove the imidazole for further purification (Figure 5.4 B lane 1).

Table 5.1 260/280 ratios for MARV VP30$_{60-329}$

Eluate	260/280
150 mM	0.86
300 mM	0.42
500 mM	0.45

5.2.2.5 Affinity Chromatography 2

Cleaved protein (Figure 5.4 B, lane 1) was purified on a second HisTrap in order to remove the purification tag from VP35. VP35 appeared slightly 'sticky' here and needed to be eluted with 25-50 mM imidazole (Figure 5.4 B, lane 3 and 4 (common in the Edwards group), some remaining till the final 500 mM wash, eluting with the purification tag (Figure 5.4 B, lane 6). The washes were combined to optimise protein yield. SEC was used to further purify VP35.

5.2.2.6 Size-Exclusion Chromatography

Sample was concentrated to 10 mL (from 50 mL) and injected onto a HiLoad 26/600 Superdex S-200 size exclusion column attached to an AKTA purifier pump measuring absorbance at 280 nm. The chromatogram revealed 4 peaks. Only 2 peaks were expected, 1 for homogenous VP35 (monomer ~29 kDa or an oligomer (trimer (~87 kDa) to hexamer (~174 kDa)) and second for the purification tag (45 kDa). Peak 1 represents aggregated protein eluted at the column void volume. Whereas the following peaks were estimated to contain, larger oligomers of MARV VP35$_{60-329}$ (peak 2) and hexameric VP35$_{60-329}$ (peak 3). Peak 4 contained the His6-MBP purification tag based on the correct size band on SDS-PAGE (Figure 5.4, D).

MARV VP35$_{60-329}$ purification techniques produced moderate yields (~5 mg of protein per 4 L of expression culture) of protein reproducibly and not optimised further.

Fractions corresponding to peak 3 at ~145 mL gave rise to MARV VP35$_{60-329}$ bands on SDS PAGE with no other contaminating bands visible. In comparison to the calibration chromatogram, this oligomer of MARV VP35$_{60-329}$ elutes with an expected molecular weight of ~177 kDa (antilog 2.25 (Figure 5.3 F)), suggesting that VP35 is forming hexamers in solution (Figure 5.4 E). MARV VP35$_{60-329}$ elutes at a higher elution volume than expected for a trimer or tetramer. However, size exclusion elution volumes are influenced by molecular shape as well as size. The observed elution volume could potentially be due to a non-spherical shape of VP35. The coiled-coil oligomerisation domain may elongate the protein such that it is non-spherical, and this could influence elution volume and therefore the interpretation of oligomeric state. Long rod-shaped proteins are known to elute at higher than expected volumes from SEC columns (Hong, Koza and Bouvier, 2012). Moreover, electrostatic interactions may also play role here. Electrostatic interactions may occur between protein and packing materials, at low ionic strength Superdex can cause retention of proteins due to negatively charged groups. However, 500 mM NaCl present in the SEC buffer makes this unlikely. Oligomeric state was therefore further investigated using complementary biophysical techniques.

A

1) marker
2) soluble
3) flow-through
4) wash 1
5) wash 2
6) high salt wash 1
7) high salt wash 1
8) 150 mM imidazole
9) 300 mM imidazole
10) 500 mM imidazole

150 kDa
100 kDa
75 kDa
50 kDa
37 kDa
25 kDa
20 kDa
10 kDa

fusion protein

B

1) marker
2) dialysed
3) flow-through
4) 25 mM imidazole
5) 50 mM imidazole
6) 500 mM imidazole

150 kDa
100 kDa
75 kDa
50 kDa
37 kDa
25 kDa
20 kDa
10 kDa

cleaved VP35

C

1) marker
2) loaded sample
3) peak 1
4 5) peak 2
6 7 8) peak 3
9 10 11) peak 4

150 kDa
100 kDa
75 kDa
50 kDa
37 kDa
25 kDa
20 kDa
10 kDa

peak 3 - pure VP35

Figure 5.3 Large Scale Purification of MARV VP35$_{60-329}$

A- 1st HisTrap HP column; 1) marker, 2) soluble lysate, 3) flow-though, 4 & 5) washes, 6 & 7) high salt washes, 8) 150 mM imidazole, 9) 300 mM imidazole, 10) 500 mM imidazole

B- 2nd HisTrap HP column; 1) marker, 2) dialysed, 3) flow-through, 4) 25 mM imidazole wash, 5) 50 mM imidazole elute, 6) 500 mM imidazole wash

C- SEC; 1) marker, 2) inject, 3) peak 1, 4 & 5) peak 2, 6-8) peak 3, 9-11) peak 4

D- SEC chromatogram. 10 mL of cleaved MARV VP35$_{60-329}$ was injected onto a pre-equilibrated Superdex S200 26/600 column. Absorbance recorded at 280 nm. 3 peaks were seen after the void volume (90 mL, labelled 2, 3 and 4). Pure protein was seen in peak 3 confirmed in C.

E- SEC calibration. Superdex S200 26/600 column calibrated with 25 mM Tris-HCl pH 8.0, 250 mM NaCl, 2.5% glycerol; 1) blue dextran 2000 20,000 kDa, 2) thyroglobulin 669 kDa, 3) ferritin 440 kDa, 4) aldolase 158 kDa, 5) conalbumin 75 kDa, 6) ovalbumin 43kDa

F- SEC calibration curve. Log10 kDa plotted against elution volume for each standard protein in E. Lines represent elution volume of MARV VP35$_{60-329}$

Fusion protein = 75 kDa

Purification tag = 45 kDa

MARV VP35$_{60-329}$ = 29 kDa

5.2.3 Oligomeric State Determination

In the literature, the crystal structure of the MARV VP35$_{60-130}$ oligomerisation domain models the domain as an elongated trimer. This contradicts our data for MARV VP35$_{60-329}$ suggesting that MARV VP35 is a hexamer.

VP35 and VP30 are known to interact with each other (Biedenkopf *et al.*, 2016), however it is unknown whether this interaction is dependent on RNA. VP30 has also been suggested to be a hexamer in solution (Hartlieb *et al.*, 2007). VP35 could therefore interact with hexameric VP30 in a 1:1 manner if a hexamer or a 2:1 manner if a trimer. However, structural and functional similarities (i.e. evolutionary similarities) between hRSV P and M2-1 vs VP35 and VP30 would suggest VP35 and VP30 might both be tetramers if the interaction between VP35 and VP30 mimicked that of hRSV P and M2-1. The aim of the experiments described below was to resolve these arguments.

5.2.3.1 Mass-Spectrometry (MS)

To confirm the oligomeric state of VP35 in solution, mass spectrometry analysis was performed as a service by Rachel George (Mass Spectrometry Facility, University of Leeds). Native liquid chromatography electro-spray ionisation mass spectrometry (LC-ESI-MS) was performed, with values reported as mass-to-charge (M/Z) ratio. The predominant peak at 120 kDa corresponded to a tetramer of VP35 (monomer 29.5 kDa) (Figure 5.4 A and B).

5.2.3.2 Size Exclusion Chromatography coupled to Multi Angle Laser Light Scattering (SEC-MALLS)

The oligomeric state of VP35 was also analysed by SEC-MALLS. SEC-MALLS combines size exclusion chromatography with multi angle laser light scatting analysis to determine molar mass in solution with a 10% inaccuracy rate and was performed as a service by Maria Nikolova (PhD student, University of Leeds). SEC-MALLS showed 1 oligomeric peak for MARV VP35$_{60-329}$ at 108 kDa ± 10%, representative of a tetramer (Figure 5.4 C and D).

Based on combined data from, MS and SEC-MALLS, MARV VP35$_{60-329}$ is tetrameric.
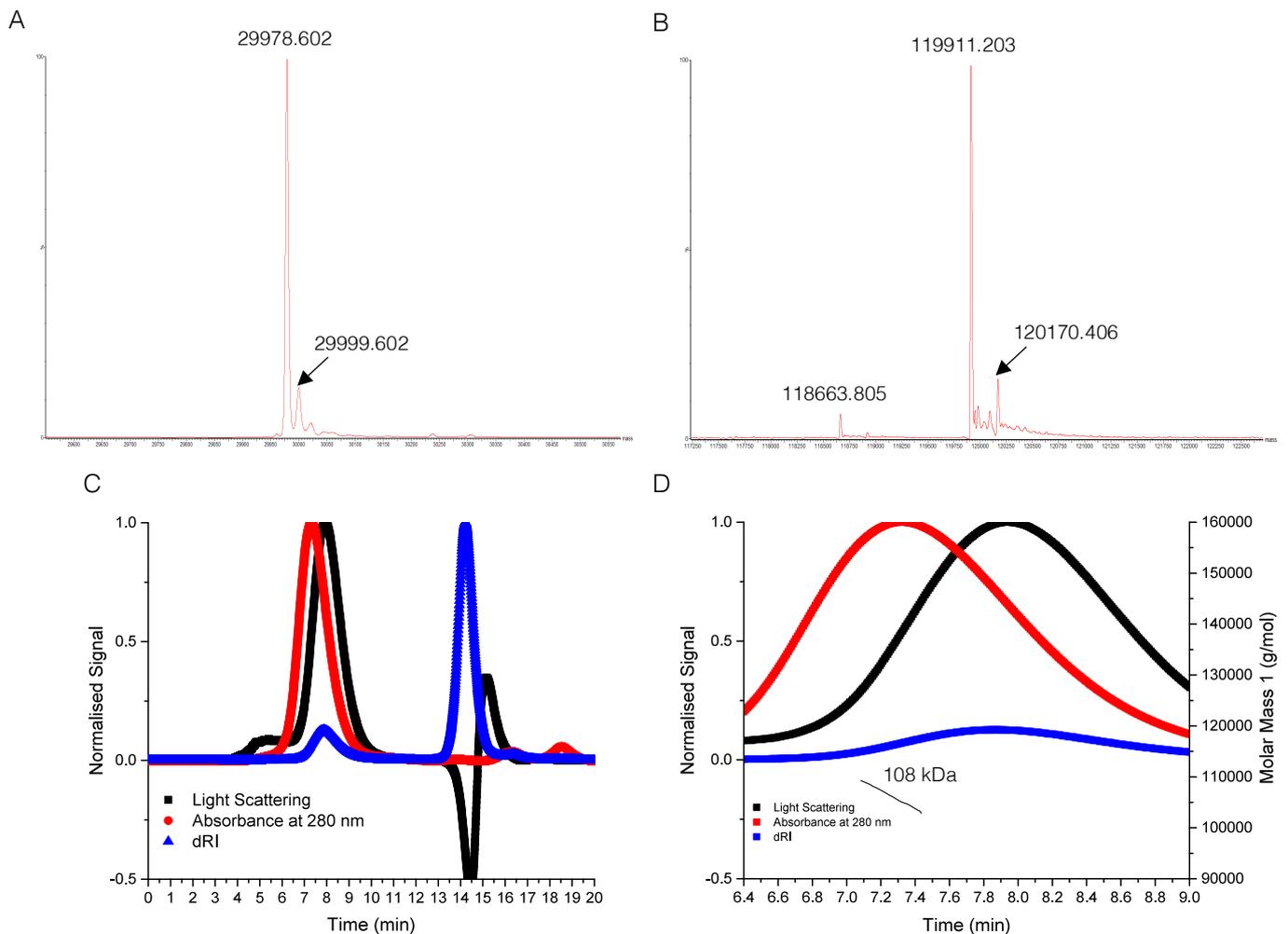
Figure 5.4 Oligomeric State Determination of MARV VP35$_{60-329}$

A and B- M/Z spectrum using electrospray ionization mass spectrometer, performed by Rachel George in the Mass Spectrometry facility, University of Leeds.
A- non-native mass spectrometry shows monomeric VP35 at 29.5 kDa.
B- native mass spectrometry shows 120 kDa tetramers present in the protein sample
C and D – SEC-MALLS chromatograms using Superdex 200 5/150 column
C- SEC-MALLS showed a single oligomeric state present in the protein sample
D- peak fitting showed presence of an oligomer of 108 kDa (± 10%)
dRI = differential refractive index

5.2.4 Functional Studies

VP35's IFN-antagonist function directly correlates with the ability of VP35 to bind dsRNA. Double-stranded viral RNA is produced during the replication cycle of EBOV and MARV. Functionally active VP35 is also a component of the Filovirus replication complex acting as the polymerase co-factor. In the literature VP35 has been shown to bind pC ssRNA weakly, and poly-IC dsRNA with high affinities (Cárdenas *et al.*, 2006), therefore 3'FI poly-IC (12 nucleotides) was used to determine direct binding affinities.

5.2.4.1 RNA Binding

In a fluorescence anisotropy assay, MARV VP35$_{60-329}$ was assessed for binding to dsRNA using 3'Fl labelled poly-I-C RNA. The poly-I-C 12mer RNA forms dsRNA that is either 6 basepairs long (1 segment of RNA interacts with itself) or 12 bases long due to dimerization with a second RNA strand forming dsRNA. MARV VP35$_{60-329}$ was concentrated to high concentrations of 10 mg/mL (86 μM) however no binding was seen (Figure 5.5 A). MARV VP35$_{60-329}$ was then subjected to a longer RNA (64 nucleotides, although apart from two 3' hairpin structures is largely single stranded) EBOV RNA$_{54-118}$ and pC 13 mer. Again, no RNA binding was seen (Figure 5.5 B). SV NP direct binding to 3'Fl RNA pC was tested to ensure the TecanSpark10m was working correctly this day, the control showed protein:RNA binding and it was therefore assumed MARV VP35$_{60-329}$ did not bind RNA.
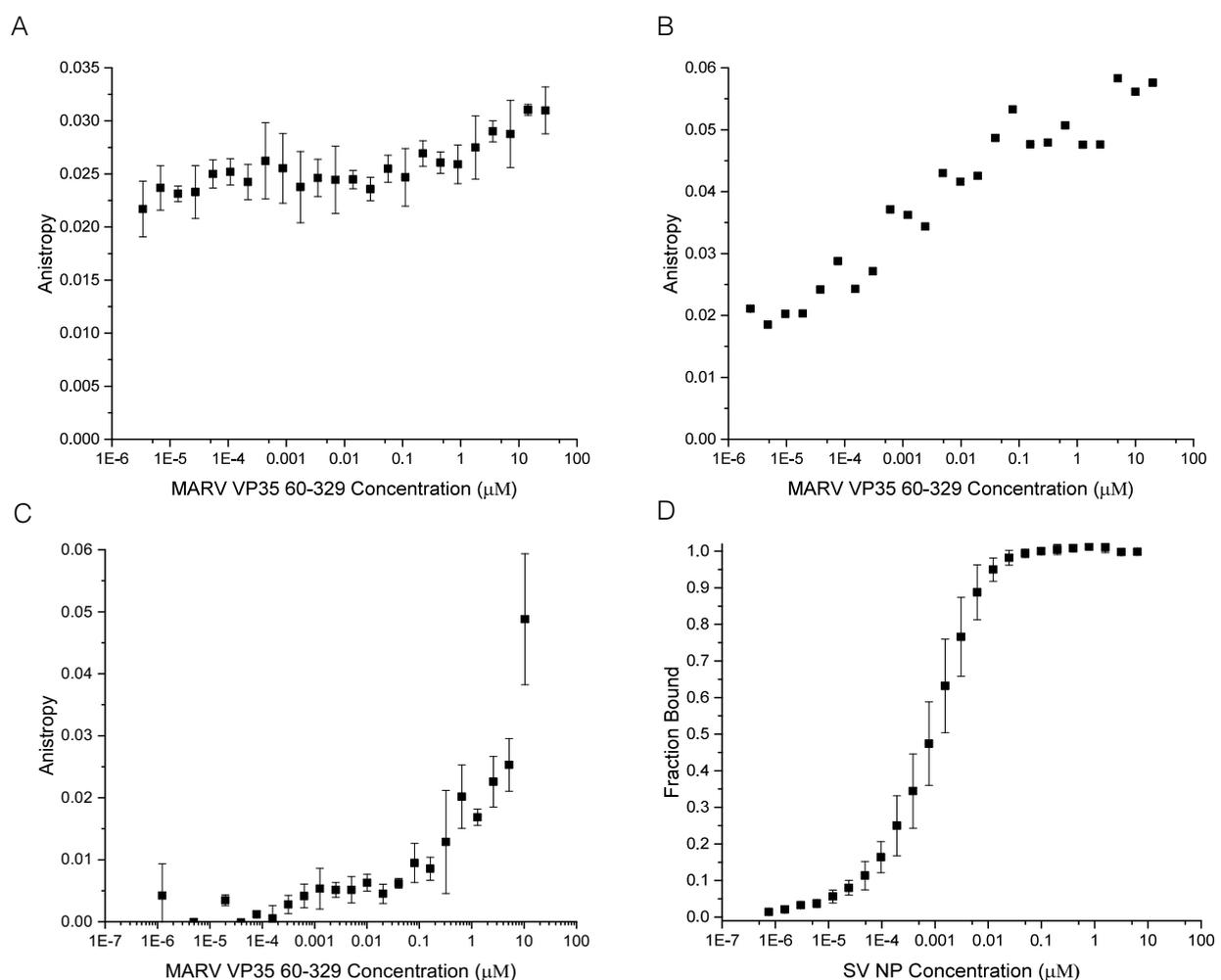

Figure 5.5 Fluorescence Anisotropy of MARV VP35$_{60-329}$ to pC and pIC RNA

A- direct binding to double stranded 3'Fl RNA poly-IC.
B- direct binding to single stranded 3'Fl EBOV$_{54-118}$ RNA.
C- direct binding to single stranded 3'Fl pC RNA.
D- SV NP direct binding control to single stranded 3'Fl RNA pC.

5.2.5 Protein Quality Check

It was suggested that the RNA binding assay between MARV VP35$_{60\text{-}329}$ and poly-IC RNA was not successful as the protein sample was not correctly folded or unstable. Therefore, circular dichroism analysis and thermal unfolding was performed to check for unfolded MARV VP35$_{60\text{-}329}$ and if it was unstable at room temperature.

5.2.5.1 Circular Dichroism and Thermal Unfolding

CD was performed as a service by Nasir Khan, University of Leeds. Pure protein was diluted to 0.1 mg/mL in 50 mM Na$_3$PO$_4$ pH 7.5. Analysis revealed a typical α-helical structure (Figure 5.6 A), with troughs at 208 and 222 nm. The ratio between 222 nm and 208 nm is greater than 1 (1.16) indicating the presence of a coiled coil domain. Moreover, MARV VP35$_{60\text{-}329}$ had a melting temperature of 44°C (Figure 5.6 B). An α-helical CD, indicating the presence of the coiled-coil oligomerization domain trace along with 44°C melting temperature suggest that the protein is indeed folded. The reasoning behind the RNA binding assay not working was therefore thought to be possibly due to the length of the (ds)RNA.
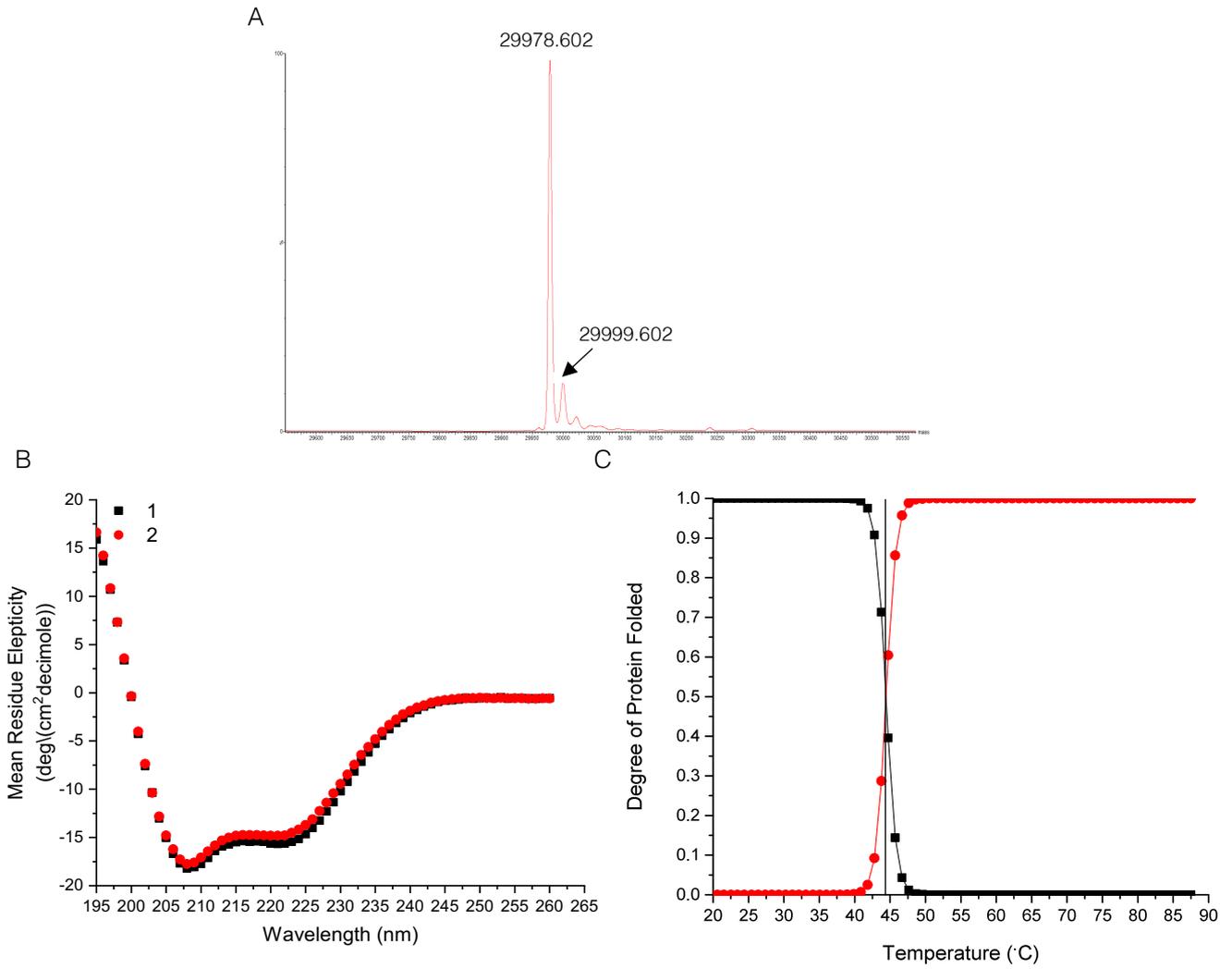
Figure 5.6 MARV VP35$_{60-329}$ Protein Quality Check

A- non-native mass spectrometry shows presence of a 29 kDa monomer in the sample.
B- circular dichroism shows folded protein that is a-helical, experiment conducted in duplicate (black and red)
C- thermal melt analysis shows 44°C melting temperature, where 1 represents complete folding and 0 completely unfolded protein.

5.2.6 Structural Studies

5.2.6.1 X-Ray Crystallography

Pure and tetrameric MARV VP35$_{60-329}$ was concentrated to 10 mg/mL for structural analysis by X-ray crystallography. Crystals were not seen in any conditions using the sitting drop vapour diffusion method. Most conditions caused protein precipitation (Figure 5.7).
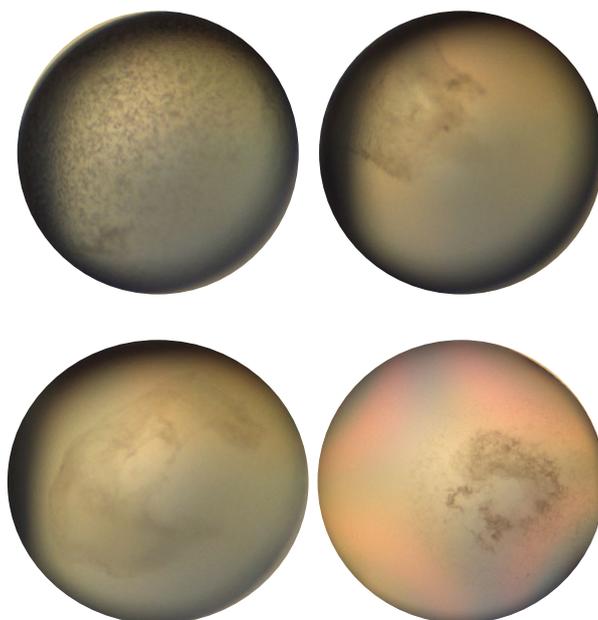

Figure 5.7 Precipitation formed in MARV VP30$_{60-329}$ Crystallisation Trials

MARV VP35$_{60-329}$ was concentrated to 10mg/mL prior to crystallization. Representative drop images from JCSG Core 1-4 crystallization trials shows protein precipitation.

5.2.6.2 Electron Microscopy

As no crystals were formed, we attempted to visualise VP35 using negative stain transmission electron microscopy due to the minimal amount of materials required. Previous electron microscopy on 106 kDa tetramer appeared to have a particle size of ~10nm (SIMV NP, unpublished data, Dr. Francis Hopkins, University of Leeds) and this was used as a guide when identifying MARV VP35$_{60-329}$ particles.

5.2.6.3 Micrograph Collection

MARV VP35$_{60-329}$ aggregated at higher concentrations 0.05-0.1 mg/mL, possibly due to the interaction with the UA on the EM-grid. Therefore sample-preparation was optimised at 0.01 mg/mL (Figure 5.8 A-C) although particles were sparse. From this grid, micrographs were collected for data analysis.

5.2.6.4 Particle Picking and Class Averages

200 particles were picked manually from 25 micrographs prior to auto-picking in RELION3 (MRC-LMB, University of Cambridge (Zivanov *et al.*, 2018)). 1388 auto-picked particles were sorted into 25 2D classes using RELION3 (Zivanov *et al.*, 2018). No classes appeared to be the correct size for a tetramer of ~116 kDa, as all classes appeared >10 nm (Figure 5.8 D).
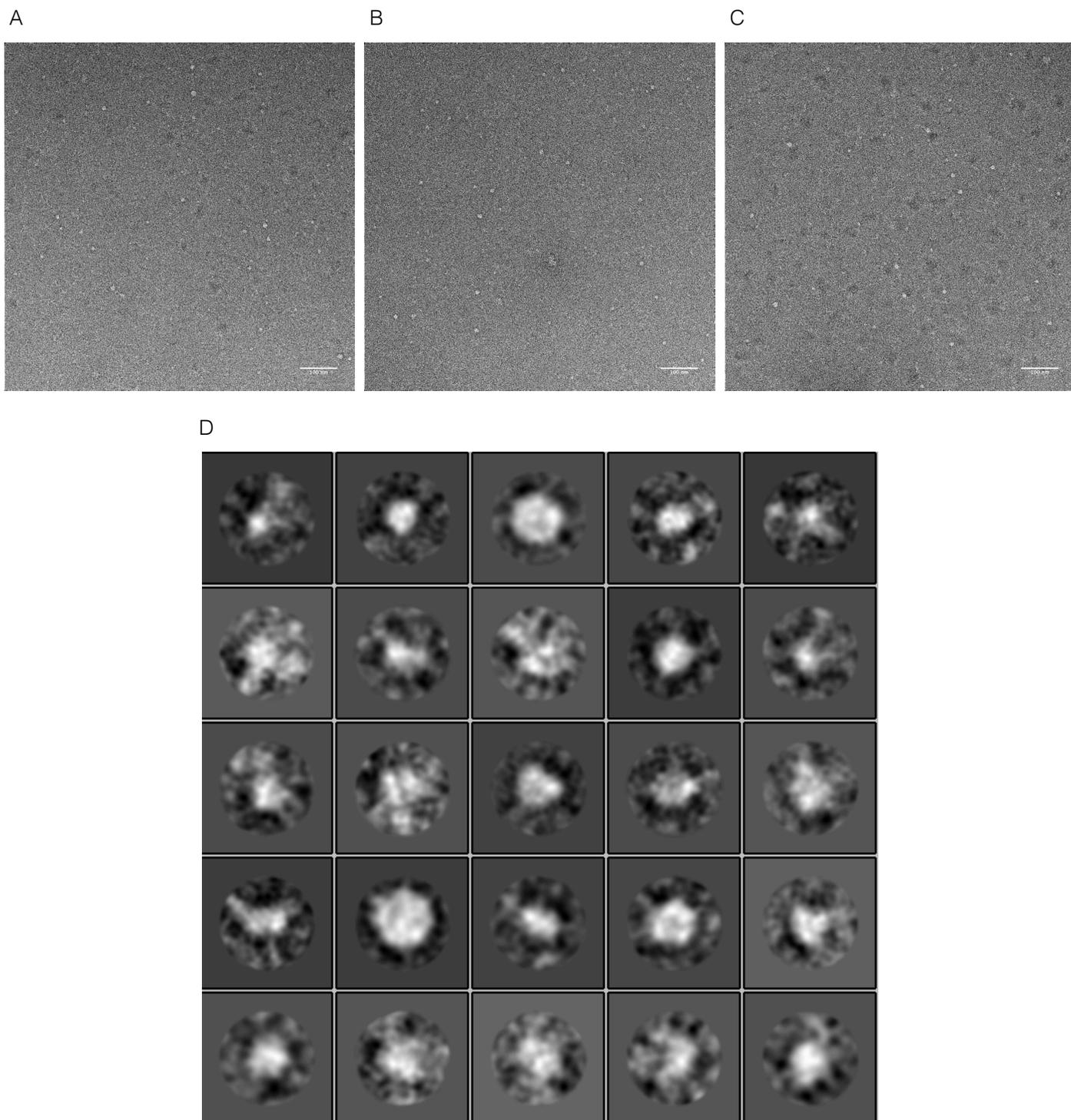
Figure 5.8 MARV VP35$_{60\text{-}329}$ Negative Stain Electron Microscopy

A-C- micrographs taken at 49,000 K, 120 kV and -2.0 defocus, with 0.01 mg/mL protein.
D- 1388 particles were 2D class averaged into 25 classes.
Scale bar 100 nm.

5.3 Discussion

5.3.1 Protein Expression and Purification of VP35

In the literature there is no purification protocol for full-length EBOV or MARV VP35, this is probably because the VP35 N-terminus is highly flexible and this potentially makes the protein insoluble. Data here suggests that VP35 is still largely insoluble even when fused to a large soluble purification tag (MBP, ~44 kDa). Thus, truncating these flexible N-terminal residues increases the solubility of VP35 (Figure 5.2 D lane 13).

We now have an optimised protocol for producing MARV VP35$_{60\text{-}329}$. Protein produced in these experiments was soluble, folded and thermally stable to 44°C. Moreover, homogeneous sample was produced as determined by SEC, SEC-MALLS and MS.

MARV VP35$_{60\text{-}329}$ purification took two days, this method was quick, and no protein degradation was seen when analysed by SDS-PAGE (Figure 5.4 A, B and C), high quality protein was also confirmed by the typical α-helical curves seen by CD and high melting temperatures (Figure 5.6 B and C respectively).

5.3.2 Elucidating the Oligomeric State of VP35

It is difficult to determine the entire structure of VP35 due to the coiled-coil oligomerisation domain and lack of regular secondary structures in larger stretches of the protein.

It was thought that a hexamer was initially eluted from SEC (Figure 5.4 D peak 3). However, due to the non-globular shape of MARV VP35 (the N-terminal oligomerisation domain structures for both EBOV and MARV VP35 reveal elongated coiled-coils), this interpretation was questioned and further investigated. MARV VP35 oligomerisation domain (60-135) was previously shown to form trimers in the crystal lattice (Bruhn *et al.*, 2017). However, data presented in this thesis gives evidence that MARV VP35$_{60\text{-}329}$ is tetrameric when analysed by native MS and SEC-MALLS. This data coincides with previously published work suggesting that EBOV VP35$_{80\text{-}340}$ is also tetrameric (Zinzula *et al.*, 2019). Further, hRSV P is also tetrameric and shares sequence homology (71% similar to MARV VP35) (Simabuco *et al.*, 2011) and functions similarly to filovirus VP35.

Tetrameric MARV VP35 is supported by the current computational model for VP35, produced from incorporating the crystal structure of EBOV VP35 NTD, oligomerisation domain and CTD, with computed RMSD of 1.20 Å, 1.97 Å and 0.75 Å respectively. Banerjee *et al.*, assessed and analysed the stability and dynamics of the EBOV VP35 structure using molecular dynamics which indicated symmetric behaviour across the tetramer (Banerjee and Mitra, 2020).
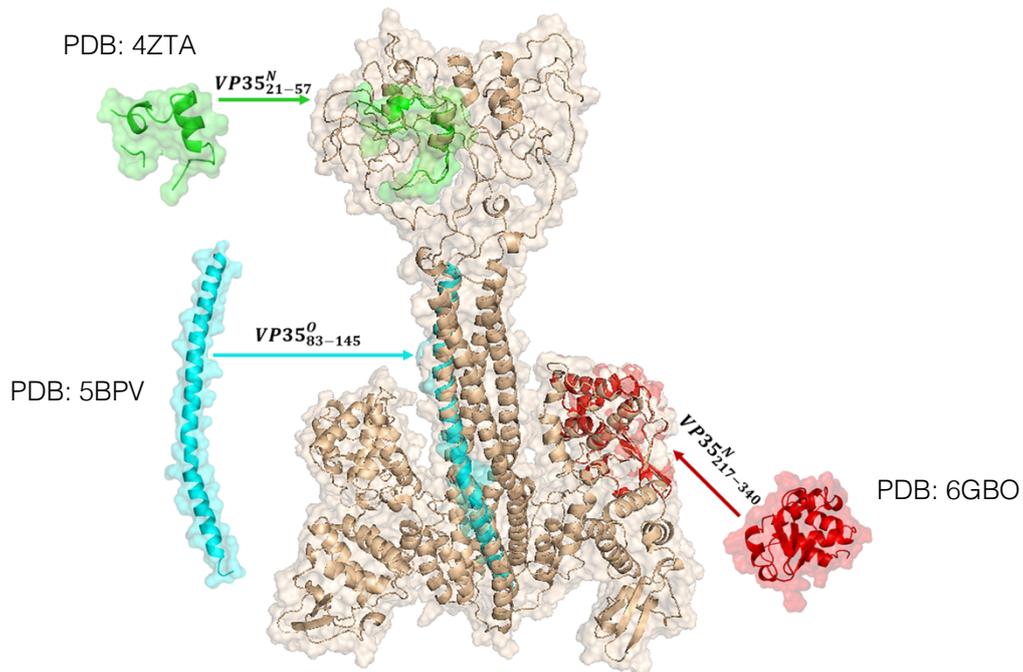
Figure 5.9 Proposed Model for Homotetrameric VP35

Adapted from Banerjee *et al*., 2020.
Computational model of homotetrameric VP35. Piecewise crystal structure of VP35$_{21-57}$ (green),
VP35$_{217-340}$ (red) and VP35$_{83-145}$ (cyan) are used to model the assembly.

5.3.3 Structural Analysis

Initial micrographs in this thesis appeared to show tetrameric particles of MARV VP35$_{60-329}$ of an estimated size of 10 nm (Figure 5.8 A-C). Upon analysis in RELION however, particle size looked much larger than that of a tetramer and particles were sparsely spread across the 25 classes. More micrographs and in turn particles would need to be picked to allow for a reliable 2D class averages and an initial 3D reconstruction prior to cryo-EM.

One interpretation of the lack of crystals and the poor-quality EM grids is that, despite the appropriate CD, MS and SEC-MALLS data, the protein is in fact not completely correctly folded and/or is aggregating post purification, and this may explain the lack of detectable RNA binding.

5.4 VP30

The crystal structure of EBOV VP30 C-terminal domain (142-272 (PDB: 2I8B)) revealed a dimer in the asymmetric unit (Hartlieb *et al.*, 2007). However, full length VP30 has been shown to form hexamers *in vitro* and *in vivo.* EBOV VP30$_{89\text{-}272}$ (~25 kDa monomer, inclusive of the CTD) eluted as a ~150 kDa hexamer during size-exclusion chromatography when compared to the marker protein aldolase (also ~150 kDa). Secondly, during a cross-linking experiment a hexamer of ~150 kDa was also formed. It is therefore thought that, in addition to the C-terminal dimer interface, another EBOV VP30 oligomerisation domain is N-terminal comprising residues 94-112 (Hartlieb *et al.*, 2007).

The aim of this part of the chapter was to further validate the oligomeric state of VP30 and gain understanding of VP30's RNA binding functionality through fluorescence anisotropy.

5.4.1 Cloning

EBOV VP30 cDNA was purchased truncated (8-272). Deletions of 7 N-terminal residues, and 16 C-terminal residues was shown to enhance stability and solubility for recombinant protein expression (Hartlieb et al. 2007). MARV VP30 was also purchased truncated (17-273), to mimic the EBOV VP30$_{8\text{-}272}$ construct.

The cDNA of VP30 from EBOV and MARV was cloned into pET-28a-SUMO (His6-SUMO tag), pGEX-6P-2 (GST tag) and pET-MAL-Precission (His6-MBP tag) expression vectors using standard cloning techniques utilising the custom designed Bam HI and Xho I restriction enzyme recognition sites (Figure 5.11 A).

Figure 5.10 I-TASSER Secondary Structure Predictions for EBOV VP30

Functional regions have been mapped to EBOV VP30. N-terminal RNA binding domain (26-40, red), overlaps with 2 serine clusters (29-31, light blue). Zinc binding domain (Cys3-His motif, yellow) and phosphorylation sites Thr143/146 (green) are deemed important for RNA binding. C-terminal domain (142-272, cyan) encompasses an M2-1 core like domain (107-197, purple) and an NP interacting domain (180-187, pink).

Figure 5.11 VP30 Cloning

A- cut and paste cloning of EBOV and MARV VP30, digested plasmid; 1) 1 kb marker, 2-5) failed DNA purification, 6) MARV VP30$_{17-273}$ pET-28a-SUMO, 7) MARV VP30$_{17-273}$ pGEX-6P-2, 8) empty plasmid, 9) MARV VP30$_{17-273}$ pET-MAL-Precission,10) EBOV VP30$_{8-272}$ pET-28a-SUMO, 11) EBOV VP30$_{8-272}$ pGEX-6P-2, 12) EBOV VP30$_{8-272}$ pET-MAL-Precission

B- PCR amplification of EBOV VP30 69-273 and 87-265 and MARV VP30 79-273 and 101-273; 1) 1 kb marker, 2) failed PCR, 3-4) EBOV VP30$_{69-272}$, 5-7) EBOV VP30$_{87-265}$, 8-10) empty lanes, 11-13) MARV VP30$_{101-273}$, 14-16) MARV VP30$_{79-273}$

C- re-digested plasmid of VP30 truncation cloning; 1) 1 kb marker, 2-8) empty plasmid, 10) EBOV VP30$_{69-272}$ pET-SUMO-28a, 11) EBOV VP30$_{69-272}$ pGEX-6P-2, 12) EBOV VP30$_{87-265}$ pET-SUMO-28a, 13) empty plasmid, 14) EBOV VP30$_{87-265}$ pGEX-6P-2, 15) MARV VP30$_{79-273}$ pET-28a-SUMO, 16) empty plasmid, 17) MARV VP30$_{79-273}$ pGEX-6P-2

5.4.1.1 PCR Amplification for VP30 Truncations

Initial small-scale expression trials highlighted EBOV VP30$_{8-272}$ to be insoluble (Figure 5.12 A and B). Therefore, truncations of VP30, based on results from J-PRED secondary structure predictions and I-TASSER structure prediction servers, were performed via PCR. Truncations were initially cloned into pET-28a-SUMO and pGEX-6P-2 as these were the most commonly used vectors in the Edwards/Barr research group; after small scale expression tests all truncations were cloned into pET-MAL-PreScission.

5.4.1.2 Restriction Digest of Ligated Constructs

To analyse whether VP30 cDNA had been successfully ligated into the expression vectors, DNA extracted from transformation colonies were re-digested (Figure 5.11 C) with Bam HI and Xho I restriction enzymes. Colonies with the correct size insert band were sequence verified.

After successful cloning was completed into expression vectors, E. coli strains optimised for protein expression were transformed for small scale expression trials.

5.4.2 Protein Expression and Purification of VP30

EBOV VP30$_{8-272}$ was previously expressed and purified from E. coli with a maltose binding protein (MBP, ~44 kDa) purification tag fusion protein (Biedenkopf *et al.*, 2016). Initial expression trials were performed using smaller purification tags including His6-SUMO (~13 kDa) and GST (~26 kDa) as these tags were commonly used with the Edwards/Barr research group and were thought would yield higher levels of folded and stable VP30 protein.

As discussed, there are several stages within the expression of recombinant protein protocol that could be optimised for higher yields of proteins. During initial expression conditions including growth temperature, IPTG concentration and addition of co-factors may produce more (or less) soluble protein.

5.4.2.1 Small Scale Expression

A single colony was used to inoculate 5 mL of starter culture. 1 mL of starter culture was used to inoculate 10 mL of media and protein expression was induced at OD$_{600}$ 0.6-0.8 with varying concentrations of IPTG (Figure 5.12 C). Due to differing concentrations of IPTG (0.1-1 mM) making a minimal difference to soluble protein production, induction with 0.4 mM IPTG and 0.05 – 1mM ZnSO$_4$ was tested as zinc was thought to help to stabilise VP30 due to its ZBD (Cys3-His motif) (Figure 5.12 D). Soluble protein extraction was analysed on a 15% SDS-PAGE gel. Truncations based on secondary structure predictions (I-TASSER) were constructed and also tested for soluble protein expression with a His6-SUMO and GST tag for both EBOV VP30 (Figure 5.12 E and F) and MARV VP30 (Figure 5.13 B and C).

Initial small-scale expression of EBOV VP30 (Figure 5.12) and MARV VP30 (Figure 5.13) using His6-SUMO and GST tags however did not reveal the soluble fusion protein expected. Although MARV VP30 constructs seemed more soluble than EBOV VP30 further optimisation was required for further experiments where higher yields are required.

5.4.2.2 Optimisation of Small-Scale Expression

Initial small-scale expression trials in R2 and Gold cells with differing IPTG concentrations, $ZnSO_4$ concentrations and growth temperatures did not produce high yields of soluble recombinant protein, and most overexpressed EBOV and MARV VP30 remained insoluble. Lemo21 cells were used in order to fine tune soluble protein expression. Lemo21 cells express T7 lysozyme from a second plasmid, when induced with rhamnose. Increased concentrations of rhamnose upon protein expression induction increases the production of T7 lysozyme which inhibits the RNA polymerase, slowing recombinant protein expression. Protein production in Lemo21 cells was induced with 0-2000 μM rhamnose, 0.5 mM IPTG and 0.1 mM $ZnSO_4$ and incubated at 37°C for 16 hours (Figure 5.14 A and B). It was initially though that some protein expression was seen when using 2000 μM of rhamnose for GST-VP30$_{8-272}$ (Figure 5.14 A, red box), however due to the size of the T7 lysozyme (34 kDa) it is difficult to draw a true comparison without and un-induced sample. No band on the SDS-PAGE for EBOV His6-SUMO-VP30$_{87-265}$ (Figure 5.14 B) was seen, possibly due to the size of the purification tag.

Soluble MARV VP30$_{17-273}$ and -$_{79-273}$ was seen for the 3 constructs in R2 and Gold cells (Figure 5.15 A and B respectively). As an attempt to further increase soluble expression, MARV VP30 was transformed into Lemo21 cells and induced with 0-2000 μM rhamnose, 0.4 mM IPTG and 0.1 mM IPTG and incubated at 37°C. However, again, only small amounts of soluble recombinant fusion protein were produced.

A — EBOV GST-VP30$_{8-272}$ (R2, Gold); marker lanes: 1) marker, 2) insoluble, 3) soluble, 4) insoluble, 5) soluble. Markers: 150 kDa, 100 kDa, 75 kDa, 50 kDa, 37 kDa, 25 kDa, 20 kDa.

B — EBOV His6-SUMO-VP30$_{8-272}$ (R2, Gold); lanes: 1) marker, 2) insoluble, 3) soluble, 4) insoluble, 5) soluble. Markers: 150 kDa, 100 kDa, 75 kDa, 50 kDa, 37 kDa, 25 kDa, 20 kDa.

C — EBOV His6-SUMO-VP30$_{8-272}$, IPTG, R2; 0.1 mM, 0.4 mM, 0.8 mM, 1 mM; lanes: 1) marker, 2) insoluble, 3) soluble, 4) insoluble, 5) soluble, 6) insoluble, 7) soluble, 8) insoluble, 9) soluble, 10) insoluble, 11) soluble, 12) insoluble, 13) soluble. Markers: 100 kDa, 75 kDa, 50 kDa, 37 kDa, 25 kDa, 20 kDa, 10 kDa.

D — EBOV His6-SUMO-VP30$_{8-272}$, Zinc, Gold; 0.05 mM, 0.1 mM, 0.5 mM, 1 mM; lanes: 1) marker, 2) insoluble, 3) soluble, 4) insoluble, 5) soluble, 6) insoluble, 7) soluble, 8) insoluble, 9) soluble. Markers: 150 kDa, 100 kDa, 75 kDa, 50 kDa, 37 kDa, 25 kDa, 20 kDa, 10 kDa.

E — EBOV VP30$_{69-272}$; His6-SUMO (R2, Gold), GST (R2, Gold); lanes: 1) marker, 2) insoluble, 3) soluble, 4) insoluble, 5) soluble, 6) insoluble, 7) soluble, 8) insoluble, 9) soluble. Markers: 150 kDa, 100 kDa, 75 kDa, 50 kDa, 37 kDa, 25 kDa, 20 kDa, 10 kDa.

F — EBOV VP30$_{87-265}$; GST (R2, Gold), His6-SUMO (R2, Gold); lanes: 1) marker, 2) insoluble, 3) soluble, 4) insoluble, 5) soluble, 6) insoluble, 7) soluble, 8) insoluble, 9) soluble. Markers: 150 kDa, 100 kDa, 75 kDa, 50 kDa, 37 kDa, 25 kDa, 20 kDa.

Figure 5.12 EBOV VP30 Small Scale Expression

10 mL of 2XYT was inoculated with 0.5 mL of starter culture. Induced at 0.6-0.8 OD_{600}
A-D EBOV $VP30_{8-272}$
A- 0.5 mM IPTG 18°C overnight, (pGEX-6P-2). 1) maker, 2) insoluble (R2), 3) soluble (R2), 4) insoluble (Gold), 5) soluble (Gold)
B- 0.5 mM IPTG 18°C overnight, (pET-28a-SUMO). 1) maker, 2) insoluble (R2), 3) soluble (R2), 4) insoluble (Gold), 5) soluble (Gold).
C- 0.1-1 mM IPTG 18°C overnight (pET-28a-SUMO, R2). 1) marker, 2) insoluble 0.1 mM IPTG, 3) soluble 0.1 mM IPTG, 5) insoluble 0.2 mM IPTG, 6) soluble 0.2 mM IPTG, 7) insoluble 0.4 mM IPTG, 8) soluble 0.4 mM IPTG, 9) insoluble 0.6 mM IPTG, 10) soluble 0.6 mM IPTG, 11) insoluble 0.8 mM 12) soluble 0.8 mM IPTG, 13) insoluble 1 mM IPTG, 14) soluble 1 mM IPTG).
D- 0.05-1 mM ZnS04, 18°C overnight (pET-28a-SUMO, Gold). 1) marker, 2) insoluble 0.05 mM ZnS04, 3) soluble 0.05 mM ZnS04, 4) insoluble 0.1 mM ZnS04 , 5) soluble 0.1 mM ZnS04, 6) insoluble 0.5 mM ZnS04 , 7) soluble 0.5 mM ZnS04, 8) insoluble 1 mM ZnS04, 9) soluble 1 mM ZnS04.
E-F EBOV $VP30_{87-265}$
E- EBOV $VP30_{69-272}$ small scale expression 18°C overnight, 0.1 mM ZnS04, 0.5 mM IPTG; 1) marker, 2) R2 insoluble His6-SUMO, 3) R2 soluble His6-SUMO, 4) Gold insoluble His6-SUMO, 5) Gold soluble His6-SUMO, 6) R2 insoluble GST, 7) R2 soluble GST, 8) Gold insoluble GST, 9) Gold soluble GST.
F- EBOV $VP30_{87-265}$ small scale expression 18°C overnight, 0.1 mM ZnS04, 0.5 mM IPTG; 1) marker, 2) R2 insoluble GST, 3) R2 soluble GST, 4) Gold insoluble GST, 5) Gold soluble GST, 6) R2 insoluble His6-SUMO, 7) R2 soluble His6-SUMO, 8) Gold insoluble His6-SUMO, 9) Gold soluble His6-SUMO.
EBOV $VP30_{8-272}$ = 30 kDa.
EBOV $VP30_{69-272}$ = 23 kDa.
EBOV $VP30_{87-265}$ = 20 kDa.
His-MBP = 45 kDa.
GST = 25 kDa.
His-SUMO=13 kDa.
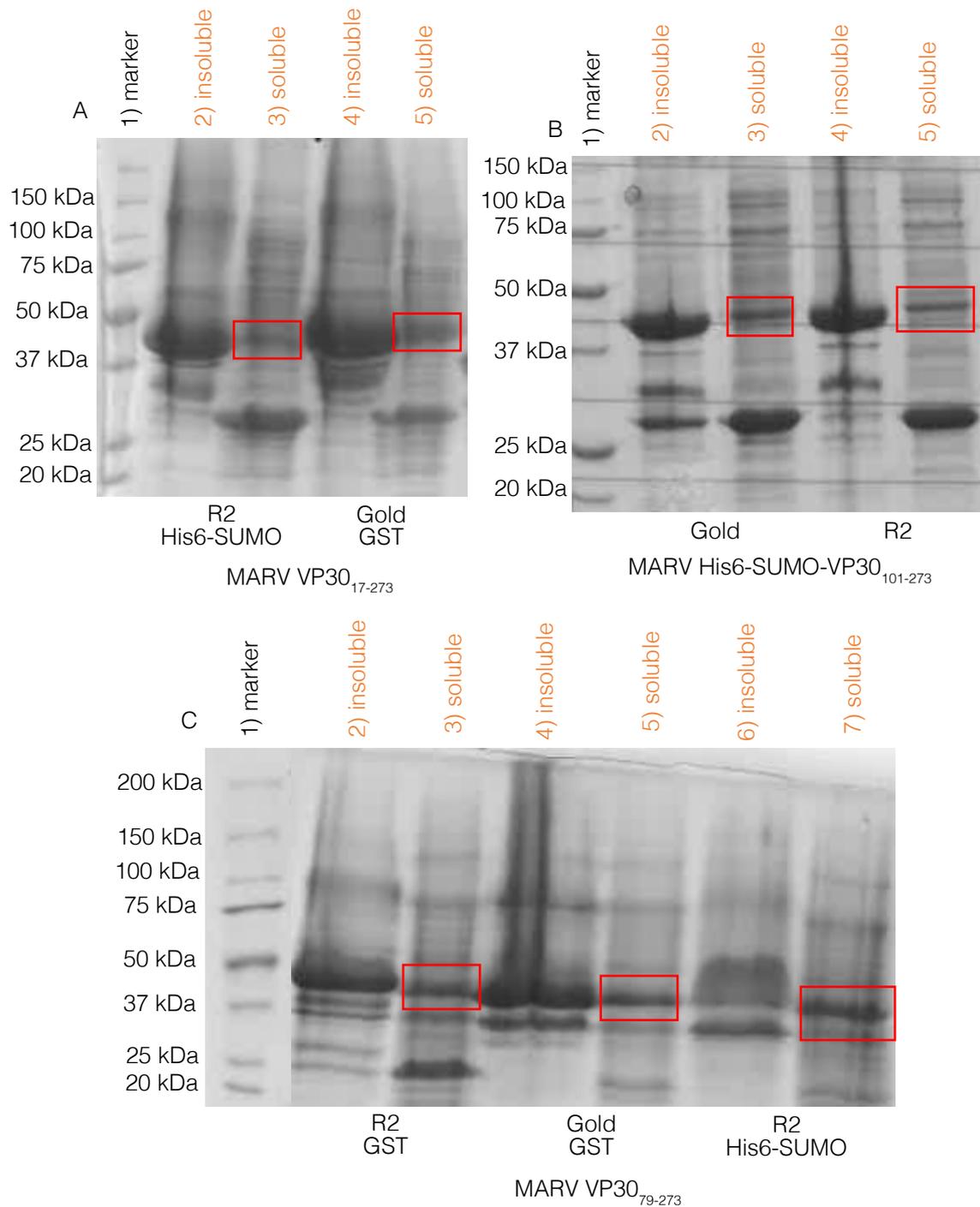
Figure 5.13 MARV VP30 Small Scale Expression

10 mL of 2XYT was inoculated with 0.5 mL of starter culture. Induced at 0.6-0.8 OD_{600} with 0.1 mM Zinc and 0.4 mM IPTG. Soluble protein highlighted by red box.
A- Marburg VP30 17-273 1) marker, 2) insoluble R2 (sumo), 3) soluble R2 (sumo), 4) insoluble Gold (pgex), 5) soluble Gold (pgex)
B- Marburg 101-273 (sumo) 1) marker, 2) insoluble Gold, 2) soluble Gold, 3) insoluble R2, 4) soluble R2
C- Marburg VP30 79-273 1) marker, 2) insoluble R2 (pgex) 3) insoluble R2 (pgex), 4) insoluble Gold (pgex), 5) soluble Gold (pgex), 6) insoluble R2 (sumo), 7) soluble R2 (sumo).
MARV $VP30_{17-273}$ = 28 kDa.
MARV $VP30_{79-273}$ = 22 kDa.
MARV $VP30_{101-273}$ = 19 kDa.
GST = 25 kDa.
His-SUMO = 13 kDa.
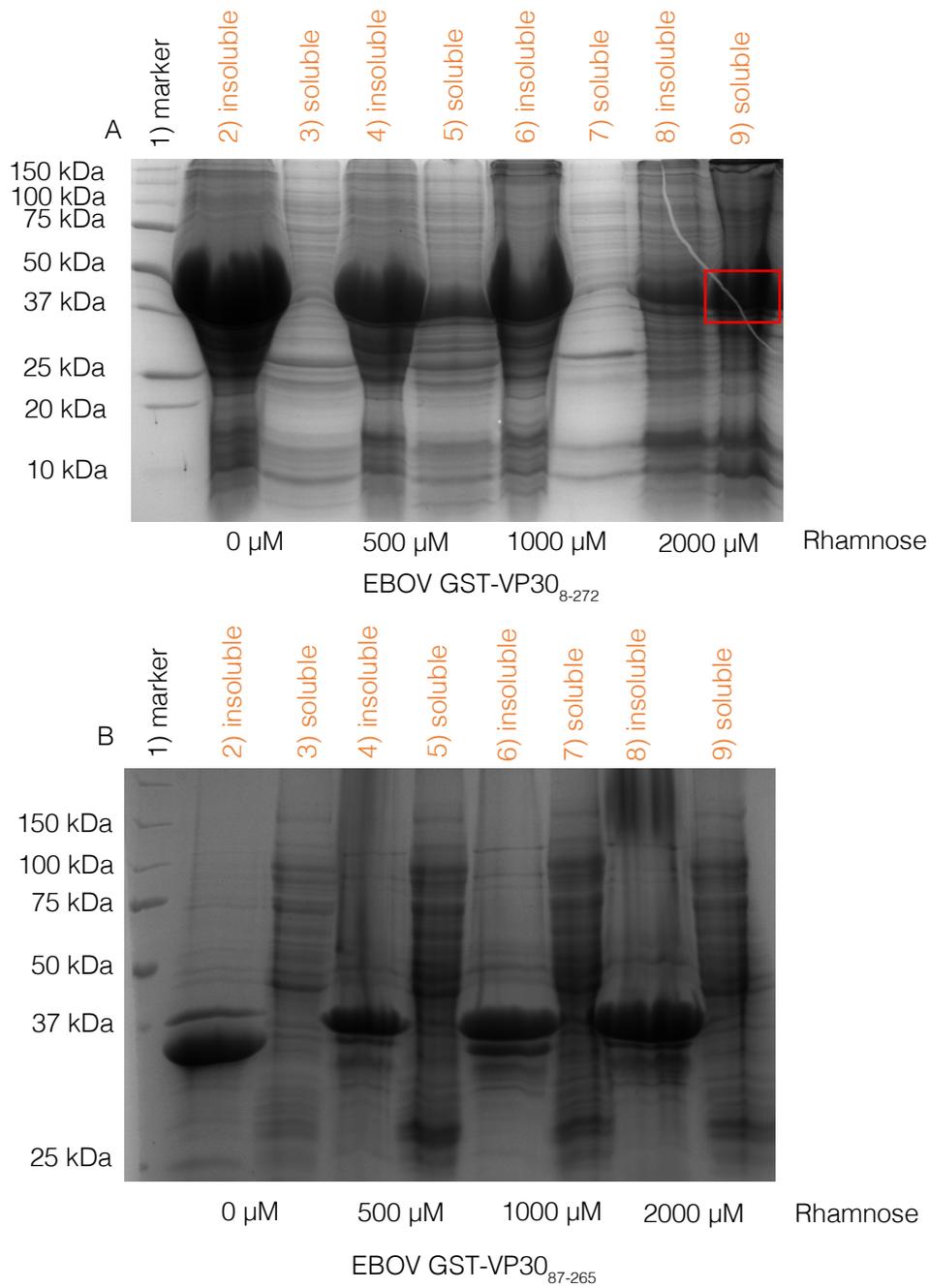
Figure 5.14 EBOV VP30 Expression Optimisation in Lemo21 Cells

10 mL of 2XYT was inoculated with 0.5 mL of starter culture and induced at OD_{600} 0.6-0.8 with 0.5 mM IPTG and 0-2000 µM rhamnose. Incubated at 37.C overnight.
A- GST-VP30$_{8-272}$ 1) marker, 2) insoluble, 3) soluble, 4) insoluble, 5) soluble, 6) insoluble, 7) soluble, 8) insoluble, 9) soluble
B- His6-SUMO-VP30$_{87-265}$ 1) marker, 2) insoluble, 3) soluble, 4) insoluble, 5) soluble, 6) insoluble, 7) soluble, 8) insoluble, 9) soluble
EBOV VP30$_{8-272}$ = 30 kDa.
EBOV VP30$_{87-265}$ = 20 kDa.
GST = 25 kDa.
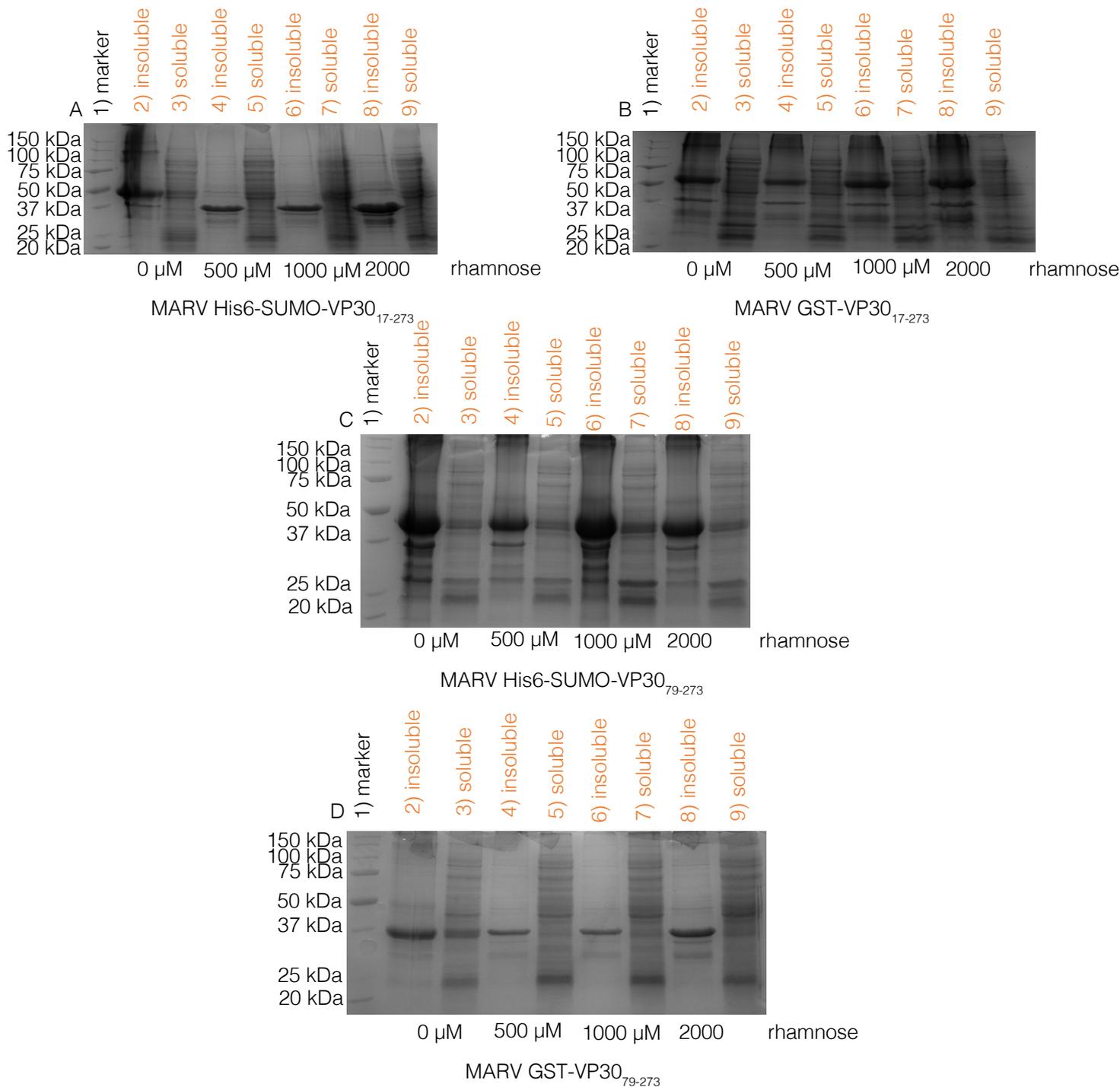
Figure 5.15 MARV VP30 Expression Optimisation in Lemo21 Cells

10 mL of 2XYT was inoculated with 0.5 mL of starter culture. Induced with 0.5 mM IPTG and 0-2000 uM rhamnose at OD_{600} 0.6-0.8.
A- $VP30_{17-273}$ pET-28a-SUMO vector. 1) marker 2) insoluble, 3) soluble, 4) insoluble, 5) soluble, 6) insoluble, 7) soluble, 8) insoluble, 9) soluble

B- $VP30_{17-273}$ pGEX-6P-2 vector. 1) marker, 2) insoluble, 3) soluble, 4) insoluble, 5) soluble, 6) insoluble, 7) soluble, 8) insoluble, 9) soluble

C- $VP30_{79-273}$ pET-28a-SUMO vector. 1) marker 2) insoluble, 3) soluble, 4) insoluble, 5) soluble, 6) insoluble, 7) soluble, 8) insoluble, 9) soluble

D- $VP30_{79-273}$ pGEX-6P-2 vector. 1) marker, 2) insoluble, 3) soluble, 4) insoluble, 5) soluble, 6) insoluble, 7) soluble, 8) insoluble, 9) soluble

MARV $VP30_{17-273}$ = 28 kDa.
MARV $VP30_{79-273}$ = 22 kDa.
GST = 25 kDa.
His-SUMO = 13 kDa.

5.4.2.3 Large Scale Purification from Lemo21 E. *coli* Cells

In the small-scale expression trial of EBOV GST-VP30$_{8-272}$ in Lemo21 cells induced with 2000 μM rhamnose, the correct size band in SDS-PAGE was seen for the fusion protein from the soluble sample. 2 L of expression cultures were grown in the same way and soluble lysate extracted using the freeze-thaw method (experiment performed prior to the purchase of the Avestin C3 Cell Disruptor). Soluble supernatant was applied to pre-equilibrated column with GS4B resin and washed with 75 mL of lysis buffer and 50 mL of high salt buffer to remove non-specific proteins and any host-bound RNA. GST-EBOV VP30$_{8-272}$ was eluted (Figure 5.17 lane 9 red box) and cleaved for 16 hours at 4°C with agitation and 0.5 mg PreScission (3C) protease. The cleaved solution turned cloudy during cleavage with protein aggregates forming white 'clumps' suggesting that the VP30 was not correctly folded or stable.
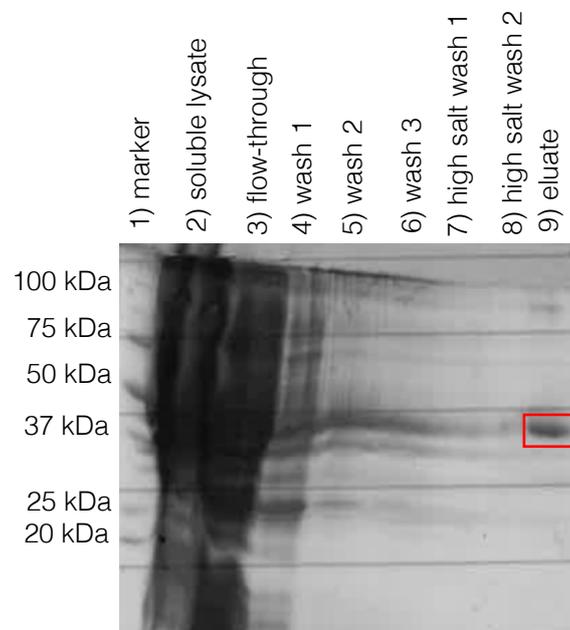

Figure 5.16 EBOV VP308-272 Large Scale Purification from Lemo21 E. coli Cells

Soluble lysate from 6 L of expression growth of GST-VP30$_{8-272}$ was applied to pre-equilibrated column of GS4B resin, washed 3 times to remove non-specific proteins and washed with 1 M NaCl to remove E.coli bound RNA prior elution with 10 mM glutathione.
C- 1) marker, 2) soluble, 3) flow through, 4-6) wash 1-3, 7-8) high salt wash 1-2, 9) eluted GST-VP30 (red box)

5.4.2.4 Large Scale Purification from Auto-induction Media

Another method to optimise soluble protein expression used was auto-induction; auto-induction media has limited glucose and once depleted lactose is converted into the inducer allolactose, allolactose releases the lac repressor allowing the production of the T7 polymerase which in turn switches on production of the target gene, VP30. 2L of auto-induction media was inoculated with 10 mL of MARV VP30$_{101-273}$ starter culture and grown at 18°C for 48 hours. Bacterial cells were harvested by centrifugation and lysed. Soluble lysate was applied to a pre-equilibrated column containing GS4B resin (Figure 5.17). GST-VP30 was cleaved on column with 0.5 mg PreScission (3C) protease for 16 hours with agitation at 4°C. Flow-through was collected and concentrated to 5 mL. VP30 appeared to cleave as the correct size band for GST (~26 kDa) was present on SDS-PAGE (Figure 5.17 B lane 3). However, the band for VP30 was not present. It was assumed that VP30 may have precipitated during cleavage, or 'stuck' to the glass column.
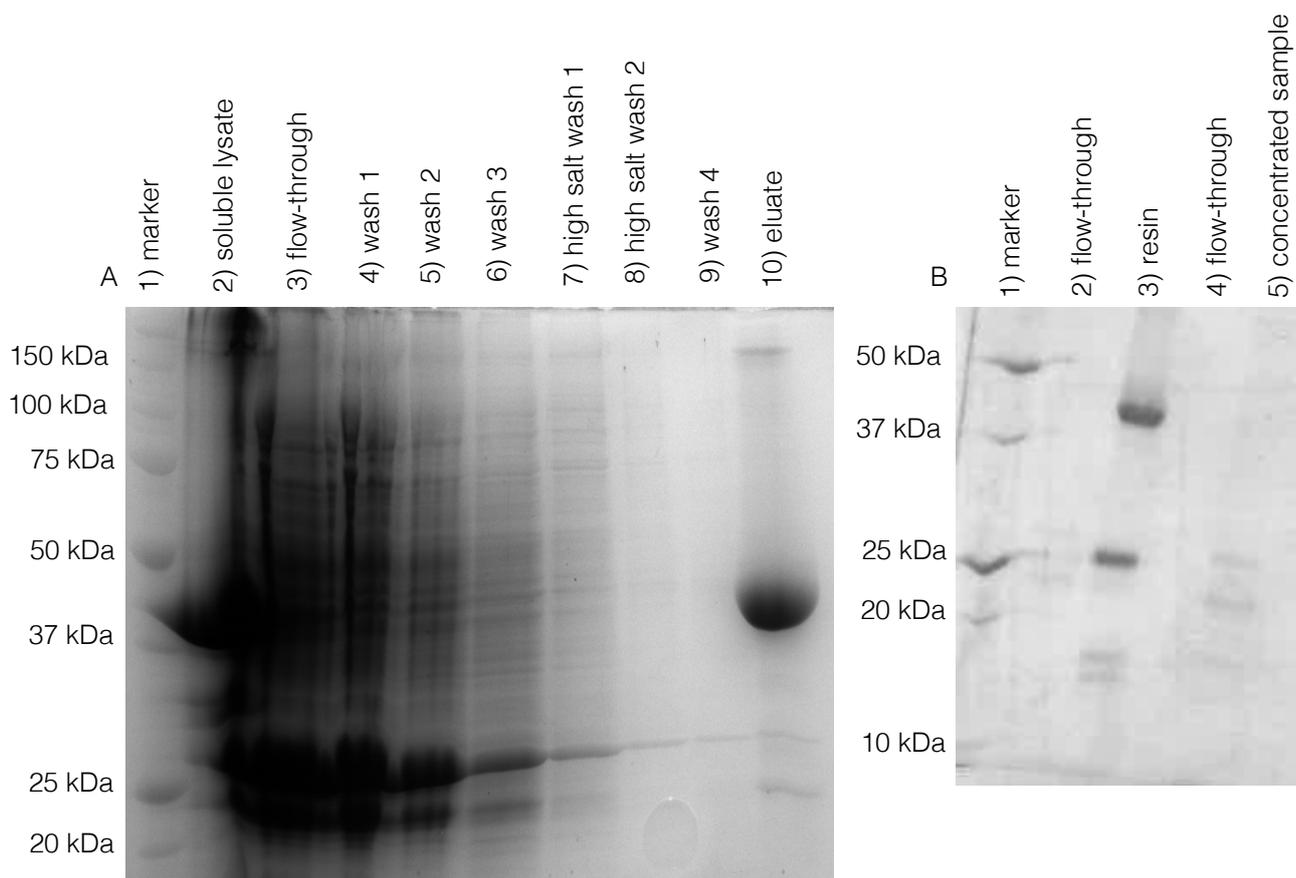

Figure 5.17 Large-Scale Purification of MARV VP30101-273 grown in Auto-Induction Media

2 L of auto-induction media was inoculated with 10 mL of starter culture. Bacteria was grown at 18.C for 52 h. Soluble lysates were applied to GS4B resin, washed 3 times to remove non-specific proteins and 2 further washes to remove *E. coli* bound RNA.
A- affinity chromatography; 1) marker, 2) soluble, 3) flow-through, 4-6) wash 1-3, 7-8) high salt wash 1-2, 9) wash 4, 10) resin
B- on-column cleavage; 1) marker, 2) flow-through, 3) resin, 4) concentrator flow-through, 5) concentrated 5 mL sample
GST- VP30$_{101-273}$ fusion = ~45 kDa

5.4.3 His6-MBP Tag

Due to the large size of maltose-binding-protein (MBP: ~44 kDa) it was suggested that this may allow high yields of VP30 purification by 'pulling' VP30 into solution. This allows us to work with VP30 and perform biophysical assays, however VP30 may remain insoluble due this phenomenon of MBP. VP30 cDNA was therefore cloned into pET-MAL-Precision; a custom-made vector by Dr. Huw Jenkins, University of Leeds, which utilises a pET-28a backbone with a 6-histidine tag and MBP tag (His6-MBP) cleavable with precession (3C) protease. Cloning was performed via cut and paste as before with restriction sites Bam HI and Xho I. As EBOV VP30$_{8-272}$ had been previously purified using an MBP tag, small scale expression trials were not performed, and recombinant VP30 protein was purified from 1 L of expression media.

5.4.3.1 Large Scale Purification of His6-MBP-VP30 in R2 cells

5.4.3.1.1   Affinity Chromatography

1 L of expression media was inoculated at 37°C with 5 mL of overnight culture. Expression culture was induced with 0.5 mM IPTG and 0.1 mM ZnSO$_4$ at 18°C for 16 hours. Bacteria were harvested by centrifugation for 10 minutes at 5000 rpm. Bacterial cells were lysed using a cell disrupter and soluble lysate separate by centrifugation; soluble lysate was applied to a pre-equilibrated HisTrap HP column. The column was washed with 25 mL washes with increasing imidazole to remove non-specific proteins and 50 mL high salt wash (1 M) to remove any host-bound RNA. VP30 constructs were eluted with 150-500 mM imidazole (EBOV Figure 5.18 C and E, and MARV Figure 5.19 A, C and E).

Each eluate was tested for a 260/280 ratio below 1 indicating RNA free protein (Table 5.2).

Table 5.2 260/280 Ratio of His6-MBP-VP30 Eluates

	260/280		
Construct	150 mM	300 mM	500 mM
EBOV VP30$_{8-272}$	0.61	0.68	0.62
EBOV VP30$_{69-272}$	/	0.59	0.53
EBOV VP30$_{87-265}$	/	0.62	0.61
MARV VP30$_{17-273}$	/	0.59	0.67
MARV VP30$_{79-273}$	0.59	0.63	0.61
MARV VP30$_{101-273}$	/	0.60	0.63

Eluates for each construct were pooled. 50% was removed for cleavage of the His6-MBP purification tag. The remaining 50% concentrated for size-exclusion chromatography (SEC) as a fusion protein.

5.4.3.1.2 Cleavage

50% of each pooled eluate was cleaved in dialysis buffer to remove the high concentrations of imidazole in preparation for secondary affinity chromatography. During cleavage precipitation occurred for all six constructs and the decision was made to use the fusion protein for further experiments.
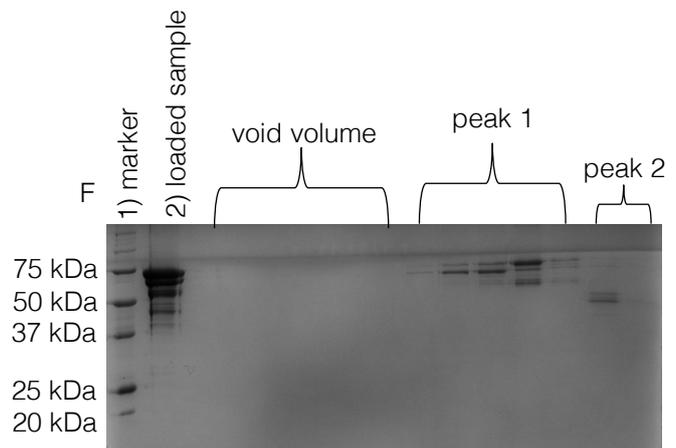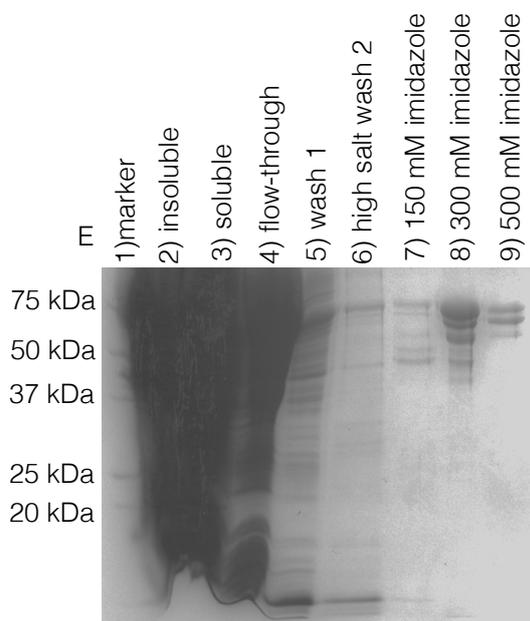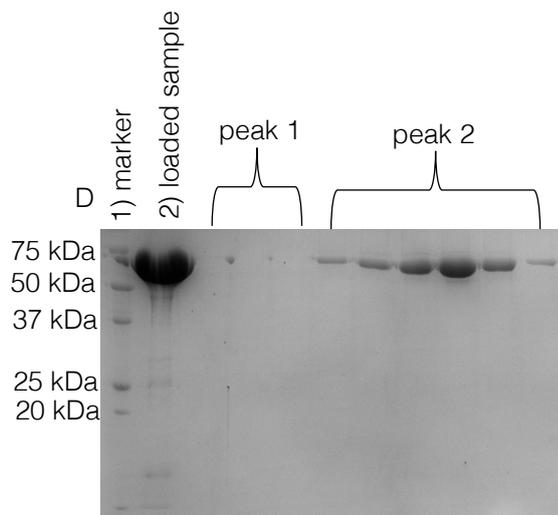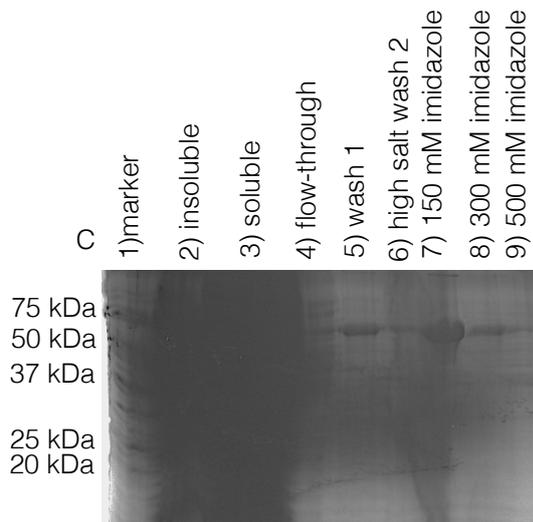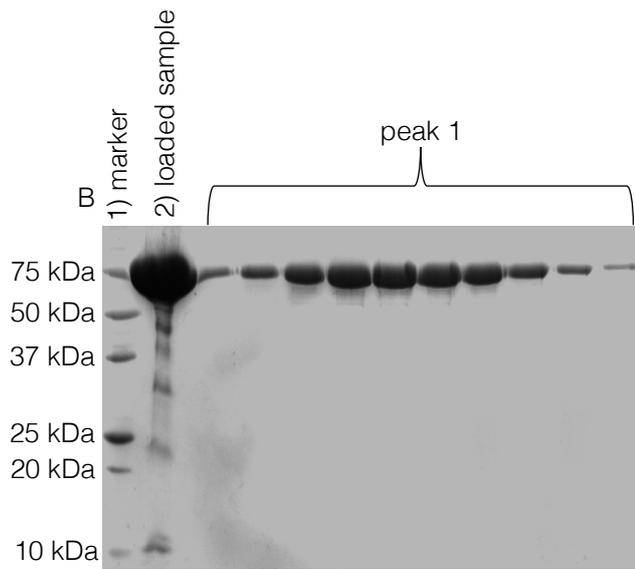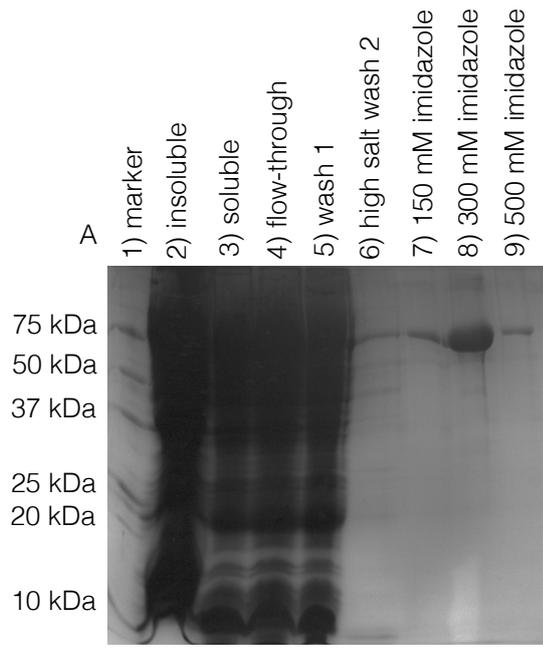
Figure 5.18 Large Scale Purification of EBOV VP30 in pET-MAL-PreScission

A- VP30$_{8\text{-}272}$ affinity chromatography; 1) marker, 2) insoluble, 3) soluble, 4) flow-through, 5) wash 1, 6) high salt wash 2), 7) 150 mM imidazole, 8) 300 mM imidazole, 9) 500 mM imidazole.
B- VP30$_{8\text{-}272}$ size exclusion chromatography; 1) marker, 2) loaded sample, 3-12) peak 1.
C- VP30$_{69\text{-}272}$ affinity chromatography; 1) marker, 2) insoluble, 3) soluble, 4) flow-through, 5) wash 1, 6) high salt wash 2), 7) 150 mM imidazole, 8) 300 mM imidazole, 9) 500 mM imidazole.
D- VP30$_{69\text{-}272}$ size exclusion chromatography; 1) marker, 2) loaded sample, 3-5) peak 1, 6-11) peak 2.
E- VP30$_{87\text{-}265}$ affinity chromatography; 1) marker, 2) insoluble, 3) soluble, 4) flow-through, 5) wash 1, 6) high salt wash 2), 7) 150 mM imidazole, 8) 300 mM imidazole, 9) 500 mM imidazole.
F- VP30$_{87\text{-}265}$ size exclusion chromatography; 1) marker, 2) loaded sample, 3-6) peak 1, 7-11) peak 2, 12-13) peak 3.
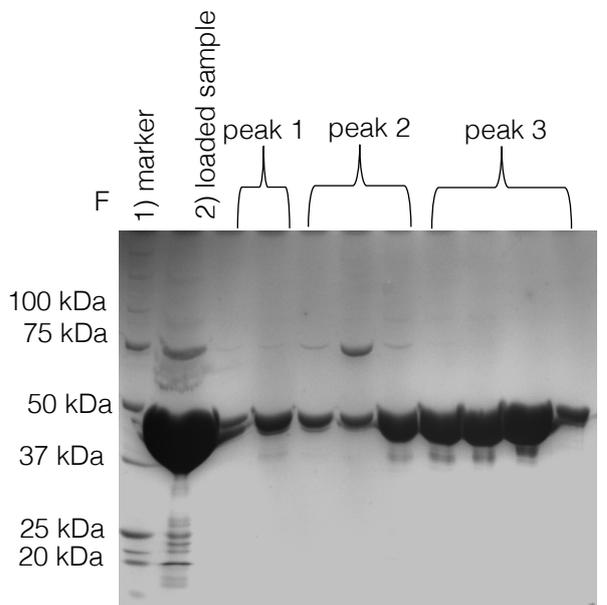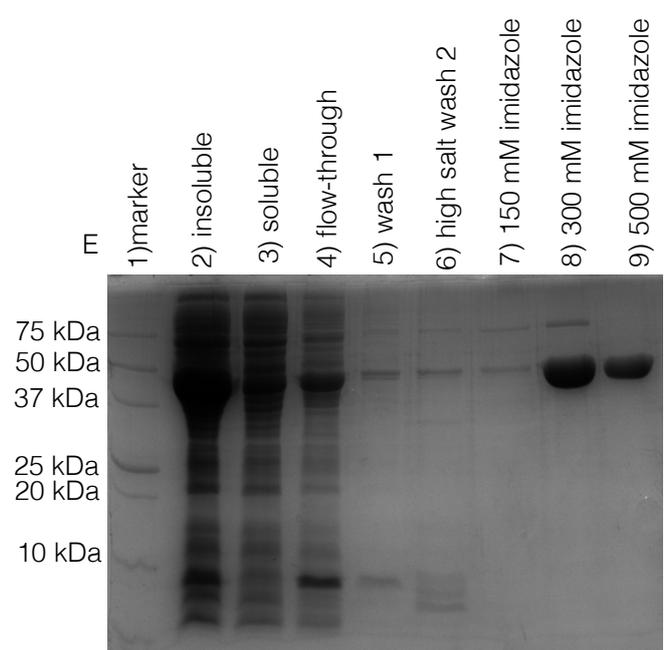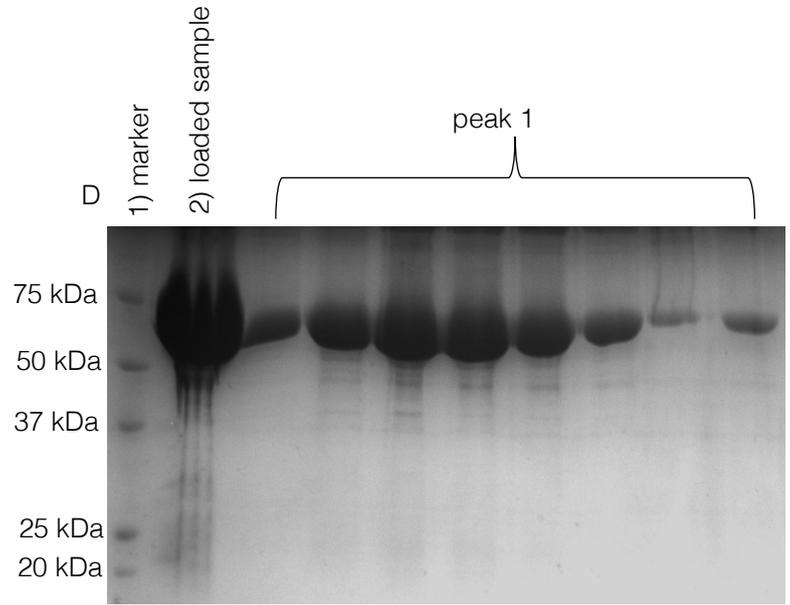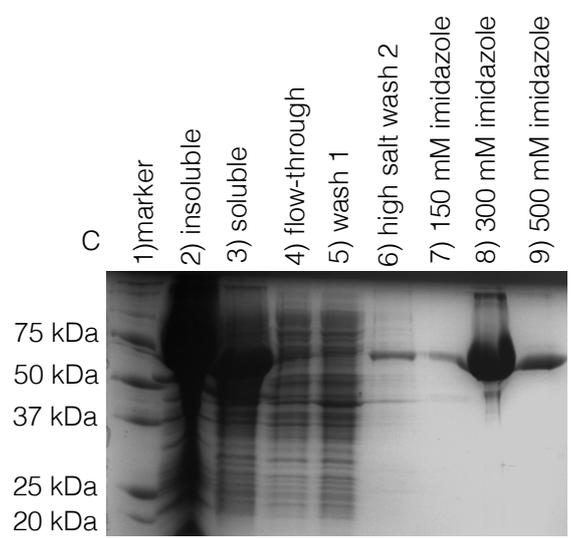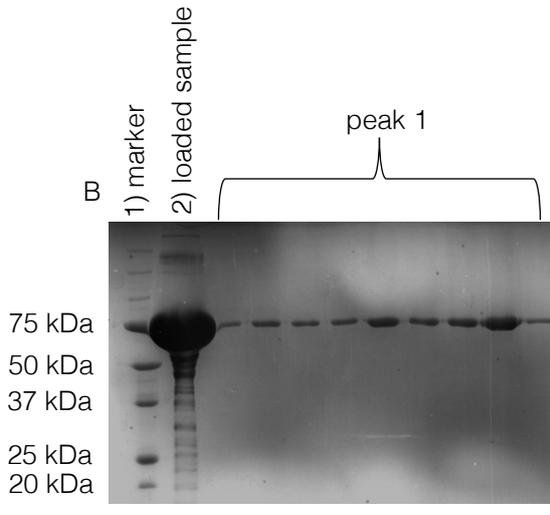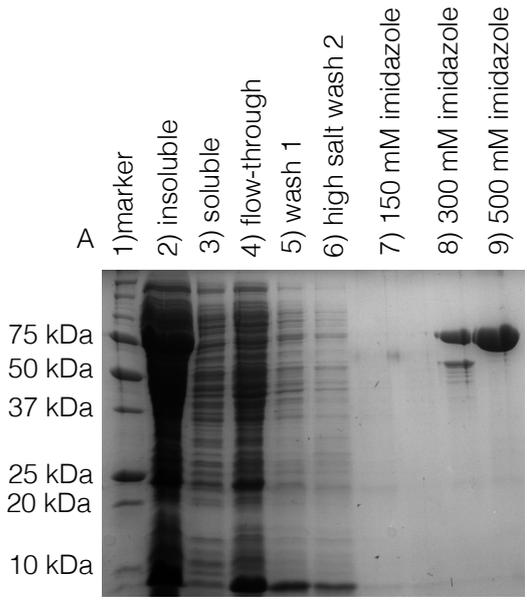
Figure 5.19 Large Scale Purification of MARV VP30 in pET-MAL-PreScissionission

A- VP30$_{17-273}$ affinity chromatography; 1) marker, 2) insoluble, 3) soluble, 4) flow-through, 5) wash 1, 6) high salt wash 2), 7) 150 mM imidazole, 8) 300 mM imidazole, 9) 500 mM imidazole.
B- VP30$_{17-273}$ size exclusion chromatography; 1) marker, 2) loaded sample, 3-11) peak 1, 12-14) peak 2.
C- VP30$_{73-273}$ affinity chromatography; 1) marker, 2) insoluble, 3) soluble, 4) flow-through, 5) wash 1, 6) high salt wash 2), 7) 150 mM imidazole, 8) 300 mM imidazole, 9) 500 mM imidazole.
D- VP30$_{73-273}$ size exclusion chromatography; 1) marker, 2) loades sample, 3-10) peak 1.
E- VP30$_{101-273}$ affinity chromatography; 1) marker, 2) insoluble, 3) soluble, 4) flow-through, 5) wash 1, 6) high salt wash 2), 7) 150 mM imidazole, 8) 300 mM imidazole, 9) 500 mM imidazole
F- VP30$_{101-273}$ size exclusion chromatography; 1) marker, 2) loaded sample, 3-7) peak 1, 8-11) peak 2.
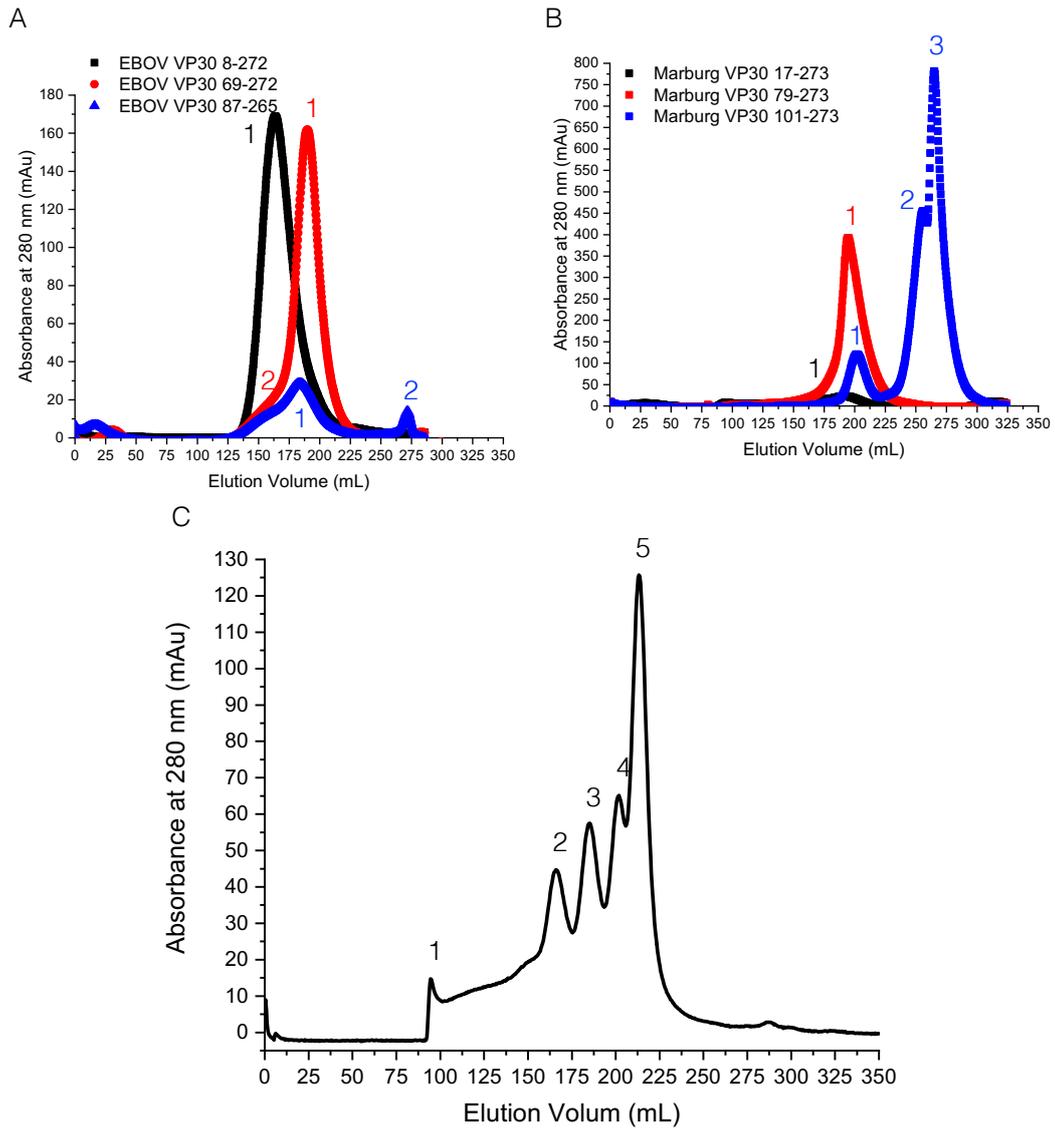
Figure 5.20 Size-Exclusion Chromatography of Hia6-MBP-VP30 Fusion Proteins

10 mL of concentrated VP30 was injected onto a HiPrep 26/600 Sephacryl S-400 HR column using and AKTA purifier pump. A- EBOV VP30. B- MARV VP30. Each VP30 construct eluted as a hexamer. A smalerl construct elution profile is shifted to the right.
C- SEC calibration. HiPrep 26/600 Sephacryl S-400 HR column calibrated with 25 mM Tris-HCl pH 8.0, 200 mM NaCl, 2.5% glycerol; 1) blue dextran 2000 20,000 kDa, 2) thyroglobulin 669kDa, 3) ferritin 440kDa, 4) aldolase 158 kDa, 5) conalbumin 75kDa, 6) ovalbumin 43kDa

5.4.3.1.3   Size-Exclusion Chromatography (SEC)

His6-MBP-VP30 constructs were concentrated to 5 mL and injected onto a 26/600 Sephacryl S-400 size-exclusion column using an AKTA prime pump. A protein peak was seen at elution volume ~185 mL corresponding to a molecular weight of ~450 kDa when compared to the calibration chromatogram for the same column.

5.4.3.2 X-Ray Crystallography

Purified fusion protein was concentrated to 12 mg/mL before crystallisation trials and used as apo protein or mixed in 1:1 μM ratio with 3'Fl labelled EBOV RNA; 3'Fl EBOV RNA was used as it was the only EBOV specific RNA available. Although the 3'Fl is not ideal for crystallography, it was thought that the EBOV RNA may help stabilise VP30 hexamer, aiding crystal packing. Precipitation was seen in various conditions using the sitting drop vapour diffusion method (Figure 5.21).
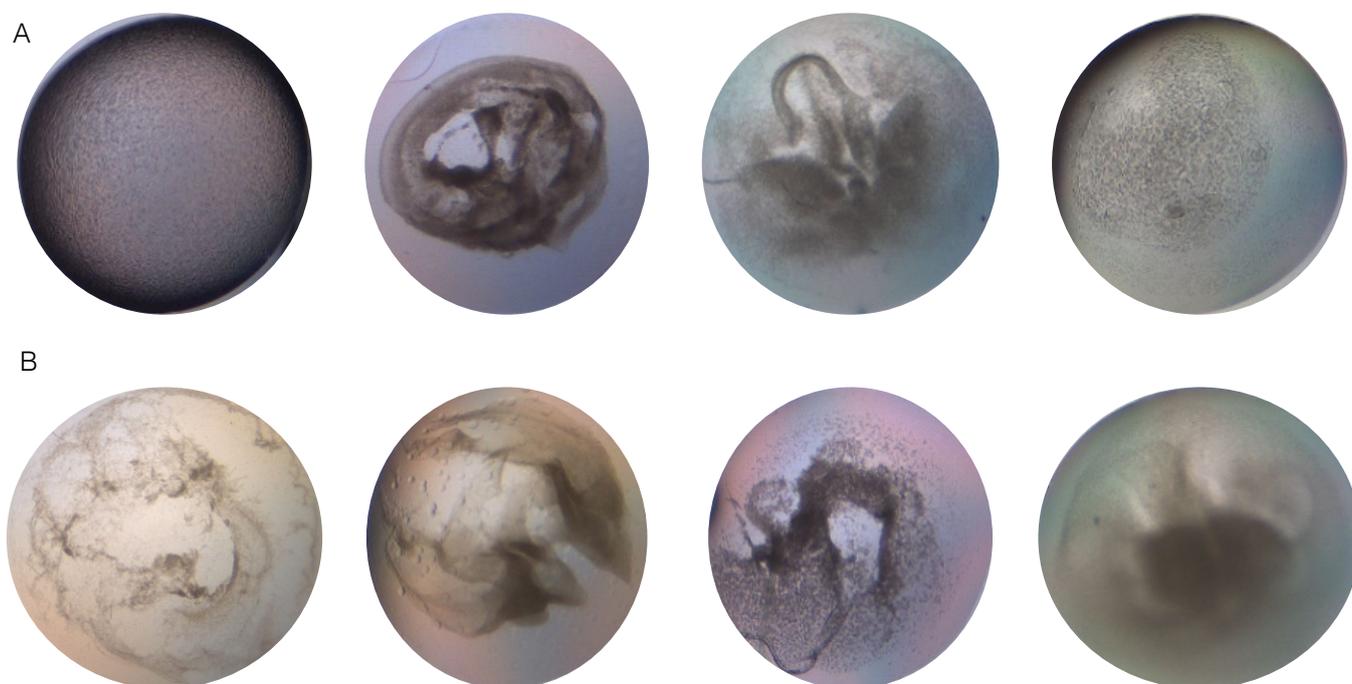
Figure 5.21 Precipitation formed in Crystallisation Trials

EBOV VP30$_{8-272}$ was concentrated to 12mg/mL prior to crystallization.
A- apo protein
B- VP30:RNA was mixed in a 1:1 uM ratio prior to to crystallization.
Representative drop images from JCSG Core 1-4 screens shows protein precipitation.

5.4.3.3 Limited Proteolysis

As an attempt to identify region(s) of VP30 that was soluble for X-ray crystallography limited proteolysis experiments were performed using trypsin. MARV VP30$_{17-273}$ fusion protein was digested with 6 mg/mL (Figure 5.22 A) or 2.5 mg/mL (Figure 5.22 B) trypsin and samples taken from 0-10 minutes. Higher concentrations of trypsin did not show any VP30 peptides, however the lower (2.5 mg/mL) trypsin digest revealed a small peptide (>10 kDa). This was not further analysed by mass-spectrometry and structural studies as it was assumed that this peptide would not provide further structural information from the previously published EBOV VP30 CTD.
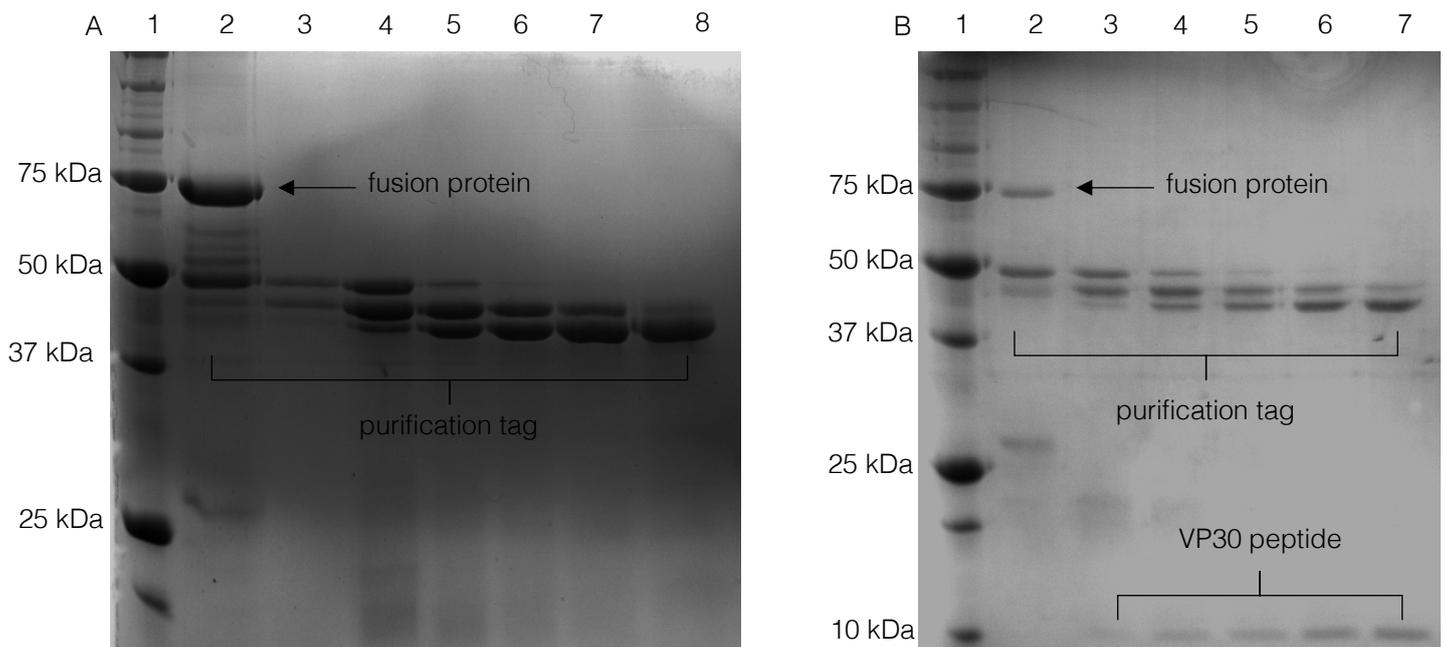
Figure 5.22 Limited Proteolysis Assay of MARV VP30$_{17-273}$

A- 6 mg/mL trypsin digest; 1) marker, 2) 0 seconds, 3) 30 seconds, 4) 1 minute,
5), 2 minutes, 6) 4 minutes, 7) 6 minutes, 8) 10 minutes
B- 2.5 mg/mL trypsin digest; 1) marker, 2) 0 seconds, 3) 30 seconds, 4) 1 minute,
5), 2 minutes, 6) 4 minutes, 7) 6 minutes, 8) 10 minutes

5.4.4 Oligomeric State Determination

To further confirm the oligomeric state of EBOV and MARV VP30 the constructs 8-272 and 17-273 respectively were subjected to mass spectrometry analysis and SEC-MALLS.

5.4.4.1 Size Exclusion Chromatography coupled to Multi Angle Laser Light Scattering (SEC-MALLS)

The oligomeric state of His6-MBP-VP30 was also analysed by SEC-MALLS. SEC-MALLS for EBOV and MARV VP30 showed a single oligomeric peak in the protein sample. EBOV VP30$_{8-272}$ showed a 463 kDa oligomer ($\pm$ 9% Figure 5.23 A and B), strongly suggesting that this protein is hexameric in solution. In contrast, oligomers of MARV VP30$_{17-273}$ were recorded between 351-626 kDa, suggesting MARV VP30 may also be hexameric in solution but higher oligomers are forming, possibly due to hexameric aggregation (Figure 5.23 C and D).
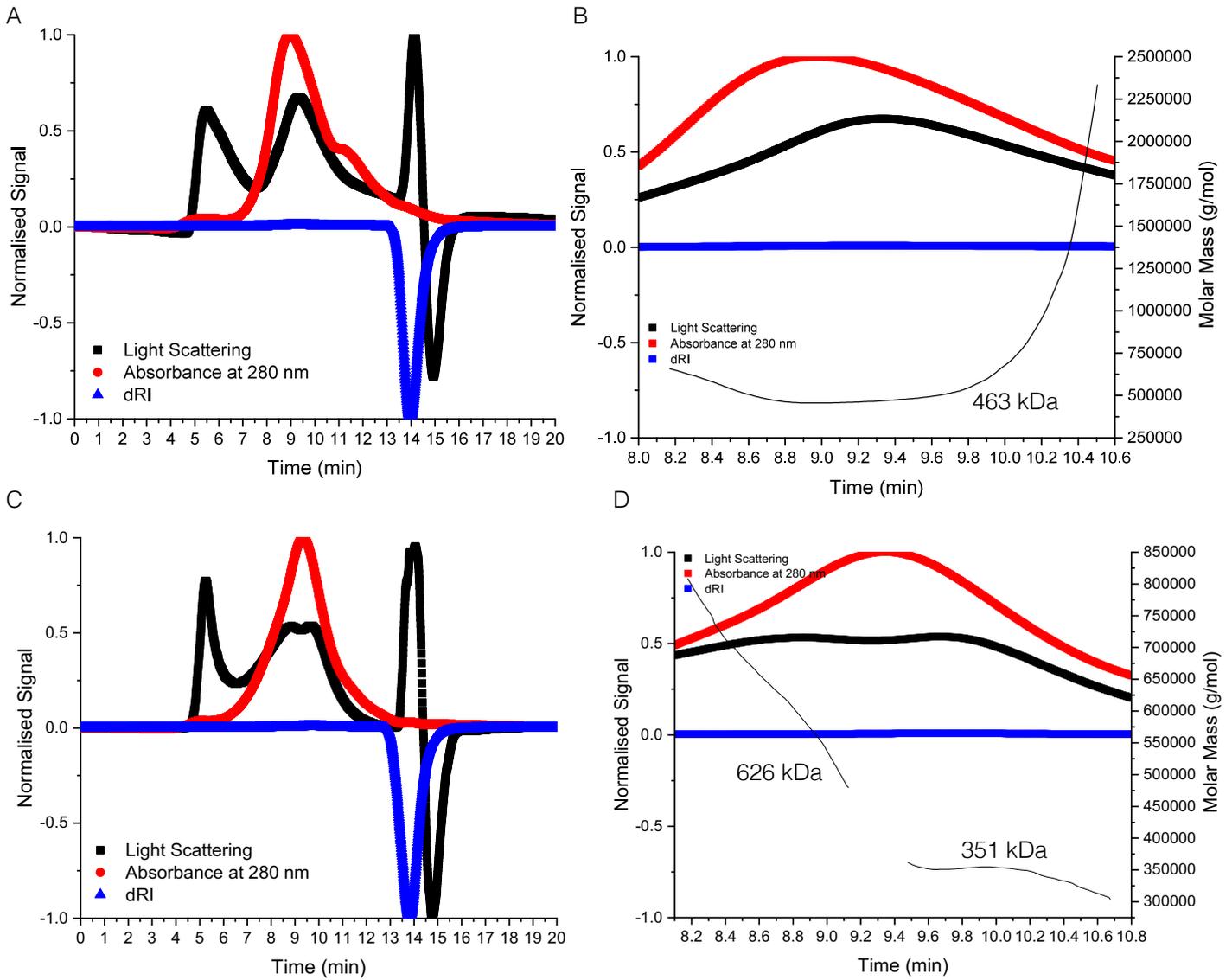
Figure 5.23 Oligomeric State Confirmation of VP30 by SEC-MALLS

A- EBOV VP30$_{8-272;}$ SEC-MALLS showed a single oligomeric peak in the protein sample
B- EBOV VP30$_{8-272}$; SEC-MALLS analysis showed 463 kDa oligomer (black line) (± 9%)
C- MARV VP30$_{17-273;}$ SEC-MALLS showed a single oligomeric peak in the protein sample
D- MARV VP30$_{17-273}$; SEC-MALLS analysis showed 351-626 kDa oligomers (black lines) (± 5%)

5.4.5 VP30:RNA Binding Interaction

VP30's viral transcription activator function is dependent on ssRNA binding. In the literature it is thought that VP30 recognises long stretches of RNA and RNA binding is optimal in the presence of a hairpin structure.

To aid understanding of the VP30:RNA binding interaction and RNA sequence specificity, direct binding affinities were determined by fluorescence anisotropy (FA) using short (10 oligonucleotides) and long (64 nucleotide) EBOV specific RNA segments (Table 5.3).

Table 5.3 3'Fl RNA Sequences

3'Fl RNA	Sequence
EBOV$_{54-118}$ 64 mer	AUGAGGAAGAUUAAUAAUUUUCCUCUCAUUGA AAUUUAUAUCCCAAUUUAAAUUGAAAUUAUUAC
pA 10 mer	AAAAAAAAAA
pC 10 mer	CCCCCCCCCC
pU 10 mer	UUUUUUUUUU

To exclude the possibility that the His6-MBP purification tag was contributing to RNA binding, purified His6-MBP was also tested for RNA binding of EBOV$_{54-118}$ and pA 3'Fl RNA (Figure 5.24 E); no binding was seen to either RNA.

RNA binding was visualised for EBOV$_{54-118}$ 3'Fl RNA (Figure 5.24 and Table 5.4) and pA oligonucleotide 3'Fl RNA (Figure 5.24 A and B and Table 5.4) with no binding observed to pC or pU under the conditions of this experiment.

As expected, the VP30 truncations, EBOV VP30$_{87-265}$ and MARV VP30$_{101-273}$ that exclude the previously established RNA binding domain and the ZBD bind with weaker affinity than the larger constructs (Figure 5.24 C and D respectively, and Table 5.4).

Two serine clusters (comprising EBOV residues 26-40) represent the previously identified N-terminal RNA binding domain. Whilst not essential for RNA binding (apparent K_D for EBOV VP30$_{69-272}$: 177 nM) the presence of residues 8-68 provided optimal binding, increasing affinity by 3 orders of magnitude (apparent K_D for EBOV VP30$_{8-272}$ is 0.13 nM).

EBOV VP30$_{87-265}$ was designed as homologous to the core domain of M2-1 which includes the RNA binding domain. This also suggests that the ZBD of M2-1 is not essential for RNA binding, but again is required for optimal RNA binding.

Table 5.4 Binding Affinities of 3'Fl RNA to VP30

| | Apparent K_D nM | | | |
| | 3'Fl RNA | | | |
	$EBOV_{54-118}$	pA	pC	pU
EBOV VP30$_{8-272}$	0.131 ± 0.028	58.1 ± 10.37	/	/
EBOV VP30$_{69-272}$	177.010 ± 52.690	/	/	/
EBOV VP30$_{87-265}$	743.620 ± 3095.000	/	/	/
MARV VP30$_{17-273}$	0.489 ± 0.045	9.43 ± 0.462	/	/
MARV VP30$_{79-273}$	583.450 ± 929.180	/	/	/
MARV VP30$_{101-273}$	/	/	/	/
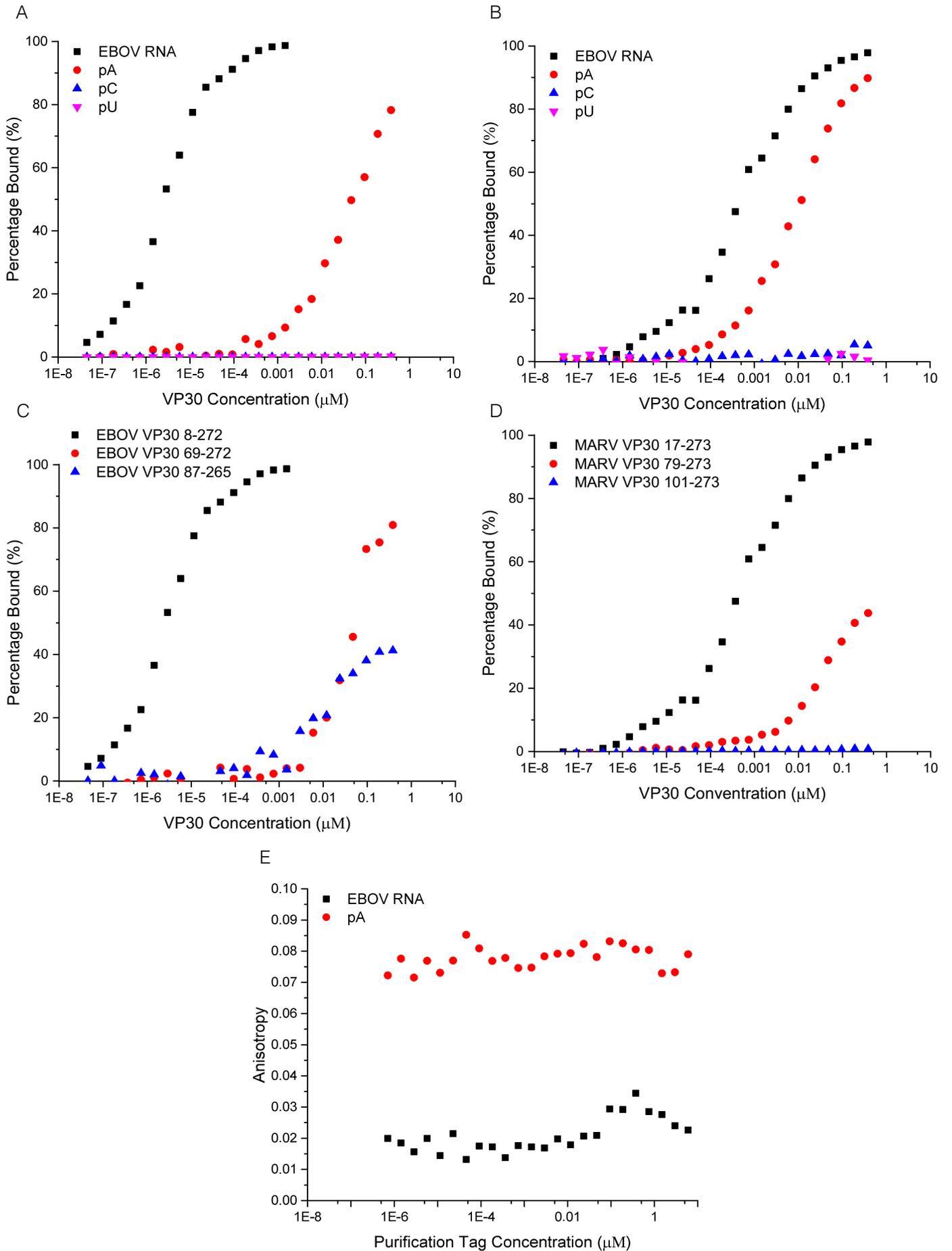
Figure 5.24 Direct Binding of VP30 to RNA using Fluoresce Anisotropy

Log concentration curve of 3' Fl RNA represented as percentage bound.
Binding affinities (EC50's) were determined from the fitted curve and per
A- EBOV VP30$_{8-272}$ direct binding to EBOV$_{54-118}$ and pA, -C, -U 3'Fl RNA
B- MARV VP30$_{17-273}$ direct binding to EBOV$_{54-118}$ and pA, -C, -U 3'Fl RNA
C- EBOV VP30 constructs direct binding to EBOV$_{54-118}$ 3'Fl RNA
D- MARV VP30 constructs direct binding to EBOV$_{54-118}$ 3'Fl RNA
E- His6-MBP direct binding to EBOV$_{54-118}$ and poly- A 3'Fl RNA

5.4.6 VP35:VP30 Interaction

In the literature it is unclear whether the interaction between VP35 and VP30 requires RNA. The homologous interaction in hRSV (between P and M2-1) does not require RNA for an interaction, as the RNA:P binding domain on M2-1 overlaps (with M2-1 preferentially binding RNA) (Tanner *et al.*, 2014; Selvaraj *et al.*, 2018). Thereby, assuming the VP35 and VP30 interact similarly, RNA would not be needed for the two protein to interact.

SEC was performed on a Superdex S-200 15/50 column attached to an AKTA purifier pump. 50 μL of MARV VP35$_{60-329}$ (250 μM) was mixed with 50 μL of His6-MBP-VP30$_{17-273}$ (150 μM) prior to SEC. VP35 was added in excess as it is unclear how the tetramer and hexamer hetero-oligomerise. Peak 1 (Figure 5.25 A and B (lane 4-6)) contained both VP35 and VP30. However, due to the large size of the VP30 hexamer (~440 kDa) the stoichiometry of the complex was not able to be identified. The complex however, was eluted after the void volume (~1.3 mL) indicating a soluble complex had been formed without RNA.
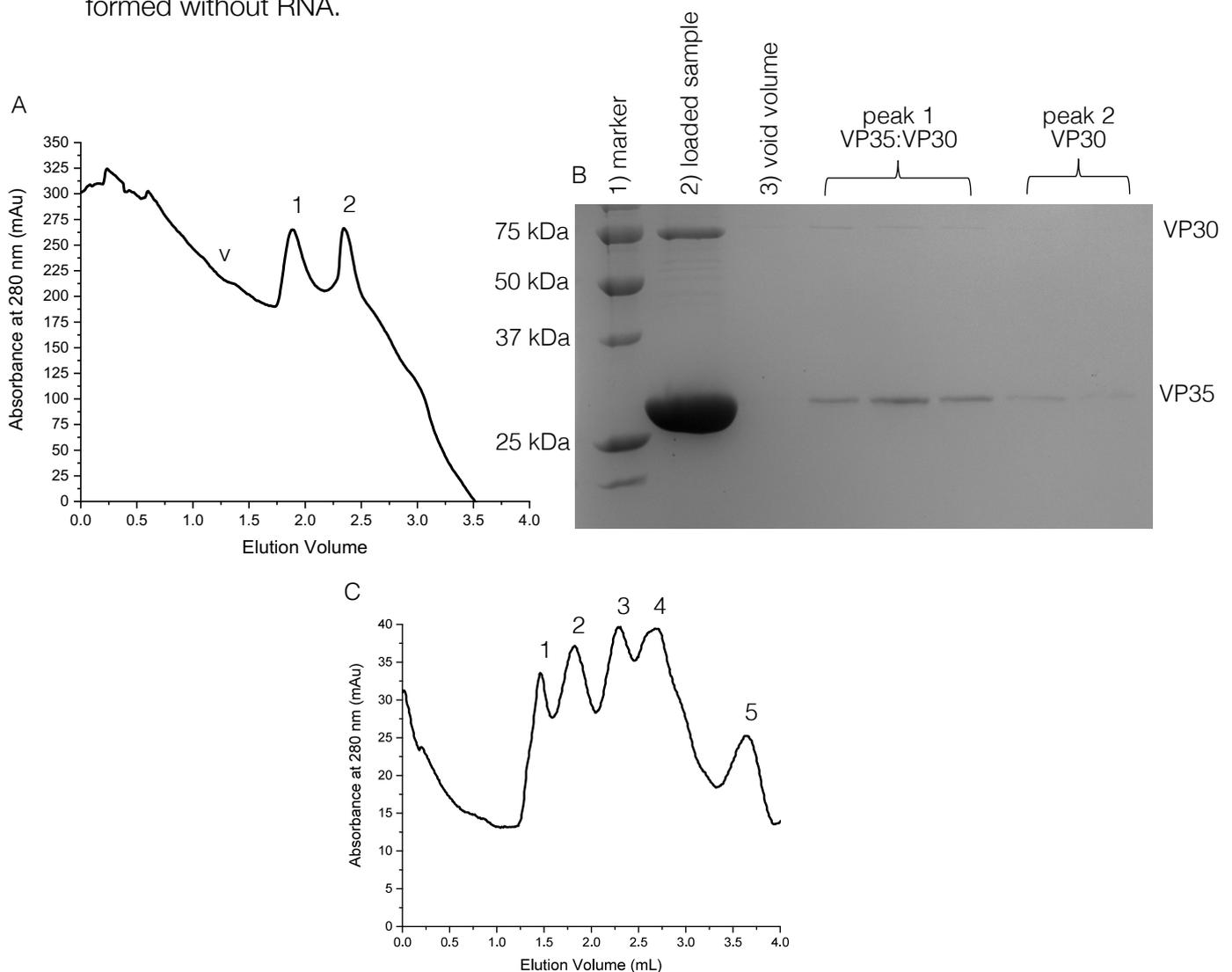
Figure 5.25 Size-Exclusion Chromatography of VP35 and VP30

A- SEC. 100 μL of MARV VP35$_{60-329}$:MARV His6-MBP-VP30 was injected onto a Superdex S-200 15/50 column with an AKTA purifier pump. Absorbance was measured at 280 nm and 100 uL fractions collected.

B- 1) marker, 2) inject, 4) void, 4-6) peak 1, 7-8) peak 2

C- SEC Calibration. HiLoad Superdex S-200 15/50 column calibrated with 25 mM Tris-HCl pH 8.0, 300 mM NaCl, 2.5% glycerol; 1) blue dextran 2000 kDa 2) ferritin 440 kDa, 3) aldolase 158 kDa, 4) conalbumin 75 kDa, Ribonuclease A, 13.7 kDa

5.5 Discussion

5.5.1 Protein Expression and Purification of VP30

There was no protocol available for the purification of VP30 proteins that produced high protein yields. A new purification method was therefore required that also did not result in protein degradation. Unfortunately, the expression of both MARV and EBOV VP30 was not was unsuccessful in obtaining soluble protein without a solubility protein (fusion) tag. A possible reason for this could be the presence of cysteines (within the Cys3-His ZBD). Larger proteins (>25 kDa) with cysteines present as problematic for protein folding and, folding yields of <20% can be expected (Palmer and Wingfield, 2004). However, the substitutions of cysteines (often for serines in protein production) is not possible here as the ZBD is essential for correct protein folding of VP30, similarly to that of hRSV M2-1(Tanner *et al.*, 2014).

The *E.coli* is still the dominant host for recombinant protein production due to cost, ease and scale. However, aggregation of recombinant proteins into insoluble inclusion bodies is a main limiting factor, along with the improper formation of disulfide bonds and the absence of chaperones for correct folding (Costa *et al.*, 2014). Dr Brian Jackson (protein production facility manger, University of Leeds) attempted to source GroEL and GroES co-expression chaperone plasmids (structurally and functionally identical to Hsp60 and Hsp10 respectively (Nielsen *et al.*, 1999)) that help with protein folding but was unsuccessful. Secondly, lower expression temperatures could also reduce protein aggregation, by slowing down the rate of protein synthesis and folding kinetics, decreasing hydrophobic interactions involved in self-aggregation (Costa *et al.*, 2014). To do this, ArticExpress (DE3) competent cells (Agilent Technologies) could be utilised. ArticExpress cells have been engineered to improve protein expression at low temperatures by co-expressing cold-adapted chaperonins Cpn10 and Cpn60 75% and 54% amino acid sequence identity to GroEL and GroES respectively) from *Oleispira Antarctica* (psychrophilic bacterium), which show high protein refolding activities at 4-12°C (Agilent Technologies, 2012). Unfortunately, due to the on-going renovations of the Garstang building (University of Leeds) the bacterial growth room (Garstang 8.54) faced temperature issues, and the Infors HT incubators struggled to reduce temperatures to 20°C making conditions for ArticExpress cells unachievable and the option of moving an incubator into a cold room was not possible.

In order to overcome aggregation, inclusion bodies could be recovered from cell lysates by low speed centrifugation with 8 M guanidine hydrochloride, producing a 'washed' protein pellet. The challenge is to then solubilise the recombinant protein and fold it into native and biologically active protein. 8 M guanidine hydrochloride is used to solubilise the washed protein pellet here by disrupting protein-protein interactions and unfolding the recombinant protein. The recombinant protein is then re-folded slowly by reducing the concentration of guanidine hydrochloride. Purifying recombinant proteins this way however, is often problematic as unfolded proteins are susceptible to chemical modifications including oxidations of methionine and cysteine and protease activity. Due to the ZBD being essential for protein function and folding, it was decided that a

preparation and extraction of insoluble recombinant protein from inclusion-bodies would not be suitable for VP30.

Production expression and purification of VP30 could also be stabilised by the addition of RNA. This was attempted for X-ray crystallography (5.4.3.2), however the only RNA available was 3'Fl-tagged EBOV RNA and the quantities needed for protein expression was unknown. After discussion with Dr. Thomas Edwards, it was agreed that this method would be material intensive and not cost-effective, as we wanted to reserve the RNA for FA experiments.

The final purification protocol for EBOV and MARV VP30 here used a His6-MBP fusion tag that was not cleaved to maintain solubility and, purification could be performed in two days which limited protein degradation. This purification method allowed for ~12.5 mg of pure fusion protein from 1 L of expression culture. A potential problem with using MBP as a fusion tag is that micelle-like structures may form, whereby misfolded recombinant proteins are sequestered and protected from the solvent and soluble protein domains face outward (Costa *et al.*, 2014). However, as EBOV and MARV VP30 was able to bind RNA in FA it was thought that these micelle-like structures were not forming as VP30:RNA interaction was still able to occur.

5.5.2 Elucidating the Oligomeric State of VP30

EBOV VP30 CTD (142-272) was previously shown to form dimers in the unit cell, and still remains the only structural model for VP30. On the other-hand, EBOV VP30$_{89-272}$ iss hexameric in solution when subjected to SEC and cross-linking analysis. Data here supports previous findings that VP30 is hexameric however it still remains unclear if VP30 is a trimer of dimers (supported by the CTD crystal structure (Hartlieb *et al.*, 2007)) or six monomers. As data in this thesis further supports the model that VP30 is hexameric, perhaps the EBOV VP30 CTD dimer might dissociate via helix seven; interactions are seen between the loop region connecting helix six and seven, allowing for conformational plasticity and higher oligomer conformers.

The N-terminus is involved in regulating transcription via phosphorylation (Hartlieb *et al.*, 2003; Martínez *et al.*, 2008; Biedenkopf, Lier and Becker, 2016; Schlereth *et al.*, 2016), and also might regulate the conformational plasticity of VP30 via the N-terminal oligomerisation domain. The leucine-zipper like motif (residues 100-103 in EBOV VP30 and 106-109 in MARV VP30) previously highlighted as important for hexamerisation (Hartlieb *et al.*, 2003). This is perhaps why the longer construct EBOV VP30$_{89-272}$, and other constructs used in this thesis, form hexamers in solution. VP30 conformational state may be concentration-dependant, similar to the bacterial transcription factor TyrR that also forms dimers and hexamers (Dixon *et al.*, 2002; Yang *et al.*, 2004), suggesting that the balance between dimers and hexamers controls transcription and replication.

5.5.3 VP30 Preferentially Binds A-rich RNA sequences and RNA Binding is Not Limited to the N-terminus

There has been speculation regarding the RNA-binding domain of the VP30 protein. The RNA binding domain previously established for EBOV VP30 encompassed residues 26-40 (and 40-51 in MARV VP30), saturated with positively charged residues but which forms two serine clusters that are phosphorylated upon activation of VP30, potentially increasing interactions with RNA for transcriptional activation. In order to test this hypothesis phosphomimetic mutants (using aspartic acid) can be assessed. Moreover, the ZBD (a Cys3-His motif), which coordinates Zn^{2+} incorporation for correctly folded and stable protein, does provide increased affinity to RNA (EBOV VP30$_{69272}$ and MARV VP30 $_{79-273}$ Figure 5.24 C and D respectively and Table 5.4). It still remains unknown how VP30 activates transcription, assuming VP30 binds viral RNA it is unclear if this is the negative-sense genome, the positive sense RNA anti-genome essential for replication, or transcribed viral mRNAs.

RNA binding data here suggests that the two serine clusters alone do not contribute to RNA binding. It is probable that this domain along with the ZBD contribute to the high affinities seen here. In order to fully elucidate this, RNA binding assays with point mutations need to be assessed.

VP30 and M2-1 share structural and functional similarities therefore it was not surprising that VP30 preferentially bound pA over -C and -U 3'FI RNA. Previous findings within our group provide evidence that M2-1 recognises mRNA end sequences that are A-rich in sequence, rather than the hRSV positive-sense genomic template (or anti-genome) (Tanner *et al.*, 2014). However, A composition throughout the hRSV genome for each transcript encompasses a pA tract of at least 4 nucleotides every ~200 nucleotides. Suggesting that M2-1 could recognise these intragenic A rich sequences throughout the length of the gene and prevent anti-termination this way. This would support data shown here and previous EMSAs that showed that VP30 preferentially binds the negative-sense genomic RNA rather than the anti-genome and the higher binding affinities were seen for the negative-sense genomic EBOV RNA$_{54-118}$ than the pA 10 3'FI RNA (Biedenkopf *et al.*, 2016; Schlereth *et al.*, 2016).

5.5.4 Interaction with VP35

The interaction between VP35:VP30 was only analysed by SEC, but does suggest the complex is stable and able to be purified. To further characterise this interaction, pull-down experiments can be performed in order to identify the corresponding interacting domains for each protein. In hRSV, P$_{90-110}$ is known to interact with M2-1. Sequence alignment between hRSV P$_{90-110}$ and EBOV VP35 shows that EBOV VP35$_{61-81}$ is the only stretch of similar residues (Figure 5.26). This comprises the NTD of EBOV VP35 which is also known to interact with NP. Perhaps, both VP30 and NP compete for the same binding region and this is another control switch between transcription and replication; whereby NP binding promotes replication (as NP is readily available for the

encapsidation of newly synthesised RNA) and VP30 binding promotes transcription, possibly acting as a recruiter protein for the transcription complex.

```
EBOV VP35 61 QTKPNPKTRNSQTQTDPICNHSFEEVVQTLA  81
hRSV P    90 --DPTPS--------DNPFSKLYKETIETFD 110
             .*.*. * .: ::*.::*:
```

Figure 5.26 Alignment between EBOV VP35 and hRSV P$_{90-110}$

Alignment made using Clustal Omega.
* Conserved residue.
: conservation between groups of strongly similar properties.
. conservation between groups of weakly similar properties.

Lastly, as VP35 has the potential to be a transformer protein, it makes biological sense that VP30 also has this ability, and perhaps their interaction is dependent on both being in complimentary states, for example, trimeric VP35 can interact with hexameric VP30, promoting transcription in a 2:1 manner. Whereas tetrameric VP35 does not interact with dimeric VP30 and supports replication. Further structural characterisation of both proteins is needed to elucidate this hypothesis. Perhaps, EM is the most sensible method to use here as EM could potentially capture both 'forms' of proteins and the VP35:VP30 interaction.

5.6 Chapter Summary

This chapter outlined a new purification method for MARV VP35$_{60-329}$ the longest MARV VP35 construct to be expressed and purified to allow oligomeric state determination. Here we also produced high yields of His6-MBP-VP30 fusion proteins for oligomeric state confirmation and RNA binding assays. Through mass spectrometry and SEC-MALLS analysis VP35 was identified as a tetramer and VP30 as a hexamer.

How these known binding partners interact still remains elusive. The SEC presented here suggests that RNA is not needed for the interaction, however the stoichiometry is unclear. In hRSV the M2-1:P interaction is a 1:1 as both are tetramers. It is possible that VP30:VP35 interactions are 1:1 also as VP35's N-terminal residues (1-59 in MARV and 1-79 in EBOV) are flexible and this region binds to VP30. It is the N-terminus that overlaps with P's M2-1 binding region (residues 90-110). Residues 28-35 in MARV VP35 are 75% identical to residues 96-103 in hRSV P. Interestingly, EBOV VP35 and P$_{90-110}$ do not align, this may be due to the interaction requiring RNA.

5.7 Future Directions

For both VP35 and VP30 crystal growth conditions will need to be optimised in order to obtain crystal structures. For VP30, protein purification may need further optimisation to not include the His6-MBP purification tag as this is linked to VP30 by a flexible linker that is often not optimal for crystal packing and therefore might be the reason no crystals were formed during initial trials. Cryo-EM may possibly be a more fruitful direction.

VP30:RNA binding assays should be carried out to further identify whether VP30 has a preference for genomic or antigenomic sequences. Phosphomimetic mutants can also be made to further validate the importance of the N- and C-terminal phosphorylation sites. Constructs should also be tested within the EBOV replicon to see if they support EBOV transcription.

Chapter 6 Characterising the M2-1:P Interaction of human Orthopneumovirus (hRSV)

6.1 Chapter Introduction

hRSV was first isolated in 1956 and infects 64 million individuals annually (Chanock et al. 1957; Nair et al. 2011). With approximately 253,000 deaths a year, up to 79% of deaths occur in children below 5, and 99% of deaths occur in developing countries (Chung *et al.*, 2013; Kiss *et al.*, 2014). Although hRSV is highly prevalent and extensive research has been carried out over the last ~60 years there is no suitable treatment apart from the broad-spectrum anti-viral ribavirin and a humanised mouse monoclonal neutralising antibody palivizumab that are both expensive and largely ineffective and only given to high-risk patients. Structure based drug design has arisen as a method to develop antiviral drug candidates.

hRSV has a 15.2 kb genome that encodes eleven proteins from ten genes. Of these proteins seven are structural and two non-structural. Replication and transcription of the hRSV genome occurs within virally-induced inclusion bodies. Inclusion bodies, sometimes referred to as elementary bodies, are cytoplasmic aggregations of viral proteins and hijacked host-cell machinery needed for replication and transcription such as ribosomes. Inclusion bodies are composed of viral RNA which is encapsidated by NP, the polymerase co-factor P, the RNA-dependent RNA polymerase L, and M2-1 required only for transcription and the production of full-length mRNAs (Rincheval *et al.*, 2017).

M2-1 is essential to the hRSV viral life cycle and function and an anti-termination transcription factor. In the absence of M2-1 short non-functional mRNAs are produced which rarely possess 3' pA tracts (Tanner *et al.*, 2014). Established binding partners of M2-1 include the polymerase co-factor P and RNA which interact directly and competitively for M2-1 P/RNA overlapping binding site (Tanner *et al.*, 2014; Selvaraj *et al.*, 2018). At present, it is unclear if this interaction occurs simultaneously on M2-1 protomers or is mutually exclusive, whilst there are four binding sites in the tetramer suggesting that M2-1 (tetrameric in the infected cell) could bind to both RNA and protein at the same time in the replication complex.

The phosphoprotein polymerase co-factor 'P' is also essential for virus life cycle. In the context of replication, at its N-terminus P chaperones monomeric 'free' NP such that NP does not aggregate and is close by to readily encapsidate newly synthesised RNA, a necessary step to avoid innate immunity. P also uses its C-terminus to bind NP that encapsidates RNA in the RNPs (residues 233-241). P binds L here too (residues 203-241) acting as a bridge between L and RNA via interacting with the RNP. By contrast, during transcription, the nascent mRNA strand emerging from L is not encapsidated (Leyrat *et al.*, 2014; Selvaraj *et al.*, 2018).

A hRSV replicon systems has provided a quantitative measure of the functionality of the hRSV polymerase complex. Mammalian cells were transfected with cDNAs expressing

the protein components of the hRSV polymerase complex, NP, P, L and M2-1 and a fifth cDNA expressing a bicistronic minigenome with the reporter gene encoding GFP. GFP expression is therefore dependent on the functionality of the polymerase complex including the M2-1:P interaction. M2-1 mutants R126E and L148A resulted in significant disruption, reducing GFP expression by 30%. R126 forms electrostatic interactions with multiple P residues (Figure 6.1 B) including E104 and E107. L148A mutation is consistent with the observed hydrophobic interaction with P L101. P mutants F98A, Y102A and T105A resulted in reduced minigenome activity too, consistent with their role in mediating the M2-1:P interaction shown in the crystal structure (discussed below). The RNA binding site on M2-1 is also at the P binding site, either directly overlapping or at least sharing some of the surface at the P binding site. Fluorescence anisotropy competition assays showed pA RNA outcompete FI-P$_{90-100}$ (EC50 1.7 μM) (Figure 6.2B) (Selvaraj *et al.*, 2018).
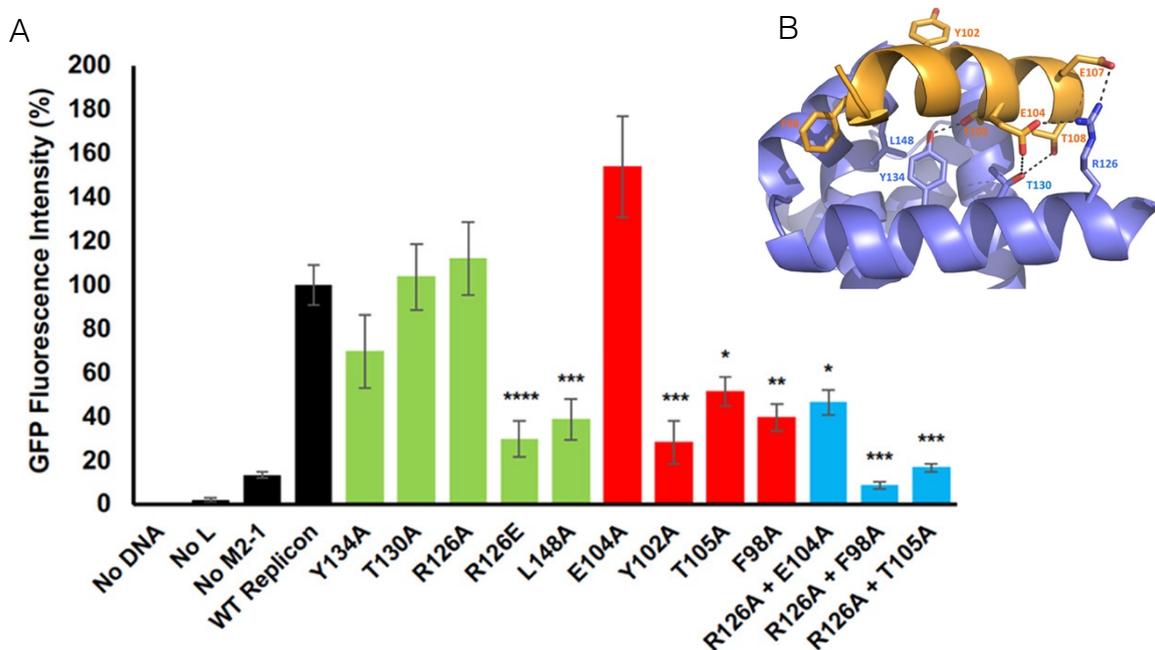
Figure 6.1 Examination for the role of M2-1 and P residues that form a Functional M2-1:P complex

A- Residues comprising the M2-1:P binding interface were mutated to glutamic acid or alanine in the hRSV replicon system. The M2-1 (green), P (red) mutants or double M2-1/P (blue) mutants ability to form a functional transcriptase complex and support transcription of a GFP reporter gene from the supplied minigenome was quantified by counting GFP expression intensity s. Histogram shows relative GFP intensity, normalized to GFP expression from cells transfected with wild-type minigenome components.
Significance values: **** P < 0.0001; *** P < 0.001; ** P < 0.01; * P < 0.05.
B- details of the M2-1:P$_{90-110}$ interactions revealed from the crystal structure (PDB: 6Y0G). M2-1 (purple) and P$_{90-110}$ (orange) residues labelled.
Adapted from Selvaraj *et al.*, 2018
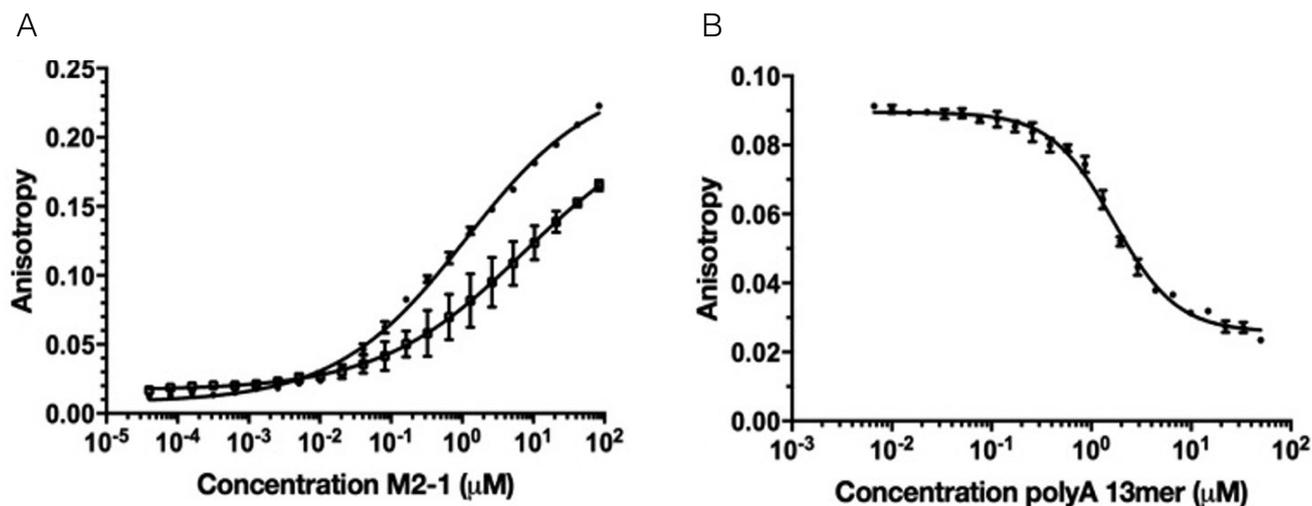
Figure 6.2 Fluorescence Anisotropy of M2-1:P

A- direct binding of M2-1 protein and fluorescein labeled P_{90-110} peptide.
B- competition binding in which Fl-P_{90-110} was outcompeted for M2-1 binding by unlabeled pA 13.
Adapted from Selvaraj et al., 2018

Moreover M2-1 has been shown to also interact with the matrix protein (M) whereby M2-1 mediates the interaction between M and RNP complexes to inhibit viral transcriptase activity and initiate viral assembly and budding via GP interaction (D. Li *et al.*, 2008). M2-1 has also been proposed to interact directly with M in mature virions by cryo-tomography at regular spacings of 12.6 nm, however this hypothesis still needs to be tested through nano-gold labelled viral proteins (Kiss *et al.*, 2014)

6.1.1 Objectives

A greater understanding of M2-1 and its binding partners is required in order to fully elucidate M2-1 functions within the viral life cycle. It remains unclear how M2-1:P interaction occurs; both viral proteins are tetrameric in solution, however the stoichiometry of this interaction is unknown, and how the interaction forms in the context of viral transcription. Moreover, the interaction between M2-1:M remains elusive. The M binding site on M2-1 is unknown and if/how M2-1 undergoes conformational changes (similarly to hMPV M2-1 (Leyrat *et al.*, 2014)). In this chapter I aim to optimise the expression of M2-1, express and purify P_{90-160}, and express and purify M. With pure, homogenous and functional proteins, the aim is to investigate the interactions of M2-1:P and M:M2-1:P using structural methods.

Proteins of interest were over-expressed in E. coli cells in a variety of vectors, which included different purification tags. Full details can be found in (3.2.11). Table 6.1 highlights the different purification tags used.

Table 6.1 Plasmids and Purification Tags for hRSV Proteins

Protein	Plasmid	Purification Tag	Cleavage	Size
M2-1	pGEX-6P-2	N-terminal GST fusion	Precision 3C protease	M2-1 = ~25 kDa GST = ~26 kDa Fusion = ~51 kDa
P_{90-160}	pET-28a backbone	C-terminal His6 fusion	N/A	Fusion = ~9 kDa
M	pET-28a-SUMO	N-terminal His6-SUMO fusion	SUMO protease	M = ~25 kDa His6-SUMO = ~13 kDa (runs at ~19 kDa) Fusion = ~38 kDa – 44 kDa

6.2 Protein Expression and Purification

6.2.1 M2-1

M2-1 was previously expressed as a GST-fusion protein (Tanner *et al.*, 2014). Using the published method of M2-1 purification, protein yields were inconsistent and relatively low. This presented future hurdles when performing material intensive structural studies such as X-ray crystallography. Moreover, the previous method was time-consuming often taking up to five days for purification resulting in protein loss by degradation. In order to avoid issues with batch-to-batch experimental reproducibility, it was necessary to optimise the M2-1 purification protocol to produce larger yields and in a time-saving manner to avoid protein degradation.

6.2.1.1 Optimisation of hRSV M2-1 Purification

There were several stages in the published purification strategy that were highlighted as problematic and in turn responsible for low protein yields and poor protein quality. These included; lysis buffer composition, lysis method and chromatography techniques used.

M2-1 was cloned into expression vector pET-28a-SUMO which produced a His6-SUMO-M2-1 fusion protein upon expression however low yields of soluble protein were expressed in small-scale cultures (data not shown). Therefore, the GST-M2-1 (Tanner *et al.*, 2014) construct was used to optimise M2-1 expression and purification.

The method of lysis used previously was sonication. This is a harsh method and if not set up correctly may result in bacterial lysate warming which can also cause protein unfolding, aggregation and/or degradation. Instead cell disruption was used as this is deemed gentle and temperature controlled at 4°C. GST-M2-1 was then allowed to bind to GS4B resin for 16 hours. It was thought that this caused protein degradation. In the optimised method of purification, GST-M2-1 binding occurred for 1 hour at 4°C and was immediately washed with lysis buffer followed by high salt buffer to remove non-

specific proteins (Figure 6.3 A lane 4 and 5). Lastly, SEC was also performed after ion exchange chromatography of M2-1 (Figure 6.3 B) to ensue tetrameric M2-1 was separated away from any aggregates (Figure 6.3 C and D). A summary of the changes made is presented in Table 6.2 below. Ion exchange chromatography was not performed in the published protocol as was introduced the purification of M2-1 in this thesis to increase the purity of M2-1, ensuring no GST was co-purified.

Table 6.2 Optimisation of hRSV M2-1 Purification

Published Protocol	Alteration	Improvement
No glycerol	5% glycerol to all purification buffers	Aggregation reduction
No DNase	10 µM DNase in lysis buffer	Removal of any host bound DNA
Lysis by sonication	Cell disruption	Reduce degradation, aggregation and temperature control at 4°C
GST-M2-1 binding to GS4B overnight	1 hour and extensive washing at 4°C	Reduced degradation
No SEC	SEC	Removal of M2-1 aggregates and buffer exchange from salt eluates during ion exchange

In short, 2 L of bacterial culture was harvested at 4°C for 10 minutes at 4000 pm. Bacterial pellets were re-suspended in lysis buffer prior to lysis via high pressure homogenisation. Lysates were clarified by centrifugation at 30,000 g for 1 hour at 4°C and soluble lysate applied to a pre-equilibrated column containing GS4B beads. GST-M2-1 was incubated with beads for 1 hour at 4°C followed by extensive washing with lysis buffer, and high-salt buffer to remove any E. coli host bound RNA. M2-1 was cleaved from GST on-column by PreScission (3C) protease at 4°C (Figure 6.3 A lane 6). Eluted M2-1 was diluted from 150 mM NaCl to 50 mM NaCl for cation exchange on SP Sepharose beads; elution of M2-1 with increasing concentrations of NaCl was performed as a stepwise elution (0-100% NaCl increasing by 10% at each step). M2-1 eluted at 60% NaCl (600 mM) (Figure 6.3 B lane 4-7). These eluates were concentrated for final purification step on a HiLoad 26/600 Superdex S-75 column attached to an AKTA prime purification pump. Absorbance was recorded at 280 nm and factions containing tetrameric M2-1 (Figure 6.3 C and D peak 2, lane 6-8) according bands seen in SDS-PAGE and comparison of elution volume to the calibration chromatogram. Tetrameric M2-1 was flash-frozen in liquid nitrogen and stored at -80°C. The published experimental protocol yielded 5 mg of M2-1 per 1 L of expression culture, the optimised protocol improved this 2-fold.
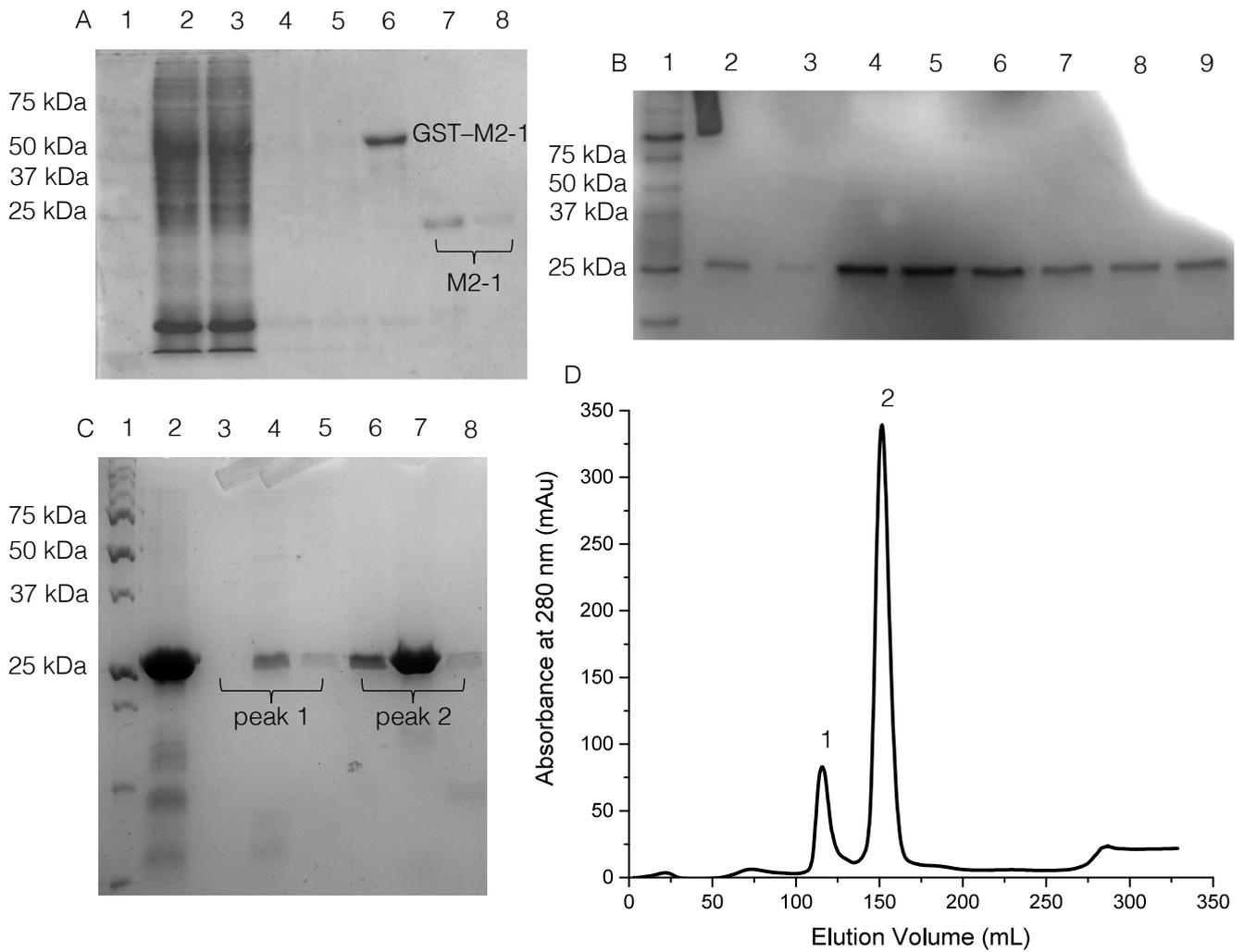
Figure 6.3 Purification of hRSV M2-1

A- affinity chromatography; 1) marker, 2) soluble, 3) resin bound, 4) 50 mL wash, 5) high salt 50 mL wash, 6) resin sample, 7) 50 mL elute 1, 8) 50 mL elute 2
B- ion exchange; 1) marker, 2) flow-through, 3) low salt wash, 4-7) 60% NaCl eluate, 8) high salt eluate
C and D- size exclusion chromatography separated M2-1 on a HiLoad 26/600 Superdex S-75 column; 1) marker, 2) inject, 3-5) peak 1, 6-8) peak 2

6.2.1.2 Protein Quality Check

Tetrameric M2-1 from the gel filtration peak runs as a single band on SDS PAGE. To validate that the M2-1 protein produced was of good quality for further structural studies, CD was performed to ensure M2-1 was correctly folded and FA to ensure M2-1 was functionally active.

6.2.1.2.1 Circular Dichroism (CD)

CD was performed as a service by Nasir Khan, University of Leeds. Pure protein was diluted to 0.1 mg/mL in 50 mM Na_3PO_4 pH 7.5. Analysis revealed a typical α-helical structure due to the characteristic CD spectrum seen in Figure 6.4 A, seen by troughs at 208 nm and 222 nm.

6.2.1.2.2 Fluorescence Anisotropy

It was previously shown that M2-1 preferentially binds A-rich RNA (Tanner *et al.*, 2014) therefore to ensure M2-1 was fully functional FA was performed to check M2-1's RNA binding activity. M2-1s RNA binding domain also overlaps with M2-1s P binding domain and it was therefore assumed that if M2-1 produced here could bind RNA, P would bind too.

M2-1 bound to 3' Fl pA 13 with an apparent K_D of 1.79 ± 0.65 μM (Figure 6.4 B) similar to previously published (apparent K_D 1.9 ± 0.33 μM) (Tanner *et al.*, 2014).
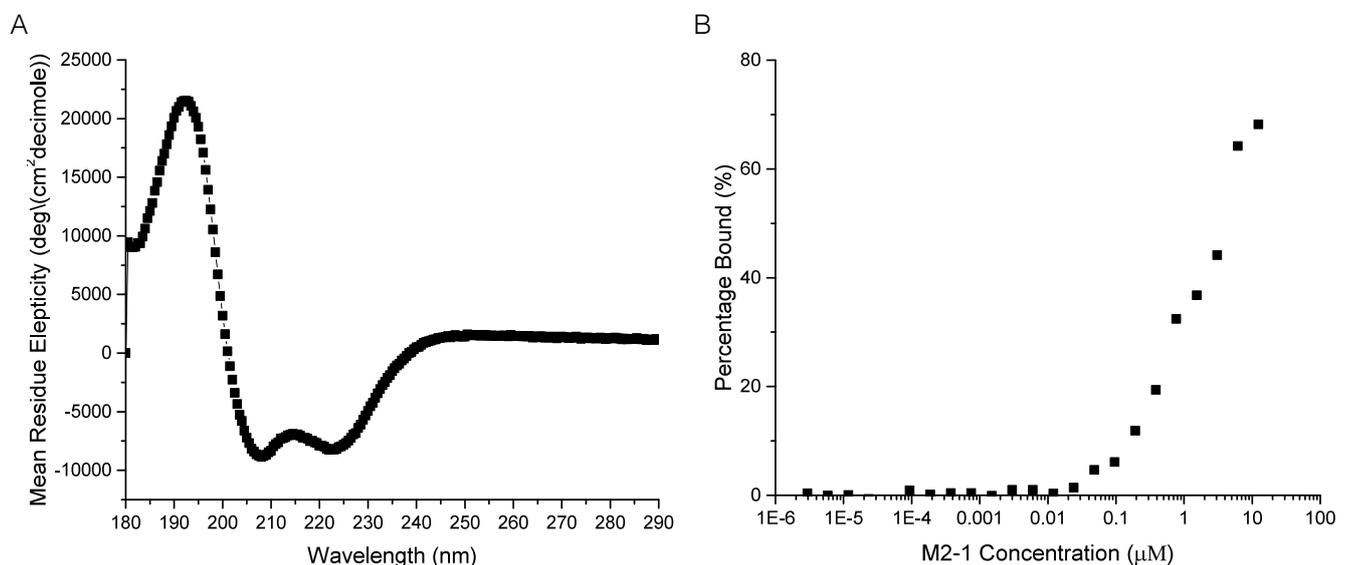

Figure 6.4 hRSV M2-1 Protein Quality Check

A- circular dichroism analysis of hRSV M2-1 in phosphate buffer pH 7.4 shows folded protein that is a-helical
B- direct binding of M2-1 to 3'Fl pA 13 RNA using fluorescence anisotropy

6.2.2 P$_{90-160}$

The M2-1:P$_{90-110}$ crystal structure was solved by Dr. Selvaraj Muniyandi (postdoc, University of Leeds) using M2-1 protein purified in this thesis. In the M2-1:P$_{90-110}$ crystal structure, P$_{90-110}$ presented itself as a single α-helix that resides within an M2-1 cleft formed from M2-1 α-helices 7, 8 and 9 (Figure 6.5). Moreover, the orientation of M2-1:P interaction was seen whereby both N-termini face each other (Figure 6.5). M2-1 favours RNA binding over P and it has been suggested that a protomer of P is displaced during RNA transcription to allow RNA binding to occur on an M2-1 protomer (Selvaraj *et al.*, 2018).

The M2-1 binding domain on P was proposed to include residues 100-120. However, a single binding site at P$_{90-110}$ interacts with M2-1 in GST pulldowns and the peptide directly binds M2-1 in fluorescence anisotropy assays; this was also the smallest peptide that still bound with high affinity (Selvaraj *et al.*, 2018). The coiled-coil domain of P (120-160) is also thought to be the oligomerisation domain and therefore the construct P$_{90-160}$ was used for further analysis. Moreover, P is punctuated by predicted unstructured regions, probably making full-length protein unsuitable for crystallography.
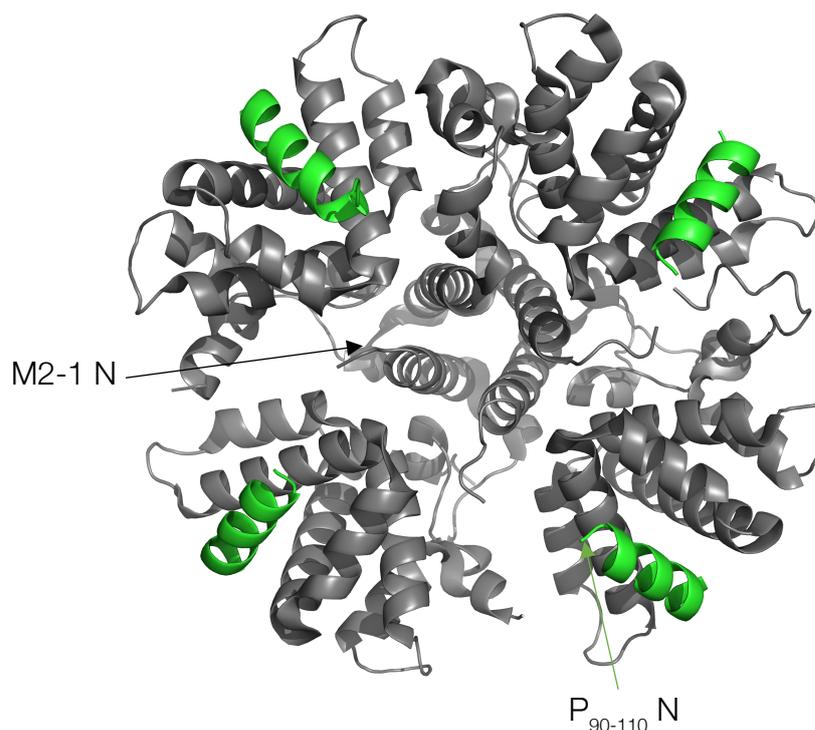

Figure 6.5 Crystal Structure of hRSV M2-1:P$_{90-110}$

M2-1 tetramer (grey) in complex with P90-110 peptide (green). N-termini (N) face in the same direction to the center of M2-1.
PDB: 6Y0G
Figure made in PyMol.

The structure of tetrameric M2-1 and tetrameric P may be able to reveal more information about this interaction and to make structural comparisons to M2-1:RNA.

Moreover, using techniques such as cryo electron-microscopy may be able to capture M2-1:RNA:P complex and offer insight into whether or not P is displaced by RNA.

The cDNA of P_{90-160} from hRSV (A2 strain) was ordered codon optimised for E. coli expression and cloned into the backbone of pET-28a with a custom designed non-cleavable C-terminal six histidine (His6) purification tag. A non-cleavable tag was used to aid purification (making the construct 10 kDa), ensuring bans would be visible on SDS-PAGE, for the concentrators used the Edwards group and the columns used for SEC. It was thought that an additional six non-native residues would not impact P_{90-160} tetramerisation and function.

6.2.2.1 Protein Expression

hRSV P_{90-110} was previously expressed as a GST fusion protein without complications and a similar buffer composition was adopted for this new construct (150 mM NaCl and 25 mM Tris). The His6 tag allowed ease during purification and the larger GST-fusion was not necessary as the new P construct (90-160) included the tetramerization domain producing oligomeric protein that was large enough for protein concentrators and size-exclusion columns already available in the Edwards/Barr group. The purification protocol was adapted to include imidazole (pH 8) (Table 3.4) to allow for P_{90-160}His6 elution during affinity chromatography.

Standard growth temperatures for E. coli BL21 (DE3) Gold cells were used and P_{90-160} expression was induced with 0.5 mM IPTG. E. coli cells were harvested by centrifugation and lysed using cell disruption.

6.2.2.1.1 Affinity Chromatography
C-terminally tagged P_{90-160} was purified from the soluble lysate via immobilised metal affinity chromatography, using nickel ions in a HisTrap HP column. The column was washed and P_{90-160} eluted with increasing concentrations of imidazole (Figure 6.6 A). Eluted fractions at 300 and 500 mM imidazole were pooled and concentrated to 10 mL for SEC.

6.2.2.1.2 Size-Exclusion Chromatography

Concentrated sample was injected onto a HiLoad 26/600 Superdex S-75 column attached to an AKTA purifier pump measuring absorbance at 280 nm. The chromatogram (Figure 6.6 C) revealed a single peak at 250 mL corresponding to a tetrameric P_{90-160} (~40 kDa) when compared to the calibration chromatogram for this column (ovalbumin ~43 kDa at 250 mL).

P_{90-160} expression produced soluble tetrameric protein reproducibly and no further optimisation was required.
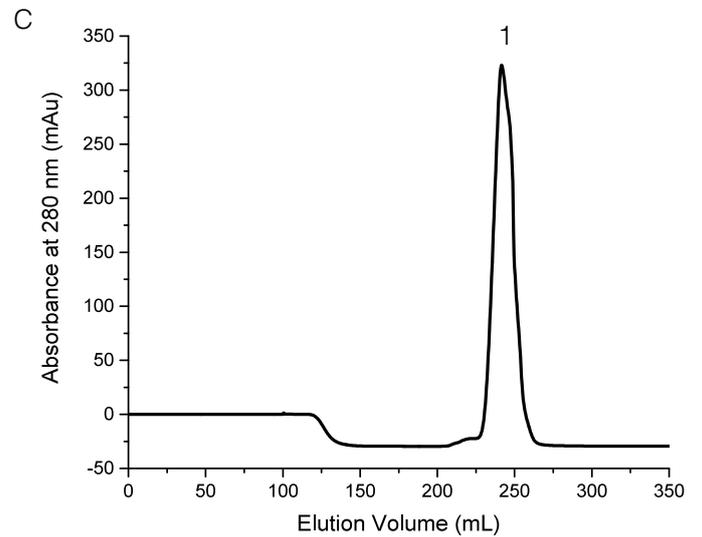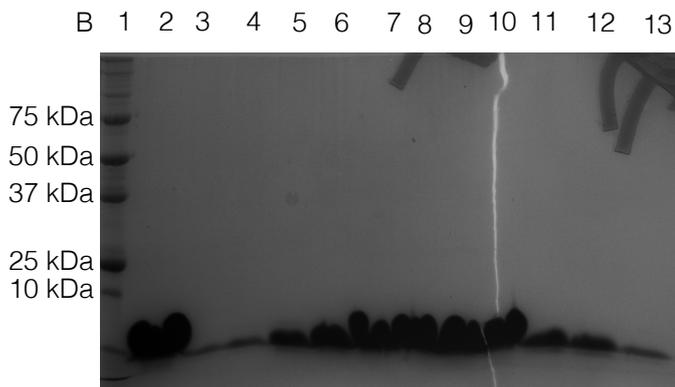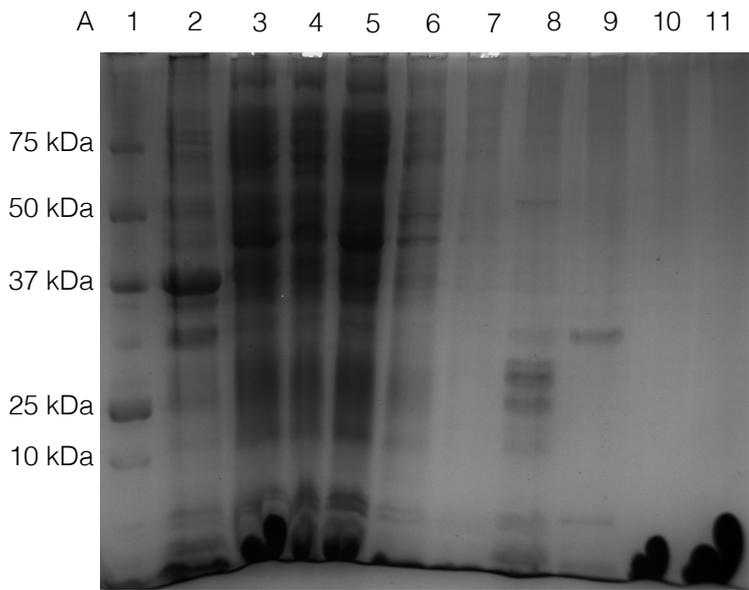
Figure 6.6 Purification of hRSV P$_{90\text{-}160}$

A- affinity chromatography; 1) marker, 2) insoluble, 3) soluble, 4) flow-through, 5-6) washes 7-8) high salt washes, 9) 150 mM imidazole elute, 10) 300 mM imidazole elute, 11) 500 mM imidazole elute
B and C- size exclusion chromatography separated P$_{90\text{-}160}$ on a HiLoad 26/600 Superdex S-75 column; 1) marker, 2) inject, 3-13) peak 1

6.2.3 M Protein

The cDNA of hRSV M protein was codon optimised for E. coli expression, synthesised and cloned into pET-28a-SUMO by Genewiz. The purification protocol was optimised from that published by Förster et al., 2015. M was previously expressed a His6 fusion protein (Förster *et al.*, 2015).

6.2.3.1 Optimisation of Expression and Purification

M was previously grown as 0.25 L cultures at 30°C for 5 hours, induced with 0.4 mM IPTG and temperature reduced to 25°C for 4 hours. Growth temperatures in the optimised protocol were altered to 37°C for optimal E. coli growth and protein expression induced with 0.5 mM IPTG for 16 hours at 18°C. Protein expression was induced at OD_{600} 0.6-0.8 to ensure E. coli were in the log phase for optimal protein expression.

E. coli cells were harvested by centrifugation and lysed by cell disruption as opposed to sonication. The addition of 0.25 % CHAPS (3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate) prevented M sticking to E. coli membranes. Lysate was clarified by centrifugation.

The above changes are outlined below in (Table 6.3). The following protocol was only adapted slightly for the different columns used in the Edwards/Barr research groups.

Table 6.3 Optimisation of hRSV M Expression and Purification

Published Protocol	Alteration	Improvement
5 hours at 30°C reduced to 25°C for 4 hours after induction	37°C until OD_{600} 0.6-0.8 reduced to 18°C for 16 hours	Optimal growth temperature for E. coli, inducing protein in the correct growth stage (log stage) optimal for protein expression and slower growth at 18°C for protein expression
Lysis by sonication	Lysis by cell disruption	Gentle and temperature controlled at 4°C

6.2.3.2 Affinity Chromatography

Soluble lysate was applied to a pre-equilibrated HisTrap HP column and washed to remove non-specific proteins. His6-SUMO-M was eluted with increasing concentrations of imidazole (150-300 mM) (Figure 6.7 A). The His6-SUMO purification tag was cleaved by SUMO protease during buffer exchange dialysis for 16 hours and separated on a HisTrap HP column. Soluble M was concentrated to 10 mL for SEC.

6.2.3.3 Size Exclusion Chromatography

SEC was performed on a HiLoad 26/600 Superdex S-200 column attached to an AKTA purifier pump recording UV absorbance at 280 nm. M was expected to elute as a dimer (~ 56kDa, monomer ~28 kDa) (Figure 6.7 B). Dimeric M eluted at ~225 mL (Figure 6.7 C) when compared to the calibration chromatogram for this column (conalbumin 75kDa, ~175 mL and ovalbumin 43kDa ~200 mL)

M expression produced soluble dimeric protein reproducibly and no further optimisation was required.
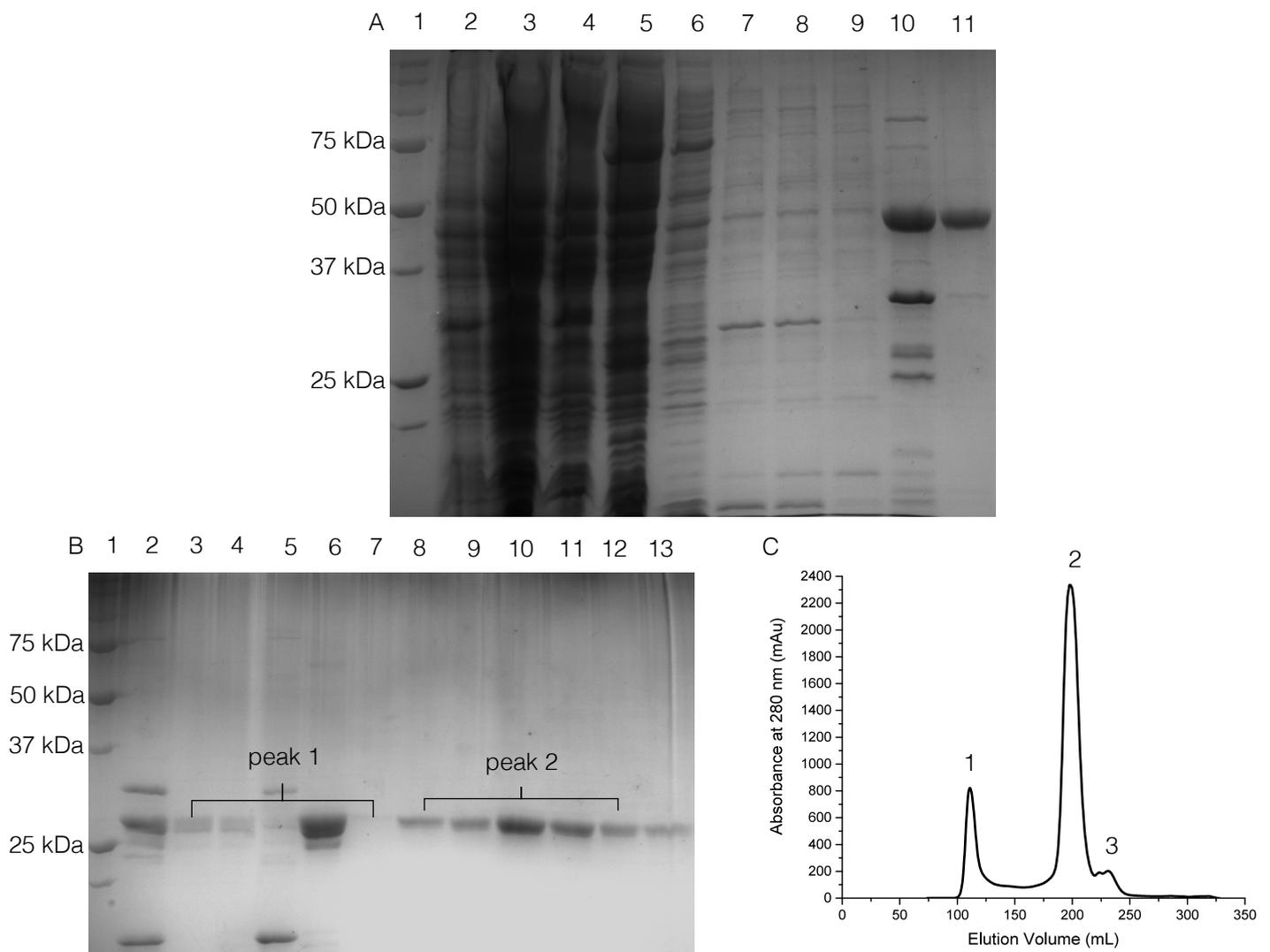
Figure 6.7 Purification of hRSV M

A- affinity chromatography 1) marker, 2) insoluble, 3) soluble, 4) flow-through 5-8) washes, 9) 150 mM imidazole eluate, 10) 300 mM imidazole eluate, 11) 500 mM imidazole eluate
B and C- size exclusion chromatography separated M on a HiLoad 26/600 Superdex S-75 column; B- 1) marker, 2) inject, 3-7) peak 1, 8-12) peak 2, 9) shoulder peak 3

6.3 Co-Complex Purification and Crystallisation

6.3.1 M2-1:P$_{90-160}$

It has been previously shown that M2-1 specifically binds P$_{90-110}$ and it was assumed that M2-1 would also bind P$_{90-160}$. To ensure this interaction occurred prior to crystal trials the co-complex was purified by SEC.

M2-1 was incubated with P$_{90-160}$ in a 1:1 molar ratio (60 µM) before SEC. SEC was performed on a HiLoad Superdex S-200 15/50 attached to an AKTA purifier pump. 3 peaks were seen on the chromatogram. 2 peaks were seen on the chromatogram after the void volume (Figure 6.8 A). The third peak eluted at ~2.25 mL and appeared to have the correct stochiometric ratio of M2:1:P and was therefore used for structural studies.
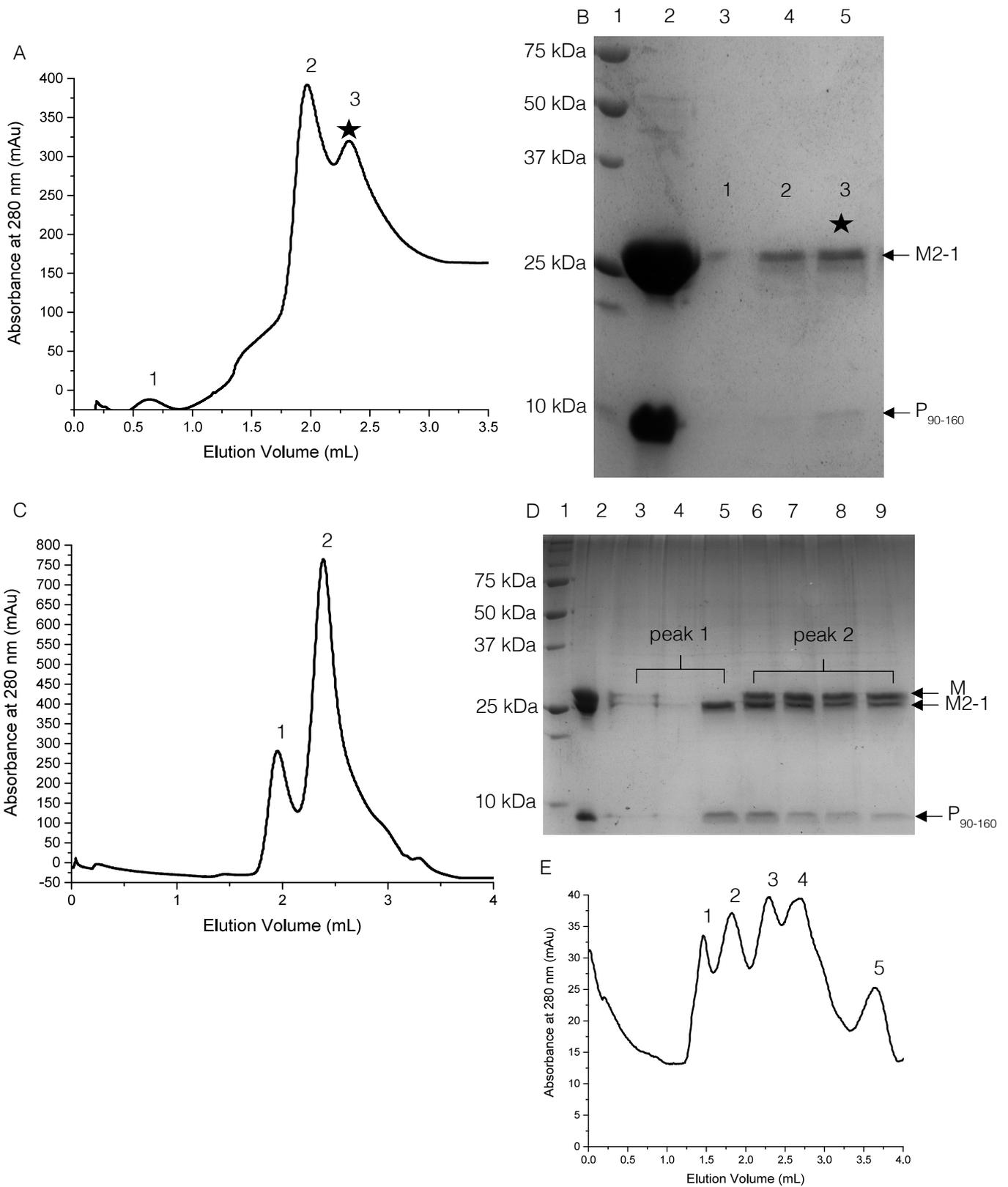
Figure 6.8 hRSV Proteins Complex Purification

Co-complex purification was performed on a HiLoad 26/600 Superdex S-200 column
A and B- M2-1 and P_{90-160}. B- 1) marker, 2) inject 3) peak 1, 4) peak 2, 5) peak 3
D and C- M2-1, M and P_{90-160}. B- 1) marker, 2) inject, 3-5) peak 1, 6-9) peak 2
E- S-200 calibration with 25 mM Tris-HCl pH 8..0, 300 mM NaCl, 2.5% glycerol;. 1) blue dextran 2000 kDa 2) ferritin 440 kDa, 3) aldolase 158 kDa, 4) conalbumin 75 kDa, Ribonuclease A, 13.7 kDa

6.3.1.1 Co-Crystal Trials of M2-1: $P_{90\text{-}160}$

Co-purified M2-1:$P_{90\text{-}160}$ was concentrated to 10 mg/mL (concentration based on a tetramer:tetramer complex interaction of 140 kDa) before crystal trials. Spherulite-like crystals were seen in various conditions using the sitting drop vapour diffusion (Figure 6.9 A-D) method but compared to previous crystal trials with apo M2-1, an alternative morphology was seen (Rachel Dods, PhD thesis 2018, University of Leeds); previously, M2-1 crystallised in a plate morphology, and needle-like crystals were observed for M2-1:$P_{90\text{-}110.}$ Different morphologies may have been due to $P_{90\text{-}160}$ binding and altering the space group.

6.3.1.1.1 Optimisation of M2-1:$P_{90\text{-}160}$ Spherulite Crystals

The spherulite-like crystals were further analysed by performing a fine screen based on the PEG 8000 concentration and 0.1 sodium cacodylate pH of the mother liquor. No crystals were seen in the optimised conditions.
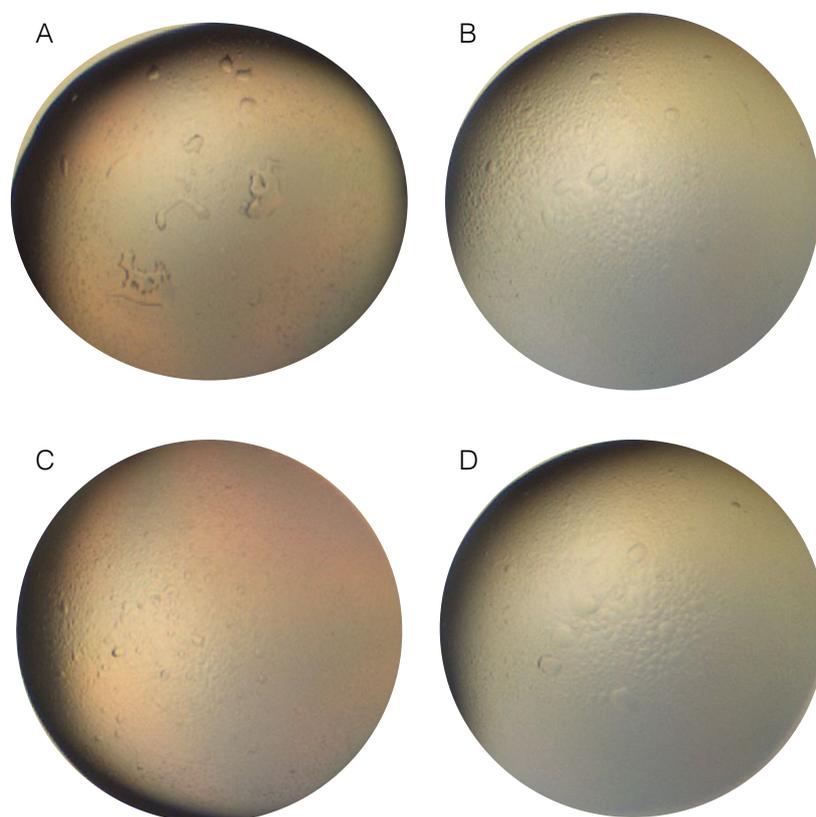
Figure 6.9 hRSV M2-1:$P_{90\text{-}160}$ Complex Crystal Morphology

10 mg/mL of co-purified M2-1:$P_{90\text{-}160}$ was used for initial crystallization trials using
A- 0.1 M MES 5 pH, 20 % w/v PEG 6K
B- 0.2 M KH2PO4, 20 %w/v PEG 3350
C- 0.16 M magnesium acetate, 0.08 M sodium cacodylate 6.5 pH, 16 %w/v
PEG 8K, 20 % v/v glycerol
D- 0.2 M calcium acetate, 0.1 M sodium cacodylate 6.5 pH, 18 % w/v PEG 8K

6.3.1.2 M2-1:P_{90-160} Negative Stain Electron Microscopy

Due to the structure of the M2-1:P_{90-160} binary complex not being solvable by X-ray crystallography (due to no crystals), we attempted to visualise the complex using negative stain transmission electron microscopy.

In the negative stain micrographs at 0.05 mg/mL, the M2-1:P_{90-160} complex appeared aggregated (Figure 6.10 panel A, white arrows). Although particles were sparse at 0.01 mg/mL less protein aggregation was seen (Figure 6.10 panel B, white arrows).
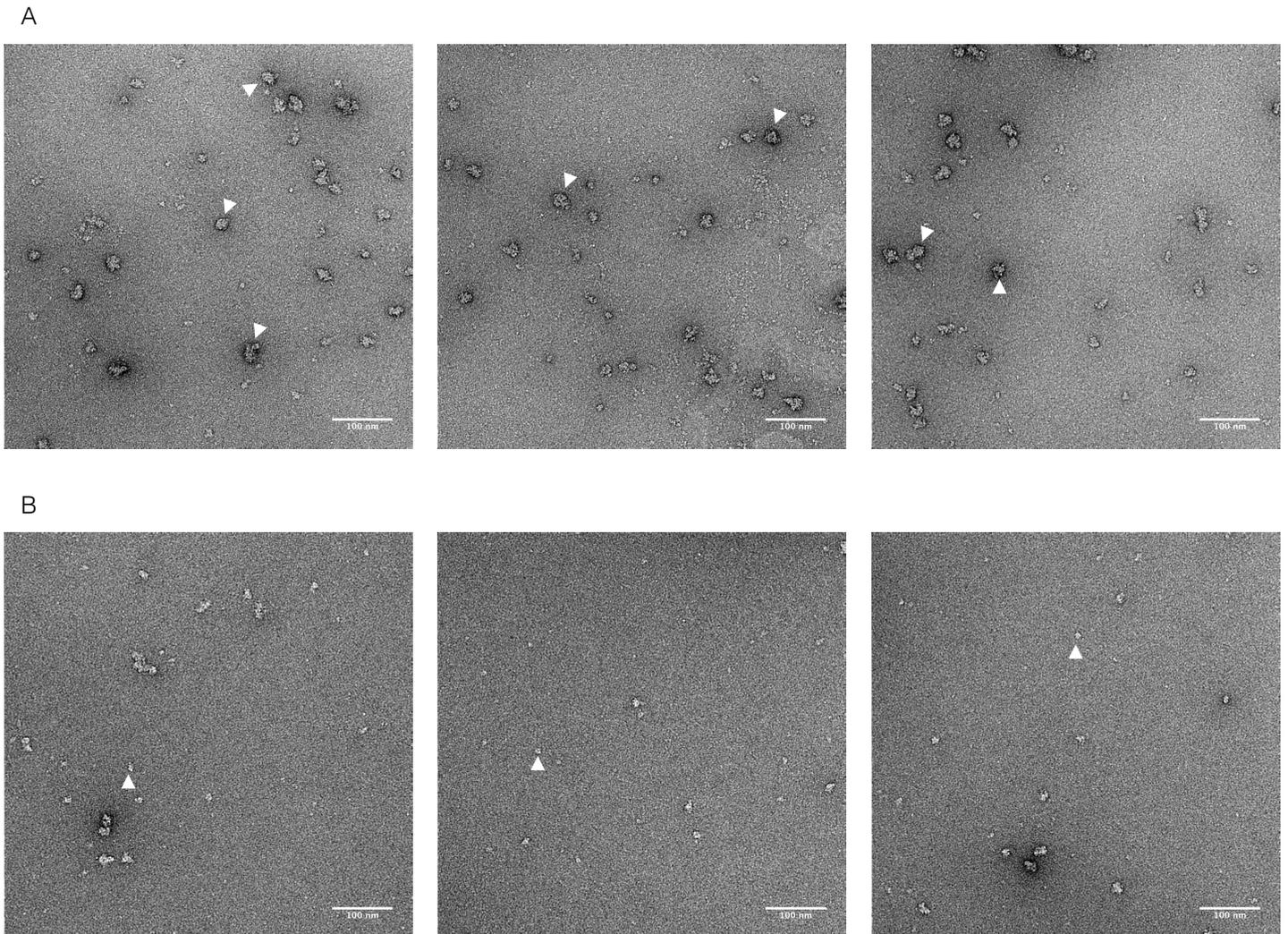
Figure 6.10 hRSV M2-1:P_{90-160} Complex Negative Stain Electron Microscopy

Micrographs taken at 49,000 K, 120 keV and -1.5 μm defocus
A- 0.05 mg/mL. Aggregates formed, highlighted by white arrows
B- 0.01 mg/mL. Tetramer- looking species highlighted by white arrows

6.3.1.2.1 Micrograph Collection

Micrographs were collected for data analysis from the grid coated with 0.01 mg/mL M2-1:P$_{90-160}$. Micrographs were collected at 49,000 x, 120 keV and a defocus of -1.5 µm.

6.3.1.2.2 Particle Picking and Class Averages

1000 particles were picked manually from 100 micrographs prior to auto-picking using RELION (regularised likelihood optimisation, MRC-LMB, University of Cambridge) software and sorted into 100 initial 2D classes (Figure 6.11 A). Classes that appeared tetrameric (~140 kDa, ~10 nm) were further analysed and sorted into 25 classes (Figure 6.11 B) and a further 10 classes were formed using 6914 particles (Figure 6.11 C).

6.3.1.2.3 Initial 3D Model

From the 10 classes made (Figure 6.11C), 2 classes (red boxes) were taken forward for initial 3D modelling. Class 1 (Figure 6.11 D) formed from 834 particles and class 2 (Figure 6.11 E) from 653 particles. Initial models were built in Chimera (UCSF Resource for Biocomputing, Visualisation and Informatics, Chimera-1.164 (Pettersen et al.,2004)). Both classes showed density of a suitable size for a tetramer:tetramer interaction (Figure 6.12 and Figure 6.13). The crystal structure of M2-1:P_{90-110} ((Selvaraj *et al.*, 2018) PDB: 6G0Y) fit into the EM density, and more density was filled for class 2 (Figure 6.13 B and C). In both models however, there are areas of EM density that were not filled. This 'extra' EM density could be for the ~224 amino acids of P_{90-160} tetramer and the non-cleavable His6 tag (~40kDa), however this density looked globular and the P_{90-160} protein tetramer is thought to be more elongated due to the coiled coil domain. More particles are needed to build a more detailed model, with higher resolution. This looks rather promising; however, we would need to proceed to single particle cryo-EM in order to solve the structure of the M2-1:P binary complex at an amino acid level.
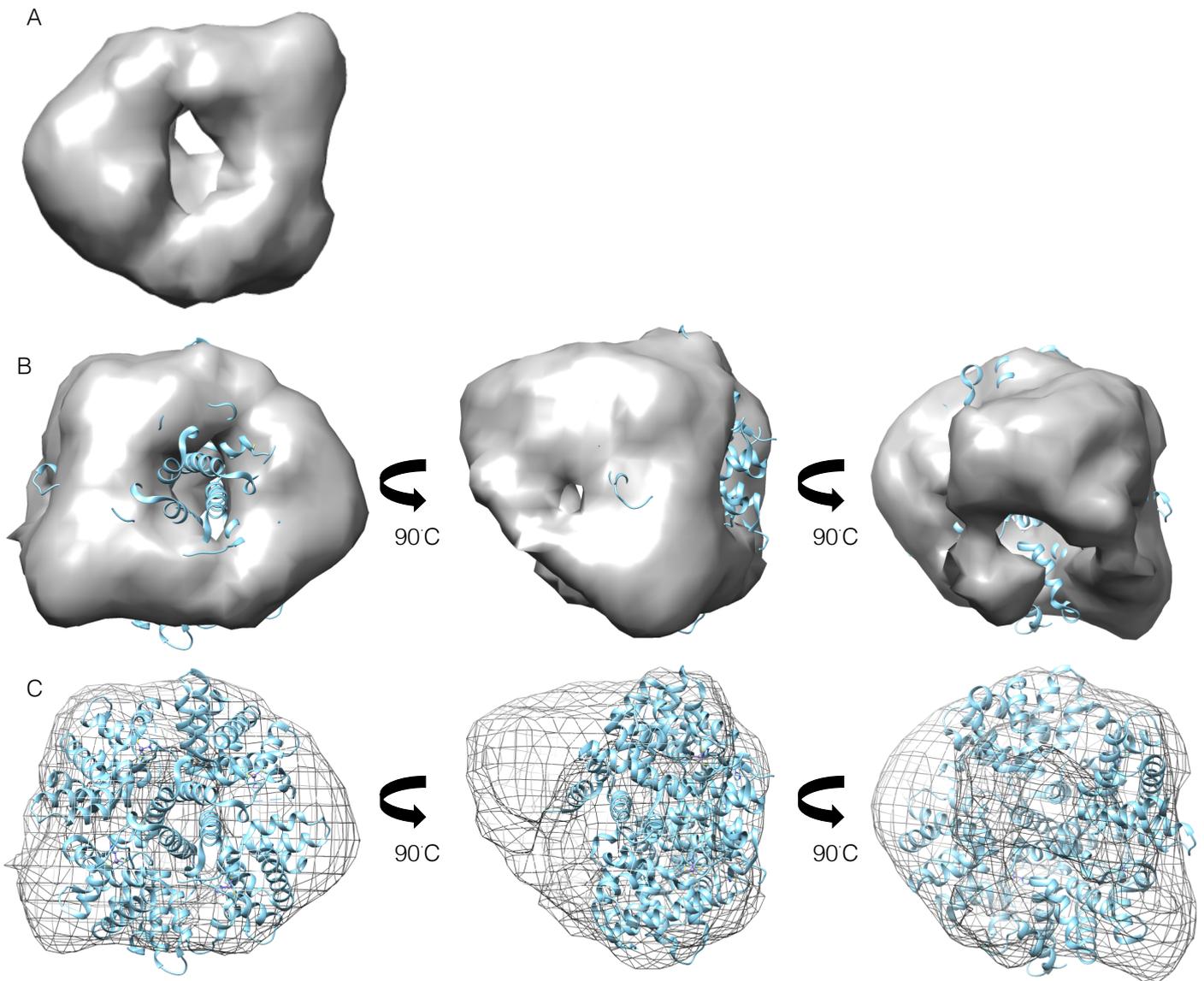
Figure 6.12 3D Model Fitting of hRSV M2-1:P_{90-160} Complex from Class 1

A- negative stain EM density
B- and C- fitting of M2-1:P_{90-110} crystal structure (PDB: 6G0Y) into negative stain EM density. B- surface view and C- mesh view
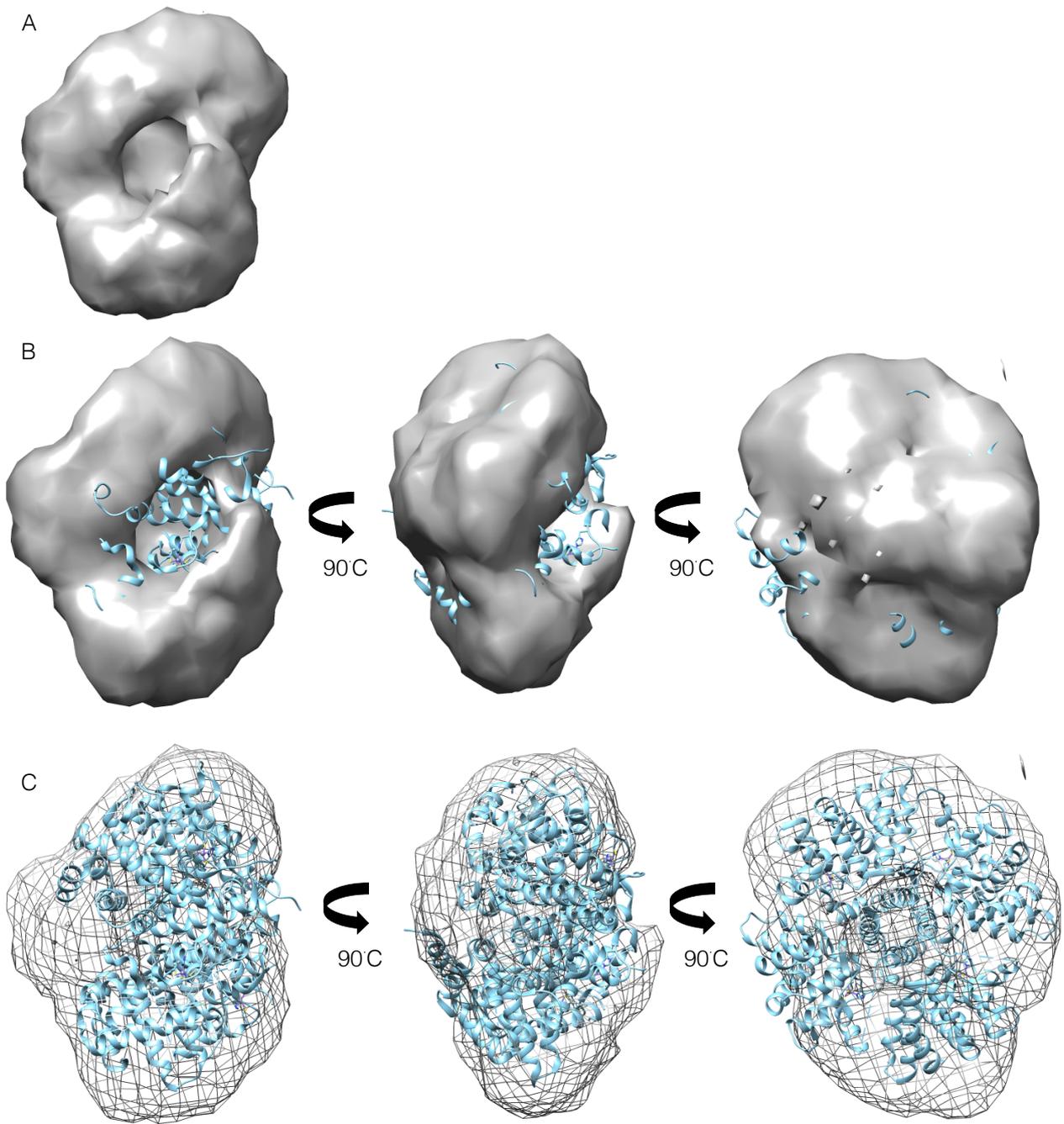
Figure 6.13 3D Model Fitting of hRSV M2-1:P$_{90-160}$ Complex from Class 2

A- initial model of negative stain EM map
B- and C- fitting of M2-1:P$_{90-110}$ crystal structure (PDB: 6G0Y) into negative stain EM density. B- surface view and C- mesh view

6.3.2 M:M2-1:P90-160

It has been previously shown that M2-1 forms a link between M and RNPs and the distance seen between M2-1 density between M and RNP being 12.6 nm. The regulatory of spacing was thought to be due to linkage associated with M or the M2-1:P interaction from the RNP complex (Kiss *et al.*, 2014). In addition to its anti-termination function, M2-1 colocalises with cytoplasmic inclusions of viral NC-proteins (L, NP and P) and viral genome associating with the RNP complex through interactions with P and the N-terminal domain of M (Mason *et al.*, 2003; D. Li *et al.*, 2008).

We therefore wanted to see the ternary complex between M, M2-1 and P. The P_{90-160} construct was chosen here as it was readily soluble with a small non-cleavable N-terminal 6His tag and included the M2-1 binding domain and tetramerisation domain.

The stoichiometry of the M:M2-1 interaction was unknown and so a 2:1:1 ratio was used whereby 2 dimers of M binds to a tetramer of M2-1 and a tetramer of P. Co-complex purification was performed on a HiLoad Superdex 15/50 S-200 column attached to an AKTA purifier pump. 1 peak contained all 3 proteins (Figure 6.8 C and D peak 2) and fractions from this peak were used for further structural studies.

6.3.2.1 Co-Complex Crystal Trials

Co-purified M:M2-1:P_{90-160} was concentrated to 12 mg/mL before crystal trials. Crystals were seen in various conditions using the sitting drop vapour diffusion (Figure 6.14).

Crystal 'hits' were seen to centre around calcium chloride ($CaCl_2$), sodium acetate and 2-methyl-2,4-pentaneiol (MPD) for M:M2-1:P_{90-160} crystals. An optimisation screen was made using differing concentrations of $CaCl_2$ and MPD and differing pH of Na Acetate at 0.1 M (Figure 6.15).
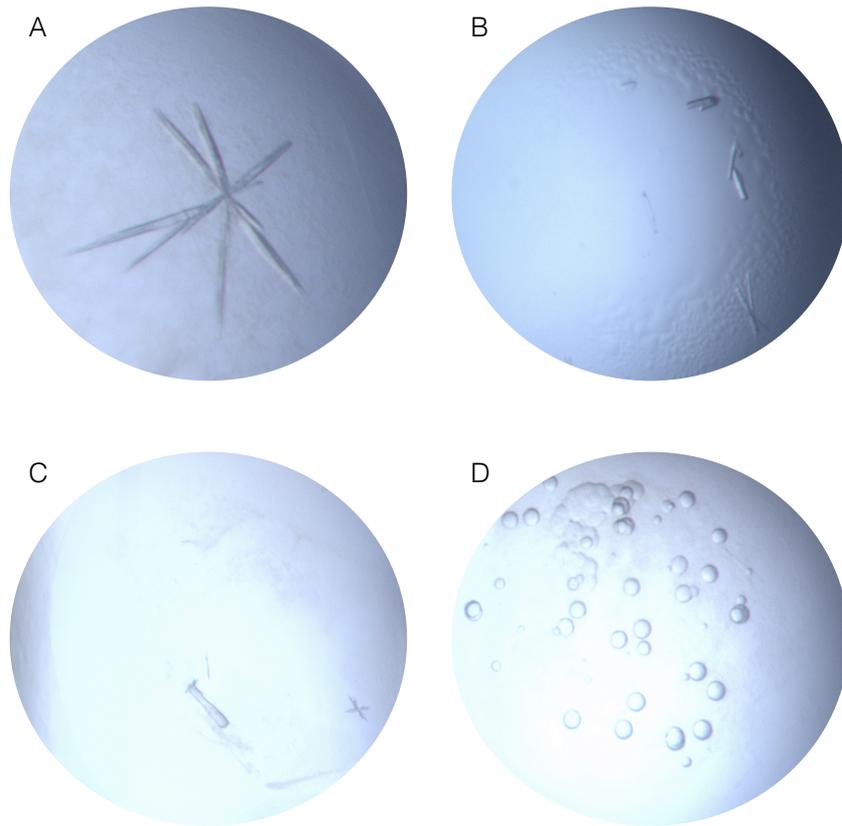
Figure 6.14 hRSV M2-1:M:P$_{90-160}$ Complex Crystal Morphology

12 mg/mL of co-purified M2-1:M:P$_{90-160}$ was used for initial crystallisation trials.
A- 0.2 M calcium chloride, 0.1 M sodium acetate pH 4.6, 30% w/v MPD.
B- 0.2 M sodium chloride, 0.1 M CHAPS pH 10.5, 20% w/v PEG 8000.
C- 0.2 M sodium chloride, 0.1 M TRIS pH 7, 35% w/v MPD.
D- 0.2 M calcium chloride, 0.1 M sodium acetate pH 4.6, 20% v/v 2-propanol.
Crystals of different morphology were visualised.
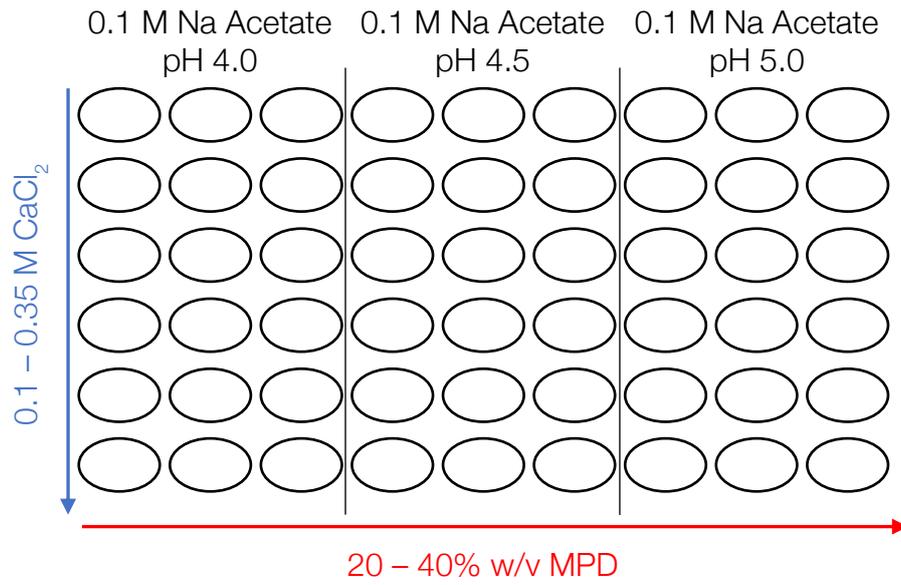
Figure 6.15 Optimisation Conditions for M:M2-1:P_{90-160} Crystallisation Trials

Optimisation conditions varied in pH for sodium acetate (pH 4-5), concentration of calcium chloride (0.1-0.35) and concentration of MPD (20-40%). Mother liquor drops were made to a final volume of 1 mL and crystal trials utilised the hanging-drop vapour diffusion method.

6.3.2.1.1 X-Ray Diffraction Data Collection from Optimisation Plates

A subset of 8 crystals were harvested from optimisation plates and tested for diffraction on the i24 beamline (DLS) (Table 6.4). Each crystal was tested for diffraction at 100 % transmission, 90° oscillations to a resolution of 1.9 Å.

Table 6.4 Crystal ID for Optimised Crystal Hits X-Ray Diffraction Data Collection

Crystal ID	Optimisation Plate Drop	Conditions
1	A1	0.05 M CaCl2 (Salt) 0.1 M Na Acet 4 pH (Buffer) 30 %v/v MPD (Precipitant)
2	A2	0.05 M CaCl2 (Salt) 0.1 M Na Acet 4 pH (Buffer) 36.667 %v/v MPD (Precipitant)
3	A3	0.05 M CaCl2 (Salt) 0.1 M Na Acet 4 pH (Buffer) 43.333 %v/v MPD (Precipitant)
4	A4	0.05 M CaCl2 (Salt) 0.1 M Na Acet 4 pH (Buffer) 50 %v/v MPD (Precipitant)
5	A1	0.05 M CaCl2 (Salt) 0.1 M Na Acet 4 pH (Buffer) 30 %v/v MPD (Precipitant)
6	A1	0.05 M CaCl2 (Salt) 0.1 M Na Acet 4 pH (Buffer) 30 %v/v MPD (Precipitant)
7	A4	0.05 M CaCl2 (Salt) 0.1 M Na Acet 4 pH (Buffer) 50 %v/v MPD (Precipitant)
8	B2	0.114 M CaCl2 (Salt) 0.1 M Na Acet 4 pH (Buffer) 36.667 %v/v MPD (Precipitant)

The diffraction pattern seen appeared similar to diffraction pattern of salt and no further data was collected.

6.3.2.1.2 X-Ray Diffraction Data for Initial Crystal Hits Collection from Initial Co-Complex Screens

As crystals from optimisation plates did not diffract (Figure 6.17), as a last resort the initial crystals from the initial co-complex crystal screening trial (Figure 6.14) were harvested for diffraction data collection and flash frozen in liquid nitrogen. Multiple crystals from the same condition were picked for the best 3 conditions (Figure 6.14 A, B and C). Crystals were shipped to the DLS for screening and data collection (Table 6.5).
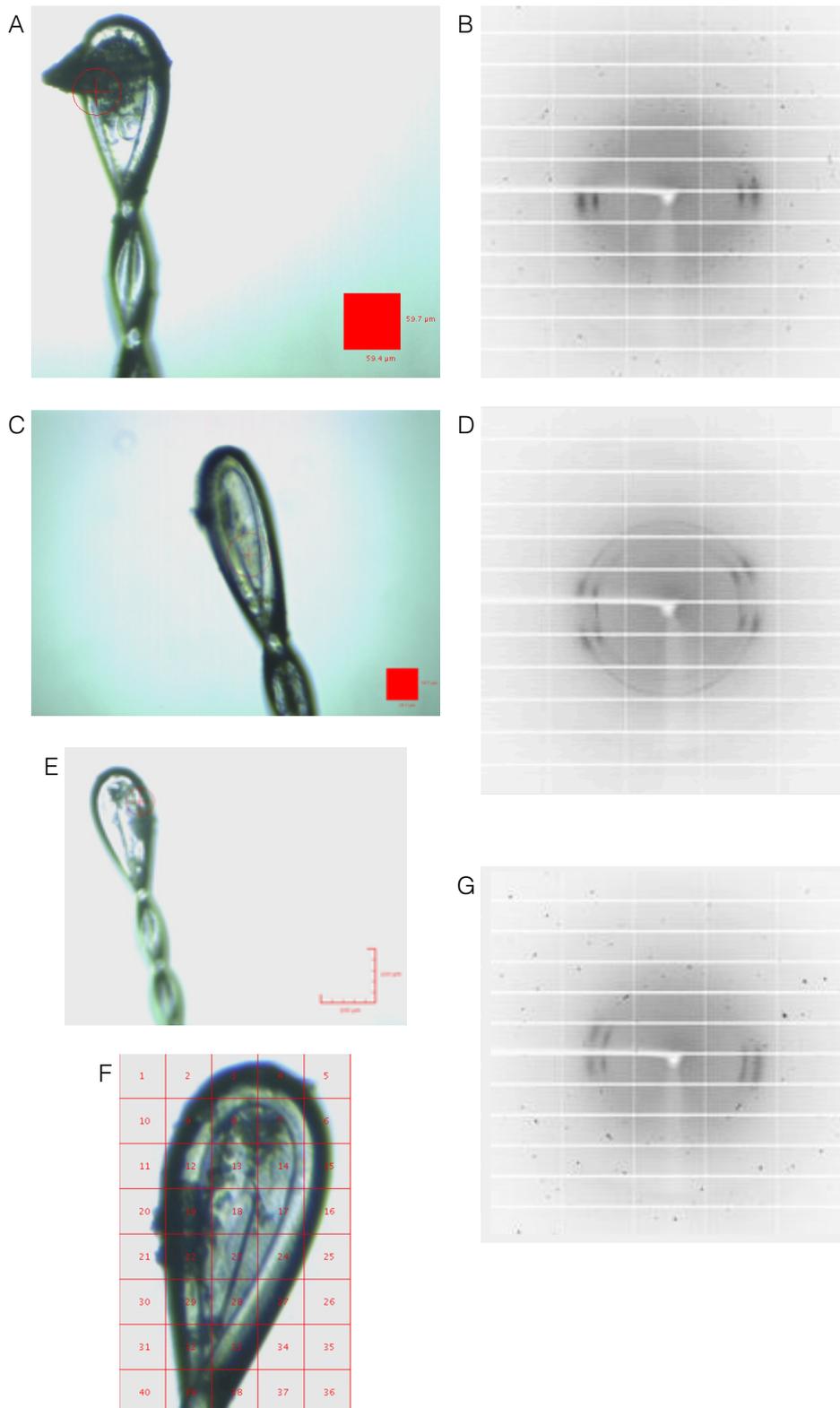
Figure 6.16 hRSV M2-1:M:P90-160 Complex Crystal Harvesting and Data Collection

A- crystal 1
B- diffraction image of crystal 1 shown in A. Crystal 1 did not diffract.
C- crystal 4
D- diffraction image of crystal 4 shown in C. Crystal 4 did not diffract
E- crystal 8
F- line scan analysis of crystal 8
G- diffraction image of crystal 8 shown in E. Crystal 8 did not diffract

Table 6.5 Crystal ID for X-Ray Diffraction Data Collection

Crystal ID	Plate	Condition
1	JCSG Core 1 G12	0.02 M CaCl2 (Salt) 0.1 M Na Acet 4.6 pH (Buffer) 30 %v/v MPD (Precipitant)
2	JCSG Core 1 G12	0.02 M CaCl2 (Salt) 0.1 M Na Acet 4.6 pH (Buffer) 30 %v/v MPD (Precipitant)
3	JCSG Core 1 G12	0.02 M CaCl2 (Salt) 0.1 M Na Acet 4.6 pH (Buffer) 30 %v/v MPD (Precipitant)
4	JCSG Core 2 A1	0.2 M NaCl (Salt) 0.1 M CAPS 10.5 pH (Buffer) 20 %w/v PEG 8K (Precipitant)
5	JCSG Core 2 A1	0.2 M NaCl (Salt) 0.1 M CAPS 10.5 pH (Buffer) 20 %w/v PEG 8K (Precipitant)
6	JCSG Core 2 A1	0.2 M NaCl (Salt) 0.1 M CAPS 10.5 pH (Buffer) 20 %w/v PEG 8K (Precipitant)
7	JCSG Core 3 C12	0.2 M NaCl (Salt) 0.1 M TRIS 7 pH (Buffer) 35 %v/v MPD (Precipitant)
8	JCSG Core 3 C12	0.2 M NaCl (Salt) 0.1 M TRIS 7 pH (Buffer) 35 %v/v MPD (Precipitant)

The pins used for harvesting crystal 6 and 8 were broken during the process and screening for these crystals could not proceed.

Diffraction was not seen for crystals 1-3, 5 and 7 (Figure 6.17). A data set was collected for crystal 4 at 100% transmission, 0.2° oscillations for a total of 360°, an exposure time of 2 seconds and to a maximum resolution of 1.70 Å.
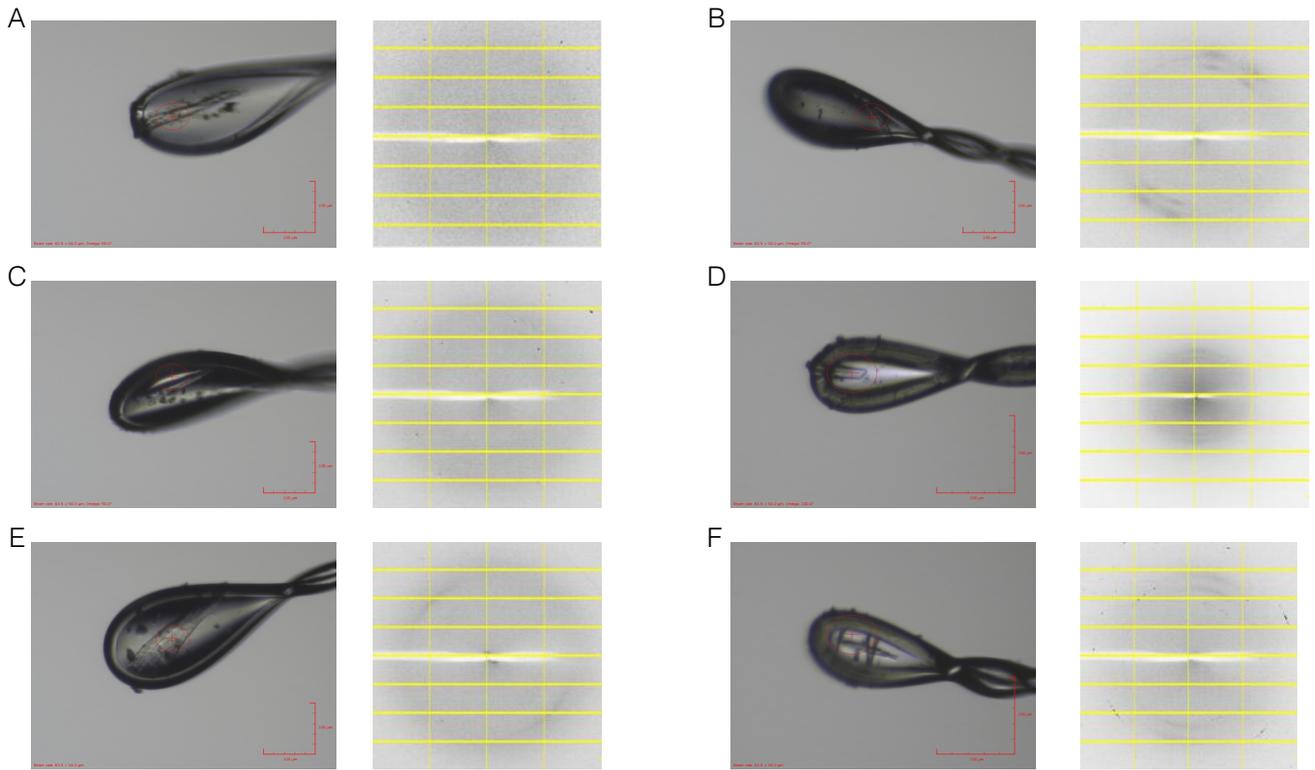
Figure 6.17 hRSV M2-1:M:P90-160 Complex Crystal Harvesting and Data Collection

A- crystal 1 and diffraction image of crystal 1. Crystal 1 did not diffract.
B- crystal 2 and diffraction image of crystal 2. Crystal 2 did not diffract.
C- crystal 3 and diffraction image of crystal 3. Crystal 3 did not diffract.
D- crystal 4 and diffraction image of crystal 4. Crystal 4 diffracted weakly.
E- crystal 5 and diffraction image of crystal 5. Crystal 5 did not diffract.
F- crystal 7 and diffraction image of crystal 7. Crystal 7 did not diffract

Auto-processing software in ISPyB database (Information Systems for Protein Crystallography Beamlines) (SynchWeb interface, DLS) (Delageniere *et al.*, 2011) defined the space group for crystal 4 data collection as C121 giving a small unit cell (Table 6.6). This was the same space group for the published structure of hRSV M (PDB: RV23, 1.70 Å) and this model was used to solve the structure via molecular replacement using Phaser against the xia2 3dii auto-processed data (resolution 1.99 Å). An initial model was built and showed a dimer of M per asymmetric unit, as expected from the published structure. No additional density was seen for hRSV M2-1 and P_{90-160} (Figure 6.18). As the published structure of hRSV M2-1 was 1.70 Å full refinement was not performed.

Table 6.6 Data Collection Statistics for Crystal 4

	Auto-Processing Software		
	xia2 dials	xia2 3dii	fast_dp
Wavelength (Å)	0.98	0.98	0.98
High Resolution Limit	65.70	65.67	27.43
Low Resolution Limit	2.04	1.99	2.13
Completeness	98.8	98.6	98.3
Multiplicity	7.07	6.98	7.1
CC-Half	0.99	0.99	0.998
I/sigma	9.15	5.23	12.10
Rmerge (I)	0.13	0.12	0.11
Anomalous Completeness (%)	98.35	98.89	97.8
Anomalous Multiplicity	3.60	3.56	3.6
A	52.72	52.71	52.71
B	79.51	79.53	79.52
C	66.05	66.03	66.03
α	90.00	90.00	90.00
β	95.95	95.95	95.95
γ	90.00	90.00	90.00
Space Group	C121	C121	C121
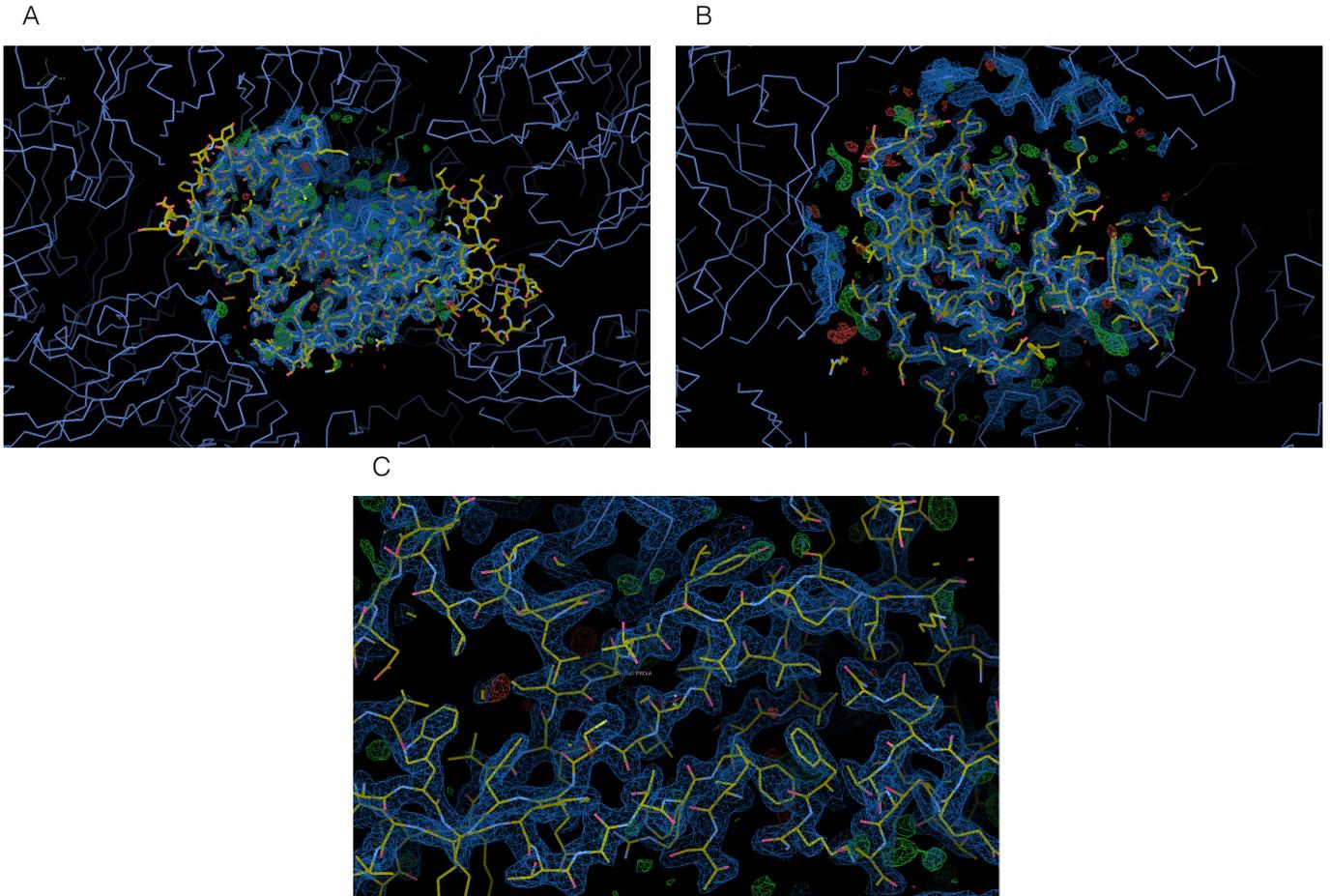
Figure 6.18 Electron Density Maps for Crystal 4

2Fo-Fc (1.0 sigma) electron density map. Red density shows what is an overfit, not accounted for by the mode. Green and blue density shows built electron density (blue) around the model and density that needs to be built (green).
A- 2 monomers are seen per asymmetric unit.
B- electron density map show good coverage and minimal red and green density.
C- electron density map to show density is filled by M amino acids entirely.

6.4 Discussion

6.4.1 Protein Purification

In order to produce high yields of M2-1 for material intensive methods such as X-ray crystallography, there was a need for an improved purification method. The previous and published method of purification was time consuming and resulted in protein degradation. Moreover, non-homogenous protein sample was detrimental to crystal formation and often resulted in 'negative hits'. This new purification method outlined in

Table 6.2 resulted in ~35 mg of pure protein from 2 L of E. coli culture. Further, the optimised expression protocol for hRSV M, decreased the total time for expression and optimised E. coli growth for protein expression, the changes are also outlined in Table 6.3. Unfortunately, the shift in the elution profile of M:M2-1:P_{90-160} compared to M2-1:P_{90-160} was small (~0.5 mL) making it difficult to elucidate the oligomeric action of the M:M2-1 interaction (Figure 6.8 E). In future, the M2-1-P_{90-160} complex, and M:M2-1:P_{90-160} complex should be subjected to native MS to support the formation of complexes, unfortunately due to time restraints this was not possible.

6.4.2 Structural Studies

The purification improvements allowed X-ray crystallography experiments. The vapour diffusion method with sitting drops was initially used for M2-1: P_{90-160} and M:M2-1:P_{90-160} was initially used for crystallisation trials with commercially available factorial screens. Optimisation of crystal; 'hits' utilised vapour diffusion method with hanging drops with and increased drop size (0.1 μL to 1 μL). Interestingly, for M2-1:P_{90-160} protein crystals were not seen for corresponding conditions utilising the sitting drop method. However, the 'crystals' formed in the initial crystallisation trials were 'quasi-like' and perhaps the addition of P_{90-160} did not allow for the ideal surface tension conditions for crystal formation. It has been previously shown that the addition of RNA to M2-1 results in protein aggregation, potentially due to the RNA binding to more than one tetramer (Leyrat *et al.*, 2014). A similar binding pattern may also occur when using P_{90-160} construct as this is also tetrameric. It has previously been proposed that the M2-1:P interaction may be simultaneous thus allowing for P tetramers to bind multiple M2-1 tetramers. Whilst SEC was used to rule these interactions out (the elution volume observed was appropriate for the size of a tetramer:tetramer complex) and a 1:1 ratio was visualised, during preparation for crystal trials the co-complex was concentrated, and aggregates may have formed here. This may also explain the aggregation observed during negative stain electron microscopy. Grid-making conditions have been optimised as with lower concentrations of the complex less aggregation was seen. In order to produce a model, an increased number of micrographs will need to be collected in order to increase the number of particles picked. The model produced in *Figure 6.12* looks promising for the binary complex of M2-1:P_{90-160}, however, the limitations of negative-stain electron microscopy probably wouldn't allow for the further conclusions. To perform analysis cryo-EM will need to be performed in order to assess to M2-1:P interaction at amino acid level, this will also need to be optimised. High-

resolution cryo-EM data is also important in obtaining any further structural information for M2-1 also, as a 2.2 Å crystal structure of M2-1 solved by Rachel Dods (PhD student, University of Leeds, unpublished data) did not reveal any further details; the 2.2 Å model did not resolve further the C-terminal residues 175-194 and the phosphorylation loops encompassing Ser 58 and 61 when compared to Tanner's 2.5 Å structure (Tanner *et al.*, 2014).

Disappointingly, the structural studies undertaken in this chapter failed to produce a suitable model for the interaction of M:M2-1:P_{90-160} using X-ray crystallography or EM. The high concentrations of NaCl (300 mM) in the SEC buffer of M:M2-1:P_{90-160} was maybe the cause of salt crystals and the lack of protein diffraction during diffraction pattern collection. Co-complex purification would need to be further optimised to reduce an NaCl concentration in the SEC buffer but keep M dimeric in order to prevent the formation of long M filaments (Förster *et al.*, 2015).

6.5   Chapter Summary

This chapter outlined an optimised purification method for hRSV M2-1, P_{90-160} and M to produce high quality protein in large quantities for structural studies.

How these known binding partners interact still remains elusive in the complex of the virion and the transcription complex. M2-1:P is known to be a 1:1 tetramer interaction however in transcription it is unknown whether interactions at individual protomers allow for M2-1 to bind the newly synthesised RNA as well as P.

6.6   Future Directions

To aid the understanding of M2-1 and binding partners. the initial X-ray crystallography methods and negative stain electron microscopy undertaken in this chapter will need to be optimised to produce co-purified protein complexes with low salt in the SEC buffer whilst still maintaining M in a dimeric or even monomeric state to prevent large oligomers forming and potential aggregation.

While the negative stain EM structure here clearly shows a tetrameric protein. more particles will be needed to produce a model that is reliable. Moreover, an M2-1 only negative stain EM structure will also aid the docking of P_{90-160} within the structure and for comparisons to be made on their (if any) difference to M2-1 and how/if it moves to accommodate tetrameric P.

Chapter 7 Concluding Remarks

This thesis presents the characterisation of *Orthonunyavirus* NP in complex with RNA. This was possible due to previously published protocols for purification, which were optimised, and subsequent protocols developed to optimise how these NPs were taken forward for analysis by X-ray crystallography and FA. Despite the high resolution of crystal structures presented in this thesis, the specific interactions involved in RNA binding still remains elusive. The binding data here however, confirms the functionality of residues that were selected for mutagenesis. High resolution structures for both protein and RNA would elucidate the specific interactions with RNA which facilitate RNP formations, while still permitting essential access of replication machinery to the RNA.

The 2013-2016 outbreak of EBOV in Africa highlighted the need for an effective therapeutic to EVD and since then efforts are being made to implement the Ad5-EBOV vaccine and the rVSV-ZBOV vaccine, which was administered in 303,000 patients in the 2018-2020 outbreak. VP35 is an essential component of the RNP complex associating with L and NP and interacting with viral RNA. VP30 is the essential activator of the viral transcription complex and is regulated by phosphorylation. Exactly how VP35:VP30 interaction occurs currently remains elusive and is thought to be mediated by RNA. This thesis presents the optimised purification protocols for both VP35 and VP30 and elucidated the oligomeric state of both proteins. VP35 is now believed to be tetrameric supported by SEC-MALLS and MS data. Initial structural analysis by negative-stain EM revealed tetrameric-like particles and classes, however further optimisation of grids is needed to produce reliable 2D class averages and 3D reconstructions. The proposed RNA binding domain of VP30 did not solely contribute this function of VP30, rather the ZBD and hexamerisation of VP30 play key roles in higher affinity binding, this can be assessed further by residue and phosphoablative mutants. Moreover, a novel finding included that VP30 preferentially binds A-rich RNA, similarly to the structural homologue M2-1 from hRSV; proposing two mechanisms for VP30 functioning: either VP30 binds the newly synthesis mRNA pA tails or recognises A-rich regions of genomic EBOV RNA (negative-sense) to support transcription.

hRSV poses a major global disease burden resulting in lower respiratory tract infections, pneumonia and asthma in later life. Premature death in infants, the elderly and the immunocompromised also occur and mortality is linked to developing countries where access to therapeutics is limited to cost. The crystal structure of M2-1 bound to P_{90-110} laid the groundwork for research presented in this thesis. We wanted to characterise the interaction between tetrameric M2-1 and tetrameric P (90-160) which may have given more insights into the events that occur during viral transcription. We confirmed the interaction between M and M2-1 which hypothesis a role for M2-1 in viral assembly. Structural studies were unsuccessful in obtaining this information, however this thesis lays the foundations for cryo-electron microscopy analysis which will reveal information at the atomic level. To further assess this hypothesis M2-1 truncation assays and residue ablations can also be performed into order to characterise the M2-1:M interaction.

SBDD is a rational for drug development, which has become possible due to scientific advancements in structure determination (X-ray crystallography automatic, synchrotron power and cryo-EM advancements) paired with computational tools. The iterative process required collaborative efforts across structural and cellular biology, and medicinal chemistry. SBDD requires a high-resolution model of a protein of interest in order to allow identification of atoms and a target site. Virtual docking has allowed the prediction of ligand conformation within the target site, this is then evaluated by prediction of favourable intermolecular interactions including hydrogen bonds and hydrophobic contacts. Successful docking is improved with co-complex structures highlighting target sites. Successful 'hit' drug candidates are further analysed for binding and functional activity, in order to develop structure-activity relationships. A structure with the lead drug candidate in complex with the protein of interest is desirable, allowing for future development of a more potent drug candidate allowing direct visualisation of the binding mechanism (Kitchen *et al.*, 2004). Structural information of these aforementioned proteins would in turn aid SBDD, targeting groups of viruses which pose an increasing threat to human and animal health and for which there is currently a paucity of available therapies. It is hoped that this thesis as laid the foundation for such high-resolution data to be collected and processed in the future.

<div align="center">Appendix</div>

A.1 Protein Bibliography

Table A.1 Molecular Weight of Proteins

Viral Protein	Molecular Weight (kDa)
AKAV NP	26
OROV NP	26
SIMV NP	26
EBOV VP35$_{1-340}$	35
EBOV VP35$_{80-340}$	28
MARV VP35$_{1-329}$	35
MARV VP35$_{60-329}$	29
EBOV VP30$_{1-288}$	32
EBOV VP30$_{8-272}$	30
EBOV VP30$_{69-272}$	23
EBOV VP30$_{87-265}$	20
MARV VP30$_{1-281}$	30
MARV VP30$_{17-278}$	28
MARV VP30$_{79-278}$	22
MARV VP30$_{101-278}$	19
hRSV M2-1$_{1-194}$	25
hRSV P$_{1-241}$	27
hRSV P$_{90-160}$	9
hRSV M$_{1-256}$	28

A.2 Protein Sequences

A.2.1 Bunyavirales

A.2.1.1 AKAV NP

```
MANQFIFNDVPQRNAATFNPDAGYVAFISKYGQQFNFTVARVFFLNQKKAKMVLHKTPQPSV
DLTFAGVKFTVVNNHFPQYTANPVSDTAFTLHRISGYLARWVAEQCKANQIKFAEAAATIVM
PLAEVKGCTWSDGYAMYLGFAPGAEMFLETFEFYPLVIDMHRVIKDGMDVNFMRKVLRQRYG
QLTAEEWMTSKLDAVKAAFGSVAQISWAKSGFSPAARAFLAQFGIQI
```

A.2.1.2 OROV NP

```
MSEFIFNDVPQRTTSTFDPEAAYVSFEARYGQVLNAGVVRVFFLNQKKAKDVLRKTSRPMVD
LTFGGVQFAMVNNHFPQFQSNPVPDNGLTLHRLSGYLARWAFTQMRSPIKQAEFRATVVVPL
AEVKGCTWNDGDAMYLGFAAGAEMFLQTFTFFPLVIEMHRVLKDGMDVNFMKKVLRQRYGQK
TAEQWMREEIVAVRAAFEAVGTLAWARTGFSPAARDFLRQFGIGI
```

A.2.1.3   SIMV NP

```
MANQFIFEDVPQRNLSTFSPEAGYVAFIGRYGQQLNFSVVRVFFLNQKKAKMVLHKTAQPNV
DLTFGGVKFTLVNNNFPQYTANPVPDNALTLHRLSGYLARWTAEQVKNNQVKLAEATAAIVM
PLAEVKGCTWNDGYTMYLGFAPGAEMFLETFEFFPLVIDMHRVLKDGMDVNFMRKALRQRYG
LLTAEQWMTQKIVEVKAAFDAVGQIAWAKSGFSPAARAFLQQFGFTG
```

A.2.2   Filoviruses

A.2.2.1   EBOV VP35$_{1-340}$

```
MTTRTKGRGHTVATTQNDRMPGPELSGWISEQLMTGRIPVNDIFCDIENNPGLCYASQMQQT
KPNPKMRNSQTQTDPICNHSFEEVVQTLASLATVVQQQTIASESLEQRITSLENGLKPVYDM
AKTISSLNRVCAEMVAKYDLLVMTTGRATATAAATEAYWAEHGQPPPGPSLYEESAIRGKIE
SRDETVPQSVREAFNNLDSTTSLTEENFGKPDISAKDLRNIMYDHLPGFGTAFHQLVQVICK
LGKDSNSLDIIHAEFQASLAEGDSPQCALIQITKRVPIFQDAAPPVIHIRSRGDIPRACQKS
LRPVPPSPKIDRGWVCVFQLQDGKTLGLKI
```

A.2.2.2   EBOV VP35$_{80-340}$

```
CNHSFEEVVQTLASLATVVQQQTIASESLEQRITSLENGLKPVYDMAKTISSLNRVCAEMVA
KYDLLVMTTGRATATAAATEAYWAEHGQPPPGPSLYEESAIRGKIESRDETVPQSVREAFNN
LNSTTSLTEENFGKPDISAKDLRNIMYDHLPGFGTAFHQLVQVICKLGKDSNSLDIIHAEFQ
ASLAEGDSPQCALIQITKRVPIFQDAAPPVIHIRSRGDIPRACQKSLRPVPPSPKIDRGWVC
VFQLQDGKTLGLKI
```

A.2.2.3   MARV VP35$_{1-329}$

```
MWDSSYMQQVSEGLMTGKVPIDQVFGANPLEKLYKRRKPKGTVGLQCSPCLMSKATSTDDIV
WDQLIVKKTLADLLIPINRQISDIQSTLNEVTTRVHEIERQLHEITPVLKMGRTLEAISKGM
SEMLAKYDHLVISTGRTTAPAAAFDAYLNEHGVPPPQPAIFKDLGVAQQACSKGTMVKNETT
DAADKMSKVLELSEETFSKPNLSAKDLALLLFTHLPGNNTPFHILAQVLSKIAYKSGKSGAF
LDAFHQILSEGENAQAALTRLSRTFDAFLGVVPPVIRVKNFQTVPRPCQKSLRAVPPNPTID
KGWVCVYSSEQGETRALKI
```

A.2.2.4   MARV VP35$_{60-329}$

```
DIVWDQLIVKKTLADLLIPINRQMSDIQSTLSEMTTKVHEIERQLHDITPVVKMGKTLEAIS
KGMSEMLAKYDHLVISTGRTTAPAAAFDAYLNEHGVPPPQPAIFKDLGVAQQAYSQKTMVKN
QTTDAADKMSKVLELSEETFSKPNLSAKDLALLLFTHLPGNNTPFHILAQVLSKIAYKSGKS
```

GAFLDAFHQILSEGENAQAALTRLSRTFDAFLGAVPPVIKVKNFQTVPRPCQKSLRAVPPNP
TIDKGWVCVYSSEQGETRALKI

A.2.2.5 EBOV VP30$_{1-288}$

MEASYERGRPRAARQHSRDGHDHHVRARSSSRENYRGEYRQSRSASQVRVPTVFHKKRVEPL
TVPPAPKDICPTLKKGFLCDSSFCKKDHQLESLTDRELLLLIARKTCGSVEQQLNITAPKDS
RLANPTADDFQQEEGPKITLLTLIKTAEHWARQDIRTIEDSKLRALLTLCAVMTRKFSKSQL
SLLCETHLRREGLGQDQAEPVLEVYQRLHSDKGGSFEAALWQQWDRQSLIMFITAFLNIALQ
LPCESSAVVVSGLRTLVPQSDNEEASTNPGTCSWSDEGTP

A.2.2.6 EBOV VP30$_{8-272}$

GRPRAARQHSRDGHDHHVRARSSSRENYRGEYRQSRSASQVRVPTVFHKKRVEPLTVPPAPK
DICPTLKKGFLCDSSFCKKDHQLESLTDRELLLLIARKTCGSVEQQLNITAPKDSRLANPTA
DDFQQEEGPKITLLTLIKTAEHWARQDIRTIEDSKLRALLTLCAVMTRKFSKSQLSLLCETH
LRREGLGQDQAEPVLEVYQRLHSDKGGSFEAALWQQWDRQSLIMFITAFLNIALQLPCESSA
VVVSGLRTLVPQSDNEE

A.2.2.7 EBOV VP30$_{69-272}$

KDICPTLKKGFLCDSSFCKKDHQLESLTDRELLLLIARKTCGSVEQQLNITAPKDSRLANPT
ADDFQQEEGPKITLLTLIKTAEHWARQDIRTIEDSKLRALLTLCAVMTRKFSKSQLSLLCET
HLRREGLGQDQAEPVLEVYQRLHSDKGGSFEAALWQQWDRQSLIMFITAFLNIALQLPCESS
AVVVSGLRTLVPQSDNEE

A.2.2.8 EBOV VP30$_{87-265}$

KKDHQLESLTDRELLLLIARKTCGSVEQQLNITAPKDSRLANPTADDFQQEEGPKITLLTLI
KTAEHWARQDIRTIEDSKLRALLTLCAVMTRKFSKSQLSLLCETHLRREGLGQDQAEPVLEV
YQRLHSDKGGSFEAALWQQWDRQSLIMFITAFLNIALQLPCESSAVVVSGLRTLV

A.2.2.9 MARV VP30$_{1-281}$

MQQPRGRSRTRNHQAIPSIYHETQLPSKPNYTNHHPRARSMSSTRSSTESSPTNHIPRARPP
STFNLSKPPPPPKDMCRNMKIGLPCTDLTCNRDHDLDNLTNRELLLLMARKMLPNTDKAFKS
PQDCGSPSLSKGLSKDKQEQTKDVLTLENLGHILNYLHRSEIGKLDETSLRAALSLTCAGIR
KTNRSLINTMTELHINHENLPQDQNGVIKQTYTGIHLDKGGQFEAALWQGWDKRSISLFVQA
ALYVMNNIPCESSISVQASYDHFILPQSQGKGQ

A.2.2.10 MARV VP30$_{17-273}$

PSIYHETQLPSKPNYTNHHPRARSMSSTRSSTESSPTNHIPRARPPSTFNLSKPPPPPKDMC
RNMKIGLPCTDLTCNRDHDLDNLTNRELLLLMARKMLPNTDKAFKSPQDCGSPSLSKGLSKD
KQEQTKDVLTLENLGHILNYLHRSEIGKLDETSLRAALSLTCAGIRKTNRSLINTMTELHIN
HENLPQDQNGVIKQTYTGIHLDKGGQFEAALWQGWDKRSISLFVQAALYVMNNIPCESSISV
QASYDHFIL

A.2.2.11 MARV VP30$_{79-273}$

RNMKIGLPCTDLTCNRDHDLDNLTNRELLLLMARKMLPNTDKAFKSPQDCGSPSLSKGLSKD
KQEQTKDVLTLENLGHILNYLHRSEIGKLDETSLRAALSLTCAGIRKTNRSLINTMTELHIN
HENLPQDQNGVIKQTYTGIHLDKGGQFEAALWQGWDKRSISLFVQAALYVMNNIPCESSISV
QASYDHFIL

A.2.2.12 MARV VP30$_{101-273}$

LTNRELLLLMARKMLPNTDKAFKSPQDCGSPSLSKGLSKDKQEQTKDVLTLENLGHILNYLH
RSEIGKLDETSLRAALSLTCAGIRKTNRSLINTMTELHINHENLPQDQNGVIKQTYTGIHLD
KGGQFEAALWQGWDKRSISLFVQAALYVMNNIPCESSISVQASYDHFIL

A.2.3 hRSV

A.2.3.1 M2-1$_{1-194}$

MSRRNPCKFEIRGHCLNGKRCHFSHNYFEWPPHALLVRQNFMLNRILKSMDKSIDTLSEISG
AAELDRTEEYALGVVGVLESYIGSINNITKQSACVAMSKLLTELNSDDIKKLRDNEELNSPK
IRVYNTVISYIESNRKNNKQTIHLLKRLPADVLKKTIKNTLDIHKSITINNPKESTVSDTND
HAKNNDTT

A.2.3.2 P$_{1-241}$

MEKFAPEFHGEDANNRATKFLESIKGKFTSPKDPKKKDSIISVNSIDIEVTKESPITSNSTI
INPTNETDDTAGNKPNYQRKPLVSFKEDPTPSDNPFSKLYKETIETFDNNEEESSYSYEEIN
DQTNDNITARLDRIDEKLSEILGMLHTLVVASAGPTSARDGIRDAMIGLREEMIEKIRTEAL
MTNDRLEAMARLRNEESEKMAKDTSDEVSLNPTSEKLNNLLEGNDSDNDLSLEDF

A.2.3.3 P$_{90-160}$

DPTPSDNPFSKLYKETIETFDNNEEESSYSYEEINDQTNDNITARLDRIDEKLSEILGMLHT
LVVASAGPT

A.2.3.4 M$_{1-256}$

METYVNKLHEGSTYTAAVQYNVLEKDDDPASLTIWVPMFQSSMPADLLIKELANVNILVKQI
STPKGPSLRVMINSRSAVLAQMPSKFTICANVSLDERSKLAYDVTTPCEIKACSLTCLKSKN
MLTTVKDLTMKTLNPTHDIIALCEFENIVTSKKVIIPTYLRSISVRNKDLNTLENITTTEFK
NAITNAKIIPYSGLLLVITVTDNKGAFKYIKPQSQFIVDLGAYLEKESIYYVTTNWKHTATR
FAIKPMED

A.3　　Modified Vector Maps

A.3.1　　pGEX-6P-2


```
              PreScission™ Protease
┌──────────────────────────────────────────────┐        ┌
Leu Glu Val Leu Phe Gln↓Gly Pro Leu Gly Ser Pro Gly Ile Pro Gly Ser Thr Arg Ala Ala Ala Ser
CTG GAA GTT CTG TTC CAG GGG CCC CTG GGA TCC CCA GGA ATT CCC GGG TCG ACT CGA GCG GCC GCA TCG
                              BamH I        EcoR I  Sma I   Sal I   Xho I    Not I
```
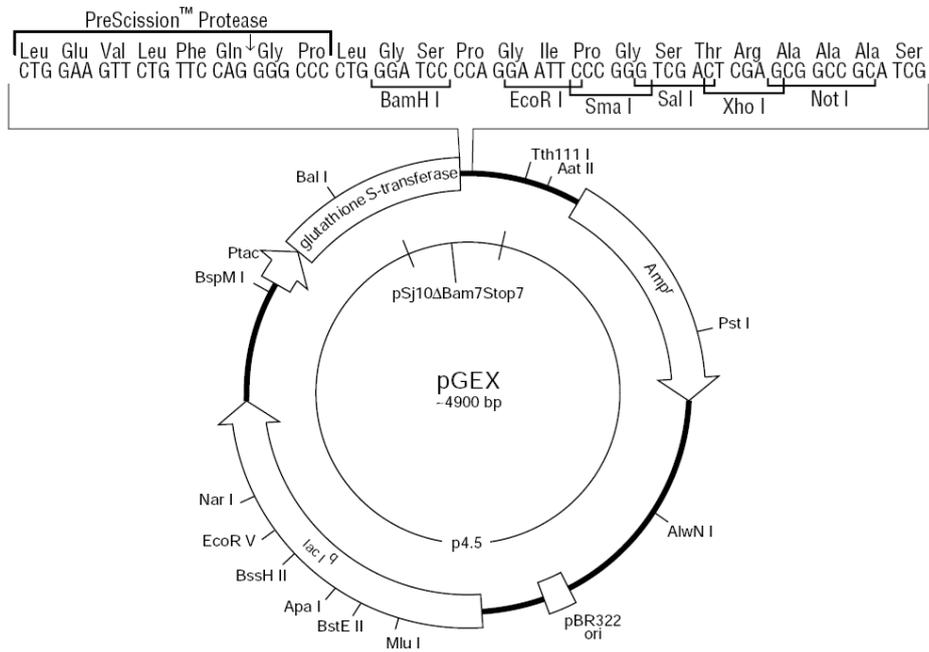
A.3.2　　pET-28a-SUMO
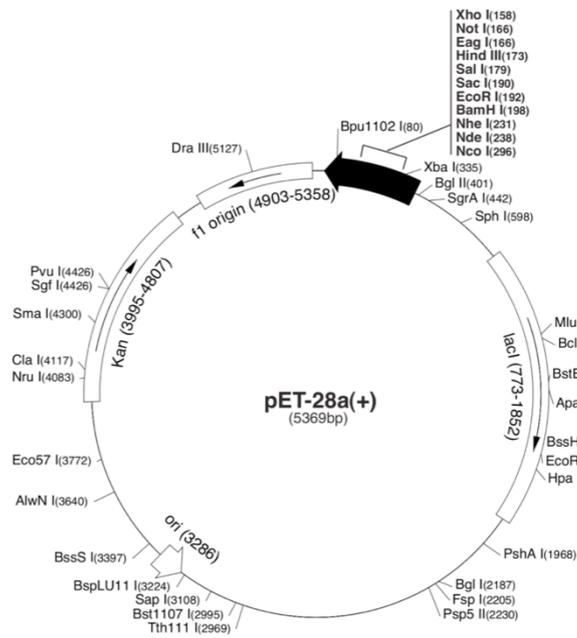
Bibliography

Adams, M. J. *et al.* (2017) 'Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2017)', *Archives of Virology*, 162(8), pp. 2505–2538. doi: 10.1007/s00705-017-3358-5.

Adu-Gyamfi, E. *et al.* (2014) 'A loop region in the N-terminal domain of Ebola virus VP40 is important in viral assembly, budding, and egress.', *Viruses*. Multidisciplinary Digital Publishing Institute (MDPI), 6(10), pp. 3837–54. doi: 10.3390/v6103837.

Adu-Gyamfi, E. *et al.* (2013) 'The Ebola Virus Matrix Protein Penetrates into the Plasma Membrane'. doi: 10.1074/jbc.M112.443960.

Agilent Technologies (2012) *ArcticExpress Competent Cells and ArcticExpress (DE3) Competent Cells Instruction Manual For Research Use Only.* Available at: www.genomics.agilent.com (Accessed: 14 February 2021).

Ahmadian, G., Randhawa, J. S. and Easton, A. J. (2000) 'Expression of the ORF-2 protein of the human respiratory syncytial virus M2 gene is initiated by a ribosomal termination-dependent reinitiation mechanism.', *The EMBO journal*, 19(11), pp. 2681–9. doi: 10.1093/emboj/19.11.2681.

Albariño, C. G. *et al.* (2018) 'Transcriptional analysis of viral mRNAs reveals common transcription patterns in cells infected by five different filoviruses', *PLOS ONE*. Edited by A. N. Freiberg. Public Library of Science, 13(8), p. e0201827. doi: 10.1371/journal.pone.0201827.

Amarasinghe, G. K. *et al.* (2019) 'Taxonomy of the order Mononegavirales: update 2019', *Archives of Virology*. Springer, 164(7), pp. 1967–1980. doi: 10.1007/s00705-019-04247-4.

Anderson, L. J. *et al.* (2013) 'Strategic priorities for respiratory syncytial virus (RSV) vaccine development', *Vaccine*, 31, pp. B209–B215. doi: 10.1016/j.vaccine.2012.11.106.

Ariza-Heredia, E. J. *et al.* (2012) 'Clinical and radiological features of respiratory syncytial virus in solid organ transplant recipients: a single-center experience.', *Transplant infectious disease : an official journal of the Transplantation Society*, 14(1), pp. 64–71. doi: 10.1111/j.1399-3062.2011.00673.x.

Ariza, A. *et al.* (2013) 'Nucleocapsid protein structures from orthobunyaviruses reveal insight into ribonucleoprotein architecture and RNA polymerization.', *Nucleic acids research*. Oxford University Press, 41(11), pp. 5912–26. doi: 10.1093/nar/gkt268.

Asenjo, A., González-Armas, J. C. and Villanueva, N. (2008) 'Phosphorylation of human respiratory syncytial virus P protein at serine 54 regulates viral uncoating', *Virology*,

380(1), pp. 26–33. doi: 10.1016/j.virol.2008.06.045.

Bai, C.-Q. *et al.* (2016) 'Clinical and Virological Characteristics of Ebola Virus Disease Patients Treated With Favipiravir (T-705)—Sierra Leone, 2014', *Clinical Infectious Diseases*. Narnia, 63(10), pp. 1288–1294. doi: 10.1093/cid/ciw571.

Bailly, B. *et al.* (2016) 'Targeting human respiratory syncytial virus transcription anti-termination factor M2-1 to inhibit in vivo viral replication', *Scientific Reports*. Nature Publishing Group, 6(1), p. 25806. doi: 10.1038/srep25806.

Baker, L. A. and Rubinstein, J. L. (2010) 'Radiation Damage in Electron Cryomicroscopy', in *Methods in enzymology*, pp. 371–388. doi: 10.1016/S0076-6879(10)81015-8.

Baker, L. E. *et al.* (2016) 'Molecular architecture of the nucleoprotein C-terminal domain from the Ebola and Marburg viruses', *Acta Crystallographica Section D Structural Biology*, 72(1), pp. 49–58. doi: 10.1107/S2059798315021439.

Baker, M. (2010) 'Mass spectrometry for biologists', *Nature Methods*. Nature Publishing Group, 7(2), pp. 157–161. doi: 10.1038/nmeth0210-157.

Bakker, S. E. *et al.* (2013) 'The respiratory syncytial virus nucleoprotein-RNA complex forms a left-handed helical nucleocapsid.', *The Journal of general virology*. Microbiology Society, 94(Pt 8), pp. 1734–8. doi: 10.1099/vir.0.053025-0.

Balmer, S. *et al.* (2014) 'Serosurveillance of Schmallenberg virus in Switzerland using bulk tank milk samples.', *Preventive veterinary medicine*. Prev Vet Med, 116(4), pp. 370–9. doi: 10.1016/j.prevetmed.2014.03.026.

Banerjee, A. and Mitra, P. (2020) 'Ebola Virus VP35 Protein: Modeling of the Tetrameric Structure and an Analysis of Its Interaction with Human PKR', *Cite This: J. Proteome Res*, 19, p. 4542. doi: 10.1021/acs.jproteome.0c00473.

Barik, S., McLean, T. and Dupuy, L. C. (1995) 'Phosphorylation of Ser232 directly regulates the transcriptional activity of the P protein of human respiratory syncytial virus: phosphorylation of Ser237 may play an accessory role.', *Virology*, 213(2), pp. 405–12. doi: 10.1006/viro.1995.0013.

Barr, J. N., Rodgers, J. W. and Wertz, G. W. (2005) 'The Bunyamwera Virus mRNA Transcription Signal Resides within both the 3′ and the 5′ Terminal Regions and Allows Ambisense Transcription from a Model RNA Segment', *Journal of Virology*. American Society for Microbiology (ASM), 79(19), p. 12602. doi: 10.1128/JVI.79.19.12602-12607.2005.

Barr, J. N., Rodgers, J. W. and Wertz, G. W. (2006) 'Identification of the Bunyamwera bunyavirus transcription termination signal', *Journal of General Virology*, 87(1), pp. 189–198. doi: 10.1099/vir.0.81355-0.

Barr, J. N. and Wertz, G. W. (2004) 'Bunyamwera bunyavirus RNA synthesis requires cooperation of 3'- and 5'-terminal sequences.', *Journal of virology*, 78(3), pp. 1129–38. doi: 10.1128/jvi.78.3.1129-1138.2004.

Barr, J. N. and Wertz, G. W. (2005) 'Role of the conserved nucleotide mismatch within 3'- and 5'-terminal regions of Bunyamwera virus in signaling transcription.', *Journal of virology*. American Society for Microbiology (ASM), 79(6), pp. 3586–94. doi: 10.1128/JVI.79.6.3586-3594.2005.

Barr, J. N., Whelan, S. P. and Wertz, G. W. (1997) 'cis-Acting signals involved in termination of vesicular stomatitis virus mRNA synthesis include the conserved AUAC and the U7 signal for polyadenylation.', *Journal of virology*, 71(11), pp. 8718–25. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=192336&tool=pmcentrez&rendertype=abstract (Accessed: 19 May 2016).

Barry, G. *et al.* (2014) 'NSs protein of Schmallenberg virus counteracts the antiviral response of the cell by inhibiting its transcriptional machinery', *Journal of General Virology*, 95(Pt_8), pp. 1640–1646. doi: 10.1099/vir.0.065425-0.

Battisti, A. J. *et al.* (2011) 'Structural Studies of Hantaan Virus', *Journal of Virology*, 85(2), pp. 835–841. doi: 10.1128/JVI.01847-10.

Becker, S. *et al.* (1998) 'Interactions of Marburg virus nucleocapsid proteins.', *Virology*, 249(2), pp. 406–17. doi: 10.1006/viro.1998.9328.

Beniac, D. R. *et al.* (2012) 'The organisation of Ebola virus reveals a capacity for extensive, modular polyploidy.', *PloS one*. Public Library of Science, 7(1), p. e29608. doi: 10.1371/journal.pone.0029608.

Bernstein, D. I. *et al.* (2012) 'Phase 1 study of the safety and immunogenicity of a live, attenuated respiratory syncytial virus and parainfluenza virus type 3 vaccine in seronegative children.', *The Pediatric infectious disease journal*, 31(2), pp. 109–14. doi: 10.1097/INF.0b013e31823386f1.

Bharat, T. A. M. *et al.* (2011) 'Cryo-Electron Tomography of Marburg Virus Particles and Their Morphogenesis within Infected Cells', *PLoS Biol*, 9(11). doi: 10.1371/journal.pbio.1001196.

Bharat, T. A. M. *et al.* (2012) 'Structural dissection of Ebola virus and its assembly determinants using cryo-electron tomography.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 109(11), pp. 4275–80. doi: 10.1073/pnas.1120453109.

Biedenkopf, N. *et al.* (2016) 'RNA Binding of Ebola Virus VP30 Is Essential for Activating Viral Transcription.', *Journal of virology*. American Society for Microbiology, 90(16), pp. 7481–96. doi: 10.1128/JVI.00271-16.

Biedenkopf, N., Lier, C. and Becker, S. (2016) 'Dynamic Phosphorylation of VP30 Is Essential for Ebola Virus Life Cycle.', *Journal of virology*. American Society for Microbiology, 90(10), pp. 4914–25. doi: 10.1128/JVI.03257-15.

Bishop, D. H. L. *et al.* (1980) 'Bunyaviridae', *Intervirology*, 14(3–4), pp. 125–143. doi: 10.1159/000149174.

Blakqori, G., van Knippenberg, I. and Elliott, R. M. (2009) 'Bunyamwera orthobunyavirus S-segment untranslated regions mediate poly(A) tail-independent translation.', *Journal of virology*. American Society for Microbiology Journals, 83(8), pp. 3637–46. doi: 10.1128/JVI.02201-08.

Blanco, J. C. G. *et al.* (2010) 'New insights for development of a safe and protective RSV vaccine.', *Human vaccines*, 6(6), pp. 482–92. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2965816&tool=pmcentrez&rendertype=abstract (Accessed: 21 April 2016).

Blondot, M.-L. *et al.* (2012) 'Structure and functional analysis of the RNA- and viral phosphoprotein-binding domain of respiratory syncytial virus M2-1 protein.', *PLoS pathogens*, 8(5), p. e1002734. doi: 10.1371/journal.ppat.1002734.

Bornholdt, Z. A. *et al.* (2013) 'Structural Rearrangement of Ebola Virus VP40 Begets Multiple Functions in the Virus Life Cycle', *Cell*. Cell Press, 154(4), pp. 763–774. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0092867413008866 (Accessed: 6 January 2017).

Bowden, T. A. *et al.* (2013) 'Orthobunyavirus Ultrastructure and the Curious Tripodal Glycoprotein Spike', *PLoS Pathogens*. Edited by T. C. Pierson. Public Library of Science, 9(5), p. e1003374. doi: 10.1371/journal.ppat.1003374.

Brackney, D. E. *et al.* (2010) 'C6/36 Aedes albopictus Cells Have a Dysfunctional Antiviral RNA Interference Response', *PLoS Neglected Tropical Diseases*. Edited by S. L. O'Neill, 4(10), p. e856. doi: 10.1371/journal.pntd.0000856.

Bragg, W. H. and Bragg, W. L. (1913) 'The Reflection o f X-rays by Crystals', *Proc. R. Soc. Lond*, A(88), pp. 428–438. doi: 10.1098/rspa.1913.0040.

Brauburger, K. *et al.* (2015) 'Transcriptional Regulation in Ebola Virus: Effects of Gene Border Structure and Regulatory Elements on Gene Expression and Polymerase Scanning Behavior'. doi: 10.1128/JVI.02341-15.

Briese, T. *et al.* (2006) 'Batai and Ngari viruses: M segment reassortment and association with severe febrile disease outbreaks in East Africa.', *Journal of virology*. American Society for Microbiology, 80(11), pp. 5627–30. doi: 10.1128/JVI.02448-05.

Brook, C. E. *et al.* (2020) 'Accelerated viral dynamics in bat cell lines, with implications for zoonotic emergence', *eLife*, 9. doi: 10.7554/eLife.48401.

Bruhn, J. F. *et al.* (2017) 'Crystal Structure of the Marburg Virus VP35 Oligomerization Domain.', *Journal of virology*. American Society for Microbiology, 91(2), pp. e01085-16. doi: 10.1128/JVI.01085-16.

Calain, P., Monroe, M. C. and Nichol, S. T. (1999) 'Ebola Virus Defective Interfering Particles and Persistent Infection', *Virology*, 262(1), pp. 114–128. doi: 10.1006/viro.1999.9915.

Calisher, C. H. (1996) 'History, Classification, and Taxonomy of Viruses in the Family Bunyaviridae', in *The Bunyaviridae*. MA: Springer US, pp. 1–17. doi: 10.1007/978-1-4899-1364-7_1.

Campbell, M. G. *et al.* (2014) 'Near-atomic resolution reconstructions using a mid-range electron microscope operated at 200kV', *Journal of Structural Biology*, 188(2), pp. 183–187. doi: 10.1016/j.jsb.2014.09.008.

Cárdenas, W. B. *et al.* (2006) 'Ebola virus VP35 protein binds double-stranded RNA and inhibits alpha/beta interferon production induced by RIG-I signaling.', *Journal of virology*. American Society for Microbiology Journals, 80(11), pp. 5168–78. doi: 10.1128/JVI.02199-05.

Cartee, T. L. and Wertz, G. W. (2001) 'Respiratory Syncytial Virus M2-1 Protein Requires Phosphorylation for Efficient Function and Binds Viral RNA during Infection', *Journal of Virology*, 75(24), pp. 12188–12197. doi: 10.1128/JVI.75.24.12188-12197.2001.

Cautain, B. *et al.* (2015) 'Components and regulation of nuclear transport processes', *FEBS Journal*, 282(3), pp. 445–462. doi: 10.1111/febs.13163.

CDC (2019) *Outbreak Table | Marburg Hemorrhagic Fever | CDC*. Available at: https://www.cdc.gov/vhf/marburg/resources/outbreak-table.html (Accessed: 15 September 2020).

Chanock, R. and Finberg, L. (1957) 'Recovery from infants with respiratory illness of a virus related to chimpanzee coryza agent (CCA). II. Epidemiologic aspects of infection in infants and young children.', *American journal of hygiene*, 66(3), pp. 291–300. Available at: http://www.ncbi.nlm.nih.gov/pubmed/13478579 (Accessed: 6 June 2019).

Chanock, R., Roizman, B. and Myers, R. (1957) 'Recovery from infants with respiratory illness of a virus related to chimpanzee coryza agent (CCA). I. Isolation, properties and characterization.', *American journal of hygiene*, 66(3), pp. 281–90. Available at: http://www.ncbi.nlm.nih.gov/pubmed/13478578 (Accessed: 6 June 2019).

Cheng, E. and Mir, M. A. (2012) 'Signatures of host mRNA 5' terminus for efficient hantavirus cap snatching.', *Journal of virology*. American Society for Microbiology Journals, 86(18), pp. 10173–85. doi: 10.1128/JVI.05560-11.

Cheng, Y. *et al.* (2015) 'A primer to single-particle cryo-electron microscopy.', *Cell*. Cell, 161(3), pp. 438–449. doi: 10.1016/j.cell.2015.03.050.

Chernov, A. A. (1997) 'Crystals built of biological macromolecules', *Physics Reports*. North-Holland, 288(1–6), pp. 61–75. doi: 10.1016/S0370-1573(97)00021-5.

Cho, A. *et al.* (2012) 'Synthesis and antiviral activity of a series of 1'-substituted 4-aza-7,9-dideazaadenosine C-nucleosides.', *Bioorganic & medicinal chemistry letters*. Elsevier, 22(8), pp. 2705–7. doi: 10.1016/j.bmcl.2012.02.105.

Chung, D.-H. *et al.* (2013) 'A cell based high-throughput screening approach for the discovery of new inhibitors of respiratory syncytial virus.', *Virology journal*. BioMed Central, 10(1), p. 19. doi: 10.1186/1743-422X-10-19.

Collins, P. L. and Graham, B. S. (2008) 'Viral and host factors in human respiratory syncytial virus pathogenesis.', *Journal of virology*, 82(5), pp. 2040–55. doi: 10.1128/JVI.01625-07.

Corti, D. *et al.* (2013) 'Cross-neutralization of four paramyxoviruses by a human monoclonal antibody', *Nature*, 501(7467), pp. 439–443. doi: 10.1038/nature12442.

Costa, S. *et al.* (2014) 'Fusion tags for protein solubility, purification, and immunogenicity in Escherichia coli: the novel Fh8 system'. doi: 10.3389/fmicb.2014.00063.

Coupeau, D. *et al.* (2013) 'Characterization of messenger RNA termini in Schmallenberg virus and related Simbuviruses', *Journal of General Virology*, 94(Pt_11), pp. 2399–2405. doi: 10.1099/vir.0.055954-0.

Cuesta, I. *et al.* (2000) 'Structural Phosphoprotein M2-1 of the Human Respiratory Syncytial Virus Is an RNA Binding Protein', *Journal of Virology*, 74(21), pp. 9858–9867. doi: 10.1128/JVI.74.21.9858-9867.2000.

Cunningham, C. and Szilágyi, J. F. (1987) 'Viral RNAs synthesized in cells infected with germiston bunyavirus', *Virology*. Academic Press, 157(2), pp. 431–439. doi: 10.1016/0042-6822(87)90285-6.

Curran, J. and Kolakofsky, D. (2008) 'Nonsegmented negative-strand RNA virus RNA synthesis in vivo', *Virology*, 371(2), pp. 227–230. doi: 10.1016/j.virol.2007.11.022.

Datta, K. *et al.* (2013) 'Characterization of PA-N terminal domain of Influenza A polymerase reveals sequence specific RNA cleavage.', *Nucleic acids research*. Oxford University Press, 41(17), pp. 8289–99. doi: 10.1093/nar/gkt603.

Davey, R. T. *et al.* (2016) 'A Randomized, Controlled Trial of ZMapp for Ebola Virus Infection', *New England Journal of Medicine*. Massachusetts Medical Society, 375(15), pp. 1448–1456. doi: 10.1056/NEJMoa1604330.

Delageniere, S. *et al.* (2011) 'ISPyB: an information management system for synchrotron macromolecular crystallography', *Bioinformatics*. Oxford Academic, 27(22), pp. 3186–3192. doi: 10.1093/bioinformatics/btr535.

Dessau, M. A. and Modis, Y. (2011) 'Protein Crystallization for X-ray Crystallography', *Journal of Visualized Experiments : JoVE*. MyJoVE Corporation, (47). doi: 10.3791/2285.

Devignot, S. *et al.* (2015) 'A Virus-Like Particle System Identifies the Endonuclease Domain of Crimean-Congo Hemorrhagic Fever Virus', *Journal of Virology*. Edited by R. W. Doms, 89(11), pp. 5957–5967. doi: 10.1128/JVI.03691-14.

DeVincenzo, J. P. *et al.* (2014) 'Oral GS-5806 Activity in a Respiratory Syncytial Virus Challenge Study', *New England Journal of Medicine*. Massachusetts Medical Society , 371(8), pp. 711–722. doi: 10.1056/NEJMoa1401184.

DiCarlo, A. *et al.* (2011) 'Phosphorylation of Marburg virus NP region II modulates viral RNA synthesis.', *The Journal of infectious diseases*, 204 Suppl 3(suppl_3), pp. S927-33. doi: 10.1093/infdis/jir319.

Dixon, M. P. *et al.* (2002) 'The central domain of Escherichia coli TyrR is responsible for hexamerization associated with tyrosine-mediated repression of gene expression.', *The Journal of biological chemistry*. J Biol Chem, 277(26), pp. 23186–92. doi: 10.1074/jbc.M112184200.

Dolnik, O. *et al.* (2015) 'Marburg virus inclusions: A virus-induced microcompartment and interface to multivesicular bodies and the late endosomal compartment', *European Journal of Cell Biology*. Urban & Fischer, 94(7–9), pp. 323–331. doi: 10.1016/J.EJCB.2015.05.006.

Domachowske, J. B. and Rosenberg, H. F. (1999) *Respiratory Syncytial Virus Infection: Immune Response, Immunopathogenesis, and Treatment*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC88919/pdf/cm000298.pdf (Accessed: 17 December 2019).

Duffy, S., Shackelton, L. A. and Holmes, E. C. (2008) 'Rates of evolutionary change in viruses: patterns and determinants', *Nature Reviews Genetics*. Nature Publishing Group, 9(4), pp. 267–276. doi: 10.1038/nrg2323.

Duijsings, D., Kormelink, R. and Goldbach, R. (2001) 'In vivo analysis of the TSWV cap-snatching mechanism: single base complementarity and primer length requirements.', *The EMBO journal*. European Molecular Biology Organization, 20(10), pp. 2545–52. doi: 10.1093/emboj/20.10.2545.

Dziuba, P. J. *et al.* (2014) 'The structure of the C-terminal domain of the Zaire ebolavirus nucleoprotein', *Acta Cryst*, 70, pp. 2420–2429. doi: 10.1107/S1399004714014710.

Dziubańska, P. J. *et al.* (2014) 'The structure of the C-terminal domain of the *Zaire ebolavirus* nucleoprotein', *Acta Crystallographica Section D Biological Crystallography*. International Union of Crystallography, 70(9), pp. 2420–2429. doi: 10.1107/S1399004714014710.

Ebbert, J. O. and Limper, A. H. (2005) 'Respiratory syncytial virus pneumonitis in immunocompromised adults: clinical features and outcome.', *Respiration; international review of thoracic diseases*, 72(3), pp. 263–9. doi: 10.1159/000085367.

Eifan, S. A. and Elliott, R. M. (2009) 'Mutational analysis of the Bunyamwera orthobunyavirus nucleocapsid protein gene.', *Journal of virology*. American Society for Microbiology (ASM), 83(21), pp. 11307–17. doi: 10.1128/JVI.01460-09.

Elliott, R. M. (1989) 'Nucleotide sequence analysis of the large (L) genomic RNA segment of Bunyamwera virus, the prototype of the family Bunyaviridae.', *Virology*, 173(2), pp. 426–36. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2596023 (Accessed: 8 July 2019).

Elliott, R. M. (2014) 'Orthobunyaviruses: recent genetic and structural insights', *Nature Reviews Microbiology*, 12(10), pp. 673–685. doi: 10.1038/nrmicro3332.

Emsley, P. and Cowtan, K. (2004) 'Coot: model-building tools for molecular graphics.', *Acta crystallographica. Section D, Biological crystallography*. Acta Crystallogr D Biol Crystallogr, 60(Pt 12 Pt 1), pp. 2126–32. doi: 10.1107/S0907444904019158.

Eshita, Y. *et al.* (1985) 'Analyses of the mRNA transcription processes of snowshoe hare bunyavirus S and M RNA species.', *Journal of virology*, 55(3), pp. 681–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/4020962 (Accessed: 9 July 2019).

Evans, D. K., Goldstein, M. and Popova, A. (2015) 'Health-care worker mortality and the legacy of the Ebola epidemic', *The Lancet Global Health*, 3(8), pp. e439–e440. doi: 10.1016/S2214-109X(15)00065-0.

Evans, P. (2006) 'Scaling and assessment of data quality.', *Acta crystallographica. Section D, Biological crystallography*. Acta Crystallogr D Biol Crystallogr, 62(Pt 1), pp. 72–82. doi: 10.1107/S0907444905036693.

Falasca, L. *et al.* (2015) 'Molecular mechanisms of Ebola virus pathogenesis: focus on cell death', *Cell Death and Differentiation*, 22, pp. 1250–1259. doi: 10.1038/cdd.2015.67.

Fazakerley, J. K. and Ross, A. M. (1989) 'Computer analysis suggests a role for signal sequences in processing polyproteins of enveloped RNA viruses and as a mechanism of viral fusion.', *Virus genes*, 2(3), pp. 223–39. doi: 10.1007/BF00125340.

FDA (2019) *Ervebo*. Available at: https://www.fda.gov/vaccines-blood-biologics/ervebo (Accessed: 27 January 2021).

FDA (2020) *First FDA-approved vaccine for the prevention of Ebola virus disease, marking a critical milestone in public health preparedness and response | FDA*. Available at: https://www.fda.gov/news-events/press-announcements/first-fda-approved-vaccine-prevention-ebola-virus-disease-marking-critical-milestone-public-health (Accessed: 27 January 2021).

Fearns, R. (2013) 'Respiratory Syncytial Virus Polymerase Protein and Promoter Interactions'. Available at: http://grantome.com/grant/NIH/R01-AI074903-04 (Accessed: 21 April 2016).

Fearns, R. and Collins, P. L. (1999) 'Role of the M2-1 transcription antitermination protein of respiratory syncytial virus in sequential transcription.', *Journal of virology*, 73(7), pp. 5852–64. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10364337 (Accessed: 5 June 2019).

Fearns, R., Peeples, M. E. and Collins, P. L. (2002) 'Mapping the transcription and replication promoters of respiratory syncytial virus.', *Journal of virology*, 76(4), pp. 1663–72. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=135899&tool=pmcentrez&rendertype=abstract (Accessed: 19 May 2016).

Feldmann, H. and Geisbert, T. W. (2011) 'Ebola haemorrhagic fever.', *Lancet (London, England)*. NIH Public Access, 377(9768), pp. 849–62. doi: 10.1016/S0140-6736(10)60667-8.

Fernández-García, Y. *et al.* (2016) 'Atomic Structure and Biochemical Characterization of an RNA Endonuclease in the N Terminus of Andes Virus L Protein', *PLOS Pathogens*. Edited by F. A. Rey. Public Library of Science, 12(6), p. e1005635. doi: 10.1371/journal.ppat.1005635.

Ferron, F. *et al.* (2011) 'The Hexamer Structure of the Rift Valley Fever Virus Nucleoprotein Suggests a Mechanism for its Assembly into Ribonucleoprotein Complexes', *PLoS Pathogens*. Edited by F. A. Rey. Public Library of Science, 7(5), p. e1002030. doi: 10.1371/journal.ppat.1002030.

Filone, C. M. *et al.* (2006) 'Development and characterization of a Rift Valley fever virus cell–cell fusion assay using alphavirus replicon vectors', *Virology*, 356(1–2), pp. 155–164. doi: 10.1016/j.virol.2006.07.035.

Fontana, J. *et al.* (2008) 'The unique architecture of Bunyamwera virus factories around the Golgi complex', *Cellular Microbiology*, 10(10), pp. 2012–2028. doi: 10.1111/j.1462-5822.2008.01184.x.

Förster, A. *et al.* (2015) 'Dimerization of Matrix Protein Is Required for Budding of Respiratory Syncytial Virus', *Journal of Virology*, 89(8), pp. 4624–4635. doi: 10.1128/JVI.03500-14.

Fowler, T. *et al.* (2005) 'Inhibition of Marburg virus protein expression and viral release

by RNA interference.', *The Journal of general virology*, 86(Pt 4), pp. 1181–8. doi: 10.1099/vir.0.80622-0.

Fuentes, S. *et al.* (2007) 'Function of the Respiratory Syncytial Virus Small Hydrophobic Protein', *Journal of Virology*, 81(15), pp. 8361–8366. doi: 10.1128/JVI.02717-06.

Galloux, M. *et al.* (2015) 'Identification and Characterization of the Binding Site of the Respiratory Syncytial Virus Phosphoprotein to RNA-Free Nucleoprotein', *Journal of Virology*. Edited by R. M. Sandri-Goldin, 89(7), pp. 3484–3496. doi: 10.1128/JVI.03666-14.

Gan, S.-W. *et al.* (2012) 'The Small Hydrophobic Protein of the Human Respiratory Syncytial Virus Forms Pentameric Ion Channels', *Journal of Biological Chemistry*, 287(29), pp. 24671–24689. doi: 10.1074/jbc.M111.332791.

Gao, Y. *et al.* (2020) 'Structure of the Human Respiratory Syncytial Virus M2-1 Protein in Complex with a Short Positive-Sense Gene-End RNA', *Structure*. Cell Press, 28(9), pp. 979-990.e4. doi: 10.1016/J.STR.2020.07.001.

Garcin, D. *et al.* (1995) 'The 5' ends of Hantaan virus (Bunyaviridae) RNAs suggest a prime-and-realign mechanism for the initiation of RNA synthesis.', *Journal of virology*. American Society for Microbiology (ASM), 69(9), pp. 5754–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7637020 (Accessed: 9 July 2019).

Gasteiger, E. *et al. Protein Analysis Tools on the ExPASy Server 571 571 From: The Proteomics Protocols Handbook Protein Identification and Analysis Tools on the ExPASy Server*. Available at: http://www.expasy.org/tools/. (Accessed: 4 January 2021).

Gaunt, E. R. *et al.* (2011) 'Molecular epidemiology and evolution of human respiratory syncytial virus and human metapneumovirus.', *PloS one*. Public Library of Science, 6(3), p. e17427. doi: 10.1371/journal.pone.0017427.

GC, J. B., Gerstman, B. S. and Chapagain, P. P. (2017) 'Membrane association and localization dynamics of the Ebola virus matrix protein VP40', *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1859(10), pp. 2012–2020. doi: 10.1016/j.bbamem.2017.07.007.

Gerlach, P. *et al.* (2015) 'Structural Insights into Bunyavirus Replication and Its Regulation by the vRNA Promoter', *Cell*, 161(6), pp. 1267–1279. doi: 10.1016/j.cell.2015.05.006.

Ghildyal, R. *et al.* (2005) 'Nuclear import of the respiratory syncytial virus matrix protein is mediated by importin beta1 independent of importin alpha.', *Biochemistry*, 44(38), pp. 12887–95. doi: 10.1021/bi050701e.

Gilman, M. S. A. *et al.* (2019) 'Structure of the Respiratory Syncytial Virus Polymerase Complex', *Cell*. Cell Press, 179(1), pp. 193-204.e14. Available at:

https://linkinghub.elsevier.com/retrieve/pii/S009286741930902X (Accessed: 14 October 2019).

Glenn, G. M. *et al.* (2016) 'A Randomized, Blinded, Controlled, Dose-Ranging Study of a Respiratory Syncytial Virus Recombinant Fusion (F) Nanoparticle Vaccine in Healthy Women of Childbearing Age', *Journal of Infectious Diseases*, 213(3), pp. 411–422. doi: 10.1093/infdis/jiv406.

Glezen, W. P. *et al.* (1986) 'Risk of primary infection and reinfection with respiratory syncytial virus.', *American journal of diseases of children (1960)*, 140(6), pp. 543–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3706232 (Accessed: 4 June 2019).

Goeijenbier, M. *et al.* (2014) 'Ebola virus disease: a review on epidemiology, symptoms, treatment and pathogenesis.', *The Netherlands journal of medicine*, 72(9), pp. 442–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25387613 (Accessed: 29 December 2016).

Gonzalez-Scarano, F. *et al.* (1985) 'An avirulent G1 glycoprotein variant of La Crosse bunyavirus with defective fusion function.', *Journal of virology*. American Society for Microbiology (ASM), 54(3), pp. 757–63. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3889368 (Accessed: 9 July 2019).

Gonzalez-Scarano, F., Pobjecky, N. and Nathanson, N. (1984) 'La Crosse bunyavirus can mediate pH-dependent fusion from without.', *Virology*, 132(1), pp. 222–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6695500 (Accessed: 9 July 2019).

Gould, P. S. and Easton, A. J. (2005) 'Coupled translation of the respiratory syncytial virus M2 open reading frames requires upstream sequences.', *The Journal of biological chemistry*, 280(23), pp. 21972–80. doi: 10.1074/jbc.M502276200.

Gouzil, J. *et al.* (2017) 'Nonstructural Protein NSs of Schmallenberg Virus Is Targeted to the Nucleolus and Induces Nucleolar Disorganization.', *Journal of virology*. American Society for Microbiology Journals, 91(1), pp. e01263-16. doi: 10.1128/JVI.01263-16.

Greenfield, N. J. (2006) 'Using circular dichroism spectra to estimate protein secondary structure.', *Nature protocols*. NIH Public Access, 1(6), pp. 2876–90. doi: 10.1038/nprot.2006.202.

Gregory, S. M. *et al.* (2011) 'Structure and function of the complete internal fusion loop from Ebolavirus glycoprotein 2.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 108(27), pp. 11211–6. doi: 10.1073/pnas.1104760108.

Gregory, S. M. *et al.* (2014) 'Ebolavirus entry requires a compact hydrophobic fist at the tip of the fusion loop.', *Journal of virology*. American Society for Microbiology, 88(12), pp. 6636–49. doi: 10.1128/JVI.00396-14.

Grosse-Kunstleve, R. W. *et al.* (2002) 'The *Computational Crystallography Toolbox* :

crystallographic algorithms in a reusable software framework', *Journal of Applied Crystallography*. International Union of Crystallography, 35(1), pp. 126–136. doi: 10.1107/S0021889801017824.

Gutiérrez-Ortega, A., Sánchez-Hernández, C. and Gómez-García, B. (2008) 'Respiratory syncytial virus glycoproteins uptake occurs through clathrin-mediated endocytosis in a human epithelial cell line.', *Virology journal*. BioMed Central, 5(1), p. 127. doi: 10.1186/1743-422X-5-127.

Hardy, R. W. and Wertz, G. W. (2000) 'The Cys3-His1 Motif of the Respiratory Syncytial Virus M2-1 Protein Is Essential for Protein Function', *Journal of Virology*, 74(13), pp. 5880–5885. doi: 10.1128/JVI.74.13.5880-5885.2000.

Harris, K. A. *et al.* (2014) 'Impact of Schmallenberg virus on British sheep farms during the 2011/2012 lambing season.', *The Veterinary record*. Vet Rec, 175(7), p. 172. doi: 10.1136/vr.102295.

Hartlieb, B. *et al.* (2003) 'Oligomerization of Ebola Virus VP30 Is Essential for Viral Transcription and Can Be Inhibited by a Synthetic Peptide* Downloaded from', *THE Journal of Biological Chemistry*, 278(43), pp. 41830–41836. doi: 10.1074/jbc.M307036200.

Hartlieb, B. *et al.* (2007) 'Crystal structure of the C-terminal domain of Ebola virus VP30 reveals a role in transcription and nucleocapsid association', *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 104(2), pp. 624–629. doi: 10.1073/pnas.0606730104.

He, S. and Scheres, S. H. W. (2017) 'Helical reconstruction in RELION', *Journal of Structural Biology*. Academic Press, 198(3), pp. 163–176. doi: 10.1016/J.JSB.2017.02.003.

Hepojoki, J. *et al.* (2012) 'Hantavirus structure - molecular interactions behind the scene', *Journal of General Virology*, 93(Pt_8), pp. 1631–1644. doi: 10.1099/vir.0.042218-0.

Herzik, M. A., Wu, M. and Lander, G. C. (2017) 'Achieving better-than-3-Å resolution by single-particle cryo-EM at 200 keV', *Nature Methods*. Nature Publishing Group, 14(11), pp. 1075–1078. doi: 10.1038/nmeth.4461.

Hoenen, T., Groseth, A., Falzarano, D., Feldmann, Heinz, *et al.* (2006) 'Ebola virus: unravelling pathogenesis to combat a deadly disease.', *Trends in molecular medicine*. Elsevier, 12(5), pp. 206–15. doi: 10.1016/j.molmed.2006.03.006.

Hoenen, T., Groseth, A. and Feldmann, H. (2019) 'Therapeutic strategies to target the Ebola virus life cycle', *Nature Reviews Microbiology*. Nature Publishing Group, 17(10), pp. 593–606. doi: 10.1038/s41579-019-0233-2.

Hoffmann, B. *et al.* (2012) 'Novel Orthobunyavirus in Cattle, Europe, 2011', *Emerging

Infectious Diseases, 18(3), pp. 469–472. doi: 10.3201/eid1803.111905.

Holm, T. *et al.* (2018) 'Biochemical and structural studies reveal differences and commonalities among cap-snatching endonucleases from segmented negative-strand RNA viruses', *Journal of Biological Chemistry*, 293(51), pp. 19686–19698. doi: 10.1074/jbc.RA118.004373.

Hong, P., Koza, S. and Bouvier, E. S. P. (2012) 'Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and their Aggregates.', *Journal of liquid chromatography & related technologies*. Taylor & Francis, 35(20), pp. 2923–2950. doi: 10.1080/10826076.2012.743724.

Horne, K. M. and Vanlandingham, D. L. (2014) 'Bunyavirus-vector interactions.', *Viruses*. Multidisciplinary Digital Publishing Institute (MDPI), 6(11), pp. 4373–97. doi: 10.3390/v6114373.

Hover, S. *et al.* (2016) 'Modulation of Potassium Channels Inhibits Bunyavirus Infection.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 291(7), pp. 3411–22. doi: 10.1074/jbc.M115.692673.

Huiskonen, J. T. *et al.* (2010) 'Electron Cryotomography of Tula Hantavirus Suggests a Unique Assembly Paradigm for Enveloped Viruses', *Journal of Virology*, 84(10), pp. 4889–4897. doi: 10.1128/JVI.00057-10.

Hutchinson, K. L., Peters, C. J. and Nichol, S. T. (1996) 'Sin Nombre Virus mRNA Synthesis', *Virology*, 224(1), pp. 139–149. doi: 10.1006/viro.1996.0515.

Ilinykh, P. A. *et al.* (2014) 'Role of Protein Phosphatase 1 in Dephosphorylation of Ebola Virus VP30 Protein and Its Targeting for the Inhibition of Viral Transcription', *Journal of Biological Chemistry*, 289(33), pp. 22723–22738. doi: 10.1074/jbc.M114.575050.

Jacoby, D. R. *et al.* (1993) 'Expression of the La Crosse M Segment Proteins in a Recombinant Vaccinia Expression System Mediates pH-Dependent Cellular Fusion', *Virology*, 193(2), pp. 993–996. doi: 10.1006/viro.1993.1213.

Jeffree, C. E. *et al.* (2007) 'Ultrastructural analysis of the interaction between F-actin and respiratory syncytial virus during virus assembly', *Virology*. Academic Press, 369(2), pp. 309–323. doi: 10.1016/J.VIROL.2007.08.007.

Jin, H. and Elliott, R. M. (1993) 'Characterization of Bunyamwera virus S RNA that is transcribed and replicated by the L protein expressed from recombinant vaccinia virus.', *Journal of virology*. American Society for Microbiology Journals, 67(3), pp. 1396–404. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8437222 (Accessed: 9 July 2019).

John, S. P. *et al.* (2007) 'Ebola virus VP30 is an RNA binding protein.', *Journal of virology*. American Society for Microbiology (ASM), 81(17), pp. 8967–76. doi: 10.1128/JVI.02523-06.

Johnson, K. A. *et al.* (2016) 'The Ebola Virus matrix protein, VP40, requires phosphatidylinositol 4,5-bisphosphate (PI(4,5)P2) for extensive oligomerization at the plasma membrane and viral egress', *Scientific Reports*. Nature Publishing Group, 6, p. 19125. doi: 10.1038/srep19125.

Karron, R. A. *et al.* (2015) 'A gene deletion that up-regulates viral gene expression yields an attenuated RSV vaccine with improved antibody responses in children.', *Science translational medicine*. NIH Public Access, 7(312), p. 312ra175. doi: 10.1126/scitranslmed.aac8463.

Kelly, S. M. and Price, N. C. (2000) *The Use of Circular Dichroism in the Investigation of Protein Structure and Function*, *Current Protein and Peptide Science*. Available at: https://ctrstbio.org.uic.edu/manuals/kelly.pdf (Accessed: 4 January 2021).

Khataby, K. *et al.* (2016) 'Ebola Virus's Glycoproteins and Entry Mechanism'. doi: 10.5772/64032.

Kirchdoerfer, R. N. R. *et al.* (2016) 'The Ebola Virus VP30-NP Interaction Is a Regulator of Viral RNA Synthesis', *PLOS Pathogens*. Edited by M. J. Schnell. Public Library of Science, 12(10), p. e1005937. Available at: www.rcsb.org (Accessed: 8 December 2016).

Kiss, G. *et al.* (2014) 'Structural analysis of respiratory syncytial virus reveals the position of M2-1 between the matrix protein and the ribonucleoprotein complex.', *Journal of virology*, 88(13), pp. 7602–17. doi: 10.1128/JVI.00256-14.

Klemm, C. *et al.* (2013) 'Systems To Establish Bunyavirus Genome Replication in the Absence of Transcription', *Journal of Virology*, 87(14), pp. 8205–8212. doi: 10.1128/JVI.00371-13.

Van Knippenberg, I. and Elliott, R. M. (2015) 'Flexibility of Bunyavirus Genomes: Creation of an Orthobunyavirus with an Ambisense S Segment'. doi: 10.1128/JVI.03595-14.

Kohl, A. *et al.* (2003) 'Bunyamwera virus nonstructural protein NSs counteracts interferon regulatory factor 3-mediated induction of early cell death.', *Journal of virology*. J Virol, 77(14), pp. 7999–8008. doi: 10.1128/jvi.77.14.7999-8008.2003.

Kohl, A. *et al.* (2004) 'Complementarity, sequence and structural elements within the 3' and 5' non-coding regions of the Bunyamwera orthobunyavirus S segment determine promoter strength', *Journal of General Virology*, 85(11), pp. 3269–3278. doi: 10.1099/vir.0.80407-0.

Kruse, T. *et al.* (2018) 'The Ebola Virus Nucleoprotein Recruits the Host PP2A-B56 Phosphatase to Activate Transcriptional Support Activity of VP30', *Molecular Cell*. Cell Press, 69(1), pp. 136-145.e6. doi: 10.1016/J.MOLCEL.2017.11.034.

Krzyzaniak, M. A. *et al.* (2013) 'Host cell entry of respiratory syncytial virus involves macropinocytosis followed by proteolytic activation of the F protein.', *PLoS pathogens*. Edited by A. Pekosz, 9(4), p. e1003309. doi: 10.1371/journal.ppat.1003309.

Kuo, L., Fearns, R. and Collins, P. L. (1996) 'The structurally diverse intergenic regions of respiratory syncytial virus do not modulate sequential transcription by a dicistronic minigenome.', *Journal of virology*, 70(9), pp. 6143–50. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8709239 (Accessed: 5 June 2019).

Kurt-Jones, E. A. *et al.* (2000) 'Pattern recognition receptors TLR4 and CD14 mediate response to respiratory syncytial virus', *Nature Immunology*, 1(5), pp. 398–401. doi: 10.1038/80833.

Lambert, D. M. (1988) 'Role of oligosaccharides in the structure and function of respiratory syncytial virus glycoproteins', *Virology*. Academic Press, 164(2), pp. 458–466. doi: 10.1016/0042-6822(88)90560-0.

Lappin, D. F. *et al.* (2019) *Localization of Bunyamwera bunyavirus G1 glycoprotein to the Golgi requires association with G2 but not with NSm*, *Journal of General Virology*. Available at: www.microbiologyresearch.org (Accessed: 8 July 2019).

Lees, J. F., Pringle, C. R. and Elliott, R. M. (1986) 'Nucleotide sequence of the Bunyamwera virus M RNA segment: conservation of structural features in the Bunyavirus glycoprotein gene product.', *Virology*, 148(1), pp. 1–14. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3753629 (Accessed: 8 July 2019).

Leung, D. W. *et al.* (2009) *Structure of the Ebola VP35 interferon inhibitory domain*, *PNAS*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2626716/pdf/zpq411.pdf (Accessed: 24 September 2019).

Leung, D. W. *et al.* (2010) 'Ebolavirus VP35 is a multifunctional virulence factor.', *Virulence*. Taylor & Francis, 1(6), pp. 526–31. doi: 10.4161/viru.1.6.12984.

Leung, D. W. *et al.* (2015) 'An Intrinsically Disordered Peptide from Ebola Virus VP35 Controls Viral RNA Synthesis by Modulating Nucleoprotein-RNA Interactions', *Cell Reports*, 11(3), pp. 376–389. doi: 10.1016/j.celrep.2015.03.034.

Leyrat, C. *et al.* (2011) 'Structure of the Vesicular Stomatitis Virus N0-P Complex', *PLoS Pathogens*. Edited by F. A. Rey, 7(9), p. e1002248. doi: 10.1371/journal.ppat.1002248.

Leyrat, C. *et al.* (2014) 'Drastic changes in conformational dynamics of the antiterminator M2-1 regulate transcription efficiency in Pneumovirinae.', *eLife*. eLife Sciences Publications Limited, 3, p. e02674. doi: 10.7554/eLife.02674.

Li, B. *et al.* (2013) 'Bunyamwera virus possesses a distinct nucleocapsid protein to facilitate genome encapsidation.', *Proceedings of the National Academy of Sciences of*

the United States of America. National Academy of Sciences, 110(22), pp. 9048–53. doi: 10.1073/pnas.1222552110.

Li, D. *et al.* (2008) 'Association of respiratory syncytial virus M protein with viral nucleocapsids is mediated by the M2-1 protein.', *Journal of virology*. American Society for Microbiology Journals, 82(17), pp. 8863–70. doi: 10.1128/JVI.00343-08.

Li, J. *et al.* (2008) 'A conserved motif in region v of the large polymerase proteins of nonsegmented negative-sense RNA viruses that is essential for mRNA capping.', *Journal of virology*. American Society for Microbiology Journals, 82(2), pp. 775–84. doi: 10.1128/JVI.02107-07.

Li, L. *et al.* (2012) 'Oral versus inhaled ribavirin therapy for respiratory syncytial virus infection after lung transplantation', *The Journal of Heart and Lung Transplantation*, 31(8), pp. 839–844. doi: 10.1016/j.healun.2012.04.002.

Li, S. *et al.* (2016) 'A Molecular-Level Account of the Antigenic Hantaviral Surface.', *Cell reports*. Elsevier, 15(5), pp. 959–967. doi: 10.1016/j.celrep.2016.03.082.

Lier, C., Becker, S. and Biedenkopf, N. (2017) 'Dynamic phosphorylation of Ebola virus VP30 in NP-induced inclusion bodies', *Virology*, 512, pp. 39–47. doi: 10.1016/j.virol.2017.09.006.

Liljeroos, L. *et al.* (2013) 'Architecture of respiratory syncytial virus revealed by electron cryotomography.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 110(27), pp. 11133–8. doi: 10.1073/pnas.1309070110.

Lötfering, B. *et al.* (1999) 'The nucleoprotein of Marburg virus is target for multiple cellular kinases.', *Virology*, 255(1), pp. 50–62. doi: 10.1006/viro.1998.9577.

Lozach, P.-Y. *et al.* (2011) 'DC-SIGN as a receptor for phleboviruses.', *Cell host & microbe*. Lippincott Williams & Wilkins, Philadelphia, 10(1), pp. 75–88. doi: 10.1016/j.chom.2011.06.007.

Lozach, P.-Y., Huotari, J. and Helenius, A. (2011) 'Late-penetrating viruses', *Current Opinion in Virology*, 1(1), pp. 35–43. doi: 10.1016/j.coviro.2011.05.004.

Madara, J. J. *et al.* (2015) 'The multifunctional Ebola virus VP40 matrix protein is a promising therapeutic target.', *Future virology*. NIH Public Access, 10(5), pp. 537–546. doi: 10.2217/fvl.15.6.

Madoff, D. H. and Lenard, J. (1982) 'A membrane glycoprotein that accumulates intracellularly: Cellular processing of the large glycoprotein of LaCrosse virus', *Cell*, 28(4), pp. 821–829. doi: 10.1016/0092-8674(82)90061-7.

Maes, P. *et al.* (2019) 'Taxonomy of the order Bunyavirales: second update 2018', *Archives of Virology*. Springer, 164(3), pp. 927–941. doi: 10.1007/s00705-018-04127-

3.

Mahaffy, A. F., Haddow, A. J. and Smithburn, K. C. (1946) 'A Neurotropic Virus Isolated from Aedes Mosquitoes Caught in the Semliki Forest', *The American Journal of Tropical Medicine and Hygiene*, s1-26(2), pp. 189–208. doi: 10.4269/ajtmh.1946.s1-26.189.

Mallipeddi, S. K., Lupiani, B. and Samal, S. K. (1996) 'Mapping the domains on the phosphoprotein of bovine respiratory syncytial virus required for N-P interaction using a two-hybrid system', *Journal of General Virology*, 77(5), pp. 1019–1023. doi: 10.1099/0022-1317-77-5-1019.

Mardani, M. *et al.* (2009) 'Crimean-Congo hemorrhagic fever virus as a nosocomial pathogen in Iran', *The American journal of tropical medicine and hygiene*. Am J Trop Med Hyg, 81(4). doi: 10.4269/AJTMH.2009.09-0051.

Martínez, M. J. *et al.* (2008) 'Role of Ebola virus VP30 in transcription reinitiation.', *Journal of virology*. American Society for Microbiology, 82(24), pp. 12569–73. doi: 10.1128/JVI.01395-08.

Mason, S. W. *et al.* (2003) 'Interaction between human respiratory syncytial virus (RSV) M2-1 and P proteins is required for reconstitution of M2-1-dependent RSV minigenome activity.', *Journal of virology*. American Society for Microbiology Journals, 77(19), pp. 10670–6. doi: 10.1128/jvi.77.19.10670-10676.2003.

Mastrangelo, P. and Hegele, R. (2013) 'RSV Fusion: Time for a New Model', *Viruses*, 5(3), pp. 873–885. doi: 10.3390/v5030873.

McCoy, A. J. *et al.* (2007) 'Phaser crystallographic software.', *Journal of applied crystallography*, 40(Pt 4), pp. 658–674. doi: 10.1107/S0021889807021206.

McMullan, G. *et al.* (2009) 'Detective quantum efficiency of electron area detectors in electron microscopy', *Ultramicroscopy*, 109(9), pp. 1126–1143. doi: 10.1016/j.ultramic.2009.04.002.

McPherson, A. and Gavira, J. A. (2014) 'Introduction to protein crystallization', *Acta Crystallographica. Section F, Structural Biology Communications*. International Union of Crystallography, 70(Pt 1), p. 2. doi: 10.1107/S2053230X13033141.

Meanger, J. *et al.* (2002) 'Respiratory syncytial virus matrix protein associates with nucleocapsids in infected cells', *Journal of General Virology*, 83(4), pp. 753–757. doi: 10.1099/0022-1317-83-4-753.

Mehedi, M. *et al.* (2011) 'A New Ebola Virus Nonstructural Glycoprotein Expressed through RNA Editing', *JOURNAL OF VIROLOGY*, 85(11), pp. 5406–5414. doi: 10.1128/JVI.02190-10.

Mendoza, E. J., Racine, T. and Kobinger, G. P. (2017) 'The ongoing evolution of

antibody-based treatments for Ebola virus infection', *Immunotherapy*. Future Medicine Ltd London, UK , 9(5), pp. 435–450. doi: 10.2217/imt-2017-0010.

Merk, A. *et al.* (2016) 'Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery', *Cell*. Elsevier, 165(7), pp. 1698–1707. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27238019 (Accessed: 12 February 2020).

Messaoudi, I., Amarasinghe, G. K. and Basler, C. F. (2015) 'Filovirus pathogenesis and immune evasion: insights from Ebola virus and Marburg virus', *Nature Publishing Group*, 13. doi: 10.1038/nrmicro3524.

Meyer, G., Deplanche, M. and Schelcher, F. (2008) 'Human and bovine respiratory syncytial virus vaccine research and development', *Comparative Immunology, Microbiology and Infectious Diseases*. Pergamon, 31(2–3), pp. 191–225. doi: 10.1016/J.CIMID.2007.07.008.

Modrof, J. *et al.* (2002) 'Phosphorylation of VP30 impairs ebola virus transcription.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 277(36), pp. 33099–104. doi: 10.1074/jbc.M203775200.

Modrof, J., Becker, S. and Mühlberger, E. (2003) 'Ebola virus transcription activator VP30 is a zinc-binding protein.', *Journal of virology*. American Society for Microbiology, 77(5), pp. 3334–8. doi: 10.1128/JVI.77.5.3334-3338.2003.

Mohl, B.-P. and Barr, J. N. (2009) 'Investigating the specificity and stoichiometry of RNA binding by the nucleocapsid protein of Bunyamwera virus', *RNA*, 15(3), pp. 391–399. doi: 10.1261/rna.1367209.

Moller-Tank, S. *et al.* (2015) 'Ebola Virus Entry: A Curious and Complex Series of Events', *PLOS Pathogens*. Edited by R. E. Dutch. Public Library of Science, 11(4), p. e1004731. doi: 10.1371/journal.ppat.1004731.

Möller, P. *et al.* (2005) 'Homo-oligomerization of Marburgvirus VP35 is essential for its function in replication and transcription.', *Journal of virology*, 79(23), pp. 14876–86. doi:10.1128/JVI.79.23.14876-14886.2005.

Money, V. A. *et al.* (2009) 'Surface features of a Mononegavirales matrix protein indicate sites of membrane interaction.', *Proceedings of the National Academy of Sciences of the United States of America*, 106(11), pp. 4441–6. doi: 10.1073/pnas.0805740106.

Mühlberger, E. *et al.* (1992) 'The nucleotide sequence of the L gene of Marburg virus, a filovirus: homologies with paramyxoviruses and rhabdoviruses.', *Virology*, 187(2), pp. 534–47. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1546452 (Accessed: 12 January 2017).

Mühlberger, E. *et al.* (1996) 'Termini of All mRNA Species of Marburg Virus: Sequence and Secondary Structure', *Virology*, 223(2), pp. 376–380. doi: 10.1006/viro.1996.0490.

Mühlberger, E. *et al.* (1999) 'Comparison of the transcription and replication strategies of marburg virus and Ebola virus by using artificial replication systems.', *Journal of virology*. American Society for Microbiology (ASM), 73(3), pp. 2333–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9971816 (Accessed: 19 September 2016).

Mühlberger, E. (2007) 'Filovirus replication and transcription', *Future Virology*. Future Medicine Ltd London, UK , 2(2), pp. 205–215. doi: 10.2217/17460794.2.2.205.

Mulangu, S. *et al.* (2019) 'A Randomized, Controlled Trial of Ebola Virus Disease Therapeutics', *New England Journal of Medicine*. Massachusetts Medical Society, 381(24), pp. 2293–2303. doi: 10.1056/NEJMoa1910993.

Murphy, B R *et al.* (1986) 'Effect of age and preexisting antibody on serum antibody response of infants and children to the F and G glycoproteins during respiratory syncytial virus infection.', *Journal of clinical microbiology*. American Society for Microbiology (ASM), 24(5), pp. 894–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3771779 (Accessed: 6 June 2019).

Murphy, Brian R *et al.* (1986) *Serum and Nasal-Wash Immunoglobulin G and A Antibody Response of Infants and Children to Respiratory Syncytial Virus F and G Glycoproteins Following Primary Infection*, *JOURNAL OF CLINICAL MICROBIOLOGY*. Available at: http://jcm.asm.org/ (Accessed: 6 June 2019).

Murphy, B. R. *et al.* (1988) 'Passive transfer of respiratory syncytial virus (RSV) antiserum suppresses the immune response to the RSV fusion (F) and large (G) glycoproteins expressed by recombinant vaccinia viruses.', *Journal of virology*, 62(10), pp. 3907–10. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3047432 (Accessed: 6 June 2019).

Murshudov, G. N. *et al.* (2011) 'REFMAC5 for the refinement of macromolecular crystal structures.', *Acta crystallographica. Section D, Biological crystallography*. International Union of Crystallography, 67(Pt 4), pp. 355–67. doi: 10.1107/S0907444911001314.

Nair, H. *et al.* (2011) 'Global burden of respiratory infections due to seasonal influenza in young children: a systematic review and meta-analysis.', *Lancet (London, England)*, 378(9807), pp. 1917–30. doi: 10.1016/S0140-6736(11)61051-9.

Nakanishi, K., Berova, N. and Woody, R. (1994) *Circular dichroism : principles and applications*. VCH. Available at: https://books.google.co.uk/books?id=aNWjQgAACAAJ&redir_esc=y (Accessed: 4 January 2021).

Nielsen, K. L. *et al.* (1999) 'A single-ring mitochondrial chaperonin (Hsp60-Hsp10) can substitute for GroEL-GroES in vivo.', *Journal of bacteriology*. American Society for Microbiology Journals, 181(18), pp. 5871–5. doi: 10.1128/JB.181.18.5871-5875.1999.

Noda, T. *et al.* (2006) 'Assembly and Budding of Ebolavirus', *PLoS Pathogens*. Public

Library of Science, 2(9), p. e99. doi: 10.1371/journal.ppat.0020099.

Noda, T. *et al.* (2007) 'Mapping of the VP40-binding regions of the nucleoprotein of Ebola virus.', *Journal of virology*. American Society for Microbiology, 81(7), pp. 3554–62. doi: 10.1128/JVI.02183-06.

Noton, S. L. and Fearns, R. (2015) 'Initiation and regulation of paramyxovirus transcription and replication', *Virology*, 479–480, pp. 545–554. doi: 10.1016/j.virol.2015.01.014.

Le Nouën, C. *et al.* (2017) 'Genetic stability of genome-scale deoptimized RNA virus vaccine candidates under selective pressure.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 114(3), pp. E386–E395. doi: 10.1073/pnas.1619242114.

Novoa, R. R. *et al.* (2005) 'Key Golgi factors for structural and functional maturation of bunyamwera virus.', *Journal of virology*. American Society for Microbiology (ASM), 79(17), pp. 10852–63. doi: 10.1128/JVI.79.17.10852-10863.2005.

Ogino, M. *et al.* (2004) 'Cell Fusion Activities of Hantaan Virus Envelope Glycoproteins', *Journal of Virology*, 78(19), pp. 10776–10782. doi: 10.1128/JVI.78.19.10776-10782.2004.

El Omari, K. *et al.* (2011) 'Structures of respiratory syncytial virus nucleocapsid protein from two crystal forms: details of potential packing interactions in the native helical form.', *Acta crystallographica. Section F, Structural biology and crystallization communications*, 67(Pt 10), pp. 1179–83. doi: 10.1107/S1744309111029228.

Orlova, E. V and Saibil, H. R. (2011) 'Structural analysis of macromolecular assemblies by electron microscopy.', *Chemical reviews*. American Chemical Society, 111(12), pp. 7710–48. doi: 10.1021/cr100353t.

Osborne, J. C. and Elliott, R. M. (2000) 'RNA binding properties of bunyamwera virus nucleocapsid protein and selective binding to an element in the 5' terminus of the negative-sense S segment.', *Journal of virology*, 74(21), pp. 9946–52. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11024122 (Accessed: 28 April 2017).

Ouizougun-Oubari, M. *et al.* (2015) 'A Druggable Pocket at the Nucleocapsid/Phosphoprotein Interaction Site of Human Respiratory Syncytial Virus.', *Journal of virology*. American Society for Microbiology Journals, 89(21), pp. 11129–43. doi: 10.1128/JVI.01612-15.

Overby, A. K. *et al.* (2008) 'Insights into bunyavirus architecture from electron cryotomography of Uukuniemi virus', *Proceedings of the National Academy of Sciences*, 105(7), pp. 2375–2379. doi: 10.1073/pnas.0708738105.

Padula, P. *et al.* (1998) 'Hantavirus pulmonary syndrome outbreak in Argentina: molecular evidence for person-to-person transmission of Andes virus', *Virology*.

Virology, 241(2). doi: 10.1006/VIRO.1997.8976.

Palmer, I. and Wingfield, P. T. (2004) 'Preparation and extraction of insoluble (inclusion-body) proteins from Escherichia coli.', *Current protocols in protein science*. NIH Public Access, Chapter 6, p. Unit 6.3. doi: 10.1002/0471140864.ps0603s38.

Patterson, J. L., Holloway, B. and Kolakofsky, D. (1984) 'La Crosse virions contain a primer-stimulated RNA polymerase and a methylated cap-dependent endonuclease.', *Journal of virology*, 52(1), pp. 215–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6481853 (Accessed: 9 July 2019).

Patterson, J. L. and Kolakofsky, D. (1984) 'Characterization of La Crosse virus small-genome transcripts.', *Journal of virology*. American Society for Microbiology (ASM), 49(3), pp. 680–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6321757 (Accessed: 9 July 2019).

Pattyn, S. *et al.* (1977) 'Isolation of Marburg-like Virus from a case of Haemorrhagic Fever in Zaire', *The Lancet*, 309(8011), pp. 573–574. doi: 10.1016/S0140-6736(77)92002-5.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, F. TE (2004) 'UCSF Chimera'. Journal of Computational Chemistry, pp. 25(13):1605–12. Available at: https://www.cgl.ucsf.edu/chimera/docs/credits.html (Accessed: 29 October 2019).

Pettersson, R. F. and von Bonsdorff, C. H. (1975) 'Ribonucleoproteins of Uukuniemi virus are circular.', *Journal of virology*, 15(2), pp. 386–92. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1167604 (Accessed: 8 July 2019).

Phane Hausmann, S. *et al.* (1999) *The Versatility of Paramyxovirus RNA Polymerase Stuttering*, Journal of Viologt. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC112614/pdf/jv005568.pdf (Accessed: 17 December 2019).

Plassmeyer, M. L. *et al.* (2005) 'California serogroup Gc (G1) glycoprotein is the principal determinant of pH-dependent cell fusion and entry.', *Virology*, 338(1), pp. 121–32. doi: 10.1016/j.virol.2005.04.026.

Plassmeyer, M. L. *et al.* (2007) 'Mutagenesis of the La Crosse Virus glycoprotein supports a role for Gc (1066-1087) as the fusion peptide.', *Virology*. NIH Public Access, 358(2), pp. 273–82. doi: 10.1016/j.virol.2006.08.050.

Plowright, R. K. *et al.* (2016) 'Transmission or Within-Host Dynamics Driving Pulses of Zoonotic Viruses in Reservoir–Host Populations', *PLOS Neglected Tropical Diseases*. Edited by J. V. Remais. Public Library of Science, 10(8), p. e0004796. doi: 10.1371/journal.pntd.0004796.

Pollard, T. D. (2010) 'A Guide to Simple and Informative Binding Assays', *Molecular

Biology of the Cell. American Society for Cell Biology, 21(23), p. 4061. doi: 10.1091/MBC.E10-08-0683.

Porterfield, J. (1973) *Bunyaviruses and Bunyaviridae*, *Intervirology*. Available at: https://www.karger.com/Article/PDF/149433 (Accessed: 8 July 2019).

Potterton, L. *et al.* (2018) '*CCP* 4 *i* 2: the new graphical user interface to the *CCP* 4 program suite', *Acta Crystallographica Section D Structural Biology*. International Union of Crystallography, 74(2), pp. 68–84. doi: 10.1107/S2059798317016035.

Prins, K. C. *et al.* (2010) 'Basic Residues within the Ebolavirus VP35 Protein Are Required for Its Viral Polymerase Cofactor Function', *JOURNAL OF VIROLOGY*, 84(20), pp. 10581–10591. doi: 10.1128/JVI.00925-10.

Punch, E. K. *et al.* (2018) 'Potassium is a trigger for conformational change in the fusion spike of an enveloped RNA virus', *Journal of Biological Chemistry*, 293(26), pp. 9937–9944. doi: 10.1074/jbc.RA118.002494.

Qanungo, K. R. *et al.* (2004) *Two RNA polymerase complexes from vesicular stomatitis virus-infected cells that carry out transcription and replication of genome RNA*, *PNAS*. Available at: www.pnas.orgcgidoi10.1073pnas.0401449101 (Accessed: 12 June 2019).

Radhakrishnan, A. *et al.* (2010) 'Protein Analysis of Purified Respiratory Syncytial Virus Particles Reveals an Important Role for Heat Shock Protein 90 in Virus Particle Assembly', *Molecular & Cellular Proteomics*, 9(9), pp. 1829–1848. doi: 10.1074/mcp.M110.001651.

Rahman Oany, A. *et al.* (2011) 'Highly conserved regions in Ebola virus RNA dependent RNA polymerase may be act as a universal novel peptide vaccine target: a computational approach'. doi: 10.1186/s40203-015-0011-4.

Raymond, D. D. *et al.* (2010) 'Structure of the Rift Valley fever virus nucleocapsid protein reveals another architecture for RNA encapsidation', *Proceedings of the National Academy of Sciences*, 107(26), pp. 11769–11774. doi: 10.1073/pnas.1001760107.

Reguera, J. *et al.* (2013) 'Structural basis for encapsidation of genomic RNA by La Crosse Orthobunyavirus nucleoprotein.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 110(18), pp. 7246–51. doi: 10.1073/pnas.1302298110.

Reguera, J., Weber, F. and Cusack, S. (2010) 'Bunyaviridae RNA Polymerases (L-Protein) Have an N-Terminal, Influenza-Like Endonuclease Domain, Essential for Viral Cap-Dependent Transcription', *PLoS Pathogens*. Edited by F. A. Rey, 6(9), p. e1001101. doi: 10.1371/journal.ppat.1001101.

Reid, S. P. *et al.* (2006) 'Ebola Virus VP24 Binds Karyopherin 1 and Blocks STAT1

Nuclear Accumulation', *Journal of Virology*, 80(11), pp. 5156–5167. doi: 10.1128/JVI.02349-05.

Rincheval, V. *et al.* (2017) 'Functional organization of cytoplasmic inclusion bodies in cells infected by respiratory syncytial virus', *Nature Communications*, 8(1). doi: 10.1038/s41467-017-00655-9.

Rovid Spickler, A. (2017) *Akabane Disease*. doi: Retrieved from http://www.cfsph.iastate.edu/DiseaseInfo/ factsheets.php.

Ruigrok, R. W. . *et al.* (2000) 'Structural characterization and membrane binding properties of the matrix protein VP40 of ebola virus', *Journal of Molecular Biology*, 300(1), pp. 103–112. doi: 10.1006/jmbi.2000.3822.

Saeed, M. F. *et al.* (2001) *Phylogeny of the Simbu serogroup of the genus Bunyavirus*, *Journal of General Virology*. Available at: https://www.microbiologyresearch.org/docserver/fulltext/jgv/82/9/0822173a.pdf?expir es=1611746639&id=id&accname=guest&checksum=0EFAC969C81AA6007C6204EF A5D884A7 (Accessed: 27 January 2021).

Sakkas, H. *et al.* (2018) 'Oropouche Fever: A Review.', *Viruses*. Multidisciplinary Digital Publishing Institute (MDPI), 10(4). doi: 10.3390/v10040175.

Sanchez, A. *et al.* (1993) 'Sequence analysis of the Ebola virus genome: organization, genetic elements, and comparison with the genome of Marburg virus', *Virus Research*. Elsevier, 29(3), pp. 215–240. doi: 10.1016/0168-1702(93)90063-S.

Santos, R. I. M. *et al.* (2008) 'Oropouche virus entry into HeLa cells involves clathrin and requires endosomal acidification', *Virus Research*, 138(1–2), pp. 139–143. doi: 10.1016/j.virusres.2008.08.016.

Saravia, J. *et al.* (2015) 'Respiratory Syncytial Virus Disease Is Mediated by Age-Variable IL-33', *PLOS Pathogens*. Edited by P. G. Thomas. Public Library of Science, 11(10), p. e1005217. doi: 10.1371/journal.ppat.1005217.

Scarff, C. A. *et al.* (2018) 'Variations on Negative Stain Electron Microscopy Methods: Tools for Tackling Challenging Systems', *Journal of Visualized Experiments*, (132). doi: 10.3791/57199.

Schlereth, J. *et al.* (2016) 'RNA binding specificity of Ebola virus transcription factor VP30', *RNA Biology*, 13(9), pp. 783–798. doi: 10.1080/15476286.2016.1194160. Schnell, J. R. and Chou, J. J. (2008) 'Structure and mechanism of the M2 proton channel of influenza A virus', *Nature*, 451(7178), pp. 591–595. doi: 10.1038/nature06531.

Selvaraj, M. *et al.* (2018) 'The Structure of the Human Respiratory Syncytial Virus M2-1 Protein Bound to the Interaction Domain of the Phosphoprotein P Defines the Orientation of the Complex.', *mBio*. American Society for Microbiology, 9(6), pp.

e01554-18. doi: 10.1128/mBio.01554-18.

Shabman, R. S. *et al.* (2013) 'An Upstream Open Reading Frame Modulates Ebola Virus Polymerase Translation and Virus Replication', *PLoS Pathogens*. Edited by S. P. J. Whelan, 9(1), p. e1003147. doi: 10.1371/journal.ppat.1003147.

Shah, J. N. and Chemaly, R. F. (2011) 'Management of RSV infections in adult recipients of hematopoietic stem cell transplantation', *Blood*, 117(10), pp. 2755–2763. doi: 10.1182/blood-2010-08-263400.

Shchetinin, A. M. *et al.* (2015) 'Genetic and Phylogenetic Characterization of Tataguine and Witwatersrand Viruses and Other Orthobunyaviruses of the Anopheles A, Capim, Guamá, Koongol, Mapputta, Tete, and Turlock Serogroups.', *Viruses*. Multidisciplinary Digital Publishing Institute (MDPI), 7(11), pp. 5987–6008. doi: 10.3390/v7112918.

Shi, X. *et al.* (2006) 'Requirement of the N-terminal region of orthobunyavirus nonstructural protein NSm for virus assembly and morphogenesis.', *Journal of virology*. American Society for Microbiology Journals, 80(16), pp. 8089–99. doi: 10.1128/JVI.00579-06.

Shi, X. *et al.* (2007) 'Role of the cytoplasmic tail domains of Bunyamwera orthobunyavirus glycoproteins Gn and Gc in virus assembly and morphogenesis.', *Journal of virology*. American Society for Microbiology Journals, 81(18), pp. 10151–60. doi: 10.1128/JVI.00573-07.

Shi, X. *et al.* (2009) 'Functional analysis of the Bunyamwera orthobunyavirus Gc glycoprotein.', *The Journal of general virology*. Microbiology Society, 90(Pt 10), pp. 2483–92. doi: 10.1099/vir.0.013540-0.

Shi, X., Brauburger, K. and Elliott, R. M. (2005) 'Role of N-linked glycans on bunyamwera virus glycoproteins in intracellular trafficking, protein folding, and virus infectivity.', *Journal of virology*. American Society for Microbiology Journals, 79(21), pp. 13725–34. doi: 10.1128/JVI.79.21.13725-13734.2005.

Shingai, M. *et al.* (2008) 'Soluble G protein of respiratory syncytial virus inhibits Toll-like receptor 3/4-mediated IFN-beta induction', *International Immunology*, 20(9), pp. 1169–1180. doi: 10.1093/intimm/dxn074.

Sigurs, N. *et al.* (2005) 'Severe Respiratory Syncytial Virus Bronchiolitis in Infancy and Asthma and Allergy at Age 13', *American Journal of Respiratory and Critical Care Medicine*, 171(2), pp. 137–141. doi: 10.1164/rccm.200406-730OC.

Simabuco, F. M. *et al.* (2011) 'Structural analysis of human respiratory syncytial virus p protein: identification of intrinsically disordered domains.', *Brazilian journal of microbiology : [publication of the Brazilian Society for Microbiology]*. Brazilian Society of Microbiology, 42(1), pp. 340–5. doi: 10.1590/S1517-83822011000100043.

Sissoko, D. *et al.* (2016) 'Experimental Treatment with Favipiravir for Ebola Virus

Disease (the JIKI Trial): A Historically Controlled, Single-Arm Proof-of-Concept Trial in Guinea', *PLOS Medicine*. Edited by M. Lipsitch, 13(3), p. e1001967. doi: 10.1371/journal.pmed.1001967.

Siuzdak, G. (2004) 'An introduction to mass spectrometry ionization: An excerpt from The Expanding Role of Mass Spectrometry in Biotechnology, 2nd ed.; MCC Press: San Diego, 2005', *Journal of the Association for Laboratory Automation*. SAGE PublicationsSage CA: Los Angeles, CA, 9(2), pp. 50–63. doi: 10.1016/j.jala.2004.01.004.

Smyth, M. S. and Martin, J. H. J. (2000) 'x Ray crystallography', *Molecular Pathology*. BMJ Publishing Group, 53(1), p. 8. doi: 10.1136/MP.53.1.8.

Some, D. *et al.* (2019) 'Characterization of Proteins by Size-Exclusion Chromatography Coupled to Multi-Angle Light Scattering (SEC-MALS)', *Journal of Visualized Experiments*, (148), p. e59615. doi: 10.3791/59615.

Sourimant, J. *et al.* (2015) 'Fine Mapping and Characterization of the L-Polymerase-Binding Domain of the Respiratory Syncytial Virus Phosphoprotein', *Journal of Virology*. Edited by D. S. Lyles, 89(8), pp. 4421–4433. doi: 10.1128/JVI.03619-14.

Spiegel, M., Plegge, T. and Pöhlmann, S. (2016) 'The Role of Phlebovirus Glycoproteins in Viral Entry, Assembly and Release', *Viruses*. Multidisciplinary Digital Publishing Institute (MDPI), 8(7). doi: 10.3390/V8070202.

Stillman, E. A. and Whitt, M. A. (1997) 'Mutational analyses of the intergenic dinucleotide and the transcriptional start sequence of vesicular stomatitis virus (VSV) define sequences required for efficient termination and initiation of VSV transcripts.', *Journal of virology*, 71(3), pp. 2127–37. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=191313&tool=pmcentrez&rendertype=abstract (Accessed: 19 May 2016).

Stokes, H. L., Easton, A. J. and Marriott, A. C. (2003) 'Chimeric pneumovirus nucleocapsid (N) proteins allow identification of amino acids essential for the function of the respiratory syncytial virus N protein', *Journal of General Virology*, 84(10), pp. 2679–2683. doi: 10.1099/vir.0.19370-0.

Sugrue, R. J. *et al.* (2002) 'Caveolin-1 is incorporated into mature respiratory syncytial virus particles during virus assembly on the surface of virus-infected cells', *Journal of General Virology*, 83(3), pp. 611–621. doi: 10.1099/0022-1317-83-3-611.

Swale, C. *et al.* (2016) 'Structural characterization of recombinant IAV polymerase reveals a stable complex between viral PA-PB1 heterodimer and host RanBP5', *Scientific Reports*. Nature Publishing Group, 6(1), p. 24727. doi: 10.1038/srep24727.

Swanson, K. A. *et al.* (2011) 'Structural basis for immunization with postfusion respiratory syncytial virus fusion F glycoprotein (RSV F) to elicit high neutralizing antibody titers', *Proceedings of the National Academy of Sciences*, 108(23), pp. 9619–

9624. doi: 10.1073/pnas.1106536108.

Talmon, Y. *et al.* (1987) 'Electron microscopy of vitrified-hydrated La Crosse virus.', *Journal of virology*, 61(7), pp. 2319–21. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3586135 (Accessed: 8 July 2019).

Tang, H.-B. *et al.* (2019) 'Isolation, genetic analysis of the first Akabane virus from goat in China.', *The Journal of veterinary medical science*. Japanese Society of Veterinary Science, 81(10), pp. 1445–1449. doi: 10.1292/jvms.18-0602.

Tang, R. S. *et al.* (2001) 'Requirement of cysteines and length of the human respiratory syncytial virus M2-1 protein for protein function and virus viability.', *Journal of virology*, 75(23), pp. 11328–35. doi: 10.1128/JVI.75.23.11328-11335.2001.

Tanner, S. J. *et al.* (2014) 'Crystal structure of the essential transcription antiterminator M2-1 protein of human respiratory syncytial virus and implications of its phosphorylation.', *Proceedings of the National Academy of Sciences of the United States of America*, 111(4), pp. 1580–5. doi: 10.1073/pnas.1317262111.

Tawar, R. G. *et al.* (2009) 'Crystal Structure of a Nucleocapsid-Like Nucleoprotein-RNA Complex of Respiratory Syncytial Virus', *Science*, 326(5957), pp. 1279–1283. doi: 10.1126/science.1177634.

Tayyari, F. *et al.* (2011) 'Identification of nucleolin as a cellular receptor for human respiratory syncytial virus', *Nature Medicine*, 17(9), pp. 1132–1135. doi: 10.1038/nm.2444.

Teng, M. N. *et al.* (2000) 'Recombinant Respiratory Syncytial Virus That Does Not Express the NS1 or M2-2 Protein Is Highly Attenuated and Immunogenic in Chimpanzees', *Journal of Virology*, 74(19), pp. 9317–9321. doi: 10.1128/JVI.74.19.9317-9321.2000.

Timmins, J. *et al.* (2003) 'Ebola virus matrix protein VP40 interaction with human cellular factors Tsg101 and Nedd4.', *Journal of molecular biology*, 326(2), pp. 493–502. doi: 10.1016/s0022-2836(02)01406-7.

Tran, T.-L. *et al.* (2009) 'The respiratory syncytial virus M2-1 protein forms tetramers and interacts with RNA and P in a competitive manner.', *Journal of virology*, 83(13), pp. 6363–74. doi: 10.1128/JVI.00335-09.

Tremaglio, C. Z. *et al.* (2013) 'Respiratory syncytial virus polymerase can initiate transcription from position 3 of the leader promoter.', *Journal of virology*, 87(6), pp. 3196–207. doi: 10.1128/JVI.02862-12.

Trunschke, M. *et al.* (2013) 'The L–VP35 and L–L interaction domains reside in the amino terminus of the Ebola virus L protein and are potential targets for antivirals', *Virology*, 441(2), pp. 135–145. doi: 10.1016/j.virol.2013.03.013.

Umashankar, M. *et al.* (2008) 'Differential cholesterol binding by class II fusion proteins determines membrane fusion properties.', *Journal of virology*. American Society for Microbiology (ASM), 82(18), pp. 9245–53. doi: 10.1128/JVI.00975-08.

Varela, M. *et al.* (2013) 'Schmallenberg Virus Pathogenesis, Tropism and Interaction with the Innate Immune System of the Host', *PLoS Pathogens*. Edited by F. Weber, 9(1), p. e1003133. doi: 10.1371/journal.ppat.1003133.

Volchkova, V. A. *et al.* (2016) 'RNA Editing of the GP Gene of Ebola Virus is an Important Pathogenicity Factor'. doi: 10.1093/infdis/jiv309.

de Waal, L. *et al.* (2018) 'Transmission of Human Respiratory Syncytial Virus in the Immunocompromised Ferret Model.', *Viruses*. Multidisciplinary Digital Publishing Institute (MDPI), 10(1). doi: 10.3390/v10010018.

Walker, P. J. *et al.* (2020) 'Changes to virus taxonomy and the Statutes ratified by the International Committee on Taxonomy of Viruses (2020)', *Archives of Virology*. Springer, 165(11), pp. 2737–2748. doi: 10.1007/s00705-020-04752-x.

Walter, C. T. *et al.* (2011) 'Amino acid changes within the Bunyamwera virus nucleocapsid protein differentially affect the mRNA transcription and RNA replication activities of assembled ribonucleoprotein templates.', *The Journal of general virology*. Microbiology Society, 92(Pt 1), pp. 80–4. doi: 10.1099/vir.0.024240-0.

Wan, W. *et al.* (2017) 'Structure and assembly of the Ebola virus nucleocapsid', *Nature*. Nature Publishing Group, 551(7680), pp. 394–397. doi: 10.1038/nature24490.

Wang, X. *et al.* (2016) 'Molecular basis for the formation of ribonucleoprotein complex of Crimean-Congo hemorrhagic fever virus', *Journal of Structural Biology*, 196(3), pp. 455–465. doi: 10.1016/j.jsb.2016.09.013.

Watanabe, S., Noda, T. and Kawaoka, Y. (2006) 'Functional Mapping of the Nucleoprotein of Ebola Virus', *JOURNAL OF VIROLOGY*, 80(8), pp. 3743–3751. doi: 10.1128/JVI.80.8.3743–3751.2006.

Weik, M. *et al.* (2002) 'Ebola virus VP30-mediated transcription is regulated by RNA secondary structure formation.', *Journal of virology*. American Society for Microbiology, 76(17), pp. 8532–9. doi: 10.1128/JVI.76.17.8532-8539.2002.

Weissenhorn, W. *et al.* (1998) 'Crystal structure of the Ebola virus membrane fusion subunit, GP2, from the envelope glycoprotein ectodomain.', *Molecular cell*, 2(5), pp. 605–16. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9844633 (Accessed: 30 December 2016).

Whitfield, A. E., Ullman, D. E. and German, T. L. (2005) 'Tomato spotted wilt virus glycoprotein GC is cleaved at acidic pH', *Virus Research*, 110(1–2), pp. 183–186. doi: 10.1016/j.virusres.2005.01.007.

Whitmore, L. and Wallace, B. A. (2004) 'DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data', *Nucleic Acids Research*, 32, pp. 668–673. doi: 10.1093/nar/gkh371.

WHO (2020) *New Ebola outbreak detected in northwest Democratic Republic of the Congo; WHO surge team supporting the response.* Available at: https://www.who.int/news-room/detail/01-06-2020-new-ebola-outbreak-detected-in-northwest-democratic-republic-of-the-congo-who-surge-team-supporting-the-response (Accessed: 15 September 2020).

Winter, G. *et al.* (2018) 'DIALS: implementation and evaluation of a new integration package', *research papers Acta Cryst*, 74, pp. 85–97. doi: 10.1107/S2059798317017235.

Wu, W. *et al.* (2012) 'The interactome of the human respiratory syncytial virus NS1 protein highlights multiple effects on host cell biology.', *Journal of virology*, 86(15), pp. 7777–89. doi: 10.1128/JVI.00460-12.

Yang, J. *et al.* (2004) 'Mode of action of the TyrR protein: repression and activation of the tyrP promoter of Escherichia coli.', *Molecular microbiology*. Mol Microbiol, 52(1), pp. 243–56. doi: 10.1111/j.1365-2958.2003.03965.x.

Yang, Z. *et al.* (1998) 'Distinct cellular interactions of secreted and transmembrane Ebola virus glycoproteins.', *Science (New York, N.Y.)*, 279(5353), pp. 1034–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9461435 (Accessed: 8 January 2017).

Yunus, A. S., Khattar, S. K. and Samal, S. K. (2001) 'Mapping the domains on the phosphoprotein of bovine respiratory syncytial virus required for N–P and P–L interactions using a minigenome system', *Journal of General Virology*, 82(4), pp. 775–779. doi: 10.1099/0022-1317-82-4-775.

Zhou, H., Cheng, X. and Jin, H. (2003) 'Identification of amino acids that are critical to the processivity function of respiratory syncytial virus M2-1 protein.', *Journal of virology*, 77(9), pp. 5046–53. doi: 10.1128/jvi.77.9.5046-5053.2003.

Zhou, P. *et al.* (2016) 'Contraction of the type I IFN locus and unusual constitutive expression of IFN-α in bats.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 113(10), pp. 2696–701. doi: 10.1073/pnas.1518240113.

Zhou, Z. *et al.* (2016) 'Codon usage is an important determinant of gene expression levels largely through its effects on transcription.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 113(41), pp. E6117–E6125. doi: 10.1073/pnas.1606724113.

Zhu, F.-C. *et al.* (2017) 'Safety and immunogenicity of a recombinant adenovirus type-5 vector-based Ebola vaccine in healthy adults in Sierra Leone: a single-centre, randomised, double-blind, placebo-controlled, phase 2 trial', *The Lancet*. Elsevier,

389(10069), pp. 621–628. doi: 10.1016/S0140-6736(16)32617-4.

Zinzula, L. *et al.* (2019) 'Structures of Ebola and Reston Virus VP35 Oligomerization Domains and Comparative Biophysical Characterization in All Ebolavirus Species', *Structure/Folding and Design*, 27, pp. 39-54.e6. doi: 10.1016/j.str.2018.09.009.

Zivanov, J. *et al.* (2018) 'New tools for automated high-resolution cryo-EM structure determination in RELION-3', *eLife*, 7. doi: 10.7554/eLife.42166.